

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΠΡΟΓΝΩΣΤΙΚΑ ΜΟΝΤΕΛΑ ΕΠΙΠΕΔΟΥ
ΧΟΛΗΣΤΕΡΙΝΗΣ ΣΤΟ ΑΙΜΑ ΜΕ ΒΑΣΗ
ΔΗΜΟΓΡΑΦΙΚΑ, ΣΩΜΑΤΟΜΕΤΡΙΚΑ ΚΑΙ
ΔΙΑΤΡΟΦΙΚΑ ΔΕΔΟΜΕΝΑ**

Βαρβάρα Ν. Λαγού

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
Ιανουάριος 2015

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΠΡΟΓΝΩΣΤΙΚΑ ΜΟΝΤΕΛΑ ΕΠΙΠΕΔΟΥ
ΧΟΛΗΣΤΕΡΙΝΗΣ ΣΤΟ ΑΙΜΑ ΜΕ ΒΑΣΗ
ΔΗΜΟΓΡΑΦΙΚΑ, ΣΩΜΑΤΟΜΕΤΡΙΚΑ ΚΑΙ
ΔΙΑΤΡΟΦΙΚΑ ΔΕΔΟΜΕΝΑ**

Βαρβάρα Ν. Λαγού

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
Ιανουάριος 2015

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

Τα μέλη της Επιτροπής ήταν:

- Επίκουρος Καθηγητής: Γεώργιος Τζαβελάς (Επιβλέπων)
- Επίκουρος Καθηγήτρια: Γεωργία Βερροπούλου
- Καθηγητής: Δημοσθένης Παναγιωτάκος

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



**DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN
APPLIED STATISTICS**

**PREDICTIVE MODELS FOR THE CHOLESTEROL
LEVEL IN BLOOD BASED ON DEMOGRAPHIC,
SOMATOMETRIC AND NUTRITION DATA**

By

Varvara N. Lagou

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

January 2015

Στη Μητέρα μου, Αγγελική

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω θερμά τον επιβλέπων καθηγητή μου, κύριο Γεώργιο Τζαβελά, η συμβολή του οποίου ήταν καθοριστικής σημασίας. Με την καθοδήγηση του και την υπομονή του με βοήθησε να ολοκληρώσω σύντομα την εργασία αυτή.

Στη συνέχεια, θα ήθελα να ευχαριστήσω όλους τους καθηγητές μου, που και στο προπτυχιακό αλλά και στο μεταπτυχιακό μου, μου έδωσαν τις απαραίτητες γνώσεις ώστε να μπορέσω να ολοκληρώσω επιτυχώς τις σπουδές μου.

Ακόμη, θα ήθελα να ευχαριστήσω τον πατέρα μου και τους συγγενείς μου, οι οποίοι ήταν δίπλα μου σε κάθε μου βήμα.

Τέλος, θα ήθελα να ευχαριστήσω τους συμφοιτητές μου. Δεν ξεχνώ τις συναντήσεις της ομάδας του ΜΕΣ12 όπου λύναμε απορίες και εμπνυχώναμε ο ένας τον άλλο.

Περίληψη

Στην παρούσα εργασία έγινε μία προσπάθεια εύρεσης των δημογραφικών, σωματομετρικών και διατροφικών παραγόντων που σχετίζονται με το επίπεδο της χοληστερόλης. Για το σκοπό αυτό χρησιμοποιήθηκε η βάση δεδομένων από την επιδημιολογική μελέτη ΑΤΤΙΚΗ. Το ενδιαφέρον μας εστιάστηκε στην εύρεση των παραγόντων εκείνων που σχετίζονται με την LDL – χοληστερόλη, επειδή αυτή είναι κυρίως η ποσότητα που επηρεάζει τη συνολική χοληστερόλη.

Αναλυτικά, στο πρώτο κεφάλαιο δίνεται η περιγραφή της νόσου, ο τρόπος υπολογισμού της συνολικής χοληστερόλης καθώς και η μεθοδολογία που ακολουθήθηκε στα επόμενα κεφάλαια. Στο δεύτερο κεφάλαιο δίνεται η περιγραφική ανάλυση των μεταβλητών που διαθέτουμε, ενώ το τρίτο κεφάλαιο δίδεται η στατιστική ανάλυση η οποία περιλαμβάνει τις συσχετίσεις και τον έλεγχο ανεξαρτησίας των μεταβλητών με την χοληστερόλη, παρουσιάζεται ο στατιστικός έλεγχος υποθέσεων και πιο συγκεκριμένα ο έλεγχος t – test. Στο τέταρτο κεφάλαιο εφαρμόζουμε στα δεδομένα μας το πολλαπλό γραμμικό μοντέλο, το λογιστικό μοντέλο καθώς και η διαχωριστική ανάλυση και συγκρίνουμε την διακριτική τους ικανότητα. Στο πέμπτο κεφάλαιο έχουμε την ανάλυση της καμπύλης ROC .

Στο τελευταίο κεφάλαιο παρουσιάζονται αναλυτικά τα συμπεράσματα στα οποία καταλήξαμε.

Abstract

The aim of this paper is to find demographic, somatometric and nutrition factors responsible for the cholesterol level in blood. For this purpose we used data from ATTICA data base. We concentrated mainly on finding the factors responsible for the total cholesterol as well as for the LDL – cholesterol level in blood.

More specifically, in Chapter 1 we give the description of the disease, the way the cholesterol level is calculated as well as the methodologies which are used for the statistical analysis. In Chapter 2, the descriptive analysis of the variables is presented. In Chapter 3 the statistical analysis is presented which includes the correlations between cholesterol and all other variables and the tests of statistical hypotheses of the cholesterol level and the other variables(t-test, Chi-squared tests). In Chapter 4 with the use of multiple linear regression, discriminant analysis and logistic regression, predictive models for the cholesterol level in blood are constructed. In Chapter 5 with the use of the Receiver Operating Curve (ROC) we compare the models. In Chapter 6 we summarize our findings .

