

**Ανάλυση δεδομένων:
Μη μπαγιεσιανή προσέγγιση[#]**

Π. Γαλάνης,^{##}
Λ. Σπάρος

Εργαστήριο Κλινικής Επιδημιολογίας,
Τμήμα Νοσηλευτικής,
Πανεπιστήμιο Αθηνών, Αθήνα

Data analysis:
Non Bayesian approach

Abstract at the end of the article

1. Εισαγωγή
2. Σφάλματα μέτρησης
 - 2.1. Τυχαία σφάλματα μέτρησης
 - 2.2. Συστηματικά σφάλματα μέτρησης
 - 2.3. Συγχυτές
3. Έλεγχοι της στατιστικής σημαντικότητας
 - 3.1. Η έννοια του παρατηρούμενου επιπέδου σημαντικότητας σύμφωνα με τον Fisher
 - 3.2. Παρατηρούμενο επίπεδο σημαντικότητας
 - 3.3. Σφάλματα τύπου I και II
4. Στατιστική εκτίμηση
 - 4.1. Σημειακή εκτίμηση
 - 4.2. Διαστήματα εμπιστοσύνης
5. Διαστήματα εμπιστοσύνης και έλεγχος της στατιστικής σημαντικότητας
6. Η συνάρτηση του παρατηρούμενου επιπέδου σημαντικότητας
7. Σύνοψη

Λέξεις ευρετηρίου

Διάστημα εμπιστοσύνης
Σημειακή εκτίμηση
Συστηματικό σφάλμα
Τυχαίο σφάλμα

Υποβλήθηκε 14.1.2005
Εγκρίθηκε 8.2.2005

1. ΕΙΣΑΓΩΓΗ

Τα τελευταία χρόνια έχει μεταβληθεί ο ρόλος της στατιστικής στην ανάλυση των δεδομένων μιας επιδημιολογικής μελέτης.¹ Μέχρι το 1970, ο ρόλος της στατιστικής στην επιδημιολογία εστιαζόταν στο αν η «τύχη» (chance) ή το «τυχαίο σφάλμα» (random error) ευθυνόταν αποκλειστικά για την παρατηρηθείσα σχέση* σε μια μελέτη. Για να ελεγχθεί αυτό, οι συνηθέστερες μέθοδοι ήταν οι παραδοσιακοί έλεγχοι της «στατιστικής σημαντικότητας», που χρησιμοποιούνταν κυρίως από τους Βρετανούς ερευνητές, και ο «έλεγχος της υπόθεσης» των Neyman-Pearson, που χρησιμοποιούνταν κυρίως από τους Αμερικανούς ερευνητές.^{1,2} Επειδή οι δύο αυτές μέθοδοι παρουσιάζουν σημαντικές ομοιότητες, χρησιμοποιείται ο όρος «έλεγχος της στατιστικής σημαντικότητας» (statistical significance testing) για την περιγραφή και των δύο.

Οι επιδημιολόγοι και οι στατιστικοί άσκησαν έντονη κριτική στην αξία αυτών των μεθόδων κατά την περίοδο 1970-1980. Η κριτική εστιαζόταν κυρίως στο ότι η επιδημιολογική έρευνα δεν αρκείται απλά στο να διαπι-

στώσει αν μια παρατηρηθείσα σχέση οφείλεται αποκλειστικά στην τύχη. Πολύ μεγαλύτερη σημασία έχουν ο υπολογισμός του μεγέθους μιας σχέσης, καθώς και η εκτίμηση της ακρίβειας με την οποία υπολογίζεται το μέγεθος της σχέσης. Για την επίτευξη αυτού του στόχου, οι περισσότεροι προτείνουν τη χρήση των διαστημάτων εμπιστοσύνης (confidence intervals ή interval estimation) έναντι των ελέγχων της στατιστικής σημαντικότητας.³⁻¹⁸ Το διάστημα εμπιστοσύνης (ΔΕ) είναι ένα εύρος τιμών γύρω από μια σημειακή εκτίμηση (point estimation), που δείχνει το βαθμό στατιστικής ακρίβειας της εκτίμησης, με την προϋπόθεση ότι η διαφορά ανάμεσα στην παρατηρούμενη τιμή σε μια μελέτη και την πραγματική τιμή της μελετώμενης σχέσης οφείλεται σε τυχαία μεταβλητότητα. Το ΔΕ, δηλαδή, αφορά στην περίπτωση κατά την οποία σε μια μελέτη υπάρχει μόνο τυχαίο σφάλμα.¹

* Σχέση ή, καλύτερα, συνάρτηση συχνότητας (occurrence function). Οι γνωστικές συναρτήσεις, αιτιογνωστική, διαγνωστική ή προγνωστική συνάρτηση, είναι συναρτήσεις παλινδρόμησης (regression function) και αποτελούν το αντικείμενο (object) των επιδημιολογικών μελετών.

[#] Μερική χρηματοδότηση: Ειδικός Λογαριασμός Κονδυλίων Έρευνας Πανεπιστημίου Αθηνών

^{##} Ο Π. Γαλάνης είναι υπότροφος του Ιδρύματος Α.Σ. ΩΝΑΣΗ

Σήμερα, ο ρόλος της στατιστικής στην «ανάλυση των επιδημιολογικών δεδομένων»* εστιάζεται σε δύο σημεία: το πρώτο είναι η εκτίμηση της μεταβλητότητας των δεδομένων και το δεύτερο είναι η εκτίμηση του «αποτελέσματος»** μιας μελέτης, αφού προηγουμένως έχουν εκτιμηθεί και ελεγχθεί όλα τα συστηματικά σφάλματα.¹⁹

2. ΣΦΑΛΜΑΤΑ ΜΕΤΡΗΣΗΣ

Σε κάθε επιδημιολογική μελέτη πραγματοποιούνται μετρήσεις. Για να είναι αξιόπιστη και έγκυρη η μέτρηση,*** θα πρέπει στο εξαγόμενο αποτέλεσμα να μην υπεισέρχονται σφάλματα.¹⁹ Σε μια επιδημιολογική μελέτη μπορεί να υπάρχουν τρία είδη σφάλματος: το τυχαίο σφάλμα, το συστηματικό σφάλμα και οι συγχυτές.^{1,19-21}

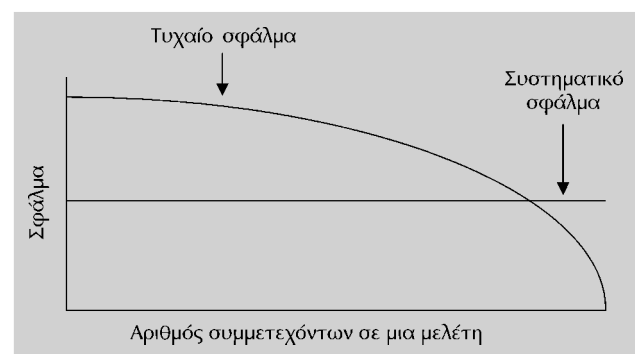
Με τον όρο σφάλμα μέτρησης νοείται κάθε απόκλιση της τιμής μιας (βιολογικής) μεταβλητής από την πραγματική τιμή, που δεν οφείλεται σε λάθος (όπως άγνοια, λαθεμένη αρχή της μεθόδου κ.ά.). Οι βιολογικές μεταβλητές, ανεξάρτητα από τη φύση τους, υπόκεινται εγγενώς κατά τη «μέτρησή» τους στον κίνδυνο σφάλματος. Από το μέγεθος του σφάλματος καθορίζεται και η ποιότητα της μέτρησης. Τα σφάλματα μέτρησης ταξινομούνται σε τυχαία (random) και συστηματικά (systematic).

Στην πλειοψηφία των επιδημιολογικών μελετών, σκοπός είναι ο υπολογισμός ενός μέτρου αποτελέσματος (measure of effect), όπως π.χ. είναι ο λόγος των επιπτώσεων-πυκνοτήτων (incidence-density ratio).¹⁹ Σε μια μελέτη, π.χ., στην οποία σκοπός είναι να υπολογιστεί ο λόγος των επιπτώσεων-πυκνοτήτων του καρκίνου του πνεύμονα σε εκείνους που καπνίζουν σε σχέση με εκείνους που δεν καπνίζουν, η εμπειρική τιμή του λόγου που διαπιστώνεται αποτελεί εκτίμηση της πραγματικής τιμής του λόγου των επιπτώσεων-πυκνοτήτων (εκτίμηση της παραμέτρου της σχέσης).¹⁹ Όσο πιο κοντά είναι η τιμή που προκύπτει από τη μελέτη στην πραγματική τιμή,

τόσο μικρότερο είναι το σφάλμα που υπεισέρχεται στη μελέτη και τόσο πιο αυθεντική (accurate)* είναι η μελέτη αυτή. Αντίθετα, εάν η τιμή που προκύπτει από τη μελέτη διαφέρει σημαντικά από την πραγματική τιμή, τότε η μελέτη θεωρείται μη αυθεντική (inaccurate). Το πρόβλημα βέβαια συνίσταται ότι δεν είναι γνωστή η πραγματική (θεωρητική) τιμή του λόγου των επιπτώσεων-πυκνοτήτων για τον καρκίνο του πνεύμονα σε εκείνους που καπνίζουν σε σχέση με εκείνους που δεν καπνίζουν. Το μόνο γνωστό είναι το αποτέλεσμα μιας μελέτης. Προφανώς, αυτό δεν ισχύει μόνο στη συγκεκριμένη περίπτωση, καθώς οι πραγματικές τιμές των μέτρων αποτελέσματος, αλλά και των μέτρων συχνότητας, είναι πάντα άγνωστες. Ακριβώς επειδή οι πραγματικές τιμές είναι άγνωστες, δεν είναι δυνατόν να καθοριστεί με ακρίβεια το σφάλμα σε μια μελέτη. Για το λόγο αυτόν, ο σχεδιασμός των επιδημιολογικών μελετών απαιτεί ιδιαίτερη προσοχή, έτσι ώστε να περιορίζονται, όσο το δυνατόν περισσότερο, τα σφάλματα.

Υπάρχει ένας απλός τρόπος για να διαχωρίζονται τα τυχαία από τα συστηματικά σφάλματα. Όσο ο αριθμός των συμμετεχόντων αυξάνει συνεχώς, τείνοντας προς το άπειρο, τόσο μειώνεται το τυχαίο σφάλμα. Όταν ο αριθμός αυτός, θεωρητικά, φτάσει το άπειρο, τότε το τυχαίο σφάλμα μηδενίζεται.¹⁹ Αντίθετα, η αύξηση του αριθμού των συμμετεχόντων δεν επηρεάζει το συστηματικό σφάλμα, το οποίο παραμένει σταθερό (εικ. 1).¹⁹

Για την καλύτερη κατανόηση της διαφοράς ανάμεσα στα τυχαία και τα συστηματικά σφάλματα θεωρήστε το



Εικόνα 1. Σχέση τυχαίου και συστηματικού σφάλματος και του αριθμού των συμμετεχόντων σε μια μελέτη.¹⁹

* Η έκφραση «ανάλυση δεδομένων» (data analysis) τείνει να εγκαταλειφθεί ως λαθεμένη, δεδομένου (α) ότι δεν υπάρχουν, σύμφωνα με τη σύγχρονη επιστημολογία, καθαρά δεδομένα στις εμπειρικές (πραγματολογικές) επιστήμες που δεν είναι θεωρητικά φορτισμένα και (β) τα «δεδομένα» αυτά δεν αναλύονται, αλλά συντίθενται. Εάν, δηλαδή, διατηρηθεί ο όρος δεδομένα, τότε η έκφραση «ανάλυση δεδομένων» θα πρέπει να αντικατασταθεί από την έκφραση «σύνθεση δεδομένων».

** Ο όρος «αποτέλεσμα» στην επιστήμη σημαίνει τη μεταβολή της παραμέτρου της έκβασης που προκαλείται από ένα συγκεκριμένο αίτιο. Στη στατιστική, ο όρος «αποτέλεσμα» σημαίνει την τιμή της παραμέτρου μιας σχέσης.

*** Μέτρηση είναι ο εμπειρικός καθορισμός των τιμών της αριθμητικής συνάρτησης.

* Η αυθεντικότητα (accuracy) αναφέρεται σε ποσοτικά (μετρικά) και μόνο μεγέθη (ή έννοιες) και σημαίνει την απουσία τόσο τυχαίων όσο και συστηματικών σφαλμάτων. Είναι λάθος να χαρακτηρίζεται το αποτέλεσμα μιας εργαστηριακής δοκιμασίας ως αυθεντικό (accurate), όταν εκφράζεται ως θετικό ή αρνητικό (ποιοτικό χαρακτηριστικό ή έννοια).

