Εθνικο Καποδιστριακο Πανεπιστημιο
Αθηνων

Τμημα Μαθηματικο

Μεταπτυχιακο Προγραμμα Σπουδων στη Στατιστικη
και Επιχειρησιακη Ερευνα

Μετα-ανάλυση με μέθοδο Bootstrap

Συγγραφέας:
ΚΥΡΙΟΣ ΒΑΣΙΛΗΣ

Επιβλέπων:
ΣΙΑΝΝΗΣ ΦΩΤΙΟΣ

Αθήνα
Ιούνιος 2012

Στο σημείο αυτό θα ήθελα να αναφέρω ότι ο κώδικας (αλγόριθμος) που δημιουργήθηκε στη γλώσσα προγραμματισμού «R» είναι διαθέσιμος για κάθε ενδιαφερόμενο εφόσον ζητηθεί.

Τέλος επιθυμώ να δηλώσω ότι η ευθύνη για τα λάθη που μπορεί να υπάρχουν είναι αποκλειστικά δική μου.

Περίληψη

Σε μια μετα-ανάλυση, οι άγνωστες παράμετροι συχνά υπολογίζονται με μέγιστη πιθανοφάνεια, και τα συμπεράσματα βασίζονται στην ασυμπτωτική θεωρία. Θεωρείται ότι, προϋπόθεση για τα χαρακτηριστικά των μελετών που περιλαμβάνονται στο μοντέλο, η κατανομή μεταξύ των μελετών και οι κατανομές των σφαλμάτων δειγματοληψίας είναι κανονικές. Στην πράξη, όμως, τα δείγματα είναι πεπερασμένα, και η υπόθεση της κανονικότητας μπορεί να παραβιαστεί, ενδεχομένως να προκληθούν μεροληπτικές εκτιμήσεις και ακατάλληλα τυπικά σφάλματα. Σε αυτή την εργασία, προτείνουμε μια παραμετρική και δύο μη παραμετρικές μεθόδους bootstrap που μπορούν να χρησιμοποιηθούν για την προσαρμογή των αποτελεσμάτων της μέγιστης πιθανοφάνειας σε μετα-ανάλυση.

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Μετα-ανάλυση, Μέγεθος Αποτελέσματος, Μοντέλο σταθερών επιδράσεων, Μοντέλο τυχαίων επιδράσεων, *Bootstrap*, Προσομοίωση

Abstract

In a meta-analysis, the unknown parameters are often estimated using maximum likelihood, and inferences are based on asymptotic theory. It is assumed that, conditional on study characteristics included in the model, the between-study distribution and the sampling distributions of the effect sizes are normal. In practice, however, samples are finite, and the normality assumption may be violated, possibly resulting in biased estimates and inappropriate standard errors. In this master thesis, we propose one parametric and two nonparametric bootstrap methods that can be used to adjust the results of maximum likelihood estimation in meta-analysis.

KEY WORDS: Meta-analysis, Effect size, Fixed Effects Model, Random Effects Model, Bootstrap, Simulation

Περιεχόμενα

Κατάλογος Πινάκων

Κατάλογος Σχημάτων

Κεφάλαιο 1

Εισαγωγή

1.1 Τι είναι η μετα-ανάλυση;

Σήμερα, περισσότερο από ποτέ, ο ρυθμός με τον οποίο παράγεται και εμπλουτίζεται η γνώση είναι ταχύτατος. Η πληθώρα των δημοσιεύσεων και η παροχή μεγάλου όγκου πληροφοριών καθιστούν δύσκολη την προσπάθεια των επιστημόνων υγείας να αξιολογήσουν ότι τους ενδιαφέρει και να ασκήσουν την βασισμένη σε ενδείξεις κλινική πρακτική. Κάτω από αυτές τις συνθήκες καθίσταται πλέον επιτακτική η ανάγκη διασφάλισης της ποιότητας των επιστημονικών δημοσιεύσεων. Συγγραφείς και εκδότες είναι υποχρεωμένοι να ακολουθούν συγκεκριμένες οδηγίες και μεθόδους προκειμένου να αποκτούν αξία και εγκυρότητα οι πληροφορίες που προσφέρουν.

Η αναγνώριση της ανάγκης παροχής έγκυρων πληροφοριών απεικονίζεται στην προσπάθεια του διεθνούς οργανισμού *Cochrane Collaboration*, ο οποίος δημιουργήθηκε το 1993 με σκοπό την εξασφάλιση της εγκυρότητας των συστηματικών ανασκοπήσεων και μετα-αναλύσεων για θέματα φροντίδας υγείας. Οι ανασκοπήσεις αποτελούν δευτερογενή δημοσιεύματα και διακρίνονται σε περιγραφικές και συστηματικές. Η συστηματική ανασκόπηση, σε αντίθεση με την

περιγραφική, αποτελεί μια ερευνητική εργασία και η διεξαγωγή της βασίζεται σε συγκεκριμένη επιστημονική μεθοδολογία. Η μετα-ανάλυση είναι ουσιαστικά μια ποσοτική συστηματική ανασκόπηση. Με τον όρο μετα-ανάλυση εννοείται η ενοποίηση και η στατιστική ανάλυση δεδομένων, τα οποία προέρχονται από ανεξάρτητες μελέτες με σκοπό την εξαγωγή σαφέστερων συμπερασμάτων.

Η πρώτη μετα-ανάλυση πραγματοποιήθηκε από τον *Karl Pearson* το 1904, σε μια προσπάθεια να αντιμετωπίσει το πρόβλημα του μικρού μεγέθους δείγματος στη διαδικασία της έρευνας ενώ ο όρος μετα-ανάλυση χρησιμοποιήθηκε για πρώτη φορά το 1976 από τον ψυχολόγο *Gene Glass*. Στη δεκαετία του 1970 αναπτύχθηκαν στατιστικές τεχνικές μετα-ανάλυσης και άρχισαν να δημοσιεύονται σχετικά άρθρα, ενώ από την δεκαετία του 1980 και μετά ο αριθμός των συστηματικών ανασκοπήσεων/μετα-αναλύσεων αυξάνει με ταχύτατους ρυθμούς. Η διεξαγωγή συστηματικών ανασκοπήσεων/μετα-αναλύσεων είναι ιδιαίτερα χρήσιμη καθώς συμβάλλουν στην συνεχιζόμενη εκπαίδευση των επαγγελμάτων υγείας, στη λήψη κλινικών αποφάσεων, στη διατύπωση νέων ερευνητικών υποθέσεων και στον άρτιο σχεδιασμό πρωτοκόλλων. Το κόστος είναι ελάχιστο, συγκρινόμενο με αυτό της βασικής έρευνας, ενώ η ζήτηση από το αναγνωστικό κοινό γίνεται ολοένα και μεγαλύτερη.

Η μετα-ανάλυση περιλαμβάνει όλα τα βήματα της συστηματικής ανασκόπησης (διατύπωση του ερωτήματος, ενδελεχή αναζήτηση της σχετικής βιβλιογραφίας, αξιολόγηση και επιλογή μελετών), καθώς και δύο επιπλέον βήματα που είναι η στατιστική σύνθεση των δεδομένων των επιμέρους μελετών και η ερμηνεία των αποτελεσμάτων. Μετα-ανάλυση πραγματοποιείται όταν υπάρχουν επαρκή δεδομένα για ένα συγκεκριμένο επιστημονικό θέμα, όταν υπάρχουν μελέτες που δε διαφέρουν σημαντικά μεταξύ τους ως προς τα βασικά χαρακτηριστικά (π.χ. η έκθεση και η έκβαση έχουν μετρηθεί με παρόμοιο τρόπο). Αντίθετα, δε συνίσταται μετα-ανάλυση σε μελέτες που μειονεκτούν ως προς την μεθοδολογία,

υπόκεινται σε συστηματικά σφάλματα και από την ανάλυση προκύπτουν ασύμβατα αποτελέσματα. Η πραγματοποίηση μετα-αναλύσεων είναι ιδιαίτερα χρήσιμη ενώ στα πλεονεκτήματα της συγκαταλέγονται η αύξηση της ισχύος και της ακρίβειας, η μείωση του διαστήματος εμπιστοσύνης των εκτιμώμενων μέτρων, η ανίχνευση ή η εκτίμηση και ο συνυπολογισμός της ετερογένειας αλλά και η απάντηση σε ερωτήματα που δεν έχουν δώσει οι πρωτογενείς μελέτες.

1.2 Στάδια μετα-ανάλυσης

Τα τελευταία χρόνια, ένας αριθμός ερευνητών έχουν προτείνει συγκεκριμένες οδηγίες για την πραγματοποίηση και την παρουσίαση συστηματικής ανασκόπησης και μετα-ανάλυσης. Τα στάδια-βήματα, τα οποία, σε κάθε περίπτωση πρέπει να ακολουθούνται, περιγράφονται παρακάτω και συνοψίζονται στον πίνακα 1.1

Στάδιο 1. Διατύπωση ερευνητικού ερωτήματος

Το πιο σημαντικό βήμα, όπως ισχύει σε κάθε έρευνα, είναι η διατύπωση του ερευνητικού ερωτήματος στο οποίο ζητείται να δοθεί απάντηση με τη συγκεκριμένη ανασκόπηση ή και μετα-ανάλυση της βιβλιογραφίας. Το ερευνητικό ερώτημα πρέπει να είναι σαφές, επιστημονικά τεκμηριωμένο και κλινικά σημαντικό. Συνήθως, το ερώτημα/αρχική υπόθεση στηρίζεται σε ευρήματα προηγούμενων σχετικών μελετών και προκύπτει μετά από κριτική θεώρηση του συγκεκριμένου θέματος.

Στάδιο 2. Καθορισμός κριτηρίων εισαγωγής και αποκλεισμού μιας μελέτης

Ο καθορισμός του πληθυσμού-στόχου, καθώς και των κριτηρίων εισόδου ή αποκλεισμού μιας μελέτης στην ανασκόπηση/μετα-ανάλυση αποτελεί επίσης καθοριστικό βήμα. Με την εφαρμογή των κριτηρίων που έχουν καθοριστεί *a priori* προκύπτει ο κατάλληλος πληθυσμός. Τα συγκεκριμένα κριτήρια αναφέρονται

τόσο στα χαρακτηριστικά των συμμετεχόντων, στο είδος της παρέμβασης, στις μεθόδους σύγκρισης, στην έκβαση όσο και στο είδος των μελετών.

Στάδιο 3. Αναζήτηση βιβλιογραφίας

Αναγκαία προϋπόθεση για τη συστηματική αναζήτηση σχετικών και κατάλληλων δημοσιεύσεων είναι ο καθορισμός όρων ευρετηριασμού. Χρησιμοποιούνται λέξεις−κλειδιά και τηρείται αναλυτικός αλγόριθμος αναζήτησης και απεικόνισης των βημάτων της ανασκόπησης της βιβλιογραφίας. Προκειμένου να αυξηθεί το αποτέλεσμα της αναζήτησης και ο αριθμός των προς αξιολόγηση μελετών, μπορούν να χρησιμοποιηθούν συνώνυμες φράσεις ή και συνδυασμός λέξεων με τη χρήση των όρων «και» ή «όχι». Η αναζήτηση μπορεί να πραγματοποιηθεί σε διάφορες βάσεις δεδομένων, κάτι το οποίο πρέπει να αναφέρεται με σαφήνεια στη μεθοδολογία και να απεικονίζεται στον αλγόριθμο αναζήτησης. Οι κυριότερες ηλεκτρονικές βάσεις όπου μπορεί να πραγματοποιηθεί αναζήτηση της βιβλιογραφίας για βιοιατρικές μελέτες είναι οι παρακάτω:

$-Medline(http://www.ncbi.nlm.nih.gov/sites/entrez)$

$-Scopus(http://www.scopus.com/home.url)$

$-Embase(http://www.embase.com)$

$-CochraneLibrary(http://www.cochrane.org)$

$-CINAHL(http://www.ebscohost.com/cinahl)$

Επιπλέον, αναζήτηση μπορεί να πραγματοποιηθεί σε βιβλιογραφικές αναφορές των δημοσιεύσεων που ανακτήθηκαν, σε αρχεία περιλήψεων από συνέδρια, σε αρχεία ιδιωτικών και κρατικών οργανισμών έρευνας, καθώς και σε αρχεία φαρμακευτικών εταιρειών.

Στάδιο 4. Αξιολόγηση και επιλογή μελετών

Μετά από τη συλλογή της βιβλιογραφίας ακολουθεί η αξιολόγηση των άρθρων βάσει κριτηρίων. Γίνεται αποτίμηση της μεθοδολογικής αρτιότητας των μελετών και επιλογή αυτών που απαντούν στο ερευνητικό ερώτημα. Για την

αποφυγή της μεροληψίας του ερευνητή, η όλη διαδικασία πρέπει να πραγματοποιείται τουλάχιστον από δύο ανεξάρτητους ερευνητές, ενώ σε περίπτωση διαφωνίας επιλύεται το πρόβλημα από τρίτο ερευνητή.

Στάδιο 5. Καταγραφή των δεδομένων

Τα βασικά χαρακτηριστικά των μελετών καταγράφονται σε μια προσχεδιασμένη φόρμα ώστε να είναι δυνατή η σύγκριση και η εκτίμηση του βαθμού ομοιότητας μεταξύ τους. Η εκτίμηση του βαθμού ομοιότητας περιγράφεται με τον όρο «εκτίμηση της ετερογένειας των μελετών». Τα στοιχεία των πρωτογενών μελετών που καταγράφονται, προκύπτουν και πάλι από τη συνεργασία δυο ερευνητών για ελαχιστοποίηση των σφαλμάτων.

Στάδιο 6. Στατιστική ανάλυση

Η στατιστική ανάλυση των δεδομένων είναι το αναγκαίο βήμα για να συνεχίσει κάποιος από τη συστηματική ανασκόπηση στη μετα-ανάλυση. Σε αυτό το στάδιο, ουσιαστικά συντίθενται όλα τα αποτελέσματα των υπαρχουσών μελετών για να δοθεί στους αναγνώστες ένα συγκεντρωτικό αποτέλεσμα. Αν οι συγγραφείς παρουσιάσουν τα ευρήματα σχετικών μελετών χωριστά, χωρίς να δώσουν ένα συγκεντρωτικό αποτέλεσμα—που προκύπτει μέσα από στατιστικές διαδικασίες—τότε γίνεται αναφορά σε συστηματική ανασκόπηση της βιβλιογραφίας και όχι σε μετα-ανάλυση.

Στάδιο 7. Παρουσίαση αποτελεσμάτων

Γίνεται η παρουσίαση των βασικών χαρακτηριστικών των μελετών σε πίνακα. Παρατίθεται ο ακριβής αριθμός των ανακτηθέντων άρθρων που προέκυψε κατά τη συστηματική αναζήτηση. Στη συνέχεια, παρέχεται ο αριθμός και οι αιτίες αποκλεισμού άρθρων μετά από έλεγχο του τίτλου ή της περίληψης, καθώς και εκείνων που αποκλείστηκαν μετά από πλήρη μελέτη. Τέλος, αν έχει πραγματοποιηθεί και μετα-ανάλυση, τότε τα αποτελέσματα μπορούν να αποτυπωθούν σε γράφημα, το οποίο είναι ιδιαίτερα χρήσιμο γιατί παρέχει μια άμεση εκτίμηση

αναφορικά με την ύπαρξη ή μη ετερογένειας μεταξύ των μελετών.

Στάδιο 8. Ερμηνεία αποτελεσμάτων

Το τελευταίο βήμα της όλης προσπάθειας αποτελεί η ερμηνεία των αποτελεσμάτων, καθώς και ο έλεγχος της συνέπειας του συμπεράσματος της μετα-ανάλυσης.

Πίνακας 1.1: Βήματα της συστηματικής ανασκόπησης και μετα-ανάλυσης

Διατύπωση ερευνητικού ερωτήματος
Καθορισμός των κριτηρίων εισόδου και αποκλεισμού
Αναζήτηση σχετικής βιβλιογραφίας
Αξιολόγηση και επιλογή των μελετών
Καταγραφή των δεδομένων
Στατιστική ανάλυση (αναφέρεται μόνο στη μετα-ανάλυση)
Παρουσίαση αποτελεσμάτων
Ερμηνεία αποτελεσμάτων

1.3 Ο ρόλος της μετα-ανάλυσης στην ιατρική επιστήμη

Τα επιτεύγματα της μετα-ανάλυσης στη σφαίρα της κλινικής έρευνας είναι εντυπωσιακά. Αν και το πλήθος των κλινικών μελετών που δημοσιεύονται στα ιατρικά περιοδικά έχει αυξηθεί, πολλές δοκιμές αποτυγχάνουν να καταλήξουν σε στατιστικά σημαντικά και γενικεύσιμα συμπεράσματα για την επίδραση μιας θεραπείας. Αυτό οφείλεται κυρίως στο ότι οι μελέτες αυτές είναι είτε πολύ μικρές

είτε πολύ περιορισμένες, λόγω των κριτηρίων που ορίζονται στο πρωτόκολλο. Με τον συνδυασμό πληροφοριών από διαφορετικές μελέτες, η μετα-ανάλυση αποκτά μεγαλύτερη στατιστική ισχύ για να ανιχνεύσει την επίδραση της θεραπείας, σε σχέση με μια ανάλυση που βασίζεται μόνο σε μια κλινική μελέτη.

Συχνά μελέτες που αντιμετωπίζουν το ίδιο ιατρικό πρόβλημα καταλήγουν σε διαφορετικά συμπεράσματα, τα οποία μπορούν να προκαλέσουν σύγχυση σε εκείνους, που με βάση τη βιβλιογραφία, επιδιώκουν την καθοδήγηση στο ιατρικό ζήτημα. Η μετα-ανάλυση αποτελεί μια ελκυστική λύση αφού μπορεί να χρησιμοποιηθεί για να υπολογίσει την μέση επίδραση της θεραπείας στο σύνολο των κλινικών μελετών ή για να προσδιορίσει ένα υποσύνολο μελετών οι οποίες συνδέονται με μια ευεργετική-θετική επίδραση της θεραπείας.

Η έκρηξη των ερευνητικών στοιχείων τις τελευταίες δεκαετίες, υπό μορφή δημοσιευμένων δοκιμών, δεν επιτρέπει την αφομοίωση τους χωρίς επίσημη ανασκόπηση. Για να αξιολογηθούν τα οφέλη μιας θεραπευτικής αγωγής, θα πρέπει να βασιστούμε στο σύνολο των αξιόπιστων στοιχείων, όπως για παράδειγμα το σύνολο των τυχαιοποιημένων κλινικών δοκιμών. Επιπλέον είναι ανάγκη να εντοπιστούν εκείνες οι ιατρικές περιοχές στις οποίες υπάρχει έλλειψη επαρκών στοιχείων προκειμένου να διεξαχθούν επιπρόσθετες κλινικές μελέτες.

Στις βιοστατιστικές οδηγίες η μετα-ανάλυση θεωρείται ως επίσημη αξιολόγηση των ποσοτικών στοιχείων τα οποία προέρχονται από δύο ή περισσότερες κλινικές δοκιμές και αφορούν το ίδιο ζητούμενο. Οι τεχνικές της μετα-ανάλυσης αναγνωρίζονται και εφαρμόζονται για την διατύπωση ενός γενικού συμπεράσματος σε σχέση με την αποτελεσματικότητα ενός φαρμάκου, αλλά και για την ανάλυση των λιγότερων συχνών εκβάσεων στην αξιολόγηση ασφάλειας.

1.4 Το περιεχόμενο της διπλωματικής

Στην παρούσα διπλωματική παρουσιάζεται το μεθοδολογικό πλαίσιο της μετα-ανάλυσης, δίνοντας όμως μεγαλύτερη έμφαση όχι στις κλασσικές στατιστικές μεθόδους της μετα-ανάλυσης (βλέπε κεφάλαιο 2) αλλά στην ανάλυση με νέες μεθόδους (όπως η μέθοδος *bootstrap*). Στο επόμενο κεφάλαιο περιγράφονται οι μέθοδοι για την εκτίμηση της επίδρασης της θεραπείας σε κάθε μελέτη και παρουσιάζονται και συγκρίνονται τα δύο στατιστικά μοντέλα μετα-ανάλυσης (μοντέλο σταθερών και μοντέλο τυχαίων επιδράσεων), τα οποία έχουν προταθεί για τον υπολογισμό της ολικής επίδρασης της θεραπείας. Οι μεθοδολογίες αναφέρονται στην περίπτωση που η κάθε μελέτη περιλαμβάνει δυο ομάδες θεραπείας με σκοπό την σύγκριση της νέας θεραπευτικής μεθόδου με το αντίστοιχο *control*. Στο τρίτο κεφάλαιο γίνεται μια εισαγωγή στην μέθοδο *bootstrap* και στην ανάλυση γραμμικών μοντέλων με τη μέθοδο αυτή καθώς και την κατασκευή *bootstrap* διαστημάτων εμπιστοσύνης. Τέλος στο τέταρτο κεφάλαιο προτείνονται τρείς μέθοδοι μετα-ανάλυσης με τη μέθοδο *bootstrap*, γίνεται σύγκριση των μεθόδων αυτών με την κλασσική μέθοδο μετα-ανάλυσης με τη βοήθεια της προσομοίωσης και εξάγονται συμπεράσματα.