Περιεχόμενα

| | |
|-----------------------------|----|
| Περιεχόμενα | xv |
| Κατάλογος Πινάκων | xv |
| Κατάλογος Διαγραμμάτων..... | xv |

ΚΕΦΑΛΑΙΟ 1: Εισαγωγή

| | | |
|-------|-----------------------------------------------------------------|---|
| 1.1 | Εισαγωγή..... | 1 |
| 1.2 | Τι είναι η νόσος..... | 2 |
| 1.3 | Υπολογισμός της συνολικής χοληστερόλης..... | 4 |
| 1.4 | Φυσιολογικές τιμές..... | 4 |
| 1.5 | Μεθοδολογία..... | 5 |
| 1.5.1 | Ο συντελεστής συσχέτισης του Pearson..... | 5 |
| 1.5.2 | Το διάγραμμα διασποράς (Scatter Plot)..... | 6 |
| 1.5.3 | Ο έλεγχος ανεξαρτησίας χ^2 – τετράγωνο..... | 6 |
| 1.5.4 | Έλεγχος για τη διαφορά των μέσων δύο ανεξάρτητων πληθυσμών..... | 7 |
| 1.5.5 | Ο έλεγχος των Kolmogorov – Smirnov..... | 8 |
| 1.5.6 | Κλίμακα μεσογειακής διατροφής..... | 8 |
| 1.5.7 | Περιγραφή δείγματος..... | 8 |

ΚΕΦΑΛΑΙΟ 2: Περιγραφική ανάλυση

| | | |
|-----|-------------------------------------------------|----|
| 2.1 | Περιγραφική Ανάλυση Συνεχών Μεταβλητών..... | 9 |
| 2.2 | Περιγραφική Ανάλυση Κατηγορικών Μεταβλητών..... | 10 |

ΚΕΦΑΛΑΙΟ 3: Συσχετίσεις – Έλεγχος ανεξαρτησίας

| | | |
|-----|-----------------------------------------------------------------------------------------------------------------------|----|
| 3.1 | Εισαγωγή..... | 13 |
| 3.2 | Συντελεστής συσχέτισης του Pearson για τη μεταβλητή Συνολική Χοληστερόλη και τα αντίστοιχα διαγράμματα διασποράς..... | 15 |

| | | |
|-----|-----------------------------------------------------------------------------|----|
| 3.3 | Συντελεστής συσχέτισης του Pearson για τη μεταβλητή LDL Χοληστερόλη..... | 16 |
| 3.4 | Έλεγχος χ^2 – τετράγωνο για την μεταβλητή Έπαρξη Χοληστερόλης..... | 31 |
| 3.5 | T – Test για τη μεταβλητή Συνολική Χοληστερόλη..... | 32 |
| 3.6 | T – Test για τη μεταβλητή LDL - Χοληστερόλη..... | 33 |
| 3.7 | Μη παραμετρικός έλεγχος για τη μεταβλητή Έπαρξη Χοληστερόλης..... | 34 |
| 3.8 | Μη παραμετρικός έλεγχος για τη μεταβλητή LDL - Χοληστερόλης..... | 35 |

ΚΕΦΑΛΑΙΟ 4: Προγνωστικά μοντέλα – Πολλαπλή γραμμική και λογιστική παλινδρόμηση

| | | |
|-------|--------------------------------------------------------|----|
| 4.1 | Εισαγωγή..... | 37 |
| 4.2 | Πολλαπλή γραμμική παλινδρόμηση..... | 38 |
| 4.2.1 | Πολυσυγγραμμικότητα..... | 39 |
| 4.3 | Βέλτιστα μοντέλα..... | 40 |
| 4.3.1 | Βέλτιστο μοντέλο για τη Συνολική Χοληστερόλη..... | 41 |
| 4.3.2 | Βέλτιστο μοντέλο για την LDL Χοληστερόλη..... | 42 |
| 4.4 | Πολλαπλή λογιστική παλινδρόμηση..... | 43 |
| 4.4.1 | Βέλτιστο μοντέλο για τη Συνολική Χοληστερόλη..... | 44 |
| 4.4.2 | Βέλτιστο μοντέλο για την LDL Χοληστερόλη..... | 45 |
| 4.5 | Διαχωριστική ανάλυση..... | 46 |
| 4.6 | Διαχωριστική ανάλυση για την LDL Χοληστερόλη..... | 47 |
| 4.7 | Διαχωριστική ανάλυση για την Συνολική Χοληστερόλη..... | 47 |

ΚΕΦΑΛΑΙΟ 5: Ανάλυση Καμπύλης ROC

| | | |
|------------|---------------------------|----|
| 5.1 | Εισαγωγή..... | 57 |
| 5.2 | Ανάλυση καμπύλης ROC..... | 58 |

ΚΕΦΑΛΑΙΟ 6: Συμπεράσματα

| | |
|---------------------------|----|
| Βιβλιογραφία | 62 |
|---------------------------|----|

Κατάλογος διαγραμμάτων

| | | |
|--------------|--------------------------------------------------------------------------------------------|----|
| 2.2.1 | Κυκλικό διάγραμμα για την μεταβλητή Φύλο..... | 10 |
| 2.2.2 | Κυκλικό διάγραμμα για την μεταβλητή Οικογενειακή Κατάσταση. | 11 |
| 2.2.3 | Κυκλικό διάγραμμα για την μεταβλητή Ύπαρξη Χοληστερόλης.... | 12 |
| 3.1.1 | Ιστόγραμμα συχνοτήτων για την μεταβλητή Ύπαρξη Χοληστερόλης..... | 14 |
| 3.1.2 | Ιστόγραμμα συχνοτήτων για την μεταβλητή Συνολική Χοληστερόλης..... | 15 |
| 3.2.1 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και την Ηλικία..... | 16 |
| 3.2.2 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και το Βάρος..... | 16 |
| 3.2.3 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και τον Δείκτη Μάζας Σώματος..... | 17 |
| 3.2.4 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και την Περιφέρεια Μέσης..... | 17 |
| 3.2.5 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και την Περιφέρεια Γοφών..... | 18 |
| 3.2.6 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και τον Αριθμό τσιγάρων ανά ημέρα..... | 18 |
| 3.2.7 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και τον αριθμό ετών καπνίσματος..... | 19 |
| 3.2.8 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και τη Συστολική Πίεση..... | 19 |