ακόλουθο παράδειγμα.¹⁹ Ένας ερευνητής θέλει να υπολογίσει το μέσο ανάστημα των γυναικών μιας πόλης, στην οποία υπάρχουν συνολικά 500.000 γυναίκες. Για το σκοπό αυτό, λαμβάνει ένα τυχαίο δείγμα 100 γυναικών από αυτή την πόλη και πραγματοποιεί 100 μετρήσεις αναστήματος, χρησιμοποιώντας πάντοτε την ίδια μετροταινία. Το μέσο ανάστημα αυτών των 100 γυναικών αποτελεί μια εκτίμηση του μέσου αναστήματος όλων των γυναικών της πόλης. Το ερώτημα που τίθεται είναι ποια σφάλματα επηρεάζουν την εκτίμηση αυτή. Η ακριβής και έγκυρη μέτρηση του αναστήματος εξαρτάται από διάφορους παράγοντες, όπως το ποιος πραγματοποιεί τη μέτρηση, ο τρόπος με τον οποίο ο ερευνητής κρατάει τη μετροταινία, ο τρόπος με τον οποίο ο ερευνητής διαβάζει το αποτέλεσμα της μέτρησης κ.ά. Ορισμένα από τα σφάλματα αυτά, όπως π.χ. ο τρόπος με τον οποίο ο ερευνητής κρατάει τη μετροταινία, μπορεί να είναι τυχαία. Στην περίπτωση αυτή, το ανάστημα ορισμένων γυναικών θα βρεθεί, εσφαλμένα, μεγαλύτερο από την πραγματική του τιμή, ενώ το ανάστημα ορισμένων άλλων θα μετρηθεί, και πάλι εσφαλμένα, χαμηλότερο από την πραγματική του τιμή. Στην περίπτωση όμως ενός τυχαίου σφάλματος δεν είναι δυνατόν όλες οι μετρήσεις να είναι πάνω ή κάτω από την πραγματική τιμή. Εάν το δείγμα των γυναικών αυξηθεί από 100 σε 10.000 ή 100.000, τότε το τυχαίο σφάλμα θα ελαττωθεί, καθώς όσο μεγαλύτερος είναι ο αριθμός των γυναικών του δείγματος, τόσο μικρότερη είναι η ασυμφωνία μεταξύ του μέσου αναστήματος των γυναικών του δείγματος και του μέσου αναστήματος όλων των γυναικών της πόλης. Είναι προφανές ότι, αν ο αριθμός των γυναικών του δείγματος πλησιάσει τις 500.000, τότε το τυχαίο σφάλμα μηδενίζεται.

Τα συστηματικά σφάλματα, όπως αναφέρθηκε, δεν επηρεάζονται από το μέγεθος του δείγματος. Αν η μετροταινία με την οποία πραγματοποιούνται οι μετρήσεις είναι π.χ. υφασμάτινη και έχει πρόσφατα πλυθεί, χωρίς να το γνωρίζει ο ερευνητής, τότε, επειδή βραχύνεται, όλες οι μετρήσεις είναι μη έγκυρες και στη μελέτη υπάρχει συστηματικό σφάλμα, δηλαδή η μελέτη είναι μεροληπτική (biased). Και αυτό γιατί όσο κι αν αυξηθεί ο αριθμός των γυναικών του δείγματος, η μετροταινία δεν πρόκειται να αποκτήσει το αρχικό της μήκος, οπότε κάθε μέτρηση καθίσταται μεροληπτική.

2.1. Τυχαία σφάλματα μέτρησης

Τυχαίο σφάλμα* καλείται η διαφορά μεταξύ της εμπειρικής (παρατηρήσιμης) τιμής μιας μεταβλητής που προκύπτει έπειτα από μια μέτρηση και της μέσης τιμής των

εμπειρικών τιμών που προκύπτουν έπειτα από ένα σύνολο μετρήσεων. Τα τυχαία σφάλματα δεν είναι προβλέψιμα κατά την εκτέλεση της μέτρησης.²⁰ Είναι δι-κατευθυνόμενα, δηλαδή οι καθέκαστες τιμές είναι άλλοτε μεγαλύτερες και άλλοτε μικρότερες της μέσης τιμής, και για τη διαπίστώσή τους είναι αναγκαία η επανάληψη της μέτρησης. Δηλαδή, η διαπίστωση του τυχαίου σφάλματος δεν είναι δυνατή όταν υπάρχει μία μόνο μέτρηση. Δεν είναι όμως απαραίτητη η γνώση της πραγματικής τιμής, όπως συμβαίνει στα συστηματικά σφάλματα. Όταν σε μια μέτρηση δεν υπάρχουν τυχαία σφάλματα, τότε η μέτρηση λέγεται ότι έχει ακρίβεια (precision), αξιοπιστία (reliability), αναπαραγωγιμότητα (reproducibility), επαναληπτικότητα (repeatability) ή συνέπεια (consistency). Η παρουσία τυχαίων σφαλμάτων καθιστά τη μέτρηση ακριβή, αναξιόπιστη κ.λπ. Η μέτρηση ή, καλύτερα, η εκτίμηση του τυχαίου σφάλματος είναι συνάρτηση της φύσης (ποιοτικής ή ποσοτικής) της μεταβλητής που μετράται.

Στην περίπτωση των ποσοτικών (μετρικών) μεγεθών (ενοιών), η εκτίμηση του τυχαίου σφάλματος γίνεται με τους παρακάτω δείκτες:²⁰

- Σταθερή απόκλιση
- Συντελεστής μεταβλητότητας
- Μέση απόκλιση
- Σχετική μέση απόκλιση
- Εύρος ή σχετικό εύρος
- Συντελεστής συσχέτισης της αξιοπιστίας
- Συντελεστής συσχέτισης ημικλάστου
- Συντελεστής συσχέτισης ισοδύναμων τύπων
- Αξιοπιστία εξέτασης-επανεξέτασης.

Το συχνότερα χρησιμοποιούμενο μέτρο του τυχαίου σφάλματος είναι ο συντελεστής μεταβλητότητας (coefficient of variation).

Στην περίπτωση των ποιοτικών μεταβλητών,** η εκτίμηση του τυχαίου σφάλματος γίνεται με αξιολογήσεις (μετρήσεις με την ευρεία έννοια) της ίδιας μεταβλητής από δύο, τουλάχιστον, διαφορετικούς εξεταστές.²⁰ Οι εξεταστές θα πρέπει να διαθέτουν το ίδιο περίπου επίπεδο γνώσεων, ώστε το ποσοστό συμφωνίας να εκτιμάει το τυχαίο σφάλμα. Το μέτρο που χρησιμοποιείται αποκλειστικά σήμερα για τη μέτρηση του τυχαίου σφάλματος στα ποιοτικά δεδομένα είναι ο δείκτης κάππα, που υπολογίζεται από τον τύπο:

* Η περιγραφή του τυχαίου σφάλματος στην παρούσα εργασία θα περιοριστεί σε απλά ποσοτικά μεγέθη και όχι σε επιδημιολογικά (μέτρα συχνότητας ή μέτρα σχέσης), για την καλύτερη κατανόηση.

** Καταχρηστικά, η έννοια του σφάλματος αναφέρεται σε ποιοτικές (ή ταξινομητικές) έννοιες, αφού εξ ορισμού οι έννοιες αυτές δεν μετρώνται.

$$\text{Δείκτης κάππα} = \frac{O-C}{1-C}$$

όπου O είναι το ποσοστό της παρατηρηθείσας συμφωνίας και C το ποσοστό της τυχαίας συμφωνίας.

2.2. Συστηματικά σφάλματα μέτρησης

Συστηματικό σφάλμα μέτρησης καλείται η διαφορά μεταξύ της πραγματικής τιμής μιας μεταβλητής και της μέσης τιμής των εμπειρικών τιμών. Η εκτίμηση των συστηματικών σφαλμάτων προϋποθέτει τη γνώση της πραγματικής τιμής της μετρούμενης μεταβλητής.²⁰ Αυτό επιβάλλει την ύπαρξη μιας μεθόδου (μέθοδος αναφοράς ή χρυσός κανόνας) που μπορεί (ή θεωρείται, κατόπιν συμφωνίας) να μετρά την πραγματική τιμή. Η απουσία συστηματικών σφαλμάτων σε μια μέτρηση ονομάζεται εγκυρότητα (validity),* ενώ η παρουσία τους καθιστά τη μέτρηση μεροληπτική (bias). Στην περίπτωση των ποσοτικών μεταβλητών, η εκτίμηση των συστηματικών σφαλμάτων γίνεται με τους παρακάτω δείκτες:²⁰

- Η απόσταση της μέσης τιμής επανειλημμένων μετρήσεων από την πραγματική τιμή (Δ)
- Η προτυπωμένη μεροληψία (standardized bias). Είναι ο λόγος του προηγούμενου μεγέθους διά της σταθερής απόκλισης
- Ο συντελεστής συσχέτισης των «ατελών» τιμών και των πραγματικών τιμών. Ένα ενδιαφέρον μέγεθος είναι το τετράγωνο του συντελεστή αυτού
- Διάφοροι δείκτες προκύπτουν από τη διαίρεση (προτύπωση) της απόστασης Δ από τη μέση απόκλιση.

Όταν οι ποιοτικές (ταξινομικές) μεταβλητές είναι δυαδικές, τα συχνότερα χρησιμοποιούμενα μέτρα εκτίμησης των συστηματικών σφαλμάτων είναι οι λόγοι πιθανοφάνειας (likelihood ratio), θετικός και αρνητικός.

2.3. Συγχυτές

Ως σύγχυση (confounding), σε μια αιτιολογική ή προγνωστική επιδημιολογική μελέτη, ορίζεται η μεταβολή ενός μέτρου αποτελέσματος** που προκαλείται από έναν εξωγενή προσδιοριστή και συμβαίνει μόνο μέσα στα

* Απαιτείται προσοχή στη χρησιμοποίηση του όρου accuracy (αυθεντικότητα), που μέχρι πρόσφατα ήταν συνώνυμος της εγκυρότητας ή και ορθότητας (correctness).²⁰ Τα τελευταία χρόνια, οι περισσότεροι συγγραφείς χρησιμοποιούν τον όρο accuracy, για να περιλάβουν την απουσία τόσο των συστηματικών όσο και των τυχαίων σφαλμάτων.

** Επιδημιολογικά μέτρα αποτελέσματος μιας αδρής σχέσης ή συνάρτησης (crude function) είναι η διαφορά ή ο λόγος των δύο μέτρων συχνότητας που παρατηρούνται στις δύο κατηγορίες του προσδιοριστή.

πλαίσια μιας συγκεκριμένης επιδημιολογικής μελέτης. Έτσι, ένας προσδιοριστής που είναι συγχυτής σε μια μελέτη, μπορεί να μην είναι συγχυτής της ίδιας σχέσης σε μια άλλη μελέτη.²²⁻²⁵ Ο συγχυτής (confounder) πρέπει να έχει τα παρακάτω χαρακτηριστικά:^{1,19,20,26,27}

- Να σχετίζεται με τη συχνότητα της νόσου (να αποτελεί αιτία της νόσου, αλλά όχι αποτέλεσμα της νόσου)
- Να σχετίζεται με το μελετώμενο προσδιοριστή
- Να μην είναι αποτέλεσμα της έκθεσης. Π.χ., θεωρήστε μια δίαιτα πλούσια σε κορεσμένα λίπη, η οποία προκαλεί αύξηση των επιπέδων της λιποπρωτεΐνης χαμηλής πυκνότητας (LDL) στο αίμα. Η αυξημένη LDL ενισχύει την πιθανότητα εμφάνισης της αθηρωματικής νόσου. Έτσι, η αυξημένη LDL σχετίζεται τόσο με τη νόσο (αθηρωματική νόσο) όσο και με την έκθεση (δίαιτα πλούσια σε κορεσμένα λίπη). Ωστόσο, η αυξημένη LDL δεν είναι συγχυτής της σχέσης μεταξύ δίαιτας και αθηρωματικής νόσου, γιατί αποτελεί αποτέλεσμα της έκθεσης, καθώς η πλούσια σε κορεσμένα λίπη δίαιτα προκαλεί αύξηση της LDL.

Χαρακτηριστικό παράδειγμα σύγχυσης αποτελεί η μελέτη, στην οποία άτομα που καταναλώναν μεγάλη ποσότητα οίονοπνεύματος είχαν υψηλότερο κίνδυνο εμφάνισης καρκίνου του πνεύμονα από εκείνα που δεν καταναλώναν οίονοπνευμα. Η αιτία γι' αυτή την αδρή σχέση* ήταν ότι τα άτομα της πρώτης κατηγορίας ήταν σε μεγάλη αναλογία βαρείς καπνιστές. Η σύγχυση από το κάπνισμα εξηγούσε την παρατηρηθείσα σχέση μεταξύ κατανάλωσης οίονοπνεύματος και καρκίνου του πνεύμονα.¹⁶ Το κάπνισμα είναι συγχυτής της σχέσης μεταξύ οίονοπνεύματος και καρκίνου του πνεύμονα, καθώς έχει τα παρακάτω χαρακτηριστικά:

- Το κάπνισμα σχετίζεται με τη νόσο, καθώς αποτελεί μια αιτία του καρκίνου του πνεύμονα
- Το κάπνισμα σχετίζεται με την έκθεση, καθώς τα άτομα που καταναλώνουν μεγάλη ποσότητα οίονοπνεύματος είναι συνήθως άτομα που καπνίζουν περισσότερο
- Το κάπνισμα δεν αποτελεί αποτέλεσμα της έκθεσης, δηλαδή της κατανάλωσης οίονοπνεύματος.

Μια καλά σχεδιασμένη μελέτη απαιτεί την εξουδετέρωση των συγχυτών, η οποία μπορεί να γίνει με διάφορους τρόπους, όπως η εξομοίωση, η διαστρωμάτωση, η πολυπαραμετρική ανάλυση, η προτύπωση, η περιοριστική μέθοδος και το πείραμα.¹⁶ Η εξουδετέρωση όλων

* Αδρή σχέση ή συνάρτηση είναι η σχέση που μελετάται, π.χ. η σχέση οίονοπνεύματος και καρκίνου του πνεύμονα. Εξωγενή χαρακτηριστικά (καταχρηστικώς μεταβλητές) είναι στην προκειμένη περίπτωση το κάπνισμα, το φύλο, η ηλικία κ.ά., που σχετίζονται με τη συχνότητα του καρκίνου του πνεύμονα (δυσνητικό συγχυτές).

