

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ  
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ  
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΑΝΑΛΥΣΗ ΕΠΑΝΑΛΑΜΒΑΝΟΜΕΝΩΝ  
ΜΕΤΡΗΣΕΩΝ: ΕΦΑΡΜΟΓΗ ΣΕ ΜΙΑ  
ΚΛΙΝΙΚΗ ΔΟΚΙΜΗ ΚΑΡΚΙΝΟΥ ΤΟΥ  
ΜΑΣΤΟΥ**

**Κάρεν-Άννα Ι. Τζιγκουνάκη**

*Διπλωματική Εργασία*

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής  
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των  
απαιτήσεων για την απόκτηση του Μεταπτυχιακού  
Διπλώματος Ειδίκευσης στην Εφαρμοσμένη Στατιστική

Πειραιάς  
Αύγουστος 2009

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από τη ΓΣΕΣ του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ ..... συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική

Τα μέλη της Επιτροπής ήταν:

- Μ. Κατέρη (Επιβλέπων)
- Γ. Πιτσέλης
- Γ. Τζαβέλας

Η έγκριση της Διπλωματικής Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνώμων του συγγραφέα.

**UNIVERSITY OF PIRAEUS**



**DEPARTMENT OF STATISTICS  
AND INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN  
APPLIED STATISTICS**

**REPEATED MEASURES ANALYSIS:  
APPLICATION ON A CLINICAL TRIAL  
OF BREAST CANCER**

By  
**Karen-Anna J. Tzigkounaki**

MSc Dissertation

submitted to the Department of Statistics and Insurance  
Science of the University of Piraeus in partial fulfillment of  
the requirements for the degree of Master of Science in  
Applied Statistics.

Piraeus, Greece  
August 2009

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

Στους γονείς μου  
Γιάννη και Μαργαρίτα

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

# Ευχαριστίες

Για την πραγματοποίηση αυτής της εργασίας βοήθησαν πολλοί. Πρώτα από όλα οι γονείς μου, Γιάννης και Μαργαρίτα Τζιγκουνάκη, όπου χωρίς την δική τους βοήθεια και στήριξη, δεν θα είχα πολλές από τις ευκαιρίες που μου έχουν δοθεί ως σήμερα.

Οι καθηγητές μου, Μάρκος Κούτρας και Μαρία Κατέρη, του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης, για την εμπιστοσύνη και υπομονή που έδειξαν μέχρι την παρουσίαση και παράδοση αυτής της εργασίας. Καθώς επίσης και οι καθηγητές Πιτσέλης Γεώργιος και Τζαβέλας Γεώργιος για τις παρατηρήσεις τους.

Ο φίλος μου Χρήστος Τζερεφός, για την επιμονή του και την ηθική στήριξή του όλους αυτούς τους μήνες.

Επίσης θα ήθελα να ευχαριστήσω την Κυρία Ουρανία Δαφνή, καθηγήτρια στο Πανεπιστήμιο Αθηνών, για την ευγενική παραχώρηση των δεδομένων προς ανάλυση.

Όλους αυτούς λοιπόν, τους ανθρώπους, που συντέλεσαν στην διεκπεραίωση αυτής της εργασίας, θα ήθελα να τους ευχαριστήσω θερμά για όλο το κουράγιο και τη βοήθεια που μου προσέφεραν.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ



## Περίληψη

Οι επαναλαμβανόμενες μετρήσεις, πλέον αποτελούν τη δομή δεδομένων πολλών επιστημονικών πεδίων και ιδίως της βιοιατρικής. Για αυτό το λόγο, οι μέθοδοι ανάλυσής τους ιδίως τα τελευταία χρόνια, αναπτύσσονται και εξελίσσονται συνεχώς. Οι μέθοδοι ανάλυσης επικεντρώνονται κυρίως σε συνεχείς μεταβλητές απόκρισης, προσφέροντας έτσι ένα ευρύ φάσμα θεωρίας και λογισμικού. Τα τελευταία χρόνια, οι έρευνες έχουν επεκταθεί και για την περίπτωση των διακριτών και συγκεκριμένα διατάξιμων μεταβλητών απόκρισης.

Στην συγκεκριμένη εργασία, επιχειρείται η ανάλυση δεδομένων επαναλαμβανόμενων μετρήσεων με διατάξιμη μεταβλητή απόκρισης. Πρόκειται για μια μελέτη ενός δείγματος γυναικών που πάσχουν από καρκίνο του μαστού. Αρχικά, γίνεται παρουσίαση της φύσης των δεδομένων καθώς και μια περιγραφική διερεύνηση αυτών. Επίσης, γίνεται αναφορά στη μέθοδο και στο μέσο συγκέντρωσης αυτών των δεδομένων. Στη συνέχεια δίνεται το θεωρητικό υπόβαθρο της ανάλυσης διατάξιμων επαναλαμβανόμενων μετρήσεων, καταλήγοντας στη διατύπωση των βασικών υποθέσεων και των βημάτων ανάλυσης που πρέπει να ακολουθηθούν. Τέλος ακολουθεί η ανάλυση των δεδομένων στο στατιστικό πακέτο SPSS, ενώ συζητούνται και εναλλακτικά λογισμικά που είναι διαθέσιμα.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

## **Abstract**

Repeated measures, constitute the main type of data for many scientific fields and especially biometrics. For this reason, the methodology of their analysis, over the last years has been developed, leading to advanced models and tools. The research methodology mostly aims in the analysis of continuous responses, providing a very wide spectrum of theory and software. The last few decades, research has been extended into discrete responses and mostly ordinal.

In this project, the main goal is to analyze data of repeated measures with an ordinal response. Our application is a study on women with breast cancer. We present the nature of the data and a descriptive analysis of them. There is also, an extended part on the way and the instrument the data were collected. Afterwards, the theoretical part of the analysis of ordinal repeated measures is presented, concluding to the basic assumptions and steps that have to be made. Finally, the analysis of the data using the statistical software SPSS follows, while discussing the use of other alternative available software is discussed.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

# Περιεχόμενα

<i>Ευχαριστίες</i>	<i>vii</i>
<i>Περίληψη</i>	<i>ix</i>
<i>Abstract</i>	<i>xi</i>
<i>Περιεχόμενα</i>	<i>xiii</i>
<i>Εισαγωγή- Παρουσίαση Προβλήματος</i>	<i>1</i>
1.1 Διαχρονικά Δεδομένα	1
1.2 Πειραματικοί Σχεδιασμοί (Between – Within subjects)	2
1.3 Ελλιπείς Παρατηρήσεις	3
1.4 Ελλιπείς Παρατηρήσεις: Τρόποι Αντιμετώπισης	5
1.4.1 Listwise Διαγραφή	5
1.4.2 Διαγραφή κατά ζεύγη	5
1.4.3 Hot deck αντικατάσταση	6
1.4.4 Αντικατάσταση από το μέσο	6
1.4.5 Εκτίμηση Παλινδρόμησης	7
1.4.6 EM Αλγόριθμος	7
1.5 Περιγραφή Δεδομένων	7
1.6 Ερωτηματολόγιο	10
1.6.1 Χρήσεις του EQ-5D	10
Οι χρήσεις του EQ-5D μπορούν να συνοψιστούν στα παρακάτω σημεία:	11
1.6.2 Οδηγίες για ανάλυση και παρουσίαση αποτελεσμάτων του EQ-5D	12
1.6.3 Η Ελληνική απόδοση του EQ-5D	15
1.6.4 Άρθρα και μελέτες με το EQ-5D	16
<i>Εφαρμογή και Ταυτοποίηση Προβλήματος</i>	<i>17</i>
2.1 Περιγραφή της Μελέτης	17
2.2 Περιγραφικά Στατιστικά του Δείγματος	19
2.3 Δεδομένα Προς Ανάλυση	21
<i>Παρουσίαση Μεθόδου Ανάλυσης</i>	<i>23</i>
3.1 Εισαγωγή	23
3.1.1 Περιθώρια Μοντέλα	23
3.1.2 Μοντέλα Τυχαίων Επιδράσεων	24
3.1.3 Μοντέλα Μεταβάσεων	24

3.1.4 Σύγκριση των τριών μεθόδων	25
3.2 Περιθώρια Μοντέλα για Διαχρονικά Δεδομένα	25
3.3 Γενικευμένες Εξισώσεις Εκτιμήσεις (GEE)	27
3.3.1 Γενικά στοιχεία	27
3.3.2 Ιστορική Αναδρομή	29
3.4 Μεθοδολογία GEE	30
3.5 Βασικά Βήματα	31
3.5.1 Κατανομή της μεταβλητής απόκρισης	31
3.5.2 Συνάρτηση σύνδεσης	32
3.5.3 Διακύμανση αποκρίσεων ως συνάρτηση του μέσου	32
3.5.4 Δομή της συσχέτισης της εξαρτημένης μεταβλητής	32
3.5.5 Γενικευμένες Εξισώσεις Εκτίμησης	34
3.5.6 Έλεγχοι Υποθέσεων	37
<i>Διερεύνηση Λογισμικού</i>	39
4.1 SPSS	39
4.2 S-Plus	39
4.3 R	40
<i>Ανάλυση Δεδομένων – Εφαρμογή</i>	43
5.1 Λογισμικό	43
5.2 Κατανομή μεταβλητής απόκρισης	43
5.3 Συνάρτηση σύνδεσης	44
5.4 Δομή πίνακα συνδιακύμανσης	45
5.5 Ανάλυση και ορισμός παραμέτρων	46
5.5.1 Unstructured Working Correlation Matrix	51
5.5.2 Exchangeable Working Correlation Matrix	57
5.5.3 Επιλογή πίνακα συνδιακυμάνσεων	58
5.5.4 Αξιολόγηση Μοντέλου	59
<i>Επίλογος</i>	65
<i>Παράρτημα</i>	67
Πίνακας Δεδομένων	67
Ερωτηματολόγιο για την Υγεία	73
<i>Βιβλιογραφία</i>	75

# Κ Ε Φ Α Λ Α Ι Ο 1

## Εισαγωγή- Παρουσίαση Προβλήματος

Η παρούσα εργασία πραγματεύεται την ανάλυση επαναλαμβανόμενων διατάξιμων δεδομένων. Κύριος άξονας ανάπτυξης της εργασίας είναι η ανάλυση των δεδομένων μιας κλινικής δοκιμής, που είχε ως αντικείμενο τη σύγκριση τριών θεραπειών που εφαρμόστηκαν σε ασθενείς με μεταστατικό καρκίνο του μαστού. Στόχος είναι η μελέτη της επιβίωσης και της ποιότητας ζωής των ασθενών και η σύγκριση των τριών θεραπειών ως προς αυτούς τους τομείς. Οι τρεις θεραπείες συγκρίνονται σε τρεις χρονικές περιόδους: πριν την χημειοθεραπεία, κατά την ολοκλήρωση της χημειοθεραπείας και 6 μήνες μετά την χημειοθεραπεία. Συνεπώς, για την ανάλυση απαιτείται η χρήση μεθόδων ανάλυσης επαναλαμβανόμενων μετρήσεων. Τα δεδομένα μας διατέθηκαν από το τμήμα Νοσηλευτικής του Πανεπιστημίου Αθηνών.

### 1.1 Διαχρονικά Δεδομένα

Τα διαχρονικά δεδομένα προκύπτουν όταν για κάποια υποκείμενα, δίνεται ένα διάλυσμα μετρήσεων που αντιπροσωπεύει την ίδια μεταβλητή, η οποία μετράται στα ίδια άτομα σε διαφορετικές χρονικές στιγμές. Αυτά τα δεδομένα συνδυάζουν χαρακτηριστικά πολυμεταβλητών δεδομένων και δεδομένων χρονοσειρών. Διαφέρουν στο ότι αφορούν μόνο μια μεταβλητή και περιλαμβάνουν πολλές μικρές χρονοσειρές, μια για κάθε υποκείμενο, αντί για μια μεγάλη χρονοσειρά. Τέτοιου είδους δεδομένα, μπορούν να συλλεχθούν είτε προοπτικά (*prospectively*), σχεδιάζοντας τη μελέτη και παρακολουθώντας τα υποκείμενα στο χρόνο που ακολουθεί, είτε αναδρομικά (*retrospectively*), συλλέγοντας μετρήσεις για κάθε υποκείμενο από παρελθοντικές καταγραφές. Τα δεδομένα αυτά είναι γνωστά και ως επαναλαμβανόμενες μετρήσεις (*repeated measures data*), στις κοινωνικές επιστήμες, τα οποία προκύπτουν από διαχρονική μελέτη των υποκειμένων κάτω από διαφορετικές συνθήκες. Η ανάλυση που θα ακολουθήσει αφορά μια κλινική δοκιμή στην οποία η απόκριση καταγράφεται επαναληπτικά και υπάρχουν ελλείπουσες μετρήσεις που οφείλονται είτε σε μη

επίσκεψη του ασθενους για μέτρηση είτε στον θάνατο του ασθενούς, οι οποίες θα αξιολογηθούν σε επόμενο κεφάλαιο. Ο επιστημονικός στόχος αυτής της κλινικής δοκιμής είναι η συγκριτική αξιολόγηση της επίδρασης της κάθε θεραπείας στη διάρκεια του χρόνου.

Η μελέτη επαναλαμβανόμενων μετρήσεων αφορά στην μελέτη της απόκρισης, ενός υποκειμένου, επαναληπτικά. Σε εφαρμογές που αφορούν την υγεία, συνήθως η απόκριση είναι κατηγορική μεταβλητή και πολλές φορές μάλιστα δίτιμη (επιτυχία-αποτυχία θεραπείας). Μια τέτοια μελέτη πραγματοποιείται όταν υπάρχει ανάγκη διαχρονικής παρακολούθησης της πορείας ασθενών, είτε όταν επιλέγεται να γίνει σύγκριση διαφορετικών μεθόδων αντιμετώπισης του ίδιου προβλήματος.

Καθώς οι μετρήσεις πάνω στο ίδιο υποκείμενο παρουσιάζουν συσχέτιση, χρησιμοποιούνται ειδικές στατιστικές μέθοδοι ανάλυσης τέτοιων δεδομένων. Αυτή η συσχέτιση πρέπει να ληφθεί υπόψη για να είναι τα συμπεράσματα της ανάλυσης, έγκυρα και αξιόπιστα. Ιδανικά δεδομένα, θεωρούνται αυτά για τα οποία ο αριθμός των μετρήσεων είναι ίδιος και αυτές πραγματοποιούνται σε ισότιμα χρονικά διαστήματα για όλα τα υποκείμενα της μελέτης. Μια μελέτη με αυτά τα χαρακτηριστικά λέγεται ισορροπημένη (*balanced*). Πολλές φορές όμως και για διάφορους λόγους, ο αριθμός των μετρήσεων για τα υποκείμενα μπορεί να είναι διαφορετικός και σε διαφορετικά χρονικά διαστήματα (*unbalanced*). Τα μη ισορροπημένα δεδομένα μπορεί να προκύψουν λόγω του σχεδιασμού της μελέτης, είτε λόγω ελλιπών μετρήσεων (Everitt, 2006).

## 1.2 Πειραματικοί Σχεδιασμοί (Between – Within subjects)

Σε κάθε διαχρονική μελέτη, τα υποκείμενα, που στην συγκεκριμένη μελέτη είναι ασθενείς, μελετώνται σε παραπάνω από μια διαφορετικές περιστάσεις μέτρησης: θεραπείες, χρονικές στιγμές ... και κατηγοριοποιούνται σε κάθε μέτρηση με βάση μια αποκριτική μεταβλητή. Τα ερωτήματα που αναζητούν απαντήσεις σε τέτοιου είδους μελέτες είναι κάποια από τα παρακάτω,

- Υπάρχουν διαφορές στους υποπληθυσμούς αναλογίζοντας την μέση απόκριση σε κάθε συνθήκη μέτρησης;
- Υπάρχουν διαφορές μεταξύ των συνθηκών μέτρησης, αναλογίζοντας τη μέση κατανομή των υποπληθυσμών;

Επομένως σε μια διαχρονική μελέτη, στόχος είναι η μοντελοποίηση των πιθανών



αποκριτικών προφίλ (*curve, trend*) των υποκειμένων, με βάση κάποια ερμηνευτικά χαρακτηριστικά. Αυτά τα χαρακτηριστικά μπορεί να είναι είτε σταθερά στη διάρκεια της μελέτης για κάθε ασθενή (φύλο), οπότε μετρώνται μόνο κατά την έναρξη της μελέτης, είτε μεταβλητά στο χρόνο (διάθεση, ψυχολογία, πόνος) οπότε μετρώνται σε κάθε χρονική στιγμή της μελέτης. Οι μεταβλητές που παραμένουν σταθερές, διαφοροποιούνται μεταξύ των υποκειμένων (*between-subject*), ενώ οι άλλες διαφοροποιούνται εντός των υποκειμένων (*within-subject*).

Σε διάφορες μελέτες, υποκείμενα επιλέγονται τυχαία από κάποιους πληθυσμούς και εισάγονται πάλι τυχαία σε μία από δύο ή περισσότερες ομάδες θεραπείας. Η σύγκριση των ομάδων, συνεπάγεται την επίδραση της κάθε θεραπείας. Στους *Between-Subjects (B-S)* σχεδιασμούς, οι αποκρίσεις ενός υποκειμένου, εμφανίζονται μόνο σε μια ομάδα. Η μεταβλητότητα μέσα σε κάθε ομάδα, είναι η διαφορά μεταξύ των υποκειμένων και οφείλεται στην τυχαιότητα με την οποία εισήχθη το κάθε υποκείμενο στην ομάδα. Η μεταβλητότητα των μέσων αποκρίσεων μεταξύ των ομάδων, είναι και η διαφορά μεταξύ των υποκειμένων και η διαφορά μεταξύ των θεραπειών.

Όταν ένα υποκείμενο εμφανίζεται σε περισσότερες από μια ομάδες, γίνεται αναφορά σε σχεδιασμούς με επαναλαμβανόμενες μετρήσεις ενός υποκειμένου και λέγονται *Within-Subjects (W-S)*. Αυτοί οι σχεδιασμοί όταν περιλαμβάνουν παρατηρήσεις ενός υποκειμένου πριν από μια θεραπεία, κατά την θεραπεία και μετά την θεραπεία, λέγονται σχεδιασμοί επαναλαμβανόμενων μετρήσεων. Σε αυτούς τους σχεδιασμούς, κάθε υποκείμενο παρατηρείται πριν και μετά τη θεραπεία, με τις παρατηρήσεις αυτές να μην είναι ανεξάρτητες.

Τα υποκείμενα γενικά διαφέρουν μεταξύ τους. Γι'αυτό στους *B-S* σχεδιασμούς αυτές οι διαφορές δεν είναι ελεγχόμενες και αντιμετωπίζονται ως σφάλματα. Στους *W-S*, οι διαφορές των υποκειμένων μπορούν να μετρηθούν και να διαχωριστούν από τα σφάλματα.

### **1.3 Έλλιπείς Παρατηρήσεις**

Σε μια διαχρονική μελέτη, συλλέγεται μια σειρά από παρατηρήσεις για κάθε υποκείμενο. Κάποιες φορές οι μετρήσεις που αντιστοιχούν σε κάθε ασθενή είναι λιγότερες από τις επιθυμητές. Η ανάλυση συνεχών επαναλαμβανόμενων μετρήσεων με ελλείπουσες τιμές έχουν κατηγοριοποιηθεί από πολλούς με βάση τον τρόπο δημιουργίας των ελλειπουσών

τιμών. Τα ελλιπή δεδομένα είναι ένα κομμάτι κάθε έρευνας και πρέπει να προαποφασισθεί πως θα αντιμετωπισθούν. Είναι πολλοί οι λόγοι για τους οποίους μπορεί να υπάρχουν ελλιπή δεδομένα.

Παρατηρήσεις μπορεί να λείπουν λόγω του σχεδιασμού (Little and Rubin, 1987), είτε γιατί για κάποια άτομα οι παρατηρήσεις δεν έλαβαν χώρα. Σε κάποια άρθρα που ασχολούνται με την πολυμεταβλητή ανάλυση (Afifi and Elashoff, 1966, Anderson, 1957, Hartley and Hocking, 1971, Hocking and Smith, 1968, Wilks, 1932) η υπόθεση για την διαδικασία με την οποία προκύπτουν οι ελλείπουσες τιμές, είναι ότι κάθε τιμή των δεδομένων έχει την ίδια πιθανότητα να μην υπάρχει. Σε άλλα άρθρα που ασχολούνται με ανάλυση διακύμανσης (Hartley, 1956, Healy and Westmacott, 1956, Rubin, 1972, Wilkinson, 1958) η υπόθεση είναι ότι τιμές της εξαρτημένης μεταβλητής λείπουν ανεξαρτήτως αυτών που θα παρατηρούνταν. Η βιβλιογραφία επίσης αναφέρεται σε περιπτώσεις που υπάρχουν ελλιπείς παρατηρήσεις εσκεμμένα (Hocking and Smith, 1972, Trawinski and Bargmann, 1964).

Μπορεί να υπάρχουν λόγω ασθένειας του ατόμου, είτε λόγω μη καταχώρησης δεδομένων. Τότε τα δεδομένα λείπουν τελείως τυχαία (*Missing completely at random*, MCAR). Δηλαδή η πιθανότητα να λείπει μια παρατήρηση είναι ανεξάρτητη από την τιμή της μεταβλητής που λείπει ή οποιασδήποτε άλλης. Για παράδειγμα, ελλιπή δεδομένα εισοδήματος δεν θα θεωρούνταν MCAR αν οι άνθρωποι με χαμηλά εισοδήματα ήταν λιγότερο πιθανό να δηλώσουν το εισόδημά τους από ότι άνθρωποι με υψηλά εισοδήματα. Βεβαίως, εάν κάποιος για λόγους υγείας δεν μπορούσε να παρευρεθεί στη συλλογή των δεδομένων τότε η απουσία των παρατηρήσεων θα θεωρούνταν MCAR. Στην περίπτωση αυτού του είδους ελλειπουσών παρατηρήσεων, κάθε στοιχείο των δεδομένων έχει την ίδια πιθανότητα να μην καταγραφεί. Σημαντική παρατήρηση είναι ότι σημαντική είναι η τιμή της παρατήρησης που λείπει και όχι η απουσία της αυτή καθαυτή. Η φύση των δεδομένων με τελείως τυχαία ελλείπουσες τιμές, προσφέρει ανάλυση που παραμένει αμερόληπτη ακόμα και με την απουσία δεδομένων. Μπορεί να χαθεί η δυναμική της ανάλυσης αλλά οι εκτιμώμενες παράμετροι δεν θα είναι μεροληπτικές από την απουσία των δεδομένων.

Κάποιες φορές τα δεδομένα δεν λείπουν τελείως τυχαία και έτσι χαρακτηρίζονται ως δεδομένα που λείπουν τυχαία (*Missing at random*, MAR). Για παράδειγμα, άνθρωποι που είναι σε κατάθλιψη μπορεί να μην θέλουν να δηλώσουν το εισόδημά τους και έτσι το δηλωθέν εισόδημα θα σχετίζεται με την κατάθλιψη. Ακόμα οι άνθρωποι που πάσχουν από κατάθλιψη μπορεί να έχουν γενικότερα χαμηλά εισοδήματα, που σημαίνει ότι αν υπάρχει

μεγάλο ποσοστό ελλιπών δεδομένων εισοδήματος ανθρώπων σε κατάθλιψη, ο μέσος όρος εισοδήματος μπορεί να είναι χαμηλότερος από ότι θα ήταν αν δεν έλειπαν τα δεδομένα. Τα δεδομένα μπορεί να θεωρηθούν ότι λείπουν τυχαία αν πληρούν την προϋπόθεση ότι η απουσία δεν εξαρτάται από την τιμή της μεταβλητής αφού ελέγχετε από κάποια άλλη μεταβλητή (κατάθλιψη). Υπάρχουν τρόποι αντιμετώπισης τέτοιων δεδομένων, παράγοντας αμερόληπτους και σημαντικούς εκτιμητές.

Εάν τα ελλιπή δεδομένα δεν λείπουν τυχαία ή τελείως τυχαία τότε λείπουν μη τυχαία (*Missing not at random*, MNAR). Αυτό συμβαίνει όταν κάποιοι με χαμηλά εισοδήματα είναι λιγότερο πιθανό να δηλώσουν το εισόδημά τους. Τέτοιας φύσης ελλιπή δεδομένα αποτελούν πρόβλημα. Θα πρέπει το μοντέλο που θα παραχθεί, να μοντελοποιεί και την απουσία, κάτι που θα το μετατρέψει σε πολύ περίπλοκο μοντέλο.

## 1.4 Ελλιπείς Παρατηρήσεις: Τρόποι Αντιμετώπισης

Υπάρχουν διάφορες τεχνικές αντιμετώπισης ελλιπών παρατηρήσεων. Παρόλα αυτά όμως η διενέργεια ανάλυσης χρησιμοποιώντας μονάχα τις πλήρεις περιπτώσεις, γίνεται χρησιμοποιώντας γνωστά στατιστικά πακέτα (Carriere, 1998).

### 1.4.1 Listwise Διαγραφή

Ο πιο συνήθης τρόπος αντιμετώπισης ελλειπυσών τιμών είναι η εξαίρεση των περιπτώσεων με ελλιπή δεδομένα και η διενέργεια της ανάλυσης. Έτσι, αν σε μια ομάδα υποκειμένων, για 5 άτομα λείπουν κάποιες παρατηρήσεις, τότε η ομάδα αυτή θα έχει μέγεθος κατά 5 λιγότερα υποκείμενα (*listwise deletion or complete case analysis*). Αν και συχνά μια τέτοια μέθοδος μειώνει σημαντικά το μέγεθος του δείγματος, έχει σημαντικά πλεονεκτήματα. Κάτω από την υπόθεση ότι τα δεδομένα λείπουν τελείως τυχαία, οδηγεί σε αμερόληπτους εκτιμητές. Φυσικά ακόμα και όταν τα δεδομένα λείπουν τελείως τυχαία, μειώνεται η ισχύς της ανάλυσης. Οι εναλλακτικές μέθοδοι που παρατίθενται παρακάτω, θα πρέπει να θεωρηθούν ως αντικαταστάσεις της listwise διαγραφής, αν και σε πολλές περιπτώσεις είναι καλύτερο απλά να εξαιρούνται οι προβληματικές περιπτώσεις (Howell, 2009).

