

Κεφάλαιο 68

Σχεδιασμός κλινικών μελετών και διαχείριση δεδομένων έρευνας

Γ. Η. Πανάγος

ΟΡΘΗ ΠΡΑΚΤΙΚΗ ΔΙΕΞΑΓΩΓΗΣ ΚΛΙΝΙΚΩΝ ΜΕΛΕΤΩΝ

Η ορθή πρακτική διεξαγωγής των κλινικών δοκιμών (GCP) είναι ένα διεθνές πρότυπο που καθορίζει τον δεοντολογικό και επιστημονικά ποιοτικό σχεδιασμό, τη διεξαγωγή, την καταγραφή και την αναφορά των κλινικών δοκιμών οι οποίες αφορούν στη συμμετοχή ανθρώπων. Η συμμόρφωση προς το πρότυπο αυτό παρέχει στο κοινό (ιατρική κοινότητα, υγειονομικές αρχές, γενικό κοινό) τη βεβαιότητα ότι τα δικαιώματα, η ασφάλεια και η καλή κατάσταση των ατόμων της κλινικής δοκιμής προστατεύονται σύμφωνα με τις αρχές που προέκυψαν από τη Διακήρυξη του Ελσίνκι, και ότι τα δεδομένα της κλινικής δοκιμής είναι αξιόπιστα.

Ο σκοπός αυτής της Οδηγίας ICH GCP (Οδηγία Διεθνούς Διάσκεψης Εναρμόνισης Ορθής Κλινικής Πρακτικής) είναι να δώσει ένα ενιαίο πρότυπο για την Ευρωπαϊκή Ένωση, την Ιαπωνία και τις Ηνωμένες Πολιτείες της Αμερικής, ώστε να διευκολύνεται η αμοιβαία αποδοχή των κλινικών δεδομένων από τις Κρατικές Αρχές Υγείας των αντίστοιχων χωρών. Η οδηγία διαμορφώθηκε σύμφωνα με την τρέχουσα ορθή πρακτική διεξαγωγής των κλινικών δοκιμών στην Ευρωπαϊκή Ένωση, την Ιαπωνία και τις Ηνωμένες Πολιτείες, καθώς και στην Αυστραλία, τον Καναδά, τις Σκανδιναβικές χώρες και τον Παγκόσμιο Οργανισμό

Υγείας.

Η οδηγία πρέπει να τηρείται όταν συλλέγονται δεδομένα κλινικών δοκιμών, τα οποία πρόκειται να υποβληθούν στις Υγειονομικές Αρχές. Οι αρχές που καθορίζονται στην οδηγία αυτή μπορούν επίσης να εφαρμόζονται και σε άλλες κλινικές δοκιμές, οι οποίες ενδέχεται να συνεπάγονται κάποια επίδραση στην ασφάλεια και την καλή κατάσταση των ανθρώπων που συμμετέχουν σ' αυτές.

ΑΡΧΕΣ ΤΗΣ ΔΙΕΘΝΟΥΣ ΔΙΑΣΚΕΨΗΣ ΕΝΑΡΜΟΝΙΣΗΣ ΟΡΘΗΣ ΚΛΙΝΙΚΗΣ ΠΡΑΚΤΙΚΗΣ

1. Οι κλινικές δοκιμές πρέπει να διεξάγονται σύμφωνα με τις δεοντολογικές αρχές που προέκυψαν από τη Διακήρυξη του Ελσίνκι και να είναι συμβατές με την Ορθή πρακτική κλινικών δοκιμών (GCP) και τη σχετική ρυθμιστική διάταξη.
2. Πριν από την έναρξη μιας κλινικής δοκιμής, οι προβλέψιμοι κίνδυνοι και οι δυσχέρειες πρέπει να εξετάζονται σε αντιπαράθεση με την αναμενόμενη ωφέλεια που θα προκύψει τόσο για το συγκεκριμένο άτομο της κλινικής δοκιμής, όσο και για την κοινωνία.
3. Μια κλινική δοκιμή πρέπει να αρχίζει και να συνεχίζεται μόνον εάν τα αναμενόμενα πλεονεκτήματα δικαιολογούν τους κινδύνους.
4. Τα δικαιώματα, η ασφάλεια και η κα-

- λή κατάσταση (well being) των ατόμων της κλινικής δοκιμής είναι τα πιο σημαντικά ζητήματα, τα οποία πρέπει να υπερισχύουν έναντι των συμφερόντων της επιστήμης και της κοινωνίας.
5. Οι διαθέσιμες μη-κλινικές και κλινικές πληροφορίες για κάποιο διερευνούμενο προϊόν θα πρέπει να είναι επαρκείς για να υποστηρίξουν την προτεινόμενη κλινική δοκιμή.
 6. Οι κλινικές δοκιμές θα πρέπει να είναι επιστημονικώς τεκμηριωμένες και να περιγράφονται σαφώς και λεπτομερώς στο πρωτόκολλο της κλινικής δοκιμής.
 7. Κάθε κλινική δοκιμή πρέπει να διεξάγεται σύμφωνα με το πρωτόκολλο της κλινικής δοκιμής το οποίο προηγουμένως επικυρώθηκε, κρίθηκε ευνοϊκά από το Συμβούλιο Ελέγχου του Ιδρύματος και την Ανεξάρτητη Επιτροπή Ηθικής και Δεοντολογίας (Independent Ethics Committee - IEC).
 8. Η παρεχόμενη ιατρική περίθαλψη και οι ιατρικές αποφάσεις που λαμβάνονται για λογαριασμό των ατόμων της κλινικής δοκιμής πρέπει να γίνονται πάντοτε υπό την ευθύνη ιατρού ή οδοντιάτρου (όπου αυτό έχει εφαρμογή), που διαθέτει τα αναγκαία προσόντα.
 9. Κάθε πρόσωπο που συμμετέχει στη διεξαγωγή μιας κλινικής δοκιμής θα πρέπει να έχει τα κατάλληλα προσόντα από απόψεως εκπαίδευσης, εξασκήσεως και εμπειρίας, προκειμένου να ασκήσει τα αντίστοιχα καθήκοντα.
 10. Πρέπει να υπάρχει συγκατάθεση κατόπιν ενημερώσεως, παρεχόμενη με την ελεύθερη βούληση, κάθε ατόμου της κλινικής δοκιμής, πριν από τη συμμετοχή του στην κλινική δοκιμή.
 11. Όλες οι πληροφορίες της κλινικής δοκιμής πρέπει να καταχωρούνται, να τυχαίνουν χειρισμού και να φυλάσσονται κατά τρόπο που επιτρέπει την ακριβή αναφορά, ερμηνεία και επαλήθευσή τους.
 12. Η εμπιστευτικότητα των αρχείων τα οποία θα μπορούσαν να αποκαλύψουν την ταυτότητα ατόμων της κλινικής δοκιμής οφείλει να τηρείται, ώστε να γίνονται σεβαστοί οι κανόνες της ιδιωτικής ζωής και των εμπιστευτικών πληροφοριών, σύμφωνα με τις ισχύουσες ρυθμιστικές διατάξεις.
 13. Τα διερευνούμενα προϊόντα πρέπει να παρασκευάζονται και να τυχαίνουν τέτοιου χειρισμού και αποθηκεύσεως, που να είναι σε συμφωνία με την επί του προκειμένου ορθή πρακτική παρασκευής (GMP, Good Manufacturing Practice). Θα πρέπει ακόμη αυτά να χρησιμοποιούνται σύμφωνα με το εγκεκριμένο πρωτόκολλο της κλινικής δοκιμής.
 14. Πρέπει να εφαρμόζονται συστήματα με διαδικασίες που εξασφαλίζουν την ποιότητα σε κάθε πλευρά της κλινικής δοκιμής.

ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ ΚΛΙΝΙΚΩΝ ΜΕΛΕΤΩΝ

Προετοιμασία για την ανάλυση δεδομένων

Πριν από την ανάλυση ενός συνόλου δεδομένων είναι σημαντικό να ελεγχθεί στον μεγαλύτερο δυνατό βαθμό εάν τα δεδομένα είναι σωστά, Σφάλματα ενδέχεται να γίνουν κατά την πραγματοποίηση των μετρήσεων, κατά την αρχική καταγραφή των δεδομένων, κατά τη μεταγραφή τους από την αρχική πηγή (ό-

πως π.χ. από τα φύλλα νοσηλείας) ή κατά την πληκτρολόγησή τους στον ηλεκτρονικό υπολογιστή (Η/Υ) Η διαδικασία αυτή λέγεται έλεγχος (ή ξεκαθάρισμα) των δεδομένων. Είναι σημαντικό, επίσης, να γίνει σχολαστικός έλεγχος για τη διαπίστωση τυχόν ύπαρξης χαρακτηριστικών τα οποία ενδέχεται να προκαλέσουν δυσχέρειες κατά την ανάλυση. Πρέπει να εξεταστούν κυρίως τρία σημεία: τα στοιχεία που λείπουν, οι τιμές που διαφέρουν πολύ και η πιθανότητα να υπάρχει ανάγκη μετασχηματισμών.

Έλεγχος δεδομένων

Τα σφάλματα σε καταχωρημένα δεδομένα είναι συνήθη. Ο έλεγχος των δεδομένων αποσκοπεί στην αναγνώριση των σφαλμάτων και, εάν είναι δυνατόν, στη διόρθωσή τους. Είναι προφανές ότι τα σφάλματα στα αρχικά δεδομένα δεν μπορούν να διορθωθούν, αλλά τα λάθη που παρεισέφησαν σε κάποιο μετέπειτα στάδιο μπορούν να διορθωθούν, εάν συμβουλευτούμε την αρχική καταγραφή των δεδομένων. Ένα σημαντικό πρώτο βήμα είναι να ελέγξουμε εάν τα δεδομένα πληκτρολογήθηκαν σωστά στον Η/Υ. Για μεγάλους φακέλους η διπλή καταχώρηση είναι η καλύτερη μέθοδος. Με τη μέθοδο αυτή τα δεδομένα πληκτρολογούνται ξανά και συγκρίνονται με την πρώτη καταχώρηση, κατά προτίμηση με τη βοήθεια ενός προγράμματος του Η/Υ, το οποίο σχεδιάστηκε ειδικά για τον σκοπό αυτό. Για μικρά σύνολα δεδομένων, ο απλούστερος τρόπος είναι ένα πρόσωπο να διαβάσει δυνατά τα δεδομένα από την οθόνη του Η/Υ και ένα άλλο να τα ελέγχει σε σύγκριση με τα αρχικά δεδομένα (αποδεικτική ανάγνωση).

Προκειμένου για τις μεταβλητές κα-

τηγοριών, ο απλούστερος τρόπος είναι να ελέγξουμε αν όλες οι καταχωρημένες τιμές δεδομένων είναι οι σωστές, επειδή υπάρχει σταθερός αριθμός προκαθορισμένων τιμών. Σε ό,τι αφορά τις συνεχείς μετρήσεις, δεν μπορούμε συνήθως να διακρίνουμε ποιες τιμές είναι σωστές και ποιες δεν είναι, και δεν είναι σημαντικό να το κάνουμε. Εντούτοις, θα μπορούσαμε να προσδιορίσουμε τα κατώτερα και τα ανώτερα λογικά όρια σε ό,τι αφορά τη συγκεκριμένη μεταβλητή.

Τιμές εκτός ορίων

Ο έλεγχος των δεδομένων συνεχών μεταβλητών μπορεί να αποκαλύψει τιμές εκτός ορίων, οι οποίες είναι ασυμβίβαστες με τα υπόλοιπα δεδομένα. Εάν δεν υπάρχουν ενδείξεις σφάλματος και η τιμή είναι εύλογη, τότε η τιμή αυτή δεν πρέπει να μεταβληθεί ή να εξαιρεθεί. Μια εξαίρεση στον κανόνα αυτόν είναι η περίπτωση όπου η τιμή είναι σωστή, αλλά η έρευνα αποκαλύπτει ότι υπάρχει κάτι ιδιαίτερο στο άτομο αυτό, όπως π.χ. όταν συνυπάρχει κάποιο νόσημα. Εδώ ενδέχεται να είναι λογικός ο αποκλεισμός αυτής της τιμής. Είναι πολύ επικίνδυνη η αφαίρεση τιμών που απλώς είναι οι μέγιστες ή οι ελάχιστες. Υπάρχουν στατιστικές μέθοδοι που μπορούν να χρησιμοποιηθούν για την ανίχνευση των ύποπτων τιμών, αλλά δεν πρέπει να χρησιμοποιούνται για τον προσδιορισμό του τι συμβαίνει με αυτές.

Οι εκτός ορίων τιμές είναι ιδιαίτερα σημαντικές, επειδή μπορούν να επηρεάσουν σημαντικά τα αποτελέσματα μιας στατιστικής ανάλυσης. Επειδή εξ ορισμού είναι ακραίες τιμές, το να συμπεριληφθούν ή να αποκλειστούν μπορεί να έχει έκδηλη επίδραση στα αποτελέσματα μιας ανάλυσης. Οι εκτός ορίων τιμές μπορούν να επηρεάσουν πολλούς

τύπους στατιστικής ανάλυσης, προκαλώντας συχνά αύξηση της μεταβλητότητας ενός συνόλου παρατηρήσεων και αλλοιώνοντας κατά συνέπεια το αποτέλεσμα που ενδιαφέρει. Μια χρήσιμη στρατηγική που μπορεί να υιοθετηθεί κατά την ανάλυση δεδομένων είναι η πραγματοποίηση της ανάλυσης τόσο με το να συμπεριληφθούν, όσο και με το να αποκλειστούν οι ύποπτες τιμές. Εάν υπάρχει μικρή διαφορά στα αποτελέσματα, τότε οι εκτός ορίων τιμές είχαν ελάχιστη επίδραση, αλλά εάν ο αποκλεισμός τους συνεπάγεται σημαντική διαφορά, τότε θα είναι καλύτερο να βρεθεί μια εναλλακτική μέθοδος ανάλυσης. Οι μέθοδοι της κατάταξης, οι οποίες εισήχθησαν αργότερα, μπορεί να αποτελούν μια καλή προσέγγιση στις περιπτώσεις αυτές.

Ελλείπονταδεδομένα

Σε περίπτωση απουσίας δεδομένων είναι απαραίτητο να προβληματιστούμε γιατί λείπουν οι τιμές αυτές. Ιδιαίτερα θα πρέπει να γνωρίζουμε εάν υπάρχει κάποιος λόγος που έχει σχέση με τη φύση της δοκιμής. Συχνά, λείπουν τιμές, συνήθως κατά τύχη, για λόγους που δεν σχετίζονται με τη φύση της δοκιμής. Σε μερικούς ασθενείς, για παράδειγμα, έχει παραλειφθεί να γίνει μια συγκεκριμένη ερώτηση ή ένα δείγμα αίματος χάθηκε ή καταστράφηκε. Η απουσία δεδομένων μπορεί εντούτοις να είναι αποκαλυπτική των αδυναμιών στο σχεδιασμό της μελέτης. Σε μια δοκιμή όπου οι πληροφορίες για τον ασθενή συλλέγονται σε διαφορετικές στιγμές, η έλλειψη πληροφοριών στον μετέπειτα χρόνο μπορεί να οφείλεται στην απόσυρση του ασθενούς από τη δοκιμή, λόγω παρενεργειών. Συνήθως χρησιμοποιούνται ειδικοί αριθμητικοί κώδικες απόσυρσης 9, 99, 999 ή

99,9 (αναλόγως της φύσης της μεταβλητής) αν και μερικά προγράμματα Η/Υ επιτρέπουν το «+» ή κάποιο άλλο σύμβολο το οποίο υποδηλώνει την παρατήρηση που λείπει. Εάν χρησιμοποιείται μια αριθμητική τιμή, είναι απαραίτητο, πριν από την ανάλυση των δεδομένων, να χαρακτηριστεί η τιμή ως ελλείπουσα τιμή στο στατιστικό πρόγραμμα.

