ΚΕΦΑΛΑΙΟ 9

Κατανομή δειγματοληψίας του μέσου

Πρόλογος

Αυτό το κεφάλαιο εστιάζει στην πιο σημαντική έννοια στην επαγωγική στατιστική – την έννοια της κατανομής δειγματοληψίας. Η κατανομή δειγματοληψίας εξυπηρετεί ως πλαίσιο αναφοράς για κάθε αποτέλεσμα, μεταξύ όλων των πιθανών αποτελεσμάτων, που θα μπορούσε να προκύψει από τύχη. Επανεμφανίζεται σε όλα τα επόμενα κεφάλαια ως το κλειδί για την κατανόηση με ποιον τρόπο, αφού αξιολογηθεί η μεταβλητότητα, μπορούμε να γενικεύσουμε πέρα από ένα περιορισμένο σύνολο πραγματικών παρατηρήσεων. Για να χρησιμοποιήσουμε μια κατανομή δειγματοληψίας, πρέπει να προσδιορίσουμε τον μέσο, την τυπική της απόκλιση και το σχήμα της – ένα φαινομενικά δύσκολο έργο το οποίο χάρη στη θεωρία της στατιστικής μπορεί να εκτελεστεί με τη βοήθεια του μέσου πληθυσμού, της τυπικής απόκλισης πληθυσμού και της καμπύλης κανονικής κατανομής αντίστοιχα.

Κάποιοι από εσάς μπορεί να συμμετείχατε σε εξετάσεις SAT και, αν ναι, μάλλον θα θυμάστε τον βαθμό σας. Έστω ότι η βαθμολογία σε εξετάσεις μαθηματικών SAT για όλους τους υποψήφιους φοιτητές την τελευταία περίοδο εξέτασης κατανέμεται γύρω από έναν μέσο 500 με τυπική απόκλιση 110. Ένας ερευνητής σε ένα πανεπιστήμιο σχεδιάζει να εξετάσει την πρόταση ότι, κατά μέσο όρο, η βαθμολογία σε εξετάσεις μαθηματικών SAT για τους πρωτοετείς σε τοπικό επίπεδο ισούται με τον εθνικό μέσο όρο 500. Το έργο του θα ήταν αρκετά απλό αν υπήρχαν διαθέσιμοι οι βαθμοί εξετάσεων μαθηματικών για όλους τους ντόπιους πρωτοετείς. Τότε, αφού θα υπολόγιζε τη μέση βαθμολογία για όλους τους ντόπιους πρωτοετείς, μια άμεση σύγκριση θα έδειχνε αν, κατά μέσο όρο, οι ντόπιοι πρωτοετείς πετυχαίνουν βαθμούς κάτω, ακριβώς ή πάνω από τον εθνικό μέσο όρο.

Ας υποθέσουμε ότι δεν είναι δυνατόν να πάρουμε τους βαθμούς για όλη την τάξη των πρωτοετών. Αντίθετα, παίρνουμε τους βαθμούς εξετάσεων μαθηματικών SAT για ένα τυχαίο δείγμα 100 φοιτητών από τον τοπικό πληθυσμό πρωτοετών και ο μέσος βαθμός αυτού του δείγματος ισούται με 533. Αν κάθε δείγμα ήταν ακριβές αντίγραφο του πληθυσμού, οι γενικεύσεις από το δείγμα στον πληθυσμό δεν θα ήταν κάτι δύσκολο. Αφού παρατηρούσαμε έναν μέσο βαθμό 533 για το δείγμα των 100 πρωτοετών φοιτητών, θα μπορούσαμε να συμπεράνουμε, χωρίς καθυστέρηση, ότι ο μέσος βαθμός σε εξετάσεις μαθηματικών για όλη την τάξη πρωτοετών επίσης ισούται με 533 και, επομένως, υπερβαίνει τον εθνικό μέσο όρο.

9.1 Τι είναι η κατανομή δειγματοληψίας;

Τυχαία δείγματα σπάνια εκπροσωπούν ακριβώς τον υπό εξέταση πληθυσμό. Ακόμα και μια μέση βαθμολογία σε εξετάσεις μαθηματικών (533) θα μπορούσε να προέρχεται, ακόμα κι από τύχη, από έναν πληθυσμό πρωτοετών των οποίων ο μέσος ισούται με τον εθνικό μέσο όρο του 500. Αναλόγως, οι γενικεύσεις από ένα μόνο δείγμα σε έναν πληθυσμό είναι πολύ πιο αβέβαιες. Πράγματι, οι γενικεύσεις δεν βασίζονται απλώς σε έναν μόνο δειγματικό μέσο 533, αλλά επίσης στην κατανομή του – μια κατανομή δειγματικών μέσων για όλα τα πιθανά τυχαία δείγματα. Εκπροσωπώντας το μοντέλο του στατιστικολόγου για τυχαία αποτελέσματα,

Κατανομή δειγματοληψίας του μέσου
Κατανομή πιθανοτήτων των μέσων για όλα τα πιθανά τυχαία δείγματα ενός δεδομένου μεγέθους από κάποιον πληθυσμό.

η *κατανομή δειγματοληψίας του μέσου* αναφέρεται στην κατανομή πιθανοτήτων των μέσων για όλα τα πιθανά τυχαία δείγματα ενός δεδομένου μεγέθους από κάποιον πληθυσμό.

Ουσιαστικά, αυτή η κατανομή αναπαριστά τη μεταβλητότητα μεταξύ δειγματικών μέσων που θα μπορούσαν να προκύψουν μόνο από τύχη και, επομένως, δημιουργεί ένα πλαίσιο αναφοράς για μια γενίκευση από έναν μόνο δειγματικό μέσο σε έναν μέσο πληθυσμού.

Η κατανομή δειγματοληψίας του μέσου μάς επιτρέπει να διαπιστώσουμε αν, δεδομένης της μεταβλητότητας μεταξύ όλων των πιθανών δειγματικών μέσων, ο παρατηρούμενος δειγματικός μέσος μπορεί να θεωρηθεί *κοινό* αποτέλεσμα ή *σπάνιο* αποτέλεσμα (από μια στοιχισμένη στο κέντρο κατανομή, εν προκειμένω κοντά στο 500). Εάν ο δειγματικός μέσος 533 θεωρείται *κοινό* αποτέλεσμα σ' αυτήν την κατανομή δειγματοληψίας, τότε η διαφορά μεταξύ 533 και 500 δεν είναι αρκετά μεγάλη σε σχέση με τη μεταβλητότητα όλων των πιθανών δειγματικών μέσων έτσι ώστε να σηματοδοτεί ότι κάτι ιδιαίτερο συμβαίνει στον υποκείμενο πληθυσμό. Επομένως, μπορούμε να συμπεράνουμε ότι η μέση βαθμολογία σε μια εξέταση μαθηματικών για όλη την τάξη πρωτοετών θα μπορούσε να είναι ίδια με τον εθνικό μέσο όρο 500. Από την άλλη πλευρά, αν ο δειγματικός μέσος 533 θεωρείται *σπάνιο αποτέλεσμα* σ' αυτήν την κατανομή δειγματοληψίας, τότε η διαφορά μεταξύ 533 και 500 είναι αρκετά μεγάλη σε σχέση με τη μεταβλητότητα όλων των πιθανών δειγματικών μέσων έτσι ώστε να σηματοδοτεί ότι κάτι ιδιαίτερο συμβαίνει στον υποκείμενο πληθυσμό. Επομένως, μπορούμε να συμπεράνουμε ότι η μέση βαθμολογία σε μια εξέταση μαθηματικών για όλη την τάξη πρωτοετών θα μπορούσε πιθανώς να υπερβαίνει τον εθνικό μέσο όρο 500.

Όλα τα πιθανά τυχαία δείγματα

Όταν επιχειρούμε να γενικεύσουμε από έναν μόνο δειγματικό μέσο σε έναν μέσο πληθυσμού, πρέπει να συμβουλευόμαστε την κατανομή δειγματοληψίας του μέσου. Στη συγκεκριμένη περίπτωση, αυτή η κατανομή βασίζεται σε *όλα τα πιθανά τυχαία δείγματα*, όπου καθένα έχει μέγεθος 100 και μπορεί να εξαχθεί από τον ντόπιο πληθυσμό πρωτοετών. *Όλα τα πιθανά τυχαία δείγματα* δεν είναι ο αριθμός των δειγμάτων μεγέθους 100 που απαιτούνται προκειμένου να *ερευνηθεί πλήρως* ο ντόπιος πληθυσμός πρωτοετών, αλλά ο αριθμός διαφορετικών τρόπων με τους οποίους ένα *μόνο* δείγμα μεγέθους 100 μπορεί να επιλεχθεί απ' αυτόν τον πληθυσμό.

«Όλα τα πιθανά τυχαία δείγματα» συνήθως είναι ένας αρκετά μεγάλος αριθμός. Για παράδειγμα, αν ο ντόπιος πληθυσμός περιλάμβανε τουλάχιστον 1.000 πρωτοετείς, ο συνολικός αριθμός πιθανών τυχαίων δειγμάτων μεγέθους 100 θα ήταν αστρονομικός. Τα 301 ψηφία αυτού του αριθμού θα έκαναν ακόμα και το εθνικό χρέος να φαντάζει μικρό. Ακόμα και με τη βοήθεια υπολογιστών, το έργο της κατασκευής αυτής της κατανομής δειγματοληψίας από την αρχή θα ήταν τρομακτικό, καταγράφοντας κάθε μέσο για όλα τα πιθανά τυχαία δείγματα.

Ευτυχώς, η στατιστική θεωρία μάς παρέχει σημαντικές πληροφορίες για την κατανομή δειγματοληψίας του μέσου, όπως θα περιγράψουμε στο υπόλοιπο κεφάλαιο. Εφοδιασμένοι μ' αυτά τα στοιχεία για κατανομές δειγματοληψίας, θα επιστρέψουμε στο συγκεκριμένο παράδειγμα στο επόμενο κεφάλαιο και θα ελέγξουμε τον ισχυρισμό ότι ο μέσος βαθμός σε εξετάσεις μαθηματικών για τον ντόπιο πληθυσμό πρωτοετών ισούται με τον εθνικό μέσο όρο 500. Μόνο σε εκείνο το σημείο –και όχι στο τέλος αυτού του κεφαλαίου– θα κατανοήσετε πλήρως τον ρόλο των κατανομών δειγματοληψίας σε πρακτικές εφαρμογές.

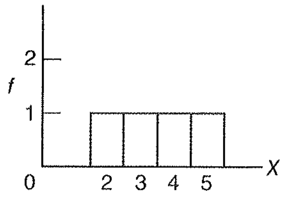
9.2 Δημιουργία κατανομής δειγματοληψίας από την αρχή

Θα προσπαθήσουμε να ορίσουμε ακριβώς το τι σημαίνει η κατανομή δειγματοληψίας δημιουργώντας μία από την αρχή, σε εξαιρετικά απλοποιημένες συνθήκες. Έστω ότι έχουμε έναν απίθανα μικρό πληθυσμό που αποτελείται από τέσσερις παρατηρήσεις με τιμές 2, 3, 4 και 5, όπως βλέπετε στο Σχήμα 9.1. Στη συνέχεια θα καταγράψουμε χωριστά όλα τα πιθανά τυχαία δείγματα τα οποία έχουν όλα μέγεθος 2 που θα μπορούσαν να ληφθούν απ' αυτόν τον πληθυσμό. Υπάρχουν τέσσερις πιθανότητες στην πρώτη εξαγωγή από τον πληθυσμό και άλλες τέσσερις πιθανότητες στη δεύτερη εξαγωγή από τον πληθυσμό, όπως δείχνει ο Πίνακας 9.1.[24] Τα δύο σύνολα πιθανοτήτων συνδυάζονται έτσι ώστε να δώσουν συνολικά 16 πιθανά δείγματα. Σας θυμίζουμε ότι προσπαθούμε να αποσαφηνίσουμε την έννοια της κατανομής δειγματοληψίας του μέσου. Στην πράξη λαμβάνουμε μόνο ένα τυχαίο δείγμα και όχι 16 πιθανά δείγματα, από τον πληθυσμό – το μέγεθος του δείγματος θα ήταν πολύ μικρό σε σχέση με έναν πολύ μεγαλύτερο πληθυσμό και, βέβαια, δεν θα ήταν γνωστές όλες οι παρατηρήσεις στον πληθυσμό.

Για καθένα από τα 16 πιθανά δείγματα, ο Πίνακας 9.1 παρουσιάζει επίσης έναν δειγματικό μέσο (προκύπτει από το άθροισμα των δύο παρατηρήσεων και τη διαίρεση διά 2) και την πιθανότητά του να υπάρχει (εκφράζεται ως $\frac{1}{16}$, επειδή καθένα από τα 16 πιθανά δείγματα είναι εξίσου πιθανό). Με μετατροπή σε σχετική συχνότητα ή κατανομή πιθανοτήτων, όπως στον Πίνακα 9.2, οι 16 δειγματικοί μέσοι απαρτίζουν την κατανομή δειγματοληψίας του μέσου, η οποία προηγουμένως είχε οριστεί ως η κατανομή πιθανοτήτων των μέσων για όλα τα πιθανά τυχαία δείγματα ενός δεδομένου μεγέθους από κάποιον πληθυσμό. Δεν έχουν όλες οι τιμές για τον δειγματικό μέσο ίσες πιθανότητες να εμφανιστούν στον Πίνακα 9.2, επειδή κάποιες τιμές εμφανίζονται περισσότερες από μία φορές

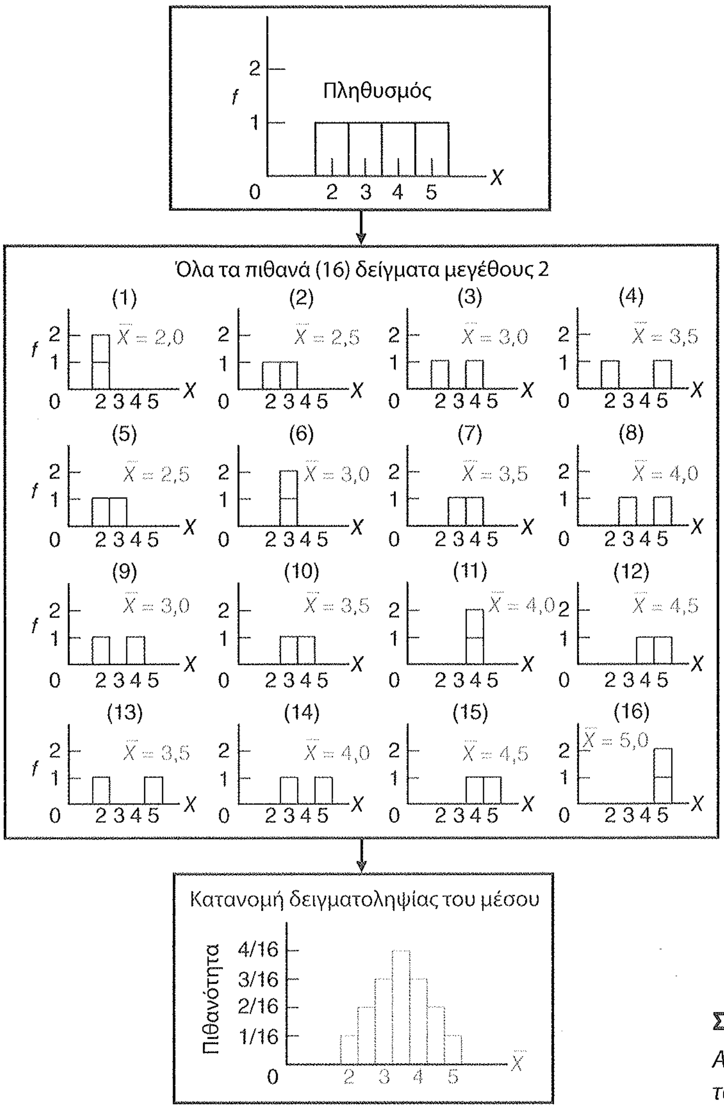
Πίνακας 9.1
ΟΛΑ ΤΑ ΠΙΘΑΝΑ ΔΕΙΓΜΑΤΑ ΜΕΓΕΘΟΥΣ 2 ΑΠΟ ΕΝΑΝ ΠΟΛΥ ΜΙΚΡΟ ΠΛΗΘΥΣΜΟ

	Όλα τα πιθανά δείγματα	Μέσος ($\overline{X}$)	Πιθανότητα
(1)	2,2	2,0	1/16
(2)	2,3	2,5	1/16
(3)	2,4	3,0	1/16
(4)	2,5	3,5	1/16
(5)	3,2	2,5	1/16
(6)	3,3	3,0	1/16
(7)	3,4	3,5	1/16
(8)	3,5	4,0	1/16
(9)	4,2	3,0	1/16
(10)	4,3	3,5	1/16
(11)	4,4	4,0	1/16
(12)	4,5	4,5	1/16
(13)	5,2	3,5	1/16
(14)	5,3	4,0	1/16
(15)	5,4	4,5	1/16
(16)	5,5	5,0	1/16



ΣΧΗΜΑ 9.1
Διάγραμμα ενός πολύ μικρού πληθυσμού.

24. Κανονικά, λαμβάνουμε ένα δείγμα από μια μεμονωμένη παρατήρηση μόνο μία φορά, δηλαδή η δειγματοληψία γίνεται *χωρίς αντικατάσταση*. Αν ωστόσο χρησιμοποιούνταν με το συγκεκριμένο και εξαιρετικά απλοποιημένο παράδειγμα, η δειγματοληψία χωρίς αντικατάσταση θα μεγέθυνε μια ασήμαντη τεχνική προσαρμογή.



ΣΧΗΜΑ 9.2

Ανάδειξη της κατανομής δειγματοληψίας του μέσου από όλα τα πιθανά δείγματα.

Πίνακας 9.2
ΚΑΤΑΝΟΜΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΤΟΥ ΜΕΣΟΥ (ΔΕΙΓΜΑΤΑ ΜΕΓΕΘΟΥΣ 2 ΑΠΟ ΕΝΑΝ ΠΟΛΥ ΜΙΚΡΟ ΠΛΗΘΥΣΜΟ)

Δειγματικός μέσος ($\overline{X}$)	Πιθανότητα
5,0	1/16
4,5	2/16
4,0	3/16
3,5	4/16
3,0	3/16
2,5	2/16
2,0	1/16

στα 16 πιθανά δείγματα. Για παράδειγμα, η τιμή 3,5 για έναν δειγματικό μέσο εμφανίζεται σε 4 από τις 16 πιθανότητες και έχει πιθανότητα $^4/_{16}$.

Πιθανότητα ενός συγκεκριμένου δειγματικού μέσου

Η κατανομή στον Πίνακα 9.2 μπορεί να μας βοηθήσει να προσδιορίσουμε την πιθανότητα λήψης ενός συγκεκριμένου δειγματικού μέσου ή ενός συνόλου δειγματικών μέσων. Για παράδειγμα, η πιθανότητα 5 για έναν τυχαία επιλεγμένο δειγματικό μέσο ισούται με $^1/_{16}$ ή 0,0625. Σύμφωνα με τον προσθετικό κανόνα για αμοιβαία αποκλειόμενα αποτελέσματα, όπως είδαμε στο Κεφάλαιο 8, η πιθανότητα που έχει ένας τυχαία επιλεγμένος δειγματικός μέσος να είναι 5,0 ή 2,0 ισούται με $^1/_{16} + {}^1/_{16} = {}^2/_{16} = 0{,}1250$. Αυτός ο τύπος εύρεσης της πιθανότητας, ο οποίος βασίζεται σε μια κατανομή δειγματοληψίας, έχει ουσιώδη ρόλο στην επαγωγική στατιστική και θα εμφανιστεί πολλές φορές στο υπόλοιπο βιβλίο.

Ανασκόπηση

Το Σχήμα 9.2 συνοψίζει την προηγούμενη συζήτηση. Αποτυπώνει την ανάδειξη της κατανομής δειγματοληψίας του μέσου από το σύνολο όλων των πιθανών (16) δειγμάτων μεγέθους 2 βάσει του πολύ μικρού πληθυσμού των τεσσάρων παρατηρήσεων. Μελετήστε καλά αυτό το σχήμα, καθώς θα αναφερθούμε ξανά σ' αυτό.

Πίνακας 9.3
ΣΥΜΒΟΛΑ ΓΙΑ ΤΟΝ ΜΕΣΟ ΚΑΙ ΤΗΝ ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ ΤΡΙΩΝ ΤΥΠΩΝ ΚΑΤΑΝΟΜΩΝ

Τύπος κατανομής	Μέσος	Τυπική απόκλιση
Δείγμα	$\overline{X}$	s
Πληθυσμός	μ	σ
Κατανομή δειγματοληψίας του μέσου	$\mu_{\overline{X}}$	$\sigma_{\overline{X}}$ (τυπικό σφάλμα του μέσου)

Υπενθύμιση:
Προτείνουμε να αποστηθίσετε τα σύμβολα του Πίνακα 9.3.

Έλεγχος προόδου *9.1 Έστω ότι έχουμε έναν πολύ απλό πληθυσμό που αποτελείται από μόλις πέντε παρατηρήσεις: 2, 4, 6, 8, 10.

(α) Παραθέστε όλα τα πιθανά δείγματα μεγέθους 2.

(β) Κατασκευάστε έναν πίνακα σχετικών συχνοτήτων που αποτυπώνει την κατανομή δειγματοληψίας του μέσου.

Απαντήσεις στη σελίδα 541.

9.3 Μερικά σημαντικά σύμβολα

Αφού καταλάβουμε επακριβώς τι είναι η κατανομή δειγματοληψίας σε εξαιρετικά απλοποιημένες συνθήκες, μπορούμε να παρουσιάσουμε τα ειδικά σύμβολα που προσδιορίζουν τον μέσο και την τυπική απόκλιση της κατανομής δειγματοληψίας του μέσου. Ο Πίνακας 9.3 παρουσιάζει επίσης τα αντίστοιχα σύμβολα για το δείγμα και για τον πληθυσμό. Προτείνουμε να αποστηθίσετε αυτά τα σύμβολα.

Γνωρίζετε ήδη τα αγγλικά γράμματα $\overline{X}$ και s, τα οποία αναπαριστούν τον μέσο και την τυπική απόκλιση οποιουδήποτε δείγματος, αλλά και τα ελληνικά γράμματα μ και σ, τα οποία αναπαριστούν τον μέσο και την τυπική απόκλιση οποιουδήποτε πληθυσμού. Νέα είναι τα ελληνικά γράμματα $\mu_{\overline{X}}$ και $\sigma_{\overline{X}}$, τα οποία αναπαριστούν τον μέσο και την τυπική απόκλιση αντίστοιχα της κατανομής δειγματοληψίας του μέσου. Για να προλάβουμε οποιαδήποτε σύγχυση, ο τελευταίος όρος, το $\sigma_{\overline{X}}$, αναφέρεται συχνά ως *τυπικό σφάλμα του μέσου* ή απλώς *τυπικό σφάλμα*.

Σημαντικότητα ελληνικών γραμμάτων

Σημειώστε ότι τα ελληνικά γράμματα χρησιμοποιούνται για την περιγραφή χαρακτηριστικών τόσο πληθυσμών όσο και κατανομών δειγματοληψίας, υποδηλώνοντας την ύπαρξη ενός κοινού χαρακτηριστικού. Αμφότεροι οι τύποι κατανομής σχετίζονται με όλες τις πιθανότητες, δηλαδή με *όλες τις πιθανές παρατηρήσεις* στον πληθυσμό, ή με τους *μέσους όλων των πιθανών τυχαίων δειγμάτων* στην κατανομή δειγματοληψίας του μέσου.

Μπορούμε τώρα να εστιάσουμε στα τρία πιο σημαντικά χαρακτηριστικά της κατανομής δειγματοληψίας του μέσου: Τον μέσο, την τυπική απόκλιση και το σχήμα του. Σε επόμενα κεφάλαια, αυτά τα τρία χαρακτηριστικά θα αποτελέσουν τη βάση για την εφαρμοσμένη λειτουργία τους στην επαγωγική στατιστική.

Έλεγχος προόδου *9.2 Αξιοποιώντας όσα έχετε μάθει, προσθέστε τα ειδικά σύμβολα για τον μέσο πληθυσμού __(α)__, τον μέσο της κατανομής δειγματοληψίας του μέσου __(β)__, τον δειγματικό μέσο __(γ)__, το τυπικό σφάλμα του μέσου __(δ)__, την τυπική απόκλιση δείγματος __(ε)__ και την τυπική απόκλιση πληθυσμού __(στ)__.

Απαντήσεις στη σελίδα 541.

9.4 Μέσος όλων των δειγματικών μέσων ($\mu_{\overline{X}}$)

Ο ίδιος ο δειγματικός μέσος μιας κατανομής έχει δικό του μέσο.

Ο *μέσος της κατανομής δειγματοληψίας του μέσου* ισούται πάντα με τον μέσο του πληθυσμού.

Όταν χρησιμοποιούμε σύμβολα, έχουμε

Μέσος της κατανομής δειγματοληψίας

$$\mu_{\bar{X}} = \mu \tag{9.1}$$

Μέσος της κατανομής δειγματοληψίας του μέσου ($\mu_{\bar{X}}$)
Ο μέσος όλων των δειγματικών μέσων ισούται πάντα με τον μέσο του πληθυσμού.

όπου το $\mu_{\bar{X}}$ αναπαριστά τον μέσο της κατανομής δειγματοληψίας και το μ αναπαριστά τον μέσο του πληθυσμού.

Εναλλάξιμοι μέσοι

Επειδή ο μέσος όλων των δειγματικών μέσων ($\mu_{\bar{X}}$) ισούται πάντα με τον μέσο του πληθυσμού (μ), μπορούμε να χρησιμοποιούμε αμφότερους αυτούς τους όρους στην επαγωγική στατιστική για να εκφράσουμε την ίδια έννοια. Οποιοσδήποτε ισχυρισμός για τον μέσο πληθυσμού μπορεί να μεταφερθεί ευθέως στον μέσο της κατανομής δειγματοληψίας και το αντίστροφο. Εάν, σύμφωνα με τα ισχυριζόμενα, ο μέσος βαθμός σε εξετάσεις μαθηματικών για τον ντόπιο πληθυσμό πρωτοετών ισούται με τον εθνικό μέσο όρο 500, τότε ο μέσος της κατανομής δειγματοληψίας επίσης θα ισούται αυτόματα με 500. Για τον ίδιο λόγο, μπορείτε να θεωρήσετε τον παρατηρούμενο δειγματικό μέσο 533 ως την απόκλιση είτε από τον μέσο της κατανομής δειγματοληψίας είτε από τον μέσο πληθυσμού. Θα πρέπει, επομένως, να είναι προφανές ότι, *αν μια έκφραση περιλαμβάνει το $\mu_{\bar{X}}$ ή το μ, αντανακλά ως επί το πλείστον μια διαφορά στην έμφαση που δίνεται στην κατανομή δειγματοληψίας ή στον πληθυσμό αντίστοιχα και όχι οποιαδήποτε διαφορά στην αριθμητική τιμή.*

Εξήγηση

Αν και είναι σημαντικό, δεν είναι ιδιαίτερα εντυπωσιακό το γεγονός ότι ο μέσος όλων των δειγματικών μέσων ισούται με τον μέσο πληθυσμού. Όπως μπορούμε να δούμε στο Σχήμα 9.2, τα δείγματα δεν είναι ακριβή αντίγραφα του πληθυσμού και οι περισσότεροι μέσοι δειγμάτων είναι είτε μεγαλύτεροι είτε μικρότεροι από τον μέσο πληθυσμού (είναι 3,5 στο Σχήμα 9.2). Παίρνοντας όμως τον μέσο όλων των δειγματικών μέσων, ουσιαστικά εξουδετερώνετε τις τυχαίες διαφορές που μπορεί να υπάρχουν μεταξύ των δειγματικών μέσων και διατηρείτε μια τιμή ίση με τον μέσο πληθυσμού.

Έλεγχος προόδου *9.3 Διευκρινίστε αν οι παρακάτω προτάσεις είναι σωστές ή λάθος. Ο μέσος όλων των δειγματικών μέσων, $\mu_{\bar{X}}$, …
(α) ισούται πάντα με την τιμή ενός συγκεκριμένου δειγματικού μέσου.
(β) ισούται με 100 αν ο μέσος πληθυσμού ισούται πράγματι με 100.
(γ) συνήθως ισούται με την τιμή ενός συγκεκριμένου δειγματικού μέσου.
(δ) μπορεί να χρησιμοποιηθεί ως όρος αντί του μέσου πληθυσμού.
Απαντήσεις στη σελίδα 541.

9.5 Τυπικό σφάλμα του μέσου ($\sigma_{\bar{X}}$)

Η κατανομή δειγματικών μέσων έχει επίσης τυπική απόκλιση, η οποία αναφέρεται ως τυπικό σφάλμα του μέσου.

Το *τυπικό σφάλμα του μέσου* ισούται με την τυπική απόκλιση του πληθυσμού διά της τετραγωνικής ρίζας του μεγέθους του δείγματος.

Τυπικό σφάλμα του μέσου ($\sigma_{\bar{X}}$)
Ένα γενικό μέτρο της μέσης ποσότητας κατά την οποία οι δειγματικοί μέσοι αποκλίνουν από τον μέσο της κατανομής δειγματοληψίας ή από τον μέσο πληθυσμού.

Όταν χρησιμοποιούμε σύμβολα, έχουμε

Τυπικό σφάλμα του μέσου

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \tag{9.2}$$

όπου το $\sigma_{\bar{X}}$ αναπαριστά το τυπικό σφάλμα του μέσου, το σ αναπαριστά την τυπική απόκλιση του πληθυσμού και το n αναπαριστά το μέγεθος του δείγματος.

Ειδικός τύπος τυπικής απόκλισης

Το τυπικό σφάλμα του μέσου αποτελεί έναν ειδικό τύπο τυπικής απόκλισης που υπολογίζει τη μεταβλητότητα στην κατανομή δειγματοληψίας. Μας παρέχει ένα *πρότυπο*, όπως μια βαθμολογημένη ράβδο για το μέτρο, το οποίο περιγράφει την ποσότητα κατά την οποία οι δειγματικοί μέσοι αποκλίνουν από τον μέσο της κατανομής δειγματοληψίας ή από τον μέσο πληθυσμού. Το *σφάλμα* στην έννοια του τυπικού σφάλματος είναι ένας όρος που δεν αναφέρεται σε λάθη πράξεων, αλλά σε λάθη στις γενικεύσεις που αποδίδονται στο γεγονός ότι, μόνο από τύχη, τα περισσότερα τυχαία δείγματα δεν είναι ακριβή αντίγραφα του πληθυσμού.

Θα ήταν χρήσιμο να σκεφτείτε το τυπικό σφάλμα του μέσου ως ένα γενικό μέτρο της μέσης ποσότητας κατά την οποία οι δειγματικοί μέσοι αποκλίνουν από τον μέσο της κατανομής δειγματοληψίας ή από τον μέσο πληθυσμού.

Στον βαθμό που το σχήμα της κατανομής των δειγματικών μέσων προσομοιάζει με καμπύλη κανονικής κατανομής, όπως θα δούμε στην επόμενη ενότητα, περίπου το 68% όλων των δειγματικών μέσων αποκλίνει λιγότερο από ένα τυπικό σφάλμα από τον μέσο της κατανομής δειγματοληψίας, ενώ μόλις περίπου το 5% όλων των δειγματικών μέσων αποκλίνει περισσότερο από δύο τυπικά σφάλματα από τον μέσο αυτής της κατανομής.

Επίδραση μεγέθους δείγματος

Μια ιδιαίτερα σημαντική επιπλοκή του Τύπου 9.2 είναι ότι, κάθε φορά που το μέγεθος δείγματος ισούται με 2 ή είναι μεγαλύτερο από το 2, η μεταβλητότητα της κατανομής δειγματοληψίας είναι μικρότερη από τη μεταβλητότητα του πληθυσμού. Μπορείτε να δείτε ένα δείγμα αυτής της επίδρασης στο Σχήμα 9.2, όπου οι μέσοι όλων των πιθανών δειγμάτων συσσωρεύονται πιο κοντά στον μέσο πληθυσμού (3,5) από τις τέσσερις αρχικές παρατηρήσεις στον πληθυσμό. Η επίδραση φαίνεται καλύτερα όταν τα μεγέθη δειγμάτων είναι μεγαλύτερα. Νωρίτερα σ' αυτό το κεφάλαιο, για παράδειγμα, δόθηκε η τιμή 110 ως τιμή του σ, της τυπικής απόκλισης πληθυσμού για βαθμολογίες σε εξετάσεις SAT. Πολύ μικρότερη είναι η μεταβλητότητα στην κατανομή δειγματοληψίας των μέσων βαθμών SAT, όπου καθεμία βασίζεται σε δείγματα 100 πρωτοετών. Σύμφωνα με τον Τύπο 9.2, στο συγκεκριμένο παράδειγμα,

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{110}{\sqrt{100}} = \frac{110}{10} = 11$$

υπάρχει δεκαπλάσια μείωση στη μεταβλητότητα, από 110 ως 11, όταν απομακρυνόμαστε από τον πληθυσμό και επικεντρωνόμαστε στην κατανομή δειγματοληψίας.

Σύμφωνα με τον Τύπο 9.2, οποιαδήποτε αύξηση στο μέγεθος δείγματος μεταφράζεται σε ένα μικρότερο τυπικό σφάλμα και, επομένως, σε μια *νέα* κατανομή δειγματοληψίας με μικρότερη μεταβλητότητα. Με μεγαλύτερο μέγεθος δείγματος, οι δειγματικοί μέσοι συσσωρεύονται πιο κοντά στον μέσο της κατανομής δειγματοληψίας και στον μέσο του πληθυσμού και, επομένως, επιτρέπουν πιο ακριβείς γενικεύσεις από τα δείγματα στους πληθυσμούς.

Εξήγηση

Δεν μας προκαλεί έκπληξη το γεγονός ότι η μεταβλητότητα θα πρέπει να είναι μικρότερη στις κατανομές δειγματοληψίας από τους πληθυσμούς. Η τυπική απόκλιση πληθυσμού εκφράζει τη μεταβλητότητα μεταξύ *μεμονωμένων παρατηρήσεων* και επηρεάζεται άμεσα από οποιαδήποτε μεγάλη ή μικρή παρατήρηση στον πληθυσμό. Από την άλλη πλευρά, το τυπικό σφάλμα του μέσου εκφράζει τη μεταβλητότητα μεταξύ των *δειγματικών μέσων*, οι οποίοι αναπαριστούν μια συλλογή από μεμονωμένες παρατηρήσεις. Η εμφάνιση σχετικά μεγάλων ή μικρών παρατηρήσεων μέσα σε ένα συγκεκριμένο δείγμα τείνει να επηρεάζει τον δειγματικό μέσο μόνο ελαφρώς, εξαιτίας της σταθεροποιητικής παρουσίας στο ίδιο δείγμα άλλων παρατηρήσεων μεσαίου επιπέδου ή ακόμα και ακραίων παρατηρήσεων στην αντίθετη κατεύθυνση. Αυτή η σταθεροποιητική επίδραση γίνεται ακόμα πιο έντονη όταν τα μεγέθη δειγμάτων μεγαλώνουν.

Έλεγχος προόδου *9.4 Διευκρινίστε αν οι παρακάτω προτάσεις είναι σωστές ή λάθος. Το τυπικό σφάλμα του μέσου, $\sigma_{\bar{X}}$, ...

(α) μετρά γενικά τη μέση ποσότητα κατά την οποία οι δειγματικοί μέσοι αποκλίνουν από τον μέσο του πληθυσμού.
(β) μετρά τη μεταβλητότητα σε ένα συγκεκριμένο δείγμα.
(γ) αυξάνεται σε τιμή όταν τα μεγέθη δειγμάτων είναι μεγαλύτερα.
(δ) ισούται με 5, δεδομένου ότι $\sigma = 40$ και $n = 64$.

Απαντήσεις στις σελίδες 541-542.

9.6 Σχήμα της κατανομής δειγματοληψίας

Προϊόν της στατιστικής θεωρίας, όπως εκφράζεται στην απλούστερη μορφή του,

Κεντρικό οριακό θεώρημα
Ανεξάρτητα από το σχήμα του πληθυσμού, το σχήμα της κατανομής δειγματοληψίας του μέσου προσομοιάζει με καμπύλη κανονικής κατανομής εάν το μέγεθος του δείγματος είναι επαρκώς μεγάλο.