### 1.4.2 Διαγραφή κατά ζεύγη

Πολλά στατιστικά μοντέλα χρησιμοποιούν αυτή τη μέθοδο αντιμετώπισης των

ελλειπουσών τιμών. Κατά τον υπολογισμό του πίνακα συσχέτισης, κάποιο υποκείμενο που έχει ελλιπή δεδομένα σε κάποιον παράγοντα, δεν υπολογίζεται στον υπολογισμό των συσχετίσεων που περιλαμβάνουν αυτόν τον παράγοντα. Αυτό έχει ως αποτέλεσμα οι παράμετροι του μοντέλου να βασίζονται σε διαφορετικές ομάδες δεδομένων, με άλλα πλήθη, και άλλα τυπικά σφάλματα. Αυτά τα προβλήματα μπορεί να «καταστρέψουν» τελείως μια ανάλυση. Προτείνεται, αν υπάρχουν λίγες περιπτώσεις με ελλείπουσες τιμές, να χρησιμοποιείται άφοβα η μέθοδος διαγραφής κατά ζεύγη (*pairwise deletion*), αλλά βέβαια με την ίδια λογική μπορεί κάλλιστα να χρησιμοποιηθεί και η *listwise* διαγραφή.

Για τις επόμενες μεθόδους, που κάποιες φορές χρησιμοποιούνται, είναι σημαντικό να γνωρίζει κανείς κάποια πράγματα ώστε να γίνει σαφές γιατί δεν πρέπει να χρησιμοποιούνται.

### **1.4.3 Hot deck αντικατάσταση**

Παλαιότερα, την δεκαετία του '40 και '50, οι περισσότεροι πολίτες ένιωθαν ότι είναι ευθύνη τους να συμμετέχουν σε έρευνες και έτσι ήταν ελάχιστες οι περιπτώσεις ελλειπουσών τιμών. Όταν υπήρχε κάποια περίπτωση κατά την οποία δεν υπήρχε η σχετική πληροφορία για κάποιο άτομο, γυναίκα ηλικίας 35 που κατοικεί σε συγκεκριμένη περιοχή για παράδειγμα, οι ερευνητές συνέλλεγαν τις μετρήσεις γυναικών ηλικίας 35 που κατοικούσαν στην συγκεκριμένη περιοχή, επέλεγαν τυχαία μία από αυτές και οι μετρήσεις αυτές αντικαθιστούσαν τις ελλιπείς μετρήσεις της έρευνας. Αυτή η μέθοδος αντιμετώπισης ελλειπουσών τιμών (*hot deck imputation*) δεν ήταν τόσο εξωφρενική τότε, καθότι δεν υπήρχαν πολλές περιπτώσεις ελλειπουσών παρατηρήσεων και η επιλογή των παρατηρήσεων «αντικατάστασης» γινόταν τυχαία από ένα σύνολο υποκειμένων με τα ίδια χαρακτηριστικά. Τώρα πια αυτή η μέθοδος δεν χρησιμοποιείται ιδιαίτερα.

### **1.4.4 Αντικατάσταση από το μέσο**

Μια παλαιά μέθοδος ήταν η αντικατάσταση της ελλείπουσας τιμής, από τον μέσο. Αν για παράδειγμα μετράτε η πίεση του αίματος, και για κάποιο άτομο δεν έχει καταγραφεί τιμή, τότε του δίνω την τιμή της μέσης πίεσης του αίματος. Αυτή η μέθοδος (*mean substitution*) δημιουργεί κάποια προβλήματα στην ανάλυση, καθώς ο μέσος όρος δεν θα αλλάξει (αφού χρησιμοποιήθηκε η τιμή του για την συμπλήρωση της τιμής που έλλειπε και δεν προστέθηκε καινούρια πληροφορία) και εφόσον το δείγμα πλέον θα μεγαλώσει, με αποτέλεσμα την υποεκτίμηση του τυπικού σφάλματος.

#### 1.4.5 Εκτίμηση Παλινδρόμησης

Ένας άλλος τρόπος να συμπληρωθεί η τιμή που λείπει, είναι να προβλεφθεί μέσα από μία ανάλυση παλινδρόμησης χρησιμοποιώντας τα στοιχεία τα οποία είναι πλήρη (*regression substitution*). Είναι μία καλύτερη μέθοδος από την προηγούμενη, δίνοντας μια τιμή η οποία έχει υπολογισθεί με βάση κάποια χαρακτηριστικά του υποκειμένου και όχι απλά δίνοντας τον αντίστοιχο μέσο όρο αγνοώντας όλα τα άλλα χαρακτηριστικά του. Βέβαια το πρόβλημα του τυπικού σφάλματος παραμένει. Το στατιστικό πακέτο SPSS έχει βρει έναν τρόπο ώστε να μειωθεί αυτό το πρόβλημα, προσθέτοντας κάποιο σφάλμα τυχαιότητας στην διαδικασία της αντικατάστασης.

#### 1.4.6 EM Αλγόριθμος

Ο αλγόριθμος EM (*Expectation Maximization*) είναι μια μέθοδος για την επεξεργασία ελλειπόντων δεδομένων. Αρχικά αντικαθιστά τις ελλείπουσες τιμές με κατ' εκτίμηση τιμές και εκτιμά τις παραμέτρους. Στη συνέχεια, επανεκτιμά τις ελλείπουσες τιμές υποθέτοντας ότι οι νέες εκτιμήσεις είναι σωστές και τέλος επανεκτιμά τις παραμέτρους. Η διαδικασία αυτή επαναλαμβάνεται έως τη σύγκλιση. Κάθε επανάληψη του αλγορίθμου, αποτελείται από δύο βήματα. Στο πρώτο βήμα (*expectation*) προκύπτει κάθε φορά η αναμενόμενη τιμή του λογαρίθμου της πιθανοφάνειας των πλήρη δεδομένων. Στο δεύτερο βήμα (*maximization*), μεγιστοποιείται η αναμενόμενη τιμή του λογαρίθμου της πιθανοφάνειας των πλήρη δεδομένων. Η διαδικασία επαναλαμβάνεται έως ότου η πιθανοφάνεια που παράγεται σε δυο συνεχόμενες επαναλήψεις, είναι η ίδια (Little and Rubin, 1987).

### 1.5 Περιγραφή Δεδομένων

Οι ασθενείς που μελετήθηκαν ήταν 343 γυναίκες με μεταστατικό καρκίνο του μαστού. Διαχωρίστηκαν με τυχαίο τρόπο σε τρεις ομάδες, όπου η κάθε μια θα λάμβανε ξεχωριστή θεραπεία (χημειοθεραπεία). Οι θεραπείες ήταν: Paclitaxel και Carboplatin (P+Cb), Docetaxel και Gemcitabine (G+Doc) ή Paclitaxel (Pw). Η μεταβλητή 'Θεραπεία' με επίπεδα 1 (για P+Cb), 2 (για G+Doc) και 3 (για Pw) αποτελεί τη βασική ανεξάρτητη μεταβλητή, που όπως θα δούμε παρακάτω στο κεφάλαιο 5 διαμορφώνει τον *between* παράγοντα που διαχωρίζει τα υποκείμενα (ασθενείς) μεταξύ τους.

Καταγράφηκαν μετρήσεις των ασθενών σε τρεις χρονικές στιγμές. Το σύνολο των δεδομένων περιέχει δημογραφικά στοιχεία των ασθενών καθώς και μετρήσεις της ποιότητας ζωής (*Quality of Life*) στις τρεις χρονικές στιγμές με χρήση του δείκτη EQ-5D ([www.euroqol.org](http://www.euroqol.org)).

Οι μεταβλητές των δεδομένων χωρίζονται σε επεξηγηματικές (ανεξάρτητες) και σε μεταβλητές απόκρισης (εξαρτημένες). Επεξηγηματικές είναι οι: ‘Θεραπεία’, ‘Μέτρηση’, ‘Κάπνισμα’, ‘Απασχόληση’, ‘Μετεκπαίδευση’ και ‘Διπλώματα’, ενώ οι μεταβλητές απόκρισης είναι οι: ‘Κινητικότητα’, ‘Αυτοεξυπηρέτηση’, ‘Συνηθισμένες δραστηριότητες’, ‘Πόνος/Δυσφορία’, ‘Άγχος/Θλίψη’ και η ‘Κατάσταση της υγείας’ που αποτελεί και την κύρια μεταβλητή που θα αποτελεί την εξαρτημένη μεταβλητή στην ανάλυση που ακολουθεί. Οι ανεξάρτητες μεταβλητές, καταγράφονται μια φορά κατά την πρώτη μέτρηση, ενώ οι 6 αποκριτικές αποτελούν τις επαναλαμβανόμενες μετρήσεις.

Τα περιγραφικά στοιχεία περιλαμβάνουν πληροφορίες για την ηλικία, το κάπνισμα, την κύρια δραστηριότητα (υπάλληλος ή ελεύθερος επαγγελματίας, συνταξιούχος, οικιακά, σε αναζήτηση εργασίας ή άλλο), την πιθανή κατοχή πτυχίου και τη συνέχιση ή μη της εκπαίδευσης του ασθενή πέραν της υποχρεωτικής εκπαίδευσης. Παρακάτω ακολουθεί ένας πίνακας (ΠΙΝΑΚΑΣ 1-1) με τις μεταβλητές που υπάρχουν για τους ασθενείς, καθώς και μία σύντομη επεξήγησή τους. Ο πλήρης πίνακας των δεδομένων βρίσκεται στο παράρτημα.

Σε κάθε χρονική στιγμή της μελέτης, οι ασθενείς συμπλήρωσαν το ερωτηματολόγιο (EUROQOL EQ-5D) για την κατάσταση της υγείας τους. Το ερωτηματολόγιο αυτό περιγράφεται αναλυτικότερα στο εδάφιο που ακολουθεί.

Στην περίπτωση μας λοιπόν, πρόκειται για ένα πρόβλημα διαχρονικής μελέτης με μονοδιάστατη απόκριση σε 3 χρονικές στιγμές, όχι όμως ισαπέχουσες (η θεραπεία γίνεται σε διαφορετικό αριθμό κύκλων για τον κάθε ασθενή). Στην ανάλυση της συγκεκριμένης εφαρμογής θα χρησιμοποιηθούν οι ασθενείς για τους οποίους έχουμε πλήρη δεδομένα (για τις τρεις χρονικές στιγμές), όπως θα δούμε αναλυτικότερα στο κεφάλαιο 2.

### ΠΙΝΑΚΑΣ 1-1(a)

Περιγραφή δεδομένων (γενικές πληροφορίες)

Μεταβλητές	Επεξήγηση	Τύπος μεταβλητής
<b>rnumber</b>	Κωδικός ασθενή	
<b>aa</b>	A/A	
<b>eligible</b>	Κατάλληλη για την έρευνα (Κατάλληλοι, Μη κατάλληλοι)	Δίτιμη
<b>never_st</b>	Δεν ξεκίνησαν ποτέ την θεραπεία (Έλαβαν θεραπεία, Δεν ξεκίνησαν ποτέ)	Δίτιμη
<b>informda</b>	Ημερομηνία ενημέρωσης	Ημερομηνία
<b>entrydat</b>	Ημερομηνία εισαγωγής στην ομάδα θεραπείας	Ημερομηνία
<b>cyclespa</b>	Συνολικός αριθμός κύκλων θεραπείας	
<b>treatmst</b>	Κατάσταση θεραπείας (Ολοκλήρωση, Μη συνέχιση)	Δίτιμη
<b>last_con</b>	Τελευταία επικοινωνία	Ημερομηνία
<b>dead</b>	Θάνατος (Ναι, Όχι)	Δίτιμη
<b>date_of_</b>	Ημερομηνία θανάτου	Ημερομηνία
<b>cause_of</b>	Αιτία θανάτου (Λόγω της ασθένειας, Όγκος, Τοξικότητα, Λόγω της θεραπείας/Σχετικό με την ασθένεια, Άλλος λόγος)	Κατηγορική (5)

### ΠΙΝΑΚΑΣ 1-1(b)

Περιγραφή δεδομένων (καταγράφηκαν κατά την 1<sup>η</sup> μέτρηση)

Μεταβλητές	Επεξήγηση	Τύπος μεταβλητής
<b>grouptr</b>	Ομάδα θεραπείας με τυχαία ανάθεση (P+Cb, G+Doc, Pw )	Κατηγορική (3)
<b>age</b>	Ηλικία	Συνεχής
<b>sex</b>	Φύλο	Κατηγορική (1)
<b>smoking</b>	Κάπνισμα (Καπνιστής, Πρώην καπνιστής, Δεν υπήρξα ποτέ)	Κατηγορική (3)
<b>occupati</b>	Ποιο από τα παρακάτω περιγράφει καλύτερα την κύρια δραστηριότητά σας (Υπάλληλος ή ελεύθερος επαγγελματίας, Συνταξιούχος, Οικιακά, Σπουδαστής/φοιτητής, Σε αναζήτηση εργασίας, Άλλο)	Κατηγορική (6)
<b>moreeduc</b>	Η εκπαίδευσή σας συνεχίστηκε και μετά την ελάχιστη ηλικία που μπορεί κανείς να σταματήσει το σχολείο? (Ναι, Όχι)	Δίτιμη
<b>certific</b>	Έχετε κάποιο πτυχίο ή ισοδύναμο επαγγελματικό προσόν? (Ναι, Όχι)	Δίτιμη

### ΠΙΝΑΚΑΣ 1-1(c)

Περιγραφή δεδομένων (καταγράφηκαν επαναλαμβανόμενα)

Μεταβλητές	Επεξήγηση	Τύπος μεταβλητής
<b>informno</b>	Χρονική στιγμή ενημέρωσης (1, 2, 3)	Κατηγορική (3)
<b>mobility</b>	Κινητικότητα (Δεν έχω κανένα πρόβλημα, Έχω μερικά προβλήματα, Είμαι καθηλωμένη)	Διατάξιμη(3)
<b>selfcare</b>	Αυτοεξυπηρέτηση (Δεν έχω κανένα πρόβλημα, Έχω μερικά προβλήματα, Είμαι ανίκανη να ντυθώ ή να πλυθώ)	Διατάξιμη(3)
<b>usualact</b>	Συνηθισμένες δραστηριότητες (Δεν έχω κανένα πρόβλημα, Έχω μερικά προβλήματα, Είμαι ανίκανη να εκτελώ τις συνηθισμένες δραστηριότητες)	Διατάξιμη(3)
<b>paindisc</b>	Πόνος/Δυσφορία (Δεν έχω καθόλου, Έχω μέτριο, Έχω υπερβολικό)	Διατάξιμη(3)
<b>anxiousd</b>	Άγχος/Θλίψη (Δεν έχω καθόλου, Έχω μέτριο, Έχω υπερβολικό)	Διατάξιμη(3)
<b>healthst</b>	Κατάσταση της υγείας (0-100)	Συνεχής-Διατάξιμη

## 1.6 Ερωτηματολόγιο

Το ερωτηματολόγιο EuroQol 5-διαστάσεων (EQ-5D) είναι ένα εργαλείο μέτρησης που παρέχει ένα περιγραφικό προφίλ, με βάση πέντε διαστάσεις της υγείας, καθώς και ένα δείκτη της κατάστασης της υγείας του ασθενή. Η ιδιαιτερότητά του είναι ότι συμπληρώνεται από τον ίδιο τον ασθενή. Οι πέντε διαστάσεις της υγείας που μετρώνται με το εργαλείο αυτό είναι η κινητικότητα, η αυτοεξυπηρέτηση, οι συνηθισμένες δραστηριότητες, ο πόνος ή η δυσφορία και το άγχος ή η κατάθλιψη, όπως αξιολογούνται από τον ίδιο τον ασθενή (κανένα πρόβλημα, μερικά προβλήματα, ανικανότητα/σοβαρό πρόβλημα). Τέλος περιλαμβάνεται και μια κλίμακα από το 0 έως το 100 στην οποία ο ασθενής αξιολογεί και σημειώνει την κατάσταση της δικής του υγείας σήμερα, με τιμές κοντά στο μηδέν να ανταποκρίνονται στην χειρίστη πιθανή κατάσταση της υγείας και με τιμές κοντά στο 100 στην βέλτιστη ([www.euroqol.org](http://www.euroqol.org)). Το πλήρες ερωτηματολόγιο βρίσκεται στο Παράρτημα.

### 1.6.1 Χρήσεις του EQ-5D

Το EQ-5D είναι ένα εργαλείο που χρησιμοποιείται για την μέτρηση της κατάστασης της υγείας. Εφαρμόσιμο σε διάφορες καταστάσεις υγείας και θεραπείας, προσφέρει ένα απλό περιγραφικό προφίλ και ένα δείκτη για την κατάσταση της υγείας που μπορεί να χρησιμοποιηθεί στην κλινική και οικονομική αξιολόγηση της θεραπείας. Είναι σχεδιασμένο για την συμπλήρωσή του από τον ίδιο τον ενδιαφερόμενο. Είναι πολύ απλό και απαιτεί μονάχα λίγα λεπτά η συμπλήρωσή του. Αν και είναι σχεδιασμένο για αυτοσυμπλήρωση, έχει χρησιμοποιηθεί σε πολλές έρευνες που γίνονται τηλεφωνικά, είτε με παρατηρητές.

Πρέπει να γίνει διαχωρισμός μεταξύ του EQ-5D αυτοσυμπληρούμενου ερωτηματολογίου, που είναι σχεδιασμένο ως ένα εργαλείο μέτρησης της κατάστασης της υγείας και του EQ-5D ερωτηματολογίου αξιολόγησης που συλλέγει αξιολογήσεις που ορίζονται από το πρώτο.

Το αυτοσυμπληρούμενο EQ-5D ερωτηματολόγιο αποτελείται από 2 σελίδες. Η πρώτη περιλαμβάνει το περιγραφικό σύστημα και η δεύτερη το EQ-VAS. Το περιγραφικό μέρος καταγράφει το επίπεδο προβλημάτων που αντιμετωπίζει ο ασθενής σε 5 διαστάσεις (κινητικότητα, αυτοεξυπηρέτηση, συνηθισμένες δραστηριότητες, πόνος/δυσφορία, και άγχος/θλίψη). Κάθε διάσταση διαχωρίζεται σε 3 επίπεδα (κανένα πρόβλημα, κάποια/μέτρια προβλήματα, ανικανότητα). Αυτές οι πληροφορίες μπορούν να χρησιμοποιηθούν με τους εξής ακόλουθους τρόπους: σαν ένα EQ-5D προφίλ υγείας για άτομα ή ομάδες, είτε σε μια



χρονική στιγμή είτε σε ένα διάστημα χρόνου. Οι διαφορές σε τέτοια προφίλ μπορούν να περιγράψουν την κατάσταση της υγείας ως αποτέλεσμα μιας θεραπείας.

Η κατάσταση της υγείας όπως προκύπτει από αυτές τις 5 διαστάσεις, μπορεί να μετατραπεί σε ένα ζυγισμένο δείκτη υγείας, δίνοντας κάποιο σκορ σε ομάδες παρατηρήσεων. Για παράδειγμα δίνοντας τις τιμές 1,2,3 σε κάθε μία από τις 5 διαστάσεις, μπορεί να δημιουργηθεί ένα μέτρο χρησιμοποιώντας 5-ψήφιους αριθμούς από το 11111 μέχρι το 33333 (συνολικά  $3^5$  ή 243 καταστάσεις). Επομένως το 12121 αντιστοιχεί στην κατάσταση «κανένα κινητικό πρόβλημα, κάποια προβλήματα στο ντύσιμο, κανένα πρόβλημα στις συνηθισμένες δραστηριότητες, κάποιος πόνος και δυσφορία, όχι άγχος και θλίψη».

Το EQ-VAS αποτελεί την βαθμολόγηση της κατάστασης της υγείας από τον ίδιο τον ασθενή, με τιμές από το 0 έως το 100. Τιμές κοντά στο μηδέν αντιστοιχούν στην χειρίστη κατάσταση της υγείας και τιμές κοντά στο 100, στη βέλτιστη.

Το EQ-5D ερωτηματολόγιο αξιολόγησης, είναι η ολοκληρωμένη μορφή του ερωτηματολογίου, περιλαμβάνοντας ακόμη 4 σελίδες. Αυτές οι σελίδες αποτελούν μια μέθοδο αξιολόγησης της υγείας και έχουν χρησιμοποιηθεί σε μελέτες πληθυσμών. Τέτοιες αξιολογήσεις μπορούν να χρησιμοποιηθούν στα αποτελέσματα του περιγραφικού μέρους του EQ-5D ερωτηματολογίου, για την παραγωγή ενός ζυγισμένου δείκτη μέτρησης της υγείας. Υπάρχουν ομάδες τέτοιων τιμών για πολλές χώρες.

Τέλος οι 2 τελευταίες σελίδες περιλαμβάνουν κάποια δημογραφικά ερωτήματα, όπως ηλικία, φύλο, επίπεδο μόρφωσης κ.α. για μια πιο γενική εικόνα του ασθενή. (Brooks, Rabin, and De Charro, 2003).

Οι χρήσεις του EQ-5D μπορούν να συνοψιστούν στα παρακάτω σημεία:

- § Έλεγχος της κατάστασης της υγείας ομάδας ασθενών σε διαφορετικές χρονικές στιγμές, δηλαδή στην παραπομπή, αποδοχή, απαλλαγή και συνέχιση των ασθενών.
- § Αξιολόγηση και έλεγχος της φροντίδας μετρώντας αλλαγές στην κατάσταση της υγείας κάθε ασθενή και ομάδων ασθενών.
- § Αξιολόγηση της σημαντικότητας των συνθηκών σε διαφορετικές χρονικές στιγμές.
- § Προσφορά πληροφορίας σχετικά με την κατανομή των πόρων σε διάφορα επίπεδα.
- § Βοήθεια στην απόδειξη σχετικά με την ιατρική αποτελεσματικότητα σε διαδικασίες όπου το φάρμακο ή η διαδικασία πρέπει να εγκριθεί.
- § Καθορισμός πληθυσμιακών επιπέδων υγείας, τοπικά και εθνικά.

## 1.6.2 Οδηγίες για ανάλυση και παρουσίαση αποτελεσμάτων του EQ-5D

Το EQ-5D χρησιμοποιείται σε πολλά πεδία έρευνας, το κάθε ένα με το δικό του επιστημονικό ενδιαφέρον. Αυτά τα ενδιαφέροντα αντιστοιχούν σε διαφορετικούς τρόπους ανάλυσης, αξιολόγησης και παρουσίασης των αποτελεσμάτων. Κάθε ερευνητής πρέπει να προσδιορίσει το σημείο ενδιαφέροντός του ώστε να αποφασίσει ποιο ή ποια από τα τρία μέρη (EQ-5D, VAS, Index) του ερωτηματολογίου θα χρησιμοποιήσει.

### 1.6.2.1 Πεδία μελέτης

Τρία είναι τα πεδία που χρησιμοποιείται το EQ-5D ευρέως. Κλινικές δοκιμές, οικονομικές μελέτες, και μελέτες υγείας πληθυσμών. Κάθε μία μελέτη επικεντρώνεται στην απάντηση ενός συγκεκριμένου ερωτήματος.

#### § Κλινικές δοκιμές

Οι κλινικές δοκιμές επικεντρώνονται κυρίως στην αποτελεσματικότητα ιατρικών επεμβάσεων. Μετρήσεις ποιότητας ζωής που σχετίζονται με την υγεία (*Health related quality of life* - HRQOL), ολοκληρώνονται σε τέτοιου είδους μελέτες για την καταγραφή παγκόσμιων και γενικών επίπεδων υγείας των ασθενών. Οι μετρήσεις του περιγραφικού μέρους του EQ-5D χρησιμοποιούνται για την περιγραφή της δυναμικής της ποιότητας ζωής του ασθενή κατά τη θεραπεία και την μετέπειτα κατάστασή του (*follow up*).

#### § Οικονομικές μελέτες

Οι οικονομικές μελέτες γίνονται παράλληλα με τις κλινικές δοκιμές. Αυτές οι μελέτες έχουν ως κύριο ενδιαφέρον τη σύγκριση μεταξύ εναλλακτικών ιατρικών παρεμβάσεων. Απαιτείται η κατανομή, μέτρηση και αξιολόγηση, των δεδομένων με τα οποία ξεκινά μια μελέτη και των αποτελεσμάτων της. Τα αποτελέσματα μπορούν να μετρηθούν σε χρηματική αξία (κόστος - κέρδος), σε φυσικές μονάδες, όπως ζωές που σώθηκαν (ανάλυση κόστους - αποτελεσματικότητας) ή σε όρους ωφέλειας (ανάλυση κόστους - ωφέλειας). Πολλοί αναλυτές δεν θέλουν να χρησιμοποιούν την ανάλυση κόστους - ωφέλειας και κάνουν την ανάλυσή τους θεωρώντας την, κόστους - αποτελεσματικότητας. Ανεξαρτήτως της ορολογίας που χρησιμοποιείται, η επίδραση μιας ιατρικής παρέμβασης μετριέται και αξιολογείται από την κατάσταση της υγείας πριν και μετά την παρέμβαση. Η αποτίμηση της υγείας λαμβάνεται ορίζοντας μια μοναδική αριθμητική τιμή μεταξύ των 0 (π.χ. θάνατος) και 1 (π.χ. τελείως υγιής). Το εργαλείο EQ-5D έχει αναπτυχθεί για εφαρμογή σε τέτοιου είδους οικονομικές αναλύσεις.

## § Μελέτες υγείας πληθυσμού

Αυτές οι μελέτες δεν επικεντρώνονται στην αποτελεσματικότητα ή στη σχέση κόστους - αποτελεσματικότητας. Το κύριο ενδιαφέρον τους βρίσκεται στην ποσοτικοποίηση των ασθενειών, στην παρακολούθηση των τάσεων και αλλαγών των ασθενειών και στην σύγκριση αυτών ανά περιοχές και χώρες.

### 1.6.2.2 Συλλογή Δεδομένων

#### § EQ-5D

Σε κλινικές, οικονομικές και πληθυσμιακές μελέτες υγείας, το εργαλείο EQ-5D μπορεί να εφαρμοσθεί για την μέτρηση της κατάστασης της υγείας. Για οικονομικές μελέτες, κόστους – ωφέλειας, χρησιμοποιείται ένα σημαντικό χαρακτηριστικό του EQ-5D, το οποίο είναι ότι η κατάσταση υγείας του ατόμου μπορεί να μετατραπεί σε αριθμό. Παρόλα αυτά υπάρχουν πολλές περιπτώσεις όπου η απλή πληροφορία των αποκρίσεων των 5 διαστάσεων του ερωτηματολογίου προσφέρει αρκετή πληροφόρηση. Σε αυτές τις περιπτώσεις το EQ-5D χρησιμοποιείται σαν ένα εργαλείο γενικής πληροφόρησης της ποιότητας ζωής μέσα από την κατάσταση της υγείας.

#### § Περιγραφικό σύστημα του EQ-5D

Σε αυτό το μέρος του εργαλείου, ο ασθενής καλείται να υποδείξει την κατάσταση της υγείας του σημειώνοντας την πιο κατάλληλη δήλωση για την κάθε διάσταση. Οι 5 διαστάσεις περιλαμβάνουν 3 επίπεδα, δημιουργώντας 243 θεωρητικές καταστάσεις υγείας.