Κατανομή των παρατηρήσεων

Οι περισσότερες στατιστικές μέθοδοι (παραμετρικές μέθοδοι) για την ανάλυση συνεχών δεδομένων περιλαμβάνουν την παραδοχή ότι τα δεδομένα προέρχονται από έναν πληθυσμό όπου οι τιμές είναι κανονικά κατανομημένες. Συχνά παρατηρούμε ότι η εφαρμογή μετασχηματισμού των δεδομένων δίνει κατανομή που είναι πολύ πλησιέστερη προς μια κανονική κατανομή. Ο πλέον συχνός μετασχηματισμός είναι ο λογαριθμικός. Άλλοι μετασχηματισμοί που χρησιμοποιούνται συχνά είναι ο μετασχηματισμός της τετραγωνικής ρίζας και ο αντίστροφος μετασχηματισμός. Εναλλακτικές μέθοδοι που δεν κάνουν αυτή την παραδοχή και βασίζονται σε σειρές είναι συνήθως διαθέσιμες, αλλά έχουν ορισμένα μειονεκτήματα.

Είναι συνεπώς σημαντικό να γνωρίζουμε την κατανομή των δεδομένων πριν ξεκινήσουμε μια ανάλυση που βασίζεται στην παραδοχή της κανονικότητας. Για κάθε συνεχή μεταβλητή πρέπει να υπολογίζεται η μέση τιμή (mean) και η τυπική απόκλιση (standard deviation). Εάν είναι δυνατόν, πρέπει να σχεδιάζεται ένα ιστόγραμμα για να γίνει αντιληπτό το σχήμα της κατανομής. Αν αυτό δεν είναι δυνατόν τότε εξετάζονται τα ποσοστιαία σημεία της κατανομής, για να ελεγχθεί κατά πόσον η κατανομή φαίνεται συμμετρική. Ιδιαίτερα για μι-

κρά δείγματα η εκτίμηση του βαθμού της κανονικότητας ενός συνόλου δεδομένων μπορεί να είναι δυσχερής.

Ένας τρόπος μέτρησης της "μη κανονικότητας" είναι ο υπολογισμός αυτών που αποκαλούνται «ροπές» της κατανομής των δεδομένων. Οι πρώτες δύο ροπές έχουν ήδη αναφερθεί και είναι η μέση τιμή (ροπή τάξης ένα περί την αρχή) και η διακύμανση (κεντρική ροπή δεύτερης τάξης). Εντούτοις, οι τιμές αυτές δεν μας δίνουν πληροφορίες σχετικά με το σχήμα της κατανομής. Υπάρχει δυνατότητα προσδιορισμού του σχήματος κατανομής με τη βοήθεια των εξής παραμέτρων: (α) ασυμμετρία ή λοξότητα η οποία ορίζει το μέτρο της ασυμμετρίας, και (β) κύρτωση η οποία ορίζει πόσο αποπλευρωμένη ή αιχμηρή είναι η καμπύλη. Οι τιμές τους μπορούν να συγκριθούν με τις θεωρητικές τιμές της κανονικής κατανομής.

Σχεδιάζοντας την έρευνα

Ο στατιστικός σχεδιασμός σχετίζεται κυρίως με (α) τη δομή της δοκιμής, (β) τη συλλογή των δεδομένων, (γ) την επιλογή των μετρήσεων (δ) τη συχνότητά τους και (ε) τη μεθοδολογία τους. Η έρευνα μπορεί αδρά να διαιρεθεί σε κλινικές δοκιμές παρατήρησης και σε πειραματικές κλινικές δοκιμές. Στις κλινικές δοκιμές παρατήρησης συλλέγουμε πληροφορίες σχετικά με μία ή περισσότερες ομάδες ανθρώπων, αλλά δεν κάνουμε τίποτε που θα τους επηρεάσει. Ένα σχετικό παράδειγμα είναι μια δοκιμή για τη διαπίστωση της συχνότητας διαταραχών της ακοής σε μικρά παιδιά. Οι κλινικές δοκιμές παρατήρησης μπορεί να είναι προοπτικές, όταν τα άτομα «στρατολογούνται» και κατόπιν συλλέγονται δεδομένα συμβαμάτων που θα επακολουθήσουν, ή αναδρομικές, όταν συλλέγο-

νται πληροφορίες σχετικά με συμβάντα του παρελθόντος που μπορούν να αποκτηθούν από ήδη υπάρχουσες πηγές, π.χ. από τους νοσοκομειακούς φακέλους. Οι κλινικές δοκιμές παρατήρησης διακρίνονται σε απογραφές, έρευνες/επισκοπήσεις, κλινικές δοκιμές ασθενών-μαρτύρων, και κλινικές δοκιμές ομάδων πληθυσμού (cohort studies).

Πειραματικές κλινικές δοκιμές είναι εκείνες στις οποίες ο ερευνητής επηρεάζει (ελέγχει) το τι θα συμβεί σε όλα ή σε μερικά από τα άτομα της δοκιμής. Οι πειραματικές κλινικές δοκιμές συνήθως αποσκοπούν στο να γίνουν συγκρίσεις μεταξύ δύο ή περισσότερων ομάδων. Οι κλινικές δοκιμές παρατήρησης μπορούν επίσης να είναι συγκριτικές, αλλά συχνά είναι στην ουσία περιγραφικές. Υπάρχουν δύο στοιχεία που χρειάζονται ιδιαίτερη προσοχή από την άποψη αυτή. Πρώτον, το υπό δοκιμή δείγμα πρέπει να είναι αντιπροσωπευτικό του πληθυσμού της δοκιμής. Δεύτερον, οι συγκρινόμενες ομάδες πρέπει να μοιάζουν όσο είναι δυνατόν περισσότερο, πλην των χαρακτηριστικών αμέσου ενδιαφέροντος. Αυτό ισχύει ιδιαίτερα για τις πειραματικές κλινικές δοκιμές, αλλά αφορά και πολλές κλινικές δοκιμές παρατήρησης, όπως οι κλινικές δοκιμές ελέγχου περιπτώσεων.

Κατηγορίες σχεδιασμού της έρευνας

Ως προς την «κατά σχεδιασμό» κατάταξη η έρευνα ταξινομείται ως εξής: 1) παρατήρησης ή πειραματική, 2) προοπτική ή αναδρομική, 3) διαρκείας ή εγκάρσιας τομής. Η πρώτη ταξινόμηση σχετίζεται άμεσα με τον σκοπό της δοκιμής, ενώ οι άλλες δύο περιγράφουν τον τρόπο συλλογής των δεδομένων.

Έρευνα διαρκείας ή εγκάρσιας τομής

Μελέτες διαρκείας είναι εκείνες οι οποίες ερευνούν μεταβολές σε συνάρτηση με τον χρόνο, πιθανώς σε σχέση με κάποια παρέμβαση. Για όλες ή κάποιες από τις μεταβλητές αυτές μπορεί να πραγματοποιηθούν περισσότερες από μία μετρήσεις αλλά δεν είναι απαραίτητο να χρησιμοποιηθούν όλες κατά την ανάλυση. Οι κλινικές δοκιμές είναι δοκιμές διαρκείας. επειδή αποσκοπούν στην αξιολόγηση του αποτελέσματος μιας θεραπείας στη διάρκεια του χρόνου. Κλινικές δοκιμές εγκάρσιας τομής είναι εκείνες όπου η παρατήρηση των ατόμων γίνεται μόνο μια φορά.

Κλινικές δοκιμές

Η κλινική δοκιμή είναι μία προγραμματισμένη ερευνητική διαδικασία σε ανθρώπους, η οποία σχεδιάζεται για την αξιολόγηση της αποτελεσματικότητας μιας ή περισσότερων μορφών θεραπείας (όπως φάρμακα, χειρουργικές επεμβάσεις, φυσικοθεραπεία, δίαιτα, εκπαίδευση σε θέματα υγείας κ.ο.κ.). Οι φαρμακευτικές κλινικές δοκιμές ταξινομούνται σε μια από τις ακόλουθες τέσσερις κατηγορίες μελετών:

1. Φάσης I: Κλινική φαρμακολογία και τοξικότητα του μελετωμένου φαρμάκου
2. Φάσης II: Αρχική κλινική έρευνα του μελετωμένου φαρμάκου
3. Φάσης III: Ευρείας κλίμακας αξιολόγηση της μελετωμένης θεραπείας
4. Φάσης IV: Επίβλεψη μετά την κυκλοφορία του φαρμάκου.

Στην πράξη, στη μεγάλη τους πλειονότητα οι συγκριτικές κλινικές δοκιμές έχουν ορισμένα κοινά χαρακτηριστικά, τα οποία επιτρέπουν να δοθούν γενικές οδηγίες σχεδιασμού, ανάλυσης και ερμηνείας. Ίσως για τον λόγο αυτό, οι κλινικές δοκιμές είναι πιθανώς η περιο-

χή της κλινικής έρευνας όπου η ολοκλήρωση των ιδεών και της μεθοδολογίας της στατιστικής υπήρξε περισσότερο επιτυχής. Η κεντρική ιδέα μιας κλινικής δοκιμής είναι ότι θέλουμε να συγκρίνουμε ομάδες ασθενών, οι οποίοι διαφέρουν μόνο ως προς το είδος της θεραπείας στην οποία υποβάλλονται. Εάν οι ομάδες διαφέρουν και ως προς κάποια άλλη παράμετρο η σύγκριση των θεραπευτικών μεθόδων είναι μεροληπτική.

Τυχαία κατανομή

Η ευρύτερα διαδεδομένη μέθοδος κατανομής της θεραπείας που αποτρέπει τα σφάλματα από μεροληψία είναι η τυχαία κατανομή, η οποία δίνει σε όλα τα άτομα της δοκιμής ισοδύναμη πιθανότητα να λάβουν το κάθε είδος θεραπείας. Ένας άλλος σημαντικός λόγος για τη χρησιμοποίηση της τυχαίας δειγματοληψίας είναι ότι οι μέθοδοι στατιστικής ανάλυσης βασίζονται στο τι αναμένεται να συμβεί σε τυχαία δείγματα πληθυσμών με καθορισμένα χαρακτηριστικά.

Απλή τυχαιοποίηση

Η απλούστερη μέθοδος τυχαίας κατανομής είναι το «στρίψιμο» ενός νομίσματος: «κορώνα» για τη θεραπεία Α και «γράμματα» για τη θεραπεία Β. Μια ισοδύναμη μέθοδος είναι η χρησιμοποίηση ενός πίνακα τυχαίων αριθμών. Τέτοιοι πίνακες συνοδεύουν όλα τα εγχειρίδια Στατιστικής ως προσαρτήματα. Μπορούν επίσης να κατασκευασθούν από Η/Υ με την χρήση απλών και εύχρηστων προγραμμάτων σε γλώσσα BASIC. Στους πίνακες αυτούς κάθε αριθμός απαντάται με την ίδια συχνότητα. Η κατάταξη των αριθμών είναι τυχαία και συνεπώς απρόβλεπτη. Το πρώτο βήμα είναι να αποφασίσουμε την αντιστοιχία μεταξύ των τυχαίων αριθμών

και των διαφόρων ομάδων της έρευνας. Για παράδειγμα, θα μπορούσαμε να πάρουμε τους περιττούς αριθμούς να αντιστοιχούν στο ένα είδος θεραπείας και τους άρτιους στο άλλο. Πρέπει τότε να διαλέξουμε ένα σημείο εκκίνησης και αυτό μπορεί να γίνει με τη χρησιμοποίηση μιας καρφίτσας ή κάποιας εξίσου αυθαίρετης μεθόδου. Επίσης, μπορούμε να διαλέξουμε την κατεύθυνση ανάγνωσης του πίνακα.

Ας υποθέσουμε ότι οι πρώτοι διψήφιοι αριθμοί είναι: 12 19 20 52 81 30 74 κ.λ.π. Εάν ορίσουμε τους περιττούς αριθμούς για τη θεραπεία Α και τους άρτιους για τη θεραπεία Β, τότε αυτοί οι αριθμοί μας δίνουν μια ακολουθία Β Α Β Β Α Β Β για τους πρώτους επτά ασθενείς. Υπάρχουν άπειρες πιθανές στρατηγικές που μπορούν να εφαρμοστούν. Βασικός κανόνας για την αποφυγή μεροληψίας είναι να έχουν επιλεγεί οι συγκεκριμένοι πίνακες που θα χρησιμοποιηθούν πριν την στρατολόγηση του πρώτου αρρώστου και να έχει προβλεφθεί ικανός αριθμός τυχαίων αριθμών. Προσθήκη αριθμών ή επέκταση του πίνακα μετά την έναρξη της έρευνας συνιστά μεροληψία.

Τετραγωνισμένη τυχαιοποίηση

Η τετραγωνισμένη τυχαιοποίηση χρησιμοποιείται για να παραμείνει σταθερός ο αριθμός των ατόμων στις διάφορες ομάδες, πάντοτε σε στενή ισορροπία αναλογίας θεραπειών. Για παράδειγμα, εάν κατανείμουμε τα άτομα σε ομάδες τεσσάρων κάθε φορά, υπάρχουν έξι τρόποι κατανομής της θεραπείας, ώστε να υπάρχουν στην κάθε ομάδα από δύο άτομα με τη θεραπεία Α και δύο με τη θεραπεία Β:

1. Α Α Β Β
2. Α Β Α Β
3. Α Β Β Α

4. Β Β Α Α

5. Β Α Β Α

6. Β Α Α Β

Για να δημιουργήσουμε την αλληλουχία της τυχαιοποίησης διαλέγουμε τις ομάδες στην τύχη. Έστω ότι χρησιμοποιούμε την προηγούμενη τυχαία σειρά αριθμών: 12 19 20 52 81 30 74. Παραλείποντας τα ψηφία που βρίσκονται εκτός των ορίων 1 έως 6, παίρνουμε για το 1: Α Α Β Β, για το 2: Α Β Α Β, για το 3: Α Β Β Α, για το 4: Β Β Α Α, για το 5: Β Α Β Α, για το 6: Β Α Α Β, από τα οποία συντάσσουμε την αλληλουχία τετραγωνισμένης τυχαιοποίησης:

Α Α Β Β, Α Β Α Β, Α Α Β Β, Α Β Α Β, Β Α Β Α, Α Β Α Β, Α Α Β Β, Α Β Β Α, Β Β Α Α

Τυχαιοποίηση κατά στρώματα

Ενώ η απλή τυχαιοποίηση απομακρύνει τους συγχυτικούς παράγοντες από τη διαδικασία της τυχαιοποίησης δεν εγγυάται ότι τα άτομα κάθε ομάδας έχουν παρόμοια χαρακτηριστικά. Οποιοσδήποτε τυχόν διαφορές μεταξύ των ομάδων οφείλονται στην τύχη, αλλά παρόμοιες διαφορές είναι τουλάχιστον άβολες και μπορούν να οδηγήσουν σε αμφιβολίες κατά την ανάλυση των αποτελεσμάτων της δοκιμής. Εφόσον η τροποποίηση της ανάλυσης είναι δυνατή ώστε να λάβει υπ' όψιν τυχόν διαφορές μεταξύ των ομάδων με την έναρξή της, είναι προτιμότερο να ελέγξουμε το πρόβλημα στο στάδιο του σχεδιασμού.

