

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΤΗΣ
ΕΛΛΗΝΙΚΗΣ ΒΙΟΤΡΑΠΕΖΑΣ ΓΙΑ
ΤΗ ΝΟΣΟ ΤΟΥ ΠΑΡΚΙΝΣΟΝ**

Δάφνη Πούλα

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
Νοέμβριος 2014

Πανεπιστήμιο Πειραιώς

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΤΗΣ
ΕΛΛΗΝΙΚΗΣ ΒΙΟΤΡΑΠΕΖΑΣ ΓΙΑ
ΤΗ ΝΟΣΟ ΤΟΥ ΠΑΡΚΙΝΣΟΝ**

Δάφνη Πούλα

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
Νοέμβριος 2014

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίσθηκε 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:

- Πολίτης Κωνσταντίνος, Αναπληρωτής Καθηγητής (Επιβλέπων)
- Αντζουλάκος Δημήτριος, Αναπληρωτής Καθηγητής
- Ευαγγελάρας Χαράλαμπος, Επίκουρος Καθηγητής

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**

**ANALYSIS OF DATA FROM
THE HELLENIC BIO-BANK ON
PARKINSON'S DISEASE**

By

Dafni Poula

MSc Dissertation

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

Piraeus, Greece
November 2014

Πανεπιστήμιο Πειραιώς

Ευχαριστίες

Θα ήθελα να ευχαριστήσω ιδιαίτερος τον Αναπληρωτή Καθηγητή του Τμήματος Στατιστικής και Ασφαλιστικής κ. Πολίτη Κωνσταντίνο, για τις γνώσεις που μου μετέδωσε κατά τη διεξαγωγή της εργασίας, καθώς και για τις χρήσιμες συμβουλές και την καθοδήγησή του. Τον ευχαριστώ πολύ που με εμπιστεύτηκε να φέρω εις πέρας την εργασία αυτή και με υπομονή με βοήθησε να την ολοκληρώσω.

Θα ήθελα επίσης να ευχαριστήσω θερμά τον κ. Αντζουλάκο Δημήτριο, καθώς και τον κ. Ευαγγελάρα Χαράλαμπο για τις συμβουλές και τις διορθώσεις τους στο κείμενο της εργασίας, ως μέλη της τριμελούς επιτροπής.

Θα ήθελα επίσης να ευχαριστήσω τον ερευνητή του Ιδρύματος Ιατροβιολογικών Ερευνών της Ακαδημίας Αθηνών κ. Βασιλάτη Δημήτριο για την εμπιστοσύνη που μου έδειξε και την παροχή της πληροφορίας από τη βιοτράπεζα, ώστε να διεκπαιρωθεί η παρούσα διπλωματική εργασία.

Τέλος, θα ήθελα να ευχαριστήσω τον ειδικό λειτουργικό επιστήμονα του Ιδρύματος Ιατροβιολογικών Ερευνών της Ακαδημίας Αθηνών κ. Μιχαλόπουλο Ιωάννη για την πολύτιμη βοήθειά του και τις υποδείξεις του κατά τη διάρκεια της εκπόνησης της διπλωματικής μου εργασίας.

Πανεπιστήμιο Πειραιώς

Περίληψη

Βιοστατιστική είναι η επιστήμη που μελετά την εφαρμογή στατιστικών μεθόδων σε ένα ευρύ φάσμα θεμάτων ιατρικής και βιολογίας. Αντικείμενό της είναι η σχεδίαση πειραμάτων, η συλλογή, η σύνοψη και η ανάλυση των δεδομένων με σκοπό την ερμηνεία και τη διεξαγωγή συμπερασμάτων διαμέσου των αποτελεσμάτων.

Ο ιδιοπαθής Παρκινσονισμός πρόκειται για μια διαταραχή των κινητικών νεύρων. Με την πάροδο της ασθένειας, τα συμπτώματά της προοδευτικά γίνονται εντονότερα. Προσβάλλει κυρίως άτομα ηλικίας μεγαλύτερης των 60 ετών, δεν είναι λίγα, όμως, τα περιστατικά που έχουν διαγνωσθεί με νεανικής μορφής Παρκινσονισμό.

Σκοπός της μελέτης αυτής είναι να ερευνήσει τους παράγοντες, περιβαλλοντικούς και γενετικούς, που μπορεί να επηρεάζουν την εμφάνιση της νόσου καθώς και να καταγράψει κλινικά, εργασιακά και δημογραφικά χαρακτηριστικά των πασχόντων ασθενών.

Στα πλαίσια της μελέτης εξετάζονται περιγραφικοί δείκτες, γραφήματα και κατανομές καταγραφής της δομής του ερωτηματολογίου μέσω του οποίου συλλέχθηκε η πληροφορία. Χρησιμοποιήθηκαν πίνακες διπλής εισόδου για τον έλεγχο της ανεξαρτησίας των μεταβλητών. Μέθοδοι πολυμεταβλητής ανάλυσης για την συσταδοποίηση των ασθενών σε επιμέρους ομάδες, καθώς και λογιστική παλινδρόμηση για την δόμηση προβλεπτικών μοντέλων για την σχετική πιθανότητα εμφάνισης της Νόσου Πάρκινσον.

Πανεπιστήμιο Πειραιώς

Abstract

Biostatistics is the science which studies the application of statistical methods to a wide range of topics of medicine and biology. Its objective is to design experiments, collect, summarize and analyze the data in order to interpret and draw conclusions throughout the results.

Idiopathic Parkinsonism is a disorder of the motor nerves . Over the course of the disease, its symptoms become progressively stronger. It affects mainly people older than 60 years, there are few, however, cases that have been diagnosed with a form of juvenile Parkinsonism.

This study aims to investigate the factors, both genetic and environmental, that can possibly affect the appearance of the disease and also record clinical, genetic and demographic characteristics of affected patients.

In course of the study, descriptive indexes, graphs, as well as the distribution of the structure of the questionnaire that has been used for the data collection are examined. Methods of multivariate analysis were used for clustering the patients in individual groups, while the logistic regression was applied so as to construct predictive models for the relative risk of developing Parkinson's disease (PD).

Πανεπιστήμιο Πειραιώς

Περιεχόμενα

Περίληψη	iix
Abstract.....	xi
Κατάλογος Πινάκων	xvii
Κατάλογος Διαγραμμάτων	xix
ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ.....	1
1.1 Γενικά για την Νόσο Πάρκινσον.....	1
1.2 Σκοπός της μελέτης.....	1
ΚΕΦΑΛΑΙΟ 2: ΓΕΝΙΚΑ ΣΤΟΙΧΕΙΑ ΓΙΑ ΤΗ ΝΟΣΟ ΤΟΥ ΠΑΡΚΙΝΣΟΝ	3
2.1 Εισαγωγή.....	3
2.2 Βάση Δεδομένων Ελλήνων Παρκινσονικών (Βιοτράπεζα).....	3
2.3 Η Νόσος Πάρκινσον	4
2.3.1 Νεανικής μορφής παρκινσονισμός.....	6
2.4 Ενοποιημένη Κλίμακα Βαθμολόγησης της Νόσου Πάρκινσον (Unified Parkinson’s Disease Rating Scale –UPDRS-).....	8
2.4.1 Κλίμακα UPDRS	8
2.4.2 Κλίμακα Hoehn and Yahr.....	10
2.4.3 Κλίμακα Schwab and England των καθημερινών δραστηριοτήτων (Activities of Daily Living –ADL-).....	11
2.4.4 Τυποποιημένη εξέταση της μινι-ψυχικής κατάστασης (Standardized mini-mental Status Examination).....	12
ΚΕΦΑΛΑΙΟ 3: ΠΕΡΙΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ.....	13
3.1 Εισαγωγή.....	13
3.1.1 Φύση των δεδομένων	13
3.2 Περιγραφική Ανάλυση για το σύνολο των εθελοντών	14
□ Φύλο	14
□ Ηλικία.....	15
□ Θρήσκευμα.....	16
□ Φυλή.....	17
□ Χώρα Γέννησης.....	18
□ Τύπος Τόπου Διαμονής	18

□ Συγγένεια εξ αίματος των γονέων	20
□ Οικογενειακό ιστορικό νόσου Πάρκινσον	20
□ Δεξιοχειρία / Αριστεροχειρία	22
□ Κάπνισμα.....	22
□ Κατανάλωση καφέ	28
□ Έκθεση σε ζιζανιοκτόνα ή/και φυτοφάρμακα	31
3.3 Περιγραφική Ανάλυση για το σύνολο των ασθενών	35
□ Έτος έναρξης της νόσου.....	35
□ Αρχικό σύμπτωμα	36
□ Πλευρά έναρξης	37
□ Συμπτώματα που εμφάνισαν	38
□ Αξιολόγηση των ασθενών	38
ΚΕΦΑΛΑΙΟ 4: ΠΙΝΑΚΕΣ ΣΥΝΑΦΕΙΑΣ.....	42
4.1 Εισαγωγή.....	42
4.2 Θεωρητικό υπόβαθρο.....	42
4.2.1 Έλεγχος κανονικότητας Kolmogorov-Smirnov για ένα πληθυσμό.....	42
4.2.2 Έλεγχος Mann-Whitney	43
4.2.3 Έλεγχος ισότητας των διακυμάνσεων δύο πληθυσμών	44
4.2.4 Έλεγχος για τη διαφορά των μέσων τιμών δύο πληθυσμών (t-test).....	44
4.2.4 Έλεγχος ανεξαρτησίας χ^2 του Pearson.....	45
4.2.5 Διαστήματα εμπιστοσύνης για τον λόγο σχετικών πιθανοτήτων	47
4.3 Στατιστική ανάλυση για τη σύγκριση ασθενών και μαρτύρων.....	48
4.3.1 Στατιστική ανάλυση μεταξύ ασθενών και μαρτύρων στο σύνολο της βάσης	48
4.3.2 Στατιστική ανάλυση μεταξύ ασθενών και μαρτύρων ανά νοσοκομείο εξέτασης.....	50
4.4 Στατιστική ανάλυση για τους ασθενείς με νεανικής μορφής Παρκινσονισμό...59	
4.4.1 Στατιστική ανάλυση μεταξύ των ασθενών με ηλικία έναρξης της νόσου μικρότερη ή/και ίση των 50 ετών και των ασθενών με ηλικία έναρξης μεγαλύτερη των 50 ετών	59
4.5 Συμπεράσματα	63
ΚΕΦΑΛΑΙΟ 5: ΑΝΑΛΥΣΗ ΚΑΤΑ ΣΥΣΤΑΔΕΣ.....	65
5.1 Εισαγωγή.....	65
5.2 Θεωρητικό υπόβαθρο.....	65

5.2.1 Μέτρα απόστασης και μέτρα ομοιότητας.....	66
5.3 Ιεραρχικές μέθοδοι ομαδοποίησης.....	69
5.3.1 Συσσωρευτικές μέθοδοι.....	69
5.3.2 Διαιρετικές μέθοδοι.....	70
5.3.3 Επιλογή του πλήθους των ομάδων.....	71
5.4 Μη ιεραρχικές μέθοδοι ομαδοποίησης.....	73
5.4.1 Μέθοδος K-Means.....	73
5.4.2 Partitioning around Medoids (PAM).....	73
5.4.3 Fuzzy Analysis Clustering (Fanny).....	74
5.5. Γραφική απεικόνιση και αξιολόγηση των συστάδων.....	74
5.5.1 Διάγραμμα Silhouette.....	74
5.5.2 Όψεις Chernoff.....	75
5.6 Εφαρμογή στα δεδομένα.....	77
5.7 Συμπεράσματα.....	84
ΚΕΦΑΛΑΙΟ 6: ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ.....	85
6.1 Εισαγωγή.....	85
6.2 Θεωρητικό υπόβαθρο.....	85
6.2.1 Εισαγωγή στα Γενικευμένα Γραμμικά Μοντέλα (ΓΓΜ).....	85
6.2.2 Λογιστική Παλινδρόμηση.....	88
6.3 Εφαρμογή στα δεδομένα.....	91
ΚΕΦΑΛΑΙΟ 7: ΣΥΜΠΕΡΑΣΜΑΤΑ.....	99
ΠΑΡΑΡΤΗΜΑ.....	101
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	104

Πανεπιστήμιο Πειραιώς

Κατάλογος Πινάκων

	Σελ.
Πίνακας 3.1: Ποσοστά ανδρών-γυναικών ανά κατάσταση εθελοντή	14
Πίνακας 3.2: Ποσοστά θρησκευμάτων ανά κατάσταση εθελοντή	17
Πίνακας 3.3: Ποσοστά τύπου τόπου διαμονής κατά την παιδική ηλικία και την τώρα	18
Πίνακας 3.4: Ποσοστά τύπου τόπου διαμονής κατά την παιδική ηλικία και την τώρα ανά κατάσταση εθελοντή	19
Πίνακας 3.5: Ποσοστά συγγένειας εξ' αίματος των γονέων ανά κατάσταση εθελοντή	20
Πίνακας 3.6: Ποσοστά οικογενειακού ιστορικού ΝΠ ανά κατάσταση εθελοντή	21
Πίνακας 3.7: Ποσοστά έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα ανά κατάσταση εθελοντή	33
Πίνακας 3.8: Ποσοστά εμφάνισης αρχικού συμπτώματος	37
Πίνακας 3.9: Ποσοστά εμφάνισης συμπτωμάτων	38
Πίνακας 3.10: Ποσοστά εμφάνισης κινητικών επιπλοκών με διακυμάνσεις ή δυσκινησίες	39
Πίνακας 4.1: Σχέση εμφάνισης ΝΠ και παράγοντα	50
Πίνακας 4.2: Σχέση εμφάνισης ΝΠ και παράγοντα ανά νοσοκομείο εξέτασης	52
Πίνακας 4.3: Κανονικότητα μεταβλητών ανά νοσοκομείο εξέτασης	53
Πίνακας 4.4: Ομοσκεδαστικότητα και ισότητα μέσων των κανονικών μεταβλητών	53
Πίνακας 4.5: Ισότητα κατανομών και ομοσκεδαστικότητα για τις μη κανονικές μεταβλητές	54
Πίνακας 4.6: Συντελεστές συσχέτισης των συνεχών μεταβλητών	58
Πίνακας 4.7: Σχέση εμφάνισης νεανικής μορφής παρκινσονισμού και παράγοντα	61
Πίνακας 4.8: Κανονικότητα μεταβλητών ανά ηλικιακή κατηγορία εμφάνισης νόσου του Πάρκινσον	62
Πίνακας 4.9: Ισότητα κατανομών και ομοσκεδαστικότητα για τις μη κανονικές μεταβλητές	62
Πίνακας 5.1: Κωδικοποίηση μεταβλητών που χρησιμοποιήθηκαν στην ομαδοποίηση	79
Πίνακας 6.1: Πίνακας με τις κυριότερες συναρτήσεις σύνδεσης και τις αντίστροφές τους	89
Πίνακας 6.2: Πίνακας τριπλής εισόδου για την ερμηνεία του όρου της αλληλεπίδρασης	97
Πίνακας 6.3: Πίνακας διπλής εισόδου για τα άτομα α)με οικογενειακό ιστορικό ΝΠ, β)χωρίς οικογενειακό ιστορικό ΝΠ	97

Πανεπιστήμιο Πειραιώς

Κατάλογος Διαγραμμάτων

	Σελ.
Διάγραμμα 3.1:	Κυκλικό διάγραμμα για την κατανομή του δείγματος των εθελοντών 14
Διάγραμμα 3.2:	Σχετικές συχνότητες ανδρών-γυναικών ανά κατάσταση εθελοντή 15
Διάγραμμα 3.3:	α)Ιστόγραμμα για την ηλικία του συνόλου του δείγματος, β)Θηκόγραμμα για την ηλικία του συνόλου του δείγματος 15
Διάγραμμα 3.4:	α)Ιστόγραμμα για την ηλικία των ασθενών, β) Θηκόγραμμα για την ηλικία των ασθενών 16
Διάγραμμα 3.5:	Σχετικές συχνότητες θρησκευμάτων ανά κατάσταση εθελοντή 17
Διάγραμμα 3.6:	Κυκλικό διάγραμμα για την κατανομή των τόπων διαμονής α)κατά την παιδική ηλικία και β)τόρα 18
Διάγραμμα 3.7:	Σχετικές συχνότητες τύπων τόπου διαμονής α)κατά την παιδική ηλικία και β) τώρα, ανά κατάσταση εθελοντή 19
Διάγραμμα 3.8:	Κυκλικό διάγραμμα για την κατανομή εμφάνισης οικογενειακού ιστορικού ως προς την νόσο Πάρκινσον 21
Διάγραμμα 3.9:	Σχετικές συχνότητες οικογενειακού ιστορικού ΝΠ ανά κατάσταση εθελοντή 22
Διάγραμμα 3.10:	Κυκλικό διάγραμμα για την κατανομή της χειρίας των εθελοντών 22
Διάγραμμα 3.11:	Σχετικές συχνότητες καπνίσματος κατά τις τελευταίες 30 ημέρες ανά κατάσταση εθελοντή 24
Διάγραμμα 3.12:	Σχετικές συχνότητες καπνίσματος 100 ή περισσότερων τσιγάρων στη ζωή τους, ανά κατάσταση εθελοντή 24
Διάγραμμα 3.13:	α)Ιστόγραμμα του μέσου ημερήσιου αριθμού τσιγάρων που καταναλώθηκαν από τους εθελοντές, β)Θηκόγραμμα του μέσου ημερήσιου αριθμού τσιγάρων που καταναλώθηκαν από τους εθελοντές 25
Διάγραμμα 3.14:	α)Ιστόγραμμα του μέσου ημερήσιου αριθμού τσιγάρων που καταναλώθηκαν από τους ασθενείς, β)Θηκόγραμμα του μέσου ημερήσιου αριθμού τσιγάρων που καταναλώθηκαν από τους ασθενείς 25
Διάγραμμα 3.15:	α)Ιστόγραμμα για τη διάρκεια καπνίσματος για τους εθελοντές, β)Θηκόγραμμα για τη διάρκεια καπνίσματος για τους εθελοντές 26
Διάγραμμα 3.16:	α)Ιστόγραμμα για τη διάρκεια καπνίσματος για τους ασθενείς, β)Θηκόγραμμα για τη διάρκεια καπνίσματος για τους ασθενείς 26
Διάγραμμα 3.17:	α)Ιστόγραμμα για την ηλικία παύσης καπνίσματος για τους εθελοντές, β)Θηκόγραμμα για την ηλικία παύσης 26

	καπνίσματος για τους εθελοντές	
Διάγραμμα 3.18:	α)Ιστόγραμμα για την ηλικία παύσης καπνίσματος για τους ασθενείς, β)Θηκόγραμμα για την ηλικία παύσης καπνίσματος για τους ασθενείς	27
Διάγραμμα 3.19:	α)Ιστόγραμμα για τη διάρκεια παύσης καπνίσματος για τους εθελοντές, β)Θηκόγραμμα για τη διάρκεια παύσης καπνίσματος για τους εθελοντές	27
Διάγραμμα 3.20:	α)Ιστόγραμμα για τη διάρκεια παύσης καπνίσματος για τους ασθενείς, β)Θηκόγραμμα για τη διάρκεια παύσης καπνίσματος για τους ασθενείς	28
Διάγραμμα 3.21:	Κυκλικό διάγραμμα για την κατανομή της κατανάλωσης καφέ σε α)τωρινό και β)παρελθόντα χρόνο	29
Διάγραμμα 3.22:	Σχετικές συχνότητες κατανάλωσης καφέ ανά κατάσταση εθελοντή σε α)τωρινό και β) παρελθόντα χρόνο	29
Διάγραμμα 3.23:	Ραβδόγραμμα για τον μέσο αριθμό φλιτζανιών καφέ που καταναλώνονται α)από τους εθελοντές, β)από τους ασθενείς	30
Διάγραμμα 3.24:	α)Ιστόγραμμα για τη διάρκεια κατανάλωσης καφέ για τους εθελοντές, β)Θηκόγραμμα για τη διάρκεια κατανάλωσης καφέ για τους εθελοντές	30
Διάγραμμα 3.25:	α)Ιστόγραμμα για τη διάρκεια κατανάλωσης καφέ για τους ασθενείς, β) Θηκόγραμμα για τη διάρκεια κατανάλωσης καφέ για τους ασθενείς	31
Διάγραμμα 3.26:	α)Ιστόγραμμα για την ηλικία παύσης κατανάλωσης καφέ για τους εθελοντές, β)Θηκόγραμμα για την ηλικία παύσης κατανάλωσης καφέ για τους εθελοντές	31
Διάγραμμα 3.27:	α)Ιστόγραμμα για την ηλικία παύσης κατανάλωσης καφέ για τους ασθενείς, β) Θηκόγραμμα για την ηλικία παύσης κατανάλωσης καφέ για τους ασθενείς	32
Διάγραμμα 3.28:	Κυκλικό διάγραμμα για την κατανομή έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα των εθελοντών	32
Διάγραμμα 3.29:	Σχετικές συχνότητες έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα ανά κατάσταση εθελοντή	33
Διάγραμμα 3.30:	Κυκλικό διάγραμμα για την κατανομή του τύπου έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα των εθελοντών	34
Διάγραμμα 3.31:	Σχετικές συχνότητες τύπου έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα ανά κατάσταση εθελοντή	34
Διάγραμμα 3.32:	α)Ιστόγραμμα του μέσου χρόνου έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα για το σύνολο των εθελοντών, β)Θηκόγραμμα του μέσου χρόνου έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα για το σύνολο των εθελοντών	35
Διάγραμμα 3.33:	α)Ιστόγραμμα του μέσου χρόνου έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα για τους ασθενείς, β)Θηκόγραμμα του	35

	μέσου χρόνου έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα για τους ασθενείς	
Διάγραμμα 3.34:	α)Ιστόγραμμα για την ηλικία έναρξης της ΝΠ, β) Θηκόγραμμα για την ηλικία έναρξης της ΝΠ	36
Διάγραμμα 3.35:	Σχετικές συχνότητες εμφάνισης εκάστοτε αρχικού συμπτώματος	37
Διάγραμμα 3.36:	Κυκλικό διάγραμμα για την κατανομή της πλευράς έναρξης της ΝΠ	38
Διάγραμμα 3.37:	α)Ιστόγραμμα για την απεικόνιση της αξιολόγησης των ασθενών σύμφωνα με την κλίμακα UPDRS I έως και III, β) Θηκόγραμμα για την απεικόνιση της αξιολόγησης των ασθενών σύμφωνα με την κλίμακα UPDRS I έως και III	39
Διάγραμμα 3.38:	Ραβδόγραμμα για την απεικόνιση της αξιολόγησης των ασθενών υπό την κλίμακα UPDRS α) IVA β) IVB	40
Διάγραμμα 3.39:	Ραβδόγραμμα για την απεικόνιση της αξιολόγησης των ασθενών σύμφωνα με την κλίμακα α) Hoehn & Yahr, β) Schwab & England	41
Διάγραμμα 3.40:	Ραβδόγραμμα για την απεικόνιση της αξιολόγησης των ασθενών σύμφωνα με την κλίμακα Standardized mini-mental Status Examination	41
Διάγραμμα 4.1:	Θηκόγραμμα για την απεικόνιση της αξιολόγησης ασθενών σύμφωνα με την κλίμακα UPDRS I έως και III που εξετάστηκαν α)στο Αττικό νοσοκομείο, β)στο νοσοκομείο Σύρου	55
Διάγραμμα 4.2:	Θηκόγραμμα για την απεικόνιση της ηλικίας έναρξης της νόσου των ασθενών που εξετάστηκαν α)στο Αττικό νοσοκομείο, β)στο νοσοκομείο Σύρου	56
Διάγραμμα 4.3:	Θηκόγραμμα για την απεικόνιση των ετών καπνίσματος των ασθενών που εξετάστηκαν α)στο Αττικό νοσοκομείο, β)στο νοσοκομείο Σύρου	57
Διάγραμμα 4.4:	Θηκόγραμμα για την απεικόνιση της ηλικίας παύσης καπνίσματος των ασθενών που εξετάστηκαν α)στο Αττικό νοσοκομείο, β)στο νοσοκομείο Σύρου	57
Διάγραμμα 4.5:	Διαγράμματα διασποράς μεταξύ των συνεχών μεταβλητών	59
Διάγραμμα 4.6:	Θηκόγραμμα για την απεικόνιση της αξιολόγησης ασθενών σύμφωνα με την κλίμακα UPDRS I έως και III με ηλικία έναρξης α)μέχρι 50 έτη, β)μεγαλύτερη από 50 έτη	63
Διάγραμμα 4.7:	Θηκόγραμμα για την απεικόνιση των ετών κατανάλωσης καφέ των ασθενών με ηλικία έναρξης α)μέχρι 50 έτη, β)μεγαλύτερη από 50 έτη	64
Διάγραμμα 5.1:	Δενδροδιάγραμμα	74
Διάγραμμα 5.2:	Δενδροδιάγραμμα που αφορά την συσταδοποίηση των ασθενών σε 2 ομάδες	80

Διάγραμμα 5.3:	Δενδροδιάγραμμα που αφορά την συσταδοποίηση των ασθενών σε 4 ομάδες, ανεξαρτήτως νοσοκομείου εξέτασης	81
Διάγραμμα 5.4:	Silhouette διάγραμμα που αφορά την συσταδοποίηση των ασθενών σε 2 ομάδες	83
Διάγραμμα 5.5:	Silhouette διάγραμμα που αφορά την συσταδοποίηση των ασθενών σε 4 ομάδες	84
Διάγραμμα 5.6:	Chernoff faces για τα 20 πρώτους ασθενείς της βάσης δεδομένων και ομαδοποίησή τους σύμφωνα με την συσταδοποίηση σε 2 ομάδες	85
Διάγραμμα 5.7:	Chernoff faces για τα 20 πρώτους ασθενείς της βάσης δεδομένων και ομαδοποίησή τους σύμφωνα με την συσταδοποίηση σε 4 ομάδες	85
Διάγραμμα 6.1	Διάγραμμα με τα κατάλοιπα του Pearson	97

Πανεπιστήμιο Πειραιώς

Πανεπιστήμιο Πειραιώς

Πανεπιστήμιο Πειραιώς

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

1.1 Γενικά για την Νόσο Πάρκινσον

Η νόσος Πάρκινσον αποτελεί μια χρόνια και προοδευτική διαταραχή του νευρικού συστήματος η οποία επηρεάζει την κίνηση. Εμφανίζεται κυρίως σε άτομα που έχουν συμπληρώσει το 60^ο έτος ζωής σε συχνότητα περίπου 3%. Πρόκειται για την νόσο των ηλικιωμένων όμως υπάρχουν και περιπτώσεις όπου τα συμπτώματά της εντοπίζονται σε άτομα ηλικίας μικρότερης των 50 ετών.

Τα συμπτώματα την ασθένειας αφορούν κυρίως σε κινητικές δυσκαμψίες των άκρων και του κορμού και προέρχονται από την προοδευτική νέκρωση των ζωτικών νευρικών κυττάρων κυρίως σε μια περιοχή του εγκεφάλου που ονομάζεται μέλαινα ουσία. Ορισμένα κύτταρα από αυτά που νεκρώνονται είναι υπεύθυνα για την παραγωγή της χημικής ουσίας ντοπαμίνης, η οποία στέλνει μηνύματα προς το μέρος του εγκεφάλου που ελέγχει την κίνηση και τον συντονισμό της. Το περισσότερο σύνθηες σύμπτωμα είναι ο τρόμος (τρέμουλο), ενώ κατά την πάροδο της ασθένειας και καθώς η ποσότητα ντοπαμίνης στον εγκέφαλο μειώνεται, ο ασθενής δεν μπορεί να ελέγξει κανονικά την κίνησή του.

Τα αίτια που προκαλούν την νόσο Πάρκινσον είναι μέχρι και σήμερα άγνωστα. Παρόλα αυτά, έχει ανακαλυφθεί πως η έκθεση σε ορισμένους περιβαλλοντικούς παράγοντες μπορεί είτε να αυξήσει είτε να μειώσει τον κίνδυνο της νόσου καθώς και τα συμπτώματά της. Επιπλέον, ερευνητές έχουν προσδιορίσει συγκεκριμένες γενετικές μεταλλάξεις οι οποίες μπορεί να επηρεάζουν την εμφάνιση της ασθένειας.

1.2 Σκοπός της μελέτης

Στόχος της παρούσας εργασίας είναι να μελετήσει διάφορους παράγοντες οι οποίοι πιστεύεται ότι μπορούν να επηρεάσουν την εμφάνιση ή μη της νόσου αλλά και την καθυστέρηση των συμπτωμάτων της. Ακόμα, θα μελετηθεί η συνεισφορά τους ή μη στην εμφάνιση νεανικής μορφής παρκινσονισμού. Σε κάθε περίπτωση γίνεται αναφορά στα στοιχεία αντίστοιχων μελετών ή/και της διεθνούς βιβλιογραφίας ενώ παράλληλα παρατίθεται και η γραφική τους απεικόνιση.

Για τον σκοπό αυτό, θα χρησιμοποιήσουμε μεθόδους περιγραφικής στατιστικής, πίνακες διπλής εισόδου, μεθόδους πολυμεταβλητής ανάλυσης για την συσταδοποίηση των δεδομένων καθώς και λογιστική παλινδρόμηση για τη δημιουργία μοντέλων

πρόβλεψης σχετικών πιθανοτήτων. Οι έλεγχοι που θα διεξάγουμε έχουν πραγματοποιηθεί σε επίπεδο σημαντικότητας 5%, αλλά σε κάθε περίπτωση παρέχεται και η τιμή της στατιστικής συνάρτησης.

Τα δεδομένα μας προέρχονται από την Ελληνική Βιοτράπεζα για την Νόσο του Πάρκινσον και μας παραχωρήθηκαν από τον κ.Βασιλάτη, ερευνητή Γ' στο Ίδρυμα Ιατροβιολογικών Ερευνών Ακαδημίας Αθηνών (ΙΙΒΕΑ).

Η συλλογή των δεδομένων έγινε με την συμπλήρωση ενός ερωτηματολογίου από τους ιατρούς Δρ. Στεφανή και Δρ. Μπόζη, στο Αττικό νοσοκομείο και το νοσοκομείο Σύρου, αντίστοιχα, και αφορούσε δημογραφικά, γενετικά και ιατρικά στοιχεία τόσο από ασθενείς που πάσχουν από την νόσο Πάρκινσον, αλλά και από μάρτυρες και μη πάσχοντες συγγενείς αυτών.

Τέλος, το στατιστικό πακέτο που χρησιμοποιήθηκε κατά την εκπόνηση της διπλωματικής εργασίας είναι το R *version 2.15.2*.

ΚΕΦΑΛΑΙΟ 2: ΓΕΝΙΚΑ ΣΤΟΙΧΕΙΑ ΓΙΑ ΤΗ ΝΟΣΟ ΤΟΥ ΠΑΡΚΙΝΣΟΝ

2.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζονται βασικά στοιχεία για τη νόσο του Πάρκινσον καθώς και στοιχεία για την Βιοτράπεζα Παρκινσονικών Ελλάδας από όπου και λάβαμε τα δεδομένα που θα αναλυθούν στη συνέχεια.

2.2 Βάση Δεδομένων Ελλήνων Παρκινσονικών (Βιοτράπεζα)

Η ερευνητική μελέτη μιας ασθένειας είτε στα πλαίσια «βασικής έρευνας» είτε όταν γίνεται «κλινική ή εφαρμοσμένη έρευνα» προϋποθέτει την ύπαρξη μιας ισχυρής βάσης δεδομένων η οποία να περιλαμβάνει ιατρικά ιστορικά ασθενών και δείγματα βιολογικού υλικού. Ο συνδυασμός της βάσης δεδομένων με τα αντίστοιχα δείγματα βιολογικού υλικού θα αποτελούν μία Βιοτράπεζα. Ο όρος Βιοτράπεζα είναι ένας γενικός όρος για την αποθήκευση βιολογικού υλικού. Οι Βιοτράπεζες μπορούν να ταξινομηθούν σύμφωνα με μια ποικιλία χαρακτηριστικών, συμπεριλαμβανομένου του τύπου ιστού, τον σκοπό, την ιδιοκτησία, την ομάδα ασθενών και το μέγεθος. Τα δείγματα και τα ιατρικά ιστορικά τα οποία συλλέγονται προέρχονται από εθελοντές με ή χωρίς τη νόσο, τα οποία χρησιμοποιούνται για τον εντοπισμό γονιδίων ή βιοδεικτών που συμβάλλουν στην κατανόηση της ασθένειας.

Τα οφέλη από μια Βιοτράπεζα είναι πολλαπλά και μπορούν να είναι πολλαπλασιαστικά. Τέτοια, δηλαδή, ώστε να μπορούν να διευκολύνουν την έρευνα διαφόρων τύπων γενετικής προδιάθεσης που οδηγούν σε πολλές πολυπαραγοντικές ασθένειες. Επιπλέον, ο συνδυασμός με ιατρικά ιστορικά, δημογραφικές πληροφορίες καθώς και πληροφορίες ως προς τον τρόπο ζωής ασθενών ή μαρτύρων είναι δυνατόν να μας υποδείξουν σχέσεις περιβαλλοντικών παραγόντων με είτε γενετικούς, είτε βιοχημικούς δείκτες. Με αυτόν τον τρόπο, μπορούν να είναι οι βάσεις μέσω των οποίων θα μπορούν να απαντηθούν ερωτήματα τα οποία θα ήταν αδύνατο ή σχεδόν αδύνατο να απαντηθούν αν δεν υπήρχαν.

Για τους ανωτέρω λόγους, πρωταρχικής σημασίας ήταν η δημιουργία μιας Πανελληνίας Βιοτράπεζας για τη Νόσο Πάρκινσον. Η δημιουργία της Βάσης Δεδομένων Ελλήνων Παρκινσονικών από το Ίδρυμα Ιατροβιολογικών Ερευνών Ακαδημίας Αθηνών (ΙΙΒΕΑΑ) μας παρέχει τη δυνατότητα ανάλυσης μερικών χαρακτηριστικών των ασθενών αυτών στη χώρα μας.

Εκμεταλλευτήκαμε το γεγονός ότι υπήρχαν δύο κυρίως κέντρα συλλογής δειγμάτων αίματος και ιατρικών ιστορικών, πρόσβαση στα οποία είχαν ασθενείς με διαφορετικά χαρακτηριστικά. Πρώτο είναι το Αττικό Νοσοκομείο στην Αθήνα, ένα μητροπολιτικό κέντρο, στο οποίο έχουν πρόσβαση ασθενείς με ποικίλα χαρακτηριστικά λόγω της πολυμορφίας του πληθυσμού που διαμένει σε αυτήν. Δεύτερο είναι το Νοσοκομείο Σύρου στην Ερμούπολη, το οποίο αποτελεί το κύριο νευρολογικό κέντρο των Κυκλάδων και στο οποίο έχουν πρόσβαση οι σχετικά απομονωμένοι πληθυσμοί των νησιών, οι οποίοι είναι πιθανό να έχουν ιδιαίτερα χαρακτηριστικά τόσο λόγω του διαφορετικού τρόπου ζωής όσο και λόγω του πιθανώς πιο περιορισμένου γενετικού υπόβαθρου. Στην εργασία αυτή, μελετήθηκαν μοντέλα ενημερωμένης συγκατάθεσης εθελοντών και ερωτηματολόγια συμπληρωμένα από τους ιατρούς Δρ. Στεφανή και Δρ. Μπόζη, αντίστοιχα, κατά την πρώτη τους επίσκεψη στα ανωτέρω νοσοκομεία.

Σε αυτό το σημείο, πρέπει να σημειωθεί ότι στη Βάση Δεδομένων Ελλήνων Παρκινσονικών περιλαμβάνονται μέχρι στιγμής τα ιστορικά εθελοντών μόνο από το Αττικό Νοσοκομείο και το Νοσοκομείο Σύρου καθώς τα δείγματα από το Πανεπιστήμιο Πατρών και το Πανεπιστήμιο Θεσσαλίας είναι παλαιότερα και είχαν συλλεχθεί πριν οριστικοποιηθεί το ερωτηματολόγιο το οποίο συμπληρώνεται από τους ιατρούς που συμμετέχουν στην προσπάθεια αυτή.

2.3 Η Νόσος Πάρκινσον

Η νόσος του Πάρκινσον (γνωστή και ως ιδιοπαθής Παρκινσονισμός, ή σύνδρομο υποκινητικής ακαμψίας, ή τρομώδης παράλυση) είναι μια εκφυλιστική διαταραχή του κεντρικού νευρικού συστήματος. Τα κινητικά συμπτώματα τα οποία προέρχονται από αυτήν εμφανίζονται σαν αποτέλεσμα της προοδευτικής νέκρωσης των κυττάρων που παράγουν το νευροδιαβιβαστή ντοπαμίνη. Η αιτία αυτής της νέκρωσης είναι ακόμα άγνωστη (www.wikipedia.org).

Τα πιο εμφανή συμπτώματα της νόσου σχετίζονται με την κίνηση και περιλαμβάνουν τρόμο (τρέμουλο) που εντοπίζεται στα χέρια και στα πόδια, αλλά και στο σαγόι, δυσκαμψία των άκρων και του κορμού, βραδυκινησία που σημαίνει ότι ελαττώνεται η ταχύτητα των κινήσεων και χάνονται οι αυθόρμητες και αυτόματες κινήσεις και, τέλος, δυσκολία στο βάδισμα. Αργότερα, μπορεί να προκύψουν προβλήματα σκέψης και συμπεριφοράς, με την άνοια συνήθως να εμφανίζεται στα

πιο προχωρημένα στάδιά της, ενώ η κατάθλιψη είναι το πιο κοινό ψυχιατρικό σύμπτωμα. Άλλα συμπτώματα περιλαμβάνουν διαταραχές αίσθησης, ύπνου και συναισθημάτων (www.wikipedia.org).

Ο μέσος όρος της ηλικίας εμφάνισης είναι τα 60 έτη. Σύμφωνα με στοιχεία της Ευρωπαϊκής Εταιρείας για τη Ν. Πάρκινσον (E.P.D.A.,2014), περίπου το 3% των ανθρώπων σε ηλικία άνω των 60 ετών παρουσιάζει τη νόσο. Όμως, στο 10% περίπου των ασθενών η διάγνωση γίνεται πριν τα 50. Υπάρχει και η νεανική μορφή της νόσου που εμφανίζεται σε άτομα κάτω των 40 ετών και αφορά ποσοστό 5-10% των συνολικών ασθενών με νόσο Πάρκινσον.

Η επίπτωση της νόσου στον γενικό πληθυσμό αντικατοπτρίζεται σε διάφορες μελέτες και στην σημερινή εποχή αυτό μεταφράζεται σε οικονομοτεχνικά κριτήρια θεραπευτικής προσπέλασης. Η νόσος του Πάρκινσον, αποτελεί κυρίως νόσο των ηλικιωμένων και έχει συχνότητα 1% στον γενικό πληθυσμό. Η νοσηρότητα στον πληθυσμό των ηλικιωμένων είναι 3 φορές μεγαλύτερη απ' ό,τι στον φυσιολογικό πληθυσμό και επιδημιολογικές μελέτες αναφέρουν ότι περί το έτος 2040 η νόσος του Πάρκινσον θα αποτελεί την 2η συχνότερη αιτία θανάτου μετά τον καρκίνο (www.parkinson.gr).

Η αύξηση του προσδόκιμου επιβίωσης (στην Ελλάδα είναι κατά μέσο όρο 72 έτη για τους άνδρες και 76 έτη για τις γυναίκες) και η πολύ πρόσφατη επιστημονική ανακοίνωση της αποκρυπτογράφησης του ανθρωπίνου γονιδιώματος (DNA), μολονότι είναι γεγονότα που επιφέρουν επιστημονική προσδοκία, στην νόσο του Parkinson λόγω του γεγονότος της εκφύλισης των βασικών γαγγλίων με την ηλικία, επιφέρουν ακόμη μεγαλύτερη αγωνία για νέες θεραπευτικές προσπελάσεις, διότι αναμένεται αύξηση του αριθμού των πασχόντων. Η μέχρι σήμερα ευρέως ακολουθούμενη φαρμακευτική αγωγή με συνδυασμό φαρμάκων, συνεπάγεται μελλοντικά μεγαλύτερο κοινωνικό και επακόλουθο οικονομικό κόστος φαρμακευτικής αγωγής (www.parkinson.gr).