Κεφάλαιο 2

Μετα-ανάλυση

Κλινική μελέτη ονομάζεται κάθε έρευνα που διεξάγεται σε ανθρώπους με στόχο να δώσει απαντήσεις σε συγκεκριμένες ερωτήσεις για μια νέα ιατρική έρευνα (εμβόλια, νέες θεραπείες). Οι κλινικές μελέτες χρησιμοποιούνται για να τεκμηριώσουν την ασφάλεια και την αποτελεσματικότητα νέων φαρμάκων ή νέων ιατρικών αγωγών. Πρώτο βασικό στοιχείο σε μια κλινική δοκιμή είναι η θεραπευτική αγωγή (*treatment*). Η θεραπευτική αγωγή είναι ο τρόπος παρέμβασης στον ασθενή, με σκοπό τη βελτίωση της κατάστασης του. Σε κάθε κλινική δοκιμή υπάρχει τουλάχιστον μια θεραπευτική αγωγή. Δεύτερο βασικό στοιχείο είναι η ομάδα ελέγχου (*control/placebo groups*). Ως ψευδοφάρμακο ή αδρανής θεραπεία (*placebo*), ορίζεται η αγωγή κατά την οποία δε γίνεται ουσιαστική θεραπεία. Μια τέτοια ομάδα αναφοράς, δεν λαμβάνει ουσιαστικά καμία θεραπεία, όμως στην πράξη είναι ένα εικονικό φάρμακο, για ψυχολογικούς κυρίως λόγους.

Σε επιστημονικά πειράματα είναι συχνά χρήσιμο να γνωρίζουμε όχι μόνο αν το πείραμα έχει ένα αποτέλεσμα στατιστικά σημαντικό αλλά και το μέγεθος κάθε παρατηρούμενου αποτελέσματος. Επίδραση της θεραπείας ($Effect\ Size$-ES) είναι ένα όνομα που δίνεται σε μια οικογένεια δεικτών που μετρούν το

μέγεθος του αποτελέσματος μιας θεραπείας.

Το είδος του ES εξαρτάται από το είδος των δεδομένων (της εξαρτημένης μεταβλητής):

- Αν η μεταβλητή απόκρισης είναι συνεχής, τότε χρησιμοποιούμε δυο κυρίως μέτρα για την σύγκριση των θεραπειών: την διαφορά μέσων τιμών ($Raw\ mean\ difference$) και την τυποποιημένη διαφορά μέσων τιμών ($Standardized\ mean\ difference$).

- Αν η μεταβλητή απόκρισης είναι δίτιμη (αφορά κυρίως κλινικές δοκιμές), υπάρχουν κυρίως τρία μέτρα: η διαφορά ποσοστών ($Risk\ difference$), ο λόγος ποσοστών ($Relative\ risk$) και τέλος ο λόγος σχετικών πιθανοτήτων ($Odds\ ratio$).

- Αν η μελέτη αφορά συσχετίσεις μεταξύ δυο μεταβλητών (συνεχών) χρησιμοποιούμε στατιστικά όπως ο συντελεστής συσχέτισης του $Pearson$ ή τον z-μετασχηματισμό του $Fisher$.

2.1 Συνεχής μεταβλητή απόκρισης

Έστω ότι έχουμε k ανεξάρτητες μελέτες και ότι σε κάθε μια από αυτές υπάρχουν δυο ομάδες ασθενών, η ομάδα στην οποία εφαρμόζεται η θεραπευτική αγωγή *treatment* και στην άλλη η θεραπεία σύγκρισης *control*. Αν η μεταβλητή απόκρισης είναι συνεχής, τότε μπορούμε να υποθέσουμε ότι ακολουθεί κανονική κατανομή (βέβαια αυτή η υπόθεση αρκετές φορές δεν είναι ρεαλιστική). Επομένως μπορούμε να θεωρήσουμε ότι τα δεδομένα των ασθενών οι οποίοι ανήκουν στην ομάδα *treatment* προέρχονται από κανονική κατανομή με μέση τιμή μ_T και τυπική απόκλιση σ_T και ότι τα αντίστοιχα δεδομένα της ομάδας *control* προέρχονται από κανονική κατανομή με μέση τιμή μ_C και τυ-

πική απόκλιση σ_C . Αν n_T το πλήθος των ασθενών της ομάδας *treatment* και n_C το πλήθος της ομάδας *control* θα έχουμε ότι $y_{jT} \sim N(\mu_T, \sigma_T^2)$ και y_{jC} $\sim N(\mu_C, \sigma_C^2)$ όπου y_{jT} με j=1,2,...,n_T είναι οι αποκρίσεις των ασθενών της ομάδας *treament* και y_{jC} της ομάδας *control* για j=1,2,...,n_C.

2.1.1 Διαφορά μέσων τιμών

Ας υποθέσουμε ότι θέλουμε να εκτιμήσουμε την διαφορά των μέσων τιμών των δυο ομάδων, δηλαδή να εκτιμήσουμε την παράμετρο θ=$\mu_T - \mu_C$. Ο εκτιμητής μέγιστης πιθανοφάνειας (MLE) της παραμέτρου θ είναι ο:

$$\hat{\theta} = \bar{y}_T - \bar{y}_C \tag{2.1}$$

όπου $\bar{y}_T$, $\bar{y}_C$ οι δειγματικές μέσες τιμές των δυο ομάδων και s_T, s_C οι αντίστοιχες δειγματικές τυπικές αποκλίσεις των δυο ομάδων. Υποθέτοντας όπως συνηθίζεται στις κλινικές μελέτες ότι οι διακυμάνσεις των δυο ομάδων θεραπείας είναι άγνωστες αλλά ίσες μεταξύ τους, τότε η εκτίμηση της διακύμανσης του εκτιμητή $\hat{\theta}$ είναι:

$$V(\hat{\theta}) = \frac{n_T + n_C}{n_T \cdot n_C} s_p^2 \tag{2.2}$$

όπου

$$s_p = \sqrt{\frac{(n_T - 1)s_T^2 + (n_C - 1)s_C^2}{n_T + n_C - 2}}$$

Η διασπορά s_p^2 ονομάζεται σταθμισμένη διασπορά (*pooled variance*) και είναι αμερόληπτη εκτιμήτρια της κοινής διασποράς σ^2. Η διαφορά των μέσων τιμών έχει το πλεονέκτημα ότι μπορεί να ερμηνευτεί εύκολα, ενώ είναι κατάλληλη όταν σε όλες τις μελέτες έχει χρησιμοποιηθεί η ίδια κλίμακα μέτρησης στην καταγραφή της απόκρισης.

2.1.2 Τυποποιημένη διαφορά μέσων τιμών

Έστω τώρα ότι θέλουμε να εκτιμήσουμε την τυποποιημένη διαφορά μέσων τιμών των δυο ομάδων, δηλαδή να εκτιμήσουμε την παράμετρο θ=$\frac{\mu_T - \mu_C}{\sigma}$ όπου σ^2 είναι η κοινή διασπορά. Επειδή το θ είναι στην ουσία μια τυποποιημένη τιμή η ποσότητα $\Phi(\theta)$ αντιπροσωπεύει το ποσοστό των τιμών της ομάδας ελέγχου που είναι μικρότερες ή ίσες από την μέση τιμή της ομάδας θεραπείας. Για παράδειγμα αν θ=0.5, τότε $\Phi(\theta)$=0.69 και αυτό σημαίνει ότι οι τιμές για τα άτομα από την ομάδα θεραπείας έχουν πιθανότητα 0.69 να διαφέρουν κατά 0.5 από εκείνες των ατόμων της ομάδας ελέγχου. Ο εκτιμητής περιορισμένης μέγιστης πιθανοφάνειας ($REML$) της παραμέτρου θ είναι ο:

$$\hat{\theta} = \frac{\bar{y}_T - \bar{y}_C}{s_p} \tag{2.3}$$

όπου s_p η εκτίμηση της κοινής τυπικής απόκλισης σ. Ο εκτιμητής $\hat{\theta}$ έχει επικρατήσει στην βιβλιογραφία με το όνομα Cohen's d. Έχει διαπιστωθεί ότι για μικρό μέγεθος δείγματος ($n \leq 10$) ο εκτιμητής $\hat{\theta}$ είναι μεροληπτικός. Για αυτό έχει προταθεί αμερόληπτος εκτιμητής ο οποίος ονομάζεται Hedge's g και υπολογίζεται ως:

$$\hat{\theta} = \frac{\bar{y}_T - \bar{y}_C}{s_p}\Big(1 - \frac{3}{4(n_T + n_C) - 9}\Big) \tag{2.4}$$

όπου j ο συντελεστής διόρθωσης:

$$j = 1 - \frac{3}{4(n_T + n_C) - 9}$$

Ο συντελεστής διόρθωσης j παίρνει τιμές από 0 έως 1 και προσεγγίζει την μονάδα καθώς το μέγεθος δείγματος αυξάνει. Η εκτίμηση της διακύμανσης του εκτιμητή θ είναι:

$$V(\hat{\theta}) = \frac{n_T + n_C}{n_T \cdot n_C} + \frac{\hat{\theta}^2}{2(n_T + n_C)} \tag{2.5}$$

Η τυποποιημένη διαφορά μπορεί να χρησιμοποιηθεί όταν πρέπει να συνδυαστούν μετρήσεις για την ίδια μεταβλητή απόκρισης αλλά που έχουν χρησιμοποιηθεί διαφορετικές κλίμακες μέτρησης.

2.2 Δίτιμη μεταβλητή απόκρισης

Έστω δυο ανεξάρτητες ομάδες ασθενών, η ομάδα στην οποία εφαρμόζεται η θεραπευτική αγωγή που μας ενδιαφέρει (*treatment*) και η ομάδα στην οποία εφαρμόζεται η θεραπεία σύγκρισης (*control*). Μια δίτιμη μεταβλητή παίρνει δυο τιμές που συχνά αναφέρονται ως επιτυχία και αποτυχία. Έστω p_T η πιθανότητα επιτυχίας στην θεραπευτική αγωγή ενός ατόμου το οποίο ανήκει στην ομάδα θεραπείας (*treament*) και p_C η αντίστοιχη πιθανότητα ενός ατόμου το οποίο ανήκει στην ομάδα σύγκρισης (*control*). Αν γνωρίζουμε το σύνολο των ασθενών των δυο ομάδων έστω n_T και n_C αντίστοιχα και τον αριθμό των ασθενών στις δυο ομάδες με θετικό αποτέλεσμα στην θεραπεία, τότε έχουμε επαρκή πληροφορία για να εκτιμήσουμε την επίδραση της θεραπείας.
Έστω ότι τα δεδομένα μιας μελέτης συνοψίζονται σε πίνακα της παρακάτω μορφής:

Πίνακας 2.1: Συνοπτικά στατιστικά για δίτιμες μεταβλητές

	Επιτυχία	Αποτυχία	Σύνολο
Ομάδα *Treatment*	a	b	n_T
Ομάδα *Control*	c	d	n_C

Στις κλινικές μελέτες, όταν η αποκριτική μεταβλητή είναι δίτιμη, χρησιμοποιούμε συνήθως τρία μέτρα για την σύγκριση των θεραπειών: την διαφορά ποσοστών

($Risk\ difference$), τον λόγο ποσοστών ($Relative\ risk$) και τέλος τον λόγο σχετικών πιθανοτήτων ($Odds\ ratio$).

2.2.1 Διαφορά ποσοστών

Αν σε μια μελέτη θέλουμε να εκτιμήσουμε την διαφορά των ποσοστών επιτυχίας των δυο ομάδων, δηλαδή θέλουμε να εκτιμήσουμε την παράμετρο θ=p_T−p_C, ο εκτιμητής μέγιστης πιθανοφάνειας (MLE) του θ είναι:

$$\hat{\theta} = \frac{a}{n_T} - \frac{c}{n_C} \tag{2.6}$$

και η εκτιμώμενη διασπορά είναι:

$$V(\hat{\theta}) = \frac{a \cdot b}{n_T^3} + \frac{c \cdot d}{n_C^3} \tag{2.7}$$

Μια αρνητική διαφορά δηλώνει ότι η πιθανότητα εμφάνισης του γεγονότος είναι μεγαλύτερη στην ομάδα ελέγχου, ενώ μια θετική διαφορά δηλώνει ακριβώς το αντίθετο.

2.2.2 Λόγος ποσοστών

Αν υποθέσουμε ότι θέλουμε να εκτιμήσουμε τον λόγω ποσοστών των δυο ομάδων (RR), δηλαδή θέλουμε να εκτιμήσουμε το $\theta = \frac{p_T}{p_C}$. Ο εκτιμητής μέγιστης πιθανοφάνειας (MLE) του θ είναι:

$$\hat{\theta} = \frac{\frac{a}{n_T}}{\frac{c}{n_C}} \Rightarrow \hat{\theta} = \frac{a \cdot n_C}{c \cdot n_T}$$

Επειδή ο λογάριθμος του παραπάνω εκτιμητή, συγκλίνει πιο γρήγορα στην κανονική κατανομή, συνήθως το μέτρο σύγκρισης που χρησιμοποιούμε είναι το θ'=$\log\left(\frac{p_T}{p_C}\right)$ με αντίστοιχο εκτιμητή τον:

$$\hat{\theta'} = \log\left(\frac{a \cdot n_C}{c \cdot n_T}\right) \tag{2.8}$$

Η εκτίμηση της διακύμανσης του μέτρου θ' είναι:

$$V(\hat{\theta'}) = \frac{1}{a} - \frac{1}{n_T} + \frac{1}{c} - \frac{1}{n_C} \tag{2.9}$$

Αν θ'=0 (ή αντίστοιχα θ=1) τότε οι δυο θεραπείες είναι ισοδύναμες, ενώ αν $\theta' > 0$ (ή αντίστοιχα $\theta > 1$) η θεραπευτική αγωγή έχει μεγαλύτερο ποσοστό επιτυχίας από την ομάδα *control*.

2.2.3 Λόγος σχετικών πιθανοτήτων

Τέλος έστω ότι θέλουμε να εκτιμήσουμε τον λόγο σχετικών πιθανοτήτων (OR) των δυο ομάδων, δηλαδή θέλουμε να εκτιμήσουμε το θ=$\frac{p_T/(1-p_T)}{p_C/(1-p_C)}$. Ο εκτιμητής μέγιστης πιθανοφάνειας (MLE) του λόγου των σχετικών πιθανοτήτων είναι:

$$\hat{\theta} = \frac{a \cdot d}{b \cdot c}$$

Όπως και με τον λόγω ποσοστών, έτσι και εδώ ο λογάριθμος του λόγου των σχετικών πιθανοτήτων συγκλίνει πιο γρήγορα στην κανονική κατανομή. Επομένως το μέτρο σύγκρισης των θεραπειών που χρησιμοποιείται συνήθως είναι το θ'=$\log\left(\frac{p_T/(1-p_T)}{p_C/(1-p_C)}\right)$ με αντίστοιχο εκτιμητή τον:

$$\hat{\theta'} = \log\left(\frac{a \cdot d}{b \cdot c}\right) \tag{2.10}$$

Η εκτίμηση της διακύμανσης του θ' είναι:

$$V(\hat{\theta'}) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \tag{2.11}$$

Αν OR=1 οι δυο θεραπείες είναι ισοδύναμες. Ένα OR μεγαλύτερο από 1 δηλώνει ότι το γεγονός είναι πιο πιθανό στην 1η ομάδα ενώ ένα OR μικρότερο της μονάδας δείχνει ότι το γεγονός είναι λιγότερο πιθανό στην 1η ομάδα. Οι τιμές που παίρνει το OR πρέπει να είναι μη αρνητικές. Εδώ θα πρέπει να προσθέσουμε ότι για να εκτιμήσουμε την διασπορά θα πρέπει a, b, c, $d \neq 0$, μια

καλή πρακτική αν δεν ισχύει κάτι τέτοιο είναι να προσθέσουμε το 0.5 σε κάθε κελί πριν προχωρήσουμε στην ανάλυση. Σε περιπτώσεις που τα φαινόμενα που μελετάμε είναι σπάνια, όπως συμβαίνει αρκετές φορές σε μελέτες που αφορούν ασθένειες, ο εκτιμητής OR τείνει να γίνει ίσος με τον RR. Όταν δηλαδή αν οι τιμές a, b είναι μικρές, τότε ισχύει ότι:

$$OR \approx RR$$

2.3 Συντελεστής συσχέτισης του Pearson

Ο συντελεστής συσχέτισης r του Pearson είναι το πιο ευρέως χρησιμοποιούμενο ES όταν πρόκειται για έλεγχο συσχετίσεων μεταξύ ποσοτικών μεταβλητών. Έστω δυο τυχαίες μεταβλητές (συνεχείς) Χ και Υ με διασπορά σ_X^2 και σ_Y^2 αντίστοιχα και συνδιασπορά $\sigma_{XY} = cov(X, Y)$ ο συντελεστής συσχέτισης ρ ορίζεται ως:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Ο συντελεστής συσχέτισης ρ δεν εξαρτάται από την μονάδα μέτρησης των Χ και Υ και είναι συμμετρικός ως προς τις Χ και Υ. Η σημειακή εκτίμηση του συντελεστή συσχέτισης ρ του πληθυσμού από το δείγμα n ζευγαρωτών παρατηρήσεων των Χ και Υ γίνεται με την αντικατάσταση της συνδιασποράς σ_{XY} και των διασπορών σ_X^2 και σ_Y^2 από τις αντίστοιχες εκτιμήσεις από το δείγμα δηλαδή:

$$\hat{\rho} \equiv r = \frac{s_{XY}}{s_X s_Y} \tag{2.12}$$

Το στατιστικό r του Pearson παίρνει τιμές στο διάστημα [-1,1]. Οι χαρακτηριστικές τιμές του r ερμηνεύονται ως εξής:

- r=1: υπάρχει τέλεια θετική συσχέτιση μεταξύ των Χ και Υ,
- r=0: δεν υπάρχει καμία (γραμμική) συσχέτιση μεταξύ των Χ και Υ,

- r=-1: υπάρχει τέλεια αρνητική συσχέτιση μεταξύ των X και Υ.

Για να εκφράσουμε την συσχέτιση δυο τ.μ. χρησιμοποιούμε επίσης την ποσότητα r^2 που λέγεται και συντελεστής προσδιορισμού. Ο συντελεστής προσδιορισμού δίνει το ποσοστό μεταβλητότητας των τιμών της Υ που υπολογίζεται από τη X και είναι ένας χρήσιμος τρόπος να συνοψίσουμε την συσχέτιση δυο μεταβλητών. Αποδεικνύεται ότι ο συντελεστής r ακολουθεί ασυμπτωτικά την κανονική κατανομή με μέση τιμή ρ και διασπορά $\frac{(1-\rho^2)^2}{n}$, δηλαδή $r \sim N\Big(\rho, \frac{(1-\rho^2)^2}{n}\Big)$. Σε περίπτωση που ο συντελεστής συσχέτισης διαφέρει σημαντικά από το μηδέν, τότε όσο μικρότερο είναι το πλήθος των παρατηρήσεων και όσο μεγαλύτερη είναι η απόλυτη τιμή του συντελεστή τόσο η κατανομή του r αποκλίνει από την κανονική. Έτσι ορίζεται ο z-μετασχηματισμός του $Fisher$ με σκοπό να επιτευχθεί η κατά προσέγγιση κανονικοποίηση της κατανομής του r. Ο μετασχηματισμός αυτός είναι:

$$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \tag{2.13}$$

Ο μετασχηματισμός αυτός για τον συντελεστή συσχέτισης του $Pearson$ σταθεροποιεί την διασπορά και ελαττώνει την λοξότητα. Συνήθως χρησιμοποιούμε το μετασχηματισμό αυτό για δείγματα με πλήθος άνω των 10 παρατηρήσεων. Η κατανομή του z είναι κατά προσέγγιση κανονική με:

$$z \sim N\left(\frac{1}{2}\ln\left(\frac{1+r}{1-r}\right), \frac{1}{n-3}\right) \tag{2.14}$$

2.4 Μοντέλα μετα-ανάλυσης

Έστω ότι σε κάθε μια από τις k μελέτες έχουμε υπολογίσει (εκτιμήσει) τις ποσότητες $\hat{\theta}_i$, i=1,2,.....,k οι οποίες εκφράζουν την εκτιμώμενη επίδραση της θεραπείας σε σχέση με το $control$. Στη συνέχεια θα πρέπει να συνδυάσουμε όλη αυτή την πληροφορία προκειμένου να υπολογίσουμε την ολική επίδραση θ

της θεραπείας. Ο εκτιμητής της ολικής επίδρασης της θεραπείας προκύπτει με βάση τον σταθμισμένο μέσο όρο των επιδράσεων της θεραπείας στις k μελέτες.

Στην ενότητα αυτή θα παρουσιάσουμε τα δυο βασικά μοντέλα μετα-ανάλυσης με τα οποία μπορούμε να συνδυάσουμε τις αρχικές μελέτες: το μοντέλο σταθερών επιδράσεων ($fixed\text{-}effects\ models$) και το μοντέλο τυχαίων επιδράσεων ($random\text{-}effects\ models$). Στο μοντέλο σταθερών επιδράσεων υποθέτουμε ότι η επίδραση της θεραπείας είναι ίδια σε όλες τις μελέτες, ενώ στο μοντέλο τυχαίων επιδράσεων η επίδραση της θεραπείας στις k μελέτες διαφέρει.