Μπορούμε να χρησιμοποιήσουμε την τυχαιοποίηση κατά στρώματα, ώστε να επιτύχουμε κατά προσέγγιση ισορροπία των σημαντικών χαρακτηριστικών, χωρίς να θυσιάσουμε το πλεονέκτημα της τυχαιοποίησης. Η μέθοδος συνίσταται στη δημιουργία χωριστού καταλόγου τετραγωνισμένης τυχαιοποίησης για κάθε υποομάδα (στρώμα). Για παράδειγμα, σε

μια δοκιμή σύγκρισης δύο εναλλακτικών μεθόδων θεραπείας του καρκίνου του μαστού, θα ήταν σημαντική η διαστρωμάτωση αναλόγως της κατάστασης από πλευράς εμμηνόπαυσης. Πρέπει να σχηματιστούν χωριστοί κατάλογοι τυχαίων αριθμών για τις προ- και μετά-εμμηνοπαυσιακές γυναίκες. Είναι σημαντικό η κατά στρώματα κατανομή της θεραπείας να βασίζεται σε τετραγωνισμένη τυχαιοποίηση μέσα σε κάθε στρώμα μάλλον, παρά σε απλή τυχαιοποίηση, γιατί χωρίς αυτή δεν θα υπήρχε έλεγχος της ισορροπίας εντός των στρωμάτων.

Η τυχαιοποίηση κατά στρώματα μπορεί να επεκταθεί σε δύο ή περισσότερες μεταβλητές στρωματοποίησης. Για παράδειγμα, θα μπορούσαμε να επεκτείνουμε τη διαστρωμάτωση στη δοκιμή του καρκίνου του μαστού, ως προς το μέγεθος του όγκου και τον αριθμό των θετικών λεμφαδένων. Θα χρειαστεί να συντάξουμε χωριστό κατάλογο τυχαιοποίησης για κάθε συνδυασμό κατηγοριών. Εάν είχαμε δύο ομάδες ως προς το μέγεθος του όγκου (π.χ. ≤ 3 και > 3 εκ.), τρεις ομάδες ως προς την επινέμηση των λεμφαδένων (0, 1 - 4, και > 4), και δύο ομάδες ως προς την εμμηνόπαυση (προ-, μετά-), τότε θα έχουμε: $2 \times 3 \times 2 = 12$ στρώματα, κάτι που μπορεί να είναι έξω από τα όρια του πρακτικού. Σημαντικό πρόβλημα με τα πολλαπλά στρώματα είναι ότι μερικοί από τους συνδυασμούς των κατηγοριών ενδέχεται να είναι σπάνιοι, έτσι ώστε να μην υπάρξει τελικά η ισορροπία θεραπείας την οποία αναμένουμε από τη χρησιμοποίηση της τετραγωνισμένης τυχαιοποίησης. Κατά συνέπεια, πρέπει να σκεφτούμε για το ποιες μεταβλητές πρέπει να χρησιμοποιηθούν για τη στρωματοποίηση, περιορίζοντας την επιλογή στις μεταβλη-

τές που γνωρίζουμε ότι είναι σημαντικές ως προς την πρόγνωση. Σε πολλές κλινικές δοκιμές η διαστρωμάτωση γίνεται με τη χρησιμοποίηση της ηλικίας και του φύλου. Σε μια πολυκεντρική δοκιμή, οι ασθενείς κάθε κέντρου θα χρειαστεί να τυχαιοποιηθούν χωριστά, εκτός εάν υπάρχει κεντρική υπηρεσία συντονισμένης τυχαιοποίησης. Έτσι το «κέντρο» αποτελεί μια μεταβλητή διαστρωμάτωσης

Εναλλακτικοί σχεδιασμοί

Ο απλούστερος σχεδιασμός για μια κλινική δοκιμή λέγεται σχεδιασμός παράλληλης ομάδας, όπου δύο διαφορετικές ομάδες ασθενών μελετώνται συγχρόνως. Η συνηθέστερη εναλλακτική μέθοδος είναι ο σχεδιασμός διασταύρωσης. Η δοκιμή διασταύρωσης είναι εκείνη κατά την οποία στην ίδια ομάδα ασθενών δίνονται και οι δύο ή όλες οι θεραπείες που μας ενδιαφέρουν διαδοχικά. Εδώ η τυχαιοποίηση χρησιμοποιείται για να προσδιοριστεί η σειρά με την οποία λαμβάνονται οι θεραπείες. Ο σχεδιασμός διασταύρωσης έχει μερικά ελκυστικά χαρακτηριστικά, με πιο σημαντικό το ότι η σύγκριση των θεραπειών γίνεται στον ίδιο ασθενή παρά μεταξύ διαφορετικών ασθενών. Υπάρχουν εντούτοις κάποια σημαντικά μειονεκτήματα:

1) Οι ασθενείς ενδέχεται να αποσυρθούν από τη δοκιμή μετά το πρώτο θεραπευτικό σχήμα, χωρίς να λάβουν το δεύτερο (θεραπευτικό σχήμα). Η απόσυρση ενδέχεται να σχετίζεται με παρενέργειες της θεραπείας. Οι περίοδοι θεραπείας πρέπει να είναι αρκετά σύντομες, για να ελαχιστοποιηθεί ο κίνδυνος της απόσυρσης για άλλους λόγους.

2) Ενδέχεται να υπάρχει μεταφερόμενη διάρκεια του θεραπευτικού αποτελέσματος από τη μία περίοδο στην άλλη,

έτσι ώστε τα αποτελέσματα της δεύτερης θεραπείας να επηρεάζονται από αυτά της πρώτης περιόδου. Με άλλα λόγια, η διαφορά που παρατηρείται μεταξύ των θεραπευτικών σχημάτων εξαρτάται από τη σειρά με την οποία αυτά εφαρμόστηκαν. Όταν υπάρχει τέτοια αλληλεπίδραση μεταξύ θεραπείας και περιόδου, τα δεδομένα της δεύτερης περιόδου μπορεί να χρειαστεί να απορριφθούν, και έτσι να εξασθενήσει σοβαρά η ισχύς της δοκιμής.

3) Οι σχεδιασμοί διασταύρωσης είναι πιο κατάλληλοι όταν το αποτέλεσμα της θεραπείας μπορεί να αξιολογηθεί γρήγορα.

4) Ενδέχεται να υπάρχει κάποια συστηματική διαφορά μεταξύ δύο περιόδων της δοκιμής. Για παράδειγμα, οι παρατηρήσεις της δεύτερης περιόδου ενδέχεται να είναι κάπως χαμηλότερες από εκείνες της πρώτης περιόδου, ανεξαρτήτως της θεραπείας. Ένα μικρό αποτέλεσμα περιόδου δεν είναι τόσο σοβαρό, αφού αφορά εξίσου και τις δύο θεραπείες.

Είναι επιθυμητό να γνωρίζουμε προκαταβολικά ότι δεν θα υπάρξει μεταφερόμενη διάρκεια του θεραπευτικού αποτελέσματος, αλλά η πληροφορία αυτή μπορεί να μην είναι διαθέσιμη. Γι' αυτό το λόγο, πολλές φορές, παρεμβάλλεται μια περίοδος «έκπλυσης» μεταξύ των περιόδων θεραπείας, στην προσπάθεια να περιοριστούν οι επιπτώσεις της μεταφερόμενης διάρκειας του θεραπευτικού αποτελέσματος

Διαδοχικοί σχεδιασμοί

Ένας άλλος τύπος σχεδιασμού δοκιμής είναι η διαδοχική δοκιμή, όπου μελετώνται παράλληλες ομάδες, αλλά η δοκιμή συνεχίζεται μέχρις ότου παρατηρηθεί ένα σαφές πλεονέκτημα από το

ένα είδος θεραπείας ή διαπιστωθεί ότι είναι απίθανο να υπάρξει κάποια διαφορά μεταξύ των ειδών θεραπείας. Οι διαδοχικές κλινικές δοκιμές έχουν το πλεονέκτημα ότι είναι βραχύτερες από τις κλινικές δοκιμές καθορισμένης διάρκειας, όταν υπάρχει μεγάλη διαφορά ως προς την αποτελεσματικότητα των δύο ειδών θεραπείας.

Σύγκριση ομάδων – Συνεχή Δεδομένα

Βασιζόμενοι στη θεωρία της εκτίμησης παραμέτρων και του ελέγχου υποθέσεων υπάρχει δυνατότητα να εξετάσουμε τις κύριες μεθόδους της ανάλυσης δεδομένων. Αν και οι μέθοδοι της αντιμετώπισης περισσότερων ομάδων παρατηρήσεων μπορούν να χρησιμοποιηθούν και για μία ή δύο ομάδες, είναι βολικό να εξετάζονται οι περιπτώσεις της μιας ή των δύο ομάδων χωριστά. Έτσι, οι μέθοδοι μπορούν να απλοποιηθούν και να μειωθούν τα προβλήματα ερμηνείας. Η περίπτωση των δύο ομάδων είναι ο πιο συνηθισμένος τύπος στατιστικής ανάλυσης. Όταν υπάρχουν δύο ή περισσότερα σύνολα παρατηρήσεων, υπάρχουν δύο τύποι σχεδιασμού που πρέπει να διακρίνονται:

1. Οι παρατηρήσεις αφορούν ανεξάρτητες ομάδες ατόμων. Για παράδειγμα, μπορούμε να έχουμε ομάδες ασθενών με διαφορετικά νοσήματα. Το μέγεθος του δείγματος μπορεί να ποικίλλει από ομάδα σε ομάδα.

2. Κάθε σύνολο παρατηρήσεων γίνεται στην ίδια ομάδα ατόμων. Για παράδειγμα, μπορούμε να έχουμε προγεννητική και μεταγεννητική παρακολούθηση της αρτηριακής πίεσεως σε μια ομάδα γυναικών. Αποκαλούμε τα δεδομένα αυτά «κατά ζεύγη» για να δείξουμε ότι οι παρατηρήσεις αφορούν τα ίδια άτομα και όχι ανεξάρτητα δείγματα. Μερικές φο-

ρές, μελετώνται δύο διαφορετικές ομάδες ατόμων, όπου κάθε άτομο «αντιστοιχείται» με ένα της άλλης ομάδας. Εδώ τα δεδομένα είναι σαφώς συνδεδεμένα και θα πρέπει να τα επεξεργαστούμε σαν να ήταν κατά ζεύγη παρατηρήσεις σε μια ομάδα.

Για ανεξάρτητες ομάδες, οι παραμετρικές μέθοδοι απαιτούν οι παρατηρήσεις μέσα σε κάθε ομάδα να εμφανίζουν μια περίπου κανονική κατανομή και οι τυπικές αποκλίσεις σε κάθε ομάδα να είναι παρόμοιες. Εάν τα αδρά δεδομένα δεν ικανοποιούν τις προϋποθέσεις αυτές, ένας μετασχηματισμός ενδέχεται να είναι χρήσιμος. Αλλιώς, θα πρέπει να χρησιμοποιηθούν μη παραμετρικές μέθοδοι. Ο υπολογισμός των μέσων τιμών στην ανάλυση συνεχών δεδομένων παίζει σημαντικό ρόλο. Λόγω του κεντρικού οριακού θεωρήματος (central limit theorem) είναι λογικό να υποθέσουμε ότι η κατανομή των μέσων τιμών για μεγάλα δείγματα είναι η κανονική κατανομή. Εν τούτοις, χρειαζόμαστε να εξετάσουμε την κατανομή των μέσων τιμών για μικρότερα δείγματα.

Η κατανομή t

Στις αρχές του 20^{ου} αιώνα αποδείχθηκε από τον W. S. Gossett, ο οποίος έγραφε με το Ψευδώνυμο «Student», ότι η μέση τιμή ενός δείγματος μιας κανονικής κατανομής με άγνωστη διακύμανση έχει κατανομή παρόμοια αλλά όχι ταυτόσημη με την κανονική κατανομή. Την αποκάλυψε κατανομή t και αναφερόμαστε σ' αυτήν ως κατανομή Student's t. Χρησιμοποιούμε την κατανομή t για την εκτίμηση και τον έλεγχο της υπόθεσης σε σχέση με τις μέσες τιμές ενός ή δύο δειγμάτων. Αν και μπορούμε να χρησιμοποιήσουμε την κανονική κατανομή για μεγάλα δείγματα, οι δύο μέθοδοι δί-

νουν στην πράξη ταυτόσημες απαντήσεις. Η κατανομή t έχει μια παράμετρο που λέγεται βαθμοί ελευθερίας. Γενικά, οι βαθμοί ελευθερίας υπολογίζονται ως το μέγεθος του δείγματος μείον τον αριθμό των υπολογιζόμενων παραμέτρων. Οι βαθμοί ελευθερίας για την κατανομή t σχετίζονται με την υπολογιζόμενη τυπική απόκλιση, η οποία υπολογίζεται ως διακύμανση μεταξύ μιας εκτιμώμενης μέσης τιμής. Συνεπώς, για ένα απλό δείγμα n παρατηρήσεων έχουμε n-1 βαθμούς ελευθερίας.