των δυνατών, γνωστών και άγνωστων, συγχυτών απο-τελεί απαραίτητη προϋπόθεση για την ανάδειξη μιας σχέσης σε αιτιακή. Η μη εξουδετέρωση όλων των συγχυτών εισάγει σφάλμα στην έρευνα και καθιστά το αποτέλεσμα μη αυθεντικό.

3. ΕΛΕΓΧΟΙ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ ΣΗΜΑΝΤΙΚΟΤΗΤΑΣ

Τα τελευταία χρόνια, έχει ασκηθεί έντονη κριτική στη χρήση των ελέγχων της στατιστικής σημαντικότητας (ΕΣΣ) ως πρωταρχικού «εργαλείου» για την εκτίμηση της δειγματικής μεταβλητότητας (variability) στις επιδημιολογικές μελέτες. Παρά το γεγονός αυτό, αρκετά επιστημονικά περιοδικά, ακόμη και σήμερα, απαιτούν τη χρήση των ΕΣΣ για τη δημοσίευση των μελετών.¹

Ο έλεγχος της επιστημονικής υπόθεσης* περιλαμβάνει δύο είδη υποθέσεων, τη μηδενική υπόθεση (null hypothesis) και την εναλλακτική (alternative).^{1,19,28} Συνηθώς, σε μια επιδημιολογική μελέτη, ο ερευνητής θέλει να μελετήσει αν υπάρχει σχέση ενός προσδιοριστή** και μιας έκβασης.*** Σε μια τέτοια περίπτωση, η μηδενική υπόθεση υποστηρίζει ότι δεν υπάρχει σχέση ανάμεσα στον προσδιοριστή και την έκβαση. Η μηδενική υπόθεση λέγεται επίσης υπόθεση της μη διαφοράς και διατυπώνεται, με σκοπό να αναιρεθεί. Η συμπληρωματική της μηδενικής υπόθεσης καλείται εναλλακτική υπόθεση. Στην παραπάνω περίπτωση, η εναλλακτική υπόθεση είναι ότι υπάρχει σχέση μεταξύ του προσδιοριστή και της έκβασης. Με το στατιστικό έλεγχο που πραγματοποιείται, η μηδενική υπόθεση είτε απορρίπτεται είτε όχι. Σε

καμιά όμως περίπτωση τα δεδομένα μιας μελέτης δεν μπορούν να προσφέρουν απόδειξη ότι η μηδενική υπόθεση είναι αληθής. Αν δεν απορριφθεί η μηδενική υπόθεση, τότε ισχύει ότι τα δεδομένα πάνω στα οποία στηρίζεται ο έλεγχος, δηλαδή τα δεδομένα της μελέτης, δεν επαρκούν για την απόρριψή της. Τονίζεται και πάλι ότι ο στατιστικός έλεγχος των υποθέσεων δεν οδηγεί στην απόδειξη της υπόθεσης, αλλά απλά παρέχει την πληροφορία εάν τα διαθέσιμα δεδομένα στηρίζουν την υπόθεση. Όταν η μηδενική υπόθεση δεν απορρίπτεται, δεν σημαίνει ότι είναι και αληθινή. Αντίθετα, η απόρριψη της μηδενικής υπόθεσης σημαίνει ότι, σύμφωνα με τα δεδομένα μιας συγκεκριμένης μελέτης, υπάρχει σχέση μεταξύ προσδιοριστή και έκβασης.

Η εναλλακτική υπόθεση μπορεί να είναι «μονής κατεύθυνσης» (one-sided), δηλαδή να υποστηρίζεται ότι υπάρχει θετική σχέση μεταξύ προσδιοριστή και έκβασης, ή «διπλής κατεύθυνσης» (two-sided), δηλαδή να υπάρχει σχέση μεταξύ προσδιοριστή και έκβασης, χωρίς όμως να καθορίζεται αν είναι θετική ή αρνητική.^{1,19} Εάν ο ερευνητής γνωρίζει, πριν από τη διεξαγωγή μιας μελέτης, ότι υπάρχει θετική ή αρνητική σχέση μεταξύ προσδιοριστή και έκβασης, τότε η εναλλακτική υπόθεση είναι μονής κατεύθυνσης. Για παράδειγμα, αν μια μελέτη διερευνά τη σχέση ανάμεσα στην καπνισματική συνήθεια (προσδιοριστής) και τη συχνότητα εμφάνισης του καρκίνου του πνεύμονα (έκβαση αποτελεί η εμφάνιση του καρκίνου του πνεύμονα), τότε η εναλλακτική υπόθεση μπορεί να είναι μονής κατεύθυνσης, ότι, δηλαδή, το κάπνισμα αυξάνει τη συχνότητα εμφάνισης του καρκίνου του πνεύμονα. Αντίθετα, αν ο ερευνητής δεν γνωρίζει, πριν από τη διεξαγωγή μιας μελέτης, το είδος της σχέσης, τότε η εναλλακτική υπόθεση είναι διπλής κατεύθυνσης.

Για την καλύτερη κατανόηση των παραπάνω θεωρήστε μια μελέτη, στην οποία ο ερευνητής θέλει να διαπιστώσει αν υπάρχει σχέση μεταξύ της χρήσης κινητών τηλεφώνων (προσδιοριστής) και της συχνότητας εμφάνισης καρκίνου του εγκεφάλου (έκβαση). Σε μια τέτοια μελέτη, η μηδενική υπόθεση είναι ότι δεν υπάρχει σχέση μεταξύ της χρήσης κινητών τηλεφώνων και της συχνότητας εμφάνισης καρκίνου του εγκεφάλου. Η εναλλακτική υπόθεση διπλής κατεύθυνσης είναι ότι υπάρχει σχέση μεταξύ κινητών τηλεφώνων και καρκίνου του εγκεφάλου, χωρίς όμως να καθορίζεται αν η σχέση αυτή είναι θετική ή αρνητική. Η εναλλακτική υπόθεση μονής κατεύθυνσης είναι ότι υπάρχει θετική σχέση μεταξύ της χρήσης κινητών τηλεφώνων και καρκίνου του εγκεφάλου, καθώς ο ερευνητής πιστεύει ότι ο καρκίνος του εγκεφάλου είναι συχνότερος σ' εκείνους που χρησιμο-

* Η επιστημονική υπόθεση είναι καταφατική αποφαντική πρόταση και εισάγει μια νέα πληροφορία. Η «μηδενική» ή «άκυρη» υπόθεση δεν αποτελεί καν επιστημονική πρόταση, αφού δεν υποστηρίζει το διαψεύσιμο ισχυρισμό για την ύπαρξη σχέσης μεταξύ προσδιοριστή και συχνότητας του νοσήματος. Ειδικότερα, η «διπλής κατεύθυνσης» (two-sided) υπόθεση αποτελεί, όπως τονίζει ένας ερευνητής, «εννοιολογικό τέρας» (conceptual monster), που έχει περιεχόμενο στη στατιστική όχι όμως και στις (πραγματολογικές) επιστήμες.

** Παράγοντας κινδύνου (risk factor) ή προσδιοριστής (determinant), όπως τελικά επικράτησε να λέγεται σήμερα, είναι το χαρακτηριστικό (συγγενές, περιβαλλοντικό ή συμπεριφοράς) των ατόμων, από το οποίο εξαρτάται (σχετίζεται ή συναρτάται) η συχνότητα ενός νοσήματος (ή του θανάτου ή, γενικότερα, ενός συμβάντος ή μιας κατάστασης υγείας).²⁰ Η ακαταλληλότητα του όρου παράγοντας κινδύνου προέκυψε από τη λέξη factor (παράγοντας), που σημαίνει αιτία, ενώ ο παράγοντας κινδύνου δεν σχετίζεται κατ' ανάγκη αιτιακά με τη συχνότητα του νοσήματος. Πρέπει να τονιστεί ότι ο προσδιοριστής συναρτάται όχι με την κατανομή του νοσήματος, αλλά με την κατανομή της συχνότητας του νοσήματος. Η εννοιολογική διαφορά είναι σημαντική.

*** Ο όρος «έκβαση» στην επιδημιολογία χρησιμοποιείται για να δηλώσει άλλοτε την εμφάνιση ενός νοσήματος (health outcome) και άλλοτε την ίαση ή το θάνατο ή την εμφάνιση καταλοίπων από ένα νόσημα (illness outcome).

ποιούν κινητά τηλέφωνα σε σχέση μ' εκείνους που δεν χρησιμοποιούν. Ανάλογα με τα δεδομένα της μελέτης, πραγματοποιείται ο κατάλληλος στατιστικός έλεγχος και απορρίπτεται ή όχι η μηδενική υπόθεση. Αν απορριφθεί η μηδενική υπόθεση, τότε, σύμφωνα με τα δεδομένα αυτής της μελέτης, υπάρχει σχέση μεταξύ κινητών τηλεφώνων και καρκίνου του εγκεφάλου. Αντίθετα, αν δεν απορριφθεί η μηδενική υπόθεση, τότε δεν σημαίνει ότι δεν υπάρχει σχέση μεταξύ κινητών τηλεφώνων και καρκίνου του εγκεφάλου. Σημαίνει απλά ότι τα δεδομένα αυτής της μελέτης δεν πρόσφεραν ένδειξη για την απόρριψη της μηδενικής υπόθεσης. Είναι πολύ πιθανό μια άλλη μελέτη να καταλήξει στο ότι υπάρχει σχέση. Προφανώς, όσες περισσότερες μελέτες απορρίψουν τη συγκεκριμένη μηδενική υπόθεση, τόσο μεγαλύτερη ένδειξη υπάρχει ότι η χρήση κινητών τηλεφώνων σχετίζεται με τη συχνότητα εμφάνισης καρκίνου του εγκεφάλου. Δεν είναι όμως δυνατόν να επαληθευτεί* η αιτιακή σχέση μεταξύ κινητών τηλεφώνων και καρκίνου του εγκεφάλου. Η μη απόρριψη, εξάλλου, της μηδενικής υπόθεσης σε μια μελέτη δεν σημαίνει ότι είναι αληθής και ότι δεν υπάρχει σχέση μεταξύ προσδιοριστή και έκβασης.

3.1. Η έννοια του παρατηρούμενου επιπέδου σημαντικότητας σύμφωνα με τον Fisher

Η έννοια του παρατηρούμενου επιπέδου σημαντικότητας εισήχθη από τον Fisher και για την καλύτερη κατανόησή της θα αναφερθεί, αναλυτικά, το κλασικό παράδειγμα της ρίψης ενός νομίσματος.²⁹ Στο παράδειγμα αυτό, η μηδενική υπόθεση είναι ότι το νόμισμα είναι «αμερόληπτο»** (fair). «Αμερόληπτο» καλείται το νόμισμα εκείνο στο οποίο, σε κάθε ρίψη, η πιθανότητα εμφάνισης «κεφαλής» ισούται με την πιθανότητα εμφάνισης «γραμμάτων». Όσες ρίψεις και αν πραγματοποιηθούν είναι ανεξάρτητες μεταξύ τους και σε κάθε ρίψη η πιθανότητα εμφάνισης «κεφαλής», όπως βέβαια και η πιθανότητα εμφάνισης «γραμμάτων», είναι ίση με 50%. Για τον έλεγχο της μηδενικής υπόθεσης πραγματοποιούνται 20 ρίψεις και καταγράφονται τα αποτελέσματα. Η ανάλυση των αποτελεσμάτων, σύμφωνα με τον Fisher, πραγματοποιείται σε τέσσερα βήματα:

* Ο όρος verification αποδίδεται με τον όρο «επαλήθευση», ενώ ο όρος confirmation αποδίδεται με τον όρο «επιβεβαίωση» ή «επικύρωση». Τέλος, ο όρος support αποδίδεται με τον όρο «στήριξη».

** Αμερόληπτο νόμισμα: Το νόμισμα που είναι συμμετρικό και ομοιογενές.

*** Πείραμα τύχης: Το πείραμα που, μολοντί εκτελείται κάτω από τις ίδιες συνθήκες, δεν οδηγεί πάντοτε στο ίδιο αποτέλεσμα.

