



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΙΑΤΡΙΚΗ

Μεθοδολογίες Μετα-ανάλυσης γενετικών δεδομένων

Νικόλαος Κακούρας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
Επιβλέπων
Καθηγητής Παντελής Μπάγκος

Λαμία, 2019



UNIVERSITY OF THESSALY

SCHOOL OF SCIENCE

INFORMATICS AND COMPUTATIONAL BIOMEDICINE

Methodology for meta-analysis in genetic data

Nikolaos Kakouras

Master thesis

Professor Pantelis Bagos

Lamia

2019



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΔΙΑΤΜΗΜΑΤΙΚΟ ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ
ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΙΑΤΡΙΚΗ
ΚΑΤΕΥΘΥΝΣΗ**

«ΥΠΟΛΟΓΙΣΤΙΚΗ ΙΑΤΡΙΚΗ ΚΑΙ ΒΙΟΛΟΓΙΑ»

Μεθοδολογίες Μετα-ανάλυσης γενετικών δεδομένων

Νικόλαος Κακούρας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
Επιβλέπων
Καθηγητής Παντελής Μπάγκος

Λαμία, 2019

«Υπεύθυνη Δήλωση μη λογοκλοπής και ανάληψης προσωπικής ευθύνης»

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, και γνωρίζοντας τις συνέπειες της λογοκλοπής, δηλώνω υπεύθυνα και ενυπογράφως ότι η παρούσα εργασία με τίτλο «Μεθοδολογίες Μετα-ανάλυσης γενετικών δεδομένων» αποτελεί προϊόν αυστηρά προσωπικής εργασίας και όλες οι πηγές από τις οποίες χρησιμοποίησα δεδομένα, ιδέες, φράσεις, προτάσεις ή λέξεις, είτε επακριβώς (όπως υπάρχουν στο πρωτότυπο ή μεταφρασμένες) είτε με παράφραση, έχουν δηλωθεί κατάλληλα και ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής.

Ο ΔΗΛΩΝ

Ημερομηνία

30 Οκτωβρίου 2019

Υπογραφή

Μεθοδολογίες Μετα-ανάλυσης γενετικών δεδομένων

Νικόλαος Κακούρας

Τριμελής Επιτροπή:

Καθηγητής Παντελής Μπάγκος

Αναπληρώτρια Καθηγήτρια Μαρία Αδάμ

Επίκουρος Καθηγήτρια Γεωργία Μπράλιου

Πίνακας Περιεχομένων

Περιεχόμενα.....	Σελ. viii
1. Εισαγωγή.....	Σελ. 1
1.2. Μετα – Ανάλυση γενετικών μελετών.....	Σελ. 2
1.2.1 Μελέτες με τη χρήση των ασθενών – μαρτύρων.....	Σελ. 2
1.2.2 Αντιστοίχιση – Ταύτιση των ασθενών – μαρτύρων.....	Σελ. 2
1.2.3 Συλλογή δεδομένων και Ανάλυση.....	Σελ. 2
1.2.4 Μετα – Ανάλυση ασθενών - μαρτύρων με χρήση λογιστικής παλινδρόμησης.....	Σελ. 3
1.2.5 Μετα-Ανάλυση μελετών με συνεχή δεδομένα	Σελ. 5
2. Υλικά και Μέθοδοι.....	Σελ. 7
2.1. Μετα- Ανάλυση γενετικών μελετών και πολύ-μεταβλητή μετα-ανάλυση κάνοντας χρήση συνοπτικών δεδομένων.....	Σελ. 7
2.1.1 Το μοντέλο της λογιστικής παλινδρόμησης.....	Σελ. 7
2.1.2 Αξιολόγηση της ετερογένειας.....	Σελ. 10
2.1.3 Επιτρεπτή ετερογένεια.....	Σελ. 11
2.1.4 Δοκιμές γενετικών μοντέλων.....	Σελ. 12
2.1.5 Γενικό πολύ-μεταβλητό μοντέλο μετα-ανάλυσης για διακριτά δεδομένα	Σελ. 13
2.1.6 Γενικευμένο πολυμεταβλητό μοντέλο μετα-ανάλυσης για συνεχή δεδομένα.....	Σελ. 15
2.1.7 Προσέγγιση χωρίς γονιδιακό μοντέλο.....	Σελ. 18
2.2. Πολυμεταβλητή Μετα-Ανάλυση με τη χρήση δίτιμων δεδομένων..	Σελ. 18
2.3. Η εφαρμογή και τα μέσα τα οποία χρειάστηκαν.....	Σελ. 20
2.4. PHP και Design Patterns.....	Σελ. 22
2.4.1 Η γλώσσα PHP.....	Σελ. 22
2.4.2 Design Patterns.....	Σελ. 24
2.4.2.1 Singleton Pattern.....	Σελ. 25
2.4.2.2 Decorator Pattern.....	Σελ. 27
2.4.2.3 Façade Pattern.....	Σελ. 30
2.4.2.4 Iterator Pattern.....	Σελ. 32
2.5. HTML και Javascript.....	Σελ. 35
2.5.1 Front-End και Back-End προγραμματισμός.....	Σελ. 35
2.5.1.1 Front-End.....	Σελ. 35
2.5.1.2 Back-End.....	Σελ. 36
2.5.2 HTML.....	Σελ. 37
2.5.3 Javascript.....	Σελ. 37
3. Αποτελέσματα και Συζήτηση.....	Σελ. 39
3.1. Χρόνοι απόκρισης εφαρμογής.....	Σελ. 39

3.2. Εφαρμογή.....	Σελ. 42
3.3. Συμπεράσματα.....	Σελ. 45
ΠΑΡΑΡΤΗΜΑ Ι.....	Σελ. 46
ΠΑΡΑΡΤΗΜΑ ΙΙ.....	Σελ. 47
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	Σελ. 48

1. Εισαγωγή

Ένας από τους πιο γρήγορα αναπτυσσόμενους τομείς των Βιοεπιστημών είναι η γενετική επιδημιολογία. Όπως πολύ εύκολα μπορεί να καταλάβει κανείς, η γενετική επιδημιολογία είναι ένας τομέας που έχει προκύψει από τη συνεργασία δύο άλλων, της παραδοσιακής γενετικής και της επιδημιολογίας. Οι δύο αυτοί τομείς έχουν ως κύριο μέλημά τους τη μελέτη των γενετικών στοιχείων των ασθενειών, τον τρόπο με τον οποίο γενετικοί παράγοντες δρουν, κι ακόμη τον τρόπο με τον οποίο οι περιβαλλοντολογικοί παράγοντες δρουν πάνω σε μεγάλους πληθυσμούς. Τα αποτελέσματα των ως άνω μελετών είναι μείζονος σημασίας, όταν τα δεδομένα που χρησιμοποιήσαμε συσχετίζονται μεταξύ τους. Δε θα είχε όμως νόημα να αναφερόμαστε σε συσχετισμένα δεδομένα αν δεν μπορούσαμε και να τα διαχειριστούμε κατάλληλα. Για να το κατορθώσουμε αυτό, να διαχειριστούμε αυτού του είδους τα δεδομένα, έπρεπε να αναπτύξουμε – σχεδιάσουμε ορισμένες νέες στατιστικές μεθόδους, οι οποίες θα μας βοηθούσαν σε αυτό το εγχείρημά μας. Οι στατιστικές αυτές μέθοδοι που αναπτύχθηκαν και που αυξάνονται σε αριθμό με την πάροδο των ετών, έχουν βασιστεί σε παλαιότερα στατιστικά μοντέλα με τη μόνη διαφορά πως έχουν παραλλαχθεί για τις ειδικές περιπτώσεις τις οποίες καλούμαστε να διαχειριστούμε.

Στην πορεία του χρόνου γίνεται φανερό πως ο τομέας της γενετικής επιδημιολογίας έχει μια ιδιαίτερα μεγάλη άνθιση σε ό,τι αφορά τη συγγραφή και δημοσίευση διαφόρων μελετών. Αυτό σημαίνει πως δημιουργείται μια πληθώρα από μελέτες οι οποίες εμπεριέχουν πληροφορίες (γενετικές και μη) που αφορούν μια συγκεκριμένη γενετική ασθένεια. Για να μπορέσουμε να αξιοποιήσουμε όσο το δυνατόν καλύτερα, όλες αυτές τις πληροφορίες που μας παρέχονται από τις παραπάνω μελέτες, δημιουργήθηκε μια διαδικασία που είναι γνωστή με τον όρο: Μετα-Ανάλυση. Η μετα-ανάλυση μας παρέχει μια συνολική εκτίμηση, όπως αυτή παρέχεται από τις προαναφερθείσες μελέτες, κι αυτό μας δίνει τη δυνατότητα να έχουμε μια καλύτερη εικόνα για το θέμα πάνω στο οποίο δουλεύουμε. Η μετα-ανάλυση μελετών που άπτονται του τομέα της γενετικής επιδημιολογίας, θα μπορούσε να πει κανείς πως δύναται να χωριστεί σε δύο κατηγορίες μεθόδων, σε αυτές που συνδυάζουν στοιχεία από μελέτες που αφορούν την ίδια οικογένεια μελετών και σε αυτές που κάνουν συνδυασμό των αποτελεσμάτων της μετα-ανάλυσης από μελέτες που βασίζονται στον πληθυσμό, αλλά και στην οικογένεια στην οποία ανήκουν. Ένα σημαντικό χαρακτηριστικό των μελετών που αφορούν πληθυσμιακές μελέτες είναι πως γίνεται συχνά χρήση του μοντέλου «Ασθενών - Μαρτύρων» (case-control). Με την πάροδο των ετών έχουν υπάρξει κι άλλα μοντέλα που κάνουν χρήση πιο εξελιγμένων μεθόδων χρησιμοποιώντας πολυμεταβλητά συστήματα μετα-ανάλυσης, αλλά υπάρχουν και οι πιο παραδοσιακές μέθοδοι μετα-ανάλυσης τέτοιου τύπου μελετών, οι οποίες τροποποιούνται κατάλληλα για την ορθή διαχείριση των διαφόρων γονοτυπικών φαινομένων. Τέλος, θα ήταν καλό να αναφέρουμε μια ιδιαίτερα σημαντική μέθοδο, που ονομάζεται «Μενδελική Τυχαιοποίηση» (“Mendelian Randomization”), η οποία θα οδηγήσει τον ερευνητικό κόσμο της γενετικής επιδημιολογίας σε σημαντικές ανακαλύψεις σε ζητήματα όπως τα πολλαπλά ζεύγη συσχετίσεων μεταξύ του φαινότυπου με τον γονότυπο και του γονότυπου με κάποια συγκεκριμένη ασθένεια, ώστε τελικά να κατανοήσουμε τη σχέση που υπάρχει μεταξύ του φαινότυπου και μιας ασθένειας.

1.2. Μετα – Ανάλυση γενετικών μελετών

1.2.1 Μελέτες με τη χρήση των ασθενών - μαρτύρων

Κατά καιρούς δημοσιεύονται μελέτες στις οποίες γίνεται χρήση του μοντέλου ασθενών - μαρτύρων. Μέσω αυτού του μοντέλου ο ερευνητής έχει τη δυνατότητα να πραγματοποιήσει μια έρευνα σχετικά με το αν μια έκθεση για παράδειγμα σε κάποια μόλυνση-ασθένεια συσχετίζεται με ένα αποτέλεσμα, το οποίο αποτέλεσμα μπορεί να είναι είτε η εκδήλωση κάποιας ασθένειας είτε γενικότερα κάποια κατάσταση ενδιαφέροντος. Το γενικότερο πλαίσιο των μελετών αυτών, είναι η δημιουργία δύο ομάδων ατόμων, όπου η μία ομάδα θα αφορά τους ασθενείς, τα οποία σχετίζονται με άτομα στα οποία γνωρίζουμε ότι υπάρχουν αποτελέσματα. Η δεύτερη ομάδα που δημιουργούμε αφορά τους μάρτυρες, στην οποία υπάρχουν άτομα μέσω των οποίων γνωρίζουμε πως είναι ανεξάρτητα του αποτελέσματος ή αλλιώς είναι τα άτομα τα οποία είναι υγιή. Στη συνέχεια κάνουμε μια αναδρομή στο παρελθόν του ατόμου ή του «ατόμου» της κάθε ομάδας, ώστε να συγκρίνουμε τη συχνότητα της έκθεσής του, σε κάποιον παράγοντα που είναι υπεύθυνος για κάποια ασθένεια, σε κάθε μία από τις προαναφερθείσες ομάδες. Θα μπορούσαμε να αναφέρουμε ότι οι μελέτες τύπου ασθενών - μαρτύρων έχουν δύο ιδιαίτερα σημαντικά πλεονεκτήματα. Το πρώτο πλεονέκτημα που χαρακτηρίζει τις μελέτες κατάλληλες είναι η δυνατότητα που παρέχετε για τη διερεύνηση μιας για παράδειγμα ασθένειας πάνω στην έξαρσή της. Το δεύτερο ήσσονος σημασίας, είναι ότι σε τέτοιου είδους μελέτες έχουμε τη δυνατότητα της μελέτης σπάνιων ασθενειών. Ένα ακόμη σημείο στο οποίο πρέπει να σταθούμε, είναι ότι μας δίνονται γρηγορότερα τα αποτελέσματα τα οποία χρειαζόμαστε, μιας και οι πληθυσμοί τους οποίους χρησιμοποιούμε γνωρίζουμε τα αποτελέσματά τους.

1.2.2 Αντιστοίχιση – Ταύτιση των ασθενών - μαρτύρων

Ένα συμπέρασμα το οποίο μπορούμε να εξάγουμε από τα παραπάνω είναι πως οι δύο ομάδες που δημιουργήσαμε –η μία με τους ασθενείς και η δεύτερη με τους μάρτυρες, εμπεριέχουν δεδομένα τα οποία μοιάζουν μεταξύ τους. Αυτή η ομοιότητα μεταξύ των δύο ομάδων μερικές φορές δεν αντιστοιχεί σε ένα προς ένα ταύτιση, αλλά σε ένα προς πολλά ή όπως αλλιώς αναφέρεται έχουμε μια «υπέρ- ταύτιση». Αυτή λοιπόν η «υπέρ- ταύτιση» μπορεί να δυσκολέψει την εύρεση αρκετών μαρτύρων.

Μια ενδιαφέρουσα τεχνική, ώστε να ισχυροποιήσουμε τέτοιου είδους μελέτες, είναι η προσθήκη περισσότερων του ενός ασθενούς, για την κάθε περίπτωση. Η προσθήκη περισσότερων ασθενών αλλά και μαρτύρων είναι ιδιαίτερα σημαντική, μιας και δίνει καλύτερα στατιστικά αποτελέσματα.

1.2.3 Συλλογή δεδομένων και Ανάλυση

Έχοντας καθορίσει με ακρίβεια ποιόι θα είναι οι ασθενείς και ποιόι οι μάρτυρες, στη συνέχεια θα πρέπει να τα συλλέξουμε. Ένα χαρακτηριστικό που αφορά τη συλλογή των δεδομένων είναι ότι πρέπει να γίνει μια αναζήτηση του ιστορικού του κάθε ατόμου, μέσω του οποίου θα ενημερωθεί ο ερευνητής για παλαιούς παράγοντες που μπορεί να είναι υπαίτιοι, για παράδειγμα της ασθένειας την οποία εξετάζουμε. Το βήμα που περιγράφηκε είναι ιδιαίτερα σημαντικό μιας και το αποτέλεσμα μας είναι ήδη γνωστό. Ένα πρόβλημα που προβάλλουν αυτού του είδους οι μελέτες αφορά τη μεροληψία των ερευνητών κατά τη συλλογή των δεδομένων. Η εξάλειψη του προβλήματος αυτού δε μπορεί να υπάρξει, σίγουρα όμως μπορεί να μετριαστεί, αποκρύπτοντας το (γνωστό για τον ερευνητή) αποτέλεσμα από το άτομο το οποίο λαμβάνεται το αντίστοιχο

ιστορικό. Αναφέραμε νωρίτερα πως ο εκάστοτε ερευνητής θα πρέπει να λαμβάνει και το αντίστοιχο ιστορικό από τον ασθενή. Αυτό το ιστορικό, εφόσον αναφερόμαστε σε βιολογικά – ιατρικά δεδομένα μπορεί να αφορά τη διατροφή του ασθενούς ή τη χρήση ορισμένων φαρμάκων και ούτω καθ' εξής. Φυσικά αυτό δεν είναι πάντοτε δυνατό, μιας και ο ασθενής δεν είναι πάντα σε θέση να αναμοχλεύσει λεπτομέρειες από τη μνήμη του για το ιστορικό του.

Σε ό,τι αφορά την ανάλυση των αποκτηθέντων δεδομένων, θα πρέπει να υπολογίσουμε τη συχνότητα εμφάνισης του κάθε δεδομένου – μεταβλητή που ελήφθη σε κάθε μία από τις ομάδες που αναφέραμε παραπάνω. Για να ενισχύσουμε, όσο περισσότερο μπορούμε, τη συσχέτιση μεταξύ της έκθεσης του ασθενούς σε κάποιους παράγοντες με τα δεδομένα, οι μελέτες αυτές κάνουν χρήση του λόγου συμπληρωματικών πιθανοτήτων ή όπως είναι πιο ευρέως διαδεδομένο ως Odds Ratio. Τα odds ratio αφορούν το λόγο μεταξύ των δεδομένων της ομάδας των ασθενών, με αυτά της ομάδας των μαρτύρων. Το επόμενο και πολύ σημαντικό βήμα είναι ο υπολογισμός του διαστήματος εμπιστοσύνης για κάθε λόγο πιθανότητας. Μέσα στο διάστημα εμπιστοσύνης που υπολογίσαμε νωρίτερα, αν συμπεριλαμβάνεται το 1.0, τότε αυτό θα σημαίνει ότι δεν υπάρχει στατιστική σημαντικότητα. Τέλος, θα ήταν πολύ σημαντικό να αναφέρουμε ότι οι μελέτες ασθενών – μαρτύρων δε μπορούν να παρέχουν πληροφορίες σχετικά με τη συχνότητα ή την επικράτηση μιας νόσου, επειδή δεν πραγματοποιούνται μετρήσεις σε δείγμα που βασίζεται στον πληθυσμό.

1.2.4 Μετα – Ανάλυση ασθενών – μαρτύρων με χρήση λογιστικής παλινδρόμησης

Το πιο σύνηθες παράδειγμα που μπορεί κάποιος να δώσει, όταν αναφερόμαστε σε μετα-ανάλυση γενετικών δεδομένων, είναι αυτό των δύο αλληλομόρφων. Αν υποθέσουμε λοιπόν, πως έχουμε δύο αλληλόμορφα A και B, τότε οι πιθανοί γονότυποι που μπορούν να προκύψουν θα είναι AA, AB και BB. Συνεχίζοντας την παραπάνω υπόθεσή μας, θεωρούμε ότι το αλληλόμορφο B εμπεριέχει έναν παράγοντα κινδύνου που εμφανίζεται σε μία ασθένεια. Σε αυτό το σημείο μπορούμε να επισημάνουμε ότι οι περισσότερες μετα-αναλύσεις γίνονται, έχοντας ως βάση τις πιο παραδοσιακές μεθόδους, μέσα από τις οποίες παρατηρούμε την ύπαρξη δύο σημείων με δύο μεταβλητές (A, B), τα οποία παρουσιάζονται με τον διαχωρισμό των γονότυπων σε δύο κατηγορίες, υποθέτοντας κάθε φορά ένα συγκεκριμένο γενετικό μοντέλο.

	ΓΟΝΟΤΥΠΟΣ					
	ΕΛΕΓΧΟΙ			ΠΕΡΙΠΤΩΣΕΙΣ		
ΜΕΛΕΤΕΣ	AA	AB	BB	AA	AB	BB
1	38	52	6	72	78	22
2	44	27	11	38	45	17
3	33	34	8	21	26	11
4	45	53	16	53	87	51

Πίνακας 1: ο παραπάνω πίνακας παρουσιάζει τις τιμές που λαμβάνει ο κάθε γονότυπος στις αντίστοιχες μελέτες

Αξιοσημείωτη είναι η αντίθεση των γονότυπων του BB και AB σε σχέση με τον γονότυπο AA, όπως φαίνεται στον παραπάνω πίνακα. Αυτή η αντίθεση μπορεί να υποδηλώσει ένα επικρατές μοντέλο κληρονομικότητας. Συνεχίζοντας τη σύντομη παρατήρησή μας βλέπουμε πως ο γονότυπος BB σε αντίθεση με τους γονότυπους AB και AA αντιστοιχεί σε ένα επικρατές μοντέλο. Αντίστοιχα, ο γονότυπος AB σε αντίθεση με τους γονότυπους AA και BB αντιστοιχεί σε ένα υπερ-επικρατές μοντέλο και ούτω καθεξής.

Μια άλλη προσέγγιση που παρατίθεται είναι η ίδια με την προηγούμενη με μια επιπλέον προσθήκη. Η προσέγγιση αυτή αφορά την πλειοψηφία των περιπτώσεων που εξετάζουμε. Μέσω αυτής της προσέγγισης κάνουμε σύγκριση των συχνοτήτων των αλληλομόρφων. Οι συγκρίσεις των προαναφερθέντων συχνοτήτων γίνονται μεταξύ των συχνοτήτων των ασθενών (cases) και αυτών των μαρτύρων (controls).

Τέλος, μια ακόμη προσέγγιση η οποία έχει προταθεί είναι η χρήση των πολλαπλών συγκρίσεων. Παρ' όλο που η τελευταία θα μπορούσε να χαρακτηριστεί ως ικανοποιητική, εν τούτοις μέσω αυτής της προσεγγίσεως ελλοχεύει ο κίνδυνος ύπαρξης ενός μεγάλου ποσοστού σφάλματος Τύπου I.