Μία ομάδα παρατηρήσεων

Η απλούστερη περίπτωση προς εξέτασή είναι όταν θέλουμε να συγκρίνουμε τη μέση τιμή μιας μόνο ομάδας παρατηρήσεων με μια ειδική τιμή. Για παράδειγμα, ας υποθέσουμε ότι θέλουμε να συγκρίνουμε τη μέση διαιτητική πρόσληψη σε μια συγκεκριμένη ομάδα ατόμων προς τη συνιστώμενη ημερήσια πρόσληψη. Ο πίνακας 1 δείχνει τη μέση ημερήσια πρόσληψη ενέργειας 11 υγιών γυναικών ηλικίας 20 - 30 ετών επί 10 ημέρες. Η μέση ημερήσια πρόσληψή τους ήταν 6.753,6 kJ. Το μικρό δείγμα παρατηρήσεων δεν παρουσιάζει προφανή ασυμμετρία. Τι μπορούμε να πούμε σχετικά με την ενεργειακή πρόσληψη αυτών των γυναικών, σε σύγκριση με τη συνιστώμενη ημερήσια πρόσληψη 7.752 kJ;

Κατά μέσον όρο οι 11 γυναίκες είχαν ημερήσια πρόσληψη μικρότερη από το συνιστώμενο επίπεδο των 7.725 kJ, δηλαδή το μέσο έλλειμμα ήταν 7.725 - 6.753,6 = 971,4 kJ. Η τυπική απόκλιση των 11 ημερησίων προσλήψεων ήταν 1.142,1 kJ, έτσι το τυπικό σφάλμα της μέσης πρόσληψης είναι:

$$SE = SD / \sqrt{n} = 1.142,1 / \sqrt{11} = 344,4 \text{ kJ}$$

Πίνακας 1: Μέση ημερήσια πρόσληψη ενέργειας (kJ) 11 υγιών γυναικών επί 10 ημέρες

Υποκείμενο	Μέση ημερήσια πρόσληψη ενέργειας
1	5.260
2	5.470
3	5.640
4	6.180
5	6.390
6	6.515
7	6.805
8	7.515
9	7.515
10	8.230
11	8.770
Μέση τιμή	6.753,6
Τυπική απόκλιση (SD)	1.142,1

Χρησιμοποιούμε την κατανομή t για να υπολογίσουμε το διάστημα εμπιστοσύνης (CI) για τη μέση ημερήσια πρόσληψη. Για 95% CI χρειαζόμαστε την τιμή του t που αντιστοιχεί στην περιοχή της ουράς του 0,05, που καθορίζεται ως $t_{0,975}$, με $11 - 1 = 10$ βαθμούς ελευθερίας. Από τον πίνακα των τιμών t , που είναι διαθέσιμος ως παράρτημα σε κάθε βιβλίο στατιστικής, η τιμή του t που χρειαζόμαστε είναι 2.228. Έτσι, το 95% CI για τη μέση πρόσληψη είναι:

$$\text{Μέση τιμή} \pm t_{0,975} \times \text{Τυπικό σφάλμα} = 6.753,6 \pm 2.228 \times 344,4 = [5.986, 7.521].$$

Η διακύμανση αυτή δεν περιλαμβάνει το συνιστώμενο επίπεδο των 7.725 kJ. Εάν υποθέσουμε ότι οι γυναίκες αυτές είναι αντιπροσωπευτικό δείγμα, τότε μπορούμε να συμπεράνουμε ότι για όλες τις γυναίκες αυτής της ηλικίας η μέση ημερήσια κατανάλωση ενέργειας είναι μικρότερη από αυτή που συνιστάται. Η ερμηνεία όμως αυτή θα ήταν παρακινδυνευμένη, επειδή βασίζεται σε ένα τόσο μικρό δείγμα, ιδιαίτερα επειδή δεν γνωρίζουμε πώς έγινε η επιλογή του

δείγματος.

Δοκιμασία t ενός δείγματος

Μπορούμε επίσης να ελέγξουμε τη μηδενική υπόθεση ότι δηλαδή τα δεδομένα μας αποτελούν ένα δείγμα πληθυσμού με μέση τιμή ίση προς την «υποτιθέμενη» μέση τιμή του γενικού πληθυσμού. Η δοκιμασία λέγεται δοκιμασία t ενός δείγματος και η τιμή του t υπολογίζεται ως:

$$t = (\text{μέση τιμή δείγματος} - \text{υποτιθέμενη μέση τιμή}) / \text{τυπικό σφάλμα της μέσης τιμής του δείγματος}$$

Έτσι στο παράδειγμά μας μπορούμε να υπολογίσουμε την τιμή του t ως:

$$t = (6.753,6 - 7.725) / (1.142,1 / \sqrt{11}) = -2,82.$$

Χρησιμοποιώντας 10 βαθμούς ελευθερίας, βρίσκουμε στον αντίστοιχο πίνακα την τιμή του p που σχετίζεται με την παραπάνω τιμή του t , ως $p < 0,02$. Έτσι, η διαιτητική πρόσληψη των γυναικών αυτών ήταν σημαντικά μικρότερη από το συνιστώμενο επίπεδο, με τη χρησιμοποίηση του συνήθους κριτηρίου στατιστικής σημαντικότητας του $p < 0,05$. Στο σημείο αυτό πρέπει να τονισθεί ότι η στατιστική σημαντικότητα δεν δίνει πληροφορίες σχετικά με το μέγεθος του ενεργειακού ελλείμματος ή/και την αββαιότητα της εκτίμησης αυτής.

Οι μέθοδοι που χρησιμοποιούν την κατανομή t για τον υπολογισμό του διαστήματος εμπιστοσύνης (CI) ή την εφαρμογή μιας δοκιμασίας t χρειάζονται δεδομένα που είναι κατά προσέγγιση κανονικά κατανομημένα. Σε διαφορετική περίπτωση πρέπει να εφαρμόζεται η δοκιμασία της μη παραμετρικής υπόθεσης για ένα μεμονωμένο δείγμα. Υπάρχουν δύο μέθοδοι προς συζήτηση: η δοκιμασία ένδειξης και η δοκιμασία ένδειξης κατά Wilcoxon.

Δοκιμασία ένδειξης

Εάν δεν υπήρχε διαφορά στον μέσο όρο μεταξύ των τιμών του δείγματος και της υποτιθέμενης μέσης τιμής, θα περιμέναμε ίσο αριθμό παρατηρήσεων (τιμών) επάνω και κάτω από τη μέση τιμή. Μπορούμε έτσι να δούμε πόσο πιθανό θα ήταν να έχουμε παρατηρήσει τα δεδομένα μας όταν αληθεύει η μηδενική υπόθεση, υπολογίζοντας την πιθανότητα των συχνοτήτων που παρατηρήθηκαν επάνω και κάτω από τη μέση τιμή. Χρησιμοποιούμε τη διωνυμική κατανομή ή την κανονική προσέγγιση σ' αυτή, για να αξιολογήσουμε την πιθανότητα των συχνοτήτων που παρατηρήθηκαν, όταν η αληθινή πιθανότητα υπέρβασης της αναμενόμενης πρόσληψης είναι $p = 1/2$. Στο παράδειγμά μας, η υποτιθέμενη τιμή πρόσληψης που μας ενδιαφέρει ήταν 7.725 kJ. Δύο γυναίκες είχαν ημερήσιες προσλήψεις επάνω από την τιμή των 7.725 kJ και εννέα κάτω από αυτή.

Ο γενικός τύπος για τη στατιστική δοκιμασία είναι:

Στατιστική δοκιμασία = (παρατηρηθείσα τιμή-υποτιθέμενη τιμή)/τυπικό σφάλμα παρατηρηθείσας τιμής.

Εδώ ενδιαφερόμαστε για τη διωνυμική κατανομή με $p = 1/2$, και $n = 11$. Οι παρατηρήσεις μας είναι είτε $r = 2$ είτε $r = 9$ (δεν έχει σημασία ποιες θα χρησιμοποιήσουμε, λόγω της συμμετρίας της κατανομής όταν $p = 1/2$). Ο αναμενόμενος αριθμός παρατηρήσεων, εάν υποθέσουμε ότι αληθεύει η μηδενική υπόθεση, είναι: $np = 11 \times 1/2 = 5,5$. Το τυπικό σφάλμα του r είναι:

$$\sqrt{[np(1-p)]} = \sqrt{(11 \times 1/2 \times 1/2)} = 1,658.$$

Θα μπορούσαμε να χρησιμοποιήσουμε την ακριβή διωνυμική κατανομή, αλλά η κανονική προσέγγιση στη διωνυμική εί-

ναι εύλογη όταν το $p = 1/2$, ακόμη και για μικρά δείγματα και είναι απλούστερη στη χρήση.

Υπολογίζουμε το στατιστικό κριτήριο z ως:

$$z = (r - np) / \sqrt{np(1-p)} = (2 - 5,5) / 1,658 = -2,11.$$

Ανατρέχοντας σε πίνακα κανονικής κατανομής βλέπουμε ότι η αμφίπλευρη περιοχή των άκρων κατανομής της κανονικής κατανομής που αντιστοιχεί στο $z = 2,11$ είναι $p = 0,035$. Εάν είχαμε χρησιμοποιήσει το $r = 2$, θα είχαμε καταλήξει στο $z = -2,11$, το οποίο μας δίνει την ίδια αμφίπλευρη τιμή p . Έτσι, η διαφορά μεταξύ των δεδομένων που παρατηρήσαμε και της συνιστώμενης τιμής είναι στατιστικά σημαντική στο επίπεδο του 3% περίπου και συμπεραίνουμε ότι η μέση ημερησία πρόσληψη αυτών των γυναικών είναι πραγματικά χαμηλότερη από το συνιστώμενο επίπεδο.

Ένα σχόλιο είναι απαραίτητο σχετικά με τη δοκιμασία ένδειξης. Όταν χρησιμοποιείται μια συνεχής κατανομή ως προσέγγιση σε μη συνεχή δεδομένα, είναι προτιμότερο να ενσωματώνεται μια συνεχής συσχέτιση στη δοκιμασία. υπολογίζοντας ξανά τη στατιστική της δοκιμασίας με τη διόρθωση συνεχείας, έχουμε: $z = [r - np] - 1/2 / \sqrt{[np(1-p)]}$. Αλλιώς, οι υπολογισμοί είναι κάπως αισιόδοξοι, ευνοώντας την απόρριψη της μηδενικής υπόθεσης.

Δοκιμασία ένδειξης αθροίσματος σειράς κατά Wilcoxon

Η μέθοδος έχει τρία στάδια:

1. Υπολογίζουμε τη διαφορά μεταξύ κάθε παρατήρησης και της τιμής που ενδιαφέρει.

2. Αγνοώντας τα πρόσημα των διαφορών, τα κατατάσσουμε με σειρά μεγέθους.

3. Υπολογίζουμε το άθροισμα των σειρών όλων των αρνητικών (ή θετικών) τιμών, που αντιστοιχούν στις παρατηρήσεις κάτω (ή επάνω) από την επιλεγείσα υποθετική μέση τιμή.

Η μέθοδος αυτή δεν κάνει υποθέσεις σχετικά με την ιδιαίτερη μορφή της κατανομής των παρατηρήσεων, προϋποθέτει όμως ότι προέρχονται από πληθυσμό με συμμετρική κατανομή. Για μικρά δείγματα (σύνολο παρατηρήσεων μέχρι 25) οι τιμές του p μπορούν να ληφθούν από κατάλληλους πίνακες που είναι διαθέσιμοι. Για μεγαλύτερα δείγματα η στατιστική δοκιμασία έχει μια κατά

προσέγγιση κανονική κατανομή με μέση τιμή: $n(n+1)/4$ και μεταβλητότητα: $n(n+1)(2n+1)/24$. Όπως συμβαίνει με τη δοκιμασία ένδειξης, οι διαφορές μηδέν παραλείπονται από τους υπολογισμούς, έτσι στον τύπο αυτό το n είναι ο αριθμός των μη μηδενικών διαφορών και έτσι ενδέχεται να είναι μικρότερο από το μέγεθος του δείγματος.

Ο πίνακας 2 δείχνει τις διαφορές των διαιτητικών προσλήψεων από τη συνιστώμενη πρόσληψη και τη σειρά κατάταξής τους, αγνοώντας τα πρόσημά τους.

Πίνακας 2. Ημερήσια πρόσληψη ενέργειας 11 υγιών γυναικών επί 10 ημέρες, διαφορές των διαιτητικών προσλήψεων από τη συνιστώμενη πρόσληψη και σειρά κατάταξής τους.

Υποκείμενο	Ημερήσια πρόσληψη ενέργειας	Διαφορά από την τιμή 7.725	Κατάταξη των διαφορών
1	5.260	2.465	11
2	5.470	2.255	10
3	5.460	2.085	9
4	6.180	1.545	8
5	6.390	1.335	7
6	6.515	1.210	6
7	6.805	920	4
8	7.515	210	1,5
9	7.515	210	1,5
10	8.230	-505	3
11	8.770	-1.045	5

Το άθροισμα των σειρών των δύο παρατηρήσεων που είναι επάνω από τη συνιστώμενη τιμή των 7.725 kJ είναι: $3+5=8$, και ανατρέχοντας στον κατάλληλο πίνακα βλέπουμε ότι το $p < 0,05$, Θα μπορούσαμε εξίσου καλά να είχαμε χρησιμοποιήσει το άθροισμα των σειρών των τιμών που είναι κάτω από τη συνιστώμενη θερμιδική πρόσληψη, το οποίο είναι: $1,5+1,5+4+6+7+8+9+10+11=58$, το οποίο δίνει επίσης $p < 0,05$. Ένα σημαντικό γενικό σημείο είναι ότι δεν αναμένουμε οι διαφορετικές δοκιμασίες να

δώσουν την ίδια απάντηση όταν εφαρμοστούν στα ίδια δεδομένα. Δεν κάνουν τις ίδιες παραδοχές και χρησιμοποιούν διαφορετικές απόψεις των δεδομένων. Γενικά, δύο έγκυρες μέθοδοι θα δώσουν παρόμοιες απαντήσεις, Σε μικρά δείγματα, εντούτοις, οι μη παραμετρικές μέθοδοι υστερούν σε ισχύ και έτσι, στο παραπάνω παράδειγμα, τείνουν να δίνουν μια λιγότερο σημαντική (μεγαλύτερη) τιμή p από την ισοδύναμη παραμετρική δοκιμασία. Στην πράξη συνήθως κάνουμε μόνο μια ανάλυση ενός συνόλου

δεδομένων, επιλέγοντας είτε παραμετρική ή μη παραμετρική εναλλακτική μέθοδο. Συνήθως χρησιμοποιούμε μια παραμετρική μέθοδο, εκτός εάν υπάρχει κάποια σαφής ένδειξη ότι δεν είναι έγκυρη, δηλαδή ότι δεν εκπληρώνονται οι τεθείσες παραδοχές.

Κατά ζεύγη παρατηρήσεις δύο ομάδων

Όταν έχουμε περισσότερες από μία ομάδα παρατηρήσεων, είναι ζωτικής σημασίας η διάκριση της περίπτωσης όπου τα δεδομένα είναι κατά ζεύγη από εκείνη όπου οι ομάδες είναι ανεξάρτητες.

Με δεδομένα κατά ζεύγη ενδιαφερόμαστε για τη μέση διαφορά μεταξύ των παρατηρήσεων σε κάθε άτομο και για τη διακύμανση των διαφορών αυτών. Δηλαδή, μας ενδιαφέρει η διακύμανση των διαφορών που παρατηρούνται σε κάθε άτομο μάλλον, παρά οι μεταξύ ατόμων διαφορές, Εξετάζοντας διαφορές μειώνουμε ουσιαστικά την ανάλυση σε πρόβλημα ενός δείγματος, ούτως ώστε μπορούμε να χρησιμοποιήσουμε για την εκτίμηση και τον έλεγχο της υπόθεσης μεθόδους παρόμοιες με αυτές που αναλύθηκαν στο προηγούμενο κεφάλαιο. Αφού εξετάζουμε τις μεταξύ καθενός ατόμου διαφορές ως ένα δείγμα, τότε πρέπει οι διαφορές αυτές να έχουν μια κατά προσέγγιση κανονική κατανομή. Δεν υπάρχει ανάγκη κάθε σύνολο δεδομένων να παρουσιάζει κανονική κατανομή.

Δύο ανεξάρτητες ομάδες παρατηρήσεων

Οι συνηθέστερες στατιστικές αναλύσεις είναι εκείνες που χρησιμοποιούνται για τη σύγκριση δύο ανεξάρτητων ομάδων παρατηρήσεων καθώς οι περισσότερες κλινικές δοκιμές δημιουργούν δε-

δομένα αυτού του τύπου. Ενδιαφερόμαστε πάλι για τη μέση διαφορά μεταξύ των ομάδων, αλλά η μεταβλητότητα μεταξύ των ατόμων γίνεται σημαντική. Τόσο το διάστημα εμπιστοσύνης όσο και ο έλεγχος t των δύο δειγμάτων βασίζονται στην παραδοχή ότι κάθε σύνολο παρατηρήσεων λαμβάνεται δειγματοληπτικά από πληθυσμό με κανονική κατανομή και ότι οι διασπορές των δύο πληθυσμών είναι ίσες.

