ΚΕΦΑΛΑΙΟ

11 Περισσότερα για τον στατιστικό έλεγχο υποθέσεων

Πρόλογος

Με βάση την πεποίθηση ότι τα πάντα θα μπορούσαν να συμβούν μόνο από τύχη –με άλλα λόγια, βάσει της θεωρίας της κατανομής δειγματοληψίας–, οι στατιστικοί έλεγχοι υποθέσεων μας επιτρέπουν να εξαγάγουμε συμπεράσματα που ξεπερνούν ένα περιορισμένο σύνολο πραγματικών παρατηρήσεων. Αυτό το κεφάλαιο περιγράφει γιατί η απόρριψη της μηδενικής υπόθεσης είναι ισχυρότερη από τη διατήρησή της και γιατί ένας μονόπλευρος έλεγχος είναι πιο πιθανός από έναν αμφίπλευρο για τον εντοπισμό μιας ψευδούς μηδενικής υπόθεσης.

Κάνουμε εικασίες για την επιτυχία του στατιστικού ελέγχου υποθέσεων όταν υποθέτουμε ότι η μηδενική υπόθεση είναι αληθής και έπειτα ότι είναι ψευδής. Οι δύο τύποι εσφαλμένων αποφάσεων –απόρριψη μιας αληθούς μηδενικής υπόθεσης (εσφαλμένος συναγερμός) ή διατήρηση μιας ψευδούς μηδενικής υπόθεσης (αστοχία)– μπορούν να ελεγχθούν από την επιλογή που θα κάνουμε για το επίπεδο σημαντικότητας και για το μέγεθος του δείγματος.

11.1 Γιατί γίνονται στατιστικοί έλεγχοι υποθέσεων;

Υπάρχει μια κρίσιμη σύνδεση μεταξύ των στατιστικών ελέγχων υποθέσεων και της ανάγκης των ερευνητών και των δημοσκόπων να κάνουν γενικεύσεις πέρα από τα υπάρχοντα δεδομένα. Αν οι 100 πρωτοετείς στο παράδειγμα των εξετάσεων SAT του προηγούμενου κεφαλαίου δεν ήταν δείγμα αλλά *απογραφή* όλης της τάξης των πρωτοετών, δεν θα υπήρχε η ανάγκη να γίνει γενίκευση πέρα από τα υπάρχοντα δεδομένα και θα ήταν άτοπο να διεξαχθεί στατιστικός έλεγχος υποθέσεων. Τώρα, η παρατηρούμενη διαφορά μεταξύ του καινούργιου παρατηρούμενου μέσου πληθυσμού 533 και του εθνικού μέσου όρου 500 θα αρκούσε από μόνη της ώστε να εξαχθεί το συμπέρασμα ότι ο μέσος βαθμός σε εξετάσεις μαθηματικών SAT για όλους τους ντόπιους πρωτοετείς είναι μεγαλύτερος του εθνικού μέσου όρου. Πράγματι, *οποιαδήποτε* παρατηρούμενη διαφορά υπέρ των ντόπιων πρωτοετών, ανεξάρτητα από το μέγεθός της, θα υποστήριζε αυτό το συμπέρασμα.

Αν πρέπει να κάνουμε γενικεύσεις πέρα από τους 100 πρωτοετείς σε έναν μεγαλύτερο τοπικό πληθυσμό, όπως μάλιστα είχε διατυπωθεί αρχικά η υπόθεση, η παρατηρούμενη διαφορά μεταξύ των 533 και 500 δεν μπορεί να ερμηνευτεί στην ονομαστική αξία. Το βασικό πρόβλημα είναι πως ο δειγματικός μέσος για ένα δεύτερο τυχαίο δείγμα 100 πρωτοετών πιθανώς θα ήταν διαφορετικός, μόνο από τύχη, από τον δειγματικό μέσο 533 για το πρώτο δείγμα. Αναλόγως, όταν επιχειρούμε να αποφασίσουμε αν η παρατηρούμενη διαφορά μεταξύ 533 και 500 είναι πραγματική ή παροδική, πρέπει να εξετάζεται η μεταβλητότητα μεταξύ δειγματικών μέσων.

Η σημασία του τυπικού σφάλματος

Για να εκτιμήσουμε την επίδραση της τύχης, χρησιμοποιούμε την έννοια της κατανομής δειγματοληψίας, δηλαδή την έννοια των δειγματικών μέσων για όλα τα πιθανά τυχαία αποτελέσματα. Ένα βασικό στοιχείο αυτής της θεωρίας είναι το τυπικό σφάλμα του μέσου, ένα μέτρο της μέσης ποσότητας κατά την οποία διαφέρουν οι δειγματικοί μέσοι, λόγω τύχης, από τον μέσο πληθυσμού. Διαιρώντας την παρατηρούμενη διαφορά (533−500) διά του τυπικού σφάλματος (11) για να πάρουμε μια τιμή του *z* (3), βρίσκουμε την αρχική παρατηρούμενη διαφορά σε μια κλίμακα του *z* είτε κοινών αποτελεσμάτων (που αποδίδονται εύλογα στην τύχη) είτε σπάνιων αποτελεσμάτων (που δεν αποδίδονται εύλογα στην τύχη). Αν ο λόγος της παρατηρούμενης διαφοράς προς το τυπικό σφάλμα, όταν εκφράζεται ως *z*, είναι αρκετά μικρός για να αποδοθεί εύλογα στην τύχη, διατηρούμε την H_0, αλλιώς, αν ο λόγος της παρατηρούμενης διαφοράς προς το τυπικό σφάλμα είναι πολύ μεγάλος για να αποδοθεί εύλογα στην τύχη, όπως στο παράδειγμα των εξετάσεων SAT, απορρίπτουμε την H_0.

Πριν κάνουμε γενικεύσεις πέρα από τα υπάρχοντα δεδομένα, πρέπει να μετράμε πάντα την επίδραση της τύχης, δηλαδή πρέπει να παίρνουμε μια τιμή για το τυπικό σφάλμα. Για να κατανοήσετε τον ζωτικό ρόλο του τυπικού σφάλματος στο παράδειγμα εξετάσεων SAT, αυξήστε την τιμή του από 11 σε 33 και παρατηρήστε ότι, ακόμα κι αν η παρατηρούμενη διαφορά παραμένει ίδια (533 − 500), θα διατηρούσαμε και δεν θα απορρίπταμε την H_0, επειδή τώρα το *z* θα ισούνταν με 1 (και όχι με 3) και θα ήταν μικρότερο από το κρίσιμο *z* 1,96.

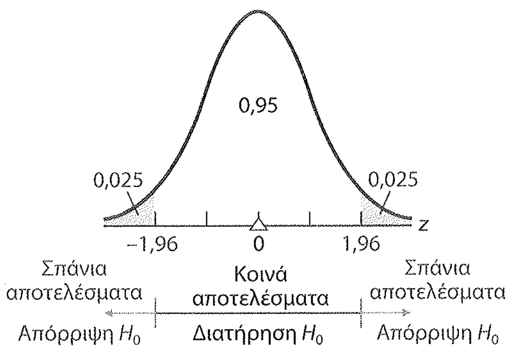
Πιθανότητα εσφαλμένων αποφάσεων

Αφού λάβουμε μια απόφαση για τη μηδενική υπόθεση, δεν γνωρίζουμε ποτέ απόλυτα αν αυτή η απόφαση είναι σωστή ή λάθος, εκτός βέβαια αν ερευνήσουμε όλον τον πληθυσμό. Ακόμα κι αν η H_0 είναι αληθής (και, επομένως, η υποθετική κατανομή του *z* γύρω από την H_0 είναι επίσης αληθής), υπάρχει μια *μικρή* πιθανότητα και από τύχη το ένα παρατηρούμενο *z* να προέρχεται πράγματι από κάποια σκιασμένη περιοχή απόρριψης της υποθετικής κατανομής του *z*, προκαλώντας έτσι την απόρριψη της αληθούς H_0. Αυτός ο τύπος εσφαλμένης απόφασης –απόρριψη μιας αληθούς H_0– αναφέρεται ως *σφάλμα τύπου I* ή *εσφαλμένος συναγερμός*.

Ως πρώτη αντίδραση θα φαινόταν λογικό να θέλουμε να καταργήσουμε τις σκιασμένες περιοχές απόρριψης από την υποθετική κατανομή δειγματοληψίας προκειμένου να εξασφαλίσουμε ότι μια αληθής H_0 δεν θα απορρίπτεται ποτέ. Ωστόσο, μια πολύ ατυχής επίπτωση αυτής της στρατηγικής είναι ότι καμία H_0, ούτε και μια θεμελιωδώς εσφαλμένη H_0, δεν θα μπορεί να απορριφθεί ποτέ. Αυτός ο δεύτερος τύπος εσφαλμένης απόφασης –διατήρηση εσφαλμένης H_0– αναφέρεται ως *σφάλμα τύπου II* ή *αστοχία*. Τόσο τα σφάλματα τύπου I όσο και τα σφάλματα τύπου II περιγράφονται πιο αναλυτικά παρακάτω σ' αυτό το κεφάλαιο.

Ελαχιστοποίηση εσφαλμένων αποφάσεων

Οι κλασικές διαδικασίες στατιστικών ελέγχων υποθέσεων, όπως αυτή που παρουσιάζουμε στο Σχήμα 11.1, τείνουν να ελαχιστοποιούν αμφότερους τους τύπους εσφαλμένων αποφάσεων. *Αν η H_0 είναι αληθής*, υπάρχει μεγά-



ΣΧΗΜΑ 11.1
Αναλογίες περιοχής που συσχετίζεται με κοινά και σπάνια αποτελέσματα (α = 0,05).

λη πιθανότητα ότι το παρατηρούμενο *z* θα μπορεί να θεωρηθεί κοινό αποτέλεσμα σύμφωνα με την υποθετική κατανομή δειγματοληψίας και ότι η αληθής H_0 θα διατηρηθεί. [Στο Σχήμα 11.1, αυτή η πιθανότητα ισούται με την αναλογία της λευκής περιοχής (0,95) στην υποθετική κατανομή δειγματοληψίας.]

Από την άλλη πλευρά, *αν η H_0 είναι σημαντικά εσφαλμένη*, επειδή ο υποθετικός μέσος πληθυσμού διαφέρει κατά πολύ από τον πραγματικό μέσο πληθυσμού, υπάρχει επίσης μεγάλη πιθανότητα ότι το παρατηρούμενο *z* θα μπορεί να θεωρηθεί σπάνιο αποτέλεσμα σύμφωνα με όσα ορίζει η υποθετική κατανομή και ότι η εσφαλμένη H_0 θα απορριφθεί. (Στο Σχήμα 11.1, αυτή η πιθανότητα δεν μπορεί να προσδιοριστεί, επειδή, σ' αυτήν την περίπτωση, η υποθετική κατανομή δειγματοληψίας δεν εκφράζει στην πραγματικότητα τη σωστή κατανομή δειγματοληψίας. Θα πούμε περισσότερα γι' αυτό παρακάτω στο κεφάλαιο.)

Ακόμα κι αν δεν γνωρίζουμε ποτέ πραγματικά αν μια συγκεκριμένη απόφαση είναι σωστή ή λάθος, είναι καθησυχαστικό να γνωρίζουμε μακροπρόθεσμα ότι *οι περισσότερες* αποφάσεις θα είναι σωστές – αν θεωρήσουμε ότι οι μηδενικές υποθέσεις είναι *είτε αληθείς είτε σημαντικά ψευδείς*.

11.2 Ισχυρές ή ασθενείς αποφάσεις

Η διατήρηση της H_0 είναι μια ασθενής απόφαση

Υπάρχουν λεπτές αλλά σημαντικές διαφορές στην ερμηνεία των αποφάσεων για τη διατήρηση της H_0 *και την απόρριψη* της H_0. Η H_0 διατηρείται όποτε το παρατηρούμενο *z* θεωρείται κοινό αποτέλεσμα βάσει της υπόθεσης ότι η H_0 είναι αληθής. Επομένως, η H_0 *θα μπορούσε* να είναι αληθής. Ωστόσο, το ίδιο παρατηρούμενο αποτέλεσμα θα μπορούσε επίσης να θεωρηθεί κοινό αποτέλεσμα όταν η αρχική τιμή στην H_0 (500) αντικατασταθεί με κάποια ελαφρώς διαφορετική άλλη. Έτσι, η διατήρηση της H_0 πρέπει να θεωρηθεί σχετικά ασθενής απόφαση. Εξαιτίας αυτής της αδυναμίας, πολλοί στατιστικολόγοι προτιμούν να περιγράφουν αυτήν την απόφαση ως απλώς μια *αποτυχία απόρριψης της H_0* παρά ως διατήρηση της H_0. Σε κάθε περίπτωση, η διατήρηση της H_0 δεν μπορεί να ερμηνευτεί ως απόδειξη ότι η H_0 είναι αληθής. Αν η H_0 είχε διατηρηθεί στο συγκεκριμένο παράδειγμα, θα μπορούσαμε να συμπεράνουμε ότι ο μέσος βαθμός σε εξετάσεις μαθηματικών SAT για όλους τους ντόπιους πρωτοετείς δεν ισούται με τον εθνικό μέσο όρο, αλλά ότι *θα μπορούσε* να ισούται με τον εθνικό μέσο όρο, όπως και με πολλές άλλες πιθανές τιμές κοντά στον εθνικό μέσο όρο.

Η απόρριψη της H_0 είναι μια ισχυρή απόφαση

Από την άλλη πλευρά, η H_0 απορρίπτεται κάθε φορά που το παρατηρούμενο *z* θεωρείται σπάνιο αποτέλεσμα –ένα αποτέλεσμα που θα μπορούσε να συμβεί μόνο από τύχη με πιθανότητα 0,05 ή ακόμα μικρότερη– με βάση την υπόθεση ότι η H_0 είναι αληθής. Αυτό το ύποπτα σπάνιο αποτέλεσμα υπονοεί ότι η H_0 είναι προφανώς ψευδής (και, αντιστρόφως, ότι η H_1 είναι προφανώς αληθής). Επομένως, η απόρριψη της H_0 μπορεί να θεωρηθεί ισχυρή απόφαση. Όταν η H_0 απορρίφθηκε στο συγκεκριμένο παράδειγμα, δεν θα ήταν σωστό να αναφέρουμε ένα οριστικό συμπέρασμα ότι ο μέσος βαθμός σε εξετάσεις μαθηματικών SAT για όλους τους ντόπιους πρωτοετείς προφανώς είναι μεγαλύτερος από τον εθνικό μέσο όρο. Συνοπτικά,

η απόφαση για τη διατήρηση της H_0 δεν υπονοεί ότι η H_0 είναι πιθανώς αληθής, αλλά μόνο ότι η H_0 *θα μπορούσε* να είναι αληθής, ενώ η απόφαση για την απόρριψη της H_0 υπονοεί ότι η H_0 είναι *προφανώς* ψευδής (και ότι η H_1 είναι *προφανώς* αληθής).

Επειδή οι περισσότεροι ερευνητές ευελπιστούν ότι θα απορρίψουν την H_0 υπέρ της H_1, η σχετική αδυναμία της απόφασης για διατήρηση της H_0 συνήθως δεν θέτει σοβαρά προβλήματα.

Γιατί η υπόθεση έρευνας δεν ελέγχεται άμεσα

Ακόμα κι αν η H_0, η μηδενική υπόθεση, αποτελεί το επίκεντρο του ενδιαφέροντος ενός στατιστικού ελέγχου, αποτελεί συνήθως δευτερεύον πρόβλημα για τον ερευνητή. Ωστόσο, υπάρχουν πολλοί λόγοι, αν και όχι πρωτεύουσας σημασίας, που κάνουν την υπόθεση έρευνας να συνταυτίζεται με την H_1 και να ελέγχεται έμμεσα.

Στερείται της απαραίτητης ακρίβειας

Η υπόθεση έρευνας, αλλά όχι η μηδενική υπόθεση, στερείται της απαραίτητης ακρίβειας για να ελεγχθεί άμεσα.

Για να υποβληθεί σε έλεγχο, μια υπόθεση πρέπει να καθορίζει έναν μόνο αριθμό γύρω από τον οποίο μπορεί να αναπτυχθεί η υποθετική κατανομή δειγματοληψίας. *Επειδή καθορίζει έναν μόνο αριθμό, η μηδενική υπόθεση, και όχι η υπόθεση έρευνας, ελέγχεται άμεσα.* Στο παράδειγμα με τις εξετάσεις SAT, η μηδενική υπόθεση καθορίζει ότι μια ακριβής τιμή (αυτή για τον εθνικό μέσο όρο 500) περιγράφει τον μέσο για τον τρέχοντα πληθυσμό ενδιαφέροντος (όλοι οι ντόπιοι πρωτοετείς). Θεωρητικά, η υπόθεση έρευνας στερείται της αναγκαίας ακρίβειας. Απλώς καθορίζει ότι υπάρχει κάποια ανισότητα μεταξύ της υποθετικής τιμής (500) και του μέσου για τον τρέχοντα πληθυσμό ενδιαφέροντος (όλοι οι ντόπιοι πρωτοετείς).

Υποστηρίζεται από μια ισχυρή απόφαση για απόρριψη

Λογικές σκέψεις επίσης συνηγορούν υπέρ του έμμεσου ελέγχου της υπόθεσης έρευνας και του άμεσου ελέγχου της μηδενικής υπόθεσης.

Επειδή η υπόθεση έρευνας συνταυτίζεται με την εναλλακτική υπόθεση, η απόφαση για απόρριψη της μηδενικής υπόθεσης, εάν πρέπει να ληφθεί, θα παρέχει *ισχυρή* υποστήριξη στην υπόθεση έρευνας, ενώ η απόφαση να διατηρηθεί η μηδενική υπόθεση, εάν πρέπει να ληφθεί, θα παρέχει το πολύ *ασθενή* υποστήριξη στη μηδενική υπόθεση.

Όπως έχουμε αναφέρει, η απόφαση για απόρριψη της μηδενικής υπόθεσης είναι ισχυρότερη από την απόφαση για τη διατήρησή της. Λογικά, μια πρόταση όπως «όλες οι αγελάδες έχουν τέσσερα πόδια» δεν μπορεί να αποδειχθεί ποτέ, παρά τη σταθερή ροή θετικών περιπτώσεων. Αρκεί μία μόνο αρνητική περίπτωση –μια αγελάδα με τρία πόδια– για να αναιρεθεί η πρόταση. Με την ίδια λογική, μία θετική περίπτωση (κοινό αποτέλεσμα) δεν αποδεικνύει τη μηδενική υπόθεση, αλλά μία αρνητική περίπτωση (σπάνιο αποτέλεσμα) αναιρεί τη μηδενική υπόθεση. (Για την ακρίβεια όμως, επειδή ένα σπάνιο αποτέλεσμα συνεπάγεται ότι η μηδενική υπόθεση είναι προφανώς *αλλά όχι αναμφίβολα* ψευδής, υπάρχει πάντοτε μια πολύ μικρή πιθανότητα ένα σπάνιο αποτέλεσμα να εκφράζει μια αληθή μηδενική υπόθεση.)