#### § EQ-VAS

Το EQ-VAS προσφέρει μια απλή μέθοδο με την οποία αποκτάται και βαθμολογείται από τον ίδιο τον ασθενή η τρέχουσα κατάσταση της υγείας του. Αυτό γίνεται με μια κλίμακα 20 cm όπου σε συνδυασμό με τις 5 διαστάσεις του περιγραφικού συστήματος, σχηματίζουν ένα προφίλ για τον κάθε ασθενή.

#### § Δείκτης EQ-5D

Μία άλλη μεταβλητή που μπορεί να παραχθεί από το περιγραφικό σύστημα είναι ο δείκτης EQ-5D. Οι αποκρίσεις λοιπόν στις 5 διαστάσεις μπορούν να μετατραπούν σε ένα μοναδικό δείκτη, σταθμίζοντάς τον με βάση κάποιες τιμές που έχουν προέλθει από γενικά δείγματα πληθυσμού.

### 1.6.2.3 Ελλιπή Δεδομένα

Η αξιολόγηση της κατάστασης της υγείας σταδιακά γίνεται ένα αναπόσπαστο κομμάτι πολλών ερευνών. Όμως η δυσκολία στη συλλογή των δεδομένων πολλές φορές είναι σημαντική για την εφαρμογή των αποτελεσμάτων.

Για το EQ-5D υπάρχουν 2 βασικοί τύποι ελλিপών δεδομένων.

1. Η έλλειψη μιας τιμής του ερωτηματολογίου.
2. Η έλλειψη ολόκληρου του ερωτηματολογίου για κάποιο άτομο.

Στη συλλογή δεδομένων επαναλαμβανόμενων μετρήσεων, 3 διαφορετικές εκδοχές υπάρχουν,

- § Καθυστερημένη εισαγωγή στη μελέτη.
- § Ελλιπείς πληροφορίες.
- § Εγκατάλειψη της μελέτης.

Προσπάθειες ελαχιστοποίησης ελλিপών τιμών και των μεροληπτικών στοιχείων που προκαλούν αυτές τις ελλείψεις, είναι σημαντικό κομμάτι το οποίο πρέπει να απασχολεί κάθε έρευνα.

Επειδή το EQ-5D περιγραφικό σύστημα βασίζεται μόνο σε 5 παρατηρήσεις, συνήθως όταν υπάρχει έστω και μια τιμή που λείπει, εξαιρείται τελείως αυτό το ερωτηματολόγιο. Μόνο σε διαχρονικές και επαναλαμβανόμενες μελέτες υπάρχουν κάποιες λύσεις για τις ελλείπουσες τιμές.

### 1.6.2.4 Μέθοδοι Ανάλυσης

Για την ανάλυση δεδομένων διαχρονικών δεδομένων, όπως αυτά προκύπτουν από ερωτηματολόγια σαν το EQ-5D, πολλά και διάφορα στατιστικά μοντέλα είναι διαθέσιμα για την ανάλυση και τον έλεγχο διαφόρων υποθέσεων. Η ανάλυση διακύμανσης (ANOVA) χρησιμοποιείται για τον έλεγχο της μέσης διαφοράς μεταξύ ομάδων. Επίσης, όταν μελετώνται κλινικές δοκιμές, χρησιμοποιείται ευρέως η ανάλυση συνδιακύμανσης (ANCOVA). Για τις περιπτώσεις πολλαπλών μετρήσεων του ίδιου υποκειμένου, είναι απαραίτητη η πολυμεταβλητή ανάλυση διακύμανσης. Παρόμοια μπορεί να χρησιμοποιηθεί ανάλυση παλινδρόμησης με δείκτρια κωδικοποίηση. Καθότι όμως η μέθοδος της κωδικοποίησης, ειδικά στη μέθοδο MANOVA είναι πολύ τεχνική και προϋποθέτει καλή γνώση της θεωρίας, η

χρήση της προτείνεται μόνο για εκπαιδευμένους αναλυτές. Η ανάλυση παλινδρόμησης προσφέρει ένα μοντέλο πρόβλεψης, όπου αυτό ζητείται. Συνήθως μια τέτοια ανάλυση γίνεται με τη μέθοδο ελάχιστων τετραγώνων, αλλά ένας πιο εξελιγμένος τρόπος, που χρησιμοποιείται σε μελέτες ποιότητας ζωής, είναι με τη χρήση πολυεπίπεδου ή ιεραρχικού μοντέλου. Χρησιμοποιώντας λοιπόν ένα τέτοιο μοντέλο μπορούν να προσδιοριστεί καλύτερα η δομή των δεδομένων, και να αντιμετωπιστούν οι ελλείπουσες τιμές.

Η μέτρηση της διαφοράς, σε μελέτες, έχει υπάρξει στο επίκεντρο συζητήσεων και διαφωνίας πολλών άρθρων. Ο κύριος στόχος μιας μελέτης, θεωρείται από πολλούς να είναι η μέτρηση της αλλαγής στην κατάσταση της υγείας. Για πολλά χρόνια υπάρχει διαφωνία για την αποτελεσματικότητα και χρησιμότητα μεταξύ των score διαφοράς και των πραγματικών δεδομένων. Αν και η μέτρηση της διαφοράς μέσω scores και η μεθοδολογία μέσω της οποίας ερμηνεύονται χρησιμοποιείται κατά κόρον, η ανάλυση που βασίζεται σε μετρήσεις πριν και μετά είναι περισσότερο κατάλληλη και στατιστικά πιο ικανή για την αναγνώριση στατιστικά σημαντικών επιδράσεων.

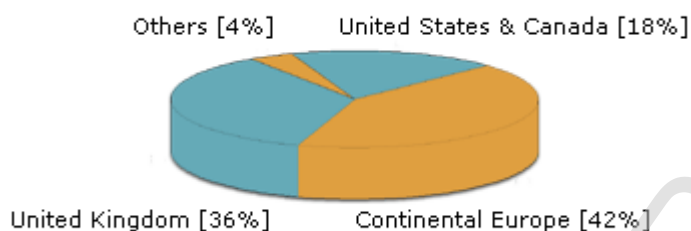
### **1.6.3 Η Ελληνική απόδοση του EQ-5D**

Η ελληνική απόδοση του ερωτηματολογίου EQ-5D, δεν είναι μια απλή μετάφραση του πρωτότυπου εργαλείου EQ-5D. Είναι η ορθή απόδοσή του με βάση ειδική επιτροπή του EuroQoI. Σκοπός ήταν από τη μια μεριά η μετάφρασή του και από την άλλη η διασφάλιση ότι θα είναι κατανοητό, λειτουργικό, έγκυρο, αξιόπιστο και ότι θα πλησιάζει αρκετά παρόμοια συστήματα αξιολόγησης της υγείας όπως αυτά έχουν αναπτυχθεί σε άλλες χώρες ήδη. Μετά από πιλοτική έρευνα σε 30 άτομα και μια έρευνα σε δείγμα 487 ατόμων από όλη την Ελλάδα, διαπιστώθηκε ότι η απόδοση αυτού του εργαλείου στην Ελληνική γλώσσα ήταν επιτυχής, επιτρέποντας την χρήση του σε μελέτες που αποσκοπούν στην μέτρηση και αξιολόγηση της ποιότητας της ζωής (Υφαντόπουλος, 2001).

Τα δύο επόμενα σχήματα απεικονίζουν σε ποιες χώρες χρησιμοποιείται το EQ-5D και σε τι είδη ερευνών. ([www.euroqol.org](http://www.euroqol.org)).

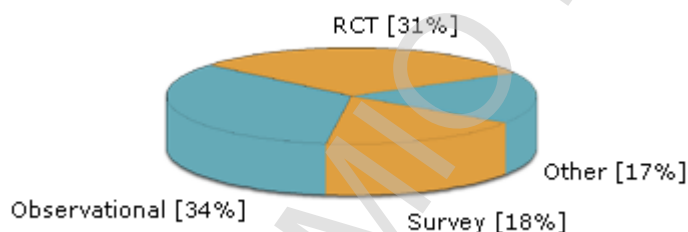
### ΣΧΗΜΑ 1-1(a)

Χώρες που χρησιμοποιείται το EQ-5D



### ΣΧΗΜΑ 1-1(b)

Έρευνες που χρησιμοποιείται το EQ-5D



#### 1.6.4 Άρθρα και μελέτες με το EQ-5D

Σε πολλές μελέτες έχει χρησιμοποιηθεί το συγκεκριμένο εργαλείο, είτε ως απλώς εργαλείο είτε για τη σύγκρισή του με κάποιο άλλο. Επίσης σε πολλά άρθρα αξιολογείται η απόδοσή του σαν εργαλείο μέτρησης της ποιότητας ζωής.

Στην εργασία των Xia, Hwang, Chang, Osenenko, Alejandro, Yan, Toomey and Srinivas, 2005, εξετάστηκε η ισχύς, η αξιοπιστία καθώς και η αποτελεσματικότητά του να εντοπίζει αλλαγές μέσα στο χρόνο. Τα αποτελέσματα έδειξαν ότι το EQ-5D είναι ένα ισχυρό και αξιόπιστο εργαλείο. Μια ακόμα εργασία των Kimman, Dirksen, Lambin and Boersma, 2009, έδειξε ότι αυτό το εργαλείο είναι αποτελεσματικό, διακρίνοντας τις ομάδες ασθενών με βελτίωση και χωρίς.

# Κ Ε Φ Α Λ Α Ι Ο 2

## Εφαρμογή και Ταυτοποίηση Προβλήματος

### 2.1 Περιγραφή της Μελέτης

Από τους 343 ασθενείς που αρχικά επιλέχθηκαν για μελέτη, κατάλληλοι για την εισαγωγή τους στην κλινική δοκιμή και με καταγεγραμμένες τιμές για την κατάσταση της δικής τους υγείας, βρέθηκαν οι 325. Τα κριτήρια για την καταλληλότητά τους, αφορούσαν κυρίως το ιατρικό ιστορικό των γυναικών και την ηλικία τους.

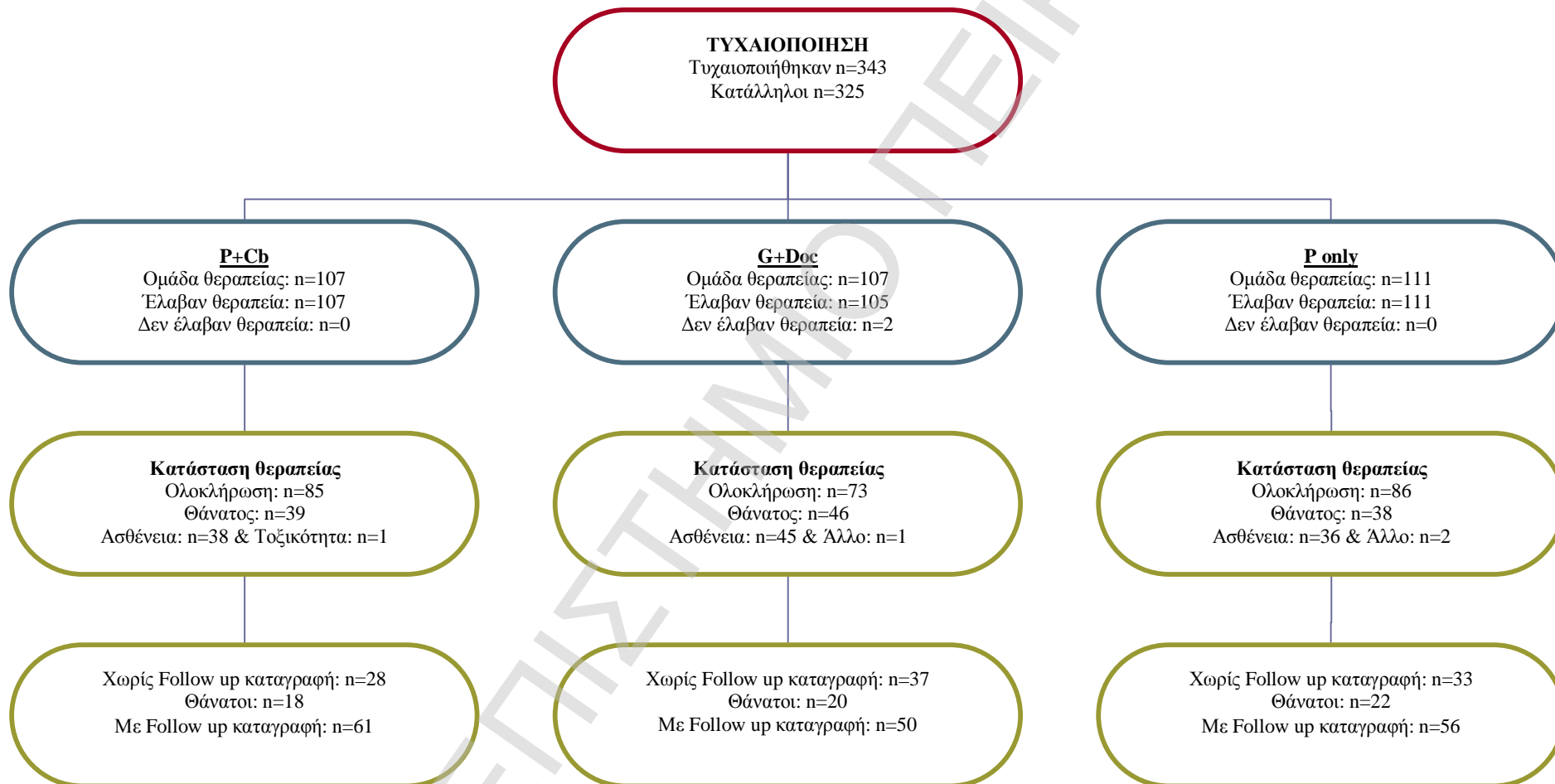
Αυτοί οι ασθενείς διαχωρίστηκαν τυχαία σε τρεις ομάδες, όπου η κάθε ομάδα θα λάμβανε μια ξεχωριστή θεραπεία. Την θεραπεία με τα φάρμακα Paclitaxel και Carboplatin (P+Cb), με τα Docetaxel και Gemcitabine (G+Doc) και με Paclitaxel (P) ακολούθησαν 107, 107 και 111 ασθενείς, αντίστοιχα.

Οι ασθενείς που ολοκλήρωσαν την θεραπεία ήταν 244. Για τους 201 ασθενείς υπάρχουν μετρήσεις και για τις τρεις χρονικές στιγμές. Για τους υπόλοιπους ασθενείς δεν καταγράφηκαν μετρήσεις σε κάποια ή κάποιες από τις τρεις χρονικές στιγμές. Η δυσκολία των κλινικών δοκιμών και κάθε είδους διαχρονικής μελέτης που πραγματοποιείται σε μεγάλο εύρος χρόνου, έγκειται κυρίως στην συλλογή και καταγραφή πληροφορίας ανά συγκεκριμένα χρονικά διαστήματα από τα ίδια άτομα.

Στο παρακάτω Διάγραμμα 2-1 απεικονίζεται η κατανομή των ασθενών. Όπως φαίνεται και από το διάγραμμα υπήρχαν ασθενείς για τους οποίους δεν υπήρχε follow up καταγραφή. Αυτοί εξαιρέθηκαν από τη μελέτη που ακολουθεί. Το τελικό μέγεθος που εξετάζεται αποτελείται από ασθενείς που είχαν παρατηρήσεις και για τις τρεις χρονικές στιγμές (πριν τη θεραπεία, κατά τη διάρκεια της θεραπείας και 6 μήνες follow-up), συμπεριλαμβάνοντας και 55 ασθενείς για τους οποίους επήλθε ο θάνατος.

## ΔΙΑΓΡΑΜΜΑ 2-1

### Κατηγοριοποίηση Δεδομένων





## 2.2 Περιγραφικά Στατιστικά του Δείγματος

Στον Πίνακα 2-1 δίνονται κάποια στατιστικά (περιγραφικά) στοιχεία που αφορούν τους ασθενείς (κατά την εισαγωγή τους στη μελέτη-πριν τη θεραπεία).

**ΠΙΝΑΚΑΣ 2-1**

Κατανομή δημογραφικών στοιχείων

Καπνιστής			Έχετε κάποιο πτυχίο ή ισοδύναμο επαγγελματικό προσόν;		
Καπνίστρια	52	17%	Όχι	224	74%
Πρώην καπνίστρια	53	17%	Ναι	78	26%
Δεν υπήρξα ποτέ	196	64%	Missing	2	1%
Missing	3	1%	Σύνολο	304	100%
Σύνολο	304	100%			

Ποιο από τα παρακάτω περιγράφει καλύτερα την κύρια δραστηριότητά σας			Η εκπαίδευσή σας συνεχίστηκε και μετά την ελάχιστη ηλικία που μπορεί κανείς να σταματήσει το σχολείο;		
Υπάλληλος ή ελεύθερος επαγγελματίας	68	22%	Όχι	191	63%
Συνταξιούχος	66	22%	Ναι	112	37%
Οικιακά	138	45%	Missing	1	0%
Σε αναζήτηση εργασίας	1	0%	Σύνολο	304	100%
Άλλο	26	9%			
Missing	5	2%			
Σύνολο	304	100%			

Σχετικά με το κάπνισμα, οι 105 (34,5%) υπήρξαν ή εξακολουθούν να είναι καπνιστές. Οι 134 ασθενείς (44%) εργάζονται ή είναι συνταξιούχοι και το 45% ασχολείται με τα οικιακά. Οι 78 (26%) έχουν κάποιο πτυχίο ή ισοδύναμο επαγγελματικό προσόν και οι 112 (37%) έχουν συνεχίσει την εκπαίδευσή τους μετά την υποχρεωτική. Από αυτά τα στοιχεία φαίνεται ότι το μορφωτικό επίπεδο των ασθενών δεν είναι ιδιαίτερα υψηλό. Το 22% εργάζεται τη στιγμή έναρξης της μελέτης, κάτι που είναι λογικό καθώς η ηλικία των ασθενών είναι κατά μέσο όρο 58 ετών και κυμαίνεται από 25 μέχρι 82 ετών με τυπική απόκλιση 11,5 και συχνότερη την ηλικία των 57. Το 72% των ασθενών έχει ηλικία άνω των 50 ετών.

Οι θάνατοι που επήλθαν, κατά τη διάρκεια της μελέτης ήταν 122. Οι λόγοι θανάτου ήταν οι εξής: Όγκος (119), Τοξικότητα (1) και άλλος λόγος (2) (ποσοστό θνησιμότητας 37,5%).

Τα στοιχεία αξιολόγησης της ποιότητας ζωής των ασθενών συνοψίζονται στον Πίνακα 2-2 που ακολουθεί. Αξίζει να σημειωθεί ότι για τους περισσότερους ασθενείς τα προβλήματα της καθημερινότητας και ο πόνος μειώθηκαν στη διάρκεια της έρευνας, αντίθετα με το άγχος το

οποίο παρέμεινε, με μια μικρή βελτίωση.

## ΠΙΝΑΚΑΣ 2-2

Κατανομή ασθενών ως προς τις 5 διαστάσεις του EQ-5D και το χρόνο μέτρησης

		Στιγμή συμπλήρωσης ερωτηματολογίου					
		1η φορά		2η φορά		3η φορά	
			%		%		%
Κινητικότητα	Δεν έχω κανένα πρόβλημα στο περπάτημα	193	63,7%	153	50,3%	157	69,8%
	Έχω μερικά προβλήματα στο περπάτημα	104	34,3%	96	31,6%	65	28,9%
	Είμαι καθηλωμένη στο κρεβάτι	6	2,0%	5	1,6%	3	1,3%
	<b>Σύνολο</b>	<b>303</b>		<b>254</b>		<b>225</b>	
Αυτοεξυπηρέτηση	Δεν έχω κανένα πρόβλημα με την αυτοεξυπηρέτησή μου	242	79,9%	192	63,2%	185	82,2%
	Έχω μερικά προβλήματα στο να πλύνω και να ντύνομαι	55	18,2%	58	19,1%	36	16,0%
	Είμαι ανίκανη να πλυθώ ή να ντυθώ	6	2,0%	4	1,3%	4	1,8%
	<b>Σύνολο</b>	<b>303</b>		<b>254</b>		<b>225</b>	
Συνηθισμένες Δραστηριότητες	Δεν έχω κανένα πρόβλημα στο να εκτελώ τις συνηθισμένες δραστηριότητές μου	181	59,7%	141	55,5%	149	66,2%
	Έχω μερικά προβλήματα στο να εκτελώ τις συνηθισμένες δραστηριότητές μου	105	34,7%	99	39,0%	66	29,3%
	Είμαι ανίκανη να εκτελώ τις συνηθισμένες δραστηριότητές μου	17	5,6%	14	5,5%	10	4,4%
	<b>Σύνολο</b>	<b>303</b>		<b>254</b>		<b>225</b>	
Πόνος/Δυσφορία	Δεν έχω καθόλου πόνο ή δυσφορία	126	41,6%	118	38,8%	137	60,9%
	Έχω μέτριο πόνο ή δυσφορία	149	49,2%	124	40,8%	82	36,4%
	Έχω υπερβολικό πόνο ή δυσφορία	28	9,2%	11	3,6%	6	2,7%
	<b>Σύνολο</b>	<b>303</b>		<b>253</b>		<b>225</b>	
Άγχος/Θλίψη	Δεν έχω άγχος ή θλίψη	72	23,7%	72	28,3%	80	35,6%
	Έχω μέτριο άγχος ή θλίψη	159	52,3%	148	58,3%	118	52,4%
	Έχω υπερβολικό άγχος ή θλίψη	73	24,0%	34	13,4%	27	12,0%
	<b>Σύνολο</b>	<b>304</b>		<b>254</b>		<b>225</b>	

Στον Πίνακα 2-3 που ακολουθεί γίνεται μια διερεύνηση της κλίμακας EQ (*Virtual Analogue Scale*). Η ομάδα θεραπείας με το φάρμακο P φαίνεται να παρουσιάζει υψηλότερες τιμές της κλίμακας που σημαίνει καλύτερη κατάσταση υγείας (πάντα με βάση τον ασθενή). Παρόλα αυτά και για τις τρεις ομάδες θεραπείας παρουσιάζεται βελτίωση με το πέρασμα του χρόνου και την εφαρμογή της θεραπείας, κάτι που φαίνεται και από τις τιμές του μέσου όρου της κλίμακας και από τις ελάχιστες τιμές που αυξάνονται στη διάρκεια της έρευνας.

## ΠΙΝΑΚΑΣ 2-3

Περιγραφικά μέτρα αναλογικής κλίμακας ανά ομάδα θεραπείας και μέτρηση

		EQ visual analogue scale		
		Ενημέρωση		
Ομάδα Θεραπείας		Πριν τη θεραπεία	Κατά τη θεραπεία	Follow up
	P+Cb	N	100	87
Mean ± SD		66±21	69±17	74±19
Minimum		10	10	30
Maximum		100	100	100
G+Doc	N	101	79	70
	Mean ± SD	67±20	70±16	75±17
	Minimum	20	30	35
	Maximum	100	100	100
P only	N	103	89	78
	Mean ± SD	72±19	72±18	80±14
	Minimum	10	20	30
	Maximum	100	100	100

## 2.3 Δεδομένα Προς Ανάλυση

Στη φάση αυτή, στην ανάλυση της συγκεκριμένης εφαρμογής θα χρησιμοποιηθούν οι ασθενείς για τους οποίους έχουμε πλήρη δεδομένα (για τις τρεις χρονικές στιγμές), δηλαδή θα περιοριστούμε στους 201 από τους 325 ασθενείς που βρέθηκαν κατάλληλοι για την έρευνα. Στον Πίνακα 2-4 δίνεται μια αρχική περιγραφική διερεύνηση των δεδομένων απόκρισης αυτών των ασθενών ως προς τις 5 μεταβλητές-διαστάσεις που καταγράφονται επαναληπτικά και ως προς τη στιγμή της μέτρησης.

**ΠΙΝΑΚΑΣ 2-4**

Κατανομή ασθενών που θα αναλυθούν ως προς τις 5 διαστάσεις του EQ-5D

		Στιγμή συμπλήρωσης ερωτηματολογίου					
		1η φορά		2η φορά		3η φορά	
			%		%		%
Κινητικότητα	Δεν έχω κανένα πρόβλημα στο περπάτημα	130	64,7%	123	61,2%	134	66,7%
	Έχω μερικά προβλήματα στο περπάτημα	69	34,3%	74	36,8%	63	31,3%
	Είμαι καθηλωμένη στο κρεβάτι	2	1,0%	4	2,0%	3	1,5%
	<b>Σύνολο</b>	<b>201</b>		<b>201</b>		<b>200</b>	
Αυτοεξυπηρέτηση	Δεν έχω κανένα πρόβλημα με την αυτοεξυπηρέτησή μου	159	79,1%	155	77,1%	161	80,1%
	Έχω μερικά προβλήματα στο να πλένομαι και να ντύνομαι	38	18,9%	44	21,9%	35	17,4%
	Είμαι ανίκανη να πλυθώ ή να ντυθώ	3	1,5%	2	1,0%	4	2,0%
	<b>Σύνολο</b>	<b>200</b>		<b>201</b>		<b>200</b>	
Συνηθισμένες Δραστηριότητες	Δεν έχω κανένα πρόβλημα στο να εκτελώ τις συνηθισμένες δραστηριότητές μου	122	60,7%	115	57,2%	126	62,7%
	Έχω μερικά προβλήματα στο να εκτελώ τις συνηθισμένες δραστηριότητές μου	68	33,8%	75	37,3%	64	31,8%
	Είμαι ανίκανη να εκτελώ τις συνηθισμένες δραστηριότητές μου	10	5,0%	11	5,5%	10	5,0%
	<b>Σύνολο</b>	<b>200</b>		<b>201</b>		<b>200</b>	
Πόνος/Δυσφορία	Δεν έχω καθόλου πόνο ή δυσφορία	89	44,3%	101	50,2%	118	58,7%
	Έχω μέτριο πόνο ή δυσφορία	100	49,8%	89	44,3%	76	37,8%
	Έχω υπερβολικό πόνο ή δυσφορία	11	5,5%	10	5,0%	6	3,0%
	<b>Σύνολο</b>	<b>200</b>		<b>200</b>		<b>200</b>	
Άγχος/Θλίψη	Δεν έχω άγχος ή θλίψη	50	24,9%	55	27,4%	70	34,8%
	Έχω μέτριο άγχος ή θλίψη	109	54,2%	119	59,2%	103	51,2%
	Έχω υπερβολικό άγχος ή θλίψη	42	20,9%	27	13,4%	27	13,4%
	<b>Σύνολο</b>	<b>201</b>		<b>201</b>		<b>200</b>	

Από τα δεδομένα αυτά φαίνεται ότι οι περισσότεροι ασθενείς δεν έχουν προβλήματα στη διάρκεια του χρόνου στην κινητικότητα, αυτοεξυπηρέτηση και στην εκπλήρωση των συνηθισμένων δραστηριοτήτων τους. Βεβαίως συνολικά οι περισσότεροι φαίνεται να υποφέρουν από πόνο και άγχος καθ' όλη τη διάρκεια της μελέτης. Για τις δύο διαστάσεις του EUROQOL EQ-5D, κινητικότητα και αυτοεξυπηρέτηση, υπάρχουν πολύ λίγες αποκρίσεις (2%) που αναφέρονται στην τρίτη και χειρίστη κατάσταση (Είμαι καθηλωμένη στο κρεβάτι / Είμαι ανίκανη να πλυθώ ή να ντυθώ). Επιλέχθηκαν λοιπόν αυτοί οι δύο παράγοντες να μετασχηματισθούν σε δίτιμες αποκρίσεις (Δεν έχω πρόβλημα / Έχω κάποια προβλήματα).