Διάστημα εμπιστοσύνης για διαφορές μεταξύ μέσων τιμών

Το απαιτούμενο τυπικό σφάλμα λαμβάνεται από έναν πιο περίπλοκο τύπο, σε σύγκριση με εκείνον της απλής περίπτωσης, αλλά χρησιμοποιεί μόνο τη μέση διακύμανση και το μέγεθος του δείγματος κάθε ομάδας. Πρώτα υπολογίζουμε τη μεικτή διακύμανση s^2 , από τον τύπο:

$$s^2 = [(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2] / (n_1 + n_2 - 2)$$

όπου s_1 και s_2 είναι οι τυπικές αποκλίσεις των δύο ομάδων, οι οποίες έχουν πλήθος n_1 και n_2 αντιστοίχως. Χρησιμοποιώντας τα σύμβολα x_1 και x_2 για τις μέσες τιμές των δύο δειγμάτων και το s για τη μεικτή τυπική απόκλιση έχουμε:

$$se(x_1 - x_2) = \sqrt{(1/n_1 + 1/n_2)}$$

Έτσι, το 95% διάστημα εμπιστοσύνης της διαφοράς μεταξύ των μέσων τιμών δίδεται από τον τύπο:

$$x_1 - x_2 \pm t_{0,975} \cdot se(x_1 - x_2)$$

όπου η τιμή του t έχει $n_1 + n_2 - 2$ βαθμούς ελευθερίας.

Ο πίνακας 3 παρουσιάζει την 24ωρη κατανάλωση ενέργειας σε ομάδες ισχίων και παχύσαρκων γυναικών.

Η μεικτή τυπική απόκλιση είναι:

$$s = \sqrt{\{[(12 \times 1,238^2) + (8 \times 1,398^2)] / 20\}} = 1,3044 \text{ MJ ημερησίως.}$$

Το τυπικό σφάλμα της διαφοράς των μέσων τιμών κατανάλωσης ενέργειας

δίδεται από τον τύπο:

$$se = 1,3044 \times \sqrt{(1/13 + 1/9)} = 0,5656 \text{ MJ ημερησίως.}$$

Η διαφορά των μέσων τιμών κατανάλωσης των δύο ομάδων είναι 2,232 MJ ημερησίως και από τον κατάλληλο πίνακα βλέπουμε ότι η τιμή του $t_{0,975}$ με 20 βαθμούς ελευθερίας είναι 2,086. Έτσι, το 95% διάστημα εμπιστοσύνης για την μέση διαφορά 24ωρης κατανάλωσης ενέργειας μεταξύ παχύσαρκων και ισχνών γυναικών είναι:

$$[2,232 \pm 2,086 \times 0,5656] = [1,05, 3,41] \text{ MJ ημερησίως.}$$

Πίνακας 3. Εικοσιτετράωρη ολική κατανάλωση ενέργειας (MJ/ημέρα) σε ομάδες ισχνών και παχύσαρκων γυναικών.

Υποκείμενο	Ισχνές (n = 13)	Παχύσαρκες (n = 9)
1	6,13	8,79
2	7,05	9,19
3	7,48	9,21
4	7,48	9,68
5	7,53	9,69
6	7,58	9,97
7	7,90	11,51
8	8,08	11,85
9	8,09	12,79
10	8,11	
11	8,40	
12	10,15	
13	10,88	
Μέση τιμή	8,066	10,928
Τυπική απόκλιση	1,238	1,398

Δοκιμασία t για δύο δείγματα

Υπάρχει επίσης η δοκιμασία t για την σύγκριση δύο ανεξάρτητων ομάδων δεδομένων. Η τιμή του t λαμβάνεται από τον τύπο:

$$t = x_1 - x_2 / se (x_1 - x_2)$$

με την κατανομή του t με n_1+n_2-2 βαθμούς ελευθερίας. Στο παράδειγμά μας έχουμε:

$t = 2,232/0,5656$, που για 20 βαθμούς ελευθερίας μας δίνει $p < 0,001$. Μπορούμε κατά συνέπεια να πούμε ότι η ολική ενεργειακή κατανάλωση των παχύσαρκων γυναικών είναι στατιστικά σημαντικά μεγαλύτερη από την αντίστοιχη των ισχνών γυναικών.

Μη παραμετρική σύγκριση δύο ομάδων – Δοκιμασία Mann-Whitney

Η δοκιμασία Mann-Whitney είναι μη παραμετρική εναλλακτική προς τη δοκιμασία t μέθοδος σύγκρισης δεδομένων δύο ανεξάρτητων ομάδων. Απαιτεί την κατάταξη όλων των παρατηρήσεων σαν να αφορούσαν μία ενιαία ομάδα. Στη συνέχεια υπολογίζεται το άθροισμα των τιμών της ενοποιημένης ομάδας και αναζητείται η τιμή στον κατάλληλο πίνακα. Ο πίνακας 4 δείχνει τα δεδομένα κατανάλωσης ενέργειας επεξεργασμένα με την μέθοδο Mann-Whitney. Τα αθροίσματα των τιμών των δύο σειρών είναι 103 και 150 αντίστοιχα. Η στατιστική παράμετρος U υπολογίζεται με τον τύπο:

$$U = n_1n_2 + \frac{1}{2}n_1(n_1+1) - T$$

Η παράμετρος T είναι απλά το άθροισμα των τιμών της μικρότερης ομάδας που στο συγκεκριμένο παράδειγμα είναι 103. Έτσι: $U = 117+91-103=58$, που μας δίνει τιμή $p < 0,01$ και οδηγεί στην απόρριψη της μηδενικής υπόθεσης της ίσης ολικής κατανάλωσης ενέργειας από τις παχύσαρκες και τις ισχνές γυναίκες. Για δείγματα μεγαλύτερα των 10 τιμών σε κάθε ομάδα η παράμετρος U έχει σχεδόν κανονική κατανομή.

Άνισες διασπορές

Μερικές φορές θέλουμε να συγκρίνουμε δύο ομάδες παρατηρήσεων, όπου η παραδοχή της κανονικής κατανομής είναι εύλογη, αλλά η μεταβλητότητα

μεταξύ των δύο ομάδων είναι έκδηλα διαφορετική. Στην περίπτωση αυτή μπορούμε να ελέγξουμε τη μηδενική υπόθεση χρησιμοποιώντας την δοκιμασία F. Η δοκιμασία F είναι πολύ απλή. Με τη μηδενική υπόθεση ότι δύο φυσιο-

λογικά κατανεμημένοι πληθυσμοί έχουν ίσες μεταβλητότητες, αναμένουμε ότι το πηλίκον της μεταβλητότητας των δύο δειγμάτων θα έχει κατανομή δείγματος γνωστή ως F-κατανομή.

Πίνακας 4. Υπολογισμοί για τη δοκιμασία U Mann-Whitney των δεδομένων κατανάλωσης ενέργειας (ΕΕ).

Ισχνές (n=13) κατάταξη	ΕΕ	Παχύσαρκες (n=9) κατάταξη	ΕΕ
1	6,13		
2	7,05		
3,5	7,48		
3,5	7,48		
5	7,53		
6	7,58		
7	7,90		
8	8,08		
9	8,09		
10	8,11		
11	8,40		
		12	8,79
		13	9,19
		14	9,21
		15	9,68
		16	9,69
		17	9,97
18	10,15		
19	10,88		
		20	11,51
		21	11,85
		22	12,79
Άθροισμα	103	Άθροισμα	150

Το πηλίκον των μεταβλητοτήτων είναι το πηλίκον των μεταβλητοτήτων των δειγμάτων ή το τετράγωνο των τυπικών αποκλίσεων των δειγμάτων. Υπολογίζουμε το πηλίκον της μεταβλητότητας των δύο δειγμάτων, διαιρώντας τη μεγαλύτερη τιμή μεταβλητότητας με τη μικρότερη και αναζητούμε το τετράγωνο της τιμής του πηλίκου στον αντίστοιχο πίνακα. Η κατανομή της παραμέτρου F έχει δύο τιμές βαθμών ελευθερίας, ανά μία για κάθε μεταβλητότητα.

Αν και είναι απίθανο τα δύο δείγματα

να προέρχονται από πληθυσμούς με την ίδια μεταβλητότητα, θα μπορούσαμε να χρησιμοποιήσουμε τη δοκιμασία U Mann-Whitney ή μια τροποποίηση της δοκιμασίας t της περίπτωσης άνισων μεταβλητοτήτων, γνωστή ως δοκιμασία Welch.

Τρεις ή περισσότερες ομάδες παρατηρήσεων

Όταν έχουμε περισσότερες ομάδες παρατηρήσεων είναι προφανώς δυνατό να συγκρίνουμε κάθε ζεύγος ομάδων,

χρησιμοποιώντας κάθε φορά τη δοκιμασία t . Αυτή όμως δεν είναι η ιδανική προσέγγιση. Είναι πολύ καλύτερο να χρησιμοποιήσουμε μία ανάλυση που μας επιτρέπει να δούμε όλα τα δεδομένα μαζί, η μέθοδος που χρησιμοποιούμε στην περίπτωση αυτή λέγεται ανάλυση διακύμανσης (σε συντομογραφία ANOVA). Όπως δηλώνει και το όνομά της, η μονόδρομη ανάλυση διακύμανσης είναι ο απλούστερος τύπος που χρησιμοποιείται, όταν υπάρχει ένας τρόπος ταξινόμησης των παρατηρήσεων. Όταν υπάρχουν δύο παράγοντες που ταξινομούν τις παρατηρήσεις χρειαζόμαστε ανάλυση διακύμανσης δύο δρόμων και ούτω καθεξής.

Η αρχή της ανάλυσης διακύμανσης είναι το μοίρασμα της ολικής διακύμανσης ενός συνόλου δεδομένων σε τμήματα, λόγω διαφορετικών πηγών της διακύμανσης. Για παράδειγμα, η διακύμανση των δεδομένων κατανάλωσης ενέργειας στον πίνακα 2 θα μπορούσε να μοιραστεί σε εκείνη που οφείλεται σε διακυμάνσεις μεταξύ των ατόμων μέσα σε κάθε ομάδα και σε εκείνη που οφείλεται σε συστηματική διαφορά μεταξύ των ομάδων. Πράγματι, επειδή η μηδενική μας υπόθεση είναι ότι δεν υπάρχει διαφορά μεταξύ των ομάδων, η δοκιμασία βασίζεται στη σύγκριση της διακύμανσης που παρατηρήθηκε μεταξύ των ομάδων (δηλ. μεταξύ των μέσων τιμών των ομάδων) με την αναμενόμενη από την παρατηρηθείσα μεταβλητότητα μεταξύ των ατόμων. Η σύγκριση παίρνει τη γενική μορφή μιας δοκιμασίας F για τη σύγκριση τιμών μεταβλητότητας, αλλά για δύο ομάδες η δοκιμασία t οδηγεί ακριβώς στην ίδια απάντηση. Τα δείγματα δεν χρειάζεται να έχουν το ίδιο μέγεθος.

Η ανάλυση διακύμανσης θα έπρεπε κατά προτίμηση να γίνει χρησιμοποιώντας ένα πρόγραμμα Η/Υ. Ένα πρόγραμμα στατιστικής θα έδινε τα αριθμητικά αποτελέσματα, αλλά είναι σημαντικό να κατανοήσουμε τις αρχές που εμπλέκονται στη διαδικασία.

1. Η ανάλυση βασίζεται στην παραδοχή ότι τα δείγματα προέρχονται από πληθυσμούς με κανονική κατανομή με την ίδια τυπική απόκλιση (ή μεταβλητότητα). Η κανονική κατανομή και η ισότητα των διασπορών δεν πρέπει να υποτίθενται, αλλά θα πρέπει να πιστοποιούνται.

2. Επειδή δεχόμαστε ότι τα δείγματα προέρχονται από πληθυσμούς με την ίδια διακύμανση, η διακύμανση της κάθε ομάδας είναι κατ'επίσημη ίση με τη διακύμανση του πληθυσμού. Έτσι, ενώ νουμε τις διασπορές των δειγμάτων (με τον ίδιο τρόπο όπως κάναμε με τη δοκιμασία t δύο δειγμάτων) για να πάρουμε μια εκτίμηση της μεταβλητότητας του πληθυσμού.

3. Μπορούμε να χρησιμοποιήσουμε την ενωμένη μεταβλητότητα για να υπολογίσουμε το διάστημα εμπιστοσύνης της διαφοράς οποιουδήποτε από τα ζεύγη των μέσων τιμών.

4. Μπορούμε να πραγματοποιήσουμε μια δοκιμασία υπόθεσης βασισμένη στη μηδενική υπόθεση ότι τα δείγματα είναι από πληθυσμούς με τη ίδια διακύμανση. Μπορούμε έτσι να συγκρίνουμε τη διακύμανση μεταξύ των μέσων τιμών των παρατηρηθέντων δειγμάτων με αυτή που θα αναμέναμε από τυχαία δείγματα εάν η μηδενική υπόθεση ήταν αληθινή. Με άλλα λόγια, μπορούμε να υπολογίσουμε την πιθανότητα να παρατηρήσουμε τέτοια διακύμανση μεταξύ δειγμάτων τα οποία ελήφθησαν στην τύχη

από τον ίδιο πληθυσμό. Η σύγκριση παίρνει τη μορφή του λόγου των διασπορών που υπολογίζεται από τις μέσες τιμές των ομάδων (τη μεταξύ των ομάδων διακύμανση) και της διακύμανσης μεταξύ των ατόμων μέσα στις ομάδες. Όπως είδαμε νωρίτερα, χρησιμοποιούμε τους πίνακες της κατανομής F για να ελέγξουμε την ισότητα των δύο διασπορών

5. Αφού φέρουμε σε πέρας την ανάλυση διακύμανσης θα μπορούσαμε να εξετάσουμε τη διακύμανση (μεταβλητότητα) των επί μέρους παρατηρήσεων γύρω από τη μέση τιμή του δείγματός τους. Για κάθε άτομο η μέση τιμή της ομάδας του είναι η τιμή που προσαρμόζεται στο μοντέλο, και η διαφορά μεταξύ τιμών που παρατηρήθηκαν και εκείνων που αναλογούν καλείται υπολειμματική. Τη διακύμανση αυτών των υπολειμματικών χρησιμοποιούμε ως μέτρο εκτίμησης της μεταβλητότητας μεταξύ ατόμων, και με αυτήν αξιολογούμε τη διαφορά μεταξύ των ομάδων. Μπορούμε να σχεδιάσουμε μια γραφική παράσταση των υπολειμματικών για να εκτιμήσουμε εάν ευσταθεί η παραδοχή της κανονικής κατανομής. Εάν η γραφική παράσταση δεν είναι ικανοποιητική, πρέπει να αναλύσουμε εκ νέου τα δεδομένα πιθανώς αφού τα μετασχηματίσουμε ή με τη βοήθεια μιας μη παραμετρικής εναλλακτικής μεθόδου.

Η ανάλυση διακύμανσης είναι μια γενικότερη μορφή της δοκιμασίας t , έτσι υπάρχει η μορφή της μη παραμετρικής δοκιμασίας Mann-Whitney. Η δοκιμασία Kruskal-Wallis είναι μια προφανής μαθηματική επέκταση της δοκιμασίας Mann-Whitney, με τα ίδια προβλήματα ερμηνείας, όπως αυτά εξετάστηκαν στη μονόδρομη ανάλυση της μεταβλητότητας.