> ***Υπενθύμιση:***
> *Η απόρριψη της H_0 συνεπάγεται ότι προφανώς είναι μια ψευδής υπόθεση, ενώ η διατήρηση της H_0 συνεπάγεται απλώς ότι θα μπορούσε να είναι μια αληθής υπόθεση.*

Επομένως, θα ήταν λογικό να συνταυτίσουμε την υπόθεση έρευνας με την εναλλακτική υπόθεση. Εάν, όπως ευελπιστεί ο ερευνητής, τα δεδομένα συνηγορούν υπέρ της υπόθεσης έρευνας, ο έλεγχος θα παραγάγει ισχυρή υποστήριξη στο προαίσθημά σας: Είναι *προφανώς* αληθής. Αν τα δεδομένα δεν τάσσονται υπέρ της υπόθεσης έρευνας, ο στατιστικός έλεγχος υποθέσεων θα παραγάγει το πολύ ασθενή υποστήριξη για τη μηδενική υπόθεση: *Θα μπορούσε* να είναι αληθής. *Η ασθενής υποστήριξη για τη μηδενική υπόθεση έχει ελάσσονες συνέπειες, καθώς αυτή η υπόθεση –ότι δεν συμβαίνει κάτι ιδιαίτερο στον πληθυσμό– συνήθως αποτελεί απλώς έναν πρακτικό μηχανισμό ελέγχων.*

11.3 Μονόπλευροι και αμφίπλευροι έλεγχοι

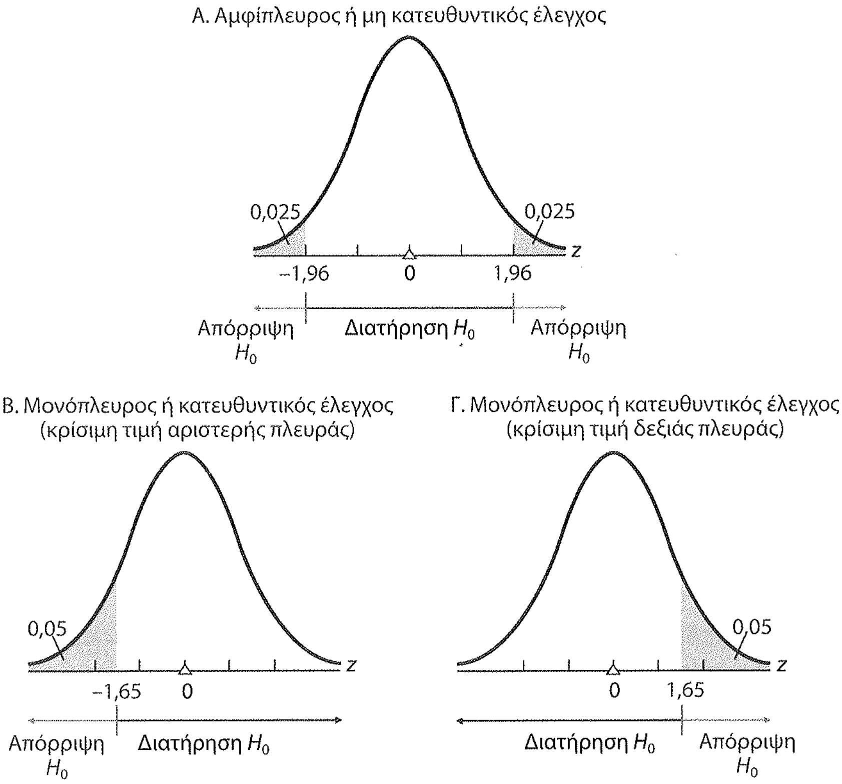
Θα μελετήσουμε τώρα μερικές τεχνικές που συμβάλλουν έτσι ώστε ο στατιστικός έλεγχος υποθέσεων να ανταποκρίνεται καλύτερα σε ειδικές περιπτώσεις.

Αμφίπλευρος έλεγχος

Γενικά, η εναλλακτική υπόθεση, H_1, είναι το συμπλήρωμα της μηδενικής υπόθεσης, H_0. Υπό τυπικές συνθήκες, η μορφή της H_1 μοιάζει με αυτήν που δείξαμε για το παράδειγμα SAT, δηλαδή

$$H_1 : \mu \neq 500$$

Αυτή η εναλλακτική υπόθεση υποστηρίζει ότι η μηδενική υπόθεση θα πρέπει να απορριφθεί αν ο μέσος βαθμός για τον πληθυσμό των ντόπιων πρωτοετών διαφέρει, προς οποιαδήποτε κατεύθυνση, από τον εθνικό μέσο όρο



ΣΧΗΜΑ 11.2
Τρεις διαφορετικοί τύποι ελέγχων ($a = 0{,}05$).

500. Ένα παρατηρούμενο z θα θεωρείται σπάνιο αποτέλεσμα αν αποκλίνει πάρα πολύ, πάνω ή κάτω, από τον εθνικό μέσο όρο. Το πλαίσιο Α του Σχήματος 11.2 παρουσιάζει περιοχές απόρριψης που σχετίζονται και με τις δύο πλευρές της υποθετικής κατανομής δειγματοληψίας. Ο αντίστοιχος κανόνας απόφασης, με το ζεύγος των κρίσιμων z-τιμών $\pm 1{,}96$, αναφέρεται ως **αμφίπλευρος** ή **μη κατευθυντικός έλεγχος**.

Αμφίπλευρος ή μη κατευθυντικός έλεγχος
Οι περιοχές απόρριψης βρίσκονται και στις δύο πλευρές της κατανομής δειγματοληψίας.

Μονόπλευρος έλεγχος (κρίσιμη τιμή αριστερής πλευράς)

Ας υποθέσουμε ότι η υπόθεση έρευνας για την περίπτωση των βαθμών στις εξετάσεις μαθηματικών SAT βασίστηκε σε παράπονα των εκπαιδευτών για την κακή προετοιμασία των ντόπιων πρωτοετών. Θα υποθέσουμε επίσης ότι, αν η έρευνα υποστηρίζει αυτά τα παράπονα, θα πρέπει να θεσμοθετηθεί ένα επανορθωτικό πρόγραμμα. Υπό αυτές τις συνθήκες, ο ερευνητής ενδεχομένως να προτιμούσε έναν στατιστικό έλεγχο υποθέσεων ο οποίος θα σχεδιαζόταν ειδικά έτσι ώστε να διαπιστώνει αν ο μέσος πληθυσμού για τους βαθμούς σε εξετάσεις μαθηματικών όλων των ντόπιων πρωτοετών είναι *μικρότερος* από τον εθνικό μέσο όρο.

Αυτή η εναλλακτική υπόθεση μας λέει ότι:

$$H_1 : \mu \leq 500$$

Βλέπουμε ότι εκφράζεται η ανησυχία πως η μηδενική υπόθεση θα πρέπει να απορρίπτεται μόνο αν ο μέσος πληθυσμού για τους βαθμούς σε εξετάσεις μαθηματικών όλων των ντόπιων πρωτοετών είναι μικρότερος από τον εθνικό μέσο όρο 500. Αναλόγως, ένα παρατηρούμενο z προκαλεί την απόφαση για απόρριψη της H_0 μόνο αν το z αποκλίνει πάρα πολύ κάτω από τον εθνικό μέσο όρο. Το πλαίσιο Β του Σχήματος 11.2 αποτυπώνει μια περιοχή απόρριψης η οποία συσχετίζεται μόνο με την αριστερή πλευρά της υποθετικής κατανομής δειγματοληψίας. Ο αντίστοιχος κανόνας απόφασης, με το κρίσιμο z να είναι $-1{,}65$, αναφέρεται ως **μονόπλευρος** ή **κατευθυντικός έλεγχος** *με την κρίσιμη τιμή αριστερής πλευράς*. Χρησιμοποιήστε τον Πίνακα Α στο Παράρτημα Γ για να επαληθεύσετε ότι, αν το κρίσιμο z ισούται με $-1{,}65$, τότε το 0,05 όλης της περιοχής κάτω από την κατανομή του z εκχωρείται στην αριστερή περιοχή απόρριψης. Παρατηρήστε ότι το επίπεδο σημαντικότητας, a, ισούται με 0,05 γι' αυτόν τον μονόπλευρο έλεγχο και επίσης για τον αρχικό αμφίπλευρο έλεγχο.

Μονόπλευρος ή κατευθυντικός έλεγχος
Η περιοχή απόρριψης βρίσκεται μόνο στη μία πλευρά της κατανομής δειγματοληψίας.

Υπερβολική ευαισθησία μονόπλευρου ελέγχου

Αυτός ο νέος μονόπλευρος έλεγχος είναι ιδιαίτερα ευαίσθητος σε οποιαδήποτε μείωση του μέσου πληθυσμού των ντόπιων πρωτοετών κάτω από τον εθνικό μέσο όρο. Αν η H_0 είναι ψευδής επειδή υπήρξε μια μείωση, τότε το παρατηρούμενο z θα είναι πιο πιθανό να αποκλίνει προς τα κάτω από τον εθνικό μέσο όρο. Όπως μπορείτε να δείτε στα πλαίσια Α και Β του Σχήματος 11.2, μια παρατηρούμενη απόκλιση προς την κατεύθυνση που μας ενδιαφέρει –κάτω από τον εθνικό μέσο όρο–είναι πιο πιθανό να εισχωρήσει στην ευρύτερη περιοχή απόρριψης για τον μονόπλευρο έλεγχο απ' ό,τι για τον αμφίπλευρο έλεγχο. Επομένως, η απόφαση για την απόρριψη μιας *εσφαλμένης* H_0 (υπέρ της υπόθεσης έρευνας) είναι πιο πιθανό να συμβεί στον μονόπλευρο παρά στον αμφίπλευρο έλεγχο.

Μονόπλευρος έλεγχος (κρίσιμη τιμή δεξιάς πλευράς)

Το πλαίσιο Γ του Σχήματος 11.2 παρουσιάζει έναν **μονόπλευρο** ή **κατευθυντικό έλεγχο** *με την κρίσιμη τιμή της δεξιάς πλευράς*. Αυτός ο μονόπλευρος έλεγχος είναι η κατοπτρική εικόνα του προηγούμενου ελέγχου. Τώρα, η εναλλακτική υπόθεση έχει ως εξής:

$$H_1 : \mu > 500$$

και το κρίσιμο z ισούται με 1,65. Αυτός ο έλεγχος σχεδιάζεται ειδικά έτσι ώστε να διαπιστώνει αν ο μέσος πληθυσμού για τους βαθμούς σε εξετάσεις μαθηματικών όλων των ντόπιων πρωτοετών είναι *μεγαλύτερος* από τον εθνικό μέσο όρο. Για παράδειγμα, η υπόθεση έρευνας για τη συγκεκριμένη υπόθεση θα μπορούσε να έχει προκύψει από κάποια σκέψη πιθανής κατάργησης ενός υπάρχοντος επανορθωτικού προγράμματος για τα μαθηματικά αν μπορεί να αποδειχθεί ότι, κατά μέσο όρο, η βαθμολογία σε εξετάσεις μαθηματικών SAT όλων των ντόπιων πρωτοετών είναι μεγαλύτερη από τον εθνικό μέσο όρο.

Μία ή δύο πλευρές;

Πριν από έναν στατιστικό έλεγχο υποθέσεων, αν υπάρχει ανησυχία σχετικά με το αν ο πραγματικός μέσος πληθυσμού διαφέρει από τον υποθετικό μέσο πληθυσμού *μόνο* για μια συγκεκριμένη κατεύθυνση, χρησιμοποιήστε τον κατάλληλο μονόπλευρο ή κατευθυντικό έλεγχο για επιπλέον ευαισθησία. Διαφορετικά, χρησιμοποιήστε τον πιο συνηθισμένο αμφίπλευρο ή μη κατευθυντικό έλεγχο.

Αφού επιλέξετε τον μονόπλευρο έλεγχο με τη δική του μία περιοχή απόρριψης, πρέπει να διατηρήσετε την H_0, ανεξάρτητα από το πόσο πολύ το παρατηρούμενο z αποκλίνει από τον υποθετικό μέσο πληθυσμού στην κατεύθυνση που δεν προκαλεί ανησυχίες. Για παράδειγμα, αν είχε χρησιμοποιηθεί μονόπλευρος έλεγχος με την κρίσιμη τιμή αριστερής πλευράς με τα δεδομένα για τους 100 πρωτοετείς του παραδείγματος SAT, η H_0 θα είχε διατηρηθεί, επειδή, ακόμα κι αν το παρατηρούμενο z ισούται με την εντυπωσιακή τιμή 3, αποκλίνει στην κατεύθυνση που προκαλεί ανησυχία – σ' αυτήν την περίπτωση, πάνω από τον εθνικό μέσο όρο. Είναι σαφές ότι θα πρέπει να χρησιμοποιείτε μονόπλευρο έλεγχο μόνο όταν δεν υπάρχει καμία απολύτως ανησυχία για αποκλίσεις, ακόμα και για πολύ μεγάλες αποκλίσεις, προς μία κατεύθυνση. Αν υπάρχει έστω και ένας μικρός προβληματισμός γι' αυτές τις αποκλίσεις, προτιμήστε έναν αμφίπλευρο έλεγχο.

Η επιλογή μονόπλευρου ή αμφίπλευρου ελέγχου θα πρέπει να γίνεται πριν από τη συλλογή των δεδομένων. Μη ρίχνετε ποτέ κρυφές ματιές στην τιμή του παρατηρούμενου z για να διαπιστώσετε αν η περιοχή απόρριψης για έναν μονόπλευρο έλεγχο βρίσκεται στη δεξιά ή στην αριστερή πλευρά της κατανομής του z. Για να θεωρηθεί μονόπλευρος έλεγχος, η θέση της περιοχής απόρριψης πρέπει να αντανακλά την ανησυχία που μπορεί να έχει ο ερευνητής μόνο για αποκλίσεις προς μια συγκεκριμένη κατεύθυνση *πριν από οποιαδήποτε επιθεώρηση των δεδομένων*. Πράγματι, ο ερευνητής θα πρέπει να είναι σε θέση να καταλήξει σε μια πειστική αιτία που θα βασίζεται στην κατανόηση της υπόθεσης έρευνας προκειμένου να υποστηρίξει την κατεύθυνση του μονόπλευρου ελέγχου.

Νέα μηδενική υπόθεση για μονόπλευρους ελέγχους

Όταν οι έλεγχοι είναι μονόπλευροι, μια ολοκληρωμένη πρόταση της μηδενικής υπόθεσης θα πρέπει επίσης να περιλαμβάνει όλες τις πιθανές τιμές του μέσου πληθυσμού προς την καθησυχαστική κατεύθυνση. Για παράδειγμα, δεδομένου ενός μονόπλευρου ελέγχου με την κρίσιμη τιμή αριστερής πλευράς, όπως H_1: $\mu < 500$, η πλήρης μηδενική υπόθεση θα πρέπει να δηλωθεί ως H_0: $\mu \geq 500$ αντί για H_0: $\mu = 500$. Με την ίδια λογική, δεδομένου ενός μονόπλευρου ελέγχου με την κρίσιμη τιμή δεξιάς πλευράς, όπως H_1: $\mu > 500$, η πλήρης μηδενική υπόθεση θα πρέπει να δηλωθεί ως H_0: $\mu \leq 500$.

Στην πραγματικότητα, η πλήρης H_0 περιγράφει όλους τους μέσους πληθυσμού που θα μπορούσαν να είναι σωστοί αν ένας μονόπλευρος έλεγχος οδηγεί στη διατήρηση της μηδενικής υπόθεσης. Για παράδειγμα, αν ένας μονόπλευρος έλεγχος με την κρίσιμη τιμή αριστερής πλευράς οδηγεί στη διατήρηση της H_0: $\mu \geq 500$, η πλήρης H_0 αντανακλά με ακρίβεια το γεγονός ότι όχι απλώς το $\mu = 500$ θα μπορούσε να είναι σωστό, αλλά και οποιαδήποτε άλλη τιμή του μέσου πληθυσμού προς την κατεύθυνση που δεν προκαλεί ανησυχίες, δηλαδή $\mu > 500$. (Μην ξεχνάτε ότι, αν ο έλεγχος είναι μονόπλευρος, ακόμα και ένα ιδιαίτερα αποκλίνον αποτέλεσμα προς την κατεύθυνση που δεν προκαλεί ανησυχίες –και πιθανώς αντανακλά έναν μέσο πολύ μεγαλύτερο από 500– θα μας κατηύθυνε προς την απόφαση να διατηρήσουμε την H_0.) Εφεξής, κάθε φορά που θα επιλέγετε να γίνει μονόπλευρος έλεγχος, γράφετε την H_0 έτσι ώστε να περιλαμβάνει τιμές του μέσου πληθυσμού προς την κατεύθυνση που δεν προκαλεί ανησυχίες – *ακόμα κι αν ο μοναδικός αριθμός στην πλήρη H_0 που προσδιορίζεται με το σύμβολο ισότητας είναι μια τιμή γύρω από την οποία επικεντρώνεται η υποθετική κατανομή δειγματοληψίας και, επομένως, η μοναδική τιμή που χρησιμοποιείται πράγματι στον στατιστικό έλεγχο υποθέσεων.*

Υπενθύμιση:
Όταν δεν υπάρχουν επαρκείς λόγοι για να γίνει μονόπλευρος έλεγχος, πραγματοποιήστε αμφίπλευρο έλεγχο.

Έλεγχος προόδου *11.1 Οι παρακάτω προτάσεις θα μπορούσαν να είναι η αφετηρία ενός στατιστικού ελέγχου υποθέσεων. Δεδομένων μόνο των πληροφοριών που δίνει κάθε πρόταση, θα χρησιμοποιούσατε αμφίπλευρο (ή μη κατευθυντικό) έλεγχο, μονόπλευρο (ή κατευθυντικό) έλεγχο με την κρίσιμη τιμή της αριστερής πλευράς ή μονόπλευρο (ή κατευθυντικό) έλεγχο με την κρίσιμη τιμή της δεξιάς πλευράς; Διευκρινίστε την απόφασή σας βρίσκοντας τα κατάλληλα H_0 και H_1. Επιπλέον, κάθε φορά που συμπεραίνετε ότι ο έλεγχος είναι μονόπλευρος, υποδείξτε τη λέξη (ή λέξεις) της πρότασης που αιτιολογεί τον μονόπλευρο έλεγχο.

(α) Ένας ερευνητής θέλει να διαπιστώσει αν για ένα δείγμα ναρκομανών το μέσο αποτέλεσμα στην κλίμακα κατάθλιψης ενός τεστ προσωπικότητας διαφέρει από το αποτέλεσμα 60, το οποίο σύμφωνα με την τεκμηρίωση του τεστ αναπαριστά το μέσο αποτέλεσμα για τον γενικό πληθυσμό.

(β) Για να αυξηθεί η βροχόπτωση, πρέπει να διεξαχθούν διεξοδικά πειράματα σποράς νεφών και τα αποτελέσματα να συγκριθούν με την ποσότητα αναφοράς 0,54 ίντσες βροχής (για συγκρίσιμες περιόδους, χωρίς σπορά νεφών).

