



ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ



Φιλωττία Πέτσα

Έλεγχοι ισοδυναμίας ή μη κατωτερότητας
διαγνωστικών ελέγχων

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

Ιωάννινα, 2025



UNIVERSITY OF IOANNINA
Department of Mathematics



Filottia Petsa

Tests for equivalence or non-inferiority of diagnostic
tests

Master's Thesis

Ioannina, 2025

Αφιερώνεται στην οικογένειά μου.

Η παρούσα Μεταπτυχιακή Διατριβή εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στον κλάδο Στατιστική και Επιχειρησιακή Έρευνα, που απονέμει το Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

Εγκρίθηκε την 01/07/2025 από την εξεταστική επιτροπή:

Ονοματεπώνυμο	Βαθμίδα
Κωνσταντίνος Ζωγράφος	Καθηγητής
Δημήτριος Μπάγκαβος	Αν. Καθηγητής
Απόστολος Μπατσίδης	Αν. Καθηγητής (επιβλέπων)

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ

“Δηλώνω υπεύθυνα ότι η παρούσα διατριβή εκπονήθηκε κάτω από τους διεθνείς ηθικούς και ακαδημαϊκούς κανόνες δεοντολογίας και προστασίας της πνευματικής ιδιοκτησίας. Σύμφωνα με τους κανόνες αυτούς, δεν έχω προβεί σε ιδιοποίηση ξένου επιστημονικού έργου και έχω πλήρως αναφέρει τις πηγές που χρησιμοποίησα στην εργασία αυτή.”

Φιλωττία Πέτσα

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα μεταπτυχιακή διατριβή εκπονήθηκε στο πλαίσιο του Προγράμματος Μεταπτυχιακών Σπουδών του Τομέα Πιθανοτήτων, Στατιστικής και Επιχειρησιακής Έρευνας του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, με Επιβλέποντα τον Αναπληρωτή Καθηγητή κ. Απόστολο Μπατσίδη και μέλη της Τριμελούς Εξεταστικής Επιτροπής, τον Καθηγητή κ. Κωνσταντίνο Ζωγράφο και τον Αν. Καθηγητή κ. Δημήτριο Μπάγκαβο.

Πρώτα απ'όλα, οφείλω ένα ειλικρινές ευχαριστώ στον Επιβλέποντα Καθηγητή μου, κ. Απόστολο Μπατσίδη, για την εμπιστοσύνη που μου έδειξε κατά τη διάρκεια εκπόνησης της διατριβής. Η συνεχής διαθεσιμότητα του, η επιστημονική υποστήριξη καθώς και η υπομονή του, αποτέλεσαν για εμένα πηγή έμπνευσης και ώθηση να αναζητήσω το βέλτιστο σε κάθε στάδιο της εργασίας, επιδιώκοντας όχι απλώς ορθά, αλλά και ποιοτικά αποτελέσματα. Η καθοδήγηση του δεν περιορίστηκε στην επίλυση επιστημονικών αποριών, αλλά και στην εξέλιξη μου τόσο σε προπτυχιακό όσο και σε μεταπτυχιακό επίπεδο.

Θερμές ευχαριστίες προς τον κ. Κωνσταντίνο Ζωγράφο, Καθηγητή του Τμήματος Μαθηματικών και τον κ. Δημήτριο Μπάγκαβο, Αναπληρωτή Καθηγητή του Τμήματος Μαθηματικών, για τη συμμετοχή τους στην Επιτροπή και τις εύστοχες παρατηρήσεις. Επιπλέον, θα ήθελα να ευχαριστήσω όλους τους Καθηγητές του Τομέα Πιθανοτήτων, Στατιστικής και Επιχειρησιακής Έρευνας για τις γνώσεις που μου προσέφεραν καθ'όλη τη διάρκεια των σπουδών μου.

Τέλος, ένα μεγάλο ευχαριστώ στους γονείς μου, Στέφανο και Μαρία και στις αδελφές μου, Βασιλική και Δήμητρα για τη διαρκή συμπαράσταση και την πίστη τους σε εμένα. Η κατανόηση τους στις δύσκολες στιγμές και η αγάπη τους, μου έδινε κουράγιο να συνεχίσω, ακόμη και όταν οι απαιτήσεις ήταν υψηλές.

ΠΕΡΙΛΗΨΗ

Η διάγνωση, όπως και η θεραπεία, αποτελεί κρίσιμο στοιχείο της κλινικής πρακτικής, και στατιστικές μελέτες στον χώρο της ιατρικής διεξάγονται με στόχο την ενίσχυση των διαγνωστικών μεθόδων. Διαγνωστικοί έλεγχοι (diagnostic tests) ονομάζονται διάφοροι ιατρικοί έλεγχοι που γίνονται με σκοπό τον εντοπισμό και την πρόληψη ασθενειών. Ενδεικτικά παραδείγματα διαγνωστικών ελέγχων είναι οι αιματολογικές εξετάσεις, το τεστ Παπανικολάου, η μαστογραφία, ο υπέρηχος άνω και κάτω κοιλίας κ.ά. Σκοπός της διενέργειας των διαγνωστικών ελέγχων είναι ο εντοπισμός ασθενειών που δεν μπορούν να φανούν με άλλον τρόπο ή ο εντοπισμός της νόσου σε πρώιμο στάδιο, δηλαδή πριν την εμφάνιση συμπτωμάτων, ώστε να αντιμετωπιστεί εγκαίρως και αποτελεσματικά. Κοινό χαρακτηριστικό των διαγνωστικών ελέγχων είναι ότι η εμφάνιση θετικού αποτελέσματος σημαίνει ότι είναι πιθανό το άτομο που υποβλήθηκε στον έλεγχο να έχει την υπό εξέταση νόσο και αυτό σημαίνει ότι θα πρέπει να υποβληθεί σε εξετάσεις μεγαλύτερης ακρίβειας ή ακόμη και κατευθείαν σε θεραπεία.

Από τα παραπάνω συμπεραίνουμε ότι η χρησιμότητα ενός διαγνωστικού ελέγχου προσδιορίζεται μεταξύ άλλων από την ακρίβεια με την οποία διακρίνει τα άτομα σε αυτά με ή χωρίς το υπό διερεύνηση νόσημα. Είναι φανερό ότι προϋπόθεση για την αξιολόγηση της διακριτικής ικανότητας ενός ελέγχου είναι η σίγουρη γνώση της ομάδας ασθενών, η οποία γνώση εξασφαλίζεται μέσω των αποτελεσμάτων σε μία διαγνωστική μέθοδο που θεωρείται σημείο αναφοράς (gold standard) και η οποία αποκαλύπτει την πραγματική κλινική κατάσταση του ασθενή. Χαρακτηριστικό παράδειγμα μεθόδου αναφοράς είναι η βιοψία. Στη σημερινή εποχή είναι όλο και πιο δύσκολο να αναπτυχθούν νέοι διαγνωστικοί έλεγχοι με μεγαλύτερη διακριτική ικανότητα από τη μέθοδο αναφοράς (βλέπε Syed et al. (2024) και τις εκεί αναφορές). Επιπρόσθετα, οι μέθοδοι αναφοράς τις περισσότερες φορές είτε έχουν υψηλό κόστος ή είναι χρονοβόρες και δεν είναι εύκολες στην εφαρμογή τους. Αυτό έχει οδηγήσει να εισαχθούν στη βιβλιογραφία στατιστικοί τρόποι ελέγχου ότι μία διαγνωστική μέθοδος έχει ισοδύναμη ή μη κατώτερη αποτελεσματικότητα σε σχέση με μία υπάρχουσα καθιερωμένη.

Η παρουσίαση των παραπάνω στατιστικών ελέγχων όταν τα διαθέσιμα δε-

δομένα είναι εξαρτημένα δίτιμα (ασθενής-όχι ασθενής) αποτελεί το βασικό αντικείμενο μελέτης αυτής της μεταπτυχιακής διατριβής. Στο παραπάνω πλαίσιο, η διάρθρωση του υπολοίπου της διατριβής έχει ως εξής. Στο Κεφάλαιο 1 (Εισαγωγή) παρουσιάζονται οι έννοιες της ισοδυναμίας και της μη κατωτερότητας μεταξύ δύο διαγνωστικών ελέγχων, καθώς και η πρακτική αξία τους στους διάφορους κλάδους της ιατρικής. Τα Κεφάλαια 2 έως 4 περιλαμβάνουν μια ανασκόπηση των βασικών μεθόδων που προτάθηκαν στη βιβλιογραφία για την αξιολόγηση της ισοδυναμίας ή της μη κατωτερότητας μιας νέας διαγνωστικής μεθόδου, σε σύγκριση με μια αποδεδειγμένα αποτελεσματική και ευρέως χρησιμοποιούμενη μέθοδο. Ειδικότερα, οι μέθοδοι ταξινομούνται στα τρία αυτά κεφάλαια ως προς την παράμετρο που χρησιμοποιείται για τη σύγκριση της απόδοσης των διαγνωστικών ελέγχων (διαφορά πιθανοτήτων, λόγος πιθανοτήτων κ.ο.κ.). Στο Κεφάλαιο 5 (Συγκριτικές μελέτες) παρατίθενται και αναλύονται τα αποτελέσματα των συγκριτικών μελετών, που έχουν εμφανιστεί στη βιβλιογραφία με σκοπό την αξιολόγηση και τη σύγκριση της επίδοσης των υπό εξέταση μεθόδων. Στο Κεφάλαιο 6 (Επίλογος) συνοψίζονται τα συμπεράσματα αυτής της μεταπτυχιακής διατριβής και δίνονται προτάσεις για περαιτέρω έρευνα. Τέλος, η μεταπτυχιακή διατριβή ολοκληρώνεται με τη Βιβλιογραφία.

ABSTRACT

Diagnosis of a disease is a critical element of clinical practice, and medical statistical studies are conducted to enhance diagnostic methods. Various medical tests performed for the purpose of disease detection and prevention are called diagnostic tests. Examples of diagnostic tests include blood tests, the Pap test, mammography, upper and lower abdominal ultrasound, etc. The purpose of carrying out diagnostic tests is either the identification of a disease that cannot be detected in any other way or the detection of a disease at an early stage. A positive result in a diagnostic test typically indicates a possible presence of the disease, suggesting the need for further testing or the initiation of treatment.

The usefulness of a diagnostic test is determined, among other factors, by its accuracy in distinguishing between individuals with and without the disease. It is clear that evaluating the discriminative power of a test requires reliable knowledge of the patient's disease status, which is typically established using a reference (gold standard) diagnostic method, such as a biopsy. Today, it is increasingly difficult to develop new diagnostic tests with a higher discriminative power than the reference method (see Syed et al. (2024)). In addition, reference methods are most often either expensive or time-consuming and not easy to implement. This prompted the introduction of statistical techniques of examining whether a diagnostic method has equivalent or non-inferior efficacy to an existing established method.

The presentation of the available statistical methodology when the available data are paired binary (patient- non patient) is the main objective of the study in this master's thesis. In the above context, the structure of the remainder of the thesis is as follows. Chapter 1 (Introduction) gives an initial idea of the concepts of equivalence and non-inferiority between two diagnostic tests, and their practical value in different branches of medicine. Chapters 2-4 include a review of the main methods proposed in the literature for assessing the equivalence or non-inferiority of a new diagnostic method, compared with a proven and widely used method, classified according to the parameter of

evaluation used. Chapter 5 analyses the results of the comparative studies to evaluate the performance of the methods presented in the previous chapters. Chapter 6 summarizes the conclusions of this thesis and provides suggestions for further research. Finally, the thesis concludes with the Bibliography.

ΠΕΡΙΕΧΟΜΕΝΑ

Περίληψη

Abstract

1	Εισαγωγή	3
1.1	Χαρακτηριστικά των διαγνωστικών ελέγχων	3
1.2	Δείκτες αξιολόγησης-Έλεγχος McNemar	6
1.3	Υπό μελέτη πρόβλημα	12
2	Έλεγχοι για τη διαφορά των πιθανοτήτων θετικής διά- γνωσης	19
2.1	Οι έλεγχοι των Lu and Bean (1995)	19
2.2	Οι έλεγχοι του Nam (1997)	24
2.3	Οι έλεγχοι των Liu et al. (2002)	31
2.4	Οι έλεγχοι των Chen et al. (2023)	41
2.5	Οι έλεγχοι του Tango (1998)	43
3	Έλεγχοι με βάση το πηλίκο των πιθανοτήτων ασύμβατων ζευγών	47
3.1	Οι έλεγχοι των Liu et al. (2005)	47
3.2	Οι έλεγχοι των Jin et al. (2016) και Wang et al. (2018)	52
3.3	Οι έλεγχοι των Chen and Pan (2022)	57
4	Έλεγχοι με βάση το πηλίκο των πιθανοτήτων θετικής	

διάγνωσης	59
4.1 Οι έλεγχοι των Lachenbruch and Lynch (1998)	59
4.2 Οι έλεγχοι των Nam and Blackwelder (2002)	66
4.3 Οι έλεγχοι των Tang et al. (2003)	71
5 Συγκριτικές μελέτες	77
5.1 Συγκρίσεις ελέγχων δευτέρου κεφαλαίου	77
5.2 Συγκρίσεις ελέγχων τρίτου κεφαλαίου	81
5.3 Συγκρίσεις ελέγχων τέταρτου κεφαλαίου	86
6 Επίλογος	89
Βιβλιογραφία	93

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

Στο εισαγωγικό αυτό κεφάλαιο, αφού τεκμηριωθεί η σημασία της ύπαρξης διαγνωστικών ελέγχων στην ιατρική πράξη και οριστεί η έννοια του χρυσού κανόνα (gold standard), θα παρουσιαστούν, εν συντομία, οι κυριότεροι δείκτες αξιολόγησης της διακριτικής ικανότητας ενός διαγνωστικού ελέγχου καθώς και της συμφωνίας ή ασυμφωνίας δύο διαγνωστικών ελέγχων. Τέλος, θα παρουσιαστεί το πρόβλημα που αποτελεί αντικείμενο μελέτης των υπόλοιπων κεφαλαίων της μεταπτυχιακής διατριβής και η αναγκαιότητα μελέτης του.

1.1 Χαρακτηριστικά των διαγνωστικών ελέγχων

Οι ιατρικοί διαγνωστικοί έλεγχοι είναι μία σειρά διαδικασιών και εξετάσεων που χρησιμοποιούνται για την αξιολόγηση της υγείας ενός ατόμου. Υπάρχουν πολλοί τύποι διαγνωστικών ελέγχων, καθένας από τους οποίους έχει συγκεκριμένες εφαρμογές. Ορισμένα, ενδεικτικά, παραδείγματα διαγνωστικών εξετάσεων είναι οι αιματολογικές εξετάσεις, οι απεικονιστικές μέθοδοι, οι κλινικές εξετάσεις και οι βιοψίες. Σε όσα ακολουθούν στο πλαίσιο αυτής της μεταπτυχιακής διατριβής επικεντρωνόμαστε σε εκείνους τους διαγνωστικούς ελέγχους με δύο δυνατά αποτελέσματα: θετικό και αρνητικό. Η παραπάνω υπόθεση δεν είναι καθόλου περιοριστική, αφού πολλές φορές μπορούμε να θέσουμε σε μία ποσοτική μεταβλητή ένα διαχωριστικό όριο για να διακρίνουμε τους πάσχοντες από τους υγιείς. Για παράδειγμα, για τη διάγνωση της μόλυνσης από το μυκοβακτήριο της φυματίωσης ως διαχωρίζουσα μεταβλητή λαμβάνεται το μέγεθος της διήθησης μετά από ένεση στο χόριο του δέρματος φυματίνης και ως διαχωριστικό όριο τα 10 mm. Όσον αφορά διαγνωστικούς ελέγχους με άλλους τύπους δεδομένων παραπέμπουμε σε όσα θα αναφερθούν στον επίλογο αυτής της μεταπτυχιακής διατριβής.

Από τα παραπάνω είναι σαφές ότι κοινό χαρακτηριστικό των διαγνωστικών

ελέγχων είναι ότι αν κάποιο άτομο εμφανίσει θετικό αποτέλεσμα, τότε αυτό σημαίνει ότι είναι πιθανό να έχει την υπό εξέταση ασθένεια και για αυτόν τον λόγο πρέπει το άτομο να υποβληθεί σε επιπρόσθετες και πιο μεγάλης ακρίβειας εξετάσεις ή ακόμα και κατευθείαν σε θεραπευτική αγωγή, αν ο διαγνωστικός έλεγχος είναι πολύ ακριβής, δηλαδή αν το αποτέλεσμα του θεωρείται σίγουρο. Στις μέρες μας, οι διαγνωστικοί έλεγχοι είναι βασικό εργαλείο στην καθημερινή ιατρική πρακτική, καθώς μέσω της εφαρμογής τους, οι ιατροί μπορούν να αξιολογήσουν την υγεία ενός ατόμου, να εντοπίσουν σε πρώιμο στάδιο ασθένειες, όπως ο καρκίνος, τα καρδιολογικά νοσήματα και οι ενδοκρινολογικές παθήσεις, να παρακολουθήσουν την εξέλιξη μιας πάθησης, να διαγνώσουν διάφορες ασθένειες και να εκτιμήσουν την αποτελεσματικότητα των θεραπευτικών παρεμβάσεων. Η σπουδαιότητα, λοιπόν, των διαγνωστικών ελέγχων έγκειται στο γεγονός ότι μπορούν να εντοπίσουν μια ασθένεια σε πρώιμο στάδιο, ακόμη και πριν από την εμφάνιση συμπτωμάτων. Η έγκαιρη αυτή ανίχνευση της ασθένειας έχει ως επακόλουθη συνέπεια τη σημαντική αύξηση της πιθανότητας επιτυχούς θεραπείας και ως αντίκτυπο την καλύτερη ποιότητα ζωής του ατόμου.

Η μεγάλη σπουδαιότητα των διαγνωστικών ελέγχων στην ιατρική πρακτική έχει ως επακόλουθη συνέπεια ότι αυτοί θα πρέπει να πληρούν ορισμένα κριτήρια ώστε να είναι αξιόπιστοι, να προσφέρουν ακριβείς και χρήσιμες πληροφορίες για τη διάγνωση και την παρακολούθηση της υγείας ενός ασθενούς, αλλά και να μπορούν να χρησιμοποιηθούν στην πράξη από την ιατρική κοινότητα. Ειδικότερα, ιδανικά, ένας διαγνωστικός έλεγχος θα πρέπει (βλέπε, ενδεικτικά, Altman and Bland, 1994b; Landis and Koch, 1977; Peeling et al., 2017; Bossuyt et al., 2015; Smith and Doe, 2018, και τις εκεί αναφορές)

- να είναι ακριβής, δηλαδή να μπορεί να διακρίνει σωστά μεταξύ υγιών ατόμων και εκείνων που έχουν τη συγκεκριμένη πάθηση,
- να είναι αξιόπιστος, δίνοντας τα ίδια αποτελέσματα ανεξάρτητα από το ποιο άτομο τον διεξάγει,
- να είναι εύκολος στην εφαρμογή και την ερμηνεία των αποτελεσμάτων του,
- να μην απαιτεί πολύ χρόνο για την εξαγωγή των αποτελεσμάτων,
- να είναι ασφαλής για τον/την ασθενή, χωρίς να προκαλεί βλάβες ή δυσάρεστες παρενέργειες,
- να είναι οικονομικά προσιτός, προσφέροντας μια καλή σχέση μεταξύ κόστους και αποτελεσματικότητας, ώστε να μπορεί να εφαρμοστεί ευρέως χωρίς να δημιουργεί σημαντικό οικονομικό βάρος για το σύστημα υγείας ή τον/την ασθενή, και

- να έχει τη δυνατότητα να εντοπίζει πρώιμα την ασθένεια και να παρακολουθεί την εξέλιξή της ή την ανταπόκριση στη θεραπεία.

Με την πάροδο του χρόνου και με στόχο πάντοτε την έγκαιρη διάγνωση κάποιων παθήσεων, έχουν προταθεί διαφορετικοί διαγνωστικοί έλεγχοι. Άρα, ένα πρώτο κρίσιμο σημείο είναι η συγκριτική τους αξιολόγηση. Ωστόσο, για να μπορεί να αξιολογηθεί ένας διαγνωστικός έλεγχος, απαραίτητη προϋπόθεση αποτελεί η σίγουρη γνώση της κατάστασης της υγείας του ατόμου. Η γνώση αυτή εξασφαλίζεται μέσω μιας εξέτασης αναφοράς, η οποία είναι ένας έλεγχος μέγιστης ακριβείας. Στην ιατρική, η εξέταση αυτή έχει καθιερωθεί να αναφέρεται με τον όρο “χρυσός κανόνας” ή “gold standard” και είναι ουσιαστικά εκείνη η μέθοδος ή η διαγνωστική εξέταση που θεωρείται πιο αξιόπιστη και ακριβής για την εκτίμηση ή τη διάγνωση μιας ασθένειας ή κατάστασης. Στην πράξη, όταν λέμε ότι μία μέθοδος ή μία εξέταση είναι ο “χρυσός κανόνας”, εννοούμε ότι είναι η αναγνωρισμένη καλύτερη επιλογή, συγκριτικά με άλλες μεθόδους, η οποία προσδιορίζει την αλήθεια και χρησιμοποιείται ως σημείο αναφοράς για την αξιολόγηση άλλων διαγνωστικών ελέγχων. Η ύπαρξη ενός χρυσού κανόνα βοηθά επίσης στην κατανόηση των περιορισμών και των δυνατοτήτων άλλων μεθόδων, επιτρέποντας στο ιατρικό προσωπικό να επιλέγει την πιο κατάλληλη προσέγγιση για τη διάγνωση. Για παράδειγμα, για τον καρκίνο του μαστού η βιοψία του ιστού θεωρείται ο χρυσός κανόνας για την επιβεβαίωση ύπαρξης ή όχι καρκίνου, για τον σακχαρώδη διαβήτη η εξέταση της γλυκοζυλιωμένης αιμοσφαιρίνης (HbA1c) είναι ο χρυσός κανόνας για τη διάγνωση και την παρακολούθηση της νόσου. Ωστόσο, υπάρχουν πολλές παθήσεις για τις οποίες δεν υπάρχει χρυσός κανόνας. Για παράδειγμα, για τον αλκοολισμό η διάγνωση βασίζεται σε κλινικά κριτήρια και συνεντεύξεις, αλλά δεν υπάρχει μία εξέταση που να μπορεί να θεωρηθεί ο χρυσός κανόνας για τη διάγνωσή του.

Από τα παραπάνω, είναι σαφές ότι η ύπαρξη ενός χρυσού κανόνα παρέχει ένα πλαίσιο για τη σύγκριση και την αξιολόγηση νέων ή παλιότερων διαγνωστικών ελέγχων. Για παράδειγμα, στην ανίχνευση του HIV, η δοκιμή PCR (Polymerase Chain Reaction) θεωρείται ο χρυσός κανόνας, ενώ άλλες μέθοδοι, όπως οι ορολογικές εξετάσεις (ειδικές εξετάσεις αίματος που ανιχνεύουν αντισώματα ή αντιγόνα), συγκρίνονται με αυτήν για να αξιολογηθεί η απόδοσή τους. Επιπρόσθετα, καθώς ο χρυσός κανόνας άλλοτε είναι περίπλοκος ή έχει μεγάλο κόστος ή είναι επικίνδυνος ή επίπονος (ανακαλέστε τον καρκίνο του μαστού και τον χρυσό κανόνα που είναι η βιοψία), η ύπαρξη αξιόπιστων διαγνωστικών ελέγχων είναι ύψιστης σημασίας. Η απόφαση για το αν ένας διαγνωστικός έλεγχος είναι ακριβής μπορεί να επιτευχθεί μέσω της αξιολόγησής του σε σχέση με τη σίγουρη γνώση που προσφέρει ο χρυσός κανόνας. Στην ενότητα που ακολουθεί, παρουσιάζονται

εν συντομία, οι κυριότεροι τρόποι αξιολόγησης ενός διαγνωστικού ελέγχου.

1.2 Δείκτες αξιολόγησης-Έλεγχος McNemar

Για την αξιολόγηση μιας διαγνωστικής μεθόδου, οι ασθενείς αρχικά ταξινομούνται σε δύο ομάδες ανάλογα με τα αποτελέσματα μίας διαγνωστικής μεθόδου που θεωρείται σημείο αναφοράς-χρυσός κανόνας και η οποία δείχνει την πραγματική νοσολογική κατάσταση του ατόμου. Στη συνέχεια, εφαρμόζεται στα ίδια άτομα η προς αξιολόγηση διαγνωστική μέθοδος και συγκρίνονται τα αποτελέσματα. Από τα παραπάνω γίνεται αντιληπτό ότι στην πράξη έχουμε στη διάθεσή μας ζευγαρωτά δίτιμα δεδομένα (paired binary data), καθώς το αποτέλεσμα τόσο του χρυσού κανόνα όσο και του διαγνωστικού ελέγχου είναι επιτυχία/αποτυχία, θετικό/αρνητικό αποτέλεσμα, ναι/όχι και εφαρμόζονται στις ίδιες πειραματικές μονάδες. Σε όσα ακολουθούν, η τιμή 0 αναπαριστά το αρνητικό αποτέλεσμα και η τιμή 1 το θετικό αποτέλεσμα.

Σε αυτό το πλαίσιο, έστω $X_i, i = 1, \dots, n$ οι δίτιμες τυχαίες μεταβλητές που παριστάνουν το αποτέλεσμα της διεξαγωγής του διαγνωστικού ελέγχου στο i -οστό άτομο, με $P(X_i = 1) = 1 - P(X_i = 0) = p_X$, και $Y_i, i = 1, \dots, n$ οι δίτιμες τυχαίες μεταβλητές που παριστάνουν το αποτέλεσμα της διεξαγωγής του χρυσού κανόνα στο i -οστό άτομο, με $P(Y_i = 1) = 1 - P(Y_i = 0) = p_Y$. Έστω επιπρόσθετα ότι x_{00}, x_{01}, x_{10} και x_{11} οι παρατηρούμενοι αριθμοί των αντίστοιχων ζευγών $(X = 0, Y = 0), (X = 0, Y = 1), (X = 1, Y = 0)$ και $(X = 1, Y = 1)$. Σε όσα ακολουθούν, x_{00} και x_{11} είναι το πλήθος των **συμβατών ζευγών (concordant pairs)**, ενώ x_{01} και x_{10} είναι το πλήθος των **ασύμβατων ζευγών (discordant pairs)**. Πιο συγκεκριμένα, x_{00} είναι ο αριθμός ζευγών όπου και οι δύο διαδικασίες έδωσαν αρνητικό αποτέλεσμα, ενώ x_{11} είναι ο αριθμός ζευγών όπου και οι δύο διαδικασίες έδωσαν θετικό αποτέλεσμα. Από την άλλη, με x_{01} συμβολίζουμε τον αριθμό των ζευγών όπου το αποτέλεσμα της διαγνωστικής μεθόδου είναι αρνητικό, ενώ εκείνο του χρυσού κανόνα είναι θετικό. Παρόμοια, με x_{10} συμβολίζουμε τον αριθμό των ζευγών όπου η δοκιμαστική δοκιμασία έδωσε θετικό αποτέλεσμα, ενώ ο χρυσός κανόνας αρνητικό. Τα δεδομένα αυτά καθώς και οι πιθανότητες εμφάνισής τους καταγράφονται σε έναν πίνακα που αναφέρεται στη βιβλιογραφία ως πίνακας συνάφειας (contingency table) ή ως πίνακας διπλής εισόδου με δύο γραμμές και δύο στήλες (2×2 πίνακας) ή ως τετράπτυχος πίνακας. Ο Πίνακας 1.1 (βλέπε McNemar, 1947) αναπαριστά τη γενική μορφή ενός τέτοιου πίνακα, όπου παρατίθενται τα αποτελέσματα ενός διαγνωστικού τεστ σε σχέση με τον χρυσό κανόνα.

	Χρυσός κανόνας		
Διαγν. Έλεγχος	+ (1)	- (0)	Σύνολο
+ (1)	$x_{11}(p_{11})$	$x_{10}(p_{10})$	$x_{11} + x_{10}(p_{1.})$
- (0)	$x_{01}(p_{01})$	$x_{00}(p_{00})$	$x_{01} + x_{00}(p_{0.})$
Σύνολο	$x_{11} + x_{01}(p_{.1})$	$x_{10} + x_{00}(p_{.0})$	n

Πίνακας 1.1: Τετράπτυχος πίνακας αποτελεσμάτων διαγνωστικού ελέγχου και χρυσού κανόνα, με αντίστοιχες πιθανότητες.

Ειδικότερα, στον Πίνακα 1.1, x_{11} είναι το πλήθος των αληθώς θετικών (true positives) περιπτώσεων, x_{00} το πλήθος των αληθώς αρνητικών περιπτώσεων (true negatives), ενώ x_{10} και x_{01} είναι το πλήθος των ψευδώς θετικών (false positives) και ψευδώς αρνητικών (false negative) περιπτώσεων, αντίστοιχα.

Στη βάση ενός πίνακα της μορφής του Πίνακα 1.1 έχουν εισαχθεί στη βιβλιογραφία διάφοροι δείκτες αξιολόγησης της απόδοσης της διακριτικής ικανότητας της διαγνωστικής μεθόδου, σε σχέση με τη σίγουρη γνώση που προσφέρει ο χρυσός κανόνας. Στη συνέχεια, θα δοθούν οι ορισμοί αυτών των δεικτών και η πρακτική τους σημασία, καθώς κάθε ένας από αυτούς τους δείκτες παρέχει μοναδικές πληροφορίες σχετικά με την απόδοση των διαγνωστικών ελέγχων.

Η **ευαισθησία (sensitivity)** ορίζεται να είναι το ποσοστό των πραγματικά θετικών περιπτώσεων που αναγνωρίζονται σωστά από τη νέα διαγνωστική μέθοδο. Με άλλα λόγια, η ευαισθησία ορίζεται ως το ποσοστό των ατόμων που πάσχουν από μια συγκεκριμένη ασθένεια και έχουν θετικό αποτέλεσμα στο διαγνωστικό τεστ. Επομένως, η ευαισθησία αναφέρεται στην ικανότητα ενός διαγνωστικού ελέγχου να ανιχνεύει σωστά τα άτομα που έχουν μια συγκεκριμένη πάθηση ή χαρακτηριστικό. Στην πράξη ορίζεται ως το πηλίκο των αληθώς θετικών περιπτώσεων προς το σύνολο των ατόμων που έχουν την πάθηση, δηλαδή προς το άθροισμα των αληθώς θετικών και ψευδώς αρνητικών περιπτώσεων. Τέλος, συμβολίζεται με Se και δίνεται από τη σχέση:

$$Se = \frac{x_{11}}{x_{11} + x_{01}}.$$

Η ευαισθησία είναι ιδιαίτερα σημαντική σε καταστάσεις όπου είναι κρίσιμο να ανιχνευτούν όλοι οι ασθενείς, όπως σε αναγνωρίσιμες ασθένειες ή μολύνσεις. Μια υψηλή ευαισθησία σημαίνει ότι το τεστ είναι αποτελεσματικό στο να μην παραλείπει τους ασθενείς που έχουν την πάθηση. Παρότι δεν υπάρχουν απόλυτες τιμές αναφοράς, διαγνωστικοί έλεγχοι με ευαισθησία μεγαλύτερη από 0.9 θεωρούνται εξαιρετικοί, ενώ αυτοί με τιμές από 0.7-0.9 αρκετά καλοί.

Ένα άλλο θεμελιώδες μέτρο της διαγνωστικής ακρίβειας μίας διαγνωστικής μεθόδου είναι η ειδικότητα. Η **ειδικότητα (specificity)** ορίζεται ως το ποσοστό των πραγματικά αρνητικών αποτελεσμάτων που αναγνωρίζονται σωστά από μια διαγνωστική μέθοδο. Με άλλα λόγια, η ειδικότητα ορίζεται ως το ποσοστό των ατόμων που δεν πάσχουν από μια συγκεκριμένη ασθένεια και έχουν αρνητικό αποτέλεσμα στο διαγνωστικό τεστ. Επομένως, η ειδικότητα αναφέρεται στην ικανότητα ενός διαγνωστικού ελέγχου να ανιχνεύει σωστά τα άτομα που δεν έχουν μια συγκεκριμένη πάθηση ή χαρακτηριστικό. Στην πράξη ορίζεται ως το πηλίκο των αληθώς αρνητικών περιπτώσεων προς το σύνολο των ατόμων που δεν έχουν την πάθηση (αληθώς αρνητικών και ψευδώς θετικών), συμβολίζεται με Sp και δίνεται από τη σχέση:

$$Sp = \frac{x_{00}}{x_{10} + x_{00}}.$$

Μια υψηλή ειδικότητα σημαίνει ότι η διαγνωστική μέθοδος είναι αποτελεσματική στο να αποφεύγει ψευδώς θετικά αποτελέσματα, δηλαδή περιπτώσεις όπου άτομα χωρίς την πάθηση εσφαλμένα διαγιγνώσκονται ως θετικά. Παρότι δεν υπάρχουν απόλυτες τιμές αναφοράς, διαγνωστικοί έλεγχοι με ειδικότητα μεγαλύτερη από 0.9 θεωρούνται εξαιρετικοί, ενώ αυτοί με τιμές από 0.7-0.9 αρκετά καλοί.

Ιδανικά σε έναν διαγνωστικό έλεγχο, θα ήταν χρήσιμη η ταυτόχρονη μεγιστοποίηση τόσο της ευαισθησίας όσο και της ειδικότητας. Μολονότι αυτή η απαίτηση είναι εξ'ορισμού αδύνατη, καθώς όσο αυξάνεται η ευαισθησία τόσο μειώνεται η ειδικότητα, επιλέγεται σημείο διαχωρισμού που επιφέρει ισορροπία. Η ROC (Receiver Operating Characteristic) καμπύλη απεικονίζει αυτή τη σχέση, καθώς δεν είναι τίποτε άλλο παρά το γράφημα της ευαισθησίας και του συμπληρώματος της ειδικότητας, ήτοι του ρυθμού ψευδώς θετικών αποτελεσμάτων που ορίζεται ως $1 - Sp$. Το εμβαδό της περιοχής κάτω από την καμπύλη ROC αποτελεί έναν δείκτη διαγνωστικής ακρίβειας (Area Under the Curve, AUC). Ο δείκτης αυτός λαμβάνει τιμές στο διάστημα (0,1). Μια τιμή κοντά στο 0.5 σημαίνει ότι ο έλεγχος δεν διακρίνει καλύτερα από το τυχαίο, τιμές πάνω από 0.8 θεωρούνται καλές, ενώ τιμές πάνω από 0.9 υποδηλώνουν πολύ καλή διακριτική ικανότητα (βλέπε Hanley and McNeil, 1982).

Δύο άλλες παράμετροι αξιολόγησης ενός διαγνωστικού ελέγχου είναι η Θετική Προβλεπτική Αξία και η Αρνητική Προβλεπτική Αξία, οι οποίες ποσοτικοποιούν την πιθανότητα του ελέγχου να καταλήξει σε σωστή πρόβλεψη-πρόγνωση. Ειδικότερα, η **Θετική Προβλεπτική Αξία (Positive Predictive Value)** αναφέρεται στην πιθανότητα ένας/μία ασθενής που έχει θετική ένδειξη (π.χ. θετικό αποτέλεσμα σε ένα διαγνωστικό τεστ) να έχει στην πραγματικότητα την ασθένεια που εξετάζεται. Δηλαδή, η Θετική Προβλεπτική Αξία είναι ίση με την

πιθανότητα το άτομο να έχει όντως την ασθένεια, δεδομένου ότι είναι θετικό το αποτέλεσμα του διαγνωστικού ελέγχου. Στην πράξη ορίζεται ως το πηλίκο των αληθώς θετικών περιπτώσεων προς το σύνολο των ατόμων που έχουν διαγνωσθεί από το τεστ ως πάσχοντες (αληθώς θετικών και ψευδώς θετικών), συμβολίζεται με PPV και δίνεται από τη σχέση:

$$PPV = \frac{x_{11}}{x_{11} + x_{10}}.$$

Η Θετική Προβλεπτική Αξία είναι σημαντική γιατί βοηθά τους επαγγελματίες υγείας να κατανοήσουν την αξιοπιστία ενός τεστ στην πράξη, καθώς αντικατοπτρίζει το ποσοστό των ατόμων με θετικό αποτέλεσμα στο τεστ που πραγματικά έχουν την ασθένεια. Επομένως, είναι σαφές ότι όσο υψηλότερη είναι η PPV , τόσο πιο αξιόπιστο θεωρείται το θετικό αποτέλεσμα του τεστ.

