Π Α Ν Ε Π Ι Σ Τ Η Μ Ι Ο Π Ε Ι Ρ Α Ι Ω Σ

Σχολή Χρηματοοικονομικής και Στατιστικής

Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ **ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

ΨΥΧΟΚΟΙΝΩΝΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΑΣΘΕΝΩΝ ΜΕ ΕΠΙΛΗΨΙΑ ΣΤΗΝ ΕΛΛΑΔΑ

Αλέξανδρος Καμπόλης

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
Ιούλιος 2021

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:

- Αναπληρωτής Καθηγητής Πολίτης Κωνσταντίνος (Επιβλέπων)
- Αναπληρωτής Καθηγητής Τζαβελάς Γεώργιος
- Αναπληρωτής Καθηγητής Γκατζώνης Στέργιος-Στυλιανός

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

UNIVERSITY OF PIRAEUS

School of Finance and Statistics

Department of Statistics and Insurance Science

POSTGRADUATE PROGRAM IN
APPLIED STATISTICS

PSYCHOLOGICAL AND SOCIAL CHARACTERISTICS OF EPILEPTIC PATIENTS IN GREECE

By

Alexandros Kampolis

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
July 2021

Στη μητέρα μου, Σοφία, η οποία
έφυγε πολύ νωρίς και άδικα από κοντά μας

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, Αναπληρωτή Καθηγητή του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς κ. Πολίτη Κωνσταντίνο, για τις χρήσιμες συμβουλές και την καθοδήγησή του κατά την διεξαγωγή της διπλωματικής μου εργασίας. Τον ευχαριστώ πολύ που με εμπιστεύτηκε να φέρω εις πέρας την εργασία αυτή και με υπομονή με βοήθησε να την ολοκληρώσω.

Παράλληλα, θα ήθελα να ευχαριστήσω τον κ. Τζαβελά Γεώργιο, Αναπληρωτή Καθηγητή του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς, καθώς και τον κ. Γκατζώνη Στέργιο-Στυλιανό, Αναπληρωτή Καθηγητή του Τμήματος Ιατρικής του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών, για την συμμετοχή τους στην τριμελή επιτροπή.

Δήλωση συγγραφέα

Έχω διαβάσει και κατανοήσει τους κανόνες του ΠΜΣ που περιέχονται στον Οδηγό Συγγραφής ΔΕ και ιδιαίτερα όσα συνιστούν λογοκλοπή. Δηλώνω ότι η παρούσα διπλωματική αποτελεί προϊόν αποκλειστικά δικής μου προσπάθειας, υπό την καθοδήγηση του επιβλέποντος καθηγητή, ενώ για όλες τις πηγές που χρησιμοποιήθηκαν περιλαμβάνονται οι αντίστοιχες αναφορές.

Περίληψη

Στην παρούσα διπλωματική εργασία αναλύθηκε ένα μέρος ενός ερωτηματολογίου που απαντήθηκε από 96 επιληπτικούς ασθενείς στο νοσοκομείο «Ευαγγελισμός», με στόχο τον εντοπισμό των ψυχοκοινωνικών και κλινικών χαρακτηριστικών των ασθενών με επιληψία στην Ελλάδα, την εύρεση των κυριότερων διαφορών που έχουν παρουσιαστεί σε σχέση με μια αντίστοιχη μελέτη (Νικολάκης, 2010) που αφορούσε δεδομένα της περιόδου 2008-2009 και την πρόβλεψη της εμφάνισης κρίσεων μέσα στο επόμενο δίμηνο από την ημερομηνία συμπλήρωσης του ερωτηματολογίου. Οι ερωτήσεις του ερωτηματολογίου που αναλύθηκαν καλύπτουν την κοινωνική ζωή, τις δεξιότητες, τις διαπροσωπικές σχέσεις και τα δημογραφικά, κοινωνικά, ψυχολογικά και κλινικά χαρακτηριστικά των επιληπτικών ασθενών.

Οι στατιστικές μέθοδοι που χρησιμοποιήθηκαν στην παρούσα εργασία ήταν η μονοδιάστατη περιγραφική ανάλυση, οι έλεγχοι υποθέσεων για την διαφορά των μέσων τιμών δύο ανεξάρτητων πληθυσμών, οι έλεγχοι υποθέσεων για την διαφορά των ποσοστών δύο ανεξάρτητων πληθυσμών, οι έλεγχοι ανεξαρτησίας X^2, η λογιστική παλινδρόμηση, η κατασκευή δέντρου αποφάσεων με τον αλγόριθμο Conditional Inference Decision Trees και η εξόρυξη κανόνων συσχετίσεων με τον αλγόριθμο Apriori.

Τα αποτελέσματα της ανάλυσης έδειξαν ότι η επιληψία δεν αποτελεί εμπόδιο στη ζωή των Ελλήνων ασθενών, ότι υπάρχει σημαντική βελτίωση στην ποιότητα ζωής τους την τελευταία δεκαετία, καθώς και ότι ο κυριότερος παράγοντας που επηρεάζει την εμφάνιση κρίσεων μέσα στο επόμενο δίμηνο είναι η ύπαρξη κρίσεων τον περασμένο χρόνο.

Abstract

In this thesis we analysed a part of a questionnaire answered from 96 epileptic patients at the «Evaggelismos» Hospital, in order to detect the psychological, social and clinical characteristics of epileptic patients in Greece, to identify the main differences presented as compared to a similar research (Nikolakis, 2010) which related to data for the period 2008-2009 and to predict the occurrence of epileptic seizures in the bimester after filling the questionnaire. The questions of the questionnaire which were analysed cover the social life, the skills, the interpersonal relations and the demographic, social, psychological and clinical characteristics of epileptic patients.

The statistical methods which were used in this thesis were univariate descriptive analysis, hypothesis tests for the difference of two independent populations, hypothesis tests for the difference between two independent population proportions, chi square tests, logistic regression, decision tree with the algorithm Conditional Inference Decision Trees and association rule mining with the algorithm Apriori.

The results of this analysis showed that epilepsy is not obstacle for patients' life in Greece, that the quality of patients' life have improved significantly during the last decade and also that the main factor which affects the occurrence of epileptic seizures in the bimester after filling the questionnaire is the occurrence of epileptic seizures in the last year.

Περιεχόμενα

Κατάλογος Πινάκων

Κατάλογος Διαγραμμάτων

ΚΕΦΑΛΑΙΟ 1
Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζουμε κάποιες βασικές πληροφορίες για την επιληψία και τους στόχους της παρούσας εργασίας.

1.1. Πληροφορίες για την επιληψία

1.1.1. Ορισμός και συμπτώματα της επιληψίας

Η επιληψία είναι μια νευρολογική πάθηση που χαρακτηρίζεται από σύντομα επεισόδια δυσλειτουργίας μιας εγκεφαλικής περιοχής (ή και ολόκληρου του εγκεφάλου). Η δυσλειτουργία αυτή οφείλεται σε μια βραχεία παθολογική ηλεκτρική δραστηριότητα νευρικών κυττάρων. Τα κύτταρα αυτά μπορεί να είναι λίγα ή περισσότερα ή ακόμη και όλα τα εγκεφαλικά κύτταρα. Όταν συμβαίνει κάτι τέτοιο, τότε το άτομο καταλαμβάνεται από επιληπτική κρίση. Δυνητικά οποιοσδήποτε μπορεί να καταληφθεί από μια τέτοια κρίση, αλλά οι περισσότεροι άνθρωποι είναι πολύ υψηλής ανθεκτικότητας στις κρίσεις. Όταν ένα άτομο έχει μικρή αντίσταση και οι κρίσεις επαναλαμβάνονται, τότε λέμε ότι το άτομο αυτό πάσχει από επιληψία.

Η επιληπτική κρίση εκδηλώνεται κυρίως με:

- σπασμούς και τρόμο στα χέρια και στα πόδια
- γρήγορο ανοιγοκλείσιμο ματιών ή συσπάσεις στα μάτια που φαίνεται σαν να γυρίζουν
- μούδιασμα
- δυσκολία στην ομιλία και λεκτικές διαταραχές
- απώλεια ούρων ή/και κοπράνων
- ατονική κρίση όπου χάνεται αιφνίδια ο μυϊκός τόνος σε όλο το σώμα

(Πηγές: https://gatzonhs.gr/xrisima/ και https://nevrologos.gr/epilipsia/)

1.1.2. Αιτιολογία της επιληψίας

Τα αίτια της επιληψίας μπορεί να είναι συγγενή ή επίκτητα.

Τα περισσότερα συγγενή αίτια παραμένουν άγνωστα αλλά μπορεί να οφείλονται σε κληρονομικούς παράγοντες.

Ως επίκτητα αίτια θεωρούνται οι παράγοντες που διαταράσσουν την ανατομική συνοχή του εγκεφαλικού ιστού και τις φυσικοχημικές παραμέτρους της βιοηλεκτρικής λειτουργίας του εγκεφάλου. Τέτοιοι παράγοντες είναι:

- αλκοολισμός
- λήψη φαρμάκων ή οπιοειδών ουσιών

- εγκεφαλική αιμορραγία
- όγκοι εγκεφάλου
- εγκεφαλικά επεισόδια
- μηνιγγίτιδα
- χτύπημα στο κεφάλι

(Πηγή: https://www.iatronet.gr/ygeia/nevrologia/article/546/epilipsia.html)

1.1.3. Αντιμετώπιση της επιληψίας

Το 70% των ατόμων με επιληψία μπορούν να ελεγχθούν ικανοποιητικά λαμβάνοντας ένα αντιεπιληπτικό φάρμακο. Περίπου 20% των ασθενών βελτιώνονται με την προσθήκη δεύτερου ή τρίτου φαρμάκου και μόνο το 10% των επιληπτικών δεν ανταποκρίνονται, με αποτέλεσμα να αλλάζουν φάρμακα ή να αναζητούν άλλους τρόπους θεραπείας όπως η κετογονική δίαιτα (διατροφή πλούσια σε λιπαρά και φτωχή σε υδατάνθρακες) και η χειρουργική επέμβαση στον εγκέφαλο.
(Πηγή: https://www.noesi.gr/book/syndrome/epilepsy)

1.2. Στόχοι της παρούσας εργασίας – Περιγραφή των δεδομένων – Δομή εργασίας

Στην παρούσα εργασία θα αναλύσουμε ένα μέρος ενός ερωτηματολογίου[1] που απαντήθηκε από 96 επιληπτικούς ασθενείς κατά την διάρκεια του έτους 2019 στο νοσοκομείο «Ευαγγελισμός». Ο στόχος μας είναι να εντοπίσουμε τα ψυχοκοινωνικά και κλινικά χαρακτηριστικά των επιληπτικών ασθενών της Ελλάδας στην σημερινή εποχή, να βρούμε τις κυριότερες διαφορές που έχουν παρουσιαστεί σε σχέση με μια αντίστοιχη έρευνα (Νικολάκης, 2010) η οποία αφορούσε δεδομένα της περιόδου 2008-2009, αλλά και να προβλέψουμε την εμφάνιση κρίσεων μέσα στο επόμενο δίμηνο από την ημερομηνία συμπλήρωσης του ερωτηματολογίου.

Οι ερωτήσεις του ερωτηματολογίου που θα αναλύσουμε[2] καλύπτουν την κοινωνική ζωή, τις δεξιότητες, τις διαπροσωπικές σχέσεις και τα δημογραφικά, κοινωνικά, ψυχολογικά και κλινικά χαρακτηριστικά των επιληπτικών ασθενών. Πιο συγκεκριμένα, αφορούν τους εξής τομείς:

- φύλο
- ηλικία
- τόπος μόνιμης κατοικίας μέχρι την ηλικία των 18 ετών

[1] Οι ερωτήσεις που περιέχονται στο ερωτηματολόγιο επιλέχθηκαν ύστερα από συνεργασία του νευρολόγου του νοσοκομείου Ευαγγελισμός κ. Σ. Γκατζώνη με τον ψυχίατρο του Αιγινήτειου νοσοκομείου κ. Χ. Παπαγεωργίου.

[2] Οι ερωτήσεις παρουσιάζονται αναλυτικά στο ΠΑΡΑΡΤΗΜΑ Π1.

- τόπος μόνιμης τωρινής κατοικίας
- ηλικία πρώτης κρίσης
- επανάληψη κρίσεων ίδιου τύπου
- αριθμός κρίσεων το τελευταίο δίμηνο
- συνολικός αριθμός κρίσεων
- αριθμός επισκέψεων στον γιατρό τον τελευταίο χρόνο
- νοσηλεία λόγω των κρίσεων τον τελευταίο χρόνο
- ιστορικό επιληψίας στο οικογενειακό περιβάλλον
- συχνότητα κρίσεων τον περασμένο χρόνο
- ύπαρξη άλλου νοσήματος εκτός από τις κρίσεις
- ικανότητα ελέγχου των κρίσεων
- βαθμός αποδοχής της επιληψίας
- κοινωνική αντιμετώπιση
- επαγγελματική κατάσταση
- βαθμός δυσκολίας ανεύρεσης εργασίας εξαιτίας της νόσου
- μορφωτικό επίπεδο
- διακοπή των σπουδών εξαιτίας της νόσου
- οικογενειακή κατάσταση
- ύπαρξη παιδιών
- συχνότητα κατανάλωσης κρασιού
- συχνότητα κατανάλωσης ούζου, ουίσκι, βότκας, τζιν, κονιάκ
- συχνότητα νυχτερινών εξόδων
- δίπλωμα οδήγησης
- οδήγηση
- βαθμός μοναξιάς
- συχνότητα επισκέψεων στον γιατρό
- ανάγκη για περισσότερη ενημέρωση για την νόσο
- βαθμός φόβου των κρίσεων
- βαθμός ανασφάλειας για το μέλλον εξαιτίας των κρίσεων
- επιδίωξη απόκτησης νέων φίλων
- βαθμός επίδρασης της νόσου στις σχέσεις με το άλλο φύλο

Η δομή της παρούσας εργασίας είναι η εξής:

- Στο Κεφάλαιο 2 παρουσιάζουμε τα αποτελέσματα από την περιγραφική ανάλυση των μεταβλητών της έρευνας, τις μεταβολές της τελευταίας δεκαετίας και τα αποτελέσματα της σύγκρισης της έρευνας με παρόμοιες έρευνες της Κολομβίας και της Νέας Ζηλανδίας.

- Στο Κεφάλαιο 3 παραθέτουμε τα αποτελέσματα των ελέγχων συσχέτισης μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και των υπόλοιπων μεταβλητών της έρευνας.
- Στο Κεφάλαιο 4 εφαρμόζουμε λογιστική παλινδρόμηση με σκοπό την πρόβλεψη της εμφάνισης κρίσεων μέσα στο επόμενο δίμηνο.
- Στο Κεφάλαιο 5 κατασκευάζουμε δέντρο απόφασης με τον αλγόριθμο Conditional Inference Decision Trees και εφαρμόζουμε τον αλγόριθμο εξόρυξης κανόνων συσχετίσεων Apriori με σκοπό την πρόβλεψη της εμφάνισης κρίσεων μέσα στο επόμενο δίμηνο.
- Στο Κεφάλαιο 6 παρουσιάζουμε τα κυριότερα συμπεράσματα της εργασίας.

Για την ανάλυση που θα ακολουθήσει στα επόμενα κεφάλαια χρησιμοποιήθηκε η γλώσσα προγραμματισμού R 4.0.5, ενώ μόνο για το Κεφάλαιο 3 χρησιμοποιήθηκε το στατιστικό πρόγραμμα IBM SPSS 25.

ΚΕΦΑΛΑΙΟ 2
Περιγραφική ανάλυση

Στο παρόν κεφάλαιο παρουσιάζουμε τα περιγραφικά στοιχεία των μεταβλητών που χρησιμοποιήθηκαν στην παρούσα εργασία. Επιπλέον, παρουσιάζουμε τα αποτελέσματα που προέκυψαν από την σύγκριση της παρούσας έρευνας με μια αντίστοιχη έρευνα που πραγματοποιήθηκε την περίοδο 2008-2009 και ελέγχουμε την στατιστική σημαντικότητά τους.

2.1. Περιγραφικά στοιχεία των μεταβλητών

Φύλο

Το 61.5% των συμμετεχόντων στην έρευνα είναι άνδρες και το 38.5% είναι γυναίκες.

Ηλικία

Η ηλικία για τους 88 συμμετέχοντες που απάντησαν την ερώτηση (8 συμμετέχοντες δεν απάντησαν) κυμαίνεται από 16 έως 76 ετών, η μέση ηλικία τους είναι 38.45 ετών, ενώ η διάμεση ηλικία τους είναι 38 ετών.

Είναι πιο αποτελεσματικό να ομαδοποιήσουμε τις τιμές της ηλικίας. Για τον σκοπό αυτό θα δημιουργήσουμε 4 ομάδες ηλικίας χρησιμοποιώντας τα τεταρτημόρια της ηλικίας. Με αυτόν τον τρόπο εξασφαλίζουμε ότι οι ομάδες μας θα έχουν παρόμοιο αριθμό παρατηρήσεων. Οι ομάδες ηλικίας που δημιουργούμε είναι οι εξής:

1. από 16 έως και 28 ετών
2. από 29 έως και 38 ετών
3. από 39 έως και 45 ετών
4. από 46 έως και 76 ετών

Τόπος μόνιμης κατοικίας μέχρι την ηλικία των 18 ετών

Στην ερώτηση που αφορά τον τόπο μόνιμης κατοικίας μέχρι την ηλικία των 18 ετών, οι δυνατές απαντήσεις ήταν οι εξής:

1. σε χωριό μέχρι 3000 κατοίκους
2. σε κωμόπολη μέχρι 10000 κατοίκους
3. σε πόλη με πάνω από 10000 κατοίκους
4. σε Αθήνα ή Θεσσαλονίκη
5. αλλού

Από τους 96 συμμετέχοντες στην έρευνα μόνο ένας δεν απάντησε την συγκεκριμένη ερώτηση.

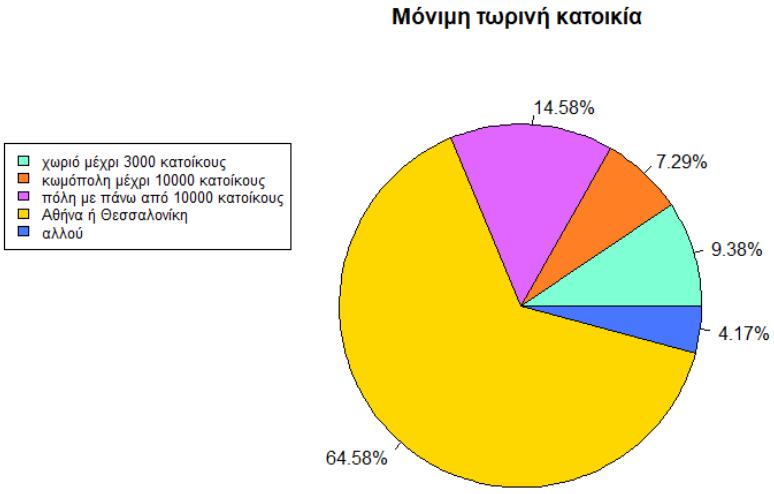
Διάγραμμα 2.1: Κυκλικό διάγραμμα για τον τόπο μόνιμης κατοικίας μέχρι την ηλικία των 18 ετών

Από το Διάγραμμα 2.1 παρατηρούμε ότι η πλειοψηφία (ποσοστό 57.89%) των ερωτηθέντων κατοικούσαν στην Αθήνα ή Θεσσαλονίκη μέχρι την ηλικία των 18 ετών, το 16.84% σε πόλη με πάνω από 10000 κατοίκους, το 14.74% σε χωριό μέχρι 3000 κατοίκους, το 6.32% σε κωμόπολη μέχρι 10000 κατοίκους και ένα μικρό ποσοστό (4.21%) απάντησαν «αλλού».

Τόπος μόνιμης τωρινής κατοικίας

Στην ερώτηση που αφορά τον τόπο μόνιμης τωρινής κατοικίας, οι δυνατές απαντήσεις ήταν οι εξής:

1. σε χωριό μέχρι 3000 κατοίκους
2. σε κωμόπολη μέχρι 10000 κατοίκους
3. σε πόλη με πάνω από 10000 κατοίκους
4. σε Αθήνα ή Θεσσαλονίκη
5. αλλού



Διάγραμμα 2.2: Κυκλικό διάγραμμα για τον τόπο μόνιμης τωρινής κατοικίας

Από το Διάγραμμα 2.2 παρατηρούμε ότι η πλειοψηφία (ποσοστό 64.58%) των ερωτηθέντων κατοικούν μόνιμα στην Αθήνα ή Θεσσαλονίκη, το 14.58% σε πόλη με πάνω από 10000 κατοίκους, το 9.38% σε χωριό μέχρι 3000 κατοίκους, το 7.29% σε κωμόπολη μέχρι 10000 κατοίκους και ένα μικρό ποσοστό (4.17%) απάντησαν «αλλού».

Ηλικία πρώτης κρίσης

Η ηλικία πρώτης κρίσης για τους συμμετέχοντες στην έρευνα κυμαίνεται από 1 έως 73 ετών, η μέση ηλικία πρώτης κρίσης είναι 17.86 ετών, ενώ η διάμεση ηλικία πρώτης κρίσης είναι 15 ετών.

Είναι πιο αποτελεσματικό να ομαδοποιήσουμε τις τιμές της ηλικίας πρώτης κρίσης. Για τον σκοπό αυτό θα δημιουργήσουμε 4 ομάδες ηλικίας πρώτης κρίσης χρησιμοποιώντας τα τεταρτημόρια της ηλικίας πρώτης κρίσης. Με αυτόν τον τρόπο εξασφαλίζουμε ότι οι ομάδες μας θα έχουν παρόμοιο αριθμό παρατηρήσεων. Οι ομάδες ηλικίας πρώτης κρίσης που δημιουργούμε είναι οι εξής:

1. από 1 έως και 10 ετών
2. από 11 έως και 15 ετών
3. από 16 έως και 21 ετών
4. από 22 έως και 73 ετών

Επανάληψη κρίσεων ίδιου τύπου

Στην ερώτηση που αφορά την επανάληψη κρίσεων ίδιου τύπου, το 76.84% των ερωτηθέντων απάντησαν αρνητικά και το 23.16% θετικά. Από τους 96 συμμετέχοντες στην έρευνα μόνο ένας δεν απάντησε την συγκεκριμένη ερώτηση.

Αριθμός κρίσεων το τελευταίο δίμηνο

Ο αριθμός των κρίσεων το τελευταίο δίμηνο για τους 93 συμμετέχοντες που απάντησαν την ερώτηση (3 συμμετέχοντες δεν απάντησαν) κυμαίνεται από 0 έως 38 κρίσεις και η μέση τιμή είναι 2 κρίσεις. Επιπλέον, η πλειοψηφία (ποσοστό 60.22%) των ερωτηθέντων δεν είχαν καμία κρίση το τελευταίο δίμηνο.

Θα ομαδοποιήσουμε τον αριθμό κρίσεων το τελευταίο δίμηνο χρησιμοποιώντας τα τεταρτημόρια της μεταβλητής, όπως ακριβώς κάναμε και με τις προηγούμενες ποσοτικές μεταβλητές. Οι 3 ομάδες που δημιουργούμε είναι οι εξής:

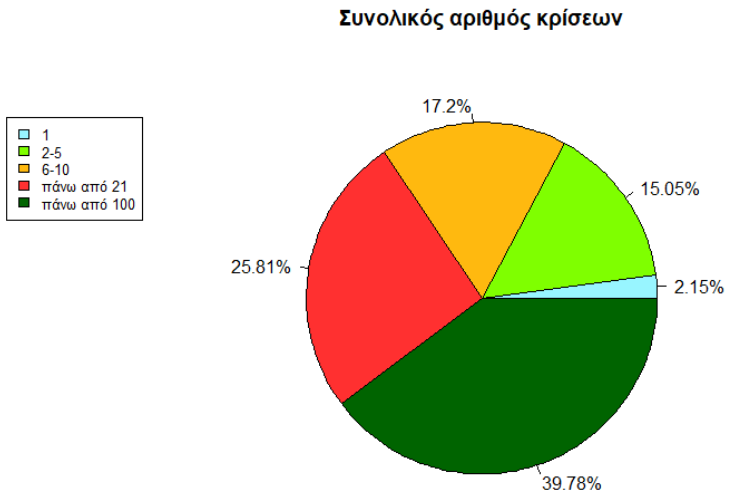
1. 0 κρίσεις
2. 1-2 κρίσεις
3. 3-38 κρίσεις

Συνολικός αριθμός κρίσεων

Στην ερώτηση που αφορά τον συνολικό αριθμό κρίσεων, οι δυνατές απαντήσεις ήταν οι εξής:

1. 1
2. 2-5
3. 6-10
4. πάνω από 21
5. πάνω από 100

Από τους 96 συμμετέχοντες στην έρευνα οι 3 δεν απάντησαν την συγκεκριμένη ερώτηση.



Διάγραμμα 2.3: Κυκλικό διάγραμμα για τον συνολικό αριθμό κρίσεων

Από το Διάγραμμα 2.3 παρατηρούμε ότι η πλειοψηφία (ποσοστό 39.78%) των ερωτηθέντων έχουν κάνει συνολικά πάνω από 100 κρίσεις στη ζωή τους, το 25.81% πάνω από 21 κρίσεις, το 17.2% 6-10 κρίσεις, το 15.05% 2-5 κρίσεις και ένα μικρό ποσοστό (2.15%) έχουν κάνει μόνο 1 κρίση.

Αριθμός επισκέψεων στον γιατρό τον τελευταίο χρόνο

Ο αριθμός επισκέψεων στον γιατρό τον τελευταίο χρόνο για τους 92 συμμετέχοντες που απάντησαν την ερώτηση (4 συμμετέχοντες δεν απάντησαν) κυμαίνεται από 0 έως 9 επισκέψεις και η μέση τιμή είναι 2 επισκέψεις.

Θα ομαδοποιήσουμε τον αριθμό επισκέψεων στον γιατρό τον τελευταίο χρόνο χρησιμοποιώντας τα τεταρτημόρια της μεταβλητής, όπως ακριβώς κάναμε και με τις προηγούμενες ποσοτικές μεταβλητές. Οι 3 ομάδες που δημιουργούμε είναι οι εξής:

1. 0-1 επισκέψεις
2. 2 επισκέψεις
3. 3-9 επισκέψεις

Νοσηλεία λόγω των κρίσεων τον τελευταίο χρόνο

Στην ερώτηση που αφορά την νοσηλεία λόγω των κρίσεων τον τελευταίο χρόνο, η πλειοψηφία (ποσοστό 88.54%) των ερωτηθέντων απάντησαν αρνητικά και το 11.46% θετικά.

Ιστορικό επιληψίας στο οικογενειακό περιβάλλον

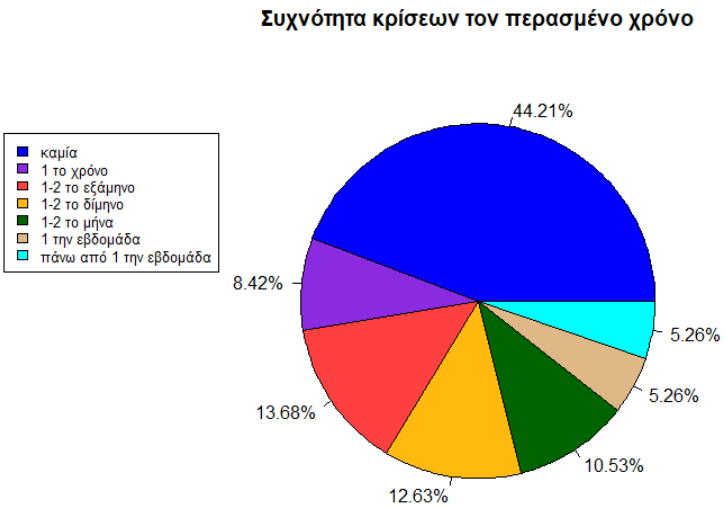
Στην ερώτηση που αφορά το ιστορικό επιληψίας στο οικογενειακό περιβάλλον, η πλειοψηφία (ποσοστό 88.42%) των ερωτηθέντων απάντησαν αρνητικά και το 11.58% θετικά. Από τους 96 συμμετέχοντες στην έρευνα μόνο ένας δεν απάντησε την συγκεκριμένη ερώτηση.

Συχνότητα κρίσεων τον περασμένο χρόνο

Στην ερώτηση που αφορά την συχνότητα των κρίσεων τον περασμένο χρόνο, οι δυνατές απαντήσεις ήταν οι εξής:

1. καμία
2. 1 το χρόνο
3. 1-2 το εξάμηνο
4. 1-2 το δίμηνο
5. 1-2 το μήνα
6. 1 την εβδομάδα
7. πάνω από 1 την εβδομάδα

Από τους 96 συμμετέχοντες στην έρευνα μόνο ένας δεν απάντησε την συγκεκριμένη ερώτηση.



Διάγραμμα 2.4: Κυκλικό διάγραμμα για την συχνότητα κρίσεων τον περασμένο χρόνο

Από το Διάγραμμα 2.4 παρατηρούμε ότι η πλειοψηφία (ποσοστό 44.21%) των ερωτηθέντων δεν είχαν καμία κρίση τον περασμένο χρόνο, το 13.68% είχαν 1-2 κρίσεις το εξάμηνο, το 12.63% 1-2 κρίσεις το δίμηνο, το 10.53% 1-2 κρίσεις το μήνα, το 8.42% 1 κρίση το χρόνο, το 5.26% 1 κρίση την εβδομάδα και το ίδιο ποσοστό (5.26%) είχαν πάνω από 1 κρίση την εβδομάδα.

Ύπαρξη άλλου νοσήματος εκτός από τις κρίσεις

Στην ερώτηση που αφορά την ύπαρξη άλλου νοσήματος εκτός από τις κρίσεις, η πλειοψηφία (ποσοστό 69.47%) των ερωτηθέντων απάντησαν αρνητικά και το 30.53% θετικά. Από τους 96 συμμετέχοντες στην έρευνα μόνο ένας δεν απάντησε την συγκεκριμένη ερώτηση.

Από τους 29 συμμετέχοντες στην έρευνα (ποσοστό 30.53%) που απάντησαν ότι πάσχουν και από άλλο νόσημα εκτός από τις κρίσεις μόνο ένας δεν ανέφερε ποιο είναι αυτό. Τα νοσήματα από τα οποία πάσχουν οι επιληπτικοί ασθενείς είναι τα εξής:

- 5 συμμετέχοντες πάσχουν από διαταραχή του θυρεοειδούς αδένα
- 2 συμμετέχοντες έχουν νοητική υστέρηση
- 2 συμμετέχοντες πάσχουν από σακχαρώδη διαβήτη
- 2 συμμετέχοντες πάσχουν από θυρεοειδίτιδα Χασιμότο

- 1 συμμετέχοντας πάσχει από κατάθλιψη και από αλλεργική ρινίτιδα λόγω σκόνης
- 1 συμμετέχοντας πάσχει από ημικρανία και από ινομυαλγία
- 1 συμμετέχοντας έχει υποβληθεί σε χειρουργείο για ανεύρυσμα ανιούσας αορτής και φέρει βηματοδότη
- 1 συμμετέχοντας πάσχει από ψυχιατρικό νόσημα, από πρόπτωση μιτροειδούς βαλβίδας και έχει έλλειψη του ενζύμου G-6-PD
- 1 συμμετέχοντας πάσχει από υπέρταση
- 1 συμμετέχοντας πάσχει από ηπατίτιδα C
- 1 συμμετέχοντας πάσχει από ρευματοειδή αρθρίτιδα
- 1 συμμετέχοντας πάσχει από υπνική άπνοια και έχει ατροφία οπτικού νεύρου
- 1 συμμετέχοντας πάσχει από συστηματικό ερυθηματώδη λύκο
- 1 συμμετέχοντας πάσχει από οστεοπόρωση
- 1 συμμετέχοντας πάσχει από κρίσεις πανικού
- 1 συμμετέχοντας πάσχει από δερματικές αλλεργίες
- 1 συμμετέχοντας πάσχει από κατάθλιψη
- 1 συμμετέχοντας πάσχει από ενδομητρίωση
- 1 συμμετέχοντας έχει υποβληθεί σε χειρουργείο αντικατάστασης αορτικής βαλβίδας
- 1 συμμετέχοντας έχει πρόβλημα με τα ισχία του
- 1 συμμετέχοντας πάσχει από ηπατίτιδα B

Ικανότητα ελέγχου των κρίσεων

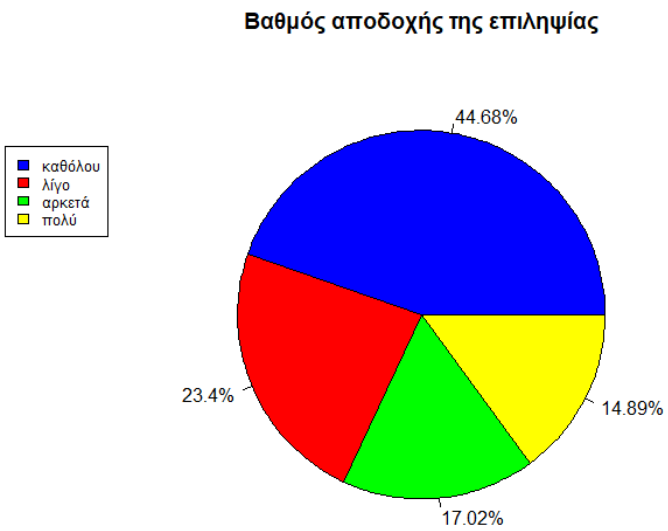
Στην ερώτηση που αφορά την ικανότητα ελέγχου των κρίσεων, η πλειοψηφία (ποσοστό 84.04%) των ερωτηθέντων απάντησαν αρνητικά και το 15.96% θετικά. Από τους 96 συμμετέχοντες στην έρευνα οι 2 δεν απάντησαν την συγκεκριμένη ερώτηση. (Δεν αναλύσαμε την ερώτηση που αφορά τον τρόπο με τον οποίο οι επιληπτικοί ασθενείς μπορούν να εμποδίσουν την κρίση).

Βαθμός αποδοχής της επιληψίας

Στην ερώτηση σχετικά με το αν ενοχλούνται οι ασθενείς να τους αποκαλούν «επιληπτικούς», οι δυνατές απαντήσεις ήταν οι εξής:

1. καθόλου
2. λίγο
3. αρκετά
4. πολύ

Από τους 96 συμμετέχοντες στην έρευνα οι 2 δεν απάντησαν την συγκεκριμένη ερώτηση.



Διάγραμμα 2.5: Κυκλικό διάγραμμα για τον βαθμό αποδοχής της επιληψίας

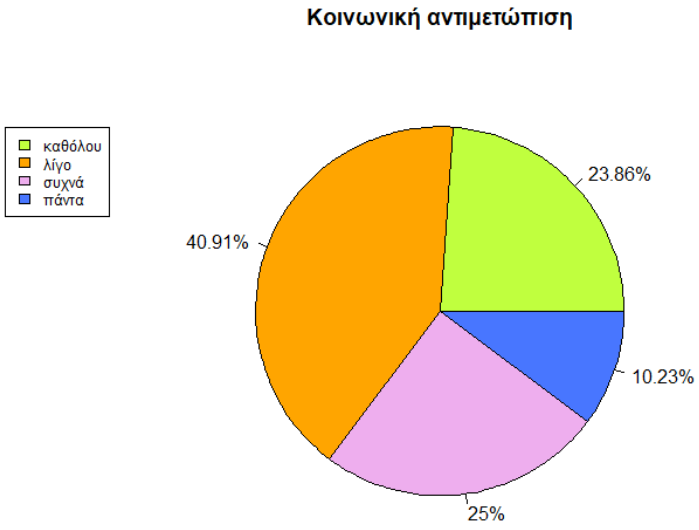
Από το Διάγραμμα 2.5 παρατηρούμε ότι η πλειοψηφία (ποσοστό 44.68%) των ερωτηθέντων απάντησαν ότι δεν τους ενοχλεί καθόλου να τους αποκαλούν «επιληπτικούς», το 23.4% ότι τους ενοχλεί λίγο, το 17.02% ότι τους ενοχλεί αρκετά και το 14.89% ότι τους ενοχλεί πολύ.

Κοινωνική αντιμετώπιση

Στην ερώτηση σχετικά με το αν η λέξη «επιληπτικός» κάνει τους άλλους να βλέπουν τους ασθενείς με προκατάληψη, οι δυνατές απαντήσεις ήταν οι εξής:

1. καθόλου
2. λίγο
3. συχνά
4. πάντα

Από τους 96 συμμετέχοντες στην έρευνα οι 8 δεν απάντησαν την συγκεκριμένη ερώτηση.