(γ) Τα στατιστικά στοιχεία για τη δημόσια υγεία δείχνουν ότι το βάρος των αμερικανών αντρών αυξάνεται κατά μέσο όρο 23 λίβρες την 20ετή περίοδο μετά την ηλικία των 40. Ένα φιλόδοξο πρόγραμμα μείωσης βάρους, το οποίο εκτείνεται σε 20 έτη, ελέγχεται με ένα δείγμα αντρών ηλικίας 40 ετών.

(δ) Αν δεν υποβληθούν σε θεραπεία όσο ζουν, ποντίκια με καρκίνο έχουν μέση διάρκεια ζωής 134 ημέρες. Για να διαπιστωθούν οι επιδράσεις ενός φαρμάκου που θα μπορούσε ενδεχομένως να παρατείνει τη διάρκεια της ζωής (και να επιβραδύνει τον ρυθμό εξάπλωσης του καρκίνου), ορίζεται μια μέση διάρκεια ζωής για μια ομάδα ποντικιών στα οποία χορηγείται το φάρμακο.

Έλεγχος προόδου *11.2 Για τις παρακάτω περιπτώσεις, διευκρινίστε αν η H_0 πρέπει να διατηρηθεί ή να απορριφθεί.

Δεδομένου ενός μονόπλευρου ελέγχου, η κρίσιμη τιμή αριστερής πλευράς με $\alpha = 0{,}01$ και

(α) $z = -2{,}34$ (β) $z = -5{,}13$ (γ) $z = 4{,}04$

Δεδομένου ενός μονόπλευρου ελέγχου, η κρίσιμη τιμή δεξιάς πλευράς με $\alpha = 0{,}05$ και

(δ) $z = 2{,}00$ (ε) $z = -1{,}80$ (στ) $z = 1{,}61$

Απαντήσεις στη σελίδα 543.

11.4 Επιλογή επιπέδου σημαντικότητας (α)

Το επίπεδο σημαντικότητας δείχνει πόσο σπάνιο πρέπει να είναι ένα παρατηρούμενο z για να είναι δυνατή η απόρριψη της H_0. Η απόρριψη της H_0 στο επίπεδο σημαντικότητας 0,05 σημαίνει ότι θα έχει επέλθει το παρατηρούμενο z, από τύχη, με πιθανότητα μόλις 0,05 (μία πιθανότητα στις 20) *ή ακόμα λιγότερο.*

Το επίπεδο σημαντικότητας αναδεικνύει επίσης τον έμφυτο κίνδυνο που υπάρχει στον στατιστικό έλεγχο υποθέσεων, δηλαδή τον κίνδυνο απόρριψης μιας αληθούς H_0. Όταν το επίπεδο σημαντικότητας ισούται με 0,05,

Πίνακας 11.1
ΚΡΙΣΙΜΕΣ z-ΤΙΜΕΣ

Τύπος ελέγχου	**Επίπεδο σημαντικότητας (a)** 0,05	0,01
Αμφίπλευρος ή μη κατευθυντικός έλεγχος (H_0: $\mu =$ κάποιος αριθμός) (H_1: $\mu \neq$ κάποιος αριθμός)	$\pm 1{,}96$	$\pm 2{,}58$
Μονόπλευρος ή κατευθυντικός έλεγχος, κρίσιμη τιμή αριστερής πλευράς (H_0: $\mu \geq$ κάποιος αριθμός) (H_1: $\mu <$ κάποιος αριθμός)	$-1{,}65$	$-2{,}33$
Μονόπλευρος ή κατευθυντικός έλεγχος, κρίσιμη τιμή δεξιάς πλευράς (H_0: $\mu \leq$ κάποιος αριθμός) (H_1: $\mu >$ κάποιος αριθμός)	$+1{,}65$	$+2{,}33$

υπάρχει πιθανότητα 0,05 ότι, ακόμα κι αν η H_0 είναι αληθής, το παρατηρούμενο z θα ξεστρατίσει στην περιοχή απόρριψης και θα προκαλέσει την απόρριψη μιας αληθούς H_0.

Ποιο επίπεδο σημαντικότητας;

Όταν η απόρριψη μιας αληθούς H_0 είναι ιδιαίτερα σοβαρή, μπορεί να επιλεχθεί ένα μικρότερο επίπεδο σημαντικότητας. Για παράδειγμα, το επίπεδο σημαντικότητας 0,01 σημαίνει ότι, πριν απορριφθεί η H_0, το παρατηρούμενο z πρέπει να πετύχει βαθμό σπανιότητας που ισούται με 0,01 (μία πιθανότητα στις εκατό) *ή ακόμα μικρότερη* και επίσης περιορίζει, σε μια πιθανότητα 0,01, τον κίνδυνο απόρριψης μιας αληθούς H_0. Το επίπεδο 0,01 θα μπορούσε να χρησιμοποιηθεί σε έναν στατιστικό έλεγχο υποθέσεων όπου η απόρριψη μίας αληθούς H_0 θα προκαλούσε την παρουσίαση ενός ακριβού καινούργιου προγράμματος εκπαίδευσης, ακόμα κι αν ο μέσος βαθμός του πληθυσμού σε εξετάσεις μαθηματικών για όλους τους ντόπιους πρωτοετείς στην πραγματικότητα θα ήταν ίσος με τον εθνικό μέσο όρο. Ένα ακόμα μικρότερο επίπεδο σημαντικότητας, όπως το επίπεδο 0,001, θα μπορούσε να χρησιμοποιηθεί όταν η απόρριψη μιας αληθούς H_0 θα είχε τρομακτικές συνέπειες – για παράδειγμα, η θεραπεία σοβαρών ασθενειών, όπως το AIDS, αποκλειστικά με ένα καινούργιο και πολύ ακριβό φάρμακο που όχι απλώς δεν αξίζει αλλά έχει επίσης σοβαρές παρενέργειες.

Αν και είναι δυνατό να υπάρχουν πολλά διαφορετικά επίπεδα σημαντικότητας, οι περισσότεροι πίνακες για ελέγχους υποθέσεων προσανατολίζονται στα επίπεδα 0,05 και 0,01. Σ' αυτό το βιβλίο, εμείς θα ορίζουμε το επίπεδο σημαντικότητας, αλλά σε πραγματικές εφαρμογές εσείς, ως ερευνητές, ενδεχομένως να πρέπει να επιλέξετε επίπεδο σημαντικότητας. *Αν δεν υπάρχουν προφανείς λόγοι για να επιλέξετε ένα μεγαλύτερο ή ένα μικρότερο επίπεδο σημαντικότητας, χρησιμοποιείτε το σύνηθες επίπεδο του 0,05* – το μεγαλύτερο επίπεδο σημαντικότητας που αναφέρεται στα περισσότερα επαγγελματικά περιοδικά.

Όταν δοκιμάζετε υποθέσεις με τον έλεγχο z, θα σας φανεί χρήσιμος ο Πίνακας 11.1, ο οποίος εμφανίζει τις κρίσιμες z-τιμές για μονόπλευρους και αμφίπλευρους ελέγχους στα επίπεδα σημαντικότητας 0,05 και 0,01. Αυτές οι z-τιμές ελήφθησαν από τον Πίνακα Α στο Παράρτημα Γ.

Έλεγχος προόδου *11.3 Καθορίστε τον κανόνα απόφασης για τις παρακάτω περιπτώσεις (ανατρέξτε στον Πίνακα 11.1 για να βρείτε κρίσιμες z-τιμές):
(α) αμφίπλευρος έλεγχος με $a = 0{,}05$
(β) μονόπλευρος έλεγχος, κρίσιμη τιμή δεξιάς πλευράς, με $a = 0{,}01$
(γ) μονόπλευρος έλεγχος, κρίσιμη τιμή αριστερής πλευράς, με $a = 0{,}05$
(δ) αμφίπλευρος έλεγχος με $a = 0{,}01$
Απαντήσεις στη σελίδα 543.

11.5 Στατιστικός έλεγχος υπόθεσης για τη βιταμίνη C

Θα εξετάσουμε καλύτερα τα τέσσερα πιθανά αποτελέσματα ενός στατιστικού ελέγχου υποθέσεων εστιάζοντας σε μια μελέτη που επιχειρεί να διαπιστώσει αν η βιταμίνη C αυξάνει το νοητικό χάρισμα μαθητών γυμνασίου. Αφού επιλεχθούν τυχαία από κάποια μεγάλη σχολική περιφέρεια, οι 36 μαθητές λαμβάνουν ημερησίως μια δόση 90 γραμμαρίων βιταμίνης C για δύο μήνες πριν από μια εξέταση IQ.

Κανονικά, οι δείκτες IQ για όλους τους μαθητές αυτής της σχολικής περιφέρειας προσεγγίζουν μια κανονική κατανομή με μέσο 100 και τυπική απόκλιση 15. Σύμφωνα με τη μηδενική υπόθεση, ο μέσος 100 θα αποτύπωνε έτσι κι αλλιώς την κατανομή δεικτών IQ ακόμα κι αν όλοι οι μαθητές της περιφέρειας λάμβαναν δόσεις βιταμίνης C. Επιπλέον, δεδομένου του αποκλειστικού ενδιαφέροντός μας να ανιχνεύσουμε οποιαδήποτε απόκλιση του μέσου πληθυσμού *πάνω* από 100, η μηδενική υπόθεση θα πάρει τη μορφή που είναι κατάλληλη για έναν μονόπλευρο έλεγχο με την κρίσιμη τιμή δεξιάς πλευράς, δηλαδή:

$$H_0: \mu \leq 100$$

Η απόρριψη της H_0 θα υποστήριζε την H_1, την υπόθεση έρευνας ότι κάτι ιδιαίτερο συμβαίνει στον υποκείμενο πληθυσμό (επειδή η βιταμίνη C αυξάνει το νοητικό χάρισμα), δηλαδή:

$$H_0: \mu > 100$$

Ο έλεγχος z είναι η κατάλληλη επιλογή

Για να βρούμε αν ο δειγματικός μέσος για το IQ των 36 μαθητών θεωρείται κοινό ή σπάνιο αποτέλεσμα σύμφωνα με όσα ορίζει η μηδενική υπόθεση, θα διενεργήσουμε έναν έλεγχο *z*. Ο έλεγχος *z* για έναν μέσο πληθυσμού είναι η κατάλληλη επιλογή επειδή, για δείκτες IQ, η τυπική απόκλιση πληθυσμού, ως γνωστόν, είναι 15 και το σχήμα της καμπύλης πληθυσμού, ως γνωστόν, είναι κανονικό.

Θα ήταν καλύτερο να είχαμε δύο ομάδες

Αν και δεν έχει σχεδιαστεί καλά, αυτό το πείραμα προσφέρει μια προοπτική που θα είναι ιδιαίτερα χρήσιμη σε επόμενα κεφάλαια. Ένα καλύτερα σχεδιασμένο πείραμα θα σύγκρινε τους δείκτες IQ για την ομάδα αντικειμένων που λαμβάνουν βιταμίνη C με τους δείκτες IQ για μια *ομάδα ελέγχου αντικειμένων που λαμβάνουν εικονικό φάρμακο*, δηλαδή ψεύτικη βιταμίνη C – ελέγχοντας έτσι την *«επίδραση εικονικού φαρμάκου», μια αυτοπροκαλούμενη βελτίωση στην απόδοση που προκαλείται από την επίγνωση του αντικειμένου ότι υποβάλλεται σε ειδική θεραπεία.* Οι στατιστικοί έλεγχοι υποθέσεων για πειράματα με δύο ομάδες περιγράφονται στα Κεφάλαια 14 και 15.

Το πλαίσιο στη σελίδα 280 συνοψίζει αυτά τα χαρακτηριστικά του στατιστικού ελέγχου υποθέσεων τα οποία μπορούν να προσδιοριστούν πριν από τη συλλογή οποιωνδήποτε δεδομένων.

11.6 Τέσσερα πιθανά αποτελέσματα

Ο Πίνακας 11.2 συνοψίζει τα τέσσερα πιθανά αποτελέσματα οποιουδήποτε στατιστικού ελέγχου υποθέσεων. Πριν ελέγξουμε μια υπόθεση, πρέπει να διερωτηθούμε για τα τέσσερα πιθανά αποτελέσματα επειδή δεν γνωρίζουμε αν η H_0 είναι αληθής ή ψευδής – γι' αυτόν τον λόγο ελέγχουμε την υπόθεση. Αν η H_0 είναι στην πραγματι-

Πίνακας 11.2
ΠΙΘΑΝΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΕΝΟΣ ΣΤΑΤΙΣΤΙΚΟΥ ΕΛΕΓΧΟΥ ΥΠΟΘΕΣΕΩΝ

	Κατάσταση της H_0	
Απόφαση	**Αληθής H_0**	**Ψευδής H_0**
Διατήρηση H_0	(1) Σωστή απόφαση	(3) Σφάλμα τύπου II (αστοχία)
Απόρριψη H_0	(2) Σφάλμα τύπου I (εσφαλμένος συναγερμός)	(4) Σωστή απόφαση

ΠΕΡΙΛΗΨΗ ΣΤΑΤΙΣΤΙΚΟΥ ΕΛΕΓΧΟΥ ΥΠΟΘΕΣΗΣ: ΕΛΕΓΧΟΣ z ΓΙΑ ΜΕΣΟ ΠΛΗΘΥΣΜΟΥ (*ΠΡΙΝ* ΑΠΟ ΤΟ ΠΕΙΡΑΜΑ ΓΙΑ ΤΗ ΒΙΤΑΜΙΝΗ C)

Πρόβλημα έρευνας

Η ημερήσια κατανάλωση βιταμίνης C προκαλεί αύξηση, κατά μέσο όρο, στους δείκτες IQ όλων των μαθητών στη σχολική περιφέρεια;

Στατιστικές υποθέσεις

$$H_0: \mu \leq 100$$
$$H_1: \mu > 100$$

Κανόνας απόφασης

Απόρριψη της H_0 στο επίπεδο σημαντικότητας 0,05 αν $z \geq 1{,}65$.

Υπολογισμοί

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{36}} = \frac{15}{6} = 2{,}5$$

κότητα αληθής, χωρίς εμείς να το γνωρίζουμε, ένας καλά σχεδιασμένος στατιστικός έλεγχος υποθέσεων θα τείνει να επιβεβαιώσει αυτό το γεγονός, δηλαδή θα μας αναγκάσει να διατηρήσουμε την H_0 και να συμπεράνουμε ότι η H_0 θα μπορούσε να είναι αληθής. Ένα άλλο συμπέρασμα, κάτι που είναι πάντα, έστω και λίγο, πιθανό, αντανακλά ένα σφάλμα τύπου Ι. Από την άλλη πλευρά, αν η H_0 είναι στην πραγματικότητα *σημαντικά ψευδής*, χωρίς εμείς να το γνωρίζουμε, ένας καλά σχεδιασμένος έλεγχος υπόθεσης θα τείνει επίσης να επιβεβαιώσει αυτό το γεγονός, δηλαδή θα μας αναγκάσει να απορρίψουμε την H_0 και να συμπεράνουμε ότι η H_0 είναι ψευδής. Ένα άλλο συμπέρασμα, κάτι που είναι πάντα, έστω και λίγο, πιθανό, αντανακλά ένα σφάλμα τύπου ΙΙ.

Τέσσερα πιθανά αποτελέσματα του πειράματος για τη βιταμίνη C

Θα ήταν διδακτικό να περιγράψουμε τα τέσσερα πιθανά αποτελέσματα του Πίνακα 11.2 στο πλαίσιο του πειράματος για τη βιταμίνη C.

1. Αν η H_0 είναι στην πραγματικότητα αληθής (επειδή η βιταμίνη C δεν προκαλεί αύξηση στον μέσο δείκτη IQ του πληθυσμού), τότε *είναι σωστή απόφαση να διατηρήσουμε την αληθή H_0*. Σ' αυτήν την περίπτωση, θα συμπεραίναμε ορθά ότι δεν υπάρχει απόδειξη πως η βιταμίνη C αυξάνει το IQ.
2. Αν η H_0 είναι στην πραγματικότητα αληθής, τότε *είναι* **σφάλμα τύπου Ι** *να απορρίψουμε την αληθή H_0* και να συμπεράνουμε ότι η βιταμίνη C αυξάνει το IQ όταν στην πραγματικότητα δεν το κάνει. Τα σφάλματα τύπου Ι ονομάζονται μερικές φορές *εσφαλμένοι συναγερμοί* επειδή προκαλούν άσκοπες ενέργειες για κάτι που στην πραγματικότητα δεν υπάρχει. Για παράδειγμα, ένα σφάλμα τύπου Ι θα μπορούσε να ενθαρρύνει μια σειρά από πειραματικές προσπάθειες που θα επιχειρούσαν να ανακαλύψουν ακριβώς ποια δόση της βιταμίνης C μεγιστοποιεί την ανύπαρκτη «αύξηση» στο IQ.

Σφάλμα τύπου Ι
Απόρριψη αληθούς μηδενικής υπόθεσης.

3. Αν η H_0 είναι στην πραγματικότητα ψευδής (επειδή η βιταμίνη C στην πραγματικότητα προκαλεί αύξηση στο μέσο IQ του πληθυσμού), τότε *είναι* **σφάλμα τύπου ΙΙ** *να διατηρήσουμε την ψευδή H_0* και να συμπεράνουμε ότι δεν υπάρχουν αποδείξεις πως η βιταμίνη C αυξάνει το IQ όταν στην πραγματικότητα το κάνει. Τα σφάλματα τύπου ΙΙ αναφέρονται κάποιες φορές ως *αστοχίες* επειδή δεν καταφέρνουν να ανιχνεύσουν μια δυνητικά σημαντική σχέση, όπως αυτή μεταξύ βιταμίνης C και IQ.

Σφάλμα τύπου ΙΙ
Διατήρηση ψευδούς μηδενικής υπόθεσης.

4. Αν η H_0 είναι στην πραγματικότητα ψευδής, τότε *είναι σωστή απόφαση να απορρίψουμε την ψευδή H_0* και να συμπεράνουμε ότι η βιταμίνη C αυξάνει το IQ.

Η σημασία της μηδενικής υπόθεσης

Ανατρέξτε στον Πίνακα 11.2 για την παρακάτω άσκηση, όπου θα χρειαστεί να περιγράψετε τα τέσσερα πιθανά αποτελέσματα για έναν συγκεκριμένο στατιστικό έλεγχο υποθέσεων. Για να αποφύγετε οποιαδήποτε σύγχυση μεταξύ σφαλμάτων τύπου I και II, πρώτα θα πρέπει να προσδιορίσετε τη μηδενική υπόθεση, H_0. Θεωρητικά, *η μηδενική υπόθεση βεβαιώνει ότι δεν υπάρχει επίδραση, ενάντια στην υπόθεση έρευνας.* Στην προκειμένη περίπτωση, αντίθετα από την υπόθεση έρευνας, η μηδενική υπόθεση (H_0: $\mu \leq 100$) θεωρεί ότι η βιταμίνη C δεν έχει καμία θετική επίδραση στο IQ.