Από την άλλη, η **Αρνητική Προβλεπτική Αξία (Negative Predictive Value)** αναφέρεται στην πιθανότητα ένας ασθενής που έχει αρνητική ένδειξη (π.χ. αρνητικό αποτέλεσμα σε ένα διαγνωστικό τεστ) να μην έχει πράγματι την πάθηση ή το χαρακτηριστικό που εξετάζεται. Με άλλα λόγια, η Αρνητική Προβλεπτική Αξία ισούται με την πιθανότητα το άτομο να μην έχει πραγματικά την ασθένεια, δεδομένου ότι είναι αρνητικό το αποτέλεσμα του διαγνωστικού ελέγχου. Στην πράξη ορίζεται ως το πηλίκο των αληθώς αρνητικών περιπτώσεων προς το σύνολο των ατόμων που έχουν διαγνωσθεί από το τεστ ως μη πάσχοντες (αληθώς αρνητικών και ψευδώς αρνητικών), συμβολίζεται με NPV και δίνεται από τη σχέση:

$$NPV = \frac{x_{00}}{x_{01} + x_{00}}.$$

Επομένως, είναι κατανοητό ότι η Αρνητική Προβλεπτική Αξία αντικατοπτρίζει το ποσοστό των ατόμων με αρνητικό αποτέλεσμα στο τεστ που πραγματικά δεν έχουν την ασθένεια. Άμεσα συμπεραίνουμε ότι όσο υψηλότερη είναι η NPV , τόσο πιο αξιόπιστο θεωρείται το αρνητικό αποτέλεσμα του τεστ.

Ένας άλλος δείκτης που μπορεί να χρησιμοποιηθεί για την αξιολόγηση των αποτελεσμάτων ενός διαγνωστικού ελέγχου συγκριτικά με εκείνα του χρυσού κανόνα είναι **το ποσοστό της παρατηρούμενης συμφωνίας (Proportion of observed agreement)**, που συχνά αναφέρεται ως Rand index και συμβολίζεται με A_0 . Ο δείκτης αυτός, όπως αναφέρεται και στην εργασία της Vanbelle (2024), μετράει τη συνολική παρατηρούμενη συμφωνία μεταξύ δύο διαγνωστικών διαδικασιών και υπολογίζεται ως το ποσοστό των συμφωνιών των δύο μεθόδων στο σύνολο των παρατηρήσεων. Από τον Πίνακα 1.1 άμεσα έχουμε ότι ο αριθμός των συμφωνιών των δύο μεθόδων είναι ίσος με $x_{11} + x_{00}$, ενώ το

σύνολο των παρατηρήσεων είναι $n = x_{11} + x_{00} + x_{10} + x_{01}$. Επομένως,

$$A_0 = \frac{x_{11} + x_{00}}{n}.$$

Από τα παραπάνω προκύπτει ότι ο δείκτης εκφράζει πόσο συχνά οι δύο μέθοδοι συμφωνούν για τα αποτελέσματα. Αν το ποσοστό παρατηρούμενης συμφωνίας είναι υψηλό, αυτό σημαίνει ότι οι δύο μέθοδοι συμφωνούν συχνά.

Επισημαίνεται ότι ο δείκτης A_0 παρέχει μια απλή αξιολόγηση της απόδοσης του διαγνωστικού ελέγχου, χωρίς να λαμβάνει υπόψη την πιθανότητα τυχαίας συμφωνίας. Για τον λόγο αυτόν συχνά συμπληρώνεται με τον **δείκτη Kappa του Cohen (Cohen's Kappa)**, που θα ορισθεί στη συνέχεια. Ειδικότερα, ο συντελεστής αυτός ορίζεται σε όρους της παρατηρούμενης συμφωνίας A_0 και της αναμενόμενης τυχαίας συμφωνίας A_e από τη σχέση:

$$\kappa = \frac{A_0 - A_e}{1 - A_e}.$$

Λαμβάνοντας υπόψη ότι η παρατηρούμενη συμφωνία είναι $A_0 = \frac{x_{11} + x_{00}}{n}$ και η αναμενόμενη τυχαία συμφωνία είναι

$$A_e = \frac{(x_{11} + x_{10})(x_{11} + x_{01}) + (x_{01} + x_{00})(x_{10} + x_{00})}{n^2}$$

έχουμε ότι

$$\kappa = \frac{2(x_{11}x_{00} - x_{10}x_{01})}{(x_{11} + x_{01})(x_{01} + x_{00}) + (x_{10} + x_{00})(x_{11} + x_{10})}.$$

Τιμές αυτού του δείκτη στο διάστημα (0.6, 0.8) ερμηνεύονται ως πολύ καλή συμφωνία, ενώ τιμές του δείκτη στο διάστημα (0.8, 1) υποδηλώνουν άριστη συμφωνία.

Παρατήρηση 1.1. Οι δείκτες *Εναισθησίας*, *Ειδικότητας*, *Θετικής Προβλεπτικής Αξίας*, *Αρνητικής Προβλεπτικής Αξίας* και το *Ποσοστό Παρατηρούμενης συμφωνίας* δεν έχουν καθορισμένες τιμές αναφοράς, καθώς οι ιδανικές τιμές ενδέχεται να διαφέρουν ανάλογα με το συγκεκριμένο τεστ, την πάθηση που μελετάται, καθώς και τον πληθυσμό στον οποίο εφαρμόζεται το τεστ. Ωστόσο, υπάρχουν γενικές κατευθυντήριες γραμμές. Ειδικότερα, τιμή του δείκτη *Εναισθησίας* πάνω από 80% θεωρείται καλή, ενώ τιμή του ίδιου δείκτη πάνω από 90% θεωρείται ότι είναι εξαιρετική. Ομοίως, τιμή του δείκτη της *Ειδικότητας* πάνω από 80% είναι επιθυμητή. Από την άλλη, οι τιμές της *Θετικής* αλλά και της *Αρνητικής Προβλεπτικής Αξίας* επηρεάζονται από τη επίπτωση της νόσου.

Τέλος, τιμές του δείκτη Ποσοστό Παρατηρούμενης Συμφωνίας πάνω από 80% θεωρούνται καλές (βλέπε, ενδεικτικά, Bossuyt et al., 2015; Altman and Bland, 1994a; Fleiss et al., 2003; Griner et al., 1981).

Καθώς το Ποσοστό Παρατηρούμενης Συμφωνίας που ορίστηκε πρωτίτερα δεν εστιάζει στη συμφωνία ως προς τη διάγνωση πασχόντων ή μη, έχουν εισαχθεί στη βιβλιογραφία δείκτες για την επίτευξη αυτού του στόχου. Ειδικότερα, κάτι τέτοιο επιτυγχάνεται από τη μέση θετική και τη μέση αρνητική συμφωνία, δείκτες που παρουσιάστηκαν ανεξάρτητα στη βιβλιογραφία από τους Dice (1945) και Sørensen (1948). Οι δείκτες αυτοί επεκτείνουν την έννοια της συμφωνίας και προσφέρουν μια πιο ισορροπημένη άποψη, καθώς λαμβάνουν υπόψη τόσο τα ψευδώς θετικά όσο και τα ψευδώς αρνητικά αποτελέσματα. Ειδικότερα, η μέση θετική συμφωνία είναι το ποσοστό των περιπτώσεων όπου τόσο ο διαγνωστικός έλεγχος όσο και ο χρυσός κανόνας συμφωνούν σε θετικό αποτέλεσμα (x_{11}) ως προς τον μέσο όρο των θετικών αποτελεσμάτων και των δύο ελέγχων και υπολογίζεται ως

$$APA = \frac{x_{11}}{\frac{1}{2}(x_{11} + x_{10} + x_{11} + x_{01})} = \frac{2x_{11}}{2x_{11} + x_{10} + x_{01}}.$$

Παρόμοια, η μέση αρνητική συμφωνία είναι το ποσοστό των περιπτώσεων όπου τόσο ο διαγνωστικός έλεγχος όσο και ο χρυσός κανόνας συμφωνούν σε αρνητικό αποτέλεσμα (x_{00}) ως προς τον μέσο όρο των αρνητικών αποτελεσμάτων και των δύο ελέγχων και υπολογίζεται ως

$$ANA = \frac{x_{00}}{\frac{1}{2}(x_{00} + x_{10} + x_{00} + x_{01})} = \frac{2x_{00}}{2x_{00} + x_{10} + x_{01}}.$$

Τέλος, ένας παραδοσιακός έλεγχος που χρησιμοποιείται για την αξιολόγηση της απόδοσης ενός διαγνωστικού ελέγχου όταν μας είναι δεδομένη η νοσολογική κατάσταση του ατόμου μέσω του χρυσού κανόνα είναι ο έλεγχος που φέρει το όνομα του Αμερικανού στατιστικού Quinn Michael McNemar (1900-1986), καθώς πρωτοπαρουσιάστηκε στην εργασία του McNemar (1947). Ο **έλεγχος του McNemar**, σε αντίθεση με τα μέτρα συμφωνίας, όπως το Kappa του Cohen, τα οποία ποσοτικοποιούν τη συνολική συμφωνία μεταξύ της υπό αξιολόγησης διαγνωστικής μεθόδου και του χρυσού κανόνα, επικεντρώνεται στον προσδιορισμό του κατά πόσο οι παρατηρούμενες διαφορές μεταξύ αυτών είναι στατιστικά σημαντικές. Δηλαδή, το τεστ McNemar εξετάζει αν τα αποτελέσματα μεταξύ δύο διαγνωστικών τεστ που εφαρμόζονται στα ίδια άτομα αποκλίνουν (στατιστικά) σημαντικά μεταξύ τους. Ειδικότερα, ο έλεγχος του McNemar εστιάζει στον έλεγχο της υπόθεσης ότι οι δύο περιθώριες πιθανότητες για κάθε

έκβαση είναι ίσες. Επομένως, ανακαλώντας τον συμβολισμό του Πίνακα 1.1, ο έλεγχος που προτάθηκε από τον McNemar εστιάζει στον έλεγχο ότι $p_{11} + p_{10} = p_{11} + p_{01}$ και $p_{01} + p_{00} = p_{10} + p_{00}$. Από τις δύο παραπάνω σχέσεις άμεσα προκύπτει ότι ο έλεγχος του McNemar αφορά τον έλεγχο της μηδενικής υπόθεσης $H_0 : p_{10} = p_{01}$ έναντι της εναλλακτικής $H_a : p_{10} \neq p_{01}$. Στο πλαίσιο αυτό, η στατιστική συνάρτηση ελέγχου που προτάθηκε από τον McNemar (1947) δίνεται από τη σχέση:

$$\chi^2 = \frac{(x_{10} - x_{01})^2}{x_{10} + x_{01}}. \quad (1.1)$$

Υπό τη μηδενική υπόθεση και υποθέτοντας ότι υπάρχει ένας ικανοποιητικός αριθμός ασύμβατων ζευγών (συνήθως $x_{10} + x_{01} \geq 25$), η στατιστική συνάρτηση ελέγχου ακολουθεί ασυμπτωτικά χι-τετράγωνο κατανομή με 1 βαθμό ελευθερίας. Επομένως, σε (ασυμπτωτικό) επίπεδο σημαντικότητας α , απορρίπτεται η μηδενική υπόθεση αν $\chi^2 \geq \chi_{1,\alpha}^2$. Σε περίπτωση που ο αριθμός των ασύμβατων ζευγών είναι μικρός, ($x_{10} + x_{01} < 25$) τότε η προσέγγιση από τη χι-τετράγωνο κατανομή δεν είναι ικανοποιητική. Σε μία τέτοια περίπτωση ο ακριβής διωνυμικός έλεγχος που θα περιγραφεί στη συνέχεια μπορεί να χρησιμοποιηθεί. Πιο συγκεκριμένα, παρατηρώντας ότι ο έλεγχος του McNemar εξαρτάται μόνο από το πλήθος των ασύμβατων ζευγών, ένας εναλλακτικός έλεγχος προκύπτει δεσμεύοντας ως προς $n^* = x_{10} + x_{01}$. Τότε, καθώς υπό τη μηδενική υπόθεση το x_{10} ακολουθεί διωνυμική κατανομή με παραμέτρους n^* και 0.5, η μηδενική υπόθεση απορρίπτεται όταν η τιμή του x_{10} διαφέρει αρκετά από την αναμενόμενη τιμή του $\frac{n^*}{2} = \frac{x_{10} + x_{01}}{2}$. Σε αυτήν την περίπτωση, όταν $x_{10} \geq x_{01}$, η p -τιμή του ελέγχου προσδιορίζεται από τη σχέση:

$$p - \text{value} = 2 \sum_{i=x_{10}}^n \binom{n}{i} 0.5^n.$$

Παρατήρηση 1.2. Ο *Edwards (1948)* εισήγαγε μια τροποποίηση της στατιστικής συνάρτησης που προτάθηκε από τον McNemar ως διόρθωση συνεχείας, καθώς η κατανομή μίας διακριτής τυχαίας μεταβλητής προσεγγίζεται από τη χι-τετράγωνο κατανομή. Ειδικότερα, πρότεινε τη στατιστική συνάρτηση:

$$\chi^2 = \frac{(|x_{10} - x_{01}| - 1)^2}{x_{10} + x_{01}}.$$

1.3 Υπό μελέτη πρόβλημα

Ο έλεγχος του McNemar και οι δείκτες αξιολόγησης που παρουσιάστηκαν στην προηγούμενη ενότητα προσπαθούν να απαντήσουν στην ερώτηση αν η

ακρίβεια μίας διαγνωστικής μεθόδου διαφέρει ή όχι από αυτήν του χρυσού κανόνα. Ωστόσο, πολύ συχνά η παραπάνω ερώτηση μπορεί να μην αποτελεί το υπό διερεύνηση ερευνητικό ερώτημα, καθώς σε κάποιες περιπτώσεις οι ερευνητές/τριες θέλουν να διαπιστώσουν αν μια νέα διαγνωστική μέθοδος, η οποία είναι για παράδειγμα λιγότερο τοξική, επίπονη, επεμβατική και/ή φθηνότερη, είναι εξίσου καλή με τον χρυσό κανόνα. Παράδειγμα αποτελεί η χρήση των μη επεμβατικών διαγνωστικών τεχνικών, όπως η εξέταση αίματος για βιοδείκτες, που συγκρίνονται με βιοψίες, οι οποίες παρότι θεωρούνται χρυσός κανόνας, είναι πιο επεμβατικές και επώδυνες για τον ασθενή (Bossuyt et al. (2015)). Επιπρόσθετα, άλλες φορές οι ερευνητές/τριες επιθυμούν απλά να συγκρίνουν την απόδοση μίας νέας μεθόδου με μία καθιερωμένη, η οποία ωστόσο δεν αποτελεί τον χρυσό κανόνα και επομένως δεν προσφέρει τη σίγουρη γνώση της νοσολογικής κατάστασης του ατόμου. Για παράδειγμα, όπως αναφέρεται στο σύγγραμμα των Zhou et al. (2011), όταν συγκρίνουμε την ακρίβεια της μαστογραφίας που χρησιμοποιεί ψηφιακό σταθμό εργασίας με αυτήν της απλής ακτινογραφίας, το πραγματικό ερευνητικό ερώτημα μπορεί να είναι όχι αν οι δύο μέθοδοι διαφέρουν (η υπόθεση που έχουμε εξετάσει μέχρι τώρα), αλλά αν ο ψηφιακός σταθμός εργασίας παράγει αποτελέσματα παρόμοια με αυτά της τυπικής απλής ακτινογραφίας ή ίσως αν η ψηφιακή ακτινογραφία είναι εξίσου καλή με την τυπική απλή ακτινογραφία. Σημειώνεται ότι ο όρος ψηφιακός σταθμός εργασίας (digital workstation) αναφέρεται σε ένα ολοκληρωμένο σύστημα υπολογιστών και λογισμικού που χρησιμοποιείται στην ιατρική απεικόνιση για την επεξεργασία, ανάλυση, προβολή και αρχειοθέτηση ψηφιακών ιατρικών εικόνων. Στο ίδιο πλαίσιο, μια πιθανή ερώτηση μπορεί να είναι αν η μαγνητική αγγειογραφία (MRA) καρωτίδων, η οποία πραγματοποιείται με τη χρήση μαγνητικών πεδίων και ραδιοκυμάτων για να δημιουργήσει λεπτομερείς εικόνες των αρτηριών της καρωτίδας, μπορεί να αντικαταστήσει είτε την αξονική αγγειογραφία (CTA), στην οποία χρησιμοποιούνται ακτίνες X και τεχνολογία υπολογιστών για να δημιουργηθούν λεπτομερείς εικόνες των καρωτίδων αρτηριών ή την ενδαρτηριακή ψηφιακή αγγειογραφία (IADSA), η οποία αποτελεί επεμβατική μέθοδο και δίνει την πιο λεπτομερή εικόνα της κατάστασης των καρωτίδων (χρυσός κανόνας). Ακόμη και αν η MRA δεν έχει ευαισθησία ή ειδικότητα τόσο υψηλή όσο οι άλλες μέθοδοι, η πιθανή χαμηλότερη ακρίβεια θα πρέπει να λαμβάνεται υπόψη με βάση τους κινδύνους (νοσηρότητα και θνησιμότητα) που σχετίζονται με τις άλλες μεθόδους.

Από τα παραπάνω γίνεται αντιληπτό ότι πολλές φορές, το ενδιαφέρον επικεντρώνεται (βλέπε, μεταξύ άλλων, Schumi and Wittes, 2011):

- στον έλεγχο της υπόθεσης αν η διαφορά, έστω δ , των δύο διαγνωστικών μεθόδων ως προς μία παράμετρο απόδοσης (π.χ. πιθανότητα ταξινόμησης

ατόμου ως ασθενή, ευαισθησία, ειδικότητα κ.ο.κ.) είναι εντός ενός διαστήματος της μορφής $[-\Delta, \Delta]$, με Δ μία θετική τιμή που είναι τέτοια ώστε να αντανακλά το πότε οι δύο μέθοδοι μπορούν να θεωρηθούν ισοδύναμες, και

- στον έλεγχο της υπόθεσης αν η διαφορά δ είναι μικρότερη ή ίση από $-\Delta$, δηλαδή αν η νέα διαγνωστική μέθοδος δεν είναι χειρότερη από την καθιερωμένη μέθοδο ακόμα και αν δεν αποδεικνύεται ότι είναι καλύτερη.

Επισημαίνεται ότι οι τιμές του ορίου ισοδυναμίας Δ ποικίλλουν ανάλογα με το πλαίσιο της μελέτης και την παράμετρο απόδοσης που εξετάζεται. Συνήθως, το Δ επιλέγεται έτσι ώστε να αντιπροσωπεύει τη μέγιστη αποδεκτή διαφορά μεταξύ των δύο μεθόδων, η οποία δεν επηρεάζει ουσιαστικά την κλινική απόφαση ή τη διάγνωση. Συνηθισμένες τιμές για το Δ κυμαίνονται μεταξύ 0.05 και 0.10 όταν αφορά παραμέτρους όπως η πιθανότητα σωστής ταξινόμησης. Σε περιπτώσεις όπου απαιτείται μεγαλύτερη ακρίβεια, το όριο μπορεί να είναι ακόμη μικρότερο. Αξίζει επιπρόσθετα να αναφερθεί ότι η επιλογή του Δ πρέπει να βασίζεται σε κλινικά αποδεκτά κριτήρια και να δικαιολογείται από την πρακτική σημασία της διαφοράς μεταξύ των μεθόδων (βλέπε, μεταξύ άλλων, Wellek, 2010).

Τα παραπάνω σε επίπεδο ελέγχου στατιστικών υποθέσεων διατυπώνονται μέσω του ελέγχου των υποθέσεων:

$$H_0 : |\delta| > \Delta \text{ έναντι } H_1 : |\delta| \leq \Delta,$$

και

$$H_0 : \delta \leq -\Delta \text{ έναντι } H_1 : \delta > -\Delta,$$

αντίστοιχα. Ο πρώτος εξ αυτών των ελέγχων είναι γνωστός ως έλεγχος ισοδυναμίας (equivalence testing) και δίνει απάντηση στο ερώτημα αν είναι δυνατή η σύγκριση της διαγνωστικής ακρίβειας μεταξύ του νέου και του καθιερωμένου ελέγχου. Από την άλλη, ο δεύτερος έλεγχος, ο οποίος είναι γνωστός ως έλεγχος μη κατωτερότητας (non-inferiority testing), απαντά στο ερώτημα αν είναι δυνατή η εύρεση στατιστικών αποδεικτικών στοιχείων ώστε να υποστηρίξουμε ότι το νέο τεστ δεν είναι κατώτερο από το υπάρχον.

Η επιλογή μεταξύ του ελέγχου ισοδυναμίας και μη κατωτερότητας εξαρτάται πρωτίστως από τους στόχους της μελέτης και τα χαρακτηριστικά των διαγνωστικών μεθόδων που αξιολογούνται. Από τη μια πλευρά, ο έλεγχος ισοδυναμίας καθίσταται ιδανικός, όταν στόχος είναι να αποδειχθεί ότι μια διαγνωστική μέθοδος δεν παρουσιάζει κλινικά σημαντικές διαφορές στην απόδοση σε σύγκριση με μια

καθιερωμένη μέθοδο, ενώ από την άλλη ο έλεγχος μη κατωτερότητας επιδιώκει να αποδείξει ότι μία νέα διαγνωστική μέθοδος δεν είναι στατιστικά χειρότερη από μια ήδη υπάρχουσα, εντός ενός αποδεκτού περιθωρίου $-\Delta$. Ερμηνευτικά, αυτό επιτρέπει στη νέα μέθοδο να είναι ελαφρώς λιγότερο αποτελεσματική, αρκεί να παρουσιάζει επιπλέον πλεονεκτήματα, όπως μειωμένο κόστος, λιγότερες παρενέργειες ή ευκολία στη χρήση. Αξίζει να σημειωθεί ότι ένας έλεγχος μη κατωτερότητας μπορεί να αποτελέσει το πρώτο στάδιο της έρευνας, και εφόσον απορριφθεί ενδέχεται να προχωρήσει σε έλεγχο ισοδυναμίας.

Οι έλεγχοι ισοδυναμίας και μη κατωτερότητας έχουν τεράστια σημασία στην κλινική πράξη, καθώς διασφαλίζουν την αξιοπιστία και την αποτελεσματικότητα των νέων διαγνωστικών μεθόδων πριν από την υιοθέτησή τους. Παράλληλα, παρέχουν τη δυνατότητα αποτίμησης νέων σύγχρονων τεχνολογιών, οι οποίες πληρούν τα καθιερωμένα πρότυπα ακρίβειας που απαιτούνται για την κλινική τους χρήση. Η χρήση τέτοιων ελέγχων συνεισφέρει στη βελτίωση της ποιότητας της υγειονομικής περίθαλψης για τους ασθενείς, μέσω ταχύτερων και λιγότερο δαπανηρών διαγνωστικών διαδικασιών, χωρίς να θυσιάζεται η ακρίβεια των αποτελεσμάτων. Ειδικότερα, οι έλεγχοι ισοδυναμίας και μη κατωτερότητας εφαρμόζονται στον κλάδο της φαρμακολογίας, για να αξιολογηθεί η βιοϊσοδυναμία φαρμάκων, επιβεβαιώνοντας ότι τα γενόσημα φάρμακα προσφέρουν το ίδιο επίπεδο ασφάλειας με τα πρωτότυπα. Αντίστοιχα, στην κτηνιατρική χρησιμοποιούνται για τη σύγκριση νέων διαγνωστικών τεστ που αφορούν ζωικές λοιμώξεις, εξασφαλίζοντας ότι παρέχουν την ίδια ή μεγαλύτερη αξιοπιστία σε σχέση με τις παραδοσιακές μεθόδους, κατοχυρώνοντας την υιοθέτηση καινοτόμων λύσεων για τη βελτίωση της υγειονομικής περίθαλψης των ζώων. Για να κατανοήσουμε περισσότερο την καθοριστική σημασία των ελέγχων ισοδυναμίας ή μη κατωτερότητας, παραθέτουμε ενδεικτικά κάποια παραδείγματα εφαρμογών τους.

1. Διαγνωστικοί Έλεγχοι για Λοιμώξεις. Σε κλινικές δοκιμές για την αξιολόγηση νέων ταχέων διαγνωστικών τεστ (rapid tests) για τη λοίμωξη του COVID-19, οι έλεγχοι μη κατωτερότητας χρησιμοποιούνται για να επιβεβαιώσουν ότι η ευαισθησία των νέων τεστ δεν είναι χαμηλότερη από αυτήν των καθιερωμένων μοριακών τεστ. Αυτές οι μελέτες διασφαλίζουν ότι τα νέα τεστ μπορούν να προσφέρουν γρήγορα και αξιόπιστα αποτελέσματα (βλέπε Fang et al., 2020).
2. Τηλεϊατρική. Η έλευση προηγμένης τεχνολογίας οδήγησε στη χρήση τεστ ισοδυναμίας ή μη κατωτερότητας, για την αξιολόγηση νέων συστημάτων διάγνωσης που βασίζονται σε τεχνητή νοημοσύνη ή σε εξ' αποστάσεως διαγνώσεις, επιτρέποντας τη σύγκριση με τις συμβατικές ιατρικές μεθόδους (βλέπε McKinney et al., 2020).

3. Ογκολογικοί Δείκτες. Σε έρευνες που στοχεύουν στην αξιολόγηση νέων βιοδεικτών (biomarkers) καρκίνου, οι έλεγχοι ισοδυναμίας χρησιμοποιούνται για τη σύγκριση της διαγνωστικής ακρίβειας νέων τεχνικών απεικόνισης ή ανάλυσης αίματος με τις παραδοσιακές προσεγγίσεις (βλέπε Jack et al., 2018).
4. Κτηνιατρική. Η μελέτη που διεξάχθηκε από τους Krömker et al. (2021) αποτελεί μια εξειδικευμένη εφαρμογή ελέγχου μη κατωτερότητας στην κτηνιατρική διάγνωση, η οποία αποσκοπεί στον περιορισμό της αναποτελεσματικής χρήσης αντιβιοτικών στη θεραπεία της κλινικής μαστίτιδας σε αγελάδες με μακροχρόνιες παθήσεις μαστού. Οι ερευνητές κλήθηκαν να εξετάσουν την αποτελεσματικότητα μιας θεραπείας που βασίζεται σε μη στεροειδή αντιφλεγμονώδη φάρμακα (NSAID), όπως η κετοπροφαίνη, σε σύγκριση με την παραδοσιακή αντιβιοτική θεραπεία. Η αλόγιστη χρήση αντιβιοτικών συνδέεται στενά με την ανάπτυξη ανθεκτικών βακτηρίων και αποτελεί μείζον πρόβλημα για τη δημόσια υγεία και ασφάλεια των τροφίμων. Για παράδειγμα, η χρήση αντιβιοτικών απαιτεί περισσότερους και αυστηρότερους ελέγχους για υπολείμματα φαρμάκων στο γάλα ή στο κρέας, διαδικασία τόσο ακριβή όσο και χρονοβόρα. Για τον λόγο αυτό, οι ερευνητές κατέφυγαν σε μια μέθοδο που εστιάζει κυρίως στη διαχείριση των φλεγμονωδών αντιδράσεων έναντι μιας αντιβιοτικής που στοχεύει στην εξάλειψη παθογόνου, θέλοντας να διερευνήσουν αν η διαχείριση των συμπτωμάτων αρκεί για τη θεραπεία της μη σοβαρής μαστίτιδας. Οι Krömker et al. (2021) εφάρμοσαν τον στατιστικό έλεγχο μη κατωτερότητας για να συγκρίνουν την αποτελεσματικότητα της κετοπροφαίνης έναντι των αντιβιοτικών και συμπέραναν ότι η θεραπεία με κετοπροφαίνη δεν υπολείπεται των αντιβιοτικών στην αντιμετώπιση της ήπιας έως μέτριας μαστίτιδας, γεγονός που επιτρέπει την ασφαλή χρήση της ως εναλλακτική θεραπεία.

Στη βιβλιογραφία έχουν παρουσιαστεί διάφορες στατιστικές μεθοδολογίες για τον έλεγχο ισοδυναμίας και μη κατωτερότητας διαγνωστικών ελέγχων στη βάση δίτιμων ζευγαρωτών δεδομένων. Στόχος αυτής της μεταπτυχιακής διατριβής είναι να προσφέρει μια ανασκόπηση των στατιστικών ελέγχων που έχουν προταθεί για τον έλεγχο ισοδυναμίας ή μη κατωτερότητας ενός νέου διαγνωστικού ελέγχου, σε σύγκριση με έναν ήδη υπάρχοντα αποτελεσματικό έλεγχο, επισημαίνοντας τα πλεονεκτήματα και τους περιορισμούς ώστε να καταστεί δυνατή η κατανόηση των εργαλείων που είναι διαθέσιμα στη βιβλιογραφία. Ειδικότερα, στο υπόλοιπο τμήμα αυτής της μεταπτυχιακής διατριβής θα παρουσιασθούν οι σημαντικότερες τέτοιες μεθοδολογίες ταξινομημένες ανάλογα με το μέτρο απόδοσης

που χρησιμοποιείται στον ορισμό της παραμέτρου δ . Στο πλαίσιο αυτό, στο Κεφάλαιο 2 παρουσιάζονται οι μεθοδολογίες εκείνες στις οποίες η παράμετρος δ υπολογίζεται ως διαφορά των πιθανοτήτων θετικής διάγνωσης μεταξύ του νέου και του καθιερωμένου διαγνωστικού ελέγχου. Στη συνέχεια, στο Κεφάλαιο 3 η παράμετρος δ υπολογίζεται χρησιμοποιώντας το πηλίκο των πιθανοτήτων εμφάνισης ασύμβατων ζευγαριών. Στο Κεφάλαιο 4 θα παρουσιαστούν οι μεθοδολογίες όπου η παράμετρος δ υπολογίζεται ως το πηλίκο πιθανοτήτων θετικής διάγνωσης. Στο Κεφάλαιο 5 συνοψίζονται τα αποτελέσματα των συγκριτικών μελετών που έχουν διενεργηθεί στη βιβλιογραφία, για την αξιολόγηση και τη σύγκριση της απόδοσης των μεθόδων. Στο Κεφάλαιο 6 (Επίλογος) συνοψίζονται τα συμπεράσματα αυτής της μεταπτυχιακής διατριβής και δίνονται προτάσεις για περαιτέρω έρευνα. Τέλος, η μεταπτυχιακή διατριβή ολοκληρώνεται με τη Βιβλιογραφία.

ΚΕΦΑΛΑΙΟ 2

ΕΛΕΓΧΟΙ ΓΙΑ ΤΗ ΔΙΑΦΟΡΑ ΤΩΝ ΠΙΘΑΝΟΤΗΤΩΝ ΘΕΤΙΚΗΣ ΔΙΑΓΝΩΣΗΣ

Στο κεφάλαιο αυτό αντικείμενο μελέτης αποτελούν οι κυριότερες μεθοδολογίες ελέγχου της ισοδυναμίας και μη κατωτερότητας δύο διαγνωστικών ελέγχων όταν η παράμετρος δ εκφράζεται ως διαφορά των πιθανοτήτων θετικής διάγνωσης μεταξύ του νέου και του καθιερωμένου διαγνωστικού ελέγχου. Πιο συγκεκριμένα, σε όσα ακολουθήσουν η παράμετρος δ εκφράζει τη διαφορά της πιθανότητας θετικής διάγνωσης με τη νέα διαγνωστική μέθοδο από την αντίστοιχη με την παραδοσιακή μέθοδο. Επομένως, ανακαλώντας τον συμβολισμό που δόθηκε στον Πίνακα 1.1, είναι $\delta = p_{1.} - p_{.1}$ ή ύστερα από λίγες πράξεις $\delta = p_{10} - p_{01}$. Στο πλαίσιο αυτό, θα παρουσιαστούν οι έλεγχοι που έχουν προταθεί στις εργασίες των Lu and Bean (1995), του Nam (1997), των Liu et al. (2002), των Chen et al. (2023) και του Tango (1998). Οι παραπάνω έλεγχοι θα εφαρμοστούν, σε επίπεδο σημαντικότητας $\alpha = 0.05$, στα δεδομένα του Πίνακα 2.1 (πηγή McNeil et al. (1975)), όπου παρατίθενται τα δεδομένα μιας μελέτης 344 ασθενών που υποβλήθηκαν σε σπινθηρογράφημα ήπατος για την ανίχνευση της ηπατικής νόσου, με την πραγματική νοσολογική κατάσταση να επιβεβαιώνεται από βιοψία. Τέλος, επισημαίνεται ότι σε όσα ακολουθούν $\mathbb{E}_0$ και Var_0 συμβολίζουν την αναμενόμενη τιμή και τη διακύμανση υπό τη μηδενική υπόθεση.

2.1 Οι έλεγχοι των Lu and Bean (1995)

Στην εργασία των Lu and Bean (1995) το ενδιαφέρον επικεντρώνεται στον έλεγχο της υπόθεσης ότι μία νέα διαγνωστική μέθοδος είναι το ίδιο καλή με μία υπάρχουσα. Αυτό το ερευνητικό ερώτημα διατυπώνεται μέσω του ελέγχου της μηδενικής υπόθεσης ότι η παράμετρος $\delta = p_{1.} - p_{.1} = (p_{11} + p_{10}) - (p_{11} + p_{01})$

Διαγν. Έλεγχος	Παραδοσιακός κανόνας		
	+ (1)	- (0)	Σύνολο
+ (1)	$x_{11} = 231$	$x_{10} = 32$	$x_{11} + x_{10} = 263$
- (0)	$x_{01} = 27$	$x_{00} = 54$	$x_{01} + x_{00} = 81$
Σύνολο	$x_{11} + x_{01} = 258$	$x_{10} + x_{00} = 86$	$n = 344$

Πίνακας 2.1: Τετράπτυχος πίνακας αποτελεσμάτων νέου διαγνωστικού ελέγχου και παραδοσιακού ή/χρυσού κανόνα. Πηγή: McNeil et al. (1975).

είναι ίση με δ_0 , όπου δ_0 μία προκαθορισμένη αρνητική τιμή (π.χ. -0.05) έναντι της εναλλακτικής υπόθεσης ότι $\delta = \delta_1$, με $\delta_1 > \delta_0$. Επομένως, έχουμε τον έλεγχο της

$$H_0 : p_{10} - p_{01} = \delta_0 \quad \text{έναντι} \quad H_1 : p_{10} - p_{01} = \delta_1. \quad (2.1)$$

Από τα παραπάνω είναι ξεκάθαρο ότι οι τιμές δ_1 και δ_0 είναι γνωστές και προκαθορισμένες υπό την εναλλακτική και τη μηδενική υπόθεση, αντίστοιχα.

Για τον έλεγχο της παραπάνω υπόθεσης οι Lu and Bean (1995), παρακινούμενοι από την εργασία του McNemar (1947) προτείνουν δύο διαφορετικές στατιστικές συναρτήσεις, οι οποίες και θα παρουσιασθούν στη συνέχεια.