- Το πρώτο βήμα είναι η καταγραφή των πιθανών αποτελεσμάτων (δυνατών περιπτώσεων ή απλά ενδεχομένων) του πειράματος τύχης.*** Η ρίψη του νομίσματος μπορεί να οδηγήσει στην εμφάνιση «κεφαλής» ή «γραμμάτων». Πραγματοποιούνται συνολικά 20 ρίψεις. Επομένως, οι πιθανοί συνδυασμοί είναι 2.²⁰ Στο συγκεκριμένο παράδειγμα, η τυχαία μεταβλητή είναι ο συνολικός αριθμός εμφάνισης «κεφαλής» στις 20 ρίψεις. Προφανώς, η τυχαία μεταβλητή αυτή μπορεί να λάβει τιμές από 0 (καμιά εμφάνιση «κεφαλής») έως και 20 (εμφάνιση «κεφαλής» και στις 20 ρίψεις).
- Το δεύτερο βήμα είναι ο υπολογισμός της πιθανότητας κάθε αποτελέσματος της τυχαίας μεταβλητής, με την προϋπόθεση ότι η μηδενική υπόθεση είναι αληθής. Κάθε τυχαία μεταβλητή έχει μια αντίστοιχη κατανομή πιθανότητας. Μια κατανομή πιθανότητας (probability distribution) εφαρμόζει τη θεωρία των πιθανοτήτων για να περιγράψει τη συμπεριφορά μιας τυχαίας μεταβλητής. Στο συγκεκριμένο παράδειγμα, η κατανομή πιθανότητας της τυχαίας μεταβλητής προσδιορίζει όλα τα πιθανά αποτελέσματα της τυχαίας μεταβλητής, καθώς και την πιθανότητα να συμβεί το καθένα από αυτά. Η τυχαία μεταβλητή του παραδείγματος ακολουθεί τη διωνυμική κατανομή, καθώς είναι μια δικότομη τυχαία μεταβλητή, που μπορεί να λάβει μια από δύο πιθανές τιμές. Εάν με p συμβολιστεί η πιθανότητα εμφάνισης «κεφαλής» και με q η πιθανότητα εμφάνισης «γραμμάτων», τότε η πιθανότητα να εμφανιστεί «κεφαλή» x φορές σε n ρίψεις δίνεται από την εξής ισότητα:

$$P(X=x) = \binom{n}{x} p^x q^{n-x} \quad (1)$$

Εφόσον θεωρείται ότι ισχύει η μηδενική υπόθεση, τότε $p=q=1/2$, ενώ $n=20$, καθώς πραγματοποιούνται συνολικά 20 ρίψεις. Έτσι, η ισότητα (1) μετατρέπεται ως εξής:

$$P(X=x) = \binom{20}{x} \frac{1^x}{2} \left(\frac{1}{2}\right)^{20-x} \quad (2)$$

$$P(X=x) = \binom{20}{x} \left(\frac{1}{2}\right)^{20}$$

Με $\binom{n}{x}$ συμβολίζεται ο συνδυασμός n αντικειμένων επιλεγμένων x τη φορά. Αντιπροσωπεύει τον αριθμό των τρόπων με τους οποίους x αντικείμενα μπορούν να επιλεγούν από ένα σύνολο n αντικειμένων, όταν

δεν ενδιαφέρει η σειρά τους. Με βάση την ισότητα (2) μπορούν, πλέον, να υπολογιστούν οι πιθανότητες εμφάνισης οποιουδήποτε αριθμού «κεφαλής» από 0–20, με την προϋπόθεση ότι ισχύει η μηδενική υπόθεση. Οι πιθανότητες αυτές φαίνονται στον πίνακα 1 και απεικονίζονται διαγραμματικά στην εικόνα 2.

- Το τρίτο βήμα είναι να καταγραφούν όλα τα αποτελέσματα που θα μπορούσαν να συμβούν και τα οποία, με δεδομένο ότι ισχύει η μηδενική υπόθεση, έχουν μια πιθανότητα μικρότερη ή ίση από την πιθανότητα του αποτελέσματος που όντως προέκυψε. Για παράδειγμα, θεωρήστε ότι στις 20 ρίψεις εμφανίζεται «κεφαλή» 4 φορές. Η πιθανότητα να συμβεί αυτό το αποτέλεσμα, σύμφωνα με τον πίνακα 1, είναι 0,0046. Τα αποτελέσματα που έχουν πιθανότητα μικρότερη ή ίση από 0,0046 είναι εκείνα για τα οποία $x=4, 3, 2, 1, 0$ και $x=16, 17, 18, 19, 20$. Εφόσον τα ενδεχόμενα αυτά είναι αμοιβαία αποκλειόμενα, τότε η πιθανότητα να προκύψει ένα από αυτά ισούται με το άθροισμα των επιμέρους πιθανοτήτων, οπότε:

$$P^* = 2 \times (0,0046 + 0,0011 + 2 \times 10^{-4} + 1,9 \times 10^{-5} + 9 \times 10^{-7})$$

$$P^* = 0,012$$

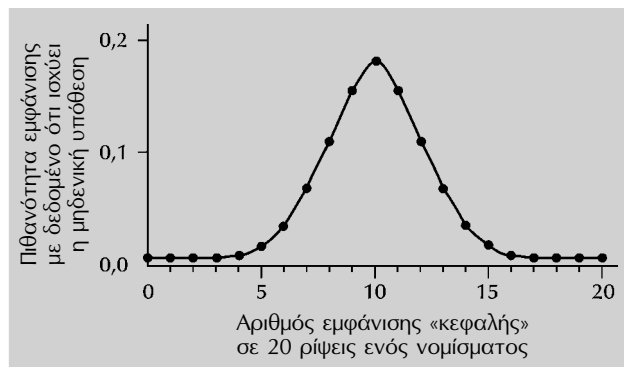
Πίνακας 1. Πιθανότητες εμφάνισης «κεφαλής» x φορές σε ένα πείραμα, το οποίο περιλαμβάνει 20 ρίψεις ενός «αμερόληπτου» νομίσματος.²⁹

Αριθμός εμφανίσεων κεφαλής (x)	Πιθανότητα εμφάνισης (P)
0	9×10^{-7}
1	$1,9 \times 10^{-5}$
2	2×10^{-4}
3	0,0011
4	0,0046
5	0,0148
6	0,0370
7	0,0739
8	0,1201
9	0,1602
10	0,1762
11	0,1602
12	0,1201
13	0,0739
14	0,0370
15	0,0148
16	0,0046
17	0,0011
18	2×10^{-4}
19	$1,9 \times 10^{-5}$
20	9×10^{-7}

- Το τέταρτο βήμα αποτελεί ουσιαστικά μια παραδοχή, σύμφωνα με την οποία η μηδενική υπόθεση θα πρέπει να απορρίπτεται όταν $P^* \leq 0,05$. Ωστόσο, ορισμένοι ερευνητές προτείνουν η μηδενική υπόθεση να απορρίπτεται όταν $P^* \leq 0,01$ ή $P^* \leq 0,001$. Η τιμή 0,05, που πρότείνει ο Fisher, είναι το προκαθορισμένο επίπεδο σημαντικότητας και συμβολίζεται με α . Εάν το πείραμα οδηγήσει σε $P^* \leq \alpha$, τότε το αποτέλεσμα καλείται σημαντικό σε επίπεδο σημαντικότητας α και η μηδενική υπόθεση απορρίπτεται.

Στο συγκεκριμένο παράδειγμα, εμφανίστηκε 4 φορές «κεφαλή» στις 20 ρίψεις, οπότε το P^* που προκύπτει είναι ίσο με 0,012. Εφόσον το P^* είναι μικρότερο από το α , απορρίπτεται η μηδενική υπόθεση στο επίπεδο σημαντικότητας 0,05. Εάν όμως εμφανιζόταν 6 φορές «κεφαλή» σε 20 ρίψεις, τότε το υπολογιζόμενο P^* θα ήταν 0,115. Σε αυτή την περίπτωση, η μηδενική υπόθεση δεν απορρίπτονταν στο επίπεδο σημαντικότητας 0,05.

Το παραπάνω παράδειγμα είναι ιδιαίτερα διευκρινιστικό όσον αφορά στην έννοια του παρατηρούμενου επιπέδου σημαντικότητας. Ωστόσο, τα πράγματα δεν είναι και τόσο απλά στην πράξη. Για παράδειγμα, θεωρήστε δύο ομάδες παιδιών, στις οποίες ο ερευνητής θέλει να ελέγξει αν υπάρχει διαφορά στο μέσο επίπεδο νοημοσύνης. Η μια ομάδα παιδιών αποτελείται μόνο από αγόρια, ενώ η άλλη ομάδα αποτελείται μόνο από κορίτσια. Το πρόβλημα στην περίπτωση αυτή είναι ότι ο ερευνητής δεν γνωρίζει τις κατανομές του επιπέδου νοημοσύνης στις δύο ομάδες παιδιών. Η μηδενική υπόθεση, εξάλλου, υποστηρίζει απλά ότι τα μέσα επίπεδα νοημοσύνης στις δύο ομάδες παιδιών είναι ίσα. Στο παράδειγμα με τη ρίψη του νομίσματος, βέβαια, ο ερευνητής γνωρίζει ότι η τυχαία μεταβλητή που μελετά ακολουθεί



Εικόνα 2. Διαγραμματική απεικόνιση των πιθανοτήτων εμφάνισης «κεφαλής» x φορές σε ένα πείραμα, το οποίο περιλαμβάνει 20 ρίψεις ενός «αμερόληπτου» νομίσματος.²⁹

τη διωνυμική κατανομή. Έτσι, δεν είναι γνωστή η κατανομή πιθανότητας του επιπέδου νοσηρότητας στις δύο ομάδες παιδιών. Το γεγονός αυτό αποτελεί ένα σημαντικό πρόβλημα, καθώς οι έλεγχοι σημαντικότητας μπορούν να πραγματοποιηθούν μόνον όταν οι κατανομές πιθανότητας των μεταβλητών είναι καθορισμένες και γνωστές, κάτι το οποίο συμβαίνει σπάνια. Ωστόσο, όταν τα μελετώμενα δείγματα είναι αρκετά μεγάλα, τότε χρησιμοποιείται ο t στατιστικός έλεγχος (t -statistic test). Στην περίπτωση αυτή, διατυπώνεται η μηδενική υπόθεση και υπολογίζεται η τιμή του t με βάση τα αποτελέσματα της μελέτης.

3.2. Παρατηρούμενο επίπεδο σημαντικότητας

Το ερώτημα που προκύπτει είναι πώς μπορεί να απορριφθεί η μηδενική υπόθεση με τη χρήση του κατάλληλου στατιστικού ελέγχου υπόθεσης. Για το σκοπό αυτό χρησιμοποιείται το παρατηρούμενο επίπεδο σημαντικότητας (τιμή p), το οποίο υπολογίζεται με τη χρήση του κατάλληλου στατιστικού μοντέλου, ανάλογα με τα δεδομένα μιας μελέτης, και το οποίο συγκρίνεται με ένα προκαθορισμένο επίπεδο σημαντικότητας (τιμή α). Το προκαθορισμένο επίπεδο σημαντικότητας καθορίζεται αυθαίρετα από τον ερευνητή και συνήθως λαμβάνει τιμές 0,05 ή 0,10.^{1,2,19,28,30}

Για λόγους ευκολίας, το παρατηρούμενο επίπεδο σημαντικότητας θα καλείται στο εξής p , ενώ το προκαθορισμένο επίπεδο σημαντικότητας θα καλείται α . Η τιμή α είναι προκαθορισμένη από τον ερευνητή και λαμβάνει μια συγκεκριμένη τιμή (συνήθως 0,05 ή 0,10) προτού ξεκινήσει η μελέτη. Η τιμή p υπολογίζεται σε σχέση με τη μηδενική υπόθεση, σύμφωνα με την οποία δεν υπάρχει σχέση μεταξύ προσδιοριστή και έκβασης.^{1,19} Η τιμή p είναι η πιθανότητα, με δεδομένο ότι η μηδενική υπόθεση είναι αληθής, τα δεδομένα μιας μελέτης να παρουσιάσουν μια σχέση –μεταξύ προσδιοριστή και έκβασης– τόσο ακραία ή πιο ακραία από αυτήν που πραγματικά λαμβάνεται.^{1,19} Για παράδειγμα, θεωρήστε μια μελέτη στην οποία διερευνάται η σχέση μεταξύ της χρήσης κινητών τηλεφώνων και της συχνότητας εμφάνισης καρκίνου του εγκεφάλου και στην οποία βρέθηκε ότι ο λόγος των οιονεί επιπτώσεων-πυκνοτήτων* είναι 2,5. Η μηδενική υπόθεση είναι ότι δεν υπάρχει σχέση μεταξύ

κινητών τηλεφώνων και καρκίνου του εγκεφάλου, οπότε ο λόγος των οιονεί επιπτώσεων-πυκνοτήτων σε μια τέτοια περίπτωση είναι ίσος με τη μονάδα. Ανάλογα με τα δεδομένα της μελέτης χρησιμοποιείται το κατάλληλο στατιστικό μοντέλο για τον έλεγχο της μηδενικής υπόθεσης, οπότε προκύπτει μια τιμή p . Αυτή η τιμή p είναι η πιθανότητα, με δεδομένο ότι ο λόγος των επιπτώσεων-πυκνοτήτων είναι ίσος με τη μονάδα, μια μελέτη να δώσει λόγο των επιπτώσεων-πυκνοτήτων ίσο ή και μεγαλύτερο από 2,5. Πρέπει να τονιστεί ότι η τιμή p δεν είναι η πιθανότητα ότι η μηδενική υπόθεση είναι αληθής. Η τιμή p υπολογίζεται με την προϋπόθεση ότι η μηδενική υπόθεση είναι αληθής.