Οι αποφάσεις συνήθως είναι σωστές

Όταν γίνονται γενικεύσεις πέρα από υφιστάμενες παρατηρήσεις, υπάρχει πάντα η πιθανότητα για ένα σφάλμα τύπου I ή τύπου II και δεν μπορούμε να είμαστε ποτέ απολύτως βέβαιοι ότι θα έχουμε λάβει τη σωστή απόφαση. Στην καλύτερη περίπτωση, μπορούμε να χρησιμοποιήσουμε μια διαδικασία ελέγχου η οποία *συνήθως* παράγει μια σωστή απόφαση όταν η H_0 είναι αληθής ή σημαντικά ψευδής. Αυτός ο ισχυρισμός θα εξεταστεί στο πλαίσιο του πειράματος για τη βιταμίνη C, αφού υποθέσουμε πρώτα ότι η H_0 είναι στην πραγματικότητα αληθής και μετά ότι η H_0 είναι στην πραγματικότητα ψευδής. Αν και αυτή η προσέγγιση ενδεχομένως να σας φαίνεται θλιβερά θεωρητική, *επειδή δεν γνωρίζουμε ποτέ αν η H_0 είναι στην πραγματικότητα αληθής ή ψευδής,* διαβάστε προσεκτικά τις επόμενες ενότητες, επειδή έχουν στενή σχέση με κάθε στατιστικό έλεγχο υποθέσεων.

Έλεγχος προόδου *11.4

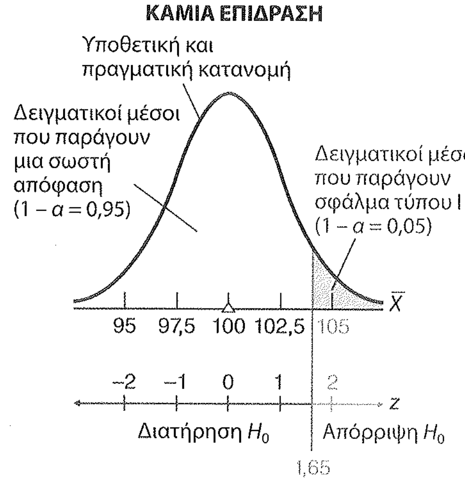
(α) Αναφέρετε τα τέσσερα πιθανά αποτελέσματα για οποιονδήποτε στατιστικό έλεγχο υποθέσεων.

(β) Σύμφωνα με τον κώδικα ποινικής δικονομίας των ΗΠΑ, ένας κατηγορούμενος θεωρείται αθώος μέχρις αποδείξεως του εναντίου. Αν θεωρήσουμε ότι μια ποινική δίκη είναι ένας στατιστικός έλεγχος υποθέσεων (με την H_0 να δηλώνει ότι ο κατηγορούμενος είναι αθώος), περιγράψτε τα τέσσερα πιθανά αποτελέσματα.

Απαντήσεις στις σελίδες 543-544.

11.7 Αν η H_0 είναι στην πραγματικότητα αληθής

Έστω ότι η H_0 είναι στην πραγματικότητα αληθής επειδή η βιταμίνη C δεν αυξάνει το μέσο IQ του πληθυσμού. Σ' αυτήν την περίπτωση, πρέπει να ενδιαφερόμαστε μόνο για το αν θα διατηρήσουμε ή θα απορρίψουμε μια αληθή H_0 (τα δύο πιο αριστερά αποτελέσματα στον Πίνακα 11.2). Θα ήταν χρήσιμο να εξετάσουμε αυτά τα δύο πιθανά αποτελέσματα σύμφωνα με την κατανομή δειγματοληψίας του Σχήματος 11.3. Γύρω από την τιμή 100, η υποθετική κατανομή δειγματοληψίας του Σχήματος 11.3 αποτυπώνει τις ιδιότητες του προβλεπόμενου μονόπλευρου ελέγχου για τη βιταμίνη C. Αν η H_0 είναι στην πραγματικότητα αληθής –και αυτό το σημείο είναι κρίσιμο–, η υποθετική κατανομή δειγματοληψίας μπορεί επίσης να θεωρηθεί ως η *σωστή* κατανομή δειγματοληψίας (από την οποία προκύπτει μάλιστα ο ένας παρατηρούμενος δειγματικός μέσος). Επομένως, ο ένας παρατηρούμενος δειγματικός μέσος (ή το z) στο πείραμα μπορεί να θεωρηθεί ως τυχαία επιλεγμένος από την υποθετική κατανομή.



ΣΧΗΜΑ 11.3

Υποθετική και πραγματική κατανομή δειγματοληψίας όταν η H_0 είναι αληθής (επειδή η βιταμίνη C δεν προκαλεί αύξηση του IQ).

Πιθανότητα σφάλματος τύπου I

Όταν από τύχη ένας τυχαία επιλεγμένος δειγματικός μέσος προέρχεται από το μικρό, σκιασμένο τμήμα της κατανομής δειγματοληψίας του Σχήματος 11.3, η z-τιμή του μέσου ισούται ή υπερβαίνει το 1,65 και γι' αυτόν τον λόγο η H_0 απορρίπτεται. Επειδή η H_0 είναι στην πραγματικότητα αληθής, αυτή η απόφαση είναι λάθος, είναι δηλαδή σφάλμα τύπου I – ένας εσφαλμένος συναγερμός υπέδειξε ως απόδειξη ότι η βιταμίνη C αυξάνει το IQ, ακόμα κι αν στην πραγματικότητα δεν το κάνει. *Η πιθανότητα ενός σφάλματος τύπου I ισούται με* **άλφα** (α), το επίπεδο σημαντικότητας. (Μην ξεχνάτε ότι το επίπεδο σημαντικότητας δείχνει την αναλογία της συνολικής περιοχής της κατανομής δειγματοληψίας στην περιοχή απόρριψης για την H_0.) Στην προκειμένη περίπτωση, η πιθανότητα ενός σφάλματος τύπου I ισούται με 0,05, όπως δείχνει το Σχήμα 11.3.

Άλφα (α)
Η πιθανότητα σφάλματος τύπου I, δηλαδή η πιθανότητα απόρριψης μιας αληθούς μηδενικής υπόθεσης.

Πιθανότητα μιας σωστής απόφασης

Όταν από τύχη ένας τυχαία επιλεγμένος δειγματικός μέσος προέρχεται από το μεγάλο λευκό τμήμα της κατανομής δειγματοληψίας του Σχήματος 11.3, η z-τιμή του μέσου είναι μικρότερη από 1,65 και η H_0 διατηρείται. Επειδή η H_0 είναι στην πραγματικότητα αληθής, αυτή είναι μια σωστή απόφαση – ανακοινώθηκε ως έλλειψη απόδειξης ότι η βιταμίνη C αυξάνει το IQ. Η πιθανότητα μιας σωστής απόφασης ισούται με $1 - \alpha$, δηλαδή 0,95.

Μείωση της πιθανότητας ενός σφάλματος τύπου I

Αν η H_0 είναι στην πραγματικότητα αληθής, ο συγκεκριμένος έλεγχος θα παραγάγει μια σωστή απόφαση με μια πιθανότητα 0,95 και ένα σφάλμα τύπου I με πιθανότητα 0,05.[25] Αν ένας εσφαλμένος συναγερμός έχει σοβαρές επιπτώσεις, η πιθανότητα ενός σφάλματος τύπου I μπορεί να μειωθεί σε 0,01 ή ακόμα και σε 0,001 χρησιμοποιώντας απλώς το επίπεδο σημαντικότητας 0,01 ή 0,001 αντίστοιχα. Ένα απ' αυτά τα επίπεδα σημαντικότητας θα μπορούσε να προτιμηθεί για τον έλεγχο της βιταμίνης C αν, για παράδειγμα, ένας εσφαλμένος συναγερμός μπορούσε να προκαλέσει την υιοθέτηση ενός ακριβού προγράμματος τροφοδότησης με άχρηστη βιταμίνη C όλων των μαθητών της περιφέρειας και ίσως τη δημιουργία ενός πιο απαιτητικού προγράμματος σπουδών που να ανταποκρίνεται στην επίπλαστη αύξηση στο νοητικό χάρισμα.

Η αληθής H_0 συνήθως διατηρείται

Υπενθύμιση:
Αν η H_0 είναι αληθής και γίνει σφάλμα, πρέπει να είναι σφάλμα τύπου I.

Αν η H_0 είναι στην πραγματικότητα αληθής, η πιθανότητα ενός σφάλματος τύπου I, α, ισούται με το επίπεδο σημαντικότητας και η πιθανότητα μιας σωστής απόφασης ισούται με $1 - \alpha$.

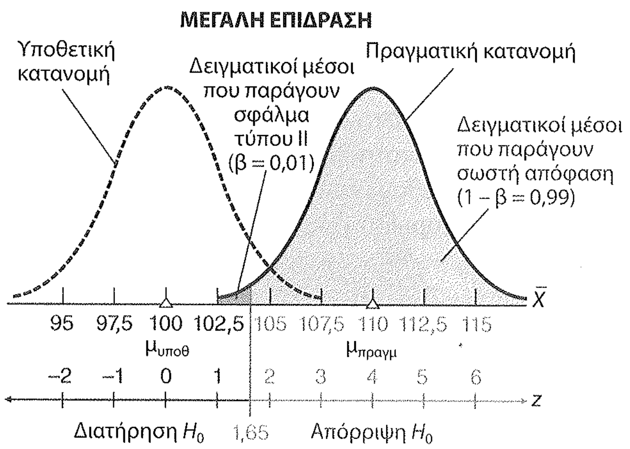
Επειδή συνήθως για το α επιλέγονται τιμές στο επίπεδο του 0,05 ή μικρότερες, μπορούμε να συμπεράνουμε ότι, αν η H_0 είναι στην πραγματικότητα αληθής, οι σωστές αποφάσεις θα είναι πιο συχνές από σφάλματα τύπου I.

Έλεγχος προόδου *11.5 Για να εξαλειφθεί το σφάλμα τύπου I, κάποιος αποφασίζει να χρησιμοποιήσει το επίπεδο σημαντικότητας 0,00. Τι λάθος έχει αυτή η διαδικασία;
Η απάντηση στη σελίδα 544.

11.8 Αν η H_0 είναι στην πραγματικότητα ψευδής εξαιτίας μιας *μεγάλης* επίδρασης

Στη συνέχεια θα θεωρήσουμε ότι η H_0 είναι στην πραγματικότητα ψευδής επειδή η βιταμίνη C αυξάνει τον μέσο πληθυσμού όχι μόνο για λίγο αλλά για *πολύ* – για παράδειγμα, κατά δέκα μονάδες. Χρησιμοποιώντας την ορολο-

25. Με απόλυτη ακρίβεια, αν η $H_0: \mu \leq 100$ είναι στην πραγματικότητα αληθής, η σωστή κατανομή δειγματοληψίας θα μπορούσε επίσης να επικεντρωθεί γύρω από κάποια τιμή μικρότερη από 100, προς την κατεύθυνση που δεν προκαλεί ανησυχίες. Σ' αυτήν την περίπτωση, οι συνέπειες του στατιστικού ελέγχου υποθέσεων θα ήταν τελικά πιο ευνοϊκές. Ουσιαστικά, επειδή η σωστή κατανομή δειγματοληψίας θα μετατοπιζόταν προς αριστερά αυτής που βλέπετε στο Σχήμα 11.3, ενώ όλα τα άλλα παραμένουν ίδια, το σφάλμα τύπου I θα είχε μικρότερη πιθανότητα από 0,05 και μια σωστή απόφαση θα είχε μεγαλύτερη πιθανότητα από 0,95.



ΣΧΗΜΑ 11.4

Υποθετική και πραγματική κατανομή δειγματοληψίας όταν η H_0 είναι ψευδής εξαιτίας μιας μεγάλης επίδρασης.

για των ερευνητών, θα μπορούσαμε να περιγράψουμε αυτήν την αύξηση ως «επίδραση δέκα μονάδων», επειδή *οποιαδήποτε διαφορά μεταξύ ενός πραγματικού και ενός υποθετικού μέσου πληθυσμού* αναφέρεται ως **επίδραση**. Αν η H_0 είναι στην πραγματικότητα ψευδής, εξαιτίας της σχετικά μεγάλης επίδρασης δέκα μονάδων της βιταμίνης C στο IQ, πρέπει να ανησυχούμε μόνο για το αν θα διατηρήσουμε ή θα απορρίψουμε μια ψευδή H_0 (τα δύο πιο δεξιά αποτελέσματα στον Πίνακα 11.2). Θα εξετάσουμε τώρα αυτά τα δύο πιθανά αποτελέσματα σε σχέση με τις κατανομές δειγματοληψίας του Σχήματος 11.4.

Επίδραση
Οποιαδήποτε διαφορά μεταξύ πραγματικού και υποθετικού μέσου πληθυσμού.

Υποθετική κατανομή δειγματοληψίας

Είναι σημαντικό να γίνεται διάκριση μεταξύ της *υποθετικής* κατανομής δειγματοληψίας και της *πραγματικής* κατανομής δειγματοληψίας (βλ. Σχήμα 11.4). Γύρω από τον υποθετικό μέσο πληθυσμού 100, η **υποθετική κατανομή δειγματοληψίας** αποτελεί τη γονική κατανομή για τον οικείο κανόνα απόφασης με κρίσιμο z 1,65 για τον προβλεπόμενο μονόπλευρο έλεγχο. Αφού προσδιοριστεί ο κανόνας απόφασης, η προσοχή μεταφέρεται από την υποθετική κατανομή δειγματοληψίας στην πραγματική κατανομή δειγματοληψίας.

Υποθετική κατανομή δειγματοληψίας
Με επίκεντρο τον υποθετικό μέσο πληθυσμού, αυτή η κατανομή χρησιμοποιείται για την παραγωγή του κανόνα απόφασης.

Πραγματική κατανομή δειγματοληψίας

Γύρω από τον πραγματικό μέσο πληθυσμού 110 (ο οποίος αποτυπώνει την επίδραση δέκα μονάδων, δηλαδή 100 + 10 = 110), η **πραγματική κατανομή δειγματοληψίας** αποτελεί τη γονική κατανομή ενός τυχαία επιλεγμένου δειγματικού μέσου (ή z) που θα παρατηρηθεί στο πείραμα. Αν τον εξετάσουμε σε σχέση με τον κανόνα απόφασης (βάσει της υποθετικής κατανομής δειγματοληψίας), ο ένας τυχαία επιλεγμένος δειγματικός μέσος (που προέρχεται από την πραγματική κατανομή δειγματοληψίας) υπαγορεύει αν θα διατηρήσουμε ή θα απορρίψουμε την ψευδή H_0.

Πραγματική κατανομή δειγματοληψίας
Με επίκεντρο τον πραγματικό μέσο πληθυσμού, αυτή η κατανομή παράγει τον έναν παρατηρούμενο μέσο (ή z).

Μικρή πιθανότητα σφάλματος τύπου II για μια μεγάλη επίδραση

Όταν από τύχη ένας τυχαία επιλεγμένος δειγματικός μέσος προέρχεται από το πολύ μικρό μαύρο τμήμα της πραγματικής κατανομής δειγματοληψίας του μέσου, η z-τιμή του μέσου είναι μικρότερη από 1,65 και, επομένως, σύμφωνα με τον κανόνα απόφασης, η H_0 διατηρείται. Επειδή η H_0 είναι στην πραγματικότητα ψευδής, αυτή η απόφαση είναι εσφαλμένη ή δείχνει ένα σφάλμα τύπου II – αστοχία, η οποία ανακοινώνεται ως έλλειψη απόδειξης ότι η βιταμίνη C αυξάνει το IQ, ακόμα κι αν στην πραγματικότητα το κάνει. Με τη βοήθεια των πινάκων για την καμπύλη κανονικής κατανομής, μπορούμε να δείξουμε ότι στην προκειμένη περίπτωση *η πιθανότητα σφάλματος τύπου II*, η οποία συμβολίζεται με το ελληνικό γράμμα **βήτα (β)**, ισούται με 0,01.

Βήτα (β)
Η πιθανότητα σφάλματος τύπου II, δηλαδή η πιθανότητα διατήρησης μιας ψευδούς μηδενικής υπόθεσης.

Το συγκεκριμένο επιχείρημα δεν προϋποθέτει ότι γνωρίζετε πώς να υπολογίζετε αυτήν την πιθανότητα 0,01 ή αυτές που θα δούμε στο υπόλοιπο κεφάλαιο. Εν συντομία, αυτές οι πιθανότητες αναπαριστούν περιοχές κάτω από την *πραγματική* κατανομή δειγματοληψίας που βρίσκονται όταν εκφράζουμε εκ νέου το κρίσιμο z ως απόκλιση από τον πραγματικό μέσο πληθυσμού [110] και όχι από τον υποθετικό μέσο πληθυσμού [100], αλλά και με μια ματιά στον Πίνακα Α στο Παράρτημα Γ, δηλαδή στον πίνακα καμπυλών κανονικής κατανομής. Όπως θα αποκαλυφθεί στην Ενότητα 11.11, όπου αυτές οι πιθανότητες –ή, για την ακρίβεια, τα συμπληρώματα $(1-\beta)$ αυτών των πιθανοτήτων– βοηθούν στην επιλογή μεγέθους δείγματος, μπορούν να υπολογιστούν πιο αποτελεσματικά μέσω ενός στατιστικού προγράμματος λογισμικού, όπως το Minitab, το οποίο ενσωματώνει τον πίνακα καμπυλών κανονικής κατανομής.

Υψηλή πιθανότητα μιας σωστής απόφασης για μια μεγάλη επίδραση

Όταν από τύχη ένας δειγματικός μέσος προέρχεται από το μεγάλο σκιασμένο τμήμα της πραγματικής κατανομής δειγματοληψίας, η z-τιμή του μέσου ισούται ή είναι μεγαλύτερη από 1,65 και η H_0 απορρίπτεται. Επειδή η H_0 είναι στην πραγματικότητα ψευδής, πρόκειται για σωστή απόφαση – ανακοινώνεται ως απόδειξη ότι η βιταμίνη C αυξάνει το IQ. Στην προκειμένη περίπτωση, η πιθανότητα μιας σωστής απόφασης, η οποία συμβολίζεται ως $1-\beta$, ισούται με 0,99.

Επισκόπηση

Αν η H_0 είναι στην πραγματικότητα ψευδής, επειδή η βιταμίνη C έχει μεγάλη επίδραση δέκα μονάδων στον μέσο δείκτη IQ του πληθυσμού, ο προβλεπόμενος μονόπλευρος έλεγχος θα έχει αρκετά καλά αποτελέσματα. Υπάρχει υψηλή πιθανότητα 0,99 ότι θα ληφθεί σωστή απόφαση και μια πολύ μικρή πιθανότητα 0,01 ότι θα γίνει σφάλμα τύπου II. Αυτό το συμπέρασμα, σε συνδυασμό με εκείνο της προηγούμενης ενότητας, αιτιολογεί τον προηγούμενο ισχυρισμό ότι οι στατιστικοί έλεγχοι υποθέσεων τείνουν να παράγουν σωστές αποφάσεις όταν η H_0 είναι στην πραγματικότητα αληθής ή όταν η H_0 είναι στην πραγματικότητα ψευδής εξαιτίας μιας μεγάλης επίδρασης.