Όπως έχουν αναφέρει χαρακτηριστικά οι Normand και Petiti σε αντίστοιχη εργασία τους, δεν έχει σημασία ποια από τις παραπάνω μεθόδους θα χρησιμοποιήσουμε, μιας και οι συνοπτικές μετρήσεις των παραδοσιακών μεθόδων μετα-ανάλυσης δύνανται να εφαρμοστούν σε μοντέλα με τυχαία δεδομένα, όπως επίσης και σε μοντέλα με σταθερά δεδομένα. Αν υποθέσουμε πως θέλουμε να κάνουμε μετα-ανάλυση κάνοντας χρήση σταθερών δεδομένων, τότε ταυτόχρονα θα πρέπει να κάνουμε και την παραδοχή ότι οι μεμονωμένες εκτιμήσεις κάθε μελέτης του $\log OR$, είναι κανονικά κατανομημένες με ένα μέσο όρο ίσο με το πραγματικό αποτέλεσμα θ και με μια διακύμανση s^2 , η οποία είναι και η εκτιμώμενη διακύμανση του $\log OR$ κάθε μελέτης. Σε αυτό όμως το σημείο αναδύεται ένα αρκετά σημαντικό πρόβλημα, αυτό της ετερογένειας. Μετά από μια σειρά πειραμάτων – ερευνών οι DerSimonian και Laird πρότειναν μια μέθοδο μέσω της οποίας θα επιλύεται το πρόβλημα της ετερογένειας. Η μέθοδος τυχαίων δεδομένων υποθέτει πως τα πραγματικά δεδομένα μεταβάλλονται μεταξύ των μελετών και μιας τυχαίας συνιστώσας από τη διακύμανση (τ^2) των μελετών που έχουμε. Η διακύμανση (τ^2) μπορεί να υπολογιστεί μη-επαναληπτικά από τη μέθοδο των στιγμών (MM = Method of Moments) ή χρησιμοποιώντας πιο σύγχρονες τεχνικές όπως η μέθοδος των ροών της μέγιστης πιθανοφάνειας (ML = Maximum Likelihood), της εμπειρικής εκτίμησης του Bayes (EB = Empirical Bayes) και της περιορισμένης μέγιστης πιθανοφάνειας (REML = Restricted Maximum Likelihood). Είναι σημαντικό να αναφέρουμε πως η προαναφερθείσα μέθοδος (DSL = DerSimonian Laird) θα μας επιστρέψει περισσότερο διαφορετικά αποτελέσματα μεταξύ των τυχαίων και σταθερών δεδομένα, όταν η ετερογένεια είναι ισχυρή (μεγάλη). Αντιθέτως, όταν η ετερογένεια είναι μικρή, τότε θα έχουμε παρόμοια αποτελέσματα. Στο σημείο αυτό πρέπει να αναφέρουμε πως υπάρχουν πολλοί υποστηρικτές της άποψης ότι ο υπολογισμός των συνοπτικών μετρήσεων είναι ακατάλληλος εφόσον υπάρχει ετερογένεια και επίσης ότι δεν υπάρχει σαφής εκτίμηση τους. Όπως χαρακτηριστικά ανέφεραν οι Thompson και Sharp, για να μειώσουμε την ετερογένεια μπορούμε να χρησιμοποιήσουμε μια μεταβλητή στο επίπεδο της μελέτης, η οποία θα χρησιμοποιείται ως ένας γραμμικός παράγοντας πρόβλεψης, με αποτέλεσμα την μετα-παλινδρόμηση των τυχαίων δεδομένων.

Στο σημείο αυτό, θα ήταν καλό να επισημάνουμε πως η παραπάνω προσέγγιση, παρ' όλο που φαίνεται ιδιαίτερα καλή, εν τούτοις υπάρχουν ορισμένα προβλήματα, τα οποία χρήζουν επίλυσης. Το πιο σημαντικό πρόβλημα που κάνει την εμφάνισή του είναι αυτό το οποίο αφορά την κατανομή των εκτιμήσεων (y_i), ιδιαίτερα όταν το πλήθος των μελετών που χρησιμοποιούμε είναι μικρό. Τέλος, τα μοντέλα των δεδομένων είτε αυτά είναι τυχαία είτε σταθερά, θεωρούν τη διακύμανση (s^2) ότι είναι γνωστή. Ένα άλλο πρόβλημα που προκύπτει, ιδιαίτερα σε μελέτες που έχουν γενετική συσχέτιση, είναι σε ό,τι έχει σχέση με τη φύση των δεδομένων, δηλαδή όταν έχουμε πολλαπλούς γονότυπους. Αυτό διότι πολλαπλοί γονότυποι θα έχουν και πολλαπλές συγκρίσεις. Ιδιαίτερο ενδιαφέρον παρουσιάζει η πρόταση που έκανε το 2005 η Minelli για τη μετα-ανάλυση γενετικών μελετών, χωρίς να κάνει χρήση από πριν κάποιου γενετικού μοντέλου, αλλά δημιουργεί ένα, έχοντας ως κύριο γνώμονα τα δεδομένα των μελετών. Συγκεκριμένα, εισήγαγε τον λογάριθμο του OR_{BB} ($\log OR_{BB}$), ο οποίος είναι το odd ratio του BB έναντι του AA, και το λ είναι το πηλίκο μεταξύ του $\log OR_{BB}$ και του $\log OR_{AB}$. Από τα παραπάνω, μπορεί

κανείς να αντιληφθεί ότι τα δύο OR συσχετίζονται. Το σπουδαιότερο σημείο στο οποίο πρέπει να σταθούμε είναι ότι η παραπάνω μέθοδος πρέπει να χρησιμοποιήσει ορισμένες επεκτάσεις προκειμένου να ενσωματωθούν καταστάσεις διαφορετικές από τα SNPs (SNP = Single Nucleotide Polymorphisms) ή στις περιπτώσεις που υφίσταται η παραπάνω επέκταση να μπορεί να ενσωματώνει μεταβλητές. Αργότερα, η Minelli, επέκτεινε την παραπάνω μέθοδο, ώσπου τα τελευταία χρόνια ο Salanti εισήγαγε μια Bayesian μέθοδο, όπως είχε κάνει και η Minelli λίγο ωρύτερα, κάνοντας χρήση του όρου του συντελεστή σταθεροποίησης.

Τέλος, θα ήταν μεγάλη παράλειψη αν δεν αναφέραμε τη μέθοδο της Μενδελικής Τυχαιοποίησης. Με την ως άνω μέθοδο έχουμε τη δυνατότητα να λάβουμε υπόψη τα ζεύγη των συσχετίσεων μεταξύ φαινότυπου-γονότυπου και γονότυπου-ασθένειας.

1.2.5 Μετα-Ανάλυση μελετών με συνεχή δεδομένα

Ένα πολύ συχνά παρατηρούμενο φαινόμενο στις μελέτες είναι πως τα δεδομένα που χρησιμοποιούμε για τη μετα-ανάλυση ταξινομούνται εκ των υστέρων στις αναλύσεις ασθενών και μαρτύρων, τα οποία προσδιορίζονται από τον γονότυπό τους και μπορούν να αναπαρασταθούν με έναν πίνακα, όπως είδαμε παραπάνω. Όμως, όπως θα δούμε και σε αυτή την παράγραφο, υπάρχουν δεδομένα όπως είναι αυτά που αφορούν τα επίπεδα της χοληστερόλης ή της συστολικής αρτηριακής πίεσης, τα οποία είναι συνεχή. Παρ' όλο που δεν είναι τόσο συνηθισμένο φαινόμενο, εν τούτοις παρουσιάζει ιδιαίτερο ενδιαφέρον. Σε αυτού του είδους τις περιπτώσεις έχουμε ταξινόμηση των μελετών που έχουν διατομιακό σχεδιασμό, το οποίο είναι και το αντιπροσωπευτικό δείγμα από τον ευρύ πληθυσμό απ' όπου συλλέχθηκε. Στη συνέχεια, λαμβάνουμε τους γονότυπους των ατόμων, οι οποίοι συγκρίνονται με τις μέσες τιμές (mean values) των συνεχών μέτρων.

Μέχρι αυτή τη στιγμή, οι περισσότερες μετα-αναλύσεις –μελετών που είναι γενετικά συνδεδεμένες – έχουν ως επίκεντρο δύο μεταβλητές (A, B). Στις μετα-αναλύσεις αυτές χωρίζουμε τον γονότυπο σε δύο κατηγορίες, θεωρώντας ένα συγκεκριμένο γενετικό μοντέλο. Αυτό μας δίνει τη δυνατότητα να κάνουμε συγκρίσεις είτε μέσω της αναλογικής πιθανότητας ($OR = Odds Ratio$), είτε κάνοντας χρήση της μέσης διαφοράς η οποία είναι περισσότερο γνωστή ως «σταθμισμένη μέση διαφορά» ($WMD = Weighted Mean Difference$).

Όταν αναφερόμαστε σε μετα-ανάλυση μελετών με συνεχή γενετικά δεδομένα, έχουμε τη δυνατότητα να κάνουμε χρήση της σταθμισμένης μέσης διαφοράς. Αυτή η προσέγγιση αφορά τη σύγκριση της συχνότητας των αλληλομόρφων μεταξύ των Ασθενών – Μαρτύρων. Αυτή η προσέγγιση μας οδηγεί στο συμπέρασμα ότι κάνουμε χρήση ενός κωδικοποιούμενου μοντέλου κληρονομικότητας. Ολοκληρώνοντας τους τρόπους μέσω των οποίων μπορούμε να διενεργήσουμε τη μετα-ανάλυση σε μελέτες με συνεχή δεδομένα, είναι και αυτός που εκτελεί πολλαπλές συγκρίσεις μεταξύ των γονοτύπων. Η τελευταία όμως αυτή προσέγγιση ελλοχεύει την εμφάνιση ποσοστού σφάλματος Τύπου I.

Όπως αναφέρουν οι Normand και Petiti στις αντίστοιχες μελέτες τους, δεν έχει σημασία ποια από τις παραπάνω προσεγγίσεις θα χρησιμοποιήσουμε, μιας και οι παραδοσιακές μέθοδοι μετα-ανάλυσης κάνοντας χρήση συνοπτικά μέτρα, εφαρμόζονται και σε αυτές που βασίζονται σε τυχαία γεγονότα, όπως επίσης και σε αυτές που βασίζονται σε μοντέλα σταθερών γεγονότων.

Στις μετα-αναλύσεις παραδοσιακών σταθερών δεδομένων, κάνουμε χρήση συνοπτικών μέτρων, κάνοντας την υπόθεση ότι οι μεμονωμένες εκτιμήσεις y_i του $\log OR$ ή αλλιώς της διαφοράς του μέσου όρου, της κάθε μελέτης είναι κανονικά κατανομημένες, με μέσο όρο ίσο με το πραγματικό αποτέλεσμα θ και τη διακύμανση s^2 , η οποία είναι η εκτιμώμενη διακύμανση του $\log OR$ (ή της μέσης διαφοράς) της συγκεκριμένης αντίθεσης κάθε μελέτης.

Ένα χαρακτηριστικό πέραν όλων των παραπάνω που αναφέραμε, είναι αυτό της ετερογένειας. Μια απ' τις πιο καλές μεθόδους για την υπερπήδηση αυτού του προβλήματος είναι αυτή των τυχαίων δεδομένων. Με τη χρήση αυτής της μεθόδου υποθέτουμε πως τα πραγματικά δεδομένα μεταβάλλονται τυχαία μεταξύ των μελετών, γι' αυτόν τον λόγο χρησιμοποιούμε μια τυχαία συνιστώσα τ^2 . Για να μπορέσουμε να υπολογίσουμε την παραπάνω συνιστώσα αρκεί να κάνουμε χρήση της μεθόδου των στιγμών, MM (Method of Moments), η οποία είναι μια μη επαναληπτική μέθοδος που έχει ήδη προταθεί από τους DerSimonian και Laird.

Στον αντίποδα έχουμε τους Thompson και Sharp, οι οποίοι προτείνουν επαναληπτικές τεχνικές, όπως αυτή των ροπών της μέγιστης πιθανοφάνειας (ML = Maximum Likelihood), της εμπειρικής εκτίμησης Bayes (EB = Empirical Bayes) και αυτή της περιορισμένης μέγιστης πιθανοφάνειας REML (REML = Restricted Maximum Likelihood). Είναι σημαντικό να αναφέρουμε ότι, αν δεν υπάρχει ετερογένεια, τότε η συνιστώσα που χρησιμοποιήσαμε παραπάνω είναι ουσιαστικά ίση με το μηδέν και επίσης οι μέθοδοι σταθερού και τυχαίου αποτελέσματος συμπίπτουν. Ιδιαίτερης σημασίας είναι το γεγονός ότι σε επίπεδο μελέτης μπορούμε να ενσωματώσουμε συγκριτικά επίπεδα σε όλες τις προαναφερθείσες τεχνικές – μεθόδους. Αυτά τα συγκριτικά επίπεδα λειτουργούν ως γραμμικοί παράγοντες πρόβλεψης, μέσω των οποίων υπάρχει κάποια μείωση της παρατηρούμενης ετερογένειας, ούτως ώστε να καταλήξουμε σε μια «μετα-παλινδρόμηση τυχαίων δεδομένων», όπως χαρακτηριστικά ανέφεραν οι Thompson, Higgins και Sharp στις αντίστοιχες μελέτες τους.

Εκτός από τις παραπάνω μεθόδους μετα-ανάλυσης που αναφέρθηκαν προηγουμένως, υπάρχει και ένα ακόμη μοντέλο το οποίο ανήκει σε ένα πιο γενικό πλαίσιο. Το τελευταίο αυτό μοντέλο περιλαμβάνει ορισμένα προκαθορισμένα βήματα, τα οποία περιλαμβάνουν δοκιμές για την ισορροπία Hardy-Weinberg (HWE). Μέσω των παραπάνω βημάτων διενεργούνται και δοκιμές που αφορούν την ετερογένεια, ώστε τελικά να αποκρυπτογραφεί το πιο πιθανό γενετικό μοντέλο. Σε αυτό το σημείο θα πρέπει να αναφέρουμε ότι στη συγκεκριμένη μέθοδο κάνουμε χρήση σταθερών και τυχαίων δεδομένων με βάση τα συνοπτικά δεδομένα. Το σημαντικότερο όμως είναι ότι τα εγγενή πολύ-μεταβλητά γενετικά δεδομένα αντιμετωπίζονται ως μονο-μεταβλητά, ώστε να εξάγουν ταυτόχρονα συμπεράσματα – αποτελέσματα.

Στον παρακάτω πίνακα (Πίνακας 2) παρατίθενται τα δεδομένα από ένα μεταλλαγμένο αλληλόμορφο, τα οποία παρουσιάζονται ως μια διμεταβλητή απάντηση των οποίων οι εκτιμήσεις συσχετίζονται. Σε αυτό το σημείο πρέπει να αναφέρουμε ότι τα παραπάνω δεδομένα μπορούν να διαμορφωθούν ανάλογα με τη συμπεριφορά των δύο logORs που προέρχονται από το παραπάνω μεταλλαγμένο αλληλόμορφο. Όπως είδαμε παραπάνω, έχουμε συσχέτιση των εκτιμήσεων που έχουν προκύψει από τα δοθέντα δεδομένα, αυτό είναι ιδιαίτερα σημαντικό μιας και η συσχέτιση αυτή μας είναι απαραίτητη, όταν θέλουμε να πάρουμε ταυτόχρονα συμπεράσματα, τα οποία σχετίζονται με τη στατιστική σημασία των δύο logORs, όπως επίσης και όταν θέλουμε να τα συγκρίνουμε, ώστε να αποκρυπτογραφηθεί το γενετικό μοντέλο της κληρονομικότητας.

Μελέτες	Ασθενείς (Cases)			Μάρτυρες (Controls)		
1	AA ₁₁	AB ₁₁	BB ₁₁	AA ₀₁	AB ₀₁	BB ₀₁
2	AA ₁₂	AB ₁₂	BB ₁₂	AA ₀₂	AB ₀₂	BB ₀₂
...
k	AA _{1k}	AB _{1k}	BB _{1k}	AA _{0k}	AB _{0k}	BB _{0k}

Πίνακας 2: Στον παραπάνω πίνακα παρουσιάζεται μια τυπική διάταξη των δεδομένων που χρησιμοποιήθηκαν σε μια μετα-ανάλυση γενετικής συσχέτισης

Μια ιδιαίτερα σημαντική παρατήρηση που μπορούμε να κάνουμε σε αυτό το σημείο είναι ότι δεν έχουν γίνει προσπάθειες προσαρμογής των γενετικών μοντέλων σε μελέτες γενετικής

συσχέτισης για τη μετα-ανάλυση πολυμεταβλητών δεδομένων. Το σημαντικότερο όμως είναι πως ο μεγάλος αριθμός των μετα-αναλύσεων που πραγματοποιούνται, χρησιμοποιούν τα πολλαπλά δεδομένα ως ανεξάρτητα είτε ως δεδομένα πολλαπλών συγκρίσεων.

Το 2005 προτάθηκε ένα πολύ ενδιαφέρον πολυμεταβλητό μοντέλο, μιας και μέχρι εκείνη τη στιγμή δεν υπήρξε κάποια πρόταση για κάποια νέα προσέγγιση. Το πολυμεταβλητό μοντέλο το οποίο προτάθηκε εισήγαγε μια κοινή μοντελοποίηση του λογαρίθμου OR_{BB} , το οποίο αντιστοιχεί στο OR (Odds Ratio) του γονότυπου BB έναντι του AA, με το λ να δηλώνει το πηλίκο των λογαρίθμων $\log OR_{BB}$ και του $\log OR_{AB}$. Με την πρώτη κιόλας ματιά, μπορεί να παρατηρήσει κανείς ότι η συγκεκριμένη προσέγγιση αναγνωρίζει πως τα δύο ORs συσχετίζονται μεταξύ τους. Σε πολύ σύντομο χρονικό διάστημα το παραπάνω μοντέλο επεκτάθηκε, προσθέτοντάς του ένα Bayesian πλαίσιο (Bayesian framework).

Στη συνέχεια η Salanti μαζί με τους συνεργάτες της εισήγαγαν κι άλλη Bayesian μέθοδο, μέσω της οποίας έχουμε άμεση ενσωμάτωση των αποκρίσεων στη μετα-ανάλυση που διενεργούμε, από το HWE (Hardy-Weinberg). Λίγο μετά έχουμε την εμφάνιση μιας διαφορετικής προσέγγισης που ακολουθήθηκε στην ήδη υπάρχουσα μέθοδο με το όνομα «Μενδελική Τυχαιοποίηση» (Mendelian Randomization) της οποίας τον σκοπό λειτουργίας της τον είδαμε παραπάνω. Μέσω της τελευταίας αυτής μεθόδου θα μπορούσαμε τελικά να αποκρυπτογραφήσουμε τη ζητούμενη σχέση μεταξύ φαινότυπου και ασθένειας, λαμβάνοντας υπόψη τις πολλαπλές πηγές της βιβλιογραφίας. Τέλος, το 2007 προτάθηκε μια διαισθητική προσέγγιση, η οποία είναι βασισμένη σε μοντέλα λογιστικής παλινδρόμησης τυχαίων επιδράσεων. Για την υλοποίηση της τελευταίας αυτής προσέγγισης έγινε χρήση του λογισμικού πακέτου του Stata.

Ολοκληρώνοντας αυτή την ενότητα είναι σημαντικό να αναφέρουμε ότι οι παραπάνω μέθοδοι, παρ' όλο που φαίνονται αξιόπιστοι, εν τούτοις η χρήση τους δεν είναι ευρέως διαδεδομένη. Ένας σημαντικός λόγος που δε χρησιμοποιούνται τόσο συχνά είναι αρχικά διότι οι Bayesian μέθοδοι, παρ' όλο που φαίνονται καταλληλότεροι, εν τούτοις είναι αρκετά δύσκολο να χρησιμοποιηθούν, μιας και οι αντίστοιχοι ερευνητές που θα θελήσουν να διενεργήσουν μια μετα-ανάλυση θα πρέπει να κατέχουν δεξιότητες σε ό,τι αφορά τη Bayesian στατιστική, όπως επίσης και προγραμματιστικές δεξιότητες στο περιβάλλον του WinBUGS. Ένα επιπρόσθετο χαρακτηριστικό το οποίο αποτρέπει τη χρήση των Bayesian μεθόδων είναι η διάθεση σημαντικού χρονικού διαστήματος, ώστε να παρακολουθηθεί η σύγκλιση της MCMC (Markov Chain Monte Carlo = είναι μέθοδοι μιας ομάδας αλγορίθμων που χρησιμοποιούνται για δειγματοληψία από μια κατανομή πιθανότητας), όπως και για την εκτέλεση των απαραίτητων διαγνωστικών. Τέλος, ένας ακόμη λόγος που οι παραπάνω μέθοδοι δε χρησιμοποιούνται αρκετά είναι διότι τα γενετικά ελεύθερα μοντέλα και οι Mendelian τυχαιοποιημένες προσεγγίσεις έχουν γραφεί σε προγράμματα, τα οποία δεν έχουν δημοσιευθεί με τις αντίστοιχες μελέτες.

2. Υλικά και Μέθοδοι

2.1 Μετα- Ανάλυση γενετικών μελετών και πολύ-μεταβλητή μετα- ανάλυση κάνοντας χρήση συνοπτικών δεδομένων

2.1.1 Το μοντέλο της λογιστικής παλινδρόμησης

Όπως εύκολα μπορούμε να παρατηρήσουμε τον Πίνακα 1, βλέπουμε πως είναι ένας πίνακας τριών ενδεχομένων. Η παρούσα μορφή του προαναφερθέντος πίνακα έχει μελετηθεί αρκετά, όσο αφορά την κατηγορηματική ανάλυση. Ένας τέτοιος πίνακας γίνεται ιδιαίτερα κατανοητός, όταν μοντελοποιηθεί κατάλληλα. Για τη μοντελοποίησή του υπάρχουν τρεις τρόποι, όπου και οι

τρεις έχουν ως βάση τις διαφορετικές υποθέσεις σε ό,τι αφορά τη διαδικασία της δημιουργίας των δεδομένων. Στη συνέχεια, θεωρούμε ένα πλαίσιο δειγματοληψίας, στο οποίο θεωρούμε ένα σταθερό συνολικό μέγεθος δείγματος, όπως επίσης και ένα κατάλληλο μοντέλο λογιστικής παλινδρόμησης, μέσω του οποίου οι διαφορετικοί γονότυποι θα θεωρούνται ανεξάρτητες μεταβλητές. Αυτό που μόλις περιγράψαμε, αντιστοιχεί στην αναδρομική πιθανοφάνεια ή όπως είναι ευρέως γνωστή ως *retrospective likelihood*, η οποία εφαρμόζεται πολύ συχνά σε μελέτες τέτοιου είδους (ασθενών - μαρτύρων). Αυτή η πιθανοφάνεια βασίζεται στην πιθανότητα έκθεσης του ασθενούς για παράδειγμα σε κάποια μόλυνση ενώ νοσεί.

Υποθέτουμε ότι οι παρατηρούμενες μετρήσεις πραγματοποιούνται από μια τυχαία μεταβλητή με τη χρήση της κατανομής Poisson, μέσω της οποίας μπορούμε να ξεχωρίσουμε ορισμένα γραμμικά μοντέλα. Σε αυτό το σημείο πρέπει να τονίσουμε ότι η τυχαία μεταβλητή αντιστοιχεί στον αριθμό των μετρήσεων, το οποίο συνεπάγεται ότι οι μελέτες αντιστοιχούν στους γονότυπους και οι καταστάσεις των ασθενών - μαρτύρων θεωρούνται ως ανεξάρτητες μεταβλητές. Ολοκληρώνοντας, θεωρούμε ακόμη ένα «διωνυμικό σχήμα δειγματοληψίας», στο οποίο οι σταθεροί αριθμοί ασθενών και μαρτύρων έχουν ελεγχθεί ανεξάρτητα. Επίσης μπορούμε να μοντελοποιήσουμε και τη δομή του πίνακα κάνοντας χρήση των μεθόδων της λογιστικής παλινδρόμησης, όπου το ασθενείς - μάρτυρες είναι η εξαρτημένη μεταβλητή και οι γονότυποι, όπως και οι ανεξάρτητες μελέτες αντιμετωπίζονται από κοινού. Αυτό που μόλις περιγράψαμε αντιστοιχεί στην περίπτωση της πιθανής πιθανοφάνειας ή όπως είναι ευρέως γνωστό ως *prospective likelihood*. Αυτή η πιθανοφάνεια βασίζεται στην πιθανότητα εμφάνισης της νόσου λόγω της έκθεσης του ασθενούς σε κάποια μόλυνση.