Αρχικά, οι Lu and Bean (1995) παρατηρούν ότι τόσο η μηδενική όσο και η εναλλακτική υπόθεση έχουν διατυπωθεί σε όρους των πιθανοτήτων p_{10} και p_{01} . Αυτό πρακτικά σημαίνει ότι όλη η πληροφορία που απαιτείται για την εφαρμογή του ελέγχου, συγκεντρώνεται στις δυαδικές παρατηρήσεις x_{10} και x_{01} και μάλιστα στη διαφορά αυτών. Επομένως, είναι λογικό ο έλεγχος της υπό εξέταση μηδενικής υπόθεσης να βασίζεται στο εμπειρικό (δειγματικό) ανάλογο της διαφοράς $\delta = p_{10} - p_{01}$, έστω $\hat{\delta}$, και στην απόκλιση του από τη δοθείσα τιμή δ_0 , δηλαδή στην ποσότητα $t = \hat{\delta} - \delta_0$. Σε αυτό το πλαίσιο, λαμβάνοντας υπόψη ότι το εμπειρικό ανάλογο της παραμέτρου δ δεν είναι τίποτε άλλο παρά η διαφορά των παρατηρούμενων ποσοστών εμφάνισης των συνδυασμών (1,0) και (0,1), αντίστοιχα, έχουμε ότι:

$$\hat{\delta} = \hat{p}_{10} - \hat{p}_{01} = \frac{x_{10}}{n} - \frac{x_{01}}{n} = \frac{x_{10} - x_{01}}{n}. \quad (2.2)$$

Σε όσα ακολουθούν θα λάβουμε υπόψη ότι η από κοινού κατανομή των $(x_{10}, x_{01}, x_{00}, x_{11})$ είναι πολυωνυμική με παραμέτρους $n, (p_{10}, p_{01}, p_{00}, p_{11})$, με τις περιθώριες κατανομές των x_{10} και x_{01} να είναι διωνυμικές με παραμέτρους (n, p_{10}) και (n, p_{01}) , αντίστοιχα. Ειδικότερα, λαμβάνοντας υπόψη ότι $\mathbb{E}(x_{10}) = np_{10}$ και

$\mathbb{E}(x_{01}) = np_{01}$, τότε, άμεσα, προκύπτει ότι:

$$\mathbb{E}(\hat{\delta}) = \frac{\mathbb{E}(x_{10})}{n} - \frac{\mathbb{E}(x_{01})}{n} = p_{10} - p_{01}, \quad (2.3)$$

και επομένως

$$\mathbb{E}(t) = \mathbb{E}(\hat{\delta} - \delta_0) = p_{10} - p_{01} - \delta_0.$$

Άρα, υπό τη μηδενική υπόθεση, είναι $\mathbb{E}_0(t) = \mathbb{E}_0(\hat{\delta} - \delta_0) = 0$. Όσον αφορά τη διακύμανση της στατιστικής συνάρτησης t έχουμε ότι

$$\begin{aligned} Var(t) &= Var\left(\frac{x_{10} - x_{01}}{n} - \delta_0\right) = Var(\hat{\delta}) = \frac{1}{n^2} Var(x_{10} - x_{01}) \\ &= \frac{1}{n^2} (Var(x_{10}) + Var(x_{01}) - 2Cov(x_{10}, x_{01})). \end{aligned}$$

Λαμβάνοντας υπόψη ότι $Var(x_{10}) = np_{10}(1 - p_{10})$, $Var(x_{01}) = np_{01}(1 - p_{01})$ και $Cov(x_{10}, x_{01}) = -np_{10}p_{01}$, με αντικατάσταση έχουμε ότι:

$$Var(t) = Var(\hat{\delta}) = \frac{p_{10} + p_{01} - (p_{10} - p_{01})^2}{n}. \quad (2.4)$$

Οπότε, καθώς, υπό τη μηδενική υπόθεση, $(p_{10} - p_{01})^2 = \delta_0^2$, έχουμε ότι:

$$Var_0(t) = \frac{p_{10} + p_{01} - \delta_0^2}{n}. \quad (2.5)$$

Παρατηρήστε ότι η διακύμανση της στατιστικής συνάρτησης t περιέχει τις άγνωστες παραμέτρους p_{10} και p_{01} . Άρα για να χρησιμοποιηθεί αυτός ο έλεγχος στην πράξη θα πρέπει να αντικατασταθούν αυτές οι άγνωστες παράμετροι από τις εκτιμήτριές τους. Ειδικότερα, οι Lu and Bean (1995) προτείνουν την αντικατάστασή τους από τις εκτιμήτριες $\hat{p}_{10} = \frac{x_{10}}{n}$ και $\hat{p}_{01} = \frac{x_{01}}{n}$, αντίστοιχα. Σημειώνουμε ότι αυτές είναι οι εκτιμήτριες οι οποίες προκύπτουν χωρίς να λαμβάνεται υπόψη ο περιορισμός που επιβάλλει η μηδενική υπόθεση. Συγκεντρωτικά, οι Lu and Bean (1995) προτείνουν τη στατιστική συνάρτηση:

$$Z_{LB}^{(1)} = \frac{\frac{x_{10} - x_{01}}{n} - \delta_0}{\sqrt{\frac{x_{10} + x_{01} - \delta_0^2}{n}}} = \frac{x_{10} - x_{01} - n\delta_0}{\sqrt{x_{10} + x_{01} - n\delta_0^2}}, \quad (2.6)$$

η οποία ακολουθεί ασυμπτωτικά, υπό τη μηδενική υπόθεση, την τυπική κανονική κατανομή. Με βάση τη στατιστική συνάρτηση $Z_{LB}^{(1)}$ της σχέσης (2.6) απορρίπτεται, σε ασυμπτωτικό επίπεδο σημαντικότητας α , η μηδενική υπόθεση

$H_0 : p_{10} - p_{01} = \delta_0$ έναντι της εναλλακτικής $H_0 : p_{10} - p_{01} = \delta_1$, με $\delta_1 > \delta_0$, αν ικανοποιείται η κρίσιμη περιοχή $Z_{LB}^{(1)} \geq z_\alpha$, όπου με z_α συμβολίζουμε εκείνο το σημείο για το οποίο $P(Z \geq z_\alpha) = \alpha$, με $Z \sim N(0, 1)$. Σε μία τέτοια περίπτωση, συμπεραίνουμε ότι η νέα διαδικασία δεν είναι (τουλάχιστον) χειρότερη από τη διαδικασία αναφοράς.

Παρατήρηση 2.1. Για $\delta_0 = 0$ η στατιστική συνάρτηση $Z_{LB}^{(1)}$ της σχέσης (2.6) ανάγεται ουσιαστικά στον έλεγχο McNemar.

Για την υλοποίηση αυτού του ελέγχου δημιουργήσαμε στη γλώσσα προγραμματισμού R μία συνάρτηση, η οποία λαμβάνει ως ορίσματα το πλήθος των ασύμμετρων ζευγών, το συνολικό μέγεθος του δείγματος, την τιμή δ_0 και το επίπεδο σημαντικότητας α του ελέγχου. Εφαρμόζοντας αυτήν τη συνάρτηση λαμβάνουμε ως αποτελέσματα την τιμή της στατιστικής συνάρτησης, την (ασυμπτωτική) p -τιμή του ελέγχου και μήνυμα αν απορρίπτεται ή όχι η μηδενική υπόθεση στο καθορισμένο επίπεδο σημαντικότητας. Στο πλαίσιο που ακολουθεί, παρατίθεται η συνάρτηση αυτή και η υλοποίησή της για τα δεδομένα του Πίνακα 2.1.

```
zequivalenceLB1 <- function(x10, x01, n, delta0, alpha)
{
  Z_LB_1 <- (x10 - x01 - n * delta0) / sqrt(x10 + x01 - n
    ↪ * delta0^2)

  pvalueZ_LB_1 <- pnorm(Z_LB_1, lower.tail=FALSE)

  reject_H0 <- Z_LB_1 >= qnorm(1 - alpha)
  return(list(Z_value = Z_LB_1, p_value = pvalueZ_LB_1,
    ↪ reject_H0 = reject_H0))
}
result <- zequivalenceLB1(32, 27, 344, -0.05, 0.05)
print(result)
$Z_value
[1] 2.911491

$p_value
[1] 0.001798541

$reject_H0
[1] TRUE
```

Η δεύτερη προσέγγιση που προτάθηκε από τους Lu and Bean (1995) βασίζεται στη δεσμευμένη προσέγγιση του McNemar, θεωρώντας τη δεσμευμένη κατανομή

του $X = x_{10} - x_{01}$ δοθέντος ότι είναι γνωστό το συνολικό πλήθος των μη συμβατών ζευγών, δηλαδή όταν είναι γνωστό το άθροισμα $x_{10} + x_{01}$. Για διευκόλυνση στην παρουσίαση όσων ακολουθούν συμβολίζουμε με n_1 το συνολικό πλήθος των ασύμβατων ζευγών, δηλαδή $n_1 = x_{10} + x_{01}$. Τότε, κανονικοποιώντας τις πιθανότητες επιτυχίας της διωνυμικής κατανομής, ώστε να αφορούν το ποσοστό των ασύμβατων ζευγών, ισχύει ότι $x_{10} \sim \text{Bin}\left(n_1, \frac{p_{10}}{p_{01} + p_{10}}\right)$ και επομένως δοθέντος ότι $x_{10} + x_{01} = n_1$, έχουμε ότι $X = x_{10} - x_{01} = 2x_{10} - n_1$ και ισχύει ότι:

$$\mathbb{E}(X|x_{10} + x_{01}) = \mathbb{E}(2x_{10} - n_1) = 2n_1 \frac{p_{10}}{p_{01} + p_{10}} - n_1 = n_1 \frac{p_{10} - p_{01}}{p_{01} + p_{10}},$$

ενώ

$$\text{Var}(X|x_{10} + x_{01}) = \text{Var}(2x_{10} - n_1) = 4\text{Var}(x_{10}) = 4n_1 \frac{p_{10}}{p_{01} + p_{10}} \frac{p_{01}}{p_{01} + p_{10}}.$$

Επομένως, υπό τη μηδενική υπόθεση, έχουμε ότι:

$$\mathbb{E}_0(X|x_{10} + x_{01}) = (x_{10} + x_{01}) \frac{\delta_0}{p_{01} + p_{10}},$$

και

$$\text{Var}_0(X|x_{10} + x_{01}) = 4(x_{10} + x_{01}) \frac{p_{10}p_{01}}{(p_{01} + p_{10})^2}.$$

Με το ίδιο σκεπτικό, όπως προηγούμενα, για να μπορεί να χρησιμοποιηθεί η στατιστική συνάρτηση στην πράξη θα πρέπει να αντικατασταθούν στις παραπάνω εκφράσεις οι άγνωστες παράμετροι p_{01} και p_{10} από τις εκτιμήτριες συναρτήσεις $\hat{p}_{01} = \frac{x_{01}}{n}$ και $\hat{p}_{10} = \frac{x_{10}}{n}$, αντίστοιχα. Πιο συγκεκριμένα, η προτεινόμενη στατιστική συνάρτηση δίνεται από τη σχέση:

$$Z_{LB}^{(2)} = \frac{x_{10} - x_{01} - \frac{(x_{10} + x_{01})\delta_0 n}{(x_{01} + x_{10})}}{\sqrt{4(x_{10} + x_{01}) \frac{n^2}{(x_{01} + x_{10})^2} \frac{x_{10}}{n} \frac{x_{01}}{n}}},$$

η οποία ισοδύναμα λαμβάνει τη μορφή

$$Z_{LB}^{(2)} = \frac{x_{10} - x_{01} - n\delta_0}{2} \sqrt{\frac{x_{10} + x_{01}}{x_{10}x_{01}}}. \quad (2.7)$$

Η στατιστική συνάρτηση $Z_{LB}^{(2)}$ ακολουθεί ασυμπτωτικά, υπό τη μηδενική υπόθεση, την τυπική κανονική κατανομή. Με βάση τη στατιστική συνάρτηση $Z_{LB}^{(2)}$

της σχέσης (2.7) απορρίπτεται, σε ασυμπτωτικό επίπεδο σημαντικότητας α , η μηδενική υπόθεση $H_0 : p_{10} - p_{01} = \delta_0$ έναντι της εναλλακτικής $H_0 : p_{10} - p_{01} = \delta_1$, αν ικανοποιείται η κρίσιμη περιοχή $Z_{LB}^{(2)} \geq z_\alpha$.

Για την υλοποίηση αυτού του ελέγχου δημιουργήσαμε στη γλώσσα προγραμματισμού R μία συνάρτηση, η οποία λαμβάνει ως ορίσματα το πλήθος των ασύμμετρων ζευγών, το συνολικό μέγεθος του δείγματος, την τιμή δ_0 και το επίπεδο σημαντικότητας α του ελέγχου. Εφαρμόζοντας αυτήν τη συνάρτηση λαμβάνουμε ως αποτελέσματα την τιμή της στατιστικής συνάρτησης, την (ασυμπτωτική) p -τιμή του ελέγχου και μήνυμα αν απορρίπτεται ή όχι η μηδενική υπόθεση στο καθορισμένο επίπεδο σημαντικότητας. Στο πλαίσιο που ακολουθεί, παρατίθεται η συνάρτηση αυτή και η υλοποίησή της για τα δεδομένα του Πίνακα 2.1.

```
zequivalenceLB2 <- function(x10, x01, n, delta0, alpha)
{
  Z_LB_2 <- ((x10 - x01 - n * delta0) / 2) * sqrt((x10 +
    ↪ x01) / (x10 * x01))

  pvalueZ_LB_2 <- pnorm(Z_LB_2, lower.tail=FALSE)

  reject_H0 <- Z_LB_2 >= qnorm(1 - alpha)
  return(list(Z_value = Z_LB_2, p_value = pvalueZ_LB_2,
    ↪ reject_H0 = reject_H0))
}
result <- zequivalenceLB2(32, 27, 344, -0.05, 0.05)
print(result)
$Z_value
[1] 2.900629

$p_value
[1] 0.001862075

$reject_H0
[1] TRUE
```

2.2 Οι έλεγχοι του Nam (1997)

Στην εργασία του Nam (1997) το ενδιαφέρον επικεντρώνεται στον έλεγχο της υπόθεσης ότι μία νέα διαγνωστική μέθοδος είναι το ίδιο καλή με μία υπάρχουσα. Επομένως, το ενδιαφέρον εστιάζεται στον έλεγχο της υπόθεσης που διατυπώθηκε στη σχέση (2.1), με τη διαφοροποίηση ότι ο Nam (1997) εξετάζει

τόσο την περίπτωση όπου $\delta_0 < \delta_1$ όσο και την περίπτωση $\delta_0 > \delta_1$.

Παρατηρώντας ο Nam (1997) ότι όλη η πληροφορία που απαιτείται για την εφαρμογή του ελέγχου, συγκεντρώνεται στις δυαδικές παρατηρήσεις x_{10} και x_{01} , θέλησε να αξιοποιήσει ότι οι τυχαίες μεταβλητές των αντίστοιχων συνδυασμών ακολουθούν από κοινού την πολυωνυμική κατανομή με πιθανότητες επιτυχίας p_{10} και p_{01} , δηλαδή ότι

$$f(x_{10}, x_{01}; p_{10}, p_{01}) = C p_{10}^{x_{10}} p_{01}^{x_{01}} (1 - p_{10} - p_{01})^{n-(x_{10}+x_{01})}, \quad (2.8)$$

όπου

$$C = \frac{n!}{x_{10}! x_{01}! (n - x_{10} - x_{01})!}.$$

Από την παραπάνω σχέση άμεσα προκύπτει ότι ο λογάριθμος της συνάρτησης πιθανοφάνειας δίνεται από τη σχέση:

$$\begin{aligned} L(p_{10}, p_{01}) &= \log(C) + x_{10} \log(p_{10}) + x_{01} \log(p_{01}) \\ &+ (n - x_{10} - x_{01}) \log(1 - p_{10} - p_{01}). \end{aligned}$$

Ωστόσο, ο λογάριθμος της συνάρτησης πιθανοφάνειας, ακολουθώντας νέα παραμετροποίηση που προκύπτει θέτοντας $p_{10} - p_{01} = \delta$, δηλαδή θέτοντας $p_{10} = p_{01} + \delta$, μπορεί ισοδύναμα να γραφεί ως:

$$\begin{aligned} L(p_{01}; \delta) &= \log(C) + x_{10} \log(p_{01} + \delta) + x_{01} \log(p_{01}) \\ &+ (n - x_{10} - x_{01}) \log(1 - 2p_{01} - \delta), \end{aligned}$$

όπου τώρα η μόνη άγνωστη παράμετρος είναι η p_{01} , καθώς η παράμετρος δ είναι πλήρως καθορισμένη τόσο υπό τη μηδενική όσο και υπό την εναλλακτική υπόθεση.

Για την εύρεση του εκτιμητή μέγιστης πιθανοφάνειας της παραμέτρου p_{01} για δοθείσα τιμή της παραμέτρου δ , υπολογίζουμε, αρχικά, την παράγωγο της συνάρτησης πιθανοφάνειας $L(p_{01}; \delta)$ ως προς p_{01} . Είναι

$$\frac{\partial L(p_{01}; \delta)}{\partial p_{01}} = \frac{x_{10}}{p_{01} + \delta} + \frac{x_{01}}{p_{01}} - \frac{2(n - x_{10} - x_{01})}{1 - 2p_{01} - \delta}.$$

Στη συνέχεια, ο εκτιμητής μέγιστης πιθανοφάνειας της παραμέτρου p_{01} για δοθείσα τιμή της παραμέτρου δ , έστω δ_0 , προκύπτει από την επίλυση της εξίσωσης

$$\left. \frac{\partial L(p_{01}; \delta)}{\partial p_{01}} \right|_{\delta=\delta_0} = 0.$$

Επομένως, αν $\tilde{p}_{01}$ είναι ο εκτιμητής μέγιστης πιθανοφάνειας της παραμέτρου p_{01} όταν $\delta = \delta_0$, τότε, μετά από λίγη άλγεβρα προκύπτει, ότι ικανοποιεί την εξίσωση:

$$a\tilde{p}_{01}^2 + b\tilde{p}_{01} + c = 0, \quad (2.9)$$

όπου $a = 2n$, $b = (2n + x_{10} - x_{01})\delta_0 - (x_{10} + x_{01})$ και $c = -x_{01}\delta_0(1 - \delta_0)$. Ειδικότερα, καθώς $0 < p_{01} < 1$, είναι:

$$\tilde{p}_{01} = \frac{-b + \sqrt{b^2 - 4ac}}{2a}. \quad (2.10)$$

Ακολούθως, για τον προσδιορισμό της στατιστικής συνάρτησης ελέγχου ο Nam (1997) χρησιμοποιεί τη μέθοδο score, η οποία βασίζεται στην παράγωγο του λογαρίθμου της συνάρτησης πιθανοφάνειας, όταν αυτή υπολογίζεται στην τιμή υπό τη μηδενική υπόθεση, έχοντας εκτιμήσει τις υπόλοιπες ενοχλητικές παραμέτρους (nuisance parameter), και η οποία δεν θα πρέπει να διαφέρει από το μηδέν περισσότερο από το δειγματικό σφάλμα. Ειδικότερα, σύμφωνα με τη μέθοδο που προτάθηκε από τον Bartlett (1953) και η οποία βασίζεται στον εκτιμητή μέγιστης πιθανοφάνειας της ενοχλητικής παραμέτρου (nuisance parameter) p_{01} για δοθείσες τιμές του δ , η στατιστική συνάρτηση ελέγχου της $H_0 : \delta = \delta_0$ έναντι της εναλλακτικής $H_1 : \delta = \delta_1$ προσδιορίζεται από τη σχέση:

$$Z_s = \frac{S(\delta_0, \tilde{p}_{01})}{\sqrt{v(\delta_0, \tilde{p}_{01})}}, \quad (2.11)$$

όπου $S(\delta, p_{01})$ είναι η πρώτη τάξης παράγωγος του λογαρίθμου της συνάρτησης πιθανοφάνειας $L(p_{01}; \delta)$ ως προς δ , και

$$v(\delta, p_{01}) = I_{11}(\delta, p_{01}) - \frac{I_{12}^2(\delta, p_{01})}{I_{22}(\delta, p_{01})},$$

με $I_{ij}(\delta, p_{01})$ να είναι το (i, j) στοιχείο του πίνακα πληροφορίας του Fisher. Ειδικότερα, λαμβάνοντας υπόψη ότι:

$$S(\delta, p_{01}) = \frac{\partial L(p_{01}; \delta)}{\partial \delta} = \frac{x_{10}}{p_{01} + \delta} - \frac{n - x_{10} - x_{01}}{1 - 2p_{01} - \delta}, \quad (2.12)$$

και ότι

$$v(\delta_0, p_{01}) = \frac{n}{2p_{01} + \delta + \delta^2}, \quad (2.13)$$

καθώς

$$I_{11}(\delta, p_{01}) = -E \left(\frac{\partial^2 L(p_{01}; \delta)}{\partial \delta^2} \right) = \frac{n(1 - p_{01})}{(p_{01} + \delta)(1 - 2p_{01} - \delta)},$$

$$I_{12}(\delta, p_{01}) = -E \left(\frac{\partial^2 L(p_{01}; \delta)}{\partial p_{01} \partial \delta} \right) = \frac{n(1 + \delta)}{(p_{01} + \delta)(1 - 2p_{01} - \delta)},$$

και

$$I_{22}(\delta, p_{01}) = -E \left(\frac{\partial^2 L(p_{01}; \delta)}{\partial p_{01}^2} \right) = \frac{n(2p_{01} + \delta - \delta^2)}{(p_{01} + \delta)p_{01}(1 - 2p_{01} - \delta)},$$

έχουμε, μετά από λίγη άλγεβρα, ότι:

$$S(\delta_0, \tilde{p}_{01}) = \frac{x_{10}\tilde{q}_{01} - (n - x_{01})\tilde{p}_{01}}{\tilde{p}_{01}(1 - \tilde{p}_{10} - \tilde{p}_{01})}, \text{ για } \tilde{p}_{01} > 0, \quad (2.14)$$

με

$$v(\delta_0, \tilde{p}_{01}) = \frac{n}{\tilde{p}_{10} + \tilde{p}_{01} - \delta_0^2}, \quad (2.15)$$

όπου $\tilde{q}_{01} = 1 - \tilde{p}_{01}$, $\tilde{p}_{10} = \tilde{p}_{01} + \delta_0$. Από τα παραπάνω, με αντικατάσταση στη σχέση (2.11), προκύπτει ότι η προτεινόμενη στατιστική συνάρτηση δίνεται από τη σχέση:

$$Z_s = \frac{x_{10}(1 - \tilde{p}_{01}) - (n - x_{01})\tilde{p}_{10}}{\tilde{p}_{10}(1 - \tilde{p}_{10} - \tilde{p}_{01})} \sqrt{\frac{\tilde{p}_{10} + \tilde{p}_{01} - \delta_0^2}{n}}. \quad (2.16)$$

Με βάση τη στατιστική συνάρτηση Z_s της σχέσης (2.16) απορρίπτεται, σε ασυμπτωτικό επίπεδο σημαντικότητας α , η μηδενική υπόθεση $H_0 : p_{10} - p_{01} = \delta_0$ έναντι της εναλλακτικής $H_1 : p_{10} - p_{01} = \delta_1$, με $\delta_1 > \delta_0$ αν ικανοποιείται η κρίσιμη περιοχή $Z_s \geq z_\alpha$. Σε μία τέτοια περίπτωση, συμπεραίνουμε ότι η νέα διαδικασία δεν είναι (τουλάχιστον) χειρότερη από τη διαδικασία αναφοράς. Παρόμοια, απορρίπτεται, σε (ασυμπτωτικό) επίπεδο σημαντικότητας α , η μηδενική υπόθεση $H_0 : p_{10} - p_{01} = \delta_0$ έναντι της εναλλακτικής $H_1 : p_{10} - p_{01} = \delta_1$, με $\delta_1 < \delta_0$ αν ικανοποιείται η κρίσιμη περιοχή $Z_s \leq -z_\alpha$ και τότε συμπεραίνουμε την υπεροχή της διαδικασίας αναφοράς. Τέλος, σε περίπτωση δίπλευρου ελέγχου $H_0 : p_{10} - p_{01} = \delta_0$ έναντι της εναλλακτικής $H_1 : p_{10} - p_{01} \neq \delta_0$, η μηδενική υπόθεση απορρίπτεται αν $|Z_s| > z_{\alpha/2}$.

Παρατήρηση 2.2. Στην ειδική περίπτωση που $\delta_0 = 0$ έχουμε, από τη σχέση (2.9), ότι ο εκτιμητής μέγιστης πιθανοφάνειας $\tilde{p}_{01}$ ικανοποιεί τη σχέση:

$$2n\tilde{p}_{01}^2 - (x_{10} + x_{01})\tilde{p}_{01} = 0$$

και επομένως:

$$\tilde{p}_{01} = \frac{x_{10} + x_{01}}{2n}. \quad (2.17)$$

Επιπρόσθετα, μετά από λίγη άλγεβρα, προκύπτει ότι σε αυτήν την περίπτωση:

$$Z_s = \frac{x_{10} - x_{01}}{\sqrt{x_{10} + x_{01}}}.$$

Ανακαλώντας τη σχέση (1.1) συμπεραίνουμε ότι σε αυτήν την περίπτωση ο έλεγχος ανάγεται ουσιαστικά σε αυτόν που προτάθηκε από τον McNemar (1947).

Για την υλοποίηση αυτού του ελέγχου δημιουργήσαμε στη γλώσσα προγραμματισμού R μία συνάρτηση, η οποία λαμβάνει ως ορίσματα το πλήθος των ασύμμετρων ζευγών, το συνολικό μέγεθος του δείγματος, την τιμή δ_0 , το επίπεδο σημαντικότητας α του ελέγχου, αλλά επιπλέον και τη δυνατότητα στο όρισμα *alternative* να δηλώσουμε αν ο έλεγχος είναι δεξιόπλευρος (γράφοντας *right*), αριστερόπλευρος (γράφοντας *left*) ή δίπλευρος (γράφοντας *twotailed*). Εφαρμόζοντας αυτήν τη συνάρτηση λαμβάνουμε ως αποτελέσματα την τιμή της στατιστικής συνάρτησης, την (ασυμπτωτική) p -τιμή του ελέγχου, μήνυμα αν απορρίπτεται ή όχι η μηδενική υπόθεση στο καθορισμένο επίπεδο σημαντικότητας, καθώς επίσης και τον τύπο του ελέγχου που διενεργήθηκε (αριστερόπλευρος, δεξιόπλευρος ή δίπλευρος). Στο πλαίσιο που ακολουθεί, παρατίθεται η συνάρτηση αυτή και η υλοποίησή της, για την περίπτωση του δεξιόπλευρου ελέγχου, για τα δεδομένα του Πίνακα 2.1.

```
z_Nam_score <- function(x10, x01, n, delta0, alpha,
  → alternative) {
  a <- 2 * n
  b <- (2 * n + x01 - x10) * delta0 - (x10 + x01)
  c <- -x01 * delta0 * (1 - delta0)

  p01_tilde <- (-b + sqrt(b^2 - 4 * a * c)) / (2 * a)
  p10_tilde <- p01_tilde + delta0

  Zs <- (x10 * (1 - p01_tilde) - (n - x01) * p10_tilde) /
    (p10_tilde * (1 - p10_tilde - p01_tilde)) *
    sqrt((p10_tilde + p01_tilde - delta0^2) / n)

  if (alternative == "right") {
    p_value <- pnorm(Zs, lower.tail = FALSE)
    reject_H0 <- Zs >= qnorm(1 - alpha)
  } else if (alternative == "left") {
    p_value <- pnorm(Zs, lower.tail = TRUE)
    reject_H0 <- Zs <= -qnorm(1 - alpha)
  } else if (alternative == "twotailed") {
    p_value <- 2 * pnorm(-abs(Zs))
```

```

    reject_H0 <- abs(Zs) >= qnorm(1 - alpha / 2)
  } else {
    stop("The 'alternative' argument must be either '
      ↪ right', 'left', or 'twotailed'")
  }
  return(list(
    Z_value = Zs,
    p_value = p_value,
    reject_H0 = reject_H0,
    alternative = alternative
  ))
}

result<- z_Nam_score(32, 27, 344, -0.05,0.05, alternative
  ↪ ="right")
print(result)
$Z_value
[1] 2.795403

$p_value
[1] 0.002591753

$reject_H0
[1] TRUE

$alternative
[1] "right"

```

Η στατιστική συνάρτηση Z_s της σχέσης (2.16) προτάθηκε από τον Nam (1997) χρησιμοποιώντας τη μέθοδο score και τη γενική θεωρία του Bartlett (1953). Ωστόσο, όπως επισημαίνεται από τον Nam (1997) και θα δούμε στη συνέχεια, στην ίδια στατιστική συνάρτηση μπορούμε να καταλήξουμε ακολουθώντας μία διαφορετική μεθοδολογία. Ειδικότερα, ανακαλώντας ότι θέλουμε, ουσιαστικά, να ελέγξουμε τη μηδενική υπόθεση $H_0 : p_{10} - p_{01} = \delta_0$, είναι λογικό η προτεινόμενη στατιστική συνάρτηση να βασίζεται στο εμπειρικό (δειγματικό) ανάλογο της διαφοράς $\delta = p_{10} - p_{01}$, έστω $\hat{\delta}$, και στην απόκλιση του από την δοθείσα τιμή δ_0 , δηλαδή να στηρίζεται στη στατιστική συνάρτηση $t = \hat{\delta} - \delta_0$. Ανακαλώντας τα αποτελέσματα που παρουσιάστηκαν στην Ενότητα 2.1 έχουμε ότι, υπό τη μηδενική υπόθεση, $\mathbb{E}_0(t) = 0$ και $Var_0(t) = \frac{p_{10} + p_{01} - \delta_0^2}{n}$. Σε αντίθεση με τους Lu and Bean (1995), ο Nam (1997) προτείνει την αντικατάσταση των άγνωστων παραμέτρων p_{01} και p_{10} όχι από τις εκτιμήτριες $\hat{p}_{01} = \frac{x_{01}}{n}$

και $\hat{p}_{10} = \frac{x_{10}}{n}$, αλλά από την εκτιμήτρια $\tilde{p}_{01}$, που προσδιορίστηκε στη σχέση (2.9), και την εκτιμήτρια $\tilde{p}_{10} = \delta_0 + \tilde{p}_{01}$, αντίστοιχα. Συνοψίζοντας τα παραπάνω η στατιστική συνάρτηση που προτάθηκε από τον Nam (1997) δίνεται από τη σχέση:

$$Z_t = \frac{x_{10} - x_{01} - n\delta_0}{\sqrt{n(\tilde{p}_{10} + \tilde{p}_{01} - \delta_0^2)}}, \quad (2.18)$$

η οποία ασυμπτωτικά ακολουθεί, υπό τη μηδενική υπόθεση, την τυπική κανονική κατανομή. Παρότι η στατιστική συνάρτηση Z_t της σχέσης (2.18) φαίνεται διαφορετική από τη στατιστική συνάρτηση Z_s της σχέσης (2.16), μετά από μία σειρά αλγεβρικών πράξεων που παρατίθενται στο Παράρτημα 2 της εργασίας του Nam (1997) προκύπτει ότι οι δύο στατιστικές συναρτήσεις είναι ισοδύναμες.

Παρατήρηση 2.3. Η στατιστική συνάρτηση Z_t της σχέσης (2.18) διαφοροποιείται από τη στατιστική συνάρτηση $Z_{LB}^{(1)}$ της σχέσης (2.6) στον τρόπο εκτίμησης των άγνωστων παραμέτρων p_{10} και p_{01} που εμφανίζονται στη διακύμανση της στατιστικής συνάρτησης $\hat{\delta}$. Πιο συγκεκριμένα, η διαφοροποίηση έγκειται στο γεγονός ότι στην περίπτωση της στατιστικής συνάρτησης Z_t οι εκτιμήτριες έχουν προκύψει υπό τη μηδενική υπόθεση, γεγονός που έχει ως συνέπεια η προσέγγιση της τυπικής κανονικής κατανομής να είναι καλύτερη.

Για την υλοποίηση αυτού του ελέγχου δημιουργήσαμε στη γλώσσα προγραμματισμού R μία συνάρτηση, η οποία λαμβάνει ως ορίσματα το πλήθος των ασύμμετρων ζευγών, το συνολικό μέγεθος του δείγματος, την τιμή δ_0 , το επίπεδο σημαντικότητας α του ελέγχου, αλλά επιπλέον και τη δυνατότητα στο όρισμα *alternative* να δηλώσουμε αν ο έλεγχος είναι δεξιόπλευρος (γράφοντας *right*), αριστερόπλευρος (γράφοντας *left*) ή δίπλευρος (γράφοντας *twotailed*). Εφαρμόζοντας αυτήν τη συνάρτηση λαμβάνουμε ως αποτελέσματα την τιμή της στατιστικής συνάρτησης, την (ασυμπτωτική) p -τιμή του ελέγχου, μήνυμα αν απορρίπτεται ή όχι η μηδενική υπόθεση στο καθορισμένο επίπεδο σημαντικότητας και μήνυμα σχετικά με τον τύπο του ελέγχου που διενεργήθηκε. Στο πλαίσιο που ακολουθεί, παρατίθεται η συνάρτηση αυτή και η υλοποίησή της, για την περίπτωση του δεξιόπλευρου ελέγχου, για τα δεδομένα του Πίνακα 2.1.

```
z_Nam_t <- function(x10, x01, n, delta0, alpha,
  ↳ alternative) {
  a <- 2 * n
  b <- (2 * n + x01 - x10) * delta0 - (x10 + x01)
  c <- -x01 * delta0 * (1 - delta0)
  p01_tilde <- (-b + sqrt(b^2 - 4 * a * c)) / (2 * a)
  p10_tilde <- p01_tilde + delta0
```

```

Zt <- (x10-x01-n*delta0)/sqrt(n*(p01_tilde+p10_tilde-
  ↪ delta0^2))
if (alternative == "right") {
  p_value <- pnorm(Zt, lower.tail = FALSE)
  reject_H0 <- Zt >= qnorm(1 - alpha)
} else if (alternative == "left") {
  p_value <- pnorm(Zt, lower.tail = TRUE)
  reject_H0 <- Zt <= -qnorm(1 - alpha)
} else if (alternative == "twotailed") {
  p_value <- 2 * pnorm(-abs(Zt))
  reject_H0 <- abs(Zt) >= qnorm(1 - alpha / 2)
} else {
  stop("The 'alternative' argument must be either 'right
  ↪ ', 'left', or 'two'")
}
return(list(
  Z_value = Zt,
  p_value = p_value,
  reject_H0 = reject_H0,
  alternative = alternative
))
}

result<- z_Nam_t(32, 27, 344, -0.05,0.05, alternative="
  ↪ right")
print(result)
$Z_value
[1] 2.795403

$p_value
[1] 0.002591753

$reject_H0
[1] TRUE
$alternative
[1] "right"

```

2.3 Οι έλεγχοι των Liu et al. (2002)

Σύμφωνα με τους Liu et al. (2002) ο έλεγχος ισοδυναμίας μεταξύ της εναλλακτικής διαγνωστικής μεθόδου και της μεθόδου αναφοράς (είτε αυτή είναι ο χρυσός κανόνας ή κάποια καθιερωμένη μέθοδος) ανάγεται στον έλεγχο των

ακόλουθων διαστηματικών υποθέσεων (interval hypotheses):

$$H_0 : p_{10} - p_{01} \geq \delta_0 \quad \text{ή} \quad p_{10} - p_{01} \leq -\delta_0, \quad H_1 : -\delta_0 < p_{10} - p_{01} < \delta_0, \quad (2.19)$$

όπου, τώρα, $\delta_0 > 0$ είναι ένα συγκεκριμένο προκαθορισμένο όριο που συνεπάγεται την κλινική ισοδυναμία των μεθόδων. Οι παραπάνω διαστηματικές υποθέσεις μπορούν να διασπασθούν στα ακόλουθα δύο σύνολα μονόπλευρων υποθέσεων:

$$H_{0\ell} : p_{10} - p_{01} \leq -\delta_0 \quad \text{έναντι} \quad H_{1\ell} : p_{10} - p_{01} > -\delta_0, \quad (2.20)$$

και

$$H_{0u} : p_{10} - p_{01} \geq \delta_0 \quad \text{έναντι} \quad H_{1u} : p_{10} - p_{01} < \delta_0. \quad (2.21)$$

Ειδικότερα, η μονόπλευρη υπόθεση της σχέσης (2.20) αναφέρεται ως έλεγχος μη κατωτερότητας.