Ο Παρκινσονισμός θεωρείται ως ένα σύνδρομο ιδιοπαθές (δεν έχει καμία γνωστή αιτία), αν και σε ορισμένες περιπτώσεις έχει γενετική προέλευση. Πολλοί παράγοντες κινδύνου και προστατευτικοί παράγοντες έχουν διερευνηθεί με σαφή ένδειξη για αυξημένο κίνδυνο εμφάνισης της νόσου σε ανθρώπους που εκτίθενται σε ορισμένα φυτοφάρμακα και μειωμένο κίνδυνο σε καπνιστές.

Η παθολογία της νόσου χαρακτηρίζεται από την συσσώρευση μιας πρωτεΐνης που ονομάζεται α-συνουκλεΐνη στα σωματίδια Lewy των νευρώνων καθώς και από τον

ανεπαρκή σχηματισμό και δράση ντοπαμίνης εντός των νευρώνων του εγκεφάλου. Τα σωματίδια Lewy είναι το σήμα κατατεθέν της ιδιοπαθούς διαταραχής και η κατανομή τους διαφέρει από το ένα άτομο στο άλλο. Η κατανομή αυτή συχνά σχετίζεται άμεσα με την έκφραση και το βαθμό των κλινικών συμπτωμάτων του ασθενούς, η διάγνωση των οποίων επιτυγχάνεται με τη μέθοδο της νευροαπεικόνισης (www.parkinsonportal.gr).

Σύγχρονες θεραπείες είναι αποτελεσματικές στη διαχείριση της νόσου κατά τα πρώιμα στάδια, κυρίως μέσω της δραστικής ουσίας λεβοντόπα και των αγωνιστών ντοπαμίνης. Καθώς η νόσος εξελίσσεται, τα φάρμακα γίνονται αναποτελεσματικά και ταυτόχρονα προκαλούν μια επιπλοκή που ονομάζεται δυσκινησία και η οποία εκφράζεται με ακούσιες κινήσεις σφασμασμού. Επιπλέον, η διατροφή και κάποιες μορφές αποκατάστασης φαίνεται να είναι αποτελεσματικές στην ανακούφιση των συμπτωμάτων, ενώ σε πολύ σοβαρές περιπτώσεις γίνονται χειρουργικές επεμβάσεις και εν τω βάθει διέγερση του εγκεφάλου. Στις μέρες μας, ερευνητικές κατευθύνσεις περιλαμβάνουν έρευνες για την πιθανή χρησιμότητα γονιδιακής θεραπείας, μεταμόσχευσης βλαστικών κυττάρων και νευροπροστατευτικών παραγόντων (www.parkinsonportal.gr).

Η ασθένεια πήρε το όνομά της από τον Άγγλο ιατρό James Parkinson, ο οποίος δημοσίευσε την πρώτη λεπτομερή περιγραφή σε ένα δοκίμιο για την «Τρομώδη Παράλυση» το 1817. Ο γάλλος νευρολόγος Jean-Martin Charcot περιέγραψε το 1876 έναν ασθενή που δεν είχε καθόλου τρόμο ενώ είχε δυσκαμψία και πρότεινε το σύνδρομο να ονομαστεί «Νόσος του Πάρκινσον» προς τιμήν του πρώτου, αντί για «Τρομώδης Παράλυση» που ίσχυε μέχρι τότε (www.parkinsonportal.gr).

Εκστρατείες ευαισθητοποίησης του κοινού περιλαμβάνουν την Ημέρα για τη νόσο του Πάρκινσον (την ημέρα γενεθλίων του James Parkinson, 11 Απριλίου) ενώ ως σύμβολο της νόσου χρησιμοποιείται η κόκκινη τουλίπα (www.parkinsonportal.gr).

2.3.1 Νεανικής μορφής παρκινσονισμός

Ο νεανικής μορφής παρκινσονισμός (Young adult onset parkinsonism disease (YOPD)) αφορά σε όλες τις μορφές της νόσου Πάρκινσον που εκδηλώνονται πριν την ηλικία των 50 ετών. Σύμφωνα με τον ιστότοπο orpha.net, τα αίτια, η κλινική παρουσίαση, η κλινική πορεία, καθώς και η διαχείριση των πρώιμης έναρξης μορφών διαφέρουν από εκείνες της κλασικής μορφής της νόσου, η οποία γενικά γίνεται

εμφανής μετά την ηλικία των 60 ετών. Η επίπτωση της YOPD έχει εκτιμηθεί σε 1.5/100000 άτομα μεταξύ 30 και 50 ετών. Εκτός από την δυσκαμψία, τον τρόπο ανάπαυσης, την ακινησία και την ασυμμετρία, η YOPD χαρακτηρίζεται από δυστονική θέση (το οποίο συχνά είναι το πρώτο σημάδι της νόσου, εμφανίζεται κυρίως στα κάτω άκρα), από έντονες αντιδράσεις, υψηλή συχνότητα βραδυκινησίας, τρέμουλο και ακαμψία.

Διάφορες μορφές YOPD έχουν ταυτοποιηθεί. Οι αυτοσωματικές υπολειπόμενες μορφές με γρηγορότερη έναρξη προκαλούνται από μεταλλάξεις στα γονίδια parkin (PARK2), DJ-1 (PARK6) και Pink1 (PARK7). Στον ευρωπαϊκό πληθυσμό, οι μεταλλάξεις στο γονίδιο parkin είναι υπεύθυνες για περισσότερο από το 50% των οικογενειακών υποθέσεων YOPD, ενώ η συχνότητα των μεταλλάξεων των PINK1 και DJ-1 είναι πολύ χαμηλότερο. Μεταλλάξεις στο γονίδιο LRRK2, μπορεί να είναι υπεύθυνες για την σποραδικές ή επικρατούσες κληρονομικές οικογενείς μορφές της νόσου.

Η διάγνωση γίνεται με ανάλυση των κλινικών συμπτωμάτων της νόσου και επιβεβαιώνεται με γενετική ανάλυση και εγκεφαλική σπινθηρογραφία των μεταφορέων ντοπαμίνης. Η διαφορική διάγνωση θα πρέπει να εξετάσει ένα ευρύ φάσμα ασθενειών, συμπεριλαμβανομένης της ευαισθησίας της L – Dopa δυστονίας, την νόσο του Wilson, το είδος της νόσου Gaucher 3, νεανικής μορφής νόσο Huntington ή Dento - Rubro - pallido - Luysian ατροφία , καθώς και κεντρικές αλλοιώσεις των βασικών γαγγλίων. Οι ασθένειες αυτές μπορεί να εκδηλωθούν ως μια μορφή Πάρκινσον διότι συνήθως συνδέονται με νευρολογικά συμπτώματα ή μορφολογικές ανωμαλίες εγκεφαλικών κυττάρων.

Γενετική συμβουλευτική συνιστάται για οικογενείς μορφές ή/και εάν υποδεικνύεται η διαλογή για συγκεκριμένες μεταλλάξεις. Ωστόσο, η παροχή συμβουλών εξακολουθεί να είναι αμφιλεγόμενη καθώς η νόσος είναι εξουθενωτική, αλλά μη θανατηφόρα, ενώ είναι διαθέσιμες αποτελεσματικές θεραπείες. Επιπροσθέτως, η διεισδυτικότητα των διαφόρων αιτιολογικών μεταλλάξεων δεν έχει ακόμη τεκμηριωθεί.

Η θεραπεία στοχεύει στην καταπολέμηση των συμπτωμάτων που συνδέονται με dopaminergic απονεύρωση, καθυστερώντας την εμφάνιση και την αντιμετώπιση των μεταγενέστερων επιπλοκών που σχετίζονται με τη θεραπεία (κινητικές διακυμάνσεις και δυσκινησίες) καθώς και στη διαχείριση των συμπτωμάτων που δεν ανταποκρίνονται στη θεραπεία dopaminergic. Η θεραπεία θα πρέπει να

συνοδεύεται από την επαγγελματική, προσωπική και οικογενειακή διαχείριση των ασθενών. Όταν η φαρμακολογική προσέγγιση δεν είναι πλέον επαρκής για να επιτρέψει μια καλή ποιότητα ζωής, μπορεί να συζητηθεί μια χειρουργική προσέγγιση που περιλαμβάνει διμερή διέγερση των υποθαλαμικών γαγγλιών.

2.4 Ενοποιημένη Κλίμακα Βαθμολόγησης της Νόσου Πάρκινσον (Unified Parkinson's Disease Rating Scale –UPDRS-)

Σύμφωνα με την Ευρωπαϊκή Εταιρεία για τη Ν. Πάρκινσον (E.P.D.A.,2014), η Ενοποιημένη Κλίμακα Βαθμολόγησης της Νόσου Πάρκινσον (UPDRS) αναπτύχθηκε σε μια προσπάθεια να συνδυάσει στοιχεία από μια σειρά από κλίμακες και να αποτελέσει ένα ολοκληρωμένο, αποτελεσματικό και ευέλικτο εργαλείο για την παρακολούθηση των επιπτώσεων της νόσου Πάρκινσον και το βαθμό της αναπηρίας που προκαλείται. Μετά από αρκετές δοκιμαστικές εκδόσεις, η κλίμακα θεσπίστηκε επίσημα το 1987 και έκτοτε έχει ενημερωθεί από ειδικούς από την Ένωση Κινητικών Διαταραχών (Movement Disorder Society) προκειμένου να συμπεριλάβει νέες εκτιμήσεις των μη κινητικών συμπτωμάτων (Fahn, Elton, 1987).

Η κλίμακα έχει τρία τμήματα τα οποία αξιολογούν τους βασικούς τομείς της αναπηρίας, μαζί με μια τέταρτη ενότητα η οποία αξιολογεί τυχόν επιπλοκές της θεραπείας. Επιπλέον, συχνά συνοδεύεται από δύο ακόμα κλίμακες. Την Hoehn and Yahr για την αξιολόγηση της έκθεσης σχετικά με τη νόσο του Πάρκινσον, και την Schwab and England για την αξιολόγηση των δραστηριοτήτων της καθημερινής ζωής. Αν και συχνά χρησιμοποιούνται σε συνδυασμό με την UPDRS, αυτές οι κλίμακες δεν αποτελούν μέρος αυτής καθ' αυτής της UPDRS.

Ακόμα, χρησιμοποιείται και η τυποποιημένη εξέταση της μινι-ψυχικής κατάστασης (Standardized mini-mental Status Examination) του ασθενούς για τη διάγνωση της γνωστικής του δυσλειτουργίας.

2.4.1 Κλίμακα UPDRS

Τα τέσσερα τμήματα καθώς και οι τομείς που αξιολογούν είναι τα εξής:

Μέρος I: Νοητική και Γνωστική Δραστηριότητα, Συμπεριφορά, Διάθεση

- Διανοητική αναπηρία
- Διαταραχή της σκέψης
- Κατάθλιψη

- Κίνητρα / Προτοβουλίες

Μέρος II: Δραστηριότητες Καθημερινής Ζωής

- Ομιλία
- Σάλιωμα
- Κατάποση
- Γραφικός χαρακτήρας
- Κοπή τροφίμων και χειρισμός σκευών
- Ντύσιμο
- Υγιεινή
- Στροφή στο κρεβάτι
- Πτώση
- Πάγωμα κατά το περπάτημα
- Περπάτημα
- Τρόμος
- Αισθητηριακές δυσκολίες

Μέρος III: Κινησιακές Λειτουργίες

- Ομιλία
- Έκφραση του προσώπου
- Τρόμος κατά την ηρεμία
- Τρόμος χεριών κατά τη δράση
- Ακαμψία
- Χτύπημα δακτύλων
- Κινήσεις των χεριών
- Γρήγορη περιστροφή των χεριών
- Ευκινησία στα πόδια
- Σήκωμα από καρέκλα
- Στάση
- Βάδισμα
- Ορθοστατική σταθερότητα
- Βραδυκινησία

Μέρος IV: Επιπλοκές Θεραπείας

- A. Δυσκινησίες
- B. Κλινικές Διακυμάνσεις
- C. Άλλες Διαταραχές

Η UPDRS εφαρμόζεται από επαγγελματίες υγείας και η αξιολόγηση αποδίδεται στις απαντήσεις κάθε ατόμου, καθώς και στην παρατήρηση και τη φυσική τους εξέταση. Τα μέρη I, II και III αποτελούνται από 44 ερωτήσεις και κάθε στοιχείο αποτιμάται με μια πενταβάθμια κλίμακα (από το 0 έως το 4), ενώ το μέρος IV περιλαμβάνει 11 ερωτήσεις με κλίμακα που κυμαίνεται μεταξύ 0 και 23. Η τελική αθροιστική αξιολόγηση κυμαίνεται από 0 έως 199 «βαθμούς», με το 0 να σημαίνει «καμία αναπηρία» και το 199 να εκφράζει «ολική αναπηρία».

2.4.2 Κλίμακα Hoehn and Yahr

Η Hoehn και Yahr κλίμακα είναι ένα σύστημα που χρησιμοποιείται, σε γενικές γραμμές, για την περιγραφή της προόδου των συμπτωμάτων της νόσου Πάρκινσον, καθώς και το σχετικό επίπεδο αναπηρίας. Αρχικά δημοσιεύθηκε το 1967 στο περιοδικό Neurology από τους Melvin Yahr και Margaret Hoehn και περιείχε τα στάδια 1 έως 5. Από τότε, έχει προστεθεί επιπλέον το στάδιο 0, ενώ έχουν προταθεί τα στάδια 1.5 και 2.5 και χρησιμοποιούνται ευρέως.

- Στάδιο 0 : Δεν υπάρχουν ενδείξεις της νόσου.
- Στάδιο 1 : Συμπτώματα σε μία μόνο πλευρά (μονομερής).
- Στάδιο 1.5 : Συμπτώματα μονομερή αλλά και με τη συμμετοχή του αυχένα και της σπονδυλικής στήλης.
- Στάδιο 2 : Συμπτώματα και στις δύο πλευρές (διμερής) αλλά χωρίς βλάβη της ισορροπίας.
- Στάδιο 2.5 : Ήπια διμερή συμπτώματα με ανάκτηση όταν η εκτελείται η εξέταση "pull" (ο γιατρός στέκεται πίσω από τον ασθενή και του ζητά να κρατήσει την ισορροπία του όταν τον τραβάει -pull- προς τα πίσω).
- Στάδιο 3 : Βλάβη βάδισης. Ήπια έως μέτρια νόσηση, σωματικά ανεξάρτητη.
- Στάδιο 4 : Σοβαρή αναπηρία, αλλά είναι σε θέση να περπατήσει ή να σταθεί χωρίς υποστήριξη.
- Στάδιο 5 : Χρειάζεται αναπηρική καρέκλα ή κρεβάτι εάν δεν υποστηρίζεται.

2.4.3 Κλίμακα Schwab and England των καθημερινών δραστηριοτήτων (Activities of Daily Living –ADL-)

Η Schwab και England κλίμακα είναι ένα μέσο αξιολόγησης της ικανότητας ενός ατόμου να εκτελεί καθημερινές δραστηριότητες, όσων αφορά την ταχύτητα και την ανεξαρτησία, εκφρασμένη ως ποσοστό, με το 100% να υποδεικνύει πλήρη ανεξαρτησία και το 0% πλήρη εξάρτηση. Η αξιολόγηση μπορεί να καθορίζεται από τον επαγγελματία ή από το άτομο το οποίο εξετάζεται σύμφωνα με τα ακόλουθα κριτήρια:

- 100% : Πλήρως ανεξάρτητος. Ικανός να κάνει όλες τις δραστηριότητες χωρίς βραδύτητα, δυσκολία ή βλάβη.
- 90% : Πλήρως ανεξάρτητος. Ικανός να κάνει όλες τις δραστηριότητες με κάποια βραδύτητα, δυσκολία ή βλάβη. Μπορεί να χρειαστεί διπλάσιο χρόνο για να τις ολοκληρώσει.
- 80% : Ανεξάρτητος στις περισσότερες δραστηριότητες. Χρειάζεται τον διπλάσιο χρόνο. Συνειδητή δυσκολία και επιβράδυνση.
- 70% : Όχι εντελώς ανεξάρτητος. Περισσότερη δυσκολία με τις δραστηριότητες. Τρεις με τέσσερις φορές περισσότερο χρόνο για ορισμένους. Μπορεί να πάρει μεγάλο μέρος της ημέρας για τις καθημερινές δραστηριότητες.
- 60% : Μερικώς εξαρτώμενος. Μπορεί να κάνει τις περισσότερες δραστηριότητες, αλλά πολύ αργά και με πολύ προσπάθεια. Μερικές καθημερινές δραστηριότητες είναι αδύνατες.
- 50% : Περισσότερο εξαρτώμενος. Απαιτείται βοήθεια για τις μισές καθημερινές δραστηριότητες. Δυσκολία με τα πάντα.
- 40% : Εξαρτώμενος σε μεγάλο βαθμό. Μπορεί να βοηθάει σε όλες τις καθημερινές δραστηριότητες αλλά λίγες μπορεί να τις εκτελέσει μόνος.
- 30% : Προσπαθώντας κάνει πότε πότε λίγες καθημερινές δραστηριότητες ή τις ξεκινάει μόνος. Χρειάζεται πολύ περισσότερη βοήθεια.
- 20% : Δεν μπορεί να κάνει τίποτα μόνος. Μπορεί να βοηθήσει λίγο σε κάποιες καθημερινές δραστηριότητες. Σοβαρά ανάπηρος.
- 10% : Απόλυτα εξαρτώμενος, ανήμπορος.
- 0% : Φυσικές λειτουργίες όπως η κατάποση, η λειτουργία της ουροδόχου κύστεως και του εντέρου υπολειτουργούν. Κλινικής.

2.4.4 Τυποποιημένη εξέταση της μινι-ψυχικής κατάστασης (Standardized mini-mental Status Examination)

Η τυποποιημένη εξέταση της μινι-ψυχικής κατάστασης (Standardized mini-mental Status Examination) αφορά ένα σύντομο ερωτηματολόγιο 30 ερωτήσεων, το οποίο μέσα σε 10 λεπτά εξετάζει λειτουργίες όπως αριθμητική, μνήμη και προσανατολισμό, για τη διάγνωση της γνωστικής δυσλειτουργίας του ασθενούς. Χρησιμοποιείται συνήθως στην ιατρική για τη διάγνωση της άνοιας, καθώς επίσης και για να εκτιμηθεί η σοβαρότητα της γνωστικής εξασθένησης και για να μπορεί να παρακολουθηθεί η πορεία των γνωστικών αλλαγών σε ένα άτομο με την πάροδο του χρόνου. Καθίσταται, έτσι, ως ένας αποτελεσματικός τρόπος τεκμηρίωσης της απόκρισης ενός ασθενούς σε μια θεραπεία.

Πανεπιστήμιο Πειραιώς

ΚΕΦΑΛΑΙΟ 3: ΠΕΡΙΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ

3.1 Εισαγωγή

Στο 3^ο κεφάλαιο της διπλωματικής εργασίας θα ασχοληθούμε με τα περιγραφικά στατιστικά στοιχεία των δεδομένων μας. Θα αναφέρουμε, αρχικά, πληροφορίες ως προς τη φύση των δεδομένων και στη συνέχεια θα προχωρήσουμε στην περιγραφική τους ανάλυση. Η ανάλυση αυτή θα αποτελείται από δύο μέρη. Το πρώτο θα αφορά κλινικά και δημογραφικά στοιχεία των εθελοντών που απαρτίζουν τη Βιοτράπεζα στο σύνολό της, ενώ το δεύτερο θα αφορά κλινικά στοιχεία και συμπτώματα που σχετίζονται με τη νόσο Πάρκινσον μόνο στο δείγμα των ασθενών.

3.1.1 Φύση των δεδομένων

Όπως έχουμε αναφέρει και στην εισαγωγή της έρευνάς μας, τα δεδομένα που είχαμε στη διάθεσή μας προέρχονται από τη βάση δεδομένων Ελλήνων Παρκινσονικών η οποία έχει δημιουργηθεί στο Ίδρυμα Ιατροβιολογικών Ερευνών Ακαδημίας Αθηνών (ΙΒΕΑΑ) υπό την επίβλεψη του ερευνητή Γ' κ.Βασιλάτη. Η συλλογή των δεδομένων έγινε με τη συμπλήρωση κατάλληλου ερωτηματολογίου από τους ιατρούς Δρ. Στεφανή και Δρ. Μπόζη, στο Αττικό Νοσοκομείο Αθηνών και στο Νοσοκομείο Σύρου, αντίστοιχα, κατά την πρώτη επίσκεψη των ασθενών σε αυτά. Ακόμα, παρόμοια ερωτηματολόγια έχουν συμπληρωθεί για «μάρτυρες» από το κοντινό περιβάλλον των ασθενών ώστε να επιτευχθεί ατομική εξομοίωση των χαρακτηριστικών των ασθενών, καθώς επίσης και για μη πάσχοντες συγγενείς αυτών ώστε να μελετηθεί και η πιθανότητα γενετικής προέλευσης της νόσου.

Το συνολικό μας δείγμα αποτελείται από 568 εθελοντές, από τους οποίους, οι 343 (ποσοστό 60.4%) είναι ασθενείς, οι 211 (ποσοστό 37.1%) είναι μάρτυρες και οι 14 (ποσοστό 2.5%) είναι μη πάσχοντες συγγενείς.

Και διαγραμματικά:



Διάγραμμα 3.1: Κυκλικό διάγραμμα για την κατανομή του δείγματος των εθελοντών

Στη συνέχεια ακολουθεί περιγραφική ανάλυση των στοιχείων που συλλέχθηκαν διαμέσου των ερωτηματολογίων, τα οποία περιέχουν δημογραφικές πληροφορίες, ιατρικό ιστορικό καθώς και πληροφορίες ως προς τον τρόπο ζωής των εθελοντών.

3.2 Περιγραφική Ανάλυση για το σύνολο των εθελοντών

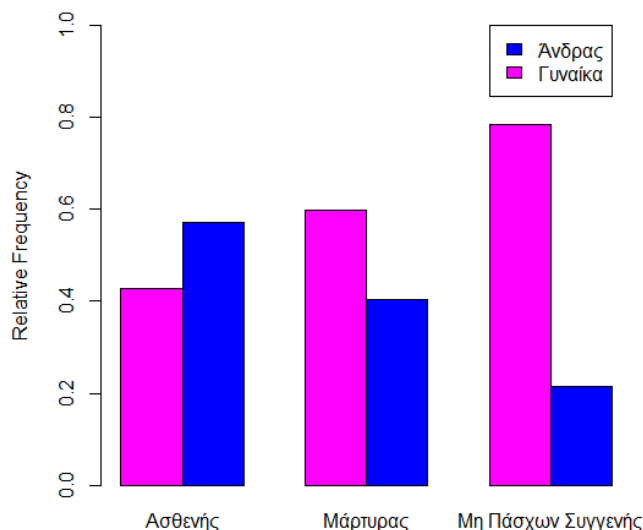
- **Φύλο**

Περισσότεροι άνδρες παρά γυναίκες διαγιγνώσκονται με ΝΠ, με αναλογία περίπου 2:1 (Miller & Cronin-Golomb, 2010), και ενώ η αναλογία αυτή υποδηλώνει την παρουσία διαφοράς μεταξύ των δύο φύλων στην νόσο, οι διαφορές μεταξύ ανδρών και γυναικών σε συμπτώματα, την πορεία και τις γνωστικές επιπτώσεις δεν έχουν μελετηθεί εκτενώς. Ο σχετικός κίνδυνος είναι 1.5 φορές μεγαλύτερος στους άνδρες σε σχέση με τις γυναίκες. Το σύνολο του δείγματός μας αποτελείται, τυχαία, ακριβώς κατά το ήμισυ από άνδρες και γυναίκες, ενώ ανά κατάσταση εθελοντή η κατανομή έχει ως εξής:

Πίνακας 3.1: Ποσοστά ανδρών-γυναικών ανά κατάσταση εθελοντή

Φύλο	Κατάσταση εθελοντή		
	Ασθενής	Μάρτυρας	Μη Πάσχων Συγγενής
Γυναίκα	42.9%	59.7%	78.6%
Άνδρας	57.1%	40.3%	21.4%

Και διαγραμματικά:

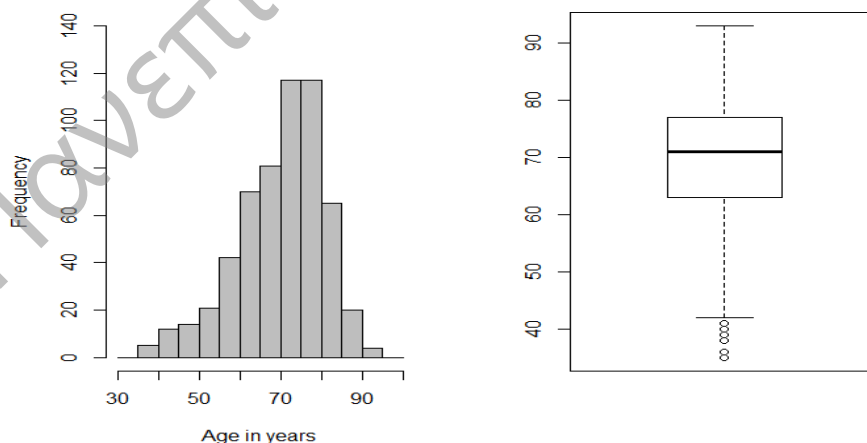


Διάγραμμα 3.2: Σχετικές συχνότητες ανδρών-γυναικών ανά κατάσταση εθελοντή

Όπως και στη διεθνή βιβλιογραφία, έτσι και στο δείγμα που αφορά άτομα που διαμένουν στην Ελλάδα, οι άνδρες τείνουν να εμφανίζουν περισσότερο συχνά τη νόσο σε σχέση με τις γυναίκες.

- **Ηλικία**

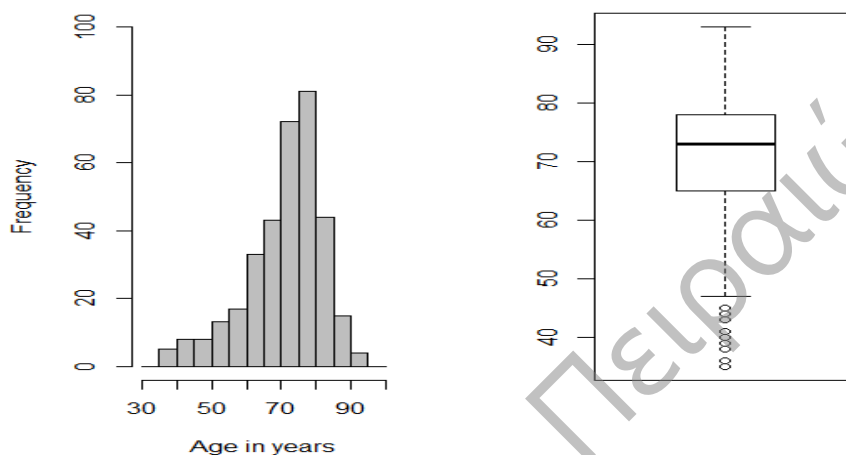
Ακόμα ένα ενδιαφέρον δημογραφικό χαρακτηριστικό για την απεικόνιση του δείγματός μας είναι η ηλικία των εθελοντών. Η κατανομή της ηλικίας των εθελοντών όπως διαμορφώθηκε στη βάση δεδομένων είναι η εξής:



Διάγραμμα 3.3: α) Ιστόγραμμα για την ηλικία του συνόλου του δείγματος, β) Θηκόγραμμα για την ηλικία του συνόλου του δείγματος

Η ηλικία των εθελοντών είναι κατά μέσο όρο τα 69.5 έτη ενώ η διάμεσος τα 71 έτη. Από το ιστόγραμμα συχνοτήτων παρατηρούμε μια αριστερή ασυμμετρία της

κατανομής των ηλικιών. Επιπλέον, το θηκόγραμμα μας ενημερώνει για κάποιες ακραίες τιμές στο δείγμα μας, οι οποίες είτε οφείλονται σε περιπτώσεις νεανικής μορφής παρκινσονισμού, είτε σε πιθανώς λανθασμένη επιλογή μαρτύρων. Προκειμένου να διαλευκάνουμε τι ακριβώς απεικονίζεται, σχεδιάζουμε τα αντίστοιχα διαγράμματα για το δείγμα των ασθενών μόνο. Τα διαγράμματα που προέκυψαν είναι τα εξής:



Διάγραμμα 3.4: α) Ιστόγραμμα για την ηλικία των ασθενών, β) Θηκόγραμμα για την ηλικία των ασθενών

Σε αυτήν την περίπτωση, η μέση ηλικία των ασθενών είναι τα 70.4 έτη ενώ η διάμεσος τα 73 έτη. Παρατηρούμε εκ νέου μια αριστερή ασυμμετρία της κατανομής των ηλικιών των ασθενών, καθώς επίσης πως υπάρχουν ακραίες τιμές στο θηκόγραμμα οι οποίες μας δίνουν ενδείξεις πως υπάρχουν περιπτώσεις νεανικής μορφής παρκινσονισμού στο μελετώμενο δείγμα.

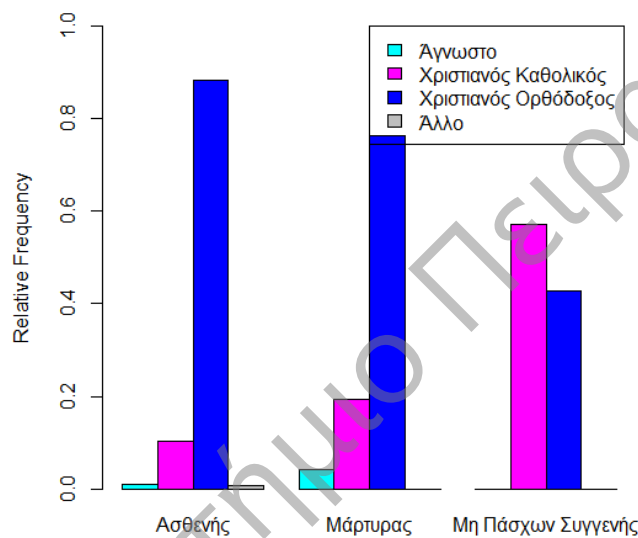
▪ Θρήσκευμα

Ένας ακόμα σημαντικός παράγοντας πιθανής εμφάνισης της ασθένειας θεωρείται το θρήσκευμα, αρχικά διότι πολλές φορές συνδέεται με ορισμένο τρόπο ζωής και επιπροσθέτως διότι πολλές φορές δεν επιτρέπεται ο γάμος μεταξύ αλλόθρησκων και συνεπώς αποτρέπεται η ανάμιξη ποικιλόμορφων γενετικών υλικών. Οι εθελοντές ρωτήθηκαν για αυτό και τα αποτελέσματα που λάβαμε ήταν τα παρακάτω:

Πίνακας 3.2: Ποσοστά θρησκευμάτων ανά κατάσταση εθελοντή

Θρήσκευμα	Κατάσταση εθελοντή		
	Ασθενής	Μάρτυρας	Μη Πάσχων Συγγενής
Χριστιανός Καθολικός	10.2%	19.4%	57.1%
Χριστιανός Ορθόδοξος	88.3%	76.3%	42.9%
Άλλο	0.6%	0.0%	0.0%
Άγνωστο	0.9%	4.3%	0.0%

Και διαγραμματικά:



Διάγραμμα 3.5: Σχετικές συχνότητες θρησκευμάτων ανά κατάσταση εθελοντή

Παρατηρούμε πως η πλειονότητα των παρατηρήσεων είναι Χριστιανοί Ορθόδοξοι καθώς και ότι υπάρχει ένα σημαντικό ποσοστό (14.8%) Χριστιανών Καθολικών, ενώ υπάρχουν και κάποιοι που είτε δεν απάντησαν είτε πιστεύουν σε κάποιο άλλο θρήσκευμα. Σε κάθε περίπτωση, θα είχε ενδιαφέρον να ελέγξουμε την πιθανότητα εμφάνισης της νόσου σε σχέση με το θρήσκευμα.

▪ Φυλή

Για λόγους παρόμοιους με αυτούς του θρησκεύματος, και η φυλή από την οποία κάποιος προέρχεται θεωρείται πιθανός παράγοντας εμφάνισης της νόσου. Δυστυχώς όμως, το δείγμα μας αποτελείται κατά 98.4% από άτομα Καυκάσιας φυλής οπότε δεν θα μπορούσαμε στη συνέχεια να εξάγουμε συμπεράσματα για τον παράγοντα αυτό.

- **Χώρα Γέννησης**

Υπολογίζοντας τα ποσοστά του δείγματος ανά χώρα καταγωγής βρήκαμε πως το 97% του δείγματος έχει γεννηθεί στην Ελλάδα, ενώ από αυτούς, το 58% είναι ασθενείς.

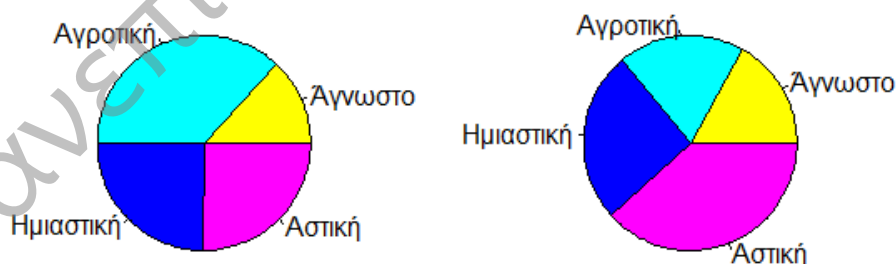
- **Τύπος Τόπου Διαμονής**

Ένας ακόμα παράγοντας που θεωρείται σημαντικός είναι ο τύπος του τόπου διαμονής τόσο κατά την παιδική ηλικία όσο και κατά την ημέρα της εξέτασης. Έτσι, οι τοποθεσίες διαχωρίστηκαν σε αγροτικές, αστικές και ημιαστικές και τα αποτελέσματα τα οποία λάβαμε ήταν τα εξής:

Πίνακας 3.3: Ποσοστά τύπου τόπου διαμονής κατά την παιδική ηλικία και την τώρα

Τύπος τοποθεσίας	Χρόνος διαμονής	
	Παιδική Ηλικία	Τώρα
Αγροτική	36.8%	19.2%
Αστική	25.2%	38.4%
Ημιαστική	24.8%	25.5%
Άγνωστο	13.2%	16.9%

Και διαγραμματικά:



Διάγραμμα 3.6: Κυκλικό διάγραμμα για την κατανομή των τόπων διαμονής α)κατά την παιδική ηλικία και β)τώρα

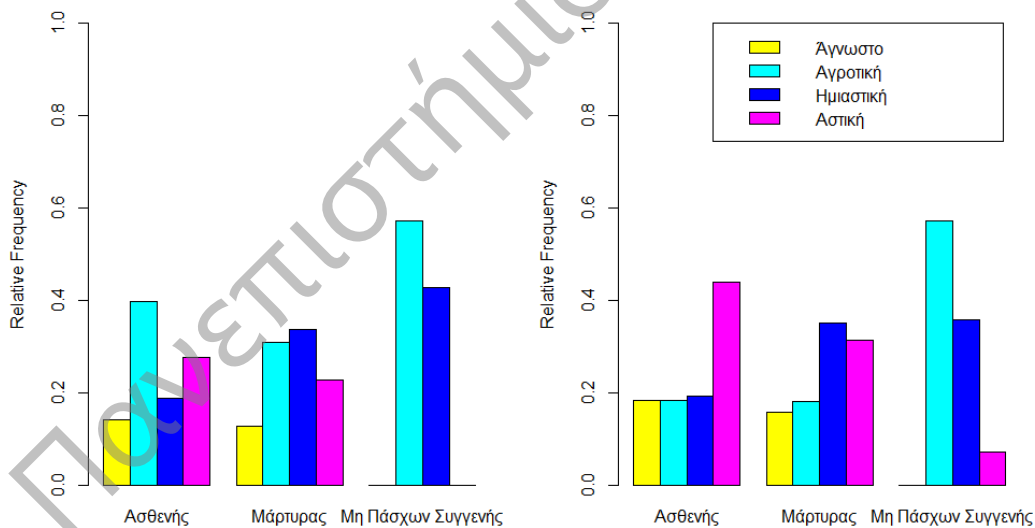
Παρατηρούμε πως το 50% των εθελοντών διέμενε ισόποσα σε αστικές και ημιαστικές τοποθεσίες κατά την παιδική ηλικία. Παρόλα αυτά, η πλειονότητα διέμενε κυρίως σε αγροτικές τοποθεσίες. Αντίθετα, ο σημερινός τόπος κατοικίας των εθελοντών φαίνεται πως είναι μόνο κατά 19% σε αγροτικές τοποθεσίες, με τις αστικές και ημιαστικές τοποθεσίες να καταλαμβάνουν ποσοστό 64%.

Επεκτείνοντας την ανάλυσή μας για κάθε κατάσταση εθελοντή έχουμε:

Πίνακας 3.4: Ποσοστά τύπου τόπου διαμονής κατά την παιδική ηλικία και την τώρα ανά κατάσταση εθελοντή

		Κατάσταση εθελοντή		
Χρόνος τοποθεσίας	Τύπος τοποθεσίας	Ασθενής	Μάρτυρας	Μη Πάσχων Συγγενής
Παιδική Ηλικία	Αγροτική	39.7%	30.8%	57.1%
	Αστική	27.7%	22.7%	0.0%
	Ημιαστική	18.6%	33.7%	42.9%
	Άγνωστο	14.0%	12.8%	0.0%
Τώρα	Αγροτική	18.4%	18.0%	7.2%
	Αστική	44.0%	31.3%	57.1%
	Ημιαστική	19.2%	35.1%	35.7%
	Άγνωστο	18.4%	15.6%	0.0%

Και διαγραμματικά:



Διάγραμμα 3.7: Σχετικές συχνότητες τύπων τόπου διαμονής α)κατά την παιδική ηλικία και β) τώρα, ανά κατάσταση εθελοντή

Εδώ, ο τόπος διαμονής κατά την παιδική ηλικία δεν φαίνεται να διαφέρει σημαντικά από τα προηγούμενα συμπεράσματά μας όσον αφορά τους ασθενείς, ενώ οι μάρτυρες φαίνεται να διαμένουν σχεδόν ισόποσα σε αγροτικές, αστικές και ημιαστικές τοποθεσίες. Από την άλλη, ο τόπος σημερινής διαμονής σε αστικές περιοχές φαίνεται να είναι διπλάσιος για τους ασθενείς σε σχέση με αυτούς που

κατοικούν σε αγροτικές ή ημιαστικές τοποθεσίες. Για τους μάρτυρες, παρατηρούμε πως ο τύπος τόπου διαμονής κατά την παιδική ηλικία δεν διαφέρει σημαντικά σε σχέση με τον σημερινό τύπο τόπου διαμονής, με εξαίρεση αυτούς που διέμεναν σε αγροτική τοποθεσία κατά την παιδική ηλικία των οποίων το ποσοστό έχει μειωθεί στο μισό περίπου όσον αφορά τη σημερινή τους διαμονή.