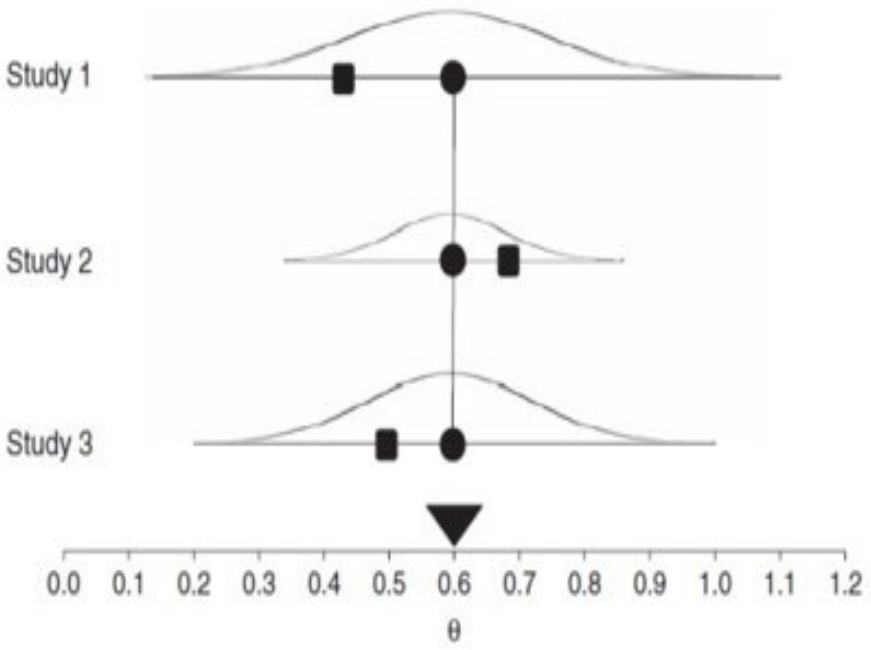
2.4.1 Μοντέλα σταθερών επιδράσεων

Στο σταθερό μοντέλο επιδράσεων ($fixed\text{-}effects\ models$) υποθέτουμε ότι η επίδραση της θεραπείας είναι ίδια σε όλες τις μελέτες. Δηλαδή το μοντέλο αυτό θεωρεί ότι όλοι οι παράγοντες που θα μπορούσαν να επηρεάσουν το μέγεθος της επίδρασης της θεραπείας είναι οι ίδιοι σε όλες τις μελέτες. Επομένως αν θ η πραγματική μέση ολική επίδραση της θεραπείας και θ_1, θ_2,...,θ_k οι πραγματικές τιμές της επίδρασης της θεραπείας για τις k μελέτες αντίστοιχα τότε: $\theta_1{=}\theta_2{=}...{=}\theta_k{=}\theta$.

Στο γράφημα (2.1) η πραγματική ολική επίδραση είναι 0.60 και εμφανίζεται στο κάτω μέρος (αναπαριστάται με ένα τρίγωνο). Η πραγματική επίδραση για κάθε μελέτη αναπαριστάται με ένα κύκλο και σύμφωνα με την παραδοχή του σταθερού μοντέλου θα είναι 0.60 για όλες τις μελέτες.

Δεδομένου ότι όλες οι μελέτες έχουν το ίδιο πραγματικό μέγεθος της επίδρασης, συνεπάγεται ότι η εκτίμηση της επίδρασης διαφέρει από την μια μελέτη στην άλλη λόγω του τυχαίου σφάλματος που υπάρχει σε κάθε μελέτη. Εάν κάθε μελέτη είχε άπειρο μέγεθος δείγματος το σφάλμα δειγματοληψίας θα είναι μηδενικό και η εκτίμηση της επίδρασης της θεραπείας θα ήταν ίδια με την πραγματική τιμή. Στην πράξη βέβαια αυτό δε συμβαίνει και έτσι υπάρχει σφάλμα

δειγματοληψίας και η εκτίμηση (στο γράφημα απεικονίζεται με τετράγωνο) δεν είναι ίδια με την πραγματική τιμή. Μια υπόθεση που συχνά γίνεται στη θεωρία για την κατανομή του δειγματικού σφάλματος είναι ότι ακολουθεί την κανονική κατανομή, με μέσο μηδέν και διακύμανση σ_i^2 για i=1,2,..k. Στο γράφημα έχουμε σχεδιάσει μια κανονική κατανομή για το πραγματικό μέγεθος της επίδρασης για κάθε μελέτη, με το πλάτος της καμπύλης να βασίζεται στην διακύμανση της κάθε μελέτης. Στη μελέτη 1 το μέγεθος του δείγματος είναι μικρό, άρα η διακύμανση θα είναι μεγάλη και η εκτίμηση της επίδρασης είναι πιθανό να πέσει σε ένα ευρύ φάσμα τιμών. Αντίθετα στη μελέτη 2, το μέγεθος του δείγματος είναι μεγάλο και η εκτίμηση είναι πιθανό να πάρει τιμές σε ένα πιο στενό φάσμα τιμών (μικρή διακύμανση).



Σχήμα 2.1: *Fixed-effect model. Πηγή: Introduction to Meta Analysis [2009]*

Επομένως η εκτίμηση της επίδρασης της θεραπείας για κάθε μελέτη είναι:

$$\hat{\theta}_i = \theta + \epsilon_i \tag{2.15}$$

όπου ϵ_i τα σφάλματα που ακολουθούν κανονική κατανομή με μέση τιμή μηδέν και διακύμανση σ_i^2 δηλαδή ισχύει $\epsilon_i \sim N(0, \sigma_i^2)$.
Άρα θα ισχύει ότι:

$$\hat{\theta}_i \sim N(\theta, \sigma_i^2)$$

Είναι λογικό ότι οι εκτιμητές $\hat{\theta}_i$ που προέρχονται από μεγάλου μεγέθους μελέτες θα είναι πιο ακριβείς και για αυτό θα έχουν μεγαλύτερη βαρύτητα. Για να εκτιμήσουμε την ολική επίδραση της θεραπείας θα υπολογίσουμε ένα σταθμισμένο μέσο, όπου το βάρος w_i για κάθε μελέτη είναι το αντίστροφο της διακύμανση της, δηλαδή:

$$w_i = \frac{1}{V(\hat{\theta}_i)} = \frac{1}{\sigma_i^2}$$

όπου σ_i^2 η κοινή διακύμανση των ομάδων θεραπείας εντός της i μελέτης.
Η ολική επίδραση της θεραπείας θ υπολογίζεται ως:

$$\hat{\theta} = \frac{\sum_{i=1}^{k} w_i \hat{\theta}_i}{\sum_{i=1}^{k} w_i} \tag{2.16}$$

Απόδειξη 1 *Από την υπόθεση ότι* $\hat{\theta}_i \sim N(\theta, \sigma_i^2)$ *η συνάρτηση πυκνότητας γράφεται ως:*

$$\begin{aligned}
f(\hat{\theta}_i; \theta) &= \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{ -\frac{(\hat{\theta}_i - \theta)^2}{2\sigma_i^2} \right\} \Rightarrow \\
L(\theta) &= \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{ -\frac{(\hat{\theta}_i - \theta)^2}{2\sigma_i^2} \right\} \Rightarrow \\
L(\theta) &= (2\pi\sigma_i^2)^{-\frac{k}{2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{k} \frac{(\hat{\theta}_i - \theta)^2}{\sigma_i^2} \right\} \Rightarrow
\end{aligned}$$

$$
\begin{aligned}
l(\theta) &= \log\left[(2\pi\sigma_i^2)^{-\frac{k}{2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{k}\frac{(\hat{\theta}_i-\theta)^2}{\sigma_i^2}\right\}\right] \Rightarrow \\
l(\theta) &= -\frac{k}{2}\log(2\pi\sigma_i^2) - \frac{1}{2}\sum_{i=1}^{k}\frac{(\hat{\theta}_i-\theta)^2}{\sigma_i^2} \Rightarrow \\
\frac{dl}{d\theta} &= \sum_{i=1}^{k}\frac{(\hat{\theta}_i-\theta)}{\sigma_i^2} \Rightarrow l'(\theta)=0 \Rightarrow \\
\sum_{i=1}^{k}\frac{\hat{\theta}_i}{\sigma_i^2} &= \hat{\theta}\sum_{i=1}^{k}\frac{1}{\sigma_i^2} \Rightarrow \hat{\theta} = \frac{\sum_{i=1}^{k}\frac{\hat{\theta}_i}{\sigma_i^2}}{\sum_{i=1}^{k}\frac{1}{\sigma_i^2}} \Rightarrow \\
\hat{\theta} &= \frac{\sum_{i=1}^{k} w_i\hat{\theta}_i}{\sum_{i=1}^{k} w_i} \qquad \alpha\nu \qquad w_i = \frac{1}{\sigma_i^2}
\end{aligned}
$$

Η διακύμανση της ολικής επίδρασης υπολογίζεται ως το αντίστροφο του αθροίσματος των βαρών, δηλαδή:

$$
V(\hat{\theta}_i) = \frac{1}{\sum_{i=1}^{k} w_i} \tag{2.17}
$$

Απόδειξη 2

$$
\begin{aligned}
V(\hat{\theta}) &= V\left(\sum_{i=1}^{k} w_i\hat{\theta}_i / \sum_{i=1}^{k} w_i\right) \Rightarrow \\
V(\hat{\theta}) &= \frac{1}{\left(\sum_{i=1}^{k} w_i\right)^2}\sum_{i=1}^{k} {w_i}^2 V(\hat{\theta}_i)
\end{aligned}
$$

και αφού για το μοντέλο σταθερών επιδράσεων ισχύει ότι $\hat{\theta}_i \sim N(\theta, \sigma_i^2)$ *ή* $\hat{\theta}_i \sim N(\theta, \frac{1}{w_i})$

$$
\begin{aligned}
V(\hat{\theta}) &= \frac{1}{\Big(\sum_{i=1}^{k} w_i\Big)^2} \sum_{i=1}^{k} w_i^2 \Big(\frac{1}{w_i}\Big) \Rightarrow \\
V(\hat{\theta}) &= \frac{\sum_{i=1}^{k} w_i}{\Big(\sum_{i=1}^{k} w_i\Big)^2} \Rightarrow \\
V(\hat{\theta}) &= \frac{1}{\sum_{i=1}^{k} w_i}
\end{aligned}
$$

Το διάστημα εμπιστοσύνης (σε επίπεδο στατιστικής σημαντικότητας α) για την ολική επίδραση της θεραπείας ($\hat{\theta}$) είναι:

$$\hat{\theta} \pm z_{a/2} \cdot \sqrt{1/\sum_{i=1}^{k} w_i} \tag{2.18}$$

2.4.2 Μοντέλα τυχαίων επιδράσεων

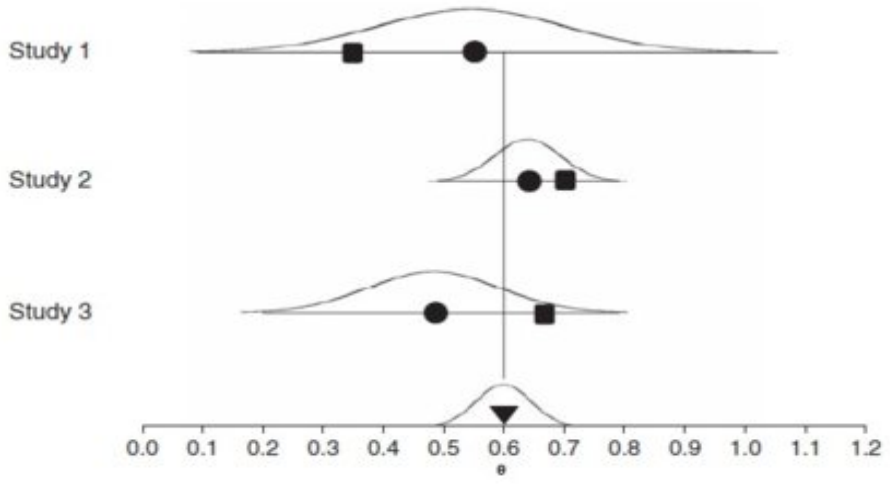
Το σταθερό μοντέλο (*fixed-effect models*) που αναφέρθηκε παραπάνω βασίζεται στην παραδοχή ότι το πραγματικό μέγεθος της επίδρασης θ_i είναι ίδιο σε όλες τις μελέτες. Ωστόσο, σε πολλές συστηματικές ανασκοπήσεις αυτή η υπόθεση είναι αβάσιμη. Όταν αποφασίζουμε να συμπεριλάβουμε μια ομάδα μελετών σε μια μετα-ανάλυση, υποθέτουμε ότι οι μελέτες έχουν αρκετά κοινά ώστε να έχει νόημα να συνθέσουμε τις πληροφορίες, αλλά δεν υπάρχει κανένας λόγος να υποθέσουμε ότι αυτές είναι ταυτόσημες με την έννοια ότι το πραγματικό μέγεθος της επίδρασης είναι ακριβώς ίδιο σε όλες τις μελέτες.

Για παράδειγμα, ας υποθέσουμε ότι έχουμε μελέτες που συγκρίνουν το ποσοστό των ασθενών που είναι θετικοί σε μια ασθένεια σε δυο ομάδες (αυτούς που έχουν εμβολιαστεί και αυτούς με το εικονικό φάρμακο) εάν το φάρμακο

είναι αποδοτικό θα περίμενε κανείς το μέγεθος της επίδρασης να είναι παρόμοιο αλλά όχι ταυτόσημο σε όλες τις μελέτες. Το μέγεθος της επίδρασης μπορεί να είναι υψηλότερο (ή χαμηλότερο), όταν οι συμμετέχοντες είναι μεγαλύτερης ηλικίας ή πιο μορφωμένοι κ.α.

Ένας τρόπος για την αντιμετώπιση αυτών των διαφορών μεταξύ των μελετών είναι να χρησιμοποιήσουμε το τυχαίο μοντέλο μετα-ανάλυσης. Στο τυχαίο μοντέλο (*random-effect models*) συνήθως υποθέτουμε ότι οι τιμές των πραγματικών επιδράσεων για τις μελέτες κατανέμονται κανονικά.

Για παράδειγμα στο γράφημα (2.2) η μέση ολική επίδραση είναι 0.60 (παριστάνεται με ένα τρίγωνο), αλλά οι επιμέρους πραγματικές επιδράσεις (παριστάνονται με κύκλο) κατανέμονται κανονικά με μέση τιμή αυτή της ολικής επίδρασης. Όπως και στο σταθερό μοντέλο λόγω του δειγματικού σφάλματος κάθε μελέτης η εκτίμηση της επίδρασης της θεραπείας (αναπαριστάται με τετράγωνο) δεν συμπίπτει με την πραγματική τιμή.



Σχήμα 2.2: *Random-effect model. Πηγή: Introduction to Meta Analysis [2009]*

Στην περίπτωση του τυχαίου μοντέλου, εκτός της παραδοχής που είναι ίδια με το σταθερό μοντέλο ότι η κατανομή του δειγματικού σφάλματος κάθε μελέτης ακολουθεί την κανονική κατανομή με μέση τιμή μηδέν και διασπορά σ_i^2, υποθέτουμε επιπλέον ότι η κατανομή του δειγματικού σφάλματος μεταξύ των μελετών ακολουθεί επίσης κανονική κατανομή με μέση τιμή μηδέν και διασπορά τ^2, που εκφράζει την ετερογένεια μεταξύ των μελετών. Επομένως η εκτίμηση της επίδρασης της θεραπείας υπολογίζεται ως:

$$\hat{\theta}_i = \theta_i + \epsilon_i$$

όπου

$$\theta_i = \theta + u_i$$

άρα

$$\hat{\theta}_i = \theta + u_i + \epsilon_i \tag{2.19}$$

με $u_i \sim N(0, \tau^2)$ και $\epsilon_i \sim N(0, \sigma_i^2)$.

Τα σφάλματα u_i και ϵ_i είναι μεταξύ τους ανεξάρτητα. Η παράμετρος τ^2 εκφράζει τη μεταβλητότητα μεταξύ των μελετών, ενώ το σ_i^2 την μεταβλητότητα εντός της i μελέτης. Επομένως για κάθε μελέτη θα ισχύει ότι:

$$\hat{\theta}_i \sim N(\theta, \tau^2 + \sigma_i^2)$$

Κάτω από το μοντέλο τυχαίων επιδράσεων το βάρος που δίνεται σε κάθε μελέτη είναι:

$$w_i^* = \frac{1}{V(\hat{\theta}_i)} = \frac{1}{\tau^2 + \sigma_i^2}$$

Ο εκτιμητής ολικής επίδρασης για το τυχαίο μοντέλο δίνεται όπως και στο

σταθερό μοντέλο από τον τύπο:

$$\hat{\theta}^* = \frac{\sum_{i=1}^{k} w_i^* \hat{\theta}_i}{\sum_{i=1}^{k} w_i^*} \tag{2.20}$$

Απόδειξη 3 *Εχοντας υποθέσει ότι* $\hat{\theta}_i \sim N(\theta, \sigma_i^2 + \tau^2)$ *η συνάρτηση πυκνότητας γράφεται ως:*

$$\begin{aligned}
f(\hat{\theta}_i; \theta, \tau^2) &= \frac{1}{\sqrt{2\pi(\sigma_i^2+\tau^2)}} \exp\left\{ -\frac{(\hat{\theta}_i - \theta)^2}{2(\sigma_i^2+\tau^2)} \right\} \Rightarrow \\
L(\theta) &= \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi(\sigma_i^2+\tau^2)}} \exp\left\{ -\frac{(\hat{\theta}_i - \theta)^2}{2(\sigma_i^2+\tau^2)} \right\} \Rightarrow \\
L(\theta) &= [2\pi(\sigma_i^2+\tau^2)]^{-\frac{k}{2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{k} \frac{(\hat{\theta}_i - \theta)^2}{(\tau^2+\sigma_i^2)} \right\} \Rightarrow \\
l(\theta) &= \log\left[[2\pi(\sigma_i^2+\tau^2)]^{-\frac{k}{2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{k} \frac{(\hat{\theta}_i - \theta)^2}{(\tau^2+\sigma_i^2)} \right\} \right] \Rightarrow \\
l(\theta) &= -\frac{k}{2} \log\left[2\pi(\sigma_i^2+\tau^2)\right] - \frac{1}{2} \sum_{i=1}^{k} \frac{(\hat{\theta}_i - \theta)^2}{(\tau^2+\sigma_i^2)} \Rightarrow \\
\frac{dl}{d\theta} &= \sum_{i=1}^{k} \frac{(\hat{\theta}_i - \theta)}{(\sigma_i^2+\tau^2)} \Rightarrow l'(\theta) = 0 \Rightarrow \\
\hat{\theta} &= \frac{\sum_{i=1}^{k} w_i^* \hat{\theta}_i}{\sum_{i=1}^{k} w_i^*} \quad \alpha\nu \qquad w_i^* = \frac{1}{\sigma_i^2+\tau^2}
\end{aligned}$$

και η διακύμανση του είναι ίση με:

$$V(\hat{\theta}) = \frac{1}{\sum_{i=1}^{k} w_i^*} \tag{2.21}$$

Απόδειξη 4 *Ομοίως με την απόδειξη για την διακύμανση στο σταθερό μοντέλο έχουμε:*

$$\begin{aligned} V(\hat{\theta}) &= V\left(\sum_{i=1}^{k} w_i^* \hat{\theta}_i / \sum_{i=1}^{k} w_i^*\right) \Rightarrow \\ V(\hat{\theta}) &= \frac{1}{\left(\sum_{i=1}^{k} w_i^*\right)^2} \sum_{i=1}^{k} w_i^{*2} V(\hat{\theta}_i) \end{aligned}$$

και αφού για το μοντέλο τυχαίων επιδράσεων ισχύει ότι $\hat{\theta}_i \sim N(\theta, \tau^2 + \sigma_i^2)$ *ή* $\hat{\theta}_i \sim N(\theta, \frac{1}{w_i^*})$

$$\begin{aligned} V(\hat{\theta}) &= \frac{1}{\left(\sum_{i=1}^{k} w_i^*\right)^2} \sum_{i=1}^{k} w_i^{*2} \left(\frac{1}{w_i^*}\right) \Rightarrow \\ V(\hat{\theta}) &= \frac{\sum_{i=1}^{k} w_i^*}{\left(\sum_{i=1}^{k} w_i^*\right)^2} \Rightarrow \\ V(\hat{\theta}) &= \frac{1}{\sum_{i=1}^{k} w_i^*} \end{aligned}$$

Το διάστημα εμπιστοσύνης (σε επίπεδο στατιστικής σημαντικότητας α) για την ολική επίδραση της θεραπείας ($\hat{\theta}$) είναι:

$$\hat{\theta} \pm z_{a/2} \cdot \sqrt{1/\sum_{i=1}^{k} w_i^*} \tag{2.22}$$

2.4.3 Εκτίμηση του τ^2

Όπως αναφέρθηκε και παραπάνω η παράμετρος τ^2 εκφράζει την μεταβλητότητα μεταξύ των μελετών. Ο εκτιμητής του τ^2 με την μέθοδο των ροπών

προτάθηκε από τους *DerSimonian* και *Laird* το 1986. Η εκτίμηση του τ^2 δίνεται από την σχέση:

$$\hat{\tau}^2 = \begin{cases} \frac{Q-(k-1)}{C} & \text{αν} \quad Q > k-1 \\ 0 & \text{αν} \quad Q \leq k-1 \end{cases} \tag{2.23}$$

όπου

$$Q = \sum_{i=1}^{k} w_i(\hat{\theta}_i - \hat{\theta})^2$$

και

$$C = \sum_{i=1}^{k} w_i - \frac{\sum_{i=1}^{k} w_i^2}{\sum_{i=1}^{k} w_i}$$

Απόδειξη 5 *Γράφουμε την εκτίμηση της διακύμανσης του* $\hat{\theta}$ *ως:*

$$\begin{aligned} V(\hat{\theta}) &= V\left(\sum_{i=1}^{k} w_i\hat{\theta}_i / \sum_{i=1}^{k} w_i\right) \Rightarrow \\ V(\hat{\theta}) &= \frac{1}{\left(\sum_{i=1}^{k} w_i\right)^2} \sum_{i=1}^{k} w_i^2 V(\hat{\theta}_i) \Rightarrow \\ V(\hat{\theta}) &= \frac{1}{\left(\sum_{i=1}^{k} w_i\right)^2} \sum_{i=1}^{k} w_i^2 \left(\frac{1}{w_i} + \tau^2\right) \Rightarrow \\ V(\hat{\theta}) &= \frac{1}{\sum_{i=1}^{k} w_i} + \frac{\tau^2 \sum_{i=1}^{k} w_i^2}{\left(\sum_{i=1}^{k} w_i\right)^2} \end{aligned}$$