Διάγραμμα 2.6: Κυκλικό διάγραμμα για την κοινωνική αντιμετώπιση

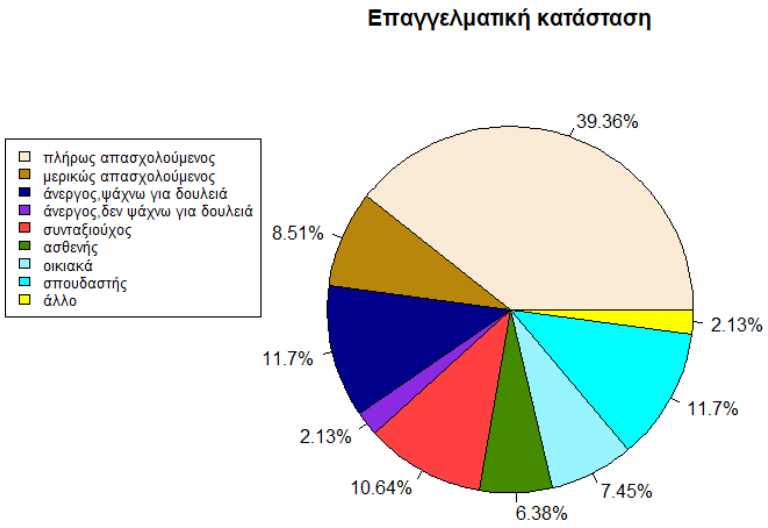
Από το Διάγραμμα 2.6 παρατηρούμε ότι η πλειοψηφία (ποσοστό 40.91%) των ερωτηθέντων απάντησαν ότι η λέξη «επιληπτικός» κάνει τους άλλους να τους βλέπουν λίγο με προκατάληψη, το 25% απάντησαν συχνά, το 23.86% απάντησαν καθόλου, ενώ το 10.23% απάντησαν πάντα.

Επαγγελματική κατάσταση

Στην ερώτηση που αφορά την επαγγελματική κατάσταση των ασθενών, οι δυνατές απαντήσεις ήταν οι εξής:

1. πλήρως απασχολούμενος
2. μερικώς απασχολούμενος
3. άνεργος, ψάχνω για δουλειά
4. άνεργος, δεν ψάχνω για δουλειά
5. συνταξιούχος
6. ασθενής
7. οικιακά
8. σπουδαστής
9. άλλο

Από τους 96 συμμετέχοντες στην έρευνα οι 2 δεν απάντησαν την συγκεκριμένη ερώτηση.



Διάγραμμα 2.7: Κυκλικό διάγραμμα για την επαγγελματική κατάσταση

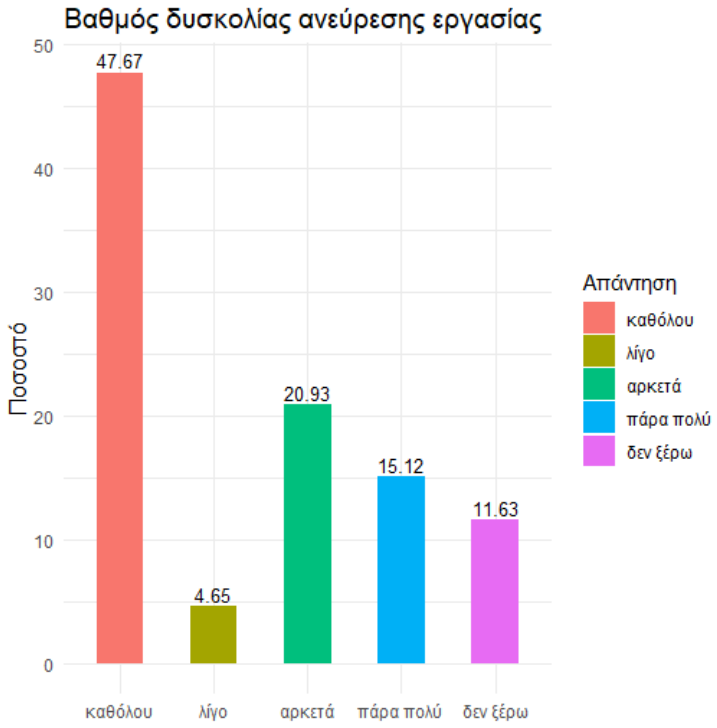
Από το Διάγραμμα 2.7 παρατηρούμε ότι η πλειοψηφία (ποσοστό 39.36%) των ερωτηθέντων απάντησαν ότι είναι πλήρως απασχολούμενοι, το 11.7% ότι είναι σπουδαστές και το ίδιο ποσοστό (11.7%) ότι είναι άνεργοι που ψάχνουν για δουλειά, το 10.64% ότι είναι συνταξιούχοι, το 8.51% ότι είναι μερικώς απασχολούμενοι, το 7.45% δήλωσαν ως επάγγελμα «οικιακά», το 6.38% απάντησαν ότι είναι ασθενείς, το 2.13% ότι είναι άνεργοι που δεν ψάχνουν για δουλειά και το ίδιο ποσοστό (2.13%) απάντησαν «άλλο».

Βαθμός δυσκολίας ανεύρεσης εργασίας εξαιτίας της νόσου

Στην ερώτηση σχετικά με το αν η επιληψία εμποδίζει τους ασθενείς στην ανεύρεση εργασίας, οι δυνατές απαντήσεις ήταν οι εξής:

1. καθόλου
2. λίγο
3. αρκετά
4. πάρα πολύ
5. δεν ξέρω

Από τους 96 συμμετέχοντες στην έρευνα οι 10 δεν απάντησαν την συγκεκριμένη ερώτηση.



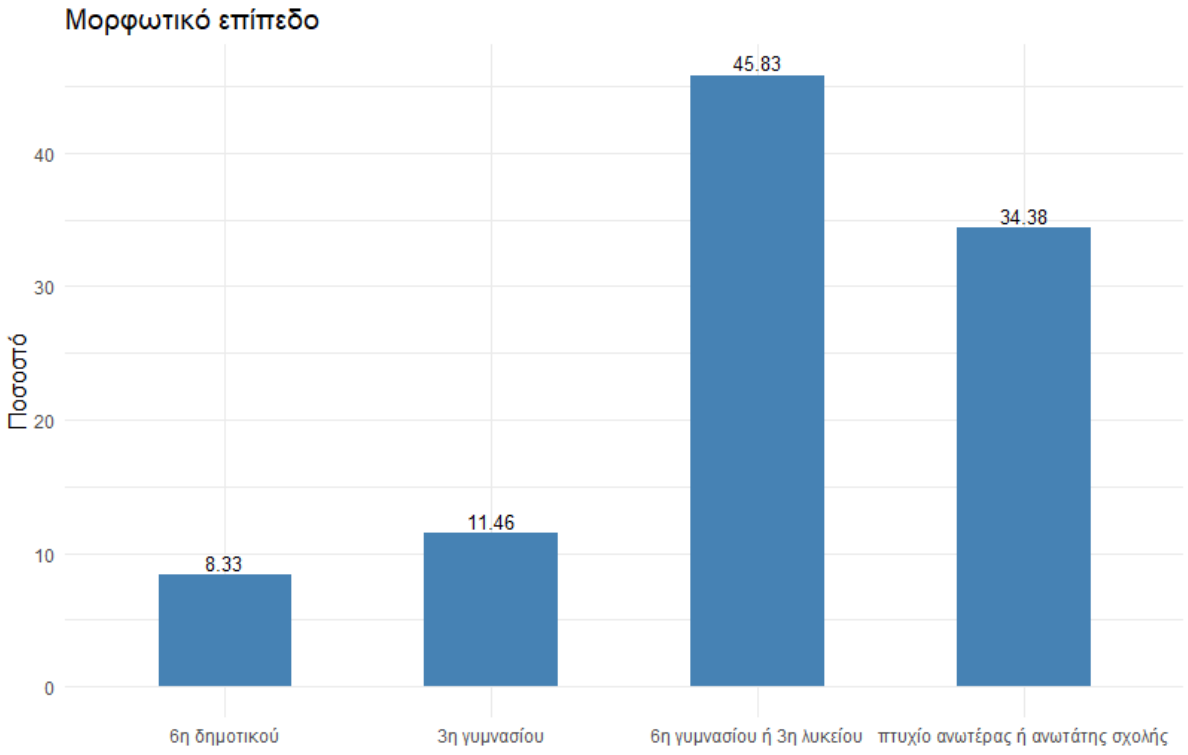
Διάγραμμα 2.8: Ραβδόγραμμα για τον βαθμό δυσκολίας ανεύρεσης εργασίας εξαιτίας της νόσου

Από το Διάγραμμα 2.8 παρατηρούμε ότι η πλειοψηφία (ποσοστό 47.67%) των ερωτηθέντων απάντησαν ότι η επιληψία δεν τους εμποδίζει καθόλου στην ανεύρεση εργασίας, το 20.93% ότι τους εμποδίζει αρκετά, το 15.12% ότι τους εμποδίζει πάρα πολύ, το 11.63% απάντησαν «δεν ξέρω» και το 4.65% απάντησαν ότι τους εμποδίζει λίγο.

Μορφωτικό επίπεδο

Στην ερώτηση που αφορά το μορφωτικό επίπεδο των ασθενών, οι δυνατές απαντήσεις ήταν οι εξής:

1. έως τρίτη δημοτικού
2. έως έκτη δημοτικού
3. έως τρίτη γυμνασίου
4. έως έκτη γυμνασίου ή τρίτη λυκείου
5. πτυχίο ανωτέρας ή ανωτάτης σχολής



Διάγραμμα 2.9: Ραβδόγραμμα για το μορφωτικό επίπεδο

Από το Διάγραμμα 2.9 παρατηρούμε ότι η πλειοψηφία (ποσοστό 45.83%) των ερωτηθέντων είναι απόφοιτοι έκτης γυμνασίου ή τρίτης λυκείου. Μεγάλο, επίσης, ποσοστό των ερωτηθέντων (34.38%) είναι πτυχιούχοι ανωτέρας ή ανωτάτης σχολής. Κατά φθίνουσα σειρά, ποσοστό 11.46% είναι απόφοιτοι τρίτης γυμνασίου και ποσοστό 8.33% είναι απόφοιτοι έκτης δημοτικού. Κανένας συμμετέχοντας στην έρευνα δεν έχει γραμματικές γνώσεις έως τρίτη δημοτικού.

<u>Διακοπή των σπουδών εξαιτίας της νόσου</u>

Στην ερώτηση σχετικά με το αν οι ασθενείς διέκοψαν τις σπουδές τους εξαιτίας της επιληψίας, η πλειοψηφία (ποσοστό 88.04%) των ερωτηθέντων απάντησαν αρνητικά και το 11.96% θετικά. Από τους 96 συμμετέχοντες στην έρευνα οι 4 δεν απάντησαν την συγκεκριμένη ερώτηση.

Οικογενειακή κατάσταση

Η πλειοψηφία (ποσοστό 64.89%) των ερωτηθέντων είναι άγαμοι, το 31.91% είναι παντρεμένοι και ένα μικρό ποσοστό (3.19%) είναι διαζευγμένοι. Από τους 96 συμμετέχοντες στην έρευνα οι 2 δεν απάντησαν την συγκεκριμένη ερώτηση.

Ύπαρξη παιδιών

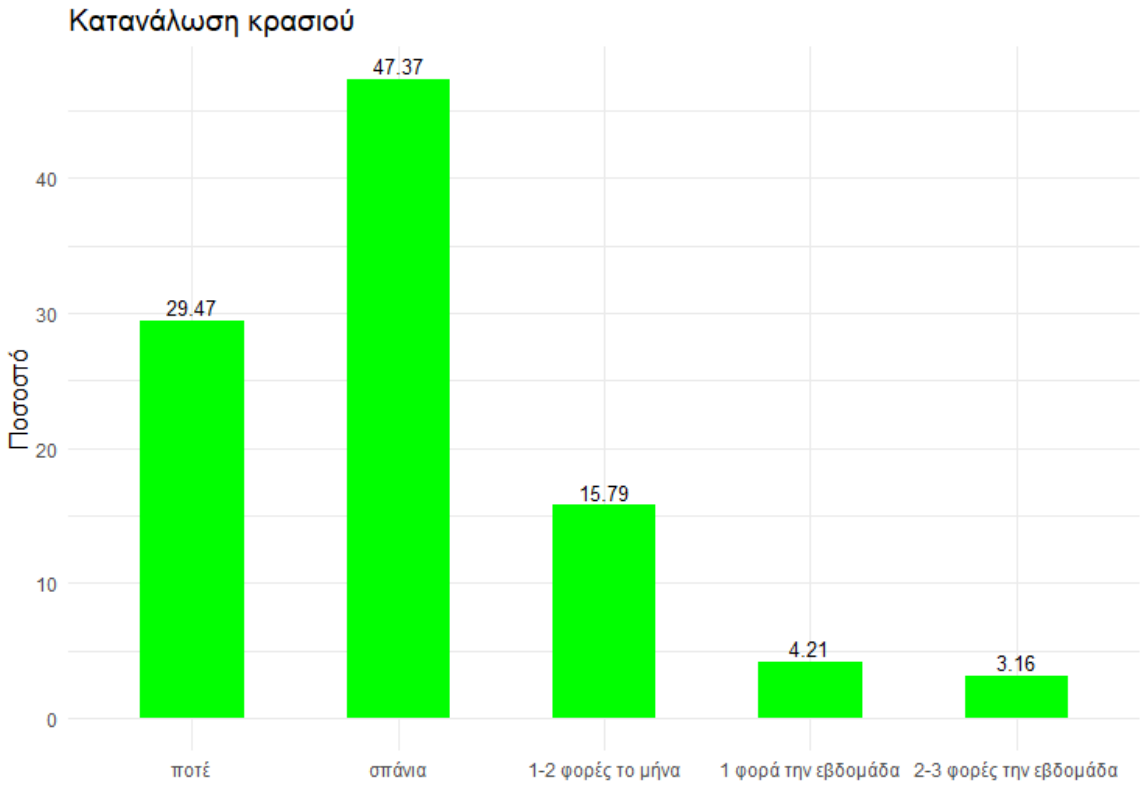
Η πλειοψηφία (ποσοστό 68.42%) των ερωτηθέντων δεν έχουν παιδιά, ενώ το 31.58% έχουν. Από τους 96 συμμετέχοντες στην έρευνα μόνο ένας δεν απάντησε την συγκεκριμένη ερώτηση. (Δεν αναλύσαμε τον αριθμό παιδιών που έχουν οι επιληπτικοί ασθενείς).

Συχνότητα κατανάλωσης κρασιού

Στην ερώτηση που αφορά την συχνότητα κατανάλωσης κρασιού, οι δυνατές απαντήσεις ήταν οι εξής:

1. ποτέ
2. σπάνια
3. μία-δύο φορές το μήνα
4. μία φορά την εβδομάδα
5. δύο-τρεις φορές την εβδομάδα
6. κάθε μέρα

Από τους 96 συμμετέχοντες στην έρευνα μόνο ένας δεν απάντησε την συγκεκριμένη ερώτηση.



Διάγραμμα 2.10: Ραβδόγραμμα για την συχνότητα κατανάλωσης κρασιού

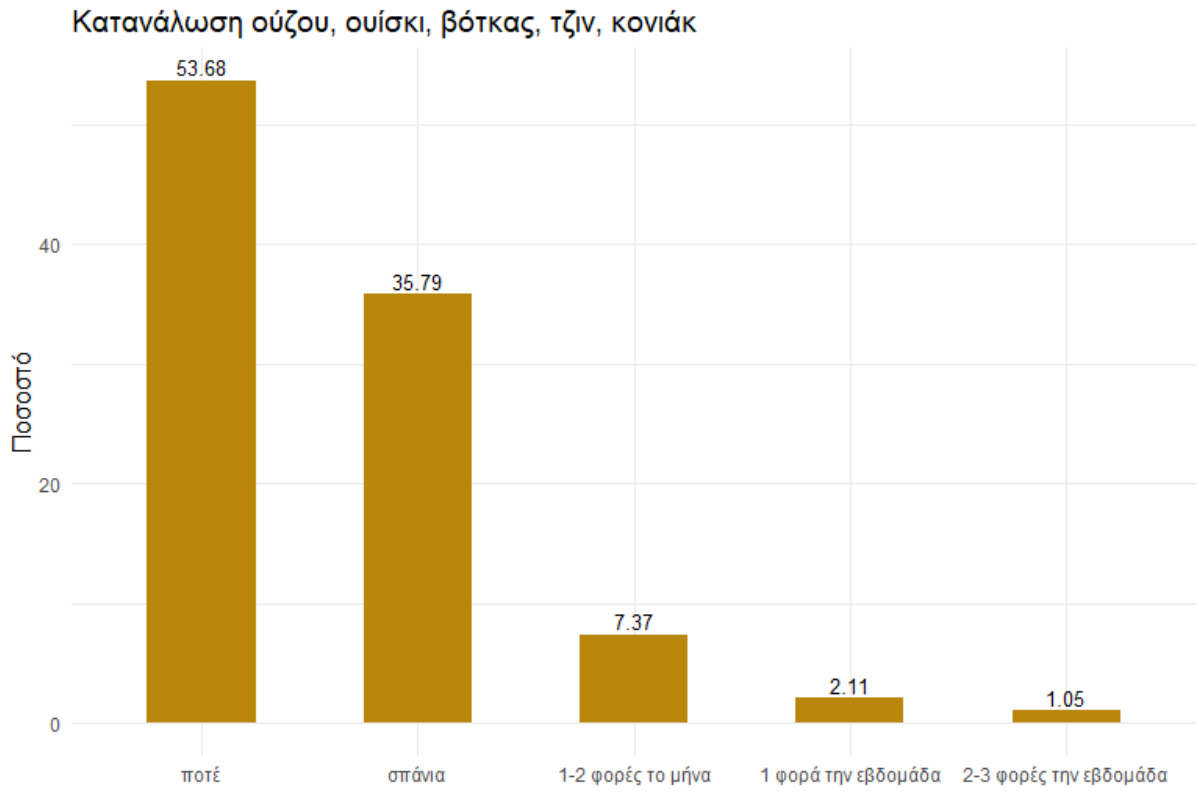
Από το Διάγραμμα 2.10 παρατηρούμε ότι η πλειοψηφία (ποσοστό 47.37%) των ερωτηθέντων σπάνια πίνουν κρασί. Μεγάλο, επίσης, ποσοστό των ερωτηθέντων (29.47%) δεν πίνουν ποτέ κρασί. Κατά φθίνουσα σειρά, ποσοστό 15.79% πίνουν κρασί μία-δύο φορές το μήνα, 4.21% μία φορά την εβδομάδα και 3.16% δύο-τρεις φορές την εβδομάδα. Κανένας συμμετέχοντας στην έρευνα δεν πίνει κρασί κάθε μέρα. (Δεν αναλύσαμε τον αριθμό ποτηριών κρασιού που πίνουν οι επιληπτικοί ασθενείς).

Συχνότητα κατανάλωσης ούζου, ουίσκι, βότκας, τζιν, κονιάκ

Στην ερώτηση που αφορά την συχνότητα κατανάλωσης ούζου, ουίσκι, βότκας, τζιν ή κονιάκ, οι δυνατές απαντήσεις ήταν οι εξής:

1. ποτέ
2. σπάνια
3. μία-δύο φορές το μήνα
4. μία φορά την εβδομάδα
5. δύο-τρεις φορές την εβδομάδα
6. κάθε μέρα

Από τους 96 συμμετέχοντες στην έρευνα μόνο ένας δεν απάντησε την συγκεκριμένη ερώτηση.



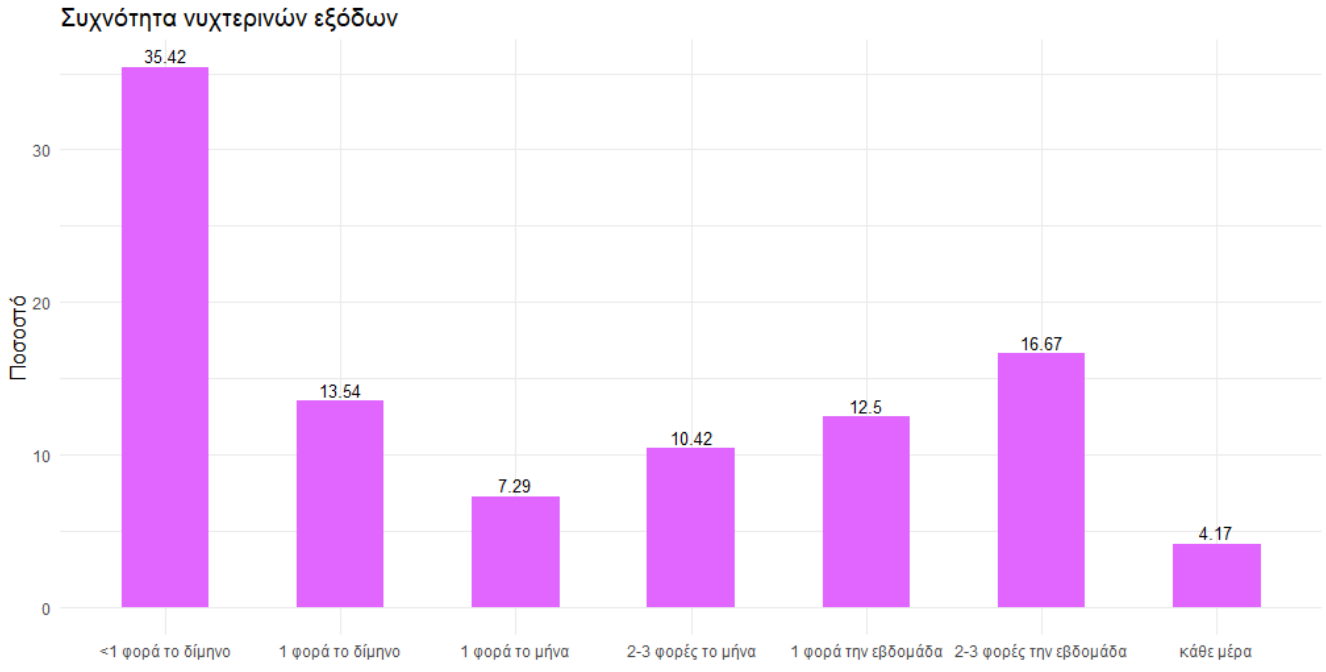
Διάγραμμα 2.11: Ραβδόγραμμα για την συχνότητα κατανάλωσης ούζου, ουίσκι, βότκας, τζιν, κονιάκ

Από το Διάγραμμα 2.11 παρατηρούμε ότι η πλειοψηφία (ποσοστό 53.68%) των ερωτηθέντων δεν πίνουν ποτέ ούζο, ουίσκι, βότκα, τζιν ή κονιάκ, το 35.79% πίνουν σπάνια, το 7.37% μία-δύο φορές το μήνα, το 2.11% μία φορά την εβδομάδα και το 1.05% δύο-τρεις φορές την εβδομάδα. Κανένας συμμετέχοντας στην έρευνα δεν πίνει αυτά τα οινοπνευματώδη ποτά κάθε μέρα. (Δεν αναλύσαμε τον αριθμό ποτηριών οινοπνευματωδών ποτών που πίνουν οι επιληπτικοί ασθενείς).

Συχνότητα νυχτερινών εξόδων

Στην ερώτηση που αφορά την συχνότητα νυχτερινών εξόδων, οι δυνατές απαντήσεις ήταν οι εξής:

1. λιγότερο από μία φορά το δίμηνο
2. μία φορά το δίμηνο
3. μία φορά το μήνα
4. 2-3 φορές το μήνα
5. μία φορά την εβδομάδα
6. 2-3 φορές την εβδομάδα
7. κάθε μέρα



Διάγραμμα 2.12: Ραβδόγραμμα για την συχνότητα νυχτερινών εξόδων

Από το Διάγραμμα 2.12 παρατηρούμε ότι η πλειοψηφία (ποσοστό 35.42%) των ερωτηθέντων απάντησαν ότι βγαίνουν έξω τα βράδια λιγότερο από μία φορά το δίμηνο, το 16.67% 2-3 φορές την εβδομάδα, το 13.54% μία φορά το δίμηνο, το 12.5% μία φορά την εβδομάδα, το 10.42% 2-3 φορές το μήνα, το 7.29% μία φορά το μήνα και το 4.17% κάθε μέρα.

Δίπλωμα οδήγησης

Το 50% των ερωτηθέντων έχουν δίπλωμα οδήγησης και το 50% δεν έχουν.

Οδήγηση

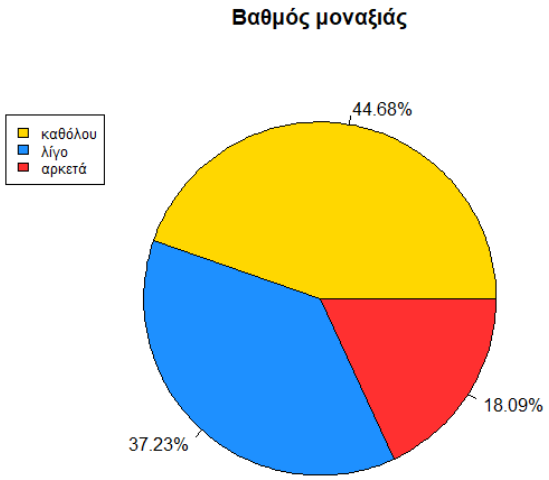
Στην ερώτηση σχετικά με το αν οδηγούν οι ασθενείς, η πλειοψηφία (ποσοστό 68.75%) των ερωτηθέντων απάντησαν αρνητικά και το 31.25% θετικά.

Βαθμός μοναξιάς

Στην ερώτηση σχετικά με το αν οι ασθενείς αισθάνονται μοναξιά, οι δυνατές απαντήσεις ήταν οι εξής:

1. καθόλου
2. λίγο
3. αρκετά

Από τους 96 συμμετέχοντες στην έρευνα οι 2 δεν απάντησαν την συγκεκριμένη ερώτηση.



Διάγραμμα 2.13: Κυκλικό διάγραμμα για τον βαθμό μοναξιάς

Από το Διάγραμμα 2.13 παρατηρούμε ότι η πλειοψηφία (ποσοστό 44.68%) των ερωτηθέντων δεν αισθάνονται καθόλου μοναξιά, το 37.23% αισθάνονται λίγη μοναξιά, ενώ το 18.09% αισθάνονται αρκετή μοναξιά.

Συχνότητα επισκέψεων στον γιατρό

Στην ερώτηση που αφορά την συχνότητα επισκέψεων στον γιατρό, οι δυνατές απαντήσεις ήταν οι εξής:

1. έχω πάνω από ένα χρόνο να πάω
2. μία φορά το χρόνο
3. μία φορά το εξάμηνο ως μία φορά το χρόνο
4. μία φορά κάθε τρεις ως έξι μήνες
5. μία φορά κάθε ένα ως τρεις μήνες
6. πιο συχνά από μία φορά το μήνα

Από τους 96 συμμετέχοντες στην έρευνα μόνο ένας δεν απάντησε την συγκεκριμένη ερώτηση.

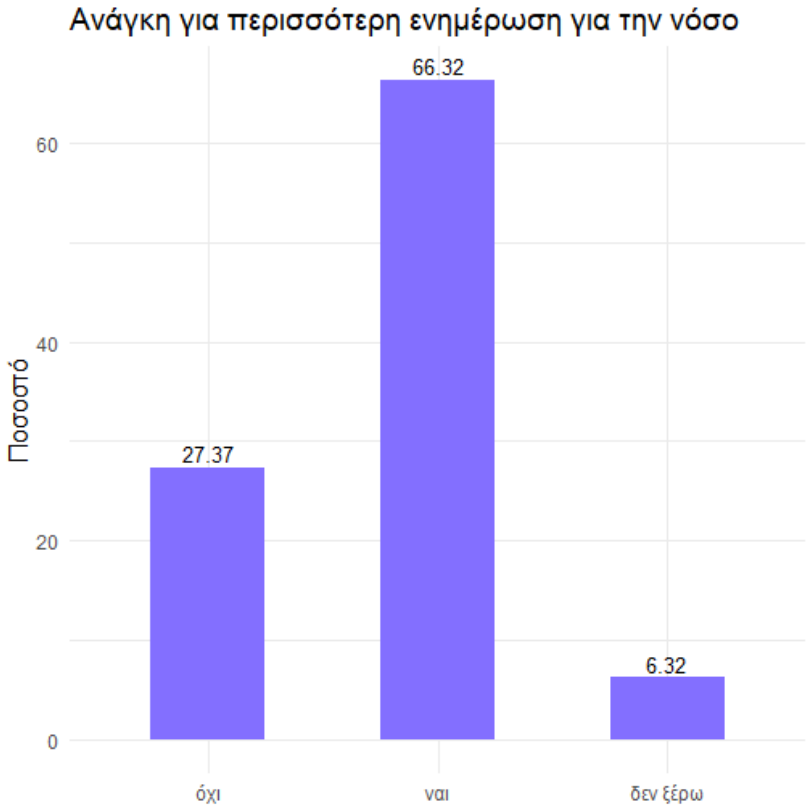
Η πλειοψηφία (ποσοστό 33.68%) των ερωτηθέντων επισκέπτονται τον γιατρό τους μία φορά το εξάμηνο ως μία φορά το χρόνο, το 21.05% μία φορά το χρόνο, το 17.89% έχουν πάνω από ένα χρόνο να επισκεφθούν τον γιατρό τους, το 15.79% μία φορά κάθε τρεις ως έξι μήνες, το 10.53% μία φορά κάθε ένα ως τρεις μήνες και ένα πολύ μικρό ποσοστό (1.05%) επισκέπτονται τον γιατρό τους πιο συχνά από μία φορά το μήνα.

Ανάγκη για περισσότερη ενημέρωση για την νόσο

Στην ερώτηση σχετικά με το αν οι ασθενείς χρειάζονται περισσότερη ενημέρωση για την νόσο τους, οι δυνατές απαντήσεις ήταν οι εξής:

1. όχι
2. ναι
3. δεν ξέρω

Από τους 96 συμμετέχοντες στην έρευνα μόνο ένας δεν απάντησε την συγκεκριμένη ερώτηση.



Διάγραμμα 2.14: Ραβδόγραμμα για την ανάγκη περισσότερης ενημέρωσης για την νόσο

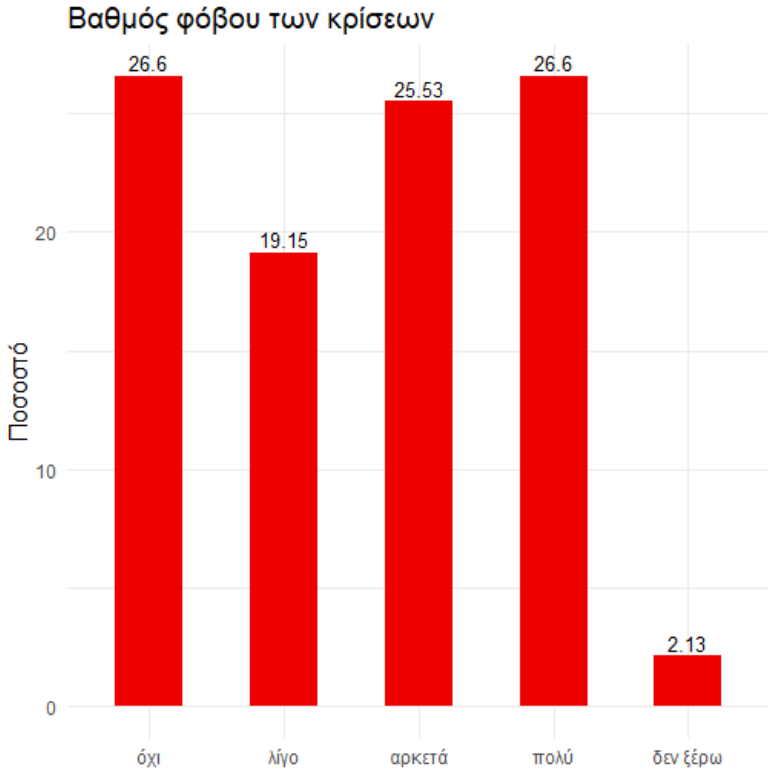
Από το Διάγραμμα 2.14 παρατηρούμε ότι η πλειοψηφία (ποσοστό 66.32%) των ερωτηθέντων απάντησαν ότι χρειάζονται περισσότερη ενημέρωση για την νόσο τους, ενώ το 27.37% ότι δεν χρειάζονται. Επίσης, ένα μικρό ποσοστό (6.32%) απάντησαν «δεν ξέρω».

Βαθμός φόβου των κρίσεων

Στην ερώτηση σχετικά με το αν οι ασθενείς φοβούνται τις κρίσεις, οι δυνατές απαντήσεις ήταν οι εξής:

1. όχι
2. λίγο
3. αρκετά
4. πολύ
5. δεν ξέρω

Από τους 96 συμμετέχοντες στην έρευνα οι 2 δεν απάντησαν την συγκεκριμένη ερώτηση.



Διάγραμμα 2.15: Ραβδόγραμμα για τον βαθμό φόβου των κρίσεων

Από το Διάγραμμα 2.15 παρατηρούμε ότι το 26.6% των ερωτηθέντων απάντησαν ότι δεν φοβούνται τις κρίσεις και το ίδιο ποσοστό (26.6%) ότι φοβούνται πολύ τις κρίσεις, το 25.53% ότι φοβούνται αρκετά τις κρίσεις, το 19.15% ότι φοβούνται λίγο τις κρίσεις, ενώ ένα μικρό ποσοστό (2.13%) απάντησαν «δεν ξέρω».

Βαθμός ανασφάλειας για το μέλλον εξαιτίας των κρίσεων

Στην ερώτηση σχετικά με το αν οι ασθενείς νιώθουν ανασφάλεια για το μέλλον εξαιτίας των κρίσεων, οι δυνατές απαντήσεις ήταν οι εξής:

1. καθόλου
2. λίγο
3. αρκετά
4. πολύ
5. δεν ξέρω

Από τους 96 συμμετέχοντες στην έρευνα μόνο ένας δεν απάντησε την συγκεκριμένη ερώτηση.

Η πλειοψηφία (ποσοστό 38.95%) των ερωτηθέντων απάντησαν ότι νιώθουν λίγη ανασφάλεια για το μέλλον εξαιτίας των κρίσεων, το 27.37% ότι δεν νιώθουν

καθόλου ανασφάλεια, το 21.05% ότι νιώθουν αρκετή ανασφάλεια, το 11.58% ότι νιώθουν πολλή ανασφάλεια, ενώ ένα πολύ μικρό ποσοστό (1.05%) απάντησαν «δεν ξέρω».

Επιδίωξη απόκτησης νέων φίλων

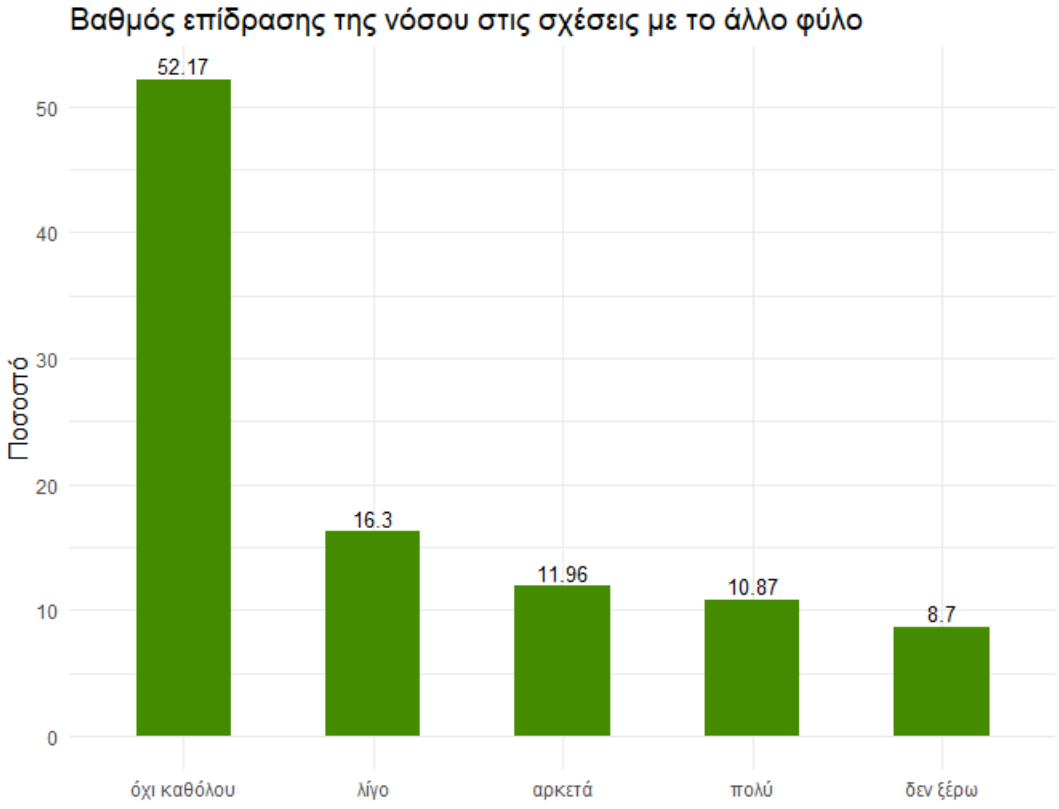
Στην ερώτηση που αφορά την επιδίωξη απόκτησης νέων φίλων, η πλειοψηφία (ποσοστό 62.11%) των ερωτηθέντων απάντησαν θετικά, το 33.68% αρνητικά, ενώ το 4.21% απάντησαν «δεν ξέρω». Από τους 96 συμμετέχοντες στην έρευνα μόνο ένας δεν απάντησε την συγκεκριμένη ερώτηση.