Όπως προαναφέρθηκε, για τον έλεγχο της μηδενικής υπόθεσης συγκρίνεται η τιμή του p με την τιμή α , η οποία είναι αυθαίρετη και καθορίζεται από τον ερευνητή. Εάν η τιμή p , που προκύπτει από τα δεδομένα μιας μελέτης, είναι μικρότερη από την τιμή α , τότε απορρίπτεται η μηδενική υπόθεση, ενώ αν η τιμή p είναι μεγαλύτερη από την τιμή α , τότε δεν απορρίπτεται η μηδενική υπόθεση, αλλά και ούτε επιβεβαιώνεται.^{1,19,28} Εάν η τιμή α καθοριστεί ως 0,05, τότε οποιαδήποτε τιμή $p < 0,05$ οδηγεί στην απόρριψη της μηδενικής υπόθεσης. Θεωρήστε μια τιμή p ίση με 0,07. Η τιμή αυτή του p οδηγεί σε απόρριψη της μηδενικής υπόθεσης εάν η τιμή α είναι 0,10, ενώ οδηγεί σε μη απόρριψη της μηδενικής όταν η τιμή α είναι 0,05. Έτσι, η ίδια τιμή p μπορεί να οδηγήσει σε ακριβώς αντίθετα συμπεράσματα ανάλογα με την τιμή α , οπότε γίνεται αντιληπτό ότι η χρήση των ελέγχων της στατιστικής σημαντικότητας και των παρατηρούμενων επιπέδων σημαντικότητας (τιμές p) για την εξαγωγή συμπερασμάτων είναι εξαιρετικά επισφαλής. Ακόμη και αν η τιμή p είναι μικρότερη από την τιμή α , αυτό δεν σημαίνει ότι η μηδενική υπόθεση δεν αποτελεί την πλέον λογική εξήγηση για τα δεδομένα μιας μελέτης.

Συχνά, δεν αναφέρεται καν η ακριβής τιμή p , παρά μόνο αν η τιμή αυτή είναι μεγαλύτερη ή μικρότερη από την τιμή α .^{1,19} Δεν υπάρχει κανένας λόγος βέβαια ένα συνεχές μέτρο, όπως είναι η τιμή p , να μετατρέπεται σε δικότομο, επειδή έτσι χάνεται πολύτιμη πληροφορία. Όταν η τιμή του α καθορίζεται ως 0,05, τότε τιμές p ίσες με 0,001 ή 0,049 οδηγούν σε απόρριψη της μηδενικής υπόθεσης. Είναι όμως πολύ μεγάλη η διαφορά ανάμεσα σε μια τιμή p ίση με 0,001 και σε μια τιμή p ίση με 0,049. Για το λόγο αυτόν, πρέπει να αναφέρεται η ακριβής τιμή p και όχι απλά αν είναι μεγαλύτερη ή μικρότερη από την τιμή α .

Παρά το γεγονός ότι, ακόμη και σήμερα, πολλοί ερευνητές χρησιμοποιούν τους ελέγχους της στατιστικής ση-

* Ο λόγος των συμπληρωματικών πιθανοτήτων (odds ratio, OR) είναι θεωρητικό μέτρο και άρα δεν μπορεί να υπολογιστεί σε μια εμπειρική μελέτη. Αυτό που υπολογίζεται σε μια αιτιολογική επιδημιολογική μελέτη είναι ο λόγος των οιονεί επιπτώσεων-πυκνοτήτων (στους εκτεθειμένους και μη εκτεθειμένους), με την προϋπόθεση ότι ο επιπολασμός του μελετώμενου νοσήματος στον πληθυσμό είναι μικρός.

μαντικότητας και τις τιμές p για την εξαγωγή συμπερασμάτων,* είναι φανερό ότι οι τιμές p αποτελούν απλά ένα συνεχές μέτρο της συμβατότητας μεταξύ της μηδενικής υπόθεσης και των δεδομένων μιας μελέτης.^{1,19} Αυθαίρετο κριτήριο αυτής της συμβατότητας αποτελεί η τιμή του α , που καθορίζεται από τον ερευνητή. Έτσι, αν η τιμή α καθοριστεί ως 0,05, τότε οποιαδήποτε τιμή $p < 0,05$ υποδηλώνει μικρή συμβατότητα της μηδενικής υπόθεσης με τα δεδομένα μιας μελέτης. Όσο μικρότερη δηλαδή είναι η τιμή p , τόσο μικρότερη είναι η συμβατότητα μεταξύ της μηδενικής υπόθεσης και των δεδομένων μιας μελέτης.

Συμπερασματικά, η χρήση των τιμών του p για την εξαγωγή ασφαλών συμπερασμάτων σχετικά με τον έλεγχο των επιστημονικών υποθέσεων είναι εξαιρετικά επισημάνης. Η τιμή p μιας μελέτης δεν πρέπει να αποτελεί κριτήριο για το αν μια μηδενική υπόθεση είναι αληθής ή όχι, καθώς απαιτείται η διεξαγωγή πολλών μελετών και η σύγκριση των αποτελεσμάτων που προκύπτουν. Η τιμή p είναι απλά ένα μέτρο της συμβατότητας μεταξύ της μηδενικής υπόθεσης και των δεδομένων μιας μελέτης και χρησιμοποιείται για την απόρριψη ή όχι της μηδενικής υπόθεσης.^{31,32} Βέβαια, η μη απόρριψη της μηδενικής υπόθεσης δεν σημαίνει ότι είναι αληθής, αλλά ότι τα δεδομένα μιας μελέτης δεν προσφέρουν αρκετή απόδειξη για την απόρριψή της. Η τιμή p παρουσιάζει ένα σημαντικό μειονέκτημα, καθώς δεν εκφράζει το μέγεθος της σχέσης μεταξύ του προσδιοριστή και της έκβασης και την ακρίβεια (απουσία τυχαίου σφάλματος) με την οποία η σχέση αυτή μετράται.³² Για την καλύτερη ερμηνεία των δεδομένων μιας μελέτης είναι εξαιρετικά σημαντικό να διαχωρίζεται η πληροφορία για το μέγεθος της σχέσης και η πληροφορία για την ακρίβεια της μελέτης. Ο διαχωρισμός αυτός επιτυγχάνεται με τη διαδικασία της εκτίμησης, όπως θα αναφερθεί στη συνέχεια.

3.3. Σφάλματα τύπου I και II

Στην εφαρμογή της τεχνικής του ελέγχου των υποθέσεων υπάρχουν δύο τύποι σφάλματος. Το πρώτο καλείται σφάλμα τύπου I (type I error) και είναι επίσης γνωστό ως σφάλμα απόρριψης (rejection error) ή σφάλμα α (α error).^{1,28} Ένα σφάλμα τύπου I συμβαίνει όταν απορρίπτεται η μηδενική υπόθεση, ενώ είναι αληθής. Η πιθανότητα να διαπραχθεί ένα σφάλμα τύπου I καθορίζεται από το προκαθορισμένο επίπεδο σημαντικότητας του ελέγχου (τιμή α). Εάν πραγματοποιηθούν κατ'επα-

νάληψη έλεγκοι της υπόθεσης διατηρώντας το επίπεδο σημαντικότητας στο 0,05, θα απορριφθεί λανθασμένα η μηδενική υπόθεση, ενώ ισχύει, πέντε φορές στους εκατό ελέγχους. Σε ορισμένες περιπτώσεις επιλέγεται η πιθανότητα 0,01, οπότε, με αυτόν τον τρόπο, απορρίπτεται –λανθασμένα– η μηδενική υπόθεση, ενώ αληθεύει, μόνο μία φορά στους εκατό ελέγχους σημαντικότητας.

Ο δεύτερος τύπος σφάλματος που μπορεί να διαπραχθεί κατά τη διάρκεια ενός ελέγχου υπόθεσης είναι το σφάλμα τύπου II (type II error) ή σφάλμα αποδοχής (acceptance error) ή σφάλμα β (β error).^{1,28} Ένα σφάλμα τύπου II διαπράττεται όταν δεν απορρίπτεται η μηδενική υπόθεση, ενώ είναι λανθασμένη. Εάν π.χ. το $\beta=0,10$, τότε η πιθανότητα μη απόρριψης της μηδενικής υπόθεσης, ενώ είναι λανθασμένη, είναι 0,10 ή 10%.

Εάν το β είναι η πιθανότητα να διαπραχθεί ένα σφάλμα τύπου II, τότε $(1-\beta)$ καλείται η ισχύς (power) του ελέγχου της υπόθεσης.^{1,28} Η ισχύς είναι η πιθανότητα να απορριφθεί η μηδενική υπόθεση, ενώ είναι λανθασμένη. Με άλλα λόγια, είναι η πιθανότητα να αποφευχθεί ένα σφάλμα τύπου II. Η ισχύς μπορεί, επίσης, να θεωρηθεί ως η πιθανότητα μια συγκεκριμένη μελέτη να διακρίνει μια απόκλιση από τη μηδενική υπόθεση, δεδομένου ότι αυτή υπάρχει.

Το συνθεότερο σφάλμα είναι να θεωρείται μια θεωρία (υπόθεση) ως αληθής, ενώ δεν είναι, και σπανιότερα το αντίθετο.

Οι ερευνητές γενικά προσπαθούν να σχεδιάσουν ελέγχους σημαντικότητας που έχουν υψηλή ισχύ. Δεν αρκεί να υπάρχει μικρή πιθανότητα να απορριφθεί η μηδενική υπόθεση όταν αληθεύει. Πρέπει να υπάρχει μεγάλη πιθανότητα να απορριφθεί η μηδενική υπόθεση όταν είναι λανθασμένη. Ένας τρόπος αύξησης της ισχύος ενός ελέγχου είναι να αυξηθεί η τιμή του α . Αύξηση της τιμής του α προκαλεί αύξηση του σφάλματος τύπου I, αλλά συγχρόνως μειώνει το σφάλμα τύπου II. Αντίθετα, μείωση της τιμής του α προκαλεί μείωση του σφάλματος τύπου I και αύξηση του σφάλματος τύπου II. Για το λόγο αυτόν, ο σχεδιασμός μιας μελέτης απαιτεί ιδιαίτερη προσοχή και ιδίως όσον αφορά στον προκαθορισμό της τιμής του α .

4. ΣΤΑΤΙΣΤΙΚΗ ΕΚΤΙΜΗΣΗ

4.1. Σημειακή εκτίμηση

Δεδομένου ότι η επιστημονική έρευνα στηρίζεται σε μετρήσεις, η ανάλυση των επιδημιολογικών δεδομένων

* Μια μελέτη δεν επιτρέπει την εξαγωγή συμπερασμάτων. Παρέχει μόνο ένδειξη (evidence) εναντίον της υπόθεσης.

μπορεί να θεωρηθεί περισσότερο ως ένα πρόβλημα μέτρησης και λιγότερο ως ένα πρόβλημα λήψης απόφασης.^{1,19} Η μέτρηση, εξάλλου, ενός μεγέθους απαιτεί λεπτομερέστερη στατιστική προσέγγιση από έναν απλό έλεγχο της στατιστικής σημαντικότητας που απλά απορρίπτει ή όχι μια υπόθεση.

Σε μια επιδημιολογική μελέτη μετράται ένα μέτρο συχνότητας (π.χ. επίπτωση-πυκνότητα ή επίπτωση-ποσοστό ή επιπολασμός) ή, συννηθέστερα, ένα μέτρο αποτελέσματος (π.χ. διαφορά ή λόγος των μέτρων συχνότητας). Για να εκτιμηθεί η σχέση μεταξύ ενός προσδιοριστή και της συχνότητας μιας έκβασης υπολογίζεται ένα μέτρο αποτελέσματος. Σε μια ιδανική επιδημιολογική μελέτη, η ανάλυση των δεδομένων και η παρουσίαση των αποτελεσμάτων θα πρέπει να περιλαμβάνει τον υπολογισμό ενός επιδημιολογικού μέτρου συχνότητας ή αποτελέσματος, καθώς και την ακρίβεια με την οποία έγινε η μέτρηση αυτή. Η στατιστική εκτίμηση (statistical estimation) παρέχει και τις δύο αυτές πληροφορίες.

Για παράδειγμα, θεωρήστε μια αιτιολογική επιδημιολογική μελέτη που διεξάγεται με σκοπό να εκτιμηθεί η σχέση μεταξύ της χρήσης κινητών τηλεφώνων και της συχνότητας εμφάνισης καρκίνου του εγκεφάλου.¹⁹ Για την εκτίμηση αυτής της σχέσης απαιτείται ο υπολογισμός του λόγου των επιπτώσεων-πυκνοτήτων του καρκίνου του εγκεφάλου σε αυτούς που χρησιμοποιούν κινητά τηλέφωνα σε σχέση με εκείνους που δεν χρησιμοποιούν. Θεωρήστε ότι ο λόγος των επιπτώσεων-πυκνοτήτων σε αυτή τη μελέτη ισούται με 2,5. Η τιμή αυτή προφανώς δεν είναι η πραγματική τιμή, αλλά αποτελεί μια εκτίμηση της πραγματικής τιμής. Όταν η εκτίμηση παρουσιάζεται ως μία και μοναδική τιμή, τότε καλείται σημειακή εκτίμηση (point estimate), ενώ η μοναδική αυτή τιμή καλείται σημειακή εκτιμήτρια (point estimator). Στο συγκεκριμένο παράδειγμα, το 2,5 είναι η σημειακή εκτιμήτρια της πραγματικής τιμής του λόγου των επιπτώσεων-πυκνοτήτων, η οποία ποσοτικοποιεί τη σχέση ανάμεσα στη χρήση κινητών τηλεφώνων και τη συχνότητα εμφάνισης του καρκίνου του εγκεφάλου. Με τον τρόπο αυτόν, τα δεδομένα μιας μελέτης χρησιμοποιούνται για την εκτίμηση ενός μέτρου αποτελέσματος. Ο υπολογισμός του μέτρου αποτελέσματος σε μια μελέτη επηρεάζεται από τυχαία και συστηματικά σφάλματα, καθώς και από τους συγχυτές. Η σημειακή εκτιμήτρια, εξάλλου, ενός μέτρου αποτελέσματος είναι σχεδόν απίθανο να είναι ίδια με την πραγματική τιμή αυτού του μέτρου αποτελέσματος. Έτσι, για τον καθορισμό της ακρίβειας της σημειακής εκτίμησης χρησιμοποιούνται τα διαστήματα εμπιστοσύνης (ΔΕ) (confidence intervals, intervals estimation), τα οποία παρέχουν ένα εύρος τιμών γύρω

από τη σημειακή εκτιμήτρια. Ένα ευρύ ΔΕ υποδηλώνει μικρή ακρίβεια, ενώ, αντίθετα, ένα στενό ΔΕ υποδηλώνει μεγάλη ακρίβεια.