Εφαρμόζοντας την μέθοδο των ροπών για το στατιστικό Q *όπου:*

$$Q = \sum_{i=1}^{k} w_i(\hat{\theta}_i - \hat{\theta})^2 = \sum_{i=1}^{k} w_i(\hat{\theta}_i - \theta)^2 - \sum_{i=1}^{k} w_i(\hat{\theta} - \theta)^2$$

$$
\begin{aligned}
E(Q) &= Q \Rightarrow E(Q) = \sum_{i=1}^{k} w_i E(\hat{\theta}_i - \theta)^2 - \sum_{i=1}^{k} w_i E(\hat{\theta} - \theta)^2 \Rightarrow \\
E(Q) &= \sum_{i=1}^{k} w_i V(\hat{\theta}_i) - \sum_{i=1}^{k} w_i V(\hat{\theta}) \Rightarrow \\
E(Q) &= \sum_{i=1}^{k} w_i \Big(\tau^2 + \frac{1}{w_i}\Big) - \sum_{i=1}^{k} w_i \left\{ \frac{1}{\sum_{i=1}^{k} w_i} + \frac{\tau^2 \sum_{i=1}^{k} w_i^2}{\Big(\sum_{i=1}^{k} w_i\Big)^2} \right\} \Rightarrow \\
E(Q) &= \tau^2 \sum_{i=1}^{k} w_i + k - 1 - \frac{\tau^2 \sum_{i=1}^{k} w_i^2}{\sum_{i=1}^{k} w_i}
\end{aligned}
$$

και αφού

$$
\begin{aligned}
E(Q) &= Q \Rightarrow \\
Q &= \tau^2 \sum_{i=1}^{k} w_i + k - 1 - \frac{\tau^2 \sum_{i=1}^{k} w_i^2}{\sum_{i=1}^{k} w_i} \Rightarrow \\
\hat{\tau}^2 &= \frac{Q - (k-1)}{\sum_{i=1}^{k} w_i - \frac{\sum_{i=1}^{k} w_i^2}{\sum_{i=1}^{k} w_i}}
\end{aligned}
$$

Επειδή με αυτή τη μέθοδο μπορεί να καταλήξουμε σε αρνητικό εκτιμητή (αυτό συμβαίνει όταν $Q \leq k-1$), στην πράξη χρησιμοποιούμε ως εκτιμητή τον $\hat{\tau}^2 = \max(0, \hat{\tau}^2)$. Όμως πλέον ο εκτιμητής $\hat{\tau}^2$ δεν είναι αμερόληπτος εκτιμητής του τ^2. Αν η τιμή του $\hat{\tau}^2$ είναι μηδέν, τότε τα βάρη w_i^* θα είναι ίσα με τα w_i. Έτσι θα μπορούσαμε να θεωρήσουμε ότι το $Fixed\ effect$ μοντέλο είναι μια περίπτωση του $Random\ effect$ μοντέλου όταν $\hat{\tau}^2 = 0$.

Εκτιμητής μέγιστης πιθανοφάνειας (ML) του τ^2

Οι εκτιμητές μέγιστης πιθανοφάνειας των παραμέτρων θ και τ, βρίσκονται χρησιμοποιώντας μια επαναληπτική διαδικασία στην οποία αρχικά θεωρούμε ότι η παράμετρος τ είναι σταθερά και υπολογίζουμε την τιμή της παραμέτρου θ η οποία μεγιστοποιεί τον λογάριθμο της πιθανοφάνειας. Στη συνέχεια υπολογίζεται η τιμή του θ που μετά θεωρείται ως σταθερά για να υπολογιστεί η νέα τιμή τ που μεγιστοποιεί τον λογάριθμο πιθανοφάνειας. Η διαδικασία αυτή συνεχίζεται μέχρι την σύγκλιση των εκτιμητών. Έτσι ο εκτιμητής του θ στην $j+1$ επανάληψη δίνεται από την σχέση:

$$\hat{\theta}_{j+1} = \frac{\sum_{i=1}^{k} w_{ij}^{*} \hat{\theta}_i}{\sum_{i=1}^{k} w_{ij}^{*}} \qquad j = 0, 1, 2, ... \tag{2.24}$$

όπου

$$w_{ij}^{*} = \frac{1}{\hat{\sigma}_i^2 + \hat{\tau}_j^2}$$

και ο εκτιμητής του τ^2 δίνεται από την σχέση:

$$\hat{\tau}_{j+1}^2 = \frac{\sum_{i=1}^{k} (w_{ij}^{*})^2 \Big\{ (\hat{\theta}_i - \hat{\theta}_{j+1})^2 - \hat{\sigma}_i^2 \Big\}}{\sum_{i=1}^{k} (w_{ij}^{*})^2} \tag{2.25}$$

Για να αρχίσει η επαναληπτική διαδικασία θέτουμε ως αρχική τιμή $\hat{\tau}^2 = 0$. Ο εκτιμητής (ML) είναι αμερόληπτος μόνο όταν το πλήθος των μελετών είναι μεγαλό, αλλιώς είναι μεροληπτικός και συνήθως υποεκτιμά την πραγματική τιμή γιατί δεν λαμβάνει υπόψη την πληροφορία η οποία χρησιμοποιήθηκε στην εκτίμηση του θ.

Εκτιμητής περιορισμένης μέγιστης πιθανοφάνειας $(REML)$ του τ^2

Ο εκτιμητής περιορισμένης μέγιστης πιθανοφάνειας $(REML)$ είναι προτιμότερος από τον εκτιμητή μέγιστης πιθανοφάνειας γιατί λαμβάνει υπόψη του την απώλεια βαθμών ελευθερίας επειδή από το ίδιο σύνολο δεδομένων εκτιμώνται και το τ^2 και το θ. Ο εκτιμητής $(REML)$ είναι αμερόληπτος και είναι και αυτός επαναληπτικός. Οι εκτιμήσεις για το θ και το τ για την $j+1$ επανάληψη δίνονται από τις σχέσεις:

$$\hat{\theta}_{j+1} = \frac{\sum_{i=1}^{k} w_{ij}^{*} \hat{\theta}_i}{\sum_{i=1}^{k} w_{ij}^{*}} \qquad j = 0, 1, 2, ... \tag{2.26}$$

όπου

$$w_{ij}^{*} = \frac{1}{\hat{\sigma}_i^2 + \hat{\tau}_j^2}$$

και

$$\hat{\tau}_{j+1}^2 = \frac{\sum_{i=1}^{k} (w_{ij}^{*})^2 \Big\{ \frac{k(\hat{\theta}_i - \hat{\theta}_{j+1})^2}{k-1} - \hat{\sigma}_i^2 \Big\}}{\sum_{i=1}^{k} (w_{ij}^{*})^2} \tag{2.27}$$

IGLS-RIGLS εκτιμητές

Οι παραπάνω εκτιμητές του τ^2 είναι παραμετρικοί και υποθέτουν κανονικότητα, ωστόσο πολλές φορές η υπόθεση αυτή δεν είναι ρεαλιστική, επομένως χρειαζόμαστε κάποιον άλλο εκτιμητή που χωρίς την υπόθεση της κανονικότητας να εκτιμάει αμερόληπτα τις παραμέτρους τ^2 και θ. Ένας τέτοιος εκτιμητής

είναι ο *IGLS* (*Iterative Generalized Least Squares*) αλλά επειδή για μικρό δείγμα δεν είναι αμερόληπτος προτιμάται ο *RIGLS* (*Restricted Iterative Generalized Least Squares*) που είναι. Έστω ότι έχουμε το κλασσικό μοντέλο τυχαίων επιδράσεων μετα-ανάλυσης:

$$\hat{\theta}_i = \theta + u_i + \epsilon_i$$

και αφού u_i, ϵ_i ανεξάρτητα έστω e_i=u_i+ϵ_i με $E(e_i)$=0 και $V(e_i)$=τ^2+σ_i^2 (δεν υποθέτουμε κανονική κατανομή για τα κατάλοιπα). Έτσι το μοντέλο μπορεί να πάρει την μορφή:

$$\hat{\theta}_i = \theta + e_i$$

ή σε μορφή πινάκων

$$\Upsilon = X \cdot \theta + e$$

όπου

$$\Upsilon = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_k \end{bmatrix} \qquad X = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \qquad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_k \end{bmatrix}$$

και ο πίνακας συνδιακύμανσης του Υ είναι

$$V(\Upsilon) = V = \begin{bmatrix} \tau^2 + \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \tau^2 + \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tau^2 + \sigma_k^2 \end{bmatrix}$$

Αν ο V γνωστός, δηλαδή στην ουσία αν το τ^2 γνωστό αφού στην μετα-ανάλυση είναι γνωστές οι εκτιμήσεις των σ_i^2 τότε ο *GLS* εκτιμητής του θ θα δίνεται από την σχέση:

$$\hat{\theta} = (X'V^{-1}X)^{-1}(X'V^{-1}Y) \tag{2.28}$$

Αν το θ είναι γνωστό και άγνωστος ο V (δηλαδή το τ^2) τότε από τα κατάλοιπα $\hat{e}$=Y-$\hat{\theta}$ αν Y^*=$\hat{e}\hat{e}'$ έχουμε ότι $E(Y^*)$=V. Έστω Y^{**} ένα διάνυσμα από τις στήλες του Y^* τότε το τ^2 εκτιμάται από το νέο γραμμικό μοντέλο:

$$E(Y^{**}) = Z^*\theta^*$$

όπου θ^*=$(\tau^2, \sigma_1^2, ..., \sigma_k^2)'$ και Z^* ο πίνακας σχεδίασης για τις τυχαίες παραμέτρους του θ^*. Ο GLS εκτιμητής για το θ^* δίνεται από τη σχέση:

$$\theta^* = (Z^{*'}V^{*-1}Z^*)^{-1}(Z^{*'}V^{*-1}Y^{**}) \tag{2.29}$$

όπου V^* ο πίνακας συνδιακύμανσης του Y^{**}. Κάτω από την υπόθεση της κανονικότητας είναι V^*=$2V \otimes V$ όπου $\otimes$ το γινόμενο του $Kronecker$ και ισχύει ότι $V(\theta^*)$=$(Z^{*\prime}V^{*-1}Z^*)^{-1}$. Σύμφωνα με τον $Goldstein$ ο εκτιμητής $RIGLS$ κάτω από την υπόθεση της κανονικότητας είναι ταυτόσημος με τον $REML$.

Όταν ούτε το θ ούτε ο V είναι γνωστά, τότε ο $IGLS$ εκτιμητής υπολογίζεται επαναληπτικά από τις σχέσεις (2.29) και (2.30), δίνοντας για αρχική τιμή του τ^2 το μηδέν. Στη μετα-ανάλυση οι τιμές ${\hat{\sigma}_i}^2$ είναι γνωστές και δεν χρειάζεται να εκτιμηθούν με την παραπάνω διαδικασία. Σύμφωνα με τον $Goldstein$ ωστόσο αυτός ο εκτιμητής είναι μεροληπτικός για μικρά μεγέθη δείγματος ενώ ο $RIGLS$ εκτιμητής είναι αμερόληπτος. Ο $RIGLS$ υπολογίζεται όπως ο $IGLS$ επαναληπτικά αλλά σε αυτή την περίπτωση ισχύει ότι:

$$E(\Upsilon^{**}) = V - X(X'V^{-1}X)^{-1}X' \tag{2.30}$$

Αν $\hat{\tau}^2$ ˂ 0 τότε θέτουμε $\hat{\tau}^2$=0.

2.5 Σύγκριση μοντέλων

Από τα παραπάνω είδαμε ότι όταν η μεταβλητότητα μεταξύ των μελετών είναι πολύ μικρή ($\hat{\tau}^2 \approx 0$) τότε τα δυο μοντέλα συμπίπτουν. Αν όμως υπάρχει ση-

μαντική (στατιστικά) μεταβλητότητα τότε μια ανάλυση που την αγνοεί, δηλαδή το μοντέλο σταθερών επιδράσεων, θα δίνει πιο μικρά διαστήματα εμπιστοσύνης σε σχέση με την προσέγγιση των τυχαίων επιδράσεων. Επίσης το μοντέλο τυχαίων επιδράσεων είναι πιο ευαίσθητο στο σφάλμα δημοσίευσης αφού δίνει μεγάλο βάρος στις μελέτες με μικρό μέγεθος δείγματος. Άλλο ένα μειονέκτημα του μοντέλου τυχαίων επιδράσεων είναι ότι αν στην μετα-ανάλυση συμπεριλαμβάνονται λίγες μελέτες δεν μπορούμε να ισχυριστούμε πάντα ότι ο εκτιμητής της μεταβλητότητας τ^2 μεταξύ των μελετών είναι αξιόπιστος.

Τέλος αν το πλήθος των μελετών είναι μικρό δηλαδή αν $k < 20$ τότε το μοντέλο τυχαίων επιδράσεων δεν δίνει συνήθως ακριβή αποτελέσματα για αυτό σε αυτές τις περιπτώσεις η ανάλυση γίνεται βασισμένη στο σταθερό μοντέλο επιδράσεων.

2.6 Ετερογένεια

Με τον όρο στατιστική ετερογένεια των μελετών δηλώνεται η μεταβλητότητα στο μέγεθος των μεταβλητών της επίδρασης της έκθεσης/θεραπείας, η οποία δεν οφείλεται στο σφάλμα δειγματοληψίας. Με άλλα λόγια, ετερογένεια σημαίνει ότι οι παράμετροι είναι διαφορετικές μεταξύ τους. Η ετερογένεια μπορεί να οφείλεται στην ποικιλομορφία μεταξύ των διαφόρων μελετών που μπορεί να είναι κλινική (συμμετέχοντες, παρέμβαση, έκβαση), μεθοδολογική (σχεδιασμός, διεκπεραίωση μελέτης) ή και στατιστική (διακύμανση των αποτελεσμάτων μεγαλύτερη από όση αναμένεται από τύχη και μόνο). Ακόμα η ετερογένεια μπορεί να διαχωριστεί σε ποιοτική όταν η κατεύθυνση της επίδρασης της θεραπείας αντιστρέφεται, δηλαδή η θεραπεία σε κάποιες μελέτες είναι ευεργετική-θετική και σε κάποιες άλλες είναι βλαβερή-αρνητική, ενώ ποσοτική ετερογένεια έχουμε

όταν ποικίλει το μέγεθος της επίδρασης της θεραπείας αλλά όχι η κατεύθυνση της, δηλαδή είναι ευεργετική (ή όχι) σε όλες τις μελέτες αλλά σε διαφορετικό βαθμό.

2.6.1 Στατιστικός έλεγχος ετερογένειας

Στατιστικά, η ετερογένεια συνήθως ελέγχεται με το στατιστικό κριτήριο Q ($Cochran's\ Q\ statistic$) και δίνεται από τη σχέση:

$$Q = \sum_{i=i}^{k} w_i(\hat{\theta}_i - \hat{\theta})^2 = \sum_{i=1}^{k} w_i\hat{\theta}_i^2 - \frac{\left(\sum_{i=1}^{k} w_i\hat{\theta}_i\right)^2}{\sum_{i=1}^{k} w_i} \tag{2.31}$$

Κάτω από την μηδενική υπόθεση ($H_o : \theta_1 = \theta_2 = ... = \theta_k$ ή ότι $\tau^2 = 0$) το στατιστικό Q ακολουθεί ασυμπτωτικά την χ^2 κατανομή με k-1 βαθμούς ελευθερίας. Έτσι η τιμή του Q αν είναι μεγαλύτερη της κριτικής τιμής της χ^2 κατανομής για κάποιο επίπεδο σημαντικότητας α, τότε η μηδενική υπόθεση θα απορρίπτεται.

Οι $Hardy$ και $Tompson$ διαπίστωσαν ότι η ισχύς του ελέγχου είναι χαμηλή, όταν τα δεδομένα είναι αραιά ή όταν κάποια μελέτη έχει πολύ μεγαλύτερο βάρος από τις υπόλοιπες. Ενώ ένα στατιστικά σημαντικό αποτέλεσμα μας φανερώνει την ύπαρξη ετερογένειας, ένα μη στατιστικά σημαντικό αποτέλεσμα δεν θα πρέπει να θεωρηθεί ως απόδειξη για την μη ύπαρξη ετερογένειας. Οι $Higgins$ και $Thompson$ πρότειναν δυο νέους δείκτες τους:

$$H^2 = \frac{Q}{k-1} \tag{2.32}$$

και

$$I^2 = \frac{H^2 - 1}{H^2} \cdot 100\% \tag{2.33}$$

όπου Q το στατιστικό του ελέγχου ετερογένειας και k ο αριθμός των μελετών. Ο δείκτης I^2 περιγράφει το ποσοστό της μεταβλητότητας μεταξύ των μελετών το οποίο οφείλεται στην ετερογένεια και όχι στο σφάλμα δειγματοληψίας. Αν $I^2 = 0$ σημαίνει ότι δεν υπάρχει ετερογένεια, τιμές μεγαλύτερες του 0.5 θεωρείται υψηλή ετερογένεια ενώ τιμές μικρότερες του 0.25 θεωρείται ήπια ετερογένεια. Ο δείκτης H^2 εκφράζει το πόσες φορές είναι μεγαλύτερο το στατιστικό ετερογένειας Q από τους αντίστοιχους βαθμούς ελευθερίας του. Αν $H = 1$ σημαίνει ότι υπάρχει ομοιογένεια μεταξύ των μελετών (αφού $E[Q] = k-1$ όταν δεν υπάρχει ετερογένεια).

Κεφάλαιο 3

Bootstrap

3.1 Εισαγωγή

Ο *Efron* είναι ο εισηγητής της τεχνικής λήψης δείγματος *bootstrap*, η οποία έχει ασκήσει σημαντική επίδραση στον τομέα της στατιστικής, που αφορά την εκτίμηση διαφόρων παραμέτρων αλλά και σε άλλους τομείς των στατιστικών εφαρμογών. Η μέθοδος *bootstrap* είναι μια από τις πρώτες υπολογιστικές, εντατικές στατιστικές τεχνικές, που αντικαθιστούν τα παραδοσιακά αλγεβρικά αποτελέσματα χρησιμοποιώντας δεδομένα βασισμένα σε προσομοιώσεις υπολογιστών. Ο *Efron* το 1979 εισήγαγε την υπολογιστική μέθοδο *bootstrap* ως τεχνική δειγματοληψίας για τον υπολογισμό της διακύμανσης εκτιμητών και της δειγματικής κατανομής ενός στατιστικού. Η μεγάλη επιτυχία της μεθόδου οφείλεται στην εύκολη και ακριβή εκτίμηση της δειγματικής κατανομής, του τυποποιημένου σφάλματος και των διαστημάτων εμπιστοσύνης με ελάχιστες ή καμία υπόθεση για την κατανομή του πληθυσμού από το οποίο λαμβάνεται το δείγμα.

Η ιδέα είναι σχετικά απλή:

Δεν ξέρουμε την κατανομή του πληθυσμού αλλά στην πραγματικότητα έ-

χουμε ένα δείγμα από αυτήν (τα δεδομένα μας). Επομένως χρησιμοποιούμε την εμπειρική κατανομή των δεδομένων μας ως μια εκτίμηση της πραγματικής, αλλά άγνωστης, κατανομής του πληθυσμού. Εμπειρική κατανομή είναι η κατανομή που δίνει πιθανότητα $1/n$ σε κάθε μια από τις n παρατηρήσεις του δείγματος μας και 0 σε οποιαδήποτε άλλη τιμή.
Δηλαδή έχουμε:

Τιμή	X_1	X_2	X_3	$\cdots$	X_n	Οποιαδήποτε άλλη τιμή
Πιθανότητα	$1/n$	$1/n$	$1/n$	$\cdots$	$1/n$	0

Για να πάρουμε ένα δείγμα από την εμπειρική κατανομή πραγματοποιούμε δειγματοληψία με επανάθεση. (Παρατήρηση: σε ένα δείγμα μεγέθους n είναι πιθανό κάποια τιμή να εμφανίζεται περισσότερες από μια φορές, έστω για παράδειγμα 2 φορές. Σε αυτή την περίπτωση είναι κατανοητό ότι έχουμε n-1 διαφορετικές τιμές και πως αυτή η τιμή έχει πια πιθανότητα $2/n$. Αυτό δεν είναι πρόβλημα καθώς είναι ισοδύναμο με το να διαλέγουμε κάθε μια από τις n παρατηρήσεις με πιθανότητα $1/n$).

Μια πλήρης περιγραφή της δειγματοληψίας είναι η εξής: Διαλέγουμε μια πρώτη τιμή από τις παρατηρήσεις μας και μετά την επιστρέφουμε και διαλέγουμε ξανά από το σύνολο των παρατηρήσεων. Επομένως ένα δείγμα *bootstrap* μπορεί να περιέχει κάποια τιμή περισσότερες από μια φορές αλλά μπορεί να μην περιέχει κάποια άλλη τιμή. Έστω ότι έχουμε τις 10 τιμές: 1, 3, 6, 8, 9, 11, 14, 16, 19, 18. Κάνοντας δειγματοληψία με επανάθεση μπορεί το δείγμα που θα προκύψει να είναι 1, 1, 3, 3, 3, 3, 8, 14, 16, 19, δηλαδή η τιμή 3 εμφανίζεται 4 φορές, η τιμή 1 εμφανίζεται 2 φορές ενώ οι τιμές 6, 9, 11, 18 δεν εμφανίστηκαν σε αυτό το δείγμα. Είναι ευνόητο ότι καθώς η μεθοδολογία υποθέτει πως παίρνουμε αρκετά τέτοια δείγματα, τελικά όλες οι παρατηρήσεις θα εμφανίζονται με την

συχνότητα που υποθέτει η εμπειρική κατανομή. Αυτό από την άλλη σημαίνει πως πρέπει να πάρουμε αρκετά δείγματα για να έχουμε μεγαλύτερη σιγουριά πως η προσέγγιση μας είναι ικανοποιητική.