Βαθμός επίδρασης της νόσου στις σχέσεις με το άλλο φύλο

Στην ερώτηση σχετικά με το αν η επιληψία επηρεάζει τις σχέσεις των ασθενών με το άλλο φύλο, οι δυνατές απαντήσεις ήταν οι εξής:

1. όχι καθόλου
2. λίγο
3. αρκετά
4. πολύ
5. δεν ξέρω

Από τους 96 συμμετέχοντες στην έρευνα οι 4 δεν απάντησαν την συγκεκριμένη ερώτηση.



Διάγραμμα 2.16: Ραβδόγραμμα για τον βαθμό επίδρασης της νόσου στις σχέσεις με το άλλο φύλο

Από το Διάγραμμα 2.16 παρατηρούμε ότι η πλειοψηφία (ποσοστό 52.17%) των ερωτηθέντων απάντησαν ότι η επιληψία δεν επηρεάζει καθόλου τις σχέσεις τους με το άλλο φύλο, το 16.3% ότι τις επηρεάζει λίγο, το 11.96% ότι τις επηρεάζει αρκετά, το 10.87% ότι τις επηρεάζει πολύ, ενώ το 8.7% απάντησαν «δεν ξέρω».

2.2. Μεταβολές την τελευταία δεκαετία

Στην ενότητα αυτή παρουσιάζουμε τις κυριότερες μεταβολές που παρατηρούνται όταν συγκρίνουμε την παρούσα έρευνα με μια αντίστοιχη έρευνα (Νικολάκης, 2010) που πραγματοποιήθηκε την περίοδο 2008-2009 και ελέγχουμε την στατιστική σημαντικότητα αυτών των μεταβολών.

2.2.1. Παρουσίαση των κυριότερων μεταβολών την τελευταία δεκαετία

- Αύξηση της μέσης ηλικίας των συμμετεχόντων κατά 7.16 έτη.
- Αύξηση της μέσης ηλικίας πρώτης κρίσης κατά 2.12 έτη.
- Μείωση του μέσου αριθμού κρίσεων το τελευταίο δίμηνο κατά 5.

- Αύξηση του ποσοστού των συμμετεχόντων που δεν είχαν καμία κρίση τον περασμένο χρόνο κατά 11.6% και ταυτόχρονη μείωση του ποσοστού των συμμετεχόντων που είχαν πάνω από 1 κρίση την εβδομάδα τον περασμένο χρόνο κατά 7.1%.
- Αύξηση του ποσοστού των συμμετεχόντων που πάσχουν από κάποιο άλλο νόσημα εκτός από τις κρίσεις κατά 8.9%.
- Αύξηση του ποσοστού των συμμετεχόντων που δεν μπορούν να εμποδίσουν την κρίση κατά 18.1%.
- Αύξηση του ποσοστού των συμμετεχόντων που δεν τους ενοχλεί καθόλου να τους αποκαλούν «επιληπτικούς» κατά 16% και ταυτόχρονη μείωση του ποσοστού των συμμετεχόντων που τους ενοχλεί πολύ να τους αποκαλούν «επιληπτικούς» κατά 14%.
- Αύξηση του ποσοστού των συμμετεχόντων που απάντησαν ότι η επιληψία δεν τους εμποδίζει καθόλου στην ανεύρεση εργασίας κατά 5.3% και ταυτόχρονη μείωση του ποσοστού των συμμετεχόντων που απάντησαν ότι η επιληψία τους εμποδίζει λίγο στην ανεύρεση εργασίας κατά 6%.
- Μείωση του ποσοστού των συμμετεχόντων που είναι άγαμοι κατά 15.3% και ταυτόχρονη αύξηση του ποσοστού των συμμετεχόντων που είναι παντρεμένοι κατά 15.4%.
- Μείωση του ποσοστού των συμμετεχόντων που δεν πίνουν ποτέ κρασί κατά 18.8% και ταυτόχρονη αύξηση του ποσοστού των συμμετεχόντων που σπάνια πίνουν κρασί κατά 16.6%.
- Αύξηση του ποσοστού των συμμετεχόντων που βγαίνουν έξω τα βράδια λιγότερο από μία φορά το δίμηνο κατά 13.4% και ταυτόχρονη μείωση του ποσοστού των συμμετεχόντων που βγαίνουν έξω τα βράδια μία φορά την εβδομάδα κατά 8.4%.
- Μείωση του ποσοστού των συμμετεχόντων που επισκέπτονται τον γιατρό τους μία φορά κάθε τρεις ως έξι μήνες κατά 13.8%, αύξηση του ποσοστού των συμμετεχόντων που επισκέπτονται τον γιατρό τους μία φορά το εξάμηνο ως μία φορά το χρόνο κατά 12.8%, αύξηση του ποσοστού των συμμετεχόντων που έχουν πάνω από ένα χρόνο να επισκεφθούν τον γιατρό τους κατά 13.5% και μείωση του ποσοστού των συμμετεχόντων που επισκέπτονται τον γιατρό τους μία φορά κάθε ένα ως τρεις μήνες κατά 7.1%.
- Αύξηση του ποσοστού των συμμετεχόντων που δεν φοβούνται τις κρίσεις κατά 10.1% και ταυτόχρονη μείωση του ποσοστού των συμμετεχόντων που φοβούνται λίγο τις κρίσεις κατά 13.9%.
- Αύξηση του ποσοστού των συμμετεχόντων που δεν νιώθουν καθόλου ανασφάλεια για το μέλλον εξαιτίας των κρίσεων κατά 9.8%, μείωση του ποσοστού των

συμμετεχόντων που νιώθουν αρκετή ανασφάλεια κατά 5.4% και μείωση του ποσοστού των συμμετεχόντων που νιώθουν πολλή ανασφάλεια κατά 5%.

- Μείωση του ποσοστού των συμμετεχόντων που επιδιώκουν να αποκτήσουν νέους φίλους κατά 9.3% και ταυτόχρονη αύξηση του ποσοστού των συμμετεχόντων που δεν επιδιώκουν να αποκτήσουν νέους φίλους κατά 11.7%.
- Αύξηση του ποσοστού των συμμετεχόντων που απάντησαν ότι η επιληψία δεν επηρεάζει καθόλου τις σχέσεις τους με το άλλο φύλο κατά 8.3% και ταυτόχρονη μείωση του ποσοστού των συμμετεχόντων που απάντησαν ότι τις επηρεάζει αρκετά κατά 5.6%.

2.2.2. Στατιστική σημαντικότητα των μεταβολών της τελευταίας δεκαετίας για τις ποσοτικές μεταβλητές

Για να ελέγξουμε την στατιστική σημαντικότητα των μεταβολών της τελευταίας δεκαετίας για τις ποσοτικές μεταβλητές, θα χρησιμοποιήσουμε ελέγχους υποθέσεων για τη διαφορά των μέσων τιμών δύο ανεξάρτητων πληθυσμών. Πιο συγκεκριμένα, μας ενδιαφέρει ο έλεγχος $H_0: \mu_1 = \mu_2$ έναντι $H_1: \mu_1 \neq \mu_2$. Στην επόμενη παράγραφο δίνουμε κάποια βασικά θεωρητικά στοιχεία για αυτόν τον έλεγχο.

Έστω ότι έχουμε ένα τυχαίο δείγμα μεγέθους n από έναν πληθυσμό με μέση τιμή μ_1 και διακύμανση σ_1^2 και ένα δεύτερο τυχαίο δείγμα μεγέθους m με μέση τιμή μ_2 και διακύμανση σ_2^2. Υπολογίζουμε τη δειγματική μέση τιμή $\bar{X}$ και τη δειγματική διακύμανση S_1^2 του πρώτου δείγματος, καθώς και τη δειγματική μέση τιμή $\bar{Y}$ και τη δειγματική διακύμανση S_2^2 του δεύτερου δείγματος. Για τον έλεγχο της διαφοράς των δύο μέσων τιμών διακρίνουμε τέσσερις περιπτώσεις:

- <u>οι δύο πληθυσμοί ακολουθούν κανονική κατανομή με γνωστές διακυμάνσεις :</u> Σε αυτήν την περίπτωση απορρίπτουμε την μηδενική υπόθεση $H_0: \mu_1 = \mu_2$ έναντι της εναλλακτικής $H_1: \mu_1 \neq \mu_2$ εάν

$$|Z| = \left| \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \right| \geq z_{\alpha/2}$$

- <u>οι δύο πληθυσμοί ακολουθούν κανονική κατανομή με άγνωστες και ίσες διακυμάνσεις:</u> Σε αυτήν την περίπτωση απορρίπτουμε την μηδενική υπόθεση $H_0: \mu_1 = \mu_2$ έναντι της εναλλακτικής $H_1: \mu_1 \neq \mu_2$ εάν

$$|t| = \left| \frac{\bar{X} - \bar{Y}}{S\sqrt{\frac{1}{n} + \frac{1}{m}}} \right| \geq t_{n+m-2;\alpha/2}, \quad S = \sqrt{\frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}}$$

- <u>οι δύο πληθυσμοί ακολουθούν κανονική κατανομή με άγνωστες και άνισες διακυμάνσεις:</u> Σε αυτήν την περίπτωση απορρίπτουμε την μηδενική υπόθεση $H_0: \mu_1 = \mu_2$ έναντι της εναλλακτικής $H_1: \mu_1 \neq \mu_2$ εάν

$$|t| = \left| \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \right| \geq t_{\nu;\alpha/2}, \quad \nu = \frac{(S_1^2/n + S_2^2/m)^2}{\frac{(S_1^2/n)^2}{n-1} + \frac{(S_2^2/m)^2}{m-1}}$$

- <u>οι δύο πληθυσμοί είναι μη κανονικοί με γνωστές ή άγνωστες διακυμάνσεις:</u> Οι τρεις περιπτώσεις ελέγχων υποθέσεων διαφοράς μέσων τιμών που είδαμε παραπάνω είχαν ως προϋπόθεση την κανονικότητα των υπό μελέτη πληθυσμών. Στην περίπτωση που η κανονικότητα δεν ισχύει, οι έλεγχοι υποθέσεων μπορούν να διεξαχθούν μόνο όταν είναι εφικτή η επίκληση του κεντρικού οριακού θεωρήματος, όταν δηλαδή τα δείγματα που έχουμε είναι αρκετά μεγάλα (συμβατικά, $n, m \geq 30$). Σε αυτήν την περίπτωση:
  - αν οι διακυμάνσεις των πληθυσμών είναι <u>γνωστές</u>, τότε απορρίπτουμε την μηδενική υπόθεση $H_0: \mu_1 = \mu_2$ έναντι της εναλλακτικής $H_1: \mu_1 \neq \mu_2$ εάν

$$|Z| = \left| \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \right| \geq z_{\alpha/2}$$

  - αν οι διακυμάνσεις των πληθυσμών είναι <u>άγνωστες</u>, τότε απορρίπτουμε την μηδενική υπόθεση $H_0: \mu_1 = \mu_2$ έναντι της εναλλακτικής $H_1: \mu_1 \neq \mu_2$ εάν

$$|Z| = \left| \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \right| \geq z_{\alpha/2}$$

Εφαρμόζοντας αυτόν τον έλεγχο υποθέσεων στα δεδομένα μας, παίρνουμε τα παρακάτω αποτελέσματα:

- <u>Ηλικία ασθενών</u>

Πληθυσμός 1:	έρευνα 2008-2009
Πληθυσμός 2:	έρευνα 2019
n:	86
m:	88
σ_1^2:	άγνωστη
σ_2^2:	άγνωστη
κανονικότητα πληθυσμού 1:	ΝΑΙ ($p - value\ ελέγχου\ Lilliefors = 0.429 > \alpha = 0.05$)
κανονικότητα πληθυσμού 2:	ΝΑΙ ($p - value\ ελέγχου\ Lilliefors = 0.3557 > \alpha = 0.05$)
ισότητα διακυμάνσεων:	ΟΧΙ ($p - value\ ελέγχου\ Levene = 0.001734 < \alpha = 0.05$)
Περίπτωση ελέγχου:	δύο κανονικοί πληθυσμοί με άγνωστες και άνισες διακυμάνσεις
p-value ελέγχου:	$\approx 0\ (< \alpha = 0.05)$
Απόφαση:	Απόρριψη H_0. Άρα στατιστική σημαντικότητα της μεταβολής της ηλικίας των ασθενών.

- <u>Ηλικία πρώτης κρίσης</u>

Πληθυσμός 1:	έρευνα 2008-2009
Πληθυσμός 2:	έρευνα 2019
n:	88
m:	96
σ_1^2:	άγνωστη
σ_2^2:	άγνωστη
κανονικότητα πληθυσμού 1:	ΝΑΙ ($p - value\ ελέγχου\ Lilliefors = 0.06466 > \alpha = 0.05$)
κανονικότητα πληθυσμού 2:	ΟΧΙ ($p - value\ ελέγχου\ Lilliefors \approx 0 < \alpha = 0.05$)
Περίπτωση ελέγχου:	δύο μη κανονικοί πληθυσμοί με άγνωστες διακυμάνσεις και μεγάλα μεγέθη δειγμάτων
p-value ελέγχου:	$0.0026\ (< \alpha = 0.05)$
Απόφαση:	Απόρριψη H_0. Άρα στατιστική σημαντικότητα της μεταβολής της ηλικίας πρώτης κρίσης.

- Αριθμός κρίσεων το τελευταίο δίμηνο

Πληθυσμός 1:	έρευνα 2008-2009
Πληθυσμός 2:	έρευνα 2019
n:	87
m:	93
σ_1^2:	άγνωστη
σ_2^2:	άγνωστη
κανονικότητα πληθυσμού 1:	ΟΧΙ ($p-value\ ελέγχου\ Lilliefors \approx 0 < \alpha = 0.05$)
κανονικότητα πληθυσμού 2:	ΟΧΙ ($p-value\ ελέγχου\ Lilliefors \approx 0 < \alpha = 0.05$)
Περίπτωση ελέγχου:	δύο μη κανονικοί πληθυσμοί με άγνωστες διακυμάνσεις και μεγάλα μεγέθη δειγμάτων
p-value ελέγχου:	$0.2990\ (> \alpha = 0.05)$
Απόφαση:	Μη απόρριψη H_0. Άρα μη στατιστική σημαντικότητα της μεταβολής του αριθμού κρίσεων το τελευταίο δίμηνο.

Συμπεραίνουμε, λοιπόν, ότι σε επίπεδο σημαντικότητας α=5% η αύξηση της μέσης ηλικίας των συμμετεχόντων κατά 7.16 έτη και η αύξηση της μέσης ηλικίας πρώτης κρίσης κατά 2.12 έτη είναι στατιστικά σημαντικές, ενώ η μείωση του μέσου αριθμού κρίσεων το τελευταίο δίμηνο κατά 5 δεν είναι στατιστικά σημαντική.

2.2.3. Στατιστική σημαντικότητα των μεταβολών της τελευταίας δεκαετίας για τις ποιοτικές μεταβλητές

Για να ελέγξουμε την στατιστική σημαντικότητα των μεταβολών της τελευταίας δεκαετίας για τις ποιοτικές μεταβλητές, θα χρησιμοποιήσουμε ελέγχους υποθέσεων για τη διαφορά των ποσοστών δύο ανεξάρτητων πληθυσμών. Πιο συγκεκριμένα, μας ενδιαφέρει ο έλεγχος $H_0: p_1 = p_2$ έναντι $H_1: p_1 \neq p_2$. Στην επόμενη παράγραφο δίνουμε κάποια βασικά θεωρητικά στοιχεία για αυτόν τον έλεγχο.

Θεωρούμε ότι έχουμε δύο πληθυσμούς, τα μέλη των οποίων κατατάσσονται σε δύο κατηγορίες. Συμβολίζουμε με p_1 και p_2 τις πιθανότητες των μελών να ανήκουν στις δύο κατηγορίες αντίστοιχα. Έστω ότι επιλέγουμε ένα τυχαίο δείγμα, μεγέθους n_1 από τον πρώτο πληθυσμό, το οποίο περιέχει X_1 άτομα που ανήκουν στην πρώτη κατηγορία. Έστω επίσης ότι επιλέγουμε ένα δεύτερο ανεξάρτητο δείγμα μεγέθους n_2 από τον δεύτερο πληθυσμό, το οποίο περιέχει X_2 άτομα που ανήκουν στη δεύτερη κατηγορία. Σε αυτήν την περίπτωση και όταν τα μεγέθη των δειγμάτων n_1 και n_2

είναι μεγάλα, απορρίπτουμε την μηδενική υπόθεση $H_0: p_1 = p_2$ έναντι της εναλλακτικής $H_1: p_1 \neq p_2$ εάν

$$|Z| = \left| \frac{X_1/n_1 - X_2/n_2}{\sqrt{\frac{X_1+X_2}{n_1+n_2}\left(1 - \frac{X_1+X_2}{n_1+n_2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right| \geq z_{\alpha/2}$$

Εφαρμόζοντας αυτόν τον έλεγχο υποθέσεων στα δεδομένα μας, παίρνουμε τα παρακάτω αποτελέσματα:

Κατηγορία	p-value ελέγχου
καμία κρίση τον περασμένο χρόνο	0.1423
πάνω από 1 κρίση την εβδομάδα τον περασμένο χρόνο	0.1483
άλλο νόσημα εκτός από τις κρίσεις	≈ 0
δεν μπορούν να εμποδίσουν την κρίση	0.007772
δεν τους ενοχλεί καθόλου να τους αποκαλούν «επιληπτικούς»	0.03886
τους ενοχλεί πολύ να τους αποκαλούν «επιληπτικούς»	0.03733
η επιληψία δεν τους εμποδίζει καθόλου στην ανεύρεση εργασίας	0.5853
η επιληψία τους εμποδίζει λίγο στην ανεύρεση εργασίας	0.2396
άγαμοι	0.03019
παντρεμένοι	0.02295
δεν πίνουν ποτέ κρασί	0.01271
πίνουν σπάνια κρασί	0.03021
βγαίνουν έξω τα βράδια λιγότερο από μία φορά το δίμηνο	0.06213
βγαίνουν έξω τα βράδια μία φορά την εβδομάδα	0.1792
επισκέπτονται τον γιατρό μία φορά κάθε τρεις ως έξι μήνες	0.0368
επισκέπτονται τον γιατρό μία φορά το εξάμηνο ως μία φορά το χρόνο	0.07305
έχουν πάνω από ένα χρόνο να επισκεφθούν τον γιατρό	0.007445
επισκέπτονται τον γιατρό μία φορά κάθε ένα ως τρεις μήνες	0.2397
δεν φοβούνται τις κρίσεις	0.1358
φοβούνται λίγο τις κρίσεις	0.04818
δεν νιώθουν καθόλου ανασφάλεια για το μέλλον εξαιτίας των κρίσεων	0.1555
νιώθουν αρκετή ανασφάλεια για το μέλλον εξαιτίας των κρίσεων	0.4958
νιώθουν πολλή ανασφάλεια για το μέλλον εξαιτίας των κρίσεων	0.4516
επιδιώκουν να αποκτήσουν νέους φίλους	0.233
δεν επιδιώκουν να αποκτήσουν νέους φίλους	0.1063
η επιληψία δεν επηρεάζει καθόλου τις σχέσεις τους με το άλλο φύλο	0.3348
η επιληψία επηρεάζει αρκετά τις σχέσεις τους με το άλλο φύλο	0.3873

Πίνακας 2.1: Αποτελέσματα ελέγχων υποθέσεων για την στατιστική σημαντικότητα των μεταβολών των ποιοτικών μεταβλητών

Από τον Πίνακα 2.1 παρατηρούμε ότι οι μεταβολές των ποιοτικών μεταβλητών που είναι στατιστικά σημαντικές σε επίπεδο σημαντικότητας α=5% (στον πίνακα είναι σημειωμένες με μπλε χρώμα) είναι οι εξής:

- η αύξηση του ποσοστού των συμμετεχόντων που πάσχουν από κάποιο άλλο νόσημα εκτός από τις κρίσεις κατά 8.9%.
- η αύξηση του ποσοστού των συμμετεχόντων που δεν μπορούν να εμποδίσουν την κρίση κατά 18.1%.
- η αύξηση του ποσοστού των συμμετεχόντων που δεν τους ενοχλεί καθόλου να τους αποκαλούν «επιληπτικούς» κατά 16%.
- η μείωση του ποσοστού των συμμετεχόντων που τους ενοχλεί πολύ να τους αποκαλούν «επιληπτικούς» κατά 14%.
- η μείωση του ποσοστού των συμμετεχόντων που είναι άγαμοι κατά 15.3%.
- η αύξηση του ποσοστού των συμμετεχόντων που είναι παντρεμένοι κατά 15.4%.
- η μείωση του ποσοστού των συμμετεχόντων που δεν πίνουν ποτέ κρασί κατά 18.8%.
- η αύξηση του ποσοστού των συμμετεχόντων που σπάνια πίνουν κρασί κατά 16.6%.
- η μείωση του ποσοστού των συμμετεχόντων που επισκέπτονται τον γιατρό τους μία φορά κάθε τρεις ως έξι μήνες κατά 13.8%.
- η αύξηση του ποσοστού των συμμετεχόντων που έχουν πάνω από ένα χρόνο να επισκεφθούν τον γιατρό τους κατά 13.5%.
- η μείωση του ποσοστού των συμμετεχόντων που φοβούνται λίγο τις κρίσεις κατά 13.9%.

2.3. Σύγκριση των αποτελεσμάτων της παρούσας έρευνας με τα αποτελέσματα παρόμοιων ερευνών που πραγματοποιήθηκαν στην Κολομβία και στη Νέα Ζηλανδία

Στην ενότητα αυτή συγκρίνουμε τα αποτελέσματα της παρούσας έρευνας με τα αποτελέσματα παρόμοιων ερευνών που πραγματοποιήθηκαν στην Κολομβία και στη Νέα Ζηλανδία.

Σύγκριση της παρούσας έρευνας με την έρευνα της Κολομβίας

Η έρευνα της Κολομβίας αφορούσε ερωτηματολόγιο, το οποίο απαντήθηκε από 354 ασθενείς που παρακολουθούνταν σε εξειδικευμένο νευρολογικό κέντρο επιληψίας κατά τη διάρκεια των ετών 2013-2016. Τα αποτελέσματα της έρευνας τα λαμβάνουμε από το άρθρο *«Clinical and sociodemographic profile of epilepsy in adults from a reference centre in Colombia»*, το οποίο δημοσιεύτηκε στις 4 Φεβρουαρίου 2017[3]. Στον Πίνακα 2.2 παρουσιάζουμε τα αποτελέσματα από την ανάλυση παρόμοιων ερωτήσεων των δύο ερευνών.

	Ελλάδα	Κολομβία
Μέση ηλικία συμμετεχόντων	38.45	37
Διάμεση ηλικία πρώτης κρίσης	15	11
Επαγγελματική κατάσταση		
εργαζόμενοι	47.87%	15%
μη εργαζόμενοι	21.28%	48%
σπουδαστές, συνταξιούχοι	22.34%	13%
Μορφωτικό επίπεδο		
μη ολοκλήρωση δευτεροβάθμιας εκπαίδευσης	11.46%	16%
δευτεροβάθμια εκπαίδευση	45.83%	16%
πανεπιστημιακή εκπαίδευση	34.38%	12.6%

Πίνακας 2.2: Αποτελέσματα ανάλυσης παρόμοιων ερωτήσεων για τις έρευνες της Ελλάδας και της Κολομβίας

Από τον Πίνακα 2.2 παρατηρούμε ότι:

- δεν υπάρχει μεγάλη διαφορά ως προς την μέση ηλικία των συμμετεχόντων στις δύο έρευνες.
- η διάμεση ηλικία πρώτης κρίσης για τους ασθενείς της Ελλάδας είναι μεγαλύτερη κατά 4 έτη από την διάμεση ηλικία πρώτης κρίσης για τους ασθενείς της Κολομβίας.

[3] Το άρθρο είναι διαθέσιμο στην ιστοσελίδα https://reader.elsevier.com/reader/sd/pii/S2173580819300185?token=72D803B011FCB7906563BB6D7A1DD11E27499611EED680909D2B4F669A4CEDFC4EDFFB3FDECD5634AD56175714905659&originRegion=eu-west-1&originCreation=20210615203231

- υπάρχουν τεράστιες διαφορές ως προς την επαγγελματική κατάσταση των επιληπτικών ασθενών στις δύο χώρες. Πιο συγκεκριμένα:
  - το ποσοστό των εργαζόμενων ασθενών στην Ελλάδα είναι αυξημένο κατά 32.87% σε σχέση με το αντίστοιχο ποσοστό στην Κολομβία.
  - το ποσοστό των μη εργαζόμενων ασθενών στην Ελλάδα είναι μειωμένο κατά 26.72% σε σχέση με το αντίστοιχο ποσοστό στην Κολομβία.
  - το ποσοστό των σπουδαστών και συνταξιούχων στην Ελλάδα είναι αυξημένο κατά 9.34% σε σχέση με το αντίστοιχο ποσοστό στην Κολομβία.
- υπάρχουν τεράστιες διαφορές ως προς το μορφωτικό επίπεδο των επιληπτικών ασθενών στις δύο χώρες. Πιο συγκεκριμένα:
  - το ποσοστό των Ελλήνων ασθενών που δεν ολοκλήρωσαν την δευτεροβάθμια εκπαίδευση είναι μειωμένο κατά 4.54% σε σχέση με το αντίστοιχο ποσοστό των Κολομβιανών ασθενών.
  - το ποσοστό των Ελλήνων ασθενών που ολοκλήρωσαν την δευτεροβάθμια εκπαίδευση είναι αυξημένο κατά 29.83% σε σχέση με το αντίστοιχο ποσοστό των Κολομβιανών ασθενών.
  - το ποσοστό των Ελλήνων ασθενών που έχουν πανεπιστημιακή εκπαίδευση είναι αυξημένο κατά 21.78% σε σχέση με το αντίστοιχο ποσοστό των Κολομβιανών ασθενών.

<u>Σύγκριση της παρούσας έρευνας με την έρευνα της Νέας Ζηλανδίας</u>

Η έρευνα της Νέας Ζηλανδίας αφορούσε ερωτηματολόγιο, το οποίο απαντήθηκε από 276 επιληπτικούς ασθενείς που παρακολουθούνταν σε διάφορα νοσοκομεία της χώρας κατά τη διάρκεια του 2018. Τα αποτελέσματα της έρευνας τα λαμβάνουμε από την ιστοσελίδα του Υπουργείου Υγείας της Νέας Ζηλανδίας[4]. Στον Πίνακα 2.3 παρουσιάζουμε τα αποτελέσματα από την ανάλυση παρόμοιων ερωτήσεων των δύο ερευνών.

[4] Τα αποτελέσματα της έρευνας είναι διαθέσιμα στην ιστοσελίδα https://www.health.govt.nz/system/files/documents/publications/epilepsy-consumer-experience-survey-2018-mar19.pdf

	Ελλάδα	Νέα Ζηλανδία
Φύλο		
άνδρας	61.5%	40.58%
γυναίκα	38.5%	56.88%
Ικανότητα ελέγχου των κρίσεων		
όχι	84.04%	37%
ναι	15.96%	63%
Επαγγελματική κατάσταση		
εργαζόμενοι	47.87%	37%
μη εργαζόμενοι	21.28%	28%

Πίνακας 2.3: Αποτελέσματα ανάλυσης παρόμοιων ερωτήσεων για τις έρευνες της Ελλάδας και της Νέας Ζηλανδίας

Από τον Πίνακα 2.3 παρατηρούμε ότι:

- στην έρευνα της Ελλάδας συμμετείχαν περισσότεροι άνδρες, ενώ στην έρευνα της Νέας Ζηλανδίας συμμετείχαν περισσότερες γυναίκες.
- οι περισσότεροι Έλληνες ασθενείς δεν μπορούν να ελέγξουν τις κρίσεις τους, ενώ οι περισσότεροι Νεοζηλανδοί ασθενείς μπορούν.
- το ποσοστό των εργαζόμενων ασθενών στην Ελλάδα είναι αυξημένο κατά 10.87% σε σχέση με το αντίστοιχο ποσοστό στη Νέα Ζηλανδία, ενώ το ποσοστό των μη εργαζόμενων ασθενών στην Ελλάδα είναι μειωμένο κατά 6.72% σε σχέση με το αντίστοιχο ποσοστό στη Νέα Ζηλανδία.

ΚΕΦΑΛΑΙΟ 3
Σχέσεις των μεταβλητών ανά δύο

Στο κεφάλαιο αυτό ελέγχουμε την ύπαρξη συσχέτισης μεταξύ των μεταβλητών της παρούσας έρευνας και της ύπαρξης κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου, αφού πρώτα δημιουργήσουμε αυτή την μεταβλητή.

3.1. Δημιουργία των μεταβλητών που αφορούν την ύπαρξη κρίσεων το τελευταίο δίμηνο και την ύπαρξη κρίσεων τον περασμένο χρόνο

Δημιουργούμε την κατηγορική μεταβλητή kriseis2, η οποία αφορά την ύπαρξη κρίσεων το τελευταίο δίμηνο και παίρνει τις εξής τιμές:

- 0, αν το άτομο δεν είχε καμία κρίση το τελευταίο δίμηνο
- 1, αν το άτομο εμφάνισε κρίσεις το τελευταίο δίμηνο

Επίσης, δημιουργούμε την κατηγορική μεταβλητή kriseis_lastyear, η οποία αφορά την ύπαρξη κρίσεων τον περασμένο χρόνο και παίρνει τις εξής τιμές:

- 0, αν το άτομο δεν είχε καμία κρίση τον περασμένο χρόνο
- 1, αν το άτομο εμφάνισε κρίσεις τον περασμένο χρόνο

3.2. Βασικά θεωρητικά στοιχεία για τους ελέγχους συσχέτισης μεταξύ ποιοτικών μεταβλητών

Η εύρεση της πιθανής σχέσης μεταξύ δύο ποιοτικών μεταβλητών επιτυγχάνεται μέσω της δημιουργίας του πίνακα συνάφειας, ο οποίος είναι διδιάστατος (στο επίπεδο) με r το πλήθος γραμμές, όσες είναι οι κατηγορίες της μίας ποιοτικής μεταβλητής, και c το πλήθος στήλες, όσες είναι οι κατηγορίες της άλλης ποιοτικής μεταβλητής. Έτσι, δημιουργούνται $r \times c$ κελιά (κυψελίδες), κάθε ένα από τα οποία παριστάνει έναν συνδυασμό των τιμών των δύο μεταβλητών και στα οποία καταγράφονται οι παρατηρούμενες συχνότητες εμφάνισής τους. Με O_{ij} συμβολίζουμε την συχνότητα εμφάνισης του κελιού (i, j), δηλαδή αυτού που δημιουργείται από την γραμμή i και την στήλη j. Τα αθροίσματα $O_{i.}$, $i = 1, \dots, r$ των γραμμών του πίνακα και τα αθροίσματα $O_{.j}$, $j = 1, \dots, c$ των στηλών του πίνακα ονομάζονται περιθώρια αθροίσματα. Ένα παράδειγμα 2×3 πίνακα συνάφειας δίνεται στον Πίνακα 3.1.

Μεταβλητή 1	Μεταβλητή 2			
	Κατηγορία 1	*Κατηγορία 2*	*Κατηγορία 3*	Σύνολα
Κατηγορία 1	O_{11}	O_{12}	O_{13}	$O_{1.}$
Κατηγορία 2	O_{21}	O_{22}	O_{23}	$O_{2.}$
Σύνολα	$O_{.1}$	$O_{.2}$	$O_{.3}$	

Πίνακας 3.1: Πίνακας συνάφειας $\mathbf{2 \times 3}$

Ο έλεγχος της ύπαρξης ή όχι ανεξαρτησίας μεταξύ δύο ποιοτικών μεταβλητών υλοποιείται με το X^2 στατιστικό τεστ του Pearson που δίνεται από τη σχέση

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

όπου E_{ij} είναι η αναμενόμενη συχνότητα του κελιού (i, j), δηλαδή ο αριθμός των περιπτώσεων του κελιού (i, j) αν οι προς μελέτη μεταβλητές ήταν στατιστικά ανεξάρτητες. Η αναμενόμενη συχνότητα E_{ij} δίνεται από τη σχέση

$$E_{ij} = \frac{\sum_{i=1}^{r} O_{ij} \sum_{j=1}^{c} O_{ij}}{\sum_{i=1}^{r}\sum_{j=1}^{c} O_{ij}} = \frac{\sum_{i=1}^{r} O_{ij} \sum_{j=1}^{c} O_{ij}}{n},$$

όπου n το μέγεθος του δείγματος. Μεγάλες αποκλίσεις των αναμενόμενων τιμών από τις παρατηρούμενες τιμές υποδηλώνει πιθανή ύπαρξη σχέσης εξάρτησης. Η μηδενική υπόθεση του ελέγχου είναι ότι οι δύο μεταβλητές είναι ανεξάρτητες, ενώ η εναλλακτική υπόθεση είναι ότι είναι εξαρτημένες. Η υπόθεση της ανεξαρτησίας απορρίπτεται, σε επίπεδο σημαντικότητας α, όταν $X^2 \geq X^2_{(r-1)(c-1),\alpha}$ (ή όταν $p - value < a$). Σε περίπτωση που η υπόθεση της ανεξαρτησίας απορρίπτεται, τότε μπορούμε να διαπιστώσουμε ποια κελιά «δημιουργούν» το πρόβλημα της εξάρτησης των δύο μεταβλητών υπολογίζοντας τις τιμές των προσαρμοσμένων τυποποιημένων καταλοίπων (adjusted standardized residuals). Τα προσαρμοσμένα τυποποιημένα κατάλοιπα δίνονται από τη σχέση

$$d_{ij} = \frac{(O_{ij} - E_{ij})/\sqrt{E_{ij}}}{\sqrt{\left(1 - \frac{O_{i.}}{n}\right)\left(1 - \frac{O_{.j}}{n}\right)}}$$

και ακολουθούν κατά προσέγγιση την κανονική κατανομή όταν οι μεταβλητές του πίνακα συνάφειας είναι ανεξάρτητες μεταξύ τους. Επομένως, μπορούν να θεωρηθούν ως z-τιμές και τιμές αυτών μεγαλύτερες κατά απόλυτη τιμή από το $1.96 = z_{0.025}$ υποδεικνύουν κελιά που διαφέρουν σαφώς από το μοντέλο της ανεξαρτησίας (για επίπεδο σημαντικότητας α=5%).

Για να μπορέσει να εφαρμοστεί ο έλεγχος ανεξαρτησίας X^2 του Pearson θα πρέπει να ισχύουν οι παρακάτω προϋποθέσεις:

1. Το μέγεθος του δείγματος δεν πρέπει να είναι μικρότερο του τετραπλασίου του αριθμού των κελιών του πίνακα συνάφειας.
2. Καμία από τις αναμενόμενες συχνότητες E_{ij} δεν πρέπει να είναι μικρότερη του 1.
3. Το ποσοστό των αναμενόμενων συχνοτήτων E_{ij} οι οποίες είναι μικρότερες του 5 δεν μπορεί να είναι μεγαλύτερο του 20% με 25%.

Στην περίπτωση που δεν ισχύουν οι παραπάνω προϋποθέσεις, τότε για 2×2 πίνακες χρησιμοποιείται ο ακριβής έλεγχος του Fisher (Fisher's exact test), ενώ σε κάθε άλλη περίπτωση υπολογίζουμε το p-value του ελέγχου ανεξαρτησίας X^2 του Pearson με προσομοίωση Monte Carlo.

Ο έλεγχος X^2 που είδαμε παραπάνω, εφαρμόζεται μόνο όταν και οι δύο ποιοτικές μεταβλητές είναι κατηγορικές. Εάν μία από τις μεταβλητές είναι διατάξιμη, τότε δεν εξετάζουμε εάν υπάρχει εξάρτηση μεταξύ των δύο μεταβλητών αλλά εάν υπάρχει γραμμική τάση. Στην ουσία δηλαδή εξετάζουμε εάν υπάρχει γραμμική τάση στα ποσοστά ως προς τις κατηγορίες της διατάξιμης μεταβλητής. Στην περίπτωση αυτή χρησιμοποιούμε τον έλεγχο X^2 για τάση. Η μηδενική υπόθεση του ελέγχου είναι ότι δεν υπάρχει γραμμική τάση στα διωνυμικά ποσοστά καθώς αυξάνονται οι κατηγορίες της διατάξιμης μεταβλητής, ενώ η εναλλακτική υπόθεση είναι ότι υπάρχει γραμμική τάση. Μαθηματικά, η μηδενική και η εναλλακτική υπόθεση μπορούν να γραφούν ως

$$H_0: p_{1|1} = p_{1|2} = \ldots = p_{1|k}$$
$$H_1: \Delta\varepsilon\nu\ \iota\sigma\chi\acute{\upsilon}\varepsilon\iota\ \eta\ H_0\ ,$$

όπου $p_{1|i} = a + bx_i$ με το x_i να είναι τα σκορ που αποδίδονται στις κατηγορίες της διατάξιμης μεταβλητής. Εάν η μεταβλητή ταξινόμησης είναι διατάξιμη, τότε ως σκορ επιλέγουμε τις τιμές $1,2,\ldots,k$, τα οποία ονομάζονται row scores, ενώ εάν η μεταβλητή ταξινόμησης είναι διαστηματική, τότε ως σκορ χρησιμοποιούμε τα κέντρα των ομάδων. Μια τρίτη επιλογή είναι να χρησιμοποιήσουμε το μέσο των τάξεων (midrank) των περιπτώσεών της στην πλήρη ταξινόμηση του δείγματος. Η στατιστική συνάρτηση του ελέγχου είναι η

$$X^2_{\tau\acute{\alpha}\sigma\eta} = \frac{\left[\sum_{i=1}^{k} r_i x_i - R\bar{x}\right]^2}{pq\left[\sum_{i=1}^{k} n_i {x_i}^2 - N\bar{x}^2\right]}\ ,$$

όπου r_i είναι η συχνότητα του «ναι» στο επίπεδο x_i, n_i είναι ο αριθμός των ατόμων στην κατηγορία αυτή, και

- $N = \sum_{i=1}^{k} n_i$
- $R = \sum_{i=1}^{k} r_i$
- $p = \frac{R}{N}$

- $q = 1 - p$
- $\bar{x} = \frac{\sum_{i=1}^{k} n_i x_i}{N}$

Η μηδενική υπόθεση απορρίπτεται, σε επίπεδο σημαντικότητας α, όταν $X^2_{τάση} \geq X^2_{1,\alpha}$ (ή όταν $p - value < a$).