Στο δείγμα των 201 ασθενών, έχουν πεθάνει οι 55 ασθενείς. Έτσι, το ποσοστό θνησιμότητας ανέρχεται σε 27,3%. Στον Πίνακα 3-5 παραθέτουμε στοιχεία για τους 55 ασθενείς για τους οποίους επήλθε ο θάνατος. Το προφίλ αυτών των ασθενών είναι κυρίως

γυναίκες που ασχολούνταν με τα οικιακά (53%) με την βασική εκπαίδευση (71%) καθώς και μη καπνίστριες (67%).

## ΠΙΝΑΚΑΣ 2-5

### Περιγραφικά ασθενών που απεβίωσαν

Η εκπαίδευση σας συνεχίστηκε και μετά την ελάχιστη ηλικία που μπορεί κανείς να σταματήσει το σχολείο;

Έχετε κάποιο πτυχίο ή ισοδύναμο επαγγελματικό προσόν;		Ναι	Όχι	Σύνολο
	Ναι	14		14 (25,5%)
	Όχι	2	39	41 (74,5%)
	Σύνολο	16 (29%)	39 (71%)	55

Ποιά απο τα παρακάτω περιγράφει καλύτερα την κύρια δραστηριότητα σας;

	Ασθενείς	%
Υπάλληλος ή ελεύθερος επαγγελματίας	8	15%
Οικιακά	29	53%
Συνταξιούχος	10	18%
Άλλο	7	13%
Άγνωστο	1	2%
Σύνολο	55	100%

Κάπνισμα

	Ασθενείς	%
Δεν υπήρξατε ποτέ καπνίστρια	37	67%
Είστε καπνίστρια	8	15%
Είστε πρώην καπνίστρια	10	18%
Σύνολο	55	100%

# Κ Ε Φ Α Λ Α Ι Ο 3

## Παρουσίαση Μεθόδου Ανάλυσης

### 3.1 Εισαγωγή

Έχουν αναπτυχθεί πολλές μέθοδοι που σχετίζονται με τα γενικευμένα γραμμικά μοντέλα για την ανάλυση επαναλαμβανόμενων μετρήσεων. Αυτές οι μέθοδοι μπορούν να εφαρμοσθούν και για διακριτές και για συνεχείς αποκριτικές μεταβλητές που ακολουθούν Κανονική, Διωνυμική, Poisson ή Γάμμα κατανομή. Τα πλεονεκτήματα αυτών των μεθόδων είναι αρκετά. Κάποια χαρακτηριστικά είναι ότι δεν χρειάζεται για κάθε υποκείμενο ο αριθμός των μετρήσεων που παρουσιάζονται να είναι σταθερός και οι χρονικές στιγμές μέτρησης να είναι ίδιες για όλα τα υποκείμενα. Τρεις είναι οι βασικές μεθοδολογίες που προτείνονται για την ανάλυση επαναλαμβανόμενων μετρήσεων: Περιθώρια μοντέλα (*Marginal Models*), Μοντέλα Τυχαίων επιδράσεων (*Random-effects Models*) και τα Μοντέλα Μεταβάσεων (*Transition Models*) (Davis, 2002).

#### 3.1.1 Περιθώρια Μοντέλα

Έστω ότι  $y_{it}$  είναι η απόκριση τη στιγμή  $t$  του  $i$  υποκειμένου. Στα περιθώρια μοντέλα, η περιθώρια αναμενόμενη τιμή  $m_{it} = E(y_{it})$  υπολογίζεται συναρτήσει των επεξηγηματικών μεταβλητών. Οι υποθέσεις που γίνονται είναι οι εξής,

1. Η περιθώρια αναμενόμενη τιμή  $m_{it}$  σχετίζεται με τις επεξηγηματικές μέσω μιας συνάρτησης σύνδεσης  $g$ ,

$$g(m_{it}) = x'_{it} \beta,$$

όπου  $x'_{it} = (x_{it1}, \dots, x_{itp})$  είναι ένα διάνυσμα συμμεταβλητών για το υποκείμενο  $i$  στο χρόνο  $t$  και  $\beta$  είναι το  $p \times 1$  διάνυσμα άγνωστων παραμέτρων.

2. Η περιθώρια διακύμανση του  $y_{it}$  σχετίζεται με την περιθώρια αναμενόμενη τιμή  $m_{it}$  μέσω της σχέσης,

$$\text{Var}(y_{it}) = fV(\mathbf{m}_i),$$

όπου  $V$  η συνάρτηση διακύμανσης και  $f$  μια γνωστή ή προς εκτίμηση παράμετρος κλίμακας (*dispersion parameter*).

### 3.1.2 Μοντέλα Τυχαίων Επιδράσεων

Στα μοντέλα τυχαίων επιδράσεων η ετερογένεια μεταξύ των υποκειμένων που προέρχεται από μη καταγεγραμμένες μετρήσεις, συνυπολογίζεται περιλαμβάνοντας στο μοντέλο τυχαίες επιδράσεις. Αυτοί οι τυχαίοι παράγοντες περιέχουν όλη την εντός των υποκειμένων (*within-subjects*) συσχέτιση. Οι υποθέσεις που γίνονται είναι οι εξής:

1. Δεδομένου ενός διανύσματος  $\mathbf{b}_i$  σταθερών επιδράσεων για το  $i$  υποκείμενο, ο δεσμευμένος μέσος του  $y_{it}$  ικανοποιεί το μοντέλο,

$$g(E(y_{it} | \mathbf{b}_i)) = \mathbf{x}'_{it} \boldsymbol{\beta} + \mathbf{z}'_{it} \mathbf{b}_i,$$

όπου  $g$  γνωστή συνάρτηση σύνδεσης και  $\mathbf{z}_{it}$  διάνυσμα συμμεταβλητών τη στιγμή  $j$  για το  $i$  υποκείμενο.

2. Τα  $y_{i1}, \dots, y_{it}$  είναι ανεξάρτητα μεταξύ τους δοσμένου του  $\mathbf{b}_i$  για κάθε  $i=1, \dots, n$ .

3. Τα  $\mathbf{b}_1, \dots, \mathbf{b}_n$  είναι ανεξάρτητα και όμοια κατανομημένα με συνάρτηση πυκνότητα πιθανότητας  $f$ .

### 3.1.3 Μοντέλα Μεταβάσεων

Στα μοντέλα μεταβάσεων για την ανάλυση επαναλαμβανόμενων μετρήσεων, οι παρατηρήσεις  $y_{i1}, \dots, y_{it}$  ενός υποκειμένου  $i$ , είναι συσχετισμένες καθώς το  $y_{it}$  εξαρτάται από τις παρελθοντικές τιμές  $y_{i1}, \dots, y_{i,t-1}$ . Οι περασμένες παρατηρήσεις χρησιμοποιούνται σαν μεταβλητές εκτίμησης. Η δεσμευμένη αναμενόμενη τιμή της παρατηρούμενης απόκρισης, δοθέντων των παρελθόντων τιμών, θεωρείται ότι ακολουθεί ένα γενικευμένο γραμμικό μοντέλο. Η γενική μορφή αυτού του μοντέλου είναι,

$$g(E(y_{it} | y_{i1}, \dots, y_{i,t-1})) = \mathbf{x}'_{it} \boldsymbol{\beta} + \sum_{r=1}^s f_r(y_{i1}, \dots, y_{i,t-1}; \mathbf{a}_1, \dots, \mathbf{a}_s),$$

όπου  $g$  είναι η συνάρτηση σύνδεσης,  $f_1, \dots, f_s$  συναρτήσεις προηγούμενων παρατηρήσεων και πιθανόν ενός διανύσματος  $\boldsymbol{\alpha} = (\mathbf{a}_1, \dots, \mathbf{a}_s)'$  άγνωστων παραμέτρων. Η δεσμευμένη

διακύμανση της  $y_{it}$  είναι ανάλογη μιας γνωστής συνάρτησης του δεσμευμένου μέσου,

$$\text{Var}(y_{it} | y_{i1}, \dots, y_{i,t-1}) = fV(E(y_{it} | y_{i1}, \dots, y_{i,t-1})),$$

όπου  $V$  η συνάρτηση διακύμανσης και  $f$  μια γνωστή ή προς εκτίμηση παράμετρος κλίμακας.

### 3.1.4 Σύγκριση των τριών μεθόδων

Στην περίπτωση γραμμικών μοντέλων για συνεχείς αποκρίσεις που ακολουθούν Κανονική κατανομή, οι τρεις μέθοδοι μπορούν να δώσουν εκτιμήσεις των παραμέτρων με ίδια ερμηνεία. Όταν όμως η απόκριση είναι κατηγορική, οι μέθοδοι δίνουν διαφορετικές ερμηνείες των παραμέτρων.

Τα μοντέλα μεταβάσεων αποδίδουν τον δεσμευμένο μέσο του  $y_{it}$  ως συνάρτηση των συμμεταβλητών και περασμένων παρατηρήσεων. Η χρήση τους πρέπει να γίνεται όταν είναι λογική η υπόθεση, ότι οι αποκρίσεις ακολουθούν μια στοχαστική διαδικασία για το κάθε υποκείμενο που εξαρτάται μόνο από τις τιμές των συμμεταβλητών.

Στα μοντέλα τυχαίων επιδράσεων, μοντελοποιείται η ετερογένεια των υποκειμένων. Οι παράμετροι του μοντέλου, επεξηγούν την επιρροή των συμμεταβλητών στην απόκριση ενός υποκειμένου και στην μέση απόκριση του πληθυσμού.

Η ερμηνεία των παραμέτρων στα περιθώρια μοντέλα, σχετίζεται με την επίδραση των συμμεταβλητών στη μέση απόκριση για όλες τις ομάδες του πληθυσμού. Όταν το επιστημονικό ενδιαφέρον είναι η απόκριση ενός υποκειμένου, τότε κατάλληλα είναι τα μοντέλα τυχαίων επιδράσεων, ενώ όταν είναι η μέση απόκριση του πληθυσμού, καταλληλότερο είναι ένα περιθώριο μοντέλο. Για παράδειγμα σε μια κλινική δοκιμή, όπου το ενδιαφέρον είναι η ανωτερότητα μιας θεραπείας έναντι καμιάς θεραπείας (*control*), καταλληλότερο είναι ένα μοντέλο περιθωρίων κατανομών (Davis, 2002).

## 3.2 Περιθώρια Μοντέλα για Διαχρονικά Δεδομένα

Τα περιθώρια μοντέλα δεν απαιτούν υποθέσεις για την κατανομή της απόκρισης, αλλά για την μέση απόκριση.

Εστω ότι  $N$  υποκείμενα παρακολουθούνται επαναληπτικά.  $Y_{ij}$  θα είναι η απόκριση του  $i$ -στου υποκειμένου, την  $j$ -στη χρονική στιγμή. Δεν θεωρούμε ότι όλα τα υποκείμενα έχουν τον ίδιο αριθμό μετρήσεων, ούτε ότι οι μετρήσεις έγιναν στις ίδιες χρονικές στιγμές για όλα

τα υποκείμενα. Επομένως θεωρούμε ότι υπάρχουν  $n_i$  μετρήσεις για κάθε υποκείμενο και ότι κάθε μέτρηση  $Y_{ij}$  πραγματοποιείται τη στιγμή  $t_{ij}$ .

Οι μετρήσεις της απόκρισης για το  $i$ -στο υποκείμενο, μπορούν να παρουσιαστούν σαν ένα  $n_i \times 1$  διάνυσμα,

$$Y_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \mathbf{M} \\ Y_{in_i} \end{pmatrix}, \quad i = 1, \dots, N,$$

όπου τα διανύσματα των αποκρίσεων θεωρούνται ανεξάρτητα μεταξύ τους (αλλά οι μετρήσεις για κάθε υποκείμενο δεν θεωρούνται ανεξάρτητες). Για κάθε απόκριση  $Y_{ij}$ , υπάρχει ένα διάνυσμα  $p \times 1$  συμμεταβλητών,

$$X_{ij} = \begin{pmatrix} X_{ij1} \\ X_{ij2} \\ \mathbf{M} \\ X_{ijp} \end{pmatrix}, \quad i = 1, \dots, N \text{ και } j = 1, \dots, n_i.$$

Κάθε υποκείμενο έχει ένα διάνυσμα συμμεταβλητών  $X_{ij}$  που σχετίζεται με την απόκριση σε κάθε χρονική στιγμή  $Y_{ij}$ . Το διάνυσμα  $X_{ij}$ , μπορεί να έχει μεταβλητές που δεν αλλάζουν από τη μια χρονική στιγμή στην άλλη (*between-subject covariates*), όπως το φύλο ή η θεραπεία, και μεταβλητές που αλλάζουν σε κάθε χρονική στιγμή (*within-subject covariates*).

Τα διανύσματα των συμμεταβλητών μπορούν να ομαδοποιηθούν σε ένα  $n_i \times p$  πίνακα συμμεταβλητών,

$$X_i = \begin{pmatrix} X'_{i1} \\ X'_{i2} \\ \mathbf{M} \\ X'_{in_i} \end{pmatrix} = \begin{pmatrix} X_{i11} & X_{i12} & \mathbf{L} & X_{i1p} \\ X_{i21} & X_{i22} & \mathbf{L} & X_{i2p} \\ \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ X_{in_i1} & X_{in_i2} & \mathbf{L} & X_{in_i p} \end{pmatrix}, \quad i = 1, \dots, N,$$

όπου οι σειρές του  $X_i$  αντιστοιχούν στις συμμεταβλητές που σχετίζονται με τις  $n_i$  διαφορετικές μετρήσεις και οι στήλες αντιστοιχούν σε κάθε μία από τις  $p$  συμμεταβλητές.

Τα περιθώρια μοντέλα μοντελοποιούν ξεχωριστά τη μέση απόκριση και την εντός των υποκειμένων συσχέτιση. Τρεις είναι οι προσδιορισμοί που πρέπει να γίνουν,

1. Ο μέσος της κάθε απόκρισης  $E(Y_{ij}/X_{ij}) = m_{ij}$  θεωρείται ότι εξαρτάται από τις



συμμεταβλητές, μέσω μιας γνωστής συνάρτησης σύνδεσης (*link function*),

$$g(\mathbf{m}_{ij}) = \mathbf{h}_{ij} = \mathbf{X}'_{ij} \mathbf{b}$$

2. Η υπό συνθήκη διακύμανση κάθε  $Y_{ij}$ , δοσμένων των συμμεταβλητών, θεωρείται ότι εξαρτάται από το μέσο, από τη σχέση,

$$\text{Var}(Y_{ij}) = fu(\mathbf{m}_{ij})$$

όπου  $u(\mathbf{m}_{ij})$  είναι γνωστή συνάρτηση διακύμανσης και  $f$  είναι μια παράμετρος κλίμακας (γνωστή ή προς εκτίμηση).

3. Η υπό συνθήκη εντός των υποκειμένων συσχέτιση στο διάνυσμα των επαναλαμβανόμενων αποκρίσεων, δοσμένων των συμμεταβλητών, θεωρείται μια συνάρτηση μιας ομάδας παραμέτρων  $\mathbf{a}$  (που επίσης εξαρτάται από τους μέσους  $\mathbf{m}_{ij}$ ).

Τα δύο πρώτα σημεία δεν κάνουν καμία υπόθεση για την κατανομή των παρατηρήσεων απόκρισης. Στο τρίτο σημείο είναι που φαίνεται η προσαρμογή των γενικευμένων γραμμικών μοντέλων σε διαχρονικά δεδομένα. Η αποφυγή υποθέσεων για την κατανομή των αποκρίσεων οδηγεί σε μια μέθοδο εκτίμησης που είναι γνωστή ως γενικευμένες εξισώσεις εκτίμησης (*Generalized Estimating Equations- GEE*), (Fitzmaurice, Laird and Ware, 2004), (Fitzmaurice, Davidian, Verbeke and Molenberghs, 2008).

### 3.3 Γενικευμένες Εξισώσεις Εκτιμήσεις (GEE)

#### 3.3.1 Γενικά στοιχεία

Στην περίπτωση δεδομένων που αφορούν στην καταλληλότητα ή ανωτερότητα μιας θεραπείας έναντι μιας άλλης (ή άλλων), οι παρατηρήσεις δεν είναι απαραίτητα ανεξάρτητες. Υπάρχει κάποια συσχέτιση που προκύπτει συνήθως από την φύση των δεδομένων (π.χ. παρατηρήσεις πάνω στο ίδιο υποκείμενο αλλά σε διαφορετικές χρονικές στιγμές).

Οι συσχετισμένες λοιπόν μετρήσεις, χωρίζονται σε συστάδες (*clusters*). Η κάθε ομάδα περιλαμβάνει τις μεταξύ τους συσχετισμένες παρατηρήσεις, ενώ οι ομάδες μεταξύ τους είναι ανεξάρτητες. Αυτές οι ομάδες μπορούν να περιλαμβάνουν είτε μετρήσεις πάνω στο ίδιο άτομο σε διαφορετικές όμως συνθήκες (ή χρονικές στιγμές) μέτρησης, είτε να περιλαμβάνουν

ένα σύνολο υποκειμένων που παρουσιάζουν μια συγκεκριμένη ομοιογένεια μεταξύ τους ως προς ένα συγκεκριμένο χαρακτηριστικό που τα διαφοροποιεί από τα υπόλοιπα υποκείμενα. Η ανάλυση τέτοιας φύσεως δεδομένων αφορά στην συμπεριφορά των περιθωρίων κατανομών αυτών των μετρήσεων. Για διαχρονικά δεδομένα, τα περιθώρια μοντέλα διαφοροποιούν την μοντελοποίηση στην μεταξύ των υποκειμένων (*between-subjects*) όπου βασίζεται στην περιθώρια μέση τιμή και στην εντός των υποκειμένων (*within-subjects*), όπου βασίζεται στις συνδιακυμάνσεις των μετρήσεων (Ballinger, 2004).

Συνήθως το επιστημονικό ενδιαφέρον επικεντρώνεται είτε στο ύψος της αλλαγής της απόκρισης με το πέρασμα του χρόνου είτε στην εξάρτηση της απόκρισης από τις επεξηγηματικές μεταβλητές. Έχοντας μια παρατήρηση για κάθε υποκείμενο, η εφαρμογή ενός γενικευμένου γραμμικού μοντέλου (*generalized linear model*) θα έδινε μια περιγραφή της σχέσης της απόκρισης (συνεχής ή διακριτής) με τις επεξηγηματικές μεταβλητές. Έχοντας όμως επαναλαμβανόμενες μετρήσεις για κάθε υποκείμενο, πρέπει να ληφθεί υπόψη και η συσχέτιση των μετρήσεων του ίδιου υποκειμένου. Υπάρχουν διάφοροι τρόποι για την ανάλυση επαναλαμβανόμενων μετρήσεων, αναλόγως με την φύση της αποκριτικής μεταβλητής. Ειδικότερα για δεδομένα που ακολουθούν Κανονική κατανομή έχουν αναπτυχθεί διάφορες μέθοδοι. Ενώ για μη κανονικά, υπάρχουν λιγότερες μέθοδοι. Συγκεκριμένα για δίτιμες αποκρίσεις (*binary*), έχουν προταθεί μοντέλα με συνάρτηση σύνδεσης (*link function*) probit, logit και log γραμμικά μοντέλα. Ένας τρόπος αντιμετώπισης διαχρονικών δεδομένων, είναι με τη χρήση γενικευμένου γραμμικού μοντέλου για την περιθώρια κατανομή της απόκρισης  $y_{it}$ . Δηλαδή, με εξισώσεις εκτίμησης που δίνουν εκτιμήσεις για τις παραμέτρους παλινδρόμησης και των διακυμάνσεών τους κάτω από όχι πολύ αυστηρές υποθέσεις για την κοινή κατανομή. Η μέθοδος αυτή καταλήγει στην μέγιστη πιθανοφάνεια όταν η απόκριση ακολουθεί κάποια κατανομή από την Εκθετική Οικογένεια Κατανομών.

Αγνοώντας και περιλαμβάνοντας στην ανάλυση την συσχέτιση των αποκρίσεων, οδηγεί σε λανθασμένες εκτιμήσεις των παραμέτρων. Οι εκτιμητές του μοντέλου είναι λιγότερο κατάλληλοι και αυτό γιατί είναι περισσότερο διεσπαρμένοι γύρω από την πραγματική τιμή του πληθυσμού από ότι θα ήταν αν είχε ληφθεί υπόψη και η εντός των υποκειμένων (*within-subjects*) συσχέτιση. Ένα εργαλείο λοιπόν το οποίο προσφέρει την αξιολόγηση αποκρίσεων που προέρχονται από διαχρονικές μελέτες και προσφέρει ασυμπτωτικά συνεπείς και αμερόληπτες (*unbiased*) εκτιμήσεις παραμέτρων χωρίς να χρειάζεται η μετατροπή των

δεδομένων, είναι οι γενικευμένες εξισώσεις εκτίμησης (*generalized estimating equations*). Αυτή η μέθοδος ανάλυσης επαναλαμβανόμενων μετρήσεων, βασίζεται στα περιθώρια μοντέλα (Liang and Zeger, 1986) και αποτελεί επέκταση της quasi - likelihood μεθόδου για διαχρονικά δεδομένα. Η μέθοδος είναι ημι -παραμετρική, καθώς δεν προσδιορίζεται πλήρως η από κοινού κατανομή των παρατηρήσεων.

Η ανάλυση βασίζεται σε υποθέσεις για τους μέσους και τις διακυμάνσεις των παρατηρήσεων, όπως και στα γενικευμένα γραμμικά μοντέλα, με τη διαφορά ότι αντί να χρειάζεται να γίνονται υποθέσεις για τη συσχέτιση μεταξύ των μετρήσεων μιας μονάδας, να γίνεται χρήση ενός χρησιμοποιούμενου πίνακα συσχετίσεων (Crowder, 1995).

### 3.3.2 Ιστορική Αναδρομή

Πριν από τα άρθρα των Liang and Zeger, 1986, οι στατιστικές μέθοδοι για την ανάλυση διακριτών διαχρονικών δεδομένων βρίσκονταν στο περιθώριο σε σχέση με τις μεθόδους για συνεχή δεδομένα. Τα πρώτα θεμέλια για την ανάπτυξη μεθόδων ανάλυσης και εκτίμησης περιθωρίων μοντέλων επαναλαμβανόμενων μετρήσεων, έγιναν από τους Grizzle, Starmer and Koch, 1969. Αν και αρχικά αυτή ήταν μια γενική μέθοδος ανάλυσης κατηγορικών δεδομένων, αργότερα διαπιστώθηκε ότι μπορεί να επεκταθεί για την ανάλυση επαναλαμβανόμενων μετρήσεων (Koch and Reinfurt 1971, Koch et al. 1977). Βασίζεται σε μια προσέγγιση σταθμισμένων ελαχίστων τετραγώνων (*weighted least squares: WLS*) με ελάχιστες υποθέσεις για την εντός των υποκειμένων σχέση των επαναλαμβανόμενων μετρήσεων. Τότε προσέφερε μια εκτίμηση για διάφορα μοντέλα επαναλαμβανόμενων κατηγορικών δεδομένων. Παρόλα αυτά υπήρχαν κάποιοι περιορισμοί για τη χρήση της. Η μέθοδος απαιτούσε όλες οι συμμεταβλητές να είναι κατηγορικές. Επίσης το υπό εξέταση δείγμα θα έπρεπε να είναι αρκετά μεγάλο, δηλαδή θα έπρεπε ο αριθμός των επαναλήψεων να μην είναι πολύ μεγάλος, ο αριθμός των κατηγοριών επίσης μικρός και ο αριθμός των υποκειμένων όσο μεγαλύτερος γινόταν ώστε να υπάρχει αντιπροσωπευτικό δείγμα σε κάθε κατηγορία. Αυτό, γιατί έπρεπε οι πολυωνυμικές πιθανότητες που θα έπρεπε να εκτιμηθούν να είναι όσο το δυνατόν λιγότερες. Αν υπήρχαν 4 κατηγορίες για την απόκριση και 6 επαναλήψεις θα έπρεπε να εκτιμηθούν  $4^6 - 1 = 4.095$  πολυωνυμικές πιθανότητες. Τέλος για να μπορέσει να εφαρμοσθεί αυτή η μέθοδος σε διαχρονική μελέτη, θα έπρεπε τα δεδομένα να είναι ισορροπημένα χρονικά, δηλαδή όλα τα υποκείμενα να έχουν μετρήσεις σε ίσα χρονικά διαστήματα.

Πλέον οι μέθοδοι εκτίμησης περιθωρίων μοντέλων μέσω πιθανοφάνειας, εξαλείφουν κάποιους από τους παραπάνω περιορισμούς, χωρίς όμως να επιτρέπουν μεγάλο αριθμό επαναλήψεων. Το 1986 οι Liang και Zeger εισήγαγαν την μέθοδο γενικευμένων εξισώσεων εκτίμησης (GEE), με την οποία οι ερευνητές και αναλυτές μπορούσαν πλέον να προσεγγίσουν με τον καλύτερο τρόπο την ανάλυση διακριτών διαχρονικών δεδομένων (Fitzmaurice, Davidian, Verbeke and Molenberghs, 2008).

### 3.4 Μεθοδολογία GEE

Αυτό το εργαλείο είναι κατάλληλο καθώς λαμβάνει υπόψη την εντός των υποκειμένων συσχέτιση των αποκρίσεων και μπορεί να εφαρμοσθεί για αποκρίσεις που δεν ακολουθούν την Κανονική κατανομή. Οι γενικευμένες εξισώσεις εκτίμησης δίνουν εκτιμητές παραμέτρων με δειγματικές κατανομές ασυμπτωτικά κανονικές και μπορούν να εφαρμοσθούν για την ερμηνεία κύριων επιδράσεων και αλληλεπιδράσεων και επίσης για την αξιολόγηση κατηγορικών ή συνεχών ανεξάρτητων μεταβλητών. Η μέθοδος αυτή των γενικευμένων εξισώσεων εκτίμησης ξεκινάει με την εκτίμηση, μέσω εκτιμητών μέγιστης πιθανοφάνειας, των παραμέτρων  $\beta$  και τον υπολογισμό της διακύμανσης μέσω μιας συνάρτησης σύνδεσης η οποία επιτρέπει στην περιθώρια μέση τιμή να μπορεί να εκφρασθεί ως ένα διάνυσμα των εκτιμώμενων παραμέτρων (Fitzmaurice, Laird and Ware, 2004).