Στο πλαίσιο αυτό, οι Liu et al. (2002), ακολουθώντας παρόμοιο σκεπτικό με αυτό του Nam (1997) για τον προσδιορισμό της στατιστικής συνάρτησης Z_t , προτείνουν στατιστικές συναρτήσεις τύπου Wald για τον έλεγχο των υποθέσεων που διατυπώνονται στις σχέσεις (2.20) και (2.21). Πιο συγκεκριμένα, από τις σχέσεις (2.3) και (2.4), προκύπτει ότι ο εκτιμητής $\hat{\delta}$ της σχέσης (2.2) ακολουθεί ασυμπτωτικά κανονική κατανομή με μέση τιμή δ και διακύμανση που προσδιορίστηκε στη σχέση (2.4), η οποία ισοδύναμα γράφεται:

$$\sigma^2 := \text{Var}(\hat{\delta}) = \frac{p_{10} + p_{01} - (p_{10} - p_{01})^2}{n}, \quad (2.22)$$

και επομένως εξαρτάται από τις άγνωστες πιθανότητες p_{10} , p_{01} και τη διαφορά αυτών $p_{10} - p_{01}$. Είναι τώρα προφανές ότι διαφορετικές στατιστικές συναρτήσεις μπορούν να προκύψουν αν χρησιμοποιηθούν διαφορετικές εκτιμήτριες των άγνωστων αυτών ποσοτήτων.

1η μέθοδος εκτίμησης της άγνωστης διακύμανσης

Σύμφωνα με αυτήν τη μέθοδο η εκτιμήτρια της διακύμανσης προκύπτει αντικαθιστώντας στη σχέση (2.22) τις άγνωστες πιθανότητες p_{10} και p_{01} από τα αντίστοιχα δειγματικά ποσοστά, δηλαδή από τους εκτιμητές μέγιστης πιθανοφάνειας αυτών χωρίς κανέναν περιορισμό. Τότε, η διακύμανση σ^2 εκτιμάται από τη σχέση:

$$\hat{\sigma}^2 = \frac{\hat{p}_{10} + \hat{p}_{01} - (\hat{p}_{10} - \hat{p}_{01})^2}{n}, \quad (2.23)$$

όπου $\hat{p}_{10} = \frac{x_{10}}{n}$ και $\hat{p}_{01} = \frac{x_{01}}{n}$. Συγκεντρωτικά, η τύπου Wald στατιστική συνάρτηση που προτείνεται από τους Liu et al. (2002) για τον έλεγχο της υπόθεσης (2.20) δίνεται από τη σχέση:

$$Z_\ell = \frac{\hat{\delta} + \delta_0}{\hat{\sigma}} = \frac{x_{10} - x_{01} + n\delta_0}{\sqrt{x_{10} + x_{01} - n\hat{\delta}^2}}, \quad (2.24)$$

η οποία ασυμπτωτικά ακολουθεί, υπό τη μηδενική υπόθεση, τυπική κανονική κατανομή. Με βάση τη στατιστική συνάρτηση Z_ℓ της σχέσης (2.24) απορρίπτεται, σε ασυμπτωτικό επίπεδο σημαντικότητας α , η μηδενική υπόθεση της σχέσης (2.20) αν $Z_\ell \geq z_\alpha$. Ωστόσο, οι Liu et al. (2002) επισημαίνουν ότι σύμφωνα με τον οδηγό διεξαγωγής κλινικών δοκιμών που συντάχθηκε από τον Lewis (1999), η εναλλακτική μεθοδολογία είναι μη κατώτερη της υπάρχουσας αν $Z_\ell \geq z_{\alpha/2}$.

Με παρόμοιο τρόπο, η τύπου Wald στατιστική συνάρτηση που προτείνεται από τους Liu et al. (2002) για τον έλεγχο της υπόθεσης (2.21) δίνεται από τη σχέση:

$$Z_u = \frac{\hat{\delta} - \delta_0}{\hat{\sigma}} = \frac{x_{10} - x_{01} - n\delta_0}{\sqrt{x_{10} + x_{01} - n\hat{\delta}^2}}, \quad (2.25)$$

η οποία ασυμπτωτικά ακολουθεί, υπό τη μηδενική υπόθεση, την τυπική κανονική κατανομή. Με βάση τη στατιστική συνάρτηση Z_u της σχέσης (2.25) απορρίπτεται, σε ασυμπτωτικό επίπεδο σημαντικότητας α , η μηδενική υπόθεση της σχέσης (2.21) αν $Z_u \leq -z_\alpha$.

Όσον αφορά τον έλεγχο της υπόθεσης της ισοδυναμίας, δηλαδή της υπόθεσης της σχέσης (2.19), χρησιμοποιώντας την αρχή της ένωσης τομής (intersection-union principle), η μηδενική υπόθεση απορρίπτεται, και επομένως συμπεραίνουμε ότι οι δύο μεθοδολογίες είναι ισοδύναμες, αν $Z_\ell \geq z_\alpha$ και $Z_u \leq -z_\alpha$, δηλαδή αν απορρίπτονται και οι δύο έλεγχοι που βασίζονται στις στατιστικές συναρτήσεις Z_ℓ και Z_u .

Για την υλοποίηση αυτού του ελέγχου δημιουργήσαμε στη γλώσσα προγραμματισμού R μία συνάρτηση, η οποία λαμβάνει ως ορίσματα το πλήθος των ασύμμετρων ζευγών, το συνολικό μέγεθος του δείγματος, την τιμή $\delta_0 > 0$, το επίπεδο σημαντικότητας α του ελέγχου, αλλά επιπλέον και τη δυνατότητα στο όρισμα `type` να δηλώσουμε αν θέλουμε να ελέγξουμε την υπόθεση της σχέσης (2.20) γράφοντας `type=noninferiority`, αν θέλουμε να ελέγξουμε την υπόθεση της σχέσης (2.21) γράφοντας `type=nonsuperiority` ή την υπόθεση της σχέσης (2.19) γράφοντας `type=equivalence`. Εφαρμόζοντας αυτήν τη συνάρτηση λαμβάνουμε ως αποτελέσματα την τιμή της στατιστικής συνάρτησης, την (ασυμπτωτική) p -τιμή του ελέγχου, μήνυμα αν απορρίπτεται ή όχι η μηδενική υπόθεση που προσ-

διορίσαμε στο όρισμα `type`, στο καθορισμένο επίπεδο σημαντικότητας καθώς και μήνυμα που αφορά τον τύπο της υπόθεσης που ελέγχθηκε. Στο πλαίσιο που ακολουθεί, παρατίθεται η συνάρτηση αυτή και η υλοποίησή της, για την περίπτωση του ελέγχου ισοδυναμίας, για τα δεδομένα του Πίνακα 2.1.

```
liu_test1 <- function(x10, x01, n, delta0, alpha = 0.05,
  ↪ type) {

  delta_hat <- (x10 - x01) / n

  Z_ell <- (x10 - x01 + n * delta0) / sqrt(x10 + x01 - n *
  ↪ delta_hat^2)
  Z_u <- (x10 - x01 - n * delta0) / sqrt(x10 + x01 - n *
  ↪ delta_hat^2)

  if (type == "noninferiority") {
    p_value <- pnorm(Z_ell, lower.tail = FALSE)
    reject_H0 <- Z_ell >= qnorm(1 - alpha)
    Z_value <- Z_ell

  } else if (type == "nonsuperiority") {
    p_value <- pnorm(Z_u, lower.tail = TRUE)
    reject_H0 <- Z_u <= -qnorm(1 - alpha)
    Z_value <- Z_u

  } else if (type == "equivalence") {
    p_value <- max(pnorm(Z_ell, lower.tail = FALSE), pnorm
    ↪ (Z_u, lower.tail = TRUE))
    reject_H0 <- (Z_ell >= qnorm(1 - alpha)) & (Z_u <= -
    ↪ qnorm(1 - alpha))
    Z_value <- c(Z_ell, Z_u)

  } else {
    stop("The 'alternative' argument must be either '
    ↪ noninferiority', 'nonsuperiority', or '
    ↪ equivalence'")
  }

  return(list(
    Z_value = Z_value,
    p_value = p_value,
    reject_H0 = reject_H0,
    type = type
  ))
}
```

```

}

result <- liu_test1(32, 27, 344, 0.05,0.05,type="
  ↪ equivalence")
print(result)
$Z_value
[1] 2.891975 -1.589284

$p_value
[1] 0.05599816

$reject_H0
[1] FALSE

$type
[1] "equivalence"

```

Επισημαίνεται ότι η παραπάνω συνάρτηση μπορεί να τροποποιηθεί έτσι ώστε να είναι εφικτή η διόρθωση συνεχείας. Σε μία τέτοια περίπτωση απλά είναι

```

Z_ell <- (x10 - x01 + n * delta0-0.5) / sqrt(x10 + x01 - n
  ↪ * delta_hat^2)
Z_u   <- (x10 - x01 - n * delta0+0.5) / sqrt(x10 + x01 - n
  ↪ * delta_hat^2)

```

2η μέθοδος εκτίμησης της άγνωστης διακύμανσης

Οι Liu et al. (2002), παρακινούμενοι από τους Lu and Bean (1995), προτείνουν έναν ακόμα τρόπο εκτίμησης της άγνωστης διακύμανσης σ^2 της σχέσης (2.22), που οδηγεί σε μία παραλλαγή των παραπάνω στατιστικών συναρτήσεων. Ειδικότερα, λαμβάνοντας υπόψη ότι, υπό τη μηδενική υπόθεση, η διαφορά $(p_{10} - p_{01})^2$ είναι ίση με δ_0^2 , προτείνουν τη χρήση της ακόλουθης εκτιμήτριας

$$\tilde{\sigma}^2 = \frac{\hat{p}_{10} + \hat{p}_{01} - \delta_0^2}{n}, \quad (2.26)$$

όπου $\hat{p}_{10} = \frac{x_{10}}{n}$ και $\hat{p}_{01} = \frac{x_{01}}{n}$. Ωστόσο, η μέθοδος αυτή δεν διατηρεί το ονομαστικό επίπεδο σημαντικότητας, και δεν προτείνεται να χρησιμοποιείται στην πράξη.

Στο πλαίσιο που ακολουθεί, παρατίθεται η συνάρτηση της R για αυτόν τον έλεγχο και η υλοποίησή του, για την περίπτωση του ελέγχου ισοδυναμίας, για τα δεδομένα του Πίνακα 2.1. Δεν δίνονται περισσότερες λεπτομέρειες, καθώς

η δομή της συνάρτησης είναι ίδια με αυτήν του πρώτου ελέγχου των Liu et al. (2002).

```
liu_test2 <- function(x10, x01, n, delta0, alpha = 0.05,
  ↪ type) {
  Z_ell <- (x10 - x01 + n * delta0) / sqrt((x01+x10-n *
  ↪ delta0^2))
  Z_u <- (x10 - x01 - n * delta0) / sqrt((x01+x10-n *
  ↪ delta0^2))

  if (type == "noninferiority") {
    p_value <- pnorm(Z_ell, lower.tail = FALSE)
    reject_H0 <- Z_ell >= qnorm(1 - alpha)
    Z_value <- Z_ell

  } else if (type == "nonsuperiority") {
    p_value <- pnorm(Z_u, lower.tail = TRUE)
    reject_H0 <- Z_u <= -qnorm(1 - alpha)
    Z_value <- Z_u

  } else if (type == "equivalence") {
    p_value <- max(pnorm(Z_ell, lower.tail = FALSE), pnorm
    ↪ (Z_u, lower.tail = TRUE))
    reject_H0 <- (Z_ell >= qnorm(1 - alpha)) & (Z_u <= -
    ↪ qnorm(1 - alpha))
    Z_value <- c(Z_ell, Z_u)

  } else {
    stop("The 'alternative' argument must be either '
    ↪ noninferiority', 'nonsuperiority', or '
    ↪ equivalence'")
  }

  return(list(
    Z_value = Z_value,
    p_value = p_value,
    reject_H0 = reject_H0,
    type = type
  ))
}

result <- liu_test2(32, 27, 344, 0.05, 0.05, type="
  ↪ equivalence")
print(result)
$Z_value
```

```
[1] 2.911491 -1.600009

$p_value
[1] 0.05479834

$reject_H0
[1] FALSE

$type
[1] "equivalence"
```

3η μέθοδος εκτίμησης της άγνωστης διακύμανσης

Οι Liu et al. (2002), παρακινούμενοι από τον Nam (1997), προτείνουν έναν ακόμα τρόπο εκτίμησης της άγνωστης διακύμανσης σ^2 της σχέσης (2.22), που οδηγεί σε μία παραλλαγή των παραπάνω στατιστικών συναρτήσεων. Πιο συγκεκριμένα, προτείνουν την αντικατάσταση της διαφοράς $(p_{10} - p_{01})^2$ από αυτήν υπό τη μηδενική υπόθεση, ήτοι από την ποσότητα δ_0^2 , και την αντικατάσταση των p_{10} και p_{01} από τις εκτιμήτριες που προκύπτουν λαμβάνοντας υπόψη ότι υπό τη μηδενική υπόθεση της σχέσης (2.20) είναι $p_{10} - p_{01} = -\delta_0$ και ότι υπό τη μηδενική υπόθεση της σχέσης (2.21) είναι $p_{10} - p_{01} = \delta_0$. Με παρόμοιο σκεπτικό με αυτό που παρουσιάστηκε στην Ενότητα 2.2 προκύπτει ότι υπό τη μηδενική υπόθεση της σχέσης (2.20) οι εκτιμήτριες $\tilde{p}_{\ell,01}$ και $\tilde{p}_{\ell,10}$ των p_{01} και p_{10} , αντίστοιχα, είναι:

$$\tilde{p}_{\ell,01} = \frac{-\tilde{b}_\ell + \sqrt{\tilde{b}_\ell^2 - 8\tilde{c}_\ell}}{4} \text{ και } \tilde{p}_{\ell,10} = \tilde{p}_{\ell,01} - \delta_0, \quad (2.27)$$

όπου

$$\tilde{b}_\ell = -\hat{\delta}(1 - \delta_0) - 2(\hat{p}_{01} + \delta_0) \text{ και } \tilde{c}_\ell = \delta_0(1 + \delta_0)\hat{p}_{01}.$$

Επομένως, η στατιστική συνάρτηση Z_ℓ τροποποιείται και λαμβάνει τη μορφή

$$Z_\ell^{(1)} = \frac{\hat{\delta} + \delta_0}{\hat{\sigma}_\ell} = \frac{x_{10} - x_{01} + n\delta_0}{\sqrt{n(\tilde{p}_{\ell,01} + \tilde{p}_{\ell,10}) - n\delta_0^2}} \quad (2.28)$$

η οποία ασυμπτωτικά ακολουθεί, υπό τη μηδενική υπόθεση, τυπική κανονική κατανομή. Με βάση τη στατιστική συνάρτηση $Z_\ell^{(1)}$ της σχέσης (2.28) απορρίπτεται, σε ασυμπτωτικό επίπεδο σημαντικότητας α , η μηδενική υπόθεση της σχέσης (2.20) αν $Z_\ell^{(1)} \geq z_\alpha$.

Παρόμοια, υπό τη μηδενική υπόθεση της σχέσης (2.21), οι εκτιμήτριες $\tilde{p}_{u,01}$ και $\tilde{p}_{u,10}$ των p_{01} και p_{10} , αντίστοιχα, είναι:

$$\tilde{p}_{u,01} = \frac{-\tilde{b}_u + \sqrt{\tilde{b}_u^2 - 8\tilde{c}_u}}{4} \text{ και } \tilde{p}_{u,10} = \tilde{p}_{u,01} + \delta_0, \quad (2.29)$$

όπου

$$\tilde{b}_u = -\hat{\delta}(1 + \delta_0) - 2(\hat{p}_{01} - \delta_0) \text{ και } \tilde{c}_u = -\delta_0(1 - \delta_0)\hat{p}_{01}.$$

Επομένως, η στατιστική συνάρτηση Z_u τροποποιείται και λαμβάνει τη μορφή

$$Z_u^{(1)} = \frac{\hat{\delta} - \delta_0}{\hat{\sigma}_u} = \frac{x_{10} - x_{01} - n\delta_0}{\sqrt{n(\tilde{p}_{u,01} + \tilde{p}_{u,10}) - n\delta_0^2}}, \quad (2.30)$$

η οποία ασυμπτωτικά ακολουθεί, υπό τη μηδενική υπόθεση, την τυπική κανονική κατανομή. Με βάση τη στατιστική συνάρτηση $Z_u^{(1)}$ της σχέσης (2.30) απορρίπτεται, σε ασυμπτωτικό επίπεδο σημαντικότητας α , η μηδενική υπόθεση της σχέσης (2.21) αν $Z_u^{(1)} \leq -z_\alpha$.

Όσον αφορά τον έλεγχο της υπόθεσης της ισοδυναμίας, δηλαδή της υπόθεσης της σχέσης (2.19), χρησιμοποιώντας την αρχή της ένωσης τομής (intersection-union principle), η μηδενική υπόθεση απορρίπτεται, και επομένως συμπεραίνουμε ότι οι δύο μεθοδολογίες είναι ισοδύναμες, αν $Z_\ell^{(1)} \geq z_\alpha$ και $Z_u^{(1)} \leq -z_\alpha$, δηλαδή αν απορρίπτονται και οι δύο έλεγχοι που βασίζονται στις στατιστικές συναρτήσεις $Z_\ell^{(1)}$ και $Z_u^{(1)}$.

Στο πλαίσιο που ακολουθεί, παρατίθεται η συνάρτηση της R για αυτόν τον έλεγχο και η υλοποίησή του, για την περίπτωση του ελέγχου ισοδυναμίας, για τα δεδομένα του Πίνακα 2.1. Δεν δίνονται περισσότερες λεπτομέρειες, καθώς η δομή της συνάρτησης είναι ίδια με αυτήν του πρώτου ελέγχου των Liu et al. (2002).

```
liu_test3 <- function(x10, x01, n, delta0, alpha = 0.05,
  ↪ type) {

  delta_hat<- (x10-x01)/n
  p10hat<- x10/n
  p01hat<- x01/n

  bltilde <- -delta_hat*(1- delta0) - 2* (p01hat + delta0)
  cltilde <- delta0 * (1+ delta0) * p01hat

  pl01tilde<- (-bltilde + sqrt(bltilde^2 - 8* cltilde))/4
```

```

pl10tilde<- pl01tilde - delta0

butilde <- -delta_hat*(1+ delta0) - 2* (p01hat - delta0)
cutilde <- -delta0 * (1- delta0) * p01hat

pu01tilde<- (-butilde + sqrt(butilde^2 - 8* cutilde))/4
pu10tilde<- pu01tilde + delta0

signal_tilde<- sqrt(n*(pl01tilde+pl10tilde)- n * delta0
  ↳ ^2)
sigmau_tilde<- sqrt(n*(pu01tilde+pu10tilde)- n * delta0
  ↳ ^2)

Z_ell1 <- (x10 - x01 + n * delta0) / signal_tilde
Z_u1 <- (x10 - x01 - n * delta0) / sigmau_tilde

if (type == "noninferiority") {
  p_value <- pnorm(Z_ell1, lower.tail = FALSE)
  reject_H0 <- Z_ell1 >= qnorm(1 - alpha)
  Z_value <- Z_ell1

} else if (type == "nonsuperiority") {
  p_value <- pnorm(Z_u1, lower.tail = TRUE)
  reject_H0 <- Z_u1 <= -qnorm(1 - alpha)
  Z_value <- Z_u1

} else if (type == "equivalence") {
  p_value <- max(pnorm(Z_ell1, lower.tail = FALSE),
  ↳ pnorm(Z_u1, lower.tail = TRUE))
  reject_H0 <- (Z_ell1 >= qnorm(1 - alpha)) & (Z_u1 <=
  ↳ -qnorm(1 - alpha))
  Z_value <- c(Z_ell1, Z_u1)

} else {
  stop("The 'alternative' argument must be either '
  ↳ noninferiority', 'nonsuperiority', or '
  ↳ equivalence'")
}

return(list(
  Z_value = Z_value,
  p_value = p_value,
  reject_H0 = reject_H0,
  type = type
))

```



```

}

result <- liu_test3(32, 27, 344, 0.05,0.05,type="
  ↪ equivalence")
print(result)
$Z_value
[1] 2.795403 -1.562518

$p_value
[1] 0.059083

$reject_H0
[1] FALSE

$type
[1] "equivalence"

```

Ο έλεγχος ισοδυναμίας μπορεί εναλλακτικά να ελεγχθεί μέσω της κατασκευής διαστήματος εμπιστοσύνης για τη διαφορά $p_{10} - p_{01}$. Πιο συγκεκριμένα, εξάγουμε, σε επίπεδο σημαντικότητας α , το συμπέρασμα της ισοδυναμίας της νέας διαγνωστικής μεθόδου με την υπάρχουσα αν και μόνο αν το $100(1 - 2\alpha)\%$ (ασυμπτωτικό) διάστημα εμπιστοσύνης είναι εντός του διαστήματος $(-\delta_0, \delta_0)$. Στο πλαίσιο αυτό, χρησιμοποιώντας την 1η μέθοδο εκτίμησης της διακύμανσης του εκτιμητή $\hat{\delta}$ έχουμε ότι ένα $100(1 - 2\alpha)\%$ ασυμπτωτικό διάστημα εμπιστοσύνης για την αληθινή διαφορά $p_{10} - p_{01}$ είναι το:

$$(\hat{p}_{10} - \hat{p}_{01} - z_\alpha \hat{\sigma}, \hat{p}_{10} - \hat{p}_{01} + z_\alpha \hat{\sigma})$$

όπου $\hat{p}_{10} = \frac{x_{10}}{n}$ και $\hat{p}_{01} = \frac{x_{01}}{n}$ και $\hat{\sigma}$ η τετραγωνική ρίζα της εκτιμώμενης διακύμανσης που προσδιορίστηκε στη σχέση (2.23). Από την άλλη, αν χρησιμοποιηθεί η 3η μέθοδος για την εκτίμηση της διακύμανσης, τότε ένα $100(1 - 2\alpha)\%$ ασυμπτωτικό διάστημα εμπιστοσύνης για την αληθινή διαφορά $p_{10} - p_{01}$ είναι το $(\hat{\delta}_\ell, \hat{\delta}_u)$ όπου

$$\hat{\delta}_\ell = \inf \left\{ \delta^* : \frac{\hat{\delta} - \delta^*}{\tilde{\sigma}_{\delta_0}} \leq z_\alpha \right\}$$

και

$$\hat{\delta}_u = \inf \left\{ \delta^* : \frac{\hat{\delta} - \delta^*}{\tilde{\sigma}_{\delta_*}} \geq -z_\alpha \right\},$$

με $\tilde{\sigma}_{\delta_*}$ να είναι η εκτίμηση της διακύμανσης υπό την τρίτη μέθοδο όταν η άγνωστη παράμετρος $\delta = \delta^*$. Σημειώνουμε ότι $\tilde{\sigma}_\ell = \tilde{\sigma}_{-\delta_0}$ και $\tilde{\sigma}_u = \tilde{\sigma}_{\delta_0}$. Το διάστημα εμπιστοσύνης με αυτήν τη μέθοδο μπορεί να προσδιοριστεί επαναληπτικά και

εξάγουμε το συμπέρασμα της ισοδυναμίας των διαγνωστικών μεθόδων αν $-\delta_0 \leq \hat{\delta}_\ell \leq \hat{\delta}_u \leq \delta_0$.

Παρατήρηση 2.4. Σε όλες τις παραπάνω στατιστικές συναρτήσεις χρησιμοποιήθηκε η προσέγγιση της διωνυμικής κατανομής από την κανονική κατανομή. Σε τέτοιες περιπτώσεις όπως αναφέρθηκε και στο Κεφάλαιο 1 προτείνεται να χρησιμοποιείται η διόρθωση συνέχειας, ειδικά όταν έχουμε διαθέσιμο ένα δείγμα με μικρό μέγεθος. Σε αυτό το πλαίσιο, η γενική μορφή των στατιστικών συναρτήσεων είναι

$$Z_\ell^{(cor)} = \frac{\hat{\delta} + \delta_0 - 0.5}{\widehat{Var}(\hat{\delta})}$$

και

$$Z_u^{(cor)} = \frac{\hat{\delta} - \delta_0 + 0.5}{\widehat{Var}(\hat{\delta})}$$

όπου $\widehat{Var}(\hat{\delta})$ ένας από τους εκτιμητές της διακύμανσης που προσδιορίστηκε είτε με την πρώτη ή την τρίτη μέθοδο.

2.4 Οι έλεγχοι των Chen et al. (2023)

Η ανάλυση των Chen et al. (2023) επικεντρώνεται στην ιδέα της fiducial κατανομής, μιας προσέγγισης που προτάθηκε αρχικά από τον Fisher (1922), ως εναλλακτική τόσο στην κλασική (frequentist) όσο και στη μπεϋζιανή στατιστική. Η κατανομή αυτή προσπαθεί να εκφράσει την αβεβαιότητα για μια παράμετρο, χωρίς τη χρήση προγενέστερης (prior) κατανομής, αλλά και χωρίς να καταφεύγει αποκλειστικά σε p -values. Η ιδέα λοιπόν ξεκινάει από μια σχέση $T(X) = A(U, \theta)$, ανάμεσα στη στατιστική συνάρτηση, την παράμετρο ενδιαφέροντος και μια βοηθητική μεταβλητή που προέρχεται από γνωστή κατανομή P_U . Σκοπός είναι, δοθέντος ότι $X = x$, να αντιστραφεί η αναφερθείσα σχέση ως $\theta = A_x^{-1}(U)$, ώστε να είναι εφικτή η εύρεση της κατανομής της άγνωστης παραμέτρου.

Ο έλεγχος μη κατωτερότητας (non-inferiority) της δοκιμαστικής διαδικασίας, διατυπώνεται, σύμφωνα με τους Chen et al. (2023), μέσω της

$$H_0 : p_{01} - p_{10} \geq \delta_0 \quad \text{έναντι} \quad H_1 : p_{01} - p_{10} < \delta_1, \quad (2.31)$$

όπου δ_0 μια προκαθορισμένη θετική τιμή, με $0 < \delta_0 < 1$, και δ_1 το προκαθορισμένο όριο μη κατωτερότητας ($\delta_1 < \delta_0$).

Σε όσα ακολουθούν, συμβολίζουμε $B_{a,b}$ την τυχαία μεταβλητή που ακολουθεί Βήτα κατανομή με παραμέτρους a και b . Γνωρίζουμε ότι η κατανομή της X_{01} είναι διωνυμική, με πλήθος επαναλήψεων n και πιθανότητα επιτυχίας p_{01} , και επομένως η fiducial κατανομή του p_{01} θα είναι:

$$p_{01} = B_{x_{01}+w_1, n-x_{01}+w_2},$$

όπου w_1, w_2 ανεξάρτητες προσομοιωμένες τιμές που ανήκουν στο διάστημα $(0,1)$. Επιπλέον, η δεσμευμένη κατανομή της τυχαίας μεταβλητής X_{10} δοθέντος της τιμής $X_{01} = x_{01}$ είναι διωνυμική με αριθμό επαναλήψεων $n - x_{01}$ και πιθανότητα επιτυχίας $\frac{p_{10}}{1-p_{01}}$, και η fiducial κατανομή της $p_{\theta} = \frac{p_{10}}{1-p_{01}}$ θα είναι:

$$p_{\theta} = B_{x_{10}+w_1, n-x_{10}-x_{01}+w_2}.$$

Η fiducial ποσότητα της παραμέτρου $\delta = p_{01} - p_{10} = p_{01} - p_{\theta}(1 - p_{01})$, μπορεί να προκύψει με αντικατάσταση των fiducial ποσοτήτων των πιθανοτήτων p_{01} και p_{10} ως εξής:

$$\delta = B_{x_{01}+w_1, n-x_{01}+w_2} - B_{x_{10}+w_1, n-x_{10}-x_{01}+w_2}(1 - B_{x_{01}+w_1, n-x_{01}+w_2}).$$

Για τον έλεγχο της υπόθεσης (2.31), η p -τιμή ορίζεται ως

$$p_f = P(\delta \geq \delta_0).$$

Επομένως, σε επίπεδο σημαντικότητας α , απορρίπτεται η μηδενική υπόθεση (2.31) όταν $p_f \leq \alpha$, και συνεπάγεται η μη κατωτερότητα της νέας δοκιμαστικής διαδικασίας.

Οι Chen et al. (2023) επισημαίνουν ότι είναι δύσκολος ο προσδιορισμός της p_f σε κλειστή μορφή. Για τον λόγο αυτό προτείνουν την εκτίμηση της μέσω Monte Carlo προσομοίωσης και παραθέτουν στο παράρτημα της εργασίας τους σχετικό κώδικα στη γλώσσα R. Επιπλέον, ένα εύλογο ερώτημα που ίσως έχει προκύψει αφορά την επιλογή των w_1 και w_2 , καθώς είναι προφανές ότι διαφορετικές επιλογές οδηγούν σε διαφορετικές στατιστικές συναρτήσεις ελέγχου. Οι Chen et al. (2023) χρησιμοποιώντας τα αποτελέσματα μίας εκτενούς μελέτης προσομοίωσης που διεξήγαγαν κατέληξαν στο συμπέρασμα ότι οι τιμές $w_1 = 0.7$ και $w_2 = 0.3$ οδηγούν σε έναν έλεγχο, ο οποίος διατηρεί το ονομαστικό επίπεδο σημαντικότητας του ελέγχου και έχει καλά αποτελέσματα ως προς την ισχύ του και είναι αυτός που προτείνεται να χρησιμοποιείται στην πράξη.

Στο πλαίσιο που ακολουθεί, γενικεύοντας τον κώδικα που δόθηκε από τους Chen et al. (2023), παραθέτουμε τη συνάρτηση της R για την υλοποίηση αυτού

του ελέγχου και την εφαρμογή του στα δεδομένα του Πίνακα 2.1. Επισημαίνουμε ότι στη συνάρτηση αυτή δηλώνουμε το πλήθος των ασύμβατων ζευγών, το συνολικό μέγεθος δείγματος, την τιμή δ_0 και το πλήθος των προσομοιώσεων για την εκτίμηση της p -τιμής στο όρισμα N .

```
library(rBeta2009)

fiducialtest <- function(x10, x01, n, delta0, N) {
  estimatep01 <- numeric(N)
  estimatep10 <- numeric(N)

  for (i in 1:N) {
    estimatep01[i] <- rbeta(1,x01 + 0.7, n - x01 + 0.3)
    estimatep10[i] <- rbeta(1, x10 + 0.7, n - x10 - x01 +
      ↪ 0.3) * (1 - estimatep01[i]) }

  pvalue73 <- sum((estimatep01 - estimatep10) >= delta0)
    ↪ / N

  return(pvalue73)
}

result<-fiducialtest(32,27,344,0.05,10000)
print(result)
[1] 0.0025
```

2.5 Οι έλεγχοι του Tango (1998)

Στην εργασία του Tango (1998), το ενδιαφέρον επικεντρώνεται στον έλεγχο της υπόθεσης ότι μια νέα διαγνωστική μέθοδος είναι το ίδιο καλή με μια υπάρχουσα, βάση της διαφοράς πιθανοτήτων θετικής διάγνωσης. Συγκεκριμένα, ο έλεγχος ισοδυναμίας διατυπώνεται ως έλεγχος της:

$$H_0 : p_{1.} - p_{.1} = -\delta_0 \quad \text{έναντι} \quad H_1 : p_{1.} - p_{.1} > -\delta_0, \quad (2.32)$$

όπου δ_0 είναι μια προκαθορισμένη θετική τιμή. Στη συνέχεια, θέτοντας $\beta = p_{1.} - (p_{.1} - \delta_0)$, άμεσα προκύπτει ότι ο έλεγχος της παραπάνω υπόθεσης ανάγεται στον έλεγχο της:

$$H_0 : \beta = 0 \quad \text{έναντι} \quad H_1 : \beta > 0. \quad (2.33)$$

Στο παραπάνω πλαίσιο, έστω ότι N_k και S_k είναι οι δίτιμες τυχαίες μεταβλητές που παριστάνουν την απόκριση του k -οστού ατόμου στο νέο και στον καθι-

ερωμένο διαγνωστικό έλεγχο, αντίστοιχα. Η κεντρική ιδέα του ελέγχου που προτάθηκε από τον Tango (1998) είναι ότι οι πιθανότητες απόκρισης του k -οστού ατόμου, τόσο με τη νέα διαγνωστική διαδικασία όσο και με την καθιερωμένη, εξαρτώνται από ένα μη παρατηρήσιμο χαρακτηριστικό, έστω θ_k , και έχουμε τις ακόλουθες δεσμευμένες πιθανότητες:

$$p_N(\theta_k) = P(N_k = 1|\theta_k) \text{ και } p_S(\theta_k) = P(S_k = 1|\theta_k).$$

Στη συνέχεια, υποθέτοντας ότι οι θ_k είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές και λαμβάνοντας υπόψη ότι

$$\mathbb{E}_\theta\{p_N(\theta)\} = p_{1.} \text{ και } \mathbb{E}_\theta\{p_S(\theta)\} = p_{.1},$$

$$Var_\theta\{p_N(\theta)\} = \sigma_{1.}^2 \text{ και } Var_\theta\{p_S(\theta)\} = \sigma_{.1}^2,$$

ενώ

$$\mathbb{E}_\theta\{p_N(\theta)p_S(\theta)\} = p_{1.}(\theta)p_{.1}(\theta) + \rho\sigma_{1.}\sigma_{.1},$$

με ρ να είναι ο συντελεστής συσχέτισης μεταξύ ασθενών, καταλήγει ότι τα διαθέσιμα δεδομένα μπορούν να μοντελοποιηθούν χρησιμοποιώντας πολυωνυμική κατανομή με πιθανότητες επιτυχίας:

$$p_{11} = p_{1.}p_{.1} + \phi,$$

$$p_{10} = p_{1.}(1 - p_{.1}) - \phi,$$

$$p_{01} = (1 - p_{1.})p_{.1} - \phi,$$

και

$$p_{00} = (1 - p_{1.})(1 - p_{.1}) + \phi,$$

με

$$\phi = \rho\sigma_{1.}\sigma_{.1}.$$

Επομένως, καθώς p_{11} , p_{10} , p_{01} και p_{00} είναι συναρτήσεις των $p_{1.}$, $p_{.1}$ και ϕ , ο λογάριθμος της συνάρτησης πιθανοφάνειας του διαθέσιμου δείγματος μπορεί να γραφεί στη μορφή

$$L(p_{1.}, p_{.1}, \phi) = C + x_{11}\log(p_{11}) + x_{10}\log(p_{10}) + x_{01}\log(p_{01}) + x_{00}\log(p_{00}),$$

με C σταθερά που δίνεται από τη σχέση:

$$C = \log\left(\frac{n!}{x_{11}!x_{10}!x_{01}!x_{00}!}\right).$$

Θέτοντας

$$\beta = p_{1\cdot} - (p_{\cdot 1} - \delta)$$

έχουμε ότι

$$L(\beta, p_{\cdot 1}, \phi) = C + x_{11} \log(p_{11}) + x_{10} \log(p_{10}) + x_{01} \log(p_{01}) + x_{00} \log(p_{00}),$$

όπου

$$\begin{aligned} p_{11} &= (\beta + p_{\cdot 1} - \delta_0)p_{\cdot 1} + \phi, \\ p_{10} &= (\beta + p_{\cdot 1} - \delta_0)(1 - p_{\cdot 1}) - \phi, \\ p_{01} &= (1 - \beta - p_{\cdot 1} + \delta_0)p_{\cdot 1} - \phi, \end{aligned}$$

και

$$p_{00} = (1 - \beta - p_{\cdot 1} + \delta_0)(1 - p_{\cdot 1}) + \phi.$$

Στη συνέχεια, ο Tango (1998) χρησιμοποιεί τη μέθοδο score για να προσδιορίσει τη στατιστική συνάρτηση του ελέγχου και ανάγει τον προσδιορισμό της στον υπολογισμό της ποσότητας:

$$\left. \frac{\partial L(\beta, \phi, p_{\cdot 1})}{\partial \phi} \right|_{p_{\cdot 1}=\hat{p}_{\cdot 1}, \phi=\hat{\phi}, \beta=0} \sqrt{\left. I_{(3,3)}^{-1} \right|_{p_{\cdot 1}=\hat{p}_{\cdot 1}, \phi=\hat{\phi}, \beta=0}}$$

όπου $\hat{p}_{\cdot 1}$ και $\hat{\phi}$ οι εκτιμητές μέγιστης πιθανοφάνειας των παραμέτρων $p_{\cdot 1}$ και ϕ υπό τη μηδενική υπόθεση, δηλαδή όταν $\beta = 0$, ενώ $I_{(3,3)}^{-1}$ είναι το στοιχείο της τρίτης γραμμής και τρίτης στήλης του αντίστροφου του πίνακα πληροφορίας του Fisher.