Συνεπώς η βασική ιδέα της μεθόδου *bootstrap* είναι πως κάνουμε δειγματοληψία με επανάθεση από το υπάρχον δείγμα και άρα θεωρούμε πως η εμπειρική κατανομή είναι μια καλή προσέγγιση της κατανομής του πληθυσμού. Αυτή η τελευταία υπόθεση είναι θεμελιώδης. Μπορεί να παρατηρήσει αμέσως κανείς πως όταν αυτό δεν ισχύει (π.χ μικρό μέγεθος δείγματος, πολυμεταβλητά προβλήματα κ.λ.π) η μέθοδος *bootstrap* είναι καταδικασμένη να μην δουλεύει καλά.

Η μέθοδος *bootstrap* χρησιμοποιείται κυρίως για στατιστική συμπερασματολογία. Μπορούμε να εκτιμήσουμε τυπικά σφάλματα, να κάνουμε ελέγχους υποθέσεων, ακόμα και σε ιδιαίτερα πολύπλοκες μηδενικές υποθέσεις, καθώς και να προσεγγίσουμε την κατανομή πολύπλοκων συναρτήσεων.

3.2 Εμπειρική συνάρτηση κατανομής

Έστω ότι έχουμε δείγμα X_1,X_2, ...,X_n (i.i.d) από μία άγνωστη κατανομή F και T_n=T(X_1, ...,X_n) εκτιμητής μιας άγνωστης παραμέτρου θ. Τότε η εμπειρική συνάρτηση κατανομής είναι η παρακάτω:

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x) \qquad \forall x \in \mathbf{R} \tag{3.1}$$

όπου

$$I(X_i \leq x) = \begin{cases} 0 & \text{αν} \quad X_i > x \\ 1 & \text{αν} \quad X_i \leq x \end{cases}$$

Από τον νόμο των μεγάλων αριθμών (N.M.A) [1] θα ισχύει ότι:

$$\hat{F}(x) \overset{n\to\infty}{\to} F(x) \tag{3.2}$$

[1] Θεώρημα: Έστω X_i μια ακολουθία από ανεξάρτητες, ισόνομες τυχαίες μεταβλητές, με μέσο $\mu < \infty$ και διασπορά $\sigma^2 < \infty$. Αν S_n=$X_1 + ... + X_n$ τότε: $\frac{1}{n} \cdot S_n \overset{n\to\infty}{\to} \mu$

Απόδειξη 6

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x) \Rightarrow$$

$$\hat{F}(x) = E\left(I(X_i \leq x)\right) = P(I(X_i \leq x) = 1) = P(X_i \leq x) \Rightarrow$$

$$\hat{F}(x) = F(x)$$

Η παραπάνω εκτίμηση της στατιστικής κατανομής της F είναι μη παραμετρική διότι δεν βασίζεται σε καμία υπόθεση για την κατανομή της F. Έτσι βλέπουμε ότι η $\sigma.\kappa.$ F μπορεί εύκολα να εκτιμηθεί χωρίς να κάνουμε καμία υπόθεση για την μορφή της, από την εμπειρική συνάρτηση κατανομής $\hat{F}$ από το δείγμα X_1,X_2,.....,X_n, οπότε για την εκτίμηση της θ από την $\sigma.\kappa.$ F, δηλαδή θ=θ_F θα χρησιμοποιήσουμε την $\theta_{\hat{F}}$ από την εμπειρική συνάρτηση κατανομής $\hat{F}$. Επομένως αν X^* [2] τυχαία μεταβλητή με κατανομή $\hat{F}$ έχουμε για i=1,2,.....,n ότι: $P(X^* = x_i) = \hat{F}(x_i) - \hat{F}(x_{i-1}) = \frac{1}{n}$ άρα $\mathrm{X}_i^* \sim \hat{F}$.

3.3 Τυπικά σφάλματα

Η μέθοδος *bootstrap* είναι πολύ χρήσιμη για τον υπολογισμό τυπικών σφαλμάτων διαφόρων εκτιμητριών και γενικά στατιστικών συναρτήσεων. Για να εκτιμήσουμε λοιπόν ένα τυπικό σφάλμα μιας ποσότητας $\hat{\theta}$=$T(X_1, X_2, ..., X_n)$ του δείγματος ακολουθούμε τα εξής βήματα:

1. Δημιουργούμε B bootstrap δείγματα, με δειγματοληψία με επανάθεση

[2] χρησιμοποιούμε το $*$ για να μην το μπερδεύουμε με το αρχικό δείγμα

2. Για κάθε bootstrap δείγμα υπολόγισε την τιμή $\hat{\theta}_i^*=T(X_1^*, X_2^*, ..., X_n^*)$ η οποία είναι η τιμή της συνάρτησης μας στο i bootstrap δείγμα για i=1,2,...B

3. Εκτίμησε το τυπικό σφάλμα της $\hat{\theta}$ ως:

$$se_B(\hat{\theta}) = \sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(\hat{\theta}_i^* - \bar{\hat{\theta}}^*)^2} \tag{3.3}$$

όπου

$$\bar{\hat{\theta}}^* = \frac{1}{B}\sum_{i=1}^{B}\hat{\theta}_i^*$$

Πρέπει να παρατηρήσουμε εδώ πως γενικά η $\bar{\hat{\theta}}^*$ διαφέρει από την $\hat{\theta}$ και αυτή την διαφορά θα τη χρησιμοποιήσουμε για να εκτιμήσουμε τη μεροληψία της εκτιμήτριας $\hat{\theta}$, που είναι:

$$bias(\hat{\theta}) = \bar{\hat{\theta}}^* - \hat{\theta} \tag{3.4}$$

Διόρθωση μεροληψίας -Bias Correction

Ο λόγος για τον οποίο εκτιμούμε την μεροληψία του $\hat{\theta}$ είναι για να διορθώσουμε την εκτίμηση του $\hat{\theta}$ ώστε να είναι λιγότερο μεροληπτικό. Η εκτίμηση του $\hat{\theta}$ με διορθωμένη την μεροληψία υπολογίζεται ως:

$$\hat{\theta}_{cor} = 2\hat{\theta} - \hat{\theta}^* \tag{3.5}$$

Βέβαια πολλές φορές στην πράξη δεν χρησιμοποιούμε την διόρθωση της μεροληψίας, γιατί διορθώνοντας την μεροληψία μπορεί να αυξήσει κατά πολύ το τυπικό σφάλμα.

3.4 Διαστήματα εμπιστοσύνης

Σε πολλές πρακτικές εφαρμογές η εκτίμηση της τιμής μιας παραμέτρου θ ενός πληθυσμού με την τιμή ενός στατιστικού μέτρου σε ορισμένο τυχαίο δείγμα

θεωρείται αρκετή. Υπάρχουν όμως εφαρμογές όπου θέλουμε να γνωρίζουμε πόση εμπιστοσύνη μπορούμε να έχουμε όταν χρησιμοποιούμε π.χ τη μέση τιμή ενός τυχαίου δείγματος για να εκτιμήσουμε την μέση τιμή του πληθυσμού. Στις περιπτώσεις αυτές εκτιμούμε ένα διάστημα τιμών $[l,u]$ στο οποίο θα βρίσκεται η τιμή της παραμέτρου με ορισμένη πιθανότητα 1-α, $0 < \alpha < 1$. Δηλαδή έχουμε:

$$P(l \leq \theta \leq u)=1\text{-}\alpha$$

Το $[l,u]$ ονομάζεται διάστημα εμπιστοσύνης με πιθανότητα ή επίπεδο εμπιστοσύνης 1-α. Το διάστημα εμπιστοσύνης για ορισμένη παράμετρο θ είναι μια συνάρτηση της κατανομής της τιμής του εκτιμητή $\hat{\theta}$ σε ορισμένο τυχαίο δείγμα και προσδιορίζεται με βάση την κατανομή του $\hat{\theta}$.

Για μεγάλο μέγεθος δείγματος ($n > 30$) η κατανομή του $\hat{\theta}$ τείνει στην κανονική κατανομή με μέση τιμή θ και τυπικό σφάλμα $se(\hat{\theta})$, από Κ.Ο.Θ[3] ισχύει:

$$\frac{\hat{\theta} - \theta}{se(\hat{\theta})} \sim N(0,1)$$

Συνεπώς το (1-α)% διάστημα εμπιστοσύνης για την παράμετρο θ μπορεί να υπολογιστεί ως:

$$\left[\hat{\theta} - z^{(1-\frac{\alpha}{2})}se(\hat{\theta}), \hat{\theta} - z^{(\frac{\alpha}{2})}se(\hat{\theta})\right] \tag{3.6}$$

όπου $z^{(a)}$ είναι το α-ποσοστιαίο σημείο της τυποποιημένης κανονικής κατανομής, π.χ $z^{(0.975)} = 1.96$ και ισχύει ότι $z^{(\alpha)}$=-$z^{(1-\alpha)}$.

Αν το δείγμα είναι μικρό ($n < 30$) τότε

[3]Κεντρικό οριακό θεώρημα:

Αν X_1,X_2...X_n είναι μια ακολουθία ανεξάρτητων και ισόνομων μεταβλητών με μέση τιμή μ και διακύμανση σ^2, τότε

$$Z_n=\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \to N(0,1)$$

$$\frac{\hat{\theta}-\theta}{se(\hat{\theta})} \sim t_{(n-1)}$$

και το αντίστοιχο δ.ε γίνεται:

$$\left[\hat{\theta} - t^{(n-1,1-\frac{a}{2})} se(\hat{\theta}), \hat{\theta} - t^{(n-1,\frac{a}{2})} se(\hat{\theta})\right] \tag{3.7}$$

όπου $t^{(n,\alpha)}$ είναι το α-ποσοστιαίο σημείο της Student's t κατανομής, π.χ $t^{(5,0.05)} = -2.01$ και ισχύει ότι $t^{(n,\alpha)}$=-$t^{(n,1-\alpha)}$.

Για παράδειγμα αν έχουμε μια τυχαία μεταβλητή x η οποία ακολουθεί κανονική κατανομή με μέση τιμή μ και διασπορά σ^2, δηλαδή $x \sim N(\mu, \sigma^2)$ και θέλουμε διάστημα εμπιστοσύνης για την τιμή μ (εδώ θ=μ) τότε η μέση τιμή $\bar{x}$ ($\hat{\theta} = \bar{x}$) της μεταβλητής x εύκολα αποδεικνύεται[4] ότι $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ επομένως η σχέση (3.6) γίνεται:

$$\left[\bar{x} - z^{(1-\frac{\alpha}{2})}\frac{s}{\sqrt{n}}, \bar{x} + z^{(1-\frac{\alpha}{2})}\frac{s}{\sqrt{n}}\right]$$

όπου η διασπορά σ^2 εκτιμάται από την δειγματική διασπορά s^2.

Παρακάτω θα δούμε πως μπορούμε να κατασκευάσουμε *bootstrap* διαστήματα εμπιστοσύνης, θα ασχοληθούμε με τέσσερις διαφορετικές μεθόδους που είναι: α) Τυπικά *bootstrap* διαστήματα εμπιστοσύνης β) $Bootstrap-t$ διαστήματα εμπιστοσύνης γ) Ποσοστιαία *Bootstrap* διαστήματα εμπιστοσύνης δ) BCa διαστήματα εμπιστοσύνης.

3.4.1 Τυπικά bootstrap διαστήματα εμπιστοσύνης

Αν υποθέσουμε ότι η κατανομή της εκτιμήτριας $\hat{\theta}$ είναι κανονική μια παραδοχή που ισχυριζόμαστε πολύ συχνά, ωστόσο αρκετές φορές στην πράξη αυτή η υπόθεση δεν είναι ρεαλιστική και αυτό μπορεί να συμβεί π.χ. αν το μέγεθος

[4] Έστω X δείγμα απο $N(\mu, \sigma^2)$ τότε:

$E(\bar{X})$=$E(\frac{\sum_{i=1}^{n} X_i}{n})$=$\frac{1}{n}E(\sum_{i=1}^{n} X_i)$=$\frac{1}{n}\sum_{i=1}^{n}(E(X_i))$=$\frac{1}{n}n\mu$=$\mu$

$V(\bar{X})$=$V(\frac{\sum_{i=1}^{n} X_i}{n})$=$\frac{1}{n^2}V(\sum_{i=1}^{n} X_i)$=$\frac{1}{n^2}\sum_{i=1}^{n} V(X_i)$=$\frac{1}{n^2}n\sigma^2$=$\frac{\sigma^2}{n}$

του δείγματος είναι μικρό ή η μορφή της $\hat{\theta}$ δεν είναι απλή, τότε ένα $(1-\alpha)\%$-διάστημα εμπιστοσύνης για τη στατιστική συνάρτηση θ είναι το:

$$\left[\hat{\theta} - z^{(1-\frac{\alpha}{2})} se_B(\hat{\theta}), \hat{\theta} - z^{(\frac{\alpha}{2})} se_B(\hat{\theta})\right] \tag{3.8}$$

όπου $\hat{\theta}$ είναι η εκτίμηση από το δείγμα, $z^{(a)}$ είναι το α-ποσοστιαίο σημείο από την τυποποιημένη κανονική κατανομή και $se_B(\hat{\theta})$ είναι το τυπικό σφάλμα της $\hat{\theta}$ υπολογισμένο με την μέθοδο *bootstrap* όπως δίνεται από την σχέση (3.3). Το διάστημα αυτό είναι συμμετρικό ως προς $\hat{\theta}$.

3.4.2 Bootstrap t-διαστήματα εμπιστοσύνης

Το προηγούμενο διάστημα εμπιστοσύνης στηρίχθηκε σε μια αυθαίρετη παραδοχή, πως η κατανομή της στατιστικής συνάρτησης που μας ενδιαφέρει είναι η κανονική και για αυτό χρησιμοποιούμε τα ποσοστιαία σημεία της κανονικής κατανομής. Έτσι αντί τα ποσοστιαία σημεία της κανονικής κατανομής θα χρησιμοποιήσουμε τα ποσοστιαία σημεία από την κατανομή που έχουμε εκτιμήσει με τη μέθοδο *bootstrap* . Η διαδικασία για την παραγωγή bootstrap-t διαστήματα εμπιστοσύνης είναι η εξής:

1. Υπολογίζουμε από το αρχικό δείγμα τις ποσότητες $\hat{\theta}$ και $se(\hat{\theta})$, όπου είναι η σημειακή εκτίμηση και το αντίστοιχο τυπικό σφάλμα της στατιστικής συνάρτησης που μας ενδιαφέρει

2. Δημιουργούμε B bootstrap δείγματα, με δειγματοληψία με επανάθεση

3. Για κάθε bootstrap δείγμα $(\mathrm{X}_1^*, X_2^*, ..., X_n^*)$ υπολόγισε την τιμή $\hat{\theta}_i^* = T(X_1^*, X_2^*, ..., X_n^*)$ και s_i^* που είναι η τιμή της συνάρτησης μας και το τυπικό σφάλμα της παραμέτρου στο i bootstrap δείγμα για i=1,2,...B

4. Για κάθε bootstrap δείγμα υπολόγισε την ποσότητα:

$$t_i^* = \frac{\hat{\theta}_i^* - \hat{\theta}}{s_i^*}$$

δηλαδή t_i^* είναι μια μορφή τυποποιημένων τιμών της θ_i^*, δηλαδή απλά εκτιμούμε το ζητούμενο ποσοστιαίο σημείο με το ποσοστιαίο σημείο των τυποποιημένων τιμών της θ_i^*

5. Τέλος αφού διατάξουμε τα σημεία t_i^* σε αύξουσα σειρά, υπολογίζουμε το $bootstrap - t$ διάστημα εμπιστοσύνης ως:

$$\left[\hat{\theta} - t^{*(1-\frac{a}{2})} se(\hat{\theta}), \hat{\theta} - t^{*(\frac{a}{2})} se(\hat{\theta})\right] \tag{3.9}$$

όπου $t^{*(a)}$ είναι το α-ποσοστιαίο σημείο της κατανομής t^*, για να υπολογιστεί αρχικά διατάζουμε κατά αύξουσα σειρά τις τιμές t_i^*, τότε το α-οστό ποσοστιαίο σημείο είναι η τιμή t_i^*, $(0 < a < 1)$ κάτω από την οποία αφήνει $B \cdot \alpha$ των παρατηρήσεων, ενώ $B \cdot (1 - \alpha)$ των παρατηρήσεων βρίσκεται πάνω από αυτή.

Για παράδειγμα αν B=1000 ο αριθμός επαναλήψεων bootstrap και α=5% τότε για να βρούμε τα ποσοστιαία σημεία αφού έχουμε υπολογίσει τις τιμές t_i^*, η $t^{*(\frac{a}{2})}$ είναι η 25^{η} διατεταγμένη παρατήρηση t_i^* και η $t^{*(1-\frac{a}{2})}$ είναι η 975^{η} τιμή αντίστοιχα. Παρατηρούμε πως το $t^{*(\frac{a}{2})}$ θα είναι αρνητικό, αυτό το διάστημα εμπιστοσύνης δεν είναι συμμετρικό κατά ανάγκη.

3.4.3 Bootstrap ποσοστιαία διαστήματα εμπιστοσύνης

Τα bootstrap διαστήματα εμπιστοσύνης βασισμένα στα ποσοστιαία σημεία είναι μια άλλη μέθοδος για να κατασκευάσουμε διαστήματα εμπιστοσύνης. Για να κατασκευάσουμε ένα $(1 - a)\%$ διάστημα εμπιστοσύνης, η διαδικασία είναι απλή:

1. Δημιουργούμε B bootstrap δείγματα, με δειγματοληψία με επανάθεση

2. Για κάθε bootstrap δείγμα $(X_1^*, X_2^*, ..., X_n^*)$ υπολόγισε την τιμή $\hat{\theta}_i^* = T(X_1^*, X_2^*, ..., X_n^*)$ για i=1,2,...B και διάταξε αυτές σε αύξουσα σειρά

3. Το διάστημα εμπιστοσύνης θα είναι:

$$[\hat{\theta}^{*(\frac{a}{2})}, \hat{\theta}^{*(1-\frac{a}{2})}] \tag{3.10}$$

όπου $\hat{\theta}^{*(\alpha)}$ είναι το α-ποσοστιαίο σημείο της κατανομής των $\hat{\theta}_i^*$

Τα διαστήματα εμπιστοσύνης βασισμένα στα ποσοστιαία σημεία έχει αποδειχθεί πως είναι ακριβή αν η κατανομή της στατιστικής συνάρτησης είναι συμμετρική, σε αντίθετη περίπτωση δεν είναι ακριβή και επομένως υποεκτιμά ή υπερεκτιμά την πραγματική πιθανότητα να περιέχουν την πραγματική τιμή.

3.4.4 BCa διαστήματα εμπιστοσύνης

Ένα BCa διάστημα εμπιστοσύνης διορθώνει προβλήματα όπως η μη κανονικότητα της εκτιμήτριας, η τυχόν μεροληψία και η διαφορετική μορφή. Έτσι τα άκρα ενός BCa διαστήματος εμπιστοσύνης είναι:

$$\left[\theta^{*(a_1)}, \theta^{*(a_2)}\right] \tag{3.11}$$

όπου $\theta^{*(a)}$ είναι το a-ποσοστιαίο σημείο της κατανομής των τιμών που έχουμε και τα a_1 και a_2 υπολογίζονται ως:

$$a_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\frac{\alpha}{2})}}{1 - \hat{a}(\hat{z}_0 + z^{(\frac{\alpha}{2})})}\right) \tag{3.12}$$

και

$$a_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\frac{\alpha}{2})}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\frac{\alpha}{2})})}\right) \tag{3.13}$$

όπου $z^{(a)}$ είναι το a-ποσοστιαίο σημείο της τυποποιημένης κανονικής κατανομής και $\Phi(a)$ είναι η συνάρτηση κατανομής της τυποποιημένης κανονικής κατανομής, δηλαδή ισχύει ότι $\Phi^{-1}(a)=z^{(a)}$. Στις παραπάνω σχέσεις υπάρχουν δυο άγνωστες ποσότητες $\hat{z}_0$, $\hat{a}$ που πρέπει να εκτιμηθούν. Η πρώτη διορθώνει ως

προς την μεροληψία ενώ η δεύτερη διορθώνει ως προς την απόκλιση από την κανονική κατανομή. Παρατηρούμε ότι αν $\hat{z}_0$=$\hat{a}$=0 τότε προκύπτει ότι a_1=a και a_2=$1-a$ και επομένως τα διαστήματα εμπιστοσύνης ταυτίζονται με αυτά της μεθόδου των ποσοστιαίων σημείων. Τα $\hat{z}_0$ και $\hat{a}$ υπολογίζονται ως:

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\hat{\theta}_i^* < \hat{\theta}}{B}\right) \tag{3.14}$$

αν η εκτιμήτρια $\hat{\theta}$ είναι η διάμεσος των bootstrap τιμών τότε $\hat{z}_0$=0.
Το $\hat{\alpha}$ που ονομάζεται και επιταχυντής υπολογίζεται ως:

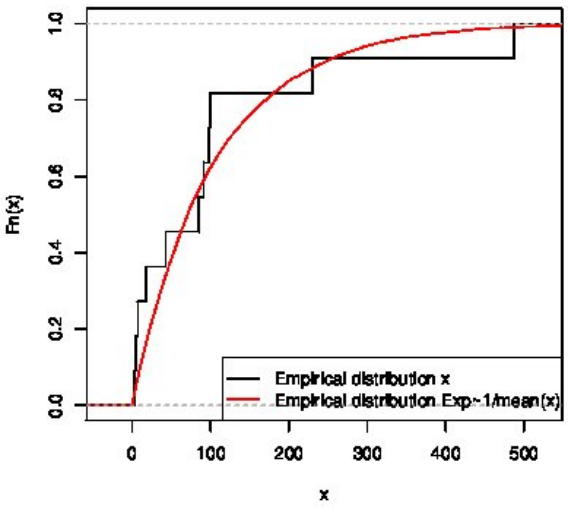
$$\hat{a} = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^3}{6\left[\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^2\right]^{\frac{3}{2}}} \tag{3.15}$$

όπου $\hat{\theta}_{(\cdot)}$ είναι η i jackknife [5] τιμή της παραμέτρου όταν αφαιρέσουμε την i παρατήρηση και $\hat{\theta}_{(\cdot)}$=$\sum_{i=1}^{n}\hat{\theta}_{(i)}$. Ένα μειονέκτημα της BCa μεθόδου είναι ότι χρειάζεται πολλές επαναλήψεις π.χ B=2000.