| | | |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------|----|
| 3.2.9 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και την Διαστολική πίεση..... | 20 |
| 3.2.10 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και το επίπεδο γλυκόζης στο αίμα..... | 20 |
| 3.2.11 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και τα Τριγλυκερίδια..... | 21 |
| 3.2.12 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και την LDL – Χοληστερόλη..... | 21 |
| 3.2.13 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και το ποσοστό πρόσληψης πρωτεϊνών..... | 22 |
| 3.2.14 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και το Ποσοστό Πρόσληψης Κορεσμένων Λιπαρών Οξέων..... | 22 |
| 3.2.15 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και το Σκορ Μεσογειακής Διατροφής..... | 23 |
| 3.2.16 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και την Εβδομαδιαία Κατανάλωση Γαλακτοκομικών..... | 23 |
| 3.2.17 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και την Εβδομαδιαία Κατανάλωση Γλυκών..... | 24 |
| 3.2.18 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και την Εβδομαδιαία Κατανάλωση Αυγών..... | 24 |
| 3.2.19 | Διάγραμμα διασποράς για την Συνολική Χοληστερόλη και την Εβδομαδιαία Κατανάλωση Κρέατος..... | 25 |
| 4.6.1 | Διάγραμμα διασποράς τριών διαστάσεων για τις μεταβλητές Δείκτης Μάζας Σώματος, Περιφέρεια Μέσης και Αριθμός Ετών Καπνίσματος..... | 45 |
| 4.7.1 | Διάγραμμα διασποράς τριών διαστάσεων για τις μεταβλητές Δείκτης Μάζας Σώματος, Ύπαρξη Χοληστερόλης και Αριθμός Ετών Καπνίσματος..... | 47 |
| 5.2.1 | Ανάλυση Καμπύλης ROC για τη μεταβλητή Συνολική Χοληστερόλη..... | 51 |
| 5.2.2 | Ανάλυση Καμπύλης ROC για τη μεταβλητή LDL Χοληστερόλη..... | 52 |

Κατάλογος πινάκων

| | | |
|----------------|-------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1.1 | Συγκεντρωτικός Πίνακας Συνεχών Μεταβλητών..... | 9 |
| 2.2.2 | Πίνακας Συχνοτήτων για την μεταβλητή Φύλο..... | 10 |
| 2.2.3 | Πίνακας Συχνοτήτων για την μεταβλητή Οικογενειακή Κατάσταση. | 11 |
| 2.2.4 | Πίνακας Συχνοτήτων για την μεταβλητή Ύπαρξη Χοληστερόλης.... | 12 |
| 3.1.1 | Έλεγχος Κανονικότητας για τις μεταβλητές LDL Χοληστερόλη, HDL Χοληστερόλη, Συνολική Χοληστερόλη και για τα Τριγλυκερίδια..... | 14 |
| 3.2.1 | Συγκεντρωτικός Πίνακας για την μεταβλητή Συνολική Χοληστερόλη..... | 25 |
| 3.3.1 | Συγκεντρωτικός Πίνακας για την μεταβλητή LDL Χοληστερόλη..... | 27 |
| 3.3.2 | Συγκεντρωτικός Πίνακας για την μεταβλητή LDL Χοληστερόλη..... | 28 |
| 3.4.2 | Συγκεντρωτικός Πίνακας για την μεταβλητή Τριγλυκερίδια..... | 68 |
| 3.6.1 | Συγκεντρωτικός Πίνακας για την μεταβλητή Ύπαρξη Χοληστερόλης..... | 29 |
| 3.6.2 | Συγκεντρωτικός Πίνακας για την μεταβλητή Ύπαρξη Χοληστερόλης..... | 30 |
| 3.7.1 | Συγκεντρωτικός Πίνακας για την μεταβλητή Συνολική Χοληστερόλη..... | 31 |
| 3.7.2 | Πίνακας μέσων τιμών και τυπικών αποκλίσεων ανά κατηγορία..... | 32 |
| 3.8.1 | Συγκεντρωτικός Πίνακας για την μεταβλητή LDL Χοληστερόλη..... | 33 |
| 3.8.2 | Πίνακας μέσων τιμών και τυπικών αποκλίσεων ανά κατηγορία..... | 33 |
| 3.8.3 | Συγκεντρωτικός Πίνακας για την μεταβλητή LDL Χοληστερόλη..... | 33 |
| 3.9.1 | Συγκεντρωτικός Πίνακας για την μεταβλητή Συνολική Χοληστερόλη... | 33 |
| 3.10.1 | Συγκεντρωτικός Πίνακας για την μεταβλητή LDL Χοληστερόλη..... | 35 |
| 4.3.1.1 | Βέλτιστο μοντέλο για την Συνολική Χοληστερόλη..... | 40 |

| | | |
|---------|----------------------------------------------------|----|
| 4.3.1.2 | F – Test για το βέλτιστο μοντέλο..... | 42 |
| 4.3.1.3 | Πίνακας ορθής πρόβλεψης..... | 42 |
| 4.3.2.1 | Βέλτιστο μοντέλο για την LDL Χοληστερόλη..... | 43 |
| 4.3.2.2 | F – Test για το βέλτιστο μοντέλο..... | 44 |
| 4.3.2.3 | Πίνακας ορθής πρόβλεψης..... | 44 |
| 4.3.2.4 | Βέλτιστο μοντέλο για την LDL Χοληστερόλη..... | 45 |
| 4.3.2.5 | F – Test για το βέλτιστο μοντέλο..... | 45 |
| 4.3.2.6 | Πίνακας ορθής πρόβλεψης..... | 45 |
| 4.4.1 | Βέλτιστο μοντέλο για την Συνολική Χοληστερόλη..... | 46 |
| 4.4.2 | Έλεγχος χ^2 – τετράγωνο για το μοντέλο..... | 47 |
| 4.4.3 | Πίνακας ορθής πρόβλεψης..... | 47 |
| 4.5.1 | Βέλτιστο μοντέλο για την LDL Χοληστερόλη..... | 48 |
| 4.5.2 | Έλεγχος χ^2 – τετράγωνο για το μοντέλο..... | 49 |
| 4.5.3 | Πίνακας ορθής πρόβλεψης..... | 49 |
| 4.6.1 | Box Test..... | 49 |
| 4.6.2 | Έλεγχος της ισότητας των μέσων τιμών..... | 49 |
| 4.6.3 | Πίνακας..... | 49 |
| 4.6.4 | Έλεγχος του Fisher..... | 49 |
| 4.6.5 | Διαχωριστική Συνάρτηση..... | 50 |
| 4.7.1 | Box Test..... | 51 |
| 4.7.2 | Έλεγχος της ισότητας των μέσων τιμών..... | 52 |
| 4.7.3 | Πίνακας..... | 53 |
| 4.7.4 | Έλεγχος της ισότητας των μέσων τιμών..... | 53 |
| 4.7.5 | Διαχωριστική Συνάρτηση..... | 53 |