Θέλοντας να διερευνηθεί η ένταση της σχέσης δύο κατηγορικών μεταβλητών είναι διαθέσιμα πλήθος στατιστικών μέτρων. Τα πιο βασικά από αυτά, τα οποία είναι όλα συναρτήσεις του X^2 στατιστικού τεστ του Pearson, είναι τα εξής:

- ο συντελεστής Φ του Pearson,

$$\Phi = \sqrt{\frac{X^2}{n}}.$$

  Παίρνει τιμές από το 0 έως το $\sqrt{q-1}$, όπου $q = \min(r, c)$. Το r είναι ο αριθμός των γραμμών και το c είναι ο αριθμός των στηλών. Για 2×2 πίνακες συνάφειας το τετράγωνο του μέτρου αυτού είναι γνωστό ως το μέτρο τ των Goodman και Kruskal, ισχύει δηλαδή $\tau = \Phi^2$. Εάν ο συντελεστής Φ πάρει την τιμή 0, τότε οι μεταβλητές είναι μεταξύ τους ανεξάρτητες.

- ο συντελεστής C του Pearson,

$$C = \sqrt{\frac{X^2}{X^2 + n}}.$$

  Παίρνει τιμές από το 0 έως το $\sqrt{\frac{q-1}{q}}$, όπου $q = \min(r, c)$. Εάν πάρει την τιμή 0, τότε οι μεταβλητές είναι μεταξύ τους ανεξάρτητες.

- ο συντελεστής V του Cramér,

$$V = \sqrt{\frac{X^2}{n * \min(r-1, c-1)}}.$$

  Παίρνει τιμές από το 0 έως το 1. Εάν πάρει:
  - την τιμή 0, τότε οι δύο μεταβλητές είναι μεταξύ τους ανεξάρτητες.
  - τιμή μέσα στο διάστημα (0, 0.10), τότε υπάρχει χαμηλή συνάφεια μεταξύ των δύο μεταβλητών.
  - τιμή μέσα στο διάστημα [0.10, 0.30], τότε υπάρχει μέτρια συνάφεια μεταξύ των δύο μεταβλητών.
  - τιμή μέσα στο διάστημα (0.30, 1], τότε υπάρχει ισχυρή συνάφεια μεταξύ των δύο μεταβλητών.

3.3. Αποτελέσματα ελέγχων συσχέτισης μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και των υπόλοιπων μεταβλητών της έρευνας

Οι μεταβλητές με τις οποίες δεν συσχετίζεται η ύπαρξη κρίσεων το τελευταίο δίμηνο, σε επίπεδο σημαντικότητας α=5%, καθώς και τα αποτελέσματα των ελέγχων παρουσιάζονται στον Πίνακα 3.2.

Μεταβλητή	Έλεγχος συσχέτισης	Μέγεθος δείγματος	p-value ελέγχου
Φύλο	X^2 του Pearson	93	0.755
Ηλικιακή ομάδα	X^2 τάσης	85	0.417
Τόπος μόνιμης κατοικίας μέχρι την ηλικία των 18 ετών	X^2 του Pearson με προσομοίωση Monte Carlo	92	0.402
Τόπος μόνιμης τωρινής κατοικίας	X^2 του Pearson με προσομοίωση Monte Carlo	93	0.894
Ηλικιακή ομάδα πρώτης κρίσης	X^2 τάσης	92	0.188
Επανάληψη κρίσεων ίδιου τύπου	X^2 του Pearson	92	0.196
Νοσηλεία λόγω των κρίσεων τον τελευταίο χρόνο	ακριβής έλεγχος του Fisher	93	0.083
Ιστορικό επιληψίας στο οικογενειακό περιβάλλον	ακριβής έλεγχος του Fisher	92	0.751
Ύπαρξη άλλου νοσήματος εκτός από τις κρίσεις	ακριβής έλεγχος του Fisher	92	1
Ικανότητα ελέγχου των κρίσεων	X^2 του Pearson	91	0.227
Βαθμός αποδοχής της επιληψίας	X^2 τάσης	91	0.117
Κοινωνική αντιμετώπιση	X^2 τάσης	85	0.901
Επαγγελματική κατάσταση	X^2 του Pearson με προσομοίωση Monte Carlo	91	0.315
Βαθμός δυσκολίας ανεύρεσης εργασίας εξαιτίας της νόσου	X^2 του Pearson με προσομοίωση Monte Carlo	83	0.091
Μορφωτικό επίπεδο	X^2 τάσης με προσομοίωση Monte Carlo	93	0.411
Διακοπή των σπουδών εξαιτίας της νόσου	ακριβής έλεγχος του Fisher	89	1
Ύπαρξη παιδιών	ακριβής έλεγχος του Fisher	92	0.257
Συχνότητα κατανάλωσης κρασιού	X^2 τάσης με προσομοίωση Monte Carlo	93	0.161

Συχνότητα κατανάλωσης ούζου, ουίσκι, βότκας, τζιν, κονιάκ	X^2 τάσης με προσομοίωση Monte Carlo	93	0.060[5]
Συχνότητα νυχτερινών εξόδων	X^2 τάσης	93	0.061[6]
Δίπλωμα οδήγησης	X^2 του Pearson	93	0.767
Οδήγηση	X^2 του Pearson	93	0.833
Βαθμός μοναξιάς	X^2 τάσης	91	0.587
Συχνότητα επισκέψεων στον γιατρό	X^2 τάσης	92	0.071
Ανάγκη για περισσότερη ενημέρωση για την νόσο	X^2 του Pearson με προσομοίωση Monte Carlo	92	0.104
Βαθμός φόβου των κρίσεων	X^2 του Pearson	91	0.529
Βαθμός ανασφάλειας για το μέλλον εξαιτίας των κρίσεων	X^2 του Pearson	92	0.337
Επιδίωξη απόκτησης νέων φίλων	X^2 του Pearson με προσομοίωση Monte Carlo	92	0.937
Βαθμός επίδρασης της νόσου στις σχέσεις με το άλλο φύλο	X^2 του Pearson με προσομοίωση Monte Carlo	90	0.830

Πίνακας 3.2: Αποτελέσματα ελέγχων συσχέτισης μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και των μεταβλητών με τις οποίες αυτή δεν συσχετίζεται

Οι μεταβλητές με τις οποίες συσχετίζεται η ύπαρξη κρίσεων το τελευταίο δίμηνο, σε επίπεδο σημαντικότητας α=5%, είναι οι εξής:

- συνολικός αριθμός κρίσεων
- αριθμός επισκέψεων στον γιατρό τον τελευταίο χρόνο
- ύπαρξη κρίσεων τον περασμένο χρόνο
- οικογενειακή κατάσταση

Στους Πίνακες 3.3-3.14 παρουσιάζονται οι πίνακες συνάφειας που προκύπτουν, καθώς και τα αποτελέσματα των ελέγχων.

[5] Το αποτέλεσμα προέκυψε μετά την συγχώνευση των εξής κατηγοριών της συχνότητας κατανάλωσης ούζου, ουίσκι, βότκας, τζιν, κονιάκ: «σπάνια» με «μία-δύο φορές το μήνα» και «μία φορά την εβδομάδα» με «δύο-τρεις φορές την εβδομάδα».

[6] Το αποτέλεσμα προέκυψε μετά την συγχώνευση των εξής κατηγοριών της συχνότητας νυχτερινών εξόδων: «λιγότερο από μία φορά το δίμηνο» με «μία φορά το δίμηνο», «μία φορά το μήνα» με «2-3 φορές το μήνα» και «μία φορά την εβδομάδα» με «2-3 φορές την εβδομάδα».

Συσχέτιση μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και του συνολικού αριθμού κρίσεων

΄Υπαρξη κρίσεων το τελευταίο δίμηνο * Συνολικές κρίσεις Crosstabulation

Count

		Συνολικές κρίσεις					
		1	2-5	6-10	πάνω από 21	πάνω από 100	Total
΄Υπαρξη κρίσεων το τελευταίο δίμηνο	όχι	2	11	13	15	12	53
	ναι	0	2	3	9	23	37
Total		2	13	16	24	35	90

Πίνακας 3.3: Πίνακας συνάφειας για την ύπαρξη κρίσεων το τελευταίο δίμηνο και τις συνολικές κρίσεις

Από τον Πίνακα 3.3 παρατηρούμε ότι:

- για την συγκεκριμένη ανάλυση χρησιμοποιήθηκαν οι απαντήσεις 90 ασθενών.
- 53 άτομα δεν εμφάνισαν κρίσεις το τελευταίο δίμηνο, ενώ 37 άτομα εμφάνισαν.
- 2 άτομα έχουν κάνει συνολικά στη ζωή τους 1 κρίση, 13 άτομα έχουν κάνει 2-5 κρίσεις, 16 άτομα έχουν κάνει 6-10 κρίσεις, 24 άτομα έχουν κάνει πάνω από 21 κρίσεις και 35 άτομα έχουν κάνει πάνω από 100 κρίσεις.
- από τα 53 άτομα που δεν εμφάνισαν κρίσεις το τελευταίο δίμηνο: 2 άτομα έχουν κάνει συνολικά στη ζωή τους 1 κρίση, 11 άτομα έχουν κάνει 2-5 κρίσεις, 13 άτομα έχουν κάνει 6-10 κρίσεις, 15 άτομα έχουν κάνει πάνω από 21 κρίσεις και 12 άτομα έχουν κάνει πάνω από 100 κρίσεις.
- από τα 37 άτομα που εμφάνισαν κρίσεις το τελευταίο δίμηνο: 2 άτομα έχουν κάνει συνολικά στη ζωή τους 2-5 κρίσεις, 3 άτομα έχουν κάνει 6-10 κρίσεις, 9 άτομα έχουν κάνει πάνω από 21 κρίσεις, 23 άτομα έχουν κάνει πάνω από 100 κρίσεις, ενώ κανένα άτομο δεν έχει κάνει μόνο 1 κρίση στη ζωή του.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	17,135[a]	4	,002
Likelihood Ratio	18,543	4	,001
Linear-by-Linear Association	15,622	1	,000
N of Valid Cases	90		

a. 2 cells (20,0%) have expected count less than 5. The minimum expected count is ,82.

Πίνακας 3.4: Αποτελέσματα ελέγχων συσχέτισης μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και των συνολικών κρίσεων

Για να ελέγξουμε αν υπάρχει συσχέτιση μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και των συνολικών κρίσεων, θα χρησιμοποιήσουμε τον έλεγχο X^2 τάσης, καθώς η μεταβλητή που αφορά τις συνολικές κρίσεις είναι διατάξιμη. Από τον Πίνακα 3.4 παρατηρούμε ότι σε επίπεδο σημαντικότητας α=5% υπάρχει ισχυρή ένδειξη γραμμικής τάσης του ποσοστού των ατόμων που είχαν κρίσεις το τελευταίο δίμηνο και του αριθμού των συνολικών κρίσεων ($p - value$ *ελέγχου* X^2 τάσης $\approx 0 < \alpha = 0.05$).

<u>Συσχέτιση μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και του αριθμού των επισκέψεων στον γιατρό τον τελευταίο χρόνο</u>

΄Υπαρξη κρίσεων το τελευταίο δίμηνο * Επισκέψεις στον γιατρό τον τελευταίο χρόνο Crosstabulation

Count

		Επισκέψεις στον γιατρό τον τελευταίο χρόνο			
		0-1	2	3-9	Total
΄Υπαρξη κρίσεων το τελευταίο δίμηνο	όχι	28	18	8	54
	ναι	12	11	13	36
Total		40	29	21	90

Πίνακας 3.5: Πίνακας συνάφειας για την ύπαρξη κρίσεων το τελευταίο δίμηνο και τον αριθμό των επισκέψεων στον γιατρό τον τελευταίο χρόνο

Από τον Πίνακα 3.5 παρατηρούμε ότι:

- για την συγκεκριμένη ανάλυση χρησιμοποιήθηκαν οι απαντήσεις 90 ασθενών.
- 54 άτομα δεν εμφάνισαν κρίσεις το τελευταίο δίμηνο, ενώ 36 άτομα εμφάνισαν.
- 40 άτομα επισκέφθηκαν τον γιατρό τους τον τελευταίο χρόνο 0-1 φορές, 29 άτομα τον επισκέφθηκαν 2 φορές και 21 άτομα τον επισκέφθηκαν 3-9 φορές.
- από τα 54 άτομα που δεν εμφάνισαν κρίσεις το τελευταίο δίμηνο: 28 άτομα επισκέφθηκαν τον γιατρό τους τον τελευταίο χρόνο 0-1 φορές, 18 άτομα τον επισκέφθηκαν 2 φορές και 8 άτομα τον επισκέφθηκαν 3-9 φορές.
- από τα 36 άτομα που εμφάνισαν κρίσεις το τελευταίο δίμηνο: 12 άτομα επισκέφθηκαν τον γιατρό τους τον τελευταίο χρόνο 0-1 φορές, 11 άτομα τον επισκέφθηκαν 2 φορές και 13 άτομα τον επισκέφθηκαν 3-9 φορές.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	5,917[a]	2	,052
Likelihood Ratio	5,867	2	,053
Linear-by-Linear Association	5,347	1	,021
N of Valid Cases	90		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 8,40.

Πίνακας 3.6: Αποτελέσματα ελέγχων συσχέτισης μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και του αριθμού των επισκέψεων στον γιατρό τον τελευταίο χρόνο

Για να ελέγξουμε αν υπάρχει συσχέτιση μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και του αριθμού των επισκέψεων στον γιατρό τον τελευταίο χρόνο, θα χρησιμοποιήσουμε τον έλεγχο X^2 τάσης, καθώς η μεταβλητή που αφορά τον αριθμό των επισκέψεων στον γιατρό τον τελευταίο χρόνο είναι διατάξιμη. Από τον Πίνακα 3.6 παρατηρούμε ότι σε επίπεδο σημαντικότητας α=5% υπάρχει ισχυρή ένδειξη γραμμικής τάσης του ποσοστού των ατόμων που είχαν κρίσεις το τελευταίο δίμηνο και του αριθμού των επισκέψεων στον γιατρό τον τελευταίο χρόνο ($p-value$ *ελέγχου* X^2 τάσης $= 0.021 < \alpha = 0.05$).

Συσχέτιση μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και της ύπαρξης κρίσεων τον περασμένο χρόνο

΄Υπαρξη κρίσεων το τελευταίο δίμηνο * ΄Υπαρξη κρίσεων τον τελευταίο χρόνο Crosstabulation

Count

		΄Υπαρξη κρίσεων τον τελευταίο χρόνο		
		όχι	ναι	Total
΄Υπαρξη κρίσεων το τελευταίο δίμηνο	όχι	38	17	55
	ναι	3	34	37
Total		41	51	92

Πίνακας 3.7: Πίνακας συνάφειας για την ύπαρξη κρίσεων το τελευταίο δίμηνο και την ύπαρξη κρίσεων τον περασμένο χρόνο

Από τον Πίνακα 3.7 παρατηρούμε ότι:

- για την συγκεκριμένη ανάλυση χρησιμοποιήθηκαν οι απαντήσεις 92 ασθενών.
- 55 άτομα δεν εμφάνισαν κρίσεις το τελευταίο δίμηνο, ενώ 37 άτομα εμφάνισαν.
- 41 άτομα δεν εμφάνισαν κρίσεις τον περασμένο χρόνο, ενώ 51 άτομα εμφάνισαν.
- από τα 55 άτομα που δεν εμφάνισαν κρίσεις το τελευταίο δίμηνο: 38 άτομα δεν εμφάνισαν κρίσεις τον περασμένο χρόνο, ενώ 17 άτομα εμφάνισαν.
- από τα 37 άτομα που εμφάνισαν κρίσεις το τελευταίο δίμηνο: 3 άτομα δεν εμφάνισαν κρίσεις τον περασμένο χρόνο, ενώ 34 άτομα εμφάνισαν.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	33,298[a]	1	,000		
Continuity Correction[b]	30,875	1	,000		
Likelihood Ratio	37,605	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	32,936	1	,000		
N of Valid Cases	92				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 16,49.

b. Computed only for a 2x2 table

Πίνακας 3.8: Αποτελέσματα ελέγχων συσχέτισης μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και της ύπαρξης κρίσεων τον περασμένο χρόνο

Για να ελέγξουμε αν υπάρχει συσχέτιση μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και της ύπαρξης κρίσεων τον περασμένο χρόνο, θα χρησιμοποιήσουμε τον έλεγχο ανεξαρτησίας X^2 του Pearson (οι προϋποθέσεις του ελέγχου ισχύουν). Από τον Πίνακα 3.8 παρατηρούμε ότι, σε επίπεδο σημαντικότητας α=5%, υπάρχει ισχυρή ένδειξη εξάρτησης μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και της ύπαρξης κρίσεων τον περασμένο χρόνο ($p-value$ ελέγχου X^2 του $Pearson \approx 0 < \alpha = 0.05$). Προκειμένου να διαπιστώσουμε ποια κελιά «δημιουργούν» το πρόβλημα της εξάρτησης των δύο αυτών μεταβλητών, θα υπολογίσουμε τα προσαρμοσμένα τυποποιημένα κατάλοιπα. Επιπλέον, θα υπολογίσουμε τα κατάλληλα μέτρα συνάφειας για να διερευνήσουμε την ένταση της σχέσης των δύο μεταβλητών.

΄Υπαρξη κρίσεων το τελευταίο δίμηνο * ΄Υπαρξη κρίσεων τον τελευταίο χρόνο Crosstabulation

Adjusted Residual

		΄Υπαρξη κρίσεων τον τελευταίο χρόνο	
		όχι	ναι
΄Υπαρξη κρίσεων το τελευταίο δίμηνο	όχι	5,8	-5,8
	ναι	-5,8	5,8

Πίνακας 3.9: Προσαρμοσμένα τυποποιημένα κατάλοιπα των κελιών του πίνακα συνάφειας για την ύπαρξη κρίσεων το τελευταίο δίμηνο και την ύπαρξη κρίσεων τον περασμένο χρόνο

Από τον Πίνακα 3.9 παρατηρούμε ότι όλα τα κελιά του πίνακα συνάφειας συμβάλλουν στη σχέση εξάρτησης μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και της ύπαρξης κρίσεων τον περασμένο χρόνο, καθώς όλα τα προσαρμοσμένα τυποποιημένα κατάλοιπα έχουν τιμές μεγαλύτερες του 1.96 κατά απόλυτη τιμή. Επιπλέον, διαπιστώνουμε τα εξής:

- στα κελιά στα οποία αντιστοιχούν αρνητικά προσαρμοσμένα τυποποιημένα κατάλοιπα υπάρχουν στατιστικά σημαντικά λιγότερες παρατηρήσεις σε σύγκριση με αυτές που αναμένονται κάτω από την υπόθεση της ανεξαρτησίας των δύο μεταβλητών.
- στα κελιά στα οποία αντιστοιχούν θετικά προσαρμοσμένα τυποποιημένα κατάλοιπα υπάρχουν στατιστικά σημαντικά περισσότερες παρατηρήσεις σε σύγκριση με αυτές που αναμένονται κάτω από την υπόθεση της ανεξαρτησίας των δύο μεταβλητών.
- είναι πιο πιθανό για κάποιον που εμφάνισε κρίσεις το τελευταίο δίμηνο να είχε εμφανίσει κρίσεις τον περασμένο χρόνο σε σχέση με κάποιον που δεν εμφάνισε κρίσεις το τελευταίο δίμηνο (γιατί $d_{22} = 5.8 > d_{21}$).

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,602	,000
	Cramer's V	,602	,000
	Contingency Coefficient	,516	,000
N of Valid Cases		92	

Πίνακας 3.10: Μέτρα συνάφειας μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και της ύπαρξης κρίσεων τον περασμένο χρόνο

Από τον Πίνακα 3.10 παρατηρούμε ότι:

- ο συντελεστής Φ του Pearson ισούται με 0.602.
- ο συντελεστής C του Pearson ισούται με 0.516.
- ο συντελεστής V του Cramér ισούται με 0.602.

Από τις τιμές των παραπάνω μέτρων συνάφειας συμπεραίνουμε ότι υπάρχει ισχυρή συσχέτιση μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και της ύπαρξης κρίσεων τον περασμένο χρόνο.

<u>Συσχέτιση μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και της οικογενειακής κατάστασης</u>

΄Υπαρξη κρίσεων το τελευταίο δίμηνο * Οικογενειακή κατάσταση Crosstabulation

Count

		Οικογενειακή κατάσταση			
		παντρεμένος	ελεύθερος	χωρισμένος	Total
΄Υπαρξη κρίσεων το τελευταίο δίμηνο	όχι	13	38	3	54
	ναι	17	20	0	37
Total		30	58	3	91

Πίνακας 3.11: Πίνακας συνάφειας για την ύπαρξη κρίσεων το τελευταίο δίμηνο και την οικογενειακή κατάσταση

Από τον Πίνακα 3.11 παρατηρούμε ότι:

- για την συγκεκριμένη ανάλυση χρησιμοποιήθηκαν οι απαντήσεις 91 ασθενών.
- 54 άτομα δεν εμφάνισαν κρίσεις το τελευταίο δίμηνο, ενώ 37 άτομα εμφάνισαν.
- 30 άτομα είναι παντρεμένα, 58 άτομα είναι άγαμα και 3 άτομα είναι διαζευγμένα.
- από τα 54 άτομα που δεν εμφάνισαν κρίσεις το τελευταίο δίμηνο: 13 άτομα είναι παντρεμένα, 38 άτομα είναι άγαμα και 3 άτομα είναι διαζευγμένα.
- από τα 37 άτομα που εμφάνισαν κρίσεις το τελευταίο δίμηνο: 17 άτομα είναι παντρεμένα, 20 άτομα είναι άγαμα και κανένα άτομο δεν είναι διαζευγμένο.

Chi-Square Tests

				Monte Carlo Sig. (2-sided)			Monte Carlo Sig. (1-sided)		
					99% Confidence Interval			99% Confidence Interval	
	Value	df	Asymptotic Significance (2-sided)	Significance	Lower Bound	Upper Bound	Significance	Lower Bound	Upper Bound
Pearson Chi-Square	6,159[a]	2	,046	,038[b]	,033	,043			
Likelihood Ratio	7,179	2	,028	,033[b]	,029	,038			
Fisher's Exact Test	5,637			,047[b]	,041	,052			
Linear-by-Linear Association	5,949[c]	1	,015	,016[b]	,012	,019	,011[b]	,008	,013
N of Valid Cases	91								

a. 2 cells (33,3%) have expected count less than 5. The minimum expected count is 1,22.

b. Based on 10000 sampled tables with starting seed 2000000.

c. The standardized statistic is -2,439.

Πίνακας 3.12: Αποτελέσματα ελέγχων συσχέτισης μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και της οικογενειακής κατάστασης

Για να ελέγξουμε αν υπάρχει συσχέτιση μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και της οικογενειακής κατάστασης, θα χρησιμοποιήσουμε το p-value του ελέγχου ανεξαρτησίας X^2 του Pearson που έχει υπολογιστεί με προσομοίωση Monte Carlo, καθώς δεν ισχύουν οι προϋποθέσεις του ελέγχου X^2 του Pearson (το ποσοστό των κελιών με αναμενόμενες συχνότητες μικρότερες του 5 είναι 33.3%). Από τον Πίνακα 3.12 παρατηρούμε ότι, σε επίπεδο σημαντικότητας α=5%, υπάρχει ισχυρή ένδειξη εξάρτησης μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και της οικογενειακής κατάστασης ($p - value = 0.038 < \alpha = 0.05$). Προκειμένου να διαπιστώσουμε ποια κελιά «δημιουργούν» το πρόβλημα της εξάρτησης των δύο

αυτών μεταβλητών, θα υπολογίσουμε τα προσαρμοσμένα τυποποιημένα κατάλοιπα. Επιπλέον, θα υπολογίσουμε τα κατάλληλα μέτρα συνάφειας για να διερευνήσουμε την ένταση της σχέσης των δύο μεταβλητών.

Ύπαρξη κρίσεων το τελευταίο δίμηνο * Οικογενειακή κατάσταση Crosstabulation

Adjusted Residual

		Οικογενειακή κατάσταση		
		παντρεμένος	ελεύθερος	χωρισμένος
Ύπαρξη κρίσεων το τελευταίο δίμηνο	όχι	-2,2	1,6	1,5
	ναι	2,2	-1,6	-1,5

Πίνακας 3.13: Προσαρμοσμένα τυποποιημένα κατάλοιπα των κελιών του πίνακα συνάφειας για την ύπαρξη κρίσεων το τελευταίο δίμηνο και την οικογενειακή κατάσταση

Από τον Πίνακα 3.13 παρατηρούμε ότι τα κελιά του πίνακα συνάφειας που συμβάλλουν στη σχέση εξάρτησης μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και της οικογενειακής κατάστασης (δηλαδή αυτά στα οποία αντιστοιχούν προσαρμοσμένα τυποποιημένα κατάλοιπα με τιμές μεγαλύτερες του 1.96 κατά απόλυτη τιμή) είναι τα εξής:

- το κελί των παντρεμένων ασθενών που δεν εμφάνισαν κρίσεις το τελευταίο δίμηνο.
- το κελί των παντρεμένων ασθενών που εμφάνισαν κρίσεις το τελευταίο δίμηνο.

Επιπλέον, διαπιστώνουμε τα εξής:

- στο κελί των παντρεμένων ασθενών που δεν εμφάνισαν κρίσεις το τελευταίο δίμηνο υπάρχουν στατιστικά σημαντικά λιγότερες παρατηρήσεις σε σύγκριση με αυτές που αναμένονται κάτω από την υπόθεση της ανεξαρτησίας των δύο μεταβλητών, καθώς στο κελί αυτό αντιστοιχεί αρνητικό προσαρμοσμένο τυποποιημένο κατάλοιπο.
- στο κελί των παντρεμένων ασθενών που εμφάνισαν κρίσεις το τελευταίο δίμηνο υπάρχουν στατιστικά σημαντικά περισσότερες παρατηρήσεις σε σύγκριση με αυτές που αναμένονται κάτω από την υπόθεση της ανεξαρτησίας των δύο μεταβλητών, καθώς στο κελί αυτό αντιστοιχεί θετικό προσαρμοσμένο τυποποιημένο κατάλοιπο.

- είναι πιο πιθανό για κάποιον που εμφάνισε κρίσεις το τελευταίο δίμηνο να είναι παντρεμένος σε σχέση με κάποιον που δεν εμφάνισε κρίσεις το τελευταίο δίμηνο (γιατί $d_{21} = 2.2 > d_{2j}, j = 2,3$).

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,260	,046
	Cramer's V	,260	,046
	Contingency Coefficient	,252	,046
N of Valid Cases		91	

Πίνακας 3.14: Μέτρα συνάφειας μεταξύ της ύπαρξης ύπαρξης κρίσεων το τελευταίο δίμηνο και της οικογενειακής κατάστασης

Από τον Πίνακα 3.14 παρατηρούμε ότι:

- ο συντελεστής Φ του Pearson ισούται με 0.26.
- ο συντελεστής C του Pearson ισούται με 0.252.
- ο συντελεστής V του Cramér ισούται με 0.26.

Από τις τιμές των παραπάνω μέτρων συνάφειας συμπεραίνουμε ότι υπάρχει μέτρια συσχέτιση μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και της οικογενειακής κατάστασης.

ΚΕΦΑΛΑΙΟ 4
Λογιστική Παλινδρόμηση

Στο παρόν κεφάλαιο παρουσιάζουμε τα μοντέλα λογιστικής παλινδρόμησης που προκύπτουν όταν χρησιμοποιούμε ως εξαρτημένη μεταβλητή αυτή που αφορά την ύπαρξη κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου. Πιο συγκεκριμένα, παρουσιάζουμε δύο μοντέλα: ένα που προκύπτει όταν αφαιρούμε τις ελλιπείς τιμές από τα δεδομένα μας και ένα δεύτερο που προκύπτει όταν εφαρμόζουμε την διαδικασία συμπλήρωσης των ελλιπών τιμών (imputation).

4.1. Βασικά θεωρητικά στοιχεία για την λογιστική παλινδρόμηση

Στην ενότητα αυτή δίνουμε το βασικό θεωρητικό πλαίσιο της λογιστικής παλινδρόμησης.

4.1.1. Εισαγωγή

Η λογιστική παλινδρόμηση είναι ένα μοντέλο παλινδρόμησης στο οποίο η εξαρτημένη μεταβλητή *Y* είναι κατηγορική, ενώ οι ανεξάρτητες μεταβλητές μπορεί να είναι είτε ποσοτικές συνεχείς, είτε κατηγορικές. Ο στόχος, λοιπόν, της λογιστικής παλινδρόμησης είναι η δημιουργία ενός μοντέλου πρόβλεψης της μέσης τιμής της υπό μελέτης κατηγορικής εξαρτημένης μεταβλητής χρησιμοποιώντας κάποιες ποσοτικές ή/και κατηγορικές ανεξάρτητες μεταβλητές. Όπως ισχύει και στην γραμμική παλινδρόμηση, όταν έχουμε μία ανεξάρτητη μεταβλητή τότε μιλάμε για απλή λογιστική παλινδρόμηση, ενώ όταν έχουμε περισσότερες ανεξάρτητες μεταβλητές μιλάμε για πολλαπλή λογιστική παλινδρόμηση. Ανάλογα με την ιδιαίτερη φύση της εξαρτημένης κατηγορικής μεταβλητής διακρίνονται τρεις τύποι λογιστικής παλινδρόμησης:

- η δίτιμη λογιστική παλινδρόμηση (binary logistic regression): Σε αυτήν την περίπτωση η εξαρτημένη μεταβλητή συνίσταται από δύο κατηγορίες, όπως για παράδειγμα είναι οι εκβάσεις επιτυχία/αποτυχία, ΝΑΙ/ΟΧΙ, γεγονός απόν/παρόν. Η μέση τιμή μιας δίτιμης τυχαίας μεταβλητής είναι η πιθανότητα εμφάνισης του χαρακτηριστικού που αντιστοιχεί στην μεταβλητή. Άρα η δίτιμη λογιστική παλινδρόμηση χρησιμοποιείται για να περιγράψει τη σχέση της πιθανότητας ενός χαρακτηριστικού (π.χ. βελτίωση της κατάστασης ενός ασθενούς) με διάφορους παράγοντες (θεραπεία, φύλο, ηλικία, κτλ.)

- η πολυωνυμική λογιστική παλινδρόμηση (multinomial logistic regression): Σε αυτήν την περίπτωση η εξαρτημένη μεταβλητή έχει τρεις ή περισσότερες κατηγορίες, οι οποίες δεν έχουν κάποια φυσική διαβάθμιση, όπως για παράδειγμα ο χαρακτηρισμός του χρώματος αντικειμένων ως ερυθρού, πράσινου ή κίτρινου.
- η διατάξιμη λογιστική παλινδρόμηση (ordinal logistic regression): Σε αυτήν την περίπτωση η εξαρτημένη μεταβλητή συνίσταται από δύο ή περισσότερες κατηγορίες μεταξύ των οποίων ισχύει η έννοια της ανισότητας, όπως για παράδειγμα σε μια ερώτηση συμφωνίας/διαφωνίας με κλίμακα καθόλου, λίγο, μέτρια, αρκετά, πολύ.

Στην παρούσα εργασία θα ασχοληθούμε μόνο με την δίτιμη λογιστική παλινδρόμηση.

4.1.2. Απλή δίτιμη λογιστική παλινδρόμηση

Έστω ότι έχουμε μια δίτιμη εξαρτημένη μεταβλητή Y που παίρνει τις τιμές 0 και 1 και μια ανεξάρτητη μεταβλητή X.

Η μέση τιμή της μεταβλητής Y ισούται με

$$p = P(Y = 1).$$

Είναι δηλαδή η πιθανότητα η τυχαία μεταβλητή Y να πάρει την τιμή 1, η οποία αντιπροσωπεύει την «επιτυχία».

Το μοντέλο της απλής δίτιμης λογιστικής παλινδρόμησης είναι το εξής:

$$\log\left(\frac{p_i}{1-p_i}\right) = b_0 + b_1 X_i \quad (1)$$

Το $\frac{p_i}{1-p_i}$, που ονομάζεται σχετική πιθανότητα «επιτυχίας» (odds), είναι το πηλίκο της πιθανότητας «επιτυχίας» προς την πιθανότητα «αποτυχίας» και εκφράζει το πόσο πιο πιθανό είναι να συμβεί η «επιτυχία» σε σχέση με το να μην συμβεί. Η σχέση (1) μπορεί να γραφεί ισοδύναμα και ως εξής:

$$p_i = \frac{e^{b_0 + b_1 X_i}}{1 + e^{b_0 + b_1 X_i}}.$$

Η ερμηνεία των παραμέτρων του μοντέλου της απλής δίτιμης λογιστικής παλινδρόμησης είναι η εξής:

- το b_0 είναι η τιμή που παίρνει ο λογάριθμος της σχετικής πιθανότητας «επιτυχίας» όταν η ανεξάρτητη μεταβλητή πάρει την τιμή 0.
- το b_1 αντιπροσωπεύει το ποσό της μεταβολής που θα επέλθει στον λογάριθμο της σχετικής πιθανότητας «επιτυχίας» εάν η ανεξάρτητη μεταβλητή αυξηθεί κατά μία μονάδα.

Εναλλακτικά, η ερμηνεία των παραμέτρων του μοντέλου της απλής δίτιμης λογιστικής παλινδρόμησης μπορεί να γίνει μέσω του λόγου σχετικών πιθανοτήτων

(odds ratio, OR). Έστω ότι η ανεξάρτητη μεταβλητή X είναι δίτιμη και παίρνει τις τιμές 0 και 1. Τότε ο λόγος σχετικών πιθανοτήτων είναι ο λόγος της σχετικής πιθανότητας «επιτυχίας» όταν η ανεξάρτητη μεταβλητή πάρει την τιμή 1 προς την σχετική πιθανότητα «επιτυχίας» όταν η ανεξάρτητη μεταβλητή πάρει την τιμή 0 και ισούται με:

$$OR = \frac{\frac{e^{b_0+b_1}}{1+e^{b_0+b_1}} / \frac{1}{1+e^{b_0+b_1}}}{\frac{e^{b_0}}{1+e^{b_0}} / \frac{1}{1+e^{b_0}}} = e^{b_1}.$$

Εάν $OR = 1$, τότε η ανεξάρτητη μεταβλητή δεν επηρεάζει την εξαρτημένη μεταβλητή.

Για την εκτίμηση των παραμέτρων του μοντέλου χρησιμοποιείται η μέθοδος μεγίστης πιθανοφάνειας (maximum likelihood method). Η πιθανοφάνεια (likelihood) είναι μια συνάρτηση των παραμέτρων του μοντέλου, η οποία εκφράζει την πιθανότητα να παρατηρηθούν τα πραγματικά δεδομένα Y γνωρίζοντας τις τιμές της ανεξάρτητης μεταβλητής X. Η πιθανοφάνεια στην περίπτωση της απλής δίτιμης λογιστικής παλινδρόμησης είναι η συνάρτηση

$$l(b_0, b_1) = \prod_{i=1}^{n} {p_i}^{Y_i}(1 - {p_i}^{Y_i}).$$

Οι εκτιμήσεις των παραμέτρων b_0 και b_1 είναι αυτές που μεγιστοποιούν τη συνάρτηση πιθανοφάνειας. Για να είναι πιο εύκολες οι πράξεις, συνήθως λογαριθμοποιούμε την πιθανοφάνεια πριν τη μεγιστοποιήσουμε. Με αυτόν τον τρόπο το γινόμενο στην έκφραση της πιθανοφάνειας μετατρέπεται σε άθροισμα. Έχουμε δηλαδή

$$\ln l(b_0, b_1) = \ln\left\{\prod_{i=1}^{n} {p_i}^{Y_i}(1 - {p_i}^{Y_i})\right\} = \sum_{i=1}^{n} Y_i \ln p_i + \left(n - \sum_{i=1}^{n} Y_i\right)\ln(1 - p_i).$$

Αντικαθιστώντας το p_i με το $\frac{e^{b_0+b_1X_i}}{1+e^{b_0+b_1X_i}}$ έχουμε

$$\ln l(b_0, b_1) = \sum_{i=1}^{n} Y_i \ln \frac{e^{b_0+b_1X_i}}{1+e^{b_0+b_1X_i}} + \left(n - \sum_{i=1}^{n} Y_i\right)\ln\left(1 - \frac{e^{b_0+b_1X_i}}{1+e^{b_0+b_1X_i}}\right).$$

Για να εκτιμήσουμε λοιπόν τις τιμές των παραμέτρων που μεγιστοποιούν την πιθανοφάνεια $l(b_0, b_1)$, παραγωγίζουμε την παραπάνω συνάρτηση ως προς b_0 και b_1 και θέτουμε κάθε μια από τις δύο σχέσεις που προκύπτουν ίση με το 0. Στη συνέχεια, λύνουμε το σύστημα εξισώσεων και έτσι εκτιμούμε τις παραμέτρους b_0 και b_1. Οι παράμετροι αυτοί ονομάζονται εκτιμητές μεγίστης πιθανοφάνειας και ακολουθούν ασυμπτωτικά την κανονική κατανομή.