[5]Η βασική ιδέα για την μέθοδο jackknife είναι η εξής: Αφαιρώντας παρατηρήσεις από το αρχικό δείγμα και εκτιμώντας ξανά την παράμετρο που μας ενδιαφέρει $\hat{\theta}_{(i)}$=$T(X_1, X_2, ..., X_{i-1}, X_{i+1}, ..., X_n)$ μπορούμε να πάρουμε πληροφορία σχετικά με την σταθερότητα και άρα την μεταβλητότητα της εκτιμήτριας. Επομένως αν αφαιρέσουμε κάθε φορά μια παρατήρηση εξετάζοντας πόσο αλλάζουν οι τιμές της εκτιμήτριας παίρνουμε μια εικόνα σχετικά με την διακύμανση της εκτιμήτριας.

Παράδειγμα 1 *Τα παρακάτω στοιχεία είναι οι ώρες (n=12) μεταξύ βλαβών του κλιματισμού σε ένα Boeing 720 αεροσκάφους και θέλουμε να εκτιμήσουμε ένα 95% διάστημα εμπιστοσύνης για το μέσο χρόνο αποκατάστασης της βλάβης.*

3	*5*	*7*	*18*	*43*	*85*	*91*	*98*	*100*	*130*	*230*	*487*



Σχήμα 3.1: Εμπειρική κατανομή

Στο παραπάνω γράφημα απεικονίζεται η εμπειρική κατανομή των δεδομένων (έστω x_i, i=1,2,...,12) καθώς και η εμπειρική κατανομή της εκθετικής κατανομής με μέση τιμή $\bar{x}$. Από την σύγκριση των παραπάνω καμπυλών θα μπορούσαμε να υποθέσουμε ότι τα δεδομένα προέρχονται από εκθετική κατανομή με μέση τιμή $\bar{x}$, δηλαδή $x_i \sim Exp(\mu = \theta)$.

Αρχικά θα υπολογίσουμε το ακριβές διάστημα εμπιστοσύνης (αφού γνωρίζουμε την κατανομή του δείγματος) και θα το συγκρίνουμε με τα bootstrap διαστήματα εμπιστοσύνης.

Αφού $x_i \sim Exp(\theta)$ *τότε*

$$\sum_{i=1}^{12} x_i \sim Gamma(n, \theta)$$

$$\frac{\sum_{i=1}^{12} x_i}{\theta} \sim Gamma(n, 1)$$

$$\frac{2\sum_{i=1}^{12} x_i}{\theta} \sim Gamma(n, 2) = \chi^2_{2n}$$

$$P\Big(\chi^{2(a/2)}_{2n} \leq \frac{2\sum_{i=1}^{12} x_i}{\theta} \leq \chi^{2(1-\frac{a}{2})}_{2n}\Big) = 1 - a$$

$$P\Big(\frac{2\sum_{i=1}^{12} x_i}{\chi^{2(1-\frac{a}{2})}_{2n}} \leq \theta \leq \frac{2\sum_{i=1}^{12} x_i}{\chi^{2(\frac{a}{2})}_{2n}}\Big) = 1 - a$$

Το ακριβές διάστημα εμπιστοσύνης για το θ *είναι: (65.90, 209.17)*
Η μέση τιμή και το τυπικό σφάλμα του δείγματος είναι:

$$\hat{\theta}\text{=}108.08 \qquad\qquad se(\hat{\theta})\text{=}39.32$$

επομένως αν υποθέσουμε ότι το δείγμα ακολουθεί $t-Student$ *κατανομή (επειδή το μέγεθος του δείγματος είναι μικρό) το αντίστοιχο 95% διάστημα εμπιστοσύνης θα είναι:*

$$(\hat{\theta} - t^{(n-1,1-\frac{a}{2})} se(\hat{\theta}), \hat{\theta} - t^{(n-1,\frac{a}{2})} se(\hat{\theta})) = (21.54, 194.62)$$

Εφαρμόζοντας τη μέθοδο bootstrap για B=1000 επαναλήψεις για την εκτίμηση του τυπικού σφάλματος της μέσης τιμής τότε το αντίστοιχο δ.ε αν θεωρήσουμε ότι η κατανομή του $\hat{\theta}$ *είναι η κανονική δίνεται από τον τύπο (3.8), όπου* $se_B(\hat{\theta})$ *υπολογίζεται από τη σχέση (3.3), εδώ είναι* $se_B(\hat{\theta})$*=37.97 και το διάστημα είναι:*

$$(\hat{\theta} - z^{(1-\frac{\alpha}{2})} se_B(\hat{\theta}), \hat{\theta} - z^{(\frac{\alpha}{2})} se_B(\hat{\theta})) = (33.66, 182.50)$$

Τα δεδομένα αυτά όμως δεν προέρχονται από κανονική κατανομή επομένως υπολογίζοντας τα διάστημα εμπιστοσύνης χωρίς την υπόθεση της κανονικότητας περιμένουμε να είναι πιο ακριβής. Επομένως ακολουθώντας τα βήματα για την κατασκευή ενός bootstrap − t διαστήματος εμπιστοσύνης βρίσκουμε ότι, $t^{*(\frac{\alpha}{2})}$*=-4.71 και* $t^{*(1-\frac{\alpha}{2})}$*=1.59 και το αντίστοιχο bootstrap − t διάστημα εμπιστοσύνης είναι:*

$$(\hat{\theta} - t^{*(1-\frac{\alpha}{2})} se(\hat{\theta}), \hat{\theta} - t^{*(\frac{\alpha}{2})} se(\hat{\theta})) = (45.55, 293.35)$$

Βάζοντας κατά αύξουσα σειρά τις τιμές $\hat{\theta}^*$ *με την μέθοδο των ποσοστιμορίων το 95% διάστημα εμπιστοσύνης θα είναι η* 25^{η} *τιμή το κάτω άκρο και η* 975^{η} *το άνω άκρο, δηλαδή:*

$$(\hat{\theta}^{*(\frac{\alpha}{2})}, \hat{\theta}^{*(1-\frac{\alpha}{2})}) = (48.16, 194.83)$$

Τέλος για να υπολογίσουμε ένα BCa διάστημα εμπιστοσύνης υπολογίζουμε αρχικά από τις σχέσεις (3.14), (3.15) τις τιμές $\hat{z}_0$*,* $\hat{a}$*, εδώ είναι* $\hat{a}$*=0.048 και* $\hat{z}_0 = \Phi^{-1}(0.542)$*=0.105 και με αντικατάσταση στις σχέσεις (3.12), (3.13) βρίσκουμε ότι* a_1*=0.054* a_2*=0.991 με αποτέλεσμα το διάστημα εμπιστοσύνης να είναι:*

$$(\theta^{*(a_1)}, \theta^{*(a_2)}) = (53.16, 207.41)$$

Τα παραπάνω αποτελέσματα συνοψίζονται στον πίνακα (3.1).

Παρατηρούμε ότι πιο κοντά στο ακριβές διάστημα εμπιστοσύνης είναι το BCa. Γεγονός που το περιμέναμε, αφού το τυπικό bootstrap διάστημα και το ποσοστημορίων είναι ακριβή για συμμετρικές κατανομές, ενώ εδώ τα δεδομένα προέρχονται από εκθετική κατανομή.

Μέθοδος	Διάστημα εμπιστοσύνης 95%
Ακριβές	65.90−209.17
$t-student$	21.54−194.62
Τυπική *bootstrap*	33.66−182.50
$Bootstrap-t$	45.55−293.35
Bootstrap ποσοστημόρια	48.16−194.83
Bca	53.16−207.41

Πίνακας 3.1: Σύγκριση διαστημάτων εμπιστοσύνης

3.5 Γραμμική παλινδρόμηση με μέθοδο bootstrap

Το πολλαπλό γραμμικό μοντέλο παλινδρόμησης χρησιμοποιείται για να μελετήσει τη σχέση μεταξύ μιας εξαρτημένης μεταβλητής (y) και διάφορων ανεξάρτητων μεταβλητών, έστω ότι έχουμε k (μαζί με την σταθερά) ανεξάρτητες μεταβλητές (x) και δείγμα n. Έτσι η μορφή του γραμμικού μοντέλου παλινδρόμησης είναι:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_{k-1} x_{ik-1} + \epsilon_i \quad i = 1, 2, ...n$$

ή σε μορφή πινάκων

$$y = X \cdot \beta + \epsilon$$

με

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k-1} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk-1} \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-1} \end{bmatrix} \qquad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

όπου ϵ τυχαίο σφάλμα με μηδενική μέση τιμή και κοινή διασπορά (ομοσκεδαστικότητα) σ^2, δηλαδή $E(\epsilon)$=0 και $V(\epsilon)$=σ^2, μια συνήθης υπόθεση είναι ότι: $\epsilon \sim N_k(0, \sigma^2)$

Ο LS (ελαχίστων τετραγώνων) εκτιμητής $\hat{\beta}$ υπολογίζεται ως

$$\hat{\beta} = (X'X)^{-1}X'y$$

Ο εκτιμητής $\hat{\beta}$ είναι αμερόληπτος και κάτω από την υπόθεση της ομοσκεδαστικότητας, δηλαδή αν $V(\epsilon_i)$=σ^2, είναι και αποτελεσματικός (έχει την μικρότερη διασπορά). Η διακύμανση του εκτιμητή $\hat{\beta}$ υπολογίζεται ως:

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

Κάτω από την υπόθεση ότι το διάνυσμα των παρατηρήσεων του διαταρακτικού όρου ϵ ακολουθεί την πολυμεταβλητή κανονική κατανομή, δηλαδή, $\epsilon \sim$N$_k(0, \sigma^2 I_k)$ τότε η κατανομή του $\hat{\beta}$ ακολουθεί την πολυμεταβλητή κανονική κατανομή με μέση τιμή $E(\hat{\beta})$=β και διακύμανση $V(\hat{\beta})$=$\sigma^2(X'X)^{-1}$, δηλαδή

$$\hat{\beta} \sim N[\beta, \sigma^2(X'X)^{-1}]$$

Η διακύμανση του διαταρακτικού όρου σ^2 μπορεί να εκτιμηθεί αμερόληπτα χρησιμοποιώντας τα κατάλοιπα $\hat{\epsilon_i}$ χρησιμοποιώντας την ακόλουθη σχέση:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{\epsilon}_i}{n-k} = \frac{\hat{\epsilon}'\hat{\epsilon}}{n-k}$$

όπου $\hat{\epsilon}$=y-$\hat{y}$ με $\hat{y}$=$X\hat{\beta}$.

Για την κατασκευή (1-α)% διαστήματος εμπιστοσύνης για την άγνωστη παράμετρο β_i ισχύει ότι:

$$\frac{\hat{\beta}_i - \beta_i}{se(\hat{\beta}_i)} \sim t_{(n-\kappa)}$$

άρα το διάστημα θα υπολογίζεται ως:

$$\left[\hat{\beta}_i - t^{(n-\kappa,1-\frac{\alpha}{2})} se(\hat{\beta}_i), \hat{\beta}_i - t^{(n-\kappa,\frac{\alpha}{2})} se(\hat{\beta}_i)\right]$$

Μια από τις υποθέσεις της γραμμικής παλινδρόμησης είναι ότι οι διαταρακτικοί όροι (σφάλματα) ϵ_i έχουν την ίδια διακύμανση η οποία είναι σταθερή για όλες τις τιμές i, δηλαδή $V(\epsilon_i)=\sigma^2$ για $i=1,2,...,n$. Αν η υπόθεση αυτή δεν ισχύει τότε υπάρχει ετεροσκεδαστικότητα στους διαταρακτικούς όρους. Αν υποθέσουμε ότι ο διαταρακτικός όρος ϵ_i παρουσιάζει ετεροσκεδαστικότητα, τότε η διακύμανση του διανύσματος των παρατηρήσεων του ϵ θα δίνεται από τον ακόλουθο πίνακα συνδιακύμανσης Σ:

$$\Sigma = V(\epsilon) = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

Όταν παρουσιάζεται το φαινόμενο της ετεροσκεδαστικότητας, ο LS εκτιμητής των συντελεστών του γραμμικού υποδείγματος είναι ακόμη αμερόληπτος κάτω από τις κλασσικές υποθέσεις, αλλά όμως δεν είναι πλέον αποτελεσματικός. Για να βρούμε λοιπόν σε αυτή τη περίπτωση άριστους αμερόληπτους εκτιμητές, εφαρμόζουμε τη γνωστή στη βιβλιογραφία με το όνομα Γενικευμένη Μέθοδο των Ελαχίστων Τετραγώνων (GLS). Έτσι ο (GLS) εκτιμητής του β είναι:

$$\hat{\beta}_{GLS} = (X'\Sigma^{-1}X)^{-1}(X'\Sigma^{-1}y)$$

και η διακύμανση του

$$V(\hat{\beta}_{GLS}) = (X'\Sigma^{-1}X)^{-1}$$

Παρακάτω θα παρουσιάσουμε τέσσερις μεθόδους *bootstrap* για την εκτιμήση των συντελεστών και των τυπικών σφαλμάτων τους ενός γραμμικού μοντέλου. Μια από αυτές είναι παραμετρική (δηλαδή υποθέτει κάποια κατανομή, την

κανονική συνήθως) και οι άλλες τρείς (*Cases*, *Error*, *Wild-bootstrap*) είναι μη παραμετρικές.

3.5.1 Παραμετρική Bootstrap-Parametric Bootstrap

Αν υποθέσουμε ότι τα κατάλοιπα ϵ_i ακολουθούν την κανονική κατανομή (ή ενδεχομένως κάποια άλλη κατανομή) τότε αρχικά εκτιμάμε με τον LS εκτιμητή (για ομοσκεδαστικότητα, αλλιώς με τον GLS) τους συντελεστές της γραμμικής παλινδρόμησης καθώς επίσης και την διασπορά. Παράγουμε n τιμές ϵ_i^* από δειγματοληψία με επανάθεση από τα ϵ_i δίνοντας σε κάθε ένα από αυτά πιθανότητα $1/n$. Έτσι δημιουργούμε νέες τιμές για την εξαρτημένη μεταβλητή ως

$$y_i^* = x_i\hat{\beta} + \epsilon_i^* \qquad \epsilon_i^* \sim N(0, \hat{\sigma}^2)$$

3.5.2 Bootstrap με βάση την αναδειγματοληψία των παρατηρήσεων-Cases bootstrap

Τα δεδομένα θα μπορούσαμε να τα δούμε σαν n ζεύγη παρατηρήσεων $(y_i, x'_{ji})'$ με i=1,2,...,n και j=1,2,...,k, έστω w_i=$(y_i, x'_{ji})'$ ένα $(k+1)\times 1$ διάνυσμα που δηλώνει τις τιμές που συνδέονται στην i-στη παρατήρηση. Επομένως το σύνολο των παρατηρήσεων μπορεί να παρασταθεί ως $(w_1, w_2, ..., w_n)$. Η bootstrap διαδικασία με βάση την αναδειγματοληψία των παρατηρήσεων έχει ως εξής:

1. Δημιούργησε n bootstrap δείγματα $(w_1^*, w_2^*, ..., w_n^*)$ με δειγματοληψία με επανάθεση από το $(w_1, w_2, ..., w_n)$ δίνοντας σε κάθε w_i πιθανότητα $\frac{1}{n}$, w_i^*=(y_i^*, x_{ji}^*) με i=1,2,...,n και j=1,2,...,k. Από αυτά δημιούργησε το διάνυσμα y^*=$(y_1^*, y_2^*, ..., y_n^*)$ και τον πίνακα X^*=$(x_{j1}^*, x_{j2}^*, ..., x_{jn}^*)'$

2. Υπολόγισε τον LS εκτιμητή $\hat{\beta}^*$ από το bootstrap δείγμα:

$$\hat{\beta}^* = (X^{*'}X^*)^{-1}X^{*'}y^*$$

3. Επανέλαβε το βήμα 1 και 2 πολλές φορές (π.χ B=1000)

Η μέθοδος αυτή σε αντίθεση με τις άλλες, θεωρεί ότι οι επεξηγηματικές μεταβλητές (x) δεν είναι σταθερές αλλά τυχαίες μεταβλητές.

3.5.3 Bootstrap με βάση την αναδειγματοληψία των κατάλοιπων-Error bootstrap

Η μέθοδος αυτή είναι μη παραμετρική και βασίζεται στην αναδειγματοληψία απο τα κατάλοιπα. Προτείνεται όταν έχουμε ομοσκεδαστικότητα αλλιώς σε περίπτωση ετεροσκεδαστικότητας καταλληλότερη είναι η μέθοδος που θα παρουσιάσουμε στην επόμενη παράγραφο. Η διαδικασία είναι:

1. Απο τα δεδομένα εκτίμησε τον $\hat{\beta}$
2. Υπολόγισε τα κατάλοιπα ως ϵ_i=y_i-$\hat{y}_i$
3. Δημιούργησε n bootstrap δείγματα $(\epsilon_1^*, \epsilon_2^*, ..., \epsilon_n^*)$ με δειγματοληψία με επανάθεση από τα ϵ_i δίνοντας σε καθένα από αυτά πιθανότητα $\frac{1}{n}$
4. Υπολόγισε τις νέες τιμές της εξαρτημένης μεταβλητής από τη σχέση

$$y^* = X'\hat{\beta} + \epsilon^*$$

5. Βρες την νέα εκτίμηση $\hat{\beta}^*$ από τη σχέση

$$\hat{\beta}^* = (X^{*'}X^*)^{-1}X^{*'}y^*$$

6. Επανέλαβε τα βήματα 3, 4, 5 πολλές φορές

3.5.4 Wild Bootstrap

Η μέθοδος αυτή μοιάζει με την Error bootstrap αλλά αλλάζουν τα βήματα 2 και 3, σε αυτή την περίπτωση οι νέες τιμές της εξαρτημένης μεταβλητής υπολογίζεται ως:

$$y_i^* = x_i\hat{\beta} + f(\hat{\epsilon}_i)v_i^*$$

όπου v_i^* τυχαία μεταβλητή με μέση τιμή 0 και διασπορά 1 και $f(\hat{\epsilon}_i)$ είναι ο μετασχηματισμός του i κατάλοιπου.
Μια καλή επιλογή για την $f(\cdot)$ είναι

$$f(\hat{\epsilon}_i) = \frac{\hat{\epsilon}_i}{\sqrt{(1-h_i)}}$$

όπου h_i είναι το i διαγώνιο στοιχείο του πίνακα hat matrix ($h = X(X'X)^{-1}X'$). Υπάρχουν διάφοροι τρόποι για να καθορίσουμε την κατανομή των v_i^* . Ο πιο απλός είναι

$$v_i^* = \begin{cases} 1 & \text{με πιθανότητα } 1/2 \\ -1 & \text{με πιθανότητα } -1/2 \end{cases}$$

Αλλά ο πιο συνηθισμένος είναι

$$v_i^* = \begin{cases} \dfrac{-(\sqrt{5}-1)}{2} & \text{με πιθανότητα } \dfrac{(\sqrt{5}+1)}{2\sqrt{5}} \\ \dfrac{(\sqrt{5}+1)}{2} & \text{με πιθανότητα } \dfrac{(\sqrt{5}-1)}{2\sqrt{5}} \end{cases}$$

Φυσικά αν έχουμε ετεροσκεδαστικότητα, αλλάζει και το βήμα 5 της προηγούμενης μεθόδου και η εκτίμηση των παραμέτρων γίνεται με τον GLS εκτιμητή.