Είναι πολύ σημαντικό να κατανοήσουμε ότι για το ποιο μοντέλο τελικά θα χρησιμοποιήσουμε, θα πρέπει να ληφθούν υπόψη ορισμένοι παράγοντες. Αυτοί οι παράγοντες θα μπορούσε να είναι η αληθοφάνεια, η δυνατότητα ευελιξίας των τεχνικών μοντελοποίησης, όπως επίσης και η δυνατότητα ερμηνείας των αποτελεσμάτων. Χαρακτηριστικό είναι το παράδειγμα του Argesti, όπως ο ίδιος το αναφέρει μέσα στη δική του μελέτη, «για να αντλήσουμε συμπεράσματα κάνοντας χρήση ένα αναδρομικό μοντέλο Poisson, πρέπει να λάβουμε υπόψη όλες τις αμφίδρομες αλληλεπιδράσεις καθώς και το 3-way του μοντέλου».

Ιδιαίτερης σημασίας είναι το αποδεικτικό στοιχείο, που προκύπτει μέσα από το χώρο της επιδημιολογίας, ότι οι εκτιμήσεις των ροπών της μέγιστης πιθανοφάνειας, οι οποίες προκύπτουν από την αναδρομική πιθανοφάνεια είναι οι ίδιες με εκείνες που προκύπτουν από την πιθανή πιθανοφάνεια. Η παραπάνω απόδειξη – παρατήρηση είχε ως αποτέλεσμα την ευρέως χρήση της προσέγγισης της τοποθέτησης των μοντέλων της λογιστικής παλινδρόμησης. Είναι πολύ σημαντικό να αναφέρουμε σε αυτό το σημείο ότι τα αποτελέσματα της πιθανής πιθανοφάνειας συγκρινόμενα με αυτά της αναδρομικής, κάνοντας χρήση της προσέγγισης Bayesian, επιβεβαίωσαν την ισότητα των δύο αυτών προσεγγίσεων.

Ένα σημείο στο οποίο θα ήταν καλό να σταθούμε είναι σε ό,τι αφορά τη λεγόμενη βιολογική αληθοφάνεια και την ευκολία της ερμηνείας των αποτελεσμάτων σε μια τέτοιου είδους ανάλυση. Στην περίπτωση των μελετών ασθενών - μαρτύρων έχουμε αντιστοιχία του παράγοντα κινδύνου με τον γονότυπο, ενώ τα δεδομένα με την ασθένεια. Στο σημείο αυτό θα πρέπει να αναφέρουμε ότι θα εξετάσουμε τον ασθενή ως εξαρτημένη μεταβλητή σε μια λογιστική παλινδρόμηση, με το γονότυπο ως ανεξάρτητη μεταβλητή.

Παρακάτω παρατίθεται ο Πίνακας 1 τροποποιημένος και έτοιμος για τη λογιστική παλινδρόμηση, όπως περιγράφηκε παραπάνω:

Μελέτη ($i=1,2,\dots,k$)	Ασθενείς ($y_{ij}=1$)	Μάρτυρες ($y_{ij}=0$)	Γονότυπος ($j=1,2,\dots,r$)
1	38	72	1
1	52	78	2
1	6	22	3
2	44	38	1
2	27	45	2
2	11	17	3
3	33	21	1
3	34	26	2
3	8	11	3
4	45	53	1
4	53	87	2
4	16	51	3

Πίνακας 3: ο Πίνακας 1, τροποποιημένος για να ξεκινήσει η λογιστική παλινδρόμηση. Η κατανομή των τριών γονοτύπων είναι σε ξεχωριστές λίστες για τους ασθενείς και τους μάρτυρες.

Όπως εύκολα μπορεί να παρατηρήσει κανείς από τον παραπάνω πίνακα, βλέπουμε ότι έχουμε τρεις ανεξάρτητες μεταβλητές ($r=3$), όπου αντιστοιχούν στους γονότυπους AA, AB, BB. Επίσης, χρειαζόμαστε και ένα σύνολο από $r-1$ δείκτες. Για να υπερκεράσουμε το πρόβλημα που αφορά την κανονικότητα των logORs, θα πρέπει να κάνουμε χρήση της δίτιμης μορφής των δεδομένων μας. Με αυτή την προσέγγιση μπορούμε να υπολογίσουμε μέσω των δεδομένων μας και τον αριθμό των μηδενικών κυττάρων της κάθε μελέτης.

Προσεγγίσεις, όπως η προαναφερθείσα, σε μερικές περιπτώσεις έχουν ορισθεί ως μέθοδοι μεμονωμένων ασθενών (IPD = Individual Patients Data), οι οποίες αρχικά χρησιμοποιήθηκαν για μετα-αναλύσεις τυχαιοποιημένων ελεγχόμενων δοκιμών (RCTs = Randomized controlled trial).

Στη συνέχεια υποθέτουμε ένα σχήμα κωδικοποίησης τριών γονοτύπων, όπως φαίνεται παρακάτω:

$$j = \begin{cases} 1, & \text{για τον γονότυπο AA,} \\ 2, & \text{για τον γονότυπο AB,} \\ 3, & \text{για τον γονότυπο BB} \end{cases}$$

όπου το y_{ij} δηλώνει την κατάσταση των ατόμων, δηλαδή αν είναι ασθενής ή μάρτυρας τα οποία έχουν γονότυπο j^{th} στην i μελέτη. Το $\pi_{ij} = P(y_{ij}=1)$, δηλώνει την πιθανότητα το άτομο που εξετάζουμε να έχει τον παράγοντα κινδύνου, το οποίο άτομο έχει j^{th} γονότυπο, στην αντίστοιχη i^{th} μελέτη.

Εφ' όσον θεωρήσαμε εξ' αρχής ότι το B αλληλόμορφο φέρει τον παράγοντα κινδύνου, αυτό θα σημαίνει αυτόματα ότι το αλληλόμορφο που έχει γονότυπο AA είναι μια αλληλουχία αναφοράς, μιας και είναι η υγιής. Επίσης θα πρέπει να δημιουργήσουμε και τις λεγόμενες ψευδο-μεταβλητές (dummy variables), ως $z_{i2}=1$ για το γονότυπο AB και $z_{i3}=1$ για το γονότυπο BB.

Επιπροσθέτως, μπορούμε να διαμορφώσουμε το π_{ij} , κάνοντας χρήση της λογιστικής παλινδρόμησης, ενσωματώνοντας ψευδο-μεταβλητές α_i , οι οποίες είναι δείκτες της μελέτης και αφορούν συγκεκριμένα σταθερά δεδομένα, όπως φαίνεται παρακάτω:

$$\text{logit}(\pi_{ij}) = \alpha_i + \theta_2 z_{i2} + \theta_3 z_{i3} = \alpha_i + \sum_{j=2}^r \theta_j z_{ij} \quad (3.1)$$

Στην παραπάνω εξίσωση μπορούμε να παρατηρήσουμε την ύπαρξη των μεταβλητών θ_2, θ_3 . Οι μεταβλητές αυτές είναι γνωστοί συντελεστές μιας λογιστικής παλινδρόμησης, μέσω της οποίας λαμβάνουμε μια συνολική εκτίμηση για το πηλίκο της πιθανότητας, η οποία σχετίζεται με τον κάθε γονότυπο, όπως παρατηρούμε και στις παρακάτω εξισώσεις:

$$OR_{AB/AA} = \exp(\theta_2) \quad (3.2)$$

και

$$OR_{BB/AA} = \exp(\theta_3) \quad (3.3)$$

Σε αυτό το σημείο θα πρέπει να τονίσουμε ότι τα ORs που υπολογίστηκαν μέσω της παρούσας προσέγγισης είναι συσχετισμένα. Αυτό παράλληλα σημαίνει πως και τα τυπικά σφάλματα θα διαφέρουν από εκείνα τα οποία κάνουν χρήση μιας πιο γενικής προσέγγισης.

2.1.2 Αξιολόγηση της ετερογένειας

Όπως είδαμε παραπάνω και συγκεκριμένα από την εξίσωση 3.1, εκτελέσαμε μια λογιστική παλινδρόμηση στην οποία διατηρήσαμε τη στρωματοποίηση των μελετών μας. Εν συνεχεία, κάνουμε την υπόθεση ότι τα δεδομένα του γονότυπου είναι σταθερά μεταξύ των μελετών. Στο σημείο αυτό πρέπει να επισημάνουμε ότι τα αποτελέσματά μας είναι ομοιογενή. Στη συνέχεια στην παραπάνω εξίσωση (3.1) προσθέτουμε έναν όρο, ο οποίος αφορά την αλληλεπίδραση μεταξύ του αποτελέσματος της μελέτης και του γονότυπου, όπως φαίνεται στην παρακάτω εξίσωση:

$$\text{logit}(\pi_{ij}) = \alpha_i + \sum_{j=2}^r \theta_j z_{ij} + \sum_{i=2}^k \sum_{j=2}^r \gamma_{ij} \alpha_i z_{ij} \quad (3.4)$$

Η εξίσωση που παρουσιάστηκε ακριβώς από πάνω, θα μπορούσε να πει κανείς ότι αντιστοιχεί στον έλεγχο της υπόθεσης ότι η επίδραση του κάθε γονότυπου ποικίλει μεταξύ των μελετών. Η υπόθεση που μόλις κάναμε δύναται να δοκιμαστεί, αν εκτελέσουμε ένα πολύ-μεταβλητό Wald τεστ.

Σε αυτό το σημείο θα ήταν καλό να υπενθυμίσουμε τι είναι το Wald τεστ. Το συγκεκριμένο τεστ είναι ένα παραμετρικό στατιστικό τεστ γενικής χρήσης το οποίο πήρε το όνομά του από τον Αβραάμ Βάλντ. Με το παρών τεστ μπορούμε να εκφράσουμε ως ένα στατιστικό μοντέλο τη σχέση μεταξύ των δεδομένων μας και των αντίστοιχων ατόμων που λαμβάνουν μέρος στη μελέτη μας. Τέλος, θα πρέπει να αναφέρουμε ότι το παρών τεστ μπορούμε να το χρησιμοποιήσουμε, ώστε να ελέγξουμε τις πραγματικές τιμές από τις παραμέτρους που βασίζονται σε εκτίμηση δείγματος.

Συνεχίζοντας τον προηγούμενο συλλογισμό μας, θεωρούμε το \mathbf{b} ότι είναι το διάνυσμα των εκτιμώμενων επιδράσεων. Αυτοί οι συντελεστές, είναι πρώτον το \mathbf{Rb} , το οποίο ισούται με το \mathbf{r} , το οποίο είναι το γραμμικό διάνυσμα του $(r-1)(k-1)$. Ο δεύτερος συντελεστής είναι ο V , μέσω του οποίου υπολογίζεται ο πίνακας της διακύμανσης-συνδιακύμανσης. Μετά από την παραπάνω υπόθεσή μας προκύπτει ο παρακάτω στατιστικός τύπος:

$$W = (\mathbf{Rb} - \mathbf{r})' (\mathbf{RVR}')^{-1} (\mathbf{Rb} - \mathbf{r}) \quad (3.4)$$

Όπου θα έχουμε και μια ασυμπτωτική κατανομή χ^2 .

Όπως είδαμε παραπάνω το τεστ Wald, το οποίο εκτελέσαμε, είχε ως σκοπό να μας δείξει την σημαντικότητα της αλληλεπίδρασης των όρων και παράλληλα αποδεικνύεται ότι είναι απολύτως ανάλογη με το τεστ του χ^2 για την ετερογένεια ή, όπως είναι ευρέως γνωστό, Cochran's Q, για το οποίο χρησιμοποιήθηκε μέθοδος για συνοπτικά δεδομένα.

Σε αυτό το σημείο θα ήταν καλό να θυμίσουμε τι είναι το Cochran's Q. Η στατιστική αυτή ανάλυση είναι ένα μη παραμετρικό στατιστικό τεστ μέσω του οποίου μπορούμε να κατανοήσουμε αν δύο ομάδες είναι ίδιες μεταξύ τους. Επίσης, χρησιμοποιείται ώστε να εκτιμηθούν τα δεδομένα, αν είναι ίδια, όταν αυτά προέρχονται από διαφορετικούς μελετητές. Τέλος, θα ήταν καλό να αναφέρουμε ότι τα αποτελέσματα που παράγονται από το παρών τεστ είναι κωδικοποιημένα με μηδέν (0) και ένα (1), όπου η σημασία του καθενός ισοδυναμεί με αποτυχία και επιτυχία αντίστοιχα.

Μια εναλλακτική στατιστική ανάλυση, πέραν του Cochran's Q τεστ, είναι αυτή το LRT (Likelihood Ratio Test). Το τελευταίο πολλές φορές μας δίνει συγκρίσιμα αποτελέσματα.

2.1.3 Επιτρεπτή ετερογένεια

Η ετερογένεια μέσα στη μετα-ανάλυση είναι ένα σύνηθες φαινόμενο, πολλές φορές όμως πρέπει να μοντελοποιηθεί. Ο κύριος λόγος της προαναφερθείσας μοντελοποίησης είναι η λήψη ενός πρόσθετου συστατικού της ετερογένειας, μέσω του οποίου πρέπει να εκτελεστεί μια λογιστική παλινδρόμηση τυχαίων δεδομένων, έτσι ώστε να επιτρέπεται στα γονοτυπικά δεδομένα να ποικίλουν μεταξύ των μελετών. Ο καλύτερος δυνατός τρόπος για να πετύχουμε το παραπάνω είναι να εισάγουμε σε μια μελέτη έναν εξειδικευμένο τυχαίο συντελεστή, ο οποίος θα αντιπροσωπεύει την απόκλιση της πραγματικής επίδρασης της μελέτης από τη συνολική μέση επίδραση. Έτσι θα έχουμε:

$$\text{logit}(\pi_{ij}) = \alpha_i + (\theta_2 + \nu_{i2})z_{i2} + (\theta_3 + \nu_{i3})z_{i3} = \alpha_i + \sum_{j=2}^r \theta_j z_{ij} + \sum_{j=2}^r \nu_{ij} z_{ij} \quad (3.5)$$

Με τους τυχαίους όρους να κατανέμονται όπως φαίνεται παρακάτω:

$$\begin{pmatrix} \nu_{i2} \\ \nu_{i3} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_2^2 & \sigma_{23} \\ \sigma_{32} & \sigma_3^2 \end{pmatrix} \right) \quad (3.6)$$

Και τη συνδιακύμανση μεταξύ των τυχαίων όρων να είναι ίση με:

$$\text{Cov}(\nu_{i2}, \nu_{i3}) = \sigma_{23} = \sigma_{32} = \rho\sigma_2\sigma_3 \quad (3.7)$$

Από την παραπάνω εξίσωση παρατηρούμε την ύπαρξη του όρου ρ , ο οποίος αφορά τη συσχέτιση των δύο τυχαίων όρων.

Ένα από τα πλέον σημαντικά πλεονεκτήματα της χρήσης των μοντέλων λογιστικής παλινδρόμησης, όπως περιγράφηκαν παραπάνω, είναι ότι μας δίνεται η δυνατότητα να προσθέσουμε κι άλλες μεταβλητές, οι οποίες ως κύριο σκοπό θα έχουν τη διερεύνηση για τυχόν πηγές ετερογένειας. Σε αυτό το σημείο θα ήταν καλό να αναφέρουμε ότι τα μοντέλα της λογιστικής παλινδρόμησης, μπορούν να χρησιμοποιηθούν για μοντέλα είτε σταθερών δεδομένων είτε τυχαίων. Με αυτόν τον τρόπο, το τελικό μοντέλο είναι ισοδύναμο με μια μετα-παλινδρόμηση. Η μετα-παλινδρόμηση χρησιμοποιείται αρκετά συχνά στις μετα-αναλύσεις. Επίσης, είναι σύνηθες φαινόμενο στις γενετικά συσχετισμένες μελέτες, η μεταβλητή την οποία θέλουμε να προσαρμόσουμε αντιστοιχεί σε ένα ομοιογενές επίπεδο μελέτης. Ένα χαρακτηριστικό παράδειγμα είναι όταν μια μεταβλητή- δείκτης η οποία αφορά την εθνικότητα του υποκείμενου πληθυσμού μας, μπορεί να επηρεάσει τα αποτελέσματά μας. Άλλα παρόμοια παραδείγματα θα μπορούσε να είναι η μέση ηλικία, η συστολική πίεση του αίματος, τα συνολικά επίπεδα χοληστερίνης και ούτω καθ' εξής. Από τα παραπάνω μπορεί κανείς να καταλάβει εύκολα ότι μπορούν να ενσωματωθούν περισσότερες από μία μεταβλητές στα παραπάνω μοντέλα, είτε αυτές είναι συνεχείς είτε διακριτές. Όμως, όσο περισσότερες μεταβλητές έχουμε, τόσο αυξάνει και ο κίνδυνος αύξησης ποσοστού σφάλματος Τύπου I.

2.1.4 Δοκιμές γενετικών μοντέλων

Όπως παρατηρήσαμε και στις προηγούμενες ενότητες, το μοντέλο της λογιστικής παλινδρόμησης έχει αρκετά πλεονεκτήματα. Ένα επιπρόσθετο είναι το γεγονός ότι έχουμε τη δυνατότητα να διαμορφώσουμε τις υποθέσεις μας σχετικά με το συγκεκριμένο γενετικό μοντέλο που εξετάζουμε. Ανεξάρτητα με το αν τα δεδομένα που εξετάζουμε είναι σταθερά ή τυχαία, έχουμε τη δυνατότητα να δημιουργήσουμε μια μηδενική υπόθεση η οποία αφορά την ισότητα των $\log ORs$, τα οποία σχετίζονται με τον κάθε γονότυπο μας. Αν υποθέσουμε ότι $\theta_2=0$ και $\theta_3>0$, τότε προτείνεται ένα υπολειπόμενο μοντέλο. Και στις δύο προηγούμενες υποθέσεις μας όπου το θ_2 ισούται με το μηδέν όπως και το θ_3 είναι σημαντικά μεγαλύτερο από το μηδέν, τότε μπορούμε να δοκιμάσουμε τη μηδενική υπόθεση μέσα από την οποία προκύπτει ότι:

$$H_0 : \Delta = \theta_2 - \theta_3 = 0 \quad (3.8)$$

Η παρούσα υπόθεση μπορεί να δοκιμαστεί κάνοντας χρήση ενός Wald τεστ, όπως ακριβώς παρουσιάστηκε παραπάνω, χρησιμοποιώντας χ^2 με ένα (1) βαθμό ελευθερίας. Ένα P-value το οποίο θα ήταν μεγαλύτερο του 0.05, θα ήταν μια σοβαρή ένδειξη σχετικά με την ισότητα των $\log ORs$. Πιο απλά θα μπορούσαμε να αναφέρουμε, ότι αυτό θα ήταν ένδειξη για την ύπαρξη ενός επικρατούς γενετικού μοντέλου. Άρα, λοιπόν στην προκειμένη περίπτωση η παραπάνω εξίσωση (3.8), η οποία αφορά τη μηδενική υπόθεση, απορρίπτεται.

Υπάρχουν όμως μερικές περιπτώσεις στις οποίες οι πληροφορίες σχετικά με το γενετικό μοντέλο διαφέρουν αρκετά μεταξύ των μελετών. Αυτό σημαίνει πως πιθανότατα θα χρειαστούμε ένα επίσημο τεστ, ώστε να αξιολογήσουμε την ομοιογένεια του γενετικού μοντέλου κληρονομικότητας.

Αυτό μπορεί να γίνει με το τεστ της μηδενικής υπόθεσης, όπως φαίνεται παρακάτω:

$$H_0 : \Delta_i = \Delta, \forall i = 1, 2, \dots, k \quad (3.9)$$

Κατ' αυτόν τον τρόπο αποκτούμε τους συντελεστές θ_{i2} και θ_{i3} μαζί με τα αντίστοιχα standard errors ($se_{\theta_{i1}}$ και $se_{\theta_{i3}}$) από την κάθε μελέτη i από τον πίνακα μεταβλητή – διακύμανση V από το μοντέλο της λογιστικής παλινδρόμησης. Έτσι λοιπόν γίνεται πιο εύκολη η λήψη μιας εκτίμησης που σχετίζεται με τη διαφορά μεταξύ των συντελεστών Δ_i και της διακύμανσης $\text{var}(\Delta_i)$, χρησιμοποιώντας τις παρακάτω εξισώσεις:

$$\Delta_i = \theta_{i2} - \theta_{i3} \quad (3.10)$$

και

$$\text{var}(\Delta_i) = \text{var}(\theta_{i2}) + \text{var}(\theta_{i3}) - 2 \text{cov}(\theta_{i2}, \theta_{i3}) \quad (3.11)$$

Άρα, μπορούμε πλέον να εκτελέσουμε ένα στάνταρ τεστ ομοιογένειας, αντίστοιχο με το τεστ του Cochran, χρησιμοποιώντας για βάρη τις αμοιβαίες διακυμάνσεις. Οπότε το στατιστικό τεστ θα είναι αυτό που παρουσιάζεται στην παρακάτω εξίσωση:

$$Q = \sum_i^k \frac{(\Delta_i - \Delta)^2}{\text{var}(\Delta_i)} \quad (3.12)$$

Είναι προφανές, ότι το στατιστικό που θα βρούμε θα πρέπει να το συγκρίνουμε με την κατανομή χ^2 με $k-1$ βαθμούς ελευθερίας. Αυτό σημαίνει πως αν υπάρξει απόρριψη του τεστ, τότε το γενετικό μοντέλο δεν είναι σταθερό μεταξύ των μελετών.

Ολοκληρώνοντας την παρούσα ενότητα, πρέπει να αναφέρουμε ότι αν χρησιμοποιούσαμε ως ομάδα αναφοράς τον γονότυπο BB, θα αποκτούσαμε εκτιμήσεις για τα ORs λιγότερα από ένα. Οπότε τα αποτελέσματα που θα λαμβάναμε, θα ήταν κυρίως αποτελέσματα τα οποία δε θα ήταν ευαίσθητα σε ό,τι αφορά την επιλογή του ομάδα αναφοράς. Επιπλέον, με την προτεινόμενη μεθοδολογία, θα μπορούσαμε να λάβουμε αξιόπιστα αποτελέσματα ακόμη και σε σπάνιες περιπτώσεις, όπου το γενετικό μοντέλο είναι υπερβολικά επικρατές.