Ακολουθώντας την παραπάνω μεθοδολογία και έπειτα από αρκετές αλγεβρικές πράξεις που παρατίθενται στο παράρτημα της εργασίας του, ο Tango (1998) καταλήγει ότι η προτεινόμενη στατιστική συνάρτηση για τον έλεγχο της (2.33) είναι η:

$$T := T(x_{10}, x_{01}, n, \delta_0) = \frac{x_{10} - x_{01} + n\delta_0}{\sqrt{n(2\hat{p}_{01} - \delta_0(\delta_0 + 1))}}, \quad (2.34)$$

όπου

$$\hat{p}_{01} = \frac{\sqrt{b^2 - 4ac} - b}{2a},$$

με $a = 2n$, $b = -x_{10} - x_{01} - (2n - x_{10} + x_{01})\delta_0$ και $c = x_{01}\delta_0(\delta_0 + 1)$.

Αποδεικνύεται ότι η στατιστική συνάρτηση T ακολουθεί, υπό τη μηδενική υπόθεση, ασυμπτωτικά τυπική κανονική κατανομή και επομένως η μηδενική υπόθεση της σχέσης (2.33) απορρίπτεται, σε ασυμπτωτικό επίπεδο σημαντικότητας α , αν $T > z_\alpha$.

Στο πλαίσιο που ακολουθεί δίνεται η συνάρτηση της R για αυτόν τον έλεγχο και η υλοποίησή της για τα δεδομένα του Πίνακα 2.1.

```

Ttest<- function( x10,x01, n, delta0, alpha){
  a<- 2*n
  b<- -x10 - x01 - (2*n -x10+ x01)*delta0
  c<- x01* delta0* (delta0 +1)

  p01hat<- ( sqrt((b^2 - 4* a* c))- b)/ (2*a)

  T<- (x10 - x01 +n*delta0)/ (sqrt(n*(2* p01hat - delta0
    ↪ *(delta0+1))))

  pvalueT<- pnorm(T, lower.tail= FALSE)

  reject_H0 <- T> qnorm(1- alpha)
  return(list(Z_value= T,
              p_value= pvalueT,
              reject_H0= reject_H0))
}
result<- Ttest( 32, 27, 344, 0.05, 0.05)
print(result)
$Z_value
[1] 2.795403

$p_value
[1] 0.002591753

$reject_H0
[1] TRUE

```

ΚΕΦΑΛΑΙΟ 3

ΕΛΕΓΧΟΙ ΜΕ ΒΑΣΗ ΤΟ ΠΗΛΙΚΟ ΤΩΝ ΠΙΘΑΝΟΤΗΤΩΝ ΑΣΥΜΒΑΤΩΝ ΖΕΥΓΩΝ

Στο κεφάλαιο αυτό αντικείμενο μελέτης αποτελούν οι κυριότερες μεθοδολογίες ελέγχου της ισοδυναμίας και μη κατωτερότητας δύο διαγνωστικών ελέγχων όταν η παράμετρος δ εκφράζεται ως το πηλίκο των πιθανοτήτων εμφάνισης ασύμβατων ζευγαριών. Πιο συγκεκριμένα, σε όσα ακολουθούν η παράμετρος δ εκφράζει το πηλίκο της πληθυσμιακής πιθανότητας εμφάνισης του ζεύγους $(1, 0)$ προς την πληθυσμιακή πιθανότητα εμφάνισης του ζεύγους $(0, 1)$. Επομένως, ανακαλώντας τον συμβολισμό που δόθηκε στον Πίνακα 1.1 είναι $\delta = \frac{p_{10}}{p_{01}}$ και το ενδιαφέρον επικεντρώνεται στον έλεγχο μη κατωτερότητας που διατυπώνεται μέσω της

$$H_0 : \delta = \frac{p_{10}}{p_{01}} \leq \delta_0 \quad \text{έναντι} \quad H_1 : \delta > \delta_0, \quad (3.1)$$

όπου δ_0 το προκαθορισμένο όριο μη κατωτερότητας. Στο πλαίσιο αυτό, θα παρουσιαστούν οι έλεγχοι που έχουν προταθεί στις εργασίες των Liu et al. (2005), των Jin et al. (2016) και των Chen and Pan (2022). Οι παραπάνω έλεγχοι θα εφαρμοστούν, σε επίπεδο σημαντικότητας $\alpha = 0.05$, στα δεδομένα του Πίνακα 2.1, δημιουργώντας κατάλληλες συναρτήσεις στην R.

3.1 Οι έλεγχοι των Liu et al. (2005)

Στην ενότητα αυτή θα παρουσιασθούν οι δύο στατιστικές συναρτήσεις που έχουν προταθεί στην εργασία των Liu et al. (2005) για τον έλεγχο της υπόθεσης της σχέσης (3.1). Η πρώτη εξ αυτών προκύπτει ακολουθώντας παρόμοιο σκεπτικό με τον έλεγχο που προτάθηκε από τους Lu and Bean (1995) και στηρίζεται στη στατιστική συνάρτηση $Z_{LB}^{(1)}$, δηλαδή κατασκευάζεται θεωρώντας

το δειγματικό ανάλογο του πηλίκου $\delta = \frac{p_{10}}{p_{01}}$ και την απόκλιση αυτού από την τιμή δ_0 . Από την άλλη, η δεύτερη στατιστική συνάρτηση προκύπτει ακολουθώντας παρόμοιο σκεπτικό με αυτό του ελέγχου που προτάθηκε στην εργασία του Nam (1997). Δηλαδή, η δεύτερη προτεινόμενη στατιστική συνάρτηση χρησιμοποιεί τη λεγόμενη μέθοδο score, και στηρίζεται στο σκεπτικό της στατιστικής συνάρτησης Z_s . Ακολουθεί η παρουσίαση αυτών στο υπόλοιπο τμήμα αυτής της ενότητας.

Αρχικά, οι Liu et al. (2005) παρατηρούν ότι ο έλεγχος της σχέσης (3.1) μπορεί να αναχθεί στον έλεγχο της υπόθεσης

$$H_0 : \log(\delta) \leq \log(\delta_0) \quad \text{έναντι} \quad H_1 : \log(\delta) > \log(\delta_0), \quad (3.2)$$

καθώς ο φυσικός λογάριθμος είναι ένα προς ένα συνάρτηση που μετασχηματίζει το πηλίκο $\frac{p_{10}}{p_{01}}$ από το διάστημα $(0, \infty)$ στο διάστημα $(-\infty, \infty)$. Στη συνέχεια, προτείνουν η στατιστική συνάρτηση να βασίζεται στο εμπειρικό (δειγματικό) ανάλογο του λογαρίθμου του πηλίκου $\delta = \frac{p_{10}}{p_{01}}$, έστω $\log(\hat{\delta})$ και στην απόκλιση του από τον λογάριθμο της δοθείσας τιμής, δηλαδή από την τιμή $\log(\delta_0)$. Δηλαδή, προτείνουν ο έλεγχος να στηρίζεται στη στατιστική συνάρτηση $t = \log(\hat{\delta}) - \log(\delta_0)$. Σε αυτό το πλαίσιο, έχουμε ότι:

$$\log(\hat{\delta}) = \log(\hat{p}_{10}) - \log(\hat{p}_{01}) = \log(x_{10}) - \log(x_{01}). \quad (3.3)$$

Λαμβάνοντας υπόψη ότι:

$$\begin{pmatrix} x_{10} \\ x_{01} \end{pmatrix} \overset{\text{ασυμπτωτικά}}{\sim} N_2 \left[\begin{pmatrix} np_{10} \\ np_{01} \end{pmatrix}, \begin{pmatrix} np_{10}(1-p_{10}) & -np_{10}p_{01} \\ -np_{10}p_{01} & np_{01}(1-p_{01}) \end{pmatrix} \right]$$

η ασυμπτωτική κατανομή της τυχαίας μεταβλητής $\log(x_{10}) - \log(x_{01})$ μπορεί να προσδιοριστεί με εφαρμογή της πολυδιάστατης δέλτα μεθόδου. Ειδικότερα, αποδεικνύεται, ότι η τυχαία μεταβλητή $\log(x_{10}) - \log(x_{01})$ ακολουθεί, ασυμπτωτικά, κανονική κατανομή με μέση τιμή ίση με

$$\log(np_{10}) - \log(np_{01}) = \log(\delta),$$

και διακύμανση που προσδιορίζεται από το παρακάτω γινόμενο:

$$\begin{pmatrix} \frac{1}{np_{10}} & -\frac{1}{np_{01}} \end{pmatrix} \begin{pmatrix} np_{10}(1-p_{10}) & -np_{10}p_{01} \\ -np_{10}p_{01} & np_{01}(1-p_{01}) \end{pmatrix} \begin{pmatrix} \frac{1}{np_{10}} \\ -\frac{1}{np_{01}} \end{pmatrix},$$

το οποίο είναι ίσο με $\frac{1}{n} \left(\frac{1}{p_{10}} + \frac{1}{p_{01}} \right)$. Επομένως,

$$\log(\hat{\delta}) \overset{\text{ασυμπτωτικά}}{\sim} N \left(\log(\delta), \frac{1}{n} \left(\frac{1}{p_{10}} + \frac{1}{p_{01}} \right) \right),$$

οπότε, υπό τη μηδενική υπόθεση, έχουμε ότι:

$$\frac{\log(\hat{\delta}) - \log(\delta_0)}{\sqrt{\frac{1}{n} \left(\frac{1}{p_{10}} + \frac{1}{p_{01}} \right)}} \underset{\text{ασυμπτωτικά}}{\sim} N(0, 1).$$

Λαμβάνοντας υπόψη ότι οι άγνωστες πιθανότητες p_{10} και p_{01} εκτιμώνται από τις (συνεπείς) εκτιμήτριες $\hat{p}_{10} = \frac{x_{10}}{n}$ και $\hat{p}_{01} = \frac{x_{01}}{n}$, αντίστοιχα, έχουμε από το Λήμμα του Slutsky ότι υπό τη μηδενική υπόθεση

$$Z_b = \frac{\log(\hat{\delta}) - \log(\delta_0)}{\sqrt{\frac{1}{x_{10}} + \frac{1}{x_{01}}}} \underset{\text{ασυμπτωτικά}}{\sim} N(0, 1). \quad (3.4)$$

Επομένως, σε (ασυμπτωτικό) επίπεδο σημαντικότητας α , η μηδενική υπόθεση απορρίπτεται, και συμπεραίνουμε τη μη κατωτερότητα της νέας διαγνωστικής μεθόδου ως προς την καθιερωμένη, αν $Z_b \geq z_\alpha$.

Για την υλοποίηση αυτού του ελέγχου δημιουργήσαμε στη γλώσσα προγραμματισμού R μία συνάρτηση, η οποία λαμβάνει ως ορίσματα το πλήθος των ασύμβατων ζευγών, την τιμή δ_0 και το επίπεδο σημαντικότητας α του ελέγχου. Εφαρμόζοντας αυτήν τη συνάρτηση λαμβάνουμε ως αποτελέσματα την τιμή της στατιστικής συνάρτησης, την (ασυμπτωτική) p -τιμή του ελέγχου και μήνυμα αν απορρίπτεται ή όχι η μηδενική υπόθεση στο καθορισμένο επίπεδο σημαντικότητας.

```
liu_test4 <- function ( x10 , x01 , delta0 , alpha ) {
  delta_hat <- x10 / x01
  Z_b <- ( log ( delta_hat ) - log ( delta0 ) ) / (sqrt ((1
    ↪ / x10) + (1 / x01) ) )
  reject_H0 <- Z_b > qnorm (1 - alpha )

  pvalueZ_b <- pnorm ( Z_b , lower.tail = FALSE )
  return ( list ( Zvalue = Z_b , pvalue = pvalueZ_b ,
    ↪ reject_H0 = reject_H0 ) )
}

result <- liu_test4 (32 ,27 ,0.8 ,0.05)
print ( result )
$Zvalue
[1] 1.504078

$pvalue
[1] 0.06628059

$reject_H0
[1] FALSE
```

Παρατηρώντας ότι η στατιστική συνάρτηση Z_b δεν λαμβάνει υπόψη το περιθώριο ισοδυναμίας στην εκτίμηση της διακύμανσης του παρονομαστή, γεγονός που μπορεί να έχει αντίκτυπο στην απόδοσή της, οι Liu et al. (2005) πρότειναν και μία ακόμη στατιστική συνάρτηση, ακολουθώντας, παρακινούμενοι από τον Nam (1997), τη μέθοδο του score (δείτε Ενότητα 2.2). Αντικείμενο μελέτης του υπόλοιπου αυτής της ενότητας θα αποτελέσει ο τρόπος κατασκευής αυτής της στατιστικής συνάρτησης.

Ανακαλώντας, από τη σχέση (2.8), την από κοινού κατανομή των (x_{10}, x_{01}) , άμεσα έχουμε ότι ο λογάριθμος της συνάρτησης πιθανοφάνειας ακολουθώντας τη νέα παραμετροποίηση που προκύπτει θέτοντας $\frac{p_{10}}{p_{01}} = \delta$, δηλαδή θέτοντας $p_{10} = p_{01}\delta$, μπορεί να γραφεί ως:

$$\begin{aligned} L(p_{01}; \delta) &= \log(C) + x_{10} \log(\delta) + (x_{10} + x_{01}) \log(p_{01}) \\ &+ (n - x_{10} - x_{01}) \log((1 - p_{01} - \delta p_{01})), \end{aligned}$$

όπου τώρα η παράμετρος δ είναι η παράμετρος ενδιαφέροντος, ενώ η παράμετρος p_{01} είναι ενοχλητική παράμετρος. Για την εύρεση του εκτιμητή μέγιστης πιθανοφάνειας της παραμέτρου p_{01} για δοθείσα τιμή της παραμέτρου δ , υπολογίζουμε, αρχικά, την παράγωγο της συνάρτησης πιθανοφάνειας $L(p_{01}; \delta)$ ως προς p_{01} . Είναι

$$\frac{\partial L(p_{01}; \delta)}{\partial p_{01}} = \frac{x_{10} + x_{01}}{p_{01}} + \frac{n - x_{10} - x_{01}}{1 - \delta p_{01} - p_{01}}(-\delta - 1).$$

Στη συνέχεια, ο εκτιμητής μέγιστης πιθανοφάνειας της παραμέτρου p_{01} , έστω $\tilde{p}_{01}$, για δοθείσα τιμή της παραμέτρου δ , έστω δ_0 , προκύπτει από την επίλυση της εξίσωσης

$$\frac{x_{10} + x_{01}}{p_{01}} = \frac{n - x_{10} - x_{01}}{1 - (\delta_0 + 1)p_{01}}(\delta_0 + 1).$$

Μετά από λίγη άλγεβρα έχουμε ότι

$$\tilde{p}_{01} = \frac{x_{10} + x_{01}}{n(\delta_0 + 1)}. \quad (3.5)$$

Ακολουθώντας, για τον προσδιορισμό της στατιστικής συνάρτησης ελέγχου οι Liu et al. (2005) χρησιμοποιούν τη μέθοδο score, η οποία βασίζεται στην παράγωγο του λογαρίθμου της συνάρτησης πιθανοφάνειας (gradient of the loglikelihood function), όταν αυτή υπολογίζεται στην τιμή υπό τη μηδενική υπόθεση, έχοντας εκτιμήσει τις υπόλοιπες ενοχλητικές παραμέτρους (nuisance parameter), και η οποία δεν θα πρέπει να διαφέρει από το μηδέν περισσότερο από το δειγματικό

σφάλμα. Ειδικότερα, η στατιστική συνάρτηση ελέγχου της υπό μελέτης μη-δενικής υπόθεσης προσδιορίζεται από τη σχέση:

$$Z_{LFM} = \frac{S(\delta_0, \tilde{p}_{01})}{\sqrt{v(\delta_0, \tilde{p}_{01})}}, \quad (3.6)$$

όπου $S(\delta, p_{01})$ είναι η πρώτης τάξης παράγωγος του λογαρίθμου της συνάρτησης πιθανοφάνειας $L(p_{01}; \delta)$ ως προς δ , και

$$v(\delta, p_{01}) = I_{11}(\delta, p_{01}) - \frac{I_{12}^2(\delta, p_{01})}{I_{22}(\delta, p_{01})},$$

με $I_{ij}(\delta, p_{01})$ να είναι το (i, j) στοιχείο του πίνακα πληροφορίας του Fisher. Ειδικότερα, λαμβάνοντας υπόψη ότι:

$$S(\delta, p_{01}) = \frac{\partial L(p_{01}; \delta)}{\partial \delta} = \frac{x_{10}}{\delta} - p_{01} \frac{n - x_{10} - x_{01}}{1 - p_{01} - \delta p_{01}}, \quad (3.7)$$

έχουμε, μετά από λίγη άλγεβρα, ότι:

$$S(\delta_0, \tilde{p}_{01}) = \frac{x_{10} - \delta_0 x_{01}}{\delta_0(\delta_0 + 1)}. \quad (3.8)$$

Καθώς

$$I_{11}(\delta, p_{01}) = -E \left(\frac{\partial^2 L(p_{01}; \delta)}{\partial \delta^2} \right) = \frac{np_{10}}{\delta^2} + \frac{np_{01}^2(1 - p_{10} - p_{01})}{(1 - p_{01} - \delta p_{01})^2},$$

$$I_{12}(\delta, p_{01}) = -E \left(\frac{\partial^2 L(p_{01}; \delta)}{\partial p_{01} \partial \delta} \right) = \frac{n(1 - p_{10} - p_{01})}{(1 - p_{01} - \delta p_{01})^2},$$

και

$$I_{22}(\delta, p_{01}) = -E \left(\frac{\partial^2 L(p_{01}; \delta)}{\partial p_{01}^2} \right) = \frac{n(p_{10} + p_{01})}{p_{01}^2} - \frac{n(1 - p_{01} - p_{10})^2(1 + \delta)^2}{(1 - p_{01} - \delta p_{01})^2},$$

έχουμε ότι:

$$v(\delta_0, \tilde{p}_{01}) = \frac{n\tilde{p}_{01}}{\delta_0(\delta_0 + 1)}. \quad (3.9)$$

Από τα παραπάνω, με αντικατάσταση στη σχέση (3.6), προκύπτει ότι η προτεινόμενη στατιστική συνάρτηση δίνεται από τη σχέση:

$$Z_{LFM} = \frac{x_{10} - \delta_0 x_{01}}{\sqrt{\delta_0(x_{10} + x_{01})}}. \quad (3.10)$$

ΚΕΦΑΛΑΙΟ 3 3.2. Οι έλεγχοι των Jin et al. (2016) και Wang et al. (2018)

Η στατιστική συνάρτηση Z_{LFM} ακολουθεί, υπό τη μηδενική υπόθεση, ασυμπτωτικά τυπική κανονική κατανομή. Επομένως, με βάση τη στατιστική συνάρτηση Z_{LFM} της σχέσης (3.10) απορρίπτεται, σε ασυμπτωτικό επίπεδο σημαντικότητας α , η μηδενική υπόθεση της σχέσης (3.1) αν ικανοποιείται η κρίσιμη περιοχή $Z_{LFM} \geq z_\alpha$. Σε μία τέτοια περίπτωση, συμπεραίνουμε ότι η νέα διαδικασία δεν είναι (τουλάχιστον) χειρότερη από τη διαδικασία αναφοράς.

Παρατήρηση 3.1. Όταν $\delta_0 = 1$ τότε ο έλεγχος Z_{LFM} είναι ισοδύναμος με τον έλεγχο McNemar.

Στο πλαίσιο που ακολουθεί, παρατίθεται η συνάρτηση της R για αυτόν τον έλεγχο και η υλοποίησή του, για τα δεδομένα του Πίνακα 2.1. Δεν δίνονται περισσότερες λεπτομέρειες, καθώς η δομή της συνάρτησης είναι ίδια με αυτήν του προηγούμενου ελέγχου.

```
liu_test5 <- function(x10, x01, delta0 , alpha= 0.05){  
  Z_LFM<- (x10-delta0*x01)/sqrt(delta0*(x10+x01))  
  reject_H0 <- Z_LFM >= qnorm(1 - alpha)  
  
  pvalueZ_LFM<- pnorm(Z_LFM,lower.tail=FALSE)  
  return(list(Zvalue = Z_LFM, pvalue=pvalueZ_LFM,  
    ↪ reject_H0 = reject_H0))  
}  
result<-liu_test5(32,27,0.8,0.05)  
print(result)  
$Zvalue  
[1] 1.513779  
  
$pvalue  
[1] 0.06504101  
  
$reject_H0  
[1] FALSE
```

3.2 Οι έλεγχοι των Jin et al. (2016) και Wang et al. (2018)

Αντικείμενο μελέτης αυτής της ενότητας αποτελεί ο (ακριβής) έλεγχος που προτάθηκε στην εργασία των Jin et al. (2016), καθώς και η τροποποίησή του που προτάθηκε στην εργασία των Wang et al. (2018). Οι έλεγχοι αυτοί βασίζονται σε μία νέα θεωρητική προσέγγιση που έχει προταθεί στη βιβλιογραφία από

τους Martin and Liu (2013a,b) και αναφέρεται ως inferential model (επαγωγικό μοντέλο). Η νέα αυτή προσέγγιση βασίζεται στην πολύ απλή και διαισθητική ιδέα του να εντοπιστεί πρώτα η υποκείμενη πηγή αβεβαιότητας, που είναι μια απουσιάζουσα αλλά προβλέψιμη ποσότητα, και στη συνέχεια να γίνει στατιστική εκτίμηση κάνοντας προβλέψεις για αυτήν την ποσότητα με στατιστικά ακριβή τρόπο.

Ειδικότερα, η μέθοδος των επαγωγικών μοντέλων βασίζεται στη συσχέτιση ανάμεσα στα δεδομένα, τις παραμέτρους και μιας βοηθητικής μεταβλητής U . Στο πλαίσιο αυτό, έστω x η παρατηρούμενη τιμή της μεταβλητής, θ η άγνωστη παράμετρος που προσπαθούμε να εκτιμήσουμε και U μια τυχαία μεταβλητή που προέρχεται από γνωστή κατανομή P_U . Η U ονομάζεται βοηθητική μεταβλητή (auxiliary variable) και επιτρέπει την εύρεση μιας σχέσης μεταξύ των παρατηρήσεων x με την παράμετρο ενδιαφέροντος θ . Συνήθως $U \sim U(0,1)$ για διευκόλυνση των υπολογισμών. Η ιδέα λοιπόν των Martin and Liu (2013b), είναι η κατασκευή μιας σχέσης της μορφής $x = a(u^*, \theta)$, για κάποιες μη παρατηρήσιμες τιμές u^* της U . Αν ήταν γνωστή η πραγματική τιμή της u^* , φυσικά θα ήταν δυνατή η επίλυση της εξίσωσης $x = a(u^*, \theta)$ και μετέπειτα, η εύρεση της τιμής θ . Σε όσα έπονται, το $\Theta_x(u)$ αντιπροσωπεύει το καλύτερο δυνατό σύνολο λύσεων της παραμέτρου θ και προσδιορίζεται από τη σχέση:

$$\Theta_x(u) = [\theta : x = a(u, \theta)], x \in X, u \in U,$$

για δεδομένη τιμή u .

Πλέον είμαστε σε θέση να παραθέσουμε το σκεπτικό των Jin et al. (2016). Αρχικά, έστω X_{10} και X_{01} είναι οι δύο τυχαίες μεταβλητές που αντιστοιχούν στα παρατηρούμενα δεδομένα x_{10} και x_{01} , αντίστοιχα. Άμεσα προκύπτει (δείτε και τη σχέση (2.8)) ότι (X_{10}, X_{01}) είναι η ελάχιστη επαρκής στατιστική συνάρτηση για τις παραμέτρους (p_{10}, p_{01}) , μέσω των οποίων ορίζεται η παράμετρος ενδιαφέροντος δ . Επιπλέον, είναι

$$P(X_{10} = x_{10}, X_{01} = x_{01}) = P(X_{10} = x_{10}, X_{10} + X_{01} = x_{10} + x_{01}),$$

ενώ η πιθανότητα $P(X_{10} = x_{10}, X_{10} + X_{01} = x_{10} + x_{01})$ είναι ίση με

$$P(X_{10} + X_{01} = x_{10} + x_{01}) \cdot P(X_{10} = x_{10} | X_{10} + X_{01} = x_{10} + x_{01}).$$

Η κατανομή του αθροίσματος των τυχαίων μεταβλητών $X_{10} + X_{01}$ ακολουθεί διωνυμική κατανομή με συνολικό αριθμό επαναλήψεων n και πιθανότητα επιτυχίας $\lambda = p_{10} + p_{01}$. Από την άλλη, η δεσμευμένη κατανομή της τυχαίας μεταβλητής X_{10} δοθείσης της τιμής του αθροίσματος $X_{10} + X_{01}$ είναι διωνυμική με αριθμό επαναλήψεων $x_{10} + x_{01}$ και πιθανότητα επιτυχίας $\theta = \frac{p_{10}}{p_{10} + p_{01}} = \frac{\delta}{1 + \delta}$.

ΚΕΦΑΛΑΙΟ 3 3.2. Οι έλεγχοι των Jin et al. (2016) και Wang et al. (2018)

Από τα παραπάνω έχουμε ότι η απο κοινού συνάρτηση πιθανότητας των X_{10} και X_{01} , δηλαδή η πιθανότητα $P(X_{10} = x_{10}, X_{10} + X_{01} = x_{10} + x_{01})$ μπορεί να γραφεί:

$$\frac{n!}{(x_{10} + x_{01})!(n - x_{10} - x_{01})!} \lambda^{x_{10} + x_{01}} (1 - \lambda)^{n - x_{10} - x_{01}} \frac{(x_{10} + x_{01})!}{x_{10}!x_{01}!} \theta^{x_{10}} (1 - \theta)^{x_{01}}.$$

Η ανάλυση αυτή που περιγράφηκε από τους Jin et al. (2016), δίνει την ευκαιρία να χωρίσουμε τη συνάρτηση πιθανότητας σε δύο τμήματα: ένα τμήμα που εξαρτάται μόνο από την παράμετρο λ και τα παρατηρούμενα δεδομένα $x_{10} + x_{01}$ και ένα δεύτερο τμήμα, όπου αγνοώντας την παράμετρο λ , μπορούμε να εξαγάγουμε συμπεράσματα για την παράμετρο θ από τη δεσμευμένη κατανομή του X_{10} δοθέντος ότι γνωρίζουμε ότι $X_{10} + X_{01} = x_{10} + x_{01}$.

Επιπρόσθετα, οι Jin et al. (2016) παρατηρούν ότι οι ανισότητες $\delta \leq \delta_0$ και $\theta = \frac{\delta}{1+\delta} \leq \frac{\delta_0}{1+\delta_0} = \theta_0$ είναι ισοδύναμες και προτείνουν έναν IM έλεγχο (Inferential Model test), ο οποίος αποτελείται από τρία βήματα, επαναπροσδιορίζοντας τον έλεγχο (3.1) ως

$$H_0 : \theta \leq \theta_0 \quad \text{έναντι} \quad H_1 : \theta > \theta_0. \quad (3.11)$$

Στη συνέχεια θα περιγραφούν τα τρία βήματα της μεθόδου. Σε όσα ακολουθούν, με $F_{a,b}$ συμβολίζεται η αθροιστική συνάρτηση κατανομής της διωνυμικής κατανομής με παραμέτρους a, b και με $G_{a,b}$ η αθροιστική συνάρτηση κατανομής της Βήτα κατανομής με παραμέτρους a, b .

A-STEP (association)

Αξιοποιώντας τη βοηθητική μεταβλητή, η δειγματοληπτική επιλογή του x_{10} γίνεται ως εξής:

$$x_{10} = \min[k : u < F_{x_{10}+x_{01},\theta}(k)]. \quad (3.12)$$

Η παραπάνω σχέση επιτρέπει τη δημιουργία συνόλων πιθανών τιμών της θ , που συμβολίζονται ως $\Theta_{x_{10},x_{01}}(u)$, και προσδιορίζονται από τη σχέση:

$$\Theta_{x_{10},x_{01}}(u) = [\theta : F_{x_{10}+x_{01},\theta}(x_{10} - 1) \leq u \leq F_{x_{10}+x_{01},\theta}(x_{10})]. \quad (3.13)$$

Η συσχέτιση αυτή, είναι η αρχή του επαγωγικού μοντέλου, και παρέχει πληροφορία για το άγνωστο θ χωρίς την εκ των προτέρων γνώση κατανομής. Ανακαλώντας τις σχέσεις που συνδέουν την αθροιστική συνάρτηση κατανομής της διωνυμικής κατανομής με την αντίστοιχη της Βήτα κατανομής, δηλαδή ότι:

$$F_{x_{10}+x_{01},\theta}(x_{10}) = G_{x_{01},x_{10}+1}(1 - \theta),$$

και

$$F_{x_{10}+x_{01},\theta}(x_{10} - 1) = G_{x_{01}+1,x_{10}}(1 - \theta),$$

προκύπτει ότι:

$$\Theta_{x_{10},x_{01}}(u) = [\theta : 1 - G_{x_{01}+1,x_{10}}^{-1}(u) \leq \theta \leq 1 - G_{x_{10},x_{01}+1}^{-1}(u)]. \quad (3.14)$$

P-STEP (prediction)

Για την εύρεση ενός κανόνα που θα απορρίπτει τη μηδενική υπόθεση, το βέλτιστο προγνωστικό τυχαίο σύνολο (optimal predictive random set) του u είναι το:

$$S = [u : u \leq U], \quad (3.15)$$

όπου U ακολουθεί ομοιόμορφη κατανομή στο διάστημα $(0, 1)$.

C-STEP (combination)

Με πρόφαση την αρχική ιδέα της IM μεθόδου, δηλαδή την εξαγωγή συμπερασμάτων για την παράμετρο ενδιαφέροντος θ , λαμβάνοντας υπόψη τόσο τα δεδομένα, όσο και την αβεβαιότητα που εισάγεται με το στοχαστικό σύνολο S , οι Jin et al. (2016) στο τελευταίο βήμα υπολόγισαν το σύνολο $\Theta_{x_{10},x_{01}}(S)$. Συγκεκριμένα, συνδυάζοντας τα αποτελέσματα της σχέσης (3.14), η οποία περιγράφει το εύρος τιμών της θ για ένα συγκεκριμένο u , και τη σχέση (3.15), που ορίζει το σύνολο από πιθανές τιμές του u , παράγεται το:

$$\Theta_{x_{10},x_{01}}(S) = [\theta : \theta \geq 1 - G_{x_{01}+1,x_{10}}^{-1}(V)]. \quad (3.16)$$

Έπειτα, για τον έλεγχο της υπόθεσης $H_0 : \theta \leq \theta_0$ έναντι της $H_1 : \theta > \theta_0$, υποστηρίχθηκε ότι ένας τρόπος για την αποδοχή ή μη αυτής, είναι η συνάρτηση ευπιστίας (plausibility):

$$\begin{aligned} pl_{x_{10},x_{01}}(\mathcal{A}) &= 1 - P[\Theta_{x_{10},x_{01}}(S) \subseteq \mathcal{A}^c] \\ &= 1 - P[\theta_0 < 1 - G_{x_{01}+1,x_{10}}^{-1}(V)] \\ &= 1 - G_{x_{01}+1,x_{10}}(1 - \theta_0) = 1 - F_{x_{10}+x_{01},\theta_0}(x_{10} - 1). \end{aligned}$$

Επομένως, σε επίπεδο σημαντικότητας α , η μηδενική υπόθεση (3.11) κατωτερότητας της νέας διαδικασίας, απορρίπτεται αν:

$$pl_{x_{10},x_{01}}(\mathcal{A}) = 1 - F_{x_{10}+x_{01},\theta_0}(x_{10} - 1) \leq \alpha. \quad (3.17)$$

Στο πλαίσιο που ακολουθεί, παρατίθεται η συνάρτηση της R για αυτόν τον έλεγχο και η υλοποίησή του, για τα δεδομένα του Πίνακα 2.1. Δεν δίνονται περισσότερες λεπτομέρειες, καθώς η δομή της συνάρτησης είναι ίδια με αυτήν του προηγούμενου ελέγχου.

```
im_test <- function ( x10 , x01 , delta0 , alpha ) {
  theta0 <- delta0 / ( 1 + delta0 )
```


ΚΕΦΑΛΑΙΟ 3 3.2. Οι έλεγχοι των Jin et al. (2016) και Wang et al. (2018)

```
pl_A <- 1 - pbeta (1 - theta0 , x01 + 1 , x10 )
reject_H0 <- pl_A <= alpha
return ( list ( pvalue = pl_A , reject_H0= reject_H0 ) )
}
result <- im_test (32 ,27 ,0.8 ,0.05)
print ( result )
$pvalue
[1] 0.08376834

$reject_H0
[1] FALSE
```

Στην εργασία των Wang et al. (2018) προτάθηκε τροποποίηση του παραπάνω IM ελέγχου. Ειδικότερα, σύμφωνα με αυτήν, η μηδενική υπόθεση κατωτερότητας της νέας διαδικασίας που διατυπώνεται στη σχέση (3.11), απορρίπτεται αν:

$$rpl_{x_{10},x_{01}}(\mathcal{A}) = 1 - (wF_{x_{10}+x_{01},\theta_0}(x_{10} - 1) + (1 - w)F_{x_{10}+x_{01},\theta_0}(x_{10})) \leq \alpha,$$

όπου w είναι μια τυχαία προσομοιωμένη τιμή που ανήκει στο διάστημα (0,1). Οι Wang et al. (2018) παραθέτουν κώδικα υλοποίησης στην R του ελέγχου που πρότειναν για την υλοποίηση ενός αριθμητικού παραδείγματος. Ο κώδικας αυτός γενικεύεται στο πλαίσιο που ακολουθεί, ενώ υλοποιείται για τα δεδομένα του Πίνακα 2.1.

```
im_test_modify <- function ( x10 , x01 , delta0 , alpha )
{
  theta0 <- delta0 / (1 + delta0)
  w <- runif (1 ,0 ,1)
  pl_Amodify <- 1 - w * pbinom ( x10 -1 , x10 + x01 ,
    ↪ theta0 ) - (1 - w ) * pbinom ( x10 , x10 +x01 ,
    ↪ theta0 )
  reject_H0 <- pl_Amodify <= alpha
  return ( list ( pvalue = pl_Amodify , reject_H0=
    ↪ reject_H0 ) )
}
result <- im_test_modify (32 ,27 ,0.8 ,0.05)
print ( result )
$pvalue
[1] 0.08045141

$reject_H0
[1] FALSE
```

Σε περίπτωση που θέλουμε να σταθεροποιήσουμε την τιμή w , τότε η παραπάνω

συνάρτηση τροποποιείται έχοντας την τιμή αυτή ως όρισμα της συνάρτησης. Ειδικότερα, έχουμε την ακόλουθη συνάρτηση:

```
im_test_modify2 <- function ( x10 , x01 , delta0 , w ,
  ↪ alpha )
{
  theta0 <- delta0 / (1 + delta0 )
  pl_Amodify2 <- 1 - w * pbinom ( x10 - 1 , x10 + x01 ,
    ↪ theta0 ) - (1 - w ) * pbinom ( x10 , x10 + x01 ,
    ↪ theta0 )
  reject_H0 <- pl_Amodify2 <= alpha
  return ( list ( pvalue = pl_Amodify2 , reject_H0 =
    ↪ reject_H0 ) )
}

result <- im_test_modify2 (32 , 27 , 0.8 , 0.3 , 0.05)
print ( result )
$pvalue
[1] 0.06044618

$reject_H0
[1] FALSE
```

3.3 Οι έλεγχοι των Chen and Pan (2022)

Στο πλαίσιο του ελέγχου μη κατωτερότητας με βάση το πηλίκο πιθανοτήτων των ασύμβατων ζευγών, οι Chen and Pan (2022) εφαρμόζουν τη μέθοδο των Fiducial Test, η οποία περιγράφηκε, διαισθητικά, στην Ενότητα 2.4. Πιο συγκεκριμένα, γνωρίζοντας ότι η κατανομή της X_{01} είναι διωνυμική, με πλήθος επαναλήψεων n και πιθανότητα επιτυχίας p_{01} , οι συγγραφείς προσπαθούν να προσδιορίσουν την κατανομή της παραμέτρου ενδιαφέροντος θ . Επιπλέον, η δεσμευμένη κατανομή της τυχαίας μεταβλητής X_{10} δοθέντος της τιμής του αθροίσματος $X_{10} + X_{01}$ είναι διωνυμική με πλήθος επαναλήψεων ίσο με $x_{10} + x_{01}$ και πιθανότητα επιτυχίας $\theta = \frac{\delta}{\delta+1}$.