Παράδειγμα 2 *Έστω το παράδειγμα από τους Efron και Tibshirani (1993) στο οποίο 82 σχολές νομικής συμμετείχαν σε μια μελέτη. Από αυτές 15 σχολές επιλέχθηκαν τυχαία για να εξεταστεί η συσχέτιση των αποτελεσμάτων της εξέτασης LSAT και του μέσου όρου (GPA) βάση της τάξης του 1973. Τα δεδομένα δίνονται στον παρακάτω πίνακα:*

Πίνακας 3.2: Δεδομένα νομικής, Πηγή: Efron, Tibshirani

school (i)	$LSAT$ (y)	GPA (x)
1	576	3.39
2	635	3.30
3	558	2.81
4	578	3.03
5	666	3.44
6	580	3.07
7	555	3.00
8	661	3.43
9	651	3.36
10	605	3.13
11	653	3.12
12	575	2.74
13	545	2.76
14	572	2.88
15	594	2.96

Το γραμμικό μοντέλο θα πάρει την μορφή:

y_i=β_1+$\beta_2 x_i$+ϵ_i για i=1,2,..,n=15

Αρχικά υποθέτουμε ότι $\epsilon_i \sim N(0,\sigma^2)$, όπου σ^2 άγνωστη παράμετρος και εκτιμάται: $\hat{\sigma}^2$=747.4756 και οι εκτιμήσεις των συντελεστών β_1, β_2 υπολογίζονται με τη μέθοδο LS

$$\hat{\beta} = [\hat{\beta}_1, \hat{\beta}_2] = [187.90, 1.33]$$

και οι αντίστοιχες τυπικές αποκλίσεις είναι

$$[se(\hat{\beta}_1), se(\hat{\beta}_2)] = [93.11, 0.30]$$

Θα δούμε πως μεταβάλλονται τα αποτελέσματα αυτά, χρησιμοποιώντας τις μεθόδους bootstrap, που αναφέρθηκαν παραπάνω.

Αρχικά η παραμετρική μέθοδος βασίζεται στην παραδοχή ότι $\epsilon_i \sim N(0,\sigma^2)$ και αφού σ^2 άγνωστη εκτιμάται από την $\hat{\sigma}^2$, άρα $\epsilon_i \sim N(0, \hat{\sigma}^2)$.
Έτσι προσομοιώνουμε n δείγμα ϵ_i^* από την $N(0, \hat{\sigma}^2)$ και με αυτά υπολογίζουμε τα νέα y_i^*, ως y_i^*=$\hat{\beta}_1$+$\hat{\beta}_2 x$+ϵ_i^* και έτσι έχουμε (y^*, X) και υπολογίζουμε τους νέους συντελεστές $\hat{\beta}^*$, η διαδικασία αυτή επαναλαμβάνεται πολλές φορές εδώ για B=1000 δίνει τα παρακάτω αποτελέσματα

$$\hat{\beta}^* = [\hat{\beta}_1^*, \hat{\beta}_2^*] = [192.59, 1.31]$$

και οι αντίστοιχες τυπικές αποκλίσεις είναι $se_B(\hat{\beta}_1^*)$=94.54 και $se_B(\hat{\beta}_2^*)$=0.30.

Η $Cases - bootstrap$ είναι η πιο απλή μέθοδος και δεν χρειάζεται καμία υπόθεση κατανομής, κάνοντας αναδειγματοληψία της εξαρτημένης και των ανεξάρτητων μεταβλητών (η μέθοδος αυτή δεν θεωρεί τις επεξηγηματικές μεταβλητές ως σταθερές) οι εκτιμήσεις των συντελεστών και οι αντίστοιχες τυπικές αποκλίσεις είναι

$$\hat{\beta} = [\hat{\beta}_1, \hat{\beta}_2] = [184.03, 1.34]$$

και

$$[se(\hat{\beta}_1), se(\hat{\beta}_2)] = [88.09, 0.29]$$

Τέλος για τις εκτιμήσεις των συντελεστών με την $Error-bootstrap$ *αφού υπολογίσουμε τα κατάλοιπα* $\hat{\epsilon}_i$ *ως* $\hat{\epsilon}_i$=$\hat{y}$-$\hat{\beta}_1$-$\hat{\beta}_2 x$, *κάνοντας αναδειγματοληψία από αυτά υπολογίζουμε τις νέες τιμές* y_i^* *από την σχέση* y_i^*=$\hat{\beta}_1 + \hat{\beta}_2 x + \epsilon_i^*$ *και εκτιμούμε τα νέα* $\hat{\beta}^*$ *για κάθε επανάληψη, εδώ για* B=*1000 έχουμε ότι*

$$\hat{\beta} = [\hat{\beta}_1, \hat{\beta}_2] = [189.19, 1.32] \quad \textit{και} \quad [se(\hat{\beta}_1), se(\hat{\beta}_2)] = [87.98, 0.28]$$

Τα παραπάνω αποτελέσματα συνοψίζονται στο παρακάτω πίνακα

Πίνακας 3.3: Σύγκριση εκτιμήσεων των συντελεστών

Μέθοδος	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$\hat{\beta}_2$	$se(\hat{\beta}_2)$
Κλασσική	187.90	93.15	1.33	0.30
Παραμετρική *bootstrap*	192.59	94.54	1.31	0.30
Cases bootstrap	184.03	88.09	1.34	0.29
Error bootstrap	189.19	87.98	1.32	0.28

Κεφάλαιο 4

Μετα-ανάλυση με bootstrap

4.1 Μέθοδοι μετα-ανάλυσης με Bootstrap

Σε μια μετα-ανάλυση, οι άγνωστες παράμετροι συνήθως υπολογίζονται με εκτιμητές μέγιστης πιθανοφάνειας (ML) και τα συμπεράσματα βασίζονται στην ασυμπτωτική θεωρία. Απαραίτητη υπόθεση για την εφαρμογή των μοντέλων ($fixed$, $random\text{-}effect$) είναι ότι η κατανομή της εκτιμώμενης επίδρασης είναι η κανονική, όπως και η κατανομή μεταξύ των μελετών. Στην πράξη βέβαια, τα δείγματα είναι πεπερασμένα και η υπόθεση της κανονικότητας μπορεί να παραβιαστεί, οδηγώντας ενδεχομένως σε μεροληπτικές εκτιμήσεις και ακατάλληλα τυπικά σφάλματα. Μια ευέλικτη τεχνική η οποία μπορεί να διορθώσει τα παραπάνω προβλήματα είναι η μέθοδος *bootstrap*. Παρακάτω θα δούμε τρείς μεθόδους *bootstrap* για μετα-ανάλυση, μια παραμετρική και δυο μη παραμετρικές (Wild,Cases).

Ας υποθέσουμε ότι έχουμε το τυχαίο μοντέλο επιδράσεων (random-effect model):

$$\hat{\theta}_i = \theta + u_i + \epsilon_i$$

Τα u_i, ϵ_i είναι μεταξύ τους ανεξάρτητα με $E(u_i)$=0, $V(u_i) = \tau^2$ και αντίστοιχα $E(\epsilon_i) = 0$, $V(\epsilon_i) = \sigma_i^2$ αν e_i=u_i+ϵ_i τότε το μοντέλο μπορεί να γραφτεί στην μορφή:

$$\hat{\theta}_i = \theta + e_i \quad \text{όπου} \quad e_i \sim N(0, \tau^2 + \sigma_i^2)$$

4.1.1 Παραμετρική bootstrap

- Βήμα 1: Αρχικά εκτιμούμε από τα αρχικά δεδομένα τις παραμέτρους θ και τ^2 με τον $REML$ ή $RIGLS$ εκτιμητή. Χρησιμοποιούμε αυτούς τους εκτιμητές γιατί είναι αμερόληπτοι εκτιμητές του τ^2

- Βήμα 2: Παράγουμε k κατάλοιπα e_i^* από την κανονική κατανομή (για αυτό και η μέθοδος λέγεται παραμετρική γιατί υποθέτει κάποια κατανομή σε αντίθεση με τις άλλες δυο μεθόδους) με μέση τιμή μηδέν και διασπορά $\hat{\tau}^2$+$\hat{\sigma_i}^2$, δηλαδή $e_i^* \sim N(0, \hat{\tau}^2 + \hat{\sigma}_i^2)$

- Βήμα 3: Προσθέτουμε τα κατάλοιπα e_i^* στην μέση εκτίμηση της επίδρασης της θεραπείας (ολική) και έτσι δημιουργούμε τις νέες εκτιμήσεις για την επίδραση της θεραπείας για κάθε μελέτη, δηλαδή

$$\hat{\theta_i}^* = \hat{\theta} + e_i^*$$

- Βήμα 4: Τέλος αφού έχουμε τις νέες εκτιμήσεις για την κάθε μελέτη, βρίσκουμε με την γνωστή διαδικασία την νέα επίδραση της θεραπείας $\hat{\theta}^*$ και την διασπορά μεταξύ των μελετών $\hat{\tau}^{*2}$

- Βήμα 5: Επαναλαμβάνουμε τα βήματα 1-4 πολλές φορές π.χ B=1000 και βρίσκουμε τις μέσες τιμές για τις δυο παραμέτρους.

4.1.2 Wild-bootstrap

- Βήμα 1: Το βήμα αυτό είναι ίδιο με της παραμετρικής, εκτιμούμε τις παραμέτρους θ και τ^2 από τα αρχικά δεδομένα
- Βήμα 2: Εκτιμούμε τα e_i ως $\hat{e}_i$=$\hat{\theta}_i$-$\hat{\theta}$
- Βήμα 3: Υπολογίζουμε τις νέες εκτιμήσεις για την επίδραση της θεραπείας σε κάθε μελέτη ως

$$\hat{\theta}_i = \theta + f(\hat{e}_i) \cdot v_i^*$$

όπου $f(\hat{e}_i)$ και v_i^* όπως αναλύθηκαν στο προηγούμενο κεφάλαιο

- Βήμα 4: Βρίσκουμε τις νέες τιμές $\hat{\theta}^*$ και $\hat{\tau}^{2*}$
- Βήμα 5: Επαναλαμβάνουμε τα βήματα 1-4 πολλές φορές π.χ B=1000 και βρίσκουμε τις μέσες τιμές για τις δυο παραμέτρους.

4.1.3 Cases-bootstrap

Η Cases-bootstrap είναι η πιο απλή μη παραμετρική μέθοδος bootstrap. Υπολογίζουμε τις εκτιμήσεις των επιδράσεων της θεραπείας $\hat{\theta}_i^*$ κάνοντας απλά αναδειγματοληψία των αρχικών επιδράσεων των εκτιμήσεων της θεραπείας, τα επόμενα βήματα είναι ίδια με των άλλων μεθόδων αφού έχουμε υπολογίσει τις νέες εκτιμήσεις, δηλαδή βρίσκουμε την ολική επίδραση της θεραπείας και επαναλαμβάνουμε την διαδικασία πολλές φορές.

4.2 Αποτελεσματικότητα εκτιμητριών

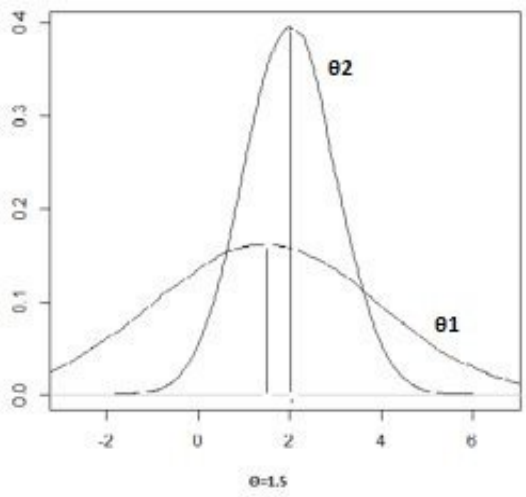
Γενικά όταν ορίζουμε μια εκτιμήτρια $\hat{\theta}$ κάποιας παραμέτρου θ ελέγχουμε αν είναι κατάλληλη ή πιο κατάλληλη από κάποια άλλη εκτιμήτρια. Για να μπορεί να θεωρηθεί η $\hat{\theta}$ ως μια καλή εκτιμήτρια της θ πρέπει να έχει μέση τιμή θ, δηλαδή η $\hat{\theta}$ να είναι αμερόληπτη εκτιμήτρια της θ και να έχει μικρή διασπορά. Δεν αρκεί όμως μόνο η αμεροληψία για να ισχυριστούμε ότι η εκτιμήτρια είναι βέλτιστη, γεγονός που φαίνεται και από το γεγονός ότι για μια παραμετρική συνάρτηση μπορούν να υπάρξουν περισσότερες από μία αμερόληπτες εκτιμήτριες. Καλύτερη από αυτές τις αμερόληπτες εκτιμήτριες θα είναι αυτή με την μικρότερη διασπορά. Επομένως ένα κριτήριο για βέλτιστη εκτιμήτρια είναι η διασπορά, η εκτιμήτρια με την μικρότερη διασπορά είναι η βέλτιστη. Το κριτήριο αυτό όμως αφορά για σύγκριση μεταξύ αμερόληπτων εκτιμητριών. Υπάρχουν όμως περιπτώσεις όπου μια μη αμερόληπτη εκτιμήτρια είναι καλύτερη από μια αμερόληπτη.

Για παράδειγμα αν για την εκτίμηση κάποιας παραμέτρου θ, έχουμε τις εκτιμήτριες $\hat{\theta}_1$, $\hat{\theta}_2$ που φαίνονται γράφημα (4.1), τότε παρατηρούμε ότι η $\hat{\theta}_2$ δεν είναι αμερόληπτη (δεν έχει μέση τιμή θ) αλλά παίρνει τιμές κοντά στο θ, ενώ η $\hat{\theta}_1$ που είναι αμερόληπτη αλλα μπορεί να πάρει τιμές πολύ μακριά από το θ (έχει μεγάλη διασπορά). Επομένως παρότι δεν είναι αμερόληπτη, καλύτερη εκτιμήτρια σε αυτή την περίπτωση είναι η $\hat{\theta}_2$.
Ένα κριτήριο σύγκρισης εκτιμητριών (αμερόληπτων ή όχι) είναι το μέσο τετραγωνικό σφάλμα (MSE) που υπολογίζεται ως:

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 \tag{4.1}$$

Βέλτιστη εκτιμήτρια είναι αυτή που έχει το μικρότερο μέσο τετραγωνικό σφάλμα. Το μέσο τετραγωνικό σφάλμα μπορεί να γραφτεί με την μορφή:

$$MSE(\hat{\theta}) = V(\hat{\theta}) + b(\hat{\theta})^2 \tag{4.2}$$



Σχήμα 4.1: Σύγκριση εκτιμητριών

όπου $b(\hat{\theta})$=$E(\hat{\theta})$-θ, δηλαδή η μεροληψία.

Απόδειξη 7

$$
\begin{aligned}
MSE(\hat{\theta}) &= E(\hat{\theta}-\theta)^2 \Rightarrow \\
MSE(\hat{\theta}) &= E\left[(\hat{\theta}-E(\hat{\theta})+E(\hat{\theta})-\theta)^2\right] \Rightarrow \\
MSE(\hat{\theta}) &= E(\hat{\theta}-E(\hat{\theta}))^2+E(E(\hat{\theta})-\theta)^2+2\cdot E\left[(\hat{\theta}-E(\hat{\theta}))\cdot(E(\hat{\theta})-\theta)\right] \Rightarrow \\
MSE(\hat{\theta}) &= V(\hat{\theta})+b(\hat{\theta})^2
\end{aligned}
$$

Από την παραπάνω σχέση προκύπτει ότι αν ένας εκτιμητής είναι αμερόληπτος το μέσο τετραγωνικό σφάλμα είναι ίσο με την διασπορά του.

4.3 Προσομοίωση

Το μέτρο του μεγέθους της επίδρασης της θεραπείας ($Effect\text{-}size$) που θα χρησιμοποιήσουμε στην μελέτη προσομοίωσης είναι η τυποποιημένη μέση διαφορά ($Standardized\ mean\ difference$), εκφράζοντας την διαφορά μεταξύ δυο ομάδων για παράδειγμα της ομάδας $treatment$ και της ομάδας $control$. Η διαδικασία έχει ως εξής:

Πρώτα παράγουμε k κατάλοιπα u_i από την κανονική κατανομή με μέση τιμή μηδέν και διασπορά τ^2 ($u_i \sim N(0,\tau^2)$) και τα προσθέτουμε στην μέση επίδραση της θεραπείας ($\theta_i=\theta+u_i$) έστω εδώ ότι $\theta=0.5$, έτσι υπολογίζουμε την πραγματική τιμή της επίδρασης της θεραπείας θ_i για τις k μελέτες . Μετά παράγουμε δεδομένα για την κάθε μελέτη και για τις δύο ομάδες από δυο κανονικές κατανομές με ίδια διακύμανση (γιατί συνήθως στις κλινικές μελέτες υποθέτουμε ότι είναι ίσες στις δυο ομάδες), δηλαδή έστω

$$y_{Ti} \sim N\Big(\frac{\theta_i}{2}, 1\Big)$$

και

$$y_{Ci} \sim N\Big(-\frac{\theta_i}{2}, 1\Big)$$

τα δεδομένα της ομάδας $treatment$ για την i μελέτη και της ομάδας $control$ αντίστοιχα. Αν θέλουμε να παράγουμε δεδομένα όχι από την κανονική κατανομή αλλά από την t_n-$Student$ κατανομή, τότε παράγουμε $u_i \sim {}^1t_n\sqrt{\frac{\tau^2}{\frac{n}{n-2}}}$ και από τη σχέση $\theta_i=\theta+u_i$ έτσι έχουμε την πραγματική τιμή της θεραπείας σε κάθε μελέτη. Έπειτα παράγουμε δεδομένα για την κάθε μελέτη και για τις δυο ομάδες από

[1] Χρησιμοποιούμε αυτό τον μετασχηματισμό γιατί:
$E(u_i)=E(t_n\sqrt{\frac{\tau^2}{\frac{n}{n-2}}})=0$, αφού μέση τιμή της $t_n=0$
$V(u_i)=V(t_n\sqrt{\frac{\tau^2}{\frac{n}{n-2}}})=\frac{\tau^2}{\frac{n}{n-2}}\frac{n}{n-2}=\tau^2$, αφού η διασπορά της $t_n=\frac{n}{n-2}$

την t_n-*Student* κατανομή ως[2]:

$$y_{Ti} \sim t_n \sqrt{\frac{1}{\frac{n}{n-2}}} + \frac{\theta_i}{2}$$

και

$$y_{Ci} \sim t_n \sqrt{\frac{1}{\frac{n}{n-2}}} - \frac{\theta_i}{2}$$

Τέλος αν θέλουμε να προσομοιώσουμε από την χ^2_n κατανομή τότε: $u_i \sim$ [3]$(\chi^2_n - n)\sqrt{\frac{\tau^2}{2n}}$ και

$$y_{Ti} \sim (\chi^2_n - n)\sqrt{\frac{1}{2n}} + \frac{\theta_i}{2}$$

και

$$y_{Ci} \sim (\chi^2_n - n)\sqrt{\frac{1}{2n}} - \frac{\theta_i}{2}$$

Στη συνέχεια από τα δυο δείγματα για την κάθε μελέτη υπολογίζουμε τις μέσες τιμές $\bar{y}_T$, $\bar{y}_C$ και τις δειγματικές διασπορές s_T και s_C όπως επίσης και την εκτίμηση της κοινής διασποράς. Επομένως η (διορθωμένη) εκτίμηση της θεραπείας για την i μελέτη και η αντίστοιχη εκτίμηση της διακύμανσης (βλέπε κεφάλαιο 2) θα δίνονται:

$$\hat{\theta}_i = \frac{\bar{y}_T - \bar{y}_C}{s_p}\Big(1 - \frac{3}{4(n_T + n_C) - 9}\Big)$$

και

$$V(\hat{\theta}_i) = \frac{n_T + n_C}{n_T \cdot n_C} + \frac{\hat{\theta}_i^{\,2}}{2(n_T + n_C)}$$

Θεωρούμε ότι $n_T = n_C = n$ όπου $n \sim U(0.4 \cdot \bar{n}, 1.6 \cdot \bar{n})$, με $\bar{n}$ το μέσο μέγεθος συμμετεχόντων (ασθενών) στην μελέτη. Για να αποκτήσουμε μια πλήρη εικόνα

[2] $E(y_{Ti})$=$E(t_n\sqrt{\frac{1}{\frac{n}{n-2}}} + \frac{\theta_i}{2})$=$\frac{\theta_i}{2}$ και $E(y_{Ci})$=$-\frac{\theta_i}{2}$
$V(y_{Ti})$=$V(t_n\sqrt{\frac{1}{\frac{n}{n-2}}} + \frac{\theta_i}{2})$=$\frac{n}{n-2}\frac{1}{\frac{n}{n-2}}$=1

[3] $E(u_i)$=$E((\chi^2_n - n)\sqrt{\frac{\tau^2}{2n}})$=$\sqrt{\frac{\tau^2}{2n}}(n-n)$=0
$V(u_i)$=$V((\chi^2_n - n)\sqrt{\frac{\tau^2}{2n}})$=$2n\frac{\tau^2}{2n}$=$\tau^2$

της αποτελεσματικότητας της μεθόδου bootstrap, χρησιμοποιούμε διαφορετικές τιμές για το μέγεθος των μελετών k, το μέγεθος των συμμετεχόντων σε κάθε μελέτη n και μεταξύ τους συνδυασμούς καθώς και την διασπορά μεταξύ των μελετών τ^2 (πίνακας 4.1).