Υπενθύμιση:
Αν η H_0 είναι ψευδής και υπάρξει σφάλμα, πρέπει να είναι σφάλμα τύπου II.

Έλεγχος προόδου *11.6 Διευκρινίστε αν οι παρακάτω προτάσεις, *οι οποίες αναφέρονται όλες στο Σχήμα 11.4*, είναι σωστές ή λάθος:

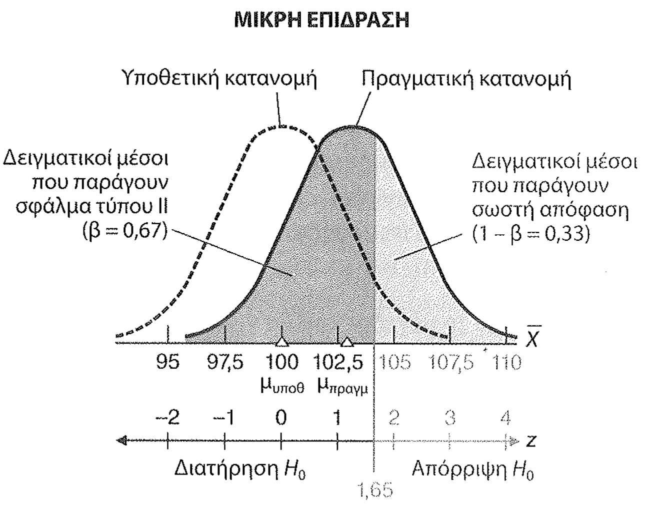
(α) Η υπόθεση ότι η H_0 είναι στην πραγματικότητα ψευδής αποτυπώνεται από τον διαχωρισμό της υποθετικής και της πραγματικής κατανομής.
(β) Στην πράξη, όταν κάνουμε στατιστικό έλεγχο υπόθεσης, δεν μπορούμε να γνωρίζουμε ότι ο πραγματικός μέσος πληθυσμού ισούται με 110.
(γ) Ο ένας παρατηρούμενος δειγματικός μέσος θεωρούμε ότι προέρχεται από την υποθετική κατανομή δειγματοληψίας.
(δ) Μια σωστή απόφαση θα λαμβανόταν αν ο ένας παρατηρούμενος δειγματικός μέσος έχει τιμή 103.
Απαντήσεις στη σελίδα 544.

11.9 Αν η H_0 είναι στην πραγματικότητα ψευδής εξαιτίας μιας *μικρής* επίδρασης

Ο προβλεπόμενος στατιστικός έλεγχος υποθέσεων δεν έχει τα ίδια καλά αποτελέσματα αν η H_0 είναι στην πραγματικότητα ψευδής επειδή η βιταμίνη C αυξάνει τον μέσο πληθυσμού IQ κατά *μόνο λίγες μονάδες* – για παράδειγμα, κατά τρεις μονάδες. Όπως δείχνει το Σχήμα 11.5, υπάρχουν και εδώ δύο διαφορετικές κατανομές δειγματικών μέσων: Η *υποθετική* κατανομή δειγματοληψίας που επικεντρώνεται γύρω από τον υποθετικό μέσο πληθυσμού 100 και η *πραγματική* κατανομή δειγματοληψίας που επικεντρώνεται γύρω από τον πραγματικό μέσο πληθυσμού 103 (ο οποίος αντανακλά την επίδραση τριών σημείων, δηλαδή $100 + 3 = 103$). Μετά τη διατύπωση του κανόνα απόφασης με τη βοήθεια της υποθετικής κατανομής δειγματοληψίας, η προσοχή μεταφέρεται στην πραγματική κατανομή δειγματοληψίας από την οποία θα προέλθει ο ένας τυχαία επιλεγμένος δειγματικός μέσος.

Χαμηλή πιθανότητα σωστής απόφασης για μια μικρή επίδραση

Σε σχέση με τον κανόνα απόφασης, η πραγματική κατανομή δειγματοληψίας παρέχει δύο τύπους τυχαία επιλεγ-



ΣΧΗΜΑ 11.5

Υποθετική και πραγματική κατανομή δειγματοληψίας όταν η H_0 είναι ψευδής εξαιτίας μιας μικρής επίδρασης.

μένων δειγματικών μέσων: Εκείνους που παράγουν σφάλμα τύπου II επειδή προέρχονται από τον μαύρο τομέα και εκείνους που παράγουν σωστή απόφαση επειδή προέρχονται από τον σκιασμένο τομέα. Εξαιτίας της μικρής επίδρασης τριών μονάδων, ο πραγματικός και ο υποθετικός μέσος πληθυσμού προσεγγίζουν περισσότερο το Σχήμα 11.5 από το Σχήμα 11.4. Κατά συνέπεια, όλη η πραγματική κατανομή δειγματοληψίας του Σχήματος 11.5 μετατοπίζεται προς την περιοχή διατήρησης για την ψευδή H_0 και αναλογικά το μεγαλύτερο μέρος αυτής της κατανομής είναι μαύρο.

Τώρα, ο προβλεπόμενος μονόπλευρος έλεγχος έχει ακόμα χειρότερα αποτελέσματα. Υπάρχει μια σχετικά υψηλή πιθανότητα 0,67 ότι θα παρατηρηθεί σφάλμα τύπου II και χαμηλή πιθανότητα 0,33 ότι θα ληφθεί η σωστή απόφαση. (Μην ξεχνάτε ότι δεν χρειάζεται να βρείτε αυτές τις πιθανότητες καμπύλης κανονικής κατανομής για να καταλάβετε το επιχείρημα.)

Η απόρριψη της ψευδούς H_0 εξαρτάται από το μέγεθος της επίδρασης

Αν η H_0 είναι στην πραγματικότητα ψευδής, η πιθανότητα ενός σφάλματος τύπου II, β, και η πιθανότητα μιας σωστής απόφασης, $1 - \beta$, εξαρτώνται από το μέγεθος της επίδρασης, δηλαδή από τη διαφορά μεταξύ του πραγματικού και του υποθετικού μέσου πληθυσμού. Όσο μικρότερη είναι η επίδραση, τόσο υψηλότερη είναι η πιθανότητα να υπάρχει σφάλμα τύπου II και τόσο χαμηλότερη είναι η πιθανότητα να ληφθεί σωστή απόφαση.

Αυτό το συμπέρασμα δεν πρέπει πραγματικά να σας εκπλήσσει. Αν η H_0 είναι στην πραγματικότητα ψευδής, θα πρέπει να υπάρχει κάποια επίδραση. Όσο μικρότερη είναι αυτή, τόσο λιγότερο πιθανό είναι να ανιχνευθεί (μέσω της ορθής απόρριψης της ψευδούς H_0) και τόσο πιο πιθανό είναι να παραλειφθεί (μέσω της εσφαλμένης διατήρησης της ψευδούς H_0). Όπως θα δούμε στην επόμενη ενότητα, αν είναι σημαντικό να επιτευχθεί ανίχνευση ακόμα και μιας σχετικά μικρής επίδρασης, η πιθανότητα μιας σωστής απόφασης μπορεί να αυξηθεί σε οποιαδήποτε επιθυμητή τιμή μέσω αύξησης του μεγέθους του δείγματος.

Έλεγχος προόδου *11.7 Διευκρινίστε αν οι παρακάτω προτάσεις, *οι οποίες όλες αναφέρονται στο Σχήμα 11.5*, είναι σωστές ή λάθος:

(α) Η τιμή του πραγματικού μέσου πληθυσμού (103) υπαγορεύει τη θέση της πραγματικής κατανομής δειγματοληψίας.
(β) Η κρίσιμη z-τιμή (1,65) βασίζεται στην πραγματική κατανομή δειγματοληψίας.
(γ) Επειδή ο υποθετικός μέσος πληθυσμού 100 είναι στην πραγματικότητα ψευδής, θα ήταν αδύνατο να παρατηρηθεί τιμή δειγματικού μέσου μικρότερη ή ίση με 100.
(δ) Θα μπορεί να ληφθεί σωστή απόφαση αν ο ένας παρατηρούμενος δειγματικός μέσος έχει τιμή 105.

Απαντήσεις στη σελίδα 544.

11.10 Επίδραση μεγέθους δείγματος

Σε κανονικές συνθήκες, ο ερευνητής ίσως να μην ενδιαφέρεται ιδιαίτερα για το χαμηλό ποσοστό ανίχνευσης, 0,33, για τη σχετικά μικρή επίδραση τριών μονάδων της βιταμίνης C στο IQ. Σε ειδικές συνθήκες όμως, αυτό το χαμηλό ποσοστό ανίχνευσης ενδεχομένως να μην είναι αποδεκτό. Για παράδειγμα, προηγούμενα πειράματα μπορεί να έχουν αποδείξει ότι η βιταμίνη C έχει πολλές θετικές επιδράσεις, συμπεριλαμβανομένης της μείωσης στη διάρκεια και στη δριμύτητα κοινών κρυολογημάτων, καθώς και καμία γνωστή αρνητική παρενέργεια.[26] Επιπλέον, τεράστιες ποσότητες βιταμίνης C θα μπορούσαν να διατεθούν στα σχολεία της περιφέρειας χωρίς κανένα κόστος. Η καθιέρωση ακόμα μίας θετικής επίδρασης, ακόμα κι αν είναι σχετικά μικρή, όπως είναι μια μικρή αύξηση στο μέσο IQ του πληθυσμού, θα μπορούσε να παγιώσει την παροχή βιταμίνης C σε όλους τους μαθητές της περιφέρειας. Επομένως, ο ερευνητής μπορεί να θέλει να χρησιμοποιήσει μια ελεγκτική διαδικασία για την οποία, αν η H_0 είναι στην πραγματικότητα ψευδής εξαιτίας μιας μικρής επίδρασης, το ποσοστό ανίχνευσης είναι αισθητά υψηλότερο από 0,33.

Για να αυξήσετε την πιθανότητα ανίχνευσης μιας ψευδούς H_0, αυξήστε το μέγεθος του δείγματος.

Αν θεωρήσουμε ότι η βιταμίνη C εξακολουθεί να έχει μόνο μια μικρή επίδραση τριών μονάδων στο IQ, μπορούμε να ελέγξουμε τις ιδιότητες του προβλεπόμενου μονόπλευρου ελέγχου όταν το μέγεθος δείγματος αυξάνεται από 36 σε 100 μαθητές. Σας θυμίζουμε τον τύπο για το τυπικό σφάλμα του μέσου, $\sigma_{\overline{X}}$,

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

Για το αρχικό πείραμα με μέγεθος δείγματος 36,

$$\sigma_{\overline{X}} = \frac{15}{\sqrt{36}} = \frac{15}{6} = 2{,}5$$

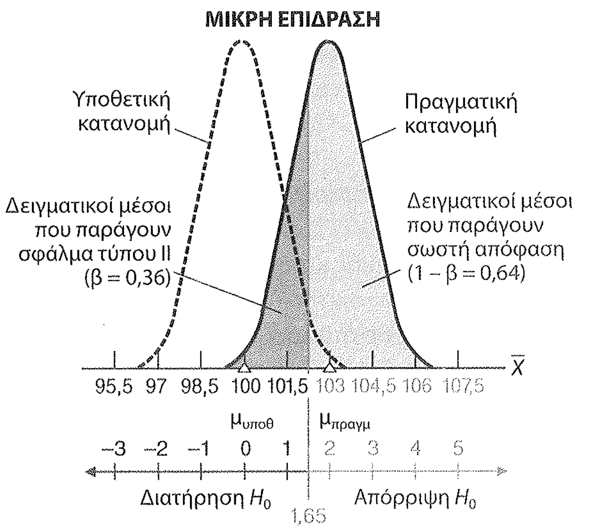
ενώ για το νέο πείραμα με μέγεθος δείγματος 100,

$$\sigma_{\overline{X}} = \frac{15}{\sqrt{100}} = \frac{15}{10} = 1{,}5$$

Είναι σαφές ότι οποιαδήποτε αύξηση στο μέγεθος δείγματος προκαλεί μείωση στο τυπικό σφάλμα του μέσου.

Οι συνέπειες της μείωσης του τυπικού σφάλματος

Όπως μπορείτε να καταλάβετε από τη σύγκριση του Σχήματος 11.5 και του Σχήματος 11.6, η μείωση του τυπικού σφάλματος από 2,5 σε 1,5 έχει δύο σημαντικές συνέπειες:



ΣΧΗΜΑ 11.6

Υποθετική και πραγματική κατανομή δειγματοληψίας όταν η H_0 είναι ψευδής εξαιτίας μιας μικρής επίδρασης αλλά το μέγεθος του δείγματος είναι σχετικά μεγάλο.

26. Δεν υπάρχουν καλά τεκμηριωμένες αποδείξεις ότι η βιταμίνη C έχει πράγματι κάποια επίδραση στο IQ. Σύμφωνα με μια πρόσφατη διεξοδική μελέτη όμως, μπορεί να υπάρχει κάποια επίδραση της βιταμίνης C στη μείωση της διάρκειας και της δριμύτητας κοινών κρυολογημάτων [Hemila, H., Chalker, E., Douglas, B. (2007). Vitamin C for preventing and treating the common cold. *Cochrane Database of Systematic Reviews*. DOI: 10.1002/14651858.CD000980.pub3].

1. Μειώνει το εμβαδόν της επάνω περιοχής διατήρησης προς τον υποθετικό μέσο πληθυσμού 100.
2. Μειώνει την πραγματική κατανομή δειγματοληψίας προς τον πραγματικό μέσο πληθυσμού 103.

Το καθαρό αποτέλεσμα είναι ότι, μεταξύ τυχαία επιλεγμένων δειγματικών μέσων για 100 μαθητές, λιγότεροι δειγματικοί μέσοι (0,36) παράγουν σφάλμα τύπου ΙΙ επειδή προέρχονται από τον μαύρο τομέα και περισσότεροι δειγματικοί μέσοι (0,64) παράγουν σωστή απόφαση –δηλαδή οδηγούν στην ανίχνευση μιας ψευδούς H_0– επειδή προέρχονται από τον σκιασμένο τομέα.

Μια εμφανής επιπλοκή είναι ότι το τυπικό σφάλμα μπορεί να μειωθεί σε οποιαδήποτε επιθυμητή τιμή μέσω απλής αύξησης του μεγέθους δείγματος. Θα είχαμε μια ακραία περίπτωση όταν το μέγεθος του δείγματος ισούται με 10.000 μαθητές (!) – τότε το τυπικό σφάλμα μειώνεται στο 0,15. Σ' αυτήν την περίπτωση, η επάνω περιοχή διατήρησης συρρικνώνεται και προσεγγίζει τον υποθετικό μέσο πληθυσμού 100 και όλη η πραγματική κατανομή δειγματοληψίας του μέσου συρρικνώνεται για να φτάσει στον πραγματικό μέσο πληθυσμού 103. Το καθαρό αποτέλεσμα είναι ότι σπάνια γίνεται σφάλμα τύπου ΙΙ και η μικρή επίδραση τριών μονάδων ανιχνεύεται σχεδόν πάντα.

Τα δείγματα μπορεί να είναι πολύ μεγάλα

Σ' αυτό το σημείο, ενδεχομένως να σκέφτεστε ότι το μέγεθος δείγματος θα πρέπει να είναι πάντα όσο το δυνατόν μεγαλύτερο, ώστε να μεγιστοποιείται η πιθανότητα ανίχνευσης μιας ψευδούς H_0, αλλά αυτό δεν ισχύει. Ένα υπερβολικά μεγάλο μέγεθος δείγματος παράγει έναν υπερβολικά ευαίσθητο στατιστικό έλεγχο υποθέσεων που ανιχνεύει ακόμα και μια πολύ μικρή επίδραση η οποία, από κάθε άποψη, στερείται σημαντικότητας. Για παράδειγμα, ένα υπερβολικά μεγάλο μέγεθος δείγματος θα μπορούσε να προκαλέσει την απόρριψη της H_0, ακόμα κι αν η βιταμίνη C πράγματι αυξάνει τον μέσο IQ του πληθυσμού κατά μόνο ½ μονάδα. Επειδή από όλες σχεδόν τις απόψεις αυτή η πολύ μικρή επίδραση στερείται σημαντικότητας, οι περισσότεροι ερευνητές θα την παρέλειπαν, δηλαδή οι περισσότεροι πιθανώς θα διατηρούσαν αυτήν την ψευδή H_0. Γι' αυτούς τους λόγους, πριν από ένα πείραμα, ένας συνετός ερευνητής επιχειρεί να επιλέξει ένα μέγεθος δείγματος το οποίο, επειδή δεν είναι υπερβολικά μεγάλο, ελαχιστοποιεί την ανίχνευση μιας μικρής, ασήμαντης επίδρασης.

Τα δείγματα μπορεί να είναι πολύ μικρά

Από την άλλη πλευρά, το μέγεθος δείγματος μπορεί να είναι πολύ μικρό. Ένα αδικαιολόγητα μικρό μέγεθος δείγματος θα παράγει έναν μη ευαίσθητο στατιστικό έλεγχο υποθέσεων (με μεγάλο τυπικό σφάλμα) που θα παραλείπει ακόμα και πολύ μεγάλες και σημαντικές επιδράσεις. Για παράδειγμα, ένα αδικαιολόγητα μικρό μέγεθος δείγματος μπορεί να προκαλέσει τη διατήρηση της H_0, ακόμα κι αν η βιταμίνη C πράγματι αυξάνει τον μέσο IQ του πληθυσμού κατά 15 μονάδες. Πριν από ένα πείραμα, ένας συνετός ερευνητής επιχειρεί να επιλέξει ένα μέγεθος δείγματος το οποίο, επειδή δεν είναι αδικαιολόγητα μικρό, μεγιστοποιεί την ανίχνευση μιας μεγάλης, σημαντικής επίδρασης.

Ούτε πολύ μικρό ούτε πολύ μεγάλο

Για όσα θέλουν να πετύχουν οι περισσότεροι ερευνητές, ένα μέγεθος δείγματος εκατοντάδων αντικειμένων είναι υπερβολικά μεγάλο, ενώ ένα με λιγότερα από πέντε αντικείμενα είναι αδικαιολόγητα μικρό. Απομένει βέβαια πολύ μεγάλο εύρος για την επιλογή του μεγέθους δείγματος μεταξύ αυτών των γενικών ακραίων τιμών. Η στατιστική παρέχει στους ερευνητές διαγράμματα, τις λεγόμενες *καμπύλες ισχύος*, ως βοήθεια για την επιλογή του κατάλληλου μεγέθους δείγματος για ένα συγκεκριμένο πείραμα.

Έλεγχος προόδου *11.8 Κρίνετε τις παρακάτω εκθέσεις πειραμάτων:

(α) Χρησιμοποιώντας μια ομάδα 4 αντικειμένων, ένας ερευνητής ανακοινώνει ότι η H_0 διατηρήθηκε στο επίπεδο σημαντικότητας 0,05.

(β) Χρησιμοποιώντας μια ομάδα 600 αντικειμένων, ένας ερευνητής ανακοινώνει ότι η H_0 απορρίφθηκε στο επίπεδο σημαντικότητας 0,05

Απαντήσεις στη σελίδα 544.