Όταν οι ομάδες είναι ταξινομημένες δεν είναι λογικό να συγκρίνουμε κάθε ζεύγος ομάδων, αλλά μάλλον θα πρέπει να μελετήσουμε την πιθανότητα να υπάρχει τάση μεταξύ των ομάδων. Για πολλούς σκοπούς θα είναι αρκετή να εξετασθεί εάν υπάρχει μια γραμμική τάση. Η προσαρμογή της γραμμικής τάσης στη μονόδρομη ανάλυση της μεταβλητότητας είναι το ισοδύναμο της γραμμικής ανάλυσης παλινδρόμησης.

Ένας μεγάλος αριθμός στατιστικών αναλύσεων γίνεται για να δοκιμαστεί η σχέση μεταξύ δύο μεταβλητών μέσα σε μια ομάδα ατόμων. Δύο κύριοι σκοποί των αναλύσεων αυτών είναι οι ακόλουθοι:

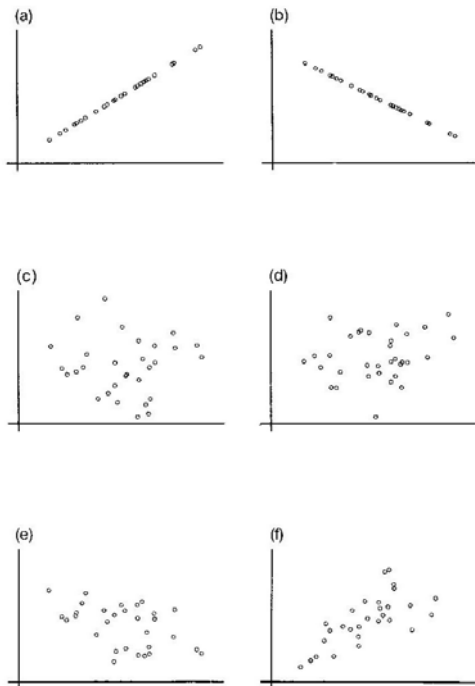
1. Η αξιολόγηση της ύπαρξης συσχέτισης των δύο μεταβλητών, δηλαδή εάν οι τιμές της μιας μεταβλητής τείνουν να είναι μεγαλύτερες (ή αντιθέτως μικρότερες) όταν αυξάνουν οι τιμές της άλλης μεταβλητής
2. Η πρόβλεψη της τιμής μιας μεταβλητής από την ήδη γνωστή τιμή της άλλης μεταβλητής.

Για την εκτίμηση της σχέσης μεταξύ συνεχών μεταβλητών, χρησιμοποιείται η μέθοδος που είναι γνωστή ως συσχέτιση. Για την πρόβλεψη της τιμής μιας συνεχούς μεταβλητής από μια άλλη, χρησιμοποιείται η τεχνική της γραμμικής παλινδρόμησης/εξάρτησης.

Συσχέτιση

Συσχέτιση είναι η μέθοδος ανάλυσης που χρησιμοποιούμε όταν μελετούμε την πιθανή σχέση μεταξύ δύο συνεχών μεταβλητών. Εάν θέλουμε να μετρήσουμε τον βαθμό συσχέτισης, αυτό μπορεί να γίνει με τον υπολογισμό του συντελεστή συσχέτισης, ο οποίος συχνά αποκαλείται απλώς συσχέτιση. Η τυπική μέθοδος οδηγεί σε μια ποσότητα που

λέγεται r , η οποία μπορεί να προσλάβει τις τιμές από -1 έως $+1$. Αυτός ο συντελεστής συσχέτισης r μετρά την "ευθύγραμμη" σχέση μεταξύ των τιμών των δύο μεταβλητών. Έτσι, οι τιμές $+1$ ή -1 σημαίνουν ότι όλα τα σημεία του διαγράμματος διακύμανσης βρίσκονται σε μια τέλεια ευθεία γραμμή, όπως φαίνεται στην παρακάτω εικόνα (εικόνα 1).



Εικόνα 1. Δεδομένα με συντελεστές συσχέτισης (a) $r=1,0$ (b) $r=-1,0$ (c) $r=0,0$ (d) $r=0,3$ (e) $r=0,5$ (f) $r=0,7$.

Στην εικόνα αυτή διακρίνονται επίσης παραδείγματα δεδομένων με ενδιάμεσες τιμές του συντελεστή r . Η συσχέτιση μεταξύ των δύο μεταβλητών είναι θετική εφόσον οι μεγαλύτερες τιμές της μιας συνοδεύονται από μεγαλύτερες τιμές της άλλης και αρνητική εφόσον η μια μεταβλητή τείνει προς μικρότερες τιμές όταν η άλλη αυξάνεται. Συντελεστής r με τιμή κοντά στο μηδέν σημαίνει ότι δεν υπάρχει γραμμική συσχέτιση μετα-

ξύ των τιμών των δύο μεταβλητών.

Παλινδρόμηση/Εξάρτηση

Αλλα προβλήματα ενδέχεται να προκύψουν όταν έχουμε ένα σύνολο δεδομένων δύο συνεχών μεταβλητών. Ιδιαίτερος, ενδέχεται να θελήσουμε να περιγράψουμε τη σχέση μεταξύ τους και έτσι να είμαστε σε θέση να προλέγουμε την τιμή της μιας μεταβλητής όταν γνωρίζουμε μόνο την τιμή της άλλης. Είναι σαφές ότι ο συντελεστής συσχέτισης δεν καλύπτει αυτές τις λειτουργίες· απλώς δείχνει την ισχύ της ως ένας απλός αριθμός. Θέλουμε να βρούμε τρόπο να περιγράψουμε τη σχέση μεταξύ των τιμών των δύο μεταβλητών, γι' αυτό το γενικό πρόβλημα χρειαζόμαστε την τεχνική που λέγεται παλινδρόμηση/εξάρτηση. Θα εξετάσουμε μόνο τη συνήθη περίπτωση όπου ενδιαφερόμαστε για μια γραμμική συσχέτιση μεταξύ των δύο μεταβλητών.

Ο πίνακας 5 και η εικόνα 2 δείχνουν δεδομένα τα οποία έχουν συλλεγεί από 24 ασθενείς με διαβήτη τύπου 1. Οι μεταβλητές είναι τιμές γλυκόζης αίματος νηστείας (mmol/l) και μέση ταχύτητα περιμετρικής βραχύνσεως ινών της αριστερής κοιλίας (Vcf) που μετράται με το υπερηχοκαρδιογράφημα. Η Vcf ενός ασθενούς δεν καταγράφηκε. Εάν μας ενδιαφέρει να προσπαθήσουμε να υπολογίσουμε (κατά προσέγγιση) την τιμή της Vcf από την τιμή της γλυκόζης του αίματος, τότε, διαφορετικά από την περίπτωση της συσχέτισης, δεν έχουμε συμμετρική σχέση μεταξύ των δύο μεταβλητών.

Μπορούμε να θεωρήσουμε ως μεταβλητή απαντήσεως (ή έκβασης) την Vc και ως προγνωστική μεταβλητή το σάκχαρο του αίματος. Αυτές συχνά αποκαλούνται και εξαρτημένη και ανεξάρτητη

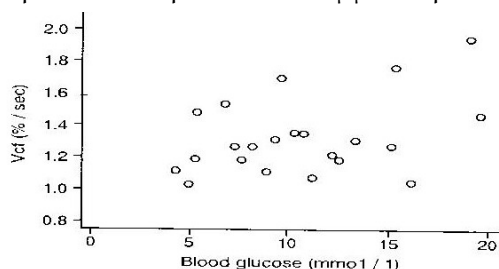
μεταβλητή, αντιστοίχως, ονόματα που προκαλούν σύγχυση και δείχνουν ποια μεταβλητή εξαρτάται από την άλλη. Η μεταβλητή απαντήσεως καταχωρείται στον κάθετο ή Υ άξονα των συντεταγμένων και η προγνωστική μεταβλητή στον οριζόντιο ή Χ άξονα, όπως απεικονίζονται στην εικόνα 2.

Πίνακας 5. Δεδομένα 24 ασθενών με σακχαρώδη διαβήτη τύπου 1.

Ασθενής	Σάκχαρο αίματος νηστείας (mmol/l)	Μέση ταχύτητα βραχύνσεως κυκλοτερούς μυϊκής ίνας (Vcf) (%/sec)
1	15,3	1,76
2	10,8	1,34
3	8,1	1,27
4	9,5	1,47
5	7,2	1,27
6	5,3	1,49
7	9,3	1,31
8	11,1	1,09
9	7,5	1,18
10	12,2	1,22
11	6,7	1,25
12	5,2	1,19
13	19,0	1,95
14	15,1	1,28
15	6,7	1,52
16	8,6	
17	4,2	1,12
18	10,3	1,37
19	12,5	1,19
20	16,1	1,05
21	13,3	1,32
22	4,9	1,03
23	8,8	1,12
24	9,5	1,70

Το πρόβλημα είναι να βρεθεί μια ευθεία γραμμή προσαρμοσμένη στα δεδομένα. η οποία Καθό κάποιο τρόπο να δίνει την "καλύτερη" πρόγνωση της τιμής του Υ για κάθε τιμή του Χ. Αυτή θα είναι μια γραμμή που θα ελαχιστοποιεί την απόσταση μεταξύ των δεδομένων και της προσαρμοσμένης γραμμής. Υπάρχουν διάφορες προσεγγίσεις του προ-

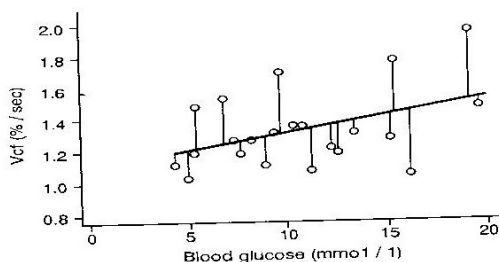
βλήματος. αλλά η τυπική συνήθης μέθοδος λέγεται παλινδρόμηση/εξάρτηση των ελαχίστων τετραγώνων. Όταν χρησιμοποιούμε αυτή τη μέθοδο για να προσαρμόσουμε μια γραμμή παλινδρόμησης ελαχιστοποιούμε το άθροισμα των τετραγώνων των καθέτων αποστάσεων από τη γραμμή. Η εικόνα 3 δείχνει τα ίδια δεδομένα με τη γραμμή παλινδρόμησης των ελαχίστων τετραγώνων, μαζί με τις αποστάσεις από τη γραμμή. Κάθε απόσταση είναι η διαφορά για κάθε ένα από τα άτομα μεταξύ της τιμής που παρατηρήθηκε και της τιμής που δίνει η γραμμή, που είναι γνωστή ως προσαρμοσμένη τιμή. Ο τεχνικός όρος για την απόσταση είναι υπολειμματική.



Εικόνα 2. Σχέση μεταξύ των τιμών γλυκόζης αίματος νηστείας και μέσης ταχύτητας περιμετρικής βραχύνσεως ίνας της αριστερής κοιλίας (Vcf). Δεδομένα από 23 ασθενείς με διαβήτη τύπου I.

Η μέθοδος των ελαχίστων τετραγώνων παράγει τη γραμμή η οποία ελαχιστοποιεί το άθροισμα των τετραγώνων των υπολειμματικών, και έτσι ελαχιστοποιεί επίσης τη διακύμανση των υπολειμματικών, η οποία είναι ακριβώς το άθροισμα των τετραγώνων, διαιρεμένο με τον αριθμό των παρατηρήσεων μείον δύο. Η διακύμανση αυτή, γνωστή ως υπολειμματική μεταβλητότητα, είναι το μέτρο της καλής προσαρμογής της γραμμής. Η υπολειμματική μεταβλητότητα είναι πολύ σημαντική για την αξιολόγηση των αποτελεσμάτων της πα-

Λινδρομικής ανάλυσης.



Εικόνα 3. Τα δεδομένα της εικόνας 2 με τη γραμμή παλινδρόμησης που δείχνει τις διαφορές μεταξύ των τιμών που παρατηρήθηκαν και των προσαρμοσμένων τιμών.

Εάν έχουμε παρατηρήσεις τιμών των δύο μεταβλητών, X (γλυκόζη αίματος) και Y (Vcf), μπορούμε να πραγματοποιήσουμε μια παλινδρόμηση/εξάρτηση του X στο Y για να πάρουμε μια ευθεία γραμμή, η οποία δίνει μια προσαρμοσμένη κατ' εκτίμηση τιμή του Y για κάθε τιμή της μεταβλητής X . Η γενική εξίσωση μιας γραμμής παλινδρόμησης είναι:

$$Y = \alpha + bX.$$

Εδώ το b είναι η κλίση της γραμμής και το α λέγεται τομή επειδή είναι η προσαρμοσμένη τιμή του Y , εκεί όπου η γραμμή τέμνει τον άξονα των Y και αντιστοιχεί στην τιμή $X=0$. Στις περισσότερες ιατρικές εφαρμογές η τιμή του α δεν έχει πρακτική σημασία, καθώς η μεταβλητή X δεν μπορεί ποθενά να είναι κοντά στο μηδέν. παραδείγματα είναι η αρτηριακή πίεση και οποιεσδήποτε μετρήσεις του σώματος.

Για τα δεδομένα των διαβητικών ασθενών η γραμμή παλινδρόμησης που καταδεικνύεται στην εικόνα 3. είναι:

$$Vcf = 1,10 + 0,0220 \times \text{γλυκόζη αίματος}.$$

Η κλίση είναι η παράμετρος που παρουσιάζει το κύριο ενδιαφέρον στην παλινδρομική ανάλυση, επειδή δείχνει την ισχύ της συσχέτισης μεταξύ των δύο μεταβλητών. Η κλίση της προσαρμοσμένης γραμμής είναι 0,022, πράγμα το ο-

ποίο σημαίνει ότι εκτιμούμε αύξηση της Vcf κατά 0,022% ανά δευτερόλεπτο για κάθε αύξηση της γλυκόζης του αίματος νηστείας κατά μία μονάδα (δηλ. 1 mmol/l). Μπορούμε να υπολογίσουμε το τυπικό σφάλμα της κλίσης, το οποίο είναι 0.0105. Η κατ' εκτίμηση κλίση b , αντιμετωπίζεται κατά το μάλλον ή ήττον με τον ίδιο τρόπο όπως η μέση τιμή ενός δείγματος. Μπορούμε να υπολογίσουμε ένα διάστημα εμπιστοσύνης για την κλίση, και μπορούμε να ελέγξουμε την υπόθεση της μη ύπαρξης καμπύλης, δηλαδή της απουσίας συσχέτισης μεταξύ της Vcf και της γλυκόζης του αίματος.

Στο σημείο αυτό είναι σημαντικό να εξετάσουμε τρεις παραδοχές που βρίσκονται στο υπόβαθρο της μεθόδου:

1. Οι τιμές της μεταβλητής έκβασης Y (π.χ. της Vcf) πρέπει να έχουν φυσιολογική κατανομή για κάθε τιμή της προγνωστικής μεταβλητής X .
2. Η μεταβλητότητα της Y , όπως αυτή αξιολογείται από τον συντελεστή μεταβλητότητας και την τυπική απόκλιση, πρέπει να είναι η ίδια για κάθε τιμή της μεταβλητής X .
3. Η συσχέτιση μεταξύ των δύο μεταβλητών πρέπει να είναι γραμμική.