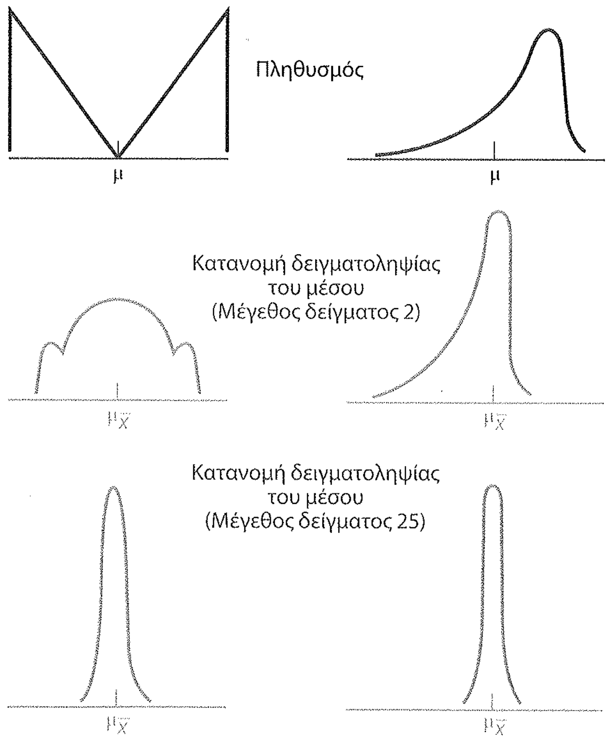
το *κεντρικό οριακό θεώρημα* δηλώνει ότι, ανεξάρτητα από το σχήμα του πληθυσμού, το σχήμα της κατανομής δειγματοληψίας του μέσου προσομοιάζει με καμπύλη κανονικής κατανομής *εάν το μέγεθος του δείγματος είναι επαρκώς μεγάλο.*

Σύμφωνα μ' αυτό το θεώρημα, δεν έχει σημασία αν το σχήμα του αρχικού πληθυσμού είναι κανονικό, θετικά ασύμμετρο, αρνητικά ασύμμετρο ή κάτι περίεργα ακανόνιστο, εφόσον το μέγεθος δείγματος είναι επαρκώς μεγάλο. Το τι σημαίνει η φράση «επαρκώς μεγάλο» εξαρτάται από το σχήμα του αρχικού γονικού πληθυσμού. Αν το σχήμα του αρχικού πληθυσμού είναι κανονικό, τότε οποιοδήποτε μέγεθος δείγματος (ακόμα και το μέγεθος δείγματος 1) θα είναι επαρκώς μεγάλο, αλλιώς, ανάλογα με τον βαθμό της μη κανονικότητας στον αρχικό πληθυσμό, ένα μέγεθος δείγματος μεταξύ 25 και 100 είναι επαρκώς μεγάλο.

Παραδείγματα

Για τον πληθυσμό με το μη κανονικό σχήμα στο επάνω πλαίσιο του Σχήματος 9.2, το σχήμα της κατανομής δειγματοληψίας στο κάτω πλαίσιο αποκαλύπτει μια αρχική μετατόπιση προς την κανονικότητα –δηλαδή ένα σχήμα που έχει κορυφή στο κέντρο και λεπταίνει σε κάθε πλευρά– ακόμα και για πολύ μικρά δείγματα μεγέθους 2. Για τους δύο μη κανονικούς πληθυσμούς του επάνω πλαισίου του Σχήματος 9.3, τα σχήματα των κατανομών δειγματοληψίας στο μεσαίο πλαίσιο δείχνουν ουσιαστικά την ίδια αρχική μετατόπιση προς την κανονικότητα όταν το μέγεθος του δείγματος ισούται με 2, ενώ τα σχήματα των κατανομών δειγματοληψίας στο κάτω πλαίσιο προσεγγίζουν αρκετά την κανονικότητα όταν το μέγεθος δείγματος ισούται με 25.

Νωρίτερα σ' αυτό το κεφάλαιο, το 533 ήταν η μέση βαθμολογία σε εξετάσεις μαθηματικών SAT για ένα τυχαίο δείγμα 100 πρωτοετών. Επειδή αυτό το μέγεθος δείγματος ικανοποιεί τις προϋποθέσεις



ΣΧΗΜΑ 9.3
Επίδραση του κεντρικού οριακού θεωρήματος.

του κεντρικού οριακού θεωρήματος, μπορούμε να θεωρήσουμε ότι ο δειγματικός μέσος 533 προκύπτει από μια κατανομή δειγματοληψίας της οποίας το σχήμα προσομοιάζει με καμπύλη κανονικής κατανομής, ακόμα κι αν δεν έχουμε πληροφορίες για το σχήμα του πληθυσμού των βαθμών σε εξετάσεις μαθηματικών για όλη την τάξη των πρωτοετών. Θα ήταν, επομένως, δυνατό να κάνουμε ακριβείς προτάσεις γι' αυτήν την κατανομή δειγματοληψίας, όπως περιγράφεται στο επόμενο κεφάλαιο, παραπέμποντας στον πίνακα για την τυποποιημένη καμπύλη κανονικής κατανομής.

Γιατί λειτουργεί το κεντρικό οριακό θεώρημα

Σε μια καμπύλη κανονικής κατανομής, όπως ίσως θα θυμάστε, οι ενδιάμεσες τιμές είναι οι πλέον επικρατούσες, ενώ οι ακραίες τιμές, μεγαλύτερες ή μικρότερες, καταλαμβάνουν τις πλευρές που εμφανίζουν μικρότερο εμβαδόν (λεπταίνουν). Γιατί, όταν το μέγεθος δείγματος είναι μεγάλο, η κατανομή δειγματοληψίας προσεγγίζει μια καμπύλη κανονικής κατανομής, ακόμα κι αν ο αρχικός πληθυσμός μπορεί να μην είναι κανονικός;

Πολλοί δειγματικοί μέσοι με ενδιάμεσες τιμές

Όταν το μέγεθος δείγματος είναι μεγάλο, είναι *πολύ πιθανό* ότι οποιοδήποτε δείγμα θα περιέχει όλο το φάσμα μικρών, ενδιάμεσων και μεγάλων αποτελεσμάτων από τον αρχικό πληθυσμό, *ανεξαρτήτως σχήματος*. Ο υπολογισμός ενός μέσου γι' αυτόν τον τύπο δείγματος τείνει να εξουδετερώνει ή να εξασθενίζει τις επιδράσεις των ακραίων αποτελεσμάτων και ο δειγματικός μέσος καταλήγει να έχει κάποια ενδιάμεση τιμή. Αναλόγως, ενδιάμεσες τιμές επικρατούν στην κατανομή δειγματοληψίας και συσσωρεύονται γύρω από μια συχνότητα κορυφής που εκπροσωπεί την πιο κοινή ή επικρατούσα τιμή του δειγματικού μέσου, όπως βλέπετε στο κάτω μέρος του Σχήματος 9.3.

Μερικοί δειγματικοί μέσοι με ακραίες τιμές

Για να καταλάβετε τις σπανιότερες τιμές των δειγματικών μέσων στις ουρές της κατανομής δειγματοληψίας, θα πρέπει να εστιάσετε σε εκείνα τα σχετικά σπάνια δείγματα τα οποία, μόνο από τύχη, περιέχουν κάτι λιγότερο από το πλήρες φάσμα των αποτελεσμάτων από τον αρχικό πληθυσμό. Μερικές φορές, εξαιτίας του σχετικά μεγάλου αριθμού ακραίων αποτελεσμάτων σε μια συγκεκριμένη κατεύθυνση, ο υπολογισμός ενός μέσου εξασθενίζει μόνο ελαφρώς την επίδρασή τους και ο δειγματικός μέσος καταλήγει να έχει κάποια ακραία τιμή. Η πιθανότητα να πάρουμε ακραίες τιμές για τον δειγματικό μέσο μειώνεται όσο πιο ακραία είναι η τιμή, παράγοντας τις εξομαλυσμένες λεπτές ουρές που χαρακτηρίζουν μια καμπύλη κανονικής κατανομής.

Έλεγχος προόδου *9.5 Διευκρινίστε αν οι παρακάτω προτάσεις είναι σωστές ή λάθος. Το κεντρικό οριακό θεώρημα

(α) δηλώνει ότι, με επαρκώς μεγάλα μεγέθη δειγμάτων, το σχήμα του πληθυσμού είναι κανονικό.
(β) δηλώνει ότι, ανεξάρτητα από το μέγεθος του δείγματος, το σχήμα της κατανομής δειγματοληψίας του μέσου είναι κανονικό.
(γ) εξασφαλίζει ότι το σχήμα της κατανομής δειγματοληψίας του μέσου ισούται με το σχήμα του πληθυσμού.
(δ) ισχύει για το σχήμα της κατανομής δειγματοληψίας – όχι για το σχήμα του πληθυσμού και όχι για το σχήμα του δείγματος.

Απαντήσεις στη σελίδα 542.

9.7 Άλλες κατανομές δειγματοληψίας

Για τον μέσο

Υπάρχουν πολλές διαφορετικές κατανομές δειγματοληψίας μέσων. Μια νέα κατανομή δειγματοληψίας δημιουργείται λόγω μιας μετάβασης σε άλλον πληθυσμό. Επιπλέον, για οποιονδήποτε πληθυσμό, υπάρχουν πολλές διαφορετικές κατανομές δειγματοληψίας, όσα είναι και τα πιθανά μεγέθη δειγμάτων. Αν και αυτές οι κατανομές δειγματοληψίας έχουν τον ίδιο μέσο, η τιμή του τυπικού σφάλματος διαφέρει πάντα και εξαρτάται από το μέγεθος του δείγματος.

Για άλλα μέτρα

Υπάρχουν κατανομές δειγματοληψίας για μέτρα άλλα εκτός από έναν μέσο. Για παράδειγμα, υπάρχουν κατανομές δειγματοληψίας για διαμέσους, αναλογίες, τυπικές αποκλίσεις, διασπορές και συσχετίσεις, καθώς και για διαφορές μεταξύ ζευγών μέσων, ζευγών αναλογιών κ.ο.κ. Θα έχουμε την ευκαιρία να μελετήσουμε μερικές απ' αυτές τις κατανομές σε επόμενα κεφάλαια.

Περίληψη

Η έννοια μιας κατανομής δειγματοληψίας είναι η πιο σημαντική έννοια στην επαγωγική στατιστική. Η κατανομή δειγματοληψίας του μέσου ορίζεται ως η κατανομή πιθανοτήτων των μέσων για όλα τα πιθανά τυχαία δείγματα ενός δεδομένου μεγέθους κάποιου πληθυσμού.

Η στατιστική θεωρία τονίζει τρία σημαντικά χαρακτηριστικά της κατανομής δειγματοληψίας του μέσου:

- Ο μέσος της κατανομής δειγματοληψίας ισούται με τον μέσο του πληθυσμού.
- Η τυπική απόκλιση της κατανομής δειγματοληψίας, δηλαδή το τυπικό σφάλμα του μέσου, ισούται με την τυπική απόκλιση του πληθυσμού διά της τετραγωνικής ρίζας του μεγέθους του δείγματος. Μια σημαντική επιπλοκή αυτού του τύπου είναι ότι ένα μεγαλύτερο μέγεθος δείγματος μεταφράζεται σε κατανομή δειγματοληψίας με μικρότερη μεταβλητότητα, επιτρέποντας πιο ακριβείς γενικεύσεις από δείγματα σε πληθυσμούς. Το τυπικό σφάλμα του μέσου συνιστά ένα γενικό μέτρο της μέσης ποσότητας κατά την οποία οι δειγματικοί μέσοι αποκλίνουν από τον μέσο της κατανομής δειγματοληψίας ή από τον μέσο πληθυσμού.
- Σύμφωνα με το κεντρικό οριακό θεώρημα, ανεξάρτητα από το σχήμα του πληθυσμού, το σχήμα της κατανομής δειγματοληψίας προσομοιάζει με καμπύλη κανονικής κατανομής αν το μέγεθος του δείγματος είναι επαρκώς μεγάλο. Ανάλογα με τον βαθμό της μη κανονικότητας στον αρχικό πληθυσμό, ένα μέγεθος δείγματος μεταξύ 25 και 100 είναι επαρκώς μεγάλο.

Για οποιονδήποτε δειγματικό μέσο μπορούμε να θεωρήσουμε ότι προέρχεται από μια κατανομή δειγματοληψίας της οποίας (1) ο μέσος ισούται με τον μέσο πληθυσμού (ανεξαρτήτως της τιμής του), (2) το τυπικό σφάλμα ισούται με την τυπική απόκλιση του πληθυσμού διά της τετραγωνικής ρίζας του μεγέθους του δείγματος και (3) το σχήμα προσομοιάζει με καμπύλη κανονικής κατανομής (αν το μέγεθος του δείγματος ικανοποιεί τις προϋποθέσεις του κεντρικού οριακού θεωρήματος).

Σημαντικοί όροι

Κατανομή δειγματοληψίας του μέσου
Μέσος της κατανομής δειγματοληψίας του μέσου ($\mu_{\bar{X}}$)
Τυπικό σφάλμα του μέσου ($\sigma_{\bar{X}}$)
Κεντρικό οριακό θεώρημα

Κύριες εξισώσεις

ΚΑΤΑΝΟΜΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΤΟΥ ΜΕΣΟΥ

$$\mu_{\bar{X}} = \mu$$

ΤΥΠΙΚΟ ΣΦΑΛΜΑ

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

Ερωτήσεις επανάληψης

9.6 Ένα τυχαίο δείγμα τείνει να μην είναι ακριβές αντίγραφο του αρχικού του πληθυσμού. Αυτό το γεγονός επιφέρει αρκετές επιπλοκές. Διευκρινίστε ποιες από τις παρακάτω προτάσεις είναι σωστές και ποιες λάθος.

(α) Όλα τα πιθανά τυχαία δείγματα μπορούν να περιλαμβάνουν μερικά δείγματα που είναι ακριβή αντίγραφα του πληθυσμού, αλλά τα περισσότερα δείγματα δεν είναι ακριβή αντίγραφα.

(β) Μπορούμε να πάρουμε ένα πιο αντιπροσωπευτικό δείγμα με διαλογή (και όχι τυχαία επιλογή) παρατηρήσεων.

(γ) Στον βαθμό που διαστρεβλώνει τον αρχικό πληθυσμό, ένα τυχαίο δείγμα μπορεί να προκαλέσει εσφαλμένη γενίκευση.

(δ) Στην πράξη, ο μέσος ενός τυχαίου δείγματος αποτιμάται σε σχέση με τη μεταβλητότητα των μέσων για όλα τα πιθανά τυχαία δείγματα.

9.7 Δώστε τον ορισμό της κατανομής δειγματοληψίας του μέσου.

9.8 Αναφέρετε τρεις σημαντικές ιδιότητες της κατανομής δειγματοληψίας του μέσου.

9.9 Διευκρινίστε αν οι παρακάτω προτάσεις είναι σωστές ή λάθος. Αν πάρουμε ένα τυχαίο δείγμα 35 αντικειμένων από κάποιον πληθυσμό, η σχετιζόμενη κατανομή δειγματοληψίας του μέσου θα έχει τις εξής ιδιότητες:

(α) Το σχήμα θα προσομοιάζει με καμπύλη κανονικής κατανομής.

(β) Ο μέσος θα ισούται με τον δειγματικό μέσο.

(γ) Το σχήμα θα προσομοιάζει με το σχήμα του πληθυσμού.

(δ) Συγκριτικά με τη μεταβλητότητα του πληθυσμού, η μεταβλητότητα μειώνεται κατά συντελεστή ίσο με την τετραγωνική ρίζα του 35.

(ε) Ο μέσος θα ισούται με τον μέσο πληθυσμού.

(στ) Η μεταβλητότητα θα είναι ίση με τη μεταβλητότητα του πληθυσμού.

9.10 Διευκρινίστε αν οι παρακάτω προτάσεις είναι σωστές ή λάθος. Η κατανομή δειγματοληψίας του μέσου

(α) κατασκευάζεται πάντα από την αρχή, ακόμα κι όταν ο πληθυσμός είναι μεγάλος.

(β) συνιστά μια γέφυρα που βοηθά την εξαγωγή γενικεύσεων από ένα δείγμα σε έναν πληθυσμό.

(γ) είναι ίδια με τον δειγματικό μέσο.

(δ) αντιπροσωπεύει πάντα το σχήμα του υποκείμενου πληθυσμού.

(ε) έχει μέσο που συμπίπτει πάντα με τον μέσο πληθυσμού.

(στ) είναι ένας μηχανισμός που χρησιμοποιείται για τον προσδιορισμό της επίδρασης της μεταβλητότητας (δηλαδή τι μπορεί να συμβεί, απλώς από τύχη, όταν τα δείγματα είναι τυχαία).

(ζ) παραμένει αναλλοίωτη ακόμα και με μετατοπίσεις σε έναν νέο πληθυσμό ή ένα νέο μέγεθος δείγματος.

(η) παρέχει ένα φάσμα πιθανοτήτων βάσει του οποίου μπορούμε να αξιολογήσουμε τον παρατηρούμενο δειγματικό μέσο.

(θ) τείνει να προσεγγίζει τον μέσο πληθυσμού όταν υπάρχουν αυξήσεις στο μέγεθος του δείγματος.

9.11 Κάποιος ισχυρίζεται ότι, επειδή ο μέσος της κατανομής δειγματοληψίας ισούται με τον μέσο πληθυσμού, οποιοσδήποτε δειγματικός μέσος πρέπει επίσης να ισούται με τον μέσο πληθυσμού. Παρακαλώ, σχολιάστε την παραπάνω άποψη.

9.12 Δεδομένου ότι η τυπική απόκλιση πληθυσμού ισούται με 24, πόσο μεγάλο πρέπει να είναι το μέγεθος δείγματος, n, ώστε το τυπικό σφάλμα να ισούται με

(α) 8;

(β) 6;

(γ) 3;

(δ) 2;

9.13 Δεδομένου ότι το μέγεθος ενός δείγματος είναι 36, πόσο μεγάλη πρέπει να είναι η τυπική απόκλιση πληθυσμού προκειμένου το τυπικό σφάλμα να είναι

(α) 1;
(β) 2;
(γ) 5;
(δ) 100;

9.14 **(α)** Λαμβάνουμε ένα τυχαίο δείγμα μεγέθους 144 από τον τοπικό πληθυσμό μαθητών δημοτικού σχολείου. Κάθε παιδί εκτιμά τον αριθμό των ωρών που δαπανά κάθε εβδομάδα στην τηλεθέαση. Σ' αυτό το σημείο, τι μπορούμε να πούμε για την κατανομή δειγματοληψίας;

(β) Ας υποθέσουμε ότι μια τυπική απόκλιση, σ, 8 ωρών αποτυπώνει τις εκτιμήσεις τηλεθέασης για τον τοπικό πληθυσμό μαθητών δημοτικού. Σ' αυτό το σημείο, τι μπορούμε να πούμε για την κατανομή δειγματοληψίας;

(γ) Ας υποθέσουμε ότι ένας μέσος, μ, 21 ωρών αποτυπώνει τις εκτιμήσεις τηλεθέασης για τον τοπικό πληθυσμό μαθητών δημοτικού. Τι μπορούμε να πούμε τώρα για την κατανομή δειγματοληψίας;

(δ) Γενικά, οι δειγματικοί μέσοι στην κατανομή δειγματοληψίας θα πρέπει να αποκλίνουν, κατά μέσο όρο, περίπου ___ ώρες από τον μέσο της κατανομής δειγματοληψίας και από τον μέσο του πληθυσμού.

(ε) Περίπου το 95% των δειγματικών μέσων σ' αυτήν την κατανομή δειγματοληψίας θα πρέπει να είναι μεταξύ ___ και ___ ωρών.

ΠΑΡΑΡΤΗΜΑ IX

Κατανομή δειγματοληψίας με τη χρήση της R

Π.IX.1 Εισαγωγή

Το παρόν Κεφάλαιο 9 συμπληρώνει το Κεφάλαιο 5, το οποίο αναφερόταν στην κανονική κατανομή και στον υπολογισμό των z-τιμών με τη χρήση της R. Προχωρά κάποια βήματα παραπάνω, καθώς αναφέρεται στην κατανομή δειγματοληψίας.

Π.IX.2 Κατανοώντας τις βασικές έννοιες – Διάνυσμα δειγματικών μέσων όρων

Στα προηγούμενα παραδείγματα του Κεφαλαίου 5, αντλήσαμε 500 τυχαίους αριθμούς από μια κανονική κατανομή με μέση τιμή 100 και τυπική απόκλιση 10. Αυτό ονομάζεται «έλξη» δείγματος μεγέθους 500 από την κανονική κατανομή με μέση τιμή 100 και τυπική απόκλιση 10. Αυτό οδηγεί σε ένα δείγμα 500 τυχαίων αριθμών. Μια άμεση ερώτηση που μπορούμε να θέσουμε είναι: «Ποιος είναι ο μέσος όρος του δείγματός μας;» Χρησιμοποιώντας τις αντίστοιχες εντολές, υπολογίσαμε ότι ο μέσος του δείγματος ήταν 99,754. Φυσικά, αν πάρουμε ένα άλλο δείγμα 500 τυχαίων αριθμών από την κανονική κατανομή με μέσο όρο 100 και τυπική απόκλιση 10, θα έχουμε ένα νέο δείγμα που έχει διαφορετικό μέσο όρο.

```
> x=rnorm(500,mean=100,sd=10)
> mean(x)
[1] 99.69969
```

Στην περίπτωση αυτή, έχουμε ένα νέο δείγμα 500 τυχαία επιλεγμένων αριθμών και ο μέσος όρος αυτού του δείγματος παρέχει ένα διαφορετικό αποτέλεσμα, δηλαδή 99,69969. Η επόμενη ερώτηση που πρέπει να τεθεί είναι: «Τι συμβαίνει εάν το κάνουμε αυτό επανειλημμένα;» Στην επόμενη δραστηριότητα, θα δοκιμάσουμε επανειλημμένα να εξαγάγουμε δείγματα από την κανονική κατανομή. Κάθε δείγμα θα επιλέξει πέντε τυχαίους αριθμούς από την κανονική κατανομή με μέση τιμή 100 και τυπική απόκλιση 10. Στη συνέχεια θα βρούμε τον μέσο όρο των πέντε αριθμών στο δείγμα μας. Θα επαναλάβουμε αυτό το πείραμα 500 φορές, συλλέγοντας τα μέσα δείγματος σε ένα διάνυσμα xbar καθώς προχωρούμε τη διαδικασία μας.

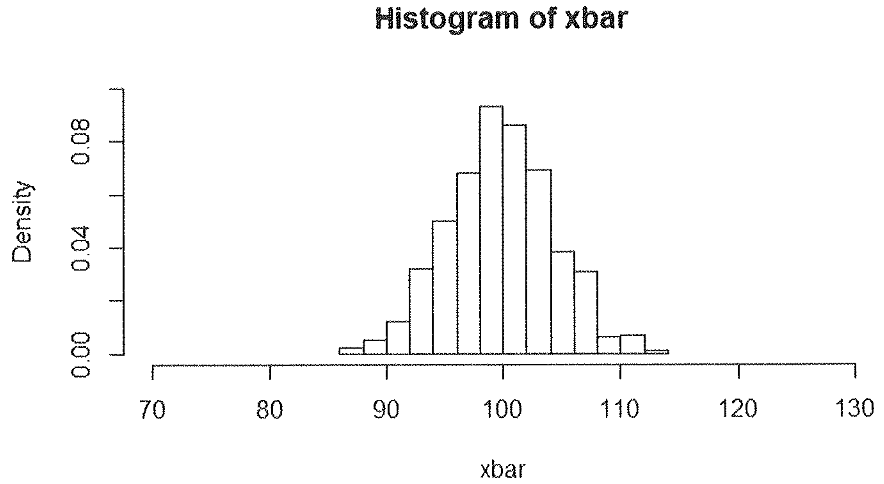
Ξεκινάμε δηλώνοντας τη μέση τιμή και την τυπική απόκλιση της κατανομής από την οποία θα αντλήσουμε τυχαίους αριθμούς. Στη συνέχεια δηλώνουμε το μέγεθος του δείγματος (ο αριθμός των τυχαίων αριθμών που έχουν επιλεγεί). Κάθε φορά που σχεδιάζουμε ένα δείγμα μεγέθους $n = 5$ από την κανονική κατανομή που έχει μέση $\mu = 100$ και τυπική απόκλιση $\sigma = 10$, χρειαζόμαστε κάπου να αποθηκεύσουμε τον μέσο του δείγματος. Επειδή σκοπεύουμε να συλλέξουμε τα μέσα των 500 δειγμάτων, δημιουργούμε ένα xbar διάνυσμα για να περιέχει αρχικά 500 μηδενικά.

Είναι εύκολο να σχεδιάσουμε ένα δείγμα μεγέθους $n = 5$ από την κανονική κατανομή που έχει μέση $\mu = 100$ και τυπική απόκλιση $\sigma = 10$. Απλώς αποδίδουμε την εντολή rnorm (n, mean = mu, sd = sigma). Για να υπολογίσουμε τον μέσο όρο αυτού του αποτελέσματος, προσθέτουμε απλώς τον μέσο προσαρμογής (rnorm (n, mean = mu, sd = sigma)). Το τελευταίο βήμα είναι να αποθηκεύσουμε αυτό το αποτέλεσμα στο διάνυσμα xbar. Στη συνέχεια πρέπει να επαναλάβουμε την ίδια διαδικασία επιπλέον 499 φορές για συνολικά 500 μέσα δειγματοληψίας. Αυτό απαιτεί τη χρήση ενός βρόχου. Η κατασκευή που χρησιμοποιείται από την R είναι παρόμοια με τα "for loops" που χρησιμοποιούνται σε πολλές γλώσσες προγραμματισμού [δείτε το i in for (i in 1:500)].

Ο δείκτης i αρχικά είναι ίσος με 1, τότε εκτελείται το «σώμα» του «βρόχου» (το τμήμα μεταξύ των αγκυλών). Στην επόμενη επανάληψη, το i είναι ίσο με 2 και το σώμα του βρόχου εκτελείται και πάλι. Ο βρόχος συνεχίζει με αυτόν τον τρόπο, αυξάνοντας τον δείκτη i κατά 1, οριστικά θέτοντας τον δείκτη i σε 500, πάνω στον οποίο εκτελείται το τελευταίο τμήμα του σώματος του βρόχου. Στη συνέχεια τερματίζεται. Στον ίδιο βρόχο έχουμε την εντολή xbar[i]=mean(rnorm(n,mean=mu,sd=sigma)). Αυτό αντλεί ένα δείγμα μεγέθους $n = 5$ από την κανονική κατανομή, υπολογίζει τη μέση τιμή του δείγματος και αποθηκεύει το αποτέλεσμα σε xbar [i], στην είσοδο x του xbar. Όταν το "for loop" ολοκληρώνει 500 επαναλήψεις, ο φορέας xbar περιέχει τους μέσους όρους των 500 δειγμάτων μεγέθους $n = 5$ που προέρχονται από την κανονική κατανομή που έχει μέση τιμή $\mu = 100$ και τυπική απόκλιση $\sigma = 10$.

Τώρα θα σχεδιάσουμε το ιστόγραμμά μας.

```
> mu=100; sigma=10
> n=5
> xbar=rep(0,500)
> for (i in 1:500) { xbar[i]=mean(rnorm(n,mean=mu,sd=sigma)) }
> hist(xbar,prob=TRUE,breaks=12,xlim=c(70,130),ylim=c(0,0.1))
```



ΕΙΚΟΝΑ Π.ΙΧ.1
Ιστόγραμμα των δειγματικών μέσων.

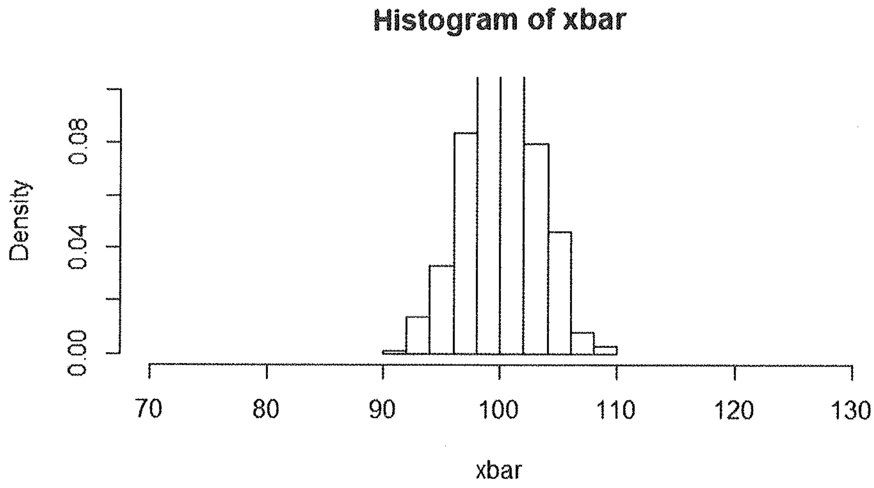
Υπάρχει ένας αριθμός σημαντικών παρατηρήσεων που πρέπει να γίνουν σχετικά με το ιστόγραμμα των δειγματικών μέσων στην Εικόνα Π.ΙΧ.1, ιδιαίτερα όταν συγκρίνεται με τα αντίστοιχα διαγράμματα του Κεφαλαίου 5. Το ιστόγραμμα της Εικόνας Π.ΙΧ.1 περιγράφει την κατανομή 500 μέσων δειγματοληψίας, το καθένα από τα οποία βρέθηκε επιλέγοντας $n = 5$ αριθμούς από την κανονική κατανομή με μέση $\mu = 100$ και τυπική απόκλιση $\sigma = 10$, υπολογίζοντας έπειτα τον μέσο όρο τους. Ο οριζόντιος άξονας τονίζει αυτό το γεγονός με την ετικέτα xbar. Είναι σημαντικό επίσης να σημειωθεί ότι η κατανομή του xbar στην Εικόνα Π.ΙΧ.1 φαίνεται «κανονική» ως προς το σχήμα. Αυτό συμβαίνει ακόμα κι αν το μέγεθος του δείγματος είναι σχετικά μικρό ($n = 5$). Φαίνεται ότι το «σημείο ισορροπίας» ή «κέντρο» της κατανομής είναι κοντά στο 100. Αυτό μπορεί να ελεγχθεί με την ακόλουθη εντολή και εκφράζει το μέσο των δειγματικών μέσων:

```
> mean(xbar)
[1] 99.77364
```

Π.ΙΧ.3 Αυξάνοντας το μέγεθος του δείγματος

Ας επαναλάβουμε το τελευταίο πείραμα, αλλά αυτή τη φορά ας τραβήξουμε δείγματα μεγέθους $n = 10$ από τον ίδιο «γονικό πληθυσμό», με την κανονική κατανομή να έχει μέση $\mu = 100$ και τυπική απόκλιση $\sigma = 10$.

```
> mu=100; sigma=10
> n=10
> xbar=rep(0,500)
> for (i in 1:500) {xbar[i]=mean(rnorm(n,mean=mu,sd=sigma))}
> hist(xbar,prob=TRUE,breaks=12,xlim=c(70,130),ylim=c(0,0.1))
```



ΕΙΚΟΝΑ Π.ΙΧ.2
Ιστόγραμμα των δειγματικών μέσων.

Όπως και παραπάνω, το ιστόγραμμα τονίζει την ύπαρξη μιας καμπύλης κανονικής κατανομής με μέσο των δειγματικών μέσων κοντά στο 100.

```
> mean(xbar)
[1] 100.0169
```

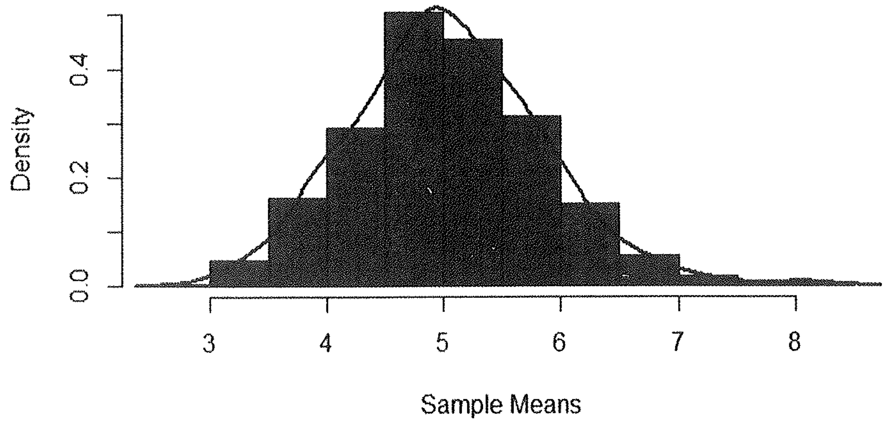
Με την αύξηση του μεγέθους των δειγμάτων μας ($n = 10$), το ιστόγραμμα των δειγμάτων σημαίνει «λιγότερο διαδεδομένο» ή «στενότερο», όπως φαίνεται σαφώς σε σύγκριση με την εξάπλωση των ιστογραμμάτων στις Εικόνες Π.ΙΧ.1 και Π.ΙΧ.2. Η συγκεκριμένη συμπεριφορά φαντάζει ως αρκετά λογική καθώς η αύξηση του δείγματος δημιουργεί ένα ιστόγραμμα πιο «συμπαγές». Θα αναμέναμε μια ακριβέστερη εκτίμηση του μέσου όρου του γονικού πληθυσμού αν ληφθεί ο μέσος όρος μεγαλύτερου μεγέθους δείγματος. Για παράδειγμα, έχετε περισσότερες πιθανότητες να εκτιμήσετε το μέσο ύψος του μαθητικού πληθυσμού σε ένα σχολείο εάν ζητήσετε από 10 μαθητές το ύψος και τον μέσο όρο τους από ό,τι εάν ζητήσετε από πέντε μόνο μαθητές το ύψος τους και τον μέσο όρο.

Π.ΙΧ.4 Αυξάνοντας αρκετά το μέγεθος του δείγματος

Στην περίπτωση αυτή προχωρούμε με τη χρήση ενός αρκετά μεγάλου δείγματος.

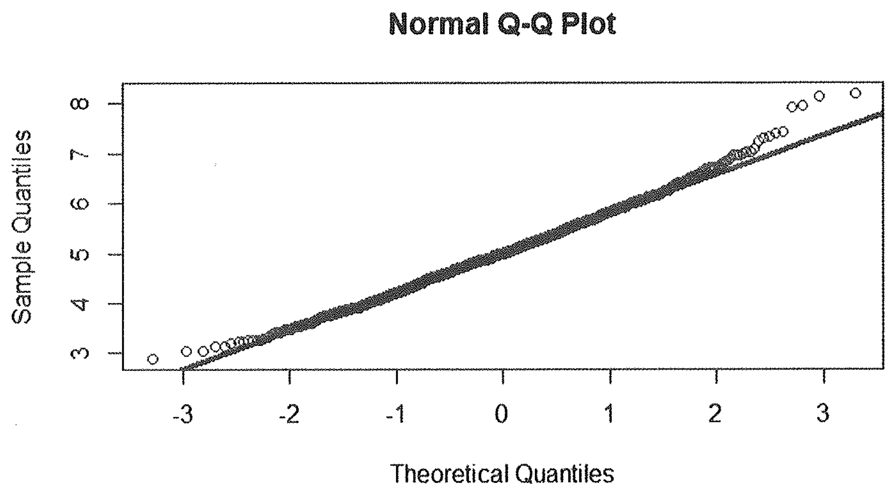
```
> sample_means = rep(NA, 1000)
> for(i in 1:1000){
+ sample_means[i] = mean(rexp(40,0.2))
+ }
> mean(sample_means)
[1] 5.029522
> ((1/0.2)^2)/40
[1] 0.625
> var(sample_means)
[1] 0.6580034
```

Υπολογίζουμε τη μέση τιμή και τη διασπορά των δειγματικών μέσων και αποτυπώνουμε σε ένα ιστόγραμμα τα αποτελέσματά μας.



ΕΙΚΟΝΑ Π.ΙΧ.3
Ιστόγραμμα των δειγματικών μέσων.

Τέλος, αποδίδουμε με ένα διάγραμμα QQ-plot την κανονικότητα του δείγματος που εξετάσαμε.



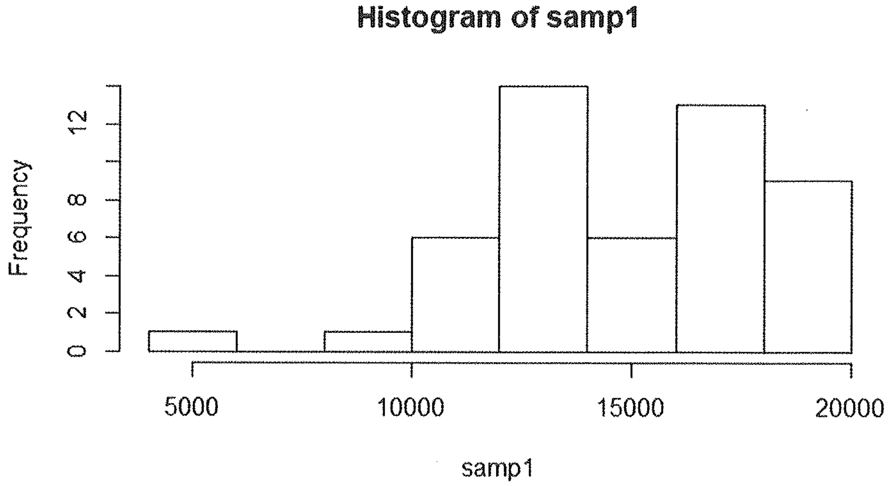
ΕΙΚΟΝΑ Π.ΙΧ.4
Διάγραμμα QQ-plot των δειγματικών μέσων.

Π.ΙΧ.5 Επιστροφή στα δεδομένα μας

Χρησιμοποιούμε για μία ακόμα φορά τη μεταβλητή των βαθμολογιών εισαγωγής για να δημιουργήσουμε ένα αρκετά μεγάλο δείγμα το οποίο θα εξετάσουμε όπως παραπάνω. Τα αποτελέσματα εμφανίζονται παρακάτω:

```
> samp1 <- sample(score, 50)
> summary(samp1)
Min. 1st Qu. Median Mean 3rd Qu. Max.
5000 13121 15050 15103 17675 19600
> hist(samp1)
> mean(samp1)
[1] 15103
> sd(samp1)
[1] 3020.719
```

Ο μέσος και η τυπική απόκλιση υπολογίζονται σε 15,103 και 3.020,719 και προφανώς διαφέρουν από τις αρχικές τιμές που έχουμε υπολογίσει. Όπως μπορούμε να παρατηρήσουμε, και το αντίστοιχο ιστόγραμμα εμφανίζεται διαφορετικό.