2.1.5 Γενικό πολύ-μεταβλητό μοντέλο μετα-ανάλυσης για διακριτά δεδομένα

Κατά τη γενική προσέγγιση για τη διμεταβλητή μετα-ανάλυση, τα δύο logORs προέρχονται από το μεταλλαγμένο αλληλόμορφο το οποίο θα μπορούσε να μοντελοποιηθεί ταυτόχρονα ως μια διμεταβλητή απόκριση. Έτσι ο λογάριθμος του OR_{AB} , όπου το συγκεκριμένο OR αντιστοιχεί στα ετερόζυγα έναντι των ομόζυγων, δίνεται από την παρακάτω εξίσωση:

$$y_{li} = \log \left(\frac{AB_{li} \cdot AA_{0i}}{AA_{li} \cdot AB_{0i}} \right) \quad (3.13)$$

Και κατά προσέγγιση η διακύμανση υπολογίζεται από την παρακάτω εξίσωση:

$$s_{1i}^2 = 1/AA_{1i} + 1/AA_{0i} + 1/AB_{1i} + 1/AB_{0i} \quad (3.14)$$

Με το ίδιο σκεπτικό, ο λογάριθμος του OR_{BB} , όπου το συγκεκριμένο OR αντιστοιχεί στο ομόζυγο για το μεταλλαγμένο αλληλόμορφο έναντι του ομόζυγου, και δίνεται από την παρακάτω εξίσωση μαζί με την αντίστοιχη διακύμανση:

$$y_{2i} = \log \left(\frac{BB_{1i}AA_{0i}}{AA_{1i}BB_{0i}} \right) \quad (3.15)$$

$$s_{2i}^2 = 1/AA_{1i} + 1/AA_{0i} + 1/BB_{1i} + 1/BB_{0i} \quad (3.16)$$

Αν υποθέσουμε τώρα, ότι τα δύο $\log ORs$ είναι κατανομημένα μετά από μια διμεταβλητή κανονική κατανομή, τότε θα ισχύει η παρακάτω εξίσωση:

$$\begin{bmatrix} y_{1i} \\ y_{2i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \beta_{1i} \\ \beta_{2i} \end{bmatrix}, \begin{bmatrix} s_{1i}^2 & \rho_{w12}s_{1i}s_{2i} \\ \rho_{w12}s_{1i}s_{2i} & s_{2i}^2 \end{bmatrix} \right\} \quad (3.17)$$

Με τις αντίστοιχες μέσες τιμές $\beta_{1i}\beta_{2i}$, οι οποίες θεωρούνται τυχαίες, κατανέμονται με τον ίδιο τρόπο, όπως φαίνεται στην παρακάτω εξίσωση:

$$\begin{bmatrix} \beta_{1i} \\ \beta_{2i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \begin{bmatrix} \tau_1^2 & \rho_{B12}\tau_1\tau_2 \\ \rho_{B12}\tau_1\tau_2 & \tau_2^2 \end{bmatrix} \right\} \quad (3.18)$$

Έτσι, το τελικό μοντέλο από το οποίο εξαγάγουμε το συμπέρασμά μας φαίνεται στην παρακάτω εξίσωση:

$$\begin{bmatrix} y_{1i} \\ y_{2i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \begin{bmatrix} s_{1i}^2 + \tau_1^2 & \rho_{w12}s_{1i}s_{2i} + \rho_{B12}\tau_1\tau_2 \\ \rho_{w12}s_{1i}s_{2i} + \rho_{B12}\tau_1\tau_2 & s_{2i}^2 + \tau_2^2 \end{bmatrix} \right\} \quad (3.19)$$

Στις παραπάνω εξισώσεις παρατηρούμε την ύπαρξη αρχικά του όρου $\rho_{w12}s_{1i}s_{2i}$, ο οποίος αντιστοιχεί στη συνδιακύμανση μεταξύ των μελετών, όπου ξεκινάει με το ρ_{w12} με τις συσχετισμένες μελέτες, οι οποίες είναι γνωστές από πριν. Εν συνεχεία συναντάμε τον όρο $\rho_{B12}\tau_1\tau_2$ ο οποίος αντιστοιχεί στη συνδιακύμανση μεταξύ των μελετών με τους τυχαίους όρους, που πρόκειται να εκτιμηθεί από τα δεδομένα, συνεπώς το ρ_{B12} αντιστοιχεί στη συσχέτιση μεταξύ των μελετών. Στο σημείο αυτό θα πρέπει να αναφέρουμε ότι στα γενικά μοντέλα διμερους μετα-ανάλυσης, η συσχέτιση μεταξύ των μελετών συνήθως θεωρείται μηδέν ή διαφορετικά θα πρέπει να δίδεται από ξεχωριστές μελέτες. Αντιθέτως, υπάρχει συσχέτιση στα δύο $\log ORs$, όταν και οι αντίστοιχες μελέτες είναι και αυτές γενετικά συσχετισμένες. Ο λόγος είναι προφανής μιας και τα δύο $\log ORs$ κάνουν μια σύγκριση έναντι της ίδιας βασικής κατηγορίας, καθώς και η συσχέτιση μεταξύ των μελετών μπορεί να υπολογιστεί αναλυτικά, κάνοντας χρήση τους αριθμούς των γονότυπων από τις μεμονωμένες μελέτες. Κατά συνέπεια, η

συνδιακύμανση μεταξύ των δύο logORs, προκύπτει από την επεξεργασία των παρατηρούμενων μετρήσεων, οι οποίες αντιστοιχούν σε ανεξάρτητες μεταβλητές Poisson με $E[Y_i] = \text{var}[Y_i] = Y_i$ και τα logORs ως αντίθεση μεταξύ των log μετρήσεων.

Παρακάτω, υπολογίζουμε τη συνδιακύμανση κάνοντας χρήση της μεθόδου δέλτα:

$$\text{cov}(y_{1i}, y_{2i}) = 1/AA_{0i} + 1/AA_{1i} \quad (3.20)$$

Και η συσχέτιση δίνεται από την παρακάτω εξίσωση:

$$\rho_{w12} = (1/AA_{0i} + 1/AA_{1i})/s_{1i}s_{2i} \quad (3.21)$$

Σε αυτό το σημείο πρέπει να αναφέρουμε ότι τα αποτελέσματά μας είναι αρκετά διαισθητικά. Η χρήση μιας παρόμοιας μεθόδου χρησιμοποιείται για τη μετα-ανάλυση επιδημιολογικών δεδομένων, χωρίς όμως να μας δίνουν λεπτομέρειες για την προέλευσή τους.

Παρακάτω παραθέτουμε ένα τμήμα του αντίστοιχου αλγορίθμου, που είναι υλοποιημένος στο λογισμικό πακέτο του Stata:

```
...
** calculation of the log(Odds Ratios)
gen b1=log(( ab1/ aa1)/( ab0/aa0))
gen b2=log(( bb1/ aa1)/( bb0/aa0))
** calculation of the variances
gen V11=1/aa0 +1/ab0+1/aa1 +1/ab1
gen V22=1/aa0 +1/aa1+1/bb0 +1/bb1
** calculation of the covariance
gen V12=1/aa0 +1/aa1
...
```

Πίνακας 4 Τμήμα κώδικα στο Stata που αντιστοιχεί στις τελευταίες εξισώσεις

2.1.6 Γενικευμένο πολυμεταβλητό μοντέλο μετα-ανάλυσης για συνεχή δεδομένα

Αν τα αποτελέσματά μας είναι συνεχή, όπως αυτά παρουσιάζονται στον παρακάτω πίνακα, τότε από κάθε μελέτη θα πάρουμε τη μέση τιμή και την τυπική απόκλιση για τα δεδομένα του κάθε γονότυπου.

Μελέτες	Γονότυποι					
	AA		AB		BB	
	mean(sd)	N	mean(sd)	N	mean(sd)	N
1	$y_{AA1}(s_{AA1})$	AA_1	$y_{AB1}(s_{AB1})$	AB_1	$y_{BB1}(s_{BB1})$	BB_1
2	$y_{AA2}(s_{AA2})$	AA_2	$y_{AB2}(s_{AB2})$	AB_2	$y_{BB2}(s_{BB2})$	BB_2
...
k	$y_{AAk}(s_{AAk})$	AA_k	$y_{ABk}(s_{ABk})$	AB_k	$y_{BBk}(s_{BBk})$	BB_k

Πίνακας 5: Δεδομένα που χρησιμοποιούνται στη μετα-ανάλυση των γενετικών μελετών με συνεχή δεδομένα

Στον παραπάνω πίνακα παρατηρούμε τη διάταξη των δεδομένων που χρησιμοποιήθηκαν σε μια μετα-ανάλυση γενετικών μελετών, που έχουν συνεχή δεδομένα ή φαινότυπους. Η κατανομή των διάφορων γονοτύπων παρατίθεται για όλους τους συμμετέχοντες στη μετα-ανάλυση (N), οι

οποίοι συνοδεύονται από το μέσο όρο και την τυπική απόκλιση από κάθε ομάδα.

Συνεχίζοντας τον αρχικό μας συλλογισμό, δηλώνουμε ότι το y_{ij} είναι το αποτέλεσμα ενός ατόμου από μια μελέτη, όπου το άτομο αντιστοιχεί στο j και η μελέτη στο i . Επίσης, χρησιμοποιούμε εικονικές μεταβλητές όπως οι z_{1ij} και z_{2ij} για τα άτομα τα οποία έχουν γονότυπο AB και BB αντίστοιχα. Κατ' αυτόν τον τρόπο μπορούμε να σχεδιάσουμε ένα γραμμικό μοντέλο, όπως αυτό παρουσιάζεται από την παρακάτω εξίσωση:

$$y_{ij} = \alpha_0 + \alpha_i + \beta_1 z_{1ij} + \beta_2 z_{2ij} \quad (3.22)$$

Αν τα δεδομένα μας είναι της μορφής του Πίνακα 5, τότε οι παρατηρήσεις μας είναι οι μελέτες του πίνακα. Αυτό επίσης σημαίνει ότι οι μελέτες πρέπει να έχουν βάρος, που θα αντιστοιχεί στον αντίστροφο της διακύμανσής τους. Σε περίπτωση που έχουμε πρόσβαση σε ατομικά δεδομένα ασθενών, αντίθετα από την προηγούμενη περίπτωση, τα χρησιμοποιούμε απευθείας στο γραμμικό μοντέλο μας. Στο μοντέλο που παρουσιάστηκε στην εξίσωση (3.22) ο όρος α_0 , αντιστοιχεί στη συνολική μέση τιμή που σχετίζεται με το γονότυπο AA, ο οποίος είναι το ομάδα αναφοράς μας. Επίσης, ο όρος α_i , αντιστοιχεί στα συγκεκριμένα δεδομένα της μελέτης. Τέλος, οι συντελεστές β_1 και β_2 αντιστοιχούν στις μέσες διαφορές που έχουν τα άτομα με τον γονότυπο AB και BB από τον AA αντίστοιχα. Μια παραλλαγή της παραπάνω εξίσωσης φαίνεται παρακάτω με τη χρήση δύο νέων συντελεστών β_{1i} και β_{2i} :

$$y_{ij} = \beta_0 + \beta_i + (\beta_{1i} + \beta_1) z_{1ij} + (\beta_{2i} + \beta_2) z_{2ij} \quad (3.23)$$

Με τους τυχαίους συντελεστές να θεωρούνται κανονικά κατανομημένοι, όπως φαίνεται παρακάτω:

$$\begin{bmatrix} \beta_{1i} \\ \beta_{2i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_2\tau_1 & \tau_2^2 \end{bmatrix} \right\} \quad (3.23)$$

Το συγκεκριμένο μοντέλο μπορεί να τοποθετηθεί σε κάθε στατιστικό πακέτο, όπως το Stata, κάνοντας χρήση της εντολής `gllamm`. Παρακάτω, μπορούμε να δούμε ένα τμήμα του κώδικα στο Stata με τη χρήση της αντίστοιχης εντολής:

```
...
**data re-arrangement in a format suitable for fitting
**the linear mixed model
rename gen11 gen1
rename gen12 gen2
rename gen22 gen3
rename x11 x1
rename sd11 sd1
rename x12 x2
rename sd12 sd2
rename x22 x3
rename sd22 sd3
reshape long gen x sd, i(study) j(z)
```

```

xi i.study i.z
**calculation of the variance
gen v=sd^2/ gen
**gllamm for stability parameterizes the residual
**variance using the logarithm of the standard error
gen s=log(sqrt(v))
**setting up one equation for each study
eq wgt: s
**creating constraints needed for gllamm in order to
**perform inverse-variance weighting
constraint define 1 [lns1]s=1
** fitting the model of Equation 3.10
eq s2: _Iz_2
eq s3: _Iz_3
gllamm x _Iz_2 _Iz_3 _Istudy* , i(study) nrf(2) eqs(s2
s3) s(wgt) constraint(1) nip(8) noest
matrix A=M_initf
gllamm x _Iz_2 _Iz_3 _Istudy* , i(study) nrf(2) eqs(s2
s3) s(wgt) constraint(1) nip(8) from(A) long adapt
...

```

Πίνακας 6 Τμήμα κώδικα υλοποίησης του γραμμικού μοντέλου με τη χρήση της εντολής gllamm

Ωστόσο, μπορούμε να διαφοροποιήσουμε την παραπάνω προσέγγιση, ώστε να πραγματοποιείται άμεσα μοντελοποίηση των διαφορών των δεδομένων μεταξύ των ζευγών, τα οποία προκύπτουν από τα άτομα που έχουν τους γονότυπους AB και BB από το γονότυπο αναφοράς AA. Αυτή η προσέγγιση φαίνεται καλύτερα από τις παρακάτω σχέσεις:

$$y_{1i} = y_{ABi} - y_{AAi} \text{ και } y_{2i} = y_{BBi} - y_{AAi}$$

Ως εκ τούτου, αποδεικνύουμε ότι οι παραπάνω διαφορές είναι κανονικά κατανοημένες, με τη μέση τιμή να ισούται με τη διαφορά των μέσων τιμών και τη διακύμανση ίση με το άθροισμα των διακυμάνσεων. Γίνεται λοιπόν προφανές ότι το διμεταβλητό μοντέλο της εξίσωσης 3.19, μπορεί να εφαρμοστεί άμεσα, ώστε να γίνει διμεταβλητή μοντελοποίηση των δύο διαφορών. Επίσης, οι δύο διαφορές συσχετίζονται, αφού και οι δύο υπολογίζουν τη διαφορά από τον γονότυπο AA (το ομάδα αναφοράς). Τέλος, πρέπει να αναφέρουμε ότι οι μεταβλητές y_{AAi} , y_{ABi} και y_{BBi} είναι αμοιβαία ανεξάρτητες και η συνδιακύμανσή υπολογίζεται από την παρακάτω εξίσωση:

$$\begin{aligned}
\text{cov}(y_{1i}, y_{2i}) &= \text{cov}(y_{ABi} - y_{AAi}, y_{BBi} - y_{AAi}) \\
&= \text{cov}(y_{ABi}, y_{BBi}) - \text{cov}(y_{ABi}, y_{AAi}) \\
&\quad - \text{cov}(y_{AAi}, y_{BBi}) + \text{cov}(y_{AAi}, y_{AAi}) \\
&= \text{var}(y_{AAi}) = \frac{sd_{AAi}^2}{AA_i}
\end{aligned} \tag{3.24}$$

Ολοκληρώνοντας την παρούσα ενότητα, πρέπει να αναφέρουμε ότι είτε έχουμε διακριτά δεδομένα, τα οποία προέρχονται από μελέτες ασθενών - μαρτύρων, είτε έχουμε συνεχή δεδομένα, τότε το γενικό μοντέλο που το περιγράφει είναι αυτό της εξίσωσης 3.19, όπως αυτό περιγράφηκε παραπάνω. Επίσης, το προαναφερθέν μοντέλο μπορεί να χρησιμοποιηθεί για να πάρουμε

ταυτόχρονα συμπεράσματα σχετικά με τη συσχέτιση των δεδομένων. Τέλος, να επισημάνουμε ότι το μοντέλο αυτό μπορεί να εισαχθεί σε οποιοδήποτε στατιστικό πακέτο επιθυμούμε. Χαρακτηριστικά, στην εφαρμογή του Stata, μπορεί να εισαχθεί κάνοντας χρήση της εντολής `mvmeta`. Σε αυτό το σημείο θα ήταν καλό να υπενθυμίσουμε ότι η εντολή `mvmeta`, εξάγει συμπεράσματα έχοντας ως βάση είτε τις ροπές της μέγιστης πιθανοφάνειας (ML), είτε την περιορισμένη μέγιστη πιθανοφάνεια (REML).

2.1.7 Προσέγγιση χωρίς γονιδιακό μοντέλο

Η προσέγγιση αυτή, θα μπορούσε να πει κανείς, πως είναι μια επέκταση του γενικού διμεταβλητού μοντέλου, κάνοντας χρήση συνοπτικών δεδομένων. Αυτή η προσέγγιση αποτελείται από μοντελοποίηση του β_2 και του λ , όπου το λ αντιστοιχεί στο πηλίκο των β_1 και β_2 . Έτσι, η εξίσωση που το περιγράφει είναι η παρακάτω:

$$\begin{bmatrix} y_{1i} \\ y_{2i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \lambda\beta_2 \\ \beta_2 \end{bmatrix}, \begin{bmatrix} s_{1i}^2 + \lambda^2\tau^2 & \rho_{w12}s_{1i}s_{2i} + \lambda\tau^2 \\ \rho_{w12}s_{1i}s_{2i} + \lambda\tau^2 & s_{2i}^2 + \tau^2 \end{bmatrix} \right\} \quad (3.25)$$

Μπορεί εύκολα να εντοπίσει κανείς ότι η παραπάνω εξίσωση είναι όμοια με την 3.19 που αναφέραμε και παραπάνω, με τη μόνη διαφορά την επανα-παραμετροποίηση των β_1 και β_2 και τη χρήση του $\lambda = \beta_1/\beta_2$. Όμως, το συγκεκριμένο μοντέλο επιβάλλει μια διακύμανση μεταξύ των μελετών τ^2 , υποδηλώνοντας ότι το β_s μοιράζεται ένα κοινό στοιχείο της διακύμανσης, ενώ το λ αντιμετωπίζεται ως παράμετρος σταθερών δεδομένων.

2.2. Πολυμεταβλητή Μετα-Ανάλυση με τη χρήση δίτιμων δεδομένων

Μια ενδιαφέρουσα προσέγγιση είναι επίσης αυτή που θα παρουσιαστεί σε αυτή την ενότητα. Η παρούσα προσέγγιση αφορά πολυμεταβλητή μετα-ανάλυση, αλλά χρησιμοποιώντας τη δίτιμη μορφή των δεδομένων που έχουμε. Η χρήση της δίτιμης φύσης των δεδομένων, σημαίνει ότι δε θα υπολογίσουμε τα αντίστοιχα logORs, όπως επίσης δε θα κάνουμε την υπόθεση ότι τα δεδομένα μας είναι κανονικά κατανομημένα. Αυτό που θα κάνουμε είναι να χρησιμοποιήσουμε απευθείας τους γονότυπους και να διενεργήσουμε μετα-ανάλυση κάνοντας χρήση της λογιστικής πλαινδρόμησης. Σε αυτό το σημείο θα ήταν καλό να αναφέρουμε ότι τα συγκεκριμένα μοντέλα, τα οποία θα δούμε, θα μπορούσαμε να τα θεωρήσουμε ως επεκτάσεις των παραπάνω μοντέλων. Με το ίδιο σκεπτικό, μπορούν να θεωρηθούν επίσης ως επεκτάσεις της γενικής πολυμεταβλητής προσέγγισης, με τη μόνη διαφορά ότι τα δύο logORs υπολογίζονται από μια σύγκριση με την ίδια βασική κατηγορία.

Στη δίτιμη περίπτωση, έχουμε αναδιάταξη των δεδομένων και οι τρεις γονότυποι (AA, AB, BB) κωδικοποιούνται σε j , με $j=0, 1, 2$. Επίσης, χρησιμοποιούμε το δ_{ij} , ώστε να υποδηλώσουμε τις καταστάσεις των ασθενών - μαρτύρων, με ασθενείς=0 και μάρτυρες=1, για τα άτομα τα οποία έχουν j^{th} γονότυπο στη μελέτη i ($i=1,2,3,\dots,k$). Στη συνέχεια, σε ό,τι αφορά το σχεδιασμό του συγκεκριμένου μοντέλου, έχουμε δύο εναλλακτικές περιπτώσεις. Η πρώτη περίπτωση, προϋποθέτει μια διμεταβλητή δειγματοληψία, όπου οι δειγματοληπτικοί δείκτες καθορίζονται από τους σταθερούς αριθμούς των ασθενών - μαρτύρων. Ο δεύτερος τρόπος σχεδιασμού, υποθέτει μια πολυωνυμική δειγματοληψία, επιλέγοντας άτομα που εξαρτώνται από την κατάσταση της ασθένειάς τους και στη συνέχεια επαληθεύεται η κατάσταση της έκθεσής τους.

Χρησιμοποιώντας την προοπτική πιθανοφάνεια (prospective likelihood), η οποία βασίζεται στην προοπτική πιθανοφάνεια εμφάνισης της αντίστοιχης ασθένειας, η κατάσταση ασθενείς - μάρτυρες είναι η εξαρτημένη μεταβλητή και οι γονότυποι αντιμετωπίζονται ως μεταβλητές.

Έπειτα, δηλώνουμε τον παράγοντα κινδύνου, ο οποίος στην περίπτωση μας θα είναι το $\pi_{ij} = P(\delta_{ij}=1)$, το οποίο στην προκειμένη περίπτωση αντιστοιχεί στην περίπτωση που ο παράγοντας κινδύνου είναι ασθενής, που αντιστοιχεί στο άτομο που έχει j^{th} γονότυπο στην i^{th} μελέτη. Όπως, έχουμε ήδη αναφέρει, το B αλληλόμορφο αντιστοιχεί στην αλληλουχία η οποία θεωρείται παράγοντας κινδύνου, αυτό σημαίνει ότι ο γονότυπος AA θα οριστεί ως κατηγορία αναφοράς και θα δημιουργηθούν εικονικές μεταβλητές όπως η $z_{1i}=1$ για τους γονότυπους AB και $z_{2i}=1$ για τους γονότυπους BB. Παρακάτω, φαίνεται η αντίστοιχη εξίσωση που περιγράφει το παραπάνω μοντέλο:

$$\begin{aligned} \text{logit}(\pi_{ij}) &= \text{logit}\left[P(\delta_{ij}=1|j)\right] \\ &= \alpha_0 + \alpha_i + (\beta_1 + \beta_{1i})z_{1i} + (\beta_2 + \beta_{2i})z_{2i} \end{aligned} \quad (4.1)$$

Η κατανομή των τυχαίων όρων από την παραπάνω εξίσωση (4.1) παρατίθεται στην παρακάτω εξίσωση:

$$\begin{pmatrix} \beta_{1i} \\ \beta_{2i} \end{pmatrix} \sim MVN \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_2\tau_1 & \tau_2^2 \end{pmatrix} \right) \quad (4.2)$$

Στην εξίσωση (4.1) παρατηρούμε τον όρο α_i . Αυτός ο όρος είναι δείκτης ο οποίος αφορά τη μελέτη συγκεκριμένων σταθερών δεδομένων, ενώ τα β_1 και β_2 είναι οι εκτιμήσεις των logORs του AB με το AA και του BB με το AA αντίστοιχα.