4.1.3. Πολλαπλή δίτιμη λογιστική παλινδρόμηση

Έστω τώρα ότι έχουμε δύο ή περισσότερες ανεξάρτητες μεταβλητές $X_1, \dots, X_k$ και θέλουμε να ελέγξουμε εάν αυτές επηρεάζουν μια δίτιμη εξαρτημένη μεταβλητή Y. Σε αυτήν την περίπτωση έχουμε πολλαπλή δίτιμη λογιστική παλινδρόμηση και το μοντέλο είναι το εξής:

$$\log\left(\frac{p_i}{1-p_i}\right) = b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots + b_k X_{ki} \quad (2)$$

Η σχέση (2) μπορεί να γραφεί ισοδύναμα και ως εξής:

$$p_i = \frac{e^{b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots + b_k X_{ki}}}{1 + e^{b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots + b_k X_{ki}}}.$$

Η ερμηνεία των παραμέτρων του μοντέλου της πολλαπλής δίτιμης λογιστικής παλινδρόμησης είναι η εξής:

- το b_0 είναι η τιμή που παίρνει ο λογάριθμος της σχετικής πιθανότητας «επιτυχίας» όταν όλες οι ανεξάρτητες μεταβλητές πάρουν την τιμή 0.
- το b_1 αντιπροσωπεύει το ποσό της μεταβολής που θα επέλθει στον λογάριθμο της σχετικής πιθανότητας «επιτυχίας» εάν η ανεξάρτητη μεταβλητή X_1 αυξηθεί κατά μία μονάδα και οι υπόλοιπες ανεξάρτητες μεταβλητές παραμείνουν σταθερές.
- το b_2 αντιπροσωπεύει το ποσό της μεταβολής που θα επέλθει στον λογάριθμο της σχετικής πιθανότητας «επιτυχίας» εάν η ανεξάρτητη μεταβλητή X_2 αυξηθεί κατά μία μονάδα και οι υπόλοιπες ανεξάρτητες μεταβλητές παραμείνουν σταθερές.
- οι υπόλοιποι συντελεστές ερμηνεύονται με παρόμοιο τρόπο.

Εναλλακτικά, η ερμηνεία των παραμέτρων του μοντέλου της πολλαπλής δίτιμης λογιστικής παλινδρόμησης μπορεί να γίνει, όπως και στην περίπτωση της απλής δίτιμης λογιστικής παλινδρόμησης, μέσω του λόγου σχετικών πιθανοτήτων (odds ratio, OR).

Η εκτίμηση των παραμέτρων του μοντέλου πολλαπλής δίτιμης λογιστικής παλινδρόμησης γίνεται πάλι με τη μέθοδο μεγίστης πιθανοφάνειας.

4.1.4. Στατιστική συμπερασματολογία για τους συντελεστές του μοντέλου δίτιμης λογιστικής παλινδρόμησης

Προκειμένου να αξιολογήσουμε τη στατιστική σημαντικότητα των παραμέτρων του μοντέλου δίτιμης λογιστικής παλινδρόμησης, χρησιμοποιούμε το κριτήριο του Wald. Οι υποθέσεις που ελέγχονται μέσω του κριτηρίου του Wald είναι οι εξής:

$$H_0: b_i = 0$$
$$H_1: b_i \neq 0$$

για $i = 1, \dots, k$. Η μηδενική υπόθεση H_0 αναφέρει ότι η i ανεξάρτητη μεταβλητή δεν ερμηνεύει τον λογάριθμο της σχετικής πιθανότητας «επιτυχίας», ενώ αντίθετα η εναλλακτική υπόθεση H_1 αναφέρει ότι η i ανεξάρτητη μεταβλητή ερμηνεύει τον λογάριθμο της σχετικής πιθανότητας «επιτυχίας».

Η στατιστική συνάρτηση του ελέγχου είναι η εξής:

$$W = \frac{\widehat{b_i}}{s(\widehat{b_i})} ,$$

η οποία, όταν ισχύει η μηδενική υπόθεση, ακολουθεί την τυπική κανονική κατανομή $N(0,1)$. Η κρίσιμη περιοχή του ελέγχου είναι η $|W| \geq z_{\alpha/2}$. Απορρίπτουμε δηλαδή την μηδενική υπόθεση ότι η i ανεξάρτητη μεταβλητή δεν είναι στατιστικά σημαντική σε επίπεδο σημαντικότητας α, εάν η στατιστική συνάρτηση W έχει τιμή που είναι μικρότερη από την τιμή $-z_{\alpha/2}$ ή μεγαλύτερη από την τιμή $z_{\alpha/2}$.

4.1.5. Έλεγχοι υποθέσεων στη δίτιμη λογιστική παλινδρόμηση

Στη λογιστική παλινδρόμηση, η βασική έννοια η οποία χρησιμοποιείται για να ελέγξουμε την καλή προσαρμογή ενός μοντέλου σε ένα σύνολο δεδομένων είναι η απόκλιση του μοντέλου (deviance). Η απόκλιση αποτελεί ένα μέτρο της «ανερμήνευτης μεταβλητότητας» της μεταβλητής απόκρισης Y, μετά την προσαρμογή του μοντέλου.

Για την δίτιμη λογιστική παλινδρόμηση υπάρχουν συνήθως δύο ειδών έλεγχοι που μας ενδιαφέρουν:

1. <u>Έλεγχος για την ολική επάρκεια του μοντέλου</u>

   Για να ελέγξουμε την ολική επάρκεια ενός μοντέλου δίτιμης λογιστικής παλινδρόμησης, χρησιμοποιούμε τον έλεγχο των Hosmer-Lemeshow (HL). Οι υποθέσεις που ελέγχονται μέσω αυτού του ελέγχου είναι οι εξής:

   H_0: *οι παρατηρηθείσες τιμές της Y δε διαφέρουν από τις εκτιμώμενες τιμές*

   H_1: *οι παρατηρηθείσες τιμές της Y διαφέρουν από τις εκτιμώμενες τιμές*

   Για να υλοποιήσουμε τον έλεγχο των Hosmer-Lemeshow ακολουθούμε τα παρακάτω βήματα:

1. Διατάσσουμε τις παρατηρήσεις ανάλογα με την προβλεπόμενη πιθανότητα επιτυχίας.
2. Χωρίζουμε τις διατεταγμένες παρατηρήσεις σε g ομάδες, με ίσο περίπου αριθμό παρατηρήσεων, και για καθεμία από αυτές καταγράφουμε τον αριθμό επιτυχιών και αποτυχιών, σχηματίζοντας έτσι έναν πίνακα $g \times 2$ (συνήθως $g = 10$).
3. Η στατιστική συνάρτηση των HL, X_{HL} είναι το X^2 του Pearson για τον παραπάνω πίνακα.

Η συνάρτηση X_{HL} ακολουθεί την κατανομή x^2_{g-2} όταν ισχύει η μηδενική υπόθεση. Απόρριψη της μηδενικής υπόθεσης δηλώνει ότι το μοντέλο μας είναι ανεπαρκές για το συγκεκριμένο επίπεδο σημαντικότητας του ελέγχου.

2. <u>Σύγκριση μεταξύ δύο (εμφωλευμένων) μοντέλων</u>

Ο έλεγχος αυτός, ο οποίος προκύπτει από τον έλεγχο του λόγου πιθανοφανειών (Likelihood Ratio Test, LRT), γίνεται για να εξετάσουμε αν δύο εμφωλευμένα μοντέλα διαφέρουν σημαντικά μεταξύ τους. Πιο συγκεκριμένα, έστω ότι έχουμε δύο εμφωλευμένα μοντέλα M_1 και M_2 (με τον όρο εμφωλευμένα εννοούμε εδώ ότι το σύνολο των επεξηγηματικών μεταβλητών του M_1 είναι υποσύνολο αυτών του M_2). Υποθέτουμε ότι οι συναρτήσεις πιθανοφάνειας των δύο μοντέλων είναι $L(M_1)$ και $L(M_2)$ αντίστοιχα, και ορίζουμε

$$l(M_1) = logL(M_1), \quad l(M_2) = logL(M_2).$$

Για να διαπιστώσουμε αν το μοντέλο M_2 είναι καλύτερο από το μοντέλο M_1, κάνουμε τον έλεγχο υπόθεσης:

$$H_0: \textit{τα δύο μοντέλα δεν διαφέρουν σημαντικά}$$
$$H_1: \textit{το μοντέλο } M_2 \text{ είναι καλύτερο από το μοντέλο } M_1$$

Η στατιστική συνάρτηση του ελέγχου δίνεται από την σχέση

$$G^2 = -2log\frac{L(M_1)}{L(M_2)} = -2(l(M_1) - l(M_2))$$

Η παρουσία του συντελεστή -2 εδώ οφείλεται στο ότι η παραπάνω ποσότητα, αν ισχύει η H_0, ακολουθεί (προσεγγιστικά) την κατανομή x^2_p, όπου:

- $p = df_1 - df_2$
- df_1 είναι ο αριθμός των βαθμών ελευθερίας μετά την προσαρμογή του μοντέλου M_1, και
- df_2 είναι ο αριθμός των βαθμών ελευθερίας μετά την προσαρμογή του μοντέλου M_2

Η μηδενική υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας α, εάν $G^2 > x^2_{p,\alpha}$.

4.1.6. Αξιολόγηση της προσαρμογής ενός μοντέλου δίτιμης λογιστικής παλινδρόμησης

Ένα άλλο δημοφιλές μέτρο αξιολόγησης ενός μοντέλου δίτιμης λογιστικής παλινδρόμησης, εκτός από τον έλεγχο των Hosmer-Lemeshow, είναι οι πίνακες ταξινόμησης (classification tables) σε συνδυασμό με τις καμπύλες ROC (Receiver Operating Characteristic curves).

Για την κατασκευή ενός πίνακα ταξινόμησης, ακολουθούμε τα παρακάτω βήματα:

1. Υπολογίζουμε όλες τις εκτιμώμενες τιμές $\widehat{p_i}$, για $i = 1,2, \dots, n$.
2. Επιλέγουμε μία πιθανότητα p_0 ως κατώφλι (ή σημείο αποκοπής, cutoff point). Για $i = 1,2, \dots, n$, αν
   - $\widehat{p_i} \geq p_0$, τότε θεωρούμε ότι για την παρατήρηση i, το υπόδειγμα προβλέπει «επιτυχία» ($\widehat{Y_i} = 1$).
   - $\widehat{p_i} < p_0$, τότε θεωρούμε ότι για την παρατήρηση i, το υπόδειγμα προβλέπει «αποτυχία» ($\widehat{Y_i} = 0$).
3. Ο πίνακας ταξινόμησης είναι ένας πίνακας δύο διαστάσεων που μας δίνει τις συχνότητες για τις επιτυχίες και τις αποτυχίες ανάμεσα στις παρατηρηθείσες και τις εκτιμώμενες τιμές. Στον Πίνακα 4.1 δίνεται η μορφή του πίνακα ταξινόμησης.

	Παρατηρούμενο αποτέλεσμα		
		Επιτυχία	*Αποτυχία*
Εκτιμώμενο αποτέλεσμα	*Επιτυχία* (πάνω από την τιμή p_0)	a	b
	Αποτυχία (κάτω από την τιμή p_0)	c	d

Πίνακας 4.1: Πίνακας ταξινόμησης

Προφανώς ισχύει $a + b + c + d = n$ (το συνολικό πλήθος των παρατηρήσεων). Διαισθητικά, όσο μεγαλύτερο είναι το άθροισμα $a + d$ (αριθμός παρατηρήσεων που το μοντέλο ταξινομεί σωστά, είτε ως επιτυχίες είτε ως αποτυχίες) σε σχέση με το άθροισμα $b + c$ (πλήθος εσφαλμένων ταξινομήσεων), τόσο μεγαλύτερη είναι η προβλεπτική αξία του υποδείγματος. Ειδικότερα, μπορούμε να αξιολογήσουμε την ευαισθησία (sensitivity), την ειδικότητα (specificity) και την ακρίβεια (accuracy) του μοντέλου:

- $ευαισθησία = \frac{a}{a+c}$ (ποσοστό επιτυχιών που «ταξινομούνται» σωστά)
- $ειδικότητα = \frac{d}{b+d}$ (ποσοστό αποτυχιών που «ταξινομούνται» σωστά)
- $ακρίβεια = \frac{a+d}{a+b+c+d}$ (είναι ένα μέτρο που συνδυάζει την ευαισθησία και την ειδικότητα)

Ιδανικά, θα θέλαμε ένα μοντέλο με υψηλή ευαισθησία και υψηλή ειδικότητα. Στην πράξη, όμως, αυτό δεν είναι εφικτό, καθώς όταν αυξάνεται το ένα μέτρο τότε το άλλο μειώνεται.

Σε μία καμπύλη ROC, για διάφορες τιμές του κατωφλίου p_0, θέτουμε στον άξονα των x την ποσότητα (1-ειδικότητα) και στον άξονα των y την ευαισθησία. Το εμβαδόν κάτω από την καμπύλη (Area Under Curve, AUC) χρησιμοποιείται ως ένα ακόμη μέτρο για την αξιολόγηση της απόδοσης του μοντέλου. Όσο πιο κοντά είναι το AUC στην μονάδα, τόσο καλύτερα ταξινομούνται οι παρατηρήσεις από το υπόδειγμα.

4.2. Εφαρμογή της δίτιμης λογιστικής παλινδρόμησης στα δεδομένα μας όταν αφαιρούμε τις ελλιπείς τιμές

Στο Κεφάλαιο 3 είδαμε ότι οι μεταβλητές με τις οποίες συσχετίζεται η ύπαρξη κρίσεων το τελευταίο δίμηνο είναι οι εξής:

- συνολικός αριθμός κρίσεων
- αριθμός επισκέψεων στον γιατρό τον τελευταίο χρόνο
- ύπαρξη κρίσεων τον περασμένο χρόνο
- οικογενειακή κατάσταση

Επομένως, θα προσαρμόσουμε, αρχικά, ένα μοντέλο δίτιμης λογιστικής παλινδρόμησης με τις παραπάνω τέσσερις μεταβλητές ως ανεξάρτητες. Προκειμένου τα δεδομένα μας να μην έχουν ελλιπείς τιμές, θα πρέπει να αφαιρέσουμε τις εξής 11 παρατηρήσεις: 11, 15, 18, 33, 34, 61, 62, 71, 74, 77, 82. Συνεπώς, θα προσαρμόσουμε το μοντέλο χρησιμοποιώντας 85 παρατηρήσεις.

Συμβολισμοί

Έστω ότι p είναι η πιθανότητα εμφάνισης κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου ενός ατόμου που πάσχει από επιληψία. Επίσης, θεωρούμε τις παρακάτω δίτιμες μεταβλητές:

- $total\_kriseis2 = \begin{cases} 1, & το\ άτομο\ έχει\ κάνει\ 6-10\ κρίσεις\ συνολικά\ στη\ ζωή\ του \\ 0, & διαφορετικά \end{cases}$
- $total\_kriseis3 =$
  $\begin{cases} 1, & το\ άτομο\ έχει\ κάνει\ πάνω\ από\ 21\ \ κρίσεις\ συνολικά\ στη\ ζωή\ του \\ 0, & διαφορετικά \end{cases}$
- $total\_kriseis4 =$
  $\begin{cases} 1, & το\ άτομο\ έχει\ κάνει\ πάνω\ από\ 100\ \ κρίσεις\ συνολικά\ στη\ ζωή\ του \\ 0, & διαφορετικά \end{cases}$
- $doctor's\_visit2 =$
  $\begin{cases} 1, & το\ άτομο\ επισκέφθηκε\ 2\ φορές\ τον\ γιατρό\ τον\ τελευταίο\ χρόνο \\ 0, & διαφορετικά \end{cases}$
- $doctor's\_visit3 =$
  $\begin{cases} 1, & το\ άτομο\ επισκέφθηκε\ 3-9\ φορές\ τον\ γιατρό\ τον\ τελευταίο\ χρόνο \\ 0, & διαφορετικά \end{cases}$
- $kriseis\_lastyear1 = \begin{cases} 1, & το\ άτομο\ εμφάνισε\ κρίσεις\ τον\ περασμένο\ χρόνο \\ 0, & διαφορετικά \end{cases}$
- $marital\_status2 =$
  $\begin{cases} 1, & το\ άτομο\ είναι\ άγαμο\ ή\ χωρισμένο \\ 0, & διαφορετικά \end{cases}$

Μοντέλο M_0

Στους Πίνακες 4.2 & 4.3 βλέπουμε τα αποτελέσματα που προκύπτουν όταν προσαρμόζουμε ένα μοντέλο δίτιμης λογιστικής παλινδρόμησης με ανεξάρτητες μεταβλητές αυτές με τις οποίες συσχετίζεται η ύπαρξη κρίσεων το τελευταίο δίμηνο[7].

Παράμετρος	Εκτίμηση	Τυπικό σφάλμα
Σταθερά	-3.3606	1.5153
total_kriseis2	0.1304	1.4712
total_kriseis3	1.5528	1.3914
total_kriseis4	2.1227	1.2853
doctor's_visit2	0.6178	0.8141
doctor's visit3	1.0623	0.8164
kriseis_lastyear1	3.6139	0.8833
marital_status2	-1.9967	0.7912

Πίνακας 4.2: Εκτιμήσεις των παραμέτρων του μοντέλου M_0

[7] Πριν την προσαρμογή του μοντέλου, συγχωνεύουμε τις εξής κατηγορίες του συνολικού αριθμού κρίσεων: «1» με «2-5». Επίσης, συγχωνεύουμε τις εξής κατηγορίες της οικογενειακής κατάστασης: «ελεύθερος-η» με «χωρισμένος-η». Αν δεν προχωρήσουμε σε αυτές τις συγχωνεύσεις, θα προκύψει ένα μοντέλο του οποίου οι εκτιμητές έχουν τεράστια τυπικά σφάλματα, βλέπε ΠΑΡΑΡΤΗΜΑ Π3.

Μοντέλο	Απόκλι-ση	Βαθμοί ελευθε-ρίας (β.ε.)	Μεταβολή στην απόκλιση	Μεταβο-λή στους β.ε.	p-value ελέγχου σημαντικό-τητας	Στατιστικά σημαντικό σε 5%
1 (Σταθερά)	115.839	84				
+total_kriseis	101.194	81	14.6450	3	0.002147	ΝΑΙ
+doctor's_visit	97.319	79	3.8746	2	0.144096	ΟΧΙ
+kriseis_lastyear	70.606	78	26.7132	1	≈ 0	ΝΑΙ
+marital_status	62.577	77	8.0289	1	0.004604	ΝΑΙ

Πίνακας 4.3: Πίνακας ανάλυσης απόκλισης του μοντέλου M_0

Από τον Πίνακα 4.3 παρατηρούμε ότι η μεταβλητή που αφορά τον αριθμό των επισκέψεων στον γιατρό τον τελευταίο χρόνο δεν είναι στατιστικά σημαντική σε επίπεδο σημαντικότητας α=5%, ενώ οι υπόλοιπες μεταβλητές είναι στατιστικά σημαντικές. Επιπλέον, εφαρμόζοντας τον έλεγχο των Hosmer-Lemeshow διαπιστώνουμε ότι το μοντέλο M_0 είναι συνολικά επαρκές σε επίπεδο σημαντικότητας α=5% ($p-value\ ελέγχου\ Hosmer-Lemeshow = 0.881 > \alpha = 0.05$). Στη συνέχεια, θα προσαρμόσουμε ένα μοντέλο (M_1) χωρίς την ανεξάρτητη μεταβλητή που αφορά τον αριθμό των επισκέψεων στον γιατρό τον τελευταίο χρόνο.

Μοντέλο M_1

Στους Πίνακες 4.4 & 4.5 βλέπουμε τα αποτελέσματα του μοντέλου που προκύπτουν όταν αφαιρούμε από το μοντέλο M_0 την ανεξάρτητη μεταβλητή που αφορά τον αριθμό των επισκέψεων στον γιατρό τον τελευταίο χρόνο.

Παράμετρος	Εκτίμηση	Τυπικό σφάλμα
Σταθερά	-2.5472	1.2377
total_kriseis2	-0.1351	1.4086
total_kriseis3	1.0572	1.2910
total_kriseis4	1.8620	1.2243
kriseis_lastyear1	3.5757	0.8303
marital_status2	-1.8810	0.7831

Πίνακας 4.4: Εκτιμήσεις των παραμέτρων του μοντέλου M_1

Μοντέλο	Απόκλι-ση	Βαθμοί ελευθε-ρίας (β.ε.)	Μεταβολή στην απόκλιση	Μεταβο-λή στους β.ε.	p-value ελέγχου σημαντικό-τητας	Στατιστικά σημαντικό σε 5%
1 (Σταθερά)	115.839	84				
+total_kriseis	101.194	81	14.6450	3	0.002147	NAI
+kriseis_lastyear	71.555	80	29.6385	1	≈ 0	NAI
+marital_status	64.331	79	7.2243	1	0.007192	NAI

Πίνακας 4.5: Πίνακας ανάλυσης απόκλισης του μοντέλου M_1

Από τον Πίνακα 4.5 παρατηρούμε ότι όλες οι μεταβλητές του μοντέλου M_1 είναι στατιστικά σημαντικές σε επίπεδο σημαντικότητας α=5%. Επιπλέον, εφαρμόζοντας τον έλεγχο των Hosmer-Lemeshow διαπιστώνουμε ότι το μοντέλο M_1 είναι συνολικά επαρκές σε επίπεδο σημαντικότητας α=5% ($p-value\ ελέγχου\ Hosmer-Lemeshow = 0.9495 > \alpha = 0.05$).

Σύγκριση των μοντέλων M_0 και M_1

Τα μοντέλα M_0 και M_1 είναι εμφωλευμένα μεταξύ τους, καθώς το σύνολο των επεξηγηματικών μεταβλητών του M_1 είναι υποσύνολο αυτών του M_0. Επομένως, για να μπορέσουμε να τα συγκρίνουμε θα χρησιμοποιήσουμε τον έλεγχο του λόγου πιθανοφανειών. Δηλαδή θα κάνουμε τον εξής έλεγχο:

H_0: *τα δύο μοντέλα δεν διαφέρουν σημαντικά*

H_1: *το μοντέλο* M_0 είναι καλύτερο από το μοντέλο M_1

Το μοντέλο M_0 έχει $df_0 = 77$ βαθμούς ελευθερίας και ο λογάριθμος της συνάρτησης πιθανοφάνειάς του ισούται με $l(M_0) = logL(M_0) = -31.28867$, ενώ το μοντέλο M_1 έχει $df_1 = 79$ βαθμούς ελευθερίας και ο λογάριθμος της συνάρτησης πιθανοφάνειάς του ισούται με $l(M_1) = logL(M_1) = -32.16558$. Συνεπώς, η στατιστική συνάρτηση του ελέγχου ισούται με $G^2 = -2\big(l(M_1) - l(M_0)\big) = -2(-32.16558 + 31.28867) = -2(-0.87691) = 1.75382$. Άρα σε επίπεδο σημαντικότητας α=5% δεν απορρίπτεται η μηδενική υπόθεση, καθώς $G^2 < x^2_{2,0.05} = 5.9915$. Οπότε τα δύο μοντέλα δεν διαφέρουν σημαντικά και για τον λόγο αυτό θα συνεχίσουμε την ανάλυσή μας με το απλούστερο μοντέλο, δηλαδή με το μοντέλο M_1.

Ερμηνεία των εκτιμητών των παραμέτρων του μοντέλου M_1

Το μοντέλο M_1 που προσαρμόστηκε έχει την εξής μορφή:

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = -2.5472 - 0.1351 * total\_kriseis2 + 1.0572 * total\_kriseis3 + 1.8620 * total\_kriseis4 + 3.5757 * kriseis\_lastyear1 - 1.8810 * marital\_status2$$

Η ερμηνεία των εκτιμητών των παραμέτρων του μοντέλου M_1 είναι η εξής:

- ο λογάριθμος της εκτιμώμενης σχετικής πιθανότητας «επιτυχίας» (δηλαδή εμφάνισης κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που έχει κάνει 1-5 κρίσεις συνολικά στη ζωή του, που επισκέφθηκε 0-1 φορές τον γιατρό τον τελευταίο χρόνο και που είναι παντρεμένο ισούται με -2.5472.
- ο λογάριθμος της εκτιμώμενης σχετικής πιθανότητας «επιτυχίας» (δηλαδή εμφάνισης κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που έχει κάνει 6-10 κρίσεις συνολικά στη ζωή του είναι μικρότερος κατά 0.1351 μονάδες σε σχέση με τον αντίστοιχο λογάριθμο για ένα άτομο που έχει κάνει 1-5 κρίσεις συνολικά στη ζωή του και που έχει ίδιο ιστορικό εμφάνισης κρίσεων τον περασμένο χρόνο και ίδια οικογενειακή κατάσταση.
- ο λογάριθμος της εκτιμώμενης σχετικής πιθανότητας «επιτυχίας» (δηλαδή εμφάνισης κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης

του ερωτηματολογίου) για ένα άτομο που έχει κάνει πάνω από 21 κρίσεις συνολικά στη ζωή του είναι μεγαλύτερος κατά 1.0572 μονάδες σε σχέση με τον αντίστοιχο λογάριθμο για ένα άτομο που έχει κάνει 1-5 κρίσεις συνολικά στη ζωή του και που έχει ίδιο ιστορικό εμφάνισης κρίσεων τον περασμένο χρόνο και ίδια οικογενειακή κατάσταση.

- ο λογάριθμος της εκτιμώμενης σχετικής πιθανότητας «επιτυχίας» (δηλαδή εμφάνισης κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που έχει κάνει πάνω από 100 κρίσεις συνολικά στη ζωή του είναι μεγαλύτερος κατά 1.8620 μονάδες σε σχέση με τον αντίστοιχο λογάριθμο για ένα άτομο που έχει κάνει 1-5 κρίσεις συνολικά στη ζωή του και που έχει ίδιο ιστορικό εμφάνισης κρίσεων τον περασμένο χρόνο και ίδια οικογενειακή κατάσταση.
- ο λογάριθμος της εκτιμώμενης σχετικής πιθανότητας «επιτυχίας» (δηλαδή εμφάνισης κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που εμφάνισε κρίσεις τον περασμένο χρόνο είναι μεγαλύτερος κατά 3.5757 μονάδες σε σχέση με τον αντίστοιχο λογάριθμο για ένα άτομο που δεν εμφάνισε κρίσεις τον περασμένο χρόνο και που έχει κάνει τον ίδιο αριθμό κρίσεων συνολικά στη ζωή του και έχει την ίδια οικογενειακή κατάσταση.
- ο λογάριθμος της εκτιμώμενης σχετικής πιθανότητας «επιτυχίας» (δηλαδή εμφάνισης κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που είναι άγαμο ή χωρισμένο είναι μικρότερος κατά 1.8810 μονάδες σε σχέση με τον αντίστοιχο λογάριθμο για ένα άτομο που είναι παντρεμένο και που έχει κάνει τον ίδιο αριθμό κρίσεων συνολικά στη ζωή του και έχει ίδιο ιστορικό εμφάνισης κρίσεων τον περασμένο χρόνο.

Στον Πίνακα 4.6 βλέπουμε τις εκτιμήσεις και τα 95% διαστήματα εμπιστοσύνης των λόγων σχετικών πιθανοτήτων (odds ratios) για τις παραμέτρους του μοντέλου M_1.

Παράμετρος	Εκτίμηση OR	95% Δ.Ε. OR
Σταθερά	0.0783031	(0.0053133, 0.7038)
total_kriseis2	0.8736516	(0.0521456, 15.2328)
total_kriseis3	2.8782469	(0.2513509, 44.9896)
total_kriseis4	6.4367297	(0.6579790, 89.1447)
kriseis_lastyear1	35.7203721	(8.4614604, 239.9105)
marital_status2	0.1524411	(0.0263212, 0.6199)

Πίνακας 4.6: Εκτιμήσεις και 95% διαστήματα εμπιστοσύνης των λόγων σχετικών πιθανοτήτων (odds ratios) για τις παραμέτρους του μοντέλου M_1

Από τον Πίνακα 4.6 συμπεραίνουμε ότι:

- η εκτιμώμενη σχετική πιθανότητα «επιτυχίας» (δηλαδή εμφάνιση κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που έχει κάνει 1-5 κρίσεις συνολικά στη ζωή του, που επισκέφθηκε 0-1 φορές τον γιατρό τον τελευταίο χρόνο και που είναι παντρεμένο ισούται με 0.0783.
- η εκτιμώμενη σχετική πιθανότητα «επιτυχίας» (δηλαδή εμφάνιση κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που έχει κάνει 6-10 κρίσεις συνολικά στη ζωή του αυξάνεται πολλαπλασιαστικά κατά 0.8736 μονάδες σε σχέση με την αντίστοιχη σχετική πιθανότητα για ένα άτομο που έχει κάνει 1-5 κρίσεις συνολικά στη ζωή του και που έχει ίδιο ιστορικό εμφάνισης κρίσεων τον περασμένο χρόνο και ίδια οικογενειακή κατάσταση.
- η εκτιμώμενη σχετική πιθανότητα «επιτυχίας» (δηλαδή εμφάνιση κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που έχει κάνει πάνω από 21 κρίσεις συνολικά στη ζωή του αυξάνεται πολλαπλασιαστικά κατά 2.8782 μονάδες σε σχέση με την αντίστοιχη σχετική πιθανότητα για ένα άτομο που έχει κάνει 1-5 κρίσεις συνολικά στη ζωή του και που έχει ίδιο ιστορικό εμφάνισης κρίσεων τον περασμένο χρόνο και ίδια οικογενειακή κατάσταση.
- η εκτιμώμενη σχετική πιθανότητα «επιτυχίας» (δηλαδή εμφάνιση κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου)

για ένα άτομο που έχει κάνει πάνω από 100 κρίσεις συνολικά στη ζωή του αυξάνεται πολλαπλασιαστικά κατά 6.4367 μονάδες σε σχέση με την αντίστοιχη σχετική πιθανότητα για ένα άτομο που έχει κάνει 1-5 κρίσεις συνολικά στη ζωή του και που έχει ίδιο ιστορικό εμφάνισης κρίσεων τον περασμένο χρόνο και ίδια οικογενειακή κατάσταση.

- η εκτιμώμενη σχετική πιθανότητα «επιτυχίας» (δηλαδή εμφάνιση κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που εμφάνισε κρίσεις τον περασμένο χρόνο αυξάνεται πολλαπλασιαστικά κατά 35.7204 μονάδες σε σχέση με την αντίστοιχη σχετική πιθανότητα για ένα άτομο που δεν εμφάνισε κρίσεις τον περασμένο χρόνο και που έχει κάνει τον ίδιο αριθμό κρίσεων συνολικά στη ζωή του και έχει την ίδια οικογενειακή κατάσταση.
- η εκτιμώμενη σχετική πιθανότητα «επιτυχίας» (δηλαδή εμφάνιση κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που είναι άγαμο ή χωρισμένο αυξάνεται πολλαπλασιαστικά κατά 0.1524 μονάδες σε σχέση με την αντίστοιχη σχετική πιθανότητα για ένα άτομο που είναι παντρεμένο και που έχει κάνει τον ίδιο αριθμό κρίσεων συνολικά στη ζωή του και έχει ίδιο ιστορικό εμφάνισης κρίσεων τον περασμένο χρόνο.

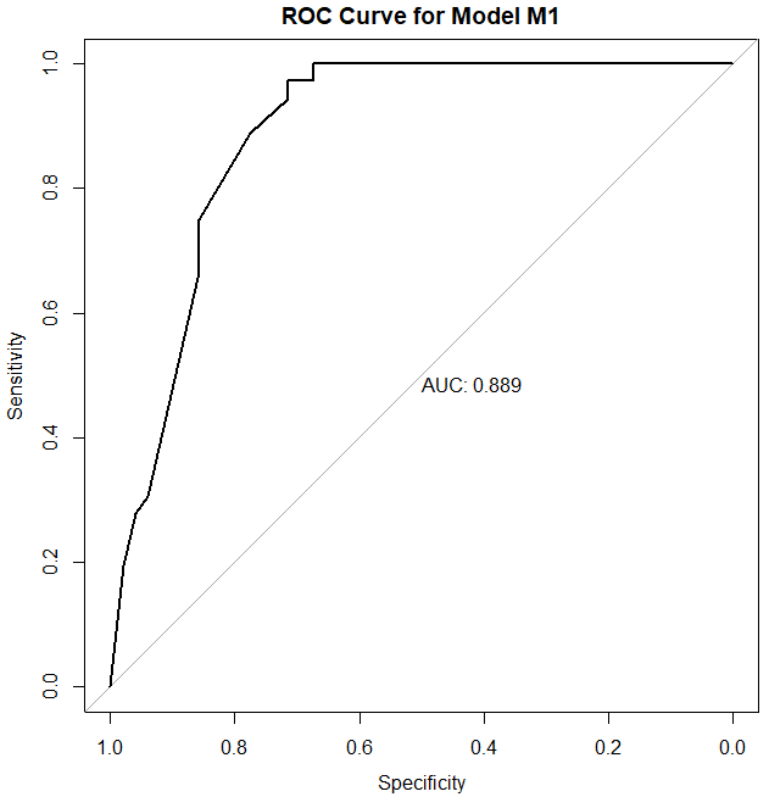
Επίσης, παρατηρούμε ότι μόνο τα odds ratios της σταθεράς, των ατόμων που εμφάνισαν κρίσεις τον περασμένο χρόνο και των ατόμων που είναι άγαμοι ή χωρισμένοι είναι στατιστικά σημαντικά, καθώς τα 95% διαστήματα εμπιστοσύνης αυτών των παραμέτρων δεν περιέχουν τη μονάδα.

Αξιολόγηση της προσαρμογής του μοντέλου M_1

Για να αξιολογήσουμε την προσαρμογή του μοντέλου M_1, θα κατασκευάσουμε τον πίνακα ταξινόμησης και την καμπύλη ROC.

	Παρατηρούμενο αποτέλεσμα		
		Επιτυχία	*Αποτυχία*
Εκτιμώμενο αποτέλεσμα	*Επιτυχία* (πάνω από την τιμή 0.5)	32	11
	Αποτυχία (κάτω από την τιμή 0.5)	4	38

Πίνακας 4.7: Πίνακας ταξινόμησης του μοντέλου M_1



Διάγραμμα 4.1: Καμπύλη ROC του μοντέλου M_1

Από τον Πίνακα 4.7 υπολογίζουμε τα παρακάτω μέτρα:

- $ευαισθησία = \frac{32}{32+4} = \frac{32}{36} = 0.89$
- $ειδικότητα = \frac{38}{11+38} = \frac{38}{49} = 0.78$
- $ακρίβεια = \frac{32+38}{32+11+4+38} = \frac{70}{85} = 0.8235$

Επίσης, από το Διάγραμμα 4.1 παρατηρούμε ότι το AUC ισούται με 0.889. Όλα αυτά μας οδηγούν στο συμπέρασμα ότι το μοντέλο M_1 είναι πολύ αποτελεσματικό.

4.3. Εφαρμογή της δίτιμης λογιστικής παλινδρόμησης στα δεδομένα μας όταν συμπληρώνουμε τις ελλιπείς τιμές (imputation)

Στην ενότητα αυτή θα βρούμε ποιες μεταβλητές, όπως έχουν προκύψει μετά την διαδικασία συμπλήρωσης των ελλιπών τιμών (imputation), συσχετίζονται με την ύπαρξη κρίσεων το τελευταίο δίμηνο και θα προσαρμόσουμε, αρχικά, ένα μοντέλο δίτιμης λογιστικής παλινδρόμησης με αυτές ως ανεξάρτητες. Η τεχνική imputation που θα χρησιμοποιήσουμε είναι η αντικατάσταση των ελλιπών τιμών της κάθε μεταβλητής με την επικρατούσα κατηγορία της.

Βασικά θεωρητικά στοιχεία για την διαδικασία συμπλήρωσης των ελλιπών τιμών (imputation) για ποιοτικές μεταβλητές

Για μερικές δειγματικές μονάδες σε δειγματοληπτικές έρευνες μπορεί να υπάρχει μερική απόκριση, με την έννοια ότι δεν αποκρίνονται σε όλες τις ερωτήσεις. Αυτό συμβαίνει όταν ο αποκρινόμενος αρνείται ή παραλείπει ή δεν μπορεί να απαντήσει σε κάποιες ερωτήσεις. Η διαδικασία συμπλήρωσης μη διαθέσιμων ή λανθασμένων στοιχείων της έρευνας με κατάλληλα στοιχεία για την δημιουργία ενός πλήρους αρχείου δεδομένων είναι γνωστή με τον όρο imputation.