- **Συγγένεια εξ αίματος των γονέων**

Σημαντικός παράγοντας κινδύνου για τη νόσο θεωρείται η συγγένεια εξ αίματος μεταξύ των γονέων καθώς σε αυτήν την περίπτωση αυξάνεται η πιθανότητα κληρονομησης σπανίων αλληλομόρφων γονιδίων που εκφράζουν την νόσο. Εξετάζοντας το δείγμα μας βλέπουμε πως στο 97.1% δεν έχει γονείς οι οποίοι να συνδέονται με εξ αίματος συγγένεια ενώ μόλις 5 άτομα (ποσοστό 0.8%) έχει. Από τα άτομα αυτά, οι τέσσερις είναι ασθενείς και ένας είναι μάρτυρας. Συγκεκριμένα, ανά κατάσταση εθελοντή έχουμε:

Πίνακας 3.5: Ποσοστά συγγένειας εξ' αίματος των γονέων ανά κατάσταση εθελοντή

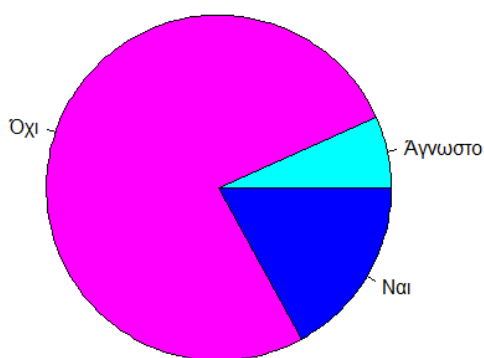
Συγγένεια εξ αίματος	Κατάσταση εθελοντή		
	Ασθενής	Μάρτυρας	Μη Πάσχων Συγγενής
Όχι	96.8%	97.2%	1.0%
Ναι	1.2%	0.5%	0%
Άγνωστο	2.0%	2.3%	0%

- **Οικογενειακό ιστορικό νόσου Πάρκινσον**

Η μελέτη του οικογενειακού ιστορικού στη νόσο του Πάρκινσον έχει οδηγήσει σε σημαντική συζήτηση για το ρόλο των γενετικών παραγόντων στην ανάπτυξη της PD. Παρά το γεγονός αυτό, το οικογενειακό ιστορικό έχει αναγνωριστεί ως ένας ανεξάρτητος παράγοντας κινδύνου για την ΝΠ (Sellbach et al., 2006).

Στη συντριπτική πλειονότητα των περιπτώσεων η νόσος του Πάρκινσον δεν κληρονομείται. Ωστόσο, επιστήμονες έχουν ανακαλύψει αρκετά γονίδια υπεύθυνα για την εμφάνιση της νόσου σε έναν πολύ μικρό αριθμό οικογενειών παγκοσμίως.

Στο δείγμα μας, το 76.2% των εθελοντών δεν έχει οικογενειακό ιστορικό για τη νόσο του Πάρκινσον ενώ το σημαντικό ποσοστό του 17.1% φαίνεται πως έχει.



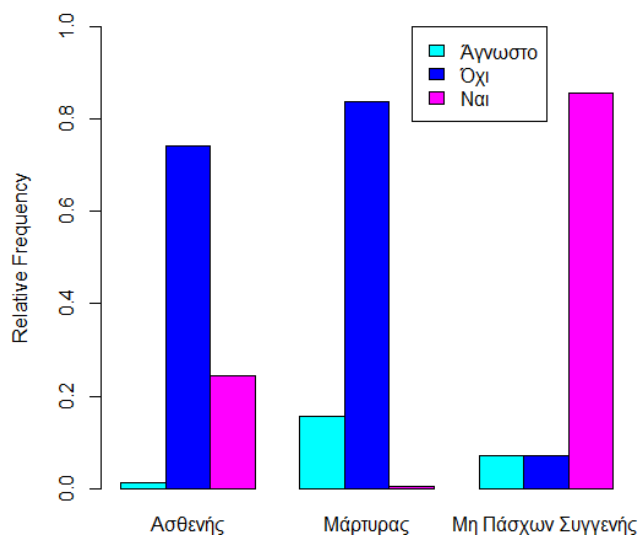
Διάγραμμα 3.8: Κυκλικό διάγραμμα για την κατανομή εμφάνισης οικογενειακού ιστορικού ως προς την νόσο Πάρκινσον

Από το ποσοστό των εθελοντών το οποίο έχει οικογενειακό ιστορικό, το 86.6% προέρχεται από το δείγμα των ασθενών και μόλις το 13.4% από αυτό των μαρτύρων ή των μη πασχόντων συγγενών, γεγονός το οποίο δίνει μια πρώτη ένδειξη πως ο παράγοντας αυτός φαίνεται να είναι στατιστικά σημαντικός ως προς την εμφάνιση της νόσου Πάρκινσον. Μελετώντας τα αντίστοιχα ποσοστά ανά κατάσταση εθελοντή, έχουμε:

Πίνακας 3.6: Ποσοστά οικογενειακού ιστορικού ΝΠ ανά κατάσταση εθελοντή

Οικογενειακό ιστορικό ΝΠ	Κατάσταση εθελοντή		
	Ασθενής	Μάρτυρας	Μη Πάσχων Συγγενής
Όχι	74.3%	83.9%	7.1%
Ναι	24.5%	0.5%	85.7%
Άγνωστο	1.2%	15.6%	7.1%

Και διαγραμματικά:



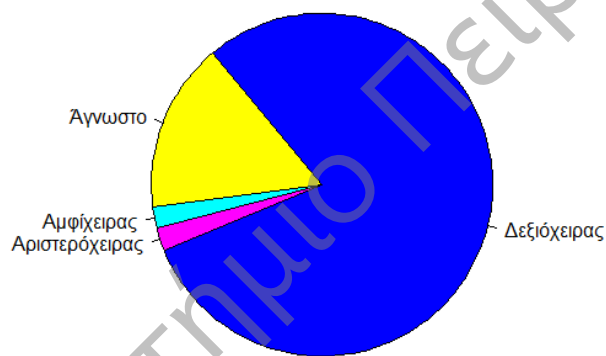
Διάγραμμα 3.9: Σχετικές συχνότητες οικογενειακού ιστορικού ΝΠ ανά κατάσταση εθελοντή

Δεδομένου ότι οι περισσότεροι μη πάσχοντες συγγενής επιλέχθηκαν από το στενό οικογενειακό περιβάλλον των ασθενών, είναι αναμενόμενο αυτοί να παρουσιάζουν τα υψηλότερα ποσοστά εμφάνισης οικογενειακού ιστορικού της νόσου Πάρκινσον.

▪ Δεξιοχειρία / Αριστεροχειρία

Καταγράφηκε επίσης το αν ο εθελοντής είναι δεξιόχειρας, αριστερόχειρας ή και αμφίχειρας. Τα αποτελέσματα που λάβαμε ήταν πως το 80% του δείγματος αποτελείται από δεξιόχειρες, το 16% δεν απάντησε, ενώ μόλις το 2.1% είναι αριστερόχειρες και ένα 1.9% είναι αμφίχειρες.

Διαγραμματικά:



Διάγραμμα 3.10: Κυκλικό διάγραμμα για την κατανομή της χειρίας των εθελοντών

▪ Κάπνισμα

Ο παράγοντας αυτός παρουσιάζει αρνητική συσχέτιση με τον κίνδυνο εμφάνισης της νόσου του Parkinson. Ειδικότερα, σχετίζεται με μείωση του σχετικού κινδύνου έως και 60%. Η δράση αυτή του καπνίσματος αποδίδεται πιθανόν στη μειωμένη δραστηριότητα του MAO-B (-πρόκειται για ένα ένζυμο που βρίσκεται στην εξωτερική μεμβράνη των μιτοχονδρίων και μεταξύ άλλων λειτουργιών είναι υπεύθυνο για τη διάσπαση της ντοπαμίνης στον ανθρώπινο οργανισμό-) στους καπνιστές είτε σε απευθείας δράση της νικοτίνης, ενώ ενδιαφέρον παρουσιάζει η υπόθεση για ενδεχόμενη τροποποίηση από το κάπνισμα του γενετικά προκαθορισμένου κινδύνου για νόσο του Πάρκινσον (Scott et al.2005)

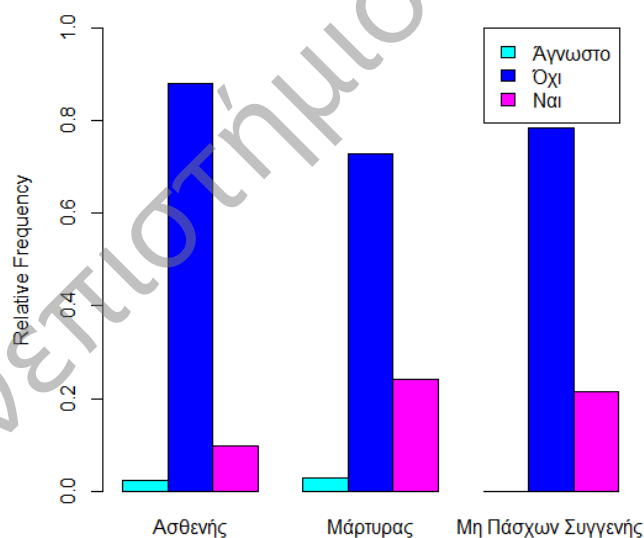
Οι μακροχρόνιοι καπνιστές προστατεύονται με κάποιο τρόπο από τη νόσο του Πάρκινσον, και αυτό δεν οφείλεται στο ότι πεθαίνουν νωρίτερα από άλλες αιτίες. Αν

και η ακριβής αιτία της προστατευτικής επίδρασης δεν είναι γνωστή, ερευνητές έχουν διαπιστώσει ότι περισσότερο ρόλο παίζει ο αριθμός των ετών που κάπνιζε κάποιος παρά τα τσιγάρα που κατανάλωνε ημερησίως ή μηνιαίως. Επιπλέον, η προστασία εξασθενεί σημαντικά μόλις οι καπνιστές κόψουν το κάπνισμα.

Το εύρημα της αντίστροφης σχέσης καπνίσματος και ΝΠ είναι σημαντικό, διότι θα μπορούσε να παράσχει ενδείξεις σχετικά με τις θεραπευτικές στρατηγικές για την προστασία από την εξουθενωτική διαταραχή της κίνησης. Ο καπνός περιέχει πολλές χημικές ουσίες που θα μπορούσαν να είναι υπεύθυνες για την εμφανή προστατευτική επίδραση. Η νικοτίνη θεωρείται ως μια χημική ουσία η οποία διεγείρει τα ντοπαμινεργικά συστήματα του εγκεφάλου και παρέχει κάποιο συμπτωματικό όφελος στην ΝΠ, ενώ παράλληλα έχει νευροπροστατευτική δράση (Quirk, 2004).

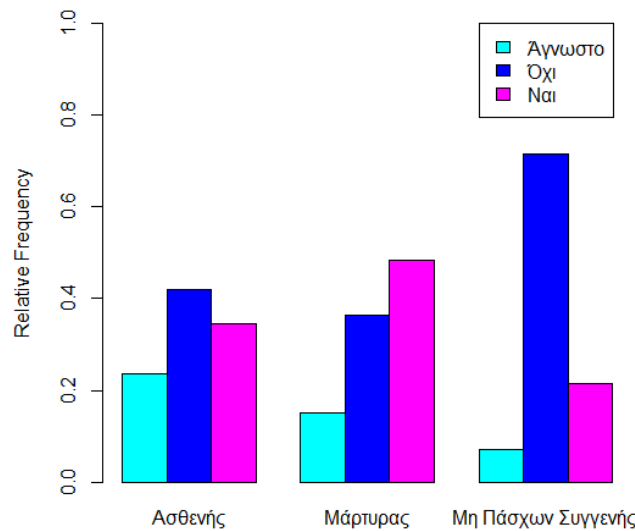
Τα αποτελέσματα που πήραμε για τη μελέτη των ανωτέρω για το σύνολο των εθελοντών ήταν τα εξής:

Το 82.2% του δείγματος δεν έχει καπνίσει τις τελευταίες 30 ημέρες, ενώ το 15.3% έχει καπνίσει. Τα ποσοστά αυτά για τους ασθενείς είναι 88% και 9.6%, αντίστοιχα.



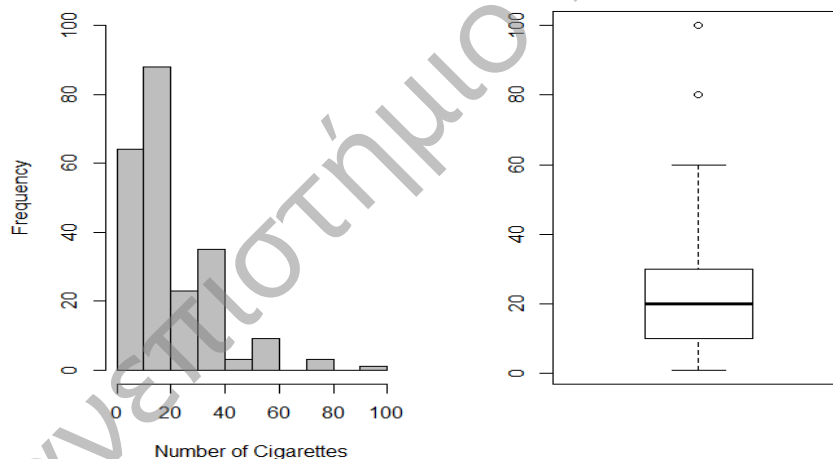
Διάγραμμα 3.11: Σχετικές συχνότητες καπνίσματος κατά τις τελευταίες 30 ημέρες ανά κατάσταση εθελοντή

Το 40.7% των εθελοντών δεν έχει καπνίσει 100 ή παραπάνω από 100 τσιγάρα στη ζωή του, ενώ το 39.3% έχουν καπνίσει. Τα ποσοστά αυτά για τους ασθενείς είναι 42% και 34.4%, αντίστοιχα.



Διάγραμμα 3.12: Σχετικές συχνότητες καπνίσματος 100 ή περισσότερων τσιγάρων στη ζωή τους, ανά κατάσταση εθελοντή

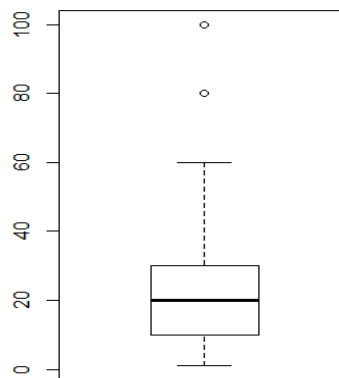
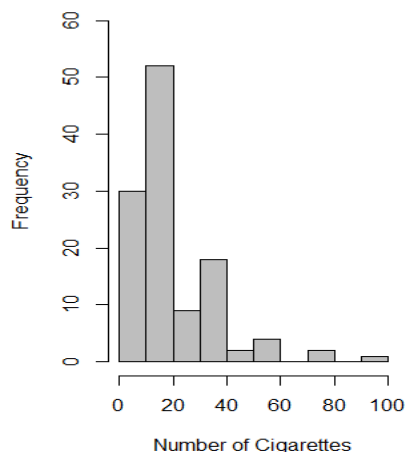
Για αυτούς που κάπνισαν τις τελευταίες 30 ημέρες ή έχουν καπνίσει παραπάνω από 100 τσιγάρα στη ζωή τους, ο μέσος αριθμός τσιγάρων που καταναλώνουν τώρα ή καταναλώναν στο παρελθόν είναι 23.13 και κατανέμεται ως εξής:



Διάγραμμα 3.13: α) Ιστόγραμμα του μέσου ημερήσιου αριθμού τσιγάρων που καταναλώθηκαν από τους εθελοντές, β) Θηκόγραμμα του μέσου ημερήσιου αριθμού τσιγάρων που καταναλώθηκαν από τους εθελοντές

Παρατηρούμε πως γενικά δεν υπάρχει ομοιομορφία στις καπνιστικές συνήθειες των εθελοντών, καθώς επίσης πως υπάρχουν και κάποιες ακραίες περιπτώσεις όπου καταναλώνονται 4 ή και 5 πακέτα (θεωρώντας πως κάθε πακέτο περιέχει 20 τσιγάρα) τσιγάρα ημερησίως.

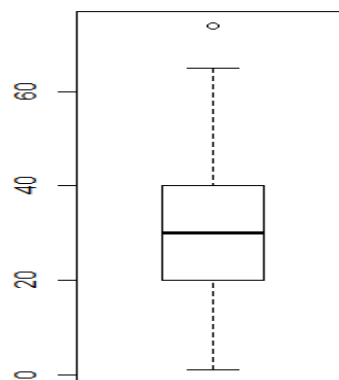
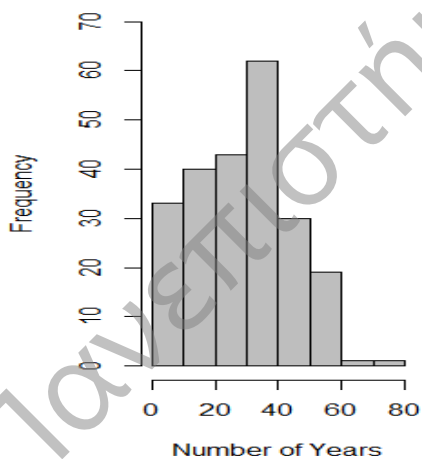
Μελετώντας τις καπνιστικές συνήθειες των ασθενών έχουμε:



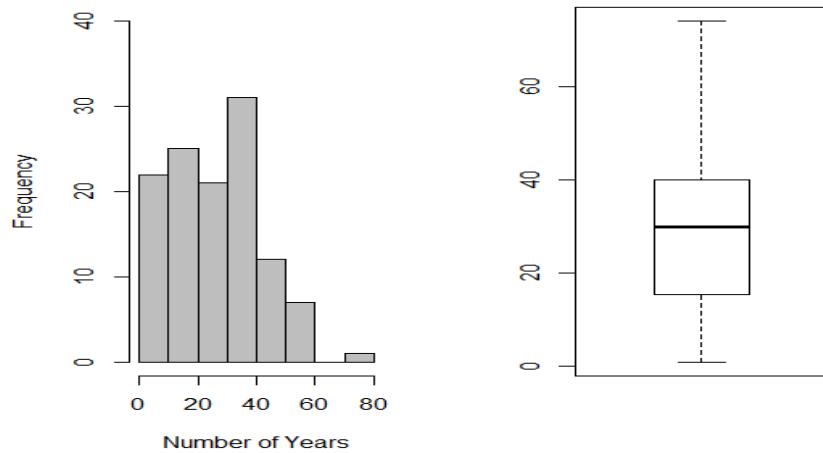
Διάγραμμα 3.14: α) Ιστόγραμμα του μέσου ημερήσιου αριθμού τσιγάρων που καταναλώθηκαν από τους ασθενείς, β) Θηκόγραμμα του μέσου ημερήσιου αριθμού τσιγάρων που καταναλώθηκαν από τους ασθενείς

Οι ασθενείς καπνίζουν ή κάπνιζαν κατά μέσο όρο 24.14 τσιγάρα ημερησίως ενώ οι ακραίες περιπτώσεις που παρατηρήθηκαν προηγουμένως αφορούν σε αυτούς.

Η καπνιστική συνήθεια στο σύνολο των εθελοντών διαρκεί κατά μέσο όρο 30.5 έτη, ενώ για το δείγμα των ασθενών ο αριθμός ανέρχεται στα 28.1 έτη.

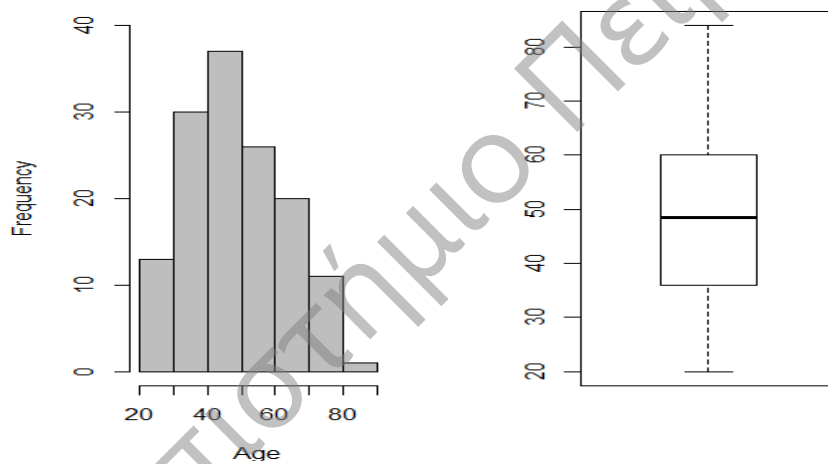


Διάγραμμα 3.15: α) Ιστόγραμμα για τη διάρκεια καπνίσματος για τους εθελοντές, β) Θηκόγραμμα για τη διάρκεια καπνίσματος για τους εθελοντές

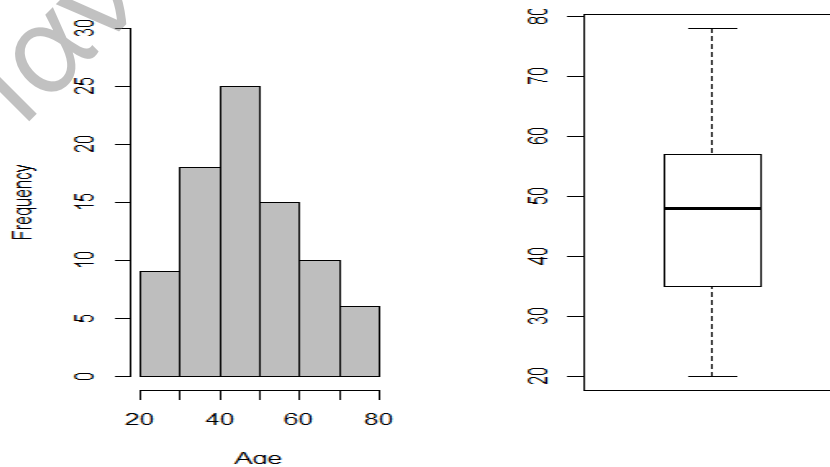


Διάγραμμα 3.16: α) Ιστόγραμμα για τη διάρκεια καπνίσματος για τους ασθενείς, β) Θηκόγραμμα για τη διάρκεια καπνίσματος για τους ασθενείς

Η μέση ηλικία παύσης καπνίσματος είναι τα 49 έτη για το σύνολο των εθελοντών, ενώ για τους ασθενείς τα 47.5 έτη.



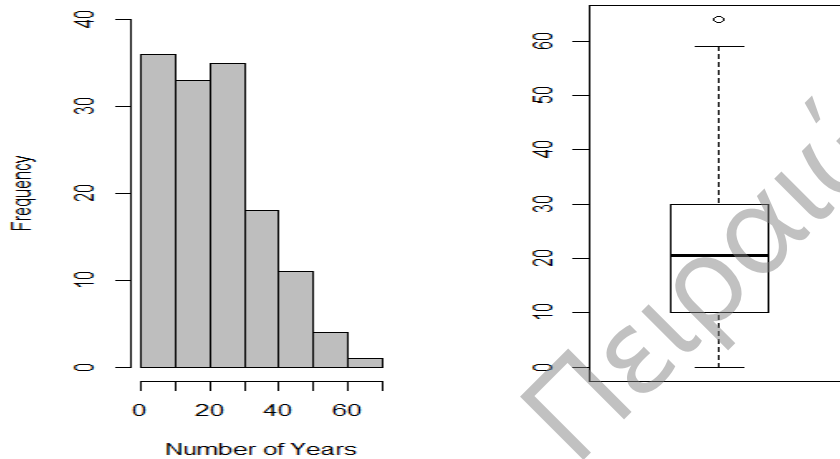
Διάγραμμα 3.17: α) Ιστόγραμμα για την ηλικία παύσης καπνίσματος για τους εθελοντές, β) Θηκόγραμμα για την ηλικία παύσης καπνίσματος για τους εθελοντές



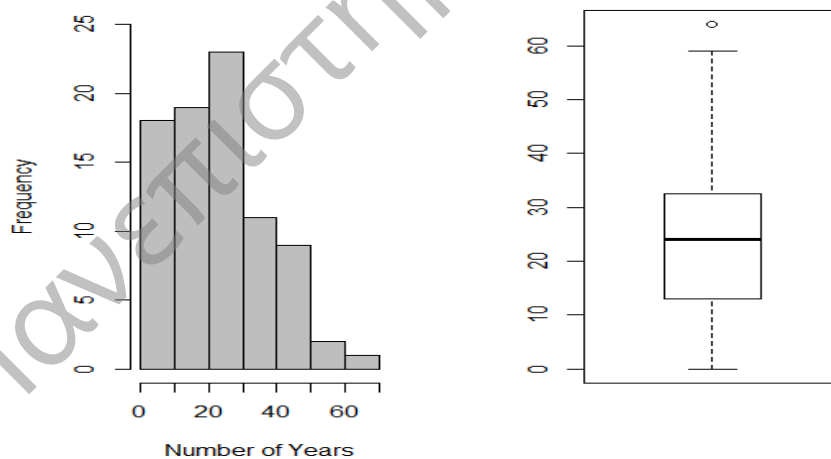
Διάγραμμα 3.18: α) Ιστόγραμμα για την ηλικία παύσης καπνίσματος για τους ασθενείς, β) Θηκόγραμμα για την ηλικία παύσης καπνίσματος για τους ασθενείς

Αφαιρώντας από την ηλικία του εθελοντή κατά την ημέρα καταχώρησής του στη βάση την ηλικία παύσης καπνίσματος, κατασκευάζουμε μια επιπλέον μεταβλητή, την Quit Smoking Age, η οποία μετρά το χρονικό διάστημα από τη διακοπή του καπνίσματος.

Άρα, για όσους έχουν κόψει το κάπνισμα, έχουν περάσει κατά μέσο όρο 22 έτη από τότε για το σύνολο των εθελοντών και 23.9 έτη για τους ασθενείς.



Διάγραμμα 3.19: α) Ιστόγραμμα για τη διάρκεια παύσης καπνίσματος για τους εθελοντές, β) Θηκόγραμμα για τη διάρκεια παύσης καπνίσματος για τους εθελοντές



Διάγραμμα 3.20: α) Ιστόγραμμα για τη διάρκεια παύσης καπνίσματος για τους ασθενείς, β) Θηκόγραμμα για τη διάρκεια παύσης καπνίσματος για τους ασθενείς

Και στα δύο παραπάνω θηκογράμματα παρατηρείται μια έκτροπη παρατήρηση. Η παρατήρηση αυτή αφορά έναν ασθενή 86 ετών, ο οποίος φαίνεται να έχει σταματήσει το κάπνισμα τα τελευταία 64 χρόνια και παρουσίασε την νόσο στην ηλικία των 77 ετών.

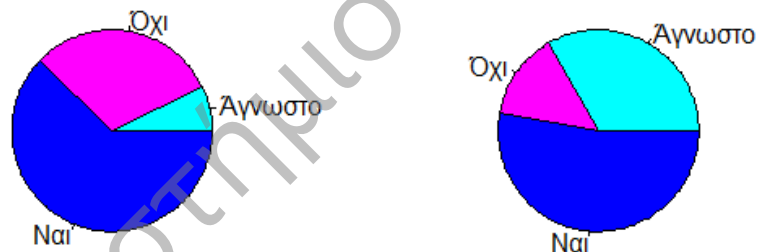
▪ Κατανάλωση καφέ

Μελέτες έχουν δείξει πως όσοι καταναλώνουν πολλή καφεΐνη είναι λιγότερο πιθανό να αναπτύξουν τη νόσο του Πάρκινσον (Hermansen et al., 2012; Qi & Li, 2013). Ακόμα, μια καναδική επιστημονική έρευνα υποδεικνύει ότι ο καφές, και γενικότερα η καφεΐνη, μπορεί να αμβλύνει την κινητική δυσλειτουργία σε όσους ήδη πάσχουν από τη νόσο (Postuma et al., 2012).

Μελετώντας τις συνήθειες των εθελοντών όσων αφορά στην κατανάλωση καφέ τώρα ή/και στο παρελθόν, τον μέσο αριθμό φλιτζανιών καφέ που καταναλώναν ημερησίως, καθώς και τα χρόνια κατανάλωσης και την ηλικία πάσης της κατανάλωσης εάν κάτι τέτοιο συμβαίνει, λάβαμε τα παρακάτω αποτελέσματα:

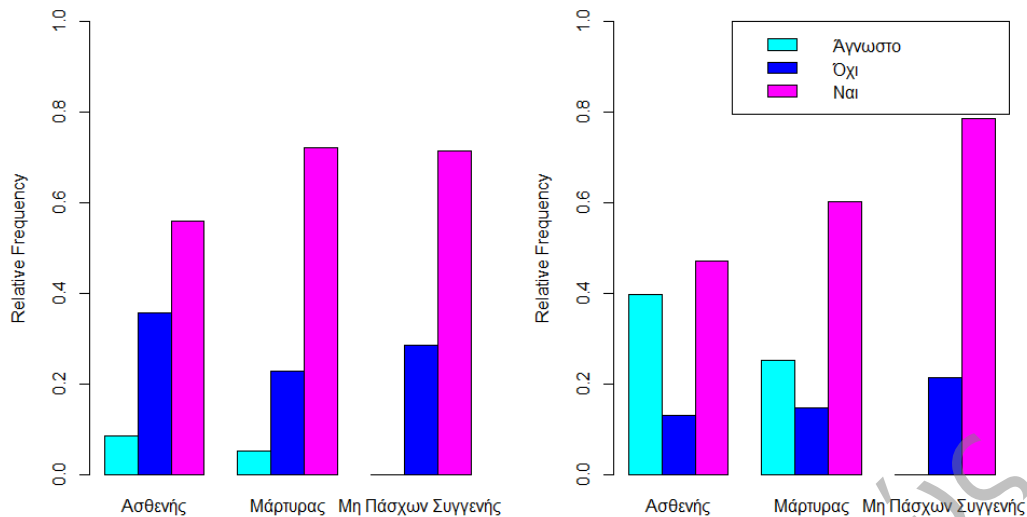
Το 62.3% των εθελοντών πίνει καφέ, ενώ το 30.6% δεν πίνει σήμερα καφέ. Ακόμα, το 52.8% συνήθιζε να καταναλώνει καφέ παλαιότερα, το 13.9% όχι, ενώ υπάρχει ένα σημαντικό ποσοστό της τάξης του 33.3% τα οποία δεν απάντησε στην ερώτηση αυτή.

Διαγραμματικά έχουμε:



Διάγραμμα 3.21: Κυκλικό διάγραμμα για την κατανομή της κατανάλωσης καφέ σε α) τωρινό και β) παρελθόντα χρόνο

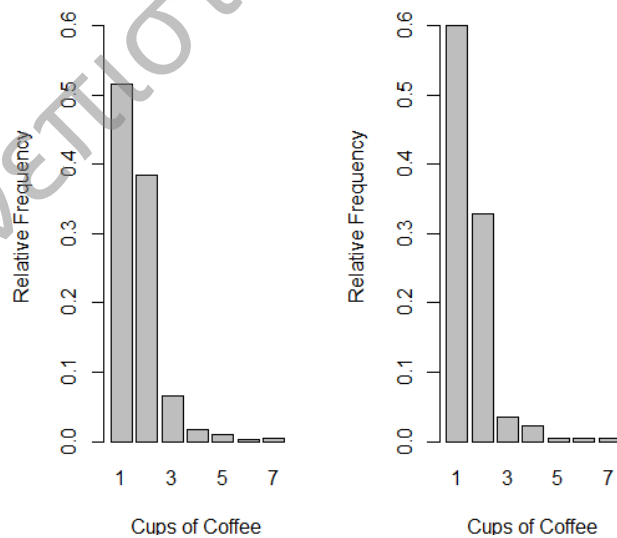
Τα αντίστοιχα διαγράμματα ανά κατηγορία εθελοντή είναι τα παρακάτω:



Διάγραμμα 3.22: Σχετικές συχνότητες κατανάλωσης καφέ ανά κατάσταση εθελοντή σε α)τωρινό και β) παρελθόντα χρόνο

Το 56% των ασθενών καταναλώνει καφέ σήμερα, ενώ το 47.2% κατανάλωνε παλαιότερα. Ακόμα, το 35.6% των ασθενών δεν πίνει καφέ σήμερα, ενώ το 13.1% δεν έπινε καφέ παλαιότερα.

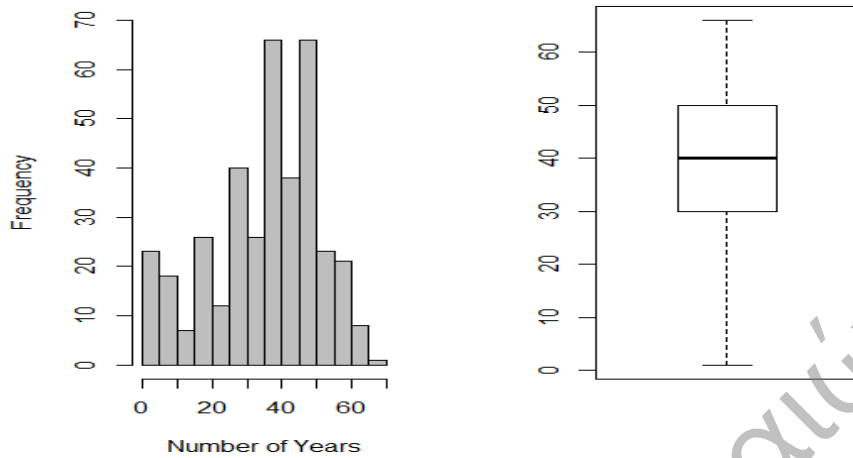
Καταγράφηκε, επίσης, ο μέσος αριθμός φλιτζανιών καφέ που καταναλώνονται τώρα ή καταναλώνονταν στο παρελθόν τόσο για το σύνολο των εθελοντών, όσο και για το δείγμα των ασθενών:



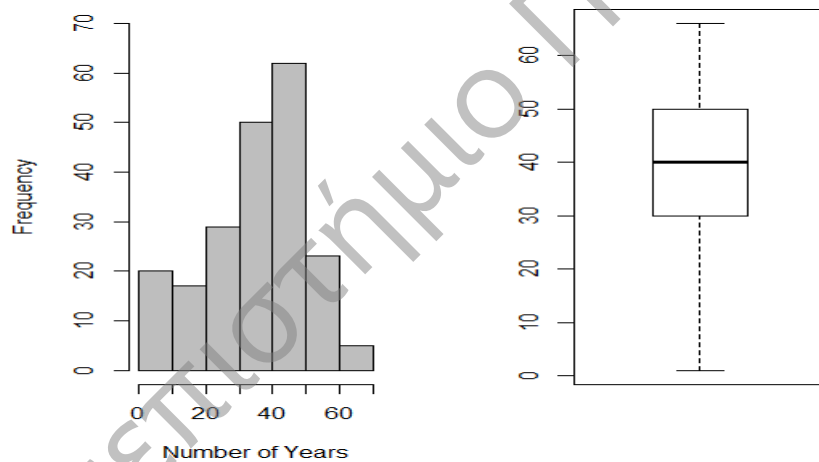
Διάγραμμα 3.23: Ραβδόγραμμα για τον μέσο αριθμό φλιτζανιών καφέ που καταναλώνονται α)από τους εθελοντές, β)από τους ασθενείς

Οι συνήθειες των δυο ομάδων δεν φαίνεται να διαφοροποιούνται σημαντικά.

Ακόμα, το σύνολο των εθελοντών καταναλώνει καφέ κατά μέσο όρο 36.8 έτη, ενώ το δείγμα των ασθενών 37.9 έτη.

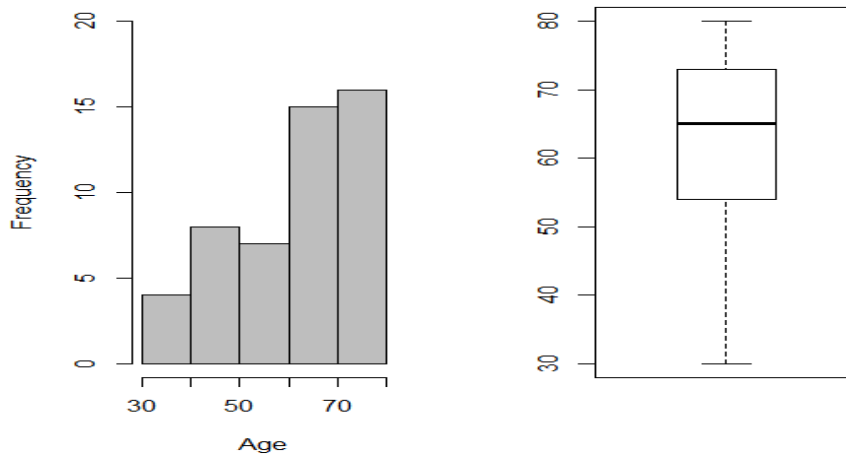


Διάγραμμα 3.24: α) Ιστόγραμμα για τη διάρκεια κατανάλωσης καφέ για τους εθελοντές, β) Θηκόγραμμα για τη διάρκεια κατανάλωσης καφέ για τους εθελοντές

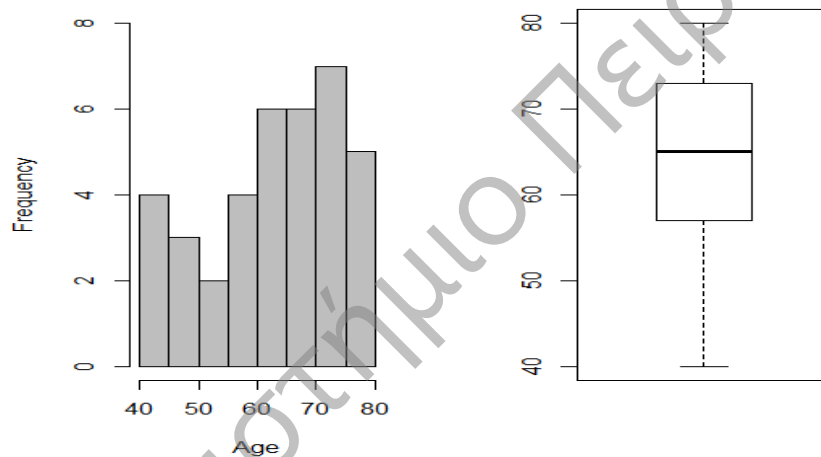


Διάγραμμα 3.25: α) Ιστόγραμμα για τη διάρκεια κατανάλωσης καφέ για τους ασθενείς, β) Θηκόγραμμα για τη διάρκεια κατανάλωσης καφέ για τους ασθενείς

Επιπλέον, καταγράφηκε η ηλικία παύσης κατανάλωσης καφέ για όσους έχουν σταματήσει. Για το σύνολο των εθελοντών η ηλικία αυτή είναι κατά μέσο όρο τα 62.4 έτη, ενώ για το δείγμα των ασθενών τα 63.6 έτη.



Διάγραμμα 3.26: α)Ιστόγραμμα για την ηλικία παύσης κατανάλωσης καφέ για τους εθελοντές, β)Θηκόγραμμα για την ηλικία παύσης κατανάλωσης καφέ για τους εθελοντές

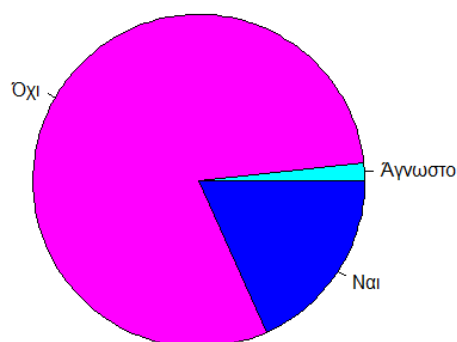


Διάγραμμα 3.27: α)Ιστόγραμμα για την ηλικία παύσης κατανάλωσης καφέ για τους ασθενείς, β)Θηκόγραμμα για την ηλικία παύσης κατανάλωσης καφέ για τους ασθενείς

- Έκθεση σε ζιζανιοκτόνα ή/και φυτοφάρμακα

Η έκθεση σε φυτοφάρμακα ή διαλύματα μπορεί ενδεχομένως να αυξάνει τον κίνδυνο εμφάνισης της νόσου Πάρκινσον. Έρευνα που δημοσιεύεται στο περιοδικό 'Neurology' (Pezzoli & Cereda, 2013) δείχνει ότι υπήρχε 33-80% αύξηση στον κίνδυνο εμφάνισης της νόσου, ανάλογα με την έκθεση. Ενδεχομένως όσοι ζουν σε αγροτικές περιοχές να αντιμετωπίζουν μεγαλύτερο κίνδυνο. Παρόλα αυτά, δεν έχει αποδειχθεί ότι τα φυτοφάρμακα άμεσα συμβάλλουν στην ανάπτυξη της νόσου του Πάρκινσον (Brown et al., 2006).