1. Η περιθώρια αναμενόμενη τιμή της απόκρισης  $E(Y_{ij}) = m_{ij}$ , εξαρτάται από τις συμμεταβλητές, μέσω μιας συνάρτησης σύνδεσης,

$$g(m_{ij}) = h_{ij} = X'_{ij} b$$

2. Η διακύμανση του κάθε  $Y_{ij}$ , δεδομένων των συμμεταβλητών εξαρτάται από το μέσο σύμφωνα με τη σχέση,

$$Var(Y_{ij}) = fu(m_{ij}),$$

όπου  $u(m_{ij})$  είναι γνωστή συνάρτηση διακύμανσης ενώ  $f$  είναι μια παράμετρος κλίμακας (γνωστή ή προς εκτίμηση), η οποία αποκαλείται και παράμετρος υπερ-διασποράς (overdispersion parameter) και εισάγεται σε ένα μοντέλο με σκοπό την μείωση της υπερ-διασποράς, δηλαδή της διαφοράς της παρατηρούμενης διασποράς από την εκτιμώμενη.

3. Η ανά δύο (*pairwise*) εντός των υποκειμένων συσχέτιση από το διάνυσμα των επαναλαμβανόμενων αποκρίσεων, δεδομένων των μεταβλητών, θεωρείται ως μια συνάρτηση των μέσων  $m_{ij}$  και μιας ομάδας παραμέτρων (εντός των υποκειμένων συσχέτισης)  $a$ . Για παράδειγμα, αν το διάνυσμα των παραμέτρων  $a$  αντιπροσωπεύει τις ανά δυο συσχετίσεις των αποκρίσεων, οι συνδιακυμάνσεις των αποκρίσεων εξαρτώνται από τα  $m_{ij}(X'_{ij}b), f$  και  $a$ . Έτσι, δίνοντας ένα μοντέλο για τις ανά δύο συσχετίσεις, ο πίνακας συνδιακυμάνσεων υπολογίζεται σαν το γινόμενο τυπικών αποκλίσεων και συσχετίσεων,

$$V_i = (A_i)^{1/2} \text{Corr}(Y_i) (A_i)^{1/2},$$

Όπου,  $A_i$  είναι ένας διαγώνιος πίνακας με  $\text{Var}(Y_{ij}) = f u(m_{ij})$  στην διαγώνιο (ο  $(A_i)^{1/2}$  είναι ένας διαγώνιος πίνακας με τις τυπικές αποκλίσεις  $\sqrt{f u(m_{ij})}$  στην διαγώνιο) και  $\text{Corr}(Y_i)$  είναι ο πίνακας συσχετίσεων (συνάρτηση των  $a$ ). Ο  $V_i$  είναι γνωστός και ως χρησιμοποιούμενος πίνακας συνδιακυμάνσεων (*working correlation matrix*) για να διαχωρίζεται από την πραγματική συνδιακύμανση των  $Y_i$ .

### 3.5 Βασικά Βήματα

#### 3.5.1 Κατανομή της μεταβλητής απόκρισης

Οι γενικευμένες εξισώσεις εκτίμησης επιτρέπουν κατανομές από την Εκθετική Οικογένεια Κατανομών (*Exponential family of distributions*), όπως η Κανονική (*Normal*), η Inverse Κανονική, η Διωνυμική (*Binomial*), η αρνητική Διωνυμική (*Negative Binomial*), η Poisson και η Γάμμα (*Gamma*). Όπως και στα γενικευμένα γραμμικά μοντέλα, η διακύμανση πρέπει να εκφρασθεί ως συνάρτηση του μέσου. Είναι πολύ σημαντικό να ορισθεί σωστά η συνάρτηση σύνδεσης και πρώτα από όλα να αναγνωρισθεί σωστά η κατανομή των δεδομένων. Παρόλα αυτά δεν χρειάζεται μεγάλη αυστηρότητα στον ορισμό των συναρτήσεων διακύμανσης, ώστε οι εκτιμητές των παραμέτρων να έχουν δειγματική κατανομή ασυμπτωτικά κανονική. Ένας γενικός κανόνας είναι ότι αν οι αποκρίσεις είναι δίτιμες, πρέπει να ορίζεται Διωνυμική κατανομή. Αν τα δεδομένα είναι μετρήσιμα, τότε επιλέγεται Poisson. Όταν τα δεδομένα είναι διατάξιμα πολυωνυμικά τότε χρησιμοποιούνται

πολυωνυμικές κατανομές (*multinomial*) (Ballinger, 2004).

### 3.5.2 Συνάρτηση σύνδεσης

Για να μοντελοποιηθεί η αναμενόμενη τιμή της περιθώριας απόκρισης του πληθυσμού  $E(Y_{ij}) = m_{ij}$ , ως γραμμικός συνδυασμός των συμεταβλητών, πρέπει να ορισθεί μια συνάρτηση σύνδεσης που θα επιτρέπει στην εξαρτημένη μεταβλητή να εκφράζεται ως ένα διάνυσμα των παραμέτρων ( $\mathbf{b}$ ),

$$g(\mathbf{m}_{ij}) = \mathbf{h}_{ij} = \mathbf{X}'_{ij} \mathbf{b} .$$

Η βασική συνάρτηση σύνδεσης είναι η ταυτοτική (*identity*), η οποία χρησιμοποιείται για δεδομένα που ακολουθούν Κανονική κατανομή. Βασικά, η κατανομή των δεδομένων είναι που οδηγεί και μειώνει τις επιλογές της σωστής συνάρτησης σύνδεσης. Αν τα δεδομένα είναι δίτιμα (*binary*) τότε η κατάλληλη συνάρτηση σύνδεσης είναι η logit. Αν τα δεδομένα είναι μετρήσιμα (*counted*) και ακολουθούν κατανομή Poisson, τότε η συνάρτηση σύνδεσης που χρησιμοποιείται είναι ο λογάριθμος του μέσου (Ballinger, 2004).

### 3.5.3 Διακύμανση αποκρίσεων ως συνάρτηση του μέσου

Το επόμενο βήμα των γενικευμένων γραμμικών εξισώσεων είναι η περιγραφή της διακύμανσης των αποκρίσεων  $Y_{ij}$  σαν συνάρτηση του μέσου,

$$Var(Y_{ij}) = fu(\mathbf{m}_{ij}),$$

όπου,  $u(\mathbf{m}_{ij})$  είναι γνωστή συνάρτηση διακύμανσης και  $f$  είναι μια παράμετρος κλίμακας (γνωστή ή προς εκτίμηση) (Davis, 2002).

Για αποκρίσεις που ακολουθούν την Κανονική κατανομή οι επιλογές θα ήταν,

$$g(\mathbf{m}_{ij}) = \mathbf{m}_{ij}, u(\mathbf{m}_{ij}) = 1, Var(Y_{ij}) = f,$$

ενώ όταν ακολουθούν την Διωνυμική κατανομή, θα ήταν,

$$g(\mathbf{m}_{ij}) = \log\left(\frac{\mathbf{m}_{ij}}{1 - \mathbf{m}_{ij}}\right), u(\mathbf{m}_{ij}) = \mathbf{m}_{ij}(1 - \mathbf{m}_{ij}), f = 1 .$$

### 3.5.4 Δομή της συσχέτισης της εξαρτημένης μεταβλητής

Στη συνέχεια πρέπει να επιλεγεί η μορφή του χρησιμοποιούμενου πίνακα συσχέτισης  $V_i$

για κάθε διάνυσμα αποκρίσεων (υποκείμενο). Για κάθε στοιχείο αυτού του πίνακα, γίνεται η υπόθεση ότι είναι η υποτιθέμενη ή εκτιμηθείς συσχέτιση των αντίστοιχων μετρήσεων. Ο  $V_i$  μπορεί να εξαρτάται από άγνωστες παραμέτρους  $a$  (ίδιες για κάθε υποκείμενο). Επομένως, ο πίνακας συσχέτισης του κάθε υποκειμένου είναι γνωστός εκτός από τις παραμέτρους  $a$  που είναι υπό εκτίμηση. Αν και αυτός ο πίνακας συσχέτισης διαφέρει από υποκείμενο σε υποκείμενο, χρησιμοποιείται ένας κοινός πίνακας συσχέτισης  $V$  που δείχνει την μέση εξάρτηση των επαναλαμβανόμενων μετρήσεων ανάμεσα στα υποκείμενα. Οι επιλογές του πίνακα συσχέτισης είναι πολλές. Μπορεί να επιλεγεί ένας πίνακας ανεξαρτησίας (*Independent*) με  $V=I$  (όπου  $I$  ο μοναδιαίος πίνακας). Στην περίπτωση που ο αριθμός των υποκειμένων είναι σημαντικά μεγαλύτερος από τον αριθμό των μετρήσεων για κάθε άτομο, η επίδραση της συσχέτισης είναι αρκετά μικρή, με αποτέλεσμα οι παράμετροι συνήθων ελάχιστων τετραγώνων να μην είναι επαρκείς. Επιλέγοντας αυτή τη μορφή του πίνακα συσχέτισης, η επίλυση των γενικευμένων εξισώσεων εκτίμησης είναι ίδια με την προσαρμογή ενός μοντέλου παλινδρόμησης για ανεξάρτητα δεδομένα. Αυτό το μοντέλο λοιπόν οδηγεί σε επαρκείς εκτιμητές και πίνακα συνδιακυμάνσεων, μόνο όταν το μοντέλο παλινδρόμησης έχει ορισθεί σωστά.

Ο ερευνητής μπορεί να θεωρήσει ότι οι εντός των υποκειμένων παρατηρήσεις είναι όμοια συσχετισμένες, (*Exchangeable structure or compound symmetry*).

$$V(a) = \begin{pmatrix} 1 & a & L & a \\ a & 1 & L & a \\ \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ a & a & L & 1 \end{pmatrix},$$

όπου,  $a$  είναι η κοινή συσχέτιση. Αυτή η μέθοδος είναι κατάλληλη όταν τα δεδομένα είναι ομαδοποιημένα γύρω από ένα υποκείμενο αλλά δεν αποτελούν χρονοσειρά.

Για δεδομένα που είναι συσχετισμένα στη διάρκεια του χρόνου εφαρμόζεται μια δομή αυτοπαλινδρόμησης συσχέτισης (*Autoregressive correlation*),  $corr(Y_i, Y_j) = a^{|i-j|}$  ( $i, j = 1, 2, \mathbf{K}, T$ ) για τον ορισμό της εντός των υποκειμένων συσχέτισης. Οι συσχετίσεις των παρατηρήσεων μειώνονται όσο περισσότερο απέχουν (χρονικά) μεταξύ τους,

$$V(\mathbf{a}) = \begin{pmatrix} 1 & \mathbf{a} & \mathbf{L} & \mathbf{a}^{T-1} \\ \mathbf{a} & 1 & \mathbf{L} & \mathbf{a}^{T-2} \\ \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ \mathbf{a}^{T-1} & \mathbf{a}^{T-2} & \mathbf{L} & 1 \end{pmatrix}$$

Επίσης μπορεί να εφαρμοσθεί και μια μη δομημένη (*Unstructured*) εκτίμηση της εντός των υποκειμένων συσχέτισης. Αυτός ο τύπος χρησιμεύει μόνο όταν υπάρχουν λίγες επαναλήψεις. Τέλος, μπορεί να θεωρηθεί ότι οι αποκρίσεις ενός υποκειμένου είναι ανεξάρτητες μεταξύ τους, καθότι δεν αξιοποιεί πλήρως τις γενικευμένες εξισώσεις εκτιμήσεων που βασικά χρησιμοποιούνται ώστε να περιλαμβάνεται στις υποθέσεις η εντός των υποκειμένων συσχέτιση (Ballinger, 2004).

### 3.5.5 Γενικευμένες Εξισώσεις Εκτίμησης

Το τελευταίο βήμα είναι η εκτίμηση των παραμέτρων  $\mathbf{b}$ . Για να βρεθεί εκτιμητής μέγιστης πιθανοφάνειας των  $\mathbf{b}$ , αρχικά γίνεται η υπόθεση ότι  $V_i$  είναι γνωστό. Πρέπει να βρεθεί η εκτίμηση των  $\mathbf{b}$ , η οποία μεγιστοποιεί την συνάρτηση log-πιθανοφάνειας. Υποθέτοντας ότι τα  $Y_i$  ακολουθούν πολυμεταβλητή Κανονική κατανομή, η συνάρτηση log-πιθανοφάνειας είναι,

$$l = -\frac{k}{2} \log(2p) - \frac{1}{2} \sum_{i=1}^N \log|V_i| - \frac{1}{2} \left\{ \sum_{i=1}^N (Y_i - \mathbf{m}_i(\mathbf{b}))' V_i^{-1} (Y_i - \mathbf{m}_i(\mathbf{b})) \right\} \quad (3.1)$$

όπου,  $k$  είναι ο συνολικός αριθμός των παρατηρήσεων. Το  $\mathbf{b}$  δεν εμφανίζεται στους δυο πρώτους όρους της συνάρτησης, επομένως στη μεγιστοποίηση της συνάρτησης, αγνοούνται αυτοί οι όροι. Επίσης, ο τρίτος όρος της συνάρτησης έχει αρνητικό πρόσημο, επομένως για την μεγιστοποίηση της συνάρτησης ως προς  $\mathbf{b}$ , αρκεί η ελαχιστοποίηση της συνάρτησης,

$$\sum_{i=1}^N \{Y_i - \mathbf{m}_i(\mathbf{b})\}' V_i^{-1} \{Y_i - \mathbf{m}_i(\mathbf{b})\}, \quad (3.2)$$

όπου,  $V_i$  θεωρείται γνωστό και  $\mathbf{m}_i$  είναι το διάνυσμα των μέσων αποκρίσεων με στοιχεία,

$$\mathbf{m}_{ij} = \mathbf{m}_{ij}(\mathbf{b}) = \mathbf{g}^{-1}(X_{ij}' \mathbf{b}) \quad (3.3)$$

Χρησιμοποιώντας αλγόριθμο, μπορεί να δειχθεί ότι εάν υπάρχει ελάχιστο της παραπάνω συνάρτησης, θα πρέπει να επιλύει τις ακόλουθες γενικευμένες εξισώσεις εκτίμησης,



$$\sum_{i=1}^N D' V_i^{-1} (y_i - \mathbf{m}_i) = 0, \quad (3.4)$$

όπου,  $V_i$  είναι ο χρησιμοποιούμενος πίνακας συνδιακύμανσης και  $D_i = \partial \mathbf{m}_i / \partial \mathbf{b}$  είναι ο παράγωγος πίνακας. Με τον όρο «χρησιμοποιούμενος», εννοείται ότι ο  $V_i$  πλησιάζει τον πραγματικό πίνακα συνδιακύμανσης των  $Y_i$ , δηλαδή  $V_i \approx \text{Cov}(Y_i)$  αλλά  $V_i \neq \text{Cov}(Y_i)$  εκτός αν τα μοντέλα για τις διακυμάνσεις και τις εντός των υποκειμένων σχέσεις είναι ορθά ορισμένα (Fitzmaurice, Laird and Ware, 2004).

Ο πίνακας  $D_i$  μπορεί να θεωρηθεί ως ένας πίνακας που μετατρέπει τα αρχικά  $Y_i$  (και  $\mathbf{m}_i$ ) σε στοιχεία  $g(\mathbf{m}_{ij})$ . Ο  $n_i \times p$  πίνακας παραγώγων δίνεται από,

$$D_i = \begin{pmatrix} \partial \mathbf{m}_{i1} / \partial b_1 & \partial \mathbf{m}_{i1} / \partial b_2 & \mathbf{L} & \partial \mathbf{m}_{i1} / \partial b_p \\ \partial \mathbf{m}_{i2} / \partial b_1 & \partial \mathbf{m}_{i2} / \partial b_2 & \mathbf{L} & \partial \mathbf{m}_{i2} / \partial b_p \\ \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ \partial \mathbf{m}_{in_i} / \partial b_1 & \partial \mathbf{m}_{in_i} / \partial b_2 & \mathbf{L} & \partial \mathbf{m}_{in_i} / \partial b_p \end{pmatrix}$$

και είναι μόνο συνάρτηση των  $\mathbf{b}$ . Από την άλλη ο πίνακας  $V_i$  είναι συνάρτηση των  $\mathbf{b}$ ,  $\mathbf{f}$  και  $\mathbf{a}$ , αφού τα διαγώνια στοιχεία του είναι διακυμάνσεις και τα μη διαγώνια στοιχεία του είναι οι «χρησιμοποιούμενες» συνδιακυμάνσεις. Δηλαδή, οι διακυμάνσεις εξαρτώνται από τους μέσους, άρα τα  $\mathbf{b}$ , μέσω της συνάρτησης διακύμανσης  $u(\mathbf{m}_{ij})$  (που επίσης εξαρτώνται από το  $\mathbf{f}$ ), και οι συνδιακυμάνσεις των  $Y_i$  εξαρτώνται από τα  $\mathbf{b}$  και  $\mathbf{a}$ . Έτσι, οι γενικευμένες εξισώσεις εκτίμησης εξαρτώνται από τα  $\mathbf{b}$  και  $\mathbf{a}$ .

Για τον λόγο αυτό, χρειάζεται η παρακάτω επαναληπτική διαδικασία εκτίμησης,

1. Με δοσμένες εκτιμήσεις των  $\mathbf{a}$  και  $\mathbf{f}$ , εκτιμάται ο  $V_i$  και δίνεται μια εκτίμηση των  $\mathbf{b}$  μέσω των γενικευμένων εξισώσεων εκτίμησης, από τη σχέση (3.4).
2. Με την εκτίμηση των  $\mathbf{b}$ , υπολογίζονται οι εκτιμήσεις των  $\mathbf{a}$  και  $\mathbf{f}$ , με βάση τα τυποποιημένα κατάλοιπα,

$$e_{ij} = (Y_{ij} - \hat{\mathbf{m}}_{ij}) / \sqrt{u(\hat{\mathbf{m}}_{ij})}.$$

Για παράδειγμα, το  $\mathbf{f}$  μπορεί να εκτιμηθεί από,

$$\hat{\mathbf{f}} = \frac{\sum_{i=1}^N \sum_{j=1}^{n_i} e_{ij}^2}{\sum_{i=1}^N n_i}$$

Επίσης η παράμετρος  $a$ , μπορεί να εκτιμηθεί με παρόμοιο τρόπο, με βάση το μοντέλο για την εντός των υποκειμένων σχέση. Αν ο σχεδιασμός είναι ισορροπημένος (*balanced*) και η σχέση εκφράζεται ως μη δομημένες συσχετίσεις, το  $a_{jk} = \text{Corr}(Y_{ij}, Y_{ik})$  μπορεί να εκτιμηθεί από τη σχέση,

$$\hat{a}_{jk} = \left( \frac{1}{\hat{f}N} \right) \sum_{i=1}^N e_{ij} e_{ik}.$$

Η επανάληψη των βημάτων 1 και 2 γίνεται έως ότου συγκλίνει η διαδικασία.

Στη σύγκλισή της, τα  $b$ , οι λύσεις των γενικευμένων εξισώσεων εκτίμησης, έχουν τα ακόλουθα χαρακτηριστικά (Fitzmaurice, Laird and Ware, 2004).

Τα  $\hat{b}$  είναι επαρκείς εκτιμητές των  $b$ . Δηλαδή τα  $\hat{b}$  με μεγάλη πιθανότητα βρίσκονται πολύ κοντά στους παραμέτρους παλινδρόμησης  $b$ . Αυτό ισχύει ακόμα και αν η εντός των υποκειμένων σχέση δεν έχει μοντελοποιηθεί σωστά. Ο λόγος για αυτό είναι ότι για να εκτιμηθούν σωστά τα  $b$ , αρκεί να έχει μοντελοποιηθεί σωστά η μέση απόκριση. Αυτό λοιπόν το ανθεκτικό στοιχείο είναι που καθιστά τις γενικευμένες εξισώσεις εκτίμησης εφαρμόσιμες σε πολλές περιπτώσεις.

Συνήθως μια πρώτη προσέγγιση για τον πίνακα συνδιακυμάνσεων των  $b$  χρησιμοποιείται ο αποκαλούμενος “*model-based*”,

$$\hat{V}(\hat{b}) = \left( \sum_{i=1}^N \left( \frac{\partial \hat{m}_i}{\partial b} \right)' \hat{V}_i^{-1} \left( \frac{\partial \hat{m}_i}{\partial b} \right) \right)^{-1}$$

Αυτός ο εκτιμητής, στην περίπτωση που δεν έχει ορισθεί σωστά το μοντέλο, δεν θα είναι συνεπής. Προτείνεται από τους Liang and Zeger (1986) η εκτίμηση του πίνακα συνδιακυμάνσεων από έναν άλλο εκτιμητή, ανθεκτικό (*robust*), ο οποίος αποτελεί συνεπή εκτιμητή ακόμα και όταν η δομή του πίνακα συσχέτισης έχει ορισθεί λάθος (Fitzmaurice, Laird and Ware, 2004),

$$V(\hat{b}) = B^{-1} M B^{-1},$$

όπου,  $B = \sum_{i=1}^N D_i V_i^{-1} D_i$  και  $M = \sum_{i=1}^N D_i V_i^{-1} \text{Cov}(Y_i) V_i^{-1} D_i$ .

Τα  $B$  και  $M$  μπορούν να εκτιμηθούν αντικαθιστώντας τα  $a$ ,  $f$  και  $b$  από τους εκτιμητές τους, καθώς και το  $\text{Cov}(Y_i)$  με το  $(Y_i - \hat{m}_i)(Y_i - \hat{m}_i)'$ . Έτσι, η παραπάνω σχέση γίνεται,

$$\hat{V}(\hat{b}) = \left( \sum_{i=1}^N \hat{D}_i' \hat{V}_i^{-1} \hat{D}_i \right)^{-1} \left\{ \sum_{i=1}^N \hat{D}_i' \hat{V}_i^{-1} (Y_i - \hat{m}_i)(Y_i - \hat{m}_i)' \hat{V}_i^{-1} \hat{D}_i \right\} \left( \sum_{i=1}^N \hat{D}_i' \hat{V}_i^{-1} \hat{D}_i \right)^{-1}$$

### 3.5.6 Έλεγχοι Υποθέσεων

Μετά την εκτίμηση του διανύσματος παραμέτρων  $\mathbf{b}$ , ακολουθούν οι έλεγχοι υποθέσεων. Έστω, η μηδενική υπόθεση  $H_0: \mathbf{C}\mathbf{b} = \mathbf{d}$  όπου  $\mathbf{C}$  ένας πίνακας σταθερών, διάστασης  $c \times p$  και  $\mathbf{d}$  ένα διάνυσμα σταθερών αριθμών διάστασης  $p$ . Το τεστ που χρησιμοποιείται είναι το στατιστικό Wald,

$$Q_C = (\mathbf{C}\hat{\mathbf{b}} - \mathbf{d})' [\mathbf{C}\mathbf{V}(\hat{\mathbf{b}})\mathbf{C}']^{-1} (\mathbf{C}\hat{\mathbf{b}} - \mathbf{d}).$$

Καθώς το διάνυσμα των παραμέτρων  $\hat{\mathbf{b}}$  ακολουθεί ασυμπτωτικά κανονική κατανομή, το στατιστικό  $Q_C$  ακολουθεί  $\chi_c^2$  κατανομή κάτω από την υπόθεση  $H_0$  (Davis, 2002).

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

# ΚΕΦΑΛΑΙΟ 4

## Διερεύνηση Λογισμικού

Για την ανάλυση δεδομένων με τη μέθοδο γενικευμένων εξισώσεων εκτίμησης, υπάρχουν διάφορα διαθέσιμα λογισμικά. Κάποια από αυτά είναι το S-Plus, το R, το MAREG και το SPSS. Τα δύο πρώτα είναι λογισμικά που λειτουργούν με εντολές. Τα τελευταία έχουν μενού από τα οποία επιλέγονται οι συγκεκριμένες λειτουργίες που ζητούνται κάθε φορά.

### 4.1 SPSS

Με το στατιστικό πακέτο SPSS μπορούν να αναλυθούν, πλέον, επαναλαμβανόμενες μετρήσεις που αφορούν και σε διακριτές μεταβλητές από την διαδικασία,

Analyze ► Generalized Linear Models ► Generalized estimating Equations

Παρακάτω, στην ανάλυση των δεδομένων θα χρησιμοποιηθεί το SPSS.

### 4.2 S-Plus

Για να χρησιμοποιηθεί το στατιστικό πακέτο S-Plus, πρέπει προηγουμένως να εγκατασταθεί η βιβλιοθήκη GEE. Στη συνέχεια αρκεί μονάχα να επιλεγεί η βιβλιοθήκη για να γίνει η ανάλυση επαναλαμβανόμενων μετρήσεων. Για να χρησιμοποιηθεί η εντολή gee θα πρέπει η μεταβλητή απόκρισης να μετατραπεί σε δίτιμη. Επομένως, αν η μεταβλητή απόκρισης έχει  $n$  επίπεδα, θα πρέπει να δημιουργηθούν  $n-1$  δίτιμες μεταβλητές και να εφαρμοσθούν  $n-1$  μοντέλα λογιστικής παλινδρόμησης με διαφορετική κάθε φορά μεταβλητή απόκρισης. Επίσης αυτή η μεταβλητή θα πρέπει να ορισθεί ως παράγοντας. Τα αποτελέσματα προσφέρουν εκτιμήσεις των παραμέτρων, τυπικά σφάλματα αυτών, καθώς και τιμές για το στατιστικό έλεγχο αυτών των εκτιμήσεων.

### 4.3 R

Παρόμοια με το S-Plus, το πακέτο R λειτουργεί και αυτό με εντολές. Για ανάλυση διαχρονικών δεδομένων είναι αναγκαία η εισαγωγή του πακέτου εντολών GEEPACK (<http://www.r-project.org/>). Οι δυο βασικές εντολές είναι οι `geese` και `ordgee` για επαναλαμβανόμενες δίτιμες και διατάξιμες, αντίστοιχα, αποκρίσεις. Η πρώτη χρησιμοποιείται παρόμοια με την εντολή `gee` του S-plus με μικρές διαφορές στην εκτίμηση του διανύσματος των παραμέτρων. Για την χρήση της δεύτερης, `ordgee`, απαιτείται η δήλωση της μεταβλητής απόκρισης ως διατάξιμη. Επίσης πρέπει τα δεδομένα να είναι ταξινομημένα ανά υποκείμενο και χρονική στιγμή. Οι δύο εντολές του R, δεν θεωρούν ελλιπής παρατηρήσεις, επομένως αν τυχόν υπάρχει κάποια θα πρέπει να εξαιρεθεί το αντίστοιχο υποκείμενο από την ανάλυση. Εδώ, η μεταβλητή απόκρισης δεν χρειάζεται να μετατραπεί σε δίτιμη. Πρέπει όμως να ορισθούν οι επεξηγηματικές μεταβλητές ως παράγοντες.