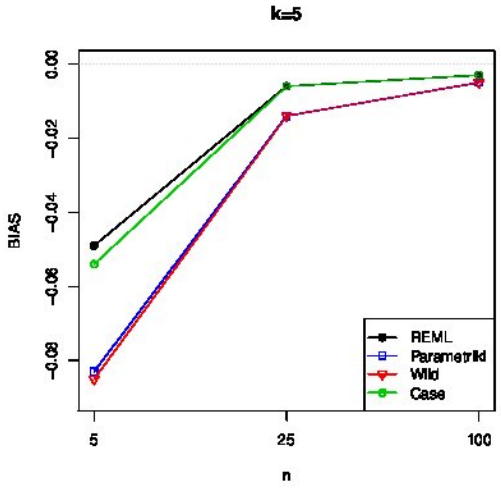
Πίνακας 4.1: Χαρακτηριστικά προσομοίωσης

Μέση (ολική) επίδραση της θεραπείας: θ=0.5
Αριθμός μελετών: k=5/10/50
Μέσο μέγεθος δείγματος μελέτης: $\bar{n}$=5/25/100
Μέγεθος δείγματος κάθε ομάδας: $n_T = n_C = n \sim U(0.4 \cdot \bar{n}, 1.6 \cdot \bar{n})$
Διασπορά μεταξύ των μελετών: τ^2=0/0.2
Κατανομές: Κανονική, t_3-$Student$, χ^2_2 (με τους μετασχηματισμους όπως αναλύθηκαν παραπάνω)

4.3.1 Αποτελέσματα για την κανονική κατανομή

Αρχικά θα συγκρίνουμε τις προηγούμενες μεθόδους *boostrap*, χρησιμοποιώντας το σταθερό μοντέλο μετα-ανάλυσης ($Fixed\text{-}effect\ model$). Σε αυτό το μοντέλο η μόνη άγνωστη παράμετρος προς εκτίμηση είναι η μέση ολική επίδραση της θεραπείας.

Όπως ήταν αναμενόμενο λόγω της εξάρτησης της επίδρασης της θεραπείας (effect-size) και της δειγματικής διασποράς κάθε μελέτης, οι αρχικές εκτιμήσεις για την μέση επίδραση της θεραπείας (θ) είναι αρνητικά μεροληπτικές, ειδικά όταν το μέγεθος δείγματος κάθε μελέτης είναι μικρό. Αύξηση των μελετών αυξάνει την μεροληψία. Από τον πίνακα (4.2) διαπιστώνεται ότι η Παραμετρική μέθοδος και η Wild μέθοδοι bootstrap, χωρίς διόρθωση της μεροληψίας αυξάνουν την αρνητική μεροληψία σε σχέση με τις αρχικές εκτιμήσεις. Αντίθετα με



Σχήμα 4.2: Μεροληψία σε σχέση με το μέγεθος των μελετών

διόρθωση της μεροληψίας για τις μεθόδους αυτές προκύπτουν σχεδόν αμερόληπτες εκτιμήσεις της μέσης επίδρασης της θεραπείας. Η Cases bootstrap δίνει πολύ μικρές αλλαγές σε σχέση με τις αρχικές εκτιμήσεις με ή χωρίς διόρθωση της μεροληψίας.

Πίνακας 4.2: Εκτίμηση για την μέση (ολική) επίδραση όταν θ=0.5 για k=5

		ΧΩΡΙΣ ΔΙΟΡΘΩΣΗ ΜΕΡΟΛΗΨΙΑΣ			ΜΕ ΔΙΟΡΘΩΣΗ ΜΕΡΟΛΗΨΙΑΣ		
n	$REML$	Παραμετρική	$Wild$	$Cases$	Παραμετρική	$Wild$	$Cases$
5	0.451	0.417	0.415	0.446	0.486	0.487	0.456
25	0.494	0.486	0.486	0.496	0.502	0.502	0.492
100	0.497	0.495	0.495	0.494	0.499	0.499	0.500

Παρόλο που από τον πίνακα (4.2) φαίνεται ότι οι μέθοδοι αυτές (οι δυο από αυτές) εκτιμούν αμερόληπτα την μέση επίδραση της θεραπείας με διόρθωση της μεροληψίας, όμως όπως φαίνεται και από τον πίνακα (4.3), ότι με διόρθωση της

μεροληψίας αυξάνεται και η τυπική απόκλιση, για μικρές μελέτες η αύξηση είναι μεγαλύτερή ενώ για μεγάλες μελέτες δεν υπάρχει μεταβολή. Όμως χωρίς διόρθωση της μεροληψίας η μέση τυπική απόκλιση για τις μεθόδους Παραμετρική, WILD μειώνεται ενώ για την CASES αυξάνεται.

Πίνακας 4.3: Μέση τιμή για την τυπική απόκλιση της μέσης επίδρασης της θεραπείας όταν θ=0.5 για k=5

		ΧΩΡΙΣ ΔΙΟΡΘΩΣΗ ΜΕΡΟΛΗΨΙΑΣ			ΜΕ ΔΙΟΡΘΩΣΗ ΜΕΡΟΛΗΨΙΑΣ		
n	$REML$	Παραμετρική	$Wild$	$Cases$	Παραμετρική	$Wild$	$Cases$
5	0.286	0.262	0.266	0.296	0.308	0.311	0.320
25	0.126	0.122	0.122	0.134	0.130	0.130	0.134
100	0.063	0.063	0.063	0.063	0.063	0.063	0.063

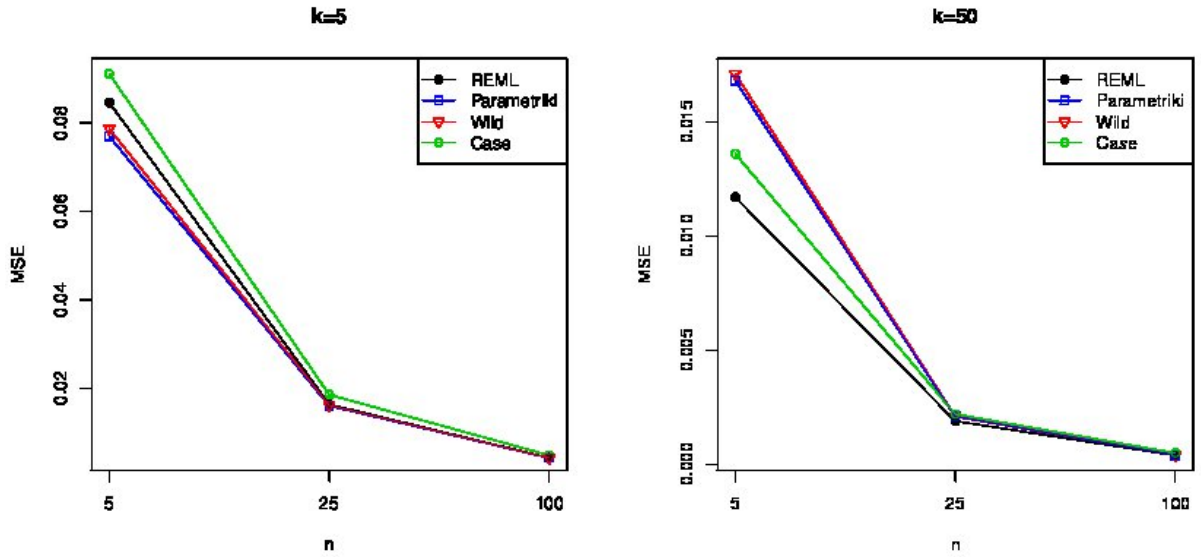
Το κριτήριο που συνδύαζει την μεροληψία και την διασπορά για την επιλογή καλύτερου εκτιμητή, όπως έχει αναλυθεί και σε προηγούμενη παράγραφο είναι το μέσο τετραγωνικό σφάλμα (MSE). Ο πίνακας (4.4) περιέχει το μέσο τετραγωνικό σφάλμα (*10000) για διάφορους συνδυασμούς των n και k.

Πίνακας 4.4: MSE*10000 για την μέση επίδραση της θεραπείας (θ=0.5)

			ΧΩΡΙΣ ΔΙΟΡΘΩΣΗ ΜΕΡΟΛΗΨΙΑΣ			ΜΕ ΔΙΟΡΘΩΣΗ ΜΕΡΟΛΗΨΙΑΣ		
n	k	$REML$	Παραμετρική	$Wild$	$Cases$	Παραμετρική	$Wild$	$Cases$
5	5	846	770	787	911	961	972	1053
	10	454	444	452	480	498	507	555
	50	117	168	171	136	98	100	125
25	5	163	160	161	185	168	168	186
	10	85	85	85	96	87	88	97
	50	19	21	21	22	18	18	21
100	5	41	40	40	49	42	42	45
	10	19	19	19	22	19	19	22
	50	4	4	4	5	4	4	5

Ένα πρώτο συμπέρασμα από τον παραπάνω πίνακα είναι ότι για μεγάλο μέγεθος δείγματος και μεγάλες μελέτες (δηλαδή n=100 και k=50) οι αρχικές εκτιμήσεις καθώς και οι τρεις *bootstrap* μέθοδοι δίνουν το ίδιο μέσο τετραγωνικό σφάλμα. Η $Cases$ μέθοδος δίνει μεγαλύτερο MSE από τις αρχικές τιμές ($REML$) για οποιοδήποτε συνδυασμό των n και k με μεγαλύτερες αυξήσεις για μικρούς συνδυασμούς. Επίσης από τον πίνακα αυτό μπορούμε να ισχυριστούμε ότι για μικρό πλήθος μελετών (k=5 ή 10) η Παραμετρική μέθοδος και η WILD χωρίς διόρθωση της μεροληψίας δίνουν μικρότερο MSE από των αρχικών με όσο μικρότερες μελέτες τόσο μεγαλύτερη διαφορά, αντίθετα με διόρθωση της μεροληψίας το MSE αυξάνεται από τις αρχικές τιμές. Τέλος για μεγάλο πλήθος μελετών το MSE είναι μικρότερο μόνο με διόρθωση της μεροληψίας αλλιώς έχουμε αύξηση σε σχέση με τις αρχικές. Άρα η διόρθωση της μεροληψίας εξαρτάται αποκλειστικά και μόνο από το πλήθος των μελετών.

Τα παραπάνω συμπεράσματα φαίνονται και στο γράφημα (4.3).



Σχήμα 4.3: Μέσο τετραγωνικό σφάλμα για την μέση ολική επίδραση σε σχέση με το μέγεθος των μελετών

Αντίθετα με το σταθερό μοντέλο, στο τυχαίο μοντέλο (*Random-effect model*) εκτός από την μέση επίδραση της θεραπείας η ετερογένεια (τ^2) είναι μια άλλη παράμετρος που θα πρέπει να εκτιμηθεί. Επειδή όπως αναφέρθηκε και παραπάνω η *Cases* μέθοδος δίνει χειρότερα αποτελέσματα από τις αρχικές εκτιμήσεις δεν θα ασχοληθούμε περισσότερο με αυτή τη μέθοδο. Για το τυχαίο μοντέλο τα αποτελέσματα της προσομοίωσης φαίνονται στους παρακάτω πίνακες (4.5, 4.6).

Όσο αφορά την εκτίμηση για την μέση εκτίμηση της θεραπείας, τα συμπεράσματα θα μπορούσαμε να πούμε ότι είναι τα ίδια με το σταθερό μοντέλο. Δηλαδή οι μέθοδοι Παραμετρική και *Wild bootstrap* χωρίς διόρθωση της μεροληψίας δίνουν καλύτερα αποτελέσματα (μικρότερο τετραγωνικό σφάλμα), όταν

το μέγεθος του δείγματος κάθε μελέτης είναι μικρό όπως και ο αριθμός των μελετών, παρατηρώντας ότι όσο μεγαλύτερη είναι η διασπορά μεταξύ των μελετών, τόσο μεγαλύτερη διαφορά σε σχέση με τις αρχικές εκτιμήσεις. Για μεγάλο αριθμό μελετών χρειάζεται διόρθωση της μεροληψίας. Όπως και στο σταθερό μοντέλο για μεγάλο μέγεθος δείγματος και μεγάλο αριθμό μελετών οι μέθοδοι *bootstrap* δεν έχουν διαφορά με τις αρχικές εκτιμήσεις.

Πίνακας 4.5: MSE^*10000 για την μέση επίδραση της θεραπείας θ=0.5, τ^2=0.2

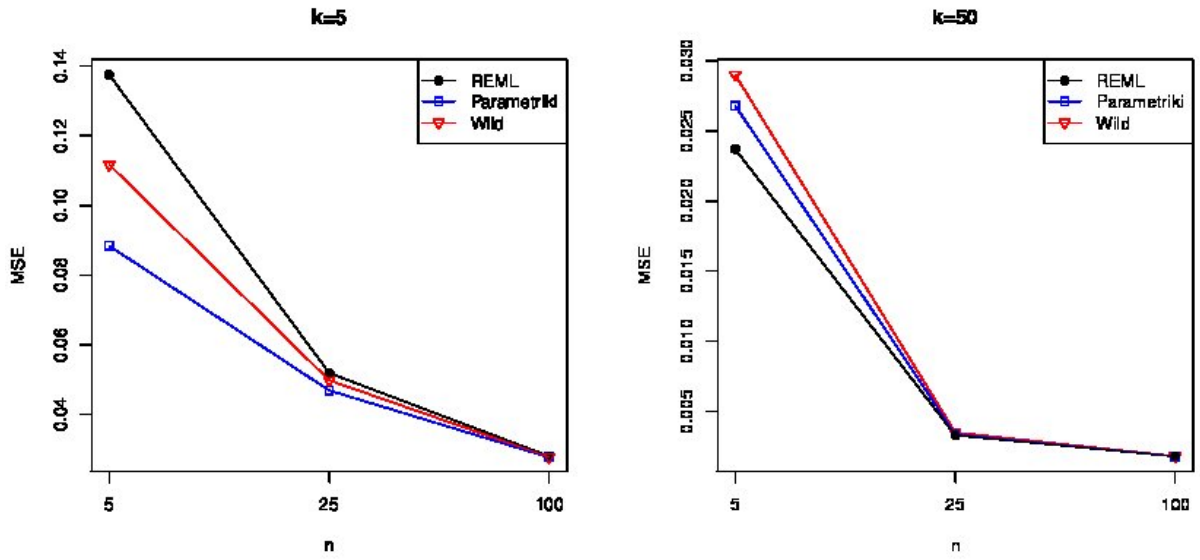
			ΧΩΡΙΣ ΔΙΟΡΘΩΣΗ ΜΕΡΟΛΗΨΙΑΣ		ΜΕ ΔΙΟΡΘΩΣΗ ΜΕΡΟΛΗΨΙΑΣ	
n	k	$REML$	Παραμετρική	$Wild$	Παραμετρική	$Wild$
5	5	1263	1163	1141	1429	1458
	10	635	605	607	701	725
	50	162	218	228	148	150
25	5	620	600	600	641	641
	10	304	294	295	316	316
	50	55	57	57	56	56
100	5	402	400	400	405	406
	10	226	224	225	228	227
	50	45	45	45	46	46

Όπως και με την εκτίμηση της μέσης επίδρασης της θεραπείας (θ), έτσι και στην εκτίμηση της διασποράς μεταξύ των μελετών (τ^2), υπάρχει αντίστροφη σχέση μεταξύ της μεροληψίας και της ακρίβειας για τις δυο μεθόδους. Δηλαδή, χωρίς διόρθωση της μεροληψίας μειώνεται η διασπορά σε σχέση με τις αρχικές εκτιμήσεις ($REML$) αυξάνεται ωστόσο η μεροληψία, ενώ με διόρθωση της με-

ροληψίας συμβαίνει το αντίθετο. Επομένως όπως φαίνεται και από τον πίνακα (4.6) το μέσο τετραγωνικό σφάλμα είναι πολύ μικρότερο από των αρχικών εκτιμήσεων για τις δυο μεθόδους χωρίς διόρθωση της μεροληψίας, αντίθετα με διόρθωση της μεροληψίας αυξάνεται. Όσο μικρότερες μελέτες και μικρό μέγεθος μελετών, τόσο μεγαλύτερη διαφορά όπως και όσο αυξάνεται η ετερογένεια. Για μεγάλο πλήθος μελετών δεν ισχύει ότι οι μέθοδοι αυτές δίνουν καλύτερα αποτελέσματα. Ένα συμπέρασμα ανάμεσα στην σύγκριση μεταξύ των δυο μεθόδων (Παραμετρική, *Wild*) είναι ότι ενώ και οι δυο μέθοδοι δεν έχουν (ή έχουν πολύ μικρές) διαφορά στην εκτίμηση της μέσης επίδρασης της θεραπείας, αλλά για την εκτίμηση της ετερογένειας όπως φαίνεται και από τον πίνακα (4.6) η Παραμετρική μέθοδος δίνει καλύτερα αποτελέσματα (κάτι που ίσως εξηγείται ότι η προσομοίωση των δεδομένων γίνεται από την κανονική κατανομή). Τα παραπάνω συμπεράσματα απεικονίζονται στο γραφήμα (4.4).

Πίνακας 4.6: MSE*10000 για την ετερογένεια θ=0.5, τ^2=0.2

			ΧΩΡΙΣ ΔΙΟΡΘΩΣΗ ΜΕΡΟΛΗΨΙΑΣ		ΜΕ ΔΙΟΡΘΩΣΗ ΜΕΡΟΛΗΨΙΑΣ	
n	k	$REML$	Παραμετρική	$Wild$	Παραμετρική	$Wild$
5	5	1375	884	1117	2069	2020
	10	583	316	382	952	889
	50	237	268	290	259	253
25	5	519	469	498	574	546
	10	183	166	173	203	197
	50	33	34	35	35	36
100	5	280	277	278	285	285
	10	116	114	113	119	120
	50	18	18	18	19	19



Σχήμα 4.4: Μέσο τετραγωνικό σφάλμα για την διασπορά μεταξύ των μελετών σε σχέση με το μέγεθος των μελετών

4.3.2 Αποτελέσματα για μη κανονική κατανομή

Από τα παραπάνω καταλήξαμε ότι οι μέθοδοι Παραμετρική, *Wild bootstrap*, δίνουν καλύτερα αποτελέσματα στην εκτίμηση της μέσης ολικής επίδρασης και της ετερογένειας από τις αρχικές εκτιμήσεις ($REML$) κυρίως όταν το δείγμα κάθε μελέτης είναι μικρό, χρησιμοποιώντας στην προσομοίωση για κατανομή των καταλοίπων την κανονική. Τώρα θα εξετάσουμε τα αποτελέσματα των μεθόδων αυτών, όταν πλέον παραβιάζεται και η υπόθεση της κανονικότητας.

Εδώ θα παράγουμε τα αρχικά δεδομένα από την t_3-*Student* και την χ^2_2 κατανομή, επιλέγουμε μικρούς βαθμούς ελευθερίας γιατί για μεγάλους βαθμούς ελευθερίας οι κατανομές αυτές προσεγγίζουν την κανονική κατανομή και εξε-

τάζουμε την περίπτωση που το μέγεθος δείγματος κάθε μελέτης καθώς και ο αριθμός των μελετών είναι μικρό (π.χ. n=k=5), γιατί για μεγάλο δείγμα από Κεντρικό Οριακό Θεώρημα η κατανομή θα συγκλίνει στην κανονική.

Γενικά τα συπμεράσματα είναι παρόμοια με αυτά, όταν τα δεδομένα προέρχονται από την κανονική κατανομή, δηλαδή οι μέθοδοι *bootstrap* μειώνουν το μέσο τετραγωνικό σφάλμα σε σχέση με τις αρχικές εκτιμήσεις. Ωστόσο το σημαντικό σε αυτές τις περιπτώσεις είναι το μέγεθος της μείωσης. Δηλαδή αν τα δεδομένα προέρχονται από κανινική κατανομή έχουμε μέιωση του μέσου τετραγωνικού σφάλματος κατά 8-10%, ενώ στην περίπτωση που προέρχονται από την t_3 κατανομή μειώνεται κατά 15-18% και στην περίπτωση της χ^2_2 η μείωση είναι 17-20%, όπωσς φαίνεται και στον πίνακα (4.7). Δηλαδή όταν παραβιάζεται η υπόθεση της κανονικότητας δίνουν πολύ καλύτερες εκτιμήσεις από τα κλασσικά μοντέλα.

Πίνακας 4.7: Ποσοστό MSE για την μέση επίδραση θ=0.5, n=k=5, τ^2=0.2

ΚΑΤΑΝΟΜΗ	Παραμετρική	$WILD$
Κανονική	-7.90%	-9.60%
t_3-*Student*	-15.2%	-18.4%
χ^2_2	-17.4%	-20.2%

Βιβλιογραφία

[1] Wim Van Den Noortage and Patrick Onghena: Parametric and non parametric bootstrap methods for meta-analysis, 2005

[2] Efron, B and Tibshirani,R.J.: An Introduction to the Bootstrap, 1994

[3] Michael Borenstein, Larry V. Hedges, Julian P.T. Higgins, Hannah R. Rothstein: Introduction to Meta-Analysis, 2009

[4] Sarah E. Brockwell and Ian R. Gordon: A comparison of statistical methods for meta-analysis, 2001

[5] Rebecca M. Turner, Rumana Z. Omar, Min Yang, Harvey Goldstein and Simon G. Thompson: A multilevel model framework for meta-analysis of clinical trials with binary outcomes, 2000

[6] Joop Hox: Multilevel Analysis Techniques and Applications, 2002

[7] Anne Whitehead: Meta-analysis of Controlled clinical Trials, 2002

[8] JAMES G. MacKINNON: Bootstrap Methods in Econometrics, 2006

[9] H. GOLDSTEIN: Multilevel mixed linear model analysis using generalized least squares, 1986

[10] Παρασκευή Ι. Στρατή: META ANALYSIS, 2007

[11] Ευριδίκη Πατελάρου, Ηρώ Μπροκαλάκη: Μεθοδολογία της Συστηματικής Ανασκόπησης και Μετα-ανάλυσης, 2010

[12] Αναστασία Γ. Ελευθεράκη: ΜΕΤΑ-ΑΝΑΛΥΣΗ Στατιστική Μεθοδολογία και Θεμελίωση, 2004

[13] Δημήτρης Καρλής: Στατιστικές υπολογιστικές μεθόδοι, 2004

[14] Ανδρομάχη Μαυροζούμη: Εισαγωγή στη Μέθοδο *Bootstrap*, 2008