Η fiducial κατανομή της θ , προέρχεται από την κατανομή Βήτα που, είναι γνωστό ότι, λειτουργεί ως συζυγής για την εκ των προτέρων διωνυμική πιθανότητα. Ωστόσο, αντί να χρησιμοποιηθεί η Μπεϋζιανή προσέγγιση, η fiducial μέθοδος αξιοποιεί μια κατανομή που εξαρτάται μόνο από τα δεδομένα και σύμφωνα με την οποία:

$$\theta \sim \text{Beta}(x_{10} + 0.5, x_{01} + 0.5).$$

Η p -τιμή για τον έλεγχο της μηδενικής υπόθεσης $\theta \leq \theta_0$, υπολογίζεται ως

$$p_f = P(\theta \leq \theta_0) = F_{x_{10}+0.5, x_{01}+0.5}(\theta_0),$$

όπου $F_{x_{10}+0.5, x_{01}+0.5}$ η αθροιστική συνάρτηση κατανομής της Βήτα κατανομής με παραμέτρους $x_{10}+0.5, x_{01}+0.5$. Η μηδενική υπόθεση του ελέγχου κατωτερότητας που διατυπώθηκε στη σχέση (3.11), απορρίπτεται σε επίπεδο σημαντικότητας α , όταν $p_f \leq \alpha$. Σε μία τέτοια περίπτωση εξάγουμε το συμπέρασμα ότι η νέα διαδικασία δεν είναι (τουλάχιστον) χειρότερη από τη διαδικασία αναφοράς.

Στο πλαίσιο που ακολουθεί παρατίθεται η συνάρτηση της R για αυτόν τον έλεγχο και η υλοποίησή του για τα δεδομένα του Πίνακα 2.1.

```
fiducialtestratio <- function ( x10 , x01 , delta0 ,
  ↪ alpha )
{
  theta0 <- delta0 / (1 + delta0)

  pvaluefid = pbeta ( theta0 , x10 +0.5 , x01 +0.5)
  reject_H0 <- pvaluefid <= alpha
  return ( list ( pvalue = pvaluefid , reject_H0=
    ↪ reject_H0 ) )
}
result <- fiducialtestratio (32 ,27 , 0.8 ,0.05)
print ( result )
$pvalue
[1] 0.06548305

$reject_H0
[1] FALSE
```

ΚΕΦΑΛΑΙΟ 4

ΕΛΕΓΧΟΙ ΜΕ ΒΑΣΗ ΤΟ ΠΗΛΙΚΟ ΤΩΝ ΠΙΘΑΝΟΤΗΤΩΝ ΘΕΤΙΚΗΣ ΔΙΑΓΝΩΣΗΣ

Στο κεφάλαιο αυτό, αντικείμενο μελέτης αποτελούν οι κυριότερες μεθοδολογίες ελέγχου της ισοδυναμίας και μη κατωτερότητας δύο διαγνωστικών ελέγχων όταν η παράμετρος δ είναι ίση με το πηλίκο των πιθανοτήτων θετικής διάγνωσης. Πιο συγκεκριμένα, και ανακαλώντας τον συμβολισμό που δόθηκε στον Πίνακα 1.1, σε όσα ακολουθούν σε αυτό το κεφάλαιο είναι $\delta = \frac{p_1}{p_{.1}}$.

Στο πλαίσιο αυτό, θα παρουσιαστούν οι έλεγχοι που έχουν προταθεί στις εργασίες των Lachenbruch and Lynch (1998), των Nam and Blackwelder (2002) και των Tang et al. (2003). Οι παραπάνω έλεγχοι θα εφαρμοστούν, σε επίπεδο σημαντικότητας $\alpha = 0.05$, στα δεδομένα του Πίνακα 2.1, δημιουργώντας κατάλληλες συναρτήσεις στην R.

4.1 Οι έλεγχοι των Lachenbruch and Lynch (1998)

Στην ενότητα αυτή το ενδιαφέρον επικεντρώνεται στους ελέγχους που προτάθηκαν από τους Lachenbruch and Lynch (1998), για τη σύγκριση μίας νέας διαγνωστικής μεθόδου με μία παραδοσιακή, καθιερωμένη μέθοδο, χωρίς να είναι διαθέσιμη η πραγματική νοσολογική κατάσταση του ασθενή από τον χρυσό κανόνα. Στο πλαίσιο αυτό, οι Lachenbruch and Lynch (1998) αντιμετωπίζουν το πρόβλημα της ισοδυναμίας των δύο μεθόδων με όρους του ποσοστού θετικών αποκρίσεων. Ειδικότερα, θεωρούν ότι οι δύο διαγνωστικές μέθοδοι είναι ισοδύναμες αν ο λόγος των θετικών αποκρίσεων τους βρίσκεται εντός ενός προκαθορισμένου αποδεκτού εύρους από δ_0 ως δ_1 , όπου $\delta_0 < 1$ και $\delta_1 > 1$. Ενδεικτικά, συνήθως, επιλέγεται $\delta_0 = 0.9$ και $\delta_1 = 1.1$ ή $\delta_1 = \frac{1}{\delta_0}$. Επομένως, οι Lachenbruch and Lynch (1998) αναγάγουν το πρόβλημα της ισοδυναμίας των

δύο διαγνωστικών μεθόδων στον έλεγχο της σύνθετης υπόθεσης:

$$H_0 : p_{1.} \leq \delta_0 p_{.1} \quad \text{ή} \quad p_{1.} \geq \delta_1 p_{.1} \quad \text{έναντι} \quad H_1 : \delta_0 p_{.1} < p_{1.} < \delta_1 p_{.1}.$$

Ανακαλώντας ότι

$$p_{1.} = p_{11} + p_{10} \quad \text{και} \quad p_{.1} = p_{01} + p_{11}$$

η παραπάνω σύνθετη υπόθεση ισοδύναμα διατυπώνεται ως:

$$H_0 : \frac{p_{11} + p_{10}}{p_{11} + p_{01}} \leq \delta_0 \quad \text{ή} \quad \frac{p_{11} + p_{10}}{p_{11} + p_{01}} \geq \delta_1, \quad \text{έναντι} \quad H_1 : \delta_0 < \frac{p_{11} + p_{10}}{p_{11} + p_{01}} < \delta_1. \quad (4.1)$$

Η παραπάνω σύνθετη υπόθεση μπορεί να διασπαστεί στα ακόλουθα δύο σύνολα μονόπλευρων υποθέσεων:

$$H_{0\ell} : \frac{p_{11} + p_{10}}{p_{11} + p_{01}} \leq \delta_0 \quad \text{έναντι} \quad H_{1\ell} : \frac{p_{11} + p_{10}}{p_{11} + p_{01}} > \delta_0 \quad (4.2)$$

και

$$H_{0u} : \frac{p_{11} + p_{10}}{p_{11} + p_{01}} \geq \delta_1 \quad \text{έναντι} \quad H_{1u} : \frac{p_{11} + p_{10}}{p_{11} + p_{01}} < \delta_1. \quad (4.3)$$

Αρχικά, οι Lachenbruch and Lynch (1998) εξετάζουν την υπόθεση της σχέσης (4.2), στην οποία, ουσιαστικά, διατυπώνεται ο έλεγχος μη κατωτερότητας. Συγκεκριμένα, προτείνουν ο έλεγχος αυτής της υπόθεσης να βασίζεται στην απόκλιση του δειγματικού ανάλογου του πληθυσμίου $\frac{p_{11} + p_{10}}{p_{11} + p_{01}}$, έστω $\hat{\delta}$, από τη δοθείσα τιμή δ_0 ή ισοδύναμα από την απόκλιση του δειγματικού ανάλογου της ποσότητας $p_{1.} = p_{11} + p_{10}$ από το γινόμενο του δειγματικού ανάλογου της ποσότητας $p_{.1} = p_{01} + p_{11}$ με τη σταθερά δ_0 . Επομένως, προτείνουν ο έλεγχος να στηριχθεί στη στατιστική συνάρτηση L_0 που προσδιορίζεται από τη σχέση:

$$L_0 = \hat{p}_{11} + \hat{p}_{10} - \delta_0(\hat{p}_{11} + \hat{p}_{01}) = \hat{p}_{10} + (1 - \delta_0)\hat{p}_{11} - \delta_0\hat{p}_{01}. \quad (4.4)$$

Για τη διεξαγωγή του ελέγχου το επόμενο βήμα είναι ο προσδιορισμός της κατανομής της στατιστικής συνάρτησης L_0 υπό τη μηδενική υπόθεση. Αρχικά, ανακαλώντας ότι $\mathbb{E}(x_{11}) = np_{11}$, $\mathbb{E}(x_{10}) = np_{10}$ και $\mathbb{E}(x_{01}) = np_{01}$, άμεσα έχουμε ότι υπό τη μηδενική υπόθεση ισχύει ότι $\mathbb{E}_0(L_0) = 0$, με $\mathbb{E}_0$ να συμβολίζει τη μέση τιμή υπό τη μηδενική υπόθεση. Επιπλέον, ισχύει ότι:

$$\begin{aligned} \text{Var}(L_0) &= \text{Var}(\hat{p}_{10} + (1 - \delta_0)\hat{p}_{11} - \delta_0\hat{p}_{01}) \\ &= \frac{1}{n^2} \text{Var}(x_{10} + (1 - \delta_0)x_{11} - \delta_0x_{01}), \end{aligned}$$

οπότε

$$\begin{aligned} n^2 Var(L_0) &= Var(x_{10}) + (1 - \delta_0)^2 Var(x_{11}) + \delta_0^2 Var(x_{01}) \\ &+ 2(1 - \delta_0)Cov(x_{10}, x_{11}) - 2\delta_0 Cov(x_{10}, x_{01}) \\ &- 2(1 - \delta_0)\delta_0 Cov(x_{11}, x_{01}). \end{aligned}$$

Λαμβάνοντας υπόψη ότι:

$$Var(x_{10}) = np_{10}(1 - p_{10}), Var(x_{11}) = np_{11}(1 - p_{11}), Var(x_{01}) = np_{01}(1 - p_{01}),$$

αλλά και

$$\begin{aligned} Cov(x_{10}, x_{11}) &= -np_{10}p_{11} \quad Cov(x_{10}, x_{01}) = -np_{10}p_{01}, \\ Cov(x_{11}, x_{01}) &= -np_{11}p_{01}, \end{aligned}$$

έπειτα από απλές αλγεβρικές πράξεις έχουμε ότι:

$$\begin{aligned} nVar_0(L_0) &= p_{10}(1 - p_{10}) + (1 - \delta_0)^2 p_{11}(1 - p_{11}) + \delta_0^2 p_{01}(1 - p_{01}) \\ &+ 2\delta_0 p_{10}p_{01} + 2\delta_0(1 - \delta_0)p_{01}p_{11} - 2(1 - \delta_0)p_{10}p_{11}. \end{aligned}$$

Είναι προφανές ότι η παραπάνω έκφραση της διακύμανσης δεν μπορεί να χρησιμοποιηθεί στην πράξη, καθώς περιέχει τις άγνωστες παραμέτρους p_{10} και p_{01} . Η παρατήρηση αυτή οδήγησε τους Lachenbruch and Lynch (1998) να προτείνουν την αντικατάσταση των άγνωστων παραμέτρων από τις εκτιμήτριες τους $\hat{p}_{10}$ και $\hat{p}_{01}$, αντίστοιχα. Ως εκ τούτου, η στατιστική συνάρτηση που προτάθηκε για τον έλεγχο της υπόθεσης (4.2), δίνεται από τη σχέση:

$$Z_\ell = \frac{L_0}{S_0} = \frac{\hat{p}_{11} + \hat{p}_{10} - \delta_0(\hat{p}_{11} + \hat{p}_{01})}{S_0}, \quad (4.5)$$

όπου

$$\begin{aligned} nS_0^2 &= \hat{p}_{10}(1 - \hat{p}_{10}) + (1 - \delta_0)^2 \hat{p}_{11}(1 - \hat{p}_{11}) + \delta_0^2 \hat{p}_{01}(1 - \hat{p}_{01}) \\ &+ 2\delta_0 \hat{p}_{10}\hat{p}_{01} + 2\delta_0(1 - \delta_0)\hat{p}_{01}\hat{p}_{11} - 2(1 - \delta_0)\hat{p}_{10}\hat{p}_{11}. \end{aligned}$$

Καθώς αποδεικνύεται ότι η στατιστική συνάρτηση Z_ℓ ακολουθεί ασυμπτωτικά, υπό τη μηδενική υπόθεση, τυπική κανονική κατανομή, έχουμε ότι, σε ασυμπτωτικό επίπεδο σημαντικότητας α , απορρίπτεται η μηδενική υπόθεση της σχέσης (4.2), αν $Z_\ell \geq z_\alpha$.

Όσον αφορά τον έλεγχο της υπόθεσης (4.3), με παρόμοιο σκεπτικό οι Lachenbruch and Lynch (1998) πρότειναν να στηρίζεται στη στατιστική συνάρτηση

$$L_1 = \hat{p}_{11} + \hat{p}_{10} - \delta_1(\hat{p}_{11} + \hat{p}_{01}) = \hat{p}_{10} + (1 - \delta_1)\hat{p}_{11} - \delta_1\hat{p}_{01}. \quad (4.6)$$

Παρατηρώντας τις σχέσεις (4.4) και (4.6) συμπεραίνουμε άμεσα ότι ο προσδιορισμός της κατανομής της στατιστικής συνάρτησης L_1 υπό τη μηδενική υπόθεση είναι παρόμοιος με αυτόν της στατιστικής συνάρτησης L_0 , με τη μόνη αλλαγή να έγκειται στη χρήση της παραμέτρου δ_1 αντί της δ_0 . Επομένως, με παρόμοια διαδικασία, η στατιστική συνάρτηση που προκύπτει για τον έλεγχο της υπόθεσης (4.3) δίνεται από τη σχέση:

$$Z_u = \frac{L_1}{S_1} = \frac{\hat{p}_{11} + \hat{p}_{10} - \delta_1(\hat{p}_{11} + \hat{p}_{01})}{S_1}, \quad (4.7)$$

όπου

$$\begin{aligned} nS_1^2 &= \hat{p}_{10}(1 - \hat{p}_{10}) + (1 - \delta_1)^2 \hat{p}_{11}(1 - \hat{p}_{11}) + \delta_1^2 \hat{p}_{01}(1 - \hat{p}_{01}) \\ &+ 2\delta_1 \hat{p}_{10} \hat{p}_{01} + 2\delta_1(1 - \delta_1) \hat{p}_{01} \hat{p}_{11} - 2(1 - \delta_1) \hat{p}_{10} \hat{p}_{11}. \end{aligned}$$

Επομένως, λαμβάνοντας επιπρόσθετα υπόψη ότι, υπό τη μηδενική υπόθεση, η στατιστική συνάρτηση Z_u ακολουθεί ασυμπτωτικά τυπική κανονική κατανομή, έχουμε ότι, σε ασυμπτωτικό επίπεδο σημαντικότητας α , απορρίπτεται η μηδενική υπόθεση της σχέσης (4.3), αν $Z_u \leq -z_\alpha$.

Λαμβάνοντας υπόψη τα παραπάνω, η σύνθετη μηδενική υπόθεση της σχέσης (4.1), δηλαδή του ελέγχου της υπόθεσης της ισοδυναμίας των δύο διαγνωστικών μεθόδων, απορρίπτεται, και διαπιστώνουμε ότι οι δύο διαδικασίες είναι ισοδύναμες, αν $Z_\ell \geq z_\alpha$ και $Z_u \leq -z_\alpha$, δηλαδή αν οι έλεγχοι που διατυπώθηκαν στις σχέσεις (4.2) και (4.3) απορρίπτονται ταυτόχρονα.

Για την υλοποίηση αυτού του ελέγχου δημιουργήσαμε στη γλώσσα προγραμματισμού R μία συνάρτηση, η οποία λαμβάνει ως ορίσματα το πλήθος των ασύμβατων ζευγών, το πλήθος του ζεύγους (+, +), το συνολικό μέγεθος του δείγματος, τις τιμές των $\delta_0 > 0$, και δ_1 , το επίπεδο σημαντικότητας α του ελέγχου, καθώς και ο τύπος του ελέγχου. Ειδικότερα, στο όρισμα type δίνεται η δυνατότητα να δηλώσουμε αν θέλουμε να ελέγξουμε την υπόθεση της σχέσης (4.2) γράφοντας type=noninferiority, αν θέλουμε να ελέγξουμε την υπόθεση της σχέσης (4.3) γράφοντας type=nonsuperiority ή την υπόθεση της σχέσης (4.1) γράφοντας type=equivalence. Εφαρμόζοντας αυτήν τη συνάρτηση λαμβάνουμε ως αποτελέσματα την τιμή της στατιστικής συνάρτησης, την (ασυμπτωτική) p -τιμή του ελέγχου και μήνυμα αν απορρίπτεται ή όχι η μηδενική υπόθεση που προσδιορίσαμε στο όρισμα type, στο καθορισμένο επίπεδο σημαντικότητας, καθώς και μήνυμα που αφορά τη μορφή του ελέγχου που διενεργήθηκε. Στο πλαίσιο που ακολουθεί, παρατίθεται η συνάρτηση αυτή και η υλοποίησή της, για την περίπτωση του ελέγχου ισοδυναμίας, για τα δεδομένα του Πίνακα 2.1.

```

lachentest <- function(x10, x01, x11, n, delta0, delta1,
  ↪ alpha, type) {

  p11hat<- x11/n
  p10hat<- x10/n
  p01hat<- x01/n

  L0<- p10hat+(1-delta0)*p11hat-delta0*p01hat
  L1<- p10hat+(1-delta1)*p11hat-delta1*p01hat

  s0<- sqrt((p10hat*(1-p10hat)+(1-delta0)^2*p11hat*(1-
  ↪ p11hat)+delta0^2*p01hat*(1-p01hat)+2*delta0* p10hat
  ↪ *p01hat+2*delta0* (1-delta0)*p11hat*p01hat-2*(1-
  ↪ delta0)*p11hat*p10hat)/n)

  s1<- sqrt((p10hat*(1-p10hat)+(1-delta1)^2*p11hat*(1-
  ↪ p11hat)+delta1^2*p01hat*(1-p01hat)
+ +2*delta1* p10hat*p01hat+2*delta1*(1-delta1)*p11hat*
  ↪ p01hat-2*(1-delta1)*p11hat*p10hat)/n)

  Z_ell <- L0/s0
  Z_u   <- L1/s1

  if (type == "noninferiority") {
    p_value <- pnorm(Z_ell, lower.tail = FALSE)
    reject_H0 <- Z_ell >= qnorm(1 - alpha)
    Z_value <- Z_ell

  } else if (type == "nonsuperiority") {
    p_value <- pnorm(Z_u, lower.tail = TRUE)
    reject_H0 <- Z_u <= -qnorm(1 - alpha)
    Z_value <- Z_u

  } else if (type == "equivalence") {
    p_value <- max(pnorm(Z_ell, lower.tail = FALSE),
  ↪ pnorm(Z_u, lower.tail = TRUE))
    reject_H0 <- (Z_ell >= qnorm(1 - alpha)) & (Z_u <= -
  ↪ qnorm(1 - alpha))
    Z_value <- c(Z_ell, Z_u)

  } else {
    stop("The 'alternative' argument must be either '
  ↪ noninferiority', 'nonsuperiority', or '
  ↪ equivalence'")
  }
}

```



```

    }

    return(list(
      Z_value = Z_value,
      p_value = p_value,
      reject_H0 = reject_H0,
      type = type
    ))
  }

result <- lachentest(32, 27, 231, 344, 0.95, 1.05, 0.05,
  ↪ type="equivalence")
print(result)
$Z_value
[1] 2.391698 -1.001980

$p_value
[1] 0.1581766

$reject_H0
[1] FALSE

$type
[1] "equivalence"

```

Επιπρόσθετα, οι Lachenbruch and Lynch (1998) πρότειναν ένα διάστημα εμπιστοσύνης για την άγνωστη παράμετρο δ χρησιμοποιώντας τον εκτιμητή σε σημείο της παραμέτρου $\delta = \frac{p_{1.}}{p_{.1}}$ και προσδιορίζοντας την ασυμπτωτική κατανομή του με τη μέθοδο δέλτα. Τα αποτελέσματα αυτά, όπως επισημαίνουν οι Tang et al. (2003), μπορούν να χρησιμοποιηθούν και στον έλεγχο υποθέσεων. Ειδικότερα, οι Lachenbruch and Lynch (1998) απέδειξαν ότι ο προτεινόμενος εκτιμητής

$$\hat{\delta} = \frac{x_{11} + x_{10}}{x_{11} + x_{01}}$$

ακολουθεί ασυμπτωτικά κανονική κατανομή με μέση τιμή δ και διακύμανση που μπορεί να εκτιμηθεί από τη σχέση

$$\hat{Var}(\hat{\delta}) = \frac{(x_{11} + x_{10})(x_{10} + x_{01})}{(x_{11} + x_{01})^3}.$$

Επομένως, προτείνεται να διεξάγεται ο έλεγχος της υπόθεσης (4.2) με χρήση

της στατιστικής συνάρτησης

$$T_{LL} = \frac{\hat{\delta} - \delta_0}{\sqrt{\hat{Var}(\hat{\delta})}} = \frac{(x_{11} + x_{10}) - \delta_0(x_{11} + x_{01})}{\sqrt{\frac{(x_{11} + x_{10})(x_{10} + x_{01})}{x_{11} + x_{01}}}}. \quad (4.8)$$

Καθώς η στατιστική συνάρτηση T_{LL} ακολουθεί, υπό τη μηδενική υπόθεση, ασυμπτωτικά τυπική κανονική κατανομή, απορρίπτεται η μηδενική υπόθεση, σε ασυμπτωτικό επίπεδο σημαντικότητας α , αν $T_{LL} \geq z_\alpha$.

Στο πλαίσιο που ακολουθεί δίνεται η συνάρτηση της R για αυτόν τον έλεγχο και η υλοποίησή της για τα δεδομένα του Πίνακα 2.1.

```
lachen2test<- function(x11, x10, x01, delta0, alpha){
  deltahat<- (x11+ x10)/(x11+ x01)
  T_LL<- ((x11+x10)- delta0*(x11+ x01))/(sqrt(((x11+x10)
  ↪ *(x10+x01))/(x11+x01)))

  pvalueT_LL <- pnorm(T_LL, lower.tail = FALSE)
  reject_H0 <- T_LL >= qnorm(1- alpha)

  return(list(Z_value= T_LL,
              p_value= pvalueT_LL,
              reject_H0= reject_H0))
}
result <- lachen2test(231, 32, 27, 0.95, 0.05)
print(result)
$Z_value
[1] 2.308123

$p_value
[1] 0.01049614

$reject_H0
[1] TRUE
```

Μια τροποποίηση της στατιστικής συνάρτησης T_{LL} προκύπτει (βλέπε Tang et al. (2003)) εκτιμώντας τη διακύμανση του εκτιμητή δ υπό τη μηδενική υπόθεση. Σε αυτήν την περίπτωση, η προτεινόμενη στατιστική συνάρτηση δίνεται από τη σχέση:

$$T_{LLM} = \frac{(x_{11} + x_{10}) - \delta_0(x_{11} + x_{01})}{\sqrt{\delta_0(x_{10} + x_{01})}}. \quad (4.9)$$

Καθώς η στατιστική συνάρτηση T_{LLM} ακολουθεί, υπό τη μηδενική υπόθεση, ασυμπτωτικά τυπική κανονική κατανομή, απορρίπτεται η μηδενική υπόθεση, σε ασυμπτωτικό επίπεδο σημαντικότητας α , αν $T_{LLM} \geq z_\alpha$.

Στο πλαίσιο που ακολουθεί δίνεται η συνάρτηση της R για αυτόν τον έλεγχο και η υλοποίησή της για τα δεδομένα του Πίνακα 2.1.

```
lachen3test<- function(x11, x10, x01, delta0, alpha){
  deltahat<- (x11+ x10)/(x11+ x01)
  T_LLM<- ((x11+x10)- delta0*(x11+ x01))/(sqrt(delta0*(
    ↪ x10+x01)))

  pvalueT_LLM <- pnorm(T_LLM, lower.tail = FALSE)
  reject_H0 <- T_LLM >= qnorm(1- alpha)

  return(list(Z_value= T_LLM,
              p_value= pvalueT_LLM,
              reject_H0= reject_H0))
}
result <- lachen3test(231, 32, 27, 0.95, 0.05)
print(result)
$Z_value
[1] 2.390921

$p_value
[1] 0.008403085

$reject_H0
[1] TRUE
```

4.2 Οι έλεγχοι των Nam and Blackwelder (2002)

Στην ενότητα αυτή θα παρουσιαστεί ο έλεγχος μη κατωτερότητας που έχει εισαχθεί στη βιβλιογραφία από τους Nam and Blackwelder (2002). Πιο συγκεκριμένα, στην εργασία αυτή το ενδιαφέρον επικεντρώνεται στον μονόπλευρο έλεγχο που διατυπώθηκε στη σχέση (4.2), με την παράμετρο δ_0 να είναι ένας προκαθορισμένος αριθμός μικρότερος της μονάδας ($\delta_0 < 1$).

Η κεντρική ιδέα των Nam and Blackwelder (2002) είναι να στηριχθούν στο εμπειρικό (δειγματικό) ανάλογο της διαφοράς $\delta - \delta_0 = p_{11} + p_{10} - \delta_0 \cdot (p_{11} + p_{01}) = p_{1.} - \delta_0 p_{.1}$, δηλαδή να στηριχθούν στη στατιστική συνάρτηση:

$$T = \hat{p}_{1.} - \delta_0 \hat{p}_{.1} = \frac{x_{11} + x_{10}}{n} - \delta_0 \frac{x_{11} + x_{01}}{n}.$$

Το επομένο βήμα είναι ο προσδιορισμός της ασυμπτωτικής κατανομής της στατιστικής συνάρτησης T , υπό τη μηδενική υπόθεση. Αρχικά, χρησιμοποιώντας

ότι $\mathbb{E}(x_{1.}) = np_{1.}$ και $\mathbb{E}(x_{.1}) = np_{.1}$, καθώς και τη σχέση $p_{1.} - \delta_0 p_{.1} = 0$ συμπεραίνουμε ότι

$$\mathbb{E}_0(T) = \frac{\mathbb{E}(\hat{p}_{1.})}{n} - \delta_0 \frac{\mathbb{E}(\hat{p}_{.1})}{n} = 0.$$

Όσον αφορά τη διακύμανση της στατιστικής συνάρτησης T έχουμε ότι:

$$Var(T) = Var(\hat{p}_{1.}) + \delta_0^2 Var(\hat{p}_{.1}) - 2\delta_0 Cov(\hat{p}_{1.}, \hat{p}_{.1}),$$

όπου λαμβάνοντας υπόψη ότι $Var_0(x_{1.}) = np_{1.}(1-p_{1.})$, $Var_0(x_{.1}) = np_{.1}(1-p_{.1})$ και $Cov_0(x_{1.}, x_{.1}) = n(p_{11} - p_{1.}p_{.1})$, με αντικατάσταση προκύπτει ότι

$$Var_0(T) = \frac{\delta_0(p_{10} + p_{01})}{n}.$$

Καθώς οι παράμετροι p_{10} και p_{01} είναι άγνωστες, για να μπορεί να χρησιμοποιηθεί η στατιστική συνάρτηση T στην πράξη, προτείνεται η αντικατάσταση των άγνωστων παραμέτρων από τις εκτιμητές τους. Στο σημείο αυτό έγκειται και η διαφοροποίηση του ελέγχου των Nam and Blackwelder (2002) σε σχέση με εκείνον των Lachenbruch and Lynch (1998), καθώς προτείνουν δύο τρόπους εκτίμησης των άγνωστων παραμέτρων p_{10} και p_{01} .

Ειδικότερα, σύμφωνα με τον πρώτο τρόπο προτείνεται η αντικατάσταση των άγνωστων πιθανοτήτων από τους εκτιμητές μέγιστης πιθανοφάνειας χωρίς κάποιον περιορισμό. Κατά αυτόν τον τρόπο προκύπτει μία στατιστική συνάρτηση τύπου Wald για τον έλεγχο της υπόθεσης (4.2), η οποία δίνεται από τη σχέση:

$$Z_W = \frac{\sqrt{n}(\hat{p}_{1.} - \delta_0 \hat{p}_{.1})}{\sqrt{\delta_0(\hat{p}_{10} + \hat{p}_{01})}}. \quad (4.10)$$

Καθώς η στατιστική συνάρτηση Z_W ακολουθεί, υπό τη μηδενική υπόθεση, ασυμπτωτικά, τυπική κανονική κατανομή, απορρίπτεται, σε ασυμπτωτικό επίπεδο σημαντικότητας α , η μηδενική υπόθεση της σχέσης (4.2) αν $Z_W \geq z_\alpha$.

Στο πλαίσιο που ακολουθεί δίνεται η συνάρτηση της R για αυτόν τον έλεγχο και η υλοποίησή της για τα δεδομένα του Πίνακα 2.1.

```
zwtest<- function(x10, x01, x11, n, delta0, alpha) {  
  
  p1dothat<- (x11+x10)/n  
  pdot1hat <- (x11+x01)/n  
  p10hat<- x10/n  
  p01hat<- x01/n  
  
  zw=sqrt(n)*(p1dothat-delta0*pdot1hat)/(sqrt(delta0*(  
    ↪ p10hat+p01hat)))
```

```

p_value <- pnorm(zw, lower.tail = FALSE)
reject_H0 <- zw >= qnorm(1 - alpha)
Z_value <- zw

return(list(
  Z_value = Z_value,
  p_value = p_value,
  reject_H0 = reject_H0))
}

result <- zwtest(32, 27, 231, 344, 0.95,0.05)
print(result)
$Z_value
[1] 2.390921

$p_value
[1] 0.008403085

$reject_H0
[1] TRUE

```

Εναλλακτικά, οι Nam and Blackwelder (2002) προτείνουν η εκτίμηση της άγνωστης διακύμανσης της στατιστικής συνάρτησης T να προκύψει λαμβάνοντας υπόψη ότι υπό τη μηδενική υπόθεση της σχέσης (4.2) είναι $\frac{p_{11} + p_{10}}{p_{11} + p_{01}} = \delta_0$. Για τον λόγο αυτό στη συνέχεια αυτής της ενότητας το ενδιαφέρον επικεντρώνεται στον προσδιορισμό αυτών των (υπό περιορισμό) εκτιμητριών συναρτήσεων.

Πιο συγκεκριμένα, αρχικά, οι Nam and Blackwelder (2002) αξιοποιώντας το γεγονός ότι η από κοινού κατανομή των τυχαίων μεταβλητών X_{ij} , $i, j = 0, 1$, ακολουθεί πολυωνυμική κατανομή με παραμέτρους επιτυχίας p_{11}, p_{10}, p_{01} και p_{00} κατέληξαν ότι ο λογάριθμος της συνάρτησης πιθανοφάνειας δίνεται από τη σχέση:

$$L(p_{10}, p_{11}, p_{01}, p_{00}) = x_{11} \log(p_{11}) + x_{10} \log(p_{10}) + x_{01} \log(p_{01}) + x_{00} \log(p_{00}) + \log(C),$$

όπου

$$C = \frac{n!}{x_{11}!x_{10}!x_{01}!x_{00}!}.$$

Στη συνέχεια ο λογάριθμος της συνάρτησης πιθανοφάνειας γράφεται ισοδύναμα με νέα παραμετροποίηση που προκύπτει θέτοντας $p_{1.} = \delta p_{.1}$. Ειδικότερα, από

την παραπάνω σχέση, λαμβάνοντας υπόψη ότι $p_{1.} = p_{11} + p_{10}$ και $p_{.1} = p_{11} + p_{01}$ προκύπτει, μετά από λίγη άλγεβρα, ότι:

$$p_{11} = \frac{p_{10} - \delta p_{01}}{\delta - 1}.$$

Επιπρόσθετα, λαμβάνοντας υπόψη ότι $p_{00} = 1 - p_{11} - p_{10} - p_{01}$ και αντικαθιστώντας την πιθανότητα p_{11} , έχουμε ότι:

$$p_{00} = \frac{\delta(1 - p_{10}) + p_{01} - 1}{\delta - 1}.$$

Ακολουθώντας τη νέα παραμετροποίηση ο λογάριθμος της συνάρτησης πιθανοφάνειας μπορεί να γραφεί ως:

$$\begin{aligned} L(p_{10}, p_{01}; \delta) &= \log(C) + x_{11} \log\left(\frac{p_{10} - \delta p_{01}}{\delta - 1}\right) + x_{10} \log(p_{10}) + x_{01} \log(p_{01}) \\ &+ x_{00} \log\left(\frac{\delta(1 - p_{10}) + p_{01} - 1}{\delta - 1}\right) \end{aligned}$$

ή ισοδύναμα ως

$$\begin{aligned} L(p_{10}, p_{01}; \delta) &= \log(C) + x_{11} \log(p_{10} - \delta p_{01}) - (x_{11} + x_{00}) \log(\delta - 1) \\ &+ x_{10} \log(p_{10}) + x_{01} \log(p_{01}) + x_{00} \log(\delta(1 - p_{10}) + p_{01} - 1), \end{aligned}$$

με τις παραμέτρους p_{10}, p_{01} να είναι ενοχλητικές παράμετροι και η παράμετρος δ να είναι η παράμετρος ενδιαφέροντος.