Υπάρχουν διάφορες μέθοδοι imputation για ποιοτικές μεταβλητές. Οι πιο βασικές από αυτές είναι οι εξής:

- επαγωγικό imputation: Η μέθοδος αυτή προσδιορίζει την τιμή που λείπει με βεβαιότητα κάνοντας χρήση λογικών περιορισμών και άλλων στοιχείων της ίδιας μονάδας, όπως για παράδειγμα στην συγκεκριμένη έρευνα που αρκετοί συμμετέχοντες δεν συμπλήρωσαν το φύλο τους, αλλά μπορέσαμε και το καταλάβαμε από επόμενη ερώτηση που απάντησαν και αφορούσε την εκπλήρωση των στρατιωτικών τους υποχρεώσεων. Είναι ο ιδεατός, αλλά λιγότερο συχνός τύπος imputation.
- ιστορικό imputation: Η μέθοδος αυτή είναι περισσότερο χρήσιμη σε διαχρονικές έρευνες, ειδικά για μεταβλητές που είναι διαχρονικά ευσταθείς. Χρησιμοποιεί

τιμές που αναφέρθηκαν από την ίδια μονάδα σε προηγούμενη μέτρηση της έρευνας.

- αντικατάσταση των ελλιπών τιμών της μεταβλητής με την επικρατούσα κατηγορία της.
- τυχαίο συνολικό hot-deck imputation: Με αυτήν την μέθοδο η τιμή που λείπει αναπληρώνεται με την τιμή ενός «δότη», που επιλέγεται τυχαία από τους αποκριθέντες.
- regression imputation: Αυτή η μέθοδος χρησιμοποιεί παλινδρόμηση της μεταβλητής για την οποία χρειάζεται imputation σε ένα σύνολο μεταβλητών για τις οποίες υπάρχει απόκριση από όλες τις μονάδες. Η εξίσωση παλινδρόμησης χρησιμοποιείται μετά για πρόβλεψη των τιμών που λείπουν.

Αποτελέσματα ελέγχων συσχέτισης μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και των υπόλοιπων μεταβλητών της έρευνας (μετά από imputation)

Τα αποτελέσματα των ελέγχων συσχέτισης μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και των υπόλοιπων μεταβλητών της έρευνας, όπως έχουν προκύψει μετά από imputation, παρουσιάζονται στον Πίνακα 4.8.

Μεταβλητή	Έλεγχος συσχέτισης	p-value ελέγχου
Φύλο	X^2 του Pearson	0.911
Ηλικιακή ομάδα	X^2 τάσης	0.508
Τόπος μόνιμης κατοικίας μέχρι την ηλικία των 18 ετών	X^2 του Pearson με προσομοίωση Monte Carlo	0.317
Τόπος μόνιμης τωρινής κατοικίας	X^2 του Pearson με προσομοίωση Monte Carlo	0.840
Ηλικιακή ομάδα πρώτης κρίσης	X^2 τάσης	0.274
Επανάληψη κρίσεων ίδιου τύπου	X^2 του Pearson	0.208
Συνολικός αριθμός κρίσεων	X^2 τάσης	≈ 0
Αριθμός επισκέψεων στον γιατρό τον τελευταίο χρόνο	X^2 τάσης	0.036
Νοσηλεία λόγω των κρίσεων τον τελευταίο χρόνο	ακριβής έλεγχος του Fisher	0.10
Ιστορικό επιληψίας στο οικογενειακό περιβάλλον	ακριβής έλεγχος του Fisher	0.745
Ύπαρξη κρίσεων τον περασμένο χρόνο	ακριβής έλεγχος του Fisher	≈ 0
Ύπαρξη άλλου νοσήματος εκτός από τις κρίσεις	ακριβής έλεγχος του Fisher	1
Ικανότητα ελέγχου των κρίσεων	X^2 του Pearson	0.304
Βαθμός αποδοχής της επιληψίας	X^2 τάσης	0.088
Κοινωνική αντιμετώπιση	X^2 τάσης	0.693
Επαγγελματική κατάσταση	X^2 του Pearson με προσομοίωση Monte Carlo	0.335
Βαθμός δυσκολίας ανεύρεσης εργασίας εξαιτίας της νόσου	X^2 του Pearson με προσομοίωση Monte Carlo	0.366
Μορφωτικό επίπεδο	X^2 τάσης με προσομοίωση Monte Carlo	0.487
Διακοπή των σπουδών εξαιτίας της νόσου	ακριβής έλεγχος του Fisher	1
Οικογενειακή κατάσταση	X^2 του Pearson με προσομοίωση Monte Carlo	0.021
Ύπαρξη παιδιών	ακριβής έλεγχος του Fisher	0.174
Συχνότητα κατανάλωσης κρασιού	X^2 τάσης με προσομοίωση Monte Carlo	0.160

Συχνότητα κατανάλωσης ούζου, ουίσκι, βότκας, τζιν, κονιάκ	X^2 τάσης με προσομοίωση Monte Carlo	0.062[8]
Συχνότητα νυχτερινών εξόδων	X^2 τάσης	0.059[9]
Δίπλωμα οδήγησης	X^2 του Pearson	0.834
Οδήγηση	X^2 του Pearson	0.843
Βαθμός μοναξιάς	X^2 τάσης	0.694
Συχνότητα επισκέψεων στον γιατρό	X^2 τάσης	0.089
Ανάγκη για περισσότερη ενημέρωση για την νόσο	X^2 του Pearson με προσομοίωση Monte Carlo	0.167
Βαθμός φόβου των κρίσεων	X^2 του Pearson	0.512
Βαθμός ανασφάλειας για το μέλλον εξαιτίας των κρίσεων	X^2 του Pearson με προσομοίωση Monte Carlo	0.307
Επιδίωξη απόκτησης νέων φίλων	X^2 του Pearson με προσομοίωση Monte Carlo	1
Βαθμός επίδρασης της νόσου στις σχέσεις με το άλλο φύλο	X^2 του Pearson με προσομοίωση Monte Carlo	0.896

Πίνακας 4.8: Αποτελέσματα ελέγχων συσχέτισης μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και των υπόλοιπων μεταβλητών της έρευνας (μετά από imputation)

Από τον Πίνακα 4.8 παρατηρούμε ότι, μετά από imputation, οι μεταβλητές με τις οποίες συσχετίζεται η ύπαρξη κρίσεων το τελευταίο δίμηνο (στον πίνακα είναι σημειωμένες με μπλε χρώμα) είναι οι ίδιες με τις οποίες συσχετίζεται όταν αφαιρούμε τις ελλιπείς τιμές, δηλαδή είναι οι εξής:

- συνολικός αριθμός κρίσεων
- αριθμός επισκέψεων στον γιατρό τον τελευταίο χρόνο
- ύπαρξη κρίσεων τον περασμένο χρόνο
- οικογενειακή κατάσταση

[8] Το αποτέλεσμα προέκυψε μετά την συγχώνευση των εξής κατηγοριών της συχνότητας κατανάλωσης ούζου, ουίσκι, βότκας, τζιν, κονιάκ: «σπάνια» με «μία-δύο φορές το μήνα» και «μία φορά την εβδομάδα» με «δύο-τρεις φορές την εβδομάδα».

[9] Το αποτέλεσμα προέκυψε μετά την συγχώνευση των εξής κατηγοριών της συχνότητας νυχτερινών εξόδων: «λιγότερο από μία φορά το δίμηνο» με «μία φορά το δίμηνο», «μία φορά το μήνα» με «2-3 φορές το μήνα» και «μία φορά την εβδομάδα» με «2-3 φορές την εβδομάδα».

Συμβολισμοί

Θα χρησιμοποιήσουμε τους ίδιους συμβολισμούς που χρησιμοποιήσαμε και στην ενότητα 4.2.

Μοντέλο M_2

Στους Πίνακες 4.9 & 4.10 βλέπουμε τα αποτελέσματα που προκύπτουν όταν, έχοντας εφαρμόσει imputation, προσαρμόζουμε ένα μοντέλο δίτιμης λογιστικής παλινδρόμησης με ανεξάρτητες μεταβλητές αυτές με τις οποίες συσχετίζεται η ύπαρξη κρίσεων το τελευταίο δίμηνο.

Παράμετρος	Εκτίμηση	Τυπικό σφάλμα
Σταθερά	-3.6904	1.3670
total_kriseis2	0.4180	1.4000
total_kriseis3	1.8901	1.2590
total_kriseis4	2.1439	1.1636
doctor's_visit2	0.9896	0.7469
doctor's_visit3	1.0423	0.6990
kriseis_lastyear1	3.6245	0.8890
marital_status2	-2.2722	0.7689

Πίνακας 4.9: Εκτιμήσεις των παραμέτρων του μοντέλου M_2

Μοντέλο	Απόκλι-ση	Βαθμοί ελευθε-ρίας (β.ε.)	Μεταβολή στην απόκλιση	Μεταβο-λή στους β.ε.	p-value ελέγχου σημαντικό-τητας	Στατιστικά σημαντικό σε 5%
1 (Σταθερά)	127.998	95				
+total_kriseis	113.803	92	14.1950	3	0.0026514	ΝΑΙ
+doctor's_visit	110.688	90	3.1147	2	0.2106912	ΟΧΙ
+kriseis_lastyear	86.179	89	24.5091	1	≈ 0	ΝΑΙ
+marital_status	74.431	88	11.7478	1	0.0006092	ΝΑΙ

Πίνακας 4.10: Πίνακας ανάλυσης απόκλισης του μοντέλου M_2

Από τον Πίνακα 4.10 παρατηρούμε ότι η μεταβλητή που αφορά τον αριθμό των επισκέψεων στον γιατρό τον τελευταίο χρόνο δεν είναι στατιστικά σημαντική σε επίπεδο σημαντικότητας α=5%, ενώ οι υπόλοιπες μεταβλητές είναι στατιστικά σημαντικές. Επιπλέον, εφαρμόζοντας τον έλεγχο των Hosmer-Lemeshow διαπιστώνουμε ότι το μοντέλο M_2 είναι συνολικά επαρκές σε επίπεδο σημαντικότητας α=5% ($p-value\ ελέγχου\ Hosmer-Lemeshow = 0.8885 > \alpha = 0.05$). Στη συνέχεια, θα προσαρμόσουμε ένα μοντέλο (M_3) χωρίς την ανεξάρτητη μεταβλητή που αφορά τον αριθμό των επισκέψεων στον γιατρό τον τελευταίο χρόνο.

Μοντέλο M_3

Στους Πίνακες 4.11 & 4.12 βλέπουμε τα αποτελέσματα του μοντέλου που προκύπτουν όταν αφαιρούμε από το μοντέλο M_2 την ανεξάρτητη μεταβλητή που αφορά τον αριθμό των επισκέψεων στον γιατρό τον τελευταίο χρόνο.

Παράμετρος	Εκτίμηση	Τυπικό σφάλμα
Σταθερά	-2.8535	1.1556
total_kriseis2	0.3380	1.3215
total_kriseis3	1.5309	1.1563
total_kriseis4	2.0211	1.0758
kriseis_lastyear1	3.4961	0.8454
marital_status2	-2.1977	0.7667

Πίνακας 4.11: Εκτιμήσεις των παραμέτρων του μοντέλου M_3

Μοντέλο	Απόκλι-ση	Βαθμοί ελευθε-ρίας (β.ε.)	Μεταβολή στην απόκλιση	Μεταβο-λή στους β.ε.	p-value ελέγχου σημαντικό-τητας	Στατιστικά σημαντικό σε 5%
1 (Σταθερά)	127.998	95				
+total_kriseis	113.803	92	14.195	3	0.0026514	ΝΑΙ
+kriseis_lastyear	88.433	91	25.369	1	≈ 0	ΝΑΙ
+marital_status	77.318	90	11.115	1	0.0008564	ΝΑΙ

Πίνακας 4.12: Πίνακας ανάλυσης απόκλισης του μοντέλου M_3

Από τον Πίνακα 4.12 παρατηρούμε ότι όλες οι μεταβλητές του μοντέλου M_3 είναι στατιστικά σημαντικές σε επίπεδο σημαντικότητας α=5%. Επιπλέον, εφαρμόζοντας τον έλεγχο των Hosmer-Lemeshow διαπιστώνουμε ότι το μοντέλο M_3 είναι συνολικά επαρκές σε επίπεδο σημαντικότητας α=5% ($p-value\ ελέγχου\ Hosmer-Lemeshow = 0.9856 > \alpha = 0.05$).

Σύγκριση των μοντέλων M_2 και M_3

Τα μοντέλα M_2 και M_3 είναι εμφωλευμένα μεταξύ τους, καθώς το σύνολο των επεξηγηματικών μεταβλητών του M_3 είναι υποσύνολο αυτών του M_2. Επομένως, για να μπορέσουμε να τα συγκρίνουμε θα χρησιμοποιήσουμε τον έλεγχο του λόγου πιθανοφανειών. Δηλαδή θα κάνουμε τον εξής έλεγχο:

$$H_0: \textit{τα δύο μοντέλα δεν διαφέρουν σημαντικά}$$

$$H_1: \textit{το μοντέλο } M_2 \text{ είναι καλύτερο από το μοντέλο } M_3$$

Το μοντέλο M_2 έχει $df_2 = 88$ βαθμούς ελευθερίας και ο λογάριθμος της συνάρτησης πιθανοφάνειάς του ισούται με $l(M_2) = logL(M_2) = -37.21549$, ενώ το μοντέλο M_3 έχει $df_3 = 90$ βαθμούς ελευθερίας και ο λογάριθμος της συνάρτησης πιθανοφάνειάς του ισούται με $l(M_3) = logL(M_3) = -38.65913$. Συνεπώς, η στατιστική συνάρτηση του ελέγχου ισούται με $G^2 = -2\big(l(M_3) - l(M_2)\big) = -2(-38.65913 + 37.21549) = -2(-1.44364) = 2.88728$. Άρα σε επίπεδο σημαντικότητας α=5% δεν απορρίπτεται η μηδενική υπόθεση, καθώς $G^2 < x^2_{2,0.05} = 5.9915$. Οπότε τα δύο μοντέλα δεν διαφέρουν σημαντικά και για τον λόγο αυτό θα συνεχίσουμε την ανάλυσή μας με το απλούστερο μοντέλο, δηλαδή με το μοντέλο M_3.

Ερμηνεία των εκτιμητών των παραμέτρων του μοντέλου M_3

Το μοντέλο M_3 που προσαρμόστηκε έχει την εξής μορφή:

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = -2.8535 + 0.3380 * total\_kriseis2 + 1.5309 * total\_kriseis3 +$$

$$2.0211 * total\_kriseis4 + 3.4961 * kriseis\_lastyear1 - 2.1977 * marital\_status2$$

Η ερμηνεία των εκτιμητών των παραμέτρων του μοντέλου M_3 είναι η εξής:

- ο λογάριθμος της εκτιμώμενης σχετικής πιθανότητας «επιτυχίας» (δηλαδή εμφάνισης κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που έχει κάνει 1-5 κρίσεις συνολικά στη ζωή του, που επισκέφθηκε 0-1 φορές τον γιατρό τον τελευταίο χρόνο και που είναι παντρεμένο ισούται με -2.8535.
- ο λογάριθμος της εκτιμώμενης σχετικής πιθανότητας «επιτυχίας» (δηλαδή εμφάνισης κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που έχει κάνει 6-10 κρίσεις συνολικά στη ζωή του είναι μεγαλύτερος κατά 0.3380 μονάδες σε σχέση με τον αντίστοιχο λογάριθμο για ένα άτομο που έχει κάνει 1-5 κρίσεις συνολικά στη ζωή του και που έχει ίδιο ιστορικό εμφάνισης κρίσεων τον περασμένο χρόνο και ίδια οικογενειακή κατάσταση.
- ο λογάριθμος της εκτιμώμενης σχετικής πιθανότητας «επιτυχίας» (δηλαδή εμφάνισης κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης

του ερωτηματολογίου) για ένα άτομο που έχει κάνει πάνω από 21 κρίσεις συνολικά στη ζωή του είναι μεγαλύτερος κατά 1.5309 μονάδες σε σχέση με τον αντίστοιχο λογάριθμο για ένα άτομο που έχει κάνει 1-5 κρίσεις συνολικά στη ζωή του και που έχει ίδιο ιστορικό εμφάνισης κρίσεων τον περασμένο χρόνο και ίδια οικογενειακή κατάσταση.

- ο λογάριθμος της εκτιμώμενης σχετικής πιθανότητας «επιτυχίας» (δηλαδή εμφάνισης κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που έχει κάνει πάνω από 100 κρίσεις συνολικά στη ζωή του είναι μεγαλύτερος κατά 2.0211 μονάδες σε σχέση με τον αντίστοιχο λογάριθμο για ένα άτομο που έχει κάνει 1-5 κρίσεις συνολικά στη ζωή του και που έχει ίδιο ιστορικό εμφάνισης κρίσεων τον περασμένο χρόνο και ίδια οικογενειακή κατάσταση.
- ο λογάριθμος της εκτιμώμενης σχετικής πιθανότητας «επιτυχίας» (δηλαδή εμφάνισης κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που εμφάνισε κρίσεις τον περασμένο χρόνο είναι μεγαλύτερος κατά 3.4961 μονάδες σε σχέση με τον αντίστοιχο λογάριθμο για ένα άτομο που δεν εμφάνισε κρίσεις τον περασμένο χρόνο και που έχει κάνει τον ίδιο αριθμό κρίσεων συνολικά στη ζωή του και έχει την ίδια οικογενειακή κατάσταση.
- ο λογάριθμος της εκτιμώμενης σχετικής πιθανότητας «επιτυχίας» (δηλαδή εμφάνισης κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που είναι άγαμο ή χωρισμένο είναι μικρότερος κατά 2.1977 μονάδες σε σχέση με τον αντίστοιχο λογάριθμο για ένα άτομο που είναι παντρεμένο και που έχει κάνει τον ίδιο αριθμό κρίσεων συνολικά στη ζωή του και έχει ίδιο ιστορικό εμφάνισης κρίσεων τον περασμένο χρόνο.

Στον Πίνακα 4.13 βλέπουμε τις εκτιμήσεις και τα 95% διαστήματα εμπιστοσύνης των λόγων σχετικών πιθανοτήτων (odds ratios) για τις παραμέτρους του μοντέλου M_3.

Παράμετρος	Εκτίμηση OR	95% Δ.Ε. OR
Σταθερά	0.057644	(0.0044979, 0.4458)
total_kriseis2	1.402158	(0.1021855, 21.3136)
total_kriseis3	4.622324	(0.5565326, 58.4423)
total_kriseis4	7.546885	(1.0879138, 83.0550)
kriseis_lastyear1	32.985283	(7.6839935, 228.8295)
marital_status2	0.111061	(0.0195588, 0.4310)

Πίνακας 4.13: Εκτιμήσεις και 95% διαστήματα εμπιστοσύνης των λόγων σχετικών πιθανοτήτων (odds ratios) για τις παραμέτρους του μοντέλου M_3

Από τον Πίνακα 4.13 συμπεραίνουμε ότι:

- η εκτιμώμενη σχετική πιθανότητα «επιτυχίας» (δηλαδή εμφάνιση κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που έχει κάνει 1-5 κρίσεις συνολικά στη ζωή του, που επισκέφθηκε 0-1 φορές τον γιατρό τον τελευταίο χρόνο και που είναι παντρεμένο ισούται με 0.05764.
- η εκτιμώμενη σχετική πιθανότητα «επιτυχίας» (δηλαδή εμφάνιση κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που έχει κάνει 6-10 κρίσεις συνολικά στη ζωή του αυξάνεται πολλαπλασιαστικά κατά 1.4021 μονάδες σε σχέση με την αντίστοιχη σχετική πιθανότητα για ένα άτομο που έχει κάνει 1-5 κρίσεις συνολικά στη ζωή του και που έχει ίδιο ιστορικό εμφάνισης κρίσεων τον περασμένο χρόνο και ίδια οικογενειακή κατάσταση.
- η εκτιμώμενη σχετική πιθανότητα «επιτυχίας» (δηλαδή εμφάνιση κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που έχει κάνει πάνω από 21 κρίσεις συνολικά στη ζωή του αυξάνεται πολλαπλασιαστικά κατά 4.6223 μονάδες σε σχέση με την αντίστοιχη σχετική πιθανότητα για ένα άτομο που έχει κάνει 1-5 κρίσεις συνολικά στη ζωή του και που έχει ίδιο ιστορικό εμφάνισης κρίσεων τον περασμένο χρόνο και ίδια οικογενειακή κατάσταση.
- η εκτιμώμενη σχετική πιθανότητα «επιτυχίας» (δηλαδή εμφάνιση κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου)

για ένα άτομο που έχει κάνει πάνω από 100 κρίσεις συνολικά στη ζωή του αυξάνεται πολλαπλασιαστικά κατά 7.5468 μονάδες σε σχέση με την αντίστοιχη σχετική πιθανότητα για ένα άτομο που έχει κάνει 1-5 κρίσεις συνολικά στη ζωή του και που έχει ίδιο ιστορικό εμφάνισης κρίσεων τον περασμένο χρόνο και ίδια οικογενειακή κατάσταση.

- η εκτιμώμενη σχετική πιθανότητα «επιτυχίας» (δηλαδή εμφάνιση κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που εμφάνισε κρίσεις τον περασμένο χρόνο αυξάνεται πολλαπλασιαστικά κατά 32.9852 μονάδες σε σχέση με την αντίστοιχη σχετική πιθανότητα για ένα άτομο που δεν εμφάνισε κρίσεις τον περασμένο χρόνο και που έχει κάνει τον ίδιο αριθμό κρίσεων συνολικά στη ζωή του και έχει την ίδια οικογενειακή κατάσταση.
- η εκτιμώμενη σχετική πιθανότητα «επιτυχίας» (δηλαδή εμφάνιση κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου) για ένα άτομο που είναι άγαμο ή χωρισμένο αυξάνεται πολλαπλασιαστικά κατά 0.111 μονάδες σε σχέση με την αντίστοιχη σχετική πιθανότητα για ένα άτομο που είναι παντρεμένο και που έχει κάνει τον ίδιο αριθμό κρίσεων συνολικά στη ζωή του και έχει ίδιο ιστορικό εμφάνισης κρίσεων τον περασμένο χρόνο.

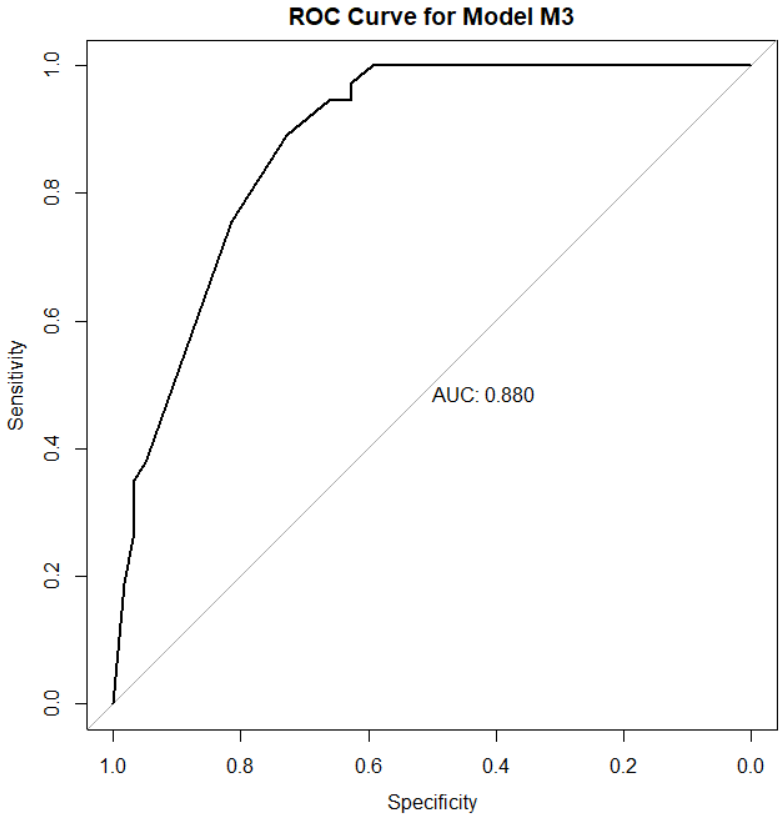
Επίσης, παρατηρούμε ότι μόνο τα odds ratios της σταθεράς, των ατόμων που έχουν κάνει πάνω από 100 κρίσεις συνολικά στη ζωή τους, των ατόμων που εμφάνισαν κρίσεις τον περασμένο χρόνο και των ατόμων που είναι άγαμοι ή χωρισμένοι είναι στατιστικά σημαντικά, καθώς τα 95% διαστήματα εμπιστοσύνης αυτών των παραμέτρων δεν περιέχουν τη μονάδα.

Αξιολόγηση της προσαρμογής του μοντέλου M_3

Για να αξιολογήσουμε την προσαρμογή του μοντέλου M_3, θα κατασκευάσουμε τον πίνακα ταξινόμησης και την καμπύλη ROC.

	Παρατηρούμενο αποτέλεσμα		
		Επιτυχία	*Αποτυχία*
Εκτιμώμενο αποτέλεσμα	*Επιτυχία* (πάνω από την τιμή 0.5)	28	11
	Αποτυχία (κάτω από την τιμή 0.5)	9	48

Πίνακας 4.14: Πίνακας ταξινόμησης του μοντέλου M_3



Διάγραμμα 4.2: Καμπύλη ROC του μοντέλου M_3

Από τον Πίνακα 4.14 υπολογίζουμε τα παρακάτω μέτρα:

- $ευαισθησία = \frac{28}{28+9} = \frac{28}{37} = 0.76$
- $ειδικότητα = \frac{48}{11+48} = \frac{48}{59} = 0.8136$
- $ακρίβεια = \frac{28+48}{28+11+9+48} = \frac{76}{96} = 0.7917$

Επίσης, από το Διάγραμμα 4.2 παρατηρούμε ότι το AUC ισούται με 0.88. Όλα αυτά μας οδηγούν στο συμπέρασμα ότι το μοντέλο M_3 είναι πολύ αποτελεσματικό.

Σύγκριση των μοντέλων M_1 και M_3

Τα αποτελέσματα που προκύπτουν όταν προσαρμόζουμε ένα μοντέλο δίτιμης λογιστικής παλινδρόμησης έχοντας αφαιρέσει τις ακραίες τιμές (M_1) είναι πολύ παρόμοια με τα αποτελέσματα του μοντέλου που προκύπτει όταν εφαρμόζουμε imputation (M_3). Πιο συγκεκριμένα, και τα δύο μοντέλα έχουν τις ίδιες ανεξάρτητες μεταβλητές, είναι συνολικά επαρκή, έχουν παρόμοιες εκτιμήσεις παραμέτρων και σχεδόν ταυτόσημη προβλεπτική ικανότητα. Η μόνη διαφορά είναι η εκτίμηση της παραμέτρου $total\_kriseis2$, καθώς με το μοντέλο M_1 η εκτίμηση είναι -0.1351, ενώ με το μοντέλο M_3 είναι 0.3380.

ΚΕΦΑΛΑΙΟ 5
Ταξινόμηση μέσω δέντρων αποφάσεων και εξόρυξη κανόνων συσχετίσεων

Στο κεφάλαιο αυτό κατασκευάζουμε δέντρο αποφάσεων με τον αλγόριθμο Conditional Inference Decision Trees για την μεταβλητή-στόχο kriseis2, που αφορά την ύπαρξη κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου. Επιπλέον, εφαρμόζουμε την μέθοδο εξόρυξης κανόνων συσχετίσεων για να βρούμε τους κανόνες με βάση τους οποίους ένα άτομο εμφάνισε κρίσεις το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου. Και για τις δύο αυτές τεχνικές εξόρυξης δεδομένων, θα χρησιμοποιήσουμε τα πλήρη δεδομένα και ως επεξηγηματικές μεταβλητές θα χρησιμοποιήσουμε αυτές με τις οποίες συσχετίζεται η ύπαρξη κρίσεων το τελευταίο δίμηνο.

5.1. Βασικά θεωρητικά στοιχεία για την ταξινόμηση μέσω δέντρων αποφάσεων και την εξόρυξη κανόνων συσχετίσεων

Στην ενότητα αυτή δίνουμε το βασικό θεωρητικό πλαίσιο της ταξινόμησης μέσω δέντρων αποφάσεων και της εξόρυξης κανόνων συσχετίσεων.

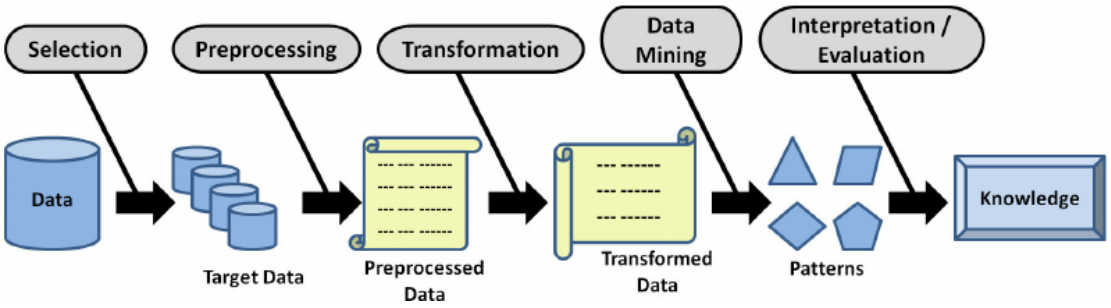
5.1.1. Εξόρυξη Δεδομένων και Ανακάλυψη Γνώσης από Βάσεις Δεδομένων

Εξόρυξη Δεδομένων (Data Mining) είναι η ανάλυση συχνά μεγάλων παρατηρούμενων συνόλων δεδομένων με σκοπό να βρούμε σχέσεις που δεν υποψιαζόμαστε και να συνοψίσουμε τα δεδομένα με καινοτόμους τρόπους, κατανοητούς και χρήσιμους για τον κάτοχο των δεδομένων. Στη βιβλιογραφία η Εξόρυξη Δεδομένων συναντάται και με τον όρο Ανακάλυψη Γνώσης από Βάσεις Δεδομένων (Knowledge Discovery in Databases, KDD), όμως στην πραγματικότητα αποτελεί ένα μόνο επιμέρους βήμα της. Τα βασικά στάδια της Ανακάλυψης Γνώσης από Βάσεις Δεδομένων (Διάγραμμα 5.1) είναι τα εξής:

1. Συλλογή Δεδομένων (Data Collection): Η συλλογή των δεδομένων συνήθως γίνεται είτε αυτόματα, όπως για παράδειγμα με χρήση αισθητήρων, είτε μη αυτόματα, όπως για παράδειγμα με χρήση ερωτηματολογίων.
2. Προεπεξεργασία Δεδομένων (Preprocessing): Η προεπεξεργασία των δεδομένων είναι το πιο σημαντικό στάδιο και γίνεται με στόχο τον καθαρισμό τους, δηλαδή την τακτοποίηση εσφαλμένων, προβληματικών ή ελλιπόντων δεδομένων. Μπορεί να απαιτήσει έως και το 60% της συνολικής προσπάθειας και αυτό διότι, αν τα

δεδομένα δεν είναι «καθαρά» και στην κατάλληλη μορφή, δεν θα προκύψουν ποιοτικά αποτελέσματα.

3. Μετασχηματισμός Δεδομένων (Transformation): Πρόκειται για τη μετατροπή των δεδομένων κάτω από ένα κοινό πλαίσιο, για επεξεργασία. Χρησιμοποιείται κυρίως για την εξομάλυνση των δεδομένων και απομάκρυνση θορύβου, για την κανονικοποίησή τους, δηλαδή την κλιμάκωση των χαρακτηριστικών του συνόλου δεδομένων σε ένα συγκεκριμένο και περιορισμένο εύρος τιμών, ή για τη δημιουργία νέων χαρακτηριστικών από τα ήδη υπάρχοντα.
4. Εξόρυξη Δεδομένων (Data Mining): Σε αυτό το στάδιο εφαρμόζεται κάποιος αλγόριθμος για την παραγωγή ενός μοντέλου, συνήθως κατηγοριοποίησης ή πρόβλεψης. Εμείς θέλουμε να χρησιμοποιήσουμε το μοντέλο αυτό, το οποίο δημιουργήθηκε με βάση κάποια γνωστά δεδομένα, έτσι ώστε να μπορεί να μας δώσει απάντηση για την τιμή μιας μεταβλητής-στόχου για νέα, άγνωστα δεδομένα.
5. Διερμηνεία και Αξιολόγηση (Interpretation/Evaluation): Σε αυτό το τελευταίο στάδιο γίνεται η διερμηνεία και η αξιολόγηση των αποτελεσμάτων (όχι του μοντέλου) που παρήχθησαν από την όλη διαδικασία.



Διάγραμμα 5.1: Βασικά στάδια της Ανακάλυψης Γνώσης από Βάσεις Δεδομένων

5.1.2. Τύποι μοντέλων που παράγονται από το στάδιο της Εξόρυξης Δεδομένων

Τα μοντέλα που παράγονται από το στάδιο της Εξόρυξης Δεδομένων διακρίνονται σε δύο βασικούς τύπους: τα μοντέλα πρόβλεψης (predictive) και τα περιγραφικά μοντέλα (descriptive).

Ένα μοντέλο πρόβλεψης έχει στόχο να προβλέψει τιμές για ένα συγκεκριμένο χαρακτηριστικό που παρουσιάζει ενδιαφέρον και που πιθανώς βασίζεται στη συμπεριφορά άλλων χαρακτηριστικών.

Ένα περιγραφικό μοντέλο βρίσκει πρότυπα (patterns) ή σχέσεις (relations) που υπάρχουν στα δεδομένα και μελετά τις ιδιότητές τους, ώστε να δοθεί μια αιτιολόγηση της συμπεριφοράς τους.

5.1.3. Μέθοδοι Εξόρυξης Δεδομένων

Υπάρχουν αρκετές μέθοδοι εξόρυξης δεδομένων. Ενδεικτικά, αναφέρουμε τις εξής:

- ταξινόμηση (classification)
- συσταδοποίηση (clustering)
- εξόρυξη κανόνων συσχετίσεων (association rule mining)
- ανίχνευση ανωμαλιών (anomaly detection)

Στην παρούσα εργασία θα ασχοληθούμε μόνο με την ταξινόμηση, και πιο συγκεκριμένα με την ταξινόμηση μέσω δέντρων αποφάσεων, και με την εξόρυξη κανόνων συσχετίσεων.

5.1.4. Ταξινόμηση

Ταξινόμηση (classification) είναι η διαδικασία εξαγωγής κανόνων από ένα σύνολο προβλεπτικών μεταβλητών που θα χρησιμοποιηθούν για την πρόβλεψη μιας κατηγορικής μεταβλητής.

Υπάρχουν διάφοροι αλγόριθμοι ταξινόμησης. Ενδεικτικά, αναφέρουμε τους εξής:

- Λογιστική Παλινδρόμηση (Logistic Regression)
- k Πλησιέστεροι Γείτονες (k Nearest Neighbours, kNN)
- Αφελής Κατηγοριοποιητής Bayes (Naïve Bayes Classifier)
- Δέντρα Αποφάσεων (Decision Trees)
- Τυχαία Δάση (Random Forests)
- Μηχανές Διανυσματικής Στήριξης (Support Vector Machines, SVM)

Στην παρούσα εργασία θα ασχοληθούμε μόνο με τα δέντρα αποφάσεων.

Πριν την εφαρμογή ενός αλγορίθμου ταξινόμησης, θα πρέπει να χωρίσουμε το δείγμα μας σε δύο σύνολα: το σύνολο εκμάθησης (training set) και το σύνολο ελέγχου (test set). Ως σύνολο εκμάθησης ορίζουμε το σύνολο των δεδομένων που

χρησιμοποιείται για την εξαγωγή των κανόνων και το οποίο είναι επισημασμένο, δηλαδή οι κατηγορίες του είναι γνωστές εξαρχής, ενώ ως σύνολο ελέγχου ονομάζουμε κάθε μη επισημασμένο σύνολο, στο οποίο οι κατηγορίες είναι άγνωστες και πρέπει να προβλεφθούν από το μοντέλο.

Τα στάδια υλοποίησης ενός αλγορίθμου ταξινόμησης είναι τα εξής:

1. Χωρισμός του δείγματος σε δύο σύνολα: το σύνολο εκμάθησης και το σύνολο ελέγχου.
2. Εκμάθηση του μοντέλου στο σύνολο εκμάθησης.
3. Εφαρμογή του μοντέλου στο σύνολο ελέγχου.
4. Αξιολόγηση της αποτελεσματικότητας του μοντέλου μέσω των πινάκων ταξινόμησης (classification tables).

5.1.5. Δέντρα αποφάσεων

Τα δέντρα αποφάσεων (decision trees) είναι ένα μοντέλο ταξινόμησης. Από θεωρητικής άποψης, τα δέντρα αποφάσεων είναι γράφοι που αποτελούνται από εσωτερικούς κόμβους και φύλλα. Εσωτερικούς κόμβους λέμε τους κόμβους που έχουν παιδιά, ενώ φύλλα λέμε τους κόμβους του κατώτερου επιπέδου, οι οποίοι δεν έχουν απογόνους. Ο τρόπος αναπαράστασης γίνεται ως εξής:

- κάθε εσωτερικός κόμβος του δέντρου ονοματίζεται με το όνομα ενός χαρακτηριστικού.
- κάθε κλαδί/σύνδεση δύο κόμβων ονοματίζεται με μια συνθήκη ή μια τιμή από αυτές που δύναται να λάβει το χαρακτηριστικό του γονικού κόμβου.
- κάθε φύλλο ονοματίζεται με το όνομα μιας κλάσης.