ΕΙΚΟΝΑ Π.ΙΧ.5
Ιστόγραμμα των βαθμολογιών εισαγωγής.

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)

ΚΕΦΑΛΑΙΟ

10 Εισαγωγή στον στατιστικό έλεγχο υποθέσεων: Ο έλεγχος z

Πρόλογος

Αυτό το κεφάλαιο περιγράφει τον πρώτο από μια σειρά στατιστικών ελέγχων υποθέσεων. Θα ήταν πολύ χρήσιμο για το υπόλοιπο βιβλίο να μάθετε τη σημασία ειδικών όρων για τους στατιστικούς ελέγχους υποθέσεων. Ωστόσο, δεν χρειάζεται να ασχοληθείτε τόσο πολύ με την ορολογία ή τους μηχανισμούς υπολογισμών ώστε να χάσετε την επαφή σας με τον ουσιώδη ρόλο της κατανομής δειγματοληψίας –το μοντέλο όλων όσων θα μπορούσαν να συμβούν μόνο από τύχη– σε οποιονδήποτε στατιστικό έλεγχο υποθέσεων.

Χρησιμοποιώντας την κατανομή δειγματοληψίας ως πλαίσιο αναφοράς, το παρατηρούμενο αποτέλεσμα χαρακτηρίζεται ως κοινό ή ως σπάνιο αποτέλεσμα. Ένα κοινό αποτέλεσμα αποδίδεται εύκολα στην τύχη και, επομένως, διατηρείται η υπόθεση ότι δεν συμβαίνει κάτι ιδιαίτερο – η μηδενική υπόθεση. Από την άλλη πλευρά, ένα σπάνιο αποτέλεσμα δεν αποδίδεται άμεσα στην τύχη και, επομένως, η μηδενική υπόθεση απορρίπτεται (συνήθως προς τέρψιν του ερευνητή).

10.1 Στατιστικός έλεγχος υπόθεσης για βαθμολογίες σε εξετάσεις SAT

Στο προηγούμενο κεφάλαιο, αναλάβαμε τον έλεγχο της υπόθεσης ότι ο μέσος βαθμός στις εξετάσεις μαθηματικών του SAT για όλους τους ντόπιους πρωτοετείς ισούται με τον εθνικό μέσο όρο 500. Τώρα, δεδομένου ενός μέσου βαθμού σε εξετάσεις μαθηματικών 533 για ένα τυχαίο δείγμα 100 πρωτοετών, θα ελέγξουμε την υπόθεση σύμφωνα με την οποία, σε σχέση με τον εθνικό μέσο όρο, δεν συμβαίνει κάτι ιδιαίτερο στον τοπικό πληθυσμό. Στον βαθμό που ένας ερευνητής υποψιάζεται συνήθως το αντίθετο –δηλαδή ότι συμβαίνει κάτι ιδιαίτερο στον τοπικό πληθυσμό– ελπίζει να απορρίψει την υπόθεση ότι δεν συμβαίνει κάτι ιδιαίτερο, κάτι που εφεξής θα αναφέρουμε ως *μηδενική υπόθεση* και θα ορίσουμε πιο επίσημα σε επόμενη ενότητα.

Υποθετική κατανομή δειγματοληψίας

Αν η μηδενική υπόθεση είναι αληθής, τότε οι δειγματικοί μέσοι της κατανομής –δηλαδή η κατανομή δειγματοληψίας του μέσου για όλα τα πιθανά τυχαία δείγματα, όπου το καθένα έχει μέγεθος 100, από τον πληθυσμό των ντόπιων πρωτοετών– θα επικεντρώνονται γύρω από τον εθνικό μέσο όρο 500. (Ο μέσος της κατανομής δειγματοληψίας ισούται πάντα με τον μέσο του πληθυσμού.) Στο Σχήμα 10.1, αυτή η κατανομή δειγματοληψίας αναφέρεται ως *υποθετική* κατανομή δειγματοληψίας, επειδή ο μέσος της ισούται με 500, τον υποθετικό μέσο βαθμό για τον πληθυσμό ντόπιων πρωτοετών.

Προεξοφλώντας τον βασικό ρόλο της υποθετικής κατανομής δειγματοληψίας στον στατιστικό έλεγχο υπόθεσης, θα εστιάσουμε σε δύο ή περισσότερες ιδιότητες αυτής της κατανομής:

1. Στο Σχήμα 10.1 εμφανίζονται κάθετες γραμμές σε διαστήματα μεγέθους 11, και από τις δύο πλευρές του υποθετικού μέσου του πληθυσμού, 500. Αυτά τα διαστήματα αντιπροσωπεύουν το μέγεθος του τυπικού σφάλματος του μέσου, $\sigma_{\overline{X}}$. Για να επαληθεύσετε αυτό το γεγονός, το οποίο παρουσιάσαμε αρχικά στο Κεφάλαιο 9, αντικαταστήστε με 110 την τυπική απόκλιση πληθυσμού σ και με 100 το μέγεθος δείγματος n στον Τύπο 9.2 για να πάρετε

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{110}{\sqrt{100}} = \frac{110}{10} = 11$$

2. Παρατηρήστε το σχήμα της υποθετικής κατανομής δειγματοληψίας στο Σχήμα 10.1, το οποίο προσομοιάζει με καμπύλη κανονικής κατανομής, επειδή το μέγεθος δείγματος 100 είναι αρκετά μεγάλο και ικανοποιεί τις προϋποθέσεις του κεντρικού οριακού θεωρήματος. Σταδιακά και με τη βοήθεια των πινάκων καμπύλης κανονικής κατανομής, θα μπορούμε να κατασκευάσουμε όρια για κοινά και σπάνια αποτελέσματα σύμφωνα με τους όρους της μηδενικής υπόθεσης.

Θεωρούμε *διστακτικά* πως η μηδενική υπόθεση ότι ο μέσος πληθυσμού για την τάξη πρωτοετών ισούται με 500 είναι αληθής. Ο έλεγχος γίνεται προκειμένου να διαπιστώσουμε αν ο παρατηρούμενος δειγματικός μέσος είναι κοινό αποτέλεσμα ή σπάνιο αποτέλεσμα στην υποθετική κατανομή δειγματοληψίας του Σχήματος 10.1.

Κοινά αποτελέσματα

Ένας παρατηρούμενος δειγματικός μέσος θεωρείται *κοινό αποτέλεσμα* αν η διαφορά μεταξύ της τιμής του και της τιμής του υποθετικού μέσου πληθυσμού είναι αρκετά μικρή και μπορεί να θεωρηθεί πιθανό αποτέλεσμα σύμφωνα με τους όρους της μηδενικής υπόθεσης.

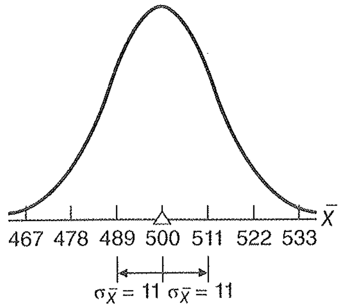
Δηλαδή ένας δειγματικός μέσος θεωρείται κοινό αποτέλεσμα αν δεν αποκλίνει πολύ από τον υποθετικό μέσο πληθυσμού, αλλά φαίνεται να προκύπτει από την πυκνή συγκέντρωση πιθανών δειγματικών μέσων στο κέντρο της κατανομής δειγματοληψίας. *Ένα κοινό αποτέλεσμα υποδηλώνει έλλειψη αποδείξεων ότι, αναφορικά με τη μηδενική υπόθεση, συμβαίνει κάτι ιδιαίτερο στον υποκείμενο πληθυσμό.* Επειδή δεν έχουμε εν προκειμένω σοβαρό λόγο να απορρίψουμε τη μηδενική υπόθεση, τη διατηρούμε.

Προσοχή:
Ο ένας παρατηρούμενος δειγματικός μέσος θεωρείται κοινό ή σπάνιο αποτέλεσμα;

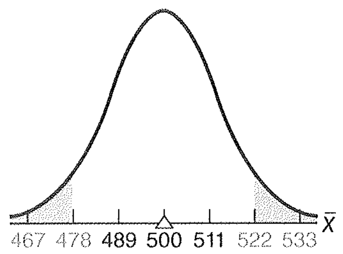
Σπάνια αποτελέσματα

Ένας παρατηρούμενος δειγματικός μέσος θεωρείται *σπάνιο αποτέλεσμα* αν η διαφορά μεταξύ της τιμής του και του υποθετικού μέσου πληθυσμού είναι πολύ μεγάλη για να θεωρηθεί πιθανό αποτέλεσμα σύμφωνα με τους όρους της μηδενικής υπόθεσης.

Δηλαδή ένας δειγματικός μέσος θεωρείται σπάνιο αποτέλεσμα αν αποκλίνει πάρα πολύ από τον υποθετικό μέσο και φαίνεται να προκύπτει από την «αραιή» συγκέντρωση πιθανών δειγματικών μέσων σε μία από τις δύο ουρές της κατανομής δειγματοληψίας. *Ένα σπάνιο αποτέλεσμα υποδηλώνει ότι,*



ΣΧΗΜΑ 10.1
Υποθετική κατανομή δειγματοληψίας του μέσου που επικεντρώνεται γύρω από έναν υποθετικό μέσο πληθυσμού 500.



ΣΧΗΜΑ 10.2

Ένα πιθανό σύνολο κοινών και σπάνιων αποτελεσμάτων (τιμές του $\overline{X}$).

αναφορικά με τη μηδενική υπόθεση, πιθανώς συμβαίνει κάτι ιδιαίτερο στον υποκείμενο πληθυσμό. Επειδή τώρα υπάρχουν λόγοι για να υποπτευθούμε ότι δεν ισχύει η μηδενική υπόθεση, την απορρίπτουμε.

Όρια για κοινά και σπάνια αποτελέσματα

Πάνω στην υποθετική κατανομή δειγματοληψίας του Σχήματος 10.2 τοποθετείται ένα πιθανό σύνολο ορίων για κοινά και σπάνια αποτελέσματα σε τιμές $\overline{X}$. (Οι τεχνικές για την κατασκευή αυτών των ορίων περιγράφονται στην Ενότητα 10.7.) Αν ο ένας παρατηρούμενος δειγματικός μέσος βρίσκεται μεταξύ 478 και 522, μπορεί να θεωρηθεί κοινό αποτέλεσμα (το οποίο αποδίδεται στη μεταβλητότητα) σύμφωνα με όσα ορίζει η μηδενική υπόθεση και η μηδενική υπόθεση θα διατηρηθεί. Αν όμως ο ένας παρατηρούμενος δειγματικός μέσος είναι μεγαλύτερος από 522 ή μικρότερος από 478, θα θεωρηθεί σπάνιο αποτέλεσμα (που δεν αποδίδεται άμεσα στη μεταβλητότητα) σύμφωνα με όσα ορίζει η μηδενική υπόθεση και η μηδενική υπόθεση θα απορριφθεί. Επειδή ο παρατηρούμενος δειγματικός μέσος 533 είναι μεγαλύτερος του 522, η μηδενική υπόθεση απορρίπτεται. Με βάση τον παρόντα έλεγχο, είναι απίθανο το δείγμα 100 πρωτοετών, με μέσο βαθμό σε εξετάσεις μαθηματικών 533, να προέρχεται από έναν πληθυσμό του οποίου ο μέσος ισούται με τον εθνικό μέσο όρο 500 και, επομένως, ο ερευνητής μπορεί να συμπεράνει ότι ο μέσος βαθμός σε εξετάσεις μαθηματικών για τον ντόπιο πληθυσμό πρωτοετών πιθανώς διαφέρει (είναι μεγαλύτερος) από τον εθνικό μέσο όρο.

10.2 Έλεγχος *z* για μέσο πληθυσμού

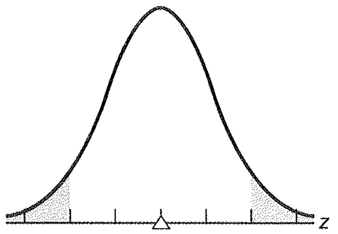
Για τον στατιστικό έλεγχο υποθέσεων με τους βαθμούς σε εξετάσεις μαθηματικών SAT, δεν συνηθίζουμε να βασίζουμε τον έλεγχο στην υποθετική κατανομή δειγματοληψίας του $\overline{X}$ που βλέπετε στο Σχήμα 10.2, αλλά στην αντίστοιχη τυποποιημένη, την υποθετική κατανομή δειγματοληψίας του *z* που βλέπετε στο Σχήμα 10.3. Τώρα το *z* αναπαριστά μια παραλλαγή στο οικείο τυποποιημένο αποτέλεσμα και εμφανίζει όλες τις ιδιότητες τυποποιημένων αποτελεσμάτων που περιγράφονται στο Κεφάλαιο 5. Επιπλέον, όπως η κατανομή δειγματοληψίας του $\overline{X}$, η **κατανομή δειγματοληψίας του *z*** *αναπαριστά την κατανομή των z-τιμών που θα παίρναμε αν υπολογιζόταν μια τιμή του z για κάθε δειγματικό μέσο για όλα τα πιθανά τυχαία δείγματα δεδομένου μεγέθους από κάποιον πληθυσμό.*

> ***Κατανομή δειγματοληψίας του z***
> *Η κατανομή των z-τιμών που θα παίρναμε αν υπολογιζόταν μια τιμή του z για κάθε δειγματικό μέσο για όλα τα πιθανά τυχαία δείγματα δεδομένου μεγέθους από κάποιον πληθυσμό.*

Η μετατροπή από $\overline{X}$ σε *z* δίνει κατανομή που προσομοιάζει με την καμπύλη τυποποιημένης κανονικής κατανομής του Πίνακα Α του Παραρτήματος Γ, επειδή, όπως δείχνει το Σχήμα 10.3, ο αρχικός υποθετικός μέσος πληθυσμού (500) εμφανίζεται ως *z*-τιμή 0 και το αρχικό τυπικό σφάλμα του μέσου (11) εμφανίζεται ως *z*-τιμή 1. Η μετατόπιση από το $\overline{X}$ στο *z* εξαλείφει τις αρχικές μονάδες μέτρησης και τυποποιεί τον στατιστικό έλεγχο υποθέσεων σε όλες τις καταστάσεις, χωρίς όμως να επηρεάζει τα αποτελέσματα του ελέγχου.

Υπενθύμιση: Μετατροπή μη επεξεργασμένου αποτελέσματος σε z-τιμή

Για να μετατρέψετε ένα μη επεξεργασμένο αποτέλεσμα σε τυποποιημένο (περιγράφεται επίσης στο Κεφάλαιο 5), εκφράζετε το μη επεξεργασμένο αποτέλεσμα ως απόσταση από τον μέσο του (αφαιρώντας τον μέσο από το



ΣΧΗΜΑ 10.3

Κοινά και σπάνια αποτελέσματα (τιμές του z).

μη επεξεργασμένο αποτέλεσμα) και έπειτα χωρίζετε αυτήν την απόσταση σε μονάδες τυπικής απόκλισης (διαιρώντας διά της τυπικής απόκλισης). Εκφράζοντας αυτόν τον ορισμό ως λεκτικό τύπο, έχουμε

$$\textit{Τυποποιημένο αποτέλεσμα} = \frac{\textit{μη επεξεργασμένο αποτέλεσμα} - \textit{μέσος}}{\textit{τυπική απόκλιση}}$$

στο οποίο βέβαια το τυποποιημένο αποτέλεσμα υποδεικνύει την απόκλιση του μη επεξεργασμένου αποτελέσματος σε μονάδες τυπικής απόκλισης, πάνω ή κάτω από τον μέσο.

Μετατροπή δειγματικού μέσου σε z

Το z στην παρούσα περίπτωση εμφανίζεται ως μικρή παραλλαγή αυτού του λεκτικού τύπου: Αντικαθιστάτε το *μη επεξεργασμένο αποτέλεσμα* με τον έναν παρατηρούμενο δειγματικό μέσο $\overline{X}$, αντικαθιστάτε τον μέσο με τον μέσο της κατανομής δειγματοληψίας, δηλαδή τον υποθετικό μέσο πληθυσμού $\mu_{υποθ}$ και αντικαθιστάτε την *τυπική απόκλιση* με το τυπικό σφάλμα του μέσου $\sigma_{\overline{X}}$. Δηλαδή

Λόγος z για έναν μόνο μέσο πληθυσμού

$$z = \frac{\overline{X} - \mu_{υποθ}}{\sigma_{\overline{X}}} \qquad (10.1)$$

όπου το z δείχνει την απόκλιση του παρατηρούμενου δειγματικού μέσου σε μονάδες τυπικού σφάλματος, πάνω ή κάτω από τον υποθετικό μέσο πληθυσμού.

Για να ελέγξουμε την υπόθεση με τους βαθμούς σε εξετάσεις SAT, πρέπει να βρούμε την τιμή του z από τον Τύπο 10.1. Δεδομένου ενός δειγματικού μέσου 533, ενός υποθετικού μέσου πληθυσμού 500 και ενός τυπικού σφάλματος 11, βρίσκουμε

$$z = \frac{533 - 500}{11} = \frac{33}{11} = 3$$

Το παρατηρούμενο z, 3, είναι μεγαλύτερο από το 1,96 που προσδιορίζεται στην υποθετική κατανομή δειγματοληψίας του Σχήματος 10.3. Έτσι, το παρατηρούμενο z θεωρείται σπάνιο αποτέλεσμα σύμφωνα με όσα ορίζει η μηδενική υπόθεση και γι' αυτόν τον λόγο η μηδενική υπόθεση απορρίπτεται. Τα αποτελέσματα αυτού του ελέγχου με το z είναι ίδια με εκείνα για τον αρχικό στατιστικό έλεγχο της υπόθεσης με το $\overline{X}$.

Έλεγχος z για μέσο πληθυσμού
Ένας στατιστικός έλεγχος υποθέσεων που εκτιμά πόσο αποκλίνει ο παρατηρούμενος δειγματικός μέσος σε μονάδες τυπικού σφάλματος από τον υποθετικό μέσο πληθυσμού.

Υποθέσεις ελέγχου z

Όταν ένας στατιστικός έλεγχος υποθέσεων εκτιμά πόσο αποκλίνει ο παρατηρούμενος δειγματικός μέσος σε μονάδες τυπικού σφάλματος από τον υποθετικό μέσο πληθυσμού, όπως σ' αυτό το παράδειγμα, *έχουμε έναν έλεγχο z* ή, με μεγαλύτερη ακρίβεια, έναν **έλεγχο z για τον μέσο πληθυσμού**. Αυτός ο έλεγχος z είναι ακριβής μόνο όταν (1) ο πληθυσμός κατανέμεται κανονικά ή το μέγεθος του δείγματος είναι αρκετά μεγάλο για να ικανοποιήσει τις προϋποθέσεις του κεντρικού οριακού θεωρήματος και (2) η τυπική απόκλιση πληθυσμού είναι γνωστή. Στο προκείμενο παράδειγμα είναι κατάλληλος ο έλεγχος z επειδή το μέγεθος δείγματος 100 είναι αρκετά μεγάλο για να ικανοποιείται το κεντρικό οριακό θεώρημα και γνωρίζουμε ότι η τυπική απόκλιση πληθυσμού είναι 110.

Έλεγχος προόδου *10.1 Υπολογίστε την τιμή του ελέγχου z για τις παρακάτω περιπτώσεις:

(α) $\overline{X} = 566$, $\sigma = 30$, $n = 36$, $\mu_{\text{υποθ}} = 560$
(β) $\overline{X} = 24$, $\sigma = 4$, $n = 64$, $\mu_{\text{υποθ}} = 25$
(γ) $\overline{X} = 82$, $\sigma = 14$, $n = 49$, $\mu_{\text{υποθ}} = 75$
(δ) $\overline{X} = 136$, $\sigma = 15$, $n = 25$, $\mu_{\text{υποθ}} = 146$

Απαντήσεις στη σελίδα 542.

10.3 Βήμα προς βήμα διαδικασία

Αφού είχαμε μια πρώτη επαφή με μερικά από τα πιο σημαντικά χαρακτηριστικά του στατιστικού ελέγχου υποθέσεων, θα εξετάσουμε λεπτομερώς τον έλεγχο για τις βαθμολογίες σε εξετάσεις SAT. Η διαδικασία ελέγχου γίνεται μια αναλυτική περιγραφή που ξεκινά με μια σύντομη δήλωση του προβλήματος που ενέπνευσε τον έλεγχο και καταλήγει σε μια ερμηνεία των αποτελεσμάτων του ελέγχου. Το παρακάτω πλαίσιο συνοψίζει τη διαδικασία βήμα προς βήμα για τον τρέχοντα στατιστικό έλεγχο υποθέσεων. Όποτε τη χρειαστούμε, θα χρησιμοποιήσουμε αυτήν τη μορφή παρακάτω στο βιβλίο.

10.4 Παρουσίαση προβλήματος έρευνας

Η διατύπωση ενός προβλήματος έρευνας εκφράζει πολλές φορές την πιο κρίσιμη και συναρπαστική φάση μιας έρευνας. Ένα σημάδι μάλιστα που μας αποκαλύπτει έναν ικανό ερευνητή είναι όταν μπορεί να εστιάσει σε ένα

ΠΕΡΙΛΗΨΗ ΣΤΑΤΙΣΤΙΚΟΥ ΕΛΕΓΧΟΥ ΥΠΟΘΕΣΕΩΝ: ΕΛΕΓΧΟΣ *z* ΓΙΑ ΕΝΑΝ ΜΕΣΟ ΠΛΗΘΥΣΜΟΥ (ΒΑΘΜΟΛΟΓΙΕΣ ΣΕ ΕΞΕΤΑΣΕΙΣ SAT)

Πρόβλημα έρευνας
Ο μέσος βαθμός σε εξετάσεις μαθηματικών SAT για όλους τους ντόπιους πρωτοετείς διαφέρει από τον εθνικό μέσο όρο 500;

Στατιστικές υποθέσεις

$$H_0: \mu = 500$$
$$H_1: \mu \neq 500$$

Κανόνας απόφασης
Απόρριψη της H_0 στο επίπεδο σημαντικότητας 0,05 αν $z \geq 1{,}96$ ή αν $z \leq -1{,}96$.

Υπολογισμοί
Δεδομένου ότι

$$\overline{X} = 533,\quad \mu_{\text{υποθ}} = 500,\quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{110}{\sqrt{100}} = 11$$

$$z = \frac{533 - 500}{11} = 3$$

Απόφαση
Απόρριψη της H_0 στο επίπεδο σημαντικότητας 0,05 επειδή το $z = 3$ είναι μεγαλύτερο από το 1,96.

Ερμηνεία
Ο μέσος βαθμός σε εξετάσεις μαθηματικών SAT για όλους τους ντόπιους πρωτοετείς δεν ισούται –είναι μεγαλύτερος– με τον εθνικό μέσο όρο 500.

σημαντικό πρόβλημα έρευνας για το οποίο μπορεί να υπάρχουν απαντήσεις. Τα παιδιά από διαλυμένες οικογένειες έχουν χαμηλότερους βαθμούς σε εξετάσεις προσωπικής προσαρμογής; Τα βίαια κινούμενα σχέδια στην τηλεόραση υποδαυλίζουν πιο επιθετική συμπεριφορά σε μαθητές προσχολικής ηλικίας; Η διανομή μερισμάτων από τα κέρδη αυξάνει την παραγωγικότητα των υπαλλήλων; Εξαιτίας της έμφασης που δίνουμε στον στατιστικό έλεγχο υποθέσεων, σ' αυτό το βιβλίο παρουσιάζονται προβλήματα έρευνας ως ολοκληρωμένα προϊόντα, συνήθως στις πρώτες προτάσεις ενός νέου παραδείγματος.

10.5 Μηδενική υπόθεση (H_0)

Αφού περιγραφεί το πρόβλημα, πρέπει να μεταφραστεί σε στατιστική υπόθεση ως προς κάποιο χαρακτηριστικό του πληθυσμού. Η μηδενική υπόθεση συμβολίζεται ως H_0 και γίνεται το σημείο εστίασης για όλη τη διαδικασία του ελέγχου (ακόμα κι αν συνήθως ευελπιστούμε ότι θα απορριφθεί). Στον έλεγχο με τις βαθμολογίες σε εξετάσεις SAT, η μηδενική υπόθεση υποστηρίζει ότι, σε σχέση με τον εθνικό μέσο όρο 500, δεν συμβαίνει κάτι ιδιαίτερο στη μέση βαθμολογία για τον ντόπιο πληθυσμό πρωτοετών. Μια αντίστοιχη πρόταση, με σύμβολα, θα ήταν:

$$H_0 : \mu = 500$$

Μηδενική υπόθεση (H_0)
Μια στατιστική υπόθεση που συνήθως υποστηρίζει ότι δεν συμβαίνει κάτι ιδιαίτερο σε σχέση με κάποιο χαρακτηριστικό του υποκείμενου πληθυσμού.

όπου το H_0 αναπαριστά τη μηδενική υπόθεση και μ είναι ο μέσος πληθυσμού για τους ντόπιους πρωτοετείς.

Γενικά, η **μηδενική υπόθεση (H_0)** *είναι μια στατιστική υπόθεση που συνήθως υποστηρίζει ότι δεν συμβαίνει κάτι ιδιαίτερο σε σχέση με κάποιο χαρακτηριστικό του υποκείμενου πληθυσμού.* Επειδή η διαδικασία του στατιστικού ελέγχου μιας υπόθεσης προϋποθέτει ότι η υποθετική κατανομή δειγματοληψίας του μέσου θα επικεντρώνεται γύρω από έναν μόνο αριθμό (500), η μηδενική υπόθεση ισούται μ' αυτόν τον αριθμό (H_0: $\mu = 500$). Επιπλέον, η μηδενική υπόθεση κάνει πάντα μια ακριβή δήλωση για ένα χαρακτηριστικό του πληθυσμού και ποτέ για ένα δείγμα. Έχουμε πει ότι σκοπός των στατιστικών ελέγχων υποθέσεων είναι να διαπιστωθεί αν ένα συγκεκριμένο αποτέλεσμα, όπως ένας παρατηρούμενος δειγματικός μέσος, θα μπορούσε να έχει προκύψει εύλογα από έναν πληθυσμό με το υποθετικό χαρακτηριστικό.

Εύρεση του μοναδικού αριθμού για την H_0

Ο ένας αριθμός που χρησιμοποιείται για τη μηδενική υπόθεση H_0 διαφέρει μεταξύ προβλημάτων. Ακόμα και για ένα δεδομένο πρόβλημα, αυτός ο αριθμός θα μπορούσε να προέρχεται από διάφορες πηγές. Για παράδειγμα, θα μπορούσε να βασίζεται σε διαθέσιμες πληροφορίες για κάποιον σχετικό πληθυσμό πλην του πληθυσμού-στόχου, όπως στο συγκεκριμένο παράδειγμα όπου το 500 είναι η μέση βαθμολογία σε εξετάσεις μαθηματικών SAT για όλους τους φοιτητές για το προηγούμενο έτος. Θα μπορούσε επίσης να βασίζεται σε κάποιο υπάρχον πρότυπο ή θεωρία – για παράδειγμα, ότι ο μέσος βαθμός σε εξετάσεις μαθηματικών για τον τρέχοντα πληθυσμό ντόπιων πρωτοετών πρέπει να ισούται με 540 επειδή αυτός ο αριθμός είναι ο μέσος βαθμός που πέτυχαν όλοι οι ντόπιοι πρωτοετείς κατά τη διάρκεια των περασμένων ετών.

Αν, όπως συμβαίνει μερικές φορές, είναι αδύνατο να προσδιοριστεί μια χρήσιμη και υπαρκτή μηδενική υπόθεση, μην αναζητήσετε τη λύση σε αυθαίρετους αριθμούς, αλλά, αντίθετα, χρησιμοποιήστε κάποια άλλη, εντελώς διαφορετική τεχνική, που είναι γνωστή ως *εκτίμηση* και την οποία περιγράφουμε στο Κεφάλαιο 12.

10.6 Εναλλακτική υπόθεση (H_1)

Στο παράδειγμα που μας απασχολεί, η εναλλακτική υπόθεση ισχυρίζεται ότι, σε σχέση με τον εθνικό μέσο όρο 500, συμβαίνει κάτι ιδιαίτερο στον μέσο βαθμό σε εξετάσεις μαθηματικών για τον ντόπιο πληθυσμό πρωτοετών (επειδή ο μέσος για τον ντόπιο πληθυσμό δεν ισούται με τον εθνικό μέσο όρο 500). Μια αντίστοιχη πρόταση, με σύμβολα, θα ήταν:

$$H_1 : \mu \neq 500$$

όπου το H_1 αναπαριστά την εναλλακτική υπόθεση, μ είναι ο μέσος πληθυσμού για την τάξη των ντόπιων πρωτοετών και το $\neq$ σημαίνει «δεν είναι ίσο με».

Η **εναλλακτική υπόθεση (H_1)** *υποστηρίζει το αντίθετο της μηδενικής υπόθεσης.* Η απόφαση για τη διατήρηση της μηδενικής υπόθεσης υποδηλώνει έλλειψη υποστήριξης για την εναλλακτική υπόθεση, ενώ η απόφαση για απόρριψη της μηδενικής υπόθεσης συνεπάγεται υποστήριξη της εναλλακτικής υπόθεσης.

Εναλλακτική υπόθεση (H_1)
Το αντίθετο της μηδενικής υπόθεσης.

Όπως θα δούμε στο επόμενο κεφάλαιο, η εναλλακτική υπόθεση μπορεί να πάρει τρεις μορφές, ανάλογα με την οπτική του ερευνητή. Στη συγκεκριμένη μορφή, το H_1 καθορίζει ένα *εύρος* πιθανών τιμών για τον *μοναδικό* αριθμό (500) που εμφανίζεται στην H_0.

Ανεξάρτητα από τη μορφή του, το H_1 συνήθως αναγνωρίζεται με την **υπόθεση έρευνας**, *την ανεπίσημη υπόθεση ή προαίσθημα ότι υπονοώντας πως συμβαίνει κάτι ιδιαίτερο στον υποκείμενο πληθυσμό γεννιέται η έμπνευση για όλη την έρευνα.* Η φράση «κάτι ιδιαίτερο» μπορεί να είναι, όπως στο συγκεκριμένο παράδειγμα, μια απόκλιση από έναν εθνικό μέσο όρο, ή θα μπορούσε να είναι, όπως σε επόμενα κεφάλαια, μια απόκλιση από κάποια συνθήκη ελέγχου που παράγεται από μια νέα μέθοδο διδασκαλίας, μια δίαιτα ή ένα εργαστήριο αυτοβελτίωσης. Σε οποιοδήποτε γεγονός, αυτή η υπόθεση έρευνας –και σίγουρα όχι η μηδενική υπόθεση– είναι που παρέχει το κίνητρο πίσω από μια έρευνα.

Υπόθεση έρευνας
Η ανεπίσημη υπόθεση ή προαίσθημα που γεννά την έμπνευση για όλη την έρευνα, και συνήθως είναι η εναλλακτική υπόθεση.

Έλεγχος προόδου *10.2 Διευκρινίστε τι λάθος έχουν οι παρακάτω στατιστικές υποθέσεις:

(α) H_0: $\mu = 155$
H_1: $\mu \neq 160$

(β) H_0: $\overline{X} = 241$
H_1: $\overline{X} = 241$

Έλεγχος προόδου *10.3 Πρώτα με λέξεις και έπειτα με σύμβολα, προσδιορίστε τη μηδενική υπόθεση για καθεμία από τις παρακάτω περιπτώσεις. (Μην ανησυχείτε για την ακριβή μορφή της εναλλακτικής υπόθεσης σ' αυτό το σημείο.)

(α) Μια σχολική σύμβουλος θέλει να διαπιστώσει αν τα αγόρια της έκτης δημοτικού στην περιφέρειά της διαφέρουν, κατά μέσο όρο, από το εθνικό πρότυπο των 10,2 πουσάπς για τα αγόρια αυτής της ηλικίας.

(β) Μια ομάδα καταναλωτών διερευνά αν, κατά μέσο όρο, τα πραγματικά βάρη των συσκευασιών κιμά που πωλούνται από μια μεγάλη αλυσίδα σούπερ μάρκετ διαφέρουν από το καθορισμένο βάρος των 16 ουγγιών.

(γ) Ένας σύμβουλος γάμου θέλει να μάθει αν, κατά τη διάρκεια μιας τυπικής συνεδρίας επίλυσης διαφορών, οι πελάτες του διαφέρουν, κατά μέσο όρο, από τις 11 προφορικές διακοπές που παρατηρούνται σε τυπικά καλά ζευγάρια.

Απαντήσεις στη σελίδα 542.

10.7 Κανόνας απόφασης

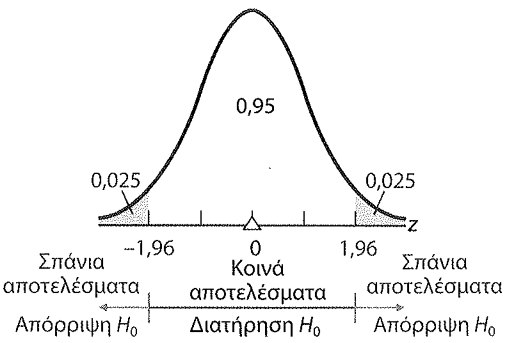
Ένας **κανόνας απόφασης** *καθορίζει ακριβώς πότε πρέπει να απορριφθεί η H_0* (επειδή το παρατηρούμενο z θεωρείται σπάνιο αποτέλεσμα). Υπάρχουν πολλοί πιθανοί κανόνες απόφασης, όπως θα δούμε στην Ενότητα 11.3. Ένας πολύ συνηθισμένος, τον οποίο έχουμε ήδη παρουσιάσει στο Σχήμα 10.3, καθορίζει ότι η H_0 θα πρέπει να *απορριφθεί* αν το παρατηρούμενο z ισούται ή είναι πιο θετικό από το 1,96 ή αν το παρατηρούμενο z ισούται ή είναι πιο αρνητικό από το –1,96. Αντίστροφα, η H_0 θα πρέπει να *διατηρηθεί* αν το παρατηρούμενη z βρίσκεται μεταξύ του –1,96 και του +1,96.

Κανόνας απόφασης
Καθορίζει ακριβώς πότε πρέπει να απορριφθεί η H_0 (επειδή το παρατηρούμενο z θεωρείται σπάνιο αποτέλεσμα).

Κρίσιμες z-τιμές

Το Σχήμα 10.4 δείχνει ότι οι z-τιμές μεταξύ ± 1,96 ορίζουν τα όρια για το μεσαίο 0,95 όλης της περιοχής (1,00) κάτω από την υποθετική κατανομή δειγματοληψίας για το z. Αυτές οι δύο z-τιμές, οι οποίες προκύπτουν από τον πίνακα καμπύλης κανονικής κατανομής, όπως μπορείτε να επαληθεύσετε ελέγχοντας τον Πίνακα Α στο Παράρτημα Γ, *διαχωρίζουν τα κοινά από τα σπάνια αποτελέσματα και ως εκ τούτου υπαγορεύουν αν η H_0 πρέπει να διατηρηθεί ή να απορριφθεί.* Εξαιτίας του ζωτικού ρόλου τους στην απόφαση για την H_0, αυτά τα αποτελέσματα αναφέρονται ως **κρίσιμες z-τιμές**.

Κρίσιμη z-τιμή
Μια z-τιμή που διαχωρίζει τα κοινά από τα σπάνια αποτελέσματα και ως εκ τούτου υπαγορεύει αν η H_0 πρέπει να διατηρηθεί ή να απορριφθεί.



ΣΧΗΜΑ 10.4
Αναλογίες εμβαδού περιοχής που σχετίζεται με κοινά και σπάνια αποτελέσματα (α = 0,05).

Επίπεδο σημαντικότητας (α)

Το Σχήμα 10.4 δείχνει επίσης την αναλογία (0,025 + 0,025 = 0,05) όλης της περιοχής που προσδιορίζεται με σπάνια αποτελέσματα. Αυτή η αναλογία, η οποία αναφέρεται συχνά ως το επίπεδο σημαντικότητας του στατιστικού ελέγχου, συμβολίζεται με το ελληνικό γράμμα *α* (άλφα) και αναλύεται στο βιβλίο στην Ενότητα 11.4. Στο συγκεκριμένο παράδειγμα, το επίπεδο σημαντικότητας, *α*, ισούται με 0,05.

Επίπεδο σημαντικότητας (α)
Ο βαθμός σπανιότητας που απαιτείται για ένα παρατηρούμενο αποτέλεσμα προκειμένου να απορριφθεί η μηδενική υπόθεση (H_0).

Το **επίπεδο σημαντικότητας (*α*)** *δείχνει τον βαθμό σπανιότητας που απαιτείται για ένα παρατηρούμενο αποτέλεσμα προκειμένου να απορριφθεί η μηδενική υπόθεση* (H_0). Για παράδειγμα, το επίπεδο σημαντικότητας 0,05 δείχνει ότι η H_0 πρέπει να απορριφθεί αν το παρατηρούμενο *z* μπορεί να συμβεί κατά τύχη με πιθανότητα μόλις 0,05 (μία πιθανότητα στις 20) *ή ακόμα μικρότερη*.

10.8 Υπολογισμοί

Μπορούμε να χρησιμοποιήσουμε πληροφορίες από το δείγμα προκειμένου να υπολογίσουμε μια τιμή για το *z*. Όπως είπαμε νωρίτερα, χρησιμοποιούμε τον Τύπο 10.1 για να μετατρέψουμε τον παρατηρούμενο δειγματικό μέσο 533 σε *z* 3.