Μελετώντας την έκθεση του δείγματός μας σε ζιζανιοκτόνα ή/και φυτοφάρμακα λάβαμε τα παρακάτω αποτελέσματα:



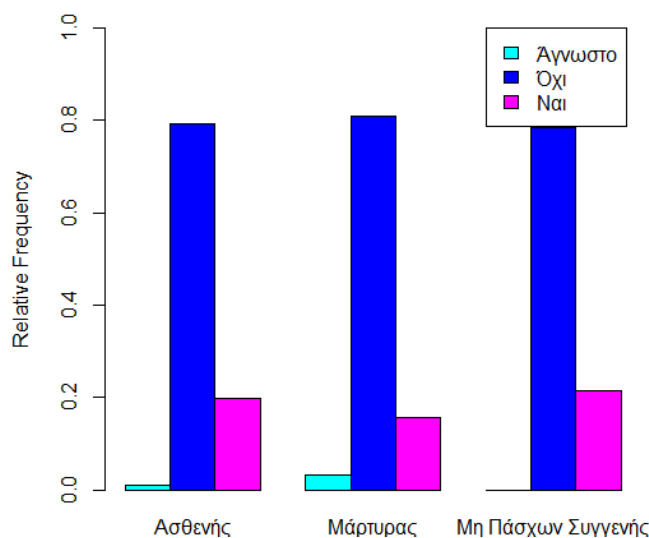
Διάγραμμα 3.28: Κυκλικό διάγραμμα για την κατανομή έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα των εθελοντών

Το 80% των εθελοντών δεν έχει εκτεθεί ποτέ σε ζιζανιοκτόνα ή/και φυτοφάρμακα κατά τη διάρκεια της ζωής του, ενώ το 18.3% έχει εκτεθεί. Από τους τελευταίους, το 65.4% αφορά σε ασθενείς της νόσου, ενώ το 34.6% σε μάρτυρες και μη πάσχοντες συγγενείς. Μελετώντας τα αντίστοιχα ποσοστά ανά κατάσταση εθελοντή, έχουμε:

Πίνακας 3.7: Ποσοστά έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα ανά κατάσταση εθελοντή

Έκθεση σε φυτοφάρμακα	Κατάσταση εθελοντή		
	Ασθενής	Μάρτυρας	Μη Πάσχων Συγγενής
Όχι	79.3%	81.0%	78.6%
Ναι	19.8%	15.6%	21.4%
Άγνωστο	0.9%	3.3%	0.0%

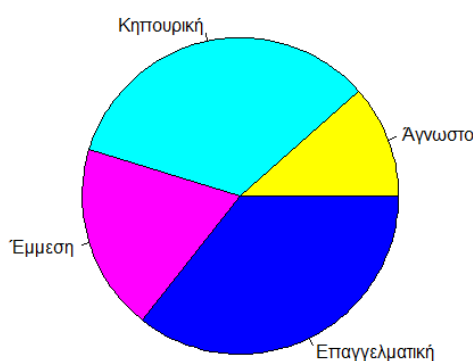
Και διαγραμματικά:



Διάγραμμα 3.29: Σχετικές συχνότητες έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα ανά κατάσταση εθελοντή

Παρατηρούμε πως οι σχετικές συχνότητες έκθεσης και μη έκθεσης ανά κατηγορία εθελοντή δεν διαφέρουν σημαντικά μεταξύ τους. Συνεπώς, λαμβάνουμε μια πρώτη ένδειξη πως ο παράγοντας της έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα δεν θα παίζει στατιστικά σημαντικό ρόλο για την εμφάνιση της νόσου.

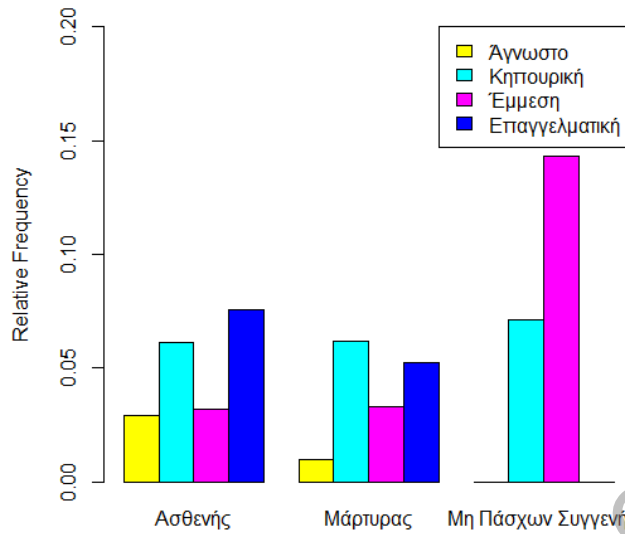
Όσοι απάντησαν θετικά για την έκθεση σε ζιζανιοκτόνα ή/και φυτοφάρμακα, ρωτήθηκαν επιπλέον για τον τύπο της έκθεσής τους, η οποία μπορούσε να ήταν είτε επαγγελματική, είτε κηπουρική, είτε έμμεση. Τα αποτελέσματα που προέκυψαν είναι τα κάτωθι:



Διάγραμμα 3.30: Κυκλικό διάγραμμα για την κατανομή του τύπου έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα των εθελοντών

Το 35.6% είχε επαγγελματική έκθεση σε ζιζανιοκτόνα ή/και φυτοφάρμακα, το 33.7% είχε κηπουρική έκθεση, ενώ το 19.2% είχε έμμεση έκθεση.

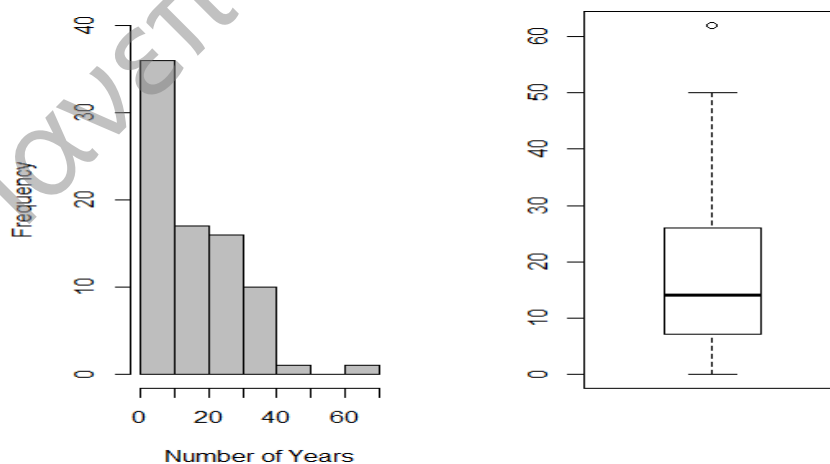
Μελετώντας ανά κατηγορία εθελοντή προέκυψαν τα εξής:



Διάγραμμα 3.31: Σχετικές συχνότητες τύπου έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα ανά κατάσταση εθελοντή

Το 38.2% των ασθενών που έχουν εκτεθεί σε ζιζανιοκτόνα ή/και φυτοφάρμακα ήταν λόγω της εργασίας τους, το 30.9% λόγω κηπουρικής ενασχόλησης, ενώ έμμεση ήταν η έκθεση για το 16.2% αυτών. Οι σχετικές αυτές συχνότητες είναι παρόμοιες με αυτές των μαρτύρων, καθώς για αυτούς τα ποσοστά είναι 33.3%, 39.4% και 21.2%, αντίστοιχα.

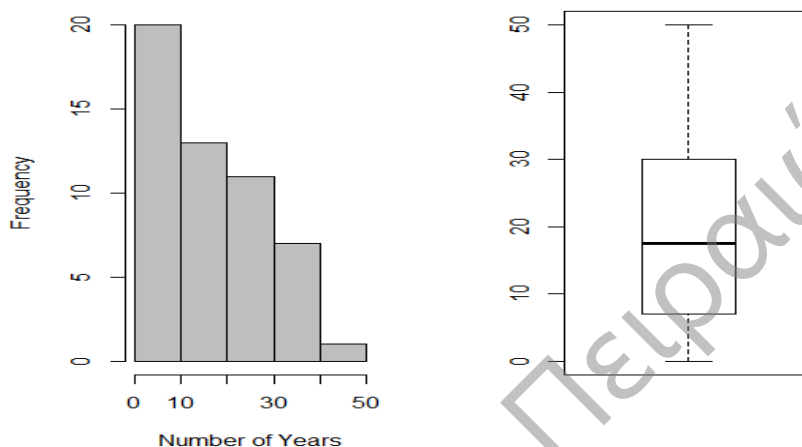
Καταγράφηκαν επίσης, τα χρόνια έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα:



Διάγραμμα 3.32: α) Ιστόγραμμα του μέσου χρόνου έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα για το σύνολο των εθελοντών, β) Θηκόγραμμα του μέσου χρόνου έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα για το σύνολο των εθελοντών

Ο χρόνος έκθεσης διαρκεί από 0 έως 26 έτη για το σύνολο των εθελοντών, με μέση διάρκεια έκθεσης να είναι τα 17.5 έτη. Ακόμα, παρατηρείται μια έκτροπη τιμή, η οποία αφορά έναν μάρτυρα 79 ετών ο οποίος διαμένει τόσο κατά την ημέρα εξέτασης όσο και την παιδική του ηλικία σε αγροτικού τύπου περιοχή, ενώ η έκθεσή του είναι επαγγελματικής φύσεως.

Μελετώντας τις έκθεση των ασθενών έχουμε:



Διάγραμμα 3.33: α) Ιστόγραμμα του μέσου χρόνου έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα για τους ασθενείς, β) Θηκόγραμμα του μέσου χρόνου έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα για τους ασθενείς

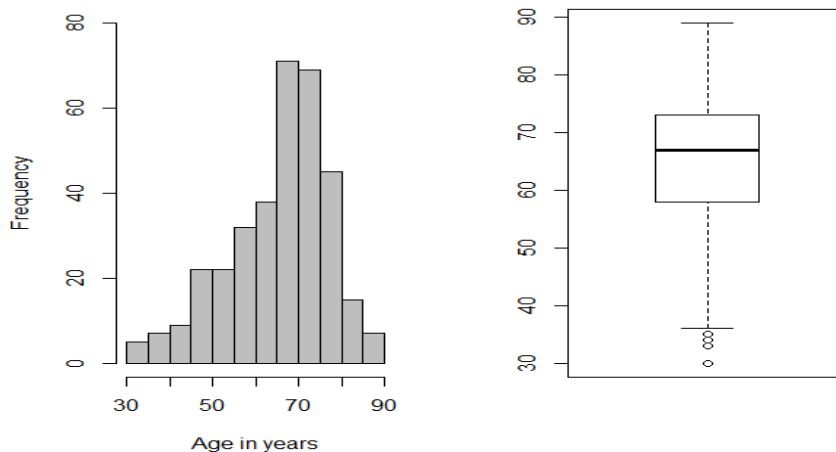
Ο χρόνος έκθεσης για τους ασθενείς διαρκεί από 0 έως 30 έτη για το σύνολο των εθελοντών, με μέση διάρκεια έκθεσης να είναι και πάλι τα 17.5 έτη.

3.3 Περιγραφική Ανάλυση για το σύνολο των ασθενών

- Έτος έναρξης της νόσου

Όπως έχουμε ήδη αναφέρει, ο μέσος όρος της ηλικίας εμφάνισης της νόσου είναι τα 60 έτη. Σύμφωνα με στοιχεία της Ευρωπαϊκής Εταιρείας για τη Νόσο του Πάρκινσον (E.P.D.A.), περίπου το 1% των ανθρώπων σε ηλικία άνω των 60 ετών παρουσιάζει την ασθένεια. Όμως, στο 10% περίπου των ασθενών η διάγνωση γίνεται πριν τα 50 έτη. Επιπλέον, υπάρχει και η νεανική μορφή παρκινσονισμού η οποία εμφανίζεται σε άτομα κάτω των 40 ετών και σε ποσοστό 5 έως 10% των συνολικών ασθενών.

Στους ασθενείς της βάσης δεδομένων μας η κατανομή της ηλικίας έναρξης της νόσου είναι η παρακάτω:



Διάγραμμα 3.34: α) Ιστογράμμο για την ηλικία έναρξης της ΝΠ, β) Θηκόγραμμα για την ηλικία έναρξης της ΝΠ

Ο μέσος όρος ηλικίας εμφάνισης της νόσου των ασθενών στη βάση δεδομένων είναι τα 64.8 έτη ενώ η διάμεσος τα 67 έτη. Το θηκόγραμμα μας ενημερώνει για κάποιες ακραίες τιμές στο δείγμα μας, οι οποίες λογικά οφείλονται σε περιπτώσεις νεανικής μορφής παρκινσονισμού. Μελετώντας τις παρατηρήσεις με ηλικία έναρξης έως 35 έτη, συμπεραίνουμε πως αφορούν ασθενείς που: κατά 80% είναι γυναίκες, που έχουν διαμείνει σε όλη τους τη ζωή σε αστικές περιοχές και δεν έχουν εκτεθεί σε φυτοφάρμακα, που κατά 60% έχουν οικογενειακό ιστορικό για τη νόσο του Πάρκινσον και κατά 80% δεν έχουν καπνίσει τις τελευταίες 30 ημέρες.

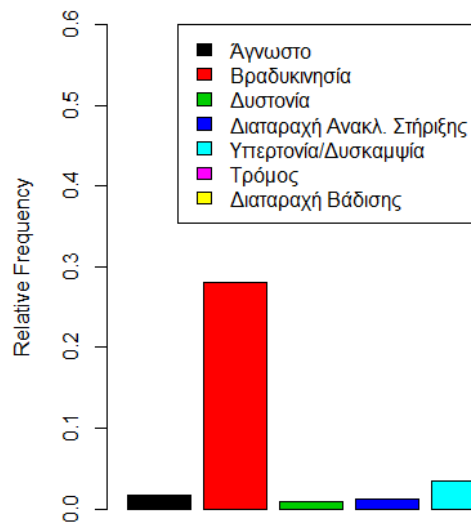
▪ Αρχικό σύμπτωμα

Βραδύτητα των κινήσεων, μυϊκή δυσκαμψία και τρέμουλο, κυρίως του ενός χεριού και κακή ισορροπία είναι να βασικά συμπτώματα της νόσου Πάρκινσον. Ο τρόμος (τρέμουλο) είναι πολύ συχνό, αλλά ένας στους τρεις ασθενείς δεν το έχει στα πρώιμα στάδια της νόσου. Το πρώτο σύμπτωμα που διαγνώστηκε στους ασθενείς είναι:

Πίνακας 3.8: Ποσοστά εμφάνισης αρχικού συμπτώματος

Αρχικό Σύμπτωμα	Ποσοστό
Τρόμος	54.8%
Βραδυκινησία	28%
Διαταραχή Βάδισης	9.9%
Υπέρτονια/Δυσκαμψία	3.5%
Διαταραχή Αντανακλαστικών Στήριξης	1.2%
Δυστονία	0.9%

Και διαγραμματικά:



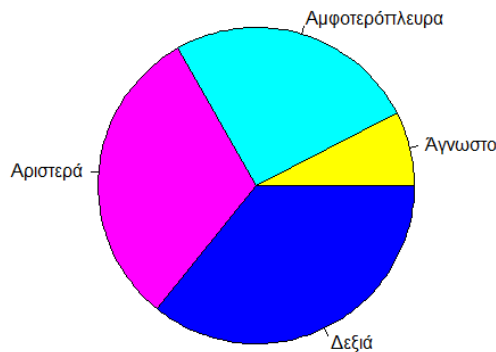
Διάγραμμα 3.35: Σχετικές συχνότητες εμφάνισης εκάστοτε αρχικού συμπτώματος

Παρατηρούμε ότι το ποσοστό του δεύτερου κατά σειρά περισσότερο παρατηρούμενου συμπτώματος είναι περίπου το μισό από αυτό του πρώτου. Συνεπώς, επιβεβαιώνεται η άποψη ότι ένας στους τρεις ασθενείς δεν παρουσιάζει το τρέμουλο σαν πρώτο σύμπτωμα εκδήλωσης της νόσου.

▪ Πλευρά έναρξης

Έχει διαπιστωθεί πως τα συμπτώματα της νόσου Πάρκινσον ξεκινούν από τη μία πλευρά και με την επιδείνωσή της ασθένειας εμφανίζονται και στη δεύτερη. Στο δείγμα της μελέτης μας, οι ασθενείς δήλωσαν πως τα αρχικά συμπτώματα εμφανίστηκαν κατά 35.9% στη δεξιά τους πλευρά, κατά 30.9% στην αριστερή τους πλευρά και κατά 25.6% αμφοτερόπλευρα.

Διαγραμματικά έχουμε:



Διάγραμμα 3.36: Κυκλικό διάγραμμα για την κατανομή της πλευράς έναρξης της ΝΠ

- **Συμπτώματα που εμφάνισαν**

Με το πέρασμα του χρόνου, η εξέλιξη της νόσου χαρακτηρίζεται από επιδείνωση, η οποία συνήθως γίνεται με αργό ρυθμό. Κάθε ασθενής μπορεί να εκδηλώσει ή να μην εκδηλώσει κάποιο από τα συμπτώματα που χαρακτηρίζουν την νόσο. Οι ασθενείς που ρωτήθηκαν διαμόρφωσαν την παρακάτω εικόνα:

Πίνακας 3.9: Ποσοστά εμφάνισης συμπτωμάτων

Σύμπτωμα	Εμφάνιση συμπτώματος	
	Ναι	Όχι
Τρόμος	77%	22.2%
Δυστονία	8.2%	90.4%
Βραδυκινήσια	95.3%	35%
Υπέρτονια/Δυσκαμψία	85.7%	12.5%
Διαταραχή Αντανακλαστικών Στήριξης	30%	66.2%
Διαταραχή Βάδισης	59.8%	31.8%
Άνοια	12.2%	86.3%
Κατάθλιψη	27.1%	70.8%
Άλλα Ψυχικά Συμπτώματα	15.2%	82.2%
Συμπτώματα από ΑΝΣ	37.9%	58%
Κινητικές Επιπλοκές	28%	70%

Όσοι από τους παραπάνω εκδήλωσαν κινητικές επιπλοκές ρωτήθηκαν επιπλέον αν παρουσίασαν τυχόν διακυμάνσεις ή και δυσκινήσιες. Τα αποτελέσματα ήταν τα εξής:

Πίνακας 3.10: Ποσοστά εμφάνισης κινητικών επιπλοκών με διακυμάνσεις ή δυσκινήσιες

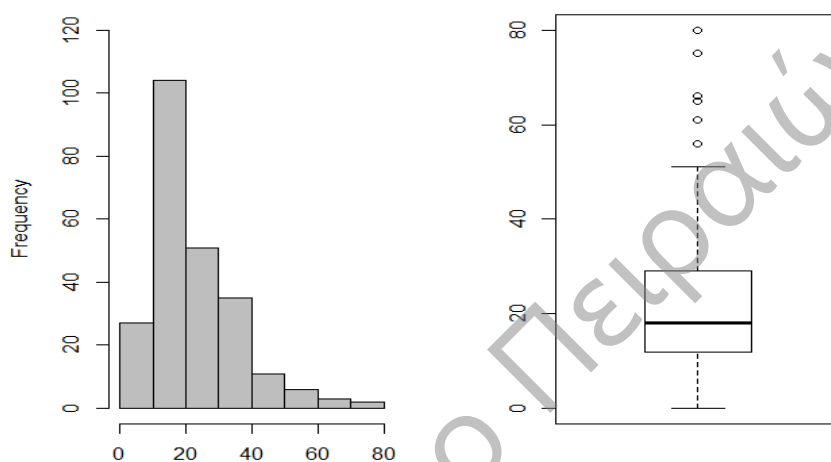
Κινητικές Επιπλοκές με:	Ναι	Όχι
Διακυμάνσεις	84.4%	11.5%
Δυσκινήσιες	65.6%	27.1%

- **Αξιολόγηση των ασθενών**

Όπως έχουμε ήδη αναφέρει, χρησιμοποιούνται τέσσερις διαφορετικές κλίμακες αξιολόγησης με στόχο την παρακολούθηση των επιπτώσεων της νόσου Πάρκινσον, το βαθμό της αναπηρίας που προκαλείται από αυτήν, την αξιολόγηση της έκθεσης σχετικά με τη νόσο, την αξιολόγηση των δραστηριοτήτων της καθημερινής ζωής, καθώς και τη διάγνωση της γνωστικής του δυσλειτουργίας.

Η κλίμακα UPDRS I έως III λαμβάνει τιμές αθροιστικά από το 0 έως το 199. Λόγω του μεγάλου εύρους τιμών, και σύμφωνα με το Κεντρικό Οριακό Θεώρημα, μπορούμε να θεωρήσουμε την μέτρηση αυτή ως συνεχή μεταβλητή. Στο δείγμα μας, η αξιολόγηση των ασθενών κυμαίνεται μεταξύ 0 και 80 «βαθμών», πράγμα που υποδεικνύει πως η αναπηρία τους δεν είχε λάβει εκτενείς διαστάσεις κατά την ημερομηνία εξέτασής τους και συνεπώς βρίσκονταν σε καλή κατάσταση.

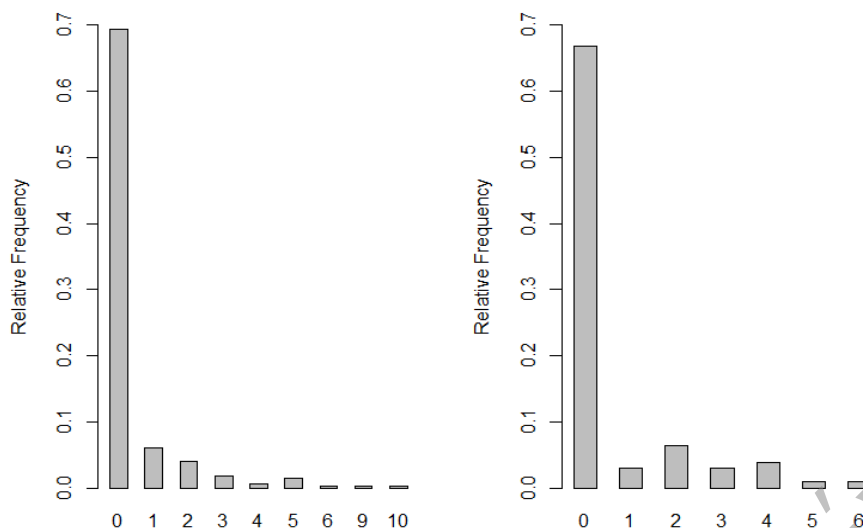
Διαγραμματικά έχουμε:



Διάγραμμα 3.37: α) Ιστόγραμμα για την απεικόνιση της αξιολόγησης των ασθενών σύμφωνα με την κλίμακα UPDRS I έως και III, β) Θηκόγραμμα για την απεικόνιση της αξιολόγησης των ασθενών σύμφωνα με την κλίμακα UPDRS I έως και III

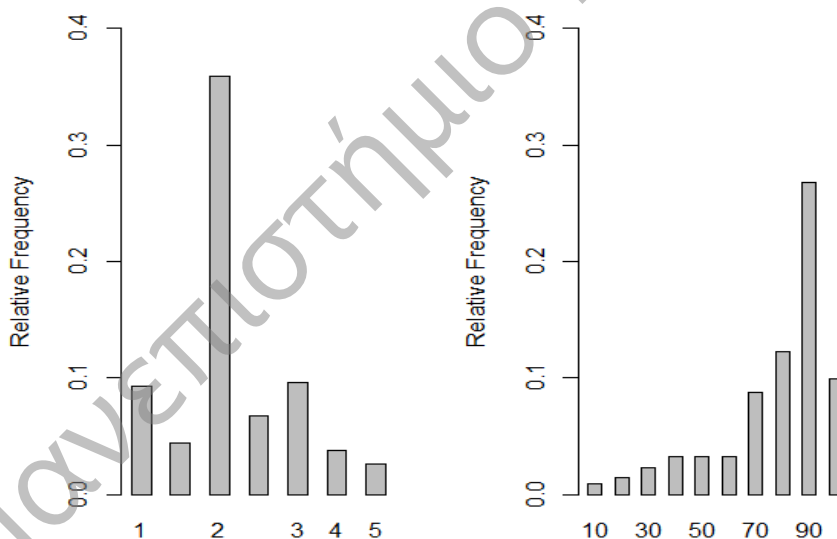
Οι ασθενείς έχουν αξιολογηθεί συχνότερα με βαθμολογίες μεταξύ 10 έως 20 «βαθμών» στην κλίμακα UPDRS I, II και III, και κατά μέσο όρο λαμβάνουν την βαθμολογία 22. Η κατανομή της μεταβλητής παρουσιάζει δεξιά ασυμμετρία, πράγμα που μας δίνει σαφείς ενδείξεις για τη μη κανονικότητά της, ενώ η διάμεσός της είναι η βαθμολογία 18.

Όσον αφορά στις υπόλοιπες κλίμακες, θεωρήσαμε πως οι βαθμολογίες στις κλίμακες αυτές είναι διακριτές μεταβλητές. Τα αποτελέσματα που λάβαμε είναι τα παρακάτω:



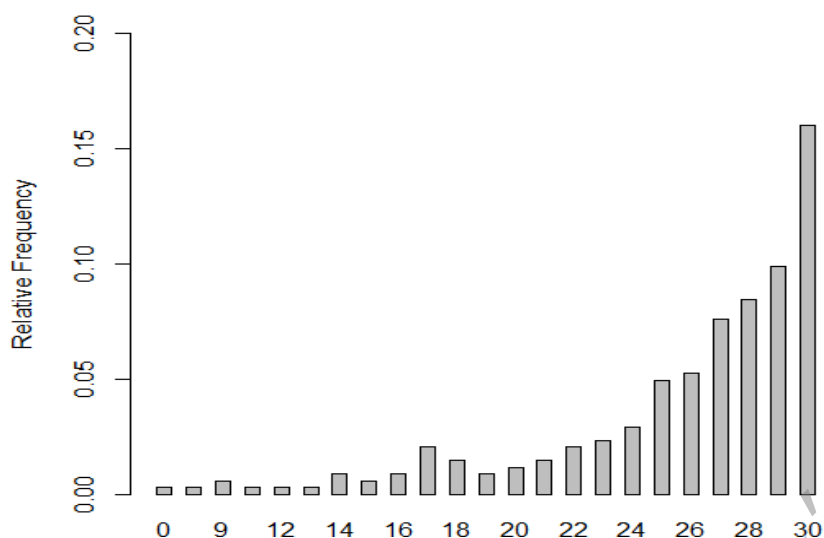
Διάγραμμα 3.38: Ραβδόγραμμα για την απεικόνιση της αξιολόγησης των ασθενών υπό την κλίμακα UPDRS α) IVA β) IVB

Η αξιολόγηση των ασθενών σύμφωνα με το μέρος IV της κλίμακας UPDRS υποδεικνύει πως δεν υπάρχουν επιπλοκές στη θεραπεία τους.



Διάγραμμα 3.39: Ραβδόγραμμα για την απεικόνιση της αξιολόγησης των ασθενών σύμφωνα με την κλίμακα α) Hoehn & Yahr, β) Schwab & England

Σύμφωνα με την αξιολόγηση της κλίμακας Hoehn & Yahr η πλειοψηφία των ασθενών βρίσκεται στο στάδιο 2, με συμπτώματα και στις δύο πλευρές αλλά χωρίς βλάβη στην ισορροπία, ενώ σύμφωνα με την κλίμακα Schwab & England οι περισσότεροι ασθενείς είναι πλήρως ανεξάρτητοι, ικανοί για όλες τις δραστηριότητες με κάποια βραδύτητα, η οποία μπορεί να απαιτεί έως και το διπλάσιο χρόνο για την ολοκλήρωσή τους.



Διάγραμμα 3.40: Ραβδόγραμμα για την απεικόνιση της αξιολόγησης των ασθενών σύμφωνα με την κλίμακα Standardized mini-mental Status Examination

Τέλος, η αξιολόγηση των ασθενών σύμφωνα την κλίμακα Standardized mini-mental Status Examination υποδεικνύει την καλή, χωρίς άνοια, γνωστική τους λειτουργία.

Τέλος, αναφέρουμε πως ο αριθμός των κλάσεων που έπρεπε να χρησιμοποιηθούν για το ιστόγραμμα της κατανομής της εκάστοτε μεταβλητής υπολογίστηκε σύμφωνα με το τύπο που προτάθηκε από τον H.A. Sturges το 1926, ο οποίος προέρχεται από μια διωνυμική κατανομή και σιωπηρά υποθέτει μια περίπου κανονική κατανομή (Sturges, H. A. 1926). Ο τύπος είναι ο εξής:

$$k = \log_{(2)} n + 1.$$

ΚΕΦΑΛΑΙΟ 4: ΠΙΝΑΚΕΣ ΣΥΝΑΦΕΙΑΣ

4.1 Εισαγωγή

Στο κεφάλαιο αυτό, θα διερευνήσουμε την εξάρτηση μεταξύ εμφάνισης της νόσου του Πάρκινσον σε σχέση με τους παράγοντες που πιστεύεται ότι την επηρεάζουν, όπως αυτό αναπτύχθηκε στο Κεφάλαιο 3.

Δημιουργήσαμε πίνακες συνάφειας μεταξύ ασθενών – μαρτύρων και του παράγοντα που εξετάζόταν κάθε φορά, και στη συνέχεια προχωρήσαμε σε στατιστικούς ελέγχους ανεξαρτησίας. Σε κάθε περίπτωση, υπολογίστηκαν ο σχετικός κίνδυνος (*relative risk, RR*), ο λόγος σχετικών πιθανοτήτων, καθώς και το 95% διάστημα εμπιστοσύνης γι' αυτόν.

Επιπλέον, προχωρήσαμε σε έλεγχο κανονικότητας, ομοσκεδαστικότητας και ισότητας των κατανομών για τους παράγοντες συνεχούς φύσης προκειμένου να εντοπίσουμε στατιστικώς σημαντικές διαφορές μεταξύ των ασθενών που εξετάστηκαν στο Αττικό νοσοκομείο και των ασθενών που εξετάστηκαν στο νοσοκομείο Σύρου.

Τέλος, διερευνήσαμε ποιοι από τους παραπάνω παράγοντες επηρεάζουν την εμφάνιση νεανικής μορφής παρκινσονισμού. Υπολογίστηκαν εκ νέου, ο σχετικός κίνδυνος, ο σχετικός λόγος πιθανοτήτων, καθώς και τα αντίστοιχα 95% διαστήματα εμπιστοσύνης.

4.2 Θεωρητικό υπόβαθρο

Ακολουθεί μια συνοπτική παρουσίαση των ελέγχων που χρησιμοποιήθηκαν για την ανάλυση των δεδομένων στο κεφάλαιο αυτό.

Το μεγαλύτερο θεωρητικό μέρος της παρουσίασης αυτής προέρχεται από τις πανεπιστημιακές σημειώσεις του κ. Αντζουλάκου «Εισαγωγή στο R» (2012) για το μάθημα «Ανάλυση δεδομένων με τη χρήση στατιστικών πακέτων».

4.2.1 Έλεγχος κανονικότητας Kolmogorov-Smirnov για ένα πληθυσμό

Έστω X_1, X_2, \dots, X_n τυχαίο δείγμα από πληθυσμό X με συνάρτηση κατανομής $F(x)$. Για τον έλεγχο της υπόθεσης:

$$H_0 : F(x) = F_0(x) \text{ με εναλλακτική την } H_1 : F(x) \neq F_0(x)$$

(η H_0 ισχύει για όλα τα x ενώ η H_1 ισχύει για τουλάχιστον ένα x), όπου $F_0(x)$ είναι μια πλήρως καθορισμένη συνάρτηση κατανομής συνεχούς τυχαίας μεταβλητής, χρησιμοποιείται συνήθως ο έλεγχος Kolmogorov-Smirnov. Η στατιστική συνάρτηση D_n του ελέγχου δίνεται από τη σχέση: $D_n = \sup_{-\infty < x < \infty} |S_n(x) - F_0(x)|$, όπου S_n είναι η εμπειρική συνάρτηση κατανομής του δείγματος. Προφανώς μεγάλες τιμές της D_n προσφέρουν ενδείξεις για την ισχύ της εναλλακτικής υπόθεσης. Η κρίσιμη περιοχή του ελέγχου σε επίπεδο σημαντικότητας α είναι η: $\mathcal{A}_n : d_n > D_{n,\alpha}$, όπου d_n είναι η παρατηρούμενη τιμή της D_n , και $D_{n,\alpha}$ το άνω α ποσοστιαίο σημείο της κατανομής του D_n . Η p -value του ελέγχου είναι ίση με $P(D_n > d_n | H_0)$.

4.2.2 Έλεγχος Mann-Whitney

Ο έλεγχος Mann-Whitney (U) για δύο ανεξάρτητα τυχαία δείγματα χρησιμοποιείται, όπως και ο έλεγχος Kolmogorov-Smirnov, για να ελέγξουμε αν αυτά προέρχονται από την ίδια συνεχή κατανομή χωρίς αυτή να προσδιορίζεται. Έτσι αν X_1, X_2, \dots, X_n και Y_1, Y_2, \dots, Y_m είναι τα δύο δείγματα, μας ενδιαφέρει ο έλεγχος της υπόθεσης:

$$H_0 : F_X(x) = F_Y(x) \text{ με εναλλακτική την } H_1 : F_X(x) \neq F_Y(x).$$

Η στατιστική συνάρτηση ελέγχου W δίνεται από τη σχέση $W = \sum_{j=1}^m S_j$, όπου S_j είναι ο βαθμός (rank) της παρατήρησης Y_j , $j = 1, 2, \dots, m$, στη διάταξη των $n+m$ παρατηρήσεων των δύο δειγμάτων από τη μικρότερη παρατήρηση έως τη μεγαλύτερη. Είναι προφανές ότι οι βαθμοί που αντιστοιχούν στα Y_j θα είναι γενικά μεγαλύτεροι από αυτούς που αντιστοιχούν στα X_i , $i = 1, 2, \dots, n$, όταν για τις διαμέσους ισχύει: $M_Y > M_X$. Ο Wilcoxon πρότεινε να αποδεχόμαστε την εναλλακτική υπόθεση για μεγάλες τιμές του W . Ανάλογα θα πρέπει να αποδεχόμαστε την εναλλακτική υπόθεση για μικρές τιμές του W . Η κρίσιμη περιοχή του ελέγχου σε επίπεδο σημαντικότητας α είναι η:

$$\mathcal{A}_n : w \leq w_{\alpha/2} \cup \mathcal{A}'_n : w \geq w'_{\alpha/2}$$

Οι τιμές $w_{\alpha/2}$ και $w'_{\alpha/2}$ υπολογίζονται κάνοντας χρήση της ακριβούς κατανομής του W . Η p -value του ελέγχου είναι ίση με:

$$2P(W^* \geq |w^*|) = 2(1 - \Phi(|w^*|)), \text{ όπου}$$

$$W^* = \frac{W - E(W)}{\sqrt{\text{Var}(W)}} = \frac{W - m(n+m+1)/2}{\sqrt{nm(n+m+1)/12}}$$

και w^* η παρατηρούμενη τιμή της.

4.2.3 Έλεγχος ισότητας των διακυμάνσεων δύο πληθυσμών

Έστω τα ανεξάρτητα δείγματα X_1, X_2, \dots, X_n και Y_1, Y_2, \dots, Y_n από κανονικούς πληθυσμούς $N(\mu_1, \sigma_1^2)$ και $N(\mu_2, \sigma_2^2)$, αντίστοιχα. Για τον έλεγχο:

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = \lambda_0 \text{ με εναλλακτική την } H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq \lambda_0$$

χρησιμοποιείται η στατιστική συνάρτηση ελέγχου $F = \frac{1}{\lambda_0} \cdot \frac{S_1^2}{S_2^2}$, η οποία ακολουθεί την κατανομή F_{n_1-1, n_2-1} όταν η H_0 είναι αληθής. Αν συμβολίσουμε με f την παρατηρούμενη τιμή της F , η κρίσιμη περιοχή του ελέγχου σε επίπεδο σημαντικότητας α είναι η

$$\mathcal{A} : f < F_{n_1-1, n_2-1; 1-\alpha/2} \cup \mathcal{A} : f > F_{n_1-1, n_2-1; \alpha/2}.$$

Η p -value του ελέγχου είναι ίση με:

$$2 \min\{P(F \leq f), P(F \geq f)\}.$$

Από μελέτες έχει προκύψει ότι η στατιστική συνάρτηση ελέγχου F είναι ευαίσθητη στην υπόθεση της κανονικότητας των πληθυσμών. Όταν όμως το μέγεθος των δύο δειγμάτων είναι ίδιο τότε ο έλεγχος είναι λιγότερο ευαίσθητος σε αποκλίσεις από την υπόθεση της κανονικότητας των κατανομών των δύο πληθυσμών. Ωστόσο μια καλή τακτική σε περιπτώσεις που αμφιβάλλουμε για την κανονικότητα των δεδομένων είναι να προτιμήσουμε τον έλεγχο του Levene για την ισότητα των διακυμάνσεων.

4.2.4 Έλεγχος για τη διαφορά των μέσων τιμών δύο πληθυσμών (t-test)

Έστω τα ανεξάρτητα δείγματα X_1, X_2, \dots, X_n και Y_1, Y_2, \dots, Y_n από κανονικούς πληθυσμούς $N(\mu_1, \sigma_1^2)$ και $N(\mu_2, \sigma_2^2)$ αντίστοιχα, με άγνωστες διακυμάνσεις. Για τον έλεγχο:

$$H_0 : \mu_1 - \mu_2 = \delta_0 \text{ με εναλλακτική την } H_1 : \mu_1 - \mu_2 \neq \delta_0$$

έχουμε τις ακόλουθες περιπτώσεις:

1. Για $\sigma_1^2 = \sigma_2^2$ (άγνωστα) η στατιστική συνάρτηση ελέγχου είναι η:

$$T = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \left(\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\right)}} = \frac{\bar{X} - \bar{Y} - \delta_0}{S_p \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

η οποία ακολουθεί την $t_{n_1+n_2-2}$ όταν η H_0 είναι αληθής. Αν συμβολίσουμε με t την παρατηρούμενη τιμή της T , η κρίσιμη περιοχή του ελέγχου σε επίπεδο σημαντικότητας α είναι

$$\{t < -t_{n_1+n_2-2; \alpha/2}\} \cup \{t > t_{n_1+n_2-2; \alpha/2}\}.$$

Η p -value του ελέγχου είναι ίση με

$$2\{P(T \geq |t| \parallel H_0) = 2(1 - P(T \leq |t| \parallel H_0))\}.$$

2. Για $\sigma_1^2 \neq \sigma_2^2$ (άγνωστα) η στατιστική συνάρτηση ελέγχου είναι η:

$$T = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

η οποία ακολουθεί την κατανομή t_v , όταν η H_0 είναι αληθής, όπου:

$$v = \frac{\left(\frac{s_1^2/n_1 + s_2^2/n_2}{(s_1^2/n_1)^2 + (s_2^2/n_2)^2}\right)^{-2}}{\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}}.$$

Η κρίσιμη περιοχή του ελέγχου σε επίπεδο σημαντικότητας α είναι:

$$\{t < -t_{v; \alpha/2}\} \cup \{t > t_{v; \alpha/2}\}.$$

Η p -value του ελέγχου είναι ίση με:

$$2\{P(T \geq |t| \parallel H_0) = 2(1 - P(T \leq |t| \parallel H_0))\}.$$

4.2.4 Έλεγχος ανεξαρτησίας χ^2 του Pearson

Ο έλεγχος αυτός χρησιμοποιείται γενικά για να εκτιμήσουμε αν δύο ή περισσότερα δείγματα τα οποία αποτελούνται από δεδομένα συχνοτήτων διαφέρουν σημαντικά μεταξύ τους. Με άλλα λόγια, χρησιμοποιείται κυρίως για την ανάλυση πινάκων διασταύρωσης ή συνάφειας με βάση δύο ποιοτικές μεταβλητές.