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

1.1 Εισαγωγή

Η χοληστερίνη είναι ένα λιπίδιο που συντίθεται κατά κύριο λόγο στο ήπαρ. Πιο συγκεκριμένα η χοληστερόλη όπως αυτή αποκαλείται, είναι μία κηρώδης στερόλη που βρίσκεται στην μεμβράνη όλων των κυττάρων των ιστών του σώματος. Όσον αφορά τη λειτουργία της, αυτή χρησιμοποιείται σε ζωικούς παρά σε φυτικούς οργανισμούς. Στην χοληστερίνη προσδίδουν λειτουργικότητα οι διαμοριακές δυνάμεις που αναπτύσσονται μεταξύ αυτής και του νερού που βρίσκεται στο κύτταρο. Το λίπος που κυκλοφορεί στο αίμα μέσω των τριγλυκεριδίων μεταφέρεται στους ιστούς του σώματος όπου αυτά είναι αποθηκευμένα για την κάλυψη των ενεργειακών αναγκών του σώματος.

Η χοληστερίνη χρησιμεύει ως πρώτη ύλη στην παρασκευή της προβιταμίνης D και στο σχηματισμό ορμονών και δεσμεύεται με πρωτεΐνες και λιποπρωτεΐνες οι οποίες είναι μία βιοχημική συνένωση αποτελούμενη από δύο πρωτεΐνες και λιπίδια. Σε μεγαλύτερη συγκέντρωση βρίσκεται σε ιστούς που έχουν πυκνές μεμβράνες όπως είναι το συκώτι, η σπονδυλική στήλη και ο εγκέφαλος, στα οποία και μεταφέρεται μέσα από το αίμα μέσω των λιποπρωτεϊνών. Η καλή λιποπρωτεΐνη η οποία είναι η HDL, στην ουσία απομακρύνει τη χοληστερίνη από τα κύτταρα των τοιχωμάτων των αρτηριών προς το συκώτι. Με αυτόν τον τρόπο μειώνεται ο κίνδυνος εμφράγματος. Είναι υψηλής πυκνότητας και μεταφέρει το 20-30% της ολικής χοληστερόλης στα διάφορα όργανα. Η κακή λιποπρωτεΐνη η οποία είναι η LDL, σχηματίζει λιπαρά εναποθέματα στα κύτταρα των τοιχωμάτων των αρτηριών. Με αυτόν τον τρόπο δημιουργείται αθηροσκληρωτική πλάκα. Είναι χαμηλής πυκνότητας και μεταφέρει το 60-70% της ολικής χοληστερόλης.

Αυξημένη χοληστερίνη στο αίμα αυξάνει τον κίνδυνο αθηροσκλήρωσης κάτι που οδηγεί σε έμφραγμα του μυοκαρδίου και σε εγκεφαλικό επεισόδιο. Το επίπεδο της χοληστερίνης στο αίμα επηρεάζεται από την παχυσαρκία, από γενετικούς ή κληρονομικούς παράγοντες και από την μειωμένη φυσική δραστηριότητα. Ο κυριότερος παράγοντας που την επηρεάζει όμως

είναι η διατροφή. Με μείωση γαλακτοκομικών και ζωικών τροφών μειώνεται και το επίπεδο χοληστερίνης στο αίμα. Ο κίνδυνος για καρδιαγγειακή νόσο αυξάνεται με την υπερχοληστερολαιμία και πιο συγκεκριμένα με μία σειρά παραγόντων οι οποίοι είναι ο σακχαρώδης διαβήτης, το κάπνισμα, η παχυσαρκία, η καθιστική ζωή, η αυξημένη λιποπρωτεΐνη και τα υπερτριγλυκερίδια. Αυξημένο επίπεδο τριγλυκεριδίων μπορεί να αυξήσει τον κίνδυνο για στεφανιαία νόσο.(www.medidiet.gr)

1.2 Τι είναι η νόσος.(www.medidiet.gr, www.doctor4all.gr)

Η χοληστερίνη αυξάνει την πιθανότητα εμφάνισης στεφανιαίας νόσου η οποία είναι η σύγχρονη μάστιγα του 20^{ου} αιώνα. Το λίπος που κυκλοφορεί στο αίμα, τα τριγλυκερίδια δηλαδή, μεταφέρεται στους διάφορους ιστούς τους σώματος. Τα τριγλυκερίδια είναι αποθηκευμένα στους ιστούς αυτούς για την κάλυψη των ενεργειακών αναγκών του σώματος. Οι τροφές που καταναλώνουν τα άτομα, όπως είναι οι πρωτεΐνες, οι υδατάνθρακες και τα λίπη, μετατρέπονται σε τριγλυκερίδια. Επομένως είναι απόλυτα φυσιολογικό να υπάρχουν τριγλυκερίδια στο αίμα. Αυτά παράγονται από τις θρεπτικές ουσίες που υπάρχουν στο διαιτολόγιο μας όταν το ολικό ποσό θερμίδων που λαμβάνουμε με την τροφή υπερβαίνει αυτό που απαιτείται για τις άμεσες ανάγκες του σώματος σε ενέργεια.