Υπάρχουν διάφοροι αλγόριθμοι κατασκευής δέντρων αποφάσεων. Οι πιο δημοφιλείς από αυτούς είναι οι εξής:

- Classification and Regression Trees (CART)
- Iterative Dichotomiser (ID3)
- C4.5
- C5
- Conditional Inference Decision Trees

Στην παρούσα εργασία θα ασχοληθούμε μόνο με τον αλγόριθμο Conditional Inference Decision Trees.

<u>Αλγόριθμος Conditional Inference Decision Trees</u>

Ο αλγόριθμος Conditional Inference Decision Trees είναι ένας αλγόριθμος κατασκευής δέντρων αποφάσεων. Είναι προτιμότερος από άλλους αλγορίθμους, γιατί εξασφαλίζει αμεροληψία. Τα στάδια υλοποίησής του είναι τα εξής:

1. Ελέγχει την μηδενική υπόθεση της ανεξαρτησίας μεταξύ της μεταβλητής απόκρισης και όλων των επεξηγηματικών μεταβλητών. Ο αλγόριθμος τερματίζεται αν η μηδενική υπόθεση δεν μπορεί να απορριφθεί. Διαφορετικά, επιλέγει την επεξηγηματική μεταβλητή που έχει την ισχυρότερη συσχέτιση με την μεταβλητή απόκρισης. Τον βαθμό συσχέτισης τον καταλαβαίνουμε από το p-value του ελέγχου ανεξαρτησίας μεταξύ της μεταβλητής απόκρισης και κάθε επεξηγηματικής μεταβλητής.
2. Εφαρμόζει δυαδικό διαχωρισμό στην επιλεγμένη επεξηγηματική μεταβλητή.
3. Εκτελεί επαναληπτικά τα Βήματα 1 & 2.

5.1.6. Εξόρυξη κανόνων συσχετίσεων

Η εξόρυξη κανόνων συσχετίσεων (association rule mining) έχει στόχο την ανακάλυψη και διατύπωση σχέσεων που υπάρχουν στα δεδομένα. Οι σχέσεις αυτές προκύπτουν από τη συχνή ταυτόχρονη εμφάνιση τιμών δεδομένων.

Το συνηθέστερο πεδίο εφαρμογής αλλά και το συνηθέστερο παράδειγμα κανόνων συσχετίσεων είναι η ανάλυση του καλαθιού αγορών (market basket analysis). Η ανάλυση του καλαθιού αγορών αναφέρεται στη μελέτη των αγορών που πραγματοποιούν οι πελάτες ενός καταστήματος. Κάθε αγορά περιλαμβάνει ένα σύνολο προϊόντων. Αναλύοντας τα σύνολα προϊόντων των πωλήσεων, μπορούν να βρεθούν ομάδες προϊόντων οι οποίες πωλούνται συχνά μαζί. Από αυτές τις ομάδες υπολογίζονται κανόνες της μορφής «εάν ένας πελάτης αγοράσει το προϊόν Α, έχει 60% πιθανότητα να αγοράσει ταυτόχρονα και το προϊόν Β, ενώ η ταυτόχρονη πώληση των προϊόντων Α και Β παρουσιάζεται στο 8% του συνόλου των πωλήσεων». Η πληροφορία αυτή είναι πολλαπλώς αξιοποιήσιμη από τον υπεύθυνο πωλήσεων. Μπορεί να χρησιμοποιηθεί για τη διαρρύθμιση και την τοποθέτηση των προϊόντων μέσα στο κατάστημα. Τοποθετώντας δύο προϊόντα, τα οποία πωλούνται συχνά μαζί, σε γειτονικές θέσεις, επιτυγχάνεται αύξηση των πωλήσεων. Ο πελάτης, ο οποίος περιπλανιέται στο κατάστημα, όταν αγοράσει το προϊόν Α, έχει περισσότερες πιθανότητες και προτρέπεται να αγοράσει το συγγενές προϊόν Β, εάν το βρει σε μια γειτονική θέση. Σε αντίθετη περίπτωση είναι πιθανόν να ξεχάσει το προϊόν Β και να το αγοράσει κάποια άλλη χρονική στιγμή από άλλο κατάστημα. Άλλος τρόπος αξιοποίησης αυτής της πληροφορίας είναι ο σχεδιασμός προσφορών. Ο υπεύθυνος πωλήσεων, προσφέροντας μια δελεαστική τιμή για το προϊόν Α, προσελκύει πελάτες και αυξάνει τις πωλήσεις του προϊόντος Α. Αν τοποθετήσει σε γειτονικά ράφια το συγγενές προϊόν Β, θα επιτύχει αύξηση πωλήσεων του Β, καθώς οι πελάτες, γνωρίζοντας ότι εξοικονόμησαν χρήματα από την αγορά του Α, πιθανόν να προβούν σε πρόσθετες αγορές.

Ορισμοί

Έστω $I = \{i_1, i_2, \dots, i_n\}$ ένα σύνολο από διακριτά στοιχεία, που αποκαλούνται αντικείμενα (items). Έστω ακόμα $D = \{t_1, t_2, \dots, t_m\}$ ένα σύνολο από συναλλαγές (transactions), όπου κάθε συναλλαγή T είναι ένα σύνολο από αντικείμενα και ονομάζεται στοιχειοσύνολο (itemset). Αν το στοιχειοσύνολο έχει k στοιχεία, τότε ονομάζεται k-στοιχειοσύνολο (k-itemset). Ισχύει ότι $T \subseteq I$. Κάθε συναλλαγή ταυτίζεται με ένα μοναδικό αναγνωριστικό που ονομάζεται TID (Transaction ID).

Ένας κανόνας συσχέτισης έχει τη μορφή $X \Rightarrow Y$, όπου $X \subseteq I$, $Y \subseteq I$ και $X \cap Y = \emptyset$. Το πρώτο μέλος του κανόνα ονομάζεται υπόθεση, ενώ το δεύτερο ονομάζεται συμπέρασμα.

Υπάρχουν δύο ποσοτικά μεγέθη που καθορίζουν πόσο ισχυρός είναι ο κανόνας $X \Rightarrow Y$ και είναι τα εξής:

1. υποστήριξη (support): Η υποστήριξη του κανόνα $X \Rightarrow Y$ είναι το ποσοστό των συναλλαγών στο D που περιέχουν και το X και το Y. Μαθηματικά, αυτό ορίζεται με τη σχέση: $supp(X \Rightarrow Y) = P(X \cup Y)$.
2. εμπιστοσύνη (confidence): Η εμπιστοσύνη του κανόνα $X \Rightarrow Y$ είναι η δεσμευμένη πιθανότητα εμφάνισης του Y, όταν εμφανίζεται το X. Με απλούστερα λόγια, επιλέγονται μόνο οι συναλλαγές που περιέχουν το X και επί αυτών των συναλλαγών υπολογίζεται το ποσοστό εκείνων που περιέχουν το Y. Μαθηματικά, αυτό ορίζεται με τη σχέση: $conf(X \Rightarrow Y) = P(Y|X)$.

Για την καλύτερη κατανόηση των εννοιών «υποστήριξη» και «εμπιστοσύνη» παραθέτουμε ένα παράδειγμα. Ο Πίνακας 5.1 περιέχει το σύνολο των συναλλαγών ενός καταστήματος. Για κάθε συναλλαγή καταγράφεται το TID και τα εμπορεύματα που πωλήθηκαν σε αυτήν τη συναλλαγή. Τα A, B, Γ, Δ, E είναι διάφορα εμπορεύματα.

TID	Εμπορεύματα
101	Α, Β, Γ
102	Γ, Δ
103	Α, Β
104	Α, Β, Δ
105	Α, Δ
106	Β, Γ

Πίνακας 5.1: Δεδομένα συναλλαγών

Θεωρούμε τον κανόνα $A \Rightarrow B$, ο οποίος σημαίνει ότι όταν κάποιος αγοράζει το προϊόν A, τότε αγοράζει και το προϊόν B. Από τον Πίνακα 5.1 παρατηρούμε ότι σε τρεις από τις συνολικά έξι συναλλαγές (101, 103, 104) πωλούνται ταυτόχρονα τα προϊόντα A και B. Άρα η υποστήριξη του κανόνα είναι 3/6, δηλαδή 50%. Επίσης, παρατηρούμε ότι το προϊόν A εμφανίζεται σε τέσσερεις συναλλαγές (101, 103, 104, 105) και ότι σε τρεις από αυτές (101, 103, 104) εμφανίζεται και το προϊόν B. Επομένως, η εμπιστοσύνη του κανόνα είναι 3/4, δηλαδή 75%.

Πρόβλημα εξαγωγής κανόνων συσχετίσεων

Προκειμένου να εξάγουμε έναν κανόνα συσχέτισης, πρέπει να ικανοποιούνται κάποια κατώτατα όρια τόσο για την υποστήριξη όσο και για την εμπιστοσύνη. Ο κανόνας πρέπει να έχει υποστήριξη μεγαλύτερη από το όριο, που ονομάζεται ελάχιστη υποστήριξη (minsup), και η εμπιστοσύνη πρέπει να είναι μεγαλύτερη από το όριο, που ονομάζεται ελάχιστη εμπιστοσύνη (minconf). Αυτοί οι δύο παράγοντες καθορίζουν τον αριθμό των κανόνων που θα προκύψουν. Τα στοιχειοσύνολα (itemsets) που έχουν υποστήριξη μεγαλύτερη από το minsup ονομάζονται συχνά.

Τα στάδια ανάπτυξης των κανόνων συσχετίσεων είναι τα εξής:

- στο πρώτο στάδιο πραγματοποιείται η ανεύρεση όλων των συχνών στοιχειοσυνόλων.
- στο δεύτερο στάδιο εκτελείται η εξαγωγή των κανόνων συσχετίσεων που υπακούν στο όριο του minconf.

Η εφαρμογή του δεύτερου σταδίου ανάπτυξης των κανόνων συσχετίσεων είναι απλή και έχει την εξής διαδικασία: για κάθε ένα από τα συχνά στοιχειοσύνολα l, βρίσκουμε όλα τα μη κενά υποσύνολά του. Για κάθε τέτοιο υποσύνολο a, παρουσιάζουμε τον κανόνα $a \Rightarrow (l - a)$, αν ο λόγος $\sup(l)/\sup(a)$, που αντιστοιχεί στην εμπιστοσύνη του κανόνα, είναι τουλάχιστον minconf (με $\sup(l)$ και $\sup(a)$ συμβολίζουμε το πλήθος των συναλλαγών που περιέχουν τα στοιχειοσύνολα l και a, αντίστοιχα).

Η δυσκολία και η προσοχή των ερευνητών έχει επικεντρωθεί στο πρώτο στάδιο ανάπτυξης των κανόνων συσχετίσεων, όπου έχουν παρουσιαστεί διάφοροι αλγόριθμοι για την επίλυσή του. Ενδεικτικά, αναφέρουμε τους εξής:

- Apriori
- Partition
- FP-Growth
- Eclat

Στην παρούσα εργασία θα ασχοληθούμε μόνο με τον αλγόριθμο Apriori.

Αλγόριθμος Apriori

Ο τρόπος λειτουργίας του αλγορίθμου Apriori έχει ως εξής: σε πρώτη φάση γίνεται το πέρασμα (διάβασμα) του πίνακα *D*. Κατά το πρώτο πέρασμα μετριέται η υποστήριξη των 1-itemsets και υπολογίζεται ποια από αυτά ικανοποιούν την συνθήκη της ελάχιστης υποστήριξης. Σε κάθε επόμενη φάση υπολογίζονται τα καινούρια itemsets που στηρίζονται στα προηγούμενα. Τα itemsets που προκύπτουν ονομάζονται υποψήφια (candidate itemsets), επειδή δεν γνωρίζουμε την υποστήριξή τους και κατά συνέπεια αν είναι συχνά. Αυτός είναι ο λόγος που βρίσκεται η υποστήριξή τους μέσω ενός περάσματος από τον αρχικό πίνακα. Το πλεονέκτημα αυτού του αλγορίθμου είναι ότι σε κάθε φάση γίνεται ένα μόνο πέρασμα από τον αρχικό πίνακα. Η διάκριση για το ποια itemsets είναι συχνά γίνεται στο τέλος, ώστε να χρησιμοποιηθούν στην επόμενη φάση.

5.2. Κατασκευή δέντρου αποφάσεων με τον αλγόριθμο Conditional Inference Decision Trees

Η ερμηνεία ενός output κατασκευής δέντρου αποφάσεων με τον αλγόριθμο Conditional Inference Decision Trees είναι η εξής: ο αριθμός στην αρχή της κάθε γραμμής αντιστοιχεί στον κόμβο του δέντρου. Όσο πιο μικρός είναι ο αριθμός, τόσο πιο ψηλά είναι ο κόμβος στην ιεραρχία του δέντρου. Τον κόμβο ακολουθεί η περιγραφή της διακλάδωσης που ξεκινά από αυτόν. Αν ο κόμβος είναι επισημασμένος με αστερίσκο, τότε είναι φύλλο και θα περιέχει το πλήθος των στοιχείων που περικλείει. Αν δεν έχει σήμανση φύλλου, τότε είναι εσωτερικός κόμβος και θα υπάρχει κι άλλη γραμμή που θα ξεκινάει από τον ίδιο αριθμό. Την πρώτη φορά που αναφέρεται κάθε διακλάδωση θα συνοδεύεται από την μεταβλητή που χρησιμοποιήθηκε για την εξαγωγή του κανόνα, από τον κανόνα και από μια στατιστική αποτύπωση της σημασίας του κόμβου.

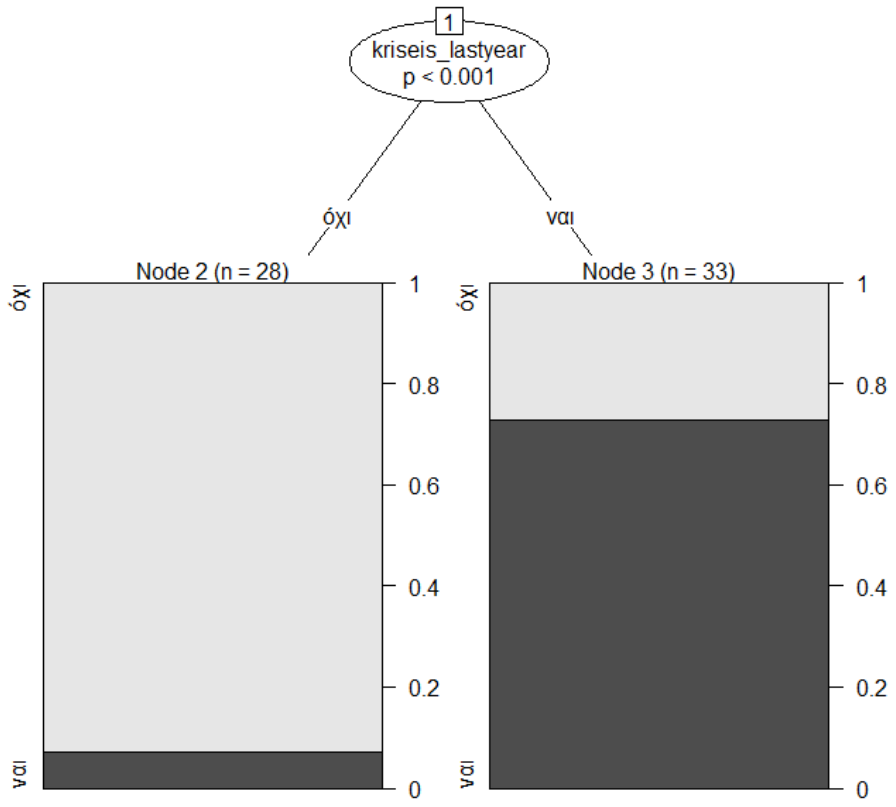
Στον Πίνακα 5.2 βλέπουμε το output κατασκευής δέντρου αποφάσεων με τον αλγόριθμο Conditional Inference Decision Trees για την μεταβλητή απόκρισης kriseis2. Για την κατασκευή του δέντρου χρησιμοποιούμε τα πλήρη δεδομένα και ως επεξηγηματικές μεταβλητές χρησιμοποιούμε αυτές με τις οποίες συσχετίζεται η ύπαρξη κρίσεων το τελευταίο δίμηνο. Επίσης, χρησιμοποιούμε το 70% των δεδομένων για το σύνολο εκμάθησης και το 30% των δεδομένων για το σύνολο ελέγχου.

```
          Conditional inference tree with 2 terminal nodes

Response:  kriseis2
Inputs:  Q17, new_Q18, kriseis_lastyear, Q60
Number of observations:  61

1) kriseis_lastyear == {όχι}; criterion = 1, statistic = 26.205
  2)*  weights = 28
1) kriseis_lastyear == {ναι}
  3)*  weights = 33
```

Πίνακας 5.2: Output κατασκευής δέντρου αποφάσεων με τον αλγόριθμο Conditional Inference Decision Trees

Από τον Πίνακα 5.2 παρατηρούμε ότι το δέντρο που κατασκευάσαμε έχει συνολικά 3 κόμβους, από τους οποίους οι δύο (2 και 3) είναι φύλλα. Για τα δεδομένα μας η βέλτιστη διάκριση απαιτεί μόνο μία μεταβλητή, αυτή που αφορά την ύπαρξη κρίσεων τον περασμένο χρόνο. Για τους επιληπτικούς ασθενείς που δεν εμφάνισαν κρίσεις τον περασμένο χρόνο καταλήγουμε στο φύλλο 2, που περιέχει 28 στοιχεία του αρχικού συνόλου, ενώ για αυτούς που εμφάνισαν κρίσεις τον περασμένο χρόνο καταλήγουμε στο φύλλο 3, που περιέχει 33 στοιχεία του αρχικού συνόλου.

Με το Διάγραμμα 5.2 οπτικοποιούμε το αποτέλεσμα του Πίνακα 5.2.



Διάγραμμα 5.2: Δέντρο αποφάσεων με τον αλγόριθμο Conditional Inference Decision Trees

Από το Διάγραμμα 5.2 παρατηρούμε ότι:

- οι περισσότεροι επιληπτικοί ασθενείς που δεν εμφάνισαν κρίσεις τον περασμένο χρόνο, δεν εμφάνισαν κρίσεις ούτε το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου.
- οι περισσότεροι επιληπτικοί ασθενείς που εμφάνισαν κρίσεις τον περασμένο χρόνο, εμφάνισαν κρίσεις και το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου.

Για να αξιολογήσουμε την αποτελεσματικότητα του δέντρου αποφάσεων με τον αλγόριθμο Conditional Inference Decision Trees, θα κατασκευάσουμε τον πίνακα ταξινόμησης.

	Παρατηρούμενο αποτέλεσμα		
		Εμφάνιση κρίσεων το τελευταίο δίμηνο	*Μη εμφάνιση κρίσεων το τελευταίο δίμηνο*
Εκτιμώμενο αποτέλεσμα	*Εμφάνιση κρίσεων το τελευταίο δίμηνο*	9	4
	Μη εμφάνιση κρίσεων το τελευταίο δίμηνο	1	10

Πίνακας 5.3: Πίνακας ταξινόμησης του δέντρου αποφάσεων με τον αλγόριθμο Conditional Inference Decision Trees

Από τον Πίνακα 5.3 υπολογίζουμε τα παρακάτω μέτρα:

- $ευαισθησία = \frac{9}{9+1} = \frac{9}{10} = 0.9$
- $ειδικότητα = \frac{10}{4+10} = \frac{10}{14} = 0.7143$
- $ακρίβεια = \frac{9+10}{9+4+1+10} = \frac{19}{24} = 0.7917$

Συμπεραίνουμε, λοιπόν, ότι το δέντρο αποφάσεων με τον αλγόριθμο Conditional Inference Decision Trees είναι πολύ αποτελεσματικό για την πρόβλεψη ύπαρξης κρίσεων το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου.

5.3. Εξόρυξη κανόνων συσχετίσεων με τον αλγόριθμο Apriori

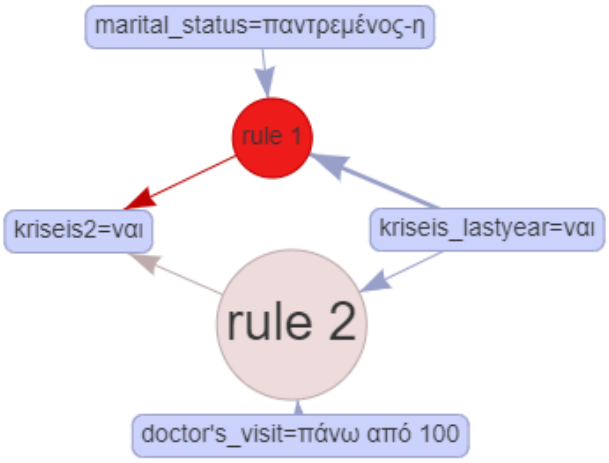
Στην ενότητα αυτή θα εφαρμόσουμε τον αλγόριθμο Apriori προκειμένου να βρούμε τους κανόνες συσχετίσεων με βάση τους οποίους ένας επιληπτικός ασθενής εμφάνισε κρίσεις το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου. Για την εφαρμογή του αλγορίθμου θα χρησιμοποιήσουμε τα πλήρη δεδομένα και ως επεξηγηματικές μεταβλητές θα χρησιμοποιήσουμε αυτές με τις οποίες συσχετίζεται η ύπαρξη κρίσεων το τελευταίο δίμηνο. Επίσης, οι κανόνες συσχετίσεων που θα υπολογίσουμε θα έχουν ελάχιστη υποστήριξη 10% και ελάχιστη εμπιστοσύνη 80% (επιλέγουμε υψηλή εμπιστοσύνη για να προκύψουν ισχυροί κανόνες).

Εφαρμόζοντας τον αλγόριθμο Apriori προκύπτουν δύο κανόνες συσχετίσεων. Πιο συγκεκριμένα, ένας επιληπτικός ασθενής εμφάνισε κρίσεις το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου εάν:

1. εμφάνισε κρίσεις τον περασμένο χρόνο και είναι παντρεμένος.
2. έχει κάνει πάνω από 100 κρίσεις συνολικά στη ζωή του και εμφάνισε κρίσεις τον περασμένο χρόνο.

Η υποστήριξη του κανόνα 1 ισούται με 0.1647 και η εμπιστοσύνη του ισούται με 0.8235, ενώ η υποστήριξη του κανόνα 2 ισούται με 0.2352 και η εμπιστοσύνη του ισούται με 0.80.

Με το Διάγραμμα 5.3 οπτικοποιούμε τους κανόνες συσχετίσεων που προέκυψαν. Ένα τέτοιο διάγραμμα ονομάζεται διάγραμμα δικτύου (network graph).



Διάγραμμα 5.3: Διάγραμμα δικτύου (network graph) για τους κανόνες συσχετίσεων

ΚΕΦΑΛΑΙΟ 6
Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάζουμε τα πιο σημαντικά αποτελέσματα της παρούσας έρευνας. Επίσης, παρουσιάζουμε τις μεταβολές που παρατηρούνται όταν συγκρίνουμε την παρούσα έρευνα με μια αντίστοιχη έρευνα που πραγματοποιήθηκε την προηγούμενη δεκαετία, καθώς και τα αποτελέσματα της σύγκρισης με παρόμοιες έρευνες της Κολομβίας και της Νέας Ζηλανδίας.

6.1. Αποτελέσματα της παρούσας έρευνας

Τα πιο σημαντικά αποτελέσματα που προέκυψαν από την περιγραφική ανάλυση των μεταβλητών της παρούσας έρευνας είναι τα εξής:

- η πλειοψηφία των ερωτηθέντων δεν επαναλαμβάνουν κρίσεις ίδιου τύπου (ποσοστό 76.84%), δεν είχαν καμία κρίση το τελευταίο δίμηνο (ποσοστό 60.22%), έχουν κάνει συνολικά πάνω από 100 κρίσεις στη ζωή τους (ποσοστό 39.78%), δεν νοσηλεύτηκαν λόγω των κρίσεων τον τελευταίο χρόνο (ποσοστό 88.54%), δεν έχουν ιστορικό επιληψίας στο οικογενειακό περιβάλλον (ποσοστό 88.42%), δεν είχαν καμία κρίση τον περασμένο χρόνο (ποσοστό 44.21%), δεν πάσχουν από κάποιο άλλο νόσημα εκτός από τις κρίσεις (ποσοστό 69.47%), δεν μπορούν να ελέγξουν τις κρίσεις (ποσοστό 84.04%), επισκέπτονται τον γιατρό τους μία φορά το εξάμηνο ως μία φορά το χρόνο (ποσοστό 33.68%) και χρειάζονται περισσότερη ενημέρωση για την νόσο τους (ποσοστό 66.32%).
- η πλειοψηφία των ερωτηθέντων απάντησαν ότι η λέξη «επιληπτικός» κάνει τους άλλους να τους βλέπουν λίγο με προκατάληψη (ποσοστό 40.91%), ότι η επιληψία δεν τους εμποδίζει καθόλου στην ανεύρεση εργασίας (ποσοστό 47.67%), ότι δεν διέκοψαν τις σπουδές τους εξαιτίας της επιληψίας (ποσοστό 88.04%), ότι δεν αισθάνονται καθόλου μοναξιά (ποσοστό 44.68%), ότι νιώθουν λίγη ανασφάλεια για το μέλλον εξαιτίας των κρίσεων (ποσοστό 38.95%), ότι επιδιώκουν να αποκτήσουν νέους φίλους (ποσοστό 62.11%) και ότι η επιληψία δεν επηρεάζει καθόλου τις σχέσεις τους με το άλλο φύλο (ποσοστό 52.17%).
- η πλειονότητα των ερωτηθέντων είναι πλήρως ή μερικώς απασχολούμενοι (ποσοστό 47.87%), είναι απόφοιτοι δευτεροβάθμιας εκπαίδευσης (ποσοστό 45.83%), είναι άγαμοι (ποσοστό 64.89%), δεν έχουν παιδιά (ποσοστό 68.42%) και δεν οδηγούν (ποσοστό 68.75%).
- η πλειοψηφία των ερωτηθέντων σπάνια πίνουν κρασί (ποσοστό 47.37%), δεν πίνουν ποτέ ούζο, ουίσκι, βότκα, τζιν ή κονιάκ (ποσοστό 53.68%) και βγαίνουν έξω τα βράδια λιγότερο από μία φορά το δίμηνο (ποσοστό 35.42%).

Τα κυριότερα αποτελέσματα που προέκυψαν από την διερεύνηση της σχέσης μεταξύ της ύπαρξης κρίσεων το τελευταίο δίμηνο και των υπόλοιπων μεταβλητών της έρευνας είναι τα εξής:

- η ύπαρξη κρίσεων το τελευταίο δίμηνο συσχετίζεται, σε επίπεδο σημαντικότητας α=5%, με τις εξής μεταβλητές:
  - συνολικός αριθμός κρίσεων
  - αριθμός επισκέψεων στον γιατρό τον τελευταίο χρόνο
  - ύπαρξη κρίσεων τον περασμένο χρόνο
  - οικογενειακή κατάσταση
- είναι πιο πιθανό για κάποιον που εμφάνισε κρίσεις το τελευταίο δίμηνο να είχε εμφανίσει κρίσεις τον περασμένο χρόνο σε σχέση με κάποιον που δεν εμφάνισε κρίσεις το τελευταίο δίμηνο.
- είναι πιο πιθανό για κάποιον που εμφάνισε κρίσεις το τελευταίο δίμηνο να είναι παντρεμένος σε σχέση με κάποιον που δεν εμφάνισε κρίσεις το τελευταίο δίμηνο.

Τα σπουδαιότερα αποτελέσματα που προέκυψαν όταν προσαρμόσαμε ένα μοντέλο δίτιμης λογιστικής παλινδρόμησης με εξαρτημένη μεταβλητή την ύπαρξη κρίσεων το τελευταίο δίμηνο και έχοντας αφαιρέσει τις ελλιπείς τιμές είναι τα εξής:

- ο συνολικός αριθμός κρίσεων, η ύπαρξη κρίσεων τον περασμένο χρόνο και η οικογενειακή κατάσταση επηρεάζουν την πιθανότητα εμφάνισης κρίσεων μέσα στο επόμενο δίμηνο σε επίπεδο σημαντικότητας α=5%.
- όταν ένα άτομο έχει κάνει πάνω από 100 κρίσεις συνολικά στη ζωή του, εμφάνισε κρίσεις τον περασμένο χρόνο και είναι παντρεμένο, τότε είναι σχεδόν βέβαιη η εμφάνιση κρίσεων μέσα στο επόμενο δίμηνο.

Τα ίδια αποτελέσματα προέκυψαν και όταν προσαρμόσαμε ένα μοντέλο δίτιμης λογιστικής παλινδρόμησης έχοντας εφαρμόσει την διαδικασία συμπλήρωσης των ελλιπών τιμών (imputation).

Τα βασικότερα αποτελέσματα που προέκυψαν όταν κατασκευάσαμε δέντρο απόφασης με τον αλγόριθμο Conditional Inference Decision Trees για την μεταβλητή που αφορά την ύπαρξη κρίσεων το τελευταίο δίμηνο είναι τα εξής:

- η βέλτιστη διάκριση για τα δεδομένα μας με βάση την οποία μπορεί να προβλεφθεί η εμφάνιση κρίσεων μέσα στο επόμενο δίμηνο απαιτεί μόνο μία μεταβλητή, αυτή που αφορά την ύπαρξη κρίσεων τον περασμένο χρόνο.
- οι περισσότεροι επιληπτικοί ασθενείς που δεν εμφάνισαν κρίσεις τον περασμένο χρόνο, δεν εμφάνισαν κρίσεις ούτε το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου.

- οι περισσότεροι επιληπτικοί ασθενείς που εμφάνισαν κρίσεις τον περασμένο χρόνο, εμφάνισαν κρίσεις και το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου.

Τέλος, εφαρμόσαμε τον αλγόριθμο Apriori προκειμένου να βρούμε τους κανόνες συσχετίσεων με βάση τους οποίους ένας επιληπτικός ασθενής εμφάνισε κρίσεις το τελευταίο δίμηνο και προέκυψαν δύο κανόνες. Πιο συγκεκριμένα, ένας επιληπτικός ασθενής εμφάνισε κρίσεις το τελευταίο δίμηνο πριν από την ημερομηνία συμπλήρωσης του ερωτηματολογίου εάν:

1. εμφάνισε κρίσεις τον περασμένο χρόνο και είναι παντρεμένος.
2. έχει κάνει πάνω από 100 κρίσεις συνολικά στη ζωή του και εμφάνισε κρίσεις τον περασμένο χρόνο.

6.2. Μεταβολές την τελευταία δεκαετία

Οι μεταβολές που παρατηρούνται όταν συγκρίνουμε την παρούσα έρευνα με μια αντίστοιχη έρευνα (Νικολάκης, 2010) που πραγματοποιήθηκε την περίοδο 2008-2009 και οι οποίες είναι στατιστικά σημαντικές είναι οι εξής:

- η αύξηση του ποσοστού των συμμετεχόντων που πάσχουν από κάποιο άλλο νόσημα εκτός από τις κρίσεις κατά 8.9%.
- η αύξηση του ποσοστού των συμμετεχόντων που δεν μπορούν να εμποδίσουν την κρίση κατά 18.1%.
- η αύξηση του ποσοστού των συμμετεχόντων που δεν τους ενοχλεί καθόλου να τους αποκαλούν «επιληπτικούς» κατά 16%.
- η μείωση του ποσοστού των συμμετεχόντων που τους ενοχλεί πολύ να τους αποκαλούν «επιληπτικούς» κατά 14%.
- η μείωση του ποσοστού των συμμετεχόντων που είναι άγαμοι κατά 15.3%.
- η αύξηση του ποσοστού των συμμετεχόντων που είναι παντρεμένοι κατά 15.4%.
- η μείωση του ποσοστού των συμμετεχόντων που δεν πίνουν ποτέ κρασί κατά 18.8%.
- η αύξηση του ποσοστού των συμμετεχόντων που σπάνια πίνουν κρασί κατά 16.6%.
- η μείωση του ποσοστού των συμμετεχόντων που επισκέπτονται τον γιατρό τους μία φορά κάθε τρεις ως έξι μήνες κατά 13.8%.
- η αύξηση του ποσοστού των συμμετεχόντων που έχουν πάνω από ένα χρόνο να επισκεφθούν τον γιατρό τους κατά 13.5%.
- η μείωση του ποσοστού των συμμετεχόντων που φοβούνται λίγο τις κρίσεις κατά 13.9%.

6.3. Σύγκριση της παρούσας έρευνας με παρόμοια έρευνα της Κολομβίας

Τα αποτελέσματα που προκύπτουν από την σύγκριση της παρούσας έρευνας με την έρευνα της Κολομβίας είναι τα εξής:

- δεν υπάρχει μεγάλη διαφορά ως προς την μέση ηλικία των συμμετεχόντων στις δύο έρευνες.
- η διάμεση ηλικία πρώτης κρίσης για τους ασθενείς της Ελλάδας είναι μεγαλύτερη κατά 4 έτη από την διάμεση ηλικία πρώτης κρίσης για τους ασθενείς της Κολομβίας.
- οι περισσότεροι Έλληνες ασθενείς εργάζονται, ενώ οι περισσότεροι Κολομβιανοί ασθενείς δεν εργάζονται.
- οι περισσότεροι Έλληνες ασθενείς ολοκλήρωσαν την δευτεροβάθμια εκπαίδευση, ενώ οι περισσότεροι Κολομβιανοί ασθενείς δεν την ολοκλήρωσαν.

6.4. Σύγκριση της παρούσας έρευνας με παρόμοια έρευνα της Νέας Ζηλανδίας

Τα αποτελέσματα που προκύπτουν από την σύγκριση της παρούσας έρευνας με την έρευνα της Νέας Ζηλανδίας είναι τα εξής:

- στην έρευνα της Ελλάδας συμμετείχαν περισσότεροι άνδρες, ενώ στην έρευνα της Νέας Ζηλανδίας συμμετείχαν περισσότερες γυναίκες.
- οι περισσότεροι Έλληνες ασθενείς δεν μπορούν να ελέγξουν τις κρίσεις τους, ενώ οι περισσότεροι Νεοζηλανδοί ασθενείς μπορούν.
- το ποσοστό των εργαζόμενων ασθενών στην Ελλάδα είναι μεγαλύτερο σε σχέση με το αντίστοιχο ποσοστό στη Νέα Ζηλανδία, ενώ το ποσοστό των μη εργαζόμενων ασθενών στην Ελλάδα είναι μικρότερο σε σχέση με το αντίστοιχο ποσοστό στη Νέα Ζηλανδία.

ΠΑΡΑΡΤΗΜΑΤΑ

Π1 ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ
Π2 ΕΝΤΟΛΕΣ ΓΙΑ ΤΗΝ R
Π3 ΠΙΝΑΚΕΣ

Π1 ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ

1. Φύλο (υπογραμμίστε το σωστό). Άνδρας Γυναίκα

2. Ηλικία …… ……

7. Πού ζήσατε μέχρι τα 18 σας χρόνια;

1. σε χωριό μέχρι 3.000 κατοίκους
2. σε κωμόπολη μέχρι 10.000 κατοίκους
3. σε πόλη με πάνω από 10.000 κατοίκους
4. σε Αθήνα ή Θεσ/νίκη
5. αλλού (γράψτε που)

8. Πού ζείτε τώρα μόνιμα;

1. σε χωριό μέχρι 3.000 κατοίκους
2. σε κωμόπολη μέχρι 10.000 κατοίκους
3. σε πόλη με πάνω από 10.000 κατοίκους
4. σε Αθήνα ή Θεσ/νίκη
5. αλλού (γράψτε που)

11. Σε ποια ηλικία είχατε την πρώτη κρίση;

……..…..

14. Μέχρι σήμερα κάνετε τις ίδιες κρίσεις;

1. όχι 2. ναι

16. Πόσες κρίσεις κάνατε το τελευταίο δίμηνο;

17. Πόσες κρίσεις έχετε κάνει συνολικά στη ζωή σας;

1
2-5
6-10
πάνω από 21
πάνω από 100

18. Πόσες φορές πήγατε στο γιατρό τον τελευταίο χρόνο;

19. Νοσηλευθήκατε τον τελευταίο χρόνο λόγω των κρίσεων;

1. όχι 2. ναι

20. Πάσχει κάποιος άλλος στην οικογένειά σας από επιληψία;

1. όχι 2. ναι

23. Πόσο συχνά είχατε κρίσεις τον περασμένο χρόνο;

1. καμία
2. 1 το χρόνο
3. 1-2 το εξάμηνο
4. 1-2 το δίμηνο
5. 1-2 το μήνα
6. 1 την εβδομάδα
7. πάνω από 1 την εβδομάδα

25. Πάσχετε από κάποιο άλλο νόσημα εκτός από τις κρίσεις;

α) όχι

β) ναι ποιο; 1)……..

2)……..