4.2. Διαστήματα εμπιστοσύνης

Η σημειακή εκτιμήτρια ενός μέτρου αποτελέσματος (ή ενός μέτρου συχνότητας) αποτελεί μία μόνο τιμή και δεν είναι δυνατόν να εκφράσει τη στατιστική μεταβλητότητα ή, διαφορετικά, το τυχαίο σφάλμα που υπεισέρχεται στην εκτίμηση.^{1,19} Εάν ο αριθμός των συμμετεχόντων σε μια μελέτη είναι μεγάλος, τότε η εκτίμηση ενός μέτρου αποτελέσματος είναι σχετικά ακριβής και το τυχαίο σφάλμα που υπεισέρχεται στην εκτίμηση είναι μικρό. Ωστόσο, αν ο αριθμός των συμμετεχόντων σε μια μελέτη είναι μικρός, τότε η εκτίμηση είναι λιγότερο ακριβής και το τυχαίο σφάλμα μεγαλύτερο. Έτσι, για να υποδηλωθεί το μέγεθος του τυχαίου σφάλματος σε μια εκτίμηση χρησιμοποιείται το ΔΕ. Το επιλεγόμενο ΔΕ είναι αυθαίρετο και καθορίζεται από τον ερευνητή. Το επίπεδο εμπιστοσύνης ισούται με $(1-\alpha)$, όπου α είναι το προκαθορισμένο επίπεδο σημαντικότητας, η τιμή του οποίου ορίζεται αυθαίρετα από τον ερευνητή. Ένα ΔΕ μπορεί να κυμαίνεται από 0–100%. Στην πλειοψηφία των περιπτώσεων λαμβάνεται το 90% ή 95% ΔΕ. Το ΔΕ είναι ένα εύρος τιμών γύρω από μια σημειακή εκτιμήτρια, που δείχνει το βαθμό στατιστικής ακρίβειας της εκτίμησης.* Το επίπεδο εμπιστοσύνης λαμβάνεται αυθαίρετα, αλλά για κάθε επίπεδο εμπιστοσύνης το εύρος του διαστήματος εκφράζει την ακρίβεια της μέτρησης. Ένα ευρύ ΔΕ υποδηλώνει μικρότερη ακρίβεια, ενώ ένα στενότερο ΔΕ υποδηλώνει μεγαλύτερη ακρίβεια. Το ανώτερο όριο (upper limit) και το κατώτερο όριο (lower limit) του ΔΕ είναι τα όρια εμπιστοσύνης (confidence limits).

Ένα 95% ΔΕ σημαίνει ότι, εάν επιλέγονταν τυχαία 100 δείγματα από τον πληθυσμό και χρησιμοποιούνταν για τον υπολογισμό 100 διαφορετικών ΔΕ για ένα μέτρο αποτελέσματος, τότε τα 95 από τα 100 ΔΕ θα κάλυπταν την πραγματική τιμή του μέτρου αποτελέσματος για το συγκεκριμένο πληθυσμό και τα 5 όχι. Πρέπει να τονιστεί ότι 95% ΔΕ δεν σημαίνει ότι τα 95% όρια εμπιστοσύνης που έχουν προκύψει από μια μελέτη περιέχουν την πραγματική τιμή του μέτρου αποτελέσματος με 95% πιθανότητα.^{1,28} Είναι προφανές ότι η συλλογή και η ανάλυση των δεδομένων και για τα 100 δείγματα

* Το διάστημα εμπιστοσύνης για την παράμετρο κατασκευάζεται από ένα συγκεκριμένο σύνολο δεδομένων. Επιπλέον, σχεδιάζεται κατά τέτοιο τρόπο, που να εξασφαλίζει ότι σε ένα ορισμένο ποσοστό τέτοιων εφαρμογών (ανάμεσα σε ένα μεγάλο αριθμό αυτών) θα περιέχει την πραγματική τιμή της παραμέτρου.

πρέπει να γίνει με τον ίδιο ακριβώς τρόπο.^{1,19} Έτσι, το μοναδικό στοιχείο που θα διαφέρει, σε αυτές τις 100 υποθετικές επαναλήψεις της μελέτης, είναι η στατιστική μεταβλητότητα ή, διαφορετικά, η τύχη. Σε μια τέτοια περίπτωση, βέβαια, πρέπει να ισχύουν οι εξής προϋποθέσεις: (α) η μεταβλητότητα των δεδομένων να περιγράφεται με τη χρήση του κατάλληλου στατιστικού μοντέλου και (β) να απουσιάζουν τα συστηματικά σφάλματα και οι συγχυτές. Μόνο έτσι μπορεί το ΔΕ να χρησιμοποιηθεί για να εκφράσει τη στατιστική μεταβλητότητα ή, αλλιώς, το τυχαίο σφάλμα μιας εκτίμησης. Ωστόσο, οι παραπάνω προϋποθέσεις είναι πρακτικά αδύνατο να επιτευχθούν, ακόμη και σε πολύ καλά σχεδιασμένες κλινικές δοκιμές (clinical trials). Ιδιαίτερα στην περίπτωση των μη πειραματικών επιδημιολογικών μελετών, τα ΔΕ παρέχουν απλά μια πρόχειρη εκτίμηση της στατιστικής μεταβλητότητας που υπάρχει στα δεδομένα μιας μελέτης. Θα πρέπει πάντως το ΔΕ να μη χρησιμοποιείται ως ένα απόλυτο μέτρο της στατιστικής μεταβλητότητας, δηλαδή του τυχαίου σφάλματος, αλλά ως ένα γενικό μέτρο της ποσότητας του σφάλματος στα δεδομένα μιας μελέτης.

5. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΣ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ ΣΗΜΑΝΤΙΚΟΤΗΤΑΣ

Τονίζεται εξαρχής ότι, σε καμιά περίπτωση, τα ΔΕ δεν πρέπει να χρησιμοποιούνται ως υποκατάστατα των ελέγχων της στατιστικής σημαντικότητας για την απόρριψη ή όχι της μηδενικής υπόθεσης. Πιο συγκεκριμένα, όπως προαναφέρθηκε, με τη μηδενική υπόθεση υποστηρίζεται ότι δεν υπάρχει σχέση μεταξύ προσδιοριστή και έκβασης. Για τον έλεγχο αυτής της υπόθεσης, στην περίπτωση που χρησιμοποιούνται οι ΕΣΣ επιλέγεται το κατάλληλο στατιστικό μοντέλο, ανάλογα με τα δεδομένα της μελέτης, και εξάγεται το παρατηρούμενο επίπεδο σημαντικότητας (τιμή p), το οποίο συγκρίνεται με το προκαθορισμένο επίπεδο σημαντικότητας (τιμή α), οπότε στη συνέχεια απορρίπτεται ή όχι η μηδενική υπόθεση.

Ορισμένοι ερευνητές χρησιμοποιούν, εσφαλμένα, τα ΔΕ ως υποκατάστατα των ΕΣΣ για την απόρριψη ή όχι της μηδενικής υπόθεσης. Πιο συγκεκριμένα, όταν τα ΔΕ χρησιμοποιούνται για τον έλεγχο της μηδενικής υπόθεσης υπολογίζεται ένα μέτρο αποτελέσματος, όπως π.χ. ο λόγος ή η διαφορά των μέτρων συχνότητας. Στην περίπτωση αυτή, το ΔΕ είναι ένα εύρος τιμών γύρω από το οποίο κυμαίνεται η τιμή του υπολογιζόμενου μέτρου αποτελέσματος. Αν το υπολογιζόμενο μέτρο αποτελέσματος είναι η διαφορά ενός μέτρου συχνότητας (π.χ. η διαφορά των επιπτώσεων-πυκνοτήτων σε εκτεθειμένους

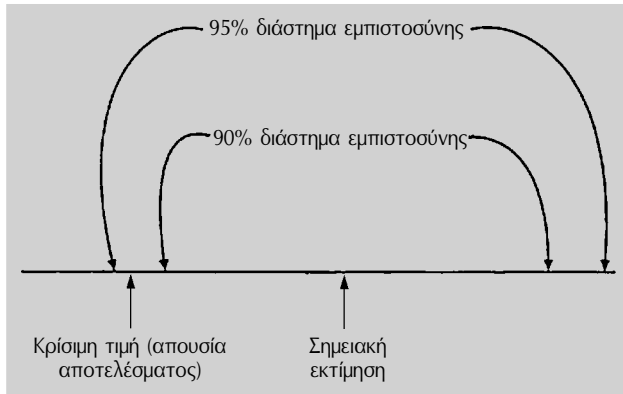
και μη εκτεθειμένους), τότε ο ερευνητής εξετάζει αν το ΔΕ περιλαμβάνει την τιμή 0. Στην περίπτωση που το ΔΕ περιλαμβάνει την τιμή 0, τότε δεν απορρίπτεται η μηδενική υπόθεση. Αντίθετα, αν το ΔΕ δεν περιλαμβάνει την τιμή 0, τότε απορρίπτεται η μηδενική υπόθεση.

Αν το υπολογιζόμενο μέτρο αποτελέσματος είναι ο λόγος ενός μέτρου συχνότητας, τότε ο ερευνητής εξετάζει αν το ΔΕ περιλαμβάνει τη μονάδα. Στην περίπτωση που το ΔΕ περιλαμβάνει τη μονάδα, τότε δεν απορρίπτεται η μηδενική υπόθεση, ενώ, αντίθετα, αν το ΔΕ δεν περιλαμβάνει τη μονάδα, τότε απορρίπτεται η μηδενική υπόθεση.

Έτσι, στην περίπτωση που το υπολογιζόμενο μέτρο αποτελέσματος είναι η διαφορά ενός μέτρου συχνότητας, η «κρίσιμη» τιμή που εξετάζει ο ερευνητής είναι το 0, ενώ αν το υπολογιζόμενο μέτρο αποτελέσματος είναι ο λόγος ενός μέτρου συχνότητας, τότε η «κρίσιμη» τιμή είναι η μονάδα. Η ερμηνεία αυτή είναι ισοδύναμη με τη διεξαγωγή ενός ελέγχου της στατιστικής σημαντικότητας της υπόθεσης ότι δεν υπάρχει σχέση. Επειδή οι έλεγχοι σημαντικότητας δίνουν λιγότερες πληροφορίες απ' ό,τι τα διαστήματα εμπιστοσύνης και συχνά παρερμηνεύονται, συστήνεται η χρήση των διαστημάτων εμπιστοσύνης στη θέση των ελέγχων σημαντικότητας για την εκτίμηση της ακρίβειας σε επιδημιολογικά δεδομένα.

Για την καλύτερη κατανόηση των παραπάνω, θεωρήστε το διάγραμμα της εικόνας 3.¹ Υποθέστε ότι διεξάγεται ένας έλεγχος της στατιστικής σημαντικότητας και ότι η τιμή α προκαθορίζεται ως 0,10. Έτσι, σύμφωνα με την εικόνα 3, το 90% ΔΕ δεν περιλαμβάνει την «κρίσιμη» τιμή, οπότε απορρίπτεται η μηδενική υπόθεση. Ωστόσο, αν ο ερευνητής προκαθόριζε την τιμή α ως 0,05, τότε το 95% ΔΕ περιλαμβάνει την «κρίσιμη» τιμή, οπότε δεν απορρίπτεται η μηδενική υπόθεση. Με βάση τα παραπάνω, γίνεται αντιληπτό ότι η μηδενική υπόθεση απορρίπτεται όταν η τιμή α είναι 0,10, ενώ δεν απορρίπτεται όταν η τιμή α είναι 0,05. Επομένως, αν διεξαγόταν ο κατάλληλος στατιστικός έλεγχος, θα οδηγούσε σε μια τιμή $p > 0,05$ και $< 0,10$. Έτσι, ανάλογα με την αυθαίρετη τιμή α , ο ερευνητής οδηγείται σε τελείως διαφορετική ερμηνεία του αποτελέσματος μιας μελέτης, γεγονός που δηλώνει την αδυναμία της χρήσης των ελέγχων της στατιστικής σημαντικότητας για σωστή ερμηνεία των αποτελεσμάτων.