Ανάλυση επιβίωσης

Δεδομένα άλλου τύπου προκύπτουν, όταν το ενδιαφέρον εστιάζεται στον χρόνο που χρειάστηκε για να συμβεί κάποιο γεγονός. Μια από τις συνηθέστερες πηγές τέτοιων δεδομένων είναι όταν καταγράφουμε τον χρόνο από κάποιο καθορισμένο χρονικό σημείο εκκίνησης, όπως π.χ. από μια εγχείρηση, μέχρι τον θάνατο του ασθενούς. Για τον λόγο αυτό, συνήθως αναφερόμαστε σε χρόνους επιβίωσης ή δεδομένα επιβίωσης και η στατιστική επεξεργασία των χρόνων επιβίωσης είναι γνωστή ως α-

νάλυση επιβίωσης. Στις κλινικές δοκιμές οι χρόνοι επιβίωσης συχνά αναφέρονται στον χρόνο θανάτου, εμφάνιση ενός ιδιαίτερου συμπτώματος ή υποτροπής ύστερα από ύφεση της νόσου. Αν και υπάρχει συνήθως σαφής καθορισμός του τέλους του χρόνου της περιόδου που ενδιαφέρει, η έναρξη μπορεί να είναι λιγότερο σαφώς καθορισμένη. Είναι, π.χ. σπάνιο να γνωρίζουμε από πόσον καιρό ο ασθενής έχει εκδηλώσει τη νόσο, και τότε ο χρόνος της διάγνωσης είναι συχνά η καλύτερη εναλλακτική λύση.

Υπάρχει ένα εγγενές χαρακτηριστικό των χρόνων επιβίωσης, το οποίο τους καθιστά ακατάλληλους για ανάλυση με οποιαδήποτε από τις μεθόδους που περιγράφηκαν στα προηγούμενα κεφάλαια, το ότι σχεδόν ποτέ δεν παρατηρούμε το γεγονός που μας ενδιαφέρει σε όλα τα άτομα. Για παράδειγμα, σε μια δοκιμή σύγκρισης της επιβίωσης ασθενών με διαφορετικές μεθόδους χειρουργικής θεραπείας καρκίνου του μαστού, αν και οι ασθενείς θα υποβληθούν σε παρακολούθηση για κάποια χρόνια θα υπάρξουν πολλές που θα είναι ακόμη στη ζωή κατά το τέλος της δοκιμής. Για τις ασθενείς αυτές δεν γνωρίζουμε πότε θα πεθάνουν, παρά μόνο ότι ήταν στη ζωή κατά το τέλος της δοκιμής. Συνεπώς, δεν γνωρίζουμε τον χρόνο επίβιωσής τους μετά την εγχείρηση, αλλά μόνο ότι ο χρόνος αυτός θα είναι μακρότερος από τον χρόνο που διήρκεσε η δοκιμή. Ονομάζουμε αυτούς τους χρόνους επιβίωσης "χρόνους περικοπής", για να δείξουμε ότι η περίοδος παρατήρησης σταμάτησε προτού συμβεί το γεγονός που μας ενδιέφερε. Ας σημειωθεί ότι, επειδή το γεγονός που ενδιαφέρει είναι κάτι δυσάρεστο, όπως ο θάνατος, το ενδιαφέρον είναι επιστημονικό και όχι κλινικό.

Από ένα σύνολο παρατηρηθέντων χρόνων επιβίωσης ενός δείγματος ατόμων μπορούμε να εκτιμήσουμε την αναλογία του πληθυσμού προς επιβίωση για ένα ορισμένο χρονικό διάστημα υπό τις αυτές συνθήκες. Για παράδειγμα, μπορούμε να χρησιμοποιήσουμε δεδομένα από μια δοκιμή ασθενών που υποβλήθηκαν σε μεταμόσχευση ήπατος για να εκτιμήσουμε την πιθανότητα επιβίωσης επί ορισμένο χρονικό διάστημα μετά τη μεταμόσχευση (με τον συνήθη όρο *proviso* της αντιπροσωπευτικότητας του αρχικού δείγματος). Η μέθοδος είναι ευφύης, επειδή όχι μόνο αφήνει περιθώρια να περιληφθούν οι παρατηρήσεις που χρόνους περικοπής, αλλά διότι χρησιμοποιεί επίσης και τις σχετικές με τα άτομα αυτά πληροφορίες, μέχρι τον χρόνο που γίνεται περικοπή. Η μέθοδος επιτρέπει τη σχεδίαση γραφικής παράστασης ή τη σύνταξη πίνακα με διάφορα ονόματα όπως: πίνακες ζωής, καμπύλη επιβίωσης, καμπύλη Kaplan-Meier. Η ανάλυση των ιατρικών δεδομένων επιβίωσης έχει γνωρίσει μεγάλη διάδοση από τις αρχές της δεκαετίας 1970, όταν αναπτύχθηκαν νέες μέθοδοι.

Κατηγορικά δεδομένα

Τα κατηγορικά δεδομένα είναι πολύ συνηθισμένα στην ιατρική έρευνα και προκύπτουν όταν άτομα εντάσσονται σε μία ή δύο ή περισσότερες αμοιβαία αποκλειστικές ομάδες. Σε ένα δείγμα ατόμων ο αριθμός που εμπίπτει σε μια ιδιαίτερη ομάδα λέγεται συχνότητα, έτσι η ανάλυση των κατηγορικών δεδομένων λέγεται ανάλυση συχνοτήτων. Όταν γίνεται σύγκριση μεταξύ δύο ή περισσότερων ομάδων τα δεδομένα συχνά παρουσιάζονται με τη μορφή πίνακα συχνοτήτων.

Ο πίνακας 6 δείχνει ένα παράδειγμα

πίνακα συχνότητων. Ένας πίνακας συχνότητας μπορεί επίσης να θεωρηθεί ως κατ' αντιπαράσταση πινακοποίηση δύο κατηγορικών μεταβλητών, που οποιαδήποτε από τις δύο ή και οι δύο μπορεί να είναι η διατεταγμένη σε τάξη μεγέθους.

Όταν υπάρχουν μόνο δύο κατηγορίες για μία από τις μεταβλητές, π.χ. εάν ο ασθενής παρουσιάζει ή όχι ένα συγκεκριμένο σύμπτωμα, τα δεδομένα μπο-

Πίνακας 6. Σχέση μεταξύ συχνότητας καισαρικής τομής και μεγέθους παπουτσιού της μητέρας.

Καισαρική τομή	Μέγεθος παπουτσιού						Σύνολο
	<35	35	36	37	38	≥39	
Ναί	5	7	6	7	8	10	43
Όχι	17	28	36	41	46	140	308
Σύνολο	22	25	42	48	54	150	351

Η προσέγγιση του πίνακα συχνότητας είναι πιο συνηθισμένη, αλλά η σύγκριση των ποσοστών είναι προτιμότερη, επειδή παρέχει ευκολότερα εκτιμήσεις και διαστήματα εμπιστοσύνης. Για μεγαλύτερους πίνακες όπου και οι δύο μεταβλητές έχουν τουλάχιστον τρεις κατηγορίες δεν υπάρχει απλή εναλλακτική λύση και χρησιμοποιούμε μεθόδους που είναι κατάλληλες για την ανάλυση πινάκων συχνότητας. Η απλούστερη περίπτωση προς εξέταση είναι όταν έχουμε μια μεμονωμένη ομάδα ατόμων και έχουμε παρατηρήσει ότι μια αναλογία από αυτά έχουν ένα ιδιαίτερο χαρακτηριστικό. Τι μπορούμε να πούμε σχετικά με την αναλογία του πληθυσμού που έχει αυτό το χαρακτηριστικό;

ΜΙΑ ΟΜΑΔΑ ΑΤΟΜΩΝ

Διάστημα εμπιστοσύνης για το πληθυσμιακό ποσοστό

Ας υποθέσουμε ότι ένας γενικός γιατρός διαλέγει στην τύχη 215 γυναίκες από τη λίστα των ασθενών του και δια-

ρούν να συνοψιστούν ως αναλογία του ολικού αριθμού ατόμων σε μια από τις κατηγορίες. Τα δεδομένα του πίνακα 6 μπορούν να εκφραστούν ως αναλογία γυναικών που υποβλήθηκαν σε καισαρική τομή σε καθεμιά από τις ομάδες μεγέθους παπουτσιού. Γι αυτόν τον τύπο δεδομένων θα περιγράψουμε την ανάλυση των κατηγορικών δεδομένων είτε ως ποσοστά είτε ως πίνακα συχνότητων.

πιστώνει ότι 39 από αυτές έχουν ιστορικό άσθματος. Θα χρησιμοποιήσουμε το σύμβολο r για τον αριθμό των ασθενών που έχουν το ειδικό χαρακτηριστικό, n το μέγεθος του δείγματος και p το ποσοστό των ασθενών με το ειδικό χαρακτηριστικό. Έτσι, στο παράδειγμά μας, $p=r/n=0,18$. Η κατανομή του ποσοστού του δείγματος των ασθενών με το ειδικό χαρακτηριστικό είναι διωνυμική. Μπορούμε, εντούτοις, να χρησιμοποιήσουμε την κανονική προσέγγιση στη διωνυμική κατανομή για να βρούμε το τυπικό σφάλμα της αναλογίας που παρατηρήθηκε και έτσι να υπολογίσουμε ένα διάστημα εμπιστοσύνης για το πληθυσμιακό ποσοστό p . Το τυπικό/σταθερό σφάλμα του ποσοστού p είναι: $se(p) = \sqrt{p(1-p)/n}$. Έτσι, το τυπικό σφάλμα της αναλογίας γυναικών με άσθμα που παρατηρήθηκε είναι: $\sqrt{[0,18 \times 0,82/215]}=0,0262$. Το 95% διάστημα εμπιστοσύνης του ποσοστού των γυναικών με άσθμα είναι κατά συνέπεια από: $0,18-1,96 \times 0,0262$ έως $0,18+1,96 \times 0,0262$, δηλαδή από 0,13 έως

0,23. Εάν υποθέσουμε ότι οι γυναίκες πελάτες αυτού του γενικού γιατρού είναι αντιπροσωπευτικό δείγμα των γυναικών της χώρας, τότε μπορούμε λογικά να είμαστε βέβαιοι, με βάση αυτό το δείγμα, ότι η συχνότητα του άσθματος στις γυναίκες σε εθνικό επίπεδο κυμαίνεται μεταξύ 13% και 23%.

Έλεγχος υποθέσεων για το πληθυσμιακό ποσοστό

Μπορούμε να ελέγξουμε τη μηδενική υπόθεση ότι το ποσοστό των ασθενών με το ειδικό χαρακτηριστικό (p) είναι μια προκαθορισμένη τιμή (p_0). Για να το κάνουμε αυτό χρησιμοποιούμε τον γενικό στατιστικό τύπο:

(παρατηρηθείσα τιμή - αναμενόμενη τιμή) / τυπικό σφάλμα της παρατηρηθείσας τιμής

που ακολουθεί την τυποποιημένη κανονική κατανομή κάτω από τη μηδενική υπόθεση ($H_0: p = p_0$). Έτσι, υπολογίζουμε:

$$z = (p - p_0) / se(p)$$

Σημειώστε ότι, επειδή ελέγχουμε τη μηδενική υπόθεση, χρησιμοποιούμε το τυπικό σφάλμα του ποσοστού που αναμένεται αν η μηδενική υπόθεση είναι αληθινή. Με άλλα λόγια, έχουμε:

$$se(p) = \sqrt{[p_0(1-p_0)/n]}$$

το οποίο διαφέρει ελαφρώς από το τυπικό σφάλμα που χρησιμοποιήσαμε για να βρούμε το διάστημα εμπιστοσύνης. Εάν θέλουμε να ελέγξουμε την προκαθορισμένη υπόθεση ότι το ποσοστό του άσθματος στις γυναίκες σε εθνική κλίμακα είναι 15%, υπολογίζουμε:

$$se(p) = \sqrt{[(0,15 \times 0,85)/215]}$$

και έτσι:

$$z = (0,18 - 0,15) / 0,0244 = 1,23$$

το οποίο, από τον πίνακα, αντιστοιχεί σε τιμή $p = 0,22 > 0,05$. Δεν μπορούμε να απορρίψουμε τη μηδενική υπόθεση ότι η

συχνότητα του άσθματος στις γυναίκες είναι 15%, και έτσι χρησιμοποιούμε το διάστημα εμπιστοσύνης που δόθηκε πιο πάνω για να δώσουμε ένα εύρος που είναι πιθανό να περιλαμβάνει την πραγματική συχνότητα. Η κανονική κατανομή προσεγγίζει καλύτερα τη διωνυμική κατανομή όταν κάνουμε μια μικρή διόρθωση $\frac{1}{2}$ στην παρατηρηθείσα συχνότητα, για να δώσουμε περιθώριο στο γεγονός ότι η μεταβλητή μπορεί να λάβει μόνο ακέραιες τιμές. Η στατιστική παράμετρος του ελέγχου, με τη διόρθωση συνεχείας είναι:

$$z_c = \{|p - p_{\text{exp}}| - \frac{1}{2}n\} / se(p)$$

όπου τα σύμβολα $| \dots |$ δείχνουν ότι το σημείο της διαφοράς μεταξύ των αναλογιών αγνοείται και το $se(p)$ παραμένει αμετάβλητο.

Δύο ομάδες ατόμων

Πιθανώς, το πιο συνηθισμένο ερώτημα στην ιατρική έρευνα αφορά τα ποσοστά που παρατηρούνται σε δύο ανεξάρτητες ομάδες. Για παράδειγμα θα εξετάσουμε δεδομένα μιας τυχαιοποιημένης κλινικής δοκιμής, η οποία αφορούσε τη διέγερση με υπέρυθρη ακτινοβολία (IRS) σε σύγκριση με εικονική θεραπεία για τον πόνο της αυχενικής οστεοαρθροπάθειας. Η εικονική θεραπεία ήταν ψεύτικος διαδερματικός ηλεκτρικός ερεθισμός και οι ασθενείς δεν ήσαν σε γνώση του είδους της θεραπείας στην οποία υποβάλλονταν. Είκοσι έξι ασθενείς εντάχθηκαν στη δοκιμή, αλλά ένας αποσύρθηκε πριν από το τέλος της δοκιμής. Εννέα από τους 12 ασθενείς της ομάδας IRS ανέφεραν βελτίωση όσον αφορά τον πόνο, σε σύγκριση με 4 από τους 13 της ομάδας εικονικής θεραπείας. Τα παρατηρηθέντα ποσοστά βελτίωσης ήταν λοιπόν 0,75 προς 0,31, με διαφορά 0,44. Για να υπολογιστεί το διάστημα

εμπιστοσύνης ή να γίνει ο έλεγχος της υπόθεσης, χρειάζεται να εξετάσουμε την κατανομή του δείγματος της διαφοράς μεταξύ των δύο ποσοστών.

Διάστημα εμπιστοσύνης για τη διαφορά δύο πληθυσμιακών ποσοστών

Το τυπικό σφάλμα της διαφοράς μεταξύ των ποσοστών p_1-p_2 που παρατηρήθηκαν είναι:

$$se(p_1-p_2) = \sqrt{[var(p_1)+var(p_2)]} \\ = \sqrt{\{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2\}}$$

Για μεγάλα δείγματα, η κατανομή της διαφοράς των πληθυσμιακών ποσοστών p_1 και p_2 θα είναι κατά προσέγγιση κανονική. Μπορούμε λοιπόν να υπολογίσουμε το 95% διάστημα εμπιστοσύνης πολύ απλά από τον τύπο: $p_1 - p_2 \pm 1,96 \times se(p_1-p_2)$.