Αρχικά πρέπει να καλέσουμε το πακέτο εντολών GEEPACK με την παρακάτω εντολή,  
`Library(geepack)`.

Για να μπορέσουν να χρησιμοποιηθούν τα δεδομένα από το πρόγραμμα, (καθώς τα αρχικά δεδομένα είναι σε αρχείο `.sav` του SPSS) πρέπει να μεταφερθούν σε EXCEL και να σωθούν με προέκταση `*.csv`. Για να καλέσουμε λοιπόν τα δεδομένα πρέπει να ορίσουμε που είναι το αρχείο και πως πρέπει να διαβαστεί από το πρόγραμμα (αν υπάρχουν ονόματα πεδίων, και πως διαχωρίζονται τα πεδία μεταξύ τους-με κόμμα, τελεία, `tab...`),

```
a<-read.table("C:/excel.csv",header=T,sep=";")
```

Τα δεδομένα μπορούν να εμφανιστούν στην οθόνη με την εντολή,

```
>a
```

και έχουν την μορφή,

	Id	time	smoke	apasx	treat	selfcare	usualact	pain	anx	health
1	110201001	1	3	1	1	1	1	1	1	5
2	110201005	1	3	3	3	1	1	1	2	5
3	110201006	1	1	3	1	1	1	1	2	4
4	110201009	1	1	3	1	2	2	3	2	4
5	110201010	1	3	3	1	2	3	3	2	4
6	110202004	1	3	1	2	1	1	2	2	5
...										
576	110209171	3	2	1	2	1	1	1	2	4

Τα δεδομένα μας είναι στη κατάλληλη μορφή για την ανάλυση που θέλουμε να διενεργήσουμε και για το λογισμικό το οποίο χρησιμοποιούμε. Οι πρώτες 192 γραμμές έχουν την πρώτη μέτρηση όλων των ασθενών, οι επόμενες 192 την δεύτερη κ.ο.κ. Έχουν εξαιρεθεί οι ασθενείς που έχουν ελλείπουσες τιμές σε κάποια μέτρηση. Εάν δεν ήταν στην κατάλληλη μορφή τα αρχικά μας δεδομένα, με τη βοήθεια της εντολής `help(ordgee)` θα μπορούσαμε να δούμε κάποιους τρόπους εναλλαγής των δεδομένων μας.

Η αποκριτική μεταβλητή εδώ έχει μετασηματιστεί σε μια κατηγορική και διατάξιμη μεταβλητή με τιμές από 1 έως 5 που αντιστοιχούν σε εύρος 20 μονάδων της αρχικής μεταβλητής (1: 0-20 πολύ άσχημα, 2: 20-40 άσχημα, 3: 40-60 μέτρια, 4: 60-80 καλά και 5: 80-100 πολύ καλά).

Ορίζοντας την αποκριτική μεταβλητή ως διατάξιμη και τις επεξηγηματικές ως παράγοντες, με την παρακάτω εντολή λαμβάνεται το μοντέλο με δομή ανεξαρτησίας για τον χρησιμοποιούμενο πίνακα συσχέτισης,

```
fit<-ordgee(ordered(health)~as.factor(time)+as.factor(treat)+as.factor(pain)+as.factor(usualact)+
as.factor(selfcare),id=id,data=a,int.const=TRUE)
```

Το τελευταίο μέρος της εντολής, είναι αυτό που καθορίζει εάν οι σταθερές παράμετροι θα είναι σταθερές χρονικά.

Τα αποτελέσματα τα παίρνουμε με την εντολή,

```
summary(fit)
```

```
Call:
ordgee(formula = ordered(health) ~ as.factor(time) + as.factor(treat) +
  as.factor(pain) + as.factor(usualact) + as.factor(selfcare), id = id, data = a, int.const = TRUE)
Mean Model:
Mean Link:          logit
Variance to Mean Relation: binomial
Coefficients:
              estimate      san.se      wald      p
Inter:1      6.58e+13  2.29e+13   8.278  0.004012
Inter:2      4.57e+00  1.30e+00  12.283  0.000457
Inter:3      1.42e+00  1.16e+00   1.505  0.219934
Inter:4     -1.31e+00  2.44e+00   0.288  0.591615
as.factor(time)2  2.67e-01  3.31e-01   0.650  0.420180
as.factor(time)3  1.29e-01  3.27e-01   0.157  0.692235
as.factor(treat)2  3.29e-01  3.16e-01   1.080  0.298754
as.factor(treat)3  3.42e-01  3.19e-01   1.152  0.283127
```

as.factor(pain)2	-2.02e+00	8.31e-01	5.895	0.015182
as.factor(pain)3	-4.37e-01	9.14e-01	0.229	0.632411
as.factor(usualact)2	-1.62e+00	5.79e-01	7.830	0.005140
as.factor(usualact)3	-1.73e+00	7.40e-01	5.454	0.019522
as.factor(selfcare)1	1.42e+00	3.93e-01	13.102	0.000295

Scale is fixed.

Correlation Model:

Correlation Structure: independence

Returned Error Value: 1

Number of clusters: 192 Maximum cluster size: 3

Από τα παραπάνω αποτελέσματα παρατηρείται ότι η θεραπεία δεν είναι στατιστικά σημαντική. Οι παράγοντες όμως pain, usualact και ειδικότερα ο selfcare φαίνεται να είναι στατιστικά σημαντικοί. Το μοντέλο που εφαρμόστηκε είναι της μορφής ,

$$P(Y_i \leq k | x_i) = a_k + b_{11}X_{11} + b_{12}X_{12} + b_{21}X_{21} + b_{22}X_{22} + b_{31}X_{31} + b_{32}X_{32} + b_{41}X_{41} + b_{42}X_{42} + b_5X_5$$

Όπου,  $X_1, X_2, X_3, X_4, X_5$  είναι αντίστοιχα οι μεταβλητές time, treat, pain, usualact, selfcare.



# Κ Ε Φ Α Λ Α Ι Ο 5

## Ανάλυση Δεδομένων – Εφαρμογή

### 5.1 Λογισμικό

Η διαδικασία της ανάλυσης των δεδομένων των γυναικών με μεταστατικό καρκίνο του μαστού με τη μέθοδο των γενικευμένων εξισώσεων εκτίμησης, εφαρμόζεται στο στατιστικό πακέτο SPSS 15.0. Το μοντέλο γενικευμένων εξισώσεων εκτίμησης (GEE) στο SPSS, επεκτείνει το γενικευμένο γραμμικό μοντέλο (GLM), προσφέροντας υποστήριξη μη ανεξάρτητων δεδομένων, όπως οι επαναλαμβανόμενες μετρήσεις.

Καθώς η αποκριτική μεταβλητή είναι διακριτή, η κατάλληλη διαδικασία είναι η Generalized Estimating Equations (GEE) του Generalized Linear Models. Υποθέτουμε ότι οι μετρήσεις είναι εξαρτημένες εντός των υποκειμένων και ανεξάρτητες μεταξύ των υποκειμένων.

### 5.2 Κατανομή μεταβλητής απόκρισης

Το πρώτο βήμα για τον ορισμό του μοντέλου γενικευμένων εξισώσεων εκτίμησης, είναι η ταυτοποίηση και επιλογή της κατανομής της αποκριτικής μεταβλητής μεταβλητής. Η εξαρτημένη μεταβλητή μπορεί να είναι μη συνεχής και έτσι να δύναται να ακολουθεί κάποια κατανομή πλην της Κανονικής (π.χ. Poisson). Ανάλογα λοιπόν με την κατανομή, εν συνεχεία επιλέγεται η συνάρτηση σύνδεσης.

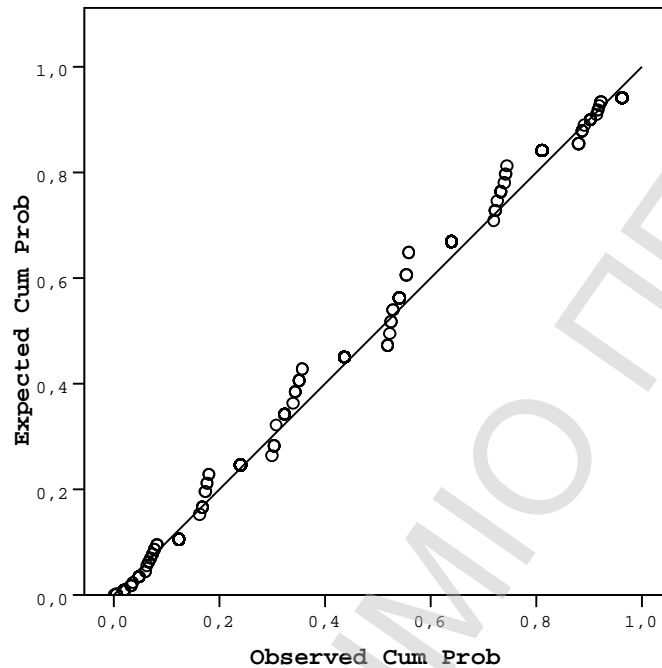
Η μεταβλητή απόκρισης είναι η ποιότητα ζωής του ασθενή (*Quality of Life*), όπως αξιολογείται από τον ίδιο τον ασθενή, σε μια κλίμακα από το 0 έως το 100 με τιμές κοντά στο μηδέν να ανταποκρίνονται στην χειρίστη πιθανή κατάσταση της υγείας του και με τιμές κοντά στο 100 στην βέλτιστη.

Από το ακόλουθο σχήμα, φαίνεται ότι η εξαρτημένη μεταβλητή ακολουθεί την Κανονική κατανομή (Σχήμα 5.1).

## ΣΧΗΜΑ 5-1

P-P plot εξαρτημένης μεταβλητής

Normal P-P Plot of Η κατάσταση της δικής σας υγείας σήμερα



### 5.3 Συνάρτηση σύνδεσης

Η συνάρτηση σύνδεσης αποτελεί τον μετασχηματισμό της αποκριτικής μεταβλητής, η οποία και μοντελοποιείται. Μετασχηματισμοί μπορεί να είναι αναγκαίοι όταν οι ανεξάρτητες μεταβλητές του μοντέλου έχουν μια μη γραμμική επίδραση στην εξαρτημένη μεταβλητή.

Γενικά, ισχύει ότι το γινόμενο των συντελεστών του μοντέλου με τις ανεξάρτητες μεταβλητές, συν το σφάλμα, ισούται με την αναμενόμενη τιμή της συνάρτησης σύνδεσης της εξαρτημένης μεταβλητής. Η επιλογή της συνάρτησης σύνδεσης εξαρτάται από την κατανομή που ακολουθεί η αποκριτική μεταβλητή.

Η συνάρτηση σύνδεσης (*Link Function*), που μπορεί να επιλεγεί έχοντας μια αποκριτική μεταβλητή που ακολουθεί Κανονική κατανομή, είναι μια από τις ακόλουθες τρεις: Η

Ταυτοτική (*Identity*)  $f(x) = x$ , η Λογαριθμική (*Log*)  $f(x) = \log(x)$  και η Δύναμη (*Power*)  $f(x) = x^a$ , με  $a \neq 0$  και  $f(x) = \log(x)$ , όταν  $a = 0$ , που χρησιμοποιούνται για κάθε κατανομή. Εφόσον λοιπόν η εξαρτημένη μεταβλητή ('Κατάσταση της δικής σας υγείας σήμερα') ακολουθεί την Κανονική κατανομή θα επιλεγεί η ταυτοτική συνάρτηση σύνδεσης (για λόγους ευκολότερης ερμηνείας). Επομένως το μοντέλο θα είναι της μορφής,

$$m_{ij} = x'_{ij} b.$$

#### 5.4 Δομή πίνακα συνδιακύμανσης

Η δομή του πίνακα συσχέτισης, προσδιορίζει πώς οι παρατηρήσεις κάθε συστάδας σχετίζονται μεταξύ τους. Η πιο συχνή επιλογή είναι του πίνακα Exchangeable (*compound symmetry*). Αυτή η δομή υποθέτει ότι οποιοσδήποτε δύο συσχετισμένες παρατηρήσεις, έχουν την ίδια συσχέτιση με οποιοσδήποτε άλλες δύο, του ίδιου υποκειμένου. Δηλαδή, για ένα υποκείμενο, η πρώτη και η τρίτη παρατήρηση έχουν την ίδια συσχέτιση με την δεύτερη και την τρίτη, αλλά πιθανότατα διαφορετική από τις παρατηρήσεις ενός άλλου υποκειμένου. Σε κάποιες περιπτώσεις, η υπόθεση της ισότητας της συσχέτισης δεν είναι αληθής. Κάποιες παρατηρήσεις μπορεί να είναι πιο ισχυρά συσχετισμένες από άλλες. Έτσι, ο πίνακας συνδιακυμάνσεων μπορεί να είναι διαφορετικός για κάποιες ομάδες υποκειμένων.

Η M-dependent δομή του πίνακα συνδιακύμανσης, υποθέτει οι συσχετίσεις των παρατηρήσεων που καταγράφονται σε ίσα χρονικά διαστήματα μεταξύ τους για τα υποκείμενα, είναι ίσες. Δηλαδή, οι συσχετίσεις οποιονδήποτε δύο παρατηρήσεων που απέχουν 6 μήνες είναι ίσες, όπως και οι συσχετίσεις οποιονδήποτε δύο παρατηρήσεων που απέχουν ένα χρόνο είναι ίσες.

Η δομή της αυτοπαλίνδρομης συσχέτισης (*Autoregressive correlation*), υποθέτει ότι η συσχέτιση μεταξύ επαναλαμβανόμενων μετρήσεων, εντός των υποκειμένων, μειώνεται εκθετικά καθώς αυτές είναι πολύ μακριά η μία από την άλλη. Συνήθως αυτό γίνεται στις διαχρονικές μελέτες.

Η δομή ανεξαρτησίας υποθέτει ότι οι επαναλαμβανόμενες μετρήσεις εντός των υποκειμένων, είναι ανεξάρτητες (Katz, 1999).

Τέλος, υπάρχει και η μη δομημένη (*Unstructured*) που δεν κάνει καμία υπόθεση για την συσχέτιση εντός των υποκειμένων. Χρησιμοποιείται κυρίως όταν οι επαναλήψεις είναι λίγες

και γενικά είναι η πιο αποτελεσματική δομή (Hedeker and Gibbons, 2006).

Ως δομή του πίνακα συσχέτισης, δεν θα επιλεγεί η δομή ανεξαρτησίας (*Independence*) καθώς οι παρατηρήσεις είναι επαναλαμβανόμενες και οι μετρήσεις θεωρείται ότι έχουν κάποια εξάρτηση. Στην πορεία της ανάλυσης θα εξεταστεί η μη δομημένη μορφή (*Unstructured*) και η Compound symmetry δομή (*Exchangeable*).

## 5.5 Ανάλυση και ορισμός παραμέτρων

Πριν την ανάλυση των δεδομένων, θα παρουσιαστούν κάποια περιγραφικά μετρά για τις μεταβλητές που εισήχθησαν στο μοντέλο. Στον Πίνακα 2.4 δίνεται η κατανομή των 5 διαστάσεων του ερωτηματολογίου EQ-5D ανά χρονική στιγμή. Για τις μεταβλητές που μετρήθηκαν μια φορά στην αρχή της μελέτης, παρουσιάζεται η κατανομή τους ανά θεραπεία στους Πίνακες 5-1(a,b,c,d), όπως και για την αποκριτική μεταβλητή και την ηλικία στον Πίνακα 5-2(a,b).

### ΠΙΝΑΚΑΣ 5-1(a)

Η εκπαίδευση σας συνεχίστηκε και μετά την ελάχιστη ηλικία που μπορεί κανείς να σταματήσει το σχολείο; \* Ομάδα Θεραπείας

		Ομάδα Θεραπείας			Total
		P+Cb	G+Doc	P only	
Όχι	Count	52	39	34	125
	% within Ομάδα Θεραπείας	72,2%	63,9%	50,7%	62,5%
Ναί	Count	20	22	33	75
	% within Ομάδα Θεραπείας	27,8%	36,1%	49,3%	37,5%
Total	Count	72	61	67	200
	% within Ομάδα Θεραπείας	100,0%	100,0%	100,0%	100,0%

### ΠΙΝΑΚΑΣ 5-1(b)

Έχετε κάποιο πτυχίο ή ισοδύναμο επαγγελματικό προσόν; \* Ομάδα Θεραπείας

		Ομάδα Θεραπείας			Total
		P+Cb	G+Doc	P only	
Όχι	Count	60	43	38	141
	% within Ομάδα Θεραπείας	83,3%	70,5%	55,9%	70,1%
Ναί	Count	12	18	30	60
	% within Ομάδα Θεραπείας	16,7%	29,5%	44,1%	29,9%
Total	Count	72	61	68	201
	% within Ομάδα Θεραπείας	100,0%	100,0%	100,0%	100,0%

### ΠΙΝΑΚΑΣ 5-1(c)

#### Καπνιστής \* Ομάδα Θεραπείας

		Ομάδα Θεραπείας			Total
		P+Cb	G+Doc	P only	
Είστε καπνίστρια	Count	9	13	13	35
	% within Ομάδα Θεραπείας	12,5%	21,3%	19,1%	17,4%
Είστε πρώην καπνίστρια	Count	14	15	10	39
	% within Ομάδα Θεραπείας	19,4%	24,6%	14,7%	19,4%
Δεν υπήρξατε ποτέ καπνίστρια	Count	49	33	45	127
	% within Ομάδα Θεραπείας	68,1%	54,1%	66,2%	63,2%
Total	Count	72	61	68	201
	% within Ομάδα Θεραπείας	100,0%	100,0%	100,0%	100,0%

### ΠΙΝΑΚΑΣ 5-1(d)

#### Ποιό απο τα παρακάτω περιγράφει καλύτερα την κύρια δραστηριότητάς σας; \* Ομάδα Θεραπείας

		Ομάδα Θεραπείας			Total
		P+Cb	G+Doc	P only	
Υπάλληλος ή ελεύθερος επαγγελματίας	Count	19	16	13	48
	% within Ομάδα Θεραπείας	27,9%	28,1%	21,3%	25,8%
Συνταξιούχος	Count	13	14	21	48
	% within Ομάδα Θεραπείας	19,1%	24,6%	34,4%	25,8%
Οικιακά	Count	36	27	26	89
	% within Ομάδα Θεραπείας	52,9%	47,4%	42,6%	47,8%
Άλλο	Count	0	0	1	1
	% within Ομάδα Θεραπείας	,0%	,0%	1,6%	,5%
Total	Count	68	57	61	186
	% within Ομάδα Θεραπείας	100,0%	100,0%	100,0%	100,0%

### ΠΙΝΑΚΑΣ 5-2(a)

#### Περιγραφικά μέτρα Ηλικίας

Πόσο χρονών είστε;		
N	Valid	200
	Missing	1
Mean		56,880
Median		57,500
Mode		57,000
Std. Deviation		11,534
Variance		133,041
Minimum		25,000
Maximum		82,000

	Πόσο χρονών είστε;		
	Ομάδα Θεραπείας		
	P+Cb	G+Doc	P only
Mean	57,79167	57,23333	55,60294
Median	59,00000	57,00000	57,00000
Mode	64,000	53,000	41,000
Std. Deviation	10,650580	11,409729	12,555714
Variance	113,435	130,182	157,646
Minimum	31,000	32,000	25,000
Maximum	82,000	80,000	75,000

**ΠΙΝΑΚΑΣ 5-2(b)**

Η κατάσταση της δικής σας υγείας σήμερα

	Ομάδα Θεραπείας								
	P+Cb			G+Doc			P only		
	Ενημέρωση			Ενημέρωση			Ενημέρωση		
	1	2	3	1	2	3	1	2	3
Mean	66,40	71,04	72,72	70,10	70,33	73,67	73,72	72,99	79,06
Median	70,00	70,00	75,00	70,00	70,00	80,00	70,00	72,50	80,00
Mode	70,00	70,00	60,00	70,00	70,00	80,00	70,00	90,00	80,00
Std. Deviation	19,38	15,21	18,71	18,19	16,15	17,47	18,53	19,13	14,79
Variance	375,51	231,36	349,95	330,97	260,69	305,16	343,27	365,87	218,77
Minimum	20,00	30,00	30,00	30,00	30,00	35,00	10,00	20,00	30,00
Maximum	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00

Οι σημαντικότερες ίσως παρατηρήσεις από τους παραπάνω πίνακες είναι: οι περισσότεροι ασθενείς δεν είχαν μόρφωση πέραν της βασικής (62,5%) και ιδίως στο δείγμα που έλαβε την πρώτη θεραπεία (P+Cb) όπου το ποσοστό ανέρχεται στο 72,2%. Ως αποτέλεσμα αυτού το 70,1% δεν είχε κάποιο πτυχίο ή επαγγελματικό προσόν.

Έτσι τα οικιακά ως κύρια απασχόληση κατέχουν το 43,1% με ακόμη ένα ποσοστό της τάξεως του 50% να εργάζεται ακόμη ή να εργαζόταν στο παρελθόν (συνταξιούχοι).

Άξιο αναφοράς αποτελεί το γεγονός ότι το 63,2% των ασθενών δεν υπήρξε ποτέ καπνιστής και επίσης 19,4% υπήρξε κάποτε αλλά όχι πια.

Η μέση ηλικία των ασθενών κατά την εισαγωγή τους στη μελέτη ήταν 56,8 έτη και η κατάσταση της υγείας των ασθενών, όπως αξιολογείται από τους ίδιους, φαίνεται κατά την τρίτη μέτρηση να είναι σε καλύτερα επίπεδα (μεταξύ 72 και 79), ανεξαρτήτως μεθόδου θεραπείας.

Αρχικά λοιπόν για την ανάλυση, ορίζεται η μεταβλητή που αποτελεί την «ταυτότητα» του κάθε ασθενή ('Ασθενής') και η μεταβλητή που διαχωρίζει της μετρήσεις εντός των υποκειμένων ('Μέτρηση'). Στη συνέχεια επιλέγεται η δομή του πίνακα συσχέτισης (Unstructured, ή Exchangeable).

Στο επόμενο βήμα, ορίζεται η εξαρτημένη – αποκριτική μεταβλητή ('Η κατάσταση της δικής σας υγείας σήμερα'), η κατανομή που ακολουθεί, καθώς και η κατάλληλη συνάρτηση σύνδεσης, αναλόγως την κατανομή που ακολουθεί η μεταβλητή απόκρισης.

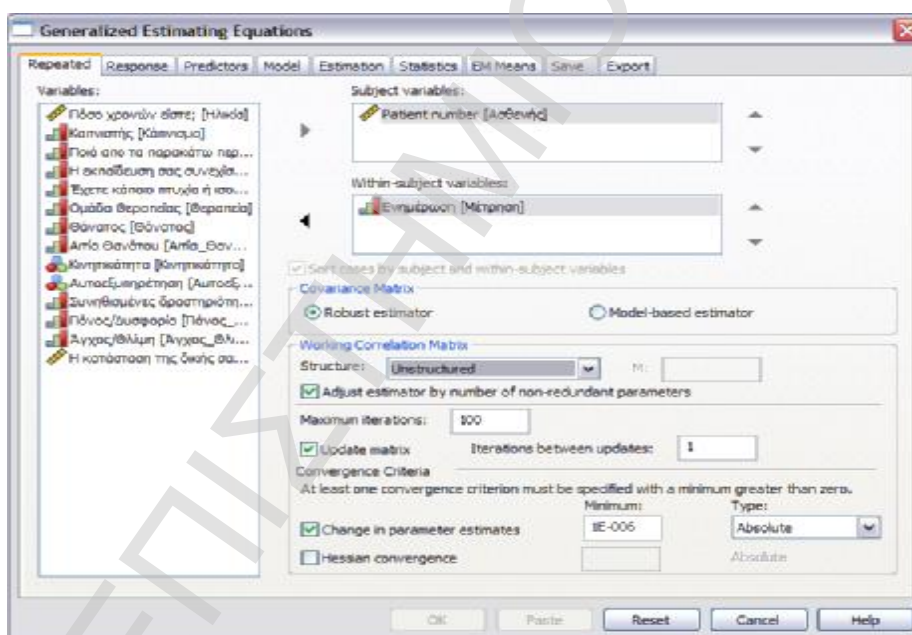
Στην καρτέλα 'Repeated' εισάγεται η μεταβλητή που χαρακτηρίζει τον ασθενή και η μεταβλητή που δείχνει την μέτρηση που αντιστοιχεί στον κάθε ασθενή (1<sup>η</sup>, 2<sup>η</sup>, 3<sup>η</sup>). Η επιλογή του ανθεκτικού εκτιμητή (*robust estimator*) για τον πίνακα συνδιακυμάνσεων, είναι ένας διορθωμένος εκτιμητής που βασίζεται στο μοντέλο και προσφέρει έναν ακριβή εκτιμητή

ακόμα και όταν ο πίνακας συσχέτισης έχει οριστεί λάθος. Ο χρησιμοποιούμενος πίνακας συσχέτισης (*Working Correlation Matrix*), ορίζει την εντός των υποκειμένων συσχέτιση. (Εικόνα 5.1).

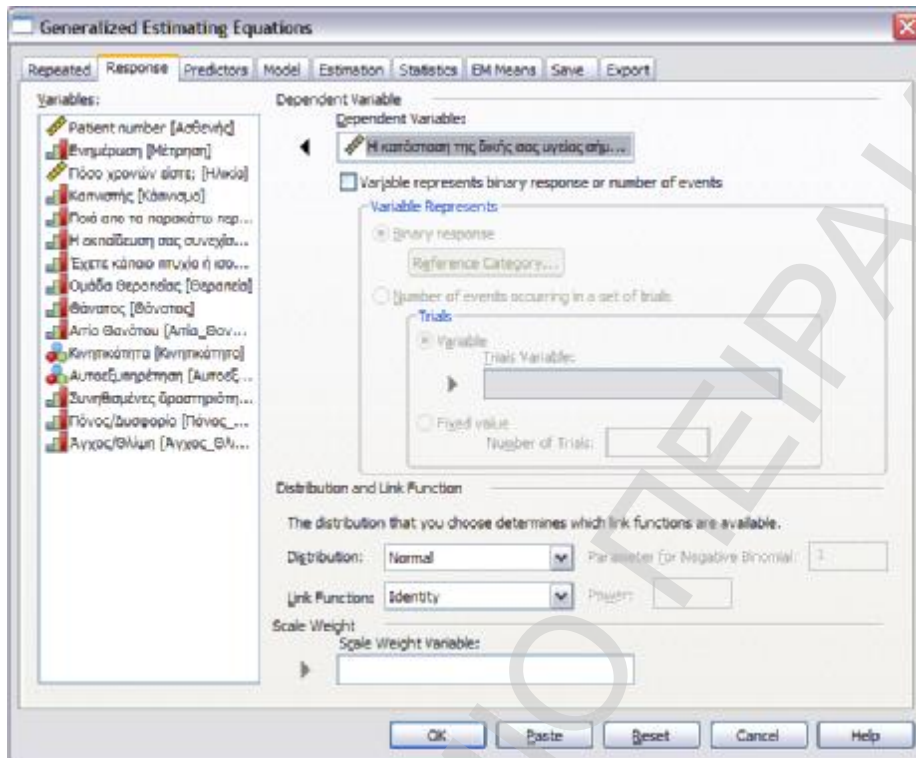
Στην καρτέλα 'Response' ορίζεται η αποκριτική μεταβλητή και η κατανομή της (Κανονική κατανομή), όπως επίσης και η συνάρτηση σύνδεσης που θα εφαρμοσθεί (Ταυτοτική, Λογαριθμική, Δύναμη), (Εικόνα 5.2).