Μια όμοια με αυτή προσέγγιση είναι αυτή που εκ των προτέρων υποθέτει ένα μοντέλο, το οποίο διατρέχει τους γονότυπους. Θα μπορούσε κανείς να αναφέρει ότι η παρούσα προσέγγιση αφορά μια ανάλυση σύμφωνα με το επικρατές μοντέλο. Άρα, το συγκεκριμένο μοντέλο θα συμπεριλαμβάνεται ως πρόβλεψη μιας μεταβλητής z η οποία θα παίρνει τιμές 0, 1 και 2 για τους γονότυπους AA, AB και BB αντίστοιχα. Παρακάτω φαίνεται η εξίσωση που αντιστοιχεί στο παραπάνω περιγραφόμενο μοντέλο:

$$\text{logit}(\pi_{ij}) = \text{logit}\left[P(\delta_{ij}=1|j)\right] = \alpha_0 + \alpha_i + (\beta + \beta_i)z_i \quad (4.3)$$

Το παραπάνω μοντέλο μας δίνει αξιόπιστα αποτελέσματα, αλλά μόνο στην περίπτωση της κυριαρχικής κληρονομικότητας. Πρέπει να αναφέρουμε ότι το συγκεκριμένο μοντέλο μπορεί να παραμετροποιηθεί θεωρώντας ένα σύστημα δειγματοληψίας το οποίο χρησιμοποιεί την αναδρομική πιθανοφάνεια (retrospective likelihood). Σε αυτή την περίπτωση έχουμε τους γονότυπους να αντιμετωπίζονται ως εξαρτημένες μεταβλητές και οι καταστάσεις των ασθενών - μαρτύρων ως μέσο πρόβλεψης σε μια πολυωνυμική λογιστική παλινδρόμηση.

Παρακάτω, βλέπουμε τη μαθηματική εξίσωση του προαναφερθέντος μοντέλου:

$$p_{ij} = P(j|\delta_{ij}) = \frac{\exp(\alpha_0 + \alpha_i + \beta_j\delta_{ij})}{\sum_{r=0}^j \exp(\alpha_0 + \alpha_i + \beta_r\delta_{ir})} \quad (4.4)$$

Επίσης, έχουμε ότι ο γραμμικός προγνωστικός παράγοντας της παραπάνω εξίσωσης είναι:

$$U_{ij} = \log\left(\frac{P_{ij}}{P_{i1}}\right) = \alpha_0 + \alpha_i + \beta_j \delta_{ij}, j = 1, 2 \quad (4.5)$$

Από την παραπάνω εξίσωση παρατηρούμε ότι τα β_1 και β_2 είναι οι εκτιμήσεις των logORs. Το $\beta_0 = 0$ για το γονότυπο $j = 0$ και χρησιμοποιείται ως κατηγορία αναφοράς μιας και αναφέρεται στο γονότυπο AA.

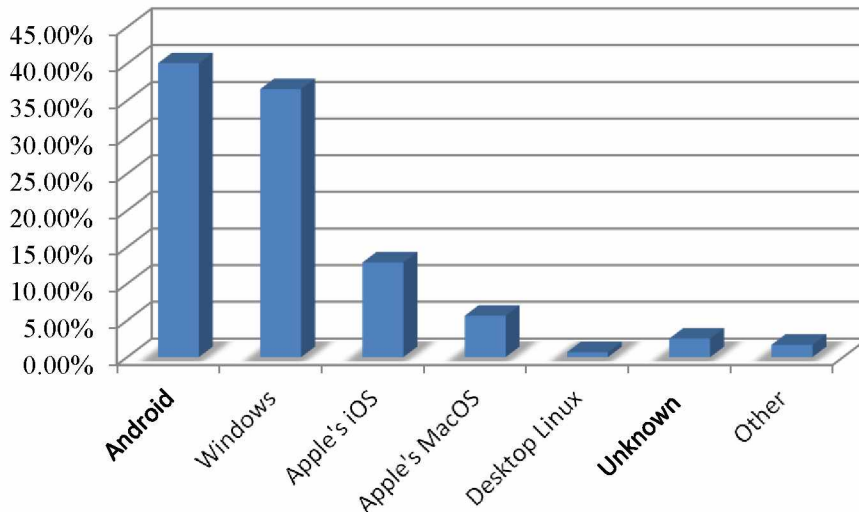
Η αναδρομική πιθανοφάνεια, θα μπορούσαμε να σημειώσουμε ότι αντιστοιχεί στην εξίσωση (4.3). Πρέπει επίσης να αναφέρουμε ότι οι εκτιμήσεις για την εκτίμηση των ροών μέγιστης πιθανοφάνειας (Maximum Likelihood), που προέρχεται από την αναδρομική πιθανότητα, είναι ίδια με εκείνες που προκύπτουν από την προοπτική πιθανοφάνεια.

2.3. Η εφαρμογή και τα μέσα τα οποία χρειάστηκαν

Η παρούσα εφαρμογή είχε ως κύριο σκοπό να διενεργήσει μετα-αναλύσεις «κάπου στο διαδίκτυο», χωρίς να απαιτεί από το χρήστη τη χρησιμοποίηση κάποιου εξειδικευμένου λογισμικού. Ολόκληρη η εφαρμογή βασίστηκε στην ιδέα μιας διαδικτυακής εφαρμογής που θα παράγει αποτελέσματα μετα-ανάλυσης με τελείως μηδενικό κόστος για τον εκάστοτε χρήστη – ερευνητή. Για την υλοποίηση της παρούσας εφαρμογής χρειάστηκαν ορισμένες άλλες εφαρμογές, όπως θα δούμε παρακάτω. Προτού αναφέρουμε σχετικά με τις εφαρμογές που χρειάστηκαν, θα ήταν ενδιαφέρον να δούμε κάτι ιδιαίτερα σημαντικό που αφορά το υπολογιστικό μας σύστημα και αυτό δεν είναι άλλο από την «καρδιά» του, η οποία είναι το λειτουργικό σύστημα.

Αρχικά, θα ήταν καλό να κατανοήσουμε τι είναι ένα λειτουργικό σύστημα. Το λειτουργικό σύστημα (ΛΣ) ή αλλιώς Operating System (OS) είναι αυτό που φέρνει σε επαφή το υλικό του υπολογιστή μας (hardware) με το λογισμικό (software) και κατ' επέκταση και με τον χρήστη. Τα λειτουργικά συστήματα θα μπορούσαμε να τα προσδιορίσουμε σε δύο μεγάλες κατηγορίες. Αυτές είναι: Windows και Unix. Αυτές οι δύο κατηγορίες εμπεριέχουν αρκετές μορφές και εκδόσεις λειτουργικών συστημάτων. Η πρώτη κατηγορία που αφορά τα λειτουργικά συστήματα των Windows, εμπεριέχει εκδόσεις όπως: Windows XP, Windows Vista, Windows 7, Windows 8 μέχρι και Windows 10 που είναι η τελευταία έκδοση. Αντίστοιχα, τα Unix συστήματα έχουν μια πληθώρα από διαφορετικά λειτουργικά τα οποία βασίζονται στη λογική των Unix. Στην κατηγορία των Unix συστημάτων ανήκουν όλες οι εκδόσεις των Linux (π.χ.: Ubuntu, OpenSuse, RedHat, κτλ), όπως επίσης και όλες οι εκδόσεις των MacOS (π.χ.: Snow Leopard, Sierra κτλ). Ένα ιδιαίτερα σημαντικό χαρακτηριστικό των Linux εκδόσεων, είναι η ελεύθερη διανομή τους, ένα χαρακτηριστικό που δεν παρέχετε από κανένα άλλο.

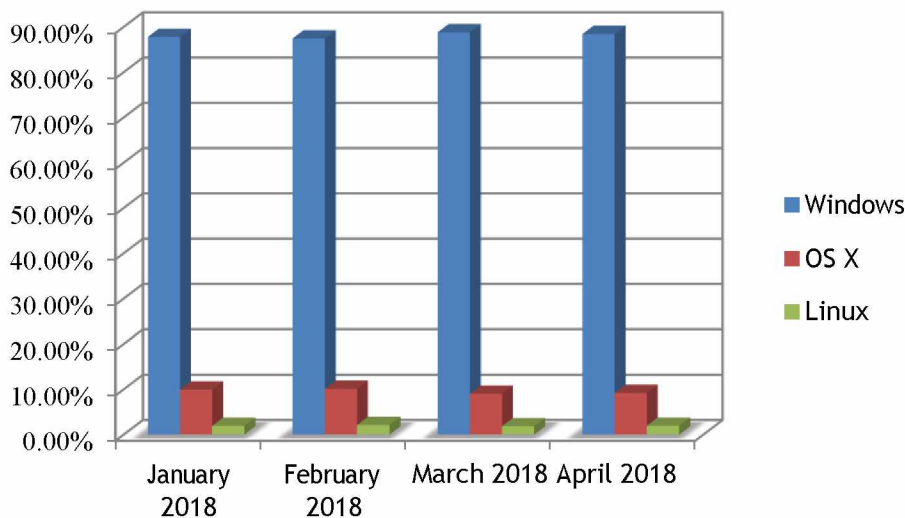
Στο παρακάτω, γράφημα παρατηρούμε το επί τις εκατό ποσοστό των χρηστών με τα αντίστοιχα λειτουργικά συστήματα:



Εικόνα 1: Στατιστικά για τη χρήση των αντίστοιχων λειτουργικών συστημάτων οικιακής χρήσης

Από το παραπάνω γράφημα, μπορεί εύκολα να παρατηρήσει κανείς ότι οι οικιακοί χρήστες προτιμούν περισσότερο λειτουργικά τα οποία είναι πιο φιλικά προς τους ίδιους. Σήμερα, ακόμη και τα «παραδοσιακά» όχι τόσο φιλικά προς το χρήστη λειτουργικά συστήματα, έχουν αλλάξει άρδην φυσικά όμως τα ποσοστιαία αποτελέσματα δεν μπορούν να αλλάξουν τόσο γρήγορα μιας και ακόμη είναι πιο ιδιαίτερα στη χρήση τους έναντι των παραδοσιακών Windows ή των Mac OS.

Στο παρακάτω γράφημα παρατηρούμε τα ποσοστά τα οποία αφορούν εταιρίες – οργανισμούς, τα οποία κάνουν χρήση των παραπάνω λειτουργικών συστημάτων.



Εικόνα 2: Χρήση των αντίστοιχων λειτουργικών συστημάτων στο πρώτο τετράμηνο του 2018

Στην παρούσα εργασία έγινε χρήση του Linux λειτουργικού συστήματος και πιο συγκεκριμένα της έκδοσης του Ubuntu 16.07.

Αφού ολοκληρώσαμε την αναφορά μας σχετικά με τι είναι το λειτουργικό σύστημα, αλλά και το ποιά χρησιμοποιήσαμε, τώρα θα δούμε τις κυριότερες εφαρμογές (software) που χρησιμοποιήσαμε.

Η πρώτη εφαρμογή ήταν το Netbeans. Η συγκεκριμένη εφαρμογή μας δίνει τη δυνατότητα να γράφουμε, να ελέγχουμε τον κώδικά μας για τυχόν σφάλματα και τέλος να τον εκτελούμε. Η συγκεκριμένη εφαρμογή είναι ιδιαίτερα χρήσιμη, μιας και μας δίνει τη δυνατότητα να γράφουμε κώδικα σε αρκετές γλώσσες προγραμματισμού.

Η επόμενη εφαρμογή η οποία ήταν πολύ σημαντική μιας και μέσω αυτής έχουμε τη δυνατότητα να «ανεβάσουμε» την εργασία μας στο διαδίκτυο. Αυτή η εφαρμογή δεν είναι άλλη από τον Nginx. Το Nginx, είναι από τους πιο δημοφιλείς servers μιας και μπορεί να χρησιμοποιηθεί από όλα τα παραπάνω λειτουργικά συστήματα. Επίσης, ο Nginx μπορεί να συνδυαστεί και με βάσεις δεδομένων, όπως η MySQL.

Τέλος, η τελευταία εφαρμογή που χρησιμοποιήθηκε για την περάτωση της παρούσης εργασίας ήταν το στατιστικό λογισμικό πακέτο του Stata. Το συγκεκριμένο πακέτο ήταν καίριας σημασίας, μιας και οι υπολογισμοί και γενικότερα οι αλγόριθμοι που χρησιμοποιήθηκαν για την υλοποίηση των μετα-αναλύσεων έγιναν μέσω αυτού. Το Stata δημιουργήθηκε το 1985 και χρησιμοποιήθηκε κυρίως από οικονομολόγους, κοινωνιολόγους, επιδημιολόγους και βιοϊατρούς. Μέσω του συγκεκριμένου πακέτου μπορούμε να διαχειριστούμε δεδομένα, να κάνουμε προσομοιώσεις, να δημιουργήσουμε γραφήματα και φυσικά να διενεργούμε στατιστικές αναλύσεις. Μέσω του Stata έχουμε τη δυνατότητα να γράψουμε τα δικά μας προγράμματα, με την αντίστοιχη γλώσσα προγραμματισμού που χρησιμοποιεί το Stata. Είναι ένα πολύ σημαντικό πλεονέκτημα, αφού μας δίνεται η δυνατότητα να διαμορφώσουμε για παράδειγμα μια μαθηματική-στατιστική εξίσωση, όπως εμείς επιθυμούμε. Σε αυτό το σημείο είναι καλό να αναφέρουμε, ότι υπάρχουν τέσσερις διαφορετικές εκδόσεις της παραπάνω εφαρμογής.

Χαρακτηριστικά αυτές είναι:

1. Stata MP,
2. Stata SE,
3. Stata IC και
4. Small Stata

Οι διαφορές μεταξύ των διαφορετικών αυτών εκδόσεων, αφορούν περισσότερο τεχνικά ζητήματα, όπως για παράδειγμα αν επιθυμούμε οι αναλύσεις μας να γίνονται με τη χρήση πολλών πυρήνων επεξεργαστών, δίνοντάς μας έτσι πολύ πιο γρήγορα τα αποτελέσματα. Τέλος, πρέπει να αναφέρουμε ότι τα μειονεκτήματα που παρατηρήθηκαν με τη χρήση του συγκεκριμένου πακέτου είναι δύο. Το πρώτο, και ίσως πολύ σημαντικό, είναι ότι δε διατίθεται δωρεάν και το δεύτερο ότι ο χρήστης του συγκεκριμένου πρέπει να μάθει να γράφει κώδικα, ώστε να μπορέσει να εκτελέσει ένα δικό του στατιστικό μοντέλο.

2.4. PHP και Design Patterns

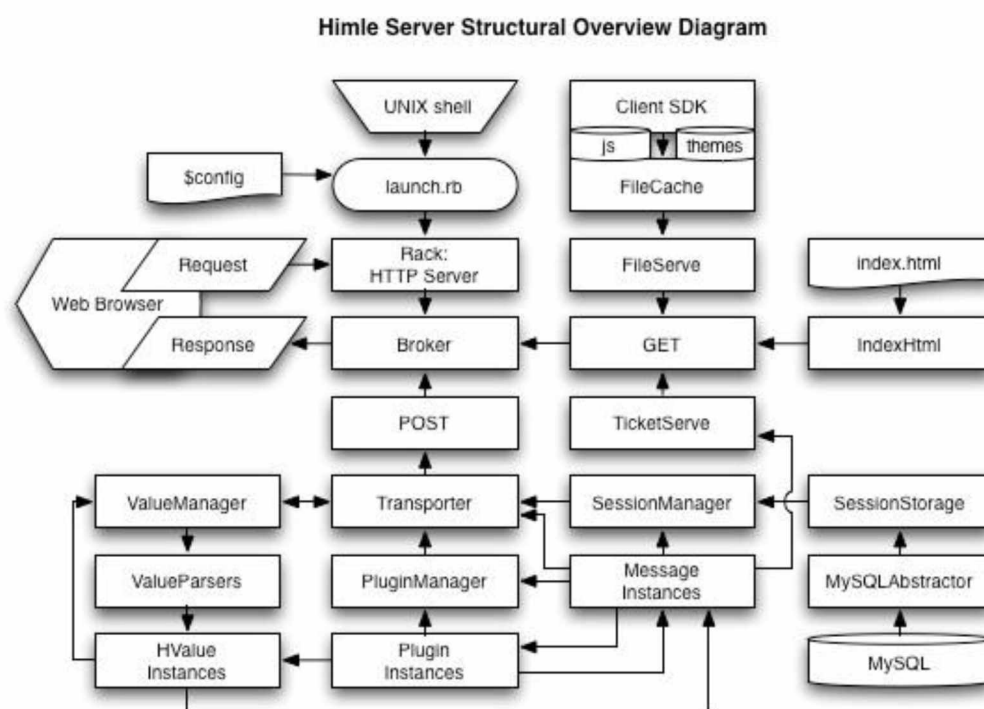
2.4.1 Η γλώσσα PHP

Η PHP έκανε την πρώτη της εμφάνιση το 1994 από τον Rasmus Lerdorf, ο οποίος τότε ήταν φοιτητής. Η PHP ξεκίνησε από ένα script που έγραψε ο Lerdorf στη γλώσσα C. Το όνομα του συγκεκριμένου script ήταν php.cgi. Ο σκοπός για τον οποίο δημιουργήθηκε ήταν για να μπορεί ο Lerdorf να κρατάει στατιστικά για τα άτομα τα οποία επισκέφθηκαν το βιογραφικό του, το οποίο ήταν online. Το αρχικό όνομα της γλώσσας ήταν PHP/FI και το 1997 εμφανίστηκε η

δεύτερη έκδοσή της (2.0). Αργότερα, το 1997 οι Andi Gutmans και Zeev Suraski ξανά έγραψαν τη γλώσσα από την αρχή, έχοντας ως πρότυπο τη δεύτερη έκδοση, όπως είδαμε νωρίτερα. Έτσι η PHP έφθασε στην έκδοση 3.0 η οποία ομοιάζει με τη σημερινή μορφή της. Το 1998 ακολούθησε η έκδοση 4.0, τον Ιούλιο του 2004 η έκδοση 5.0 και σήμερα βρισκόμαστε στην 7.0.

Η συγκεκριμένη γλώσσα προγραμματισμού είναι πολύ διαδεδομένη στις μέρες μας για τη δημιουργία δυναμικών ιστοσελίδων· αυτό σημαίνει πως δίνει τη δυνατότητα στις σελίδες μας να «επικοινωνούν» καλύτερα με τους χρήστες, αλλά και η διαχείρισή τους είναι πολύ πιο εύκολη σε σχέση με μια στατική ιστοσελίδα. Η PHP είναι γλώσσα προγραμματισμού που αφορά την πλευρά του server (server side).

Παρακάτω μπορούμε να δούμε τη δομή ενός Himle server:



Εικόνα 3: Δομή Himle server

Για να δημιουργήσουμε μια εφαρμογή με τη συγκεκριμένη γλώσσα αρκεί η επέκταση του αρχείου που έχουμε γράψει να είναι *.php, *.php4, *.phtml κ.ά. Κάνοντας χρήση της PHP έχουμε τη δυνατότητα να δημιουργήσουμε είτε μεμονωμένα αρχεία php, είτε να τοποθετήσουμε κώδικα μέσα σε ήδη υπάρχοντα HTML κώδικα, αξιοποιώντας έτσι και την εμφάνιση που μας παρέχεται μέσω της HTML, αλλά και της δυναμικότητας της PHP.

Η γλώσσα PHP είναι μια αντικειμενοστραφής γλώσσα προγραμματισμού ή όπως αλλιώς αναφέρεται Object Oriented Programming (OOP). Το χαρακτηριστικό αυτό της συγκεκριμένης γλώσσας είναι πολύ σημαντικό διότι παρέχει στον κώδικά μας μεγάλη ευελιξία. Η ευελιξία αυτή, της PHP, όπως και σε όλες τις αντικειμενοστραφείς γλώσσες όπως η Java, η C++ και άλλες, έγκειται στο γεγονός ότι μπορούμε να δημιουργήσουμε κλάσεις, αντικείμενα δίνοντας στον κώδικά μας μεγαλύτερη ευελιξία, όπως επίσης μπορούν να γίνουν πιο εύκολα τυχόν τροποποιήσεις.

Ένα επίσης πολύ σημαντικό χαρακτηριστικό της PHP είναι πως ο κώδικάς της δεν επηρεάζεται από το αντίστοιχο λειτουργικό του στον οποίο υλοποιούμε τον κώδικά μας. Αυτό είναι πολύ

ενδιαφέρον αν σκεφτεί κανείς πως η διαφορετικότητα του λειτουργικού συστήματος επηρεάζει άλλες γλώσσες, όπως η C++. Άρα λοιπόν δεν παρουσιάζεται κανένα πρόβλημα στην εκτέλεση του προγράμματός μας είτε όταν το εκτελέσουμε μέσω κάποιου Unix συστήματος (βλέπε Linux, Mac OS), είτε μέσω λειτουργικού συστήματος της Microsoft.

Παρακάτω βλέπουμε ένα πολύ μικρό τμήμα του PHP κώδικα που χρησιμοποιήθηκε στην παρούσα εφαρμογή:

```
59         foreach ($availableExperiments as $exp) {
60             if ($exp->getExperimentClass() == 'binary') {
61                 ?>
62                 <li><a href="/?experiment=?= $exp->getName() ?>"><?= $exp->getName()
63                 <?php
64             }
65         }
66         ?>
67     </ul>
68 </li>
69 <li class="dropdown">
70     <a href="#" class="dropdown-toggle" data-toggle="dropdown" role="button" aria-hasp
71     aria-expanded="false">Summary <span class="caret"></span></a>
72     <ul class="dropdown-menu">
73         <?php
74         foreach ($availableExperiments as $exp) {
75             if ($exp->getExperimentClass() != 'binary') {
76                 ?>
77                 <li><a href="/?experiment=?= $exp->getName() ?>"><?= $exp->getName()
78                 <?php
79             }
80         }
81         ?>
82     </ul>
83 </li>
84
85 </ul>
86 </div><!-- /.nav-collapse -->
87 </div>
88 </nav>
89
```

Εικόνα 4: Στην παρούσα εικόνα και πιο συγκεκριμένα εντός του κόκκινου πλαισίου παρουσιάζεται ένα πολύ μικρό τμήμα του PHP κώδικά μας.

Όπως παρατηρούμε από την παραπάνω εικόνα, είναι πολύ εύκολο να εντοπίσουμε τον κώδικα PHP μιας και ορίζεται η αρχή του κώδικα με το χαρακτηριστικό “<?php ” και με το “?” ορίζεται το τέλος του κώδικα μας.

Σε αυτό το σημείο θα ήταν καλό να παρατηρήσουμε ότι η εκμάθηση της συγκεκριμένης γλώσσας δεν είναι ιδιαίτερα δύσκολη, ιδιαίτερα αν έχουμε ξανά γράψει κώδικα σε μια γλώσσα όπως η C, Java, Python και ούτω καθ’ εξής.