Η χοληστερίνη και τα τριγλυκερίδια είναι λιπίδια του αίματος τα οποία για να κυκλοφορήσουν στο αίμα ενσωματώνονται σε ειδικά ευδιάλυτα στο πλάσμα του αίματος συμπλέγματα τα οποία ονομάζονται λιποπρωτεΐνες. Η LDL-Low Density Lipoproteins είναι η χαμηλής πυκνότητας λιποπρωτεΐνη, η λεγόμενη κακή χοληστερίνη. Αυτή μεταφέρει τα τριγλυκερίδια από το συκώτι στους ιστούς του σώματος. Η HDL-High Density Lipoproteins είναι η υψηλής πυκνότητας λιποπρωτεΐνη, η λεγόμενη καλή χοληστερίνη. Αυτή μεταφέρει τη χοληστερίνη στο συκώτι όπου αποδομείται και απεκκρίνεται από τον οργανισμό. Έτσι προστατεύονται τα αγγεία από την αθηροσκλήρυνση. Στον άνθρωπο η χοληστερίνη χρησιμεύει στην κάλυψη των αναγκών του οργανισμού για την κατασκευή κυτταρικών μεμβρανών, την σύνθεση ορμονών και το σχηματισμό χολικών αλάτων.

Η οικογενής συνδυασμένη υπερλιπιδαιμία, η οικογενής δυσβηταλιποπρωτεϊναιμία και η οικογενής υπερτριγλυκεριδαιμία είναι σοβαρές κληρονομικές διαταραχές οι οποίες αυξάνουν τα επίπεδα της χοληστερίνης στο αίμα. Όμοιες αυξήσεις προκαλεί και η αύξηση σωματικού βάρους ή η παχυσαρκία και μία σειρά φαρμάκων τα οποία είναι τα ακόλουθα:

- Τα οιστρογόνα
- Τα κορτικοστεροειδή
- Τα διουρητικά
- Τα αντικαταθλιπτικά
- Τα αντισυλληπτικά
- Οι βήτα-blockers

Καθώς επίσης και μία σειρά ασθενειών οι οποίες είναι οι ακόλουθες:

- Το σύνδρομο Cushing
- Το σύνδρομο πολυκυστικών ωοθηκών
- Διάφορες νεφροπάθειες
- Ο σακχαρώδης διαβήτης
- Και η κατάχρηση αλκοόλ

Ακόμη, οι λιπώδεις δίαιτες που περιέχουν υψηλές τιμές κορεσμένων λιπών (όπως είναι το κόκκινο κρέας, ο κρόκος αυγού, τα γαλακτοκομικά προϊόντα με υψηλά λιπαρά) και υπέρμετρα λιπαρά οξέα (τα οποία βρίσκονται στο έτοιμο και γρήγορο φαγητό). Τέλος, η έλλειψη σωματικής άσκησης και το κάπνισμα το οποίο και ελαττώνει τις τιμές της HDL-καλής χοληστερόλης.

1.3 Υπολογισμός της συνολικής χοληστερόλης(www.doctor4all.gr)

Η συνολική χοληστερόλη υπολογίζεται ως το άθροισμα τριών συστατικών. Αυτά είναι η HDL χοληστερόλη, η LDL χοληστερόλη και η VLDL χοληστερόλη. Η τελευταία είναι τα τριγλυκερίδια διαιρεμένα με το πέντε. Πιο συγκεκριμένα:

$$\text{Ολική Χοληστερόλη} = \text{HDL} + \text{LDL} + \text{VLDL όπου : VLDL} = 1/5 * \text{Τριγλυκερίδια}$$

Υπάρχουν δύο βασικές υποθέσεις για να ισχύει ο τύπος αυτός. Η πρώτη είναι ότι θα πρέπει τα δείγματα αίματος να λαμβάνονται μετά από δωδεκάωρη νηστεία, ενώ η δεύτερη είναι ότι θα πρέπει η τιμή των τριγλυκεριδίων να μην υπερβαίνει τα 400 mg/dl. Σε διαφορετική περίπτωση δεν είναι εφικτός ο υπολογισμός της ολικής χοληστερόλης.

1.4 Φυσιολογικές τιμές

Φυσιολογικές τιμές για την ολική χοληστερόλη θεωρούνται οι τιμές κάτω των 220 mg/dl. Όταν ένα άτομο έχει ολική χοληστερόλη άνω των 220 mg/dl θεωρείται ότι νοσεί και θα πρέπει να ακολουθήσει ειδική φαρμακευτική αγωγή.

Όσον αφορά για την LDL χοληστερόλη, οι φυσιολογικές της τιμές διαφέρουν ανάλογα με το εάν το άτομο ανήκει σε μία από τις τρεις ακόλουθες ομάδες. Η πρώτη ομάδα περιλαμβάνει τα άτομα που πάσχουν από στεφανιαία νόσο. Οι φυσιολογικές τιμές εδώ θεωρούνται αυτές που είναι μικρότερες των 100 mg/dl. Η δεύτερη ομάδα περιλαμβάνει τα άτομα τα οποία δεν πάσχουν από στεφανιαία νόσο και έχουν δύο ή περισσότερους προδιαθεσικούς παράγοντες κινδύνου. Εδώ φυσιολογικές τιμές θεωρούνται αυτές που είναι μικρότερες των 130 mg/dl. Η τρίτη και τελευταία ομάδα περιλαμβάνει τα άτομα που δεν πάσχουν από στεφανιαία νόσο και έχουν λιγότερους από δύο προδιαθεσικούς παράγοντες κινδύνου. Οι φυσιολογικές τιμές εδώ είναι αυτές που είναι μικρότερες των 160 mg/dl. Ως προδιαθεσικοί παράγοντες κινδύνου θεωρούνται το κάπνισμα, η υψηλή πίεση, πιο συγκεκριμένα η συστολική πίεση να είναι μεγαλύτερη του 140 και η διαστολική του 90 καθώς και οι χαμηλές τιμές της HDL χοληστερόλης, δηλαδή τιμές μικρότερες των 40 mg/dl. Ακόμη ως προδιαθεσικός παράγοντας κινδύνου θεωρείται και το εάν υπάρχει οικογενειακό ιστορικό πρόωρης καρδιοπάθειας,

δηλαδή εάν ο πατέρας ή ο αδερφός νόσησε σε ηλικία μικρότερη των 55 ετών και η μητέρα ή η αδερφή σε ηλικία μικρότερη των 65 ετών.