Σε έναν πίνακα συνάφειας με r γραμμές και c στήλες, συμβολίζουμε με O_{ij} τη συχνότητα εμφάνισης του κελιού (i, j) , δηλαδή αυτού που δημιουργείται από τη γραμμή i και τη στήλη j . Αυτό που μας ενδιαφέρει είναι να διερευνήσουμε εάν η μεταβλητή «γραμμή» και η μεταβλητή «στήλη» είναι ανεξάρτητες μεταξύ τους. Για τον έλεγχο της υπόθεσης αυτής, συγκρίνουμε τις παρατηρούμενες συχνότητες (O_{ij}) κάθε κελιού με τις αναμενόμενες συχνότητες (E_{ij}) που υπολογίζονται υπό την μηδενική υπόθεση. Πιο συγκεκριμένα, η αναμενόμενη συχνότητα κάθε κελιού εφόσον ισχύει η μηδενική υπόθεση δίνεται από τη σχέση:

$$E_{ij} = \frac{O_{i.} * O_{.j}}{n}$$

όπου $O_{i.}$ και $O_{.j}$ τα περιθώρια αθροίσματα των γραμμών και των στηλών, αντίστοιχα, και n το μέγεθος του δείγματος.

Η στατιστική συνάρτηση που χρησιμοποιείται για τον έλεγχο της μηδενικής υπόθεσης είναι η:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

η οποία, υπό την H_0 , ακολουθεί την χ^2 κατανομή με $(r-1)(c-1)$ βαθμούς ελευθερίας. Απορρίπτουμε δηλαδή την μηδενική υπόθεση και θεωρούμε ότι οι δύο μεταβλητές δεν είναι ανεξάρτητες μεταξύ τους εάν η τιμή της στατιστικής συνάρτησης είναι μεγαλύτερη ή ίση από την τιμή $\chi_{(r-1)(c-1),\alpha}^2$, δηλαδή το α-ποσοστημόριο της κατανομής χ^2 με $(r-1)(c-1)$ βαθμούς ελευθερίας.

Στην περίπτωση που το μέγεθος του δείγματος είναι μικρό, αντί του ελέγχου ανεξαρτησίας χ^2 του Pearson, χρησιμοποιείται ο ακριβής έλεγχος του Fisher (*Fisher's exact test*). Ο έλεγχος αυτός χρησιμοποιεί την κατανομή των συχνοτήτων O_{ij} δοθέντων των περιθωρίων αθροισμάτων $O_{i.}, O_{.j}$ και συγκρίνει τον πίνακα του δείγματος μόνο με εκείνους τους πίνακες που έχουν τα ίδια περιθώρια αθροίσματα, σε αντίθεση με τον έλεγχο χ^2 που τον συγκρίνει με όλους τους δυνατούς πίνακες (Σαχλάς, Μπερσίμης 2013).

4.2.5 Διαστήματα εμπιστοσύνης για τον λόγο σχετικών πιθανοτήτων

Μια σημαντική ποσότητα στην ανάλυση πινάκων συνάφειας είναι ο λόγος συμπληρωματικών πιθανοτήτων (*odd*). Πρόκειται για το πηλίκο της πιθανότητας ενός ενδεχομένου προς την πιθανότητα του συμπληρωματικού ενδεχομένου. Έχουμε δηλαδή:

$$odd = \frac{P(event)}{1 - P(event)}.$$

Ο λόγος συμπληρωματικών πιθανοτήτων παίρνει τιμές μεγαλύτερες ή ίσες του 0 και εκφράζει το πόσο πιθανό είναι να συμβεί ένα ενδεχόμενο σε σχέση με το να μην συμβεί.

Το σημαντικότερο μέτρο συνάφειας είναι ο λόγος σχετικών πιθανοτήτων (*odds ratio, OR*) επειδή προκύπτει ως ο λόγος του *odds* αυτών που έχουν εκτεθεί σε κάποιο παράγοντα προς το *odds* αυτών που δεν έχουν εκτεθεί στον παράγοντα. Εκφράζει το πόσο πιθανό είναι κάποιος που έχει εκτεθεί σε κάποιο παράγοντα να εμφανίσει την ασθένεια (γεγονός), σε σύγκριση με κάποιον που δεν έχει εκτεθεί στον παράγοντα (Σαχλάς, Μπερσίμης 2013).

Ο δειγματικός λόγος σχετικών πιθανοτήτων σε έναν 2x2 πίνακα συνάφειας προκύπτει σύμφωνα με τον τύπο:

$$OR = \hat{\theta} = \frac{n_{11}/n_{12}}{n_{21}/n_{22}} = \frac{n_{11} \cdot n_{22}}{n_{21} \cdot n_{12}}$$

ο οποίος είναι και ο εκτιμητής μεγίστης πιθανοφάνειας (ε.μ.π.) του πραγματικού λόγου σχετικών πιθανοτήτων.

Επειδή η δειγματική κατανομή του εκτιμητή $\hat{\theta}$ παρουσιάζει μεγάλη ασυμμετρία όταν το δείγμα είναι μικρό, χρησιμοποιούμε λογαριθμικό μετασχηματισμό. Η δειγματική κατανομή του (φυσικού) λογάριθμου του $\hat{\theta}$ πλησιάζει την κανονική κατανομή για μεγάλα δείγματα. Το τυπικό σφάλμα του εκτιμητή $\ln \hat{\theta}$ είναι ασυμπτωτικά ίσο με:

$$ASE(\ln \hat{\theta}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}.$$

Οπότε, για μεγάλα δείγματα, ένα $100(1-a)\%$ διάστημα εμπιστοσύνης για την παράμετρο $\ln \hat{\theta}$ είναι το:

$$\ln(\hat{\theta}) \pm z_{\alpha/2} ASE(\ln(\hat{\theta})).$$

Αντιλογαριθμίζοντας, προκύπτει ένα $100(1-\alpha)\%$ διάστημα εμπιστοσύνης για το $\hat{\theta}$ (Πολίτης 2014).

4.3 Στατιστική ανάλυση για τη σύγκριση ασθενών και μαρτύρων

4.3.1 Στατιστική ανάλυση μεταξύ ασθενών και μαρτύρων στο σύνολο της βάσης

Συγκρίνοντας το σύνολο των ασθενών με το σύνολο των μαρτύρων κάναμε ελέγχους ανεξαρτησίας για την εμφάνιση της νόσου σε σχέση με κάθε παράγοντα που έχει διακριτή φύση. Σε κάθε περίπτωση, διεξήγαμε τον έλεγχο χ^2 του *Pearson* για έλεγχο της υπόθεσης H_0 : Η εμφάνιση της νόσου Πάρκινσον είναι ανεξάρτητη του παράγοντα, έναντι της εναλλακτικής H_1 : Η εμφάνιση της νόσου Πάρκινσον εξαρτάται από τον παράγοντα. Στην περίπτωση όπου ένα ή περισσότερα κελιά είχαν αναμενόμενη συχνότητα μικρότερη του 5, αντί για τον έλεγχο χ^2 του *Pearson* χρησιμοποιήθηκε ο ακριβής έλεγχος του *Fisher* καθώς είναι περισσότερο ανθεκτικός.

Τα αποτελέσματα που λάβαμε ήταν τα εξής:

Πίνακας 4.1: Σχέση εμφάνισης ΝΠ και παράγοντα

Παράγοντας	Στατιστικό	P-value	RR	OR		
				Value	95% Conf. Interval	
					Lower	Upper
Φύλο	14.854	<0.001	1.295	1.976	1.395	2.800
Συγγένεια εξ αίματος των γονέων	0.695	0.371	1.294	2.470	0.274	22.251
Οικογενειακό ιστορικό ΝΠ	49.825	<0.001	1.674	58.306	8.043	422.674
Κάπνισμα: Κατά τις τελευταίες 30 ημέρες 100 τσιγάρα στη ζωή του	21.863	<0.001	0.593	0.330	0.204	0.533
	6.070	0.014	0.823	0.619	0.422	0.907
Κατανάλωση καφέ: Τώρα Στο παρελθόν	12.178	<0.001	0.778	0.497	0.334	0.738
	0.244	0.621	0.947	0.879	0.526	1.468
Έκθεση σε ζιζανιοκτόνα ή/και φυτοφάρμακα	1.233	0.267	1.097	1.295	0.820	2.047

Η εμφάνιση της νόσου δεν φαίνεται να επηρεάζεται από την εξ αίματος συγγένεια των γονέων, την κατανάλωση καφέ στο παρελθόν, αλλά ούτε και από την έκθεση σε ζιζανιοκτόνα ή/και φυτοφάρμακα, καθώς τα αντίστοιχα διαστήματα εμπιστοσύνης για τον λόγο σχετικών πιθανοτήτων περιέχουν τον αριθμό 1, οπότε δεν απορρίπτεται η υπόθεση περί ανεξαρτησίας.

Οι παράγοντες φύλο και οικογενειακό ιστορικό νόσου Πάρκινσον θεωρούνται παράγοντες κινδύνου για την εμφάνιση της ασθένειας διότι έχουν σχετικό κίνδυνο μεγαλύτερο από την μονάδα ($RR > 1$), ενώ το κάπνισμα, τόσο σε πρόσφατο όσο και παρελθοντικό χρονικό διάστημα, καθώς και η κατανάλωση καφέ σε πρόσφατο χρονικό διάστημα θεωρούνται προστατευτικοί παράγοντες για την εμφάνιση της νόσου ($RR < 1$). Επιπλέον, εξάγουμε τα παρακάτω συμπεράσματα:

- Η πιθανότητα εμφάνισης της νόσου δοθέντος ότι ένα άτομο είναι άνδρας είναι σχεδόν 30% μεγαλύτερη από την αντίστοιχη πιθανότητα όταν το άτομο είναι γυναίκα.
- Η πιθανότητα εμφάνισης της νόσου δοθέντος ότι ένας συγγενής πρώτου ή δεύτερου βαθμού έχει εμφανίσει νόσο του Πάρκινσον είναι περίπου 70%

μεγαλύτερη από την αντίστοιχη πιθανότητα όταν το άτομο δεν έχει οικογενειακό ιστορικό για τη νόσο.

- Η πιθανότητα εμφάνισης της νόσου δοθέντος ότι ένα άτομο έχει καπνίσει κατά τις τελευταίες 30 ημέρες είναι 41% μικρότερη από την αντίστοιχη πιθανότητα όταν το άτομο δεν έχει καπνίσει κατά τις τελευταίες 30 ημέρες. Αντίστοιχα, η πιθανότητα εμφάνισης της νόσου δοθέντος ότι ένα άτομο έχει καπνίσει παραπάνω από 100 τσιγάρα στη ζωή του είναι 18% μικρότερη από την αντίστοιχη πιθανότητα όταν το άτομο δεν έχει καπνίσει παραπάνω από 100 τσιγάρα στη ζωή του.
- Η πιθανότητα εμφάνισης της νόσου δοθέντος ότι ένα άτομο καταναλώνει σήμερα καφέ είναι σχεδόν 22% μικρότερη από την αντίστοιχη πιθανότητα όταν το άτομο δεν καταναλώνει σήμερα καφέ.

4.3.2 Στατιστική ανάλυση μεταξύ ασθενών και μαρτύρων ανά νοσοκομείο εξέτασης

Επεκτείνοντας την ανάλυσή μας, προχωρούμε στη διεξαγωγή των παραπάνω ελέγχων ξεχωριστά μεταξύ των εθελοντών που εξετάστηκαν στην Αθήνα και στη Σύρο. Τα αποτελέσματα που λάβαμε ήταν τα παρακάτω:

Πίνακας 4.2: Σχέση εμφάνισης ΝΠ και παράγοντα ανά νοσοκομείο εξέτασης

Παράγοντας	Νοσοκομείο εξέτασης			
	Αττικό		Σύρου	
	Στατιστικό	P-value	Στατιστικό	P-value
Φύλο	18.209	<0.001	0.556	0.456
Συγγένεια εξ' αίματος των γονέων	1.279	0.346	0.001	0.742
Οικογενειακό ιστορικό ΝΠ	19.436	<0.001	30.291	<0.001
Κάπνισμα:				
Κατά τις τελευταίες 30 ημέρες	19.302	<0.001	9.382	0.002
100 τσιγάρα στη ζωή του	2.908	0.088	4.713	0.030
Κατανάλωση καφέ:				
Τώρα	5.842	0.016	11.762	0.001
Στο παρελθόν	0.473	0.492	1.071	0.301
Έκθεση σε ζιζανιοκτόνα ή/και φυτοφάρμακα	3.629	0.057	0.401	0.527

Τα παραπάνω αποτελέσματα επαληθεύονται και στα επιμέρους δείγματα, με μόνη διαφορά ότι η μηδενική υπόθεση δεν απορρίπτεται για τον παράγοντα φύλο στο νοσοκομείο Σύρου, οπότε, άνδρες και γυναίκες που εξετάζονται σε αυτό νοσούν με παρόμοια πιθανότητα.

Επιπλέον, συγκρίνουμε και τις συνεχείς μεταβλητές που αφορούν τις δύο υποομάδες με στόχο να ανιχνεύσουμε στατιστικώς σημαντικές διαφορές ανάμεσα στους ασθενείς που εξετάστηκαν στο Αττικό νοσοκομείο και στους ασθενείς που εξετάστηκαν στο νοσοκομείο Σύρου. Η μεθοδολογία που ακολουθήθηκε ήταν η εξής:

- ✓ Αρχικά έγινε ο έλεγχος κανονικότητας *Kolmogorov–Smirnov* των συνεχών μεταβλητών για κάθε ομάδα ξεχωριστά.
- ✓ Εάν και τα δύο δείγματα ακολουθούσαν την κανονική κατανομή προχωρούσαμε στον έλεγχο *F* για ισότητα διακυμάνσεων. Εάν τα δείγματα ήταν και ομοσκεδαστικά, προχωρούσαμε σε *t-test* για έλεγχο ισότητας των μέσων.
- ✓ Εάν τουλάχιστον ένα εκ των δύο δειγμάτων δεν ακολουθούσε την κανονική κατανομή, τότε προχωρούσαμε στον μη παραμετρικό έλεγχο ισότητας κατανομών *Mann–Whitney* και στη συνέχεια στον έλεγχο *Levene* για ισότητα διακυμάνσεων.

Παρουσιάζουμε τα αποτελέσματα συνοπτικά:

Πίνακας 4.3: Κανονικότητα μεταβλητών ανά νοσοκομείο εξέτασης

Μεταβλητή	Νοσοκομείο εξέτασης			
	Αττικό		Σύρου	
	Στατιστικό	P-value	Στατιστικό	P-value
Ηλικία έναρξης ΝΠ	0.1075	<0.001	0.1562	<0.001
Χρόνια καπνίσματος	0.1055	0.0248	0.0692	0.9256
Ηλικία παύσης καπνίσματος	0.1376	0.0137	0.0895	0.7813
Χρόνια κατανάλωσης καφέ	0.1791	<0.001	0.1149	0.0506
Ηλικία παύσης κατανάλωσης καφέ	0.1294	0.4401	0.1536	0.4449
Χρόνια έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα	0.1438	0.0811	0.1447	0.3687
UPDRS I-III	0.1388	<0.001	0.107	0.0274

Οι μεταβλητές «Ηλικία παύσης κατανάλωσης καφέ» και «Χρόνια έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα» ακολουθούν την κανονική κατανομή σε επίπεδο σημαντικότητας 5% καθώς για αυτές παρατηρήθηκαν p-values με τιμές 0.4401 και 0.0811, αντίστοιχα, στο δείγμα του Αττικού νοσοκομείου, ενώ στο νοσοκομείο Σύρου οι αντίστοιχες τιμές ήταν 0.4449 και 0.3687. Για τις μεταβλητές αυτές θα προχωρήσουμε σε έλεγχο ομοσκεδαστικότητας και στη συνέχεια στους κατάλληλους ελέγχους ισότητας μέσων, ανάλογα με τα αποτελέσματα που θα λάβουμε.

Πίνακας 4.4: Ομοσκεδαστικότητα και ισότητα μέσων των κανονικών μεταβλητών

Μεταβλητή	Ομοσκεδαστικότητα		Ισότητα μέσων	
	Στατιστικό	P-value	Στατιστικό	P-value
Ηλικία παύσης κατανάλωσης καφέ	1.2321	0.6998	-1.1268	0.2675
Χρόνια έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα	0.9768	0.9244	-0.8466	0.4013

Και για τις δύο παραπάνω μεταβλητές ισχύει η ισότητα των διακυμάνσεων μεταξύ των δειγμάτων Αττικού και νοσοκομείου Σύρου, οπότε εφαρμόζοντας το t -test για έλεγχο της ισότητας των μέσων συμπεραίνουμε πως ισχύει η ισότητα. Συγκεκριμένα, τόσο στο Αττικό νοσοκομείο, όσο και στο νοσοκομείο Σύρου, οι ασθενείς που εξετάζονται σταματούν να καταναλώνουν καφέ κατά μέσο όρο στην ηλικία των 63.6 ετών, ενώ έχουν εκτεθεί σε ζιζανιοκτόνα ή/και φυτοφάρμακα γύρω στα 18.5 έτη.

Για τις λοιπές μεταβλητές στις οποίες τουλάχιστον ένα εκ των δύο δειγμάτων κρίθηκε ότι δεν ακολουθεί την κανονική κατανομή, προχωρήσαμε σε έλεγχο ισότητας των κατανομών τους, έχοντας ως μέτρο ισότητας την διάμεσο των κατανομών. Στη συνέχεια, ελέγξαμε και την υπόθεση ισότητας των διακυμάνσεών τους. Τα αποτελέσματα που λάβαμε παρουσιάζονται συνοπτικά ως εξής:

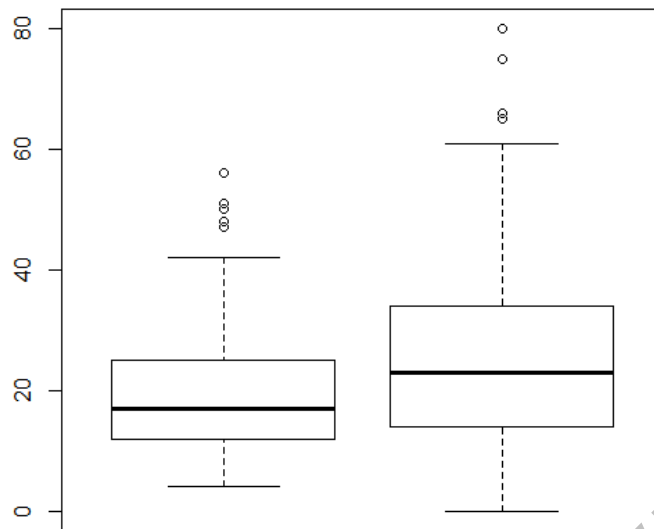
Πίνακας 4.5: Ισότητα κατανομών και ομοσκεδαστικότητα για τις μη κανονικές μεταβλητές

Μεταβλητή	Ισότητα κατανομών		Ομοσκεδαστικότητα	
	Στατιστικό	P-value	Στατιστικό	P-value
Ηλικία έναρξης ΝΠ	7675.5	<0.001	12.356	<0.001
Χρόνια καπνίσματος	1130	0.0262	0.0876	0.7678
Ηλικία παύσης καπνίσματος	536.5	0.0144	0.5042	0.4797
Χρόνια κατανάλωσης καφέ	4355	0.9625	8.5182	0.0039
UPDRS I-III	4871	0.0049	12.1253	0.0006

Τόσο στο Αττικό Νοσοκομείο, όσο και στο Νοσοκομείο Σύρου, οι ασθενείς που εξετάζονται καταναλώνουν καφέ για περίπου 38 έτη.

Αντίθετα, στατιστικώς σημαντικές διαφορές εντοπίστηκαν στα κάτωθι:

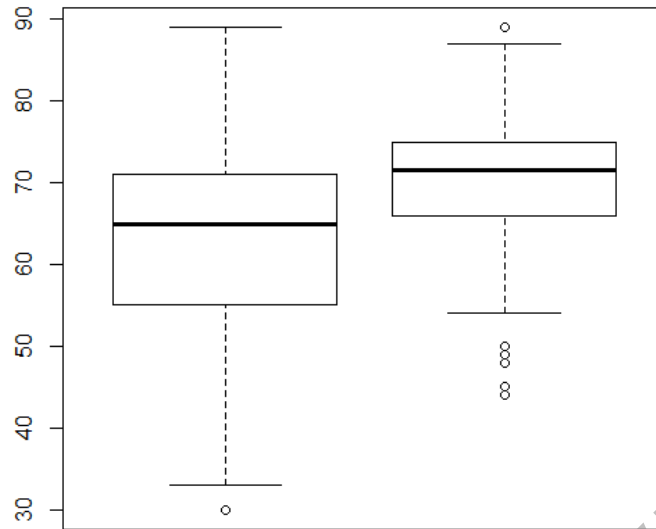
- UPDRS I-III



Διάγραμμα 4.1: Θηκόγραμμα για την απεικόνιση της αξιολόγησης ασθενών σύμφωνα με την κλίμακα UPDRS I έως και III που εξετάστηκαν α)στο Αττικό νοσοκομείο, β)στο νοσοκομείο Σύρου

Οι ασθενείς του Αττικού έχουν διάμεση τιμή αξιολόγησης 17 «βαθμούς» της κλίμακας UPDRS I-III ενώ αυτοί της Σύρου 23, με p-value για τον αντίστοιχο έλεγχο ισότητας διαμέσων ίσο με 0.0049. Ακόμα, η κατάσταση των πρώτων φαίνεται να είναι καλύτερη από αυτή των δεύτερων, καθώς το εύρος των τιμών αξιολόγησης είναι από 4 έως 56 «βαθμούς» για αυτούς, ενώ για τους ασθενείς που εξετάστηκαν στο νοσοκομείο Σύρου είναι 0 έως 80 «βαθμούς». Επιπλέον, το 75% των ασθενών του Αττικού νοσοκομείου αξιολογούνται με μέγιστη βαθμολογία τους 25 «βαθμούς» της κλίμακας UPDRS I-III, ενώ το αντίστοιχο ποσοστό για το νοσοκομείο Σύρου είναι οι 34 «βαθμοί». Οι μεγαλύτερες αξιολογήσεις υποδεικνύουν πως οι ασθενείς του νοσοκομείου Σύρου πηγαίνουν να εξεταστούν όταν τα συμπτώματα της νόσου είναι περισσότερο εμφανή.

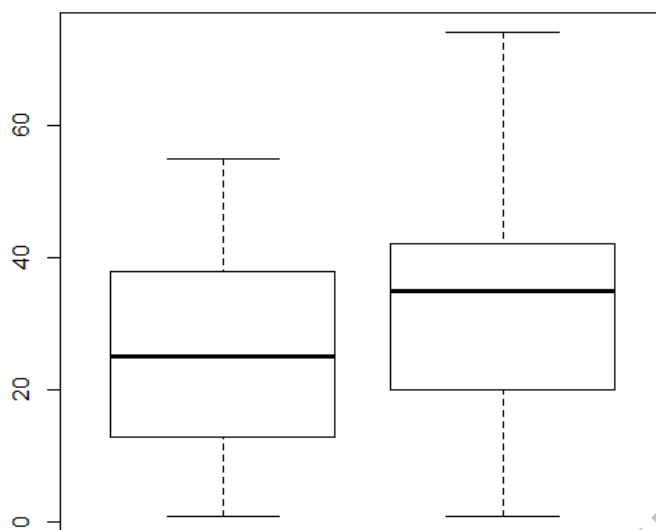
- Ηλικία έναρξης της νόσου



Διάγραμμα 4.2: Θηκόγραμμα για την απεικόνιση της ηλικίας έναρξης της νόσου των ασθενών που εξετάστηκαν α)στο Αττικό νοσοκομείο, β)στο νοσοκομείο Σύρου

Η διάμεση ηλικία έναρξης της νόσου στους ασθενείς του Αττικού Νοσοκομείου είναι 65 έτη, ενώ σε αυτούς της Σύρου τα 71.5 έτη, με p-value για τον αντίστοιχο έλεγχο μικρότερο του 10^{-3} . Οπότε, υπάρχει πάρα πολύ ισχυρή ένδειξη ότι η ηλικία έναρξης της νόσου είναι διαφορετική για τους ασθενείς στα δύο νοσοκομεία. Ακόμα, στο Αττικό εμφανίζονται συχνότερα περιστατικά νεανικής μορφής παρκινσονισμού, ενώ στο δείγμα του νοσοκομείου Σύρου, τέτοιες περιπτώσεις θεωρούνται έκτροπες παρατηρήσεις.

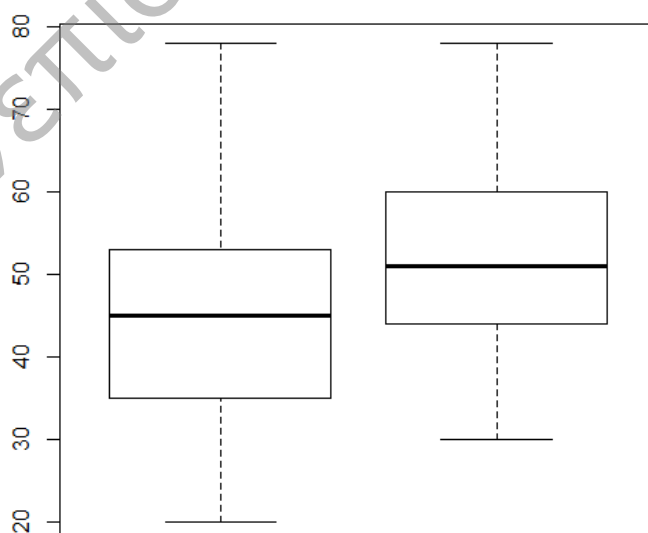
- Χρόνια καπνίσματος



Διάγραμμα 4.3: Θηκόγραμμα για την απεικόνιση των ετών καπνίσματος των ασθενών που εξετάστηκαν α)στο Αττικό νοσοκομείο, β)στο νοσοκομείο Σύρου

Οι ασθενείς του Αττικού έχουν διάμεσο χρόνο καπνίσματος ίσο με 25 έτη, ενώ οι ασθενείς του νοσοκομείου Σύρου 35 έτη, με p-value για τον αντίστοιχο έλεγχο ισότητας διαμέσων ίσο με 0.0262. Ακόμα, διαφέρει και το εύρος του χρονικού διαστήματος καπνίσματος στα δύο δείγματα. Στο πρώτο είναι 1 έως 55 έτη, ενώ στο δεύτερο 1 έως 74 έτη.

- Ηλικία παύσης καπνίσματος



Διάγραμμα 4.4: Θηκόγραμμα για την απεικόνιση της ηλικίας παύσης καπνίσματος των ασθενών που εξετάστηκαν α)στο Αττικό νοσοκομείο, β)στο νοσοκομείο Σύρου

Οι ασθενείς του Αττικού έχουν διάμεση ηλικία παύσης καπνίσματος τα 45 έτη, ενώ οι ασθενείς του νοσοκομείου Σύρου σχεδόν μία εξαετία αργότερα, με p-value για τον αντίστοιχο έλεγχο ισότητας διαμέσων ίσο με 0.0144.

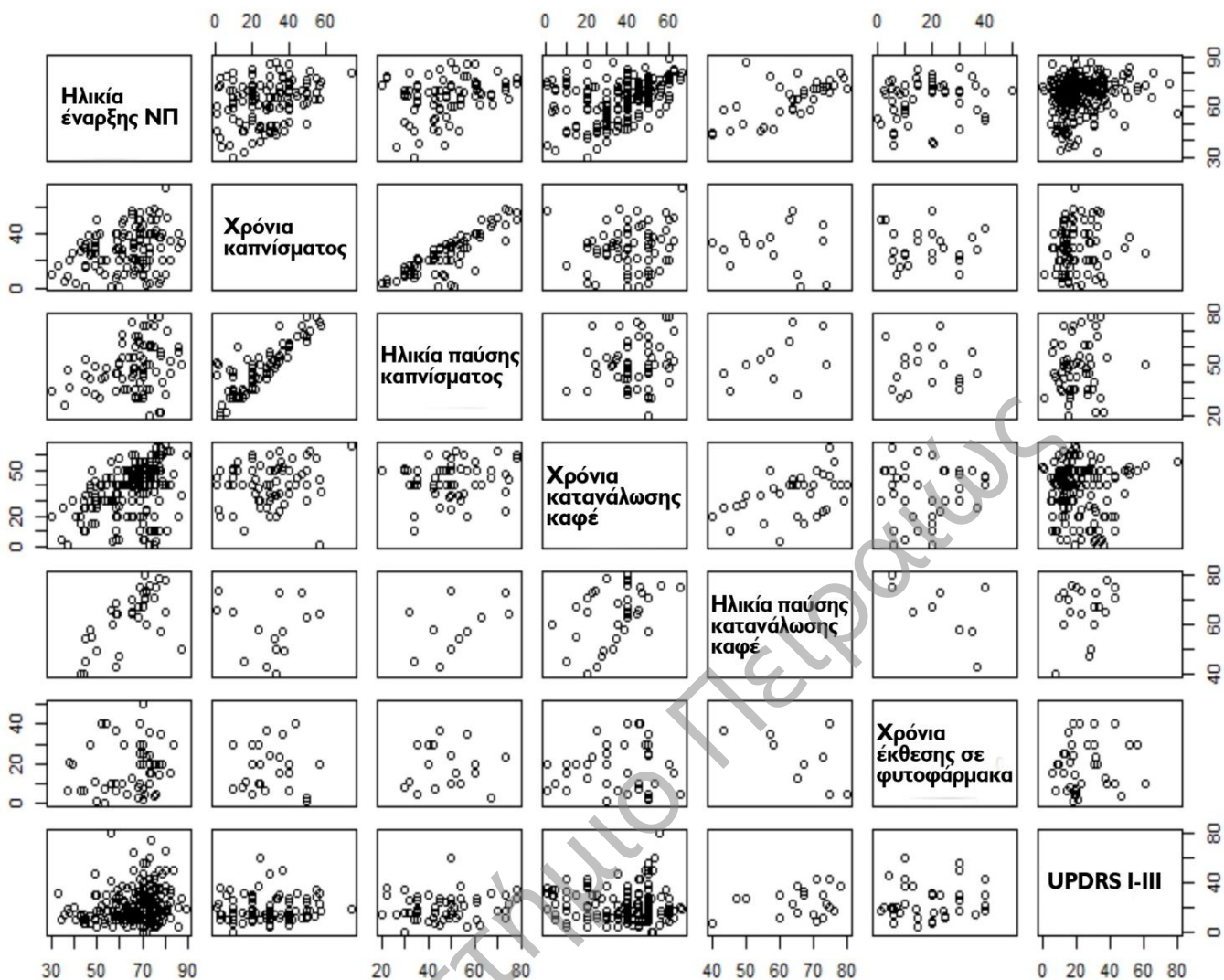
Επιπλέον, παρουσιάζουμε σε ένα συνοπτικό πίνακα τους συντελεστές συσχέτισης των συνεχών μεταβλητών.

Πίνακας 4.6: Συντελεστές συσχέτισης των συνεχών μεταβλητών

	Ηλικία έναρξης ΝΠ	Χρόνια καπνίσματος	Ηλικία παύσης καπνίσματος	Χρόνια κατανάλωσης καφέ	Ηλικία παύσης κατανάλωσης καφέ	Χρόνια έκθεσης σε φυτοφάρμακα	U PDRS I-III
Ηλικία έναρξης ΝΠ	1						
Χρόνια καπνίσματος	0.166820	1					
Ηλικία παύσης καπνίσματος	0.173570	0.688618	1				
Χρόνια κατανάλωσης καφέ	0.314165	0.092459	0.064810	1			
Ηλικία παύσης κατανάλωσης καφέ	0.515736	-0.048076	0.341095	0.344098	1		
Χρόνια έκθεσης σε φυτοφάρμακα	0.100277	-0.003780	0.033784	0.031988	-0.457142	1	
UPDRS I-III	0.156299	0.032966	0.070115	-0.031790	0.172975	0.063477	1

Η υψηλότερη συσχέτιση αφορά τα χρόνια που ένα άτομο καπνίζει και την ηλικία παύσης καπνίσματος, πράγμα λογικό. Ακόμα, η ηλικία έναρξης της νόσου σχετίζεται θετικά και ισχυρά με την ηλικία παύσης κατανάλωσης καφέ, γεγονός που επιβεβαιώνει τον ισχυρισμό ότι η κατανάλωση καφέ σε πρόσφατο χρόνο είναι προστατευτικός παράγοντας για την εμφάνιση της νόσου. Τέλος, η κλίμακα UPDRS I-III δεν φαίνεται να παρουσιάζει ισχυρή γραμμική συσχέτιση με κάποια από τις υπόλοιπες μεταβλητές.

Τα αντίστοιχα διαγράμματα διασποράς μεταξύ των μεταβλητών είναι τα παρακάτω:



Διάγραμμα 4.5: Διαγράμματα διασποράς μεταξύ των συνεχών μεταβλητών

Με εξαίρεση το διάγραμμα που αφορά τις μεταβλητές καπνίσματος όπου η γραμμική εξάρτηση είναι ξεκάθαρη, αλλά και αυτό που αφορά τις ηλικίες παύσης καπνίσματος και κατανάλωσης καφέ, τα υπόλοιπα διαγράμματα έχουν τη μορφή κυκλικού νέφους, πράγμα που σημαίνει πως δεν υπάρχει εμφανής είτε γραμμική συσχέτιση, είτε μεγάλη διακύμανση των τιμών όταν εξετάζονται ανά δύο οι μεταβλητές.

4.4 Στατιστική ανάλυση για τους ασθενείς με νεανικής μορφής Παρκινσονισμό

4.4.1 Στατιστική ανάλυση μεταξύ των ασθενών με ηλικία έναρξης της νόσου μικρότερη ή/και ίση των 50 ετών και των ασθενών με ηλικία έναρξης μεγαλύτερη των 50 ετών

Στο σύνολο της βάσης υπάρχουν 343 ασθενείς, εκ των οποίων οι 50 (ποσοστό 14.58%) εκδήλωσαν την νόσο μέχρι τα 50 έτη. Θα διερευνήσουμε την ύπαρξη συσχέτισης μεταξύ της εμφάνισης της νόσου μέχρι τα 50 έτη με κάθε παράγοντα που έχει διακριτή φύση. Σε κάθε περίπτωση, διεξήγαμε το χ^2 τεστ του *Pearson* για έλεγχο της υπόθεσης H_0 : Η εμφάνιση της νόσου Πάρκινσον μέχρι τα 50 έτη είναι ανεξάρτητη του παράγοντα, έναντι της εναλλακτικής H_1 : Η εμφάνιση της νόσου Πάρκινσον μέχρι τα 50 έτη εξαρτάται από τον παράγοντα. Επιπλέον, όπως και πρωτότερα, στην περίπτωση όπου ένα ή περισσότερα κελιά είχαν αναμενόμενη συχνότητα μικρότερη του 5, αντί για το χ^2 τεστ του *Pearson* χρησιμοποιήθηκε ο ακριβής έλεγχος του *Fisher* καθώς είναι περισσότερο ανθεκτικός. Τα αποτελέσματα που λάβαμε ήταν τα εξής:

Πίνακας 4.7: Σχέση εμφάνισης νεανικής μορφής παρκινσονισμού και παράγοντα

Παράγοντας	Στατιστικό	P-value	RR	OR		
				Value	95% Conf.Interval	
					Lower	Upper
Φύλο	0.041	0.839	0.948	0.939	0.513	1.720
Συγγένεια εξ' αίματος των γονέων	0.324	0.478	1.689	1.918	0.196	18.819
Οικογενειακό ιστορικό ΝΠ	1.605	0.205	1.423	1.522	0.792	2.927
Κάπνισμα: Κατά τις τελευταίες 30 ημέρες 100 τσιγάρα στη ζωή του	18.631	<0.001	3.388	4.940	2.260	10.800
	7.604	0.006	2.331	2.688	1.307	5.529
Κατανάλωση καφέ: Τώρα Στο παρελθόν	1.010	0.315	1.350	1.415	0.717	2.793
	0.727	0.394	1.537	1.622	0.529	4.976
Έκθεση σε ζιζανιοκτόνα ή/και φυτοφάρμακα	0.138	0.710	1.124	1.148	0.554	2.380

Ο μόνος παράγοντας που φαίνεται να επηρεάζει στατιστικώς σημαντικά την εμφάνιση νεανικής μορφής παρκινσονισμού είναι το κάπνισμα τόσο κατά το πρόσφατο παρελθόν, όσο και διαχρονικότερα. Συγκεκριμένα, ένας ασθενής που καπνίζει κατά το τελευταίο χρονικό διάστημα, πριν την εξέταση, αυξάνει την πιθανότητα να νοσήσει από 2 έως 11 φορές, ενώ κάποιος που έχει καπνίσει πάνω από 100 τσιγάρα στη ζωή του αυξάνει την πιθανότητα αυτή από 1.3 έως 5.5 φορές.

Αντίστοιχα στοιχεία δεν βρέθηκαν στη διεθνή ή/και την εγχώρια βιβλιογραφία. Θα πρέπει, συνεπώς, να επισημάνουμε πως το δείγμα των ασθενών με νεανικό Παρκινσονισμό είναι αρκετά μικρό, καθώς αποτελείται από 50 άτομα. Συνεπώς, τα αποτελέσματα αυτά θα πρέπει να αντιμετωπίζονται με αμφιβολία, ενώ χρήζουν περαιτέρω διερεύνησης.

Επεκτείνοντας την ανάλυσή μας, προχωρούμε στη σύγκριση των δυο υποομάδων με στόχο να ανιχνεύσουμε στατιστικώς σημαντικές διαφορές μεταξύ των ασθενών που εμφάνισαν τη νόσο μέχρι και τα 50 έτη σε σχέση με αυτούς που εμφάνισαν Παρκινσονισμό σε μεγαλύτερη ηλικία. Η μεθοδολογία που ακολουθήθηκε είναι ίδια με προηγουμένως, ενώ τα αποτελέσματα που λάβαμε ήταν τα παρακάτω:

Πίνακας 4.8: Κανονικότητα μεταβλητών ανά ηλικιακή κατηγορία εμφάνισης νόσου του Πάρκινσον

Μεταβλητή	Ηλικία εμφάνισης ΝΠ			
	<=50 ετών		>50 ετών	
	Στατιστικό	P-value	Στατιστικό	P-value
Χρόνια καπνίσματος	0.1249	0.4003	0.1130	0.0048
Χρόνια κατανάλωσης καφέ	0.1527	0.0636	0.1893	<0.001
Χρόνια έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα	0.2468	0.1612	0.1340	0.0457
UPDRS I-III	0.2083	0.0084	0.1290	<0.001

Παρατηρούμε πως για κάθε μεταβλητή τουλάχιστον ένα εκ των δύο δειγμάτων κρίθηκε ότι δεν ακολουθεί την κανονική κατανομή. Οπότε, προχωρήσαμε σε έλεγχο ισότητας των κατανομών τους, έχοντας ως μέτρο ισότητας την διάμεσο των κατανομών. Στη συνέχεια, ελέγξαμε και την υπόθεση ισότητας των διακυμάνσεών τους. Τα αποτελέσματα που λάβαμε παρουσιάζονται συνοπτικά ως εξής:

Πίνακας 4.9: Ισότητα κατανομών και ομοσκεδαστικότητα για τις μη κανονικές μεταβλητές

Μεταβλητή	Ισότητα κατανομών		Ομοσκεδαστικότητα	
	Στατιστικό	P-value	Στατιστικό	P-value
Χρόνια καπνίσματος	907	0.0804	4.805	0.0304
Χρόνια κατανάλωσης καφέ	1090	<0.001	2.6474	0.1053
Χρόνια έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα	124	0.1902	1.1193	0.2952
UPDRS I-III	1726.5	0.0079	2.5104	0.1144

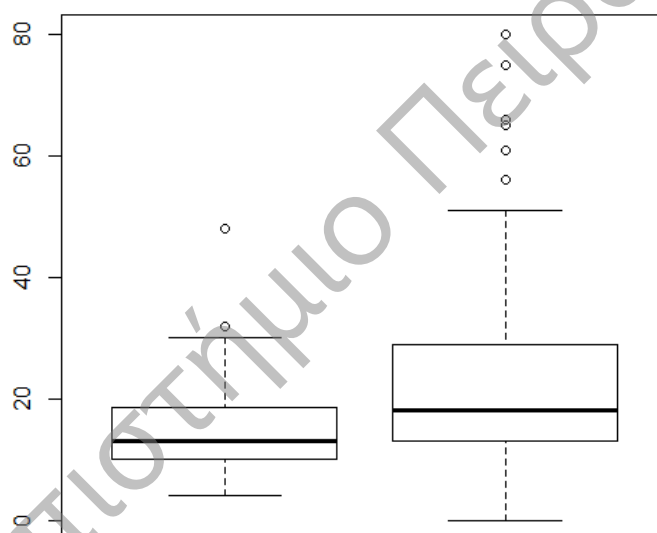
Τόσο για την μεταβλητή «Χρόνια καπνίσματος» όσο και για την μεταβλητή «Χρόνια έκθεσης σε ζιζανιοκτόνα ή/και φυτοφάρμακα» παρατηρήθηκαν p-values με τιμές μεγαλύτερες από 0.05. Συνεπώς, σε επίπεδο σημαντικότητας 5% αποφασίζουμε

πως οι κατανομές των δυο υποομάδων δεν διαφέρουν στατιστικώς σημαντικά. Συγκεκριμένα, ο διάμεσος χρόνος καπνίσματος για τις δυο ομάδες είναι 30 έτη, ενώ ο διάμεσος χρόνος έκθεσης σε φυτοφάρμακα είναι 25 έτη.