11.11 Ισχύς και μέγεθος δείγματος

Ισχύς (1 – β)
Η πιθανότητα ανίχνευσης μιας συγκεκριμένης επίδρασης.

Η **ισχύς** ενός στατιστικού ελέγχου υποθέσεων *ισούται με την πιθανότητα* (1 – β) *ανίχνευσης μιας συγκεκριμένης επίδρασης* όταν η μηδενική υπόθεση (H_0) είναι ψευδής. Η ισχύς είναι απλώς το συμπλήρωμα (1 – β) της πιθανότητας (β) αποτυχίας ανίχνευσης της επίδρασης, δηλαδή το συμπλήρωμα της πιθανότητας ενός σφάλματος τύπου II. Οι σκιασμένοι τομείς στα Σχήματα 11.4, 11.5 και 11.6 αποτυπώνουν διαφορετικούς βαθμούς ισχύος.

Στα Σχήματα 11.5 και 11.6, είχαν επιλεχθεί τα μεγέθη δειγμάτων 36 και 100, λαμβάνοντας υπόψη την ευκολία στις πράξεις, ώστε να τονιστούν οι διαφορετικοί βαθμοί ισχύος για μια μικρή επίδραση τριών μονάδων της βιταμίνης C στο IQ. Κατά προτίμηση, η επιλογή μεγέθους δείγματος θα πρέπει να εκφράζει –όσο το επιτρέπουν οι καταστάσεις– την ώριμη κρίση σας σχετικά με το τι συνιστά (1) τη μικρότερη σημαντική επίδραση και (2) έναν εύλογο βαθμό ισχύος για την ανίχνευση αυτής της επίδρασης. Για παράδειγμα, οι παρακάτω σκέψεις θα μπορούσαν να επηρεάσουν την επιλογή ενός νέου μεγέθους δείγματος για τη μελέτη της βιταμίνης C.

1. Η μικρότερη επίδραση που αξίζει ανίχνευσης θα μπορούσαμε να συμπεράνουμε ότι ισούται με επτά μονάδες. Αυτό είναι κάτι που μπορεί να εκφράζει την κρίση μας, και ενδεχομένως να υποστηρίζεται από συμβούλους εκπαίδευσης, ότι μόνο ένα μέσο IQ τουλάχιστον 107 για όλους τους μαθητές της σχολικής περιφέρειας αιτιολογεί την προσπάθεια και τα έξοδα αναβάθμισης όλου του προγράμματος σπουδών. Ακόμα μια πιθανή αιτία για την εστίαση σε μια επίδραση επτά μονάδων –όταν δεν υπάρχει σημαντικός λόγος για το αντίθετο– θα ήταν ότι, επειδή το 7 είναι περίπου το μισό της τυπικής απόκλισης 15, αποφεύγει ακραίες επιδράσεις μεγέθους, συνιστώντας μια «μέτρια» επίδραση σύμφωνα με τις ευρέως αποδεκτές οδηγίες του Jacob Cohen που περιγράφονται στην Ενότητα 14.9.
2. Ένας πιθανός βαθμός ισχύος γι' αυτήν την επίδραση επτά μονάδων θα μπορούσαμε να συμπεράνουμε ότι ισούται με 0,80. Αυτός ο βαθμός ισχύος θα ανιχνεύσει την καθορισμένη επίδραση με ένα ανεκτό ποσοστό 80%. Όταν δεν υπάρχουν ιδιαίτερες ανησυχίες για το σφάλμα τύπου II, πολλοί ερευνητές θα επέλεγαν το 0,80 ως προεπιλεγμένη τιμή για την ισχύ –μαζί με το 0,05 ως προεπιλεγμένη τιμή για το επίπεδο σημαντικότητας– προκειμένου να αποφύγουν τα μεγάλα μεγέθη δειγμάτων που απαιτούνται από υψηλούς βαθμούς ισχύος, όπως 0,95 ή 0,99.

Καμπύλες ισχύος

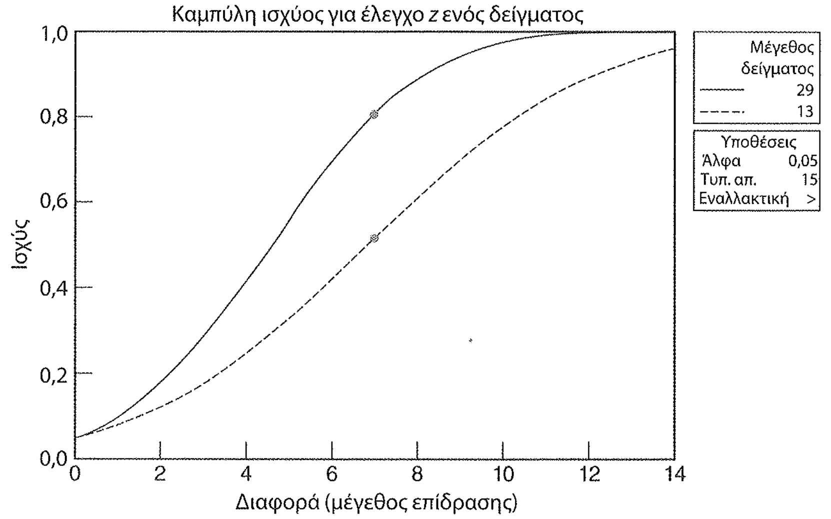
Καμπύλη ισχύος
Δείχνει πώς μεταβάλλεται η πιθανότητα ανίχνευσης οποιασδήποτε πιθανής επίδρασης για ένα σταθερό μέγεθος δείγματος.

Μια **καμπύλη ισχύος** *δείχνει πώς μεταβάλλεται η πιθανότητα ανίχνευσης οποιασδήποτε πιθανής επίδρασης* –από πολύ μικρή ως πολύ μεγάλη– *για ένα σταθερό μέγεθος δείγματος.*[27] Με λίγα μόνο πατήματα πλήκτρων στο πληκτρολόγιο, το λογισμικό *Power and Sample Size* του Minitab υπολογίζει ότι το μέγεθος δείγματος 29 θα ικανοποιήσει τις αρχικές προδιαγραφές ανίχνευσης μιας επίδρασης επτά μονάδων με ισχύ 0,80. Η επάνω (συνεχής γραμμή) καμπύλη ισχύος στο Σχήμα 11.7 βασίζεται σε μέγεθος δείγματος 29 και περιλαμβάνει μια τελεία της οποίας οι συντεταγμένες είναι μια επίδραση επτά μονάδων (διαφορά) και μια ισχύς 0,80.

Η καμπύλη ισχύος σχήματος S για ένα δείγμα 29 αντικειμένων δείχνει επίσης την αύξηση της ισχύος που παρατηρείται όταν υπάρχει αύξηση στην επίδραση του μεγέθους. Επαληθεύστε ότι η ισχύς ισούται μόνο περίπου με 0,40 για μια μικρότερη επίδραση τεσσάρων μονάδων και περίπου με 0,95 για μια μεγαλύτερη επίδραση δέκα μονάδων. Μια επίδραση μονάδων θα ανιχνεύεται μόνο περίπου στο 40% των περιπτώσεων, ενώ μια επίδραση δέκα μονάδων θα ανιχνεύεται περίπου στο 95% των περιπτώσεων.

Πραγματικά θέματα, όπως οι περιορισμοί στους πόρους (χρήματα ή εγκαταστάσεις), θα μπορούσαν να επιβάλουν μια μείωση (πάντα επώδυνη) στο προκαθορισμένο μέγεθος δείγματος 29. Αν και οι αρχικές προδιαγραφές αποτυπώνουν την καλύτερη κρίση μας για το κατάλληλο μέγεθος του δείγματος, υπάρχει συνήθως περιθώριο για συμβιβασμούς. Για παράδειγμα, αναφορικά με το Σχήμα 11.7, θα μπορούσαμε να εξετάσουμε τις ιδιότητες της κάτω καμπύλης ισχύος (με διακεκομμένη γραμμή) για ένα μικρότερο μέγεθος δείγματος 13. (Σε κανονικές συνθήκες, για να ελαχιστοποιήσουμε την απώλεια ισχύος, θα είχαμε προφανώς μελετήσει καμπύλες ισχύος για

27. Για περισσότερες πληροφορίες σχετικά με τις καμπύλες ισχύος, ανατρέξτε στο Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.



ΣΧΗΜΑ 11.7
Καμπύλη ισχύος από το Minitab για το πείραμα με τη βιταμίνη C, δεδομένου ότι n = 29 (συνεχής γραμμή) και n = 13 (διακεκομμένη γραμμή).

πιο μέτριες μειώσεις στο αρχικό μέγεθος δείγματος, όπως 28, 27 κ.λπ., αλλά επιλέξαμε την καμπύλη ισχύος για το 13, έτσι ώστε να τονίσουμε τις γραφικές διαφορές μεταξύ των καμπυλών στο Σχήμα 11.7.) Η τελεία στην καμπύλη ισχύος για δείγμα 13 δείχνει ότι θα γίνει ανίχνευση μιας επίδρασης επτά μονάδων με ισχύ περίπου 0,50. Οι περισσότεροι ερευνητές δεν θα ήταν πρόθυμοι να μειώσουν το μέγεθος δείγματος στο 13, επειδή με μια τέτοια μείωση δεν θα ήταν εύκολο να ανιχνεύουμε μια επίδραση επτά μονάδων· για την ακρίβεια, θα την ανιχνεύαμε περίπου στις μισές περιπτώσεις.

Το μέγεθος δείγματος 29 θα μπορούσε επίσης να μειωθεί έμμεσα κάνοντας συμβιβασμούς σε άλλα χαρακτηριστικά των αρχικών προδιαγραφών. Θα μπορούσαμε να μειώσουμε το προκαθορισμένο μέγεθος δείγματος *μεγεθύνοντας* τη μικρότερη σημαντική επίδραση (κατά προτίμηση με μικρές αυξήσεις πάνω από την επίδραση επτά μονάδων), *μειώνοντας* τον βαθμό ισχύος (κατά προτίμηση όχι πολύ κάτω από 0,80), *αυξάνοντας* το επίπεδο σημαντικότητας (κατά προτίμηση όχι πάνω από 0,10), *επιλέγοντας*, αν θα ήταν δυνατόν, έναν μονόπλευρο και όχι έναν αμφίπλευρο έλεγχο (αν αυτό δεν είχε γίνει ήδη στη μελέτη για τη βιταμίνη C) ή με έναν συνδυασμό όλων των παραπάνω. Για παράδειγμα, θα μπορούσαμε να μεγεθύνουμε τη μικρότερη σημαντική επίδραση από επτά σε οκτώ μονάδες. Αν και δεν φαίνεται στο Σχήμα 11.7, το Minitab υπολογίζει ότι ένα μικρότερο δείγμα 22 αντικειμένων ανιχνεύει τη μεγαλύτερη επίδραση οκτώ μονάδων με ισχύ που ισούται με 0,80.

Επειδή μια ανάλυση ισχύος εξαρτάται από πολλούς παράγοντες, συμπεριλαμβανομένης της αντικειμενικής κρίσης του ερευνητή σχετικά με το ποιο είναι *εύλογο* ποσοστό ανίχνευσης για τη *μικρότερη* σημαντική επίδραση, καθώς και της διαθεσιμότητας τοπικών πόρων και επακόλουθων συμβιβασμών, δύο εξίσου ικανοί ερευνητές μπορεί να επέλεγαν διαφορετικά μεγέθη δείγματος για την ίδια μελέτη. Σε κάθε περίπτωση, στα χέρια ενός δίκαιου ερευνητή,

η χρήση καμπυλών ισχύος αναπαριστά μια διακριτή βελτίωση σε σχέση με την αυθαίρετη επιλογή μεγέθους δείγματος, επειδή οι καμπύλες ισχύος συμβάλλουν στην αναγνώριση ενός μεγέθους δείγματος το οποίο, καθώς δεν είναι ούτε αδικαιολόγητα μικρό ούτε υπερβολικά μεγάλο, δημιουργεί έναν στατιστικό έλεγχο υποθέσεων με την κατάλληλη ευαισθησία.

Ανάλυση ισχύος μελετών άλλων

Αν υποπτεύεστε ότι η *αποτυχία απόρριψης της* H_0 που αναφέρει κάποιος ερευνητής έχει προκληθεί από ένα αδικαιολόγητα μικρό μέγεθος δείγματος, θα μπορούσατε να εξετάσετε τις καμπύλες ισχύος αναδρομικά, ώστε να αξιολογήσετε την επάρκεια των δημοσιευμένων αποτελεσμάτων. Για παράδειγμα, αν το μέγεθος του δείγματος που αναφέρεται για μια μελέτη σχετικά με τη βιταμίνη C ήταν μόνο 13 αντικείμενα, θα μπορούσατε να έχετε βασιστεί στην κάτω καμπύλη του Σχήματος 11.7 για να διαπιστώσετε ότι θα είχε ανιχνευθεί η μικρότερη σημαντική επίδραση των επτά μονάδων με μια πολύ χαμηλή ισχύ (περίπου 0,50). Θα μπορούσατε, επομένως, να υποστηρίξετε την ανάγκη για *αναπαραγωγή* της αρχικής μελέτης με ένα πιο ισχυρό και μεγάλο μέγεθος δείγματος.

Δεν είναι απαραίτητη η πρόβλεψη του σωστού μεγέθους επίδρασης

Η χρήση καμπυλών ισχύος δεν προϋποθέτει την πρόβλεψη του *σωστού* μεγέθους επίδρασης –ένα ανέφικτο έργο–, αλλά απλώς ότι καθορίζετε τη μικρότερη επίδραση η οποία, *αν υπάρχει*, αξίζει να ανιχνευθεί. Αν το πραγματικό μέγεθος της επίδρασης είναι μεγαλύτερο από την καθορισμένη επίδραση, η πραγματική ισχύς θα υπερβαίνει την καθορισμένη ισχύ – επειδή το μεγαλύτερο μέρος της πραγματικής κατανομής δειγματοληψίας επικαλύπτει την περιοχή απόρριψης για την ψευδή H_0 περισσότερο απ' όσο κάνει η κατανομή δειγματοληψίας για την καθορισμένη επίδραση. (Αν αυτό δεν είναι προφανές, συγκρίνετε τα Σχήματα 11.4 και 11.5.) Ως εκ τούτου, μια πιο σημαντική επίδραση έχει ακόμα περισσότερες πιθανότητες να ανιχνευθεί. Από την άλλη πλευρά, αν το πραγματικό μέγεθος επίδρασης στην πραγματικότητα είναι μικρότερο από την καθορισμένη επίδραση, όλη η διαδικασία λειτουργεί αντίστροφα αλλά σε κάθε περίπτωση υπέρ σας, επειδή μια ασήμαντη επίδραση, την οποία πιθανώς να παραβλέπατε, είναι ακόμα λιγότερο πιθανό να ανιχνευθεί.

Έναρξη ανάλυσης ισχύος

Σκοπός αυτού του βιβλίου δεν είναι να παράσχει αναλυτικές πληροφορίες για χειροκίνητους ή ηλεκτρονικούς υπολογισμούς ανάλυσης ισχύος. Οι χειροκίνητοι υπολογισμοί περιγράφονται στο Κεφάλαιο 8 του βιβλίου του D. C. Howell, *Statistical Methods for Psychology*, 8η έκδοση (Belmont, CA: Wadsworth, 2013). Ηλεκτρονικοί υπολογισμοί γίνονται στα τρία πακέτα στατιστικού λογισμικού –Minitab, SPSS και SAS– που χρησιμοποιούμε σ' αυτό το βιβλίο, αλλά και σε πολλούς δωρεάν ιστότοπους, όπως είναι το G*Power 3 στη σελίδα http://www.gpower.hhu.de/. Αφού αποφασίσετε ποια είναι η μικρότερη σημαντική επίδραση που χρήζει ανίχνευσης με μια συγκεκριμένη ισχύ, οι λεπτομέρειες μιας ανάλυσης ισχύος, χειροκίνητης ή ηλεκτρονικής, συνήθως είναι απλές και μπορούν να εφαρμοστούν σε οποιαδήποτε ανάλυση ισχύος που θα μπορούσατε να ξεκινήσετε μόνοι σας.

Έλεγχος προόδου *11.9 Μελετήστε τις καμπύλες ισχύος του Σχήματος 11.7 για να αξιολογήσετε τα κατά προσέγγιση ποσοστά ανίχνευσης, όπως στρογγυλοποιούνται στην εγγύτερη δεκάδα, για τις παρακάτω περιπτώσεις:
(α) μια επίδραση τριών μονάδων, με μέγεθος δείγματος 29
(β) μια επίδραση έξι μονάδων, με μέγεθος δείγματος 13
(γ) μια επίδραση δώδεκα μονάδων, με μέγεθος δείγματος 13
Απαντήσεις στη σελίδα 544.

Έλεγχος προόδου *11.10 Ένας ερευνητής συμβουλεύεται ένα διάγραμμα για να βρει το μέγεθος δείγματος που απαιτείται προκειμένου να ανιχνεύσει μια επίδραση οκτώ μονάδων με πιθανότητα 0,80. Τι συμβαίνει σ' αυτό το ποσοστό ανίχνευσης 0,80 –θα είναι *μικρότερο*, *ίδιο* ή *μεγαλύτερο*– αν, χωρίς να το γνωρίζει ο ερευνητής, η πραγματική επίδραση ισούται στην πραγματικότητα με
(α) δώδεκα μονάδες;
(β) πέντε μονάδες;
Απαντήσεις στη σελίδα 544.

Περίληψη

Η τύχη πρέπει να συνυπολογίζεται όταν λαμβάνουμε μια απόφαση για τη μηδενική υπόθεση (H_0), προσδιορίζοντας αν μια παρατηρούμενη διαφορά θεωρείται κοινό ή σπάνιο αποτέλεσμα. Ακόμα κι αν δεν γνωρίζουμε ποτέ αν μια συγκεκριμένη απόφαση για τη μηδενική υπόθεση είναι αληθής ή όχι, είναι καθησυχαστικό να γνωρίζουμε ότι, μακροπρόθεσμα, οι περισσότερες αποφάσεις θα είναι σωστές, αν θεωρήσουμε ότι οι μηδενικές υποθέσεις είναι είτε αληθείς είτε σημαντικά ψευδείς.

Η απόφαση να διατηρήσουμε την H_0 είναι ασθενής και υπονοεί απλώς ότι η H_0 *θα μπορούσε* να είναι αληθής, ενώ η απόφαση να απορρίψουμε την H_0 είναι ισχυρή και υπονοεί ότι η H_0 είναι *προφανώς* ψευδής (και, αντιστρόφως, ότι η H_1 είναι *προφανώς* αληθής).

Αν και η υπόθεση έρευνας, και όχι η μηδενική υπόθεση, είναι ιδιαίτερα σημαντική, η υπόθεση έρευνας συνήθως ταυτίζεται με την εναλλακτική υπόθεση και ελέγχεται έμμεσα για δύο λόγους: (1) Στερείται της απαραίτητης ακρίβειας και (2) βάσει της λογικής, στηριζόμενοι στο γεγονός ότι η απόρριψη της μηδενικής υπόθεσης (εξαιτίας ενός αρνητικού παραδείγματος ή ενός σπάνιου αποτελέσματος) είναι ισχυρότερη απόφαση από τη διατήρηση της μηδενικής υπόθεσης.