Στο παράδειγμά μας, η διαφορά των ποσοστών που παρατηρήθηκαν είναι:

$$p_1-p_2 = 0,7500-0,3077 = 0,4423$$

και το τυπικό σφάλμα είναι:

$$se(p_1 - p_2) = \sqrt{[0,75 \times 0,25/12 + 0,3077 \times 0,6923/13]} = 0,1789.$$

Το διάστημα εμπιστοσύνης για τη διαφορά των ποσοστών ως προς την ανακούφιση του πόνου είναι λοιπόν: $0,4423 \pm 1,96 \times 0,1789$ ή $[0,09, 0,791]$.

ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ ΓΙΑ ΤΗ ΔΙΑΦΟΡΑ ΔΥΟ ΠΛΗΘΥΣΜΙΑΚΩΝ ΠΟΣΟΣΤΩΝ p_1-p_2

Παρόμοια προσέγγιση υιοθετείται όταν κάνουμε τον έλεγχο της υπόθεσης για τη σύγκριση των δύο ποσοστών υπολογίζεται ξανά το τυπικό σφάλμα της διαφοράς των ποσοστών, αλλά επειδή αξιολογούμε την πιθανότητα των δεδομένων με την παραδοχή ότι αληθεύει η μηδενική υπόθεση, υπολογίζουμε ένα ελαφρώς διαφορετικό τυπικό σφάλμα. Εάν η μηδενική υπόθεση είναι αληθινή, τα δύο δείγματα προέρχονται από πλη-

θυσμούς που έχουν το ίδιο ποσοστό ατόμων με το χαρακτηριστικό που μας ενδιαφέρει, ας πούμε p . Δεν γνωρίζουμε το p , αλλά τόσο το p_1 όσο και το p_2 είναι εκτιμήσεις του p . Η καλύτερη εκτίμηση του p δίνεται με τον υπολογισμό του ποσοστού των ατόμων που έχουν το χαρακτηριστικό, χρησιμοποιώντας όλα τα δεδομένα και των δύο δειγμάτων σε συνδυασμό. Έτσι:

$$p = r_1+r_2 / n_1+n_2.$$

Το τυπικό σφάλμα της διαφοράς των πληθυσμιακών ποσοστών p_1 και p_2 (κάτω από τη μηδενική υπόθεση $H_0: p_1=p_2$) υπολογίζεται με την παραδοχή ότι το ποσοστό σε κάθε ομάδα είναι p .

Έτσι έχουμε:

$$se(p_1-p_2) = \sqrt{\{[p_1(1-p_1)/n_1]+[p_2(1-p_2)/n_2]\}} \\ = \sqrt{[p(1-p)(1/n_1+2/n_2)]}.$$

Η κατανομή του δείγματος της διαφοράς των ποσοστών p_1-p_2 είναι κανονική, έτσι υπολογίζουμε τη στατιστική συνάρτηση z ως:

$$z = p_1-p_2/[se(p_1-p_2)].$$

Στο παράδειγμα, η διαφορά των πληθυσμιακών ποσοστών είναι $p_1-p_2 = 0,4423$. Επειδή $p_1=9/12$ και $p_2=4/13$, η εκτίμηση του ποσοστού των ασθενών με το ειδικό χαρακτηριστικό για όλο το δείγμα είναι: $p=(9+4)/(12+13)=0,52$ και το τυπικό σφάλμα της διαφοράς των ποσοστών είναι:

$$\sqrt{[0,52 \times 0,48 \times (1/12+1/13)]}=0,20.$$

Το στατιστικό κριτήριο z λοιπόν είναι $z=0,4423/0,20=2,21$. Συνεπώς, από τον πίνακα έχουμε $p=0,027 < 0,05$. Έτσι, υπάρχει ένδειξη διαφοράς μεταξύ των σχημάτων θεραπείας. Όπως όμως είδαμε προηγουμένως, το διάστημα εμπιστοσύνης ως προς τη διαφορά είναι ευρύ, επειδή τα δείγματα είναι μικρά.

Όπως και με την περίπτωση του ενός δείγματος, συνιστάται η χρησιμοποίηση της διόρθωσης συνεχείας κατά τη σύ-

γκριση των δύο ποσοστών, ειδικά όταν τα δείγματα είναι μικρά. Ο τροποποιημένος τύπος για τον υπολογισμό του z μετά τη χρησιμοποίηση διόρθωσης συνεχειάς είναι:

$$z = \frac{|p_1 - p_2| - \frac{1}{2}(1/n_1 + 1/n_2)}{se(p_1 - p_2)}$$
 όπου το $se(p_1 - p_2)$ είναι αμετάβλητο.

Η ανάλυση των πινάκων συχνότητας

Τα ποσοστά είναι ένας τρόπος έκφρασης των συχνοτήτων όταν υπάρχουν μόνο δύο πιθανές εκβάσεις, όπως π.χ. παρουσία ή απουσία ενός συμπτώματος. Υποθέτουμε τώρα ότι τα άτομα που απαρτίζουν έναν πληθυσμό μπορούν να ταξινομηθούν σε t κατηγορίες A_1, A_2, \dots, A_t , κατά το κριτήριο A και σε s κατηγορίες B_1, B_2, \dots, B_s , κατά το κριτήριο B . Έτσι ένα άτομο θα ανήκει

στο κελί (i, j) αν ανήκει συγχρόνως στην κατηγορία A_i του A κριτηρίου και B_j του κριτηρίου B . Σε τυχαίο δείγμα n ατόμων συμβολίζουμε με n_{ij} τις συχνότητες των ατόμων που ανήκουν στο κελί (i, j) . Τα n_{ij} αποτελούν τα στοιχεία ενός πίνακα που ονομάζεται πίνακας συνάφειας διότι σε αυτόν στηρίζεται ο έλεγχος του βαθμού συσχέτισης ή εξάρτησης των δύο κριτηρίων A και B . Όταν έχουμε δύο κριτήρια και τις αντίστοιχες συχνότητες n_{ij} σε μορφή πίνακα, αυτός ονομάζεται *πίνακας συνάφειας δύο διαστάσεων*.

Υπάρχει μια ενιαία γενική προσέγγιση στην ανάλυση όλων των πινάκων συχνότητας, αλλά στην πράξη η μέθοδος ανάλυσης ποικίλλει, σύμφωνα με:

1. τον αριθμό των κατηγοριών,
2. την ταξινόμηση ή όχι των κατηγοριών κατά σειρά μεγέθους,
3. τον αριθμό των ανεξάρτητων ομάδων ατόμων, και
4. τη φύση της ερώτησης που τίθεται.

Πίνακας συνάφειας δύο διαστάσεων

Η ανάλυση των πινάκων 2×2 ακολουθεί την ίδια βασική μέθοδο που χρησιμοποιείται και σε μεγαλύτερους πίνακες, αλλά υπάρχουν μερικά ιδιαίτερα χαρακτηριστικά που πρέπει να σημειωθούν. Ο πίνακας 7 δείχνει δεδομένα από μια δοκιμή ασθενών-μαρτύρων που έγινε μεταξύ κολυμβητών, για να διαπιστωθεί η πιθανή συσχέτιση του χλωριωμένου νερού της πισίνας και της φθοράς της αδαμαντίνης των δοντιών. Μεταξύ 49 κολυμβητών με φθορά της αδαμαντίνης (οι περιπτώσεις) 32 ανέφεραν κολύμβηση έξι και πλέον ωρών την εβδομάδα, σε σύγκριση με 118 από 245 κολυμβητές χωρίς βλάβες της αδαμαντίνης (οι μάρτυρες). Μπορούμε να δούμε, αν και τα δεδομένα παρουσιάζονται με τη μορφή πίνακα συχνοτήτων 2×2 , ότι η σύγκριση των ομάδων

είναι στην πραγματικότητα σύγκριση δύο ποσοστών.

Πίνακας 7. Σύγκριση του αριθμού των ωρών κολύμβησης από κολυμβητές με ή χωρίς φθορά της αδαμαντίνης των δοντιών.

Χρόνος κολύμβησης	Φθορά της αδαμαντίνης των δοντιών		Σύνολο
	Ναι (περιπτώσεις)	Όχι (μάρτυρες)	
<6 ώρες	32	118	150
≥6 ώρες	17	127	144
Σύνολο	49	245	294

Η μηδενική υπόθεση είναι ότι η φθορά της αδαμαντίνης δεν σχετίζεται με τον χρόνο κολύμβησης (δηλ. την έκθεση στο χλωριωμένο νερό). Για να εφαρμόσουμε τη δοκιμασία χ^2 χρειάζεται να υπολογίσουμε εάν η μηδενική υπόθεση αληθεύει για τις αναμενόμενες συχνότητες. Θα βοηθήσει στους υπολογισμούς να χρησιμοποιήσουμε τα σύμβολα a, b, c και d αντί να γράψουμε τις παρατηρηθείσες συχνότητες, όπως στον πίνακα 8.

Πίνακας 8. Γενικός πίνακας 2x2

	Στήλη 1	Στήλη 2	Σύνολο
Σειρά 1	a	b	a + b
Σειρά 2	c	d	c + d
Σύνολο	a + c	b + d	N

Η αναμενόμενη συχνότητα σε ένα κελί εκτιμάται από την εξίσωση: σύνολο Σειράς/σύνολο Στήλης (που αντιστοιχούν στο κελί αυτό)/γενικό Σύνολο.

Έτσι για το κελί με την παρατηρηθείσα συχνότητα a, η αναμενόμενη τιμή εί-

Πίνακας 9. Αναμενόμενες συχνότητες και συμβολή στο κριτήριο χ^2 για τα δεδομένα του πίνακα 7

Παρατηρηθείσα συχνότητα (O)	Αναμενόμενη συχνότητα (E)	O - E	(O - E) ² / E
a = 32	E (a) = 25	7	1,960
b = 118	E (b) = 125	-7	0,392
c = 17	E (c) = 24	-7	2,042
d = 127	E (d) = 120	7	0,408
Σύνολο 294	294	0	$\chi^2 = 4,802$

Για ένα πίνακα 2x2 ο τύπος για το κριτήριο χ^2 μπορεί να απλοποιηθεί ως εξής:

$$\chi^2 = N(ad - bc)^2 / (a+b)(a+c)(b+d)(c+d)$$

Στη γενική περίπτωση ενός πίνακα c x c το κατάλληλο στατιστικό κριτήριο γράφεται απλά ως:

$$\chi^2 = \sum (O-E)^2 / E$$

Οπότε, για τα δεδομένα του πίνακα 7 έχουμε:

$$\chi^2 = 294 \times (32 \times 127 - 118 \times 17)^2 / 150 \times 49 \times 245 \times 144 = 4,802$$

Από τον πίνακα βρίσκουμε $p < 0,05$ το οποίο υποδηλώνει ότι υπάρχουν ενδείξεις που υποστηρίζουν ότι υπάρχει συσχέτιση μεταξύ του χρόνου κολύμβησης και της φθοράς της αδαμαντίνης των δοντιών.

Σχόλιο: Όταν το μέγεθος των δειγμάτων είναι μικρό χρησιμοποιούμε μια διόρθωση συνεχείας που είναι γνωστή ως διόρθωση Yates, από το όνομα του στατιστικολόγου που την επινόησε. Ο τύπος, με τη διόρθωση Yates γίνεται:

$$\chi^2 = n(|ad-bc| - n/2)^2 / (a+b)(a+c)(b+d)(c+d).$$

ναί $(a+b)(a+c)/n$.

Με τα δεδομένα του πίνακα 7 οι αναμενόμενες συχνότητες και η συμβολή τους στο κριτήριο χ^2 δίνονται στον πίνακα 9. Η διαφορά O-E (observed - expected) είναι η ίδια, εκτός από το πρόσημο. και για τα τέσσερα κελιά, και αυτό είναι αληθινό για όλους τους πίνακες 2x2. Αυτό δείχνει ότι έχουμε μία ανεξάρτητη παρατήρηση μάλλον παρά τέσσερις και συνεπώς ένα βαθμό ελευθερίας.

Μέγεθος δείγματος - Δοκιμασία Fisher

Η χρησιμοποίηση της κατανομής χ^2 για το στατιστικό κριτήριο χ^2 , βασίζεται σε μια προσέγγιση «μεγάλου δείγματος». Στο πλαίσιο των πινάκων συχνότητας υπάρχουν μερικές αρκετά σαφείς οδηγίες για το πόσο μεγάλες πρέπει να είναι οι συχνότητες, ώστε να είναι έγκυρη η μέθοδος. Σύμφωνα με τις οδηγίες αυτές το 80% των κελιών του πίνακα θα πρέπει να έχουν αναμενόμενες συχνότητες μεγαλύτερες του 5 και ότι όλα τα κελιά πρέπει να έχουν αναμενόμενες συχνότητες μεγαλύτερες του 1. Υπάρχει μια εναλλακτική προσέγγιση για πίνακες με πολύ μικρές αναμενόμενες συχνότητες η ακριβής δοκιμασία Fisher. Αν και η μέθοδος είναι διαφορετική ως προς την αρχή της από οποιαδήποτε άλλη που έχει περιγραφεί στο κεφάλαιο αυτό, βασίζεται και αυτή στα σύνολα των συχνότητων που παρατηρήθηκαν σε σειρές και στήλες. Η μέθοδος συνίσταται στην

αξιολόγηση της πιθανότητας που σχετίζεται με όλους τους δυνατούς πίνακες 2x2 που έχουν τα ίδια σύνολα σειρών και στηλών με τα δεδομένα που παρατηρήθηκαν, κάνοντας την παραδοχή ότι η μηδενική υπόθεση αληθεύει. Όπως και πριν, η μηδενική υπόθεση εδώ είναι ότι οι μεταβλητές των σειρών και των στηλών είναι άσχετες. Ο μαθηματικός τύπος για τον υπολογισμό κάθε πιθανότητας είναι μάλλον περίπλοκος, γι αυτό ο υπολογισμός γίνεται πολύ καλύτερα με τη κατάλληλου προγράμματος στον Η/Υ.

ΕΠΙΛΕΓΜΕΝΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Bradford Hill A.(1971) Principles of medical statistics, 9th ed., The Lancet Limited, London
2. Glantz S.A.(1992) Primer of biostatistics, 3rd ed., McGraw-Hill Inc., New York
3. Mosteller F., Rourke R.(1973) Sturdy statistics: Non-parametrics and older statistics, Addison-Wesley, Reading
4. Feinstein A.F.(1977) Clinical biostatistics, Mosby, St.Louis
5. Colton T.(1974) Statistics in Medicine, Little Brown, Boston
6. Mainland D.(1963) Elementary medical statistics, 2nd ed., Saunders, Philadelphia
7. Leaverton P.E.(1978) A review of biostatistics, 2nd ed., Little Brown, Boston
8. Πανάγος Γ. Priori: Πρόγραμμα σε Basic για τον καθορισμό του απαιτούμενου μεγέθους δείγματος σε κλινικές ή άλλες βιοϊατρικές μελέτες. Ογκολογικά Αρχεία 1:43-46, 1992