Το αποτέλεσμα αυτό για το κάπνισμα εξηγεί εν μέρει το μη συμβατό του προηγούμενου αποτελέσματος. Από ότι φαίνεται, στο δείγμα των ασθενών με νεανικό Παρκινσονισμό έτυχε να συμμετέχουν άτομα με πολυετή διάρκεια καπνίσματος με αποτέλεσμα να μην είναι ενδεικτική η συγκρισιμότητα τους. Σε κάθε περίπτωση, όπως αναφέρθηκε και νωρίτερα, τα αποτελέσματα αυτά θα πρέπει να αντιμετωπίζονται με αμφιβολία, ενώ χρήζουν περαιτέρω διερεύνησης.

Στατιστικώς σημαντικές διαφορές εντοπίστηκαν στα κάτωθι:

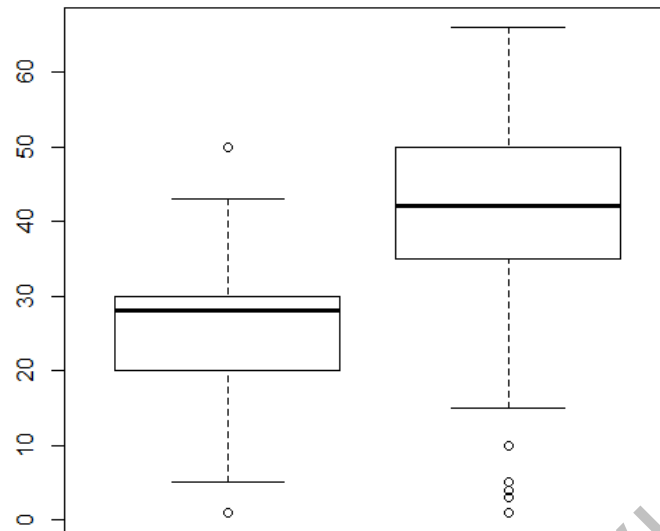
- UPDRS I-III



Διάγραμμα 4.6: Θηκόγραμμα για την απεικόνιση της αξιολόγησης ασθενών σύμφωνα με την κλίμακα UPDRS I έως και III με ηλικία έναρξης α)μέχρι 50 έτη, β)μεγαλύτερη από 50 έτη

Οι ασθενείς με νεανικής μορφής Παρκινσονισμό έχουν διάμεση τιμή αξιολόγησης 13 «βαθμούς» της κλίμακας UPDRS I-III ενώ οι ασθενείς με ηλικία έναρξης μεγαλύτερη των 50 ετών 18, με p-value για τον αντίστοιχο έλεγχο ισότητας διαμέσων ίσο με 0.0079.

- Χρόνια κατανάλωσης καφέ



Διάγραμμα 4.7: Θηκόγραμμα για την απεικόνιση των ετών κατανάλωσης καφέ των ασθενών με ηλικία έναρξης α)μέχρι 50 έτη, β)μεγαλύτερη από 50 έτη

Οι ασθενείς με νεανικής μορφής Παρκινσονισμό έχουν διάμεσο χρόνο κατανάλωσης καφέ 28 έτη, ενώ οι ασθενείς με ηλικία έναρξης μεγαλύτερη των 50 ετών έχουν διάμεσο χρόνο κατανάλωσης καφέ 42 έτη, με p-value για τον αντίστοιχο έλεγχο ισότητας διαμέσων μικρότερο από 0.001.

4.5 Συμπεράσματα

Η βασική αιτία η οποία καθιστά την νόσο Πάρκινσον ως ανίατη ασθένεια είναι το γεγονός ότι πρόκειται για μια πολυπαραγοντική ασθένεια που απορρέει ως αποτέλεσμα τόσο από την περιβαλλοντική έκθεση σε διάφορους παράγοντες όσο και στις γενετικές προδιαθέσεις που ένα άτομο μπορεί να έχει.

Από την μελέτη που διεξήγαμε, η εμφάνιση της νόσου Πάρκινσον δεν φαίνεται να επηρεάζεται από την εξ αίματος συγγένεια των γονέων, την κατανάλωση καφέ σε παρελθοντική από την εξέταση χρονική περίοδο, αλλά ούτε και από την έκθεση σε ζιζανιοκτόνα ή/και φυτοφάρμακα. Σχετικά με τον τελευταίο παράγοντα, υπάρχουν αντικρουόμενες μελέτες που χαρακτηρίζουν τον παράγοντα αυτό είτε σημαντικό για την εμφάνιση της νόσου (Galanaud JP et al., 2005) είτε υποστηρίζουν πως δεν επηρεάζει απαραίτητως την εμφάνισή της (Nuti A et al., 2004; Taylor CA et al., 1999).

Από τη άλλη, οι παράγοντες φύλο και οικογενειακό ιστορικό νόσου Πάρκινσον φαίνεται να επηρεάζουν και μάλιστα θεωρούνται παράγοντες κινδύνου για την

εμφάνιση της ασθένειας. Αντίθετα, το κάπνισμα, τόσο σε πρόσφατο όσο και παρελθοντικό χρονικό διάστημα από την ημέρα εξέτασης, καθώς και η κατανάλωση καφέ σε πρόσφατο χρονικό διάστημα παίζουν σπουδαίο ρόλο στην εμφάνιση της νόσου και θεωρούνται προστατευτικοί παράγοντες για αυτήν.

Τα παραπάνω αποτελέσματα επαληθεύθηκαν και στα επιμέρους δείγματα Αττικού νοσοκομείου και νοσοκομείου Σύρου, με μόνη διαφορά ότι άνδρες και γυναίκες που εξετάζονται στο δεύτερο νοσοούν εξίσου.

Τέλος, μόνο ο παράγοντας που κρίθηκε στατιστικά σημαντικός για την εμφάνιση νεανικής μορφής παρκινσονισμού είναι το κάπνισμα. Το δείγμα των ασθενών με νεανικό Παρκινσονισμό αποτελείται όμως μόνο από 50 άτομα και συνεπώς, τα αποτελέσματα χρήζουν περαιτέρω διερεύνησης.

Πανεπιστήμιο Πειραιώς

ΚΕΦΑΛΑΙΟ 5: ΑΝΑΛΥΣΗ ΚΑΤΑ ΣΥΣΤΑΔΕΣ

5.1 Εισαγωγή

Η ανάλυση κατά συστάδες (cluster analysis) είναι ένα βασικό εργαλείο για τμηματοποίηση δεδομένων και δημιουργία ομάδων. Είναι μια διερευνητική στατιστική τεχνική σχεδιασμένη για τον εντοπισμό εσωτερικών σχέσεων στα δεδομένα. Η ιδέα της ανάλυσης κατά συστάδες βασίζεται στην απλή λογική ότι: «τα δεδομένα μπορούν να διαχωριστούν σε ομάδες βάσει της απόστασης μεταξύ τους». Εκείνο που θα εξετάσουμε στις επόμενες παραγράφους είναι συστηματικές μεθόδους ομαδοποίησης οι οποίες έχουν κάποια μαθηματική/στατιστική βάση, καθώς και η εφαρμογή κάποιων από αυτές στο δείγμα των ασθενών μας προκειμένου να εξετάσουμε εάν υπάρχει κάποια ξεκάθαρη διαφοροποίηση μεταξύ τους.

5.2 Θεωρητικό υπόβαθρο

Σύμφωνα με τις πανεπιστημιακές σημειώσεις του κ. Κούτρα «Ανάλυση κατά συστάδες» (2013), υπάρχουν διάφορες προσεγγίσεις για το πώς μπορούμε να ομαδοποιήσουμε τα δεδομένα μας. Ένας εμπειρικός τρόπος δημιουργίας συστάδων είναι με χρήση κάποιας γραφικής παράστασης των πολυδιάστατων παρατηρήσεων (π.χ.: sun ray plots, Glyphs, Chernoff faces), και τοποθέτηση στην ίδια ομάδα των ατόμων που παρουσιάζουν παρόμοια εικόνα. Οι καμπύλες Andrews μπορούν επίσης να δώσουν ένα εναλλακτικό γραφικό τρόπο ομαδοποίησης.

Υπάρχουν επίσης, πιο συστηματικές μέθοδοι ομαδοποίησης οι οποίες έχουν ισχυρότερη μαθηματική/στατιστική βάση. Οι μέθοδοι ομαδοποίησης μπορούν, γενικά, να χωριστούν σε δύο διαφορετικές κατηγορίες ανάλογα με τον τρόπο που προχωρούν στη διαμόρφωση των ομάδων: στις ιεραρχικές και στις μη ιεραρχικές μεθόδους.

Στις ιεραρχικές μεθόδους, οι ομάδες σχηματίζονται σταδιακά είτε με συνένωση μικρότερων ομάδων σχηματίζοντας συνεχώς μεγαλύτερες ομάδες μέχρι να φτάσουμε να έχουμε όλα τα δεδομένα σε μία ομάδα (συσσωρευτικές μέθοδοι), είτε με διαίρεση ομάδων σε μικρότερες μέχρι να φτάσουμε σε μια κατάσταση όπου κάθε παρατήρηση να είναι από μόνη της μια ομάδα (διαιρετικές μέθοδοι).

Στις μη ιεραρχικές μεθόδους θεωρείται ότι ο αριθμός των ομάδων είναι γνωστός από πριν. Σε αυτές χρησιμοποιούμε έναν επαναληπτικό αλγόριθμο για να

τοποθετούμε τις παρατηρήσεις στις ομάδες ανάλογα με το ποια ομάδα είναι πιο κοντά στην εκάστοτε παρατήρηση.

Στο ερώτημα «ποιά μέθοδο θα πρέπει να χρησιμοποιήσω ώστε να πάρω καλύτερα αποτελέσματα» δεν υπάρχει ικανοποιητική γενική απάντηση. Καλό είναι όμως να έχουμε υπόψη μας πως οι ιεραρχικές μέθοδοι καλό είναι να αποφεύγονται να χρησιμοποιούνται για μεγάλο πλήθος δεδομένων αφού απαιτούν πολύ χρόνο, μνήμη και υπολογιστική ισχύ. Επίσης, υπάρχει η τάση να δημιουργούνται ομάδες με ανομοιογενές μέγεθος. Από την άλλη, οι μη ιεραρχικές μέθοδοι ενώ δουλεύουν ικανοποιητικά με μεγάλα δείγματα και δημιουργούν ομάδες παραπλήσιου μεγέθους, επηρεάζονται αρκετά από τις αρχικές τιμές που θα χρησιμοποιηθούν.

5.2.1 Μέτρα απόστασης και μέτρα ομοιότητας

Η ανάλυση κατά συστάδες είναι μια μέθοδος που σκοπό έχει να κατατάξει σε ομάδες τις υπάρχουσες παρατηρήσεις, χρησιμοποιώντας την πληροφορία που υπάρχει σε κάποιες μεταβλητές. Με άλλα λόγια η ανάλυση κατά συστάδες εξετάζει πόσο όμοιες είναι κάποιες παρατηρήσεις ως προς κάποιον αριθμό μεταβλητών με σκοπό να δημιουργήσει ομάδες από παρατηρήσεις που μοιάζουν μεταξύ τους (Καρλής 2005).

Δύο βασικές έννοιες για την ανάλυση κατά συστάδες είναι οι έννοιες της απόστασης και της ομοιότητας. Παρατηρήσεις που μοιάζουν πολύ μεταξύ τους θα πρέπει να δίνουν πολύ μικρή τιμή στην απόσταση που τις χαρακτηρίζει, ενώ παράλληλα πρέπει να δίνουν πολύ μεγάλη τιμή στο μέτρο ομοιότητας που χρησιμοποιείται. Μια επιτυχημένη ανάλυση θα πρέπει να καταλήξει σε ομάδες για τις οποίες οι παρατηρήσεις μέσα σε κάθε ομάδα να είναι όσο γίνεται πιο ομοιογενείς, αλλά παρατηρήσεις διαφορετικών ομάδων να διαφέρουν όσο γίνεται περισσότερο (Καρλής 2005).

Αν συμβολίσουμε με $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ το διάνυσμα των παρατηρήσεων (για τις p μεταβλητές) που αφορά το i άτομο ($i = 1, 2, \dots, n$). Παραδείγματα μετρικών που μπορούν να χρησιμοποιηθούν ως απόσταση μεταξύ δύο παρατηρήσεων $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ και $x_j = (x_{j1}, x_{j2}, \dots, x_{jp})$ είναι:

- Η μετρική Minkowski:
$$d_{ij} = \left[\sum_{r=1}^p |x_{ir} - x_{jr}|^m \right]^{\frac{1}{m}}.$$

➤ Για $m = 2$ παίρνουμε την Ευκλείδεια απόσταση:

$$d_{ij} = \sqrt{\sum_{r=1}^p (x_{ir} - x_{jr})^2} .$$

➤ Ενώ για $m = 1$ παίρνουμε την απόσταση Manhattan:

$$d_{ij} = \sum_{r=1}^p |x_{ir} - x_{jr}| .$$

▪ Η μετρική max ή Chebyshev: $d_{ij} = \max_{r=1,2,\dots,p} |x_{ir} - x_{jr}|$.

▪ Η μετρική Canberra: $d_{ij} = \sum_{r=1}^p \frac{|x_{ir} - x_{jr}|}{|x_{ir}| + |x_{jr}|}$.

Η απόσταση μπορεί να χρησιμοποιηθεί προκειμένου να κατασκευαστεί ένας πίνακας με n γραμμές και n στήλες, όπου κάθε στοιχείο του θα απεικονίζει την απόσταση μεταξύ του i στοιχείου (ή συστάδας) και του j στοιχείου (ή συστάδας) στην θέση (i, j) . Ο πίνακας αυτός καλείται πίνακας αποστάσεων και, προφανώς, θα έχει μηδενικά στοιχεία στη διαγώνιο του, και επιπλέον θα είναι συμμετρικός.

Μια εναλλακτική κατηγορία μέτρων που μπορούν να χρησιμοποιηθούν για να μας δείξουν αν δυο παρατηρήσεις είναι όμοιες ή ανόμοιες μεταξύ τους είναι τα λεγόμενα μέτρα ομοιότητας (similarity measures). Έχοντας εισαγάγει κατάλληλα τέτοια μέτρα, θα μπορούμε να τοποθετούμε ζεύγη παρατηρήσεων στην ίδια ή σε διαφορετική ομάδα ανάλογα με το αν η τιμή του μέτρου είναι μεγάλη ή μικρή, αντίστοιχα.

Αν υποθέσουμε ότι για κάθε ζεύγος παρατηρήσεων $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ και $x_j = (x_{j1}, x_{j2}, \dots, x_{jp})$ ορίζεται ένας πραγματικός αριθμός $s_{ij} = s(x_i, x_j)$ έτσι ώστε να ισχύουν οι επόμενες τρεις ιδιότητες:

1. $s_{ij} \geq 0$ για κάθε i, j και $i = j \Rightarrow s_{ij} = 1$

2. $s_{ij} \leq 1$

3. $s_{ij} = s_{ji}$ (συμμετρική ιδιότητα)

τότε θα λέμε ότι η συνάρτηση $s_{ij} = s(x_i, x_j)$ δίνει ένα μέτρο ομοιότητας.

Το πιο γνωστό μέτρο ομοιότητας για ποσοτικές παρατηρήσεις είναι ο (δειγματικός) συντελεστής συσχέτισης, ο οποίος δίνεται από τον τύπο:

$$s_{ij} \equiv \frac{\sum_{r=1}^p (x_{ir} - \bar{x}_{i\bullet})(x_{jr} - \bar{x}_{j\bullet})}{\left(\sum_{r=1}^p (x_{ir} - \bar{x}_{i\bullet})^2 \sum_{r=1}^p (x_{jr} - \bar{x}_{j\bullet})^2 \right)^{1/2}} ,$$

$$\text{όπου } \bar{x}_{i\bullet} = \frac{1}{p} \sum_{r=1}^p x_{ir} \text{ και } \bar{x}_{j\bullet} = \frac{1}{p} \sum_{r=1}^p x_{jr}.$$

Ένα αντίστοιχο μέτρο ομοιότητας μεταξύ δύο παρατηρήσεων i και j , για τα οποία τα χαρακτηριστικά που μας ενδιαφέρουν είναι κατηγορικά, είναι το

$$s_{ij} = \frac{u}{p}$$

(simple matching similarity) όπου u είναι ο αριθμός των συμφωνιών, δηλαδή ο αριθμός των μεταβλητών για τις οποίες οι παρατηρήσεις i και j συμβαίνει να βρίσκονται στην ίδια κατάσταση και p είναι το πλήθος των μεταβλητών.

Τέλος, στην περίπτωση που οι μεταβλητές που μελετάμε δεν είναι όλες του ίδιου τύπου (συνεχείς-κατηγορικές-δίτιμες), ο Gower πρότεινε η ομοιότητά τους να μετράται μέσω του συντελεστή

$$s_{ij} = \frac{\sum_{r=1}^p w_{ij}(r) s_{ij}(r)}{\sum_{r=1}^p w_{ij}(r)},$$

όπου:

α. αν η μεταβλητή r είναι συνεχής θέτουμε

$$s_{ij}(r) = 1 - \frac{|x_{ir} - x_{jr}|}{R_r}$$

με

$$R_r = \max_{i=1,2,\dots,p} x_{ir} - \min_{i=1,2,\dots,p} x_{ir}$$

το εύρος της r -οστής μεταβλητής.

β. αν η μεταβλητή r είναι διακριτή θέτουμε

$$s_{ij}(r) = \begin{cases} 1, & \text{για } x_{ir} = x_{jr} \\ 0, & \text{διαφορετικά} \end{cases}.$$

Τα βάρη $w_{ij}(r)$ παίρνουν την τιμή 1 ή 0, ανάλογα με το αν η σύγκριση στην r -οστή μεταβλητή έχει νόημα ή όχι.

5.3 Ιεραρχικές μέθοδοι ομαδοποίησης

5.3.1 Συσσωρευτικές μέθοδοι

Κατά την εκτέλεση συσσωρευτικών μεθόδων ακολουθείται η παρακάτω διαδικασία:

Ξεκινάμε αρχικά με N συστάδες, κάθε μία από τις οποίες περιέχει μόνο ένα στοιχείο, και κατασκευάζουμε έναν $N \times N$ πίνακα με αποστάσεις. Στη συνέχεια, βρίσκουμε στον πίνακα το ζεύγος των συστάδων με την μικρότερη απόσταση μεταξύ τους. Ενώνουμε τις συστάδες αυτές σε μια συστάδα και ανανεώνουμε τον πίνακα αποστάσεων διαγράφοντας τις γραμμές και στήλες που αντιστοιχούν σε αυτές, ενώ παράλληλα προσθέτουμε μια γραμμή και μια στήλη με τις αποστάσεις της ένωσής τους από τις υπόλοιπες συστάδες. Επαναλαμβάνουμε τα παραπάνω βήματα $(N-1)$ φορές μέχρι να καταλήξουμε στην ύπαρξη μίας μόνο συστάδας η οποία περιέχει το σύνολο των στοιχείων. Σε κάθε ένα από τα βήματα που περιγράφηκαν, καταγράφουμε τις συστάδες που δημιουργήθηκαν κατά τη διάρκεια της ένωσης (συσσώρευσης) καθώς και το επίπεδο (απόσταση) στο οποίο δημιουργήθηκε η κάθε μία.

Επιλογές για απόσταση μεταξύ συστάδων:

α) Μέθοδος της απλής συνένωσης (Single Linkage Method)

Ως απόσταση μεταξύ δύο συστάδων A και B θεωρούμε την απόσταση με την μικρότερη τιμή από όλες τις πιθανές αποστάσεις μεταξύ ενός στοιχείου (ή συστάδας) του A και ενός στοιχείου (ή συστάδας) του B .

β) Μέθοδος της πλήρους συνένωσης (Complete Linkage Method)

Ως απόσταση μεταξύ δύο συστάδων A και B θεωρούμε την απόσταση με την μεγαλύτερη τιμή από όλες τις πιθανές αποστάσεις μεταξύ ενός στοιχείου (ή συστάδας) του A και ενός στοιχείου (ή συστάδας) του B .

γ) Μέθοδος των σταθμισμένων μέσων (Weighted Average Linkage Method)

Ως απόσταση μεταξύ δύο συστάδων A και B θεωρούμε την μέση απόσταση μεταξύ των δύο συστάδων (το άθροισμα όλων των πιθανών αποστάσεων μεταξύ

ενός στοιχείου του A και ενός στοιχείου του B διά του γινομένου του πλήθους των στοιχείων της A επί του πλήθους των στοιχείων της B .

δ) Μέθοδος του Ward (Ward's Method)

Η μέθοδος αυτή διαφέρει από τις υπόλοιπες και είναι σχεδιασμένη να ελαχιστοποιεί τη διακύμανση μέσα στις ομάδες. Για κάθε συστάδα k θεωρούμε ως ESS_k το άθροισμα των τετραγώνων των αποστάσεων κάθε στοιχείου της συστάδας από τον μέσο της συστάδας και ESS το άθροισμα των ESS_k . Ως απόσταση μεταξύ δύο συστάδων A και B θεωρούμε την αύξηση που θα προκύψει στο ESS από την ένωση των δύο συστάδων.

Συγκρίνοντας τις διάφορες μεθόδους μεταξύ τους με δεδομένα προσομοίωσης έχει διαπιστωθεί πως συνήθως η καλύτερη ομαδοποίηση επιτυγχάνεται με τη μέθοδο του Ward και τη μέθοδο των σταθμισμένων μέσων (Weighted Average Linkage Method). Η μέθοδος της απλής συνένωσης είναι αυτή με τη χειρότερη επίδοση. Παρόλα αυτά σε πολλά προβλήματα δεν είναι ξεκάθαρο ποια μέθοδος είναι προτιμότερη και η καθεμία δουλεύει καλύτερη με συγκεκριμένη μορφή δεδομένων. Φυσικά, εάν οι ομάδες είναι αρκετά διαφορετικές μεταξύ τους όλες οι μέθοδοι θα βρουν τη σωστή ομαδοποίηση (Κούτρας 2013).

5.3.2 Διαιρετικές μέθοδοι

Κατά την εκτέλεση διαιρετικών μεθόδων ακολουθείται η παρακάτω διαδικασία:

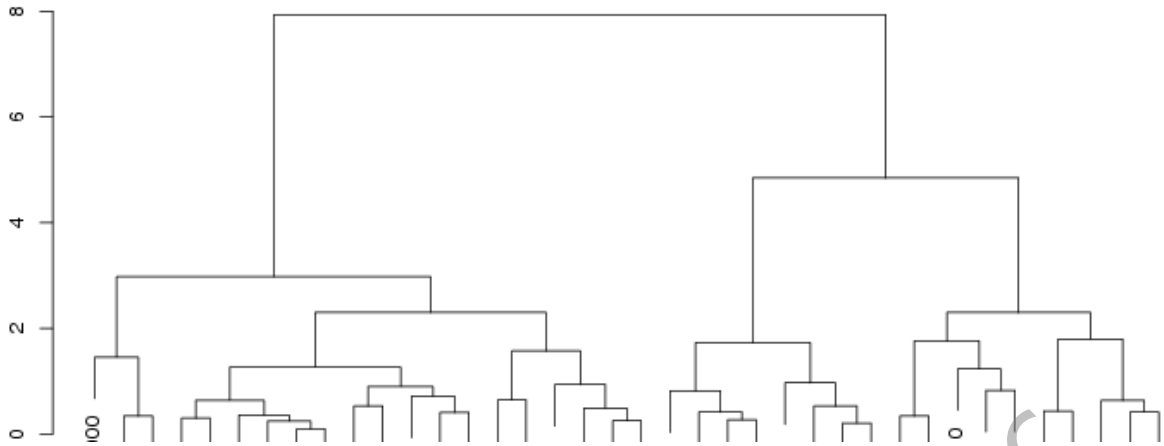
Ξεκινάμε από μία ενιαία συστάδα στην οποία περιέχονται όλα τα στοιχεία που μελετάμε. Δημιουργούμε μία νέα συστάδα επιλέγοντας εκείνο το στοιχείο το οποίο απέχει την μεγαλύτερη απόσταση από τα υπόλοιπα. Στη συνέχεια, ανακατανέμουμε τα στοιχεία μεταξύ των δύο συστάδων βάση της απόστασής τους από αυτές. Επιλέγουμε τη συστάδα με τη μεγαλύτερη διάμετρο (μεγαλύτερη απόσταση μεταξύ δυο στοιχείων της συστάδας) και από αυτήν δημιουργείται η τρίτη συστάδα επιλέγοντας εκείνο το στοιχείο το οποίο απέχει την μεγαλύτερη απόσταση από τα υπόλοιπα, όπως και προηγουμένως. Η διαδικασία μπορεί να συνεχιστεί θεωρητικά μέχρις ότου να δημιουργηθούν τόσες συστάδες όσες τα στοιχεία μας, αν και στην πράξη, η διαιρετική μέθοδος σταματά όταν έχουμε φτάσει σε έναν επιθυμητό αριθμό ομάδων, οι οποίες διαχωρίζονται ικανοποιητικά μεταξύ τους.

Ο κύριος λόγος που οι διαιρετικές μέθοδοι δεν είναι αρκετά διαδεδομένες είναι ότι απαιτούν πολύ περισσότερους υπολογισμούς από ότι οι συσσωρευτικές μέθοδοι. Πράγματι, ενώ στο πρώτο στάδιο ενός συσσωρευτικού αλγόριθμου όλες οι πιθανές ενώσεις δύο παρατηρήσεων (από τις n διαθέσιμες) είναι $n(n-1)/2$, στο πρώτο στάδιο ενός διαιρετικού αλγόριθμου, οι πιθανοί διαμερισμοί του συνόλου των n παρατηρήσεων σε δύο μη κενά σύνολα είναι $2^{n-1} - 1$. Στην πρώτη περίπτωση η αύξηση του n οδηγεί το πλήθος των δυνατών επιλογών σε αύξηση ανάλογη του τετραγώνου του n , ενώ στη δεύτερη έχουμε εκθετική αύξηση του πλήθους.

5.3.3 Επιλογή του πλήθους των ομάδων

Η ανάπτυξη λογικών κανόνων που οδηγούν στην εύρεση του βέλτιστου αριθμού των ομάδων ενός συνόλου δεδομένων έχει απασχολήσει αρκετούς ερευνητές που δραστηριοποιούνται στον κλάδο της ανάλυσης κατά συστάδες αφού είναι φανερό ότι το συγκεκριμένο πρόβλημα έχει εξαιρετικό πρακτικό ενδιαφέρον.

Ένας απλός πρακτικός τρόπος εύρεσης του πλήθους των ομάδων είναι να εξετάσουμε το δενδροδιάγραμμα που προκύπτει από μια ιεραρχική συσσωρευτική μέθοδο και από αυτό να καθορίσουμε το βέλτιστο πλήθος. Πιο συγκεκριμένα, σε εκείνο το σημείο του δενδρογράμματος που παρατηρείται η μεγαλύτερη μεταβολή της ποσότητας που καταγράφεται στον οριζόντιο άξονα (απόσταση ή μέτρο ομοιότητας) μπορούμε να φέρουμε μια παράλληλη γραμμή προς τον κατακόρυφο άξονα και να δούμε σε πόσα σημεία τέμνει το δενδρόγραμμα. Το πλήθος k , για το οποίο παρατηρούμε μεγάλες αποστάσεις συνένωσης σε σχέση με το προηγούμενο ($k-1$ ομάδες) αποτελεί μια λογική τιμή για το βέλτιστο πλήθος των ομάδων. Για παράδειγμα, με βάση το επόμενο δενδροδιάγραμμα, μπορεί κανείς να ισχυριστεί ότι το βέλτιστο πλήθος των ομάδων που υπάρχουν στα δεδομένα είναι 3 ή 2, αφού, όταν πάμε στο επόμενο επίπεδο συνένωσης (1 ομάδα), η απόσταση στην οποία πραγματοποιείται αυτή εκτινάσσεται στο 8 ενώ προηγουμένως ήταν κοντά στο 4.



Διάγραμμα 5.1: Δενδροδιάγραμμα

Είναι φανερό ότι η απόφαση με βάση τα προαναφερθέντα κριτήρια είναι σε μεγάλο βαθμό υποκειμενική. Δίνουμε στη συνέχεια κάποιες πιο αντικειμενικές τεχνικές που βασίζονται σε μεθόδους ανάλυσης διακύμανσης (ANOVA) και σε αποστάσεις από τα κέντρα των ομάδων.

A. Αν \bar{x}_j είναι ο μέσος (κέντρο βάρους) της j ομάδας ($j=1,2,\dots,k$) και \bar{x} είναι ο μέσος (κέντρο βάρους) όλων των παρατηρήσεων, τότε αναλύουμε το συνολικό άθροισμα τετραγωνικών αποκλίσεων στα εξής δύο μέρη

$$B = \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})' \quad \text{και} \quad W = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)(x_{ij} - \bar{x}_j)'$$

και επιδιώκουμε το B να είναι όσο το δυνατόν μεγαλύτερο και το W όσο το δυνατόν μικρότερο.

Επειδή το κριτήριο αυτό θα οδηγεί πάντοτε σε επιλογές με μεγάλες τιμές του k , συνήθως επιλέγουμε την ομαδοποίηση για την οποία προκύπτει η μεγαλύτερη τιμή του λόγου

$$c = \frac{\frac{tr(B)}{k-1}}{\frac{tr(W)}{k-g}}$$

B. Αφού υπολογίσουμε τις ποσότητες

$$W_k = \sum_{i \in G} d(x_i, \bar{x}_k), \quad (G = 1, 2, \dots, k) \quad \text{και} \quad P_k = \sum_{j=1}^k W_j$$

επιλέγουμε το k (πλήθος ομάδων) για το οποίο το P_k είναι ελάχιστο.

Ένα εναλλακτικό κριτήριο προκύπτει αν αθροίσουμε τις αποστάσεις όλων των παρατηρήσεων από τον ολικό μέσο

$$T = \sum_{i=1}^n d(x_i, \bar{x})$$

και επιλέξουμε το k που μεγιστοποιεί έναν από τους παρακάτω δείκτες

$$R^2 = 1 - \frac{P_k}{T}, \quad F = \frac{\frac{T - P_k}{k - 1}}{\frac{P_k}{k - g}}$$

Αξίζει να αναφερθεί ότι, η τελευταία στατιστική συνάρτηση μοιάζει με το κριτήριο F της ανάλυσης διακύμανσης (ANOVA).

Τέλος, μια άλλη μέθοδος που έχει προταθεί είναι να γίνεται προσθήκη νέας ομάδας αν η ποσότητα $(n - k + 1)P_k / p_{k+1}$ υπερβαίνει την τιμή 10, αλλιώς να θεωρείται ότι το πλήθος k των ομάδων στο οποίο φτάσαμε είναι το βέλτιστο.

5.4 Μη ιεραρχικές μέθοδοι ομαδοποίησης

5.4.1 Μέθοδος K-Means

Κατά την εκτέλεση της μεθόδου K-Means, επιλέγουμε τυχαία K στοιχεία τα οποία θα αποτελέσουν τους αρχικούς πυρήνες των συστάδων. Στη συνέχεια, κατανέμουμε κάθε στοιχείο που μελετάμε στη συστάδα εκείνη της οποίας ο πυρήνας είναι πιο κοντά σε αυτό. Έπειτα, υπολογίζονται οι νέοι πυρήνες (centroids) για τις συστάδες ως ο μέσος όρος των στοιχείων της κάθε συστάδας. Η διαδικασία επαναλαμβάνεται έως ότου κανένα στοιχείο δεν αλλάζει κατανομή μεταξύ των συστάδων.

5.4.2 Partitioning around Medoids (PAM)

Η μέθοδος Partitioning Around Medoids διαφέρει από την μέθοδο k-means στο σημείο ότι ως πυρήνας μιας συστάδας είναι πάντα ένα στοιχείο της συστάδας (medoid) και επιδιώκεται η ελαχιστοποίηση της απόστασης των υπόλοιπων στοιχείων από τον πυρήνα. Αρχικά επιλέγεται ένα καλό αρχικό σύνολο από medoids (build phase). Ακολούθως ελέγχεται κατά πόσο η εναλλαγή ενός στοιχείου με ένα medoid θα ελαχιστοποιήσει την απόσταση μεταξύ του πυρήνα και των άλλων στοιχείων και αν ναι, πραγματοποιείται (swap phase).

5.4.3 Fuzzy Analysis Clustering (Fanny)

Η μέθοδος ομαδοποίησης Fuzzy επιτρέπει σε κάθε στοιχείο να ανήκει σε περισσότερες από μια συστάδες. Αυτό επιτυγχάνεται υπολογίζοντας κάποια ποσοστά (memberships) για κάποιο στοιχείο για κάθε συστάδα τέτοια ώστε το άθροισμά τους να είναι ίσο με 1.

Στις μεθόδους αυτές, ο αριθμός των ομάδων που θα δημιουργηθούν είναι εξ αρχής γνωστός. Στην πραγματικότητα τα κριτήρια επιλογής του αριθμού των ομάδων είναι πολλά, αλλά μερικές φορές η ερμηνεία που μπορούμε να δώσουμε είναι ο καλύτερος οδηγός για να επιλέξουμε αυτόν τον αριθμό. Επίσης, για πρακτικούς σκοπούς, μπορούμε να διεξάγουμε μια ιεραρχική ανάλυση και αφού δούμε όλες τις λύσεις κι επιλέξουμε τον αριθμό των ομάδων, να προχωρήσουμε σε μια μη ιεραρχική μέθοδο συσταδοποίησης.

5.5. Γραφική απεικόνιση και αξιολόγηση των συστάδων

5.5.1 Διάγραμμα Silhouette

Το διάγραμμα Silhouette αναφέρεται σε μια μέθοδο ερμηνείας και επικύρωσης των συστάδων των δεδομένων. Η τεχνική παρέχει μια συνοπτική γραφική αναπαράσταση του πόσο καλά κάθε αντικείμενο βρίσκεται εντός της συστάδας καθώς και ποια σημεία βρίσκονται κάπου μεταξύ δύο συστάδων. Περιγράφηκε για πρώτη φορά από τον Peter J. Rousseeuw (Rousseeuw 1987) ο οποίος πρότεινε την εξής μεθοδολογία:

1. Για κάθε παρατήρηση i που ανήκει σε μία συστάδα υπολογίζεται η ποσότητα $a(i)$ η οποία είναι η μέση απόσταση του σημείου αυτού από τις παρατηρήσεις που απαρτίζουν ολόκληρη τη συστάδα και εκφράζει το πόσο καλά έχει τοποθετηθεί το άτομο αυτό στη συστάδα που ανήκει. Μικρή τιμή του $a(i)$ υποδηλώνει πως το σημείο ορθά έχει ανατεθεί στη συστάδα του, και αντίστροφα.
2. Στη συνέχεια υπολογίζεται η ποσότητα $b(i)$ ως η μικρότερη μέση απόσταση του σημείου i από τα σημεία των υπολοίπων συστάδων που έχουν σχηματιστεί και η συστάδα την οποία αφορά ονομάζεται «γειτονική». Μεγάλη τιμή της ποσότητας $b(i)$ υποδηλώνει πως η παρατήρηση i δεν ταιριάζει με την γειτονική του συστάδα.

3. Ορίζεται η ποσότητα:

$$s(i) = \frac{b(i) - \alpha(i)}{\max\{\alpha(i), b(i)\}},$$

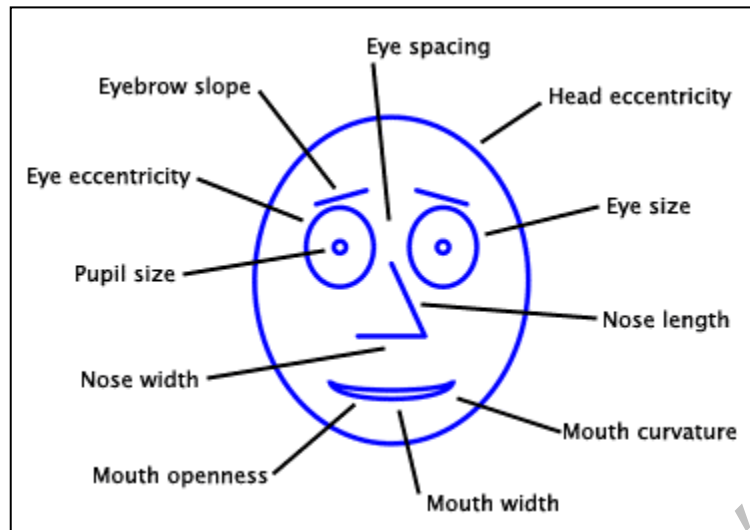
η οποία λαμβάνει τιμές στο διάστημα $[-1,1]$ και σύμφωνα με όσα αναφέρθηκαν, όταν λαμβάνει την τιμή 1 το σημείο έχει ομαδοποιηθεί σωστά, ενώ όταν λαμβάνει την τιμή -1 το σημείο ταιριάζει περισσότερο στη γειτονική του συστάδα. Μια τιμή της ποσότητας $s(i)$ κοντά στο 0 υποδηλώνει πως το σημείο βρίσκεται στα όρια των δύο ομάδων.

Τέλος, υπολογίζεται η μέση ποσότητα $s(i)$ επί του συνόλου των δεδομένων η οποία χρησιμοποιείται ως ένα μέτρο αξιολόγησης του πόσο καλά έχουν ομαδοποιηθεί τα δεδομένα μέσα στις συστάδες. Εάν υπάρχουν πάρα πολλές ή πολύ λίγες συστάδες, κάποιες από αυτές θα παρουσιάσουν στενότερο διάγραμμα silhouette από τις υπόλοιπες, κάνοντας φανερό, έτσι, το γεγονός ότι έχει επιλεγεί λάθος αριθμός των k ομάδων.