Για την εύρεση των εκτιμητών μέγιστης πιθανοφάνειας των παραμέτρων p_{01} και p_{10} , για δοθείσα τιμή της παραμέτρου δ , υπολογίζουμε αρχικά τις παραγώγους του λογαρίθμου της συνάρτησης πιθανοφάνειας $L(p_{10}, p_{01}; \delta)$ ως προς p_{01} και p_{10} , αντίστοιχα. Επομένως, είναι:

$$\frac{\partial L(p_{10}, p_{01}; \delta)}{\partial p_{01}} = -\frac{x_{11}\delta}{p_{10} - \delta p_{01}} + \frac{x_{01}}{p_{01}} - \frac{x_{00}}{\delta(1 - p_{10}) + p_{01} - 1}$$

και

$$\frac{\partial L(p_{10}, p_{01}; \delta)}{\partial p_{10}} = -\frac{x_{11}}{p_{10} - \delta p_{01}} + \frac{x_{10}}{p_{10}} - \frac{x_{00}\delta}{\delta(1 - p_{10}) + p_{01} - 1}.$$

Οι εκτιμητές μέγιστης πιθανοφάνειας των παραμέτρων p_{01} και p_{10} προκύπτουν από την επίλυση των ακόλουθων εξισώσεων

$$\frac{\partial L(p_{10}, p_{01}; \delta)}{\partial p_{10}} = 0 \quad \text{και} \quad \frac{\partial L(p_{10}, p_{01}; \delta)}{\partial p_{01}} = 0.$$

Συμβολίζοντας με $\hat{p}_{01} = \frac{x_{01}}{n}$ και $\hat{p}_{10} = \frac{x_{10}}{n}$ οι παραπάνω εξισώσεις μπορούν να γραφούν ως:

$$\tilde{p}_{11}\tilde{p}_{10}(\delta-1-\delta\tilde{p}_{10}+\tilde{p}_{01})+\hat{p}_{10}(\tilde{p}_{10}-\delta\tilde{p}_{01})(\delta-1-\delta\tilde{p}_{10}+\tilde{p}_{10})-\hat{p}_{00}\delta(\tilde{p}_{10}-\delta\tilde{p}_{01})=0$$

και

$$-\hat{p}_{11}\delta\tilde{p}_{01}(\delta-1-\delta\tilde{p}_{01})+\hat{p}_{01}(\tilde{p}_{10}-\delta\tilde{p}_{01})(\delta-1-\delta\tilde{p}_{10}+\tilde{p}_{01})+\hat{p}_{00}\tilde{p}_{01}(\tilde{p}_{10}-\delta\tilde{p}_{01})=0.$$

Από τις παραπάνω εξισώσεις, μετά από αρκετή άλγεβρα, έχουμε ότι:

$$\tilde{p}_{10} = \frac{-\hat{p}_{1.} + \delta^2(\hat{p}_{.1} + 2\hat{p}_{10}) + \sqrt{(\hat{p}_{1.} - \delta^2\hat{p}_{.1})^2 + 4\delta^2\hat{p}_{01}\hat{p}_{10}}}{2\delta(\delta + 1)} \quad (4.11)$$

και

$$\tilde{p}_{01} = \delta\tilde{p}_{10} - (\delta - 1)(1 - \hat{p}_{00}). \quad (4.12)$$

Συνοψίζοντας τα παραπάνω, η στατιστική συνάρτηση που προτείνουν οι Nam and Blackwelder (2002) λαμβάνει τη μορφή

$$Z_{CML} = \frac{\sqrt{n}(\hat{p}_{1.} - \delta_0\hat{p}_{.1})}{\sqrt{\delta_0(\tilde{p}_{10} + \tilde{p}_{01})}}, \quad (4.13)$$

όπου οι εκτιμήτριες $\tilde{p}_{10}$ και $\tilde{p}_{01}$ έχουν προσδιοριστεί στις σχέσεις (4.11) και (4.12), αντίστοιχα, με την τροποποίηση ότι $\delta = \delta_0$, ενώ $\hat{p}_{1.} = \frac{x_{11}+x_{10}}{n}$ και $\hat{p}_{.1} = \frac{x_{11}+x_{01}}{n}$. Λαμβάνοντας, επιπρόσθετα, υπόψη ότι η στατιστική συνάρτηση Z_{CML} ασυμπτωτικά ακολουθεί, υπό τη μηδενική υπόθεση, τυπική κανονική κατανομή, η μηδενική υπόθεση της σχέσης (4.2) απορρίπτεται, σε ασυμπτωτικό επίπεδο σημαντικότητας α , αν $Z_{CML} \geq z_\alpha$.

Παρατήρηση 4.1. Στην περίπτωση που $\delta_0 = 1$, συμπεραίνουμε από τις σχέσεις (4.10) και (4.13) ότι και οι δύο έλεγχοι ανάγονται στον κλασικό έλεγχο του McNemar (1947).

Στο πλαίσιο που ακολουθεί δίνεται η συνάρτηση της R για αυτόν τον έλεγχο και η υλοποίησή της για τα δεδομένα του Πίνακα 2.1.

```
zclmtest <- function(x11, x10, x01, x00, n, delta0, alpha
  ↪ ) {

  p1dothat <- (x11+x10)/n
  pdot1hat <- (x11+x01)/n
  p10hat <- x10/n
  p01hat <- x01/n
```

```

p00hat <- x00/n
p10tilde <- (-p1dothat + delta0^2 * (pdot1hat + 2*
  ↪ p10hat) + sqrt((p1dothat - delta0^2*pdot1hat)^2 +
  ↪ 4*delta0^2*p10hat*p01hat))/ (2*delta0 *(delta0
  ↪ +1))

p01tilde <- delta0* p10tilde - (delta0 -1)*(1- p00hat)

zclm <- ( sqrt(n)* (p1dothat- delta0* pdot1hat))/ (sqrt
  ↪ (delta0* (p10tilde +p01tilde)))

p_value <- pnorm(zclm, lower.tail = FALSE)
reject_H0 <- zclm >= qnorm(1-alpha)
Z_value <- zclm

return(list(
  Z_value = Z_value,
  p_value = p_value,
  reject_H0 = reject_H0))
}

result <- zclmtest(231, 32, 27, 54, 344, 0.95, 0.05)
print(result)
$Z_value
[1] 2.32424

$p_value
[1] 0.01005631

$reject_H0
[1] TRUE

```

4.3 Οι έλεγχοι των Tang et al. (2003)

Στην ενότητα αυτή το ενδιαφέρον μας θα επικεντρωθεί στην παρουσίαση των ελέγχων που προτάθηκαν στην εργασία των Tang et al. (2003) για τον έλεγχο της σύνθετης υπόθεσης της σχέσης (4.1), η οποία διασπάται στα δύο σύνολα μονόπλευρων υποθέσεων των σχέσεων (4.2) και (4.3), με τη διαφοροποίηση ότι θεωρούν την παράμετρο $\delta_1 = \frac{1}{\delta_0}$, όπου $0 < \delta_0 < 1$ ένα προκαθορισμένο όριο ισοδυναμίας (π.χ. 0.95).

Αν $\beta = p_{1.} - \delta_0 p_{.1}$, τότε η προς έλεγχο υπόθεση

$$H_{0\ell} : \frac{p_{1.}}{p_{.1}} \leq \delta_0 \quad \text{έναντι} \quad H_{1\ell} : \frac{p_{1.}}{p_{.1}} > \delta_0 \quad (4.14)$$

μπορεί ισοδύναμα να εκφραστεί ως

$$H_{0\ell} : \beta \leq 0 \quad \text{έναντι} \quad H_{1\ell} : \beta > 0. \quad (4.15)$$

Για τον έλεγχο της υπόθεσης μη κατωτερότητας που διατυπώνεται στη σχέση (4.15), οι Tang et al. (2003) κατασκεύασαν μια στατιστική συνάρτηση μέσω της score μεθόδου, χρησιμοποιώντας την παραμετροποίηση του λογαρίθμου της συνάρτησης πιθανοφάνειας που προτάθηκε από τον Tango (1998). Ανακαλώντας τα αποτελέσματα που παρουσιάστηκαν στην Ενότητα 2.5 και θέτοντας $\phi = \rho_{\sigma_1, \sigma_1}$, έχουμε ότι:

$$p_{11} = p_{1.}p_{.1} + \rho_{\sigma_1, \sigma_1} = (\beta + \delta_0 p_{.1})p_{.1} + \phi,$$

$$p_{10} = p_{1.}(1 - p_{.1}) - \rho_{\sigma_1, \sigma_1} = (\beta + \delta_0 p_{.1})(1 - p_{.1}) - \phi,$$

$$p_{01} = (1 - p_{1.})p_{.1} - \rho_{\sigma_1, \sigma_1} = (1 - \beta - \delta_0 p_{.1})p_{.1} - \phi,$$

και

$$p_{00} = (1 - p_{1.})(1 - p_{.1}) + \rho_{\sigma_1, \sigma_1} = (1 - \beta - \delta_0 p_{.1})(1 - p_{.1}) + \phi.$$

Τότε, ο λογάριθμος της συνάρτησης πιθανοφάνειας του πολωνυμικού μοντέλου, λαμβάνοντας υπόψη τη νέα παραμετροποίηση, είναι:

$$L(\beta, \phi, p_{.1}) = C + x_{11} \log(p_{11}) + x_{10} \log(p_{10}) + x_{01} \log(p_{01}) + x_{00} \log(p_{00}),$$

όπου

$$C = \log \left(\frac{n!}{x_{11}! x_{10}! x_{01}! x_{00}!} \right).$$

Από την παραπάνω σχέση προκύπτει ότι:

$$\frac{\partial L(\beta, \phi, p_{.1})}{\partial \phi} = \frac{x_{11}}{p_{11}} - \frac{x_{10}}{p_{10}} - \frac{x_{01}}{p_{01}} + \frac{x_{00}}{p_{00}},$$

$$\frac{\partial L(\beta, \phi, p_{.1})}{\partial \beta} = p_{.1} \frac{\partial L(\beta, \phi, p_{.1})}{\partial \phi} + \frac{x_{10}}{p_{10}} - \frac{x_{00}}{p_{00}},$$

και

$$\frac{\partial L(\beta, \phi, p_{.1})}{\partial p_{.1}} = (\beta + 2\delta_0 p_{.1}) \frac{\partial L(\beta, \phi, p_{.1})}{\partial \phi} + \frac{x_{10}\delta_0}{p_{10}} + \frac{x_{01}}{p_{01}} - \frac{x_{00}(\delta_0 + 1)}{p_{00}}.$$

Τότε, σύμφωνα με τη μέθοδο score, η γενική μορφή της στατιστικής συνάρτησης, που στη συνέχεια εξειδικεύτηκε από τους Tang et al. (2003), είναι η

$$Z_T = \frac{\frac{\partial L(\beta, \phi, p_{.1})}{\partial \beta} \Big|_{\beta=0, \phi=\hat{\phi}, p_{.1}=\hat{p}_{.1}}}{\sqrt{\text{Var} \left(\frac{\partial L(\beta, \phi, p_{.1})}{\partial \beta} \Big|_{\beta=0, \phi=\hat{\phi}, p_{.1}=\hat{p}_{.1}} \right)}}, \quad (4.16)$$

με $\hat{\phi}$ και $\hat{p}_{.1}$ να είναι οι εκτιμητές μέγιστης πιθανοφάνειας όταν $\beta = 0$, δηλαδή οι εκτιμητές που προκύπτουν μεγιστοποιώντας τη συνάρτηση πιθανοφάνειας $L(0, \phi, p_{.1}) := L(\phi, p_{.1})$. Μετά από αλγεβρικές πράξεις, οι οποίες δίνονται στο παράρτημα της εργασίας των Tang et al. (2003), έχουμε ότι η στατιστική συνάρτηση που προκύπτει είναι η $T_1 := T(x_{11}, x_{10}, x_{01}, n, \delta_0)$ που δίνεται από τη σχέση:

$$T_1 = \frac{x_{11} + x_{10} - (x_{11} + x_{01})\delta_0}{\sqrt{n((1 + \delta_0)\hat{p}_{01} + (x_{11} + x_{01} + x_{10})(\delta_0 - 1)/n)}}, \quad (4.17)$$

όπου

$$\hat{p}_{01} = \frac{\sqrt{b^2 - 4ac} - b}{2a},$$

με

$$a = n(1 + \delta_0), b = (x_{11} + x_{01})\delta_0^2 - (x_{11} + x_{10} + 2x_{01})$$

και

$$c = \frac{x_{01}(1 - \delta_0)(x_{11} + x_{10} + x_{01})}{n}.$$

Λαμβάνοντας επιπρόσθετα υπόψη ότι η στατιστική συνάρτηση T_1 ακολουθεί ασυμπτωτικά, υπό τη μηδενική υπόθεση της σχέσης (4.14), τυπική κανονική κατανομή, έχουμε ότι η μηδενική υπόθεση της σχέσης (4.14) απορρίπτεται, σε ασυμπτωτικό επίπεδο σημαντικότητας α , αν $T_1 \geq z_\alpha$.

Όσον αφορά τον έλεγχο της υπόθεσης

$$H_{0u} : \frac{p_{1.}}{p_{.1}} \geq \delta_1 = \frac{1}{\delta_0} \quad \text{έναντι} \quad H_{1u} : \frac{p_{1.}}{p_{.1}} < \delta_1 = \frac{1}{\delta_0} \quad (4.18)$$

που ισοδύναμα εκφράζεται ως

$$H_{0u} : \beta_1 = p_{.1} - \delta_0 p_{1.} \leq 0 \quad \text{έναντι} \quad H_{1u} : \beta_1 = p_{.1} - \delta_0 p_{1.} > 0 \quad (4.19)$$

οι Tang et al. (2003) προτείνουν τη χρήση της στατιστικής συνάρτησης που προκύπτει άμεσα με εναλλαγή του ρόλου της νέας και παραδοσιακής διαγνωστικής μεθοδολογίας. Ειδικότερα, προτείνουν τη χρήση της στατιστικής

συνάρτησης

$$T_2 := T(x_{11}, x_{01}, x_{10}, n, \delta_0) = \frac{x_{11} + x_{01} - (x_{11} + x_{10})\delta_0}{\sqrt{n((1 + \delta_0)\hat{p}_{10} + (x_{11} + x_{01} + x_{10})(\delta_0 - 1)/n)}},$$

όπου

$$\hat{p}_{10} = \frac{\sqrt{b_1^2 - 4ac_1} - b_1}{2a}$$

με

$$a = n(1 + \delta_0), b_1 = (x_{11} + x_{10})\delta_0^2 - (x_{11} + x_{01} + 2x_{10})$$

και

$$c_1 = \frac{x_{10}(1 - \delta_0)(x_{11} + x_{10} + x_{01})}{n}.$$

Επιπρόσθετα, η μηδενική υπόθεση της σχέσης (4.18) απορρίπτεται, σε ασυμπτωτικό επίπεδο σημαντικότητας α , αν $T_2 \geq z_\alpha$.

Τέλος, η σύνθετη υπόθεση

$$H_0 : \frac{p_{11} + p_{10}}{p_{11} + p_{01}} \leq \delta_0 \text{ ή } \frac{p_{11} + p_{10}}{p_{11} + p_{01}} \geq \frac{1}{\delta_0} \quad \text{έναντι} \quad H_1 : \delta_0 < \frac{p_{11} + p_{10}}{p_{11} + p_{01}} < \frac{1}{\delta_0},$$

απορρίπτεται, σε ασυμπτωτικό επίπεδο σημαντικότητας α , αν $T_1 \geq z_\alpha$ και $T_2 \geq z_\alpha$.

Παρατήρηση 4.2. Επισημαίνεται ότι ο ρόλος των δύο διαγνωστικών ελέγχων μπορεί να εναλλάσσεται για τον έλεγχο της συμπληρωματικής μονόπλευρης υπόθεσης, με τον όρο ότι η στατιστική συνάρτηση ελέγχου παραμένει έγκυρη μετά την εναλλαγή, καθώς απαιτείται και η συμμετρία των προκαθορισμένων ορίων ισοδυναμίας. Η συμμετρία αυτή πιτρέπει τη χρήση της διαδικασίας των δύο μονόπλευρων ελέγχων (*Two One-Sided Tests, TOST*), για τον σύνθετο έλεγχο ισοδυναμίας.

Στο πλαίσιο που ακολουθεί δίνεται η συνάρτηση της R για αυτόν τον έλεγχο και η υλοποίησή της για τα δεδομένα του Πίνακα 2.1.

```
T1test <- function(x10, x01, x11, n, delta0, alpha, type)
{
  a <- n*(1 +delta0)
  b <- (x11 + x01)* delta0^2 - (x11 + x10 + 2*x01)
  c <- (x01* (1- delta0)* (x11+ x10 + x01))/n

  p01hat <- ( sqrt((b^2 - 4* a* c))- b)/ (2*a)
```

```

T1 <- (x11 + x10 -(x11+ x01)* delta0)/ ( sqrt(n* ((1+
  ↳ delta0)*p01hat + (x11+ x01+ x10)* (delta0 -1)/n))
  ↳ )

b1 <- (x11 + x10)* delta0^2 - (x11 + x01 + 2*x10)
c1 <- (x10* (1- delta0)* (x11+ x10 + x01))/n

p10hat <- ( sqrt((b1^2 - 4* a* c1))- b1)/ (2*a)

T2 <- (x11 + x01 -(x11+ x10)* delta0)/ ( sqrt(n* ((1+
  ↳ delta0)*p10hat + (x11+ x01+ x10)* (delta0 -1)/n))
  ↳ )

if (type == "noninferiority") {
  p_value <- pnorm(T1, lower.tail = FALSE)
  reject_H0 <- T1 >= qnorm(1 - alpha)
  Z_value <- T1

} else if (type == "nonsuperiority") {
  p_value <- pnorm(T2, lower.tail = FALSE)
  reject_H0 <- T2 >= qnorm(1 - alpha)
  Z_value <- T2

} else if (type == "equivalence") {
  p_value <- max(pnorm(T1, lower.tail = FALSE), pnorm(
    ↳ T2, lower.tail = FALSE))
  reject_H0 <- (T1 >= qnorm(1 - alpha)) & (T2 >= qnorm
    ↳ (1 - alpha))
  Z_value <- c(T1, T2)
} else {
  stop("The 'alternative' argument must be either '
    ↳ noninferiority', 'nonsuperiority', or '
    ↳ equivalence'")
}

return(list(
  Z_value = Z_value,
  p_value = p_value,
  reject_H0 = reject_H0,
  type = type))
}

result <- T1test(32, 27, 231, 344, 0.95 , 0.05, type = "

```

```
    ↪ equivalence")  
  print(result)  
$Z_value  
[1] 2.324240 1.073933  
  
$p_value  
[1] 0.1414264  
  
$reject_H0  
[1] FALSE  
  
$type  
[1] "equivalence"
```

ΚΕΦΑΛΑΙΟ 5

ΣΥΓΚΡΙΤΙΚΕΣ ΜΕΛΕΤΕΣ

Το ενδιαφέρον στα προηγούμενα τρία κεφάλαια αυτής της μεταπτυχιακής διατριβής επικεντρώθηκε στην παρουσίαση των κυριότερων μεθοδολογιών ελέγχου της υπόθεσης της μη κατωτερότητας, της μη ανωτερότητας ή της ισοδυναμίας, αλλά και στη δημιουργία συναρτήσεων στην R για την υλοποίηση αυτών σε πρακτικές εφαρμογές. Ειδικότερα, οι μεθοδολογίες αυτές ταξινομήθηκαν σε τρεις μεγάλες κατηγορίες ανάλογα με την παράμετρο που χρησιμοποιείται για τον έλεγχο των παραπάνω υποθέσεων. Ένα εύλογο ερώτημα που ίσως να έχει προκύψει είναι αν υπάρχει κάποια αναγκαιότητα για την ύπαρξη διαφορετικών μεθοδολογιών ή μπορούμε να ισχυριστούμε ότι κάποια ή κάποιες εξ αυτών έχει ή έχουν καλύτερη απόδοση συγκριτικά με τις υπόλοιπες. Η απάντηση σε αυτό το ερώτημα μπορεί να δοθεί μέσω της διεξαγωγής προσομοιώσεων για να διαπιστωθεί αν οι προτεινόμενες διαδικασίες ελέγχου διατηρούν το ονομαστικό επίπεδο ελέγχου και αν έχουν ικανοποιητική ισχύ. Η παραπάνω διαπίστωση οδήγησε πολλούς συγγραφείς να διεξάγουν συγκριτικές μελέτες στο πλαίσιο της δημοσίευσης μίας νέας μεθόδου. Στη συνέχεια αυτού του κεφαλαίου θα παραθέσουμε τα αποτελέσματα των συγκριτικών μελετών που έχουν διενεργηθεί στη βιβλιογραφία τόσο για τη διατήρηση του επιπέδου σημαντικότητας όσο και της ισχύος. Σε κάθε περίπτωση θα αναφέρονται οι συγκρινόμενες μεθοδολογίες και τα συμπεράσματα που έχουν εξαχθεί.

5.1 Συγκρίσεις ελέγχων δευτέρου κεφαλαίου

Στην ενότητα αυτή θα παρουσιαστούν τα αποτελέσματα συγκριτικών μελετών που αφορούν την απόδοση των ελέγχων που παρουσιάστηκαν στο δεύτερο κεφάλαιο αυτής της μεταπτυχιακής διατριβής, δηλαδή των ελέγχων που ανάγονται στον έλεγχο για τη διαφορά των πιθανοτήτων θετικής διάγνωσης. Πιο συγκεκριμένα, θα σχολιασθούν τα αποτελέσματα που παρατίθενται στις εργασίες των

Nam (1997), Tango (1998) και Liu et al. (2002). Από όσο είμαστε σε θέση να γνωρίζουμε δεν έχει διεξαχθεί μία εκτεταμένη συγκριτική μελέτη για την αξιολόγηση της απόδοσης όλων αυτών των ελέγχων. Ωστόσο, έχουν πραγματοποιηθεί μελέτες προσομοίωσης που συμπεριλαμβάνουν κάποιους από αυτούς.

Αποτελέσματα της μελέτης του Nam (1997)

Στην εργασία του Nam (1997) συγκρίνονται ως προς την απόδοσή τους ο έλεγχος που στηρίζεται στη στατιστική συνάρτηση Z_s της σχέσης (2.16) και αποτέλεσε αντικείμενο μελέτης στην Ενότητα 2.2 με τον έλεγχο που στηρίζεται στη στατιστική συνάρτηση $Z_{LB}^{(1)}$ της σχέσης (2.6) και αποτέλεσε αντικείμενο μελέτης στην Ενότητα 2.1. Τα συμπεράσματα που προκύπτουν παρατίθενται στη συνέχεια.

- Η μέθοδος των Lu and Bean (1995) είναι αντισυντηρητική (anticonservative), καθώς το εμπειρικό ποσοστό σφάλματος Τύπου I υπερβαίνει το ονομαστικό επίπεδο σημαντικότητας α . Ειδικότερα, παρατηρούμε ότι για $\alpha = 0.05$, το σφάλμα Τύπου I της μεθόδου με τη σ.σ. $Z_{LB}^{(1)}$ της σχέσης (2.6) έφτασε ως και 0.114 στην περίπτωση μικρού σε μέγεθος δείγματος ($n = 30$). Η συμπεριφορά αυτή αιτιολογείται (βλέπε Nam (1997)) από το γεγονός ότι για την εκτίμηση της διακύμανσης χρησιμοποιούνται εκτιμήτριες που προκύπτουν μη λαμβάνοντας υπόψη τον περιορισμό της μηδενικής υπόθεσης. Ο Nam (1997) επισημαίνει ότι το πρόβλημα εξακολουθεί να υφίσταται ακόμη και με εφαρμογή διόρθωση συνέχειας, ιδιαίτερα για μεγαλύτερα μεγέθη δείγματος ($n = 50, n = 100$).
- Από την άλλη ο έλεγχος που στηρίζεται στη στατιστική συνάρτηση Z_s και έχει προκύψει με τη μέθοδο score με τη χρήση εκτιμητών μέγιστης πιθανοφάνειας υπό τον περιορισμό της μηδενικής υπόθεσης, διατηρεί το ποσοστό σφάλματος Τύπου I πολύ κοντά στο ονομαστικό επίπεδο σημαντικότητας. Πιο συγκεκριμένα, για μέγεθος δείγματος $n = 30, 50, 100$, οι αντίστοιχοι μέσοι όροι των εμπειρικών ποσοστών σφαλμάτων τύπου I είναι 4.2%, 4.7% και 4.8%, αντίστοιχα. Επισημαίνεται ωστόσο ότι οι τιμές του εμπειρικού ποσοστού σφάλματος τύπου I κυμαίνονται από 1.1% έως 6.2%.
- Από τα αποτελέσματα που παρατίθενται στην προαναφερθείσα εργασία και αφορούν την ισχύ των δύο μεθόδων προκύπτει ότι σε όλες τις περιπτώσεις που παρατίθενται η ισχύς της σ.σ. $Z_{LB}^{(1)}$ των Lu and Bean (1995) είναι ελαφρώς μεγαλύτερη από την αντίστοιχη της score μεθόδου του Nam (1997). Η ιδιότητα αυτή ωστόσο θεωρούμε ότι οφείλεται στο γεγονός ότι

ο έλεγχος των Lu and Bean (1995) παρουσιάζει φιλελεύθερη συμπεριφορά, δηλαδή απορρίπτει σε μεγάλο αριθμό φορών τη μηδενική υπόθεση ακόμα και αν αυτή είναι αληθής (πρώτο σημείο παραπάνω).

Από τα παραπάνω προκύπτει ότι παρότι η μέθοδος score του Nam (1997) με τη σ.σ. Z_s έχει ελαφρώς μικρότερη ισχύ, ωστόσο είναι προτιμότερη συγκριτικά με τον έλεγχο των Lu and Bean (1995), λόγω της φιλελεύθερης συμπεριφοράς του τελευταίου ελέγχου.

Αποτελέσματα της μελέτης του Tango (1998)

Στην εργασία του Tango (1998) συγκρίνεται εκ νέου η μεθοδολογία που προτάθηκε από τους Lu and Bean (1995), η οποία στηρίζεται στη σ.σ. $Z_{LB}^{(1)}$ της σχέσης (2.6) και αποτέλεσε αντικείμενο μελέτης στην Ενότητα 2.1, με τη μεθοδολογία που προτάθηκε από τον ίδιο και στηρίζεται στη στατιστική συνάρτηση T που προσδιορίστηκε στη σχέση (2.34) (βλέπε Ενότητα 2.5).

- Τα εμπειρικά επίπεδα σημαντικότητας του ελέγχου με τη σ.σ. T που προτάθηκε από τον Tango (1998) είναι πολύ κοντά στο ονομαστικό επίπεδο σημαντικότητας $\alpha = 5\%$. Η παραπάνω διαπίστωση ισχύει τόσο για μικρού μεγέθους δείγματα ($n = 30$) όσο και για περιπτώσεις όπου η πιθανότητα εμφάνισης ασύμβατων ζευγών p_{10} είναι εξαιρετικά χαμηλή.
- Καταλήγουν στο ίδιο συμπέρασμα με αυτό του Nam (1997) ότι το εμπειρικό ποσοστό σφάλματος της μεθόδου των Lu and Bean (1995) υπερβαίνει σημαντικά σε κάποιες περιπτώσεις το ονομαστικό επίπεδο σημαντικότητας, λαμβάνοντας τιμή έως 18.7%.
- Όσον αφορά την ισχύ της μεθόδου T αυξάνει όσο το μέγεθος του δείγματος αυξάνει για σταθεροποιημένη τιμή της παραμέτρου δ . Επισημαίνεται ότι η ισχύς επηρεάζεται τόσο από τη συσχέτιση όσο και από την πιθανότητα εμφάνισης ασύμβατου ζεύγους. Για τα διάφορα σενάρια που χρησιμοποιήθηκαν η ισχύς κυμαίνεται από 18.4 % έως 77.2%.

Από τα παραπάνω προκύπτει ότι η μέθοδος του Tango (1998) είναι προτιμότερη συγκριτικά με τον έλεγχο των Lu and Bean (1995), λόγω της φιλελεύθερης συμπεριφοράς του τελευταίου ελέγχου.

Αποτελέσματα της μελέτης των Liu et al. (2002)

Στην εργασία των Liu et al. (2002) συγκρίνονται ως προς την απόδοσή τους

δύο έλεγχοι. Ο πρώτος έλεγχος είναι αυτός που στηρίζεται στις στατιστικές συναρτήσεις Z_ℓ και Z_u που δόθηκαν στις σχέσεις (2.24) και (2.25), αντίστοιχα (βλέπε Ενότητα 2.3, παράγραφο 1η μέθοδος εκτίμησης της διακύμανσης). Ο δεύτερος έλεγχος είναι αυτός που στηρίζεται στις στατιστικές συναρτήσεις $Z_\ell^{(1)}$ και $Z_u^{(1)}$ των σχέσεων (2.28) και (2.30) (βλέπε Ενότητα 2.3, παράγραφο 3η μέθοδος εκτίμησης της διακύμανσης). Επιπλέον, θεώρησαν τους ίδιους ελέγχους με χρήση διόρθωση συνέχειας. Στη συνέχεια ο πρώτος έλεγχος αναφέρεται ως Sample-based και ο δεύτερος ως RMLE-based (restricted maximum likelihood-based). Τέλος, οι Liu et al. (2002) συγκρίνουν τις δύο μεθοδολογίες τόσο για τον έλεγχο της μη κατωτερότητας, όσο και για τον έλεγχο της ισοδυναμίας.

Τα συμπεράσματα από τη διεξαγωγή της μελέτης για τον έλεγχο της μη κατωτερότητας είναι τα ακόλουθα.

- Η RMLE-based μέθοδος χωρίς τη διόρθωση συνέχειας, διατηρεί τα εμπειρικά ποσοστά σφάλματος χαμηλότερα από το ονομαστικό επίπεδο $\alpha = 5\%$. Ωστόσο, επισημαίνεται ότι σε περίπτωση έλλειψης ασύμβατων ζευγών, η μέθοδος αυτή μπορεί να παρουσιάσει συντηρητική συμπεριφορά για αυστηρά όρια ισοδυναμίας ($\delta = 0.05$). Από την άλλη, η RMLE-based μέθοδος με διόρθωση συνέχειας παρουσιάζει αντισυντηρητική συμπεριφορά. Επομένως, συνίσταται να μην χρησιμοποιείται η διόρθωση συνέχειας.
- Αντίθετα, η Sample-based μέθοδος χωρίς τη διόρθωση συνέχειας παρουσίασε ποσοστά σφάλματος που ξεπερνούσαν το 20% (πολύ υψηλότερα του $\alpha = 5\%$), ιδίως όταν το μέγεθος του δείγματος ήταν αρκετά μικρό ($n = 30$), καθιστώντας την αντισυντηρητική. Από την άλλη, τα εμπειρικά επίπεδα σημαντικότητας ήταν μικρότερα από το ονομαστικό επίπεδο σημαντικότητας όταν εφαρμόζεται η διόρθωση συνέχειας. Για τον λόγο αυτό προτείνεται σε αυτήν την περίπτωση η εφαρμογή διόρθωσης συνέχειας.
- Ως προς την ισχύ, μεταξύ της RMLE-based χωρίς διόρθωση συνέχειας και της Sample-based με διόρθωση συνέχειας φαίνεται ότι παρουσιάζει μια μικρή υπεροχή η πρώτη εκ των μεθόδων.

Τα συμπεράσματα από τη διεξαγωγή της μελέτης για τον έλεγχο της ισοδυναμίας είναι τα ακόλουθα.

- Η Sample-based μέθοδος φαίνεται να έχει δυσκολία να διατηρήσει το ονομαστικό επίπεδο σημαντικότητας ακόμα και με τη διόρθωση συνέχειας, ενώ

η RMLE-based μέθοδος χωρίς τη διόρθωση συνέχειας φαίνεται να διατηρεί το επίπεδο σημαντικότητας στην πλειοψηφία των περιπτώσεων.

- Όσον αφορά την ισχύ των μεθόδων, αυτή αυξάνει όσο το δειγματικό μέγεθος αυξάνει ή όσο η πιθανότητα εμφάνισης ασύμβατου ζεύγους p_{01} μικραίνει για σταθεροποιημένη τιμή της παραμέτρου δ . Ωστόσο, η ισχύς και των δύο τεστ είναι πολύ μικρή για δείγματα μεγέθους $n = 50, 100$, όταν $\delta = 0.05, 0.1$. Από την άλλη, και οι δύο μέθοδοι επιτυγχάνουν υψηλά ποσοστά ισχύος (πάνω από 90%), όταν είναι διαθέσιμο δείγμα μεγάλου μεγέθους ($n = 200$) και έχουμε λιγότερο αυστηρά περιθώρια ισοδυναμίας ($\delta \geq 0.15$). Επισημαίνουμε ωστόσο ότι η RMLE-based μέθοδος των σ.σ. $Z_\ell^{(1)}$ και $Z_u^{(1)}$ χωρίς διόρθωση συνέχειας έδωσε καλύτερα αποτελέσματα.

Από τα παραπάνω συμπεραίνουμε ότι καμία από τις δύο μεθόδους δεν μπορεί να προταθεί όταν η πιθανότητα εμφάνισης ασύμβατου ζεύγους είναι μικρή (π.χ. μικρότερη από 0.2) ή το συνολικό μέγεθος δείγματος είναι μικρότερο από 50. Ωστόσο, η RMLE-based μέθοδος χωρίς τη διόρθωση συνέχειας φαίνεται να είναι μία αποδεκτή προσέγγιση και προτείνεται σε σχέση με τη Sample-based μέθοδο με ή χωρίς διόρθωση συνέχειας.

Τέλος, από τα προηγούμενα προκύπτει ότι είναι ανοικτό θέμα προς διερεύνηση η συγκριτική μελέτη όλων των μεθόδων που παρουσιάστηκαν στο δεύτερο κεφάλαιο της μεταπτυχιακής διατριβής.

5.2 Συγκρίσεις ελέγχων τρίτου κεφαλαίου

Στην ενότητα αυτή θα παρουσιαστούν τα αποτελέσματα συγκριτικών μελετών που αφορούν την απόδοση των ελέγχων που παρουσιάστηκαν στο τρίτο κεφάλαιο αυτής της μεταπτυχιακής διατριβής, δηλαδή των ελέγχων που ανάγονται στον έλεγχο για το πηλίκο των πιθανοτήτων ασύμβατων ζευγών. Πιο συγκεκριμένα, θα σχολιασθούν τα αποτελέσματα που παρατίθενται στις εργασίες των Liu et al. (2005), Jin et al. (2016), Wang et al. (2018) και Chen and Pan (2022).

Αποτελέσματα της μελέτης των Liu et al. (2005)

Οι Liu et al. (2005) πρότειναν, όπως είδαμε στο Κεφάλαιο 3, δύο μεθόδους για τον έλεγχο της υπόθεσης μη κατωτερότητας ενός νέου διαγνωστικού ελέγχου βάσει των στατιστικών συναρτήσεων Z_b και Z_{LFM} των σχέσεων (3.4) και (3.10), αντίστοιχα (βλέπε Ενότητα 3.1). Η σύγκριση της απόδοσης αυτών των ελέγχων

αποτέλεσε στόχο της μελέτης προσομοίωσης που διεξήγαγαν. Τα συμπεράσματα από αυτήν παρατίθενται στη συνέχεια.