2.4.2 Design Patterns

Ένα από τα πιο καίριας σημασίας πεδία που υπάρχουν στο χώρο του προγραμματισμού είναι αυτό που ονομάζεται «σχεδιαστικά μοντέλα» ή όπως είναι πιο γνωστά “Design Patterns”. Τα μοντέλα αυτά είναι από τα πιο σημαντικά τμήματα κώδικα σε ό,τι αφορά το software engineering. Ο σκοπός των συγκεκριμένων μοντέλων είναι να μας δίνουν ένα γενικό πλαίσιο για τη δημιουργία ενός όσο πιο αποδοτικού κώδικα γίνεται, ώστε να λαμβάνουμε τα αποτελέσματα μας γρηγορότερα – χωρίς περιττούς υπολογισμούς- και γενικότερα μας δίνεται η δυνατότητα να τα επαναχρησιμοποιήσουμε οποιαδήποτε στιγμή και σε διαφορετικές περιπτώσεις. Τα μοντέλα αυτά, είναι ό,τι καλύτερο υπάρχει μέχρι αυτή τη στιγμή για την

επίλυση περίπλοκων αλλά και μη προβλημάτων καθ' όσον καλούμαστε να σχεδιάσουμε μια εφαρμογή ή ένα σύστημα. Τα συγκεκριμένα μοντέλα χρησιμοποιούνται κατά το πλείστον από γλώσσες προγραμματισμού που εμπίπτουν στο χαρακτηριστικό της αντικειμενοστρέφειας, όπως η PHP που είδαμε νωρίτερα, η Java, η Python κ.ο.κ., ο λόγος είναι ότι και τα ίδια τα μοντέλα κάνουν χρήση της αντικειμενοστρέφειας, μέσω της οποίας μας γίνονται κατανοητές οι σχέσεις που υπάρχουν στις επικοινωνίες μεταξύ των κλάσεων ή των αντικειμένων, χωρίς ωστόσο να καθορίζονται ακριβώς οι τελικές κλάσεις της εφαρμογής μας ή τα αντικείμενα τα οποία χρησιμοποιούνται. Πιο συνοπτικά θα μπορούσαμε να αναφέρουμε ότι τα σχεδιαστικά μοντέλα είναι η δομημένη προσέγγιση στα ζητήματα του προγραμματισμού υπολογιστών.

Τα σχεδιαστικά μοντέλα μπορούν να ομαδοποιηθούν σε τρεις κατηγορίες: τα μοντέλα δημιουργίας (creational patterns), τα δομημένα μοντέλα (structural patterns) και τα μοντέλα συμπεριφοράς (behavior patterns).

Στον παρακάτω πίνακα διακρίνουμε από κάθε ομάδα (όπως αναφέραμε παραπάνω) τα αντίστοιχα μοντέλα:

Creational Patterns	Structural Patterns	Behavior Patterns
Abstract factory pattern	Adapter pattern	Chain of responsibility pattern
Builder pattern	Aggregate pattern	Command pattern
Factory method pattern	Bridge pattern	Externalize pattern
Prototype pattern	Composite pattern	Interpreter pattern
Singleton pattern	Decorator pattern	Iterator pattern
	Extensibility pattern	Mediator pattern
	Facade pattern	Memento pattern
	Flyweight pattern	Null object pattern
	Marker pattern	Observe pattern
	Pipes and filters	Protocol stack
	Opaque pointer	Scheduled-task pattern
	Proxy pattern	Single-serving visitor pattern
		Specification pattern
		State pattern
		Strategy pattern
		Template method pattern
		Visitor pattern

Πίνακας 7: Κατηγορίες σχεδιαστικών μοντέλων με τα αντίστοιχα μοντέλα.

Στην παρούσα εργασία έγινε χρήση των παρακάτω σχεδιαστικών προτύπων:

1. Singleton Pattern,
2. Decorator Pattern,
3. Façade Pattern και του
4. Iterator Pattern.

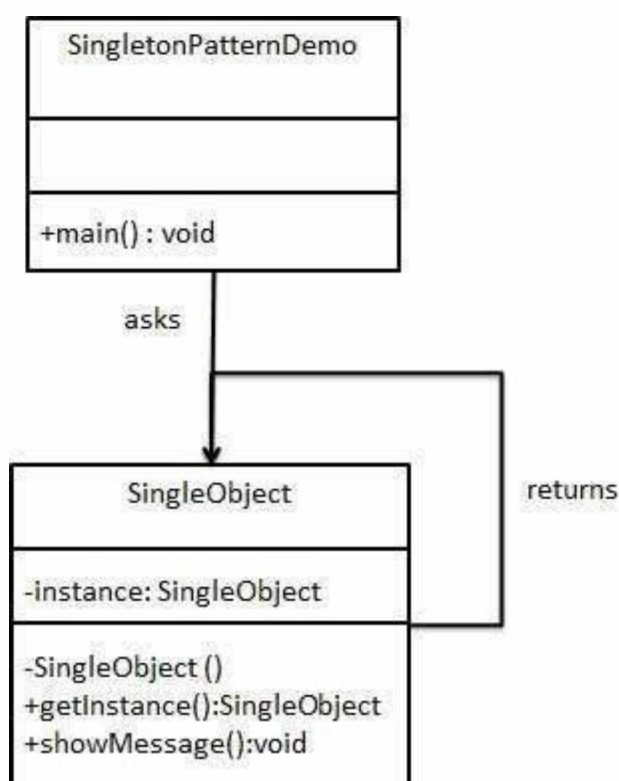
2.4.2.1 Singleton Pattern

Το πρότυπο Singleton είναι ένα από τα πιο απλά πρότυπα σχεδίασης. Το συγκεκριμένο design pattern μας δίνει τη δυνατότητα να δημιουργούμε με πολύ καλύτερους τρόπους ένα

αντικείμενο. Το παρών πρότυπο περιλαμβάνει μία κλάση η οποία είναι υπεύθυνη για τη δημιουργία ενός αντικειμένου, ενώ παράλληλα διασφαλίζεται η δημιουργία μόνο του ενός αντικειμένου. Η παραπάνω κλάση μας δίνει τη δυνατότητα να έχουμε πρόσβαση στο μοναδικό αντικείμενό της άμεσα.

Έτσι λοιπόν, φθάσαμε στη δημιουργία του παραπάνω προτύπου. Για αρχή δημιουργούμε μια κλάση που την ονομάζουμε SingleObject. Η συγκεκριμένη κλάση μεταχειρίζεται ιδιωτικά (private) τη συνάρτηση δημιουργό της (constructor) δίνοντας κατ' αυτόν τον τρόπο μια στατική εμφάνιση. Έτσι λοιπόν, η κλάση SingleObject παρέχει μια στατική μέθοδο για να αποκτήσει ένα στατικό στιγμιότυπο (instance) προς τα έξω.

Όπως μπορούμε να δούμε και στην παρακάτω εικόνα, δημιουργούμε και μια ακόμη κλάση με όνομα SingletonPatternDemo, η οποία θα χρησιμοποιήσει την πρώτη κλάση (SingleObject), ώστε να αποκτήσει ένα αντικείμενό της.



Εικόνα 5: Δομή και λειτουργία του Singleton Design Pattern

Παρακάτω, μπορούμε να δούμε μια υλοποίηση του παραπάνω μοντέλου με τη χρήση της PHP:

```

<?php
namespace DesignPatterns\Creational\Singleton;
final class Singleton
{
    /**
     * @var Singleton
     */
    private static $instance;
    /**
     * gets the instance via lazy initialization (created on first usage)
  
```

```

*/
public static function getInstance(): Singleton
{
    if (null === static::$instance) {
        static::$instance = new static();
    }
    return static::$instance;
}
/**
 * is not allowed to call from outside to prevent from creating multiple instances,
 * to use the singleton, you have to obtain the instance from
Singleton::getInstance() instead
*/
private function _construct()
{
}
/**
 * prevent the instance from being cloned (which would create a second instance
of it)
*/
private function _clone()
{
}
/**
 * prevent from being unserialized (which would create a second instance of it)
*/
private function _wakeup()
{
}
}

```

Πίνακας 8: Υλοποίηση του Singleton Design Pattern με χρήση της PHP.

2.4.2.2 Decorator Pattern

Ένα ακόμη πρότυπο το οποίο χρησιμοποιήθηκε στην παρούσα εφαρμογή ήταν το Decorator Design Pattern. Το συγκεκριμένο πρότυπο επισυνάπτει δυναμικά επιπλέον αρμοδιότητες σε ένα αντικείμενο. Το παρών πρότυπο παρέχει μια πιο ευέλικτη εναλλακτική λύση στην υπο-κλάση για την επέκταση της λειτουργικότητάς της. Άρα, έχουμε μια δυναμική τροποποίηση των αντικειμένων της κλάσης στην οποία ανήκουν, χωρίς παράλληλα να υπάρχουν προβλήματα στην υποκείμενη ή άλλες κλάσεις. Αυτό κάνει το συγκεκριμένο πρότυπο ιδιαίτερα χρήσιμο σε οικείες υποκλάσεις. Επίσης, αν έχουμε ένα σύνολο από υποκλάσεις και πρέπει αν προσθέσουμε επιπλέον δυνατότητες, οι οποίες μπορούν να εφαρμοστούν ανεξάρτητα στην κάθε υποκλάση, μπορούμε να χρησιμοποιήσουμε το παρών πρότυπο για να αποφύγουμε να γράψουμε δύο και πλέον φορές τον ίδιο κώδικα.

Το Decorator Pattern δομικά ομοιάζει με αυτό του Proxy Pattern. Το αντικείμενο του Decorator κρατά μια αναφορά σε ένα αντικείμενο και αναδημιουργεί μια ακριβώς ίδια δημόσια (κοινή για όλη την εφαρμογή - public) διεπαφή στο νέο αντικείμενο. Με το Decorator μπορούμε επίσης να

προσθέσουμε μεθόδους, επεκτείνοντας τη διεπαφή του νεοδημιουργηθέντος αντικειμένου ή επίσης μπορεί να αντικαταστήσει τις μεθόδους κατά βούληση.

Παρακάτω παραθέτουμε ένα δείγμα κώδικα σε PHP, το οποίο υλοποιεί το συγκεκριμένο πρότυπο:

```
abstract class Component
{
    protected $data;
    protected $value;

    abstract public function getData();

    abstract public function getValue();
}
class ConcreteComponent extends Component
{
    public function _construct()
    {
        $this->value = 1000;
        $this->data = "Concrete Component:\t{$this->value}\n";
    }
    public function getData()
    {
        return $this->data;
    }
    public function getValue()
    {
        return $this->value;
    }
}
abstract class Decorator extends Component
{
}
class ConcreteDecorator1 extends Decorator
{
    public function _construct(Component $data)
    {
        $this->value = 500;
        $this->data = $data;
    }
    public function getData()
    {
```

```

    return $this->data->getData() . "Concrete Decorator 1:\t{$this->value}\n";
}
public function getValue()
{
    return $this->value + $this->data->getValue();
}
}
class ConcreteDecorator2 extends Decorator
{
    public function _construct(Component $data)
    {
        $this->value = 500;
        $this->data = $data;
    }
    public function getData()
    {
        return $this->data->getData() . "Concrete Decorator 2:\t{$this->value}\n";
    }

    public function getValue()
    {
        return $this->value + $this->data->getValue();
    }
}
class Client
{
    private $component;
    public function _construct()
    {
        $this->component = new ConcreteComponent();
        $this->component = $this->wrapComponent($this->component);

        echo $this->component->getData();
        echo "Client:\t\t\t";
        echo $this->component->getValue();
    }
    private function wrapComponent(Component $component)
    {
        $component1 = new ConcreteDecorator1($component);
        $component2 = new ConcreteDecorator2($component1);
        return $component2;
    }
}
$client = new Client();

```

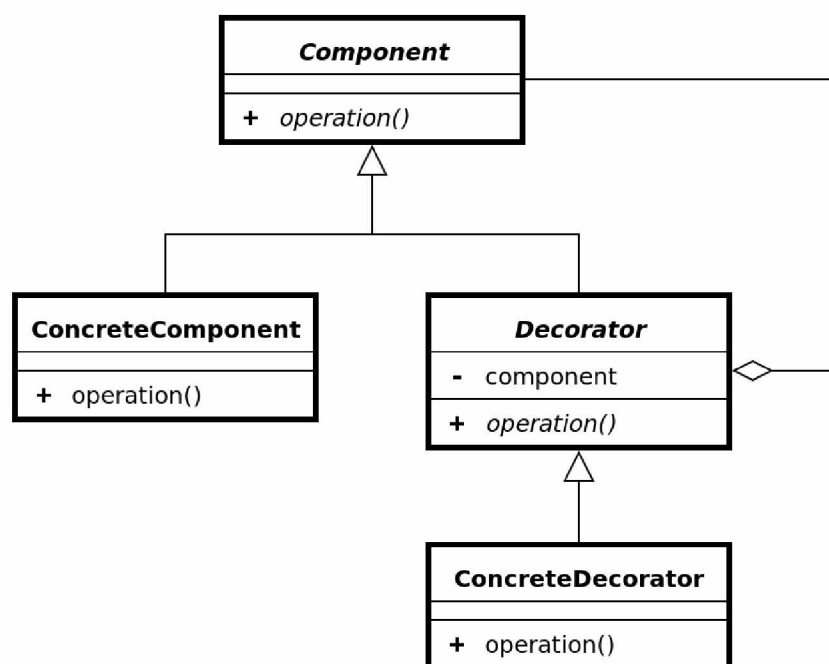
```

// Result: #quanton81
//Concrete Component:    1000
//Concrete Decorator 1:    500
//Concrete Decorator 2:    500
//Client:                2000

```

Πίνακας 9: Υλοποίηση του Decorator Pattern στη γλώσσα PHP.

Επίσης, παρακάτω παρουσιάζεται το UML σχήμα, το οποίο κάνει ακόμη πιο κατανοητό το παραπάνω πρότυπο:

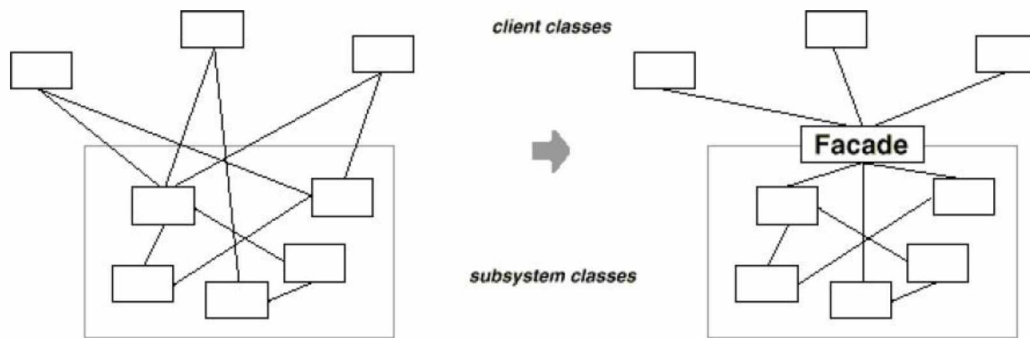


Εικόνα 6: Το UML σχήμα του Decorator Pattern.

2.4.2.3 Façade Pattern

Το πρότυπο Façade είναι ένα από τα πιο διαδεδομένα πρότυπα τα οποία χρησιμοποιούνται στον αντικειμενοστραφή προγραμματισμό (OOP, Object Oriented Programming). Το συγκεκριμένο πρότυπο είναι ένα αντικείμενο το οποίο μας παρέχει μια απλή διεπαφή, αποκρύπτοντας το μεγάλο μεγέθους κώδικα που χρησιμοποιούμε. Μέσω του προτύπου Façade μπορούμε να δώσουμε περισσότερη ευελιξία σε ό,τι αφορά το σχεδιασμό του συστήματός μας, διότι μειώνουμε τις εξαρτήσεις που υπάρχουν στον κώδικά μας με εσωτερικές λειτουργίες κάποιας βιβλιοθήκης. Επίσης, μας δίνει τη δυνατότητα η χρήση του συγκεκριμένου προτύπου να δημιουργήσουμε λογισμικό βιβλιοθηκών, δίνοντάς μας ευκολία στη χρήση τους, στην κατανόησή τους, όπως επίσης και στη δοκιμή τους αυτό συνεπάγεται ότι ο κώδικας από τις βιβλιοθήκες που δημιουργήσαμε είναι πιο ευανάγνωστος.

Παρακάτω, παρουσιάζεται μια απλή δομή του συστήματός μας, το οποίο κάνει χρήση του Façade προτύπου:



Εικόνα 7: Δομή συστήματος που κάνει χρήση το πρότυπο Façade.

Μια ενδεικτική μορφή κώδικα, όπως αυτός παρουσιάζεται από τη διαδικτυακή εγκυκλοπαίδεια της Wikipedia παρουσιάζεται παρακάτω:

```

/**
 * The complicated, underlying logic.
 */
class CPU
{
    public function freeze() { /* ... */ }
    public function jump($position) { /* ... */ }
    public function execute() { /* ... */ }
}
class Memory
{
    public function load($position, $data) { /* ... */ }
}
class HardDrive
{
    public function read($lba, $size) { /* ... */ }
}
/**
 * The facade that users would be interacting with.
 */
class ComputerFacade
{
    protected $cpu;
    protected $memory;
    protected $hd;
    public function _construct()
    {

```

```

$this->cpu = new CPU;
$this->ram = new Memory;
$this->hd = new HardDrive;
}
public function start()
{
    $this->cpu->freeze();
    $this->ram->load(BOOT_ADDRESS, $this->hd->read(BOOT_SECTOR,
SECTOR_SIZE));
    $this->cpu->jump(BOOT_ADDRESS);
    $this->cpu->execute();
}
}
/**
 * How a user could start the computer.
 */
$computer = new ComputerFacade;
$computer->start();

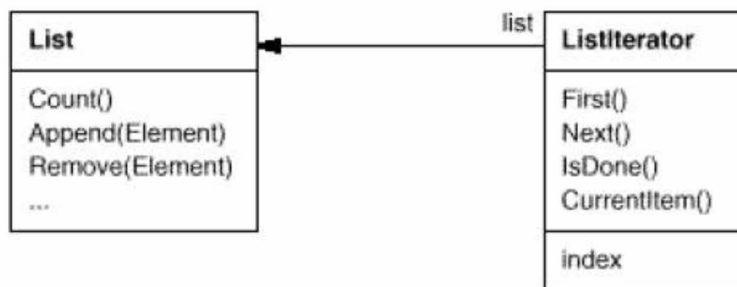
```

Πίνακας 10: Ενδεικτικός κώδικας υλοποίησης του παραπάνω προτύπου, σε γλώσσα PHP.

2.4.2.4 Iterator Pattern

Το τελευταίο πρότυπο το οποίο χρησιμοποιήσαμε στην παρούσα εργασία είναι το Iterator. Συνοπτικά, θα μπορούσαμε να αναφέρουμε ότι το συγκεκριμένο πρότυπο μας δίνει πρόσβαση στα στοιχεία ενός συνόλου αντικειμένων.

Κάνοντας χρήση ενός συγκεντρωτικού αντικειμένου, όπως για παράδειγμα μιας λίστας, μας δίνεται η δυνατότητα πρόσβασης στα αντίστοιχα στοιχεία του, χωρίς παράλληλα να μας γνωστοποιείται η εσωτερική δομή του. Επιπλέον, μπορεί να θέλουμε να διατρέξουμε ολόκληρη τη λίστα μας με διαφορετικούς τρόπους, ανάλογα με το είδος των αποτελεσμάτων που θέλουμε. Προφανώς, δεν επιθυμούμε να γεμίσουμε τη διεπαφή της λίστας μας με λειτουργίες για τους διαφορετικούς τρόπους με τους οποίους τη διατρέχουμε. Επίσης, υπάρχει η πιθανότητα να διατρέξουμε τη λίστα μας περισσότερες από μία φορές. Αν θέλουμε λοιπόν, να κάνουμε κάτι από τα παραπάνω τότε η καλύτερη λύση –προς το παρόν– είναι η χρήση του τρέχοντος προτύπου. Η βασική ιδέα του μοντέλου είναι το πρότυπο αυτό να πάρει την ευθύνη για την πρόσβασή μας και το πέρασμά μας εκτός του αντικειμένου της λίστας μας. Στην παρακάτω εικόνα παρατηρούμε την ύπαρξη της κλάσης Iterator, μέσω της οποίας ορίζεται η διεπαφή, και επίσης μέσω αυτής έχουμε πρόσβαση στα στοιχεία της λίστας μας. Ένα iterator αντικείμενο είναι υπεύθυνο για να κρατήσει το ίχνος του συγκεκριμένου στοιχείου, δηλαδή, να ξέρει από ποιά στοιχεία έχουμε ήδη περάσει.



Εικόνα 8: Δομή του Iterator προτύπου.

Προτού μπορέσουμε να εμφανίσουμε την κλάση ListIterator, πρέπει να παρέχουμε τη λίστα για να τη διατρέξουμε. Μια ακόμη φορά, έχουμε ένα στιγμιότυπο (instance) της κλάσης ListIterator, κατ' αυτόν τον τρόπο έχουμε τη δυνατότητα πρόσβασης στα στοιχεία της λίστας διαδοχικά. Η συνάρτηση CurrentElement() που υπάρχει εντός της κλάσης ListIterator, επιστρέφει το συγκεκριμένο στοιχείο στη λίστα. Πρώτα γίνεται αρχικοποίηση του τρέχοντος στοιχείου και τοποθετείται σαν πρώτο στοιχείο. Στη συνέχεια προωθείται το τρέχον στοιχείο στο επόμενο και μέσω της συνάρτησης IsDone() ελέγχουμε αν έχουμε προχωρήσει πέρα από το τελευταίο στοιχείο της λίστας μας ή πιο απλά αν διατρέξαμε όλα τα στοιχεία της λίστας μας. Παρακάτω παραθέτουμε έναν υποτυπώδη κώδικα με χρήση της γλώσσας PHP, όπως παρουσιάζεται στη διαδικτυακή εγκυκλοπαίδεια της Wikipedia:

```

<?php

// BookIterator.php
namespace DesignPatterns;
class BookIterator implements \Iterator
{
    private $i_position = 0;
    private $booksCollection;
    public function __construct(BookCollection $booksCollection)
    {
        $this->booksCollection = $booksCollection;
    }
    public function current()
    {
        return $this->booksCollection->getTitle($this->i_position);
    }
    public function key()
    {
        return $this->i_position;
    }
    public function next()
    {
        $this->i_position++;
    }
}
  
```

```

}
public function rewind()
{
    $this->i_position = 0;
}
public function valid()
{
    return !is_null($this->booksCollection->getTitle($this->i_position));
}
}

```

```

<?php

// BookCollection.php
namespace DesignPatterns;
class BookCollection implements \IteratorAggregate
{
    private $a_titles = array();
    public function getIterator()
    {
        return new BookIterator($this);
    }
    public function addTitle($string)
    {
        $this->a_titles[] = $string;
    }
    public function getTitle($key)
    {
        if (isset($this->a_titles[$key])) {
            return $this->a_titles[$key];
        }
        return null;
    }
    public function is_empty()
    {
        return empty($this->$a_titles);
    }
}

```

```
<?php
// index.php
require 'vendor/autoload.php';
use DesignPatterns\BookCollection;
$booksCollection = new BookCollection();
$booksCollection->addTitle('Design Patterns');
$booksCollection->addTitle('PHP7 is the best');
$booksCollection->addTitle('Laravel Rules');
$booksCollection->addTitle('DHH Rules');
foreach($booksCollection as $book){
    var_dump($book);
}
```

Πίνακας 11: Μορφή κώδικα για το Iterator πρότυπο.