10.9 Απόφαση

Διατηρείται ή απορρίπτεται η H_0 ανάλογα με τη θέση της παρατηρούμενης τιμής του *z* σε σχέση με τις κρίσιμες *z*-τιμές που καθορίζονται στον κανόνα απόφασης. Σύμφωνα με τον τρέχοντα κανόνα, η H_0 θα πρέπει να απορριφθεί στο επίπεδο σημαντικότητας 0,05 επειδή το παρατηρούμενο *z*, 3, είναι μεγαλύτερο από το κρίσιμο *z* 1,96 και, επομένως, θεωρείται σπάνιο αποτέλεσμα, δηλαδή ένα απίθανο αποτέλεσμα από έναν πληθυσμό που επικεντρώνεται γύρω από τη μηδενική υπόθεση.

Διατήρηση ή απόρριψη της H_0;

Αν δυσκολεύεστε να καταλάβετε αν πρέπει να διατηρήσετε ή να απορρίψετε την H_0, θυμηθείτε τη λογική πίσω από τον στατιστικό έλεγχο υποθέσεων. Απορρίπτετε την H_0 μόνο αν η παρατηρούμενη τιμή του *z* θεωρείται σπάνιο αποτέλεσμα επειδή αποκλίνει πάρα πολύ από τις ουρές της κατανομής δειγματοληψίας. Επομένως, απορρίπτετε την H_0 μόνο αν η παρατηρούμενη τιμή του *z* ισούται ή είναι μεγαλύτερη από το ανώτερο κρίσιμο *z* (1,96) ή αν ισούται ή είναι μικρότερη από το κατώτερο κρίσιμο *z* (−1,96). Πριν αποφασίσετε, θα ήταν χρήσιμο να σχεδιάσετε την υποθετική κατανομή δειγματοληψίας, μαζί με τις κρίσιμες *z*-τιμές και τις γραμμοσκιασμένες περιοχές απόρριψης, και έπειτα να χρησιμοποιήσετε κάποιο σημάδι, όπως ένα βέλος (↑), το οποίο θα ορίζει τη θέση της παρατηρούμενης τιμής του *z* (3) κατά μήκος της κλίμακας *z*. Αν αυτό το σημάδι βρίσκεται στη σκιασμένη περιοχή απόρριψης –ή μακριά απ' αυτή, όπως στο Σχήμα 10.4–, τότε η H_0 θα πρέπει να απορριφθεί.

Έλεγχος προόδου *10.4 Για τις παρακάτω περιπτώσεις, υποδείξτε αν η H_0 πρέπει να διατηρηθεί ή να απορριφθεί και δικαιολογήστε την απάντησή σας καθορίζοντας την ακριβή σχέση μεταξύ του παρατηρούμενου αποτελέσματος και της κρίσιμης z-τιμής. Θα πρέπει η H_0 να διατηρηθεί ή να απορριφθεί, δεδομένου ενός στατιστικού ελέγχου με κρίσιμες z-τιμές $\pm$ 1,96 και

(α) $z = 1{,}74$ **(β)** $z = 0{,}13$ **(γ)** $z = -2{,}51$

Απαντήσεις στη σελίδα 542.

10.10 Ερμηνεία

Τέλος, θα ερμηνεύσουμε την απόφαση σε σχέση με το αρχικό πρόβλημα έρευνας. Στο συγκεκριμένο παράδειγμα, μπορούμε να συμπεράνουμε ότι, επειδή η μηδενική υπόθεση απορρίφθηκε, ο μέσος βαθμός σε εξετάσεις μαθηματικών SAT για την τοπική τάξη πρωτοετών πιθανώς θα διαφέρει από τον εθνικό μέσο όρο 500.

Αν και δεν απορρέει απαραίτητα απ' αυτόν τον έλεγχο, μπορούμε να εξαγάγουμε ένα πιο συγκεκριμένο συμπέρασμα. Επειδή ο δειγματικός μέσος 533 (ή το αντίστοιχο z 3) εμπίπτει στην *ανώτερη* περιοχή απόρριψης της υποθετικής κατανομής δειγματοληψίας, μπορούμε να συμπεράνουμε ότι ο μέσος βαθμός σε εξετάσεις μαθηματικών SAT για τον πληθυσμό όλων των ντόπιων πρωτοετών πιθανώς *είναι μεγαλύτερος* από τον εθνικό μέσο όρο 500. Με την ίδια λογική, αν ο παρατηρούμενος δειγματικός μέσος ή το αντίστοιχο z βρισκόταν στην *κατώτερη* περιοχή απόρριψης της υποθετικής κατανομής δειγματοληψίας, θα μπορούσαμε να συμπεράνουμε ότι ο μέσος πληθυσμού για όλους τους ντόπιους πρωτοετείς πιθανώς είναι *μικρότερος* από τον εθνικό μέσο όρο.

Αν ο παρατηρούμενος δειγματικός μέσος ή το αντίστοιχο z βρισκόταν στην περιοχή διατήρησης της υποθετικής κατανομής δειγματοληψίας, θα είχαμε συμπεράνει (όχι με την ίδια βεβαιότητα, όπως θα δούμε στην Ενότητα 11.2) ότι δεν υπάρχει καμία απόδειξη πως ο μέσος πληθυσμού για όλους τους ντόπιους πρωτοετείς διαφέρει από τον εθνικό μέσο όρο 500.

Έλεγχος προόδου *10.5 Σύμφωνα με την Αμερικανική Ένωση Ψυχολόγων, τα μέλη της που έχουν διδακτορικό τίτλο και εργάζονται με πλήρες ωράριο έχουν εισόδημα, κατά μέσο όρο, \$82.500 κατ' έτος, με τυπική απόκλιση \$6.000. Ένας ερευνητής θέλει να διαπιστώσει αν το ποσό των \$82.500 είναι επίσης ο μέσος μισθός για όλες τις γυναίκες μέλη της ένωσης με διδακτορικό τίτλο και πλήρες ωράριο. Οι μισθοί λαμβάνονται από ένα τυχαίο δείγμα 100 γυναικών απ' αυτόν τον πληθυσμό και ο μέσος μισθός ισούται με \$80.100.

(α) Κάποιος ισχυρίζεται ότι η παρατηρούμενη διαφορά μεταξύ \$80.100 και \$82.500 είναι αρκετά μεγάλη, τόσο που μπορεί να υποστηρίξει το συμπέρασμα ότι οι γυναίκες ψυχολόγοι της ένωσης κερδίζουν λιγότερα από τους άντρες. Εξηγήστε γιατί είναι σημαντικό να πραγματοποιηθεί στατιστικός έλεγχος υποθέσεων.

(β) Ο ερευνητής θέλει να διεξαγάγει στατιστικό έλεγχο υποθέσεων για ποιον πληθυσμό;

(γ) Ποια είναι η μηδενική υπόθεση, H_0;

(δ) Ποια είναι η εναλλακτική υπόθεση, H_1;

(ε) Καθορίστε τον κανόνα απόφασης, χρησιμοποιώντας το επίπεδο σημαντικότητας 0,05.

(στ) Υπολογίστε την τιμή του z. (Μην ξεχάσετε να μετατρέψετε την τυπική απόκλιση σε τυπικό σφάλμα.)

(ζ) Ποια είναι η απόφασή σας για την H_0;

(η) Χρησιμοποιώντας λέξεις, ερμηνεύστε αυτήν την απόφαση σε σχέση με το αρχικό πρόβλημα.

Απαντήσεις στη σελίδα 542.

Περίληψη

Για να ελέγξετε μια υπόθεση για τον μέσο πληθυσμού, ένας μόνο παρατηρούμενος δειγματικός μέσος εξετάζεται στο πλαίσιο μιας υποθετικής κατανομής δειγματοληψίας, η οποία επικεντρώνεται γύρω από έναν μέσο πληθυσμού μηδενικής υπόθεσης. Αν ο δειγματικός μέσος φαίνεται να προκύπτει από την πυκνή συγκέντρωση πιθανών δειγματικών μέσων στο μέσον της κατανομής δειγματοληψίας, θεωρείται κοινό αποτέλεσμα και η μηδενική υπόθεση διατηρείται. Από την άλλη πλευρά, αν ο δειγματικός μέσος φαίνεται να προκύπτει από την αραιή συγκέντρωση πιθανών *δειγματικών μέσων* στις άκρες της κατανομής *δειγματοληψίας*, θεωρείται σπάνιο αποτέλεσμα και η μηδενική υπόθεση απορρίπτεται.

Οι στατιστικοί έλεγχοι υποθέσεων δεν βασίζονται στην κατανομή δειγματοληψίας του $\overline{X}$ όπως εκφράζεται στις αρχικές μονάδες μέτρησης, αλλά στην τυποποιημένη ισοδύναμη, την κατανομή δειγματοληψίας του z. Αυτός ο έλεγχος, ο οποίος αναφέρεται ως έλεγχος z για έναν μέσο πληθυσμού, είναι κατάλληλος μόνο όταν (1) ο πληθυσμός κατανέμεται κανονικά ή το μέγεθος δείγματος είναι πολύ μεγάλο για να ικανοποιήσει το κεντρικό οριακό θεώρημα και (2) η τυπική απόκλιση πληθυσμού είναι γνωστή.

Όταν κάνετε στατιστικό έλεγχο υπόθεσης, υιοθετείτε την παρακάτω αναλυτική διαδικασία:

- **Διατυπώστε το πρόβλημα έρευνας.** Χρησιμοποιώντας λέξεις, δηλώστε το πρόβλημα που πρέπει να μελετήσει η έρευνα.
- **Αναγνωρίστε τις στατιστικές υποθέσεις.** Οι στατιστικές υποθέσεις αποτελούνται από μια μηδενική υπόθεση (H_0) και μια εναλλακτική υπόθεση (ή υπόθεση έρευνας) (H_1). Η μηδενική υπόθεση παρέχει την τιμή γύρω από την οποία επικεντρώνεται η υποθετική κατανομή δειγματοληψίας.

  Ανάλογα με το αποτέλεσμα του στατιστικού ελέγχου υποθέσεων, η H_0 θα διατηρηθεί ή θα απορριφθεί. Στον βαθμό που η H_0 υποδηλώνει ότι δεν συμβαίνει κάτι ιδιαίτερο στον υποκείμενο πληθυσμό, ο ερευνητής συνήθως ελπίζει ότι τελικά θα την απορρίψει υπέρ της H_1, της υπόθεσης έρευνας. Στο συγκεκριμένο κεφάλαιο, οι στατιστικές υποθέσεις παίρνουν τη μορφή

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

  (Δύο άλλες πιθανές μορφές για στατιστικές υποθέσεις περιγράφονται στο Κεφάλαιο 11.)
- **Καθορίστε έναν κανόνα απόφασης.** Αυτός ο κανόνας υποδεικνύει ακριβώς πότε πρέπει να απορριφθεί η H_0. Η ακριβής μορφή του κανόνα απόφασης εξαρτάται από πολλούς παράγοντες, για τους οποίους θα μιλήσουμε στο Κεφάλαιο 11. Σε οποιοδήποτε γεγονός, η H_0 απορρίπτεται κάθε φορά που το παρατηρούμενο z αποκλίνει από το 0 όσο ή περισσότερο από το κρίσιμο z.

  Το επίπεδο σημαντικότητας δείχνει πόσο σπάνιο πρέπει να είναι ένα παρατηρούμενο z (αν υποθέσουμε ότι ισχύει η H_0) προτού απορριφθεί η H_0.
- **Υπολογίστε την τιμή του παρατηρούμενου z.** Εκφράστε τον έναν παρατηρούμενο δειγματικό μέσο ως παρατηρούμενο z, χρησιμοποιώντας τον Τύπο 10.1.
- **Λάβετε μια απόφαση.** Διατηρήστε ή απορρίψτε την H_0 στο καθορισμένο επίπεδο σημαντικότητας, αιτιολογώντας την απόφασή σας καταδεικνύοντας τη σχέση μεταξύ παρατηρούμενης και κρίσιμης z-τιμής.
- **Ερμηνεύστε την απόφαση.** Χρησιμοποιώντας λέξεις, ερμηνεύστε την απόφαση σε σχέση με το αρχικό πρόβλημα έρευνας. Η απόρριψη της μηδενικής υπόθεσης υποστηρίζει την υπόθεση έρευνας, ενώ η διατήρησή της δεν καταφέρνει να το κάνει.

Σημαντικοί όροι

Κατανομή δειγματοληψίας του z
Έλεγχος z για μέσο πληθυσμού
Μηδενική υπόθεση (H_0)
Εναλλακτική υπόθεση (H_1)
Υπόθεση έρευνας
Κανόνας απόφασης
Κρίσιμη z-τιμή
Επίπεδο σημαντικότητας (α)

Κύριες εξισώσεις

ΛΟΓΟΣ z

$$z = \frac{\overline{X} - \mu_{\text{υποθ}}}{\sigma_{\overline{X}}}$$

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

Ερωτήσεις επανάληψης

10.6 Υπολογίστε την τιμή του ελέγχου z για τις παρακάτω περιπτώσεις.

(α) $\overline{X} = 12$, $\sigma = 9$, $n = 25$, $\mu_{\text{υποθ}} = 15$

(β) $\overline{X} = 3.600$, $\sigma = 4.000$, $n = 100$, $\mu_{\text{υποθ}} = 3.500$

(γ) $\overline{X} = 0{,}25$, $\sigma = 0{,}10$, $n = 36$, $\mu_{\text{υποθ}} = 0{,}22$

10.7 Δεδομένου ότι οι κρίσιμες z-τιμές είναι $\pm 1{,}96$, η H_0 πρέπει να γίνει αποδεκτή ή να απορριφθεί για τις z-τιμές που υπολογίσατε στην Άσκηση 10.6;

***10.8** Για τον πληθυσμό γενικά, η κλίμακα ευφυΐας ενηλίκων Wechsler έχει σχεδιαστεί έτσι ώστε να δίνει μια κανονική κατανομή βαθμών εξετάσεων με μέσο 100 και τυπική απόκλιση 15. Οι αρμόδιοι των σχολικών περιφερειών αναρωτιούνται αν, κατά μέσο όρο, ένας δείκτης IQ διαφορετικός του 100 περιγράφει τα νοητικά χαρίσματα όλων των φοιτητών της περιφέρειάς τους. Οι δείκτες IQ στην κλίμακα Wechsler λαμβάνονται από ένα τυχαίο δείγμα 25 φοιτητών και διαπιστώνεται ότι ο μέσος δείκτης IQ είναι 105. Χρησιμοποιώντας την αναλυτική διαδικασία του κεφαλαίου, ελέγξτε τη μηδενική υπόθεση στο επίπεδο σημαντικότητας 0,05.
Απαντήσεις στη σελίδα 543.

10.9 Το κανονικό εύρος για ένα ευρέως αποδεκτό μέτρο του σωματικού μεγέθους, τον δείκτη μάζας σώματος (ΔΜΣ), κυμαίνεται από 18,5 ως 25. Χρησιμοποιώντας το μεσαίο ΔΜΣ 21,75 ως την τιμή της μηδενικής υπόθεσης για τον μέσο πληθυσμού, ελέγξτε αυτήν την υπόθεση στο επίπεδο σημαντικότητας 0,01 για ένα τυχαίο δείγμα 30 συμμετεχόντων οι οποίοι μετρούν το βάρος τους και έχουν μέσο ΔΜΣ = 22,2 και τυπική απόκλιση 3,1.

10.10 Θα υποθέσουμε ότι με το πέρασμα των χρόνων μια εξέταση άγχους που διενεργείται με χαρτί και μολύβι δίνει μέσο αποτέλεσμα 35 για όλους τους νέους πρωτοετείς φοιτητές. Θέλουμε να καταλάβουμε αν μπορούμε να θεωρήσουμε ότι τα αποτελέσματα ενός τυχαίου δείγματος 20 νέων πρωτοετών, με μέσο 30 και τυπική απόκλιση 10, προέρχονται απ' αυτόν τον πληθυσμό. Ελέγξτε στο επίπεδο σημαντικότητας 0,05.

10.11 Σύμφωνα με τον εκπαιδευτικό κώδικα της Καλιφόρνια (http://www.cde.ca.gov/ls/fa/sf/peguidemidhi.asp), οι μαθητές της δευτεροβάθμιας εκπαίδευσης θα πρέπει να παρακολουθούν για 400 λεπτά το μάθημα της φυσικής αγωγής κάθε 10 σχολικές ημέρες. Ένα τυχαίο δείγμα 48 μαθητών έχει μέσο 385 λεπτά και τυπική απόκλιση 53 λεπτά. Ελέγξτε την υπόθεση στο επίπεδο σημαντικότητας 0,05 ότι ο πληθυσμός του δείγματος ικανοποιεί αυτήν την προϋπόθεση.

10.12 Σύμφωνα με μια έρευνα του 2009 που βασίστηκε στην απογραφή πληθυσμού των ΗΠΑ (http://www.census.gov/prod/2011pubs/acs-15.pdf), ο καθημερινός χρόνος μετακίνησης αμερικανών εργαζομένων μόνο για τη μετάβαση και όχι και για την επιστροφή τους από τη δουλειά ήταν κατά μέσο όρο 25 λεπτά, με μια υποθετική τυπική απόκλιση 13 λεπτών. Ένας ερευνητής θέλει να διαπιστώσει αν ο εθνικός μέσος όρος περιγράφει τον μέσο χρόνο μετάβασης για όλους τους εργαζομένους της περιοχής του Σικάγου. Οι χρόνοι μετάβασης στη δουλειά λαμβάνονται από ένα τυχαίο δείγμα 169 εργαζομένων από τη συγκεκριμένη περιοχή και διαπιστώνεται ότι ο μέσος χρόνος είναι 22,5 λεπτά. Ελέγξτε τη μηδενική υπόθεση στο επίπεδο σημαντικότητας 0,05.

10.13 Προσθέστε τις λέξεις που λείπουν στις παρακάτω προτάσεις:
Αν ο ένας παρατηρούμενος δειγματικός μέσος μπορεί να θεωρηθεί __(α)__ αποτέλεσμα σύμφωνα με όσα ορίζει η υπόθεση, η H_0 θα __(β)__ . Διαφορετικά, αν ο ένας παρατηρούμενος δειγματικός μέσος μπορεί να θεωρηθεί __(γ)__ αποτέλεσμα σύμφωνα με την υπόθεση, η H_0 θα __(δ)__ .

Το ζεύγος z-τιμών που διαχωρίζει τα κοινά από τα σπάνια αποτελέσματα αναφέρεται ως __(ε)__ z-τιμές. Μέσα στην υποθετική κατανομή δειγματοληψίας, η αναλογία του εμβαδού της περιοχής που κατανέμεται σε σπάνια αποτελέσματα αναφέρεται ως __(στ)__ και συμβολίζεται με το ελληνικό γράμμα __(ζ)__ .

Όταν βασιζόμαστε στην κατανομή δειγματοληψίας του z, ο στατιστικός έλεγχος υποθέσεων αναφέρεται ως έλεγχος __(η)__ . Αυτός ο έλεγχος ταιριάζει σε κάθε περίπτωση όπου το μέγεθος δείγματος είναι επαρκώς μεγάλο για να ικανοποιεί το __(θ)__ και αν η __(ι)__ είναι γνωστή.

ΠΑΡΑΡΤΗΜΑ Χ

Στατιστικός έλεγχος υποθέσεων: Ο έλεγχος *z*

Γιώργος Ανδρουλάκης

Π.Χ.1 Εισαγωγή

Από το παρόν Κεφάλαιο 10 ξεκινούν να παρουσιάζονται οι ενότητες της στατιστικής συμπερασματολογίας που περιλαμβάνουν τα διαστήματα εμπιστοσύνης και τον έλεγχο υποθέσεων.

Στο Κεφάλαιο 5 παρουσιάστηκε η κανονική κατανομή και ο τρόπος με τον οποίο μπορούμε να τη χειριστούμε με το λογισμικό R. Στο παρόν κεφάλαιο θα δούμε έναν τρόπο για να σκιαγραφούμε περιοχές στο γράφημα της κανονικής κατανομής και τον τρόπο με τον οποίο το λογισμικό R μπορεί να εκτελέσει έλεγχο υποθέσεων *z* με βάση την κανονική κατανομή.

Π.Χ.2 Το γράφημα της κανονικής κατανομής

Ας δούμε πώς μπορούμε να φτιάξουμε ένα πιο σύνθετο γράφημα στην R χρησιμοποιώντας εντολές που αρχικά το δημιουργούν και σταδιακά συμπληρώνουν στοιχεία πάνω σε αυτό.

Έστω ότι έχουμε στον νου μας να σχεδιάσουμε τη συνάρτηση κανονικής κατανομής με μέση τιμή 500 και τυπική απόκλιση 11 και κατόπιν να σκιαγραφήσουμε τις περιοχές που ορίζονται: (α) Από το μείον άπειρο έως και –2 τυπικές αποκλίσεις απόσταση από τον μέσο και (β) από 2 τυπικές αποκλίσεις απόσταση από τον μέσο έως το συν άπειρο. Αρχικά ορίζουμε και δίνουμε τιμές σε μεταβλητές σύμφωνα με τις παραμέτρους του προβλήματος. Γενικά, είναι μια καλή προγραμματιστική τεχνική να χρησιμοποιούμε μεταβλητές γιατί με αυτόν τον τρόπο ο κώδικας που γράφουμε είναι επαναχρησιμοποιήσιμος, αφού αλλάζοντας μόνο τις αρχικές τιμές το πρόγραμμά μας μπορεί να ξανατρέξει με τις νέες αυτές τιμές και να δώσει αποτελέσματα που γνωρίζουμε ήδη ότι είναι σωστά.

```
> mean=500; sd=11
> lb=478; ub=522
```

Παρατηρήστε ότι, εναλλακτικά, θα μπορούσαμε να ορίσουμε τις περιοχές lb και ub ως εξής:

```
> lb=mean - 2*sd; ub=mean + 2*sd
```

το οποίο θα ήταν συνεπέστερο με την περιγραφή του προβλήματός μας, αλλά θα μας δέσμευε σε μελλοντικές επαναχρησιμοποιήσεις για υλοποίηση αποκλειστικά 2 τυπικών αποκλίσεων απόσταση από τον μέσο της κατανομής.

Επόμενο βήμα είναι οριοθετήσουμε την περιοχή σχεδίασης της κανονικής κατανομής τυπικές αποκλίσεις από τον μέσο της κατανομής. Στο σημείο αυτό να θυμίσουμε ότι, σύμφωνα με το κεντρικό οριακό θεώρημα, το 99% των παρατηρήσεων από έναν πληθυσμό που ερμηνεύεται από την κανονική κατανομή λαμβάνει τιμές τυπικές αποκλίσεις γύρω από τη μέση τιμή του, επομένως η περιοχή που θέλουμε να οριοθετήσουμε για το γράφημά μας την υπερκαλύπτει.

```
> x <- seq(-4,4,length=100)*sd + mean
```

Κάναμε χρήση της εντολής seq της R ώστε στη μεταβλητή (διάνυσμα) x να δώσουμε 100 ισοαπέχουσες τιμές από –4 έως +4 τυπικές αποκλίσεις γύρω από τον μέσο της κατανομής μας.

```
> hx <- dnorm(x,mean,sd)
```

Με την παραπάνω εντολή στο διάνυσμα hx έχουν τοποθετηθεί οι τιμές της συνάρτησης πυκνότητας πιθανότητας της κανονικής κατανομής με παραμέτρους mean και sd που αντιστοιχούν στις 100 τιμές-σημεία του διανύσματος x.

Στην αρχή της ενότητας αυτής είπαμε ότι σταδιακά θα δημιουργήσουμε και θα συμπληρώσουμε το γράφημα του παραδείγματός μας. Η παρακάτω εντολή της R φτιάχνει το αρχικό γράφημα:

```
> plot(x, hx, type="n", xlab="Values", ylab="",
+ main="Normal Distribution", axes=T)
```

Συγκεκριμένα, δημιουργεί τους άξονες έτσι ώστε να περιλαμβάνουν στον οριζόντιο άξονα τις τιμές του διανύσματος x και στον κάθετο άξονα τις τιμές του διανύσματος hx. Τοποθετεί επίσης ονόματα στο γράφημα (όρισμα main), στον άξονα x (όρισμα xlab) και κενό όνομα στον άξονα y (όρισμα ylab). Το όρισμα type χρησιμοποιείται για να καθοριστεί ο τρόπος απεικόνισης και ένωσης των σημείων του γραφήματος. Στην προκείμενη περίπτωση, με την τιμή "n" δηλώνουμε να μην απεικονίσει ούτε τα σημεία ούτε να τα ενώσει μεταξύ τους. Την ένωση των σημείων την αναλαμβάνει η εντολή lines:

```
> lines(x, hx)
```

Οι επόμενες δύο γραμμές χρησιμοποιούν την εντολή segments της R, που ζωγραφίζει ευθύγραμμα τμήματα ανάμεσα σε δύο σημεία με έναν επαναληπτικό τρόπο ώστε να δημιουργήσει την ψευδαίσθηση της σκιαγράφησης μιας περιοχής. Συγκεκριμένα, κάνοντας χρήση επαναληπτικού βρόχου, ζωγραφίζει εκατό κάθετα ευθύγραμμα τμήματα χρώματος πορτοκαλί, περίπου το ένα δίπλα στο άλλο.

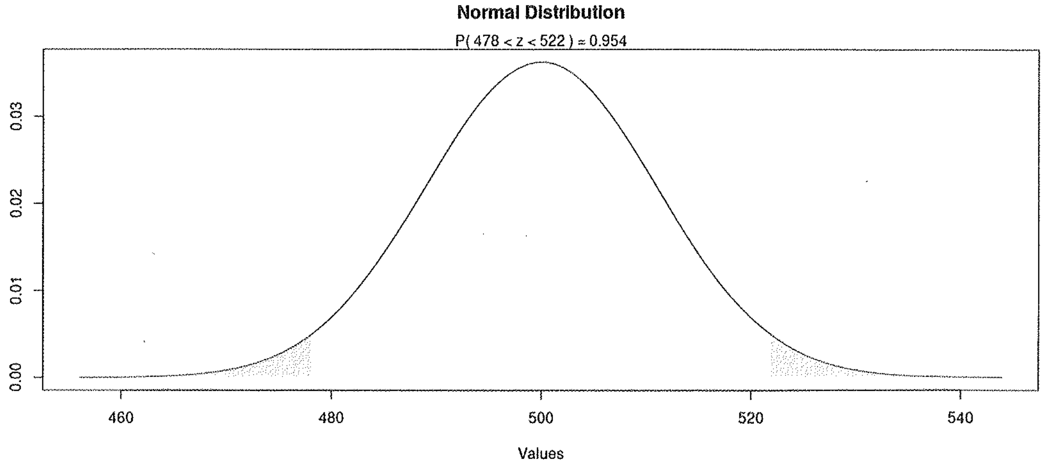
```
> for (j in seq(mean - 4*sd, lb, length=100)) {segments(j, 0, j, dnorm(j,mean,sd), col ="orange")}
> for (j in seq(ub, mean + 4*sd, length=100)) {segments(j, 0, j, dnorm(j,mean,sd), col ="orange")}
```

Οι παρακάτω τρεις γραμμές εντολών χρησιμοποιούνται η πρώτη για να υπολογιστεί το ολοκλήρωμα της περιοχής από lb έως ub κάνοντας χρήση της συνάρτησης pnorm, που την έχουμε ξανασυναντήσει στο Κεφάλαιο 5, και οι επόμενες δύο για να δομήσουν το κείμενο result συνενώνοντας κείμενο και μεταβλητές με τη χρήση της συνάρτησης paste. Να τονίσουμε ότι η συνάρτηση signif χρησιμοποιείται για να στρογγυλοποιήσει τη μεταβλητή area στα τρία δεκαδικά ψηφία.

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

Η συνάρτηση mtext, που εκτελείται παρακάτω, προσθέτει κείμενο σε επιλεγμένες θέσεις, ανάλογα με το όρισμα που χρησιμοποιούμε κάθε φορά [εκτελέστε help("mtext"), για να δείτε ένα πλήρες κείμενο βοήθειας σχετικά με αυτήν τη συνάρτηση]. Στη συγκεκριμένη περίπτωση τοποθετεί κάτω από τον τίτλο του γραφήματος το κείμενο result που κατασκευάσαμε προηγουμένως.

```
> mtext(result,3)
```



Π.Χ.3 Έλεγχος *z* για μέσο πληθυσμού

Στην ενότητα αυτή θα δούμε πώς μπορούμε να χρησιμοποιήσουμε τις έτοιμες συναρτήσεις της R ώστε να εκτελέσουμε έλεγχο υπόθεσης για τον μέσο ενός δείγματος κάνοντας μετατροπή σε αντίστοιχη *z*-τιμή.

Έστω ότι θέλουμε να ελέγξουμε τον μέσο ενός δείγματος που βρέθηκε ίσος με 533 σε σχέση με έναν υποθετικό πληθυσμό με μέσο 500 και τυπική απόκλιση 11. Ο έλεγχος που θέλουμε να κάνουμε είναι αμφίπλευρος σε επίπεδο σημαντικότητας 95%.

Παρατηρούμε ότι ο μέσος του δείγματος είναι μεγαλύτερος από τον μέσο του πληθυσμού, επομένως, αν απορριφθεί η μηδενική υπόθεση, αυτό θα οφείλεται στο ότι το *z* σημείο που αντιστοιχεί η τιμή 533 είναι πάνω από το άνω κρίσιμο σημείο του στατιστικού ελέγχου, δηλαδή βρίσκεται στην περιοχή απόρριψης της μηδενικής υπόθεσης. Αρχικά, υπολογίζουμε την πιθανότητα της κανονικής κατανομής που ορίζει το άνω κρίσιμο σημείο και την τοποθετούμε στη μεταβλητή critical.

```
> critical = 1 - 0.05/2
> critical
[1] 0.975
```

Υπολογίζουμε και εμφανίζουμε στην οθόνη το κρίσιμο σημείο του στατιστικού ελέγχου που αντιστοιχεί στην πιθανότητα critical.

```
> round(qnorm(critical), 2)
[1] 1.96
```

Μετατρέπουμε τη μέση τιμή 533 σε σημείο της τυποποιημένης κανονικής κατανομής *z* αφαιρώντας από αυτήν τον μέσο όρο και διαιρώντας με την τυπική απόκλιση. Το αποτέλεσμα το καταχωρίζουμε στη μεταβλητή myz και τυπώνουμε στην οθόνη μας για να ελέγξουμε την τιμή της.

```
> myz = (533 - mean) / sd
> myz
[1] 3
```

Τέλος, εκτελούμε λογικό έλεγχο αν η τιμή της τυποποιημένης κανονικής κατανομής myz που υπολογίσαμε προηγουμένως είναι μεγαλύτερη από το κρίσιμο σημείο. Το αποτέλεσμα του λογικού ελέγχου το τοποθετούμε σε μια μεταβλητή η οποία θα πάρει την τιμή TRUE αν ο λογικός έλεγχος ισχύει και την τιμή FALSE αν δεν ισχύει. Τυπώνουμε το αποτέλεσμα του λογικού ελέγχου στην οθόνη του υπολογιστή μας.

```
> Reject = myz > qnorm(critical)
> print(paste("z > 1.96 :", Reject))
[1] "z > 1.96 : TRUE"
```

Επομένως, βρισκόμαστε στην περιοχή απόρριψης της μηδενικής υπόθεσης.

Π.Χ.4 Έλεγχος *z* σε τυχαίο δείγμα

Ένας παρατηρητικός αναγνώστης θα έχει διαπιστώσει ότι για όλες τις συχνά χρησιμοποιούμενες στατιστικές διαδικασίες στο λογισμικό R υπάρχει και μία έτοιμη συνάρτηση. Λογικό είναι να υποθέσει ότι και για τον έλεγχο *z* θα υπάρχει μια έτοιμη συνάρτηση στη βασική υλοποίηση της R. Λάθος. Δεν υπάρχει. Η απάντηση για έναν έμπειρο στατιστικό είναι απλή. Στα πραγματικά προβλήματα σπανιότατα θα χρειαστεί να εκτελέσουμε έλεγχο υπόθεσης *z* γιατί έχει ως προϋπόθεση τη γνώση του σ. Για ποιο λόγο να θεωρήσει ένας έμπειρος ερευνητής ότι το σ που ήξερε δεν έχει μεταβληθεί; Ειδικά όταν η τυπική απόκλιση που μετράει στο δείγμα του διαφέρει από το σ που είχε από παλιότερες έρευνες. Όπως θα δούμε σε επόμενο κεφάλαιο, «αναγκάζεται» να καταφύγει σε έλεγχο υπόθεσης που δε στηρίζεται στην κανονική κατανομή, άρα δεν θα χρησιμοποιήσει έλεγχο *z*. Παρ' όλα αυτά, έστω και για εκπαιδευτικούς λόγους, θα έπρεπε να υπάρχει έτοιμη συνάρτηση στην R.

Αναζητώντας τέτοια συνάρτηση στην τεκμηρίωση για την R στον παγκόσμιο ιστό, θα βρείτε ότι στη βιβλιοθήκη BSDA της R υπάρχει μια τέτοια συνάρτηση. Επομένως, το μόνο που χρειάζεται είναι να εγκαταστήσουμε τη βιβλιοθήκη BSDA στον υπολογιστή μας και μετά να την ενεργοποιήσουμε μέσα στο περιβάλλον της R.

Για την εγκατάσταση της βιβλιοθήκης χρειάζεται ο υπολογιστής μας να έχει ενεργή πρόσβαση στο διαδίκτυο και να εκτελέσουμε την εντολή:

```
> install.packages("BSDA")
```

Η παραπάνω εντολή χρειάζεται να τρέξει μόνο μία φορά. Αν στο παρελθόν είχαμε εγκαταστήσει τη βιβλιοθήκη, τότε αυτή βρίσκεται ήδη στον υπολογιστή μας και το μόνο που μένει είναι να την ενεργοποιήσουμε στο περιβάλλον της R. (Για λόγους διαχείρισης της μνήμης του υπολογιστή μας το λογισμικό R δεν ενεργοποιεί αυτόματα όλες τις βιβλιοθήκες που έχουμε εγκαταστήσει, γιατί αυτό θα ισοδυναμούσε με μεγάλη σπατάλη υπολογιστικής ισχύος για συναρτήσεις που μπορεί να μη χρειαστούμε κάθε φορά.) Η ενεργοποίηση γίνεται με την παρακάτω εντολή:

```
> library(BSDA)
```

Για να δοκιμάσουμε τη χρήση της εντολής z.test που είναι υλοποιημένη στη βιβλιοθήκη BSDA φτιάχνουμε ένα τυχαίο δείγμα 20 αριθμών που προέρχονται από την κανονική κατανομή με μέσο 533 και τυπική απόκλιση sd, δηλαδή 11.

```
> myx = round(rnorm(20, 533, sd),0)
> myx
[1] 540 536 528 540 541 527 546 552 530 521 543 536 528 556 547 524 541 533 525 524
```

Τώρα είμαστε έτοιμοι να εκτελέσουμε αμφίπλευρο έλεγχο για το δείγμα μας myx αναφορικά με το αν μπορεί να προέρχεται από έναν πληθυσμό με μέσο 500 και τυπική απόκλιση 11.

```
> z.test(myx, y=NULL, mu=mean, sigma.x = sd)

            One-sample z-Test

data: myx
z = 14.595, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 500
95 percent confidence interval:
531.0791 540.7209
sample estimates:
mean of x
     535.9
```

Οι δύο γραμμές με έντονα γράμματα μας ενημερώνουν για το αποτέλεσμα του αντίστοιχου έλεγχου υπόθεσης παρέχοντας πληροφορία για τη *z*-τιμή του μέσου του δείγματος, για την *p*-τιμή του στατιστικού ελέγχου και για το είδος του ελέγχου. Οι υπόλοιπες εμφανιζόμενες πληροφορίες αφορούν τα διαστήματα εμπιστοσύνης, που θα δούμε σε επόμενο κεφάλαιο. Όπως θα μάθουμε, έλεγχος υπόθεσης και διαστήματα εμπιστοσύνης είναι αλληλένδετες στατιστικές διαδικασίες και, επομένως, δεν υπάρχει λόγος να εκτελούνται με διαφορετική εντολή. Ανάλογα με τη χρήση που θέλουμε να κάνουμε θα επιλέγουμε πότε την πληροφορία που αντιστοιχεί στον έλεγχο υποθέσεων και πότε την πληροφορία που αντιστοιχεί στα διαστήματα εμπιστοσύνης.

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)

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.