Για την HDL χοληστερόλη θεωρούμε ότι όταν οι τιμές της είναι επιθυμητές όταν είναι μικρότερες των 100 mg/dl, ενώ είναι οριακές όταν κυμαίνονται από 100 έως 129 mg/dl. Επίσης, τιμές από 130 έως 159 mg/dl θεωρούνται οριακά υψηλές, ενώ τιμές από 160 έως 189 mg/dl θεωρούνται υψηλές. Τέλος πολύ υψηλές τιμές θεωρούνται αυτές που είναι μεγαλύτερες των 190 mg/dl. Γενικά θα πρέπει ο λόγος της HDL προς την LDL χοληστερόλη να είναι περίπου ίσος με πέντε και όχι μεγαλύτερος. Για τα τριγλυκερίδια, τιμές άνω των 150 έως 199 mg/dl θεωρούνται πολύ υψηλές.

Οι παραπάνω τιμές είναι αυτές που προκύπτουν από την έρευνα του κ. Χάρη Καραχριστιανίδη σε άρθρο του στο site: www.mikroviologos.gr. Καθώς και του Ρίχτερ Δημητρίου, Διευθυντή της Β Καρδιολογικής Ευρωκλινικής Αθηνών, σε άρθρο του στο site: www.health4you.gr.

1.5 Μεθοδολογία

Στην παρούσα εργασία έχουν χρησιμοποιηθεί αρκετοί στατιστικοί έλεγχοι, ώστε να βγουν τα κατάλληλα συμπεράσματα. Όλοι αυτοί οι έλεγχοι παρουσιάζονται αναλυτικά παρακάτω.

1.5.1 Ο συντελεστής συσχέτισης του Pearson

Έστω δύο συνεχείς μεταβλητές X και Y για τις οποίες έχουμε καταγράψει τις τιμές x_i της X και τις αντίστοιχες τιμές y_i της, $i=1,2,\dots,n$. Ο συντελεστής συσχέτισης της X και της Y υπολογίζεται από τη σχέση:

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Εκφράζει το βαθμό γραμμικής συσχέτισης μεταξύ των μεταβλητών X και Y και παίρνει τιμές από -1 έως 1. Όσο μεγαλύτερος, σε απόλυτη τιμή, είναι ο συντελεστής συσχέτισης, τόσο μεγαλύτερη είναι η γραμμική εξάρτηση των δύο μεταβλητών.

1.5.2. Το Διάγραμμα Διασποράς – Scatter Plot

Έστω ότι έχουμε n ζεύγη παρατηρήσεων (x_i, y_i) , $i=1,2,\dots,n$ από τις μεταβλητές X και Y . Θέλοντας να αναπαραστήσουμε στο επίπεδο τα σημεία αυτά, τοποθετούμε τις παρατηρηθείσες τιμές της μεταβλητής X στον οριζόντιο άξονα και τις παρατηρηθείσες τιμές της μεταβλητής Y στον κατακόρυφο άξονα. Στην περίπτωση που τα σημεία του επιπέδου που προκύπτουν βρίσκονται «κοντά» σε μία ευθεία, μπορούμε να δεχθούμε ότι η σχέση των μεταβλητών X και Y είναι κατά προσέγγιση γραμμική. Διαφορετικά δεν μπορούμε να συμπεράνουμε ότι η σχέση των δύο μεταβλητών είναι γραμμική, χωρίς αυτό να σημαίνει ότι δεν σχετίζονται με κάποια άλλη μορφή συναρτησιακής εξάρτησης.

1.5.3. Ο έλεγχος χ^2

Η μέθοδος αυτή χρησιμοποιείται για να εκτιμήσουμε εάν δύο ή περισσότερα δείγματα τα οποία αποτελούνται από δεδομένα συχνοτήτων διαφέρουν σημαντικά μεταξύ τους. Πιο συγκεκριμένα χρησιμοποιείται για τον έλεγχο ύπαρξης ή όχι συσχέτισης μεταξύ δύο κατηγορικών μεταβλητών, αναλύοντας τον αντίστοιχο πίνακα διασταύρωσης ή συνάφειας. Η μηδενική υπόθεση υποστηρίζει ότι δεν υπάρχει σχέση μεταξύ της μεταβλητής «γραμμή» και της μεταβλητής «στήλη», σε έναν πίνακα διασταύρωσης με r γραμμές και c στήλες. Ο έλεγχος της υπόθεσης αυτής γίνεται με την σύγκριση των παρατηρούμενων συχνοτήτων (Π) κάθε κελιού με τις αναμενόμενες συχνότητες (A) που υπολογίζονται υπό την μηδενική υπόθεση. Η αναμενόμενη τιμή κάθε κελιού εφόσον ισχύει η μηδενική υπόθεση δίνεται από την ακόλουθη σχέση:

$$A = \frac{RT * CT}{N}$$

Όπου RT είναι το σύνολο της γραμμής που ανήκει το συγκεκριμένο κελί, CT το σύνολο της στήλης που ανήκει το συγκεκριμένο κελί και N ο συνολικός αριθμός των παρατηρήσεων.

Για να γίνει ο έλεγχος της μηδενικής υπόθεσης, υπολογίζουμε το στατιστικό χ^2 – τετράγωνο το οποίο συγκρίνει το σύνολο των παρατηρηθέντων συχνοτήτων με το σύνολο των αναμενόμενων συχνοτήτων μέσω του ακόλουθου τύπου:

$$x^2 = \frac{(\Pi - A)^2}{A}$$

Με την άθροιση να γίνεται για όλα τα $r \cdot c$ κελιά. Ο αριθμός των βαθμών ελευθερίας είναι ίσος με $df = (r-1) \cdot (c-1)$.