2.5. HTML και Javascript

Όπως αναφέρθηκε και στην αρχή της παρούσης εργασίας, ο σκοπός της συγκεκριμένης εργασίας-εφαρμογής ήταν η δημιουργία μιας ιστοσελίδας μέσω της οποίας ο εκάστοτε χρήστης θα είχε την ευκαιρία να εκτελέσει ελεύθερα μετα-αναλύσεις. Αυτό σημαίνει ότι για την εφαρμογή μας χρειάστηκε να αναπτύξουμε αρχικά το γραφικό περιβάλλον μέσω του οποίου επικοινωνεί ο χρήστης και εν συνεχεία το δεύτερο και πολύ σημαντικό σκέλος που αφορά την πλευρά του server, μέσω του οποίου θα εκτελεστούν οι αντίστοιχες μετα-αναλύσεις. Άρα, όπως εύκολα παρατηρούμε έγινε χρήση του λεγόμενου Front-End και του Back-End προγραμματισμού.

2.5.1 Front-End και Back-End προγραμματισμός

2.5.1.1 Front-End

Σε αυτό το είδος του Web Developing ασχολούμαστε με τη μορφή (web design) και γενικότερα τη λειτουργία της ιστοσελίδας μας. Στον Front-End προγραμματισμό έχουμε την εμφάνιση των δύο βασικότερων «γλωσσών προγραμματισμού», της HTML και του CSS. Δε θα μπορούσαμε να χαρακτηρίσουμε ως γλώσσες προγραμματισμού τα HTML και CSS, μιας και είναι περισσότερο ο τρόπος εμφάνισης της εφαρμογής μας και τίποτα παραπάνω.

Στην παρακάτω εικόνα βλέπουμε ένα τμήμα του HTML και του CSS κώδικα που υλοποιήσαμε:

```
<style>
  .delete-row {
    display: block;
    padding: 3px;
  }
</style>
</head>
<body>

<nav class="navbar navbar-inverse navbar-fixed-top">
  <div class="container">
    <div class="navbar-header">
      <button type="button" class="navbar-toggle collapsed" data-toggle="collapse" data-target="
        aria-expanded="false" aria-controls="navbar">
        <span class="sr-only">Toggle navigation</span>
        <span class="icon-bar"></span>
        <span class="icon-bar"></span>
        <span class="icon-bar"></span>
      </button>
      <a class="navbar-brand" href="#">Bioinformatics Experiment Runner</a>
    </div>
    <div id="navbar" class="collapse navbar-collapse">
      <ul class="nav navbar-nav">
        <li class="dropdown">
          <a href="#" class="dropdown-toggle" data-toggle="dropdown" role="button" aria-haspopup="
            aria-expanded="false">Binary <span class="caret"></span></a>
          <ul class="dropdown-menu">
```

Εικόνα 9: Τμήμα του HTML και του CSS κώδικα της εφαρμογής μας.

Όμως, όπως είπαμε και παραπάνω, αυτό είναι το ένα είδος προγραμματισμού. Μέχρι τώρα όμως δεν είδαμε το προγραμματιστικό σκέλος του Front-End προγραμματισμού. Ο προγραμματισμός αυτός γίνεται με τη βοήθεια της Javascript και της jQuery. Όπως θα δούμε και παρακάτω, η Javascript μας δίνει ένα πιο δυναμικό χαρακτήρα στην ιστοσελίδα μας. Όσο για τη jQuery, θα μπορούσαμε να αναφέρουμε ότι είναι μια Javascript βιβλιοθήκη, απλούστερη από τη Javascript αλλά και πιο εύκολη στην εκμάθησή της, απ' ό,τι η JavaScript.

Έτσι λοιπόν, ολοκληρώσαμε το interface και τις βασικές λειτουργίες της εφαρμογής μας, το οποίο για τον κοινό χρήστη είναι αρκετά σημαντικό.

2.5.1.2 Back-End

Ο Back-End προγραμματισμός είναι η «καρδιά» της όλης εφαρμογής. Είναι τόσο πολύ σημαντικό αυτό το είδος του προγραμματισμού, μιας και αυτό κάνει την πολύ και δύσκολη εργασία. Σε αυτό το κομμάτι του προγραμματισμού κύριο ρόλο λαμβάνουν γλώσσες όπως η PHP, η Python, η Perl και άλλες. Ο Back-End προγραμματισμός, είναι το τμήμα του web development που αφορά κυρίως τη λειτουργία σε επίπεδο server, πράγμα που σημαίνει ότι ο χρήστης δε μπορεί να δει τίποτα από όλα αυτά τα οποία συμβαίνουν στο παρασκήνιο, αλλά λαμβάνει μόνο τα τελικά αποτελέσματα. Σε αυτό το σημείο θα ήταν καλό να αναφέρουμε ότι στο πεδίο του Back-End προγραμματισμού, συμπεριλαμβάνονται και οι βάσεις δεδομένων, όπως και το λειτουργικό σύστημα, για το οποίο μιλήσαμε νωρίτερα. Οι εφαρμογές αυτού του τύπου προγραμματισμού μπορεί να διαφέρουν πάρα πολύ μεταξύ τους, αν η εφαρμογή μας χρησιμοποιεί cloud servers ή warehouses, τότε για να γίνει όπως αναφέρεται το containerization, δηλαδή η τοποθέτηση της εφαρμογής μας εντός του χώρο στον οποίο θα εκτελεί τις εργασίες μας, μαζί με τις αντίστοιχες υπηρεσίες όπως το Docker, το BaaS (Backend-as-a-Service), είτε σε APIs τα οποία αφορούν την αντικατάστασή τους από πιο σύνθετες διεργασίες.

2.5.2 HTML

Η ιστορία της HTML ξεκινάει από τη δεκαετία του '80, σε μια απ' της πιο πολύ-συζητημένες τοποθεσίες του κόσμου. Το περίφημο εργαστήριο του CERN. Στο CERN ο φυσικός Τιμ Μπέρνερς Λι ήθελε να διαμοιράσει έγγραφα σε άλλους συνεργάτες του. Έτσι λοιπόν, δημιούργησε το ENQUIRE. Το 1989 πρότεινε τη χρήση του υπερκειμένου, το οποίο ήταν ένα σύστημα που βασιζόταν στο νεοφανές διαδίκτυο. Έτσι ο Λι έφτιαξε την πρώτη μορφή της HTML, όπως επίσης και το αντίστοιχο λογισμικό του server στα τέλη του 1990. Στα τέλη του 1991 η HTML ήταν πλέον διαθέσιμη με το όνομα "HTML tags". Με το πέρασμα των ετών υπήρξαν νέες εκδόσεις, το 1995 ήταν η δεύτερη, φθάνοντας στο 2014 που έχουμε την τελευταία μέχρι σήμερα έκδοσή της που αριθμείται ως η πέμπτη (5).

Σε πιο τεχνικά χαρακτηριστικά μπορεί κανείς να παρατηρήσει ότι η HTML συντίθεται από τις λεγόμενες ετικέτες ή αλλιώς tags, οι οποίες συνήθως χρησιμοποιούνται σε ζεύγη, με την πρώτη να αναφέρεται ως ετικέτα έναρξης και τη δεύτερη ως ετικέτα λήξης. Τέτοιες ετικέτες είναι η `<html>...</html>`, η οποία σηματοδοτεί την αρχή και το τέλος του HTML κώδικα, η `<h1>...</h1>`, η οποία μας αλλάζει το μέγεθος της γραμματοσειράς του κειμένου που εμπεριέχεται στην ετικέτα έναρξης και λήξης, και άλλες με τις αντίστοιχες ιδιότητές τους. Επίσης, ένα ενδιαφέρον χαρακτηριστικό που μας παρέχει η HTML είναι ότι μπορούμε να προσθέσουμε στην ιστοσελίδα μας εικόνες, πίνακες, κινούμενη εικόνα, ήχο και πολλά άλλα, μέσω των οποίων δημιουργείται μια πιο ευχάριστη αίσθηση στο χρήστη. Επίσης, θα ήταν παράληψή μας να μην αναφέρουμε ότι η HTML μας δίνει τη δυνατότητα να κάνουμε τον ιστοχώρο μας πιο δυναμικό με τη χρήση των script γλωσσών προγραμματισμού, όπως η JavaScript, η Python και άλλες.

2.5.3 Javascript

Η JavaScript έκανε την εμφάνισή της με το όνομα Mocha, η οποία δημιουργήθηκε από τον Brendan Eich. Στη συνέχεια η Mocha μετονομάστηκε σε LiveScript για να καταλήξει στη σημερινή της ονομασία. Όπως διαφαίνεται και από το σημερινό της όνομα, η JavaScript έχει πάρει πολλά στοιχεία από τη Java, όπως αυτό της αντικειμενοστρέφειας. Η JavaScript ακόμη και σήμερα είναι μια γλώσσα η οποία χρησιμοποιείται κατά κόρον σε διαδικτυακές εφαρμογές και πιο συγκεκριμένα σε εφαρμογές που αφορούν την πλευρά του πελάτη ή αλλιώς το client-side. Η μεγάλη επιτυχία της, στο χώρο του Web-developing, ήρθε όταν άρχισε να χρησιμοποιείται η τεχνολογία Ajax. Το 2009 εμφανίστηκε το CommonJS μέσω του οποίου δημιουργείται ένα κοινό πρότυπο βιβλιοθήκης, ώστε η JavaScript να βρίσκει εφαρμογή και σε επίπεδο server. Σήμερα, γίνεται μεγάλη χρήση του Node.js, το οποίο αφορά και αυτό τον προγραμματισμό σε επίπεδο server, με τη μόνη διαφορά ότι το συγκεκριμένο είδος προγραμματισμού βασίζεται σε γεγονότα (events).

Παρακάτω, μπορούμε να δούμε ένα μέρος του κώδικά μας σε Javascript:

```
107 <!-- jQuery (necessary for Bootstrap's JavaScript plugins) -->
108 <script src="https://ajax.googleapis.com/ajax/libs/jquery/1.12.4/jquery.min.js"></script>
109 <!-- Include all compiled plugins (below), or include individual files as needed -->
110 <!-- Latest compiled and minified JavaScript -->
111 <script src="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/js/bootstrap.min.js"
112     integrity="sha384-Tc5IQib027qvyjSMfHj0MaLkfuWvXZxUPnCJA7l2mCWNIPg9mGCD8wGNIcPD7Txa"
113     crossorigin="anonymous"></script>
114 <script>
115     (function ($) {
116         $('body').on('click', '[data-action="delete-row"]', function () {
117             if ($('[data-row-id]').length <= 1) {
118                 alert("You can't delete the last row")
119                 return;
120             }
121             $(this).attr('data-target').remove();
122         });
123         $('[data-action="add-row"]').on('click', function () {
124             var $table = $('.parameters-table');
125             var html = $table.find('[data-row-id="0"]').html().replace("[data-row-id='0']", "[data-row-
126             html = '<tr data-row-id=' + $('[data-row-id]').length + '>' + html + '</tr>';
127             var $tmp = $(html);
128             $tmp.find('.study').text($('[data-row-id]').length + 1);
129             $table.append($tmp);
130         });
131     })(jQuery);
132 </script>
```

Εικόνα 10: Μέρος του κώδικα σε Javascript που χρησιμοποιήσαμε στην εφαρμογή μας.

3. Αποτελέσματα και Συζήτηση

3.1 Χρόνοι απόκρισης εφαρμογής

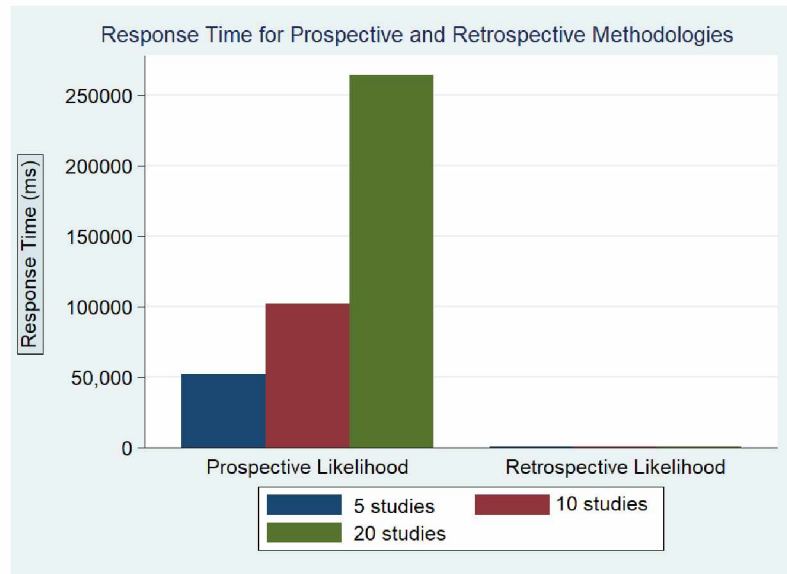
Ένα ιδιαίτερα σημαντικό χαρακτηριστικό το οποίο ενδιαφέρει τους προγραμματιστές, αλλά και τους χρήστες όλων των ειδών εφαρμογών, είναι αυτό του χρόνου απόκρισης. Όπως είναι απόλυτα λογικό μια εφαρμογή εκτός από τη χρηστικότητα που θα πρέπει να προσφέρει στο χρήστη (δηλαδή να είναι φιλική προς το χρήστη), θα πρέπει να δίνει τα αντίστοιχα αποτελέσματα που ζήτησε ο χρήστης, όσο το δυνατόν πιο γρήγορα. Φυσικά, οι χρόνοι απόκρισης της κάθε εφαρμογής διαφοροποιούνται από εφαρμογή σε εφαρμογή. Για να κατανοήσουμε καλύτερα πόσο σημαντικός είναι ο χρόνος απόκρισης ας σκεφτούμε πόσο σημαντικό είναι το gprs ενός αεροπλάνου. Αν σε περίπτωση που δε λειτουργεί, όπως πρέπει, και πολύ περισσότερο, όταν χρειαστεί περισσότερο χρόνο απ' ό,τι πρέπει, τότε αυτό μπορεί να γίνει καταστρεπτικό για το αεροπλάνο. Έτσι λοιπόν, μπορούμε να κατανοήσουμε πως έχουμε εφαρμογές που πρέπει να μας παρέχουν αποτελέσματα σχεδόν «ακαριαία», αλλά και εφαρμογές που ακόμη κι αν έχουν κάποια καθυστέρηση σχετικά με την απόκρισή τους δε θα υπάρξει κάποια ανεπανόρθωτη ζημιά, όπως για παράδειγμα αν ένας διαδικτυακός τόπος για κάποιο λόγο, π.χ. υπερφόρτωση χρηστών, δε μπορεί να ανοίξει.

Στην προκειμένη περίπτωση, η εφαρμογή η οποία δημιουργήθηκε θα πρέπει να είναι όσο το δυνατόν πιο γρήγορη μιας και ο χρήστης αναμένει στον υπολογιστή του για να λάβει τα αντίστοιχα αποτελέσματα.

Όπως προφανέστατα μπορούμε να κατανοήσουμε η ύπαρξη των επτά (7) διαφορετικών μεθοδολογιών, σηματοδοτεί και επτά διαφορετικούς χρόνους απόκρισης. Η κάθε μεθοδολογία εκτελεί το δικό της ξεχωριστό αλγόριθμο που είναι γραμμένος στην αντίστοιχη γλώσσα του Stata. Άρα, αν ο αλγόριθμος της εκάστοτε μεθοδολογίας που χρησιμοποιούμε είναι περισσότερο πολύπλοκος από κάποιον άλλο, τότε και ο αντίστοιχος χρόνος απόκρισης θα είναι μεγαλύτερος. Φυσικά, ο χρόνος απόκρισης αφορά περισσότερο τους ήδη υπάρχοντες αλγόριθμους που χρησιμοποιούμε με την εκτέλεση του Stata και όχι τόσο με τον αλγόριθμο της εφαρμογής μας.

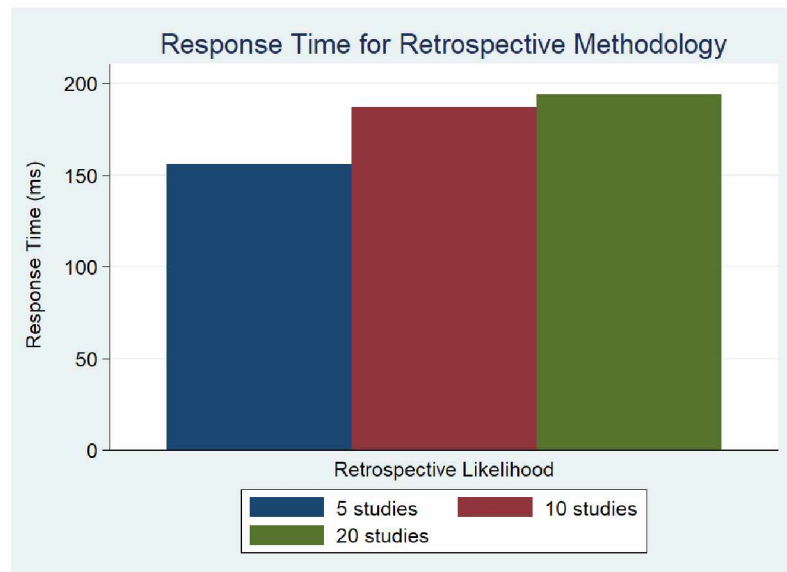
Ενδεικτικά παρουσιάζουμε παρακάτω τους αντίστοιχους χρόνους αναμονής που προκύπτουν από τις μεθοδολογίες ανά κατηγορία και είδους δεδομένων με την χρήση πέντε μελετών.

Οι χρόνοι απόκρισης για την εφαρμογή, αλλά και για τις μεθοδολογίες που κάνουν χρήση τη binary μορφή δεδομένων (Prospective και Retrospective Likelihood):



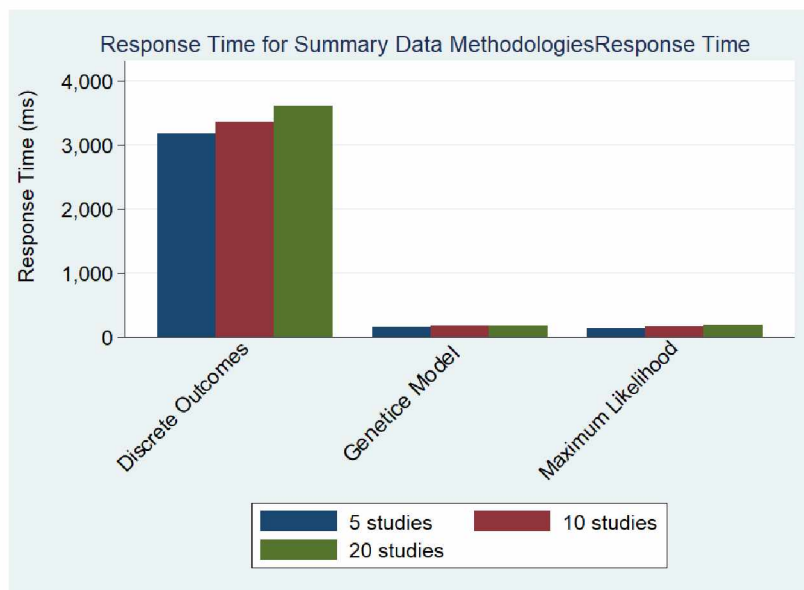
Εικόνα 11: Χρόνοι απόκρισης για binary δεδομένα

Στο παρακάτω διάγραμμα παρατηρούμε καλύτερα τους χρόνους που προκύπτουν από τη Retrospective μεθοδολογία:



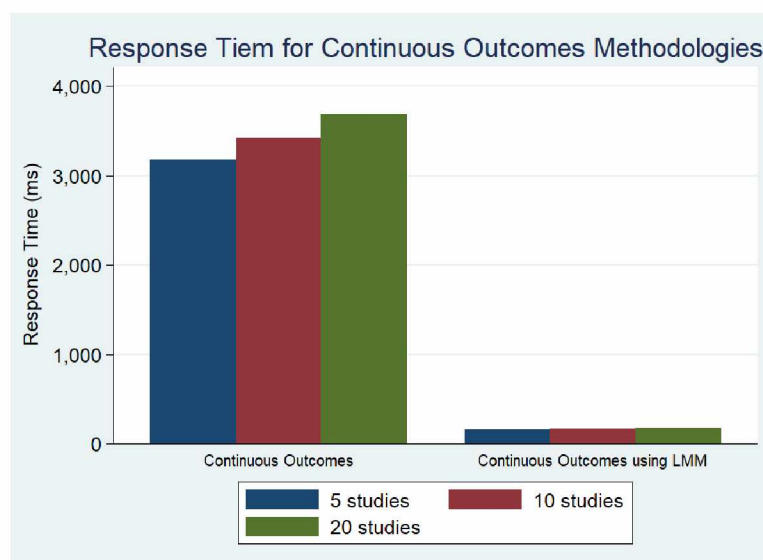
Εικόνα 12: Χρόνοι απόκρισης της Retrospective μεθοδολογίας

Στο παρακάτω διάγραμμα παρουσιάζονται οι αντίστοιχοι χρόνοι απόκρισης για τις μεθοδολογίες που κάνουν χρήση τη summary μορφή των δεδομένων:



Εικόνα 13: Χρόνοι απόκρισης για summary δεδομένα.

Στη συνέχεια, παρουσιάζονται οι χρόνοι απόκρισης για τις μεθοδολογίες Continuous Outcomes και Continuous Outcomes using Linear Mixed Model:



Εικόνα 14: Χρόνοι απόκρισης για τις Continuous μεθοδολογίες

Όπως παρατηρούμε από τα παραπάνω γραφήματα η μεθοδολογία «Prospective Likelihood» αποτελεί τη μοναδική μεθοδολογία της οποίας ο χρόνος απόκρισης είναι πολύ μεγάλος, σε σύγκριση με τις υπόλοιπες. Η ερμηνεία του συγκεκριμένου φαινομένου έγκειται στον τρόπο με τον οποίο λειτουργεί η συγκεκριμένη μεθοδολογία.