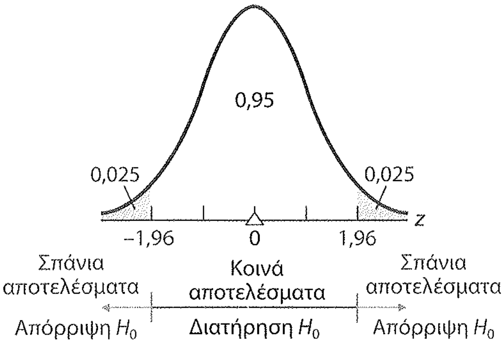
Πιθανότητα εσφαλμένων αποφάσεων

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

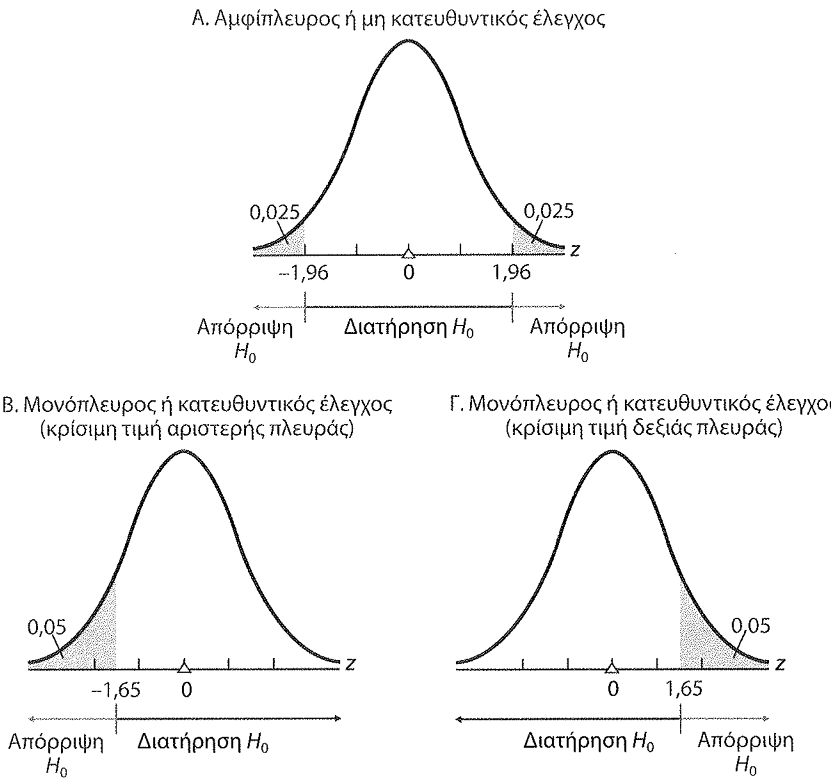
Θα μελετήσουμε τώρα μερικές τεχνικές που συμβάλλουν έτσι ώστε ο στατιστικός έλεγχος υποθέσεων να ανταποκρίνεται καλύτερα σε ειδικές περιπτώσεις.

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 ($\alpha = 0{,}05$).

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, α, 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)	±1,96	±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 I. 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 II.

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 I** *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 I ονομάζονται μερικές φορές *εσφαλμένοι συναγερμοί* επειδή προκαλούν άσκοπες ενέργειες για κάτι που στην πραγματικότητα δεν υπάρχει. Για παράδειγμα, ένα σφάλμα τύπου I 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 I
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 II** *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 II αναφέρονται κάποιες φορές ως *αστοχίες* επειδή δεν καταφέρνουν να ανιχνεύσουν μια δυνητικά σημαντική σχέση, όπως αυτή μεταξύ βιταμίνης C και IQ.

Σφάλμα τύπου II
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

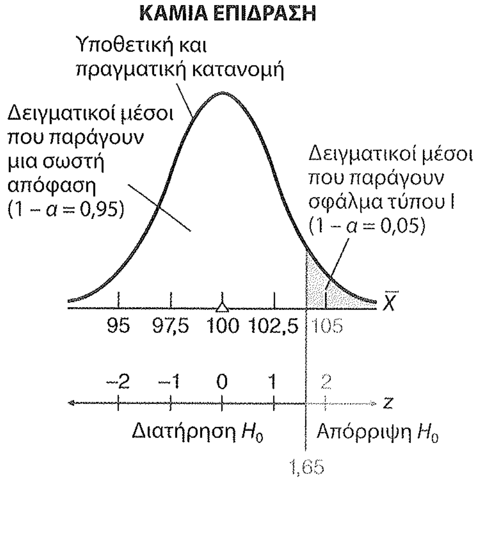
(α) Αναφέρετε τα τέσσερα πιθανά αποτελέσματα για οποιονδήποτε στατιστικό έλεγχο υποθέσεων.

(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.

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

Η αληθής H_0 συνήθως διατηρείται

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.

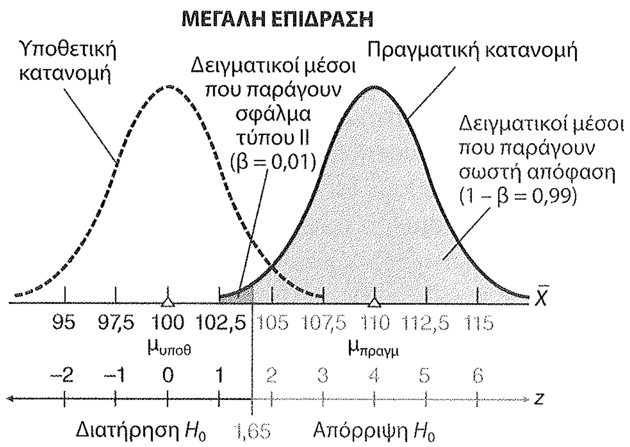
Έλεγχος προόδου *11.5 Για να εξαλειφθεί το σφάλμα τύπου I, κάποιος αποφασίζει να χρησιμοποιήσει το επίπεδο σημαντικότητας 0,00. Τι λάθος έχει αυτή η διαδικασία;

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.

γία 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.

Το συγκεκριμένο επιχείρημα δεν προϋποθέτει ότι γνωρίζετε πώς να υπολογίζετε αυτήν

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

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:

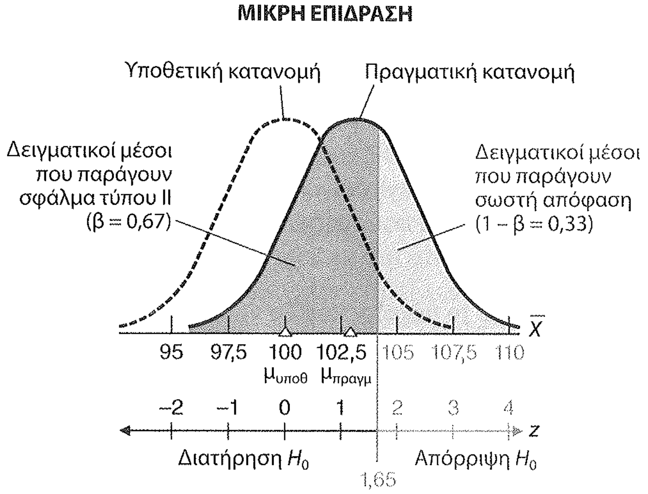
(α) Η υπόθεση ότι η H_0 είναι στην πραγματικότητα ψευδής αποτυπώνεται από τον διαχωρισμό της υποθετικής και της πραγματικής κατανομής.
(β) Στην πράξη, όταν κάνουμε στατιστικό έλεγχο υπόθεσης, δεν μπορούμε να γνωρίζουμε ότι ο πραγματικός μέσος πληθυσμού ισούται με 110.
(γ) Ο ένας παρατηρούμενος δειγματικός 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, $\sigma_{\bar{X}}$,

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

Για το αρχικό πείραμα με μέγεθος δείγματος 36,

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

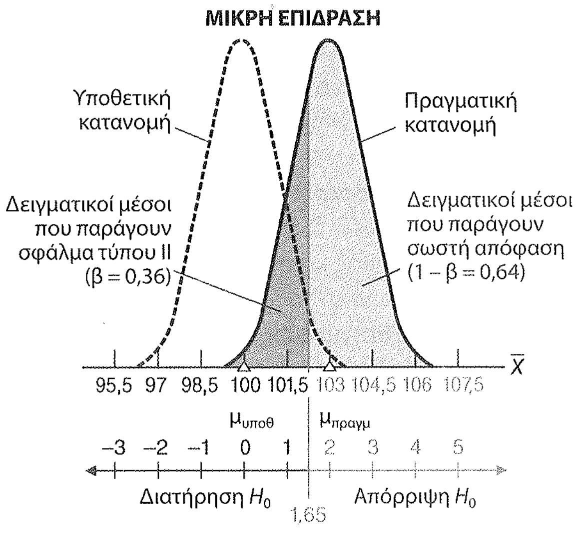
ενώ για το νέο πείραμα με μέγεθος δείγματος 100,

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

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 II 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 II 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

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 ΙΙ. Οι σκιασμένοι τομείς στα Σχήματα 11.4, 11.5 και 11.6 αποτυπώνουν διαφορετικούς βαθμούς ισχύος.

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

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 ΙΙ, 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

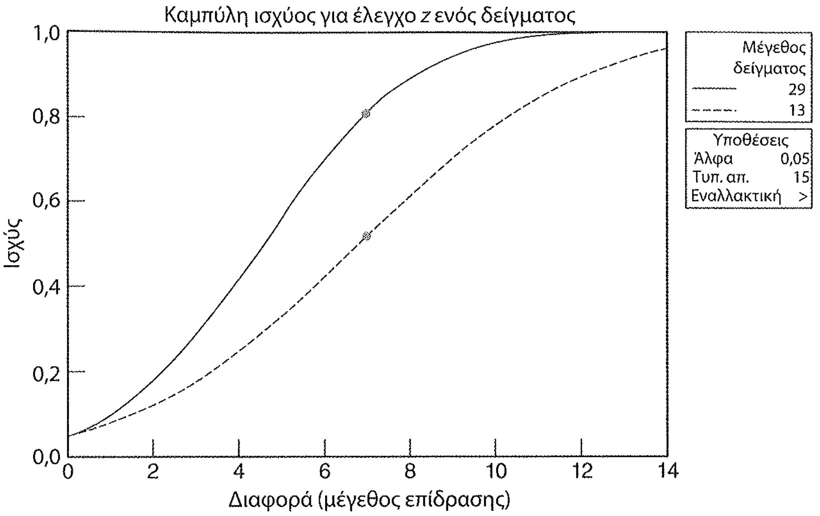
Μια **καμπύλη ισχύος** *δείχνει πώς μεταβάλλεται η πιθανότητα ανίχνευσης οποιασδήποτε πιθανής επίδρασης* -από πολύ μικρή ως πολύ μεγάλη- *για ένα σταθερό μέγεθος δείγματος.*[27] Με λίγα μόνο πατήματα πλήκτρων στο πληκτρολόγιο, το λογισμικό *Power and Sample Size* του Minitab υπολογίζει ότι το μέγεθος 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.

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

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 – *β*)
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

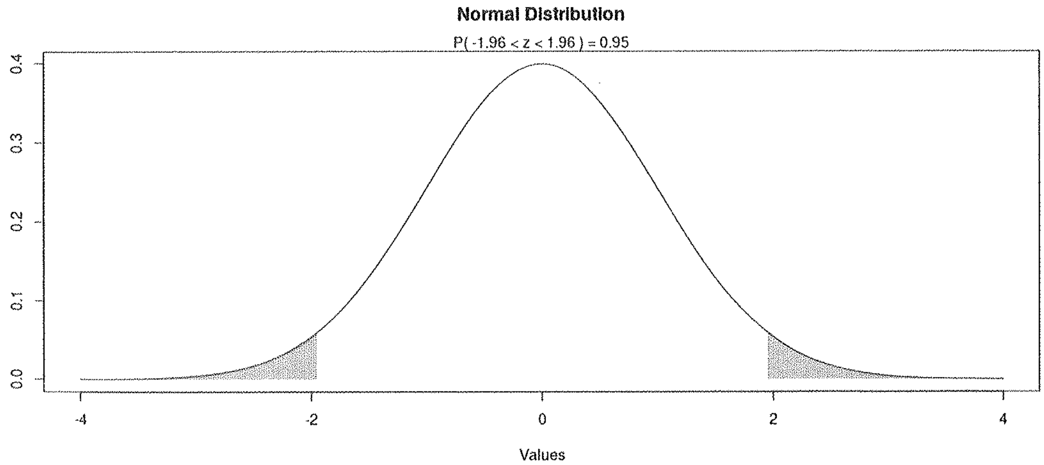
Το παρόν Κεφάλαιο 11 συμπληρώνει τα παραδείγματα που ξεκινήσαμε να παρουσιάζουμε στο Κεφάλαιο 10 εστιάζοντας στους μονόπλευρους ελέγχους υποθέσεων.

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.

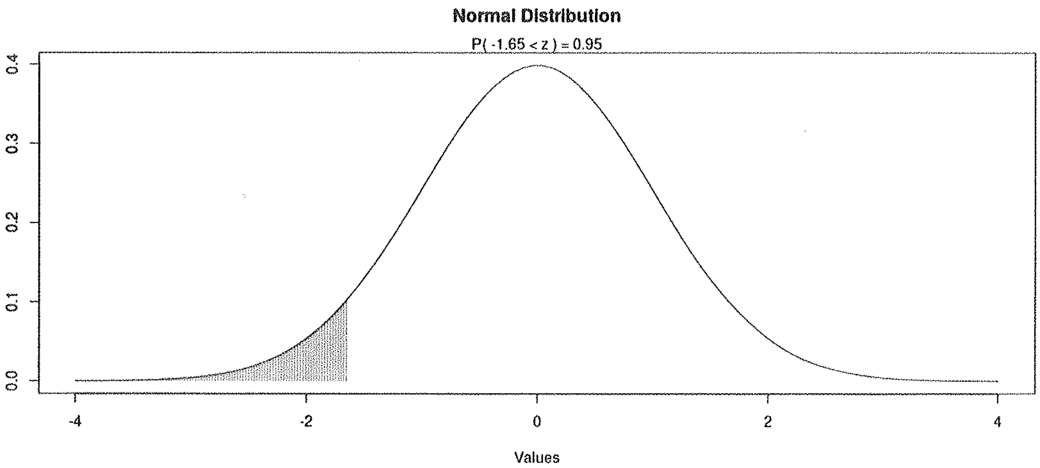


Ένα μεγάλο πλεονέκτημα στον προγραμματισμό με επαναχρησιμοποιήσιμο κώδικα είναι η ευκολία με την οποία με ελάχιστες μετατροπές σε προϋπάρχουσα υλοποίηση μπορούμε να δημιουργήσουμε κώδικα για νέες στατιστικές διεργασίες.

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:

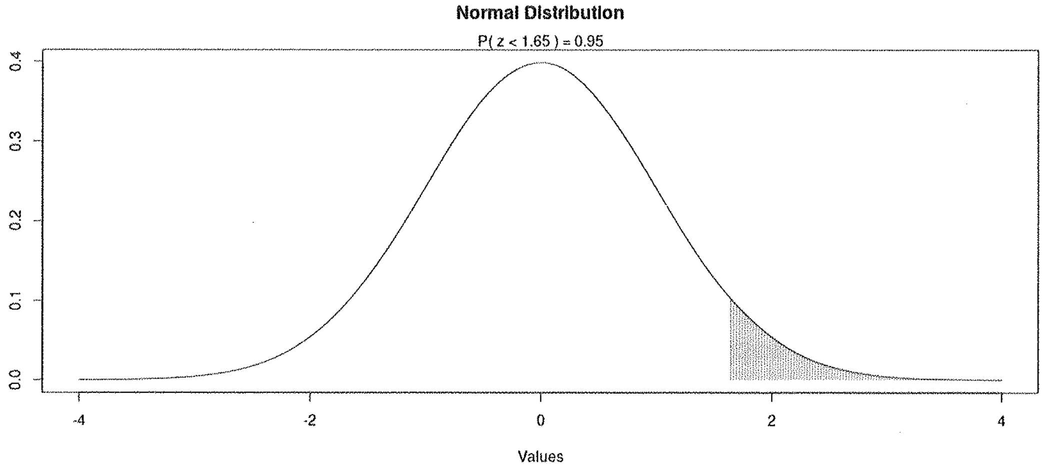


Αντίστοιχα, για να σκιαγραφήσουμε το δεξιό άκρο του σχήματος, ώστε να οπτικοποιήσουμε την κρίσιμη περιοχή απόρριψης από δεξιά, ο προσαρμοσμένος κώδικας είναι ο παρακάτω:

```
> 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)

ΚΕΦΑΛΑΙΟ

12 Εκτίμηση (διαστήματα εμπιστοσύνης)

Πρόλογος

Ένας στατιστικός έλεγχος υποθέσεων απλώς δείχνει αν υπάρχει κάποια επίδραση. Το διάστημα εμπιστοσύνης προσφέρει περισσότερες πληροφορίες, καθώς δείχνει, με έναν γνωστό βαθμό εμπιστοσύνης, το εύρος των πιθανών επιδράσεων. Τα διαστήματα εμπιστοσύνης μπορεί να είναι μεμονωμένα ή ως συνέπεια ενός ελέγχου που έχει απορρίψει τη μηδενική υπόθεση. Όσο ωριμάζει το πεδίο της έρευνας, η χρήση διαστημάτων εμπιστοσύνης διευρύνεται περισσότερο.

Στο Κεφάλαιο 10, είδαμε την περίπτωση μιας έρευνας για την ανίχνευση της διαφοράς μεταξύ του μέσου βαθμού σε εξετάσεις μαθηματικών για όλους τους ντόπιους πρωτοετείς και του εθνικού μέσου όρου. Αυτή η περίπτωση οδήγησε σε έναν έλεγχο z και στο συμπέρασμα ότι ο μέσος για τον ντόπιο πληθυσμό υπερβαίνει τον εθνικό μέσο όρο. Ως προς τον εθνικό μέσο όρο, αυτό το συμπέρασμα είναι ιδιαίτερα ενημερωτικό και μπορεί να προκαλέσει ακόμα και ευαρέσκεια στους υπευθύνους του τοπικού πανεπιστημίου. Ωστόσο, η ίδια έρευνα για τις εξετάσεις SAT θα μπορούσε να είχε ξεκινήσει από μια επιθυμία για απλή *εκτίμηση* της τιμής του μέσου του τοπικού πληθυσμού και όχι για έλεγχο μιας *υπόθεσης* που θα βασιζόταν στον εθνικό μέσο όρο. Αυτό το νέο θέμα μεταφράζεται σε ένα πρόβλημα εκτίμησης, και με τη βοήθεια των εκτιμήσεων σημείων και των διαστημάτων εμπιστοσύνης, οι πληροφορίες ενός δείγματος μπορούν να χρησιμοποιηθούν για την εκτίμηση του άγνωστου μέσου του πληθυσμού όλων των ντόπιων πρωτοετών.

12.1 Εκτίμηση σημείου για το μ

Εκτίμηση σημείου
Μία μόνο τιμή που αναπαριστά κάποιο χαρακτηριστικό του πληθυσμού, όπως τον μέσο πληθυσμού.

Μια *εκτίμηση σημείου* για το μ χρησιμοποιεί μία μόνο τιμή για να αναπαραστήσει τον άγνωστο μέσο πληθυσμού.

Αυτός είναι ο πιο άμεσος τύπος εκτίμησης. Αν ένα τυχαίο δείγμα 100 ντόπιων πρωτοετών αποκαλύπτει έναν μέσο βαθμό σε εξετάσεις SAT 533, τότε το 533 θα είναι η εκτίμηση σημείου του άγνωστου μέσου του πληθυσμού όλων των ντόπιων πρωτοετών. Η καλύτερη εκτίμηση σημείου για τον άγνωστο μέσο του πληθυσμού είναι απλώς η παρατηρούμενη τιμή του δειγματικού μέσου.

Μια βασική ατέλεια

Αν και είναι άμεσες και απλές, οι εκτιμήσεις σημείου πάσχουν από μια βασική ατέλεια. Τείνουν να είναι ανακριβείς. Εξαιτίας της μεταβλητότητας λόγω δειγματοληψίας, είναι απίθανο ένας μόνο δειγματικός μέσος, όπως το 533, να συμπίπτει με τον μέσο πληθυσμού. Επειδή οι εκτιμήσεις σημείου δεν φέρουν καμία πληροφορία για τον βαθμό ανακρίβειας εξαιτίας της μεταβλητότητας λόγω δειγματοληψίας, οι επιστήμονες της στατιστικής συμπληρώνουν τις εκτιμήσεις σημείου με έναν άλλο, πιο ρεαλιστικό τύπο εκτίμησης, τις *εκτιμήσεις διαστημάτων* ή τα *διαστήματα εμπιστοσύνης*.

Έλεγχος προόδου *12.1 Ένα τυχαίο δείγμα 200 αποφοίτων αμερικανικών πανεπιστημίων έχει μέσο ετήσιο εισόδημα $62.600. Ποια είναι η καλύτερη εκτίμηση του άγνωστου μέσου ετήσιου εισοδήματος για όλους τους αποφοίτους αμερικανικών πανεπιστημίων;

Η απάντηση στη σελίδα 545.

12.2 Διάστημα εμπιστοσύνης (CI) για το μ

Ένα *διάστημα εμπιστοσύνης* για το μ χρησιμοποιεί ένα εύρος τιμών το οποίο, σε έναν γνωστό βαθμό βεβαιότητας, περιλαμβάνει την τιμή για τον άγνωστο μέσο πληθυσμού.

Διάστημα εμπιστοσύνης (CI) *Ένα εύρος τιμών το οποίο, με έναν γνωστό βαθμό βεβαιότητας, περιλαμβάνει ένα άγνωστο χαρακτηριστικό του πληθυσμού, όπως έναν μέσο πληθυσμού.*

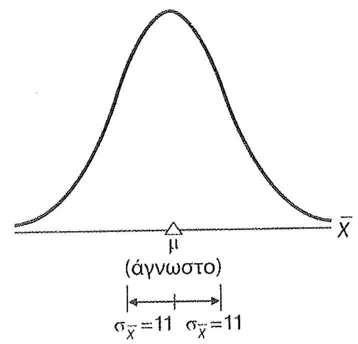
Για παράδειγμα, ο ερευνητής των εξετάσεων SAT θα μπορούσε να χρησιμοποιήσει ένα διάστημα εμπιστοσύνης για να ισχυριστεί, *με 95% βεβαιότητα*, ότι το διάστημα μεταξύ 511,44 και 554,56 περιλαμβάνει τον μέσο βαθμό του πληθυσμού σε εξετάσεις μαθηματικών για όλους τους ντόπιους πρωτοετείς. Το 95% βεβαιότητα σημαίνει ότι, αν πολλά απ' αυτά τα διαστήματα κατασκευάστηκαν για μια μακρά σειρά δειγμάτων, περίπου το 95% θα περιλάμβανε τον μέσο πληθυσμού για όλους τους ντόπιους πρωτοετείς. Μακροπρόθεσμα, το 95% αυτών των διαστημάτων εμπιστοσύνης είναι σωστό επειδή περιλαμβάνει τον άγνωστο μέσο πληθυσμού. Το υπόλοιπο 5% είναι λάθος επειδή δεν συμπεριλαμβάνει τον άγνωστο μέσο πληθυσμού.

Γιατί λειτουργούν τα διαστήματα εμπιστοσύνης

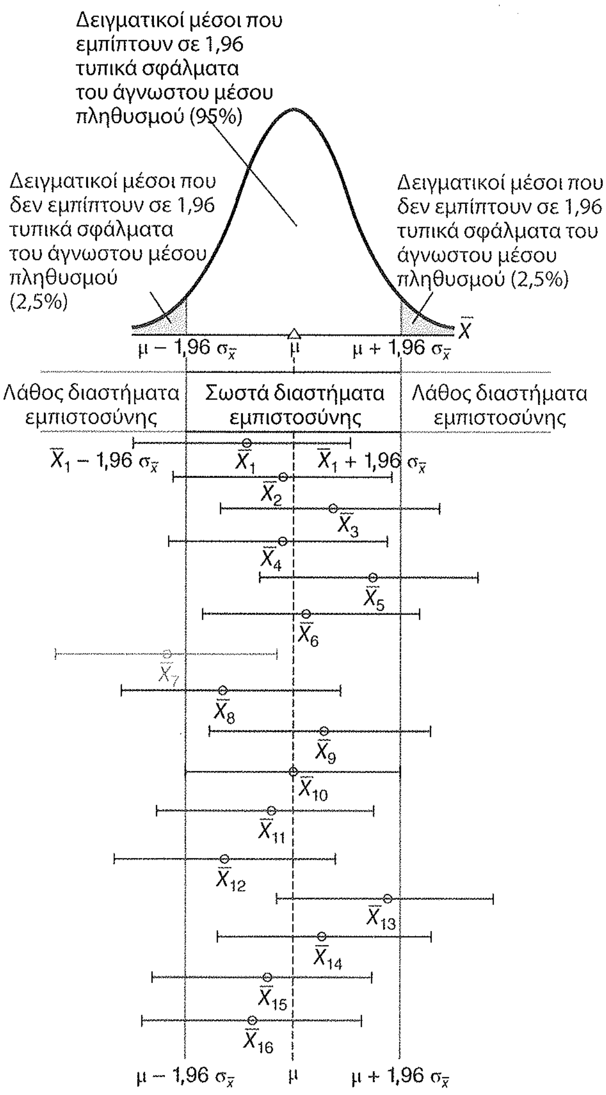
Για να κατανοήσετε την έννοια των διαστημάτων εμπιστοσύνης, πρέπει να τα μελετήσετε στο πλαίσιο τριών σημαντικών ιδιοτήτων της κατανομής δειγματοληψίας του μέσου, όπως περιγράφονται στο Κεφάλαιο 10.

Για την κατανομή δειγματοληψίας από την οποία προέρχεται ο δειγματικός μέσος 533, όπως βλέπετε στο Σχήμα 12.1, οι τρεις σημαντικές ιδιότητες είναι οι εξής:

- Ο μέσος της κατανομής δειγματοληψίας ισούται με τον άγνωστο μέσο πληθυσμού για όλους τους ντόπιους πρωτοετείς, ανεξαρτήτως της τιμής του, επειδή ο μέσος αυτής της κατανομής δειγματοληψίας ισούται πάντα με τον μέσο του πληθυσμού.
- Το τυπικό σφάλμα της κατανομής δειγματοληψίας ισούται με την τιμή (11) που παίρνουμε από τη διαίρεση της τυπικής απόκλισης πληθυσμού (110) διά της τετραγωνικής ρίζας του μεγέθους δείγματος ($\sqrt{100}$).
- Το σχήμα της κατανομής δειγματοληψίας προσομοιάζει με κανονική κατανομή επειδή το μέγεθος δείγματος 100 ικανοποιεί τις προϋποθέσεις του κεντρικού οριακού θεωρήματος.



ΣΧΗΜΑ 12.1
Κατανομή δειγματοληψίας του μέσου (βαθμοί σε εξετάσεις SAT).



ΣΧΗΜΑ 12.2

Μια σειρά από διαστήματα εμπιστοσύνης 95% (όπως αποκαλύπτονται από μια κατανομή δειγματοληψίας).

Μια σειρά διαστημάτων εμπιστοσύνης

Στην πράξη, παίρνουμε μόνο έναν δειγματικό μέσο απ' αυτήν την κατανομή δειγματοληψίας και τον χρησιμοποιούμε για να κατασκευάσουμε ένα μόνο διάστημα εμπιστοσύνης 95%. Ωστόσο, έστω ότι δεν παίρνουμε μόνο έναν αλλά μια σειρά από τυχαία επιλεγμένους δειγματικούς μέσους απ' αυτήν την κατανομή δειγματοληψίας. Εξαιτίας της μεταβλητότητας λόγω δειγματοληψίας, αυτοί οι δειγματικοί μέσοι τείνουν να διαφέρουν μεταξύ τους. Για κάθε δειγματικό μέσο, κατασκευάζετε ένα διάστημα εμπιστοσύνης 95% προσθέτοντας 1,96 τυπικά σφάλματα στον δειγματικό μέσο και αφαιρώντας 1,96 τυπικά σφάλματα από τον δειγματικό μέσο, δηλαδή χρησιμοποιείτε την έκφραση

$$\overline{X} \pm 1{,}96\sigma_{\overline{X}}$$

για να πάρετε ένα διάστημα εμπιστοσύνης 95% για κάθε δειγματικό μέσο.

Σωστά διαστήματα εμπιστοσύνης

Γιατί, σύμφωνα με τη στατιστική θεωρία, το 95% αυτών των διαστημάτων εμπιστοσύνης περιλαμβάνει τον άγνωστο μέσο πληθυσμού; Όπως δείχνει το Σχήμα 12.2, επειδή η κατανομή δειγματοληψίας είναι κανονική, το 95% όλων των δειγματικών μέσων εμπίπτει εντός των 1,96 τυπικών σφαλμάτων του άγνωστου μέσου πληθυσμού, δηλαδή το 95% όλων των δειγματικών μέσων αποκλίνει λιγότερο από *1,96 τυπικά σφάλματα από τον άγνωστο μέσο πληθυσμού*. Επομένως, και προσέξτε ιδιαίτερα σ' αυτό το σημείο, όταν οι δειγματικοί μέσοι επεκτείνονται σε διαστήματα εμπιστοσύνης –προσθέτοντας και αφαιρώντας 1,96 τυπικά σφάλματα–, σε ποσοστό 95% όλα τα πιθανά διαστήματα εμπιστοσύνης είναι σωστά επειδή περιλαμβάνουν τον άγνωστο μέσο πληθυσμού. Για να αποσαφηνίσουμε αυτό το σημείο, οι 15 από τους 16 δειγματικούς μέσους του Σχήματος 12.2 εμπίπτουν εντός των 1,96 τυπικών σφαλμάτων του άγνωστου μέσου πληθυσμού. Τα αντίστοιχα 15 διαστήματα εμπιστοσύνης έχουν εύρος που επεκτείνουν τη διακεκομμένη γραμμή για τον μέσο πληθυσμού και μπορούν, κατά συνέπεια, να θεωρηθούν σωστά διαστήματα επειδή περιλαμβάνουν την τιμή του άγνωστου μέσου πληθυσμού.

Λάθος διαστήματα εμπιστοσύνης

Το 5% όλων των διαστημάτων εμπιστοσύνης δεν περιλαμβάνεται στον άγνωστο μέσο πληθυσμού. Όπως δείχνει το Σχήμα 12.2, το 5% όλων των δειγματικών μέσων (2,5% σε κάθε πλευρά) αποκλίνει περισσότερο από 1,96 τυπικά σφάλματα από τον άγνωστο μέσο πληθυσμού. Επομένως, όταν οι δειγματικοί μέσοι επεκτείνονται σε διαστήματα εμπιστοσύνης –προσθέτοντας και αφαιρώντας 1,96 τυπικά σφάλματα–, σε ποσοστό 5% όλα τα πιθανά διαστήματα εμπιστοσύνης είναι λάθος επειδή δεν συμπεριλαμβάνουν τον άγνωστο μέσο πληθυσμού. Για να τονίσουμε αυτό το σημείο, μόνο 1 από τους 16 δειγματικούς μέσους που προβάλλονται στο Σχήμα 12.2 δεν εμπίπτει στα 1,96 τυπικά σφάλματα του άγνωστου μέσου πληθυσμού. Το τελικό διάστημα εμπιστοσύνης, το οποίο εμφανίζεται σκιασμένο, έχει εύρος που δεν επεκτείνει τη διακεκομμένη γραμμή για τον μέσο πληθυσμού και, επομένως, θεωρείται εσφαλμένο διάστημα επειδή δεν περιλαμβάνει την τιμή του άγνωστου μέσου πληθυσμού.

Διάστημα εμπιστοσύνης για το μ βάσει του z

Για να βρείτε το προαναφερθέν διάστημα εμπιστοσύνης από 511,44 ως 554,56 για τον άγνωστο μέσο βαθμό σε εξετάσεις μαθηματικών όλων των ντόπιων πρωτοετών, χρησιμοποιείτε την παρακάτω γενική έκφραση:

Διάστημα εμπιστοσύνης για το μ (βάσει του z)

$$\overline{X} \pm (z_{εμπ})(\sigma_{\overline{X}}) \qquad (12.1)$$

όπου το $\overline{X}$ αναπαριστά τον δειγματικό μέσο, το $z_{εμπ}$ αναπαριστά έναν αριθμό από τον γνωστό πίνακα κανονικής κατανομής που ικανοποιεί τις προδιαγραφές για το διάστημα εμπιστοσύνης και το $\sigma_{\overline{X}}$ αναπαριστά το τυπικό σφάλμα του μέσου.

Δεδομένου ότι το $\overline{X}$, ο δειγματικός μέσος για τους βαθμούς σε εξετάσεις μαθηματικών SAT, ισούται με 533, τότε το $z_{εμπ}$ ισούται με 1,96 (από τους πίνακες τυποποιημένης κανονικής κατανομής, όπου οι z-τιμές $\pm 1{,}96$ ορίζουν το μεσαίο 95% του χωρίου που βρίσκεται κάτω από την καμπύλη κανονικής κατανομής) και το τυπικό σφάλμα, $\sigma_{\overline{X}}$, ισούται με τον 11, και τότε ο Τύπος 12.1 γίνεται

$$533 \pm (1{,}96)(11) = 533 \pm 21{,}56 = \begin{cases} 554{,}56 \\ 511{,}44 \end{cases}$$

όπου τα 554,56 και 511,44 είναι το άνω και το κάτω όριο του διαστήματος εμπιστοσύνης. Μπορούμε τώρα να ισχυριστούμε ότι, με 95% βεβαιότητα, το διάστημα μεταξύ 511,44 και 554,56 περιλαμβάνει την τιμή του άγνωστου μέσου βαθμού σε εξετάσεις μαθηματικών για όλους τους ντόπιους πρωτοετείς.

Δύο υποθέσεις

Η χρήση του Τύπου 12.1 για την κατασκευή διαστημάτων εμπιστοσύνης προϋποθέτει ότι η τυπική απόκλιση του πληθυσμού είναι γνωστή και ότι ο πληθυσμός είναι κανονικός ή ότι το μέγεθος του δείγματος είναι επαρκώς μεγάλο –τουλάχιστον 25– για να ικανοποιηθούν οι προϋποθέσεις του κεντρικού οριακού θεωρήματος.

Έλεγχος προόδου *12.2 Παίρνουμε τους βαθμούς κατανόησης κειμένου από μια ομάδα μαθητών της τετάρτης δημοτικού. Ο βαθμός 4,0 δείχνει επίπεδο κατάλληλο για την τετάρτη δημοτικού, ένας βαθμός κάτω από 4,0 δείχνει επίπεδο κατώτερο απ' αυτό που απαιτείται για την τετάρτη δημοτικού και ένας βαθμός πάνω από 4,0 δείχνει ένα επίπεδο ανώτερο. Θα υποθέσουμε ότι η τυπική απόκλιση πληθυσμού ισούται με 0,4. Ένα τυχαίο δείγμα 64 μαθητών τετάρτης δημοτικού αποκαλύπτει μέσο βαθμό 3,82.

(α) Κατασκευάστε ένα διάστημα εμπιστοσύνης 95% για τον άγνωστο μέσο πληθυσμού. (Μην ξεχάσετε να μετατρέψετε την τυπική απόκλιση σε τυπικό σφάλμα.)

(β) Ερμηνεύστε αυτό το διάστημα εμπιστοσύνης. Με άλλα λόγια, υπάρχουν στιβαρές αποδείξεις υπερεπάρκειας ή ελλιπούς επάρκειας των μαθητών;

Απαντήσεις στη σελίδα 545.

12.3 Ερμηνεία διαστήματος εμπιστοσύνης

Ένας ισχυρισμός βεβαιότητας 95% αποτυπώνει μια μακροχρόνια αποτίμηση απόδοσης για μια εκτεταμένη σειρά διαστημάτων εμπιστοσύνης. Αν μια σειρά διαστημάτων εμπιστοσύνης κατασκευάζεται για να εκτιμηθεί ο ίδιος μέσος πληθυσμού, όπως στο Σχήμα 12.2, περίπου το 95% αυτών των διαστημάτων θα πρέπει να περιλαμβάνει τον μέσο πληθυσμού. Στην πράξη κατασκευάζεται μόνο ένα διάστημα εμπιστοσύνης και όχι μια σειρά διαστημάτων, και αυτό το ένα διάστημα είναι σωστό ή λάθος, επειδή είτε περιλαμβάνει τον μέσο πληθυσμού είτε δεν τον περιλαμβάνει. Βέβαια, *ποτέ δεν γνωρίζουμε πραγματικά αν ένα συγκεκριμένο διάστημα εμπιστοσύνης είναι σωστό ή λάθος* παρά μόνο αν διεξαγάγουμε έρευνα σε όλο τον πληθυσμό. Ωστόσο,

> **όταν το επίπεδο εμπιστοσύνης ισούται με 95% ή περισσότερο, μπορούμε να είμαστε *εύλογα βέβαιοι* ότι το ένα παρατηρούμενο διάστημα εμπιστοσύνης περιλαμβάνει τον πραγματικό μέσο πληθυσμού.**

Για παράδειγμα, μπορούμε να είμαστε *εύλογα βέβαιοι* ότι ο πραγματικός μέσος βαθμός του πληθυσμού σε εξετάσεις μαθηματικών για όλους τους ντόπιους πρωτοετείς δεν είναι ούτε μικρότερος από 511,44 ούτε μεγαλύτερος από 554,56. Ομοίως θα μπορούσαμε να πούμε ότι είμαστε *εύλογα βέβαιοι* ότι ο πραγματικός μέσος πληθυσμού για όλους τους ντόπιους πρωτοετείς είναι μεταξύ 511,44 και 554,56.

Έλεγχος προόδου *12.3 Πριν από τις εξετάσεις GRE, ένα τυχαίο δείγμα τελειοφοίτων πανεπιστημίου παρακολούθησε ειδικά μαθήματα γι' αυτές. Αφού ένας ερευνητής ανέλυσε τα αποτελέσματά τους στις εξετάσεις GRE, ανέφερε εντυπωσιακή αύξηση, σε σχέση με τον εθνικό μέσο όρο 500, όπως υποδεικνύεται από ένα διάστημα εμπιστοσύνης 95% μεταξύ 507 και 527. Οι παρακάτω ερμηνείες είναι σωστές ή λάθος;

(α) Περίπου το 95% όλων των αντικειμένων βαθμολογήθηκε από 507 ως 527.
(β) Το διάστημα από 507 ως 527 αναφέρεται σε πιθανές τιμές του μέσου πληθυσμού όλων των φοιτητών που παρακολούθησαν ειδικά μαθήματα.
(γ) Ο πραγματικός μέσος πληθυσμού είναι αναμφίβολα μεταξύ 507 και 527.
(δ) Αυτό το συγκεκριμένο διάστημα περιγράφει τον μέσο πληθυσμού περίπου στο 95% των φορών.
(ε) Στην πράξη, δεν γνωρίζουμε ποτέ πραγματικά αν το διάστημα από 507 ως 527 είναι σωστό ή λάθος.
(στ) Μπορούμε να είμαστε εύλογα βέβαιοι ότι ο μέσος πληθυσμού είναι μεταξύ 507 και 527.

Απαντήσεις στη σελίδα 545.

12.4 Επίπεδο εμπιστοσύνης

Επίπεδο εμπιστοσύνης
Το ποσοστό των περιπτώσεων στις οποίες μια σειρά διαστημάτων εμπιστοσύνης περιλαμβάνει το άγνωστο χαρακτηριστικό του πληθυσμού, όπως τον μέσο πληθυσμού.

Το ***επίπεδο εμπιστοσύνης*** *υποδεικνύει το ποσοστό των περιπτώσεων στις οποίες μια σειρά διαστημάτων εμπιστοσύνης περιλαμβάνει το άγνωστο χαρακτηριστικό του πληθυσμού, όπως τον μέσο πληθυσμού.*

Οποιοδήποτε επίπεδο εμπιστοσύνης μπορεί να αντιστοιχηθεί σε ένα διάστημα εμπιστοσύνης με απλή αντικατάσταση του $z_{εμπ}$ με μια κατάλληλη τιμή στον Τύπο 12.1. Για παράδειγμα, για να κατασκευάσετε ένα διάστημα εμπιστοσύνης 99% από τα δεδομένα για τους βαθμούς σε εξετάσεις μαθηματικών SAT, θα συμβουλευτείτε πρώτα τον Πίνακα Α στο Παράρτημα Γ προκειμένου να επαληθεύσετε ότι οι τιμές του $z_{εμπ}$ $\pm 2{,}58$ ορίζουν το μεσαίο 99% όλης της περιοχής κάτω από την καμπύλη κανονικής κατανομής. Έπειτα θα αντικαταστήσετε με αριθμούς όλα τα σύμβολα του Τύπου 12.1 για να πάρετε

$$533 \pm (2{,}58)(11) = 533 \pm 28{,}38 = \begin{cases} 561{,}38 \\ 504{,}62 \end{cases}$$

Θα μπορούσαμε να πούμε, με *99% βεβαιότητα,* ότι το διάστημα μεταξύ 504,62 και 561,38 περιλαμβάνει την τιμή του άγνωστου μέσου βαθμού σε εξετάσεις μαθηματικών για όλους τους ντόπιους πρωτοετείς. Αυτό σημαίνει ότι, μακροπρόθεσμα, το 99% αυτών των διαστημάτων εμπιστοσύνης θα περιλαμβάνει τον άγνωστο μέσο πληθυσμού.

Επίδραση στο πλάτος του διαστήματος

Παρατηρήστε ότι το διάστημα εμπιστοσύνης 99% μεταξύ 504,62 και 561,38 είναι πλατύτερο και, επομένως, λιγότερο ακριβές από το αντίστοιχο διάστημα 95% μεταξύ 511,44 και 554,56. Η μετατόπιση από το 95% στο 99% ως επίπεδο εμπιστοσύνης απαιτεί αύξηση στην τιμή του $z_{εμπ}$ από 1,96 σε 2,58. Αυτή η αύξηση προκαλεί με τη σειρά της ένα πλατύτερο και λιγότερο ακριβές διάστημα εμπιστοσύνης. Οποιαδήποτε μετατόπιση σε υψηλότερο επίπεδο εμπιστοσύνης παράγει πάντα πλατύτερο και λιγότερο ακριβές διάστημα εμπιστοσύνης, εκτός αν συνοδεύεται από μια αύξηση στο μέγεθος του δείγματος, όπως αναφέρουμε στην επόμενη ενότητα.

Επιλογή επιπέδου εμπιστοσύνης

Αν και έχουν χρησιμοποιηθεί πολλά διαφορετικά επίπεδα εμπιστοσύνης, τα 95% και 99% είναι τα πιο δημοφιλή. Γενικά, ένα μεγαλύτερο επίπεδο εμπιστοσύνης, όπως 99%, θα πρέπει να χρησιμοποιείται σε περιπτώσεις όπου ένα εσφαλμένο διάστημα θα μπορούσε να έχει ιδιαίτερα δριμείες συνέπειες, όπως η αποτυχία ενός δημοσκόπου σε μια δημοσκόπηση της κοινής γνώμης να προβλέψει τον νικητή των προεδρικών εκλογών.

12.5 Επίδραση μεγέθους δείγματος

Όσο μεγαλύτερο είναι το μέγεθος δείγματος, τόσο μικρότερο θα είναι το τυπικό σφάλμα και, επομένως, τόσο πιο ακριβές (στενότερο) θα είναι το διάστημα εμπιστοσύνης. Πράγματι, όσο μεγαλώνει το μέγεθος δείγματος, το τυπικό σφάλμα θα προσεγγίζει το μηδέν και το διάστημα εμπιστοσύνης θα συρρικνώνεται σε μια εκτίμηση σημείου. Υπό αυτήν την προοπτική, το μέγεθος δείγματος για ένα διάστημα εμπιστοσύνης, αντίθετα από έναν στατιστικό έλεγχο υποθέσεων, δεν μπορεί ποτέ να είναι τόσο μεγάλο.

Επιλογή μεγέθους δείγματος

Όπως συμβαίνει με τους στατιστικούς ελέγχους υποθέσεων, το μέγεθος δείγματος μπορεί να επιλεχθεί σύμφωνα με προδιαγραφές που καθορίζονται πριν από την έρευνα. Για να παραχθεί ένα διάστημα εμπιστοσύνης το οποίο έχει την επιθυμητή ακρίβεια (πλάτος) και συνάδει επίσης με το επιθυμητό επίπεδο εμπιστοσύνης, θα πρέπει να χρησιμοποιούνται τύποι που αφορούν το μέγεθος δείγματος και παρουσιάζονται σε άλλα βιβλία στατιστικής.[28] Για τη σωστή χρήση αυτών των τύπων θα πρέπει πριν από την έρευνα να είναι γνωστή ή να έχει εκτιμηθεί η τυπική απόκλιση πληθυσμού.

Έλεγχος προόδου *12.4 Με βάση ένα τυχαίο δείγμα 120 ενηλίκων, ένας δημοσκόπος αναφέρει, με 95% βεβαιότητα, ότι στο διάστημα μεταξύ 58% και 72% όλοι οι Αμερικανοί πιστεύουν στη ζωή μετά θάνατον.

(α) Αν αυτό το διάστημα είναι πολύ πλατύ, τι μπορούμε να κάνουμε, αν μπορούμε να κάνουμε κάτι, με τα υπάρχοντα δεδομένα προκειμένου να πάρουμε ένα στενότερο διάστημα εμπιστοσύνης;

(β) Τι μπορούμε να κάνουμε για να πάρουμε ένα στενότερο διάστημα εμπιστοσύνης 95% αν προγραμματίζεται άλλη παρόμοια έρευνα;

Απαντήσεις στη σελίδα 545.

12.6 Στατιστικοί έλεγχοι υποθέσεων ή διαστήματα εμπιστοσύνης;

Σε κανονικές συνθήκες, τα δεδομένα χρησιμοποιούνται είτε για να ελεγχθεί μια υπόθεση είτε για να κατασκευαστεί ένα διάστημα εμπιστοσύνης, αλλά όχι και για τα δύο. Οι στατιστικοί έλεγχοι υποθέσεων συνήθως προτιμώνται από τα διαστήματα εμπιστοσύνης στις επιστήμες της συμπεριφοράς, και αυτή η έμφαση εκφράζεται επίσης σ' αυτό το βιβλίο. Στην πραγματικότητα όμως, τα *διαστήματα εμπιστοσύνης τείνουν να προσφέρουν περισσότερες πληροφορίες από τους στατιστικούς ελέγχους υποθέσεων.*

Οι στατιστικοί έλεγχοι υποθέσεων απλώς δείχνουν αν υπάρχει ή όχι επίδραση, ενώ τα διαστήματα εμπιστοσύνης δείχνουν το πιθανό μέγεθός της.

Για το πείραμα της βιταμίνης C του Κεφαλαίου 11, ένας στατιστικός έλεγχος υποθέσεων απλώς δείχνει αν η βιταμίνη C επηρεάζει τους δείκτες IQ ή όχι, ενώ ένα διάστημα εμπιστοσύνης 95% δείχνει το πιθανό μέγεθος της επίδρασης της βιταμίνης C στους δείκτες IQ. Για παράδειγμα, θα μπορούσαμε να ισχυριστούμε, με 95% βεβαιότητα, ότι το διάστημα μεταξύ 102 και 112 περιλαμβάνει το πραγματικό μέσο IQ του πληθυσμού για μαθητές που λαμβάνουν βιταμίνη C. Με άλλα λόγια, η πραγματική επίδραση της βιταμίνης C βρίσκεται προφανώς κάπου ανάμεσα στις 2 και 12 μονάδες IQ (πάνω από την τιμή της μηδενικής υπόθεσης 100).

---

28. Για παράδειγμα, διαβάστε την Ενότητα 17.8 στο βιβλίο των King, B. M., & Minium, E. W. (2008). *Statistical Reasoning in the Behavioral Sciences* (5η έκδοση). Hoboken, NJ: Wiley.

Πότε χρησιμοποιούνται τα διαστήματα εμπιστοσύνης

Αν η πρωταρχική ανησυχία είναι αν υπάρχει επίδραση ή όχι –όπως συμβαίνει συχνά σε σχετικά νέους τομείς έρευνας–, χρησιμοποιείτε στατιστικό έλεγχο υποθέσεων. Για παράδειγμα, δεδομένου ότι ένας κοινωνικός ψυχολόγος δεν είναι βέβαιος σχετικά με το αν η κατανάλωση αλκοόλ από μάρτυρες αυξάνει το πλήθος των ανακριβειών στην κατάθεσή τους για μια ληστεία, θα ήταν καλύτερο να χρησιμοποιήσει έναν στατιστικό έλεγχο υποθέσεων. Διαφορετικά, δεδομένου ότι προηγούμενες έρευνες αποκαλύπτουν με σαφήνεια ότι οι ανακρίβειες στις καταθέσεις μαρτύρων αυξάνονται με την κατανάλωση αλκοόλ, ένας νέος ερευνητής θα μπορούσε να χρησιμοποιήσει ένα διάστημα εμπιστοσύνης για να εκτιμήσει τον πιθανό μέσο αριθμό αυτών των ανακριβειών.

Πράγματι, θα μπορούσατε να εξετάσετε την περίπτωση χρήσης ενός διαστήματος εμπιστοσύνης όταν ο στατιστικός έλεγχος υποθέσεων οδηγεί στην απόρριψη της μηδενικής υπόθεσης. Για παράδειγμα, σχετικά με το πείραμα της βιταμίνης C που μελετήσαμε στο Κεφάλαιο 11, αφού εξακριβωθεί (μέσω απόρριψης της μηδενικής υπόθεσης) ότι η βιταμίνη C επιδρά στους δείκτες IQ, θα είχε νόημα να εκτιμήσουμε, με διάστημα εμπιστοσύνης 95%, ότι το διάστημα μεταξύ 102 και 112 περιγράφει το πιθανό μέγεθος αυτής της επίδρασης, δηλαδή μια αύξηση (πάνω από 100) μεταξύ 2 και 12 μονάδων IQ.

12.7 Διάστημα εμπιστοσύνης για ποσοστό πληθυσμού

Περιθώριο σφάλματος
Η ποσότητα που προστίθεται και αφαιρείται από κάποια τιμή δείγματος, όπως μια αναλογία δείγματος ή ένας δειγματικός μέσος, για να ληφθούν τα όρια ενός διαστήματος εμπιστοσύνης.

Θα περιγράψουμε εν συντομία έναν τύπο διαστήματος εμπιστοσύνης –για ποσοστά ή αναλογίες πληθυσμού– που συναντούμε συχνά στα μέσα ενημέρωσης. Για παράδειγμα, ένα πρόσφατο δελτίο Τύπου ανέφερε ότι, μεταξύ ενός τυχαίου ή «επιστημονικού» δείγματος 1.500 ενήλικων Αμερικανών, το 64% ήταν υπέρ κάποιας μορφής θανατικής ποινής. Επιπλέον, το **περιθώριο σφάλματος** ισούται με ±3%, δεδομένου ότι θέλουμε να είμαστε 95% βέβαιοι για τα αποτελέσματά μας. Με λίγο διαφορετικά λόγια, θα μπορούσαμε να ισχυριστούμε, με 95% βεβαιότητα, ότι το διάστημα μεταξύ 61% και 67% (64 ± 3) περιλαμβάνει το πραγματικό ποσοστό των Αμερικανών που εγκρίνουν κάποιας μορφής θανατική ποινή.

Ουσιαστικά, αυτό το διάστημα εμπιστοσύνης 95% προέρχεται απ' αυτήν την έκφραση:

ποσοστό δείγματος ± (1,96) (*τυπικό σφάλμα του ποσοστού*)

όπου το 1,96 προέρχεται από την τυποποιημένη καμπύλη κανονικής κατανομής και το τυπικό σφάλμα του ποσοστού είναι ανάλογο με το τυπικό σφάλμα του μέσου.[29] Διαφορετικά, όλα τα προηγούμενα σχόλια σχετικά με τα διαστήματα εμπιστοσύνης για τους μέσους πληθυσμού ισχύουν για τα διαστήματα εμπιστοσύνης για ποσοστά ή αναλογίες πληθυσμού. Ως εκ τούτου, στην προκειμένη περίπτωση, μπορούμε να είμαστε εύλογα βέβαιοι ότι το πραγματικό ποσοστό του πληθυσμού είναι μεταξύ 61% και 67%.

Μέγεθος δείγματος και περιθώριο σφάλματος

Όπως βλέπουμε συχνά σε δημοσκοπήσεις πανεθνικού βεληνεκούς, το τεράστιο δείγμα των 1.500 Αμερικανών μειώνει το μέγεθος του τυπικού σφάλματος και, επομένως, εγγυάται ένα σχετικά μικρό περιθώριο σφάλματος ±3%. Αν, σύμφωνα με την κρίση του δημοσκόπου, ήταν ανεκτό ένα μεγαλύτερο περιθώριο σφάλματος, θα μπορούσαν να χρησιμοποιηθούν μικρότερα δείγματα. Για παράδειγμα, αν ένα μεγαλύτερο περιθώριο σφάλματος ±5% ήταν ανεκτό, θα μπορούσε να χρησιμοποιηθεί ένα τυχαίο δείγμα περίπου 500 ενηλίκων, ενώ, αν ήταν ανεκτό ένα ακόμα μεγαλύτερο περιθώριο σφάλματος ±10%, θα μπορούσε να χρησιμοποιηθεί ένα τυχαίο δείγμα μόλις 100 ενηλίκων.

Οι δημοσκόποι χρησιμοποιούν μεγαλύτερα δείγματα

Σε κάθε περίπτωση, οι δημοσκόποι χρησιμοποιούν συχνά δείγματα με εκατοντάδες ή ακόμα και χιλιάδες συμ-

29. Μια αναλογία (ή ένα ποσοστό, το οποίο είναι ουσιαστικά η αναλογία επί 100) είναι ένας ειδικός τύπος μέσου όπου, μετά την κωδικοποίηση όλων των παρατηρήσεων ως 0 ή 1, τα 1 προστίθενται και διαιρούνται διά του συνολικού πλήθους των παρατηρήσεων. Επομένως, αν και δεν τονίζεται σ' αυτό το βιβλίο, το τυπικό σφάλμα της αναλογίας (ή του ποσοστού) θα μπορούσε να ληφθεί από τον τύπο για το τυπικό σφάλμα του μέσου.

μετέχοντες προκειμένου να παραγάγουν στενότερα και ακριβέστερα διαστήματα εμπιστοσύνης. Σε σχέση με τα πολύ μικρότερα δείγματα των περισσότερων ερευνητών, αυτά τα μεγαλύτερα δείγματα αποτυπώνουν πολλούς παράγοντες, όπως τη σχετική ευτέλεια των παρατηρήσεων, οι οποίες μπορεί να παράγονται με τυχαία τηλεφωνήματα, καθώς και το γεγονός ότι τα δείγματα δεν μπορεί να είναι ποτέ πολύ μεγάλα σε έρευνες – αν και μπορεί να είναι πολύ μεγάλα σε πειράματα, όπως είδαμε στην Ενότητα 11.10.

Ένα τελευταίο σημείο προσοχής

Όταν βασιζόμαστε σε τυχαία επιλεγμένους συμμετέχοντες σε έρευνα, τα διαστήματα εμπιστοσύνης αντανακλούν μόνο ένα είδος σφάλματος – το *στατιστικό* σφάλμα που οφείλεται στην τυχαία μεταβλητότητα δειγματοληψίας. Υπάρχουν κι άλλα είδη *μη στατιστικών* σφαλμάτων που θα μπορούσαν να θέσουν υπό αμφισβήτηση την τιμή ενός διαστήματος εμπιστοσύνης. Για παράδειγμα, η προηγούμενη εκτίμηση ότι το 61% ως 67% όλων των Αμερικανών είναι υπέρ της θανατικής ποινής μπορεί να είχε διογκωθεί από την προσθήκη σε μια ουδέτερη ερώτηση, όπως «Είστε υπέρ της θανατικής ποινής;», μιας μεροληπτικής φράσης, όπως «σε σχέση με την πρόσφατη έξαρση των φόνων αθώων παιδιών;». Ενδεχομένως επίσης το παραπάνω διάστημα να μην είχε καταφέρει να αποτυπώσει τον στοχευμένο πληθυσμό όλων των ενήλικων Αμερικανών, επειδή το τυχαίο δείγμα στην πραγματικότητα διερευνά έναν σχετικά περιορισμένο πληθυσμό. Για παράδειγμα, μπορεί να υπάρχει σημαντικός αριθμός ανθρώπων που δεν συμμετέχουν στην έρευνα των οποίων η άποψη για τη θανατική ποινή διαφέρει αισθητά από τη στάση όσων απάντησαν. Όταν δεν υπάρχουν τέτοιες σημαντικές πληροφορίες, οι αναφορές των διαστημάτων εμπιστοσύνης θα πρέπει να ερμηνεύονται προσεκτικά.

Έλεγχος προόδου *12.5 Σε ένα πρόσφατο επιστημονικό δείγμα περίπου 900 ενήλικων Αμερικανών, το 70% ήταν υπέρ του αυστηρότερου ελέγχου όπλων, με περιθώριο σφάλματος ±4% για ένα διάστημα εμπιστοσύνης 95%. Επομένως, το διάστημα εμπιστοσύνης 95% ισούται με 66% ως 74%. Διευκρινίστε αν οι παρακάτω ερμηνείες είναι σωστές ή λάθος:

(α) Το διάστημα από 66% ως 74% αναφέρεται σε πιθανές τιμές του ποσοστού του δείγματος.
(β) Το πραγματικό ποσοστό του πληθυσμού είναι μεταξύ 66% και 74%.
(γ) Μακροπρόθεσμα, μια σειρά διαστημάτων παρόμοιων μ' αυτό δεν θα μπορούσε να συμπεριλαμβάνει το ποσοστό του πληθυσμού περίπου στο 5% των περιπτώσεων.
(δ) Μπορούμε να είμαστε εύλογα βέβαιοι ότι το ποσοστό του πληθυσμού είναι μεταξύ 66% και 74%.
Απαντήσεις στη σελίδα 545.

Άλλοι τύποι διαστημάτων εμπιστοσύνης

Τα διαστήματα εμπιστοσύνης μπορούν να κατασκευαστούν για μέσους και ποσοστά πληθυσμού, αλλά και για διαφορές μεταξύ δύο μέσων πληθυσμού, όπως θα δούμε σε επόμενα κεφάλαια. Αν και δεν αναφέρεται σ' αυτό το βιβλίο, τα διαστήματα εμπιστοσύνης μπορούν επίσης να κατασκευαστούν για άλλα χαρακτηριστικά πληθυσμών, όπως οι συντελεστές διασποράς και συσχέτισης.

Περίληψη

Αντί να ελέγξετε μια υπόθεση για έναν μόνο μέσο πληθυσμού, θα μπορούσατε να επιλέξετε να εκτιμήσετε αυτό το χαρακτηριστικό του πληθυσμού χρησιμοποιώντας μια εκτίμηση σημείου ή ένα διάστημα εμπιστοσύνης.

Στην εκτίμηση σημείου, ένα μόνο χαρακτηριστικό του δείγματος, όπως ένας δειγματικός μέσος, εκτιμά το αντίστοιχο χαρακτηριστικό του πληθυσμού. Οι εκτιμήσεις σημείου παραβλέπουν τη μεταβλητότητα δειγματοληψίας και, επομένως, τείνουν να μην είναι ακριβείς.

Τα διαστήματα εμπιστοσύνης καθορίζουν εύρη τιμών τα οποία, μακροπρόθεσμα, περιλαμβάνουν το άγνωστο χαρακτηριστικό του πληθυσμού, όπως τον μέσο, για ένα συγκεκριμένο ποσοστό των φορών. Για παράδειγμα, δεδομένου ενός διαστήματος εμπιστοσύνης 95%, τότε, μακροπρόθεσμα, περίπου το 95% όλων αυτών των διαστημάτων εμπιστοσύνης είναι πραγματικό επειδή περιλαμβάνει την άγνωστη τιμή του χαρακτηριστικού του πληθυσμού. Τα διαστήματα εμπιστοσύνης λειτουργούν επειδή είναι προϊόντα κατανομών δειγματοληψίας.

Οποιοδήποτε επίπεδο εμπιστοσύνης μπορεί να αντιστοιχηθεί με ένα διάστημα εμπιστοσύνης, αλλά συνήθως χρησιμοποιούνται τα επίπεδα 95% και 99%. Δεδομένου ενός απ' αυτά τα επίπεδα εμπιστοσύνης, ακόμα κι αν δεν μπορούμε ποτέ να γνωρίζουμε αν ένα συγκεκριμένο διάστημα εμπιστοσύνης είναι σωστό ή λάθος, μπορούμε να είμαστε εύλογα βέβαιοι ότι ένα συγκεκριμένο διάστημα όντως περιλαμβάνει την άγνωστη τιμή του χαρακτηριστικού του πληθυσμού.

Στενότερα και ακριβέστερα διαστήματα εμπιστοσύνης παράγονται από χαμηλότερα επίπεδα εμπιστοσύνης (για παράδειγμα, 95% αντί για 99%) και από μεγαλύτερα μεγέθη δειγμάτων.

Τα διαστήματα εμπιστοσύνης τείνουν να προσφέρουν περισσότερες πληροφορίες από τους στατιστικούς ελέγχους υποθέσεων. Οι έλεγχοι υποθέσεων απλώς υποδεικνύουν αν υπάρχει ή όχι κάποια επίδραση, ενώ τα διαστήματα εμπιστοσύνης δείχνουν το πιθανό μέγεθος της επίδρασης. Όποτε κρίνεται απαραίτητο –όπως κάθε φορά που έχει απορριφθεί η μηδενική υπόθεση–, μπορείτε να εξετάζετε τη χρήση διαστημάτων εμπιστοσύνης.

Τα διαστήματα εμπιστοσύνης για ποσοστά ή αναλογίες πληθυσμού είναι παρόμοια, τόσο ως προς την προέλευση όσο και ως προς την ερμηνεία, με τα διαστήματα εμπιστοσύνης για μέσους πληθυσμού.

Σημαντικοί όροι

Εκτίμηση σημείου
Διάστημα εμπιστοσύνης (CI)
Επίπεδο εμπιστοσύνης
Περιθώριο σφάλματος

Κύρια εξίσωση

ΔΙΑΣΤΗΜΑ ΕΜΠΙΣΤΟΣΥΝΗΣ

$$CI = \overline{X} \pm (z_{εμπ})(\sigma_{\overline{X}})$$

Ερωτήσεις επανάληψης

12.6 (Σωστό ή Λάθος) Θα πρέπει να εξετάζετε τη χρήση διαστήματος εμπιστοσύνης όταν
(α) η μηδενική υπόθεση έχει απορριφθεί.
(β) το θέμα είναι αν υπάρχει ή όχι επίδραση.
(γ) το θέμα περιλαμβάνει πιθανά μεγέθη επίδρασης.
(δ) δεν υπάρχει ωφέλιμη μηδενική υπόθεση.

***12.7** Στην Ερώτηση 10.5 στη σελίδα 263, συμπεράναμε ότι ο μέσος μισθός στον πληθυσμό γυναικών μελών της Αμερικανικής Ένωσης Ψυχολόγων είναι μικρότερος από τον αντίστοιχο ($82.500) όλων των συγκρίσιμων μελών που έχουν διδακτορικό τίτλο και εργάζονται με πλήρες ωράριο.
(α) Δεδομένης της τυπικής απόκλισης πληθυσμού $6.000 και του δειγματικού μέσου του μισθού $80.100 για ένα τυχαίο δείγμα 100 γυναικών μελών, κατασκευάστε ένα διάστημα εμπιστοσύνης 99% για τον μέσο μισθό για όλες τις γυναίκες μέλη της ένωσης.
(β) Δεδομένου αυτού του διαστήματος εμπιστοσύνης, υπάρχει αξιόπιστη απόδειξη ότι ο μέσος μισθός για όλες τις γυναίκες μέλη είναι μικρότερος από $82.500, τον μέσο μισθό όλων των μελών;
Απαντήσεις στη σελίδα 545.

12.8 Στην Ερώτηση επανάληψης 11.12 στη σελίδα 292, αντί για στατιστικό έλεγχο υπόθεσης, ενδεχομένως να προτιμούσατε να κατασκευάσετε ένα διάστημα εμπιστοσύνης για το μέσο βάρος όλων των συσκευασιών 2 λιβρών κατά τη διάρκεια μιας πρόσφατης αλλαγής στην παραγωγή.
(α) Δεδομένης της τυπικής απόκλισης πληθυσμού 0,30 ουγγιών και του δειγματικού μέσου του βάρους (33,09 ουγγιές) για ένα τυχαίο δείγμα 36 συσκευασιών γλυκών, κατασκευάστε ένα διάστημα εμπιστοσύνης 95%.

(β) Ερμηνεύστε αυτό το διάστημα, δεδομένης της επιθυμίας του κατασκευαστή να παραγάγει συσκευασίες γλυκών που είναι βαρύτερες, κατά μέσο όρο, από 32 ουγγιές.

12.9 Είναι δελεαστικό να ισχυριστούμε ότι, αφού έχει κατασκευαστεί ένα συγκεκριμένο διάστημα εμπιστοσύνης 95%, περιλαμβάνει το άγνωστο χαρακτηριστικό του πληθυσμού με πιθανότητα 0,95. Τι λάθος έχει αυτός ο ισχυρισμός;

***12.10** Έστω ότι ένα από τα παρακάτω διαστήματα εμπιστοσύνης 95% εκτιμά την επίδραση της βιταμίνης C στο IQ:

Διάστημα εμπιστοσύνης 95%	Κάτω όριο	Άνω όριο
1	100	102
2	95	99
3	102	106
4	90	111
5	91	98

(α) Ποιο υποστηρίζει πιο έντονα το συμπέρασμα ότι η βιταμίνη C *αυξάνει* το IQ;
(β) Ποιο υποδηλώνει το μεγαλύτερο μέγεθος δείγματος;
(γ) Ποιο υποστηρίζει πιο έντονα το συμπέρασμα ότι η βιταμίνη C *μειώνει* το IQ;
(δ) Ποιο είναι πιθανότερο να παρακινήσει τον ερευνητή να διεξαγάγει ένα επιπρόσθετο πείραμα με μεγαλύτερο μέγεθος δείγματος;

Απαντήσεις στη σελίδα 545.

12.11 Αντίθετα από τα διαστήματα εμπιστοσύνης, οι στατιστικοί έλεγχοι υποθέσεων απαιτούν τη χρήση κάποιας προκαθορισμένης τιμής του πληθυσμού για την εκτίμηση των τιμών των δειγμάτων. Μπορείτε να σκεφτείτε άλλες διαφορές μεταξύ στατιστικών ελέγχων υποθέσεων και διαστημάτων εμπιστοσύνης;

ΠΑΡΑΡΤΗΜΑ XII

Διαστήματα εμπιστοσύνης

Γιώργος Ανδρουλάκης

Π.XII.1 Εισαγωγή

Το παρόν Κεφάλαιο 12 συμπληρώνει τα δύο προηγούμενα κεφάλαια εστιάζοντας στα διαστήματα εμπιστοσύνης, τα οποία λειτουργούν συμπληρωματικά στον έλεγχο υπόθεσης. Αν θέλαμε με μία φράση να αναδείξουμε αυτήν τη συμπληρωματικότητα, θα λέγαμε ότι τα διαστήματα εμπιστοσύνης χρησιμοποιούνται για να κάνουμε αναγωγή ευρημάτων από το δείγμα προς τον πληθυσμό, ενώ ο έλεγχος υπόθεσης για να εξετάσουμε αν τα ευρήματα αυτά στο δείγμα είναι ικανά να κλονίσουν μια πεποίθησή μας για τον πληθυσμό.

Π.XII.2 Διάστημα εμπιστοσύνης (CI) για το μ

Στο προηγούμενο κεφάλαιο μελετήσαμε αναλυτικά τον τρόπο με τον οποίο με το λογισμικό R μπορούμε να εκτελέσουμε ελέγχους υποθέσεων με βάση την κανονική κατανομή. Στα παραδείγματα ελέγχου υπόθεσης που παρουσιάσαμε ο προσεκτικός αναγνώστης έχει διαπιστώσει ότι δεν χρειαστήκαμε όλη την πληροφορία που το στατιστικό πακέτο μάς έδωσε. Στο παρόν κεφάλαιο θα δούμε ότι αυτή η παρεχόμενη πληροφορία αφορά τα διαστήματα εμπιστοσύνης.

Αρχικά, στο πλαίσιο του πρώτου παραδείγματος, ας δημιουργήσουμε 100 τυχαίους ακέραιους αριθμούς με μέση τιμή 500 και τυπική απόκλιση 100 και ας τους καταχωρίσουμε στο διάνυσμα myx:

```
> myx = round(rnorm(100, 500, 100),0)
> myx
  [1] 351 300 306 567 467 356 441 455 385 644 218 467 520 407 509 514 536 519 579 399
 [21] 473 330 392 396 454 583 465 493 416 387 560 562 443 584 595 613 530 377 91 335
 [41] 586 511 571 608 497 649 523 679 599 438 528 515 626 462 391 665 491 588 507 674
 [61] 574 461 367 515 409 586 557 273 545 381 509 574 379 401 472 435 479 704 544 660
 [81] 521 532 656 490 531 339 497 496 458 518 578 570 633 502 523 606 590 477 462 526
```

Να τονίσουμε ότι, αν τρέξετε τις παραπάνω εντολές στον υπολογιστή σας, σίγουρα οι τυχαίοι αριθμοί που θα δημιουργηθούν θα είναι διαφορετικοί.

Έστω ότι θέλουμε να κατασκευάσουμε το διάστημα εμπιστοσύνης για το μέσο του διανύσματος myx σε επίπεδο σημαντικότητας 95%. Όπως αναφέραμε στο προηγούμενο κεφάλαιο, η εντολή που χρησιμοποιούμε για την κατασκευή διαστημάτων εμπιστοσύνης είναι η ίδια με αυτή για τον έλεγχο υποθέσεων. Αναφέραμε στο προηγούμενο κεφάλαιο ότι μια υλοποίηση της εντολής για έλεγχο υποθέσεων και διαστήματος εμπιστοσύνης με βάση την κανονική κατανομή έχει υλοποιηθεί στη βιβλιοθήκη BSDA της R. Η βιβλιοθήκη αυτή δεν είναι προεγκατεστημένη στην R, επομένως για να την εγκαταστήσουμε για πρώτη φορά πρέπει να τρέξουμε την ακόλουθη εντολή στον υπολογιστή μας, ο οποίος θα πρέπει να έχει ενεργή σύνδεση στο διαδίκτυο κατά την εκτέλεσή της:

```
> install.packages("BSDA")
```

Αφού έχουμε εξασφαλίσει ότι η βιβλιοθήκη είναι εγκατεστημένη στον υπολογιστή μας, για να μπορέσουμε να τη χρησιμοποιήσουμε πρέπει να την ενεργοποιήσουμε. Αυτό επιτυγχάνεται με την παρακάτω εντολή:

```
> library(BSDA)
```

Τώρα μπορούμε να εκτελέσουμε την εντολή z.test για να υπολογίσουμε το διάστημα εμπιστοσύνης για τον μέσο των 100 τυχαίων ακέραιων αριθμών που έχουμε καταχωρίσει στη μεταβλητή myx.

```
> z.test(myx, y=NULL, mu=500, sigma.x = 100)

          One-sample z-Test

data: myx
z = -0.543, p-value = 0.5871
alternative hypothesis: true mean is not equal to 500
95 percent confidence interval:
  474.9704 514.1696
sample estimates:
mean of x
    494.57
```

Για να ερμηνεύσουμε τα αποτελέσματα της παραπάνω εντολής, τώρα εστιάζουμε στις τελευταίες πέντε γραμμές του αποτελέσματος. Αρχικά, από τις δύο τελευταίες γραμμές ενημερωνόμαστε ότι η μέση τιμή του δείγματος των 100 τυχαίων ακέραιων αριθμών είναι 494,57. Το αντίστοιχο διάστημα εμπιστοσύνης σε επίπεδο σημαντικότητας 95% που κατασκευάζεται με βάση την κανονική κατανομή γύρω από τον μέσο αυτόν είναι το [474,9704, 514,1696].

Π.XII.3 Ερμηνεία διαστήματος εμπιστοσύνης

Είδαμε στη θεωρία ότι το εύρος του διαστήματος εμπιστοσύνης εξαρτάται από το επίπεδο σημαντικότητας και το μέγεθος του δείγματος. Όσο αυξάνεται το επίπεδο σημαντικότητας, μειώνεται δηλαδή η αβεβαιότητα, αυτό έχει ως αποτέλεσμα το εύρος του διαστήματος να μεγαλώνει, ενώ, αντίθετα, καθώς αυξάνεται το μέγεθος του δείγματος, αυτό έχει ως αποτέλεσμα το εύρος του διαστήματος εμπιστοσύνης να μικραίνει.

Παρακάτω θα δούμε μια πρακτική υλοποίηση των δύο παραπάνω εννοιών σε ένα γράφημα για να κατανοήσουμε και οπτικά τον ρόλο του επιπέδου σημαντικότητας και του μεγέθους του δείγματος.

Π.XII.4 Επίπεδο εμπιστοσύνης

Αρχικά, ας δημιουργήσουμε 100 τυχαίους ακέραιους αριθμούς με μέση τιμή 500 και τυπική απόκλιση 100 και ας τους καταχωρίσουμε στο διάνυσμα myx:

```
> myx = round(rnorm(100, 500, 100),0)
> myx
  [1] 586 662 424 616 537 470 556 620 417 410 496 618 447 474 502 457 285 410 550 625
 [21] 494 521 475 388 532 512 479 256 665 311 418 478 483 463 518 662 410 551 523 383
 [41] 404 465 678 566 369 433 659 578 266 459 579 619 609 290 510 532 543 447 527 561
 [61] 534 450 336 319 476 528 436 289 701 611 514 436 437 466 439 682 591 524 582 632
 [81] 370 572 227 470 626 300 513 740 511 495 550 588 660 521 514 544 397 406 433 590
```

Χρησιμοποιώντας το τέχνασμα που πρωτοαναφέρθηκε στο Κεφάλαιο 10, ας δημιουργήσουμε ένα κενό γράφημα με άξονες για το x από 0,01 έως 0,20 και για το y από 460 έως 540.

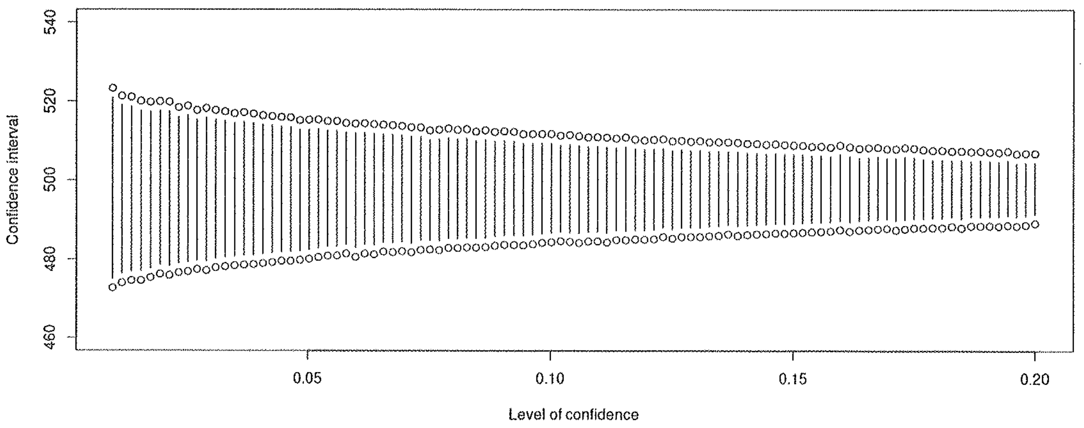
```
> plot(c(0.01,0.2),c(460,540),type="n",xlab="Level of confidence",
+     ylab="Confidence interval")
```

Ακολούθως, με τη βοήθεια ενός διπλού εμφωλευμένου βρόχου, θα υπολογίσουμε και θα οπτικοποιήσουμε το αποτέλεσμα της επίδρασης της αλλαγής του επιπέδου σημαντικότητας. Αρχικά, ο πρώτος βρόχος υλοποιείται με την πρώτη και την τελευταία γραμμή του παρακάτω κώδικα. Συγκεκριμένα, στο διάνυσμα k τοποθετούνται 100 ισαπέχουσες τιμές από 0,2 έως 0,01 (έχουμε κατά νου να απεικονίσουμε το εύρος του διαστήματος εμπιστοσύνης σε επίπεδο σημαντικότητας 80% έως 99%, άρα φτιάχνουμε 100 αριθμούς από 1–0,8 = 0,2 έως 1–0,99 = 0,01).

Για κάθε υλοποίηση θα χρειαστούμε 10.000 απεικονίσεις. Δημιουργούμε ένα διάνυσμα, το a, το οποίο έχει 10.000 θέσεις. Εκτελούμε έναν εσωτερικό βρόχο από το 1 έως το 10.000 προκειμένου να τοποθετήσουμε τιμές στο διάνυσμα a που μόλις κατασκευάσαμε. Σε κάθε θέση του διανύσματος αυτού τοποθετούμε τον μέσο ενός τυχαίου δείγματος με 100 επιλογές από το διάνυσμα myx, επιλογές που γίνονται με βάση την ομοιόμορφη κατανομή και έτσι ώστε να επιτρέπεται η επανατοποθέτηση. Τέλος, στο κενό γράφημα που δημιουργήσαμε με την εντολή points φτιάχνουμε ένα κάθετο ευθύγραμμο τμήμα για κάθε τιμή του k, το οποίο έχει άκρα το εκατοστημόριο που ορίζεται από το k έως το 1–k. Συνολικά, ο κώδικας που μόλις περιγράψαμε είναι ο παρακάτω:

```
> for (k in seq(0.2,0.01, length = 100)){
+  a<-numeric(10000)
+  for (i in 1:10000){
+   a[i]<-mean(sample(myx,100,replace=T))
+  }
+  points(c(k,k),quantile(a,c(k,1-k)),type="b")
+ }
```

Ο παραπάνω κώδικας έχει ως αποτέλεσμα το ακόλουθο γράφημα:



Παρατηρούμε ότι, καθώς αυξάνεται η αβεβαιότητα από 0,01 έως 0,20, μειώνεται δηλαδή το αντίστοιχο επίπεδο σημαντικότητας από 99% έως 80%, μειώνεται και το εύρος των μέσων των τυχαίων δειγμάτων σε αυτό το επίπεδο σημαντικότητας. Κοιτάζοντας το παραπάνω σχήμα από δεξιά προς τα αριστερά, διαπιστώνουμε και οπτικά αυτό που περιγράφει και η αντίστοιχη θεωρία, δηλαδή ότι, καθώς αυξάνεται το επίπεδο σημαντικότητας, αυξάνεται και το αντίστοιχο εύρος του διαστήματος.

Π.XII.5 Επίδραση μεγέθους δείγματος

Προκειμένου να διαπιστώσουμε την επίδραση του μεγέθους του δείγματος στο εύρος του διαστήματος, θα εκτελέσουμε ένα ανάλογο «πείραμα». Αρχικά, ας δημιουργήσουμε πάλι 100 τυχαίους ακέραιους αριθμούς με μέση τιμή 500 και τυπική απόκλιση 100, τους οποίους καταχωρίζουμε στο διάνυσμα myx:

```
> myx = round(rnorm(100, 500, 100),0)
> myx
  [1] 419 550 435 429 409 589 276 472 574 438 478 511 433 492 564 446 432 527 626 512
 [21] 542 394 592 359 509 513 492 672 473 662 553 518 228 341 610 500 566 380 388 457
 [41] 323 420 539 436 629 591 627 541 512 452 577 419 463 555 530 465 490 501 557 705
 [61] 544 428 508 394 463 507 665 427 431 440 592 432 642 536 536 551 510 373 327 551
 [81] 585 497 605 397 528 638 599 607 778 428 544 265 365 483 792 573 692 357 496 505
```

Χρησιμοποιώντας το ίδιο με προηγουμένως τέχνασμα, δημιουργούμε ένα κενό γράφημα με άξονες για το x από 0 έως 100 και για το y από 400 έως 600.