Όταν ένα ή περισσότερα κελιά έχουν αναμενόμενη συχνότητα μικρότερη του 5, ο έλεγχος υπολογίζεται σύμφωνα με τον ακριβή έλεγχο του Fisher για τους $2 \cdot 2$ πίνακες συνάφειας. Όταν μία κατηγορική μεταβλητή έχει περισσότερα από δύο επίπεδα, ο έλεγχος του Fisher υπολογίζεται με την τεχνική του Monte Carlo με 10.000 δείγματα και 95% διάστημα εμπιστοσύνης.

1.5.4 Έλεγχος για τη διαφορά των μέσων δύο ανεξάρτητων πληθυσμών

Στους στατιστικούς ελέγχους υποθέσεων επιθυμούμε να ελέγξουμε αν μία ή περισσότερες παράμετροι ενός πληθυσμού ικανοποιούν μια βασική υπόθεση, η οποία καλείται μηδενική, έναντι μιας δεύτερης, εναλλακτικής υπόθεσης. Όταν η κατανομή των δεδομένων μας είναι γνωστή, για παράδειγμα, όταν τα δεδομένα ακολουθούν την Κανονική Κατανομή, συνήθως κάνουμε ελέγχους για κάποιες παραμέτρους της και ο έλεγχος καλείται παραμετρικός. Έστω ότι θέλουμε να ελέγξουμε την υπόθεση ότι η μέση τιμή μ_1 ενός χαρακτηριστικού κάποιου κανονικού πληθυσμού με άγνωστη διακύμανση είναι ίση με τη μέση τιμή μ_2 ενός άλλου κανονικού πληθυσμού με άγνωστη διακύμανση, έναντι της εναλλακτικής ότι είναι διάφορη, σε δοσμένο επίπεδο σημαντικότητας α , δηλαδή:

$$H_0 : \mu_1 = \mu_2 \text{ vs } H_1 : \mu_1 \neq \mu_2$$

Οι ανεξάρτητοι πληθυσμοί από τους οποίους προέρχονται τα δύο ανεξάρτητα τυχαία δείγματα $(x_1, x_2, \dots, x_{n_1})$ μεγέθους n_1 και $(y_1, y_2, \dots, y_{n_2})$ μεγέθους n_2 αντίστοιχα) που θα χρησιμοποιηθούν για τον έλεγχο συνήθως αποτελούν επίπεδα κάποιας κατηγορικής μεταβλητής. Έτσι λοιπόν μπορούμε να θεωρήσουμε ότι ο συγκεκριμένος έλεγχος εξετάζει ισότητα μέσων τιμών μεταξύ δύο επιπέδων - τιμών μιας κατηγορικής μεταβλητής. Η στατιστική συνάρτηση που βασίζεται ο συγκεκριμένος έλεγχος εξαρτάται από την ισότητα των διακυμάνσεων των δύο ανεξάρτητων πληθυσμών.

1.5.5 Ο έλεγχος των Kolmogorov – Smirnov

Το κριτήριο των Kolmogorov – Smirnov χρησιμοποιεί τις πληροφορίες που παρέχονται από την εμπειρική ή δειγματική συνάρτηση κατανομής και βασίζεται στην αρχή ότι η παρατηρούμενη-δειγματική συνάρτηση κατανομής που υπολογίζεται από το δείγμα, θα προσεγγίζει την πραγματική κατανομή που αντιστοιχεί σε κάποιο πληθυσμό τον οποίο θέλουμε να ελέγξουμε και έχουμε κάνει τη δειγματοληψία μας. Έτσι, αν F είναι η συνάρτηση κατανομής του πληθυσμού και G η δειγματική συνάρτηση κατανομής, η μηδενική υπόθεση του ελέγχου διατυπώνεται ως εξής:

$$H_0 : F = G$$

δηλαδή, το δείγμα μας προέρχεται από συνεχή πληθυσμό με συνάρτηση κατανομής την F .

1.5.6 Η κλίμακα μεσογειακής διατροφής

Η κλίμακα μεσογειακής διατροφής είναι γνωστή από το 1970, και έγινε πρότυπο από τη μελέτη 7 χωρών. Γενικά σχετίζεται με τη μείωση της εμφάνισης χρόνιων νοσημάτων. Το σκορ μεσογειακής διατροφής προτάθηκε αρχικά από τον κ. Παναγιωτάκο το 2006. Ο δείκτης αυτός περιλαμβάνει τρεις ομάδες τροφίμων. Οι ομάδες αυτές βαθμολογήθηκαν χρησιμοποιώντας μονότομες ή μη μονότονες συναρτήσεις με τις τιμές τους να κυμαίνονται από 0 έως 5. Το συνολικό Σκορ (Med Diet Scor) λαμβάνει τιμές από 0 έως 55. Όσο μεγαλύτερες είναι οι τιμές του, τόσο περισσότερο βρίσκεται ότι το άτομο ακολουθεί τη Μεσογειακή διατροφή, η οποία περιλαμβάνει μη επεξεργασμένα δημητριακά, φρούτα, λαχανικά, όσπρια, ελαιόλαδο, ψάρια, κόκκινο κρέας, πουλερικά, πλήρη γαλακτοκομικά προϊόντα και μερική κατανάλωση αλκόολ.

1.5.7 Περιγραφή του δείγματος

Χρησιμοποιήθηκε δείγμα μεγέθους 1000 ατόμων που είναι μέρος της βάσης δεδομένων ΑΤΤΙΚΗ που καλύπτει την περίοδο 2001 έως 2006 και είναι μεγέθους 4016 ατόμων ηλικίας από 18 έως 81 ετών. Το δείγμα είναι σταθμισμένο ως προς το φύλο.

ΚΕΦΑΛΑΙΟ 2

Περιγραφική ανάλυση

2.1 Περιγραφική Ανάλυση Συνεχών Μεταβλητών

Στον πίνακα που ακολουθεί παρουσιάζονται αναλυτικά τα περιγραφικά μέτρα των συνεχών μεταβλητών που κρίθηκαν σημαντικές στην ανάλυση που ακολουθεί στα επόμενα κεφάλαια. Πιο συγκεκριμένα παραθέτουμε τη μέση τιμή της κάθε μεταβλητής και την τυπική της απόκλιση.