Σε αντίθεση με τις υπόλοιπες μεθοδολογίες ασθενών - μαρτύρων, η «Prospective Likelihood» χρησιμοποιεί δεδομένα τα οποία δεν είναι ταξινομημένα σε ασθενείς και μάρτυρες, αλλά δε γνωρίζουμε αν υπάρχει ο παράγοντας κινδύνου, όπως είδαμε σε προηγούμενα κεφάλαια. Άρα λοιπόν, είναι λογικό ο χρόνος απόκρισης να είναι μεγαλύτερος,

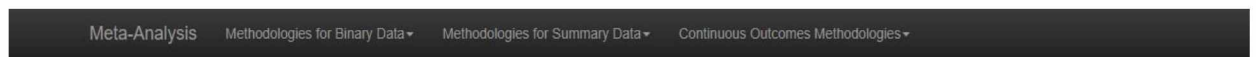
μιας και η πολυπλοκότητα του αλγορίθμου είναι μεγαλύτερη, ώστε να μας επιστρέψει όσο το δυνατόν αντικειμενικότερα αποτελέσματα.

Τέλος, διακρίνουμε μια αύξηση του χρόνου απόκρισης σε κάθε είδος μεθοδολογίας που χρησιμοποιήσαμε, μετά από την προσθήκη νέων δεδομένων από νέες μελέτες, κάτι που είναι απολύτως φυσιολογικό. Ενδεικτικά, τα αποτελέσματα που λάβαμε για τα ίδια ακριβώς δεδομένα ανά μεθοδολογία ήταν για 5, 10 και 20 διαφορετικές μελέτες.

3.2. Εφαρμογή

Η εφαρμογή μας σκοπό έχει ο εκάστοτε βιοεπιστήμονας να διενεργήσει τις μετα-αναλύσεις που επιθυμεί. Γι' αυτό το λόγο η εφαρμογή μας θα πρέπει να έχει μια αρκετά απλή διεπαφή δίνοντας στο χρήστη τη φιλικότητα που χρειάζεται, ώστε να μπορέσει να τη χρησιμοποιήσει χωρίς να ψάχνει για το ποια είναι η λειτουργία του καθενός π.χ. κουμπιού που υπάρχει.

Παρακάτω, μπορούμε να δούμε την κύρια διεπαφή (main interface) με την οποία έρχεται πρώτη φορά σε επαφή ο χρήστης:



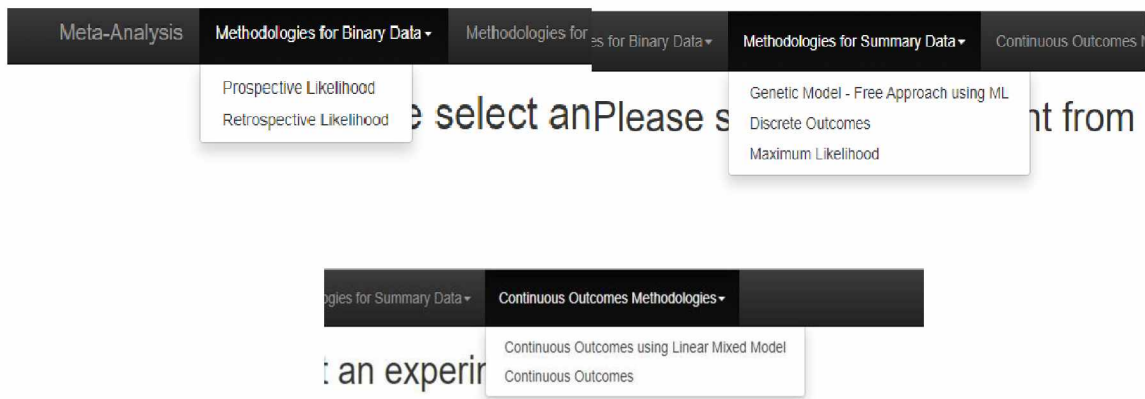
Please select an experiment from the menu

Εικόνα 15: Αρχική διεπαφή της εφαρμογής

Όπως μπορούμε να παρατηρήσουμε υπάρχουν τρία κουμπιά στο επάνω τμήμα της οθόνης μας. Το πρώτο με όνομα «Meta-Analysis» μας μεταφέρει απ' όπου κι αν βρισκόμαστε στην κεντρική οθόνη, δηλαδή στην οθόνη που είδαμε παραπάνω. Δίπλα από το κουμπί που μας μεταφέρει στην κύρια οθόνη υπάρχουν τρία ακόμη τα οποία αφορούν τις μεθοδολογίες που έχουμε ανάλογα με το είδος των δεδομένων που χρησιμοποιούμε. Πιο συγκεκριμένα, οι μεθοδολογίες που κάνουν χρήση της Binary μορφής των δεδομένων βρίσκονται στο πρώτο κουμπί με όνομα «Methodologies for binary data», αντίστοιχα δίπλα σε αυτό το κουμπί υπάρχει το αντίστοιχο που περιέχει μεθοδολογίες που κάνουν χρήση της summary μορφής των δεδομένων και των continuous.

Κάνοντας κλικ πάνω σε ένα από τα παραπάνω κουμπιά μας εμφανίζει ένα drop-down menu στο οποίο υπάρχουν οι αντίστοιχες μεθοδολογίες.

Παρακάτω, διακρίνουμε τις μεθοδολογίες των αντίστοιχων μορφών των δεδομένων:



Εικόνα 16: drop-down menu με τις μεθοδολογίες ανά είδος δεδομένων που

Εύκολα λοιπόν, ο χρήστης μπορεί να επιλέξει την αντίστοιχη μεθοδολογία που χρειάζεται και να συνεχίσει με την εισαγωγή των δεδομένων του για να ολοκληρώσει τη μετα-ανάλυση που θέλει.

Στην παρακάτω εικόνα βλέπουμε τη φόρμα στην οποία ο χρήστης μπορεί να βάλει τα δεδομένα του χειροκίνητα (ένα-προς-ένα). Υπάρχουν αντίστοιχα τα κουμπιά για την προσθήκη δεδομένων νέων μελετών για την αφαίρεση μελετών και για την εκτέλεση της μετα-ανάλυσης.

Selected experiment: **Discrete Outcomes**

Study	aa0	ab0	bb0	aa1	ab1	bb1
1	<input type="text" value="1"/>	<input type="text" value="1"/>	<input type="text" value="1"/>	<input type="text" value="1"/>	<input type="text" value="1"/>	<input type="text" value="1"/>

Note:

aa0, ab0, bb0 are the different types of genotypes from controls,
 aa1, ab1, bb1 are the different types of genotypes from cases.

gen11, gen12, gen22 are the different types of genotypes,
 x11, x12, x22 are the mean which have every genotype respectively,
 sd11, sd12, sd22 are the standard deviation which have every genotype respectively.

Εικόνα 17: φόρμες για την εισαγωγή των δεδομένων του χρήστη της αντίστοιχης μεθοδολογίας

Όπως είδαμε παραπάνω μπορούμε να βάλουμε τα δεδομένα μας ένα προς ένα, στην περίπτωση όμως που έχει ο χρήστης πολλά δεδομένα να τοποθετήσει, τότε η όλη διαδικασία γίνεται λίγο δύσκολη. Γι' αυτό το λόγο η εφαρμογή μας δίνει τη δυνατότητα στο χρήστη να κάνει upload ένα αρχείο γραμμένο σε ένα απλό «Σημειωματάριο».

Παρακάτω, βλέπουμε την αντίστοιχη φόρμα:

Upload your .txt file:

Choose your file:

Επιλογή αρχείου Δεν επιλέχθηκε... ανάνα αρχείο.

Upload

Structure of Uploaded Data:

Data for case-control studies

```
3 34 44 2 20 83
4 30 49 3 17 62
17 84 48 3 30 31
5 32 63 5 23 52
12 62 119 8 39 133
9 48 47 6 34 68
18 134 363 20 214 483
```

Data for continuous values studies

```
181 296 88 1.18 0.32 1.24 0.33 1.33 0.36
500 896 317 0.79 0.25 0.84 0.25 0.9 0.27
328 596 210 1.09 0.31 1.12 0.29 1.25 0.4
845 1236 409 1.19 0.37 1.22 .37 1.3 0.41
168 256 100 1.36 0.36 1.38 0.38 1.48 0.42
237 380 150 1.13 0.21 1.19 0.24 1.27 0.22
284 369 139 1.16 0.35 1.2 0.39 1.22 0.41
```

Εικόνα 18: φόρμα upload

Κατ' αυτόν τον τρόπο ο χρήστης ανεβάζει το αρχείο του, το οποίο έχει επέκταση .txt. Προφανώς η διάταξη που θα πρέπει να έχει το αρχείο είναι συγκεκριμένη αλλιώς δε θα μπορέσει να γίνει η μετα-ανάλυση. Γι' αυτόν ακριβώς το λόγο, ακριβώς κάτω από τη φόρμα για το «ανέβασμα» του αρχείου του, ο χρήστης λαμβάνει και τις αντίστοιχες οδηγίες για το πώς πρέπει να είναι γραμμένα τα δεδομένα του μέσα στο αντίστοιχο αρχείο που θέλει να εκτελέσει.

Παρακάτω βλέπουμε μια τυπική διάταξη δύο txt αρχείων:

uploaded_data_structure_for_continuous - Σημειωματάριο

Αρχείο Επεξεργασία Μορφή Προβολή Βοήθεια

```
181 296 88 1.18 0.32 1.24 0.33 1.33 0.36
500 896 317 0.79 0.25 0.84 0.25 0.9 0.27
328 596 210 1.09 0.31 1.12 0.29 1.25 0.4
845 1236 409 1.19 0.37 1.22 .37 1.3 0.41
168 256 100 1.36 0.36 1.38 0.38 1.48 0.42
237 380 150 1.13 0.21 1.19 0.24 1.27 0.22
284 369 139 1.16 0.35 1.2 0.39 1.22 0.41
256 335 115 0.89 0.21 0.92 0.21 1.03 0.27
1121 1693 575 0.97 0.22 1 0.23 1.07 0.24
500 797 299 1.07 0.256 1.13 0.24 1.17 0.24
```

uploaded_data_structure_for_case_control - Σημειωματάριο

Αρχείο Επεξεργασία Μορφή Προβολή Βοήθεια

```
3 34 44 2 20 83
4 30 49 3 17 62
17 84 48 3 30 31
5 32 63 5 23 52
12 62 119 8 39 133
9 48 47 6 34 68
18 134 363 20 214 483
```

Εικόνα 19: Μορφή αρχείων προς upload

Τέλος, μετά από την εκτέλεση των μεθοδολογιών που θέλουμε λαμβάνουμε τα αντίστοιχα αποτελέσματα που παράγονται από το Stata.

Παρακάτω βλέπουμε μια τυπική δομή των αποτελεσμάτων που παράγονται:

Matrixes from Discrete Outcomes Methodology :

Table with b1 and b2

	Coefficient	Standard Error	z	P> z	95% Conf.	Interval
Overall Mean						
b1	-1.07e-10	.8164996	-0.00	1.000	-1.600304	1.600304
b2	-1.07e-10	.8164996	-0.00	1.000	-1.600304	1.600304

Lambda

Group	Coefficient	Standard Error	z	P> z	95% Conf.	Interval
Lambda	9993924	7.66e+09	0.00	1.000	-1.50e+10	1.50e+10

(1)

Group	Coefficient	Standard Error	z	P> z	95% Conf.	Interval
(1)	-1.15e-14	.8164996	-0.00	1.000	-1.600304	1.600304

[Home](#) [More Results](#)

Εικόνα 20:Τυπική διάταξη αποτελεσμάτων

3.3. Συμπεράσματα

Ολοκληρώνοντας την παρούσα εργασία θα ήταν καλό να δούμε τη γενικότερη εικόνα της εφαρμογής και να εξάγουμε ορισμένα συμπεράσματα. Η συγκεκριμένη εφαρμογή, όπως είδαμε και στην αρχή της εργασίας, έχει ως κύριο μέλημά της τη διενέργεια μετα-αναλύσεων. Ο συγκεκριμένος σκοπός ολοκληρώθηκε, δίνοντας έτσι τη δυνατότητα στον κάθε χρήστη να κάνει μόνος του τις δικές του μετα-αναλύσεις σχετικές με γενετικά δεδομένα.

Παρέχεται στο χρήστη μια πληθώρα μεθοδολογιών δίνοντάς του ευελιξία κινήσεων, ανάλογα με τη μετα-ανάλυση που θέλει να κάνει. Ένα ακόμη ιδιαίτερα σημαντικό χαρακτηριστικό της συγκεκριμένης εφαρμογής είναι η απλή εμφάνισή της. Η απλή εμφάνισή της παρέχει στο χρήστη την απαιτούμενη φιλικότητα που επιζητεί η κάθε εφαρμογή. Επίσης υπάρχουν και οι απαιτούμενες σημειώσεις που ενημερώνουν το χρήστη για το ποια είναι η λειτουργία του εκάστοτε τμήματος.

Συνοψίζοντας, η εφαρμογή μας είναι ιδιαίτερα εύχρηστη και παράγει γρήγορα τα αποτελέσματα των μετα-αναλύσεων που θέλουμε, αλλά παράλληλα απευθύνεται σε συγκεκριμένου επιστημονικού πεδίου χρήστες.

Καθημερινά έχουμε νέες έρευνες και τις αντίστοιχες μελέτες τους. Ο επιστημονικός κόσμος αυξάνει τις γνώσεις του σε όλα τα πεδία. Έτσι λοιπόν μια εφαρμογή, όπως η δική μας, θα βοηθήσει το έργο πολλών επιστημόνων που ασχολούνται με τα γενετικά δεδομένα, μιας και αυτού του είδους τα δεδομένα θα βοηθήσουν τον άνθρωπο να βρει λύσεις σε βιολογικά και γονιδιωματικά προβλήματα που μέχρι σήμερα δυσκολεύουν τον άνθρωπο όχι μόνο στην υγεία του, αλλά και στις διαπροσωπικές του σχέσεις.

ΠΑΡΑΡΤΗΜΑ Ι

Πίνακες:

α/α	Τίτλος Πίνακα	Σελίδα
1	Πίνακας με τις τιμές που λαμβάνει ο κάθε γονότυπος ανά μελέτη	3
2	Πίνακας τυπικής διάταξης δεδομένων που χρησιμοποιήθηκαν για Μετα-Ανάλυση γενετικής συσχέτισης	6
3	Πίνακας κατανομής τριών γονοτύπων σε ξεχωριστές λίστες για τους ασθενείς και τους μάρτυρες αντίστοιχα	9
4	Τμήμα κώδικα Stata	15
5	Πίνακας Μετα-Ανάλυσης γενετικών μελετών με συνεχή δεδομένα	15
6	Κώδικας υλοποίησης γραμμικού μοντέλου με τη χρήση της εντολής gllamm	16
7	Κατηγορίες σχεδιαστικών μοντέλων	25
8	Singleton Design Pattern	27
9	Decorator Design Pattern	28
10	Facade Design Pattern	31
11	Iterator Design Pattern	33

ΠΑΡΑΡΤΗΜΑ ΙΙ

Εικόνες:

α/α	Τίτλος Εικόνας	Σελίδα
1	Στατιστικά για τη χρήση των αντίστοιχων λειτουργικών συστημάτων	21
2	Ποσοστιαία χρήση ανά λειτουργικό σύστημα κατά το πρώτο τετράμηνο του 2018	21
3	Δομή του Himler Server	22
4	Τμήμα του PHP κώδικα της εφαρμογής	24
5	Δομή του Singleton Design Pattern	26
6	Δομή του Decorator Design Pattern	30
7	Δομή του Façade Design Pattern	31
8	Δομή του Interator Design Pattern	33
9	Δείγμα HTML και CSS κώδικα της εφαρμογής	36
10	Τμήμα του Javascript κώδικα της εφαρμογής	38
11	Δείγμα χρόνου απόκρισης μεθοδολογιών Prospective και Retrospective Likelihood	40
12	Χρόνοι απόκρισης Retrospective μεθοδολογίας	40
13	Χρόνος απόκρισης για 5, 10 και 20 μελέτες των μεθοδολογιών που χρησιμοποιούν summary δεδομένα	41
14	Χρόνοι απόκρισης για 5, 10 και 20 μελέτες για τις μεθοδολογίες με συνεχή δεδομένα	41
15	Αρχική διεπαφή της εφαρμογής	42
16	Drop-down menu με τις μεθοδολογίες ανά είδος δεδομένων που χρησιμοποιούμε	43
17	Φόρμες για την εισαγωγή των δεδομένων του χρήστη της αντίστοιχης μεθοδολογίας	43
18	Φόρμα upload	44
19	Μορφή αρχείων προς upload	44
20	Τυπική διάταξη αποτελεσμάτων	45

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] R. Mägi and A. P. Morris, “Open Access SOFTWARE GWAMA: software for genome-wide association meta-analysis,” 2010.
- [2] E. Zintzaras and J. P. A. Ioannidis, “HEGESMA: genome search meta-analysis and heterogeneity testing,” *Bioinforma. Appl. NOTE*, vol. 21, no. 18, pp. 3672–3673, 2005.
- [3] A. X. Garg, D. Hackam, and M. Tonelli, “Special Feature: Methods Series Systematic Review and Meta-analysis: When One Study Is Just not Enough,” *Canada Clin J Am Soc Nephrol*, vol. 3, pp. 253–260, 2008.
- [4] G. Biondi-Zoccai, G. Biondi-Zoccai, M. Lotrionte, G. Landoni, and M. G. Modena, “HSR Proceedings in Intensive Care and Cardiovascular Anesthesia,” 2011.
- [5] H. Ab, “Meta-analysis in medical research,” 2010.
- [6] “Bagos-ISCB2014_Abstract_Book.pdf.” .
- [7] M. Borenstein, L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein, “Introduction to Meta-Analysis,” *Psychother. Res. J. Soc. Psychother. Res.*, vol. 19, no. 4–5, p. 421, 2009.
- [8] N. L. Dimou, M. Adam, and P. G. Bagos, “A multivariate method for meta-analysis and comparison of diagnostic tests,” *Stat. Med.*, vol. 35, no. 20, pp. 3509–3523, 2016.
- [9] O. Efthimiou, D. Mavridis, A. Cipriani, S. Leucht, P. Bagos, and G. Salanti, “An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios,” *Stat. Med.*, vol. 33, no. 13, pp. 2275–2287, 2014.
- [10] S. E. Fienberg and W. J. Van Der Linden, *Statistics for Social and Behavioral Sciences*. 2010.
- [11] J. E. Sweat, *PHP|Architect’s Guide to PHP Design Patterns*. 2005.
- [12] D. May, *Introduction to*, no. May. 2009.
- [13] E. Gamma, R. Helm, R. Johnson, and J. Vlissides, “Design Patterns – Elements of Reusable Object-Oriented Software,” *A New Perspect. Object-Oriented Des.*, p. 334, 2002.
- [14] B. J. Becker, “Multivariate Meta-Analysis: Contributions of Ingram Olkin,” *Stat. Sci.*, vol. 22, no. 3, pp. 401–406, 2007.
- [15] G. K. Nikolopoulos, “Statistical Applications in Genetics and Molecular Biology A Method for Meta-Analysis of Case-Control Genetic Association Studies Using Logistic Regression A Method for Meta-Analysis of Case-Control Genetic Association Studies Using Logistic,” vol. 6, no. 1, 2007.
- [16] P. G. Bagos, “Meta-analysis in Stata using gllamm,” *Res. Synth. Methods*, vol. 6, no. 4, pp. 310–332, 2015.
- [17] P. G. Bagos, “A unification of multivariate methods for meta-analysis of genetic association studies,” *Stat. Appl. Genet. Mol. Biol.*, vol. 7, no. 1, p. Article31, 2008.

- [18] Attia, J., Thakkinstian, A. and D'Este, C. (2003) Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. *J Clin Epidemiol*, 56, 297-303.
- [19] Becker, K.G., Barnes, K.C., Bright, T.J. and Wang, S.A. (2004) The genetic association database. *Nat Genet*, 36, 431-432.
- [20] Chen, H.Y. (2003) A note on the prospective analysis of outcome-dependent samples. *J Roy Soc B*, 65, 575-584.
- [21] Cordell, H.J. and Clayton, D.G. (2005) Genetic association studies. *Lancet*, 366, 1121-1131.
- [22] Gu, C., Province, M.A. and Rao, D.C. (2001) Meta-analysis for model-free methods. *Adv Genet*, 42, 255-272.
- [23] Hedeker, D. (2003) A mixed-effects multinomial logistic regression model. *Stat Med*, 22, 1433-1446.
- [24] Hedeker, D. and Gibbons, R.D. (1994) A random-effects ordinal regression model for multilevel analysis. *Biometrics*, 50, 933-944.
- [25] Minelli, C., Thompson, J.R., Abrams, K.R. and Lambert, P.C. (2005a) Bayesian implementation of a genetic model-free approach to the meta-analysis of genetic association studies. *Stat Med*, 24, 3845-3861.
- [26] Minelli, C., Thompson, J.R., Abrams, K.R., Thakkinstian, A. and Attia, J. (2005b) The choice of a genetic model in the meta-analysis of molecular association studies. *Int J Epidemiol*.
- [27] Minelli, C., Thompson, J.R., Tobin, M.D. and Abrams, K.R. (2004) An integrated approach to the meta-analysis of genetic association studies using Mendelian randomization. *Am J Epidemiol*, 160, 445-452.
- [28] Normand, S.L. (1999) Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med*, 18, 321-359.
- [29] Petiti, D.B. (1994) *Meta-analysis Decision Analysis and Cost-Effectiveness Analysis*. Oxford University Press.
- [30] Salanti, G., Higgins, J.P., Trikalinos, T.A. and Ioannidis, J.P. (2006) Bayesian meta-analysis and meta-regression for gene-disease associations and deviations from Hardy-Weinberg equilibrium. *Stat Med*.
- [31] Salanti, G., Sanderson, S. and Higgins, J.P. (2005) Obstacles and opportunities in meta-analysis of genetic association studies. *Genet Med*, 7, 13-20.
- [32] Thakkinstian, A., D'Este, C. and Attia, J. (2004) Haplotype analysis of VDR gene polymorphisms: a meta-analysis. *Osteoporos Int*, 15, 729-734.
- [33] Thakkinstian, A., McElduff, P., D'Este, C., Duffy, D. and Attia, J. (2005) A method for meta-analysis of molecular association studies. *Stat Med*, 24, 1291-1306.
- [34] Thakkinstian, A., McElduff, P., D'Este, C., Duffy, D. and Attia, J. (2005) A method for meta-analysis of molecular association studies. *Stat Med*, 24, 1291-1306.
- [35] Thompson, J.R., Minelli, C., Abrams, K.R., Tobin, M.D. and Riley, R.D. (2005) Meta-analysis of genetic studies using Mendelian randomization--a multivariate approach. *Stat*

- Med, 24, 2241-2254.
- [36] Thompson, S.G. and Higgins, J.P. (2002) How should meta-regression analyses be undertaken and interpreted? *Stat Med*, 21, 1559-1573.
- [37] Thompson, S.G. and Sharp, S.J. (1999) Explaining heterogeneity in metaanalysis: a comparison of methods. *Stat Med*, 18, 2693-2708.
- [38] Zou, G.Y. (2005) Statistical Methods for the Analysis of Genetic association studies. *Ann Hum Genet*, 69, 1-15.