```
> plot(c(0,100),c(400,600),type="n",xlab="Sample size",
+     ylab="Confidence interval")
```

Ακολούθως, με τη βοήθεια ενός διπλού εμφωλευμένου βρόχου, θα υπολογίσουμε και θα οπτικοποιήσουμε το αποτέλεσμα της επίδρασης της αλλαγής του μεγέθους του δείγματος. Αρχικά, ο πρώτος βρόχος υλοποιείται με την πρώτη και την τελευταία γραμμή του παρακάτω κώδικα. Συγκεκριμένα, στο διάνυσμα k τοποθετούνται 96 ακέραιες τιμές, οι αριθμοί από το 5 έως το 100 (έχουμε κατά νου να αυξάνουμε σταδιακά το μέγεθος των δειγμάτων περιλαμβάνοντας πλήθος δεδομένων 5, 6, 7 έως 100). Για κάθε υλοποίηση θα χρειαστούμε 10.000 απεικονίσεις. Δημιουργούμε ένα διάνυσμα, το a, το οποίο έχει 10.000 θέσεις. Εκτελούμε έναν εσωτερικό βρόχο από το 1 έως το 10.000 προκειμένου να τοποθετήσουμε τιμές στο διάνυσμα a που μόλις κατασκευάσαμε. Σε κάθε θέση του διανύσματος αυτού τοποθετούμε τον μέσο ενός τυχαίου δείγματος με k επιλογές από το διάνυσμα myx, επιλογές που γίνονται με βάση την ομοιόμορφη κατανομή και έτσι ώστε να επιτρέπεται η επανατοποθέτηση. Τέλος, στο κενό γράφημα που δημιουργήσαμε με την εντολή points φτιάχνουμε ένα κάθετο ευθύγραμμο τμήμα για κάθε τιμή του k, το οποίο έχει άκρα το εκατοστημόριο που ορίζεται από το 2,5% και το 97,5%. Συνολικά, ο κώδικας που μόλις περιγράψαμε είναι ο παρακάτω:

```
> for (k in seq(5,100,1)){
+  a<-numeric(10000)
+  for (i in 1:10000){
+   a[i]<-mean(sample(myx,k,replace=T))
+  }
+  points(c(k,k),quantile(a,c(.025,.975)),type="b")
+ }
```

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)

ΚΕΦΑΛΑΙΟ

13 Έλεγχος *t* για ένα δείγμα

Πρόλογος

Τα τρία επόμενα κεφάλαια περιγράφουν διάφορους ελέγχους t. Όταν, όπως συμβαίνει συνήθως, η τυπική απόκλιση πληθυσμού είναι άγνωστη, η εκτίμησή της πρέπει να γίνεται με την τυπική απόκλιση δείγματος. Η εκτίμηση της άγνωστης τυπικής απόκλισης πληθυσμού έχει σημαντικές επιπλοκές που απαιτούν τόσο τη χρήση βαθμών ελευθερίας όσο και την αντικατάσταση του ελέγχου z με τον έλεγχο t. (Ίσως χρειαστεί να κάνετε μια επανάληψη στους βαθμούς ελευθερίας – βλ. Ενότητα 4.6.)

13.1 Έρευνα για την κατανάλωση καυσίμων

Ενδεχομένως η νομοθεσία (των ΗΠΑ εν προκειμένω) να ορίσει ότι τα καινούργια αυτοκίνητα πρέπει να τηρούν κανόνες κατανάλωσης καυσίμων και να διανύουν 45 μίλια ανά γαλόνι (mpg) βενζίνης. (Στην Ελλάδα ακολουθείται ένα άλλο πρότυπο μέτρησης κατανάλωσης –λίτρα ανά 100 χιλιόμετρα– αντί του αμερικανικού προτύπου. Εν προκειμένω, τα 45 μίλια ανά γαλόνι αντιστοιχούν σε κατανάλωση 6,3 λίτρων καυσίμων ανά 100 χιλιόμετρα. Σ' αυτό το βιβλίο τηρούμε τη σύμβαση του πρωτότυπου κειμένου, μετρώντας την κατανάλωση με το πρότυπο των μιλίων ανά γαλόνι.) Επειδή θα ήταν αδύνατο να ελεγχθούν όλα τα καινούργια αυτοκίνητα, οι έλεγχοι συμμόρφωσης θα βασίζονταν σε τυχαία δείγματα από το σύνολο της παραγωγής κάθε μοντέλου αυτοκινήτου. Αν ένας στατιστικός έλεγχος υποθέσεων υποδεικνύει επιδόσεις κάτω του προτύπου, επιβάλλεται ποινή στον κατασκευαστή – έστω \$200 ανά αυτοκίνητο του συνόλου της παραγωγής.

Σ' αυτούς τους ελέγχους, η μηδενική υπόθεση δηλώνει ότι, σε σχέση με τον υποχρεωτικό μέσο των 45 mpg, δεν συμβαίνει κάτι ιδιαίτερο στον πληθυσμό για κάποιο μοντέλο αυτοκινήτου – δηλαδή δεν υπάρχουν επιδόσεις κάτω του προτύπου και ο μέσος πληθυσμού ισούται ή είναι μεγαλύτερος από 45 mpg. Η εναλλακτική υπόθεση εκφράζει μια ανησυχία ότι ο μέσος πληθυσμού είναι μικρότερος από 45 mpg. Συμβολικά, οι δύο στατιστικές υποθέσεις έχουν ως εξής:

$$H_0: \mu \geq 45$$
$$H_1: \mu < 45$$

Από την πλευρά του κατασκευαστή, ένα σφάλμα τύπου I (μια αυστηρή ποινή, ακόμα κι αν το αυτοκίνητο τηρεί το πρότυπο) είναι πολύ σοβαρό. Αναλόγως, για να ελέγξουμε το σφάλμα τύπου I, θα χρησιμοποιήσουμε το επίπεδο σημαντικότητας 0,01 αντί του συνηθισμένου 0,05. Από την πλευρά των κανονισμών, ένα σφάλμα τύπου II (να μην τιμωρείται ο κατασκευαστής ακόμα κι αν το αυτοκίνητο δεν πληροί τις προδιαγραφές) είναι επίσης σοβαρό. Στην πράξη, θα πρέπει να επιλεχθεί ένα μέγεθος δείγματος, όπως περιγράφεται στην Ενότητα 11.11, για να γίνεται έλεγχος του σφάλματος τύπου II, δηλαδή για να εξασφαλίζεται ένα εύλογο ποσοστό ανίχνευσης για τη μικρότερη μείωση (που κρίνεται σημαντική) του πραγματικού μέσου του πληθυσμού κάτω από το υποχρεωτικό 45 mpg. Για να απλοποιήσουμε τους υπολογισμούς στο προκείμενο παράδειγμα όμως, ο προβλεπόμενος μονόπλευρος έλεγχος βασίζεται σε δεδομένα από ένα πολύ μικρό δείγμα μόλις έξι τυχαία επιλεγμένων αυτοκινήτων.

Για λόγους που θα αποκαλυφθούν σύντομα, ο έλεγχος *z* πρέπει να αντικατασταθεί από έναν νέο έλεγχο υπόθεσης, τον έλεγχο *t*. Αφιερώστε λίγα λεπτά για να εξοικειωθείτε με την περίληψη που παραθέτουμε στο πλαίσιο παρακάτω σχετικά με την έρευνα για την κατανάλωση καυσίμων, παρατηρώντας τις σημαντικές ομοιότητες μεταξύ αυτής και των προηγούμενων στατιστικών ελέγχων υποθέσεων με τον έλεγχο *z*.

Κατανομή δειγματοληψίας του *t*
Η κατανομή που θα παίρναμε αν μια τιμή του t υπολογιζόταν για κάθε δειγματικό μέσο για όλα τα πιθανά τυχαία δείγματα ενός δεδομένου μεγέθους από κάποιον πληθυσμό.

Υπενθύμιση:
Οι βαθμοί ελευθερίας (df) αναφέρονται στο πλήθος των τιμών που μεταβάλλονται ελεύθερα όταν, για παράδειγμα, χρησιμοποιείται η μεταβλητότητα σε ένα δείγμα για να εκτιμηθεί η άγνωστη μεταβλητότητα στον πληθυσμό.

13.2 Κατανομή δειγματοληψίας του *t*

Όπως η κατανομή δειγματοληψίας του *z*, η **κατανομή δειγματοληψίας του *t*** *αναπαριστά την κατανομή που θα παίρναμε αν μια τιμή του t υπολογιζόταν για κάθε δειγματικό μέσο για όλα τα πιθανά τυχαία δείγματα ενός δεδομένου μεγέθους από κάποιον πληθυσμό.* Στις αρχές της δεκαετίας του 1900, ο William Gosset ανακάλυψε την κατανομή δειγματοληψίας του *t* και στη συνέχεια δημοσίευσε αυτό το επίτευγμα με τον τίτλο «Student». Στην πραγματικότητα, ο Gosset δεν ανακάλυψε απλώς μία αλλά μια ολόκληρη οικογένεια από κατανομές δειγματοληψίας *t* (τις λεγόμενες κατανομές του «Student»). Κάθε κατανομή *t* συσχετίζεται με έναν ειδικό αριθμό που αναφέρεται ως *βαθμός ελευθερίας* και έχει ήδη περιγραφεί στην Ενότητα 4.6. Η έννοια των βαθμών ελευθερίας παρουσιάστηκε επειδή χρησιμοποιήσαμε τη μεταβλητότητα σε ένα δείγμα για να εκτιμήσουμε την άγνωστη μεταβλητότητα στον πληθυσμό. Θυμίζουμε ότι, όταν οι *n αποκλίσεις γύρω από τον δειγματικό μέσο* χρησιμοποιούνται για την εκτίμηση της μεταβλητότητας στον πληθυσμό, μόνο οι $n - 1$ μεταβάλλονται ελεύθερα εξαιτίας του περιορισμού ότι το άθροισμα αυτών των αποκλίσεων πρέπει να είναι πάντα μηδέν. Επειδή ένας βαθμός ελευθερίας χάνεται εξαιτίας του περιορισμού μηδενικού αθροίσματος, υπάρχουν μόνο $n - 1$ βαθμοί ελευθερίας, δηλαδή, συμβολικά,

Βαθμοί ελευθερίας (ένα δείγμα)

$$df = n - 1 \tag{13.1}$$

όπου το *df* αναπαριστά τους βαθμούς ελευθερίας και το *n* ισούται με το μέγεθος δείγματος. Επειδή η έρευνα για την κατανάλωση καυσίμων περιλαμβάνει έξι αυτοκίνητα, ο αντίστοιχος έλεγχος *t* βασίζεται σε μια κατανομή δειγματοληψίας με πέντε βαθμούς ελευθερίας ($df = 6 - 1$).

Σε σύγκριση με την τυποποιημένη κανονική κατανομή

Το Σχήμα 13.1 παρουσιάζει τρεις κατανομές *t*. Όταν υπάρχει άπειρος (∞) αριθμός βαθμών ελευθερίας (και, επομένως, η τυπική απόκλιση δείγματος γίνεται ίδια με την τυπική απόκλιση πληθυσμού), η κατανομή του *t* είναι ίδια με την τυποποιημένη κανονική κατανομή του *z*. Παρατηρήστε ότι, ακόμα και με τέσσερις ή δέκα βαθμούς ελευθερίας, μια κατανομή *t* μοιράζεται αρκετές ιδιότητες με την κανονική κατανομή. Όλες οι κατανομές *t* είναι συμμετρικές, μονοκόρυφες και κωδωνοειδείς, με πυκνή συγκέντρωση που κορυφώνεται στη μέση (όταν το *t* ισούται με 0) και μειώνεται δεξιά και αριστερά από το μέσον (όσο το *t* γίνεται πιο θετικό ή αρνητικό αντίστοιχα). *Οι διογκωμένες πλευρές της κατανομής t, ιδιαίτερα προφανείς με μικρές τιμές του df, συνιστούν την πιο σημαντική διαφορά μεταξύ των κατανομών t και z.*

ΠΕΡΙΛΗΨΗ ΣΤΑΤΙΣΤΙΚΟΥ ΕΛΕΓΧΟΥ ΥΠΟΘΕΣΕΩΝ: ΕΛΕΓΧΟΣ *t* ΓΙΑ ΜΕΣΟ ΠΛΗΘΥΣΜΟΥ (ΕΡΕΥΝΑ ΓΙΑ ΤΗΝ ΚΑΤΑΝΑΛΩΣΗ ΚΑΥΣΙΜΩΝ)

Πρόβλημα έρευνας

Η μέση κατανάλωση καυσίμων για κάποιον πληθυσμό αυτοκινήτων πέφτει κάτω από το νομικά απαιτούμενο ελάχιστο των 45 mpg;

Στατιστικές υποθέσεις

$$H_0 : \mu \geq 45$$
$$H_1 : \mu < 45$$

Κανόνας απόφασης

Απόρριψη της H_0 στο επίπεδο σημαντικότητας 0,01 αν $t \leq -3{,}365$ (από τον Πίνακα Β, Παράρτημα Γ, δεδομένου ότι $df = n - 1 = 6 - 1 = 5$).

Υπολογισμοί

Δεδομένου ότι $\overline{X} = 43$, $s_{\overline{X}} = 0{,}89$

(Βλ. Πίνακα 13.1 στη σελίδα 322 για τις πράξεις.),

$$t = \frac{43 - 45}{0{,}89} = -2{,}25$$

Απόφαση

Διατήρηση της H_0 στο επίπεδο σημαντικότητας 0,01 επειδή το $t = -2{,}25$ είναι λιγότερο αρνητικό από το $-3{,}365$.

Ερμηνεία

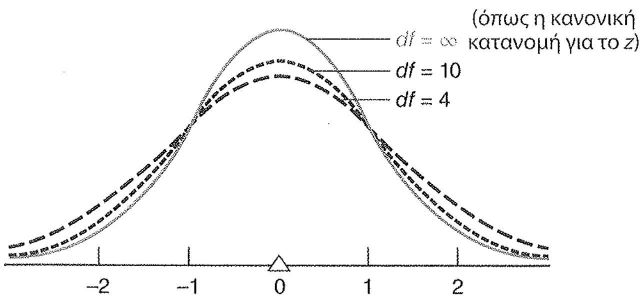
Η μέση κατανάλωση καυσίμων πληθυσμού *θα μπορούσε* να ισούται με το υποχρεωτικό 45 mpg ή να είναι μεγαλύτερη. Ο κατασκευαστής δεν πρέπει να τιμωρείται.

Πίνακας για κατανομές t

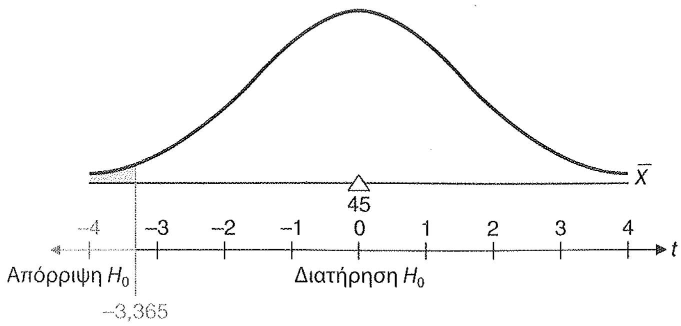
Για εξοικονόμηση χώρου, οι πίνακες για κατανομές t επικεντρώνονται μόνο στις κρίσιμες τιμές του t που αντιστοιχούν στα πιο κοινά επίπεδα σημαντικότητας. Ο Πίνακας Β του Παραρτήματος Γ παρουσιάζει τις κρίσιμες τιμές t για μονόπλευρους ή αμφίπλευρους στατιστικούς ελέγχους υποθέσεων στα επίπεδα σημαντικότητας 0,05, 0,01 και 0,001. Όλες οι κρίσιμες τιμές t που εμφανίζονται είναι θετικές και προέρχονται από το πάνω μισό κάθε κατανομής. Εξαιτίας της συμμετρίας της κατανομής t, μπορείτε να πάρετε τις αντίστοιχες κρίσιμες τιμές t για το κάτω μισό κάθε κατανομής αν απλώς προσθέσετε ένα αρνητικό πρόσημο μπροστά από κάθε εγγραφή του πίνακα.

Εύρεση κρίσιμων τιμών t

Για να βρείτε ένα κρίσιμο t στον Πίνακα Β, διαβάστε την καταχώριση στο κελί που βρίσκεται στην τομή της γραμμής για το σωστό πλήθος βαθμών ελευθερίας και της στήλης για τις προδιαγραφές ελέγχου. Για παράδειγμα, για να βρείτε το κρίσιμο t για την έρευνα σχετικά με την κατανάλωση καυσίμων, πρώτα πηγαίνετε στο δεξί πλαίσιο για έναν μονόπλευρο έλεγχο και έπειτα βρίσκετε τη γραμμή που αντιστοιχεί στους πέντε βαθμούς ελευθερίας



ΣΧΗΜΑ 13.1
Διάφορες κατανομές t.



ΣΧΗΜΑ 13.2
Υποθετική κατανομή δειγματοληψίας του t (έρευνα κατανάλωσης καυσίμων).

και τη στήλη για έναν μονόπλευρο έλεγχο στο επίπεδο σημαντικότητας 0,01. Το κελί στην τομή δείχνει 3,365. Πρέπει να προστεθεί ένα αρνητικό πρόσημο μπροστά από το 3,365, επειδή ο στατιστικός έλεγχος υποθέσεων προϋποθέτει ότι η αριστερή πλευρά θα είναι κρίσιμη. Έτσι, το –3,365 είναι το κρίσιμο *t* για την έρευνα της κατανάλωσης καυσίμων και ο αντίστοιχος κανόνας απόφασης παρουσιάζεται στο Σχήμα 13.2, όπου η κατανομή του *t* επικεντρώνεται γύρω από το μηδέν (η αντίστοιχη τιμή του *t* για την αρχική τιμή της μηδενικής υπόθεσης 45 mpg).

Αν η έρευνα για την κατανάλωση καυσίμων περιλάμβανε αμφίπλευρο έλεγχο (στο επίπεδο 0,01 με πέντε βαθμούς ελευθερίας), τότε θα επιλέγαμε το αριστερό πλαίσιο για έναν αμφίπλευρο έλεγχο και το κελί στην τομή θα είχε τιμή 4,032. Θα έπρεπε να προστεθούν θετικά και αρνητικά πρόσημα μπροστά από το 4,032, επειδή αμφότερες οι πλευρές είναι κρίσιμες. Σ' αυτήν την περίπτωση, το ±4,032 θα ήταν το ζεύγος κρίσιμων τιμών *t*.

Απουσία του df από τον Πίνακα Β του Παραρτήματος Γ

Αν το επιθυμητό πλήθος βαθμών ελευθερίας δεν εμφανίζεται στη στήλη *df* του Πίνακα Β, χρησιμοποιήστε τη γραμμή στον πίνακα με το επόμενο μικρότερο πλήθος βαθμών ελευθερίας. Για παράδειγμα, αν καθορίζονται 36 βαθμοί ελευθερίας, χρησιμοποιήστε τα στοιχεία από τη γραμμή για τους 30 βαθμούς ελευθερίας. Η στρογγυλοποίηση στο επόμενο μικρότερο *df* παράγει ένα ελαφρώς μεγαλύτερο κρίσιμο *t*, με αποτέλεσμα η απόρριψη της μηδενικής υπόθεσης να παρουσιάζει ελαφρώς μεγαλύτερη δυσκολία. Αυτή η διαδικασία εκτονώνει πιθανές διαφωνίες σχετικά με τις οριακές αποφάσεις ερευνητών ως προς την απόρριψη της μηδενικής υπόθεσης.

Έλεγχος προόδου *13.1 Βρείτε τις κρίσιμες τιμές *t* για τους παρακάτω στατιστικούς ελέγχους υποθέσεων:
(α) αμφίπλευρος έλεγχος, $a = 0{,}05$, $df = 12$
(β) μονόπλευρος έλεγχος, κρίσιμη τιμή αριστερής πλευράς, $a = 0{,}01$, $df = 19$
(γ) μονόπλευρος έλεγχος, κρίσιμη τιμή δεξιάς πλευράς, $a = 0{,}05$, $df = 38$
(δ) αμφίπλευρος έλεγχος, $a = 0{,}01$, $df = 48$
Απαντήσεις στη σελίδα 545.

13.3 Έλεγχος *t*

Συνήθως, όπως στην έρευνα για την κατανάλωση καυσίμων, *η τυπική απόκλιση πληθυσμού είναι άγνωστη και πρέπει να εκτιμηθεί από το δείγμα.* Η επακόλουθη μετάβαση από το τυπικό σφάλμα του μέσου, $\sigma_{\overline{X}}$, στην εκτίμησή του, $s_{\overline{X}}$, έχει σημαντική επίδραση σε όλο τον στατιστικό έλεγχο υπόθεσης για έναν μέσο πληθυσμού. Ο γνωστός έλεγχος *z*,

Λόγος t
Μια αντικατάσταση για τον λόγο z κάθε φορά που πρέπει να γίνει εκτίμηση της άγνωστης τυπικής απόκλισης του πληθυσμού.

$$z = \frac{\text{δειγματικός μέσος} - \text{υποθετικός μέσος πληθυσμού}}{\text{τυπικό σφάλμα}} = \frac{\overline{X} - \mu_{\text{υποθ}}}{\sigma_{\overline{X}}}$$

με την κανονική κατανομή του, πρέπει να αντικατασταθεί από έναν νέο έλεγχο *t*,

Λόγος *t* για έναν μόνο μέσο πληθυσμού

$$t = \frac{\text{δειγματικός μέσος} - \text{υποθετικός μέσος πληθυσμού}}{\text{εκτιμώμενο τυπικό σφάλμα}} = \frac{\overline{X} - \mu_{\text{υποθ}}}{s_{\overline{X}}} \qquad (13.2)$$

με τη δική του κατανομή δειγματοληψίας t και $n-1$ βαθμούς ελευθερίας. Για το παράδειγμα της κατανάλωσης καυσίμων, δεδομένου ότι ο δειγματικός μέσος της κατανάλωσης καυσίμων, $\overline{X}$, ισούται με 43, ότι ο υποθετικός μέσος πληθυσμού, $\mu_{υποθ}$, ισούται με 45 και ότι το εκτιμώμενο τυπικό σφάλμα, $s_{\overline{X}}$, ισούται με 0,89 (από τον Πίνακα 13.1), ο Τύπος 13.2 γίνεται

$$t = \frac{43 - 45}{0{,}89} = -2{,}25$$

με $df = 5$. Επειδή η παρατηρούμενη τιμή του t (–2,25) είναι λιγότερο αρνητική από την κρίσιμη τιμή του t (–3,365), η μηδενική υπόθεση διατηρείται και μπορούμε να συμπεράνουμε ότι ο κατασκευαστής αυτοκινήτων δεν πρέπει να τιμωρηθεί, επειδή η μέση κατανάλωση καυσίμων για τα αυτοκίνητα του πληθυσμού θα μπορούσε να ισούται με το υποχρεωτικό πρότυπο 45 mpg.

Μεγαλύτερη μεταβλητότητα του λόγου t

Όπως έχουμε σημειώσει, οι πλευρές της κατανομής δειγματοληψίας για το t διογκώνονται περισσότερο από εκείνες για το z, ιδιαίτερα όταν το μέγεθος δείγματος είναι μικρό.[30] Κατά συνέπεια, προκειμένου να εξυπηρετηθεί η μεγαλύτερη μεταβλητότητα του t, η κρίσιμη τιμή t πρέπει να είναι μεγαλύτερη από την αντίστοιχη κρίσιμη τιμή του z. Για παράδειγμα, δεδομένου του μονόπλευρου ελέγχου στο επίπεδο σημαντικότητας 0,01 για την έρευνα για την κατανάλωση καυσίμων, η κρίσιμη τιμή για το t (–3,365) είναι μεγαλύτερη από το z (–2,33).

13.4 Κοινό θέμα στατιστικών ελέγχων υποθέσεων

Το υπόλοιπο βιβλίο εξετάζει μια ποικιλία ελέγχων –z, t, F, U, T και H– για διάφορες περιπτώσεις. Παρά τους νέους τύπους και τα ειδικά σύμβολά τους,

> **όλοι αυτοί οι στατιστικοί έλεγχοι υποθέσεων αναπαριστούν παραλλαγές του ίδιου θέματος: Αν κάποιο παρατηρούμενο χαρακτηριστικό, όπως ο μέσος για ένα τυχαίο δείγμα, θεωρείται σπάνιο αποτέλεσμα σύμφωνα με όσα ορίζει η μηδενική υπόθεση, η υπόθεση θα απορριφθεί, αλλιώς θα διατηρηθεί.**

Προκειμένου να προσδιοριστεί αν ένα αποτέλεσμα είναι σπάνιο, το παρατηρούμενο χαρακτηριστικό μετατρέπεται σε κάποια νέα τιμή, όπως t, και συγκρίνεται με κρίσιμες τιμές από την κατάλληλη κατανομή δειγματοληψίας. Γενικά, αν η παρατηρούμενη τιμή ισούται ή είναι μεγαλύτερη από μια θετική κρίσιμη τιμή (ή αν ισούται ή είναι πιο αρνητική από μια αρνητική κρίσιμη τιμή), το αποτέλεσμα θα θεωρηθεί σπάνιο και η μηδενική υπόθεση θα απορριφθεί.

13.5 Μια επανάληψη στους βαθμούς ελευθερίας

Η έννοια των βαθμών ελευθερίας χρησιμοποιείται σε όλο το υπόλοιπο βιβλίο. Θεωρητικά, όταν χρησιμοποιείται για την εκτίμηση κάποιου άγνωστου χαρακτηριστικού του πληθυσμού, δεν μεταβάλλονται όλες οι παρατηρούμενες τιμές μέσα στο δείγμα. Για παράδειγμα, τα στοιχεία για την κατανάλωση καυσίμων αποτελούνται από έξι τιμές: 40, 44, 46, 41, 43 και 44. Ωστόσο, ο έλεγχος t γι' αυτά τα στοιχεία έχει μόλις πέντε βαθμούς ελευθερίας εξαιτίας του περιορισμού μηδενικού αθροίσματος. Μόνο πέντε απ' αυτές τις έξι παρατηρούμενες τιμές μεταβάλλονται ελεύθερα γύρω από τον μέσο τους, 43, και κατά συνέπεια παρέχουν έγκυρες πληροφορίες για την εκτίμηση. *Η έννοια των βαθμών ελευθερίας αξιοποιείται μόνο επειδή χρησιμοποιούμε παρατηρήσεις σε ένα δείγμα για να εκτιμήσουμε κάποιο άγνωστο χαρακτηριστικό του πληθυσμού.*

Σε επόμενες ενότητες θα συναντήσουμε άλλους μαθηματικούς περιορισμούς και μερικές φορές θα χάνονται αρκετοί βαθμοί ελευθερίας. Σε κάθε περίπτωση όμως, οι βαθμοί ελευθερίας υποδεικνύουν πάντα το πλήθος των

---

30. Ουσιαστικά, οι διογκωμένες πλευρές προκαλούνται από την πρόσθετη μεταβλητότητα του εκτιμώμενου τυπικού σφάλματος στον παρονομαστή του t. Για μια ολοκληρωμένη εξήγηση, διαβάστε το Κεφάλαιο 7 στο βιβλίο του Howell, D. H. (2013). *Statistical Methods for Psychology* (8η έκδοση). Belmont, CA: Wadsworth.

τιμών που μεταβάλλονται ελεύθερα, δεδομένου ενός ή περισσότερων μαθηματικών περιορισμών σε ένα σύνολο τιμών που χρησιμοποιούνται για την *εκτίμηση* κάποιου άγνωστου χαρακτηριστικού του πληθυσμού.

13.6 Λεπτομέρειες: Εκτίμηση του τυπικού σφάλματος ($s_{\overline{X}}$)

Αν η τυπική απόκλιση πληθυσμού είναι άγνωστη, πρέπει να εκτιμηθεί από το δείγμα. Αυτή η φαινομενικά ελάσσονα επιπλοκή έχει σημαντικές επιπτώσεις στον στατιστικό έλεγχο υποθέσεων – είναι μάλιστα η αιτία που ο έλεγχος *z* πρέπει να αντικατασταθεί από τον έλεγχο *t*. Τώρα το *s* αντικαθιστά το *σ* στον τύπο για το τυπικό σφάλμα του μέσου. Αντί για

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

έχουμε

Εκτιμώμενο τυπικό σφάλμα του μέσου

$$s_{\overline{X}} = \frac{s}{\sqrt{n}} \qquad (13.3)$$

όπου το $s_{\overline{X}}$ αναπαριστά το εκτιμώμενο τυπικό σφάλμα του μέσου, το *n* ισούται με το μέγεθος δείγματος και το *s* ορίζεται ως

$$s = \sqrt{\frac{SS}{n-1}} = \sqrt{\frac{SS}{df}}$$

όπου το *s* είναι η τυπική απόκλιση δείγματος, το *df* αναφέρεται στους βαθμούς ελευθερίας και το *SS* ορίζεται ως

$$SS = \Sigma(X - \overline{X})^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$$

Εκτιμώμενο τυπικό σφάλμα του μέσου ($s_{\overline{X}}$)
Το τυπικό σφάλμα του μέσου χρησιμοποιείται κάθε φορά που η άγνωστη τυπική απόκλιση του πληθυσμού πρέπει να εκτιμηθεί.

Αυτή η νέα εκδοχή του τυπικού σφάλματος, το **εκτιμώμενο τυπικό σφάλμα του μέσου**, *χρησιμοποιείται κάθε φορά που η άγνωστη τυπική απόκλιση του πληθυσμού πρέπει να εκτιμηθεί.*

Έλεγχος προόδου *13.2 Μια ομάδα καταναλωτών διαλέγει τυχαία 10 συσκευασίες μίας λίβρας κιμά που πωλούνται σε ένα σούπερ μάρκετ. Υπολογίστε **(α)** τον μέσο και **(β)** το εκτιμώμενο τυπικό σφάλμα του μέσου γι' αυτό το δείγμα, δεδομένων των παρακάτω βαρών σε ουγγιές: 16, 15, 14, 15, 14, 15, 16, 14, 14, 14.

(**Σημείωση:** Ανατρέξτε στα πλαίσια I και II του Πίνακα 13.1 για αναλυτικές οδηγίες σχετικά με τον υπολογισμό του μέσου και του εκτιμώμενου τυπικού σφάλματος γι' αυτό το νέο σύνολο δεδομένων.)

Απαντήσεις στη σελίδα 545.

13.7 Λεπτομέρειες: Υπολογισμοί για τον έλεγχο *t*

Τα τρία πλαίσια του Πίνακα 13.1 παρουσιάζουν τα βήματα υπολογισμών που παράγουν *t* –2,25 για την έρευνα σχετικά με την κατανάλωση καυσίμων.

Πλαίσιο I

Αυτό το πλαίσιο περιλαμβάνει τις περισσότερες πράξεις και παράγει τιμές για τον δειγματικό μέσο, $\overline{X}$, και την τυπική απόκλιση δείγματος, *s*. Η τυπική απόκλιση δείγματος λαμβάνεται πρώτα από τον Τύπο 4.4 (στη σελίδα 114), όπως χρησιμοποιείται για τον υπολογισμό του αθροίσματος τετραγώνων,

Υπενθύμιση:
Αντικαθιστάτε το n με το n – 1 μόνο όταν διαιρείτε το SS για να πάρετε το s.

$$SS = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$$

και έπειτα από τη διαίρεση του αθροίσματος τετραγώνων, *SS*, διά των βαθμών ελευθερίας του, $n - 1$, και εξάγοντας την τετραγωνική ρίζα.

Πλαίσιο II

Η διαίρεση της τυπικής απόκλιση δείγματος, *s*, διά της τετραγωνικής ρίζας του μεγέθους δείγματος, *n*, δίνει την τιμή για το εκτιμώμενο τυπικό σφάλμα, $s_{\overline{X}}$.

Πλαίσιο III

Τέλος, η διαίρεση της διαφοράς μεταξύ του δειγματικού μέσου, $\overline{X}$, και της τιμής της μηδενικής υπόθεσης, $\mu_{υποθ}$, διά του εκτιμώμενου τυπικού σφάλματος, $s_{\overline{X}}$, δίνει την τιμή του λόγου *t*.

Έλεγχος προόδου *13.3 Η ομάδα καταναλωτών της Ερώτησης 13.2 υποψιάζεται ότι ένα σούπερ μάρκετ κερδίζει περισσότερα πουλώντας μικρότερη ποσότητα από το καθορισμένο βάρος των 16 ουγγιών στις συσκευασίες της «μίας λίβρας» κιμά. Δεδομένου ότι ένα τυχαίο δείγμα 10 συσκευασιών δίνει μέσο 14,7 ουγγιές και εκτιμώμενο τυπικό σφάλμα του μέσου 0,26 ουγγιές, ακολουθήστε την τυπική αναλυτική διαδικασία για να ελέγξετε τη μηδενική υπόθεση στο επίπεδο σημαντικότητας 0,05 με το *t*.

Η απάντηση στη σελίδα 546.

13.8 Διαστήματα εμπιστοσύνης για το μ βάσει του *t*

Υπό ελαφρώς διαφορετικές συνθήκες, ενδεχομένως να θέλετε να εκτιμήσετε την άγνωστη μέση κατανάλωση καυσίμων για τον πληθυσμό των αυτοκινήτων αντί να ελέγξετε μια υπόθεση με βάση το πρότυπο 45 mpg. Για παράδειγμα, μπορεί να μην υπάρχει κανένα νομίμως προκαθορισμένο ελάχιστο πρότυπο στα 45 mpg, αλλά απλώς μια επιθυμία εκ μέρους του κατασκευαστή να εκτιμήσει τη μέση κατανάλωση καυσίμων για έναν πληθυσμό αυτοκινήτων – πιθανώς ως το πρώτο βήμα της σχεδίασης μιας νέας και βελτιωμένης έκδοσης του τρέχοντος μοντέλου.

Όταν η τυπική απόκλιση πληθυσμού είναι άγνωστη και πρέπει επομένως να εκτιμηθεί, όπως στην προκειμένη περίπτωση, το *t* αντικαθιστά το *z* στον νέο τύπο για ένα διάστημα εμπιστοσύνης:

Διάστημα εμπιστοσύνης για το μ βάσει του *t*

$$\overline{X} \pm (t_{εμπ})(s_{\overline{X}}) \qquad (13.4)$$

όπου το $\overline{X}$ αναπαριστά τον δειγματικό μέσο, το $t_{εμπ}$ αναπαριστά έναν αριθμό (που κατανέμεται με $n - 1$ βαθμούς ελευθερίας) από τους πίνακες *t*, ο οποίος ικανοποιεί τις προδιαγραφές εμπιστοσύνης για το διάστημα εμπιστοσύνης, και το $s_{\overline{X}}$ αναπαριστά το εκτιμώμενο τυπικό σφάλμα του μέσου, όπως ορίζεται στον Τύπο 13.3.

Εύρεση του $t_{εμπ}$

Για να βρείτε την κατάλληλη τιμή για το $t_{εμπ}$ στον Τύπο 13.4, ανατρέξτε στον Πίνακα Β του Παραρτήματος Γ. Διαβάστε την καταχώριση από το κελί που βρίσκεται στην τομή της γραμμής για το σωστό πλήθος βαθμών ελευθερίας και της στήλης για τις προδιαγραφές εμπιστοσύνης. Στην προκειμένη περίπτωση, αν το ζητούμενο είναι ένα διάστημα εμπιστοσύνης 95%, πρώτα βρείτε τη γραμμή που αντιστοιχεί σε 5 βαθμούς ελευθερίας (από την πράξη $df = n - 1 = 6 - 1 = 5$) και έπειτα εντοπίστε τη στήλη για το επίπεδο εμπιστοσύνης 95%, δηλαδή τη στήλη της

Πίνακας 13.1
ΥΠΟΛΟΓΙΣΜΟΙ ΓΙΑ ΤΟΝ ΕΛΕΓΧΟ t (ΕΡΕΥΝΑ ΓΙΑ ΤΗΝ ΚΑΤΑΝΑΛΩΣΗ ΚΑΥΣΙΜΩΝ)

I. Εύρεση των $\bar{X}$ ΚΑΙ s

(α) Ακολουθία πράξεων:

Εκχώρηση τιμής στο n (1).
Πρόσθεση όλων των αποτελεσμάτων X (2).
Αντικατάσταση με αριθμούς στον τύπο (3) και λύση ως προς $\bar{X}$.
Ύψωση στο τετράγωνο κάθε αποτελέσματος X (4), ένα κάθε φορά, και πρόσθεση όλων των τετραγώνων των αποτελεσμάτων X (5).
Αντικατάσταση με αριθμούς στον τύπο (6) και λύση ως προς s (7).

(β) Δεδομένα και πράξεις:

X	**4** X^2
40	1.600
44	1.936
46	2.116
41	1.681
43	1.849
44	1.936

1 $n = 6$ **2** $\Sigma X = 258$ **5** $\Sigma X^2 = 11.118$

3 $\bar{X} = \frac{\Sigma X}{n} = \frac{258}{6} = 43$

6 $SS = \Sigma X^2 - \frac{(\Sigma X)^2}{n} = 11.118 - \frac{(258)^2}{6} = 11.118 - \frac{66.564}{6} = 11.118 - 11.094 = 24$

7 $s = \sqrt{\frac{SS}{n-1}} = \sqrt{\frac{24}{6-1}} = \sqrt{4,8} = 2,19$

II. Εύρεση του $s_{\bar{X}}$

(α) Ακολουθία πράξεων:

Αντικατάσταση με τους αριθμούς που πήρατε πιο πάνω στον τύπο **8** και λύση ως προς $s_{\bar{X}}$.

(β) Πράξεις:

8 $s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{2,19}{\sqrt{6}} = \frac{2,19}{2,45} = 0,89$

III. Εύρεση του παρατηρούμενου t

(α) Ακολουθία πράξεων:

Εκχώρηση τιμής στο $\mu_{υποθ}$ **9**, τον υποθετικό μέσο πληθυσμού.
Αντικατάσταση με τους αριθμούς που πήρατε πιο πάνω στον τύπο **10** και λύση ως προς t.

(β) Πράξεις:

9 $\mu_{υποθ} = 45$

10 $t = \frac{\bar{X} - \mu_{υποθ}}{s_{\bar{X}}} = \frac{43 - 45}{0,89} = \frac{-2}{0,89} = -2,25$

οποίας η επικεφαλίδα είναι ένας αστερίσκος. (Ο διπλός αστερίσκος επισημαίνει τη στήλη για το επίπεδο εμπιστοσύνης 99%.) Το κελί στην τομή δείχνει ότι θα πρέπει να εισαχθεί η τιμή 2,571 στον Τύπο 13.4.[31]

Δεδομένης αυτής της τιμής για το $t_{εμπ}$, καθώς και της τιμής του 43 για το $\overline{X}$ (από τον Πίνακα 13.1), του δειγματικού μέσου κατανάλωσης καυσίμων, και του 0,89 για το $s_{\overline{X}}$, του εκτιμώμενου τυπικού σφάλματος, ο Τύπος 13.3 γίνεται

$$43 \pm (2{,}571)(0{,}89) = 43 \pm 2{,}29 = \begin{cases} 45{,}29 \\ 40{,}71 \end{cases}$$

Μπορούμε να ισχυριστούμε, με 95% βεβαιότητα, ότι το διάστημα μεταξύ 40,71 και 45,29 περιλαμβάνει την πραγματική μέση κατανάλωση καυσίμων για όλα τα αυτοκίνητα του πληθυσμού.

Ερμηνεία

Η ερμηνεία αυτού του διαστήματος εμπιστοσύνης είναι ίδια με εκείνη που βασίζεται στο z. Μακροπρόθεσμα, το 95% όλων των διαστημάτων εμπιστοσύνης, όπως εκείνου που μόλις περιγράψαμε, θα περιλαμβάνει τον άγνωστο μέσο πληθυσμού. Αν και ποτέ δεν γνωρίζουμε πραγματικά αν αυτό το συγκεκριμένο διάστημα εμπιστοσύνης είναι σωστό ή λάθος, μπορούμε να είμαστε *εύλογα βέβαιοι* ότι ο πραγματικός μέσος για όλο τον πληθυσμό αυτοκινήτων δεν είναι ούτε μικρότερος από 40,71 mpg ούτε μεγαλύτερος από 45,29 mpg.

Έλεγχος προόδου *13.4 Η ομάδα καταναλωτών (της Ερώτησης 13.3) συμπεραίνει ότι, παρά τους ισχυρισμούς του σούπερ μάρκετ, το μέσο βάρος των συσκευασιών «μίας λίβρας» κιμά είναι μικρότερο από το ορθό βάρος των 16 ουγγιών, ακόμα κι αν ληφθεί υπόψη η μεταβλητότητα δειγματοληψίας.

(α) Κατασκευάστε ένα διάστημα εμπιστοσύνης 95% για το πραγματικό βάρος όλων των συσκευασιών «μίας λίβρας» κιμά.

(β) Ερμηνεύστε αυτό το διάστημα εμπιστοσύνης.

Απαντήσεις στη σελίδα 546.

13.9 Υποθέσεις

Όταν ελέγχετε υποθέσεις ή κατασκευάζετε διαστήματα εμπιστοσύνης για μέσους πληθυσμού, χρησιμοποιείτε το t αντί για το z κάθε φορά που, *όπως συμβαίνει σχεδόν πάντα,* η τυπική απόκλιση πληθυσμού είναι άγνωστη. Με απόλυτη ακρίβεια, όταν χρησιμοποιείτε το t, πρέπει να θεωρείτε ότι ο υποκείμενος πληθυσμός κατανέμεται κανονικά. Ακόμα κι αν παραβιαστεί αυτή η υπόθεση κανονικότητας, το t διατηρεί το μεγάλο μέρος της ακρίβειάς του εφόσον το μέγεθος δείγματος δεν είναι πολύ μικρό. Αν χρησιμοποιείτε ένα πολύ μικρό δείγμα (λιγότερο από περίπου 10) *και* πιστεύετε ότι το δείγμα προέρχεται από έναν μη κανονικό πληθυσμό –πιθανώς εξαιτίας μιας έντονης θετικής ή αρνητικής κλίσης στις παρατηρήσεις στο δείγμα–, θα ήταν συνετό να αυξάνατε το μέγεθος του δείγματος πριν διεξαγάγετε στατιστικό έλεγχο υπόθεσης ή κατασκευάσετε διάστημα εμπιστοσύνης.

---

31. Οι προδιαγραφές για τα διαστήματα εμπιστοσύνης λαμβάνονται από το αριστερό πλαίσιο του Πίνακα Β επειδή τα συμμετρικά όρια ενός διαστήματος εμπιστοσύνης είναι ανάλογα με έναν αμφίπλευρο στατιστικό έλεγχο υποθέσεων. Ουσιαστικά, αμφότερες οι διαδικασίες προϋποθέτουν ότι ένα συγκεκριμένο πλήθος τυπικών σφαλμάτων θα προστεθεί και θα αφαιρεθεί σε σχέση με την τιμή της μηδενικής υπόθεσης (ώστε να ληφθεί η ανώτερη και η κατώτερη κρίσιμη τιμή για έναν αμφίπλευρο στατιστικό έλεγχο υποθέσεων) ή την τιμή του δειγματικού μέσου (ώστε να ληφθεί το ανώτερο και το κατώτερο όριο ενός διαστήματος εμπιστοσύνης).

Περίληψη

Όταν η τυπική απόκλιση πληθυσμού, σ, είναι άγνωστη, πρέπει να εκτιμηθεί με την τυπική απόκλιση δείγματος, s. Με την ίδια λογική, το τυπικό σφάλμα του μέσου, $\sigma_{\bar{X}}$, θα πρέπει τότε να εκτιμηθεί με το $s_{\bar{X}}$. Υπό αυτές τις συνθήκες, το t και όχι το z θα πρέπει να χρησιμοποιείται για τον έλεγχο μιας υπόθεσης ή για την κατασκευή ενός διαστήματος εμπιστοσύνης για τον μέσο πληθυσμού.

Ο λόγος t κατανέμεται με $n-1$ βαθμούς ελευθερίας και οι κρίσιμες τιμές t λαμβάνονται από τον Πίνακα Β στο Παράρτημα Γ. Εξαιτίας των διογκωμένων πλευρών των κατανομών δειγματοληψίας t, ιδιαίτερα όταν το μέγεθος δείγματος είναι μικρό, οι κρίσιμες τιμές του t είναι μεγαλύτερες (θετικές ή αρνητικές) από τις αντίστοιχες κρίσιμες τιμές του z.

Οι βαθμοί ελευθερίας (*df*) αναφέρονται στο πλήθος των τιμών που μεταβάλλονται ελεύθερα, δεδομένου ενός ή περισσότερων μαθηματικών περιορισμών σε ένα σύνολο τιμών που χρησιμοποιείται για την εκτίμηση κάποιου χαρακτηριστικού του πληθυσμού.

Η χρήση του t προϋποθέτει ότι ο υποκείμενος πληθυσμός κατανέμεται κανονικά. Οι παραβιάσεις αυτής της υπόθεσης είναι σημαντικές μόνο όταν οι παρατηρήσεις σε μικρά δείγματα φαίνεται να προέρχονται από μη κανονικούς πληθυσμούς.

Σημαντικοί όροι

Κατανομή δειγματοληψίας του *t*
Λόγος *t*
Βαθμοί ελευθερίας (*df*)
Εκτιμώμενο τυπικό σφάλμα του μέσου ($s_{\bar{X}}$)

Κύριες εξισώσεις

ΛΟΓΟΣ t

$$t = \frac{\bar{X} - \mu_{υποθ}}{s_{\bar{X}}}$$

$$\text{όπου} \quad s_{\bar{X}} = \frac{s}{\sqrt{n}}$$

Ερωτήσεις επανάληψης

***13.5** Ένα σύστημα βιβλιοθήκης δανείζει βιβλία για περιόδους 21 ημερών. Αυτή η πολιτική εκτιμάται εκ νέου, ώστε να ληφθεί μια απόφαση για το αν η περίοδος δανεισμού θα πρέπει να είναι μεγαλύτερη ή μικρότερη από 21 ημέρες. Προκειμένου να ληφθεί αυτή η απόφαση, οι υπεύθυνοι συμβουλεύονται τα αρχεία δανεισμού βιβλίων για να διαπιστώσουν τις πραγματικές περιόδους που χρησιμοποιούν τα μέλη της βιβλιοθήκης. Ένα τυχαίο δείγμα οκτώ εγγραφών αποκάλυψε τις εξής περιόδους δανεισμού σε ημέρες: 21, 15, 12, 24, 20, 21, 13 και 16. Ελέγξτε τη μηδενική υπόθεση με το t, χρησιμοποιώντας το επίπεδο σημαντικότητας 0,05.
Απαντήσεις στη σελίδα 546.

13.6 Θα υποθέσουμε ότι έχει αποδειχθεί πως τα ποντίκια των εργαστηρίων χρειάζονται κατά μέσο όρο 32 προσπάθειες σε έναν περίπλοκο υδάτινο λαβύρινθο πριν καταφέρουν να πετύχουν σε ένα κριτήριο εκμάθησης που απαιτεί τρεις διαδοχικές επιτυχίες. Για να διαπιστωθεί αν ένα ελαφρύ ερέθισμα για το αντίθετο έχει κάποια επίδραση στην απόδοση, ένα δείγμα επτά ποντικών υποβλήθηκε σε ήπιο ηλεκτροσόκ αμέσως πριν από κάθε δοκιμή.

(α) Δεδομένου ότι $\overline{X} = 34,89$ και $s = 3,02$, ελέγξτε τη μηδενική υπόθεση με t, χρησιμοποιώντας το επίπεδο σημαντικότητας 0,05.
(β) Κατασκευάστε ένα διάστημα εμπιστοσύνης 95% για τον πραγματικό αριθμό των δοκιμών που απαιτούνται για την εκμάθηση του λαβυρίνθου.
(γ) Ερμηνεύστε αυτό το διάστημα εμπιστοσύνης.

***13.7** Η θερμοκρασία της Γης αυξάνεται επειδή η θερμότητα παγιδεύεται από τις εκπομπές αερίων του θερμοκηπίου, όπως του διοξειδίου του άνθρακα, στην ατμόσφαιρα της Γης; Το κέντρο ερευνών για το κλίμα National Climatic Data Center αναφέρει στον ιστότοπό του (http://www.ncdc.noaa.gov/pub/data/anomalies.html) ότι οι μέσες παγκόσμιες θερμοκρασίες τα τελευταία χρόνια αποκλίνουν από τη μακροχρόνια μέση θερμοκρασία όλου του 20ού αιώνα. Σε βαθμούς Fahrenheit, οι ετήσιες *αποκλίσεις πάνω* από τη μακροχρόνια μέση θερμοκρασία για καθένα από τα δέκα τελευταία χρόνια, σε χρονολογική σειρά έως το 2015, ήταν 1,15, 1,15, 1,01, 1,03 1,15, 1,22, 1,03, 1,13, 1,21 και 1,35.
(α) Δεδομένου ότι $\overline{X} = 1,14$ και $s = 0,10$ γι' αυτά τα δέκα χρόνια, χρησιμοποιήστε το t στο επίπεδο σημαντικότητας 0,01 για να ελέγξετε τη μηδενική υπόθεση ότι η θερμοκρασία της Γης δεν αυξάνεται. Με άλλα λόγια, θα μπορούσε η απόκλιση του δειγματικού μέσου γι' αυτά τα δέκα έτη να προέρχεται από έναν πληθυσμό ετήσιων αποκλίσεων για όλο τον 20ό αιώνα που παρουσιάζει μέση απόκλιση *ίση με μηδέν*;
(β) Αν είναι δυνατόν (επειδή η μηδενική υπόθεση έχει απορριφθεί), κατασκευάστε ένα διάστημα εμπιστοσύνης 99% και ερμηνεύστε το.
Απαντήσεις στις σελίδες 546-547.

13.8 Έστω ότι, κατά μέσο όρο, οι υγιείς νεαροί ενήλικες ονειρεύονται 90 λεπτά κάθε βράδυ, όπως προκύπτει από διάφορες μετρήσεις, όπως τον ύπνο γρήγορης κίνησης των ματιών (REM). Ένας ερευνητής θέλει να διαπιστώσει αν η κατανάλωση καφέ πριν από τον ύπνο επηρεάζει τη διάρκεια των ονείρων. Μετά την κατανάλωση μιας καθορισμένης ποσότητας καφέ, η διάρκεια των ονείρων παρακολουθείται για 28 υγιείς νεαρούς άντρες σε ένα τυχαίο δείγμα. Τα αποτελέσματα δείχνουν δειγματικό μέσο, $\overline{X}$, 88 λεπτά και τυπική απόκλιση δείγματος, s, 9 λεπτά.
(α) Χρησιμοποιήστε το t για να ελέγξετε τη μηδενική υπόθεση στο επίπεδο σημαντικότητας 0,05.
(β) Αν είναι δυνατόν (επειδή η μηδενική υπόθεση έχει απορριφθεί), κατασκευάστε ένα διάστημα εμπιστοσύνης 99% και ερμηνεύστε το.

13.9 Στον έλεγχο της κατανάλωσης καυσίμων που περιγράψαμε σ' αυτό το κεφάλαιο, θα προτιμούσατε μικρότερο ή μεγαλύτερο μέγεθος δείγματος αν ήσαστε
(α) ο κατασκευαστής αυτοκινήτων; Γιατί;
(β) ένας δραστήριος εισαγγελέας υπέρ της νομικής ρύθμισης; Γιατί;

13.10 Ακόμα κι αν η τυπική απόκλιση πληθυσμού είναι άγνωστη, ένας ερευνητής χρησιμοποιεί το z αντί του πιο κατάλληλου t για να ελέγξει μια υπόθεση στο επίπεδο σημαντικότητας 0,05.
(α) Το πραγματικό επίπεδο σημαντικότητας είναι μεγαλύτερο ή μικρότερο από 0,05;
(β) Η πραγματική κρίσιμη τιμή είναι μεγαλύτερη ή μικρότερη από το κρίσιμο z;

ΠΑΡΑΡΤΗΜΑ XIII

Έλεγχος *t*

Γιώργος Ανδρουλάκης

Π.XIII.1 Εισαγωγή

Το παρόν Κεφάλαιο 13 επεκτείνει την εφαρμογή του ελέγχου υποθέσεων και των διαστημάτων εμπιστοσύνης σε πληθυσμούς με άγνωστη διακύμανση κάνοντας χρήση της κατανομής *t*.

Π.XIII.2 Η κατανομή *t*

Συνεχίζοντας τα παραδείγματα οπτικοποίησης των κατανομών που έχουμε παρουσιάσει στα τελευταία κεφάλαια, στην τρέχουσα ενότητα θα φτιάξουμε ένα συγκριτικό γράφημα της τυποποιημένης κανονικής κατανομής και της κατανομής *t* με 5 βαθμούς ελευθερίας.

Ξεκινάμε το παράδειγμά μας αποδίδοντας τιμές στις παραμέτρους μας:

```
> mean=0; sd=1
> x <- seq(-4,4,length=100)*sd + mean
```

Ορίσαμε τον μέσο ως 0 και την τυπική απόκλιση 1, για να τις χρησιμοποιήσουμε για την τυποποιημένη κανονική κατανομή. Επιπλέον, στο διάνυσμα x τοποθετούμε 100 ισαπέχουσες τιμές που εκτείνονται από το –4 έως το 4, δηλαδή 4 τυπικές αποκλίσεις απόσταση δεξιά και αριστερά από τον μέσο της τυποποιημένης κατανομής.

```
> hn <- dnorm(x,mean,sd)
> ht <- dt(x, 5)
```

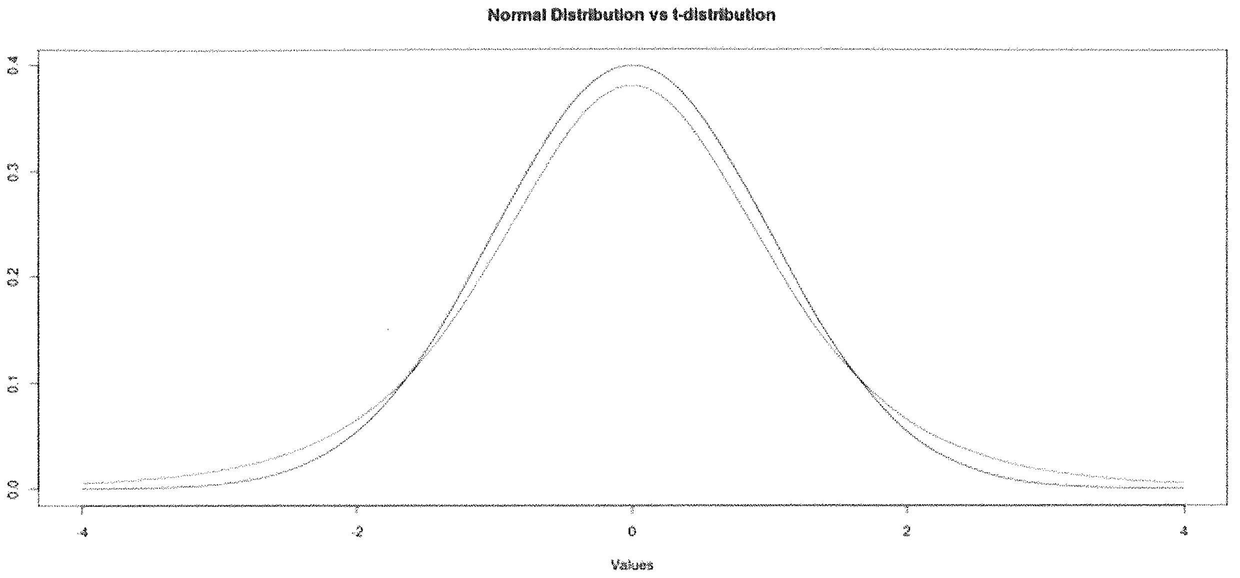
Στα διανύσματα hn και ht τοποθετούμε τις τιμές της τυποποιημένης κανονικής κατανομής και της κατανομής *t* με 5 βαθμούς ελευθερίας αντίστοιχα, υπολογισμένες στα 100 σημεία του διανύσματος x. Είμαστε έτοιμοι να φτιάξουμε το συγκριτικό μας γράφημα:

```
> plot(x, hx, type="n", xlab="Values", ylab="",
+     main="Normal Distribution vs t-distribution", axes=T)
```

Αρχικά, όπως συνηθίζουμε στα τελευταία παραδείγματα, φτιάχνουμε ένα «κενό» γράφημα που θα το χρησιμοποιήσουμε ως καμβά για το υπόλοιπο γράφημά μας. Στο σημείο αυτό ο παρατηρητικός αναγνώστης θα αναρωτηθεί γιατί να χρησιμοποιήσουμε τις τιμές της τυποποιημένης κατανομής για τη δημιουργία του κενού γραφήματος και όχι τις τιμές της κατανομής *t*. Αν το σκεφτούμε λίγο, θα καταλάβουμε ότι η απάντηση είναι απλή και θα έπρεπε να το είχαμε σκεφτεί διαβάζοντας τη θεωρία. Είδαμε στη θεωρία ότι η κατανομή *t* μοιάζει με την κανονική κατανομή, έχει δηλαδή σχήμα καμπάνας, με τη διαφορά ότι οι κεντρικές τιμές της είναι χαμηλότερα και οι τιμές της στα άκρα ψηλότερα από τις αντίστοιχες τιμές της τυποποιημένης κανονικής κατανομής. Με άλλα λόγια, η τυποποιημένη κανονική κατανομή στα άκρα του σχήματος θα έχει τιμές κάτω από τις τιμές της κατανομής *t* και στο κέντρο τιμές πάνω από την κατανομή *t*. Επομένως, έτσι εξασφαλίσαμε ότι ο «καμβάς» μας θα χωράει και το γράφημα και των δύο κατανομών. Είμαστε έτοιμοι να κατασκευάσουμε τα δύο γραφήματα

```
> lines(x, hx, type ="l", col = "blue")
> lines(x, ht, type ="l", col = "red")
```

Επιλέξαμε την εντολή lines, που τοποθετεί μικρά ευθύγραμμα τμήματα, τα οποία όμως, επειδή είναι πολύ κοντά το ένα με το άλλο, δίνουν την ψευδαίσθηση της καμπυλότητας. Η πρώτη γραμμή του παραπάνω κώδικα ενώνει με συνεχόμενα ευθύγραμμα τμήματα τα σημεία της τυποποιημένης κανονικής κατανομής και η δεύτερη της κατανομής *t* με 5 βαθμούς ελευθερίας.



Στο παραπάνω σχήμα εμφανίζεται το συγκριτικό γράφημα κανονικής κατανομής και κατανομής *t* που δημιουργήσαμε. Παρατηρούμε τα χαρακτηριστικά της κατανομής *t* όπως προκύπτουν από τη θεωρία, δηλαδή ότι έχει σχήμα καμπάνας, η κορυφή της είναι «χαμηλότερα» από την κορυφή της κανονικής κατανομής και τα άκρα της δεξιά και αριστερά «ψηλότερα» από την κανονική κατανομή.

Π.XIII.3 Έλεγχος *t*

Ας εστιάσουμε στον έλεγχο υποθέσεων με βάση την κατανομή *t*. Έστω το παράδειγμα από το ερωτηματολόγιο των φοιτητών που έχουμε δει και σε προηγούμενες ενότητες. Απομονώνουμε τις απαντήσεις των 556 φοιτητών αναφορικά με το πόσες ώρες αφιέρωναν στο διάβασμα την περίοδο που προετοιμάζονταν για τις εισαγωγικές εξετάσεις για το πανεπιστήμιο. Πριν από οποιαδήποτε στατιστική επεξεργασία, μια καλή πρακτική είναι να διερευνήσουμε τα δεδομένα μας. Για την πρώτη στατιστική προσέγγιση, η R, μέσω της έτοιμης συνάρτησης summary, εμφανίζει

```
> summary(mydata$readinghours)
   Min. 1st Qu. Median Mean 3rd Qu. Max.
    1.00 3.00 4.00 4.57 5.00 20.00
```

έξι ποσοτικούς δείκτες, την ελάχιστη τιμή 1,00, το πρώτο τεταρτημόριο με τιμή 3,00, τη διάμεσο 4,00, τη μέση τιμή 4,57, το τρίτο τεταρτημόριο με τιμή 5,00 και τη μέγιστη τιμή 20,00. Άμεσα, από τη σύγκριση των τιμών διαμέσου και μέσης τιμής, έχουμε ένδειξη ότι δεν υπάρχει συμμετρία. Για να το διαπιστώσουμε και οπτικά, ένα κατάλληλο γράφημα είναι το ιστόγραμμα. Ένα από τα μεγάλα πλεονεκτήματα του ελεύθερου λογισμικού R, εκτός από το ότι είναι ελεύθερο, είναι τα γραφήματά του. Διαθέτει πάρα πολλές έτοιμες βιβλιοθήκες γραφικών. Μια μικρή αναζήτηση στο διαδίκτυο και ο απαιτητικός αναγνώστης θα διαπιστώσει την υπερπλία της R σε αυτόν τον τομέα. Μία από τις ισχυρότερες βιβλιοθήκες για γραφικά στην R είναι η ggplot2. Όπως εξηγήσαμε αναλυτικά και σε άλλες ενότητες, οι περισσότερες βιβλιοθήκες της R δεν είναι προεγκατεστημένες, για να γίνεται εξοικονόμηση υπολογιστικής ισχύος του υπολογιστή. Επομένως, για να τη χρησιμοποιήσουμε πρέπει να την εγκαταστήσουμε στον υπολογιστή μας και, εφόσον έχει ολοκληρωθεί αυτή η διαδικασία, να την ενεργοποιήσουμε για χρήση στο περιβάλλον της R. Για την αρχική εγκατάσταση της βιβλιοθήκης ggplot2 εκτελούμε την παρακάτω εντολή έχοντας εξασφαλίσει ήδη ότι ο υπολογιστής μας έχει πρόσβαση στο διαδίκτυο ώστε να μεταφορτώσει και να εγκαταστήσει τα απαραίτητα αρχεία:

```
> install.packages("ggplot2")
```

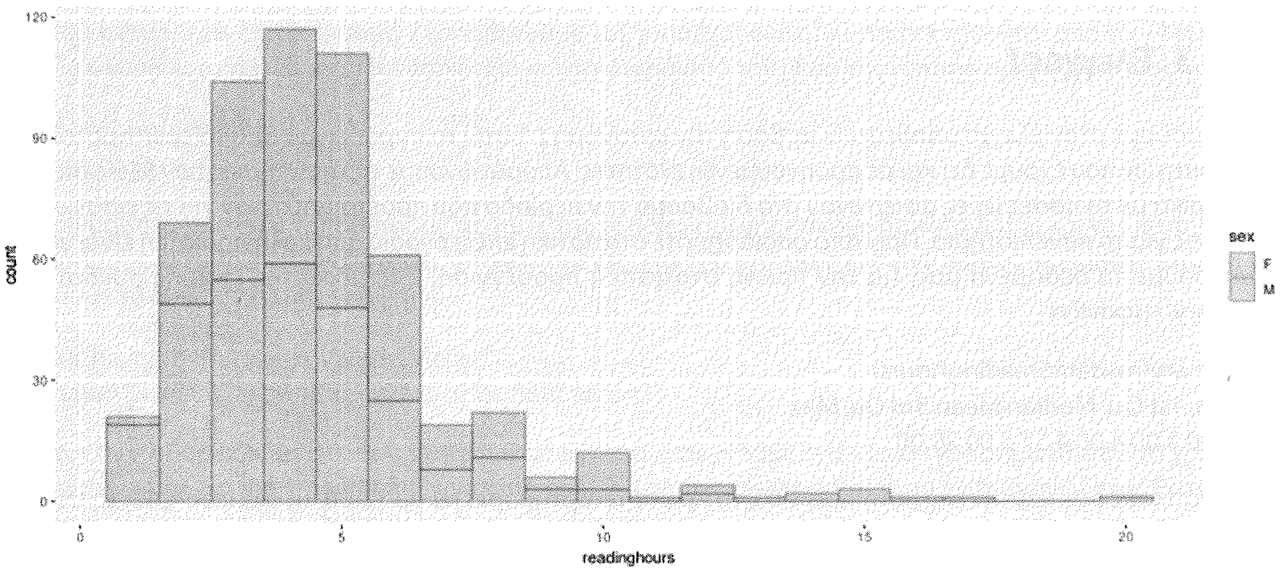
Τώρα η βιβλιοθήκη είναι πλέον εγκατεστημένη στον υπολογιστή μας, οπότε για να τη χρησιμοποιήσουμε αρκεί μόνο να την ενεργοποιήσουμε με την εντολή library, όπως παρακάτω:

```
> library(ggplot2)
```

Η φιλοσοφία της βιβλιοθήκης ggplot2 είναι τα γραφήματα να περιλαμβάνουν «διάφανα» επίπεδα το ένα πάνω στο άλλο έτσι ώστε το γράφημα να εμπλουτίζεται σταδιακά. Αρχικά, ο χρήστης ορίζει το πλαίσιο δεδομένων και τις μεταβλητές που θα χρησιμοποιηθούν και κατόπιν, ως ξεχωριστό επίπεδο, το είδος του γραφήματος που θα κατασκευαστεί με αυτές τις μεταβλητές. Επομένως, η παρακάτω εντολή ggplot, η οποία είναι γενική εντολή και ανάλογα με τα ορίσματά της φτιάχνει και διαφορετικό είδος γραφήματος,

```
> ggplot(data = FinalData, aes(x=readinghours, color=sex, fill=sex)) + geom_histogram(binwidth = 1, alpha=0.3)
```

αρχικά κατευθύνει την R να χρησιμοποιήσει το πλαίσιο δεδομένων FinalData (τα ερωτηματολόγια των φοιτητών του Πανεπιστημίου Πατρών). Μέσω του ορίσματος aes ορίζονται αισθητικές (aesthetic) παράμετροι του γραφήματος. Στο παράδειγμά μας επιλέγονται τα δεδομένα του άξονα x να είναι η μεταβλητή readinghours (υπενθυμίζουμε ότι αντιστοιχεί στο ερώτημα πόσες ώρες διάβαζαν οι φοιτητές κατά μέσο όρο κατά τη διάρκεια προετοιμασίας τους για τις πανελλήνιες εξετάσεις), τα χρώματα (color) του περιγράμματος του γραφήματος να αντιστοιχούν στη μεταβλητή sex (όσα και τα φύλα) και ο χρωματισμός των στοιχείων του γραφήματος (fill) να αντιστοιχεί πάλι στη μεταβλητή sex. Ακολουθεί το στρώμα του γραφήματος geom_histogram, το οποίο κατασκευάζει σε αυτό το επίπεδο ένα ιστόγραμμα. Τα ιδιαίτερα χαρακτηριστικά του ιστογράμματος είναι το εύρος της κάθε μπάρας να είναι 1 (binwidth=1) και ο χρωματισμός να έχει διαφάνεια 30% (alpha=0,3). Το αποτέλεσμα της εντολής μας είναι το παρακάτω:



Παρατηρήστε ότι, επειδή ζητήσαμε ο χρωματισμός να λάβει υπόψη του τα δύο φύλα, οι μπάρες περιλαμβάνουν μία μπλε περιοχή, που αντιστοιχεί στις τιμές M (male) της μεταβλητής sex, και μία κόκκινη περιοχή, που αντιστοιχεί στις τιμές F (female). Οι δύο περιοχές έχουν τοποθετηθεί η μία πάνω από την άλλη και, επομένως, οι μπάρες εκφράζουν τη συχνότητα ανεξάρτητα από το φύλο.

Για να κάνουμε έλεγχο υποθέσεων ή διάστημα εμπιστοσύνης με βάση την κατανομή *t* η εντολή της R είναι η t.test. Την εκτελούμε με μοναδικό όρισμα τη μεταβλητή readinghours και λαμβάνουμε το παρακάτω αποτέλεσμα:

```
> t.test(FinalData$readinghours)

        One Sample t-test
```

```
data: mydata$readinghours
t = 43.756, df = 555, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 4.364987 4.775301
sample estimates:
mean of x
4.570144
```

Το πρώτο πράγμα που απεικονίζεται στην οθόνη του υπολογιστή μας είναι ότι η R εκτέλεσε έλεγχο υποθέσεων ενός δείγματος. Γιατί μας ενημερώνει για κάτι τέτοιο; Η απάντηση είναι απλή, η φιλοσοφία του στατιστικού πακέτου είναι πιο σύνθετη. Οι εντολές έχουν τη δυνατότητα πίσω από το κεντρικό όνομά τους να «κρύβουν» διαφορετικές υλοποιήσεις ανάλογα με το είδος των δεδομένων και τα ορίσματα. Έτσι, ο χρήστης δεν είναι αναγκασμένος να θυμάται τις εξειδικευμένες λεπτομέρειες κάθε ορίσματος. Το κάνει το λογισμικό γι' αυτόν και στις περισσότερες περιπτώσεις το κάνει σωστά. Για να εξασφαλιστεί ο χρήστης για το αν έχει γίνει η σωστή χρήση της εντολής, το λογισμικό –στην περίπτωση αυτή των πολλαπλών διαφορετικών υλοποιήσεων– εμφανίζει στην οθόνη τι «αποφάσισε» να υλοποιήσει.

Η επόμενη γραμμή εμφανίζει σε ποια δεδομένα έτρεξε η συγκεκριμένη εντολή. Οι δύο επόμενες γραμμές στην απάντηση του λογισμικού αφορούν τον έλεγχο υποθέσεων. Συγκεκριμένα, στη δεύτερη από αυτές τις δύο γραμμές αναφέρεται η εναλλακτική υπόθεση "true mean is not equal to 0", επομένως το λογισμικό εκτελεί αμφίπλευρο έλεγχο αναφορικά με το αν ο μέσος είναι διάφορος του μηδενός. Στην πρώτη γραμμή αναφέρονται οι 555 βαθμοί ελευθερίας, ο δείκτης t, που υπολογίστηκε σε 43,756, η αντίστοιχη p-τιμή του, που υπολογίστηκε σε $2{,}2 \cdot 10^{-16}$. Επομένως, αν στοχεύαμε στο να κάνουμε έλεγχο υποθέσεων, θα απορρίπταμε τη μηδενική υπόθεση. Ακολουθεί το αντίστοιχο διάστημα εμπιστοσύνης σε επίπεδο σημαντικότητας 95%, το οποίο είναι [4,364987, 4,775301]. Η τελευταία πληροφορία που λαμβάνουμε από την εντολή t.test είναι η μέση τιμή των δεδομένων μας, 4,570144.

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)