Είναι προφανές ότι τα ΔΕ δεν πρέπει να χρησιμοποιούνται, όπως προαναφέρθηκε, ως υποκατάστατα των ΕΣΣ.^{1,19} Εάν συμβεί κάτι τέτοιο, τότε χάνονται τα πλεονεκτήματα που έχει η χρήση των ΔΕ έναντι των ΕΣΣ. Ο ΕΣΣ και η τιμή p που προκύπτει, αποτελούν ουσιαστικά ένα μέτρο



Εικόνα 3. 90% και 95% διαστήματα εμπιστοσύνης, με το δεύτερο να περιλαμβάνει την κρίσιμη τιμή.¹

συμβατότητας μεταξύ της μηδενικής υπόθεσης και των δεδομένων μιας μελέτης. Τα ΔΕ προσφέρουν επιπλέον μια εκτίμηση του μεγέθους της σχέσης μεταξύ ενός προσδιοριστή και μιας έκβασης, ενώ αποτελούν και μέτρο της ακρίβειας αυτής της σημειακής εκτίμησης. Δεν προκαλεί έκπληξη, εξάλλου, ότι τα ΔΕ είναι πιο χρήσιμα και προσφέρουν περισσότερες πληροφορίες, αν ληφθεί υπόψη ότι τα όρια εμπιστοσύνης είναι δύο αριθμοί, ενώ η τιμή του p είναι ένας αριθμός. Τα πράγματα γίνονται ακόμη δυσκολότερα, όταν σε πολλές μελέτες δεν αναφέρεται καθόλου η ακριβής τιμή του p , παρά μόνο αν είναι μεγαλύτερη ή μικρότερη από την τιμή του α . Στην περίπτωση αυτή, ένα συνεχές μέτρο, όπως είναι η τιμή p , μετατρέπεται σε διχότομο, οπότε χάνεται πολύτιμη πληροφορία. Όταν η τιμή α είναι 0,05, είτε το p είναι 0,001 είτε είναι 0,049 οδηγεί σε απόρριψη της μηδενικής υπόθεσης. Είναι προφανές όμως πόσο διαφορετικές είναι οι δύο αυτές τιμές p . Πάντως, ακόμη και στην περίπτωση που χρησιμοποιούνται τα ΔΕ για την ερμηνεία των αποτελεσμάτων, δεν θα πρέπει ο ερευνητής να εστιάζει την προσοχή του αυστηρά στα όρια εμπιστοσύνης, αλλά να λαμβάνει σοβαρά υπόψη του και το εύρος των ΔΕ. Σε κάθε περίπτωση, βέβαια, θα πρέπει η προσοχή των ερευνητών να εστιάζεται στα ΔΕ και όχι στις τιμές p .

Για την καλύτερη κατανόηση των παραπάνω, θεωρήστε τα δεδομένα του πίνακα 2.¹ Ο ερευνητής θέλει να ελέγξει τη μηδενική υπόθεση ότι δεν υπάρχει σχέση μεταξύ της έκθεσης (του προσδιοριστή) και της συχνότητας εμφάνισης ενός νοσήματος. Πραγματοποιώντας τον κατάλληλο στατιστικό έλεγχο προκύπτει ότι η τιμή του p ισούται με 0,14, οπότε σε επίπεδο σημαντικότητας ίσο με 0,05 δεν απορρίπτεται η μηδενική υπόθεση. Επομένως, σύμφωνα με τα αποτελέσματα αυτής της μελέ-

της, δεν υπάρχει σχέση μεταξύ έκθεσης και νοσήματος. Ωστόσο, εάν υπολογιστεί ο λόγος των επιπτώσεων-πυκνοτήτων, ο οποίος ισούται με 3,1, προκύπτει ότι η συχνότητα του νοσήματος είναι 3 φορές μεγαλύτερη στους εκτεθειμένους σε σχέση με τους μη εκτεθειμένους. Επομένως, η σημειακή εκτιμήτρια του μέτρου αποτελέσματος υποδηλώνει σχέση μεταξύ έκθεσης και συχνότητας του νοσήματος. Το 95% ΔΕ για την τιμή 3,1 του λόγου των επιπτώσεων-πυκνοτήτων είναι 0,7–13. Το γεγονός ότι το 95% ΔΕ περιλαμβάνει την «κρίσιμη» τιμή, δηλαδή τη μονάδα, σημαίνει ότι η εκτίμηση του αποτελέσματος δεν είναι στατιστικά σημαντική σε επίπεδο σημαντικότητας ίσο με $1-0,95=0,05$. Παρατηρώντας, ωστόσο, κάποιος τα όρια του ΔΕ, καταλήγει στο συμπέρασμα ότι υπάρχει μέτρια έως ισχυρή σχέση μεταξύ έκθεσης και νοσήματος, παρά το ότι ο στατιστικός έλεγχος υποδηλώνει απουσία σχέσης. Όσο πιο στενό, εξάλλου, είναι το ΔΕ, τόσο μεγαλύτερη είναι και η ακρίβεια της εκτίμησης, τόσο μικρότερο δηλαδή είναι το τυχαίο σφάλμα της μελέτης. Αντίθετα, όσο πιο ευρύ είναι το ΔΕ, τόσο μικρότερη είναι και η ακρίβεια της εκτίμησης, τόσο μεγαλύτερο δηλαδή είναι το τυχαίο σφάλμα της μελέτης.

6. Η ΣΥΝΑΡΤΗΣΗ ΤΟΥ ΠΑΡΑΤΗΡΟΥΜΕΝΟΥ ΕΠΙΠΕΔΟΥ ΣΗΜΑΝΤΙΚΟΤΗΤΑΣ

Συχνά, χρησιμοποιείται μια καμπύλη, η «συνάρτηση του παρατηρούμενου επιπέδου σημαντικότητας» (p value function) ή «συνάρτηση του διαστήματος εμπιστοσύνης» (confidence interval function), όπου στο ίδιο διάγραμμα απεικονίζονται τόσο οι τιμές p όσο και οι τιμές των ΔΕ.^{1,10,11,33-35} Με τη συνάρτηση αυτή διαπιστώνεται ο λόγος για τον οποίο πρέπει να προτιμάται η χρήση των ΔΕ έναντι των ΕΣΣ, για να εκφραστεί το μέγεθος μιας σχέσης και η ακρίβεια μιας εκτίμησης.

Όπως προαναφέρθηκε, η τιμή p είναι μια στατιστική ποσότητα που μπορεί να θεωρηθεί ως ένα μέτρο της συμβατότητας των δεδομένων μιας μελέτης και της μηδενικής υπόθεσης. Όταν το υπολογιζόμενο μέτρο αποτελέσματος είναι ο λόγος των επιπτώσεων-πυκνοτήτων, τότε η μηδενική υπόθεση, που συνήθως ελέγχεται, είναι

Πίνακας 2. Υποθετικά δεδομένα μιας αιτιολογικής μελέτης.¹

	Έκθεση	
	Ναι	Όχι
Περιπτώσεις νοσήματος	9	2
Πληθυσμοχρόνος (ανθρωπο-έτη)	186	128

ότι ο λόγος αυτός ισούται με τη μονάδα. Γι' αυτή τη μηδενική υπόθεση υπολογίζεται, με το κατάλληλο στατιστικό μοντέλο, η τιμή p . Αυτή η τιμή p αποτελεί μέτρο της συμβατότητας των δεδομένων μιας μελέτης και της μηδενικής υπόθεσης, ότι δηλαδή ο λόγος των επιπτώσεων-πυκνοτήτων ισούται με τη μονάδα. Ο λόγος των επιπτώσεων-πυκνοτήτων μπορεί, θεωρητικά, να λάβει οποιαδήποτε τιμή από το μηδέν έως το άπειρο, οπότε, αντίστοιχα, μπορούν να ελεγχθούν άπειρες υποθέσεις. Κάθε υπόθεση θα αντιστοιχεί σ' ένα διαφορετικό λόγο επιπτώσεων-πυκνοτήτων.¹

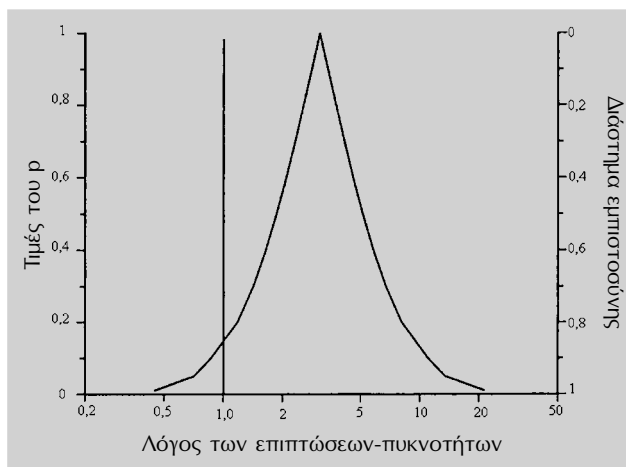
Η εικόνα 4 αναφέρεται στα δεδομένα του πίνακα 2.¹ Στην εικόνα αυτή απεικονίζονται οι τιμές p που ουσιαστικά «ελέγχουν» τη συμβατότητα μεταξύ των δεδομένων του πίνακα 2 και κάθε πιθανής τιμής του λόγου των επιπτώσεων-πυκνοτήτων. Η καμπύλη παρέχει την τιμή p , η οποία προκύπτει όταν ισχύει η μηδενική υπόθεση, όταν δηλαδή ο λόγος των επιπτώσεων-πυκνοτήτων ισούται με τη μονάδα. Αυτή η τιμή p , που προκύπτει έπειτα από την εφαρμογή του κατάλληλου στατιστικού μοντέλου, είναι 0,14. Σύμφωνα με αυτή την τιμή p , όταν η τιμή α είναι 0,05, τότε δεν απορρίπτεται η μηδενική υπόθεση, οπότε, εσφαλμένα, είναι δυνατόν να θεωρηθεί ότι δεν υπάρχει σχέση μεταξύ έκθεσης και νοσήματος. Το υψηλότερο σημείο της καμπύλης αντιστοιχεί σε σημειακή εκτιμήτρια ίση με 3,1 για το λόγο των επιπτώσεων-πυκνοτήτων. Στην εικόνα 4, εξάλλου, φαίνονται και τα ΔΕ για τιμή του λόγου των επιπτώσεων-πυκνοτήτων ίση με 3,1. Το 95% ΔΕ για σημειακή εκτιμήτρια ίση με 3,1 είναι 0,7–13. Εάν, εσφαλμένα, χρησιμοποιηθεί το ΔΕ ως υποκατάστατο του ΕΣΣ, τότε δεν απορρίπτεται η

μηδενική υπόθεση, οπότε θα μπορούσε να λεχθεί ότι δεν υπάρχει σχέση μεταξύ έκθεσης και νοσήματος. Ωστόσο, παρατηρώντας στην εικόνα 4 το εύρος των τιμών που μπορεί να λάβει ο λόγος των επιπτώσεων-πυκνοτήτων, διαπιστώνεται ότι υπάρχει σχέση μεταξύ έκθεσης και νοσήματος. Η σημειακή εκτιμήτρια είναι η τιμή εκείνη που παρουσιάζει τη μεγαλύτερη συμβατότητα με τα δεδομένα της μελέτης. Η ερμηνεία ότι δεν υπάρχει σχέση μεταξύ έκθεσης και νοσήματος είναι εσφαλμένη, απλά και μόνο αν ληφθεί υπόψη ότι η σημειακή εκτιμήτρια υποδηλώνει ότι το νόσημα είναι 3 φορές συχνότερο στους εκτεθειμένους σε σχέση με τους μη εκτεθειμένους. Σύμφωνα με την εικόνα 4, εξάλλου, οι τιμές του λόγου των επιπτώσεων-πυκνοτήτων που είναι συμβατές με τα δεδομένα της μελέτης κυμαίνονται περίπου μεταξύ 1–10, γεγονός που υποδηλώνει θετική σχέση μεταξύ έκθεσης και νοσήματος, κάτι που έρχεται σε πλήρη αντίθεση με την ερμηνεία της τιμής p . Συμπερασματικά, η χρήση της τιμής p για την εξαγωγή ασφαλών συμπερασμάτων, σχετικά με τον έλεγχο μιας υπόθεσης, είναι εξαιρετικά επισφαλής και γι' αυτό συστήνεται η χρήση των ΔΕ.

Η «συνάρτηση της τιμής p » παρέχει όλα τα πιθανά ΔΕ για μια σημειακή εκτιμήτρια που βρέθηκε σε μια μελέτη.³⁶ Η «συνάρτηση της τιμής p » μπορεί να κατασκευαστεί ακόμη και αν γνωρίζει κάποιος την ανώτερη και την κατώτερη τιμή ενός και μόνο ΔΕ. Ορισμένοι ερευνητές θεωρούν ότι η γνώση της τιμής p προσθέτει πληροφορία σε μια μελέτη και επομένως είναι απαραίτητη.³⁶ Ωστόσο, η άποψη αυτή είναι λανθασμένη, καθώς η «συνάρτηση της τιμής p » παρέχει τα δύο βασικά στοιχεία μιας εκτίμησης, κάτι το οποίο δεν ισχύει όταν χρησιμοποιείται η τιμή p .³⁶

- Το μέγεθος της σχέσης μεταξύ προσδιοριστή και έκβασης
- Την ακρίβεια της μέτρησης.