Στην καρτέλα 'Predictors' τοποθετούνται στο πλαίσιο των παραγόντων (*Factors*) οι κατηγορικοί παράγοντες του μοντέλου ('Κινητικότητα', 'Αυτοεξυπηρέτηση', 'Συνηθισμένες δραστηριότητες', 'Πόνος-δυσφορία', 'Άγχος-θλίψη', 'Μέτρηση', 'Θεραπεία', 'Καπνιστής', 'Κύρια δραστηριότητα', 'Μετεκπαίδευση', 'Διπλώματα') και στο πλαίσιο των συμμεταβλητών (*Covariates*) οι συνεχείς παράγοντες ('Ηλικία'), (Εικόνα 5-3).

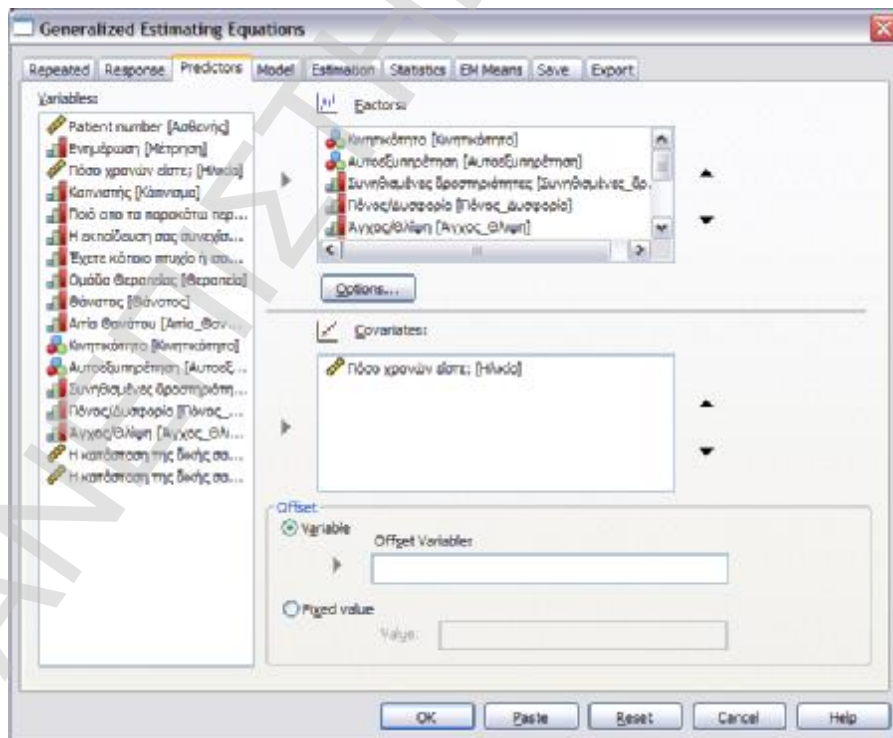
ΕΙΚΟΝΑ 5-1



EIKONA 5-2



EIKONA 5-3





### 5.5.1 Unstructured Working Correlation Matrix

Διενεργώντας λοιπόν την ανάλυση, με τους παράγοντες που αναφέρθηκαν προηγουμένως και επιλέγοντας την μη-δομημένη μορφή του χρησιμοποιούμενου πίνακα συνδιακυμάνσεων, δίνονται τα ακόλουθα αποτελέσματα.

Αρχικά όλες οι σχετικές πληροφορίες, όπως η εξαρτημένη μεταβλητή, η κατανομή που ακολουθεί, η συνάρτηση σύνδεσης, το πεδίο που ορίζει την ταυτότητα του ασθενή και αντίστοιχα αυτό που ορίζει την στιγμή της μέτρησης, καθώς και ο τύπος του χρησιμοποιούμενου πίνακα συσχέτισης (Πίνακας 5-3).

#### ΠΙΝΑΚΑΣ 5-3

Βασικά στοιχεία ορισμού του μοντέλου  
με μη-δομημένο πίνακα συνδιακυμάνσεων

Dependent Variable		Η κατάσταση της δικής σας υγείας σήμερα
Probability Distribution		Normal
Link Function		Identity
Subject Effect	1	Ασθενής
Within-Subject Effect	1	Ενημέρωση
Working Correlation Matrix Structure		Unstructured

Επίσης, δίνεται μια ανάλυση των εισαχθέντων παραγόντων (Πίνακας 5-4).

#### ΠΙΝΑΚΑΣ 5-4

Πληροφορίες σχετικά με το πλήθος των ασθενών και των μετρήσεων

Number of Levels	Subject Effect	Ασθενής	182
	Within-Subject Effect	Ενημέρωση	1
Number of Subjects			182
Number of	Minimum		1
Measurements per	Maximum		1
Correlation Matrix Dimension			1

Δίνεται ο αριθμός των υποκειμένων, έτσι όπως μετρείται με βάση τις χρονικές στιγμές. Τα επίπεδα της μεταβλητής των υποκειμένων (δηλαδή οι ασθενείς) είναι 182 και τα επίπεδα της μεταβλητής των μετρήσεων είναι 3.

Η διάσταση του πίνακα συσχέτισης θα πρέπει να είναι όσα είναι τα επίπεδα της μεταβλητής μετρήσεων, δηλαδή 3x3.

Οι ασθενείς δεν είναι 201 και η μικρότερη τιμή που παίρνει η μεταβλητή των μετρήσεων είναι 2 αντί 3, γιατί υπάρχουν κάποιες ελλείπουσες τιμές (των συμμεταβλητών) τις οποίες το πρόγραμμα εξαιρεί από μόνο του, με αποτέλεσμα κάποιοι ασθενείς που έχουν 3 μετρήσεις να καταλήγουν στην ανάλυση με δύο από αυτές.

Εφαρμόζοντας λοιπόν τον μη δομημένο (*unstructured*) πίνακα συσχέτισης, λαμβάνεται το πρώτο προς εξέταση μοντέλο με παράγοντες όλες τις κύριες επιδράσεις των παραγόντων ('Κινητικότητα', 'Αυτοεξυπηρέτηση', 'Συνηθισμένες Δραστηριότητες', 'Πόνος/Δυσφορία', 'Άγχος/Θλίψη', 'Θεραπεία', 'Μέτρηση', 'Κάπνισμα', 'Απασχόληση', 'Μετεκπαίδευση', 'Διπλώματα', 'Ηλικία').

Από τον Πίνακα 5-5(a), φαίνεται ότι η επίδραση των παραγόντων 'Μετεκπαίδευση', 'Διπλώματα' και 'Κινητικότητα' δεν είναι σημαντική με την επίδραση της ηλικίας να είναι οριακά στατιστικά σημαντική (Sig.=0,086). Στατιστικά σημαντικοί είναι οι παράγοντες, 'Αυτοεξυπηρέτηση', 'Συνηθισμένες Δραστηριότητες', 'Πόνος/Δυσφορία', 'Άγχος/Θλίψη', 'Μέτρηση', 'Θεραπεία', 'Κάπνισμα' και 'Απασχόληση'.

#### ΠΙΝΑΚΑΣ 5-5(a)

Επίδραση παραγόντων (Unstructured form)

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	183,595	1	,000
Κινητικότητα	2,888	1	,089
Αυτοεξυπηρέτηση	7,797	1	,005
Συνηθισμένες_ δραστηριότητες	32,796	2	,000
Πόνος_Δυσφορία	19,793	2	,000
Άγχος_Θλιψη	45,318	2	,000
Μέτρηση	12,516	2	,002
Θεραπεία	8,101	2	,017
Κάπνισμα	12,704	2	,002
Απασχόληση	8,162	3	,043
Μετεκπαίδευση	,381	1	,537
Διπλώματα	1,600	1	,206
Ηλικία	3,620	1	,057

Πριν εξαιρεθούν από το μοντέλο οι επιδράσεις αυτές θα πρέπει να εξεταστεί η

αλληλεπίδρασή τους με το παράγοντα ‘Θεραπεία’ και ‘Μέτρηση’ που αποτελούν 2 από τους βασικότερους και με το μεγαλύτερο ενδιαφέρον παράγοντες.

Εισάγοντας τις αλληλεπιδράσεις όλων των παραγόντων με τη ‘Θεραπεία’, προκύπτει ότι οι αλληλεπιδράσεις των παραγόντων ‘Πόνος/Δυσφορία’, ‘Άγχος/Θλίψη’, ‘Συνηθισμένες Δραστηριότητες’ και ‘Απασχόληση’ με την ‘Θεραπεία’ είναι στατιστικά σημαντικές (Πίνακας 5-5(b)). Κατά συνέπεια εξαιρούνται οι υπόλοιπες αλληλεπιδράσεις, καθώς και οι παράγοντες ‘Μετεκπαίδευση’ και ‘Διπλώματα’ και ‘Ηλικία’.

### ΠΙΝΑΚΑΣ 5-5(b)

Επίδραση παραγόντων (Unstructured form)

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	209,030	1	,000
Κινητικότητα	3,505	1	,061
Αυτοεξυπηρέτηση	4,874	1	,027
Συνηθισμένες_δραστηριότητες	31,550	2	,000
Πόνος_Δυσφορία	16,761	2	,000
Άγχος_Θλιψη	43,343	2	,000
Μέτρηση	12,420	2	,002
Θεραπεία	4,957	2	,084
Κάπνισμα	10,638	2	,005
Απασχόληση	5,071	3	,167
Μετεκπαίδευση	1,422	1	,233
Διπλώματα	3,006	1	,083
Ηλικία	2,859	1	,091
Μέτρηση * Θεραπεία	2,309	4	,679
Θεραπεία * Κινητικότητα	,703	2	,704
Θεραπεία * Αυτοεξυπηρέτηση	,635	2	,728
Θεραπεία * Συνηθισμένες_δραστηριότητες	7,021	4	,135
Θεραπεία * Πόνος_Δυσφορία	13,730	4	,008
Θεραπεία * Άγχος_Θλιψη	7,202	4	,126
Θεραπεία * Κάπνισμα	1,623	4	,805
Θεραπεία * Μετεκπαίδευση	,369	2	,831
Θεραπεία * Διπλώματα	,900	2	,638
Θεραπεία * Απασχόληση	4,242	4	,374

Εφαρμόζοντας λοιπόν αυτές τις εξαιρέσεις, και εξετάζοντας τις αλληλεπιδράσεις των επαναλαμβανόμενων στο χρόνο παραγόντων (‘Κινητικότητα’, ‘Αυτοεξυπηρέτηση’, ‘Συνηθισμένες Δραστηριότητες’, ‘Πόνος/Δυσφορία’, ‘Άγχος/Θλίψη’) με την ‘Μέτρηση’, φαίνεται να είναι στατιστικά σημαντική η αλληλεπίδραση του παράγοντα ‘Κινητικότητα’ με την ‘Μέτρηση’ καθώς και οι αλληλεπιδράσεις ‘Απασχόληση’, ‘Πόνος/Δυσφορία’ και ‘Άγχος/Θλίψη’ με την ‘Θεραπεία’ (Πίνακας 5-5(c)).

**ΠΙΝΑΚΑΣ 5-5(c)**

Επίδραση παραγόντων (Unstructured form)

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	1512,705	1	,000
Κινητικότητα	4,023	1	,045
Αυτοεξυπηρέτηση	4,144	1	,042
Συνηθισμένες_δραστηριότητες	26,231	2	,000
Πόνος_Δυσφορία	17,498	2	,000
Άγχος_Θλιψη	47,664	2	,000
Μέτρηση	1,168	2	,558
Θεραπεία	6,430	2	,040
Κάπνισμα	9,829	2	,007
Απασχόληση	8,220	3	,042
Θεραπεία * Πόνος_Δυσφορία	17,627	4	,001
Θεραπεία * Άγχος_Θλιψη	7,215	4	,125
Θεραπεία * Απασχόληση	7,815	4	,099
Θεραπεία * Συνηθισμένες_δραστηριότητες	6,303	4	,178
Μέτρηση * Κινητικότητα	5,559	2	,062
Μέτρηση * Αυτοεξυπηρέτηση	,101	2	,951
Μέτρηση * Συνηθισμένες_δραστηριότητες	6,202	4	,185
Μέτρηση * Πόνος_Δυσφορία	2,293	4	,682
Μέτρηση * Άγχος_Θλιψη	6,028	4	,197

Εξαιρώντας λοιπόν όλες τις υπόλοιπες αλληλεπιδράσεις καταλήγουμε σε ένα σύνολο 9 κύριων επιδράσεων και 4 αλληλεπιδράσεων, (Πίνακας 5-5(d)) .

**ΠΙΝΑΚΑΣ 5-5(d)**

Επίδραση παραγόντων (Unstructured form)

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	1102,464	1	,000
Κινητικότητα	2,584	1	,108
Αυτοεξυπηρέτηση	6,982	1	,008
Συνηθισμένες_δραστηριότητες	32,696	2	,000
Πόνος_Δυσφορία	19,848	2	,000
Άγχος_Θλιψη	42,434	2	,000
Μέτρηση	15,745	2	,000
Θεραπεία	3,869	2	,144
Κάπνισμα	12,474	2	,002
Απασχόληση	8,269	3	,041
Θεραπεία * Πόνος_Δυσφορία	17,493	4	,002
Θεραπεία * Άγχος_Θλιψη	10,423	4	,034
Θεραπεία * Απασχόληση	8,699	4	,069
Μέτρηση * Κινητικότητα	5,918	2	,052

Στον πίνακα που ακολουθεί (Πίνακας 5-5(e)) φαίνονται οι εκτιμήσεις των παραμέτρων

αυτών των παραγόντων, από όπου δύναται κανείς να κρίνει την σημαντικότητα αυτών των εκτιμήσεων.

### ΠΙΝΑΚΑΣ 5-5(ε)

#### Εκτίμηση παραγόντων (Unstructured form)

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	47,700	9,6754	28,736	66,663	24,305	1	,000
[Κινητικότητα=1]	2,261	2,3390	-2,323	6,845	,934	1	,334
[Κινητικότητα=2]	0	.	.	.	.	.	.
[Αυτοεξυπηρέτηση=1]	4,656	1,7620	1,202	8,109	6,982	1	,008
[Αυτοεξυπηρέτηση=2]	0	.	.	.	.	.	.
[Συνηθισμένες δραστηριότητες=1]	17,375	4,3463	8,857	25,894	15,982	1	,000
[Συνηθισμένες δραστηριότητες=2]	8,967	4,1567	,820	17,114	4,654	1	,031
[Συνηθισμένες δραστηριότητες=3]	0	.	.	.	.	.	.
[Πόνος_Δυσφορία=1]	,044	7,7529	-15,151	15,240	,000	1	,995
[Πόνος_Δυσφορία=2]	-3,790	7,4888	-18,468	10,888	,256	1	,613
[Πόνος_Δυσφορία=3]	0	.	.	.	.	.	.
[Άγχος_Όλιψη=1]	7,543	5,1210	-2,494	17,580	2,170	1	,141
[Άγχος_Όλιψη=2]	7,397	4,6810	-1,777	16,572	2,497	1	,114
[Άγχος_Όλιψη=3]	0	.	.	.	.	.	.
[Μέτρηση=1]	-5,850	1,5494	-8,887	-2,813	14,256	1	,000
[Μέτρηση=2]	-1,322	1,1765	-3,628	,984	1,263	1	,261
[Μέτρηση=3]	0	.	.	.	.	.	.
[Θεραπεία=1]	-3,259	9,7188	-22,308	15,789	,112	1	,737
[Θεραπεία=2]	-20,67	8,7464	-37,814	-3,529	5,586	1	,018
[Θεραπεία=3]	0	.	.	.	.	.	.
[Κάπνισμα=1]	-7,513	2,1306	-11,688	-3,337	12,433	1	,000
[Κάπνισμα=2]	-2,195	1,7858	-5,695	1,305	1,510	1	,219
[Κάπνισμα=3]	0	.	.	.	.	.	.
[Απασχόληση=1]	6,112	4,3692	-2,451	14,676	1,957	1	,162
[Απασχόληση=2]	9,525	3,7747	2,127	16,923	6,367	1	,012
[Απασχόληση=3]	6,752	3,3400	,205	13,298	4,086	1	,043
[Απασχόληση=5]	0	.	.	.	.	.	.
[Θεραπεία=1] * [Πόνος_Δυσφορία=1]	-3,419	9,3828	-21,809	14,971	,133	1	,716
[Θεραπεία=1] * [Πόνος_Δυσφορία=2]	-7,167	9,1102	-25,023	10,689	,619	1	,431
[Θεραπεία=1] * [Πόνος_Δυσφορία=3]	0	.	.	.	.	.	.
[Θεραπεία=2] * [Πόνος_Δυσφορία=1]	14,136	8,5395	-2,601	30,873	2,740	1	,098
[Θεραπεία=2] * [Πόνος_Δυσφορία=2]	12,765	8,1414	-3,192	28,722	2,458	1	,117
[Θεραπεία=2] * [Πόνος_Δυσφορία=3]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Πόνος_Δυσφορία=1]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Πόνος_Δυσφορία=2]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Πόνος_Δυσφορία=3]	0	.	.	.	.	.	.
[Θεραπεία=1] * [Άγχος_Όλιψη=1]	9,661	5,9208	-1,944	21,265	2,662	1	,103
[Θεραπεία=1] * [Άγχος_Όλιψη=2]	,472	5,3725	-10,058	11,002	,008	1	,930
[Θεραπεία=1] * [Άγχος_Όλιψη=3]	0	.	.	.	.	.	.
[Θεραπεία=2] * [Άγχος_Όλιψη=1]	9,945	6,0158	-1,846	21,735	2,733	1	,098
[Θεραπεία=2] * [Άγχος_Όλιψη=2]	1,896	5,4800	-8,845	12,637	,120	1	,729
[Θεραπεία=2] * [Άγχος_Όλιψη=3]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Άγχος_Όλιψη=1]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Άγχος_Όλιψη=2]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Άγχος_Όλιψη=3]	0	.	.	.	.	.	.
[Θεραπεία=1] * [Απασχόληση=1]	3,696	4,3702	-4,869	12,262	,715	1	,398
[Θεραπεία=1] * [Απασχόληση=2]	-5,642	4,7494	-14,951	3,666	1,411	1	,235
[Θεραπεία=1] * [Απασχόληση=3]	0	.	.	.	.	.	.
[Θεραπεία=2] * [Απασχόληση=1]	8,787	3,9984	,950	16,623	4,829	1	,028
[Θεραπεία=2] * [Απασχόληση=2]	-3,505	4,2973	-11,927	4,918	,665	1	,415
[Θεραπεία=2] * [Απασχόληση=3]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Απασχόληση=1]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Απασχόληση=2]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Απασχόληση=3]	0	.	.	.	.	.	.
[Μέτρηση=1] * [Κινητικότητα=1]	4,144	1,8430	,532	7,756	5,056	1	,025
[Μέτρηση=1] * [Κινητικότητα=2]	0	.	.	.	.	.	.
[Μέτρηση=2] * [Κινητικότητα=1]	-,630	1,4569	-3,486	2,225	,187	1	,665
[Μέτρηση=2] * [Κινητικότητα=2]	0	.	.	.	.	.	.
[Μέτρηση=3] * [Κινητικότητα=1]	0	.	.	.	.	.	.
[Μέτρηση=3] * [Κινητικότητα=2]	0	.	.	.	.	.	.
(Scale)	146,1	.	.	.	.	.	.

Από τον πίνακα 5-5(ε) κρίνεται απαραίτητο να εξαιρεθούν οι παράγοντες αλληλεπίδρασης

*‘Πόνος/Δυσφορία’, ‘Άγχος/Θλίψη’ με την ‘Θεραπεία’ και ‘Κινητικότητα’ με ‘Μέτρηση’ καθώς και οι κύριες επιδράσεις ‘Πόνος/Δυσφορία’, ‘Άγχος/Θλίψη’ και ‘Κινητικότητα’.* Οι εκτιμήσεις για τον παράγοντα *‘Μέτρηση’*, διαφοροποιούν στατιστικά σημαντικά την πρώτη από την τρίτη μέτρηση, με αρνητικό συντελεστή που σημαίνει ότι κατά την τρίτη μέτρηση η κατάσταση της υγείας του ασθενή είναι καλύτερη. Δηλαδή, για έναν ασθενή με τις ίδιες καταγραφές της κατάστασης της υγείας του με βάση τις πέντε διαστάσεις του ερωτηματολογίου, η κατάσταση της υγείας του καλυτερεύει από την πρώτη στην τρίτη μέτρηση κατά 5,85 μονάδες στην κλίμακα μέτρησης. Επίσης, οι εκτιμήσεις για τον παράγοντα *‘Θεραπεία’*, διαφοροποιούν στατιστικά σημαντικά την δεύτερη από την τρίτη θεραπεία, με αρνητικό συντελεστή που σημαίνει ότι η τρίτη θεραπεία είναι καλύτερη από την δεύτερη.

Πραγματοποιώντας λοιπόν αυτές τις αλλαγές, το μοντέλο που προκύπτει έχει 7 παράγοντες (μεταξύ των οποίων μια αλληλεπίδραση) με τις παρακάτω εκτιμήσεις παραμέτρων (Πίνακας 5-5(f)).

Στατιστικά σημαντικές διαφοροποιήσεις, φαίνονται στον παράγοντα *‘Αυτοεξυπηρέτηση’*, στον παράγοντα *‘Συνηθισμένες Δραστηριότητες’* μεταξύ των αποκρίσεων *‘Κανένα πρόβλημα’* και *‘Ανικανότητα εκτέλεσης Συνηθισμένων Δραστηριοτήτων’*, στον παράγοντα *‘Μέτρηση’*, στον παράγοντα *‘Κάπνισμα’* μεταξύ των αποκρίσεων *‘Καπνιστής’* και *‘Μη Καπνιστής’* και στον παράγοντα *‘Απασχόληση’*.

### ΠΙΝΑΚΑΣ 5-5(f)

Εκτίμηση παραγόντων (Unstructured form)

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	51,223	4,9591	41,504	60,943	106,692	1	,000
[Αυτοεξυπηρέτηση=1]	8,488	1,7719	5,015	11,960	22,944	1	,000
[Αυτοεξυπηρέτηση=2]	0	.	.	.	.	.	.
[Συνηθισμένες_δραστηριότητες=1]	19,170	4,4006	10,545	27,795	18,976	1	,000
[Συνηθισμένες_δραστηριότητες=2]	6,469	4,3034	-1,965	14,904	2,260	1	,133
[Συνηθισμένες_δραστηριότητες=3]	0	.	.	.	.	.	.
[Μέτρηση=1]	-5,108	1,1095	-7,283	-2,934	21,200	1	,000
[Μέτρηση=2]	-2,474	,8410	-4,123	-,826	8,657	1	,003
[Μέτρηση=3]	0	.	.	.	.	.	.
[Θεραπεία=1]	-5,199	2,9315	-10,945	,546	3,146	1	,076
[Θεραπεία=2]	-5,081	2,9274	-10,818	,657	3,012	1	,083
[Θεραπεία=3]	0	.	.	.	.	.	.
[Κάπνισμα=1]	-8,050	2,2953	-12,549	-3,551	12,300	1	,000
[Κάπνισμα=2]	-3,579	2,0518	-7,600	,443	3,042	1	,081
[Κάπνισμα=3]	0	.	.	.	.	.	.
[Απασχόληση=1]	7,207	3,6095	,132	14,281	3,986	1	,046
[Απασχόληση=2]	9,389	3,5416	2,448	16,331	7,028	1	,008
[Απασχόληση=3]	6,979	3,0179	1,064	12,894	5,349	1	,021
[Απασχόληση=5]	0	.	.	.	.	.	.
[Θεραπεία=1] * [Απασχόληση=1]	5,478	4,6495	-3,635	14,591	1,388	1	,239
[Θεραπεία=1] * [Απασχόληση=2]	-6,509	5,4538	-17,199	4,180	1,425	1	,233
[Θεραπεία=1] * [Απασχόληση=3]	0	.	.	.	.	.	.
[Θεραπεία=2] * [Απασχόληση=1]	10,531	4,4529	1,803	19,258	5,593	1	,018
[Θεραπεία=2] * [Απασχόληση=2]	-4,497	4,8653	-14,033	5,039	,854	1	,355
[Θεραπεία=2] * [Απασχόληση=3]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Απασχόληση=1]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Απασχόληση=2]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Απασχόληση=3]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Απασχόληση=5]	0	.	.	.	.	.	.
(Scale)	180,661						

#### 5.5.2 Exchangeable Working Correlation Matrix

Διενεργώντας λοιπόν την ανάλυση όπως προηγουμένως και επιλέγοντας την compound symmetry μορφή του χρησιμοποιούμενου πίνακα συνδιακυμάνσεων, προκύπτει ότι τα αποτελέσματα είναι παρόμοια, καταλήγωντας έτσι σε ένα μοντέλο με τους ίδιους παράγοντες. Οι εκτιμήσεις των παραμέτρων για το δεύτερο αυτό μοντέλο δίνονται στον Πίνακα 5-6.