Χρησιμοποιείτε έναν πιο ευαίσθητο μονόπλευρο έλεγχο μόνο όταν πριν από μια έρευνα υπάρχει προβληματισμός για τις αποκλίσεις προς μια συγκεκριμένη κατεύθυνση. Διαφορετικά, χρησιμοποιείτε έναν αμφίπλευρο έλεγχο.

Επιλέγετε τις στατιστικές υποθέσεις μεταξύ των παρακάτω τριών πιθανοτήτων:

Για έναν αμφίπλευρο, μη κατευθυντικό έλεγχο,

$$H_0 : \mu = \text{κάποιος αριθμός}$$
$$H_1 : \mu \neq \text{κάποιος αριθμός}$$

Για έναν μονόπλευρο ή κατευθυντικό έλεγχο με την κρίσιμη τιμή της αριστερής πλευράς,

$$H_0 : \mu \geq \text{κάποιος αριθμός}$$
$$H_1 : \mu < \text{κάποιος αριθμός}$$

Για έναν μονόπλευρο ή κατευθυντικό έλεγχο με την κρίσιμη τιμή της δεξιάς πλευράς,

$$H_0 : \mu \leq \text{κάποιος αριθμός}$$
$$H_1 : \mu > \text{κάποιος αριθμός}$$

Αν δεν υπάρχουν προφανείς αιτίες για την επιλογή μεγαλύτερου ή μικρότερου επιπέδου σημαντικότητας, χρησιμοποιείτε το σύνηθες επίπεδο 0,05.

Υπάρχουν τέσσερα πιθανά αποτελέσματα για οποιονδήποτε στατιστικό έλεγχο υποθέσεων:

- Αν η H_0 είναι στην πραγματικότητα αληθής, είναι σωστή απόφαση να διατηρήσετε την αληθή H_0.
- Αν η H_0 είναι στην πραγματικότητα αληθής, είναι σφάλμα τύπου I να απορρίψετε την αληθή H_0.
- Αν η H_0 είναι στην πραγματικότητα ψευδής, είναι σφάλμα τύπου II να διατηρήσετε την ψευδή H_0.
- Αν η H_0 είναι στην πραγματικότητα ψευδής, είναι σωστή απόφαση να απορρίψετε την ψευδή H_0.

Όταν γίνονται γενικεύσεις πέρα από τα υπάρχοντα δεδομένα, υπάρχει πάντα η πιθανότητα να εμφανιστεί σφάλμα τύπου I ή II. Στην καλύτερη περίπτωση, ένας στατιστικός έλεγχος υποθέσεων τείνει να παράγει μια σωστή απόφαση όταν η H_0 είναι είτε στην πραγματικότητα αληθής είτε στην πραγματικότητα ψευδής εξαιτίας μιας μεγάλης επίδρασης.

Αν η H_0 είναι στην πραγματικότητα αληθής, η πιθανότητα σφάλματος τύπου I, α, ισούται με το επίπεδο σημαντικότητας και η πιθανότητα μιας σωστής απόφασης ισούται με $1 - \alpha$.

Αν η H_0 είναι στην πραγματικότητα ψευδής, η πιθανότητα σφάλματος τύπου II, β, και η πιθανότητα σωστής απόφασης, $1 - \beta$, εξαρτώνται από το μέγεθος της επίδρασης – δηλαδή τη διαφορά μεταξύ του πραγματικού και του υποθετικού μέσου πληθυσμού. Όσο μεγαλύτερη είναι η επίδραση, τόσο μικρότερη είναι η πιθανότητα να υπάρχει σφάλμα τύπου II και τόσο μεγαλύτερη είναι η πιθανότητα να ληφθεί σωστή απόφαση.

Για να αυξήσετε την πιθανότητα ανίχνευσης ψευδούς H_0, ακόμα και μίας ψευδούς H_0 που εκφράζει μια πολύ μικρή επίδραση, χρησιμοποιείτε μεγαλύτερο μέγεθος δείγματος.

Η επιλογή μεγέθους δείγματος θα πρέπει να είναι τέτοια ώστε το μέγεθος να μην είναι αδικαιολόγητα μικρό αλλά ούτε και υπερβολικά μεγάλο. Τότε παράγεται ένας στατιστικός έλεγχος υποθέσεων με την κατάλληλη ευαισθησία.

Οι καμπύλες ισχύος βοηθούν τον ερευνητή να επιλέξει ένα μέγεθος δείγματος που εξασφαλίζει ένα εύλογο ποσοστό ανίχνευσης για τη μικρότερη σημαντική επίδραση. Αν το αρχικά καθορισμένο μέγεθος δείγματος είναι πολύ μεγάλο, μπορεί να μειωθεί αν *μεγεθύνουμε* τη μικρότερη σημαντική επίδραση, αν *μειώσουμε* τον βαθμό ισχύος, αν *αυξήσουμε* το επίπεδο σημαντικότητας, αν *επιλέξουμε*, αν είναι δυνατόν, έναν μονόπλευρο έλεγχο ή με έναν συνδυασμό των παραπάνω.

Σημαντικοί όροι

Αμφίπλευρος ή μη κατευθυντικός έλεγχος
Μονόπλευρος ή κατευθυντικός έλεγχος
Σφάλμα τύπου I
Σφάλμα τύπου II
Άλφα (α)
Επίδραση
Υποθετική κατανομή δειγματοληψίας
Πραγματική κατανομή δειγματοληψίας
Βήτα (β)
Ισχύς ($1 - \beta$)
Καμπύλη ισχύος

Ερωτήσεις επανάληψης

11.11 Δώστε δύο λόγους γιατί η υπόθεση έρευνας δεν ελέγχεται άμεσα.

11.12 Μια γραμμή παραγωγής σε ένα εργοστάσιο γλυκών έχει σχεδιαστεί έτσι ώστε να παράγει κουτιά 2 λιβρών ίδιων γλυκών των οποίων το βάρος ακολουθεί μια κανονική κατανομή με μέσο 33 ουγγιές και τυπική απόκλιση 0,30 της ουγγιάς. Ένα τυχαίο δείγμα 36 κουτιών από την πιο πρόσφατη παραγωγή αποκαλύπτει μέσο βάρος 33,09 ουγγιές. (Παρεμπιπτόντως, έχουμε εδώ μια εξαίρεση από τη συνηθισμένη κατάσταση όπου ο ερευνητής ελπίζει να απορρίψει τη μηδενική υπόθεση.)

(α) Περιγράψτε τον πληθυσμό που υποβάλλεται σε έλεγχο.
(β) Ακολουθώντας τη συνήθη διαδικασία, ελέγξτε τη μηδενική υπόθεση στο επίπεδο σημαντικότητας 0,05.
(γ) Κάποιος χρησιμοποιεί έναν μονόπλευρο έλεγχο, με την κρίσιμη τιμή της δεξιάς πλευράς, επειδή ο δειγματικός μέσος (33,09) είναι μεγαλύτερος από τον υποθετικό μέσο πληθυσμού (33). Τα σχόλιά σας;

11.13 Διαβάστε ξανά το πρόβλημα της Ερώτησης 10.5 στη σελίδα 263.
(α) Ποια μορφή πρέπει να πάρουν οι H_0 και H_1 αν ο ερευνητής ενδιαφέρεται μόνο για τη διαφορά μισθών κατά των γυναικών;
(β) Αν αυτός ο στατιστικός έλεγχος υποθέσεων υποστηρίζει το συμπέρασμα ότι υπάρχει διάκριση τιμών κατά των γυναικών, θα ξεκινήσει μια ακριβή δικαστική αγωγή κατά των πανεπιστημίων των ΗΠΑ. Υπό αυτές τις συνθήκες, προτείνετε να χρησιμοποιηθεί επίπεδο σημαντικότητας 0,05 ή 0,01; Γιατί;

***11.14** Για το πείραμα με τη βιταμίνη C του κεφαλαίου, θα μπορούσατε να περιγράψετε τη μηδενική υπόθεση με σύμβολα και λέξεις ως εξής:

$$H_0: \mu \leq 100, \text{ δηλαδή η βιταμίνη C δεν αυξάνει το IQ}$$

Ακολουθώντας τη μορφή του Πίνακα 11.2 και προσπαθώντας να είστε όσο το δυνατόν πιο συγκεκριμένοι, θα μπορούσατε να περιγράψετε τα τέσσερα πιθανά αποτελέσματα του πειράματος για τη βιταμίνη C ως εξής:

	Κατάσταση της H_0	
Απόφαση	**Αληθής H_0**	**Ψευδής H_0**
Διατήρηση της H_0	*Σωστή απόφαση:* Συμπεραίνουμε ότι δεν υπάρχει απόδειξη πως η βιταμίνη C αυξάνει το IQ όταν στην πραγματικότητα δεν το κάνει.	*Σφάλμα τύπου II:* Συμπεραίνουμε ότι δεν υπάρχει απόδειξη ότι η βιταμίνη C αυξάνει το IQ όταν στην πραγματικότητα το κάνει.
Απόρριψη της H_0	*Σφάλμα τύπου I:* Συμπεραίνουμε ότι η βιταμίνη C αυξάνει το IQ όταν στην πραγματικότητα δεν το κάνει.	*Σωστή απόφαση:* Συμπεραίνουμε ότι η βιταμίνη C αυξάνει το IQ όταν στην πραγματικότητα το κάνει.

Χρησιμοποιώντας την απάντηση για το πείραμα της βιταμίνης C ως υπόδειγμα, καθορίστε τη μηδενική υπόθεση και τα τέσσερα πιθανά αποτελέσματα των παρακάτω ασκήσεων:

***(α)** Ερώτηση 11.1(β) στη σελίδα 277.
Η απάντηση στη σελίδα 544.
(β) Ερώτηση 11.1(γ).

11.15 Πρέπει να ανησυχούμε για τα τέσσερα πιθανά αποτελέσματα *πριν* προχωρήσουμε σε στατιστικό έλεγχο της υπόθεσης.
(α) Αν υποθέσουμε ότι ο έλεγχος έχει ήδη γίνει και η *μηδενική υπόθεση διατηρείται*, για ποια από τα τέσσερα πιθανά αποτελέσματα πρέπει να ανησυχούμε ακόμα;
(β) Αν υποθέσουμε ότι ο έλεγχος έχει ήδη γίνει και η *μηδενική υπόθεση απορρίπτεται*, για ποια από τα τέσσερα πιθανά αποτελέσματα πρέπει να ανησυχούμε ακόμα;

11.16 Χρησιμοποιώντας το επίπεδο σημαντικότητας 0,05, ένας ερευνητής διατηρεί την H_0. Σύμφωνα μ' αυτόν, υπάρχει πιθανότητα 0,95 η H_0 να είναι αληθής. Τα σχόλιά σας;

11.17 Σε μια άλλη μελέτη, μια ερευνήτρια απορρίπτει την H_0 στο επίπεδο σημαντικότητας 0,01. Σύμφωνα μ' αυτή, υπάρχει μια πιθανότητα 0,99 η H_0 να είναι ψευδής. Τα σχόλιά σας;

11.18 Για έναν προβλεπόμενο μονόπλευρο έλεγχο με κρίσιμη τιμή αριστερής πλευράς, στο επίπεδο σημαντικότητας 0,05, σχεδιάστε δύο γενικά διαγράμματα. Κάθε διάγραμμα θα πρέπει να εμφανίζει τον τομέα στην πραγματική κατανομή δειγματοληψίας που παράγει ένα σφάλμα τύπου II και τον τομέα που παράγει μια ορθή απόφαση. Το ένα διάγραμμα πρέπει να αποτυπώνει την περίπτωση η H_0 να είναι στην πραγματικότητα ψευδής επειδή ο πραγματικός μέσος πληθυσμού είναι *ελαφρώς μικρότερος* από τον υποθετικό μέσο πληθυσμού και το άλλο διάγραμμα θα πρέπει να αποτυπώνει την περίπτωση η H_0 να είναι στην πραγματικότητα ψευδής επειδή ο πραγματικός μέσος πληθυσμού είναι *αισθητά μικρότερος* από τον υποθετικό μέσο πληθυσμού. (**Υπόδειξη:** Πρώτα βρείτε τον κανόνα απόφασης για τον υποθετικό μέσο πληθυσμού και έπειτα σχεδιάστε την πραγματική κατανομή δειγματοληψίας για κάθε περίπτωση.)

11.19 Πώς πρέπει να τροποποιηθεί ένας προβλεπόμενος στατιστικός έλεγχος υποθέσεων αν ενδιαφέρεστε κυρίως για
(α) το σφάλμα τύπου I;
(β) το σφάλμα τύπου II;

11.20 Ανατρέξτε στις καμπύλες ισχύος του Σχήματος 11.7 για να εκτιμήσετε το κατά προσέγγιση ποσοστό ανίχνευσης, στρογγυλοποιημένο στην εγγύτερη δεκάδα, για τις εξής περιπτώσεις:
(α) μια επίδραση τεσσάρων μονάδων, με μέγεθος δείγματος 13
(β) μια επίδραση δέκα μονάδων, με μέγεθος δείγματος 29
(γ) μια επίδραση επτά μονάδων, με μέγεθος δείγματος 18 (με παρεμβολή)

11.21 Μελετήστε το Σχήμα 11.7 για να αξιολογήσετε το κατά προσέγγιση μέγεθος της μικρότερης σημαντικής επίδρασης που θα μπορούσε να ανιχνευθεί...
(α) με πιθανότητα 0,80, δεδομένου μεγέθους δείγματος 13.
(β) με πιθανότητα 0,50, δεδομένου μεγέθους δείγματος 29.

11.22 Το Σχήμα 11.7 παρουσιάζει καμπύλες ισχύος για μεγέθη δειγμάτων 13 και 29. Χρησιμοποιώντας αυτές τις καμπύλες ως αναφορά, υποδείξτε γενικά (αν είναι μικρότερο από 13, μεταξύ 13 και 29 ή μεγαλύτερο από 29) το μέγεθος δείγματος που είναι απαραίτητο για την ανίχνευση μεγέθους επίδρασης
(α) 4 με ισχύ 0,80
(β) 8 με ισχύ 0,80
(γ) 8 με ισχύ 0,50
(δ) 10 με ισχύ 0,80

11.23 Για κάθε ομάδα εναλλακτικών προτάσεων που βλέπετε παρακάτω, ελέγξτε την προτεινόμενη προεπιλεγμένη τιμή, δηλαδή την τιμή που θα πρέπει να υιοθετήσετε, εκτός αν υπάρχει σοβαρός λόγος για το αντίθετο.

(α) μονόπλευρος έλεγχος με κρίσιμη τιμή αριστερής πλευράς ________ μονόπλευρος έλεγχος με κρίσιμη τιμή δεξιάς πλευράς ________ αμφίπλευρος έλεγχος ________

(β) επίπεδο σημαντικότητας 0,10 ________ επίπεδο σημαντικότητας 0,05 ________ επίπεδο σημαντικότητας 0,01

(γ) ισχύς 0,50 ________ ισχύς 0,80 ________ ισχύς 0,95 ________

ΠΑΡΑΡΤΗΜΑ XI

Περισσότερα για τον στατιστικό έλεγχο υποθέσεων

Γιώργος Ανδρουλάκης

Π.XI.1 Εισαγωγή

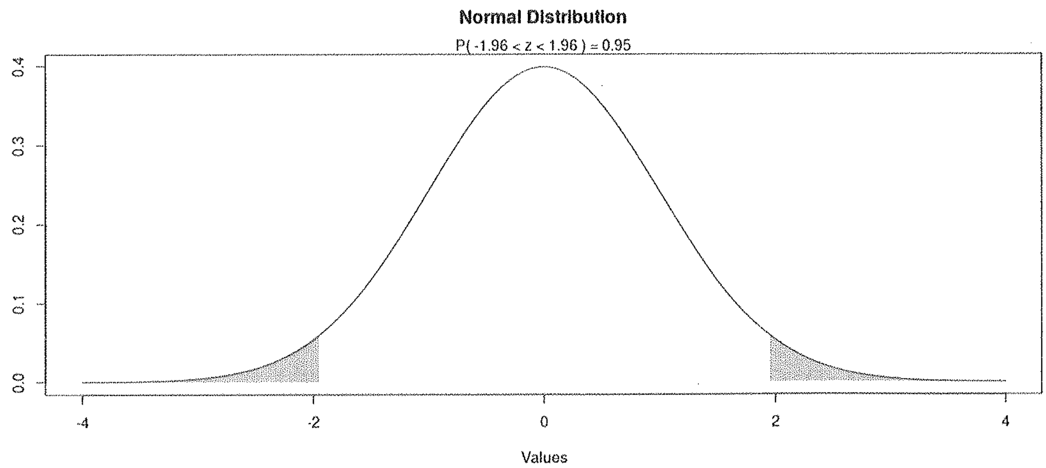
Το παρόν Κεφάλαιο 11 συμπληρώνει τα παραδείγματα που ξεκινήσαμε να παρουσιάζουμε στο Κεφάλαιο 10 εστιάζοντας στους μονόπλευρους ελέγχους υποθέσεων.

Π.XI.2 Μονόπλευροι έλεγχοι υποθέσεων

Εφαρμόζοντας τον κώδικα του παραδείγματος της προηγούμενης ενότητας και αλλάζοντας μόνο τις τιμές των παραμέτρων στις δύο πρώτες γραμμές, μπορούμε να φτιάξουμε το γράφημα της τυποποιημένης κανονικής κατανομής, δηλαδή της κατανομής με μέσο 0 και τυπική απόκλιση 1. Προκειμένου να οριοθετήσουμε την περιοχή που αφήνει τόσο από αριστερά όσο και από δεξιά το 2,5% της συνολικής περιοχής κάτω από τη συνάρτηση πιθανότητας, γνωρίζουμε ότι τα αντίστοιχα κρίσιμα σημεία είναι το –1,96 και το 1,96 αντίστοιχα. Επομένως, ο παρακάτω κώδικας της R

```
> mean=0; sd=1
> lb=-1.96; ub=1.96
> x <- seq(-4,4,length=100)*sd + mean
> hx <- dnorm(x,mean,sd)
> plot(x, hx, type="n", xlab="Values", ylab="",
+ main="Normal Distribution", axes=T)
> lines(x, hx)
> for (j in seq(mean - 4*sd, lb, length=100)) {segments(j, 0, j, dnorm(j,mean,sd), col = "red")}
> for (j in seq(ub, mean + 4*sd, length=100)) {segments(j, 0, j, dnorm(j,mean,sd), col = "red")}
> area <- pnorm(ub, mean, sd) - pnorm(lb, mean, sd)
> result <- paste("P(",lb,"< z <",ub,") =",
+             signif(area, digits=3))
> mtext(result,3)
```

θα έχει ως αποτέλεσμα τη δημιουργία του παρακάτω γραφήματος, στο οποίο οι δύο περιοχές έχουν σκιαγραφηθεί με κόκκινη διαγράμμιση.