Ουσιαστικά, η «συνάρτηση της τιμής p » αποτελεί μέτρο εκτίμησης του τυχαίου σφάλματος σε μια επιδημιολογική μελέτη. Ένα οποιοδήποτε ΔΕ υποδηλώνει το μέγεθος της σχέσης μεταξύ προσδιοριστή και έκβασης, ενώ, παράλληλα, καθορίζει και το εύρος της συνάρτησης της τιμής p , οπότε αντικατοπτρίζεται με αυτόν τον τρόπο η ακρίβεια της μέτρησης. Πιο συγκεκριμένα, η «συγκέντρωση» (concentration) της καμπύλης γύρω από τη σημειακή εκτιμήτρια υποδηλώνει την ακρίβεια της εκτίμησης. Μια «στενή» συνάρτηση της τιμής p προέρχεται από μια μελέτη με μεγάλη ακρίβεια, στην οποία συμμετείχαν πολλά άτομα, ενώ μια «ευρεία» συνάρτηση της τιμής του p προέρχεται από μια μελέτη με μικρή ακρίβεια, στην οποία συμμετείχαν λίγα άτομα. Η γνώση ενός οποιοδήποτε ΔΕ επαρκεί για να κατασκευαστεί ολόκληρη η συνάρτηση της τιμής p . Η γνώση της «συ-



Εικόνα 4. Η συνάρτηση του παρατηρούμενου επιπέδου σημαντικότητας, από την οποία προκύπτουν όλα τα διαστήματα εμπιστοσύνης για μια υποθετική μελέτη, στην οποία ο λόγος των επιπτώσεων-πυκνοτήτων εκτιμήθηκε ως 3,1.¹

νάρτησης της τιμής p » είναι απαραίτητη για την εξαγωγή συμπερασμάτων, τα οποία δεν επηρεάζονται από μία και μόνο τιμή, όπως π.χ. είναι η τιμή p που αντιστοιχεί στη μηδενική υπόθεση ότι δεν υπάρχει σχέση μεταξύ προσδιοριστή και έκβασης.

Η τιμή p που προέκυψε από τα δεδομένα του πίνακα 2 υποδηλώνει ότι δεν υπάρχει σχέση μεταξύ προσδιοριστή και έκβασης. Ωστόσο, το συμπέρασμα αυτό είναι λανθασμένο, καθώς, παρατηρώντας κάποιος τη «συνάρτηση της τιμής p » (εικ. 4), διαπιστώνει ότι υπάρχει μέτρια έως ισχυρή σχέση μεταξύ προσδιοριστή και έκβασης.¹ Τα όρια εμπιστοσύνης ενός ΔΕ, όταν λαμβάνονται ως ενδεικτικά της συνάρτησης της τιμής p , απεικονίζουν το μέγεθος και την ακρίβεια μιας εκτίμησης. Αντίθετα, μια τιμή p δεν δηλώνει το μέγεθος μιας σχέσης και την ακρίβεια της εκτίμησης, παρά μόνο το βαθμό συμβατότητας μεταξύ των δεδομένων μιας μελέτης και της μηδενικής υπόθεσης.

7. ΣΥΝΟΨΗ

Τα τελευταία χρόνια ασκείται έντονη κριτική στη χρήση των ελέγχων της στατιστικής σημαντικότητας ως πρωταρχικού «εργαλείου» (κύριου μέσου) για την εκτίμηση της δειγματικής μεταβλητότητας στις επιδημιολογικές μελέτες. Παρά το γεγονός αυτό, αρκετά επιστημονικά περιοδικά, ακόμη και σήμερα, απαιτούν τη χρήση των ΕΣΣ για τη δημοσίευση των μελετών.¹ Σε μια επιδημιολογική μελέτη, συνήθως, ο ερευνητής θέλει να μελετήσει αν υπάρχει σχέση ανάμεσα σε έναν προσδιοριστή και μια έκβαση. Σε μια τέτοια περίπτωση, η μηδενική υπόθεση, συνήθως, είναι ότι δεν υπάρχει σχέση ανάμεσα στον προσδιοριστή και την έκβαση. Είναι προφανές ότι η μηδενική υπόθεση καθορίζεται από το τι θέλει να μελετήσει ο ερευνητής. Παρά το γεγονός ότι, ακόμη και σήμερα, αρκετοί χρησιμοποιούν τους ελέγχους της στα-

τιστικής σημαντικότητας και τις τιμές p για την εξαγωγή συμπερασμάτων, είναι φανερό ότι οι τιμές p αποτελούν απλά ένα συνεχές μέτρο της συμβατότητας μεταξύ της μηδενικής υπόθεσης και των δεδομένων μιας μελέτης.^{1,19} Αυθαίρετο κριτήριο αυτής της συμβατότητας αποτελεί η τιμή α , που καθορίζεται από τον ερευνητή. Έτσι, αντί των ΕΣΣ, προτείνεται η χρήση των ΔΕ.

Το ΔΕ είναι ένα εύρος τιμών γύρω από μια σημειακή εκτιμήτρια, που δείχνει το μέγεθος της σχέσης μεταξύ προσδιοριστή και έκβασης, καθώς και το βαθμό στατιστικής ακρίβειας της εκτίμησης. Το επίπεδο εμπιστοσύνης λαμβάνεται αυθαίρετα, αλλά για κάθε επίπεδο εμπιστοσύνης το εύρος του διαστήματος εκφράζει την ακρίβεια της μέτρησης. Ένα ευρύτερο ΔΕ υποδηλώνει μικρότερη ακρίβεια, ενώ ένα στενότερο ΔΕ υποδηλώνει μεγαλύτερη ακρίβεια. Το ανώτερο όριο και το κατώτερο όριο του ΔΕ είναι τα όρια εμπιστοσύνης. Ένα 95% ΔΕ σημαίνει ότι εάν επιλέγονταν τυχαία 100 δείγματα από τον πληθυσμό και χρησιμοποιούνταν αυτά τα δείγματα για να υπολογιστούν 100 διαφορετικά ΔΕ για ένα μέτρο αποτελέσματος, τότε τα 95 από τα 100 ΔΕ θα περιλάμβαναν την πραγματική τιμή του μέτρου αποτελέσματος για το συγκεκριμένο πληθυσμό και τα 5 όχι. Πρέπει να σημειωθεί ότι 95% ΔΕ δεν σημαίνει ότι τα 95% όρια εμπιστοσύνης που έχουν προκύψει από μια μελέτη περιέχουν την αληθινή τιμή του μέτρου αποτελέσματος με 95% πιθανότητα.^{1,28} Τέλος, τα ΔΕ δεν πρέπει να χρησιμοποιούνται ως υποκατάστατα των ελέγχων της στατιστικής σημαντικότητας για την απόρριψη ή όχι της μηδενικής υπόθεσης. Ο ΕΣΣ και η τιμή του p που προκύπτει, αποτελούν ουσιαστικά ένα μέτρο της συμβατότητας μεταξύ της μηδενικής υπόθεσης και των δεδομένων μιας μελέτης. Τα ΔΕ προσφέρουν επιπλέον μια εκτίμηση του μεγέθους της σχέσης μεταξύ προσδιοριστή και έκβασης, ενώ αποτελούν και μέτρο της ακρίβειας αυτής της σημειακής εκτίμησης.

ABSTRACT

Data analysis: Non Bayesian approach

P. GALANIS, L. SPAROS

Laboratory of Clinical Epidemiology, School of Nursing, University of Athens, Athens, Greece

Archives of Hellenic Medicine 2005, 22(4):377–391

An epidemiologic study can be viewed as an exercise in measurement. As in any measurement, the goal is to obtain an accurate result, with as little error as possible. There are three types of error; random error, systematic error and confounding. Random error is nothing other than variability in the data that cannot be readily explained. If an epidemiologic study is thought of as an exercise in measurement, then the result should provide, usually, an estimate of a measure of effect. Ideally, the analysis of data and the reporting of results

should quantify the magnitude of that measure of effect and portray the degree of precision with which it is measured. Statistical estimation provides these quantities. When an estimate is presented as a single value, it is referred to as a point estimate. To indicate the precision of the point estimate, a confidence interval is used, which is a range of values around the point estimate. A wide confidence interval indicates low precision and a narrow interval indicates high precision. The reason for the use of confidence intervals is that a point estimate, being a single value, cannot express the statistical variation or random error that underlies the estimate. Confidence intervals are used to indicate the amount of random error in the estimate.

Key words: Confidence interval, Point estimate, Random error, Systematic error

Βιβλιογραφία

1. ROTHMAN KJ, GREENLAND S. Approaches to statistical analysis. In: *Modern epidemiology*. 2nd ed. Lippincott Williams & Wilkins, Philadelphia, 1998:183–199
2. GOODMAN SN. P values, hypothesis tests, and likelihood: Implications for epidemiology of a neglected historical debate. *Am J Epidemiol* 1993, 137:485–496
3. WULFF HR. Confidence limits in evaluating controlled therapeutic trials. *Lancet* 1973, ii:969–970
4. SALSBERG DS. The religion of statistics as practiced in medical journals. *Am Statist* 1985, 39:220–223
5. SIMON R, WITTES RE. Methodologic guidelines for reports of clinical trials. *Cancer Treat Rep* 1985, 69:1–3
6. LANGMAN MJS. Towards estimation and confidence intervals. *Br Med J* 1986, 292:716
7. GARDNER MA, ALTMAN DG. Confidence intervals rather than P values: Estimation rather than hypothesis testing. *Br Med J* 1986, 292:746–750
8. WALKER AM. Reporting the results of epidemiologic studies. *Am J Public Health* 1986, 76:556–558
9. POCKOCK SJ, HUGHES MD, LEE RJ. Statistical problems in the reporting of clinical trials. *N Engl J Med* 1987, 317:426–432
10. POOLE C. Beyond the confidence interval. *Am J Public Health* 1987, 77:195–199
11. POOLE C. Confidence intervals exclude nothing. *Am J Public Health* 1987, 77:492–493
12. THOMPSON WD. Statistical criteria in the interpretation of epidemiologic data. *Am J Public Health* 1987, 77:191–194
13. EVANS SJW, MILLS P, DAWSON J. The end of the P-value? *Br Heart J* 1988, 60:177–180
14. ANSCOMBE FJ. The summarizing of clinical experiments by significance levels. *Stat Med* 1990, 9:703–708
15. COHEN J. The earth is round ($P < 0.05$). *Am Psychol* 1994, 47:997–1003
16. AHLBOM A, NORELL S. Εισαγωγή στη σύγχρονη επιδημιολογία (Ελληνική μετάφραση). Ιατρικές Εκδόσεις Λίτσας, Αθήνα, 1992:109–111, 127
17. COLTON T. *Statistics in medicine*. Little, Brown & Co, London, 1974
18. FLEISS JL. Confidence intervals vs significance tests: Quantitative interpretation (letter). *Am J Public Health* 1986, 76:587
19. ROTHMAN KJ. Random error and the role of statistics. In: *Epidemiology: An introduction*. Oxford University Press, New York, 2002:113–129
20. ΣΠΑΡΟΣ ΛΔ. *Μετα-επιδημιολογία ή εφαρμοσμένη ιατρική έρευνα*. Εκδόσεις ΒΗΤΑ, Αθήνα, 2001:18, 41, 71–75
21. ΓΑΛΑΝΗΣ Π, ΣΠΑΡΟΣ ΛΔ. Διαστρωματική ανάλυση δεδομένων. *Αρχ Ελλ Ιατρ* 2004, 21:378–384
22. STROM BL. *Pharmacoepidemiology*. 2nd ed. John Wiley & Sons, New York, 1994:619–624, 695
23. STRIKE PW. *Statistical methods in laboratory medicine*. Butterworth-Heinemann, Oxford, 1991:499–502
24. CLAYTON D, HILLS M. *Statistical models in epidemiology*. Oxford University Press, Oxford, 1993:261–270
25. ΚΑΗΝ ΗΑ, ΣΕΜΠΟΣ ΣΤ. *Statistical methods in epidemiology*. Oxford University Press, Oxford, 1989:105–108
26. ΓΑΛΑΝΗΣ Π. Αλληλεπίδραση παραγόντων κινδύνου στην επιδημιολογία. Μεταπτυχιακή Διπλωματική Εργασία, Αθήνα, 2003:26–29
27. ΣΠΑΡΟΣ ΛΔ, ΓΑΛΑΝΗΣ Π, ΖΑΧΟΣ Ι, ΤΣΙΛΙΔΗΣ Κ. *Επιδημιολογία*. Ι. Εκδόσεις ΒΗΤΑ, Αθήνα, 2004:72–76
28. ΡΑΓΑΝΟ Μ, ΓΑΥΝΒΡΕΑΥ Κ. *Αρχές Βιοστατιστικής* (Ελληνική μετάφραση). Εκδόσεις «ΕΛΛΗΝ», Αθήνα, 2000:233–246
29. HOWSON C, URBACH P. *Scientific reasoning: The Bayesian approach*. 2nd ed. Open Court Publ Co, Chicago, 1996:171–183
30. GOODMAN SA. A comment on replication, p-values and evidence. *Stat Med* 1992, 11:875–879
31. GOODMAN SN, ROYALL R. Evidence and scientific research. *Am J Public Health* 1988, 78:1568–1574
32. BANDT CL, BOEN JR. A prevalent misconception about sample size, statistical significance, and clinical importance. *J Periodontol* 1972, 43:181–183
33. BIRNBAUM A. A unified theory of estimation, I. *Ann Math Stat* 1961, 32:112–135
34. MIETTINEN OS. Elements of data analysis and inference. In: *Theoretical epidemiology*. John Wiley & Sons, New York, 1985:107–128
35. SULLIVAN KM, FOSTER DA. Use of the confidence interval function. *Epidemiology* 1990, 1:39–42
36. KULLDORFF M, GRAUBARD B, VELIE E. The P-value and P-value function. *Epidemiology* 1999, 10:345–347

Corresponding author:

P. Galanis, 123 Papadiamadopoulou street, GR-115 27 Athens, Greece
e-mail: pgalanis@nurs.uoa.gr