## ΠΙΝΑΚΑΣ 5-6

Εκτίμηση παραγόντων (Exchangeable form)

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	51,268	4,9579	41,551	60,986	106,929	1	,000
[Αυτοεξυπηρέτηση=1]	8,762	1,7799	5,273	12,250	24,233	1	,000
[Αυτοεξυπηρέτηση=2]	0	.	.	.	.	.	.
[Συνηθισμένες_δραστηριότητες=1]	18,813	4,4605	10,071	27,555	17,789	1	,000
[Συνηθισμένες_δραστηριότητες=2]	6,011	4,3287	-2,474	14,495	1,928	1	,165
[Συνηθισμένες_δραστηριότητες=3]	0	.	.	.	.	.	.
[Μέτρηση=1]	-5,118	1,1095	-7,293	-2,944	21,279	1	,000
[Μέτρηση=2]	-2,454	,8417	-4,104	-,804	8,501	1	,004
[Μέτρηση=3]	0	.	.	.	.	.	.
[Θεραπεία=1]	-5,141	2,9239	-10,872	,590	3,092	1	,079
[Θεραπεία=2]	-5,108	2,9064	-10,804	,589	3,088	1	,079
[Θεραπεία=3]	0	.	.	.	.	.	.
[Κάπνισμα=1]	-8,099	2,2827	-12,573	-3,625	12,589	1	,000
[Κάπνισμα=2]	-3,516	2,0258	-7,487	,454	3,013	1	,083
[Κάπνισμα=3]	0	.	.	.	.	.	.
[Απασχόληση=1]	7,298	3,5819	,277	14,318	4,151	1	,042
[Απασχόληση=2]	9,446	3,5248	2,537	16,354	7,181	1	,007
[Απασχόληση=3]	7,052	2,9686	1,234	12,871	5,644	1	,018
[Απασχόληση=5]	0	.	.	.	.	.	.
[Θεραπεία=1] * [Απασχόληση=1]	5,480	4,6050	-3,546	14,505	1,416	1	,234
[Θεραπεία=1] * [Απασχόληση=2]	-6,261	5,3939	-16,833	4,310	1,348	1	,246
[Θεραπεία=1] * [Απασχόληση=3]	0	.	.	.	.	.	.
[Θεραπεία=2] * [Απασχόληση=1]	10,397	4,4261	1,722	19,072	5,517	1	,019
[Θεραπεία=2] * [Απασχόληση=2]	-4,347	4,8608	-13,874	5,180	,800	1	,371
[Θεραπεία=2] * [Απασχόληση=3]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Απασχόληση=1]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Απασχόληση=2]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Απασχόληση=3]	0	.	.	.	.	.	.
[Θεραπεία=3] * [Απασχόληση=5]	0	.	.	.	.	.	.
(Scale)	180,360						

Συγκρίνοντας τους πίνακες 5-5(f) και 5-6, παρατηρείται ότι οι εκτιμήσεις είναι πάρα πολύ κοντά. Επομένως το επόμενο βήμα είναι να επιλεγεί μια από τις δύο μορφές του χρησιμοποιούμενου πίνακα συνδιακυμάνσεων.

### 5.5.3 Επιλογή πίνακα συνδιακυμάνσεων

Η επιλογή του κατάλληλου μοντέλου για την περιγραφή ενός συνόλου δεδομένων, είναι πολύ σημαντική. Ένα πλεονέκτημα της μεθόδου γενικευμένων εξισώσεων εκτίμησης είναι



ότι υπολογίζει συνεπείς εκτιμητές παραμέτρων ακόμα και όταν ο χρησιμοποιούμενος πίνακας συσχέτισης έχει οριστεί λανθασμένα. Μία μέθοδος που έχει προταθεί, για την επιλογή του κατάλληλου χρησιμοποιούμενου πίνακα συνδιακυμάνσεων μέσω υπο-δειγμάτων, είναι από τους Pan and Connett, 2002. Ένα κριτήριο επιλογής είναι το Quasi-likelihood κάτω από ανεξαρτησία (QIC) που μπορεί να χρησιμοποιηθεί για την επιλογή ενός μοντέλου μεταξύ διαφορετικών δομών συσχέτισης. Αυτή με την μικρότερη τιμή του QIC είναι και η καλύτερη δομή με βάση αυτό το κριτήριο (Cui).

Αποτελεί μια προέκταση του κριτηρίου Akaike (AIC), που χρησιμοποιείται ευρέως για την επιλογή μοντέλου κυρίως σε ανάλυση παλινδρόμησης ανεξάρτητων δεδομένων. Το AIC βασίζεται στην πιθανοφάνεια και στις ασυμπτωτικές υποθέσεις για τον εκτιμητή μεγίστης πιθανοφάνειας (MLE). Καθότι στις γενικευμένες εξισώσεις εκτίμησης, δεν γίνεται καμία υπόθεση για την κατανομή και κατά συνέπεια δεν ορίζεται πιθανοφάνεια, το κριτήριο AIC δεν είναι άμεσα εφαρμόσιμο. Το κριτήριο λοιπόν QIC βασίζεται στη χρήση της πιθανοφάνειας που ορίζεται κάτω από την ανεξάρτητη δομή του χρησιμοποιούμενου πίνακα συνδιακυμάνσεων (χάριν απλότητας) και των ανθεκτικών εκτιμήσεων των συντελεστών παλινδρόμησης (Pan, 2001).

Λαμβάνοντας λοιπόν την τιμή του μέτρου QIC για τα μοντέλα με τις 2 δομές συσχέτισης, παρατίθεται ο παρακάτω πίνακας (Πίνακας 5-7).

**ΠΙΝΑΚΑΣ 5-7**

	Unstructured	Exchangeable
QIC	95.804	95.644

Από τον Πίνακα 5-7 διακρίνεται ότι η compound symmetry (exchangeable) δομή είναι η καταλληλότερη, καθώς έχει την μικρότερη τιμή για το κριτήριο QIC.

#### **5.5.4 Αξιολόγηση Μοντέλου**

Από τις εκτιμήσεις των παραμέτρων στον πίνακα 5-6 για την μέτρηση διακρίνεται ότι κατά την τρίτη μέτρηση καλύτερεύει η κατάσταση της υγείας του ασθενή (όπως αξιολογείται από τον ίδιο). Επίσης η 3<sup>η</sup> θεραπεία (P only), φαίνεται να προσφέρει στον ασθενή καλύτερη κατάσταση από τις άλλες δύο.

Για παράδειγμα, ένας ασθενής ο οποίος βρίσκεται στην 1<sup>η</sup> ομάδα θεραπείας 'P+Cb', χωρίς προβλήματα αυτοεξυπηρέτησης, συνηθισμένων δραστηριοτήτων, μη καπνιστής και ο

ο οποίος ασχολείται με τα οικιακά, με βάση το μοντέλο θα αξιολογήσει την κατάστασή του κατά την 2<sup>η</sup> μέτρηση (κατά την ολοκλήρωση της θεραπείας) με 77,37 ενώ κατά την 3<sup>η</sup> μέτρηση (6 μήνες μετά τη θεραπεία) με 79,836.

Το μοντέλο είναι της μορφής,

Κατάσταση της δικής σας υγείας σήμερα=

$$51,268 + 8,762X_1 + 18,813X_{21} + 6,011X_{22} - 5,118X_{31} - 2,454X_{32} + 7,268X_{41} + 9,446X_{42} + 7,052X_{43} - 5,141X_{51} - 5,108X_{52} - 8,099X_{61} - 3,516X_{62} + 5,480X_{71} - 6,261X_{72} - 10,397X_{73} - 4,347X_{74}$$

Όπου  $X_1, X_2, X_3, X_4, X_5, X_6, X_7$  οι παράμετροι 'Αυτοεξυπηρέτηση', 'Συνηθισμένες Δραστηριότητες', 'Μέτρηση', 'Απασχόληση', 'Θεραπεία', 'Κάπνισμα', και 'Απασχόληση\*Θεραπεία' αντίστοιχα. Οι επιμέρους αποτελούν δείκτριες μεταβλητές αυτών.

Οι επόμενοι πίνακες έχουν τις υπολογισμένες μέσες αποκρίσεις για όλες τις μετρήσεις, για όλες τις θεραπείες, καθώς και των συνδυασμών αυτών (ΠΙΝΑΚΑΣ 5-8(a,b,c)).

#### ΠΙΝΑΚΑΣ 5-8(a)

Μέση απόκριση ανά θεραπεία

Ομάδα Θεραπείας	Mean	Std. Error	95% Wald Confidence Interval	
			Lower	Upper
P+Cb	60,05781	2,121340	55,90006	64,21556
G+Doc	62,36820	1,770674	58,89774	65,83866
P only	63,47660	1,703743	60,13733	66,81588

#### ΠΙΝΑΚΑΣ 5-8(b)

Μέση απόκριση ανά μέτρηση

Ενημέρωση	Mean	Std. Error	95% Wald Confidence Interval	
			Lower	Upper
1	59,52436	1,575811	56,43583	62,61289
2	62,18849	1,687833	58,88040	65,49658
3	64,64248	1,589832	61,52646	67,75849

### ΠΙΝΑΚΑΣ 5-8(c)

Μέση απόκριση ανά θεραπεία και ανά μέτρηση

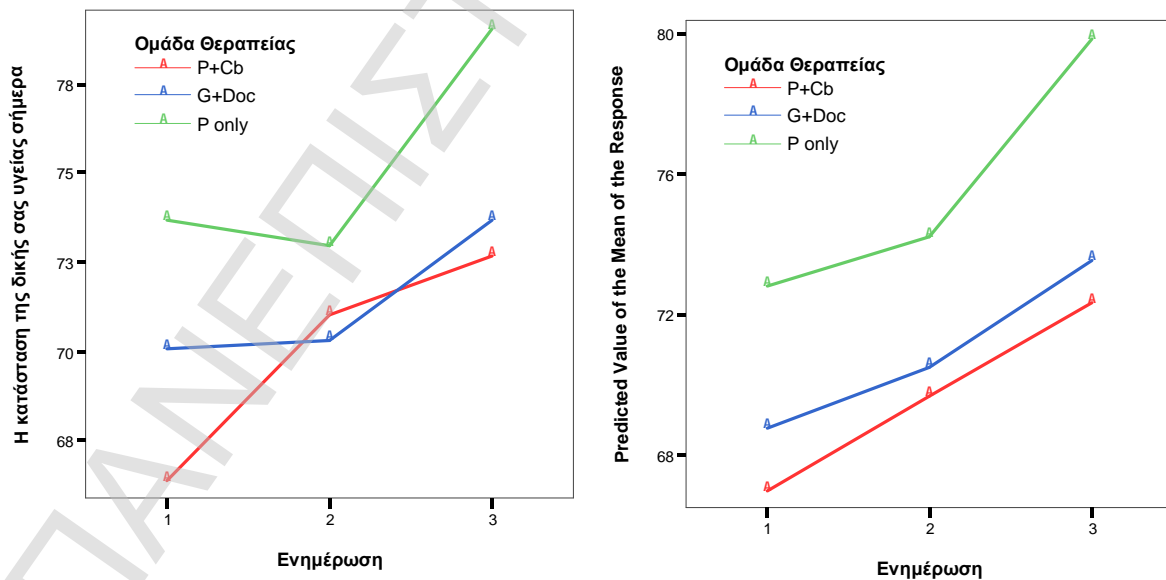
Ενημέρωση	Ομάδα Θεραπείας	Mean	Std. Error	95% Wald Confidence Interval	
				Lower	Upper
1	P+Cb	57,46372	2,212954	53,12641	61,80103
	G+Doc	59,77412	1,796551	56,25294	63,29529
	P only	60,88252	1,752324	57,44803	64,31701
2	P+Cb	60,12785	2,212896	55,79066	64,46505
	G+Doc	62,43825	1,935253	58,64522	66,23127
	P only	63,54665	1,896478	59,82962	67,26368
3	P+Cb	62,58184	2,181305	58,30656	66,85712
	G+Doc	64,89224	1,866810	61,23336	68,55112
	P only	66,00064	1,758655	62,55374	69,44754

Διακρίνουμε αρχικά, από τον Πίνακα 5-8(a) ότι η 3<sup>η</sup> θεραπεία (P only) έχει την μεγαλύτερη μέση απόκριση (καλύτερη κατάσταση υγείας). Ακολούθως από τον Πίνακα 5-8(b) φαίνεται ότι πλέον κατά την 3<sup>η</sup> μέτρηση η κατάσταση του ασθενή έχει καλυτερέψει.

Παρακάτω, δίνεται ένα σχήμα (Σχήμα 5-2), όπου απεικονίζονται οι μέσες αποκρίσεις ανά χρονική στιγμή και ανά θεραπεία.

### ΣΧΗΜΑ 5-2

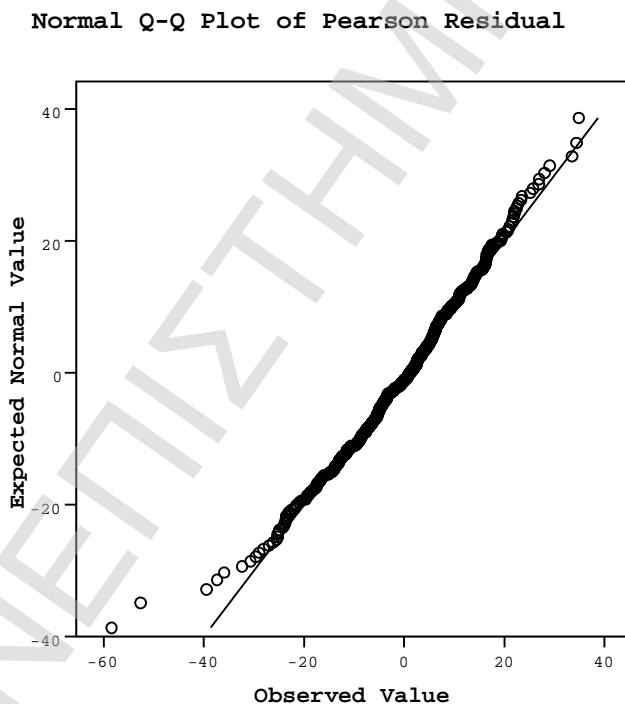
Μέσες αποκρίσεις ανά θεραπεία και μέτρηση (Πραγματική και Εκτιμηθείσα)



Από τα παραπάνω σχήματα φαίνεται ότι η πορεία της κατάστασης της υγείας των ασθενών, στη διάρκεια του χρόνου είναι πιο ομαλή με βάση τις εκτιμήσεις. Ακόμα φαίνεται ότι διαχωρίζονται πιο ξεκάθαρα οι θεραπείες P+Cb και G+Doc, όπου από το διάγραμμα των πραγματικών δεδομένων δεν διακρίνεται (ιδίως στην δεύτερη μέτρηση) ποια από τις δύο υπερσχύει. Όλες οι θεραπείες έχουν μια αυξητική τάση, που σημαίνει ότι και οι τρεις προσφέρουν στον ασθενή μια καλύτερη κατάσταση της υγείας του. Εξαρχής, αποτελεσματικότερη φαίνεται να είναι η τρίτη θεραπεία (P only).

Στο παρακάτω σχήμα (ΣΧΗΜΑ 5-3) απεικονίζονται τα κατάλοιπα του μοντέλου σε ένα Q-Q Plot, από όπου φαίνεται ότι ακολουθούν την Κανονική κατανομή.

**ΣΧΗΜΑ 5-3**  
Q-Q Plot καταλοίπων



Τέλος, στον πίνακα 5-9 δίνεται ο χρησιμοποιούμενος πίνακας συσχέτισης (exchangeable), από όπου φαίνεται ότι η συσχέτιση των μετρήσεων είναι 0,435.

### ΠΙΝΑΚΑΣ 5-9

Χρησιμοποιούμενος πίνακας συσχέτισης

Measurement	Measurement		
	[Μέτρηση = 1]	[Μέτρηση = 2]	[Μέτρηση = 3]
[Μέτρηση = 1]	1,000	,456	,456
[Μέτρηση = 2]	,456	1,000	,456
[Μέτρηση = 3]	,456	,456	1,000

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

# Κ Ε Φ Α Λ Α Ι Ο 6

## Επίλογος

Τα δεδομένα τα οποία μελετήθηκαν αποτελούσαν στοιχεία και μετρήσεις γυναικών από μια κλινική δοκιμή μελέτης του καρκίνου του μαστού. Οι ασθενείς έλαβαν με τυχαίο τρόπο μια από τρεις θεραπείες και παρακολουθήθηκαν για τρεις χρονικές στιγμές: στην αρχή της θεραπείας, κατά την ολοκλήρωσή της και έξι μήνες μετά. Σε κάθε μέτρηση, οι ασθενείς καλούνταν να συμπληρώσουν το ερωτηματολόγιο EQ-5D, για την κατάσταση της δικής τους υγείας όπως αξιολογείται από τους ίδιους.

Σε μια διαχρονική μελέτη ο απώτερος στόχος μιας ανάλυσης και της προσαρμογής ενός μοντέλου είναι η πορεία του υποκειμένου στη διάρκεια του χρόνου, καθώς και η ανωτερότητα μιας θεραπείας έναντι μιας άλλης.

Συχνά σε διαχρονικές μελέτες υπάρχουν ελλειπίς παρατηρήσεις, κάτι που κάνει περίπλοκη την ανάλυση των δεδομένων. Οι ελλειπίς παρατηρήσεις μπορεί να λείπουν εντελώς τυχαία, τυχαία ή μη τυχαία. Έχουν προταθεί διάφοροι τρόποι αντιμετώπισης αυτών. Εδώ επιλέχθηκε να περιληφθούν στην ανάλυση οι ασθενείς οι οποίοι είχαν πλήρη στοιχεία.

Η ανάλυση επαναλαμβανόμενων μετρήσεων βρίσκει πολύ συχνά εφαρμογή σε βιοιατρικές μελέτες και κλινικές δοκιμές. Ανάλογα με την φύση των δεδομένων, έχουν μελετηθεί διάφορες μέθοδοι ανάλυσης δεδομένων τέτοιας φύσεως. Η μέθοδος που χρησιμοποιήθηκε εδώ ήταν των γενικευμένων εξισώσεων εκτίμησης (GEE).

Οι βασικές υποθέσεις που πρέπει να γίνουν για την εφαρμογή ενός μοντέλου γενικευμένων εξισώσεων εκτίμησης είναι ο ορισμός της κατανομής που ακολουθεί η εξεταζόμενη απόκριση, η επιλογή της κατάλληλης συνάρτησης σύνδεσης (ανάλογα με την κατανομή) και η επιλογή της δομής του χρησιμοποιούμενου πίνακα συσχέτισης.

Εφαρμόζοντας λοιπόν τη μέθοδο γενικευμένων εξισώσεων εκτίμησης προκύπτει ένα μοντέλο το οποίο λαμβάνει υπόψη του τους παράγοντες 'Αυτοεξυπηρέτηση', 'Συνηθισμένες Δραστηριότητες', 'Μέτρηση', 'Απασχόληση', 'Θεραπεία', 'Κάπνισμα', και 'Απασχόληση\*Θεραπεία'. Από την αξιολόγηση του μοντέλου διακρίνεται ότι η τρίτη θεραπεία (P only) είναι ανώτερη από τις άλλες δύο με αισθητή τη διαφορά από την πρώτη κιάλας μέτρηση. Οι άλλες δύο θεραπείες προσφέρουν στον ασθενή μια κατάσταση υγείας

καλή και παρόμοια, με την δεύτερη θεραπεία (G+Doc) να είναι λίγο καλύτερη. Επίσης όταν ο ασθενής στην καθημερινότητά του ασχολείται με κάτι (Υπάλληλος ή Οικιακά) η κατάσταση της υγείας του φαίνεται να είναι καλύτερη από ότι είναι για κάποιον ασθενή που είναι συνταξιούχος. Αναμενόμενο εύρημα αποτελεί η πολύ καλή κατάσταση της υγείας του ασθενή όταν αυτός δεν έχει προβλήματα αυτοεξυπηρέτησης και μπορεί να κάνει τις συνηθισμένες του δραστηριότητες. Τέλος ο συνδυασμός της δεύτερης θεραπείας (G+Doc) με την απασχόληση του ασθενή ως συνταξιούχος, φαίνεται να επιδρά θετικά στην κατάσταση της υγείας του ασθενή από ότι για τις άλλες θεραπείες.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑΣ







Table with columns for ID and numerical values, containing data from 110209171 to 110206048.







## Ερωτηματολόγιο για την Υγεία

### (Greek version)

Βάζοντας ένα ÷ σε ένα κουτάκι κάθε ομάδας παρακάτω, παρακαλούμε σημειώστε ποιες δηλώσεις περιγράφουν καλύτερα την κατάσταση της υγείας σας σήμερα.

#### Κινητικότητα

- |                                       |   |
|---------------------------------------|---|
| Δεν έχω κανένα πρόβλημα στο περπάτημα | α |
| Έχω μερικά προβλήματα στο περπάτημα   | α |
| Είμαι καθλωμένος/η στο κρεβάτι        | α |

#### Αυτοεξυπηρέτηση

- |   |   |
|---|---|
| Δεν έχω κανένα πρόβλημα με την αυτοεξυπηρέτησή μου    | α |
| Έχω μερικά προβλήματα στο να πλένομαι και να ντύνομαι | α |
| Είμαι ανάκανος/η να πλυθώ ή να ντυθώ                  | α |

#### Συνηθισμένες Δραστηριότητες (π.χ. δουλειά, μελέτη, νοικοκυριό, οικογενειακές δραστηριότητες ή δραστηριότητες ελεύθερου χρόνου)

- |   |   |
|---|---|
| Δεν έχω κανένα πρόβλημα στο να εκτελώ τις συνηθισμένες δραστηριότητές μου | α |
| Έχω μερικά προβλήματα στο να εκτελώ τις συνηθισμένες δραστηριότητές μου   | α |
| Είμαι ανάκανος/η να εκτελώ τις συνηθισμένες δραστηριότητές μου            | α |

#### Πόνος/Δυσφορία

- |                                 |   |
|---------------------------------|---|
| Δεν έχω καθόλου πόνο ή δυσφορία | α |
| Έχω μέτριο πόνο ή δυσφορία      | α |
| Έχω υπερβολικό πόνο ή δυσφορία  | α |

#### Άγχος /Θλίψη

- |                              |   |
|------------------------------|---|
| Δεν έχω άγχος ή θλίψη        | α |
| Έχω μέτριο άγχος ή θλίψη     | α |
| Έχω υπερβολικό άγχος ή θλίψη | α |

Για να βοηθήσουμε κάποιον να πει πόσο καλή ή κακή είναι μια κατάσταση υγείας, ζωγραφίσαμε μια κλίμακα (σαν ένα θερμόμετρο) πάνω στην οποία η καλύτερη κατάσταση που μπορείτε να φανταστείτε έχει βαθμό 100 και η χειρότερη κατάσταση που μπορείτε να φανταστείτε έχει βαθμό 0.

Θα θέλαμε να σημειώσετε πάνω σε αυτήν την κλίμακα πόσο καλή ή κακή είναι η υγεία σας σήμερα, κατά τη γνώμη σας. Παρακαλούμε κάντε το αυτό, τραβώντας μια γραμμή από το παρακάτω τετράγωνο προς οποιοδήποτε σημείο της κλίμακας δείχνει πόσο καλή ή κακή είναι η κατάσταση της υγείας σας σήμερα.

**Η κατάσταση  
της δικής σας  
υγείας σήμερα**

Η καλύτερη  
κατάσταση υγείας  
που μπορείτε να  
φανταστείτε



Η χειρότερη  
κατάσταση υγείας  
που μπορείτε να  
φανταστείτε



## Βιβλιογραφία

- Υφαντόπουλος, Γ. (2001) *Η Ελληνική απόδοση του EuroQol EQ-5D. Αρχεία Ελληνικής Ιατρικής*, 18(2) :181-191. Πανεπιστήμιο Αθηνών, Αθήνα.
- Afifi, A.A. and Elashoff, R.M. (1966). *Missing observations in Multivariate statistics I. Review of the literature*. Journal of the American Statistical Association, 61:594-604.
- Anderson, T.W. (1957). *Maximum Likelihood estimates for a multivariate normal distribution when some observations are missing*. Journal of the American Statistical Association, 521:200-203.
- Ballinger, G.A. (2004). *Using Generalized Estimating Equations for Longitudinal Data Analysis*. Organizational Research Methods, 7:127-150.
- Brooks, R. and Rabin, R. and De Charro, F. (2003). *The measurement and valuation of health status using EQ-5D: A European perspective*. Kluwer Academic, Publishers, Netherlands.
- Carriere, K.C. (1998). *Methods for repeated measures data analysis with missing values*. Journal of Statistical Planning and Inference, 7:221-236.
- Crowder, M. (1995), *On the use of a working correlation matrix in using generalized linear models for repeated measures*. Biometrika, 82:407-410.
- Cui, J. *QIC Program and model Selection in GEE analyses*. The Strata Journal, 7:209-220.
- Davis, S.D. (2002). *Statistical Methods for the Analysis of Repeated Measurements*, Springer, New York
- Everitt, B.S. (2006). *The Cambridge Dictionary of Statistics*. Cambridge University Press, UK.
- Fitzmaurice, G.M. and Laird, N.M. and Ware, J.H. (2004). *Applied Longitudinal Analysis*. John Wiley and Sons, New York.
- Fitzmaurice, G. and Davidian, M. and Verbeke, G. and Molenberghs, G. (2008). *Longitudinal Data Analysis*. Chapman & Hall//CRC.
- Grizzle, J.E. and Starmer, C. F. and Koch, G. G. (1969). *Analysis of Categorical data by linear models*. Biometrics, 25: 489-504.
- Hartley, H.O. (1956). *Programming analysis of variance for general purpose computers*. Biometrics, 12:110-122.
- Hartley, H.O. and Hocking, R.R. (1971). *Incomplete data analysis*. Biometrics, 27:783-823.

- Healy, M.J.R. and Westmacott, M. (1956). *Missing values in experiments analyzed on automatic computers*. Applied Statistics, 5:203-206.
- Hedeker, D. and Gibbons, R.D. (2006), *Longitudinal Data Analysis*. Wiley Series in Probability and Statistics.
- Hocking, R.R. and Smith, W.B.(1968). *Estimation of parameters in the multivariate normal distribution with missing observations*. Journal of the American Statistical Association, 63:159-173.
- Howell, D.C. (2009), *Treatment of Missing Data*. University of Vermont.
- Katz, M.H. (1999), *Multivariate Analysis: A practical Guide for Clinicians*. Cambridge University Press, Cambridge.
- Kimman, M.L. and Dirksen, C.D. and Lambin, P. and Boersma, L.J. (2009). *Responsiveness of the EQ-5D in breast cancer patients in their first year after treatment*. Health and Quality of Life Outcomes, 7:11.
- Koch, G. G. and Reinfurt, D. W. (1971). *The analysis of categorical data from mixed models*. Biometrics, 27:157.
- Koch, G.G. and Landis, J.R. and Freeman, J.L. and Freeman, D.H. and Lehnen, Jr. and R.G. (1977). *A general methodology for the analysis of experiments with the repeated measurement of categorical data*. Biometrics, 33:133.
- Liang, K.Y. and Zeger, S.L. (1986). *Longitudinal Data Analysis Using Generalized Linear Models*. Biometrika, 73: 13-22.
- Little, R.J.A and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*, Chichester. Wiley.
- Pan, W. (2001). *Akaike's Information Criterion in Generalized Estimating Equations*. Biometrics, 57: 120-125.
- Pan, W. and Connett, E. (2002). *Selecting the working correlation structure in generalized estimating equations with application to the lung health study*. Statistica Sinica, 12:475-490.
- Rubin, D.B. (1972). *A non iterative algorithm for least squares estimation of missing values in any analysis of variance design*. Applied Statistics, 21:136-141.
- Wilkinson, G.N. (1958). *Estimation of missing values for the analysis of incomplete data*. Biometrics, 14: 257-286.
- Xia, Q. and Hwang, S. S. and Chang, V. T. and Osenenko, P. and Alejandro, Y. and Yan, H. and Toomey, K. and Srinivas, S. (2005). *Validity, reliability and responsiveness of Euroqol (EQ5D) in patients (Pts) receiving palliative care (PC)*. Journal of Clinical Oncology.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