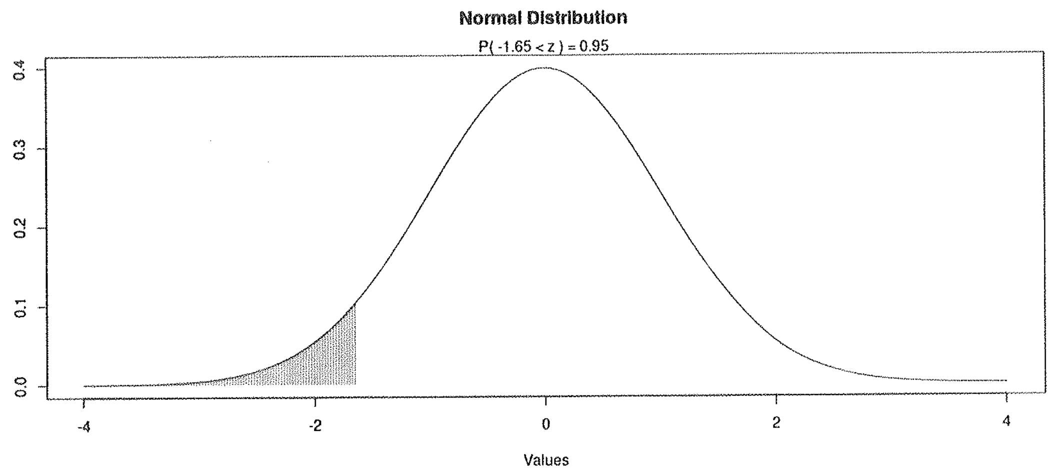


Ένα μεγάλο πλεονέκτημα στον προγραμματισμό με επαναχρησιμοποιήσιμο κώδικα είναι η ευκολία με την οποία με ελάχιστες μετατροπές σε προϋπάρχουσα υλοποίηση μπορούμε να δημιουργήσουμε κώδικα για νέες στατιστικές διεργασίες.

Έστω ότι ενδιαφερόμαστε να σκιαγραφήσουμε μόνο την περιοχή αριστερά, ώστε να οπτικοποιήσουμε την περιοχή απόρριψης για έναν μονόπλευρο έλεγχο υπόθεσης. Ένας προσεκτικός αναγνώστης εύκολα θα διαπιστώσει ότι κυριολεκτικά με τρεις αλλαγές στον προηγούμενο κώδικα υλοποίησης θα πάρουμε το επιθυμητό αποτέλεσμα. Συγκεκριμένα, η πρώτη προσαρμογή αφορά τις παραμέτρους της δεύτερης γραμμής, όπου το αριστερό άκρο της κρίσιμης περιοχής οριοθετείται στο −1,65 (επίσης δεν μας χρειάζεται στη δεύτερη γραμμή το άνω άκρο ub της κρίσιμης περιοχής). Η δεύτερη προσαρμογή αφορά τη διαγραφή της γραμμής που ήταν υπεύθυνη για τη δημιουργία της δεξιάς σκιαγράφησης και η τρίτη προσαρμογή αφορά τον τύπο υπολογισμού του ορισμένου ολοκληρώματος area. Ο νέος κώδικας είναι ο παρακάτω:

```
> mean=0; sd=1
> lb=-1.65; ub=1.65
> x <- seq(-4,4,length=100)*sd + mean
> hx <- dnorm(x,mean,sd)
> plot(x, hx, type="n", xlab="Values", ylab="",
+     main="Normal Distribution", axes=T)
> lines(x, hx)
> for (j in seq(mean - 4*sd, lb, length=100)) {segments(j, 0, j, dnorm(j,mean,sd), col = "red")}
> area <- pnorm(mean+4*sd, mean, sd) - pnorm(lb, mean, sd)
> result <- paste("P(",lb,"< z ) =",
+            signif(area, digits=3))
> mtext(result,3)
```

Ο οποίος μας δίνει το παρακάτω γράφημα:

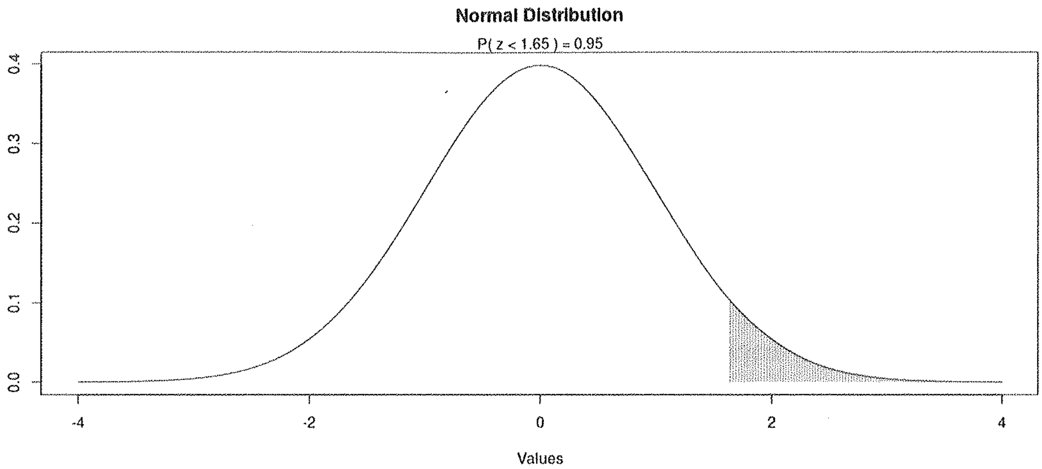


Αντίστοιχα, για να σκιαγραφήσουμε το δεξιό άκρο του σχήματος, ώστε να οπτικοποιήσουμε την κρίσιμη περιοχή απόρριψης από δεξιά, ο προσαρμοσμένος κώδικας είναι ο παρακάτω:

```
> mean=0; sd=1
> lb=-1.65; ub=1.65
> x <- seq(-4,4,length=100)*sd + mean
> hx <- dnorm(x,mean,sd)
> plot(x, hx, type="n", xlab="Values", ylab="",
+     main="Normal Distribution", axes=T)
> lines(x, hx)
> for (j in seq(ub, mean + 4*sd, length=100)) {segments(j, 0, j, dnorm(j,mean,sd), col = "red")}
```

```
> area <- pnorm(ub, mean, sd) - pnorm(mean - 4*sd, mean, sd)
> result <- paste("P( z <",ub,") =",
+           signif(area, digits=3))
> mtext(result,3)
```

Ο οποίος μας δίνει το παρακάτω γράφημα:



Π.XI.3 Επιλογή επιπέδου σημαντικότητας (*α*)

Στην ενότητα αυτή θα δούμε πώς μπορούμε να εκτελέσουμε μονόπλευρο έλεγχο υποθέσεων με το λογισμικό R, καθώς και τις διαφορές που παρουσιάζει αυτός ανάλογα με το είδος που επιλέγουμε.

Αρχικά, για τις ανάγκες του παραδείγματός μας δημιουργούμε ένα τυχαίο δείγμα 20 ακέραιων αριθμών με μέση τιμή 500 και τυπική απόκλιση sd. Για να παραγάγουμε τους τυχαίους αριθμούς χρησιμοποιούμε την έτοιμη συνάρτηση rnorm, ενώ για να εξασφαλίσουμε ότι αυτοί θα είναι ακέραιοι τους στρογγυλοποιούμε σε μηδέν δεκαδικά ψηφία, άρα τους κάνουμε ακέραιους. Επειδή η παραγωγή τυχαίων αριθμών επηρεάζεται από τη χρονική στιγμή που επιλέγουμε να τους φτιάξουμε, όταν δοκιμάσετε το παράδειγμα αυτό στον υπολογιστή σας θα δημιουργηθούν πιθανότατα τελείως διαφορετικοί τυχαίοι αριθμοί.

```
> myx = round(rnorm(20, 500, sd),0)
> myx
[1] 500 501 500 504 500 501 502 501 499 500 499 501 499 499 501 500 499 500 499 501
```

Οι τυχαίοι αριθμοί που παράχθηκαν με την παραπάνω διαδικασία τοποθετήθηκαν σε μία μεταβλητή με το όνομα myx. Σύμφωνα με το παράδειγμα της προηγούμενης ενότητας, για να εκτελέσουμε αμφίπλευρο έλεγχο χρησιμοποιούμε τη συνάρτηση z.test της βιβλιοθήκης BSDA. Στο παράδειγμά μας, ο έλεγχος υπόθεσης εκτελείται έχοντας ορίσματα μέση τιμή 500 και τυπική απόκλιση 11.

```
> z.test(myx, y=NULL, mu=500, sigma.x = 11)

        One-sample z-Test

data: myx
z = 0.12197, p-value = 0.9029
alternative hypothesis: true mean is not equal to 500
95 percent confidence interval:
  495.4791 505.1209
sample estimates:
```

```
mean of x
      500.3
```

Για να σιγουρευτούμε ότι εκτελέσαμε τον ορθό έλεγχο υπόθεσης κοιτάμε στο αποτέλεσμα που λάβαμε πώς έχει εκφραστεί η εναλλακτική υπόθεση (alternative hypothesis): Πράγματι, ο έλεγχος που εκτελέσαμε είναι αμφίπλευρος, αφού ως εναλλακτική υπόθεση αναφέρεται από το στατιστικό πακέτο ότι ο μέσος δεν είναι ίσος με 500.

Η επόμενη πληροφορία που αναζητάμε από το στατιστικό πακέτο είναι το αποτέλεσμα του στατιστικού ελέγχου. Εξετάζουμε τη γραμμή που αναφέρεται στη z-τιμή και στην αντίστοιχη p-τιμή του στατιστικού ελέγχου. Παρατηρούμε ότι οι 20 τυχαίοι αριθμοί έχουν μέση τιμή που αντιστοιχεί στην τιμή z=0,12197 της τυποποιημένης κανονικής κατανομής, η οποία εκφράζεται με την p-τιμή 0,9029. Αν υποθέσουμε ότι επιθυμούσαμε ο στατιστικός μας έλεγχος να έχει 95% στατιστική σημαντικότητα, τότε για να οδηγηθούμε σε απόρριψη της μηδενικής μας υπόθεσης θα έπρεπε ο δείκτης z να είναι στην περιοχή απόρριψης, δηλαδή στην περιοχή που οριοθετείται στο 5% του εμβαδού κάτω από την καμπύλη της τυποποιημένης κανονικής κατανομής. Επομένως, αφού δοκιμάζουμε αμφίπλευρο έλεγχο, η περιοχή απόρριψης περιλαμβάνει το 2,5% του εμβαδού αριστερά του γραφήματος και το 2,5% του εμβαδού στα δεξιά του γραφήματος. Άρα, η κρίσιμη p-τιμή του στατιστικού μας ελέγχου πρέπει να συγκριθεί με το 0,025. Στην περίπτωσή μας, η p-τιμή είναι 0,9029, τιμή δηλαδή πολύ μεγαλύτερη από το 0,025, επομένως δεν μπορούμε να απορρίψουμε τη μηδενική υπόθεση.

Ας δούμε τώρα πώς θα μπορούσαμε να εκτελέσουμε μονόπλευρο έλεγχο. Η μόνη διαφοροποίηση σε σχέση με προηγουμένως είναι το όρισμα alternative στην εκτέλεση της εντολής z.test. Παρατηρούμε ότι δίνουμε στο όρισμα την τιμή "less".

```
> z.test(myx, y=NULL, alternative = "less", mu=500, sigma.x = 11)

        One-sample z-Test

data: myx
z = 0.12197, p-value = 0.5485
alternative hypothesis: true mean is less than 500
95 percent confidence interval:
      NA 504.3458
sample estimates:
mean of x
      500.3
```

Ποιον έλεγχο εκτελέσαμε λοιπόν; Η πρόθεσή μας με το όρισμα που βάλαμε είναι ο έλεγχος που αντιστοιχεί σε εναλλακτική υπόθεση για μέσο μικρότερο του 500. Πράγματι, ορθό! Επιβεβαιώνεται και από τη γραμμή του αποτελέσματος που τυπώθηκε στην οθόνη μας και αντιστοιχεί στην εναλλακτική υπόθεση. Θα απορρίψουμε ή όχι τη μηδενική μας υπόθεση, δηλαδή ότι ο μέσος είναι μεγαλύτερος ή ίσος από 500; Η z-τιμή δεν άλλαξε. Γιατί να αλλάξει; Το διάνυσμα των 20 τυχαίων ακέραιων τιμών που έχουμε στη μεταβλητή myx δεν έχει αλλάξει, επομένως ούτε η μέση τιμή του και άρα ούτε και η z-τιμή του στην τυποποιημένη κανονική κατανομή. Γιατί όμως η αντίστοιχη p-τιμή άλλαξε; Γιατί, αφού εκτελούμε μονόπλευρο έλεγχο, έχουν διαφοροποιηθεί οι περιοχές απόρριψης ή μη απόρριψης του στατιστικού ελέγχου και, επομένως, αλλάζει η αντίστοιχη πιθανότητα η z-τιμή των δεδομένων μας να βρίσκεται στην αντίστοιχη περιοχή. Παρατηρούμε ότι η p-τιμή είναι 0,5485, το οποίο είναι μεγαλύτερο του 0,05 και άρα δεν έχουμε επαρκή στατιστική ένδειξη ώστε να απορρίψουμε τη μηδενική υπόθεση.

Είδαμε στη θεωρία ότι υπάρχουν δύο είδη μονόπλευρων ελέγχων, αυτός με το «≥» και αυτός με το «≤». Προηγουμένως είδαμε πώς εκτελούμε τον πρώτο από τους δύο αυτούς ελέγχους. Εύλογα, για να εκτελέσουμε τον δεύτερο, αρκεί μόνο να το δηλώσουμε σωστά στο όρισμα alternative της κλήσης της συνάρτησης z.test, όπως παρακάτω:

```
> z.test(myx, y=NULL, alternative = "greater", mu=500, sigma.x = 11)

        One-sample z-Test

data: myx
```

```
z = 0.12197, p-value = 0.4515
alternative hypothesis: true mean is greater than 500
95 percent confidence interval:
 496.2542       NA
sample estimates:
mean of x
    500.3
```

Παρατηρούμε από το αποτέλεσμα του παραπάνω στατιστικού ελέγχου ότι η εναλλακτική υπόθεση αφορά την περίπτωση στην οποία ο μέσος είναι μεγαλύτερος από το 500. Όπως και προηγουμένως, η *z*-τιμή δεν έχει αλλάξει, ενώ η *p*-τιμή άλλαξε σε 0,4515, τιμή που αντιστοιχεί στην τιμή που προκύπτει αν αφαιρέσουμε από τη μονάδα την *p*-τιμή που είχαμε βρει στον προηγούμενο έλεγχο. Και σε αυτόν τον στατιστικό έλεγχο δεν καταφέρνουμε να απορρίψουμε τη μηδενική υπόθεση.

Μέχρι τώρα είδαμε στατιστικούς ελέγχους σε επίπεδο σημαντικότητας 95%. Μάλιστα, στην υλοποίηση του *z* ελέγχου υπόθεσης δεν ορίζαμε πουθενά το επίπεδο σημαντικότητας, γιατί η αντίστοιχη υλοποίηση την είχε ως default. Πώς μπορούμε να αλλάξουμε το επίπεδο σημαντικότητας και να ορίσουμε κάτι διαφορετικό; Χρησιμοποιώντας το όρισμα conf.level στη συνάρτηση z.test της βιβλιοθήκης BSDA της R. Συγκεκριμένα, αν θέλουμε να χρησιμοποιήσουμε επίπεδο σημαντικότητας 99% σε αμφίπλευρο έλεγχο για τα δεδομένα μας myx, θα εκτελέσουμε την ακόλουθη εντολή:

```
> z.test(myx, y=NULL, mu=500, sigma.x = 11, conf.level = 0.99)

        One-sample z-Test

data: myx
z = 0.12197, p-value = 0.9029
alternative hypothesis: true mean is not equal to 500
99 percent confidence interval:
 493.9643 506.6357
sample estimates:
mean of x
    500.3
```

Παρατηρούμε ότι το αποτέλεσμα της εκτέλεσης της παραπάνω εντολής δεν έχει αλλάξει τις τιμές που αφορούν τον έλεγχο υπόθεσης σε σχέση με αυτές που εμφανίστηκαν όταν είχαμε εκτελέσει έλεγχο υπόθεσης σε επίπεδο σημαντικότητας 95%. Γιατί; Γιατί, όταν αλλάζουμε το επίπεδο σημαντικότητας, στην πραγματικότητα το μόνο που αλλάζει δεν είναι οι *z* και *p*-τιμές αλλά η «ευαισθησία» μας για το πότε θα απορρίψουμε ή όχι τη μηδενική υπόθεση.

Ανάλογα με προηγουμένως, αν επιθυμούσαμε να εκτελέσουμε μονόπλευρο έλεγχο σε επίπεδο σημαντικότητας 99%, θα χρησιμοποιούσαμε την παρακάτω εντολή:

```
> z.test(myx, y=NULL, alternative = "less", mu=500, sigma.x = 11, conf.level = 0.90)

        One-sample z-Test

data: myx
z = 0.12197, p-value = 0.5485
alternative hypothesis: true mean is less than 500
90 percent confidence interval:
        NA 503.4522
sample estimates:
mean of x
    500.3
```

και για τους ίδιους λόγους που εξηγήσαμε στο προηγούμενο παράδειγμα δεν έχει αλλάξει τίποτα στο αποτέλεσμα που πήραμε συγκρίνοντάς το με τον αντίστοιχο έλεγχο επιπέδου σημαντικότητας 95%.

Ομοίως, για τον εναλλακτικό μονόπλευρο έλεγχο σε επίπεδο σημαντικότητας 99% θα παίρναμε:

```
> z.test(myx, y=NULL, alternative ="greater", mu=500, sigma.x = 11, conf.level = 0.99)

One-sample z-Test

data: myx
z = 0.12197, p-value = 0.4515
alternative hypothesis: true mean is greater than 500
99 percent confidence interval:
   494.5779 NA
sample estimates:
mean of x
      500.3
```

Βιβλιογραφία

Becker, R. A., Chambers, J. M., and Wilks, A. R. (1988). *The New S Language*. Wadsworth & Brooks/Cole.

Beckerman, A. P., and Petchey, O. L. (2012). *Getting Started with R: An introduction for biologists* (Oxford University Press, Oxford) [Κεφάλαιο 3].

Crawley, M. J. (2005). *Statistics: An introduction using R* (John Wiley & Sons, Chichester).

Keen, K. J. (2010). *Graphics for Statistics and Data Analysis with R*. CRC Press.

Raykov, T., and Marcoulides, G. A. (2013). *Basic Statistics: An introduction with R* (Rowman and Littlefield, Plymouth).

Καρλής, Δ., και Ντζούφρας, Ι. (2015). *Εισαγωγή στον Προγραμματισμό και τη Στατιστική Ανάλυση με R*. (https://repository.kallipos.gr/handle/11419/2601)

Φωκιανός, Κ., και Χαραλάμπους, Χ. (2010). *Εισαγωγή στην R* – Πρόχειρες Σημειώσεις, 2η έκδοση. Τμήμα Μαθηματικών και Στατιστικής, Πανεπιστήμιο Κύπρου. (https://cran.r-project.org/doc/contrib/mainfokianoscharalambous.pdf)