- Η απόδοση της μεθόδου που στηρίζεται στη σ.σ. Z_b , δηλαδή στη μέθοδο δέλτα, ως προς τη διατήρηση του επιπέδου σημαντικότητας δεν είναι ικανοποιητική για μικρά σε μέγεθος δείγματα ($n < 100$). Ιδιαίτερα στην περίπτωση όπου η πιθανότητα p_{01} εμφάνισης ασύμβατου ζεύγους είναι μικρότερη από 0.05, οι προσομοιώσεις έδειξαν ότι τα εμπειρικά ποσοστά σφάλματος Τύπου I της μεθόδου μπορούν να φτάσουν ως και 0%. Το γεγονός αυτό καθιστά τη μέθοδο ιδιαίτερα συντηρητική (conservative). Μόνο όταν το μέγεθος δείγματος είναι μεγαλύτερο (κυρίως $n \geq 100$) και η πιθανότητα εμφάνισης $p_{01} \geq 0.10$, διατηρείται το ποσοστό σφάλματος κοντά στο ονομαστικό επίπεδο σημαντικότητας α , όντας τα εμπειρικά ποσοστά εντός του διαστήματος (4.29-5.34%).
- Από την άλλη, παρότι η μέθοδος που στηρίζεται στη σ.σ. Z_{LFM} , δηλαδή στη μέθοδο score, παρουσιάζει συντηρητική συμπεριφορά για μικρά μεγέθη δείγματος ($n < 100$), διατηρεί πιο κοντά τα ποσοστά σφάλματος Τύπου I στο 5%, ακόμη και στην περίπτωση που $n = 50$. Επιπλέον, στις περιπτώσεις που $p_{01} \geq 0.10$, η μέθοδος score διατηρεί τα ποσοστά σφάλματος εντός του αναμενόμενου εύρους, καθιστώντας την πιο αξιόπιστη επιλογή για ελέγχους μη κατωτερότητας συγκριτικά με τη μέθοδο με τη σ.σ. Z_b .
- Η ισχύς και των δύο μεθόδων είναι αύξουσα συνάρτηση του μεγέθους δείγματος, της πιθανότητας p_{10} και της απόκλισης $\delta_1 - \delta_0$.
- Η απόδοση της σ.σ. Z_{LFM} είναι καλύτερη όταν $p_{01} < 0.1$ και το μέγεθος δείγματος είναι μικρότερο του 100. Από την άλλη, αν $p_{01} > 0.05$ και το μέγεθος δείγματος είναι τουλάχιστον ίσο με 100, η ισχύς των δύο μεθόδων είναι συγκρίσιμη.

Από τα παραπάνω συμπεραίνουμε ότι για μικρά σε μέγεθος δείγματα είναι προτιμότερη η χρήση της σ.σ. Z_{LFM} , ενώ για μεγαλύτερα σε μέγεθος δείγματα η διαφοροποίηση των δύο δεν είναι έντονη.

Αποτελέσματα της μελέτης των Jin et al. (2016)

Στη μελέτη των Jin et al. (2016) συγκρίνονται ως προς την απόδοση τους οι έλεγχοι που προτάθηκαν στην εργασία των Liu et al. (2005) και στηρίζονται στις σ.σ. Z_b και Z_{LFM} των σχέσεων (3.4) και (3.10), αντίστοιχα (βλέπε Ενότητα 3.1) και ο έλεγχος που αυτοί πρότειναν χρησιμοποιώντας το λεγόμενο

επαγωγικό μοντέλο. Ο έλεγχος αυτός διεξάγεται στη βάση της συνάρτησης ευπιστίας $pl_{x_{10}, x_{01}}(A)$, όπως ορίστηκε στη σχέση (3.17) της Ενότητας 3.2. Σε όσα ακολουθούν σε αυτήν την παράγραφο οι μέθοδοι με χρήση των Z_b και Z_{LFM} θα αναφέρονται ως ασυμπτωτικές μέθοδοι, ενώ η μέθοδος των Jin et al. (2016) εν συντομία θα αναφέρεται ως IM-μέθοδος.

- Η IM-μέθοδος παρουσιάζει καλύτερη απόδοση ως προς τη διατήρηση του ονομαστικού επιπέδου σημαντικότητας. Παρότι η επίδοσή της είναι συντηρητική, καθίσταται έγκυρη υπό κάθε συνθήκη, διατηρώντας τα εμπειρικά ποσοστά σφάλματος αυστηρά κάτω από το επίπεδο $\alpha = 5\%$ (1.37- 4.58%).
- Αντίθετα, η μέθοδος score (βλέπε σχέση (3.10)), για $\delta_0 = 0.8$ παραβιάζει τα 95% όρια εμπιστοσύνης (0.0457, 0.0543) για το ονομαστικό επίπεδο $\alpha = 5\%$, και ειδικότερα υπερβαίνει το ανώτερο όριο 7 φορές. Υπό τις ίδιες συνθήκες, η δέλτα μέθοδος (βλέπε σχέση (3.4)) υπερβαίνει μόλις μια φορά το ανώτερο όριο.
- Η IM-μέθοδος, σε μικρά μεγέθη δείγματος ($n \leq 100$), εμφάνισε χαμηλότερη ισχύ από τις ασυμπτωτικές μεθόδους. Κυρίως στην περίπτωση όπου $\delta_0 = 0.8$ για μέγεθος δείγματος $n = 50$, η ισχύς της IM-μεθόδου είναι μόλις 6.54%, σχεδόν η μισή από αυτή της score μεθόδου (12.84%). Η διαφορά αυτή παρατηρείται να μειώνεται σημαντικά σε μεγαλύτερα μεγέθη δειγμάτων ($n = 400$).
- Κατά κύριο λόγο, η μέθοδος score διατήρησε την υψηλότερη ισχύ υπό κάθε συνθήκη. Η μέθοδος δέλτα οδήγησε σε ενδιάμεσες επιδόσεις για μικρά μεγέθη δείγματος ($n \leq 100$), αλλά με παρόμοια αποτελέσματα ισχύος με τη μέθοδο score για μεγαλύτερα μεγέθη.

Συμπερασματικά, η IM-μέθοδος μπορεί να προταθεί, καθώς διατηρεί το ονομαστικό επίπεδο σημαντικότητας, ιδιότητα που δεν έχουν οι άλλοι δύο έλεγχοι, έχει ισχύ που δεν είναι πολύ μικρότερη από αυτήν των άλλων δύο ελέγχων, ενώ η ισχύς της είναι συγκρίσιμη με αυτών για μεγάλα σε μέγεθος δείγματα.

Αποτελέσματα της μελέτης των Wang et al. (2018)

Μια πιο συνολική συγκριτική μελέτη διεξάγεται στο άρθρο των Wang et al. (2018), όπου αξιολογούνται τέσσερις μέθοδοι για τον έλεγχο μη κατωτερότητας με βάση το πηλίκο των πιθανοτήτων ασύμβατων ζευγών. Ειδικότερα, συγκρίνονται οι μεθοδολογίες των Liu et al. (2005) που στηρίζονται στη σ.σ. Z_b (μέθοδος δέλτα) και στη σ.σ. Z_{LFM} (score μέθοδος), η μέθοδος που προτάθηκε από τους

Jin et al. (2016) και βασίζεται στο επαγωγικό μοντέλο (IM-μέθοδος) καθώς και η προτεινόμενη από αυτούς μέθοδος του τυχαίου επαγωγικού μοντέλου (Randomized Inferential Model, RIM) με βάση τη συνάρτηση $rpl_{x_{10}, x_{01}}(\mathcal{A})$, που παρουσιάστηκε στην Ενότητα 3.2. Από τη μελέτη προσομοίωσης που διεξήγαγαν εξάγονται τα ακόλουθα συμπεράσματα.

- Λόγω του ότι βασίζονται σε ασυμπτωτικές προσεγγίσεις, τόσο η μέθοδος δέλτα όσο και η score μέθοδος, αποτυγχάνουν να ελέγξουν με ακρίβεια το σφάλμα τύπου I. Ειδικότερα, όταν το μέγεθος του δείγματος είναι αρκετά μικρό ($n \leq 30$) ή όταν η πιθανότητα εμφάνισης p_{01} είναι αρκετά χαμηλή ($p_{01} \leq 0.1$), οι δύο έλεγχοι τείνουν να ξεπερνάνε το ονομαστικό επίπεδο $\alpha = 0.05$ ή να εμφανίζονται ως συντηρητικοί, πέφτοντας κάτω από το ονομαστικό επίπεδο, αντίστοιχα. Ακόμη και σε μεγαλύτερα μεγέθη δείγματος ($n = 500$) παρατηρείται ότι οι εμπειρικές τιμές των σφαλμάτων τύπου I δεν είναι εντός των ορίων του 95% διαστήματος εμπιστοσύνης για το σφάλμα τύπου I, δηλαδή εντός του διαστήματος (0.0486, 0.0514), όταν $\alpha = 0.05$.
- Από την άλλη πλευρά, η IM-μέθοδος εμφανίζεται σταθερά συντηρητική, με κόστος τη χαμηλότερη στατιστική ισχύ συγκριτικά με τις υπόλοιπες μεθοδολογίες.
- Ωστόσο, η RIM-μέθοδος φαίνεται να διατηρεί αρκετά κοντά τα ποσοστά σφάλματος στο ονομαστικό επίπεδο $\alpha = 0.05$, εξασφαλίζοντας την εγκυρότητα της μεθόδου.
- Η ισχύς των δύο ασυμπτωτικών μεθόδων είναι ικανοποιητική σε μεγαλύτερα μεγέθη δείγματος, ωστόσο η απόδοση τους μειώνεται δραματικά όταν το μέγεθος δείγματος είναι μικρό ($n \leq 15$) ή όταν η πιθανότητα εμφάνισης ασύμβατου ζεύγους (0,1) είναι αρκετά χαμηλή.
- Παρόμοια αποτελέσματα με αυτά των δέλτα και score μεθόδων εμφανίζει και η RIM-μέθοδος για μεγάλα μεγέθη δείγματος, ενώ παράλληλα βελτιώνει σημαντικά την ισχύ και σε μικρότερα μεγέθη δείγματος. Για παράδειγμα, όταν $n = 50$ και $p_{01} = 0.05$, η ισχύς της RIM-μεθόδου είναι σχεδόν η διπλάσια από αυτή της IM-μεθόδου.

Από τα παραπάνω εξάγεται το συμπέρασμα ότι η RIM-μέθοδος υπερτερεί της μεθόδου IM. Επιπλέον, έχει σε πολλές περιπτώσεις καλύτερη απόδοση από εκείνη των ασυμπτωτικών μεθόδων και προτείνεται ειδικά για περιπτώσεις μικρών σε μέγεθος δειγμάτων.

Αποτελέσματα της μελέτης των Chen and Pan (2022)

Η συγκριτική μελέτη των Chen and Pan (2022) είναι εκείνη που συγκρίνει όλες τις διαδικασίες ελέγχου που περιγράφηκαν στο Κεφάλαιο 3. Ειδικότερα, συγκρίνονται οι μεθοδολογίες των Liu et al. (2005) που στηρίζονται στη σ.σ. Z_b (μέθοδος δέλτα) και στη σ.σ. Z_{LFM} (score μέθοδος), η μέθοδος που προτάθηκε από τους Jin et al. (2016) και βασίζεται στο επαγωγικό μοντέλο (IM-μέθοδος), η μέθοδος του τυχαίου επαγωγικού μοντέλου (Randomized Inferential Model, RMI) που προτάθηκε από τους Wang et al. (2018) και η Fiducial μέθοδος που προτάθηκε από τους Chen and Pan (2022) και αποτέλεσε αντικείμενο μελέτης της Ενότητας 3.3. Από τη μελέτη προσομοίωσης που διεξήγαγαν εξάγονται τα ακόλουθα συμπεράσματα.

- Οι δύο μεθοδολογίες που στηρίζονται στις στατιστικές συναρτήσεις Z_b και Z_{LFM} δεν διατηρούν για μικρά μεγέθη δείγματος ($n \leq 30$), το ονομαστικό επίπεδο σημαντικότητας. Ειδικότερα, παρουσιάζουν φιλελεύθερη απόδοση (μη συντηρητική), απορρίπτοντας πιο συχνά από το αναμενόμενο τη μηδενική υπόθεση, ενώ αυτή είναι αληθής. Ωστόσο, η απόδοση των δύο ελέγχων βελτιώνεται για μεγαλύτερα μεγέθη δείγματος ($n \geq 300$).
- Αντιθέτως, η IM-μέθοδος παρουσιάζεται ως συντηρητική με εμπειρικά ποσοστά σφάλματος τύπου I χαμηλότερα από 4%.
- Από την άλλη, τόσο η μέθοδος RIM όσο και η Fiducial παρείχαν ακριβή έλεγχο υπό κάθε συνθήκη, διατηρώντας τα εμπειρικά ποσοστά σφάλματος τύπου I εντός του αποδεκτού εύρους.
- Όσον αφορά την ισχύ, η score μέθοδος πέτυχε την υψηλότερη ισχύ, ειδικότερα σε μεγαλύτερα μεγέθη δείγματος ($n = 400$), ενώ η μέθοδος δέλτα παρουσίασε παρόμοια αποτελέσματα (ελαφρώς λιγότερο ισχυρά) με αυτά της score μεθόδου. Ωστόσο, στα παραπάνω συμπεράσματα πρέπει να ληφθεί υπόψη η φιλελεύθερη ιδιότητά της που προηγούμενα αναφέρθηκε.
- Από την άλλη, η IM-μέθοδος εμφάνισε τη χαμηλότερη ισχύ, γεγονός που μπορεί να εξηγηθεί λόγω της συντηρητικής συμπεριφοράς της.
- Τέλος, η Fiducial μέθοδος πέτυχε υψηλά ποσοστά ισχύος ακόμη και για μικρότερα μεγέθη δείγματος, αρκετά κοντά με τα αποτελέσματα της score μεθόδου, ενώ ξεπέρασε αυτά της RIM-μεθόδου.

Συνοψίζοντας, παρότι η μέθοδος score παρουσιάζει την υψηλότερη ισχύ, δεν είναι η μόνη που προτείνεται. Ειδικότερα, καθώς η Fiducial μέθοδος διατηρεί το

ονομαστικό επίπεδο σημαντικότητας και παρέχει συγκρίσιμα αποτελέσματα ως προς την ισχύ προτείνεται να χρησιμοποιείται και αυτή.

5.3 Συγκρίσεις ελέγχων τέταρτου κεφαλαίου

Στην ενότητα αυτή θα παρουσιαστούν τα αποτελέσματα της συγκριτικής μελέτης των Tang et al. (2003) που αφορά την απόδοση ελέγχων που παρουσιάστηκαν στο τέταρτο κεφάλαιο αυτής της μεταπτυχιακής διατριβής, δηλαδή ελέγχων που ανάγονται στον έλεγχο για το πηλίκο των πιθανοτήτων θετικής διάγνωσης. Ειδικότερα, οι Tang et al. (2003) συγκρίνουν την επίδοση της μεθόδου που βασίζεται στη σ.σ. T_1 για τον έλεγχο μη κατωτερότητας με αυτές που προτάθηκαν από τους Lachenbruch and Lynch (1998) και στηρίζονται στις σ.σ. Z_1 , T_{LL} και T_{LLM} . Παραπέμπουμε στις σχέσεις (4.5), (4.8) και (4.9), αντίστοιχα, στην Ενότητα 4.1. Σε αυτό το πλαίσιο, τα αποτελέσματα των προσομοιώσεων συνοψίζονται στα συμπεράσματα που ακολουθούν.

- Η μέθοδος των Tang et al. (2003) είναι η μόνη που διατηρεί το ονομαστικό επίπεδο σημαντικότητας, καθώς τα εμπειρικά ποσοστά σφάλματος Τύπου I είναι κοντά στο $\alpha = 5\%$. Δεν υπάρχουν περιπτώσεις που έχει φιλελεύθερη συμπεριφορά, ενώ εμφανίζει συντηρητική συμπεριφορά όταν οι πιθανότητες θετικής διάγνωσης $p_{1.}$, $p_{.1}$ λαμβάνουν υψηλές τιμές και η πιθανότητα εμφάνισης ασύμβατου ζεύγους είναι χαμηλή.
- Από την άλλη, οι άλλες τρεις μεθοδολογίες παρουσιάζουν φιλελεύθερη συμπεριφορά σε αρκετές περιπτώσεις. Για παράδειγμα, σε μικρά μεγέθη δείγματος ($n = 50$), η μέθοδος T_{LLM} στην περίπτωση όπου $p_{.1} = 0.95$ και $\phi = 0.03$, εμφάνισε ποσοστά σφάλματος τύπου I ίσα με 10.7%.
- Τέλος, η μέθοδος των Tang et al. (2003) παρουσιάζει πολύ καλή ισχύ, ειδικά σε περιπτώσεις μεγάλου μεγέθους δείγματος, ή υψηλών πιθανοτήτων ($p_{1.}$ και $p_{.1}$) ή ισχυρής συσχέτισης ϕ .

Από τα παραπάνω συμπεράσματα, είναι φανερό η υπεροχή της μεθόδου των Tang et al. (2003) σε σχέση με τις τρεις μεθόδους που συγκρίθηκε, καθώς από τα αποτελέσματα που παρατέθηκαν είναι η μόνη που διατηρεί το ονομαστικό επίπεδο σημαντικότητας.

Παρατήρηση 5.1. Θα ήταν παράλειψη να μην αναφερθεί ότι οι Nam and Blackwelder (2002) διεξήγαγαν μελέτη προσομοίωσης για την αξιολόγηση της απόδοσης των ελέγχων που πρότειναν στη βάση των στατιστικών συναρτήσεων

Z_W και Z_{CML} των σχέσεων (4.10) και (4.13), αντίστοιχα. Από τη μελέτη τους προέκυψε ότι τα εμπειρικά επίπεδα σημαντικότητας του ελέγχου που στηρίζεται στη σ.σ. Z_{CML} είναι πολύ κοντά στο ονομαστικό επίπεδο σημαντικότητας $\alpha = 5\%$, ενώ σε ορισμένες περιπτώσεις, όπως όταν το μέγεθος του δείγματος είναι μικρό ($n = 25$) παρουσίασε ελαφρώς συντηρητική συμπεριφορά. Αντίθετα, η μέθοδος που στηρίζεται στη σ.σ. Z_W είναι αντισυντηρητική, καθώς υπερβαίνει το ονομαστικό επίπεδο σημαντικότητας, με το σφάλμα Τύπου I να φτάνει ως και 0.093 στην περίπτωση μικρού μεγέθους δείγματος ($n = 25$). Τέλος, συμπεραίνουν ότι η μέθοδος που στηρίζεται στη σ.σ. Z_{CML} υπερτερεί της μεθόδου που στηρίζεται στη σ.σ. Z_W , καθώς η υψηλότερη ισχύς της δεύτερης οφείλεται στη φιλελεύθερη συμπεριφορά της.

ΚΕΦΑΛΑΙΟ 6

ΕΠΙΛΟΓΟΣ

Σε πολλές μελέτες της ιατρικής, η σύγκριση νέων διαγνωστικών μεθόδων με υπάρχουσες, και ιδίως με τον “χρυσό κανόνα”, δεν αποσκοπεί στην ανάδειξη στατιστικά σημαντικών διαφορών, αλλά στην αξιολόγηση του κατά πόσο μπορεί να θεωρηθεί ισοδύναμη ή τουλάχιστον μη κατώτερη σε σχέση με καθιερωμένες και κλινικά επαληθευμένες πρακτικές. Η ανάγκη αυτή δημιουργεί ένα επιδημιολογικό πλαίσιο, στο οποίο οι παραδοσιακοί έλεγχοι διαφοράς, όπως ο έλεγχος McNemar, είναι ανεπαρκείς. Σε αυτό το πλαίσιο, οι έλεγχοι ισοδυναμίας ή μη κατωτερότητας για ζεύγη δίτιμων δεδομένων, που αναλύθηκαν εκτενώς στα Κεφάλαια 2 έως 4, προσφέρουν ένα κατάλληλο στατιστικό εργαλείο στη μελέτη του ερευνητικού μας προβλήματος, ιδιαίτερα στις περιπτώσεις όπου η νέα διαγνωστική μέθοδος προσφέρει πλεονεκτήματα, όπως χαμηλότερο κόστος ή ταχύτερη διάγνωση. Επισημαίνεται ότι οι κώδικες της R που παρατίθενται σε κάθε ενότητα επιτρέπουν την υλοποίηση των ελέγχων σε πρακτικές εφαρμογές. Σε αυτό το πλαίσιο, εφαρμόστηκαν οι προτεινόμενες διαδικασίες στα δεδομένα του Πίνακα 2.1. Ειδικότερα, χρησιμοποιώντας τις σ.σ. που παρουσιάστηκαν συμπεραίνεται η αποτελεσματικότητα της προτεινόμενης μεθόδου, δηλαδή κατά πόσο είναι ισοδύναμη, μη κατώτερη ή μη ανώτερη από την καθιερωμένη διαδικασία με την οποία συγκρίνεται. Τα αποτελέσματα αυτά, καθώς και η στατιστική συνάρτηση που χρησιμοποιήθηκε για την αξιολόγηση κάθε μεθόδου, καταγράφονται στον Πίνακα 6.1. Τα αποτελέσματα συγκριτικών μελετών ως προς τη διατήρηση του ονομαστικού επιπέδου σημαντικότητας και ως προς την ισχύ αυτών παρουσιάστηκαν στο Κεφάλαιο 5, αναδεικνύοντας αδυναμίες και πλεονεκτήματα για κάποιους εξ αυτών των ελέγχων. Ειδικότερα, από τα αποτελέσματα των προσομοιώσεων που έχουν συμπεριληφθεί στις δημοσιευμένες εργασίες που αναφέρθηκαν στο προηγούμενο κεφάλαιο της μεταπτυχιακής διατριβής συμπεραίνουμε όσα διατυπώνονται στη συνέχεια.

- Από τους ελέγχους του Κεφαλαίου 2 προτείνεται η χρήση της μεθόδου score του Nam (1997) με στατιστική συνάρτηση Z_s , αλλά και η χρήση της

ΚΕΦΑΛΑΙΟ 6

Εργασία αναφοράς	Σ.Σ.	Αποτέλεσμα ελέγχου
Lu and Bean (1995)	$Z_{LB}^{(1)}$	Μη κατώτερη
Lu and Bean (1995)	$Z_{LB}^{(2)}$	Μη κατώτερη
Nam (1997)	Z_s	Μη κατώτερη
Nam (1997)	Z_t	Μη κατώτερη
Liu et al. (2002)	Z_ℓ, Z_u	Δεν είναι ισοδύναμη
Liu et al. (2002)	$Z_\ell^{(1)}, Z_u^{(1)}$	Δεν είναι ισοδύναμη
Chen et al. (2023)	p_f	Μη κατώτερη
Tango (1998)	T	Ισοδύναμη
Liu et al. (2005)	Z_D	Μη ανώτερη
Liu et al. (2005)	Z_{LFM}	Μη ανώτερη
Jin et al. (2016)	$pl(\mathcal{A})$	Μη ανώτερη
Wang et al. (2018)	$rpl(\mathcal{A})$	Μη ανώτερη
Chen and Pan (2022)	p_f	Μη ανώτερη
Lachenbruch and Lynch (1998)	Z_ℓ, Z_u	Δεν είναι ισοδύναμη
Lachenbruch and Lynch (1998)	T_{LL}	Μη κατώτερη
Lachenbruch and Lynch (1998)	T_{LLM}	Μη κατώτερη
Nam and Blackwelder (2002)	Z_W	Μη κατώτερη
Nam and Blackwelder (2002)	Z_{CML}	Μη κατώτερη
Tang (2003)	T_1, T_2	Δεν είναι ισοδύναμη

Πίνακας 6.1: Αποτελέσματα ελέγχων στα δεδομένα του Πίνακα 2.1.

μεθόδου του Tango (1998) με στατιστική συνάρτηση T (ιδίως για μεγάλα μεγέθη δείγματος), ενώ δεν προτείνεται η χρήση του ελέγχου των Lu and Bean (1995) που στηρίζεται στη στατιστική συνάρτηση $Z_{LB}^{(1)}$. Επισημαίνεται ωστόσο ότι είναι ανοικτό θέμα προς διερεύνηση η συγκριτική μελέτη όλων των μεθόδων που παρουσιάστηκαν σε αυτό το κεφάλαιο.

- Από τους ελέγχους του Κεφαλαίου 3 προτείνεται η χρήση τόσο της μεθόδου score των Liu et al. (2005) που στηρίζεται στη στατιστική συνάρτηση Z_{LFM} , όσο και της Fiducial μεθόδου των Chen and Pan (2022), ενώ δεν προτείνεται η χρήση της IM-μεθόδου των Jin et al. (2016).
- Από τους ελέγχους του Κεφαλαίου 4 προτείνεται η χρήση της μεθόδου των Tang et al. (2003) με στατιστική συνάρτηση T_1 , ενώ δεν προτείνεται η χρήση των μεθόδων που αναλύθηκαν από τους Lachenbruch and Lynch (1998).

Στο σημείο αυτό θα ήταν παράλειψη να μην αναφερθεί ότι σε αυτήν τη μεταπτυχιακή διατριβή παρουσιάστηκαν οι κυριότερες μεθοδολογίες ελέγχου για την περίπτωση εξαρτημένων δίτιμων δεδομένων. Αυτό σημαίνει ότι πιθανά δεν εξαντλήθηκε ο κατάλογος τέτοιων μεθοδολογιών, ενώ επισημαίνεται ότι υπάρχουν μεθοδολογίες για τον έλεγχο ισοδυναμίας ή μη κατωτερότητας ή μη ανωτερότητας για ανεξάρτητα δίτιμα δεδομένα, για συνεχή δεδομένα, για καμπύλες ROC, για πολυωνυμικά δεδομένα ή ακόμη και για δεδομένα χρόνου επιβίωσης (time-to-event). Ενδεικτικά, για ανάλογα προβλήματα με ανεξάρτητα δίτιμα δεδομένα, η εργασία των Katz et al. (1978) προσφέρει μια ολοκληρωμένη σύγκριση των μεθόδων, που βασίζονται σε διαστήματα εμπιστοσύνης για τον λόγο κινδύνου (risk ratio). Μεταξύ άλλων, αναφέρουμε τους ελέγχους των Mascha and Sessler (2011) για συνεχή δεδομένα όπου επικεντρώνονται στη μέση διαφορά μεταξύ δύο μεθόδων, ενώ έχουν αναπτυχθεί προσεγγίσεις ακόμη και με πιο εύληκτο καθορισμό περιθωρίου ισοδυναμίας, όπως στην εργασία των Sandie et al. (2022). Χαρακτηριστικό παράδειγμα εφαρμογής του ελέγχου μη κατωτερότητας σε συνεχή δεδομένα μέσω της Μπεϋζιανής (Bayesian) προσέγγισης, αποτελεί η εργασία των Ghosh et al. (2011), μιας δοκιμής τριών ομάδων που συγκρίνει ένα εικονικό φάρμακο (placebo), την καθιερωμένη θεραπεία (reference) και την προτεινόμενη μέθοδο. Σε τέτοιου σύνθετου τύπου μελέτες, εξετάζεται η υπεροχή (superiority) της καθιερωμένης θεραπείας έναντι του placebo, καθώς ταυτόχρονα γίνεται και η σύγκριση της με την προτεινόμενη διαδικασία. Αξίζει να αναφερθεί ότι για δεδομένα που αποτιμώνται μέσω ROC καμπυλών, οι εργασίες των Liu (2006) και DeLong et al. (1988) αναλύουν μεθόδους για την αξιολόγηση των διαγνωστικών ελέγχων, λαμβάνοντας υπόψη τους δείκτες της ευαισθησίας και της ειδικότητας. Επιπλέον, σε περιπτώσεις όπου οι εκβάσεις αφορούν τον χρόνο μέχρι την πραγματοποίηση του συμβάντος, έχουν προταθεί εξειδικευμένες μέθοδοι ελέγχου ισοδυναμίας για λογοκριμένα δεδομένα (censored data), όπως στη μελέτη των Com-Nougue et al. (1993) για τους διαγνωστικούς ελέγχους του λεμφώματος non Hodgkin.

Παρά τη μεθοδολογική ανάπτυξη που έχει σημειωθεί, ο έλεγχος ισοδυναμίας ή μη κατωτερότητας σε ζεύγη δίτιμων δεδομένων εξακολουθεί να αντιμετωπίζει σημαντικές προκλήσεις. Ένα βασικό ζήτημα έγκειται στην επιλογή του ορίου ισοδυναμίας, δηλαδή της παραμέτρου που καθορίζει το αποτέλεσμα του ελέγχου, αλλά συχνά βασίζεται σε εμπειρικά ή υποκειμενικά (κλινικά) κριτήρια, ενώ ταυτόχρονα η πλειοψηφία των μεθόδων δεν ενσωματώνει την αβεβαιότητα της εξάρτησης μεταξύ των ζευγών. Οι περιορισμοί αυτοί μπορούν να αρθούν με εφαρμογή εναλλακτικών στατιστικών προσεγγίσεων, όπως για παράδειγμα μέσω της Μπεϋζιανής προσέγγισης. Για περισσότερες πληροφορίες, παραπέμπουμε μεταξύ άλλων στις εργασίες των Walker and Nowacki (2011), Kelter (2020) και

Shi and Bai (2008).

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Altman, D. G. and Bland, J. M. (1994a). Diagnostic tests 1: sensitivity and specificity. *BMJ*, 308(6943):1552.
- Altman, D. G. and Bland, M. J. (1994b). *Diagnostic Tests 2: Predictive values*. BMJ Publishing Group.
- Bartlett, M. S. (1953). Approximate confidence intervals.ii. more than one unknown parameter. *Biometrika*, 40(12):306–317.
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., and et al. (2015). Stard 2015: An updated list of essential items for reporting diagnostic accuracy studies. *British Medical Journal*, 351:h5527.
- Chen, C. and Pan, H. (2022). A fiducial test for assessing the non-inferiority of odds ratio in matched-pairs design. *Communications in Statistics - Simulation and Computation*, 51(6):2978–2991.
- Chen, C., Ye, S., Xu, J., Wan, K., and Chen, Y. (2023). A fiducial approach for testing the non-inferiority of proportion difference in matched-pairs design. *Pharmaceutical Statistics*, 22(3):520–530.
- Com-Nougue, C., Rodary, C., and Patte, C. (1993). How to establish equivalence when data are censored: A randomized trial of treatments for b non-hodgkin lymphoma. *Statistics in Medicine*, 12:1353–1364.
- DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44(3):837–845.

- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.
- Edwards, A. L. (1948). Note on the correction for continuity in testing the significance of the difference between correlated proportions. *Psychometrika*, 13(3):185–187.
- Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., and Ji, W. (2020). Sensitivity of chest ct for covid-19: Comparison to rt-pcr. *Radiology*, 296(2).
- Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London*, 222:309–368.
- Fleiss, J. L., Levin, B., and Paik, M. C. (2003). *Statistical Methods for Rates and Proportions*. Wiley, Hoboken, NJ, 3rd edition.
- Ghosh, P., Nathoo, F., Gönen, M., and Tiwari, R. C. (2011). Assessing non-inferiority in a three-arm trial using the bayesian approach. *Statistics in Medicine*, 30:1795–1808.
- Griner, P. F., Mayewski, R. J., Mushlin, A. I., and Greenland, P. (1981). Selection and interpretation of diagnostic tests and procedures. principles and applications. *Annals of Internal Medicine*, 94(4 Pt 1):557–592.
- Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36.
- Jack, C. R. et al. (2018). NIA-AA research framework: Toward a biological definition of alzheimer’s disease. *Alzheimer’s & Dementia*, 14(4):535–562.
- Jin, H., Li, S., and Jin, Y. (2016). The IM-based method for testing the non-inferiority of odds ratio in matched-pairs design. *Statistics and Probability Letters*, 109:145–151.
- Katz, D., Baptista, J., Azen, S. P., and Pike, M. C. (1978). Obtaining confidence intervals for the risk ratio in cohort studies. *Biometrics*, 34(3):469–474.
- Kelter, R. (2020). Bayesian alternatives to null hypothesis significance testing in biomedical research: A non-technical introduction to bayesian hypothesis testing with jasp. *BMC Medical Research Methodology*, 20(1):1–12.
- Krömker, V. et al. (2021). Non-inferiority trial investigating the efficacy of non-steroidal anti-inflammatory drugs and antimicrobial treatment of mild

- to moderate clinical mastitis in dairy cows with long-lasting udder diseases. *Frontiers in Veterinary Science*, 8:1–10.
- Lachenbruch, P. A. and Lynch, C. J. (1998). Assessing screening tests: Extensions of McNemar’s test. *Statistics in Medicine*, 17(19):2207–2217.
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.
- Lewis, J. A. (1999). Statistical principles for clinical trials (ich e9) an introductory note on an international guideline. *Statistics in Medicine*, 18:1903–1942.
- Liu, J.-P. (2006). Tests of equivalence and non-inferiority for diagnostic accuracy based on the paired areas under roc curves. *Statistics in Medicine*, 25(7):1219–1238.
- Liu, J.-P., Chow, S.-C., and Li, G. (2002). Tests for equivalence or non-inferiority for paired binary data. *Statistics in Medicine*, 21(2):231–245.
- Liu, J. P., Fan, H. Y., and Ma, M. C. (2005). Tests for equivalence based on odds ratio for matched-pair design. *Journal of Biopharmaceutical Statistics*, 15(6):889–901.
- Lu, Y. and Bean, J. A. (1995). On the sample size for one-sided equivalence of sensitivities based upon mcnemar’s test. *Statistics in Medicine*, 14(16):1831–1839.
- Martin, R. and Liu, C. (2013a). Correction. *Journal of the American Statistical Association*, 108(503):1138–1139.
- Martin, R. and Liu, C. (2013b). Inferential models: A framework for prior-free posterior probabilistic inference. *Journal of the American Statistical Association*, 108(501):301–313.
- Mascha, E. J. and Sessler, D. I. (2011). Equivalence and noninferiority testing in regression models and repeated-measures designs. *Anesthesia & Analgesia*, 112:678–687.
- McKinney, S. M. et al. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788):89–94.
- McNeil, B. J., Keeler, E., and Adelstein, S. J. (1975). Primer on certain elements of medical decision making. *New England Journal of Medicine*, 293:211–215.

- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2):153–157.
- Nam, J.-m. (1997). Establishing equivalence of two treatments and sample size requirements in matched-pairs design. *Biometrics*, 53(4):1422–1430.
- Nam, J.-m. and Blackwelder, W. C. (2002). Analysis of the ratio of marginal probabilities in a matched-pair setting. *Statistics in Medicine*, 21:689–699.
- Peeling, R. W., Olliaro, P., Boeras, D. I., and Fongwen, N. (2017). Reassured diagnostics to inform disease control strategies, strengthen health systems and improve patient outcomes. *Nature Microbiology*, 2(8):1–6.
- Sandie, A. B., Molinari, N., Wanjoya, A., Kouanfack, C., Laurent, C., and Tchatchueng-Mbougua, J. B. (2022). Non-inferiority test for a continuous variable with a flexible margin in an active controlled trial: an application to the “Stratall ANRS 12110 / ESTHER” trial. *Trials*, 23(1):202.
- Schumi, J. and Wittes, J. T. (2011). Through the looking glass: understanding non-inferiority. *Trials*, 12(106):1–12.
- Shi, L. and Bai, P. (2008). Bayesian confidence interval for the difference of two proportions in the matched-paired design. *Communications in Statistics - Theory and Methods*, 37(13):2034–2051.
- Smith, J. A. and Doe, J. (2018). Safety considerations in diagnostic testing. *Journal of Clinical Diagnostics*, 12(3):123–130.
- Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. *Kongelige Danske Videnskabernes Selskab*, 5(4):1–34.
- Syed, S., Rahaman, A., Mondal, A., Shaligram, S., and Pawar, S. P. (2024). Diagnosis of infectious diseases: complexity to convenience. *Sensors & Diagnostics*, 3:354–380.
- Tang, N.-S., Tang, M.-L., and Chan, I. S. F. (2003). On tests of equivalence via non-unity relative risk for matched-pair design. *Statistics in Medicine*, 22:1217–1233.
- Tango, T. (1998). Equivalence test and confidence interval for the difference in proportions for the paired-sample design. *Statistics in Medicine*, 17:891–908.

- Vanbelle, S. (2024). Statistical inference for agreement between multiple raters on a binary scale. *British Journal of Mathematical and Statistical Psychology*, 77(2):245–260.
- Walker, E. and Nowacki, A. S. (2011). Understanding equivalence and noninferiority testing. *Journal of General Internal Medicine*, 26(2):192–196.
- Wang, Z., Jin, H., Lu, H., and Jin, Y. (2018). An efficient test based on the inferential model for the non-inferiority of odds ratio in matched-pairs design. *Statistical Methods in Medical Research*, 27:2831–2841.
- Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority*. CRC press.
- Zhou, X.-H., Obuchowski, N. A., and McClish, D. K. (2011). *Comparing the Accuracy of Two Diagnostic Tests*, chapter 5, pages 165–192. John Wiley and Sons, Ltd.