30. Μπορείτε να εμποδίσετε την κρίση, κι αν ναι με ποιο τρόπο;

1. όχι 2. ναι με ποιο τρόπο;

……..……

……..……

31. Σας ενοχλεί να σας αποκαλούν «επιληπτικό»;

1. καθόλου 2. λίγο 3. αρκετά 4. πολύ

33. Η λέξη «επιληπτικός» κάνει τους άλλους να σας βλέπουν με προκατάληψη;

1. καθόλου 2. λίγο 3. συχνά 4. πάντα

55. Ποια από τις παρακάτω προτάσεις ταιριάζει περισσότερο στις εργασιακές σας σχέσεις;

1. πλήρως απασχολούμενος
2. μερικώς απασχολούμενος
3. άνεργος, ψάχνω για δουλειά
4. άνεργος, δεν ψάχνω για δουλειά
5. συνταξιούχος
6. ασθενής
7. οικιακά
8. σπουδαστής
9. άλλο (γράψτε τι ……..)

56. Σας εμποδίζει η νόσος σας στην ανεύρεση εργασίας;

1. καθόλου 3. αρκετά
2. λίγο 4. πάρα πολύ
5. δεν ξέρω

58. Τι γραμματικές γνώσεις έχετε;

1. έως τρίτη δημοτικού
2. έως έκτη δημοτικού
3. έως τρίτη γυμνασίου
4. έως έκτη γυμνασίου ή τρίτη λυκείου
5. πτυχίο ανωτέρας ή ανωτάτης σχολής

59. Διακόψατε τις σπουδές σας εξ αιτίας των προβλημάτων υγείας που έχετε;

1. όχι 2. ναι 3. δεν ξέρω

60. Είσθε: 1. παντρεμένος-η, 2. ελεύθερος-η, 3. χωρισμένος-η

61. Έχετε παιδιά; 1. όχι 2. ναι πόσα; ……

63. Πίνετε κρασί;

1. ποτέ
2. σπάνια
3. μια-δυο φορές το μήνα
4. μια φορά την εβδομάδα
5. δυο-τρεις φορές την εβδομάδα
6. κάθε μέρα

γράψτε πόσα ποτήρια πίνετε κάθε φορά ……..…....

64. Πίνετε ούζο, ουίσκι, βότκα, τζιν, κονιάκ;

1. ποτέ
2. σπάνια
3. μια-δυο φορές το μήνα
4. μια φορά την εβδομάδα
5. δυο-τρεις φορές την εβδομάδα
6. κάθε μέρα

γράψτε πόσα ποτά πίνετε κάθε φορά ……..……..

65. Πόσο συχνά βγαίνετε έξω τα βράδια;

1. Λιγότερο από μία φορά το δίμηνο
2. Μία φορά το δίμηνο
3. Μία φορά το μήνα
4. 2-3 φορές το μήνα
5. Μια φορά την εβδομάδα
6. 2-3 φορές την εβδομάδα
7. Κάθε μέρα

72. Έχετε δίπλωμα οδήγησης; 1. όχι 2. ναι

73. Οδηγείτε; 1. όχι 2. ναι

74. Αισθάνεστε μοναξιά; 1. καθόλου 2. λίγο 3. αρκετά

76. Πόσο συχνά πηγαίνετε στο γιατρό;

1. έχω πάνω από ένα χρόνο να πάω
2. μια φορά το χρόνο
3. μια φορά το εξάμηνο ως μια φορά το χρόνο
4. μια φορά κάθε τρεις ως έξι μήνες
5. μια φορά κάθε ένα ως τρεις μήνες
6. πιο συχνά από μια φορά το μήνα

80. Χρειάζεστε περισσότερη ενημέρωση για τη νόσο σας;

1. όχι 2. ναι 3. δεν ξέρω

82. Φοβάστε τις κρίσεις;

1. όχι 2. λίγο 3. αρκετά 4. πολύ 5. δεν ξέρω

86. Νιώθετε ανασφάλεια για το μέλλον εξ αιτίας των κρίσεων;

1. καθόλου 2. λίγο 3. αρκετά 4. πολύ 5. δεν ξέρω

88. Επιδιώκετε ν' αποκτήσετε καινούργιους φίλους;

1. ναι 2. όχι 3. δεν ξέρω

91. Επηρεάζει τις σχέσεις σας με το άλλο φύλο το πρόβλημα υγείας που έχετε;

1. όχι καθόλου 2. λίγο 3. αρκετά 4. πολύ 5. δεν ξέρω

Π2 ΕΝΤΟΛΕΣ ΓΙΑ ΤΗΝ R

```
##################### εντολές κεφαλαίου 2 ##############################
#######################################################################
library(readxl)
a=data.frame(read_excel("dataset.xlsx"))
a$Q1=as.factor(a$Q1)
a$Q2=as.numeric(a$Q2)
a$Q7=as.factor(a$Q7)
a$Q8=as.factor(a$Q8)
a$Q11=as.numeric(a$Q11)
a$Q14=as.factor(a$Q14)
a$Q16=as.numeric(a$Q16)
a$Q17=as.factor(a$Q17)
a$Q18=as.numeric(a$Q18)
a$Q19=as.factor(a$Q19)
a$Q20=as.factor(a$Q20)
a$Q23=as.factor(a$Q23)
a$Q25=as.factor(a$Q25)
a$Q30=as.factor(a$Q30)
a$Q31=as.factor(a$Q31)
a$Q33=as.factor(a$Q33)
a$Q55=as.factor(a$Q55)
a$Q56=as.factor(a$Q56)
a$Q58=as.factor(a$Q58)
a$Q59=as.factor(a$Q59)
a$Q60=as.factor(a$Q60)
a$Q61=as.factor(a$Q61)
a$Q63=as.factor(a$Q63)
a$Q64=as.factor(a$Q64)
a$Q65=as.factor(a$Q65)
a$Q72=as.factor(a$Q72)
a$Q73=as.factor(a$Q73)
a$Q74=as.factor(a$Q74)
a$Q76=as.factor(a$Q76)
a$Q80=as.factor(a$Q80)
a$Q82=as.factor(a$Q82)
```

```
a$Q86=as.factor(a$Q86)
a$Q88=as.factor(a$Q88)
a$Q91=as.factor(a$Q91)

Q2=na.omit(a$Q2)
library(DescTools)
new_Q2=CutQ(Q2,breaks=quantile(Q2,seq(0,1,by=0.25),na.rm=TRUE),
      labels=NULL,na.rm=FALSE)

b=round((table(a$Q7)/95)*100,2)
b=paste(b,"%",sep="")
pie(table(a$Q7),labels=b,main ="Μόνιμη κατοικία μέχρι την ηλικία των 18 ετών",
      col = rainbow(length(table(a$Q7))))
legend(locator(1), c("χωριό μέχρι 3000 κατοίκους","κωμόπολη μέχρι
      10000 κατοίκους","πόλη με πάνω από 10000 κατοίκους","Αθήνα ή
      Θεσσαλονίκη","αλλού"), cex = 0.8,fill = rainbow(length(table(a$Q7))))

c=round((table(a$Q8)/96)*100,2)
c=paste(c,"%",sep="")
pie(table(a$Q8),labels=c,main ="Μόνιμη τωρινή κατοικία ",
      col = c("aquamarine","chocolate1","mediumorchid1","gold","royalblue1"))
legend(locator(1), c("χωριό μέχρι 3000 κατοίκους","κωμόπολη μέχρι
      10000 κατοίκους","πόλη με πάνω από 10000 κατοίκους","Αθήνα ή
      Θεσσαλονίκη","αλλού"),cex = 0.8,fill = c("aquamarine",
      "chocolate1","mediumorchid1","gold","royalblue1"))

library(DescTools)
new_Q11=CutQ(a$Q11,breaks=quantile(a$Q11,seq(0,1,by=0.25),na.rm=TRUE),
      labels=NULL,na.rm=FALSE)

Q16=na.omit(a$Q16)
library(DescTools)
new_Q16=CutQ(Q16,breaks=quantile(Q16,seq(0,1,by=0.25),na.rm=TRUE),
      labels=NULL,na.rm=FALSE)
```

```
d=round((table(a$Q17)/93)*100,2)
d=paste(d,"%",sep="")
pie(table(a$Q17),labels=d,main ="Συνολικός αριθμός κρίσεων",
        col = c("cadetblue1","chartreuse1","darkgoldenrod1","firebrick1",
        "darkgreen"))
legend(locator(1), c("1","2-5","6-10","πάνω από 21","πάνω από 100"),
        cex = 0.8, fill = c("cadetblue1","chartreuse1","darkgoldenrod1",
        "firebrick1","darkgreen"))

Q18=na.omit(a$Q18)
library(DescTools)
new_Q18=CutQ(Q18,breaks=quantile(Q18,seq(0,1,by=0.25),na.rm=TRUE),
        labels=NULL,na.rm=FALSE)

d=round((table(a$Q23)/95)*100,2)
d=paste(d,"%",sep="")
pie(table(a$Q23),labels=d,main ="Συχνότητα κρίσεων τον περασμένο χρόνο",
        col = c("blue1","blueviolet","brown1","darkgoldenrod1","darkgreen",
        "burlywood", "cyan"))
legend(locator(1), c("καμία","1 το χρόνο","1-2 το εξάμηνο","1-2 το δίμηνο",
        "1-2 το μήνα","1 την εβδομάδα","πάνω από 1 την εβδομάδα"),
        cex = 0.8, fill = c("blue1","blueviolet","brown1","darkgoldenrod1",
        "darkgreen", "burlywood","cyan"))

d=round((table(a$Q31)/94)*100,2)
d=paste(d,"%",sep="")
pie(table(a$Q31),labels=d,main ="Βαθμός αποδοχής της επιληψίας",
        col = c("blue","red","green","yellow"))
legend(locator(1), c("καθόλου","λίγο","αρκετά","πολύ"),cex = 0.8,
        fill = c("blue","red","green","yellow"))

d=round((table(a$Q33)/88)*100,2)
d=paste(d,"%",sep="")
pie(table(a$Q33),labels=d,main ="Κοινωνική αντιμετώπιση",
        col = c("olivedrab1","orange","plum2","royalblue1"))
legend(locator(1), c("καθόλου","λίγο","συχνά","πάντα"),cex = 0.8,
        fill = c("olivedrab1","orange","plum2","royalblue1"))
```

```
d=round((table(a$Q55)/94)*100,2)
d=paste(d,"%",sep="")
pie(table(a$Q55),labels=d,main ="Επαγγελματική κατάσταση",
      col = c("antiquewhite","darkgoldenrod","blue4","blueviolet",
      "brown1","chartreuse4","cadetblue1","cyan","yellow"))
      legend(locator(1), c("πλήρως απασχολούμενος","μερικώς
      απασχολούμενος","άνεργος,ψάχνω για δουλειά","άνεργος,δεν ψάχνω για
      δουλειά","συνταξιούχος","ασθενής","οικιακά","σπουδαστής","άλλο"),cex =
      0.8,fill=c("antiquewhite","darkgoldenrod","blue4","blueviolet","brown1",
      "chartreuse4","cadetblue1","cyan","yellow"))

Απάντηση=c("καθόλου","λίγο","αρκετά","πάρα πολύ","δεν ξέρω")
e=round(table(a$Q56)/86*100,2)
a1=data.frame(Απάντηση,e)
library(ggplot2)
a1$Απάντηση=factor(a1$Απάντηση,levels = a1$Απάντηση)
p=ggplot(data=a1, aes(x=Απάντηση, y=e,fill=Απάντηση)) +
  geom_bar(stat="identity",width=0.5)+
  geom_text(aes(label=e), vjust=-0.3, size=3.5)+
  ggtitle("Βαθμός δυσκολίας ανεύρεσης εργασίας")+
  ylab("Ποσοστό")+
  theme_minimal()
p+theme(axis.title.x = element_blank())
p + theme(
plot.title = element_text(size=14),
axis.title.y = element_text(size=12),
axis.title.x = element_blank())

Απάντηση=c("6η δημοτικού","3η γυμνασίου","6η γυμνασίου ή 3η λυκείου","πτυχίο
      ανωτέρας ή ανωτάτης σχολής")
e=round(table(a$Q58)/96*100,2)
a1=data.frame(Απάντηση,e)
library(ggplot2)
a1$Απάντηση=factor(a1$Απάντηση,levels = a1$Απάντηση)
p=ggplot(data=a1, aes(x=Απάντηση, y=e)) +
  geom_bar(stat="identity",width=0.5,fill="steelblue")+
  geom_text(aes(label=e), vjust=-0.3, size=3.5)+
```

```
  ggtitle("Μορφωτικό επίπεδο")+
  ylab("Ποσοστό")+
  theme_minimal()
p+theme(axis.title.x = element_blank())
p + theme(
plot.title = element_text(size=14),
axis.title.y = element_text(size=12),
axis.title.x = element_blank())

Απάντηση=c("ποτέ","σπάνια","1-2 φορές το μήνα","1 φορά την εβδομάδα","2-3
        φορές την εβδομάδα")
e=round(table(a$Q63)/95*100,2)
a1=data.frame(Απάντηση,e)
library(ggplot2)
a1$Απάντηση=factor(a1$Απάντηση,levels = a1$Απάντηση)
p=ggplot(data=a1, aes(x=Απάντηση, y=e)) +
  geom_bar(stat="identity",width=0.5,fill="green")+
  geom_text(aes(label=e), vjust=-0.3, size=3.5)+
  ggtitle("Κατανάλωση κρασιού")+
  ylab("Ποσοστό")+
  theme_minimal()
p+theme(axis.title.x = element_blank())
p + theme(
plot.title = element_text(size=14),
axis.title.y = element_text(size=12),
axis.title.x = element_blank())

Απάντηση=c("ποτέ","σπάνια","1-2 φορές το μήνα","1 φορά την εβδομάδα","2-3
        φορές την εβδομάδα")
e=round(table(a$Q64)/95*100,2)
a1=data.frame(Απάντηση,e)
library(ggplot2)
a1$Απάντηση=factor(a1$Απάντηση,levels = a1$Απάντηση)
p=ggplot(data=a1, aes(x=Απάντηση, y=e)) +
  geom_bar(stat="identity",width=0.5,fill="darkgoldenrod")+
  geom_text(aes(label=e), vjust=-0.3, size=3.5)+
  ggtitle("Κατανάλωση ούζου, ουίσκι, βότκας, τζιν, κονιάκ")+
```

```
  ylab("Ποσοστό")+
  theme_minimal()
p+theme(axis.title.x = element_blank())
p + theme(
plot.title = element_text(size=14),
axis.title.y = element_text(size=12),
axis.title.x = element_blank())

Απάντηση=c("<1 φορά το δίμηνο","1 φορά το δίμηνο","1 φορά το μήνα","2-3 φορές
        το μήνα","1 φορά την εβδομάδα","2-3 φορές την εβδομάδα","κάθε μέρα")
e=round(table(a$Q65)/96*100,2)
a1=data.frame(Απάντηση,e)
library(ggplot2)
a1$Απάντηση=factor(a1$Απάντηση,levels = a1$Απάντηση)
p=ggplot(data=a1, aes(x=Απάντηση, y=e)) +
  geom_bar(stat="identity",width=0.5,fill="mediumorchid1")+
  geom_text(aes(label=e), vjust=-0.3, size=3.5)+
  ggtitle("Συχνότητα νυχτερινών εξόδων")+
  ylab("Ποσοστό")+
  theme_minimal()
p+theme(axis.title.x = element_blank())
p + theme(
plot.title = element_text(size=14),
axis.title.y = element_text(size=12),
axis.title.x = element_blank())

d=round((table(a$Q74)/94)*100,2)
d=paste(d,"%",sep="")
pie(table(a$Q74),labels=d,main ="Βαθμός μοναξιάς",
col = c("gold","dodgerblue","firebrick1"))
legend(locator(1), c("καθόλου","λίγο","αρκετά"),cex = 0.8,
fill = c("gold","dodgerblue","firebrick1"))

Απάντηση=c("όχι","ναι","δεν ξέρω")
e=round(table(a$Q80)/95*100,2)
a1=data.frame(Απάντηση,e)
library(ggplot2)
```

```
a1$Απάντηση=factor(a1$Απάντηση,levels = a1$Απάντηση)
p=ggplot(data=a1, aes(x=Απάντηση, y=e)) +
  geom_bar(stat="identity",width=0.5,fill="slateblue1")+
  geom_text(aes(label=e), vjust=-0.3, size=3.5)+
  ggtitle("Ανάγκη για περισσότερη ενημέρωση για την νόσο")+
  ylab("Ποσοστό")+
  theme_minimal()
p+theme(axis.title.x = element_blank())
p + theme(
plot.title = element_text(size=14),
axis.title.y = element_text(size=12),
axis.title.x = element_blank())

Απάντηση=c("όχι","λίγο","αρκετά","πολύ","δεν ξέρω")
e=round(table(a$Q82)/94*100,2)
a1=data.frame(Απάντηση,e)
library(ggplot2)
a1$Απάντηση=factor(a1$Απάντηση,levels = a1$Απάντηση)
p=ggplot(data=a1, aes(x=Απάντηση, y=e)) +
  geom_bar(stat="identity",width=0.5,fill="red2")+
  geom_text(aes(label=e), vjust=-0.3, size=3.5)+
  ggtitle("Βαθμός φόβου των κρίσεων")+
  ylab("Ποσοστό")+
  theme_minimal()
p+theme(axis.title.x = element_blank())
p + theme(
plot.title = element_text(size=14),
axis.title.y = element_text(size=12),
axis.title.x = element_blank())

Απάντηση=c("όχι καθόλου","λίγο","αρκετά","πολύ","δεν ξέρω")
e=round(table(a$Q91)/92*100,2)
a1=data.frame(Απάντηση,e)
library(ggplot2)
a1$Απάντηση=factor(a1$Απάντηση,levels = a1$Απάντηση)
p=ggplot(data=a1, aes(x=Απάντηση, y=e)) +
  geom_bar(stat="identity",width=0.5,fill="chartreuse4")+
```

```
  geom_text(aes(label=e), vjust=-0.3, size=3.5)+
  ggtitle("Βαθμός επίδρασης της νόσου στις σχέσεις με το άλλο φύλο")+
  ylab("Ποσοστό")+
  theme_minimal()
p+theme(axis.title.x = element_blank())
p + theme(
plot.title = element_text(size=14),
axis.title.y = element_text(size=12),
axis.title.x = element_blank())

b=read.table("age_nik.txt")
b1=read.table("age_kam.txt")
age=b$V1
age1=b1$V1
library(nortest)
lillie.test(age)
lillie.test(age1)
library(lawstat)
V=c(age,age1)
G=c(rep(1,times=length(age)),rep(2,times=length(age1)))
levene.test(V,G)
t.test(age,age1,var.equal=F)

c=read.table("first_nik.txt")
c1=read.table("first_kam.txt")
first=c$V1
first1=c1$V1
library(nortest)
lillie.test(first)
lillie.test(first1)
n_1=length(first)
n_2=length(first1)
z_stat=(mean(first) - mean(first1)) / sqrt(var(first) / n_1 + var(first1) / n_2)
z_stat
library(distributions3)
Z <- Normal(0, 1)
1 - cdf(Z, 3.012741) + cdf(Z, -3.012741)
```

```
d=read.table("kriseis_nik.txt")
d1=read.table("kriseis_kam.txt")
kriseis=d$V1
kriseis1=d1$V1
library(nortest)
lillie.test(kriseis)
lillie.test(kriseis1)
n_1=length(kriseis)
n_2=length(kriseis1)
z_stat=(mean(kriseis) - mean(kriseis1)) / sqrt(var(kriseis) / n_1 + var(kriseis1) / n_2)
z_stat
library(distributions3)
Z <- Normal(0, 1)
1 - cdf(Z, 1.03846) + cdf(Z, -1.03846)

x=c(29,42)
n=c(89,95)
prop.test(x,n)

x=c(11,5)
n=c(89,95)
prop.test(x,n)

x=c(19,66)
n=c(88,95)
prop.test(x,n)

x=c(58,79)
n=c(88,94)
prop.test(x,n)

x=c(25,42)
n=c(87,94)
prop.test(x,n)
```

```
x=c(25,14)
n=c(87,94)
prop.test(x,n)

x=c(36,41)
n=c(85,86)
prop.test(x,n)

x=c(9,4)
n=c(85,86)
prop.test(x,n)

x=c(73,61)
n=c(91,94)
prop.test(x,n)

x=c(15,30)
n=c(91,94)
prop.test(x,n)

x=c(44,28)
n=c(91,95)
prop.test(x,n)

x=c(28,45)
n=c(91,95)
prop.test(x,n)

x=c(20,34)
n=c(91,96)
prop.test(x,n)

x=c(19,12)
n=c(91,96)
prop.test(x,n)
```

```
x=c(27,15)
n=c(91,95)
prop.test(x,n)
x=c(19,32)
n=c(91,95)
prop.test(x,n)

x=c(4,17)
n=c(91,95)
prop.test(x,n)

x=c(16,10)
n=c(91,95)
prop.test(x,n)

x=c(15,25)
n=c(91,94)
prop.test(x,n)

x=c(30,18)
n=c(91,94)
prop.test(x,n)

x=c(16,26)
n=c(91,95)
prop.test(x,n)

x=c(24,20)
n=c(91,95)
prop.test(x,n)

x=c(15,11)
n=c(91,95)
prop.test(x,n)
```

```
x=c(65,59)
n=c(91,95)
prop.test(x,n)

x=c(20,32)
n=c(91,95)
prop.test(x,n)

x=c(40,48)
n=c(91,92)
prop.test(x,n)

x=c(16,11)
n=c(91,92)
prop.test(x,n)


##################### εντολές κεφαλαίου 4 ##############################
########################################################################
library(readxl)
a=data.frame(read_excel("dataset1.xlsx"))
a$Q1=as.factor(a$Q1)
a$Q2=as.numeric(a$Q2)
a$Q7=as.factor(a$Q7)
a$Q8=as.factor(a$Q8)
a$Q11=as.numeric(a$Q11)
a$Q14=as.factor(a$Q14)
a$Q16=as.numeric(a$Q16)
a$Q17=as.factor(a$Q17)
a$Q18=as.numeric(a$Q18)
a$Q19=as.factor(a$Q19)
a$Q20=as.factor(a$Q20)
a$Q23=as.factor(a$Q23)
a$Q25=as.factor(a$Q25)
a$Q30=as.factor(a$Q30)
a$Q31=as.factor(a$Q31)
a$Q33=as.factor(a$Q33)
```

```
a$Q55=as.factor(a$Q55)
a$Q56=as.factor(a$Q56)
a$Q58=as.factor(a$Q58)
a$Q59=as.factor(a$Q59)
a$Q60=as.factor(a$Q60)
a$Q61=as.factor(a$Q61)
a$Q63=as.factor(a$Q63)
a$Q64=as.factor(a$Q64)
a$Q65=as.factor(a$Q65)
a$Q72=as.factor(a$Q72)
a$Q73=as.factor(a$Q73)
a$Q74=as.factor(a$Q74)
a$Q76=as.factor(a$Q76)
a$Q80=as.factor(a$Q80)
a$Q82=as.factor(a$Q82)
a$Q86=as.factor(a$Q86)
a$Q88=as.factor(a$Q88)
a$Q91=as.factor(a$Q91)
a$new_Q2=as.factor(a$new_Q2)
a$new_Q11=as.factor(a$new_Q11)
a$new_Q18=as.factor(a$new_Q18)
a$kriseis2=as.factor(a$kriseis2)
a$kriseis_lastyear=as.factor(a$kriseis_lastyear)

which(is.na(a$kriseis2))
which(is.na(a$Q17))
which(is.na(a$new_Q18))
which(is.na(a$kriseis_lastyear))
which(is.na(a$Q60))

new_data=data.frame(a$kriseis2,a$Q17,a$new_Q18,a$kriseis_lastyear,a$Q60)
names(new_data)=c("kriseis2","Q17","new_Q18","kriseis_lastyear","Q60")
new_data=new_data[-c(11,15,18,33,34,61,62,71,74,77,82),]
m=glm(kriseis2~Q17+new_Q18+kriseis_lastyear+Q60,data=new_data,
      family=binomial)
summary(m)
```

```
a=a[,-c(1,3,6,8,10,13)]
library(Hmisc)
a$Q7=impute(a$Q7,mode)
a$Q14=impute(a$Q14,mode)
a$Q17=impute(a$Q17,mode)
a$Q20=impute(a$Q20,mode)
a$Q25=impute(a$Q25,mode)
a$Q30=impute(a$Q30,mode)
a$Q31=impute(a$Q31,mode)
a$Q33=impute(a$Q33,mode)
a$Q55=impute(a$Q55,mode)
a$Q56=impute(a$Q56,mode)
a$Q59=impute(a$Q59,mode)
a$Q60=impute(a$Q60,mode)
a$Q61=impute(a$Q61,mode)
a$Q63=impute(a$Q63,mode)
a$Q64=impute(a$Q64,mode)
a$Q74=impute(a$Q74,mode)
a$Q76=impute(a$Q76,mode)
a$Q80=impute(a$Q80,mode)
a$Q82=impute(a$Q82,mode)
a$Q86=impute(a$Q86,mode)
a$Q88=impute(a$Q88,mode)
a$Q91=impute(a$Q91,mode)
a$new_Q2=impute(a$new_Q2,mode)
a$new_Q11=impute(a$new_Q11,mode)
a$new_Q18=impute(a$new_Q18,mode)
a$kriseis2=impute(a$kriseis2,mode)
a$kriseis_lastyear=impute(a$kriseis_lastyear,mode)

library(readxl)
a=data.frame(read_excel("4a.xlsx"))
a$kriseis2=as.factor(a$kriseis2)
a$Q17=as.factor(a$Q17)
a$new_Q18=as.factor(a$new_Q18)
a$kriseis_lastyear=as.factor(a$kriseis_lastyear)
a$Q60=as.factor(a$Q60)
```

```
m=glm(kriseis2~Q17+new_Q18+kriseis_lastyear+Q60,data=a,family=binomial)
summary(m)
anova(m,test="Chisq")
logLik(m)
library(ResourceSelection)
y=a$kriseis2
m=glm(y~Q17+new_Q18+kriseis_lastyear+Q60,data=a,family=binomial)
hoslem.test(m$y,fitted(m))
m=glm(kriseis2~Q17+kriseis_lastyear+Q60,data=a,family=binomial)
summary(m)
anova(m,test="Chisq")
logLik(m)
y=a$kriseis2
m=glm(y~Q17+kriseis_lastyear+Q60,data=a,family=binomial)
hoslem.test(m$y,fitted(m),g=12)

library(questionr)
odds.ratio(m,level=0.95)

library(caret)
logRegModel=train(kriseis2~Q17+kriseis_lastyear+Q60,data=a,method='glm',
        family="binomial")
logRegPrediction=predict(logRegModel,a[,-1])
logRegConfMat=confusionMatrix(logRegPrediction,a[,"kriseis2"])
logRegConfMat
library(pROC)
mod=glm(kriseis2~Q17+kriseis_lastyear+Q60,data=a,family=binomial)
th=factor(a$kriseis2)
pre=predict(mod,a[,-1],type="response")
r=roc(th,pre)
plot.roc(r,print.auc=T,main="ROC Curve for Model M1")

library(readxl)
b=data.frame(read_excel("4b.xlsx"))
b$kriseis2=as.factor(b$kriseis2)
b$Q17=as.factor(b$Q17)
b$new_Q18=as.factor(b$new_Q18)
```

```
b$kriseis_lastyear=as.factor(b$kriseis_lastyear)
b$Q60=as.factor(b$Q60)
library(Hmisc)
b$kriseis2=impute(b$kriseis2,mode)
b$Q17=impute(b$Q17,mode)
b$new_Q18=impute(b$new_Q18,mode)
b$kriseis_lastyear=impute(b$kriseis_lastyear,mode)
b$Q60=impute(b$Q60,mode)

m=glm(kriseis2~Q17+new_Q18+kriseis_lastyear+Q60,data=b,family=binomial)
summary(m)
anova(m,test="Chisq")
logLik(m)
library(ResourceSelection)
y=b$kriseis2
m=glm(y~Q17+new_Q18+kriseis_lastyear+Q60,data=b,family=binomial)
hoslem.test(m$y,fitted(m))
m=glm(kriseis2~Q17+kriseis_lastyear+Q60,data=b,family=binomial)
summary(m)
anova(m,test="Chisq")
logLik(m)
y=b$kriseis2
m=glm(y~Q17+kriseis_lastyear+Q60,data=b,family=binomial)
hoslem.test(m$y,fitted(m))

library(questionr)
odds.ratio(m,level=0.95)

library(caret)
logRegModel=train(kriseis2~Q17+kriseis_lastyear+Q60,data=b,method='glm',
        family="binomial")
logRegPrediction=predict(logRegModel,b[,-3])
logRegConfMat=confusionMatrix(logRegPrediction,b[,"kriseis2"])
logRegConfMat
library(pROC)
mod=glm(kriseis2~Q17+kriseis_lastyear+Q60,data=b,family=binomial)
th=factor(b$kriseis2)
```

```
pre=predict(mod,b[,-3],type="response")
r=roc(th,pre)
plot.roc(r,print.auc=T,main="ROC Curve for Model M3")


#################### εντολές κεφαλαίου 5 ##############################
######################################################################
library(readxl)
a=data.frame(read_excel("4a.xlsx"))
a$kriseis2=as.factor(a$kriseis2)
a$Q17=as.factor(a$Q17)
a$new_Q18=as.factor(a$new_Q18)
a$kriseis_lastyear=as.factor(a$kriseis_lastyear)
a$Q60=as.factor(a$Q60)
levels(a$kriseis2)=c("όχι","ναι")
levels(a$Q17)=c("1-5","6-10","πάνω από 21","πάνω από 100")
levels(a$new_Q18)=c("0-1","2","3-9")
levels(a$kriseis_lastyear)=c("όχι","ναι")
levels(a$Q60)=c("παντρεμένος-η","άγαμος-η ή χωρισμένος-η")
library(caret)
set.seed(10)
inTrainRows=createDataPartition(a$kriseis2,p=0.7,list=FALSE)
trainData=a[inTrainRows,]
testData=a[-inTrainRows,]
library(party)
n=ctree(kriseis2~.,data=trainData)
n
plot(n)
dt=predict(n,testData[,-1])
confusionMatrix(dt,testData[,"kriseis2"])

library(arules)
names(a)=c("kriseis2","doctor's_visit","total_kriseis","kriseis_lastyear",
        "marital_status")
ep_rules=apriori(data=a,parameter=list(supp=0.1,conf =0.80),
        appearance=list(rhs="kriseis2=ναι"))
inspect(ep_rules)
```

```
library(arulesViz)
library(visNetwork)
plot(ep_rules,method="graph",engine="htmlwidget")
```

Π3 ΠΙΝΑΚΕΣ

Παράμετρος	Εκτίμηση	Τυπικό σφάλμα
Σταθερά	-15.5947	3956.1804
total_kriseis2	12.2577	3956.1807
total_kriseis3	12.3909	3956.1806
total_kriseis4	13.8111	3956.1805
total_kriseis5	14.3639	3956.1805
doctor's visit2	0.6287	0.8148
doctor's visit3	1.0597	0.8153
kriseis_lastyear1	3.5761	0.8873
marital_status2	-1.9714	0.7922
marital_status3	-15.8697	2708.2344

Πίνακας Α1: Εκτιμήσεις των παραμέτρων του μοντέλου που προκύπτει όταν δεν συγχωνεύουμε κατηγορίες των μεταβλητών που αφορούν τον συνολικό αριθμό κρίσεων και την οικογενειακή κατάσταση

Σχόλιο για τον Πίνακα Α1

- $total\_kriseis2 = \begin{cases} 1, & το\ άτομο\ έχει\ κάνει\ 2-5\ κρίσεις\ συνολικά\ στη\ ζωή\ του \\ 0, & διαφορετικά \end{cases}$
- $total\_kriseis3 = \begin{cases} 1, & το\ άτομο\ έχει\ κάνει\ 6-10\ κρίσεις\ συνολικά\ στη\ ζωή\ του \\ 0, & διαφορετικά \end{cases}$
- $total\_kriseis4 =$
  $\begin{cases} 1, & το\ άτομο\ έχει\ κάνει\ πάνω\ από\ 21\ \ κρίσεις\ συνολικά\ στη\ ζωή\ του \\ 0, & διαφορετικά \end{cases}$
- $total\_kriseis5 =$
  $\begin{cases} 1, & το\ άτομο\ έχει\ κάνει\ πάνω\ από\ 100\ \ κρίσεις\ συνολικά\ στη\ ζωή\ του \\ 0, & διαφορετικά \end{cases}$

- $doctor's\_visit2 =$

  $\begin{cases} 1, & το\ άτομο\ επισκέφθηκε\ 2\ φορές\ τον\ γιατρό\ τον\ τελευταίο\ χρόνο \\ 0, & διαφορετικά \end{cases}$
- $doctor's\_visit3 =$

  $\begin{cases} 1, & το\ άτομο\ επισκέφθηκε\ 3-9\ φορές\ τον\ γιατρό\ τον\ τελευταίο\ χρόνο \\ 0, & διαφορετικά \end{cases}$
- $kriseis\_lastyear1 = \begin{cases} 1, & το\ άτομο\ εμφάνισε\ κρίσεις\ τον\ περασμένο\ χρόνο \\ 0, & διαφορετικά \end{cases}$
- $marital\_status2 = \begin{cases} 1, & το\ άτομο\ είναι\ άγαμο \\ 0, & διαφορετικά \end{cases}$
- $marital\_status3 =$

  $\begin{cases} 1, & το\ άτομο\ είναι\ χωρισμένο \\ 0, & διαφορετικά \end{cases}$

ΒΙΒΛΙΟΓΡΑΦΙΑ

ΕΛΛΗΝΙΚΗ

Σαχλάς Α., Μπερσίμης Σ. (2017). *Εφαρμοσμένη Στατιστική με Έμφαση στις Επιστήμες Υγείας*, Εκδόσεις Τζιόλα.

Νικολάου Χ. (2019). *Ανάλυση Δεδομένων με την R*, Εκδόσεις Δίσιγμα.

Βερύκιος Β., Κωτσιαντής Σ., Σταυρόπουλος Η., Τζαγκαράκης Μ. (2019). *Η Επιστήμη των Δεδομένων: Βασικές Αρχές, Θεωρία & Εφαρμογές με τη Γλώσσα R*, Εκδόσεις Νέων Τεχνολογιών.

Βερύκιος Β., Καγκλής Β., Σταυρόπουλος Η. (2015). *Η επιστήμη των δεδομένων μέσα από τη γλώσσα R*., Εκδόσεις Κάλλιπος

Κύρκος Ε. (2015). *Επιχειρηματική Ευφυΐα και Εξόρυξη Δεδομένων*, Εκδόσεις Κάλλιπος

Πολίτης Κ. (2020). Πανεπιστημιακές Σημειώσεις για το μεταπτυχιακό μάθημα *Γενικευμένα Γραμμικά Μοντέλα,* ΠΜΣ «Εφαρμοσμένη Στατιστική», Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς.

Μερκούρης Π. (2018). Πανεπιστημιακές Σημειώσεις για το προπτυχιακό μάθημα *Προχωρημένες Μέθοδοι Δειγματοληψίας*, Τμήμα Στατιστικής, Οικονομικό Πανεπιστήμιο Αθηνών.

Μπατσίδης Α. (2014). Πανεπιστημιακές Σημειώσεις για το προπτυχιακό μάθημα *Στατιστική Ανάλυση Δεδομένων*, Τμήμα Μαθηματικών, Πανεπιστήμιο Ιωαννίνων.

Νικολάκης Δ. Π. (2010). *Ψυχοκοινωνικό προφίλ επιληπτικών ασθενών: Μεταβολές την τελευταία δεκαετία*, Διπλωματική εργασία για το ΠΜΣ «Εφαρμοσμένη Στατιστική», Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς.

Σωφρονάς Η. (2015). *Τεχνικές εξόρυξης δεδομένων: Μελέτη εξόρυξης δεδομένων στον αθλητισμό με χρήση του λογισμικού Weka*, Διπλωματική εργασία για το Τμήμα Μηχανολόγων Μηχανικών, Εθνικό Μετσόβιο Πολυτεχνείο.

ΞΕΝΗ

Agresti A. (1996). *An Introduction to Categorical Data Analysis*, Wiley series in probability and statistics.

ΙΣΤΟΣΕΛΙΔΕΣ

https://gatzonhs.gr/xrisima/

https://nevrologos.gr/epilipsia/

https://www.iatronet.gr/ygeia/nevrologia/article/546/epilipsia.html

https://www.noesi.gr/book/syndrome/epilepsy

http://www.amarkos.gr/material/%CE%A3%CF%85%CE%BD%CE%AC%CF%86%CE%B5%CE%B9%CE%B1.pdf

https://ir.lib.uth.gr/xmlui/bitstream/handle/11615/47986/16472.pdf?sequence=1&isAllowed=y

https://www.rdocumentation.org/packages/party/versions/1.3-7/topics/Conditional%20Inference%20Trees

https://el.wikipedia.org/wiki/%CE%9A%CE%B1%CE%BD%CF%8C%CE%BD%CE%B5%CF%82_%CF%83%CF%85%CF%83%CF%87%CE%AD%CF%84%CE%B9%CF%83%CE%B7%CF%82

https://www.researchgate.net/figure/The-Knowledge-Discovery-in-Databases-KDD-process_fig1_274425359