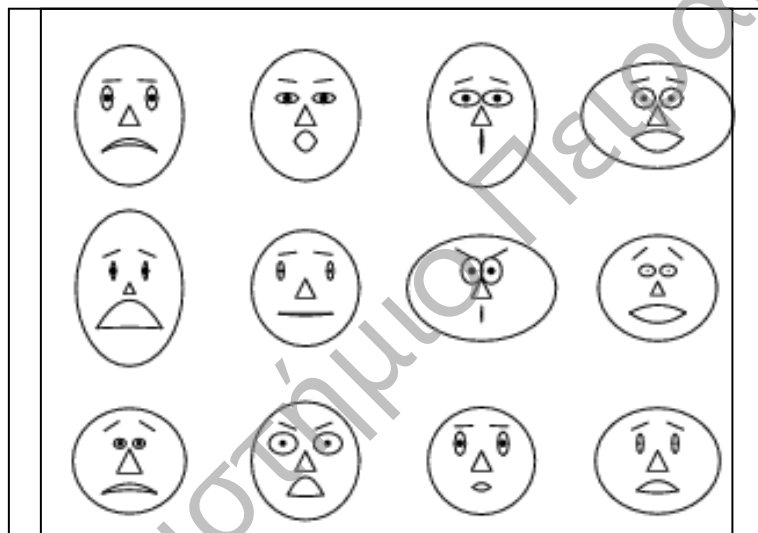
5.5.2 Όψεις Chernoff

Οι όψεις Chernoff (Chernoff faces) είναι μια γραφική αναπαράσταση των πολυμεταβλητών δεδομένων σε ανθρώπινα πρόσωπα. Το 1973, ο στατιστικός Herman Chernoff δημοσίευσε το άρθρο του “The Use of Faces to Represent Points in k -Dimensional Space Graphically” στην εφημερίδα της αμερικανικής στατιστικής ένωσης (JASA), όπου κωδικοποίησε δεδομένα σε ανθρώπινα χαρακτηριστικά προσώπου, γνωστά και ως όψεις Chernoff, με σκοπό να βοηθήσει τους ερευνητές να ανιχνεύσουν πρότυπα, ομάδες ή και συσχετίσεις σε αυτά.

Τα επιμέρους τμήματα του προσώπου, όπως τα μάτια, η μύτη, το στόμα και τα αυτιά χρησιμοποιούνται κατάλληλα ώστε να αναπαρασταθούν οι τιμές των μεταβλητών και αυτό επιτυγχάνεται διαφοροποιώντας το σχήμα, το μέγεθος, τον προσανατολισμό καθώς και την καμπυλότητα των χαρακτηριστικών αυτών. Στόχος είναι η αξιοποίηση της έμφυτης ικανότητας του ανθρώπινου νου να αναγνωρίζει τις μικρές διαφορές στα χαρακτηριστικά του ανθρώπινου προσώπου. Έτσι, κάθε μία από αρκετές μεταβλητές αντιστοιχίζεται σε κάποιο από τα παραπάνω χαρακτηριστικά (Σχήμα 1) ενώ από τον συνδυασμό τους προκύπτουν οι επιμέρους όψεις (Σχήμα 2).



Σχήμα 5.1: Διαφοροποιήσεις των χαρακτηριστικών του προσώπου



Σχήμα 5.2: Πιθανοί σχηματισμοί όψεων Chernoff

Το σκεπτικό πίσω από τον γραφικό αυτό τρόπο αναπαράστασης σημείων με δυνατό αλλά ίσως πολύπλοκο τρόπο διασύνδεσης, είναι να μπορεί ένας ερευνητής να αντιληφθεί γρήγορα τον τρόπο με τον οποίο σχετίζονται αυτά και στη συνέχεια να μπορεί να εφαρμόσει την κατάλληλη στατιστική μέθοδο (Chernoff, 1973). Είναι ένας εμπειρικός τρόπος παρατήρησης και για αυτόν τον λόγο υπάρχουν πολύ καλύτεροι τρόποι αξιολόγησης των δεδομένων. Παρόλα αυτά είναι μια ενδιαφέρουσα μέθοδος για να ασχοληθεί κανείς.

5.6 Εφαρμογή στα δεδομένα

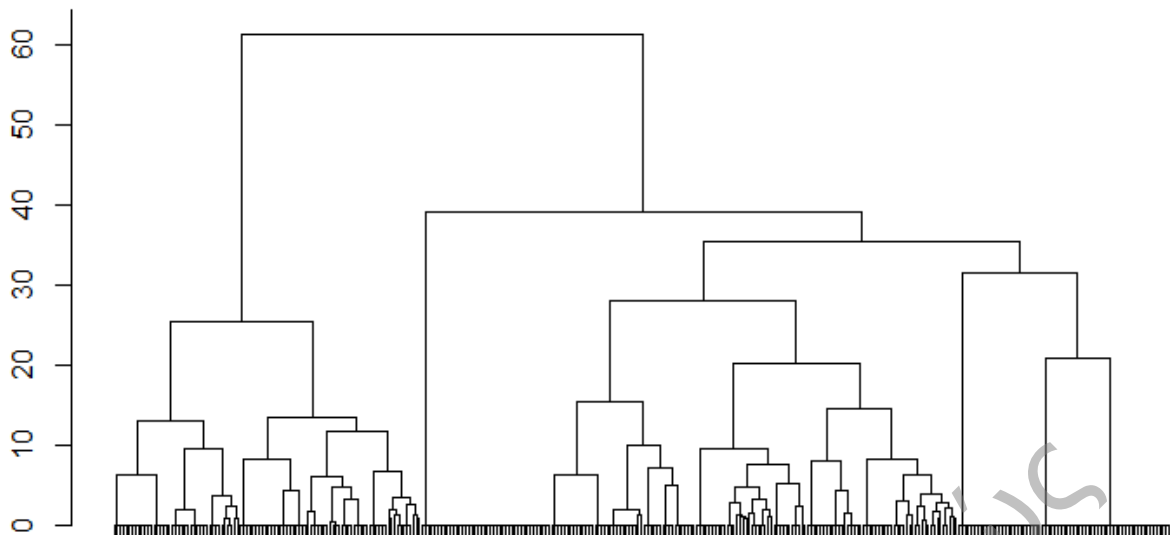
Όταν σε μια στατιστική μέθοδο υπολογίζονται αποστάσεις, θα πρέπει να είμαστε πολύ προσεκτικοί με το γεγονός ότι μπορεί οι μεταβλητές να έχουν διαφορετική μονάδα μέτρησης μεταξύ τους (π.χ.: ηλικία έναρξης της νόσου, οικογενειακό ιστορικό για την νόσο Πάρκινσον). Μεταβλητές με μεγάλες τιμές θα επηρεάζουν περισσότερο την απόσταση, συγκριτικά με μεταβλητές μικρότερων τιμών. Στην συγκεκριμένη εφαρμογή, μορφοποιήσαμε κατάλληλα τα δεδομένα ώστε τα σκορ των παραγόντων να είναι κανονικοποιημένες τιμές, και να αποφύγουμε με αυτόν τον τρόπο το πρόβλημα αυτό. Συνεπώς, το συνολικό άθροισμα των μέσων είναι 0 και, έτσι, οι θετικές τιμές υποδηλώνουν σκορ πάνω από τον μέσο όρο, ενώ αρνητικές τιμές υποδηλώνουν τιμές κάτω από τον μέσο όρο.

Ακόμα, χρησιμοποιήθηκε η εξής κωδικοποίηση για τις μεταβλητές που επιλέχθηκαν ως παράγοντες ομαδοποίησης των ασθενών:

Πίνακας 5.1: Κωδικοποίηση μεταβλητών που χρησιμοποιήθηκαν στην ομαδοποίηση

Μεταβλητή	Κωδικοποίηση
Νοσοκομείο εξέτασης	Αθήνας=0, Σύρου=1
Φύλο	Άνδρας=0, Γυναίκα=1
Θρήσκευμα	Χριστιανός Ορθόδοξος=0, Χριστιανός Καθολικός=1
Οικογενειακό ιστορικό ΝΠ	Όχι=0, Ναι=1
Κάπνισμα κατά τις τελευταίες 30 ημέρες	Όχι=0, Ναι=1
Κατανάλωση καφέ τώρα	Όχι=0, Ναι=1
Έκθεση σε ζιζανιοκτόνα ή/και φυτοφάρμακα	Όχι=0, Ναι=1
Ηλικία έναρξης ΝΠ	[30,50)=0, $\geq 50=1$

Σε πρώτο στάδιο εφαρμόστηκαν όλες οι δυνατές ιεραρχικές μέθοδοι συσταδοποίησης που αναφέρθηκαν νωρίτερα. Μετά από δοκιμές, εντοπίστηκε ότι η καλύτερη ερμηνευτικά λύση είναι εκείνη η οποία προέκυψε χρησιμοποιώντας την μέθοδο του Ward και ως μετρική την ευκλείδεια απόσταση. Αυτή θα παρουσιαστεί στη συνέχεια:



Διάγραμμα 5.2: Δενδροδιάγραμμα που αφορά την συσταδοποίηση των ασθενών σε 2 ομάδες

Σύμφωνα με το παραπάνω διάγραμμα, μπορούμε να ισχυριστούμε πως οι ασθενείς που έχουν εξεταστεί, μπορούν να ομαδοποιηθούν σε δύο ομάδες. Η πρώτη ομάδα αποτελείται από 243 ασθενείς, ενώ η δεύτερη από 100, και η σύνθεσή τους από τις επιμέρους μεταβλητές διαμορφώνεται ως εξής:

- Ομάδα 1^η

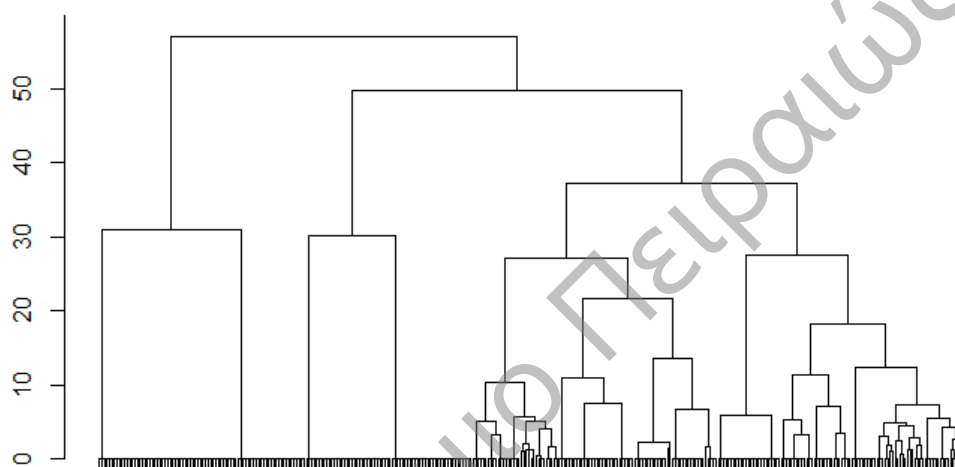
Αποτελείται κατά 97.9% από άτομα που εξετάστηκαν στο Αττικό νοσοκομείο. Το 40% που την διαμορφώνει είναι γυναίκες, ενώ μόλις το 0.41% είναι Χριστιανοί Καθολικοί. Το 23.87% έχει οικογενειακό ιστορικό για την νόσο του Πάρκινσον ενώ το 19.75% έχει εμφανίσει νεανικό Παρκινσονισμό. Το 45.68% καταναλώνει καφέ κατά το κοντινό διάστημα της εξέτασης, το 17.28% έχει εκτεθεί κάποια στιγμή της ζωής του σε ζιζανιοκτόνα ή/και φυτοφάρμακα, αλλά μόνο το 12.76% έχει καπνίσει τις τελευταίες 30 ημέρες.

- Ομάδα 2^η

Αποτελείται αποκλειστικά από ασθενείς που εξετάστηκαν στο νοσοκομείο Σύρου. Κατά το ήμισυ η ομάδα αποτελείται από γυναίκες, ενώ κατά 35% από Χριστιανούς Καθολικούς. Το 26% έχει οικογενειακό ιστορικό για την νόσο του Πάρκινσον, μόλις 2% όμως έχει εμφανίσει την νόσο πριν τα 50 του έτη. Το 40% καταναλώνει καφέ κατά το κοντινό διάστημα της εξέτασης, το 26% έχει εκτεθεί κάποια στιγμή της ζωής του σε ζιζανιοκτόνα ή/και φυτοφάρμακα, αλλά μόνο το 2% έχει καπνίσει τις τελευταίες 30 ημέρες.

Σύμφωνα με τα παραπάνω, υπάρχει ξεκάθαρος διαχωρισμός μεταξύ των ασθενών που εξετάστηκαν στα δύο νοσοκομεία. Συνεπώς, καταφέραμε να επαληθεύσουμε τα αποτελέσματα του Κεφαλαίου 4 με την μέθοδο της ομαδοποίησης.

Προκειμένου να διερευνήσουμε εκτενέστερα τα δεδομένα μας, και να εντοπίσουμε ομαδοποιήσεις σύμφωνα με τα λοιπά χαρακτηριστικά, ακολουθήσαμε την ίδια διαδικασία παραλείποντας την μεταβλητή «Νοσοκομείο εξέτασης». Σε αυτήν την περίπτωση, δημιουργήθηκαν τέσσερις ομάδες κατάταξης των ασθενών όπως φαίνεται και στο παρακάτω δενδροδιάγραμμα:



Διάγραμμα 5.3: Δενδροδιάγραμμα που αφορά την συσταδοποίηση των ασθενών σε 4 ομάδες, ανεξαρτήτως νοσοκομείου εξέτασης

Η πρώτη ομάδα αποτελείται 82 άτομα, η δεύτερη από 66, η τρίτη από 98, ενώ η τέταρτη από 97 άτομα. Η επιμέρους σύστασή τους έχει ως εξής:

- Ομάδα 1^η

Αποτελείται αποκλειστικά από Χριστιανούς Ορθόδοξους άνδρες, που δεν έχουν οικογενειακό ιστορικό της νόσου, δεν έχουν καπνίσει κατά τις τελευταίες 30 ημέρες πριν την εξέταση, δεν έχουν εκτεθεί ποτέ σε κάποιο ζιζανιοκτόνο ή/και φυτοφάρμακο, ενώ κανένας δεν έχει εμφανίσει νεανικής μορφής Παρκινσονισμό. Ακόμα, το 59.76% καταναλώνει καφέ κατά το πρόσφατο παρελθόν της εξέτασης.

▪ Ομάδα 2^η

Αποτελείται κατά 28.57% από γυναίκες, κατά 2% από Χριστιανούς Καθολικούς, κατά 12.25% από ασθενείς με οικογενειακό ιστορικό ΝΠ. Το 33.67% έχει καπνίσει τις τελευταίες 30 ημέρες πριν την εξέταση, ενώ το 49% έχει εκτεθεί σε κάποιο ζιζανιοκτόνο ή/και φυτοφάρμακο. Το 44.9% έχει εμφανίσει νεανικό Παρκινσονισμό και το 67.35% καταναλώνει καφέ.

▪ Ομάδα 3^η

Αποτελείται αποκλειστικά από Χριστιανές Ορθόδοξες γυναίκες, που δεν έχουν οικογενειακό ιστορικό της νόσου, δεν έχουν καπνίσει κατά τις τελευταίες 30 ημέρες πριν την εξέταση, δεν έχουν εκτεθεί ποτέ σε κάποιο ζιζανιοκτόνο ή/και φυτοφάρμακο, ενώ καμία δεν έχει εμφανίσει νεανικής μορφής Παρκινσονισμό. Ακόμα, το 48.48 % καταναλώνει καφέ κατά το πρόσφατο παρελθόν της εξέτασης.

▪ Ομάδα 4^η

Αποτελείται κατά 54.64% από γυναίκες, κατά 35.05% από Χριστιανούς Καθολικούς, κατά 74.23% από ασθενείς με οικογενειακό ιστορικό ΝΠ. Κανένας δεν έχει καπνίσει τις τελευταίες 30 ημέρες πριν την εξέταση, ενώ το 20.62% έχει εκτεθεί σε κάποιο ζιζανιοκτόνο ή/και φυτοφάρμακο. Το 6.20% έχει εμφανίσει νεανικό Παρκινσονισμό και το 4.12% καταναλώνει καφέ.

Από τα παραπάνω συμπεραίνουμε πως βασικό χαρακτηριστικό διαχωρισμού μεταξύ των δύο πρώτων ομάδων είναι ο παράγοντας φύλο. Στην διαμόρφωση των ομάδων 3 και 4, όλοι οι παράγοντες φαίνεται να συμβάλουν εξίσου, συνεπώς δεν μπορούμε να εξάγουμε κάποιο ξεκάθαρο συμπέρασμα.

Αντίστοιχα, χρησιμοποιώντας την μέθοδο των K-means στο δείγμα των ασθενών προχωρούμε στην σύσταση δύο ομάδων, καθεμία από τις οποίες περιέχει 187 και 156 άτομα. Το άθροισμα των τετραγώνων εντός των ομάδων είναι 10063.029 και 9481.334, αντίστοιχα, ενώ ο λόγος του αθροίσματος των τετραγώνων μεταξύ των ομάδων προς το συνολικό άθροισμα τετραγώνων ισούται με 24.5 %.

Στην περίπτωση των τεσσάρων ομάδων, σχηματίζονται ομάδες με 105, 102, 35 και 101 άτομα. Το άθροισμα των τετραγώνων εντός των ομάδων είναι 2912.717, 3779.031, 2016.033 και 4608.637, αντίστοιχα, ενώ ο λόγος του αθροίσματος των τετραγώνων μεταξύ των ομάδων προς το συνολικό άθροισμα τετραγώνων ισούται με 50.4 %.

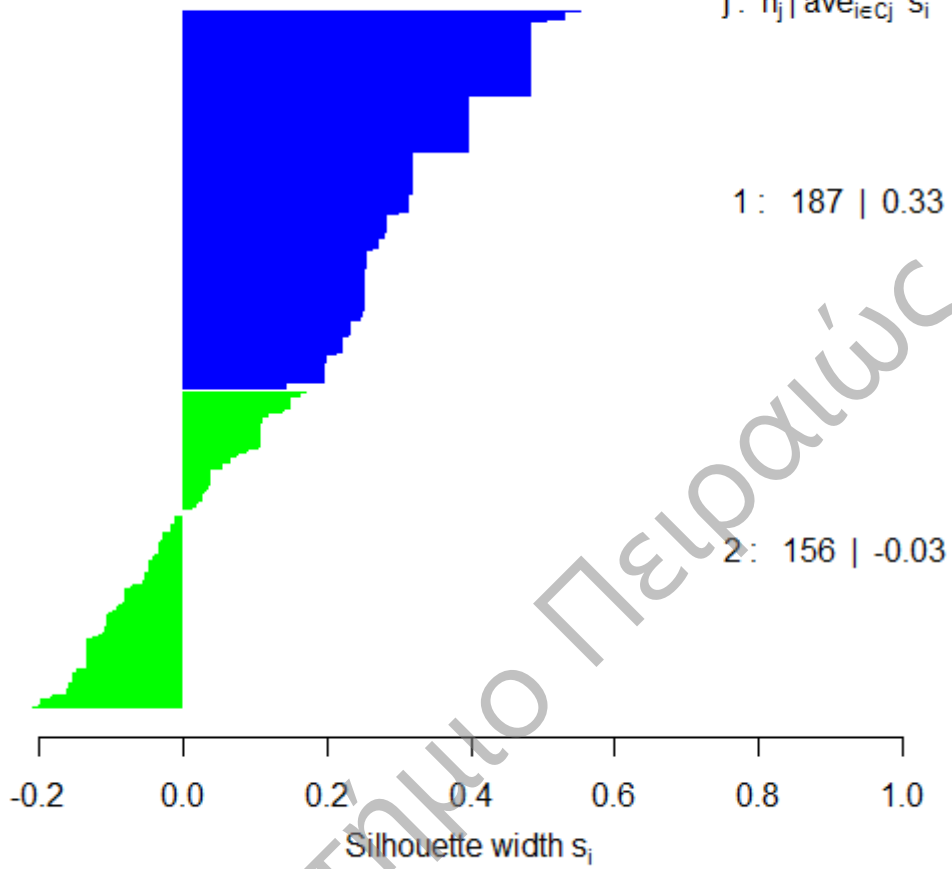
Ακόμα, παρουσιάζονται τα διαγράμματα Silhouette που αφορούν τις δυο πρώτες ομαδοποιήσεις.

Silhouette Plot k=2

n = 343

2 clusters C_j

$j: n_j | \text{ave}_{i \in C_j} s_i$



Διάγραμμα 5.4: Silhouette διάγραμμα που αφορά την συσταδοποίηση των ασθενών σε 2 ομάδες

Silhouette Plot k=4

n = 343

4 clusters C_j

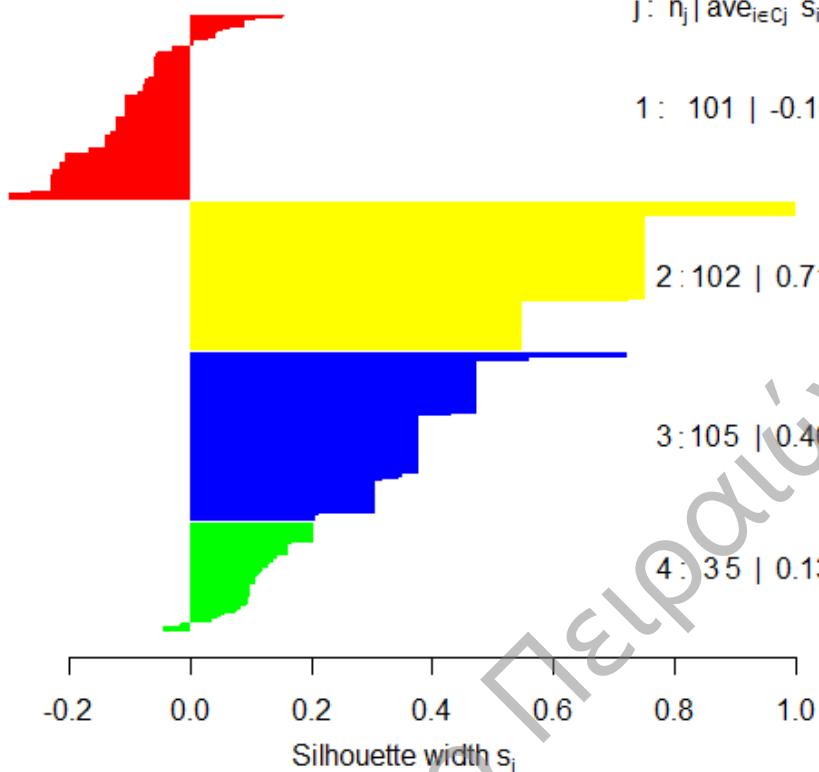
$j: n_j | \text{ave}_{i \in C_j} s_i$

1: 101 | -0.10

2: 102 | 0.71

3: 105 | 0.40

4: 35 | 0.13

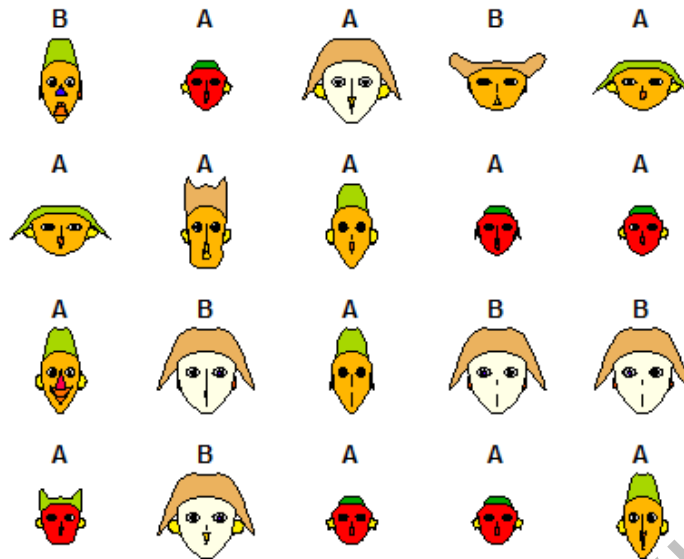


Διάγραμμα 5.5: Silhouette διάγραμμα που αφορά την συσταδοποίηση των ασθενών σε 4 ομάδες

Η συσταδοποίηση σε 2 ομάδες φαίνεται να περιγράφει καλύτερα τα δεδομένα μας καθώς η μέση ποσότητα $s(i)$ λαμβάνει μικρότερη τιμή για αυτήν, υποδηλώνοντας πως ο σωστός αριθμός ομάδων για τα δεδομένα μας είναι 2.

Παρόλα αυτά, και στις δύο περιπτώσεις τα σημεία φαίνεται να έχουν καταναμηθεί σωστά στις ομάδες τους καθώς τα επιμέρους $s(i)$ παίρνουν στην πλειονότητά τους θετικές τιμές φτάνοντας ακόμα και την τιμή 1. Τα σημεία με αρνητικές τιμές $s(i)$ παίρνουν maximum την τιμή -0.2 υποδηλώνοντας πως είναι οριακά λανθασμένα καταναμημένα στις ομάδες τους.

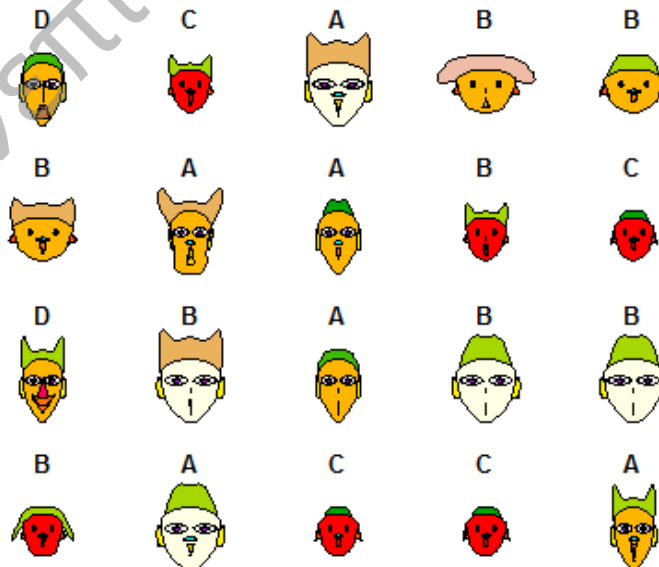
Τέλος, παρουσιάζονται οι όψεις Chernoff για τους 20 πρώτους ασθενείς της βάσης δεδομένων μας, καθώς και σε ποια ομάδα ανήκουν δεδομένης της παραπάνω συσταδοποίησης σε 2 ομάδες:



Διάγραμμα 5.6: Chernoff faces για τα 20 πρώτους ασθενείς της βάσης δεδομένων και ομαδοποίησή τους σύμφωνα με την συσταδοποίηση σε 2 ομάδες

Σύμφωνα με την κωδικοποίηση των χαρακτηριστικών που χρησιμοποιήθηκε (Παράρτημα Α.1) μπορούμε εύκολα να καταλάβουμε, για παράδειγμα, από το μήκος του προσώπου ότι ο 2^{ος}, ο 4^{ος}, ο 5^{ος}, ο 6^{ος}, ο 9^{ος}, ο 10^{ος}, ο 16^{ος}, ο 18^{ος} και ο 19^{ος} ασθενής που εξετάστηκαν είναι άνδρες, ενώ αντιλαμβανόμαστε ότι όσοι έχουν στενό πρόσωπο δεν έχουν οικογενειακό ιστορικό για την νόσο Πάρκινσον.

Αντίστοιχα, για τη συσταδοποίηση των ασθενών σε 4 ομάδες, οι όψεις Chernoff είναι:



Διάγραμμα 5.7: Chernoff faces για τα 20 πρώτους ασθενείς της βάσης δεδομένων και ομαδοποίησή τους σύμφωνα με την συσταδοποίηση σε 4 ομάδες

Σύμφωνα με την κωδικοποίηση που χρησιμοποιήθηκε σε αυτήν την περίπτωση (Παράρτημα Α.2), όσοι απεικονίζεται να έχουν ίσια μαλλιά, για παράδειγμα, δεν καταναλώνουν καφέ κατά την κοντινή στην εξέτασή τους χρονική περίοδο.

5.7 Συμπεράσματα

Αφού δοκιμάσαμε να τμηματοποιήσουμε το δείγμα των ασθενών σε k επιμέρους ομάδες, καταλήξαμε στο συμπέρασμα πως τα δεδομένα μας μπορούν να κατανεμηθούν ικανοποιητικά σε 2 ομάδες όπου βασικός παράγοντας διαφοροποίησής τους είναι το νοσοκομείο εξέτασης του ασθενούς. Όπως έχουμε αναφέρει και στο Κεφάλαιο 4, οι δύο υποπληθυσμοί διαφέρουν σημαντικά μεταξύ τους ως προς τη διάμεση ηλικία έναρξης της νόσου, τις καπνιστικές τους συνήθειες καθώς και την κινητική κατάστασή τους.

Επιπλέον, όταν δεν λαμβάνουμε υπ' όψιν τον παράγοντα νοσοκομείο εξέτασης, τα δεδομένα μας φαίνεται να κατανέμονται ικανοποιητικά σε 4 ομάδες με κύριους παράγοντες διαχωρισμού το φύλο, το οικογενειακό ιστορικό για την νόσο Πάρκινσον καθώς και το θρήσκευμα του κάθε ασθενούς.

ΚΕΦΑΛΑΙΟ 6: ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

6.1 Εισαγωγή

Οι μέθοδοι παλινδρόμησης έχουν γίνει αναπόσπαστο κομμάτι της ανάλυσης δεδομένων. Περιγράφουν τη σχέση μεταξύ μιας μεταβλητής απόκρισης με μία ή περισσότερες ερμηνευτικές μεταβλητές. Αρκετά συχνά, η μεταβλητή απόκρισης είναι διακριτής φύσεως λαμβάνοντας δύο ή περισσότερες πιθανές τιμές. Το μοντέλο της λογιστικής παλινδρόμησης είναι το πιο συχνά χρησιμοποιούμενο μοντέλο παλινδρόμησης για την ανάλυση των δεδομένων αυτών.

Αυτό που διακρίνει ένα μοντέλο λογιστικής παλινδρόμησης από το μοντέλο γραμμικής παλινδρόμησης είναι ότι η μεταβλητή απόκρισης στη λογιστική παλινδρόμηση είναι δυαδικής μορφής. Αυτή η διαφορά μεταξύ της λογιστικής και της γραμμικής παλινδρόμησης αντικατοπτρίζεται τόσο στη μορφή του μοντέλου όσο και στις υποθέσεις του (Hosmer et al, 2013).

Στο παρόν κεφάλαιο παρουσιάζεται η γενική μεθοδολογία διαχείρισης δεδομένων δυαδικής μορφής, ενώ κατά την πρακτική της εφαρμογή, παρουσιάζεται το μοντέλο το οποίο κατασκευάστηκε προκειμένου να ερμηνεύσει την πιθανότητα ασθένειας κάποιου ατόμου χρησιμοποιώντας ένα σετ επεξηγηματικών μεταβλητών. Σημαντικό μέρος των παρακάτω προέρχονται από τις πανεπιστημιακές σημειώσεις του κ.Πολίτη στα «Γενικευμένα Γραμμικά Μοντέλα» (2014) καθώς και από τον ηλεκτρονικό ιστότοπο www.Learn-Biostatistics.gr.

6.2 Θεωρητικό υπόβαθρο

6.2.1 Εισαγωγή στα Γενικευμένα Γραμμικά Μοντέλα (ΓΓΜ)

Το κλασικό μοντέλο παλινδρόμησης αντικατοπτρίζει τη γραμμική σχέση ανάμεσα στη μεταβλητή απόκρισης και τις ερμηνευτικές μεταβλητές και περιγράφεται από τη σχέση:

$$Y_i = \beta_0 + \sum_{j=1}^k \beta_j X_{ij} + \varepsilon_i \quad \text{ή ισοδύναμα} \quad E(Y_i) = \beta_0 + \sum_{j=1}^k \beta_j X_{ij} .$$

Βασική υπόθεση του μοντέλου είναι ότι τα σφάλματα ε_i είναι ανεξάρτητα ανά δύο και ακολουθούν κανονική κατανομή με μέση τιμή 0 και διακύμανση σ^2

$(\varepsilon_i \sim N(0, \sigma^2))$, πράγμα που σημαίνει πως και τα Y_i ακολουθούν κανονική κατανομή με σταθερή διακύμανση.

Στην περίπτωση που η μεταβλητή απόκρισης είναι δίτιμη μεταβλητή, η οποία λαμβάνει την τιμή 1 όταν το γεγονός πραγματοποιείται (επιτυχία) και την τιμή 0 όταν το γεγονός δεν πραγματοποιείται (αποτυχία), η υπόθεση της κανονικότητας δεν πληρείται. Για τον λόγο αυτό, αναπτύχθηκε μια ενοποιημένη θεωρία για τα γραμμικά μοντέλα, η οποία παρουσιάστηκε αρχικά από τους Nedler και Wedderburn (1972), που δεν απαιτεί την υπόθεση της κανονικότητας για τη μεταβλητή απόκρισης. Σύμφωνα με αυτή, τα γραμμικά μοντέλα μπορούν να μελετηθούν ενιαία κάτω από την υπόθεση ότι η κατανομή της μεταβλητής απόκρισης ανήκει στην εκθετική οικογένεια κατανομών.

Μια κατανομή πιθανότητας λέμε ότι ανήκει στην εκθετική οικογένεια κατανομών εάν η συνάρτηση πιθανότητας (ή πυκνότητας αν η κατανομή είναι συνεχής) της κατανομής μπορεί να γραφεί στη μορφή:

$$f_Y(y; \theta; \phi) = \exp \left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right]$$

όπου a, b, c είναι τρεις γνωστές συναρτήσεις, ενώ θ, ϕ είναι παράμετροι.

Εάν το ϕ είναι γνωστό, τότε έχουμε την εκθετική οικογένεια με μία παράμετρο και το θ αναφέρεται ως κανονική (canonical) παράμετρος της κατανομής. Εάν το ϕ είναι άγνωστο, τότε μπορούμε να το θεωρήσουμε σαν μία παράμετρο κλίμακας για την κατανομή, οπότε καλείται παράγοντας όχλησης (nuisance factor) της κατανομής.

Όλα τα γενικευμένα γραμμικά μοντέλα αποτελούνται από τρεις συνιστώσες, την τυχαία, τη συστηματική και την συνάρτηση συνδέσμου.

Η τυχαία συνιστώσα προσδιορίζει τη μεταβλητή απόκρισης Y έχοντας υποθέσει μια κατανομή πιθανότητας για αυτή.

Η συστηματική συνιστώσα πρόκειται για μια γραμμική συνάρτηση πρόβλεψης η οποία καθορίζεται από τις ερμηνευτικές μεταβλητές για το μοντέλο. Αν, για παράδειγμα, αυτές συμβολιστούν με X_j , για $j = 1, 2, \dots, k$, τότε, ο γραμμικός τους συνδυασμός γράφεται ως εξής:

$$\eta_i = \beta_0 + \sum_{j=1}^k \beta_j X_{ij}$$

όπου X_{ij} είναι η τιμή της μεταβλητής X_j για την παρατήρηση i .

Τέλος, η συνάρτηση συνδέσμου (link function) προσδιορίζει μια συνάρτηση η οποία συνδέει το στοχαστικό τμήμα του μοντέλου (την αναμενόμενη τιμή της μεταβλητής απόκρισης, έστω $\mu_i = E(Y_i)$), με το μη στοχαστικό τμήμα (γραμμικός συνδυασμός των ερμηνευτικών μεταβλητών X_j , για $j = 1, 2, \dots, k$). Η συνάρτηση αυτή, έστω $g(\cdot)$, σχετίζει το μ_i με τη συνάρτηση γραμμικής πρόβλεψης ως εξής:

$$\eta_i = g(\mu_i) = \beta_0 + \sum_{j=1}^k \beta_j X_{ij}.$$

Η συνάρτηση g θεωρούμε ότι είναι πάντα μια συνάρτηση μονότονη και διαφορίσιμη.

Μια ειδική περίπτωση συνάρτησης σύνδεσης ορίζεται από την αντίστροφη συνάρτηση της παραγώγου της b , σύμφωνα με τη σχέση:

$$g = (b')^{-1}.$$

Στην περίπτωση αυτή η g ονομάζεται κανονική συνάρτηση σύνδεσης (canonical link function).

Πίνακας 6.1: Πίνακας με τις κυριότερες συναρτήσεις σύνδεσης και τις αντίστροφές τους

Link	$\eta_i = g(\mu_i)$	$\mu_i = g^{-1}(\eta_i)$
Identity	μ_i	η_i
Log	$\log_e \mu_i$	\exp^{η_i}
Inverse	μ_i^{-1}	η_i^{-1}
Inverse-square	μ_i^{-2}	$\eta_i^{-1/2}$
Square-root	$\sqrt{\mu_i}$	η_i^2
Logit	$\log \frac{\mu_i}{1 - \mu_i}$	$\frac{1}{1 + e^{-\eta_i}}$
Probit	$\Phi^{-1}(\mu_i)$	$\Phi(\eta_i)$
Log-log	$-\log[-\log(\mu_i)]$	$\exp[-\exp(-\eta_i)]$
Complementary log-log	$\log[-\log(1 - \mu_i)]$	$1 - \exp[-\exp(\eta_i)]$

Θα πρέπει σε αυτό το σημείο να τονίσουμε πως σε ένα ΓΓΜ, η μεταβλητή απόκρισης εισέρχεται στο μοντέλο μέσω της μέσης τιμής :

$$g(\mu_i) = \beta_0 + \sum_{j=1}^k \beta_j X_{ij}$$

Άρα, καμία υπόθεση δεν γίνεται για την κατανομή των σφαλμάτων, παρά μόνο για την κατανομή της Y . Στην πραγματικότητα, τα σφάλματα ε_i δεν χρησιμοποιούνται καθόλου, παρά μόνο τα κατάλοιπα που προκύπτουν από την προσαρμογή του μοντέλου.

Απλούστερο παράδειγμα διακριτής κατανομής από την εκθετική οικογένεια κατανομών είναι η κατανομή Bernoulli. Το κλασσικό μοντέλο παλινδρόμησης για δίτιμα δεδομένα θεωρεί μια γραμμική σχέση ανάμεσα στην πιθανότητα επιτυχίας ($Y = 1$) για κάθε άτομο στο δείγμα και στις επεξηγηματικές μεταβλητές.

Αντίθετα, το λογιστικό μοντέλο παλινδρόμησης χρησιμοποιεί σαν μεταβλητή απόκρισης ένα μετασχηματισμό της μέσης τιμής της μεταβλητής Y

$$\eta_i = g(p_i) = g(E(Y_i)).$$

Η g είναι η συνάρτηση σύνδεσης και απεικονίζει το διάστημα $[0,1]$ στην πραγματική ευθεία έτσι ώστε : $g^{-1}(\eta_i) = p_i$.

Οι συναρτήσεις σύνδεσης που χρησιμοποιούνται για δίτιμα δεδομένα είναι η Logit, η Probit ή η Complementary log-log. Όταν χρησιμοποιούμε σαν συνάρτηση σύνδεσης την συνάρτηση Logit, τότε μιλάμε για ένα μοντέλο λογιστικής παλινδρόμησης.

6.2.2 Λογιστική Παλινδρόμηση

Η Λογιστική Παλινδρόμηση είναι μέλος της οικογένειας των Γενικευμένων Γραμμικών Μοντέλων. Είναι μια τεχνική σχεδιασμένη για την ανάλυση δεδομένων που αφορούν τη μελέτη και την πρόβλεψη των τιμών μιας κατηγορικής εξαρτημένης μεταβλητής με τη χρήση ποσοτικών και ποιοτικών ανεξάρτητων μεταβλητών (Agresti, 2007).

Το μοντέλο της λογιστικής παλινδρόμησης εκφράζεται από τον τύπο:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \sum_{j=1}^k \beta_j X_{ij}$$

από το οποίο εύκολα προκύπτει πως:

$$p_i = \frac{\exp(\beta_0 + \sum_{j=1}^k \beta_j X_{ij})}{1 + \exp(\beta_0 + \sum_{j=1}^k \beta_j X_{ij})}.$$

Προκειμένου να αξιολογήσουμε τη στατιστική σημαντικότητα των παραμέτρων του μοντέλου, συγκρίνουμε την τιμή της κάθε παραμέτρου με το τυπικό της σφάλμα. Το κριτήριο αυτό είναι γνωστό ως κριτήριο του *Wald*.

Οι υποθέσεις που ελέγχονται μέσω του κριτηρίου αυτού είναι οι εξής:

$$H_0 : \beta_i = 0 \text{ με εναλλακτική την } H_1 : \beta_i \neq 0.$$

Η μηδενική υπόθεση αναφέρει ότι η i ανεξάρτητη μεταβλητή δεν ερμηνεύει τον λογάριθμο του λόγου συμπληρωματικών πιθανοτήτων, άρα ουσιαστικά δεν επηρεάζει την πιθανότητα επιτυχίας, ενώ αντίθετα η εναλλακτική υπόθεση αναφέρει ότι η i ανεξάρτητη μεταβλητή τον ερμηνεύει.

Η στατιστική συνάρτηση του ελέγχου είναι η:

$$W = \frac{\hat{\beta}_i}{s(\hat{\beta}_i)}$$

η οποία όταν ισχύει η μηδενική υπόθεση, ακολουθεί ασυμπτωτικά την τυπική κανονική κατανομή $N(0,1)$. Η κρίσιμη περιοχή του ελέγχου είναι η $|W| \geq z_{\alpha/2}$.

Προκειμένου να ελέγξουμε την συνολική στατιστική σημαντικότητα του μοντέλου χρησιμοποιούμε τον λόγο της πιθανοφάνειας του μοντέλου που εκτιμήθηκε προς την πιθανοφάνεια του κορεσμένου μοντέλου. Ως κορεσμένο, ορίζεται το μοντέλο του οποίου ο αριθμός των παραμέτρων ισούται με τον αριθμό των παρατηρήσεων. Το κορεσμένο μοντέλο ταιριάζει τέλεια στα δεδομένα καθώς εξηγεί το 100% της μεταβλητότητας.

Η προσαρμογή του μοντέλου λογιστικής παλινδρόμησης στα δεδομένα ελέγχεται μέσω της απόκλισης (deviance). Η απόκλιση είναι ένα μέτρο της μεταβλητότητας που μένει ανερμήνευτη από το μοντέλο της λογιστικής παλινδρόμησης και ισούται με:

$$D = -2 \log \frac{L(M_M)}{L(M_S)} = -2(l(M_M) - l(M_S)).$$

όπου $L(M_S)$ είναι η μέγιστη πιθανοφάνεια υπό το κορεσμένο μοντέλο (saturated) και $L(M_M)$ είναι η μέγιστη πιθανοφάνεια υπό το μοντέλο του οποίου την προσαρμογή θέλουμε να ελέγξουμε. Στην παραπάνω σχέση, το σύμβολο l δηλώνει τον φυσικό λογάριθμο της πιθανοφάνειας.

Διαισθητικά, όσο πιο μικρή είναι η απόκλιση ενός μοντέλου, τόσο πιο κοντά είναι στο κορεσμένο, και αυτό είναι ένδειξη καλής προσαρμογής. Ωστόσο, η κατανομή της ποσότητας D δεν είναι γνωστή.

Σύμφωνα με τον έλεγχο πιθανοφανειών, όμως, εάν θεωρήσουμε ένα μοντέλο M_1 του οποίου οι μεταβλητές είναι υποσύνολο ενός άλλου μοντέλου M_2 με αποκλίσεις D_1 και D_2 , αντίστοιχα, καθώς και βαθμούς ελευθερίας df_1 και df_2 , αντίστοιχα, τότε η ποσότητα:

$$D_1 - D_2 = -2 \left[\log \frac{L_{\text{ReducedModel}}}{L_{\text{FullModel}}} \right] = -2 \left[l_{\text{ReducedModel}} - l_{\text{FullModel}} \right]$$

ακολουθεί προσεγγιστικά την κατανομή χ_p^2 , όπου $p = df_1 - df_2$.

Οι υποθέσεις που ελέγχονται μέσω του ελέγχου αυτού είναι οι εξής:

H_0 : Τα μοντέλα M_1 και M_2 δεν διαφέρουν, με εναλλακτική την H_1 : Τα μοντέλα M_1 και M_2 είναι διαφορετικά.

6.2.2.1 Επιλογή βέλτιστου μοντέλου

Υπάρχουν αρκετές περιπτώσεις στις οποίες, ενώ διαθέτουμε ένα μεγάλο σύνολο ανεξάρτητων μεταβλητών για την πρόβλεψη μιας μεταβλητής απόκρισης Y (εξαρτημένης μεταβλητής) θα μας ενδιέφερε η επιλογή ενός μικρού υποσυνόλου από αυτές για τη δημιουργία ενός αποτελεσματικού μοντέλου πρόβλεψης. Για το λόγο αυτό, έχει ιδιαίτερα μεγάλη πρακτική σημασία, να αναζητούμε ένα μικρό υποσύνολο των μεταβλητών που πιστεύουμε ότι επηρεάζουν την τιμή της μεταβλητής απόκρισης Y το οποίο όμως να μπορεί να την προβλέψει με αρκετή ακρίβεια.

Για τον σκοπό αυτό έχουν προκύψει κάποιες επαναληπτικές διαδικασίες (αλγόριθμοι) οι οποίες επιλέγουν με «ακολουθιακό» τρόπο τις χρήσιμες ανεξάρτητες μεταβλητές για την πρόβλεψη μιας μεταβλητής απόκρισης Y ή απορρίπτουν

διαδοχικά τις μη σημαντικές ανεξάρτητες μεταβλητές ή τέλος συνδυάζουν και τα δύο αυτά βήματα. Έτσι δημιουργείται μια αλληλουχία γραμμικών μοντέλων καθένα από τα οποία διαφέρει από το προηγούμενό του ως προς μια ανεξάρτητη μεταβλητή, μέχρις ότου να ικανοποιηθεί ένα κριτήριο τερματισμού. Το τελευταίο μοντέλο της αλληλουχίας που προκύπτει θεωρείται ως το πλέον κατάλληλο για την πρόβλεψη της μεταβλητής απόκρισης Y .

Χρήσιμο εργαλείο για την επιλογή του βέλτιστου μοντέλου περιγραφής της μεταβλητής απόκρισης από τις ανεξάρτητες μεταβλητές είναι το κριτήριο AIC .

Το κριτήριο πληροφορίας του Akaike (AIC) είναι ένα μέτρο της σχετικής ποιότητας ενός στατιστικού μοντέλου για ένα δεδομένο σύνολο δεδομένων. Ως εκ τούτου, το AIC είναι ένα μέσο για την επιλογή του βέλτιστου μοντέλου. Ο τύπος υπολογισμού του δίνεται από τη σχέση:

$$AIC = 2k - 2\ln(L)$$

όπου k είναι ο αριθμός των παραμέτρων του μοντέλου και L η μέγιστη πιθανοφάνεια του εκάστοτε μοντέλου.

6.3 Εφαρμογή στα δεδομένα

Στην παρούσα παράγραφο θα επιχειρήσουμε να κατασκευάσουμε ένα μοντέλο που να περιγράφει σε πρώτο στάδιο την πιθανότητα να νοσήσει κάποιος δεδομένων κάποιων παραγόντων. Σε δεύτερο στάδιο, θα επιχειρήσουμε να κατασκευάσουμε ένα δεύτερο μοντέλο που θα αφορά μόνο τους ασθενείς και την πιθανότητα να εμφανίσει κάποιος νεανικής μορφής Παρκινσονισμό, δεδομένων των ίδιων παραγόντων.

Εκμεταλλευόμενοι τα αποτελέσματα τόσο του Κεφαλαίου 3 όσο και του Κεφαλαίου 4, αποκλείσαμε εξ αρχής κάποιους παράγοντες από τη μελέτη μας, καθώς έχουμε ήδη αποδείξει πως δεν επηρεάζουν στατιστικώς σημαντικά την εμφάνιση της νόσου. Από την άλλη, παράγοντες με μη ξεκάθαρο ρόλο στην παρουσίαση της ΝΠ που χρήζουν περαιτέρω μελέτης, όπως η έκθεση σε φυτοφάρμακα ή/και ζιζανιοκτόνα συμπεριλήφθησαν στην κατασκευή των προβλεπτικών μας μοντέλων.

Οι παράγοντες καθώς και τα επίπεδα αναφοράς που χρησιμοποιήθηκαν για το εκάστοτε μοντέλο είναι τα εξής:

1. Νοσοκομείο εξέτασης (Hosp): Αττικό
2. Φύλο (Gender) : Άνδρας
3. Θρήσκευμα (Rel) : Χριστιανός Ορθόδοξος

4. Οικογενειακό ιστορικό ΝΠ (FH) : Όχι
5. Κάπνισμα κατά τις τελευταίες 30 ημέρες πριν την εξέταση (Smoke) : Όχι
6. Κατανάλωση καφέ τώρα (Coffee) : Όχι
7. Έκθεση σε ζιζανιοκτόνα ή/και φυτοφάρμακα (Agr) : Όχι

Αφού πρώτα δηλώσαμε τις επεξηγηματικές μας μεταβλητές ως παράγοντες, εξετάζουμε αν τελικά επηρεάζουν την πιθανότητα να νοσήσει ένας εθελοντής από την νόσο του Πάρκινσον. Δεδομένου ότι η σειρά εισαγωγής των μεταβλητών επηρεάζει τη σημαντικότητά τους, ξεκινάμε από το πλήρες μοντέλο, το οποίο περιέχει όλες τις παραπάνω μεταβλητές καθώς και τις επιμέρους αλληλεπιδράσεις τους, προκειμένου να αποφασίσουμε για τη σημαντικότητά τους. Σε κάθε περίπτωση, εξετάζουμε τη μηδενική υπόθεση H_0 : Το μοντέλο δεν διαφέρει σημαντικά από το κορεσμένο, έναντι της εναλλακτικής H_1 : Το μοντέλο διαφέρει σημαντικά από το κορεσμένο. Χρησιμοποιώντας το κριτήριο πληροφορίας του Akaike (AIC), καταλήξαμε στο παρακάτω μοντέλο:

$$\log\left(\frac{p}{1-p}\right) = 2.07 - 1.89Hosp1 - 1.15Gender1 - 0.38Rel1 + 79.7FH1 + 18Smokel - 0.91Coffeel + 18.8Agr1 + 1.05Hosp1 : FH1 - 46.5Gender1 : FH1 - 0.05Rel1 : FH1 - 38.9Hosp1 : Smokel - 39.2Gender1 : Smokel - 62.8FH1 : Smokel + 0.87Hosp1 : Coffeel + 0.61Gender1 : Coffeel - 0.15Rel1 : Coffeel - 19Smokel : Coffeel - 19Hosp1 : Agr1 - 18.7Gender1 : Agr1 - 46.7FH1 : Agr1 - 18.3Smokel : Agr1 - 1.65Hosp1 : Gender1 : Coffeel - 31Rel1 : FH1 : Coffeel + 38.5Hosp1 : Smokel : Coffeel + 38.9Gender1 : Smokel : Coffeel + 19.2Hosp1 : Gender1 : Agr1 + 61.2Gender1 : FH1 : Agr1 + 20.1Hosp1 : Smokel : Agr1$$

Προκειμένου να ελέγξουμε την επάρκεια του μοντέλου αυτού σε σχέση με το κορεσμένο μοντέλο, θα προχωρήσουμε σε έλεγχο της υπόθεσης H_0 : Το reduced model δεν διαφέρει σημαντικά από το κορεσμένο, έναντι της εναλλακτικής H_1 : Το reduced model διαφέρει σημαντικά από το κορεσμένο. Το p-value του ελέγχου με βάση τον λόγο πιθανοφανειών προέκυψε ίσο με 0.33 άρα σε επίπεδο σημαντικότητας 5% δεν απορρίπτουμε την μηδενική υπόθεση και αποφασίζουμε πως το μοντέλο μας είναι επαρκές.

Παρατηρώντας τον πίνακα ανάλυσης διακύμανσης του παραπάνω μοντέλου (Παράρτημα Β.1), βλέπουμε πως υπάρχουν αρκετές αλληλεπιδράσεις οι οποίες είναι στατιστικά μη σημαντικές καθώς λαμβάνουν τιμές μικρότερες από την τιμή της κατανομής χ_p^2 με τους αντίστοιχους βαθμούς ελευθερίας, όπως αυτή ορίστηκε παραπάνω. Ξεκινώντας από αυτή με τη μικρότερη ερμηνευτική τιμή μεταβλητότητας κατά την εισαγωγή της στο μοντέλο, αφαιρούμε διαδοχικά τις μη σημαντικές μεταβλητές.

Ο έλεγχος που πραγματοποιήσαμε σε κάθε περίπτωση ήταν ο έλεγχος της υπόθεσης H_0 : Ο όρος της αλληλεπίδρασης των μεταβλητών δεν είναι στατιστικά σημαντικός, έναντι της εναλλακτικής H_1 : Ο όρος της αλληλεπίδρασης των μεταβλητών είναι στατιστικά σημαντικός.

Ακόμα, παρατηρήσαμε πως υπάρχουν στατιστικά μη σημαντικές αλληλεπιδράσεις δεύτερης τάξης που συμμετείχαν στο παραπάνω μοντέλο, ενώ αλληλεπιδράσεις τρίτης τάξης βρέθηκαν να είναι στατιστικά σημαντικές. Ο λόγος που οι πρώτες έχουν εισαχθεί στο μοντέλο είναι η αρχή του περιθωρίου (*marginality principle*). Σύμφωνα με την αρχή του περιθωρίου, οι κύριες επιδράσεις των παραγόντων με μη μηδενικές αλληλεπιδράσεις δεν θα πρέπει να μην τεστάρονται ή να μην ερμηνεύονται. Αντίστοιχα, αλληλεπιδράσεις που περιέχονται σε αλληλεπιδράσεις ανώτερης τάξης δεν θα πρέπει να μένουν ανερμήνευτες (Martinez, 2013).

Δεδομένου όμως ότι μια αλληλεπίδραση τρίτης τάξεως θα ήταν δύσκολο να ερμηνευτεί αποφασίζουμε να παραλείψουμε τους συντελεστές αυτούς. Το μοντέλο στο οποίο καταλήξαμε είναι πολύ απλούστερο και είναι το εξής:

$$\log\left(\frac{P}{1-p}\right) = 2.16 - 1.70Hosp1 - 0.97Gender1 + 1.90FH1 - 1.25Smokel - 0.61Coffeel - 2.02FH1 : Smokel$$

Το p-value του ως προς τον έλεγχο της επάρκειάς του, με τον έλεγχο του λόγου πιθανοφανειών, ισούται με 0.15, άρα, σε επίπεδο σημαντικότητας 5% δεν απορρίπτουμε την μηδενική υπόθεση και αποφασίζουμε πως το μοντέλο μας είναι επαρκές.

Ο πίνακας ανάλυσης διακύμανσης του παραπάνω μοντέλου είναι ο:

	Df	Deviance Resid.	Df	Resid. Dev
NULL			484	641.84
Hosp	1	49.152	483	592.69
Gender	1	10.625	482	582.07
FH	1	33.201	481	548.87
Smoke	1	27.835	480	521.03
Coffee	1	6.359	479	514.67
FH:Smoke	1	4.616	478	510.05

Σύμφωνα με το μοντέλο αυτό, μπορούμε να εξάγουμε τα κάτωθι συμπεράσματα:

- Η σχετική πιθανότητα να νοσεί κάποιος που εξετάζεται στο νοσοκομείο Σύρου είναι ίση με $\exp(-1.70) = 0.18$ φορές επί την αντίστοιχη πιθανότητα να νοσεί κάποιος που εξετάζεται στο Αττικό νοσοκομείο.
- Η σχετική πιθανότητα να νοσεί μία γυναίκα είναι ίση με $\exp(-0.97) = 0.38$ φορές επί την αντίστοιχη πιθανότητα να νοσεί ένας άνδρας.
- Η σχετική πιθανότητα να νοσεί κάποιος που καταναλώνει καφέ είναι ίση με $\exp(-0.61) = 0.54$ φορές επί την αντίστοιχη πιθανότητα να νοσεί κάποιος που δεν καταναλώνει καφέ.
- Η σχετική πιθανότητα να νοσεί κάποιος που έχει οικογενειακό ιστορικό ΝΠ και καπνίζει είναι ίση με $\exp(1.90 - 1.25 - 2.02) = \exp(-1.37) = 0.25$ φορές επί την αντίστοιχη πιθανότητα να νοσεί ένας κάποιος που δεν έχει οικογενειακό ιστορικό ΝΠ και δεν καπνίζει.

Στο σημείο αυτό, θα ήταν χρήσιμο να δώσουμε την ερμηνεία της αλληλεπίδρασης δεύτερης τάξης. Δυο ερμηνευτικές μεταβλητές καθορίζουν την μεταβλητή απόκρισης με τέτοιο τρόπο ώστε η μερική επίδραση της μιας να εξαρτάται από το επίπεδο της άλλης.

Θα προσπαθήσουμε να ερμηνεύσουμε τον παραπάνω όρο αλληλεπίδρασης. Για τον σκοπό αυτό, αρχικά κατασκευάζουμε έναν πίνακα τριπλής εισόδου, με τις δύο μεταβλητές που αλληλεπιδρούν καθώς και την μεταβλητή απόκρισης. Ο πίνακας που λάβαμε είναι ο εξής:

Πίνακας 6.2: Πίνακας τριπλής εισόδου για την ερμηνεία του όρου της αλληλεπίδρασης

		PD	
FH	Smoke	Yes	No
Yes	Yes	5	4
	No	76	9
No	Yes	27	41
	No	223	131

Θα εξετάσουμε τη συσχέτιση καπνίσματος και εμφάνισης της νόσου Πάρκινσον για κάθε επίπεδο της μεταβλητής οικογενειακό ιστορικό για την ΝΠ. Παίρνουμε τους πίνακες δύο διαστάσεων:

Πίνακας 6.3: Πίνακας διπλής εισόδου για τα άτομα α) με οικογενειακό ιστορικό ΝΠ, β) χωρίς οικογενειακό ιστορικό ΝΠ

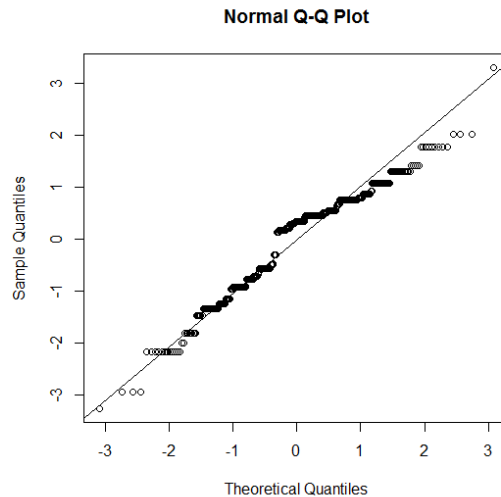
FH=Yes	PD	
Smoke	Yes	No
Yes	5	4
No	76	9

FH=No	PD	
Smoke	Yes	No
Yes	27	41
No	223	131

Για αυτούς με FH=Yes, το σχετικό ποσοστό (odds) Παρκινσονικών ισούται με $\frac{76}{9} = 8.44$ στους μη καπνιστές και είναι σχεδόν 7πλάσιο σε σχέση με το αντίστοιχο ποσοστό καπνιστών, $\frac{5}{4} = 1.25$.

Αντίστοιχα, για αυτούς με FH=No, το αντίστοιχο odds για μη καπνιστές είναι 2.5 φορές μεγαλύτερο.

Τέλος, προχωρούμε σε έλεγχο κανονικότητας των καταλοίπων του μοντέλου:



Διάγραμμα 6.1: Διάγραμμα με τα κατάλοιπα του Pearson

Τα κατάλοιπα φαίνεται να προσεγγίζουν ικανοποιητικά την ευθεία $y=x$. Οπότε μπορούμε να ισχυριστούμε πως ακολουθούν την κανονική κατανομή.

Θα προχωρήσουμε τώρα στην κατασκευή του δεύτερου μοντέλου, το οποίο θα αφορά μόνο τους ασθενείς και την πιθανότητα να εμφανίσει κάποιος νεανικής μορφής Παρκινσονισμό, δεδομένων των παραπάνω παραγόντων. Δύο σημεία τα οποία επιλέξαμε να διαφοροποιήσουμε είναι το κάπνισμα και η κατανάλωση καφέ κατά το διάστημα κοντά στην ημερομηνία εξέτασης. Σε αυτήν την περίπτωση χρησιμοποιήσαμε τα παρακάτω επίπεδα αναφοράς:

5. Κάπνισμα πάνω από 100 τσιγάρων (SmokePast) : Όχι
6. Κατανάλωση καφέ παλαιότερα (CoffeePast) : Όχι

Ξεκινώντας από το πλήρες μοντέλο, εξετάζουμε τη μηδενική υπόθεση H_0 : Το μοντέλο δεν διαφέρει σημαντικά από το κορεσμένο, έναντι της εναλλακτικής H_1 : Το μοντέλο διαφέρει σημαντικά από το κορεσμένο. Χρησιμοποιώντας το κριτήριο πληροφορίας του Akaike (AIC), καταλήξαμε στο παρακάτω μοντέλο:

$$\log\left(\frac{P}{1-p}\right) = -2.40 - 33.87Hosp1 - 0.02Gender1 - 54.16Rel1 + 3.52FH1 + 1.53SmokePast1 - 0.40CoffeePast1 + 0.38Agr1 + 37.07Gender1 : Rel1 - 3.29Gender1 : FH1 + 35.28Rel1 : FH1 - 2.84Hosp1 : SmokePast1 + 2.07Gender1 : SmokePast1 - 3.12FH1 : SmokePast1 + 34.76Hosp1 : CoffeePast1 - 35.13FH1 : Agr1$$

Προκειμένου να ελέγξουμε την επάρκεια του μοντέλου αυτού σε σχέση με το κορεσμένο μοντέλο, θα προχωρήσουμε σε έλεγχο της υπόθεσης H_0 : Το reduced

model δεν διαφέρει σημαντικά από το κορεσμένο, έναντι της εναλλακτικής H_1 : Το reduced model διαφέρει σημαντικά από το κορεσμένο με τον έλεγχο του λόγου πιθανοφαιών. Το p-value του ελέγχου προέκυψε ίσο με 0.99 άρα σε επίπεδο σημαντικότητας 5% δεν απορρίπτουμε την μηδενική υπόθεση και αποφασίζουμε πως το μοντέλο μας είναι επαρκές.

Ο πίνακας ανάλυσης διακύμανσης του παραπάνω μοντέλου διαμορφώνεται ως εξής:

	Df	Deviance Resid.	Df	Resid. Dev
NULL			173	139.61
Hosp	1	8.0798	172	131.53
Gender	1	0.0251	171	131.51
Rel	1	0.3376	170	131.17
FH	1	1.9543	169	129.22
SmokePast	1	3.9479	168	125.27
CoffeePast	1	0.7200	167	124.55
Agr	1	0.2459	166	124.30
Gender:Rel	1	1.6614	165	122.64
Gender:FH	1	0.0083	164	122.63
Rel:FH	1	1.9109	163	120.72
Hosp:SmokePast	1	1.3489	162	119.37
Gender:SmokePast	1	3.2392	161	116.14
FH:SmokePast	1	2.2988	160	113.84
Hosp:CoffeePast	1	6.6823	159	107.15
FH:Agr	1	3.7680	158	103.39

Παρατηρούμε πως υπάρχουν παράγοντες οι οποίοι είναι στατιστικά μη σημαντικοί καθώς λαμβάνουν τιμές μικρότερες από την τιμή της κατανομής χ_p^2 με τους αντίστοιχους βαθμούς ελευθερίας. Ξεκινώντας από αυτόν με τη μικρότερη ερμηνευτική τιμή μεταβλητότητας κατά την εισαγωγή του στο μοντέλο, αφαιρούμε διαδοχικά τις μη σημαντικές μεταβλητές.

Ο έλεγχος που πραγματοποιήσαμε σε κάθε περίπτωση ήταν ο έλεγχος της υπόθεσης H_0 : Ο όρος της αλληλεπίδρασης των μεταβλητών δεν είναι στατιστικά σημαντικός, έναντι της εναλλακτικής H_1 : Ο όρος της αλληλεπίδρασης των μεταβλητών είναι στατιστικά σημαντικός. Το μοντέλο στο οποίο καταλήξαμε είναι το εξής:

$$\log\left(\frac{p}{1-p}\right) = -1.86 - 1.25Hosp1 + 0.84SmokePast$$

Το μοντέλο αυτό φαίνεται να επιβεβαιώνει τα ευρήματα του Κεφαλαίου 4 και του Κεφαλαίου 5. Το κάπνισμα φαίνεται να είναι παράγοντας κινδύνου σε αυτήν την περίπτωση, ενώ το νοσοκομείο εξέτασης φαίνεται για μία ακόμη φορά να είναι καθοριστικός παράγοντας διαφοροποίησης μεταξύ των ασθενών.

Το p-value του ελέγχου του λόγου πιθανοφανειών για την επάρκεια του μοντέλου ισούται με 0.99, άρα, σε επίπεδο σημαντικότητας 5% δεν απορρίπτουμε την μηδενική υπόθεση και αποφασίζουμε πως το μοντέλο μας είναι επαρκές.

Ο πίνακας ανάλυσης διακύμανσης του παραπάνω μοντέλου είναι ο:

	Df	Deviance Resid.	Df	Resid. Dev
NULL			261	220.46
Hosp	1	9.9905	260	210.47
SmokePast	1	0.0251	259	204.98

Σύμφωνα με το μοντέλο αυτό, μπορούμε να εξάγουμε τα εξής συμπεράσματα:

- Η σχετική πιθανότητα να εμφανίσει νεανικής μορφής Παρκινσονισμό ένας ασθενής που εξετάζεται στο νοσοκομείο Σύρου είναι ίση με $\exp(0.84) = 2.32$ φορές επί την αντίστοιχη πιθανότητα ενός που εξετάζεται στο Αττικό νοσοκομείο.
- Η σχετική πιθανότητα να εμφανίσει νεανικής μορφής Παρκινσονισμό ένας ασθενής που έχει καπνίσει πάνω από 100 τσιγάρα στη ζωή του είναι ίση με $\exp(0.84) = 2.32$ φορές επί την αντίστοιχη πιθανότητα ενός ασθενούς που δεν καπνίζει.

ΚΕΦΑΛΑΙΟ 7: ΣΥΜΠΕΡΑΣΜΑΤΑ

Από το δείγμα των 568 ατόμων που έχουν καταγραφεί στην Ελληνική Βιοτράπεζα για την Νόσο Πάρκινσον και με βάση το μέρος του ερωτηματολογίου που αναλύθηκε, τα κυριότερα ευρήματα είναι τα εξής:

Επαληθεύσαμε πολλά από τα ευρήματα της διεθνούς βιβλιογραφίας όπως:

- ✓ Οι άνδρες νοσούν περισσότερο συχνά σε σχέση με τις γυναίκες, καθιστώντας τον παράγοντα φύλο έναν από τους σημαντικότερους παράγοντες εμφάνισης της νόσου Πάρκινσον.
- ✓ Η νόσος Πάρκινσον δεν οφείλεται γενικά σε κληρονομικούς παράγοντες. Παρόλα αυτά όμως, το οικογενειακό ιστορικό για την νόσο είναι ένας παράγοντας που επηρεάζει σημαντικά την ανάπτυξη της ασθένειας. Αντίθετα, η συγγένεια εξ αίματος των γονέων κρίθηκε ασήμαντη.
- ✓ Δεν έχει αποδειχθεί ότι τα φυτοφάρμακα συμβάλλουν άμεσα στην ανάπτυξη της ΝΠ ενώ πολλές μελέτες υποστηρίζουν πως δεν σχετίζεται με κανένα τρόπο. Στο δικό μας δείγμα, ο παράγοντας αυτός αποδείχθηκε ως στατιστικά μη σημαντικός.
- ✓ Το κάπνισμα και η κατανάλωση καφεΐνης αποδείχθηκαν να είναι προστατευτικοί παράγοντες για την ασθένεια, με το κάπνισμα να μειώνει την σχετική πιθανότητα εμφάνισής της κατά 40% και η κατανάλωση καφεΐνης κατά 22%.
- ✓ Πρωταρχικό σύμπτωμα που παρατηρείται τους ασθενείς είναι ο τρόμος (τρέμουλο) όμως ένας στους τρεις ασθενείς παρουσιάζει κάποιο από τα υπόλοιπα συμπτώματα σαν αρχικό.

Κατά την συσταδοποίηση του δείγματος σε επιμέρους ομάδες τόσο με ιεραρχικές όσο και με μη ιεραρχικές μεθόδους, πρωταρχικός παράγοντας διαφοροποίησης αποδείχθηκε το νοσοκομείο εξέτασης των ασθενών. Συγκεκριμένα, στατιστικώς σημαντικές διαφορές εντοπίστηκαν μεταξύ των ασθενών που εξετάστηκαν στο Αττικό νοσοκομείο και το νοσοκομείο Σύρου ως προς τα παρακάτω:

- ✓ Κλίμακα UPDRS I-III. Οι μεγαλύτερες αξιολογήσεις των ασθενών του νοσοκομείου της Σύρου υποδεικνύουν πως οι ασθενείς εκεί πηγαίνουν να εξεταστούν όταν τα συμπτώματα της νόσου είναι περισσότερο εμφανή.
- ✓ Παράλληλα, διαφέρει και η ηλικία έναρξης της νόσου με τους ασθενείς της σύρου να νοσούν 6.5 έτη αργότερα σε σχέση με τους ασθενείς του Αττικού.
- ✓ Τη μέση διάρκεια χρόνων καπνίσματος. Οι ασθενείς του Αττικού νοσοκομείου φαίνεται να σταματούν το κάπνισμα μία δετία νωρίτερα από αυτούς του νοσοκομείου Σύρου.

Ενδιαφέροντα ήταν και τα ευρήματα για τον νεανικής μορφής Παρκινσονισμό, καθώς δεν φαίνεται να έχει αναφερθεί κάτι αντίστοιχο στη διεθνή βιβλιογραφία:

- ✓ Μοναδικός παράγοντας που φαίνεται να επηρεάζει την εμφάνιση νεανικής μορφής Παρκινσονισμού είναι το κάπνισμα, και μάλιστα αυτήν τη φορά είναι παράγοντας κινδύνου για την ασθένεια.

Τέλος, συνοψίσαμε τα παραπάνω αποτελέσματα κατασκευάζοντας ένα προβλεπτικό μοντέλο με τη μέθοδο της λογιστικής παλινδρόμησης για την εκτίμηση της σχετικής πιθανότητας να εμφανίσει ένα άτομο την νόσο Πάρκινσον δεδομένων των παραγόντων που μελετήθηκαν.

ΠΑΡΑΡΤΗΜΑ

A.1 Στον πίνακα που ακολουθεί παρουσιάζεται η κωδικοποίηση που χρησιμοποιήθηκε από το στατιστικό πακέτο R για την παραγωγή των όψεων Chernoff που αντιστοιχούν στις 2 συστάδες.

Modified Item	Variable
height of face	Gender
width of face	FamilyHistory
structure of face	SmokedLast30Days
height of mouth	CoffeeNow
width of mouth	Agrochemichals
smiling	EarlyPD
height of eyes	Religion
width of eyes	City
height of hair	Gender
width of hair	FamilyHistory
style of hair	SmokedLast30Days
height of nose	CoffeeNow
width of nose	Agrochemichals
width of ear	EarlyPD
height of ear	Religion

A.2 Στον πίνακα που ακολουθεί παρουσιάζεται η κωδικοποίηση που χρησιμοποιήθηκε από το στατιστικό πακέτο R για την παραγωγή των όψεων Chernoff που αντιστοιχούν στις 4 συστάδες.

Modified Item	Variable
height of face	Gender
width of face	FamilyHistory
structure of face	SmokedLast30Days
height of mouth	CoffeeNow
width of mouth	Agrochemichals
smiling	EarlyPD
height of eyes	Religion
width of eyes	Gender
height of hair	FamilyHistory
width of hair	SmokedLast30Days
style of hair	CoffeeNow
height of nose	Agrochemichals

width of nose	EarlyPD
width of ear	Religion
height of ear	Gender

Κώδικας στην R:

```
library(aplpack)
```

```
faces(Parkinson[1:20,2:9],labels=Parkinson$Group)
```

B.1 Στον πίνακα που ακολουθεί παρουσιάζεται ο πίνακας ανάλυσης διασποράς του κορεσμένου μοντέλου που χρησιμοποιήθηκε για την κατασκευή του προβλεπτικού μοντέλου μέσω της λογιστικής παλινδρόμησης.

Analysis of Deviance Table

Model: binomial, link: logit

Response: PD

Terms added sequentially (first to last)

	Df	Deviance Resid.	Df	Resid. Dev
NULL			475	629.30
Hosp	1	46.247	474	583.05
Gender	1	11.433	473	571.62
Rel	1	1.190	472	570.43
FH	1	32.739	471	537.69
Smoke	1	27.732	470	509.95
Coffee	1	6.587	469	503.37
Agr	1	0.562	468	502.81
Hosp:Gender	1	1.121	467	501.68
Hosp:FH	1	1.419	466	500.27
Gender:FH	1	0.240	465	500.03
Rel:FH	1	0.427	464	499.60
Hosp:Smoke	1	0.014	463	499.58
Gender:Smoke	1	2.424	462	497.16

FH:Smoke	1	11.221	461	485.94
Hosp:Coffee	1	0.586	460	485.35
Gender:Coffee	1	0.135	459	485.22
Rel:Coffee	1	0.282	458	484.94
FH:Coffee	1	1.362	457	483.57
Smoke:Coffee	1	0.587	456	482.99
Hosp:Agr	1	1.312	455	481.67
Gender:Agr	1	0.306	454	481.37
FH:Agr	1	1.019	453	480.35
Smoke:Agr	1	0.002	452	480.35
Hosp:Gender:Coffee	1	5.352	451	474.99
Rel:FH:Coffee	1	0.970	450	474.02
Hosp:Smoke:Coffee	1	0.395	449	473.63
Gender:Smoke:Coffee	1	4.634	448	469.00
Hosp:Gender:Agr	1	3.062	447	465.93
Hosp:Smoke:Agr	1	4.821	446	461.11

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑΣ

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

- Αντζουλάκος, Δ. (2012). *Εισαγωγή στο R*, Πανεπιστημιακές Σημειώσεις για το ΠΜΣ στην «Εφαρμοσμένη Στατιστική», Πανεπιστήμιο Πειραιώς.
- Καρλής, Δ. (2005). *Πολυμεταβλητή Στατιστική Ανάλυση*, Εκδόσεις Σταμούλης.
- Κούτρας, Μ. (2013). *Ανάλυση κατά συστάδες*, Πανεπιστημιακές Σημειώσεις για το ΠΜΣ στην «Εφαρμοσμένη Στατιστική», Πανεπιστήμιο Πειραιώς.
- Πολίτης, Κ. (2014). *Γενικευμένα Γραμμικά Μοντέλα*, Πανεπιστημιακές Σημειώσεις για το ΠΜΣ στην «Εφαρμοσμένη Στατιστική», Πανεπιστήμιο Πειραιώς.
- Σαχλάς, Αθ. και Μπερσίμης, Σ. (2013). *Βιοστατιστική και στατιστικές μέθοδοι στην επιδημιολογία*, Πανεπιστημιακές Σημειώσεις για το ΠΜΣ στην «Εφαρμοσμένη Στατιστική», Πανεπιστήμιο Πειραιώς.

Ξένη

- Agresti, A. (2007). *An Introduction to Categorical Data Analysis*, Second Edition, John Wiley & Sons, Inc., Publication.
- Brown, T.P., Rumsby, P.C., Capleton, A.C., Rushton, L. and Levy, L.S. (2006). Pesticides and Parkinson's disease-is there a link?, *Environ Health Perspect.* 114(2): 156-64.
- Chernoff, H. (1973). The Use of Faces to Represent points in k-Dimensional Space Graphically, *Journal of the American Statistical Association*, 361.
- Fahn, S., Elton, R.L., (1987). Members of the UPDRS Development Committee. The Unified Parkinson's Disease Rating Scale. In Fahn, S., Marsden, C.D., Calne, D.B. and Goldstein, M., eds: *Recent Developments in Parkinson's Disease*, Vol. 2. Florham Park, NJ: Macmillan Healthcare Information, pp. 153-163, 293-304.
- Galanaud, J.P., Elbaz, A., Clavel, J., Vidal, J.S., Corrèze, J.R., Alperovitch, A. and Tzourio, C. (2005). Cigarette smoking and Parkinson's disease: a case-control study in a population characterized by a high prevalence of pesticide exposure, *Mov Disord.*, 20: 181-9.
- Hermansen, K., Krogholm, K.S., Bech, B.H., Dragsted, L.O., Hyldstrup, L., Jørgensen, K., Larsen, M.L. and Tjønneland, A.M. (2012). Coffee can protect against disease. *Ugeskr Laeger*, 174(39): 2293-7.
- Hosmer, D.W., Lemeshow, S. and Sturdivant, R.X. (2013). *Applied Logistic Regression*, 3rd Edition, John Wiley & Sons, Inc.
- Martinez, H.R. (2013). *Analysing interactions of fitted models*.

- Miller, I.N. and Cronin-Golomb, A. (2010). Gender Differences in Parkinson's disease: Clinical characteristics and Cognition, *Mov Disord.*, 25(16): 2695–2703.
- Nuti, A., Ceravolo, R., Dell'Agnello, G., Gambaccini, G., Bellini, G., Kiferle, L., Rossi, C., Logi, C. and Bonuccelli, U. (2004). Environmental factors and Parkinson's disease: a case-control study in the Tuscany region of Italy, *Parkinsonism Relat Disord.*, 10: 481–5.
- Pezzoli, G. and Cereda, E. (2013). Exposure to pesticides or solvents and risk of Parkinson disease, *Neurology*, 80(22): 2035-2041.
- Postuma, R.B., Lang, A.E., Munhoz, R.P., Charland, K., Pelletier, A., Moscovich, M., Filla, L., Zanatta, D., Rios Romenets, S., Altman, R., Chuang, R. and Shah, B. (2012). Caffeine for treatment of Parkinson disease: a randomized controlled trial, *Neurology*, 79(7): 651-8.
- Qi, H. and Li, S. (2014). Dose-response meta-analysis on coffee, tea and caffeine consumption with risk of Parkinson's disease. *Geriatr Gerontol Int.* 430-9.
- Quik, M. (2004). Smoking, nicotine and Parkinson's disease. *Trends Neurosci*, 27: 561-568.
- Rousseeuw, P.J. (1987). Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis, *Computational and Applied Mathematics*, 20: 53–65.
- Scott, W.K., Zhang, F., Stajich, J.M., Scott, B.L., Stacy, M.A. and Vance, J.M. (2005). Family-based case-control study of cigarette smoking and Parkinson disease, *Neurology*, 64(3): 442-7.
- Sellbach, A.N., Boyle, R.S., Silburn, P.A. and Mellick, G.D. (2006). Parkinson's disease and family history, *Parkinsonism Relat Disord.*, 12(7): 399–409.
- Taylor, C.A., Saint-Hilaire, M.H., Cupples, L.A., Thomas, C.A., Burchard, A.E., Feldman, R.G. and Myers, R.H. (1999). Environmental, medical, and family history risk factors for Parkinson's disease: a New England-based case control study. *Am J Med Genet*, 88: 742–9.

Ηλεκτρονικές πηγές

1. http://en.wikipedia.org/wiki/Parkinson's_disease
2. <http://www.epda.eu.com/en/#>
3. <http://www.parkinson.gr/>
4. <http://www.parkinsonportal.gr/>
5. www.Learn-Biostatistics.gr
6. http://www2.stat.unibo.it/montanari/Didattica/Multivariate/CA_lab.pdf
7. <http://www.stat.berkelev.edu/classes/s133/Cluster2a.html>
8. http://www.sagepub.com/upm-data/21121_Chapter_15.pdf

9. http://www.orpha.net/consor/cgi-bin/Disease_Search_Simple.php?lng=EN&diseaseGroup=parkinson
10. <https://www.google.gr/search?q=chernoff+faces&client=firefox-a&hs=uKU&rls=org.mozilla:el:official&channel=nts&biw=1366&bih=634&tbm=isch&tbo=u&source=univ&sa=X&ei=SURkVMfaMoXVOJ6dgNAG&ved=0CCwQsAQ>

Πανεπιστήμιο Πειραιώς