Πίνακας 2.1.1

Συγκεντρωτικός Πίνακας Συνεχών Μεταβλητών

| Μεταβλητές | $\mu \pm \sigma$ |
|-----------------------|----------------------|
| Βάρος | 75,58 \pm 16,027 |
| Ύψος | 169,45 \pm 9,091 |
| Δείκτης Μάζας Σώματος | 26,229 \pm 4,78530 |
| Περιφέρεια Μέσης | 90,25 \pm 16,273 |
| Περιφέρεια Γοφών | 104,22 \pm 11,294 |
| LDL - Χοληστερόλη | 48,70 \pm 16,04 |
| HDL - Χοληστερόλη | 121,60 \pm 35,909 |
| Συνολική Χοληστερόλη | 193,25 \pm 39,752 |

Παρατηρούμε ότι το πλήθος των ατόμων στο δείγμα μας έχουν φυσιολογικό βάρος και μέτριο ύψος.

2.2 Περιγραφική Ανάλυση Κατηγορικών Μεταβλητών

Δίνεται ο πίνακας συχνοτήτων για την μεταβλητή Φύλο καθώς και το αντίστοιχο διάγραμμα πίτας.

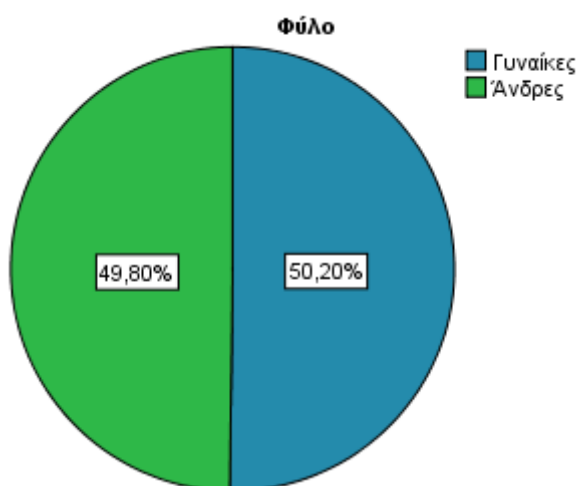
Πίνακας 2.2.1

Πίνακας Συχνοτήτων για την μεταβλητή φύλο

| | Συχνότητα | Ποσοστό | Αθροιστική Συχνότητα |
|----------|-----------|---------|----------------------|
| Γυναίκες | 502 | 50,2% | 50,2 |
| Άνδρες | 498 | 49,8% | 49,8 |
| Σύνολο | 1000 | 100,0% | 100,0 |

Διάγραμμα 2.1.1

Κυκλικό Διάγραμμα για το φύλο



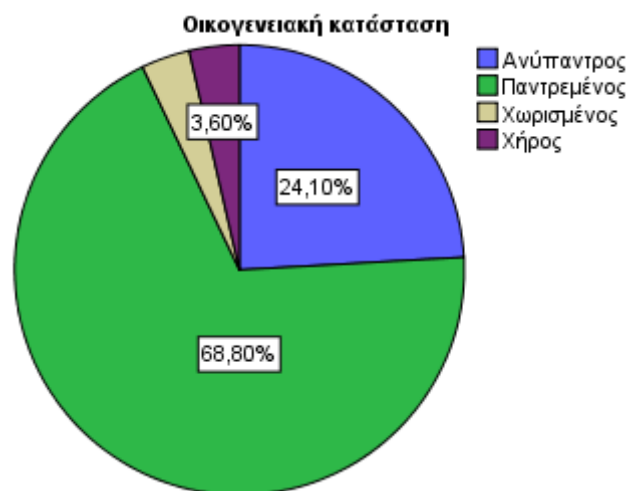
Παρατηρούμε ότι στο δείγμα μας οι άνδρες και οι γυναίκες καταλαμβάνουν σχεδόν το ίδιο ποσοστό. Πιο συγκεκριμένα οι άνδρες βρίσκονται σε ποσοστό 50,20% ενώ οι γυναίκες σε ποσοστό ίσο με 49,80%.

Όσον αφορά την οικογενειακή κατάσταση, παραθέτουμε τον πίνακα συχνοτήτων από όπου και προκύπτει ότι ποσοστό ίσο με 24.1% είναι ανύπαντροι και 68.8% είναι παντρεμένοι. Ακολουθεί και το αντίστοιχο διάγραμμα πίτας, όπου και επιβεβαιώνονται τα συμπεράσματα μας και διαγραμματικά.

Πίνακας 2.2.3
Πίνακας Συχνοτήτων για την μεταβλητή Οικογενειακή Κατάσταση

| | Συχνότητα | Ποσοστό | Αθροιστική Συχνότητα |
|-------------|-----------|---------|----------------------|
| Ανύπαντρος | 241 | 24,1% | 24,1 |
| Παντρεμένος | 688 | 68,8% | 68,8 |
| Χωρισμένος | 35 | 3,5% | 3,5 |
| Χήρος | 36 | 3,6% | 3,6 |
| Σύνολο | 1000 | 100,0% | 100,0 |

Διάγραμμα 2.2.2
Κυκλικό διάγραμμα για την οικογενειακή κατάσταση



Δίνεται επίσης ο πίνακας συχνοτήτων και το κυκλικό διάγραμμα για την μεταβλητή Ύπαρξη Χοληστερόλης.

Πίνακας 2.2.3
Πίνακας Συχνοτήτων για την μεταβλητή Ύπαρξη Χοληστερόλης

| | Συχνότητα | Ποσοστό | Αθροιστική Συχνότητα |
|--------------------|-----------|---------|----------------------|
| Ναι (≥ 220) | 308 | 78,4 | 78,4 |
| Όχι (< 220) | 85 | 21,6 | 100,0 |
| Σύνολο | 393 | 100,0 | |

Διάγραμμα 2.2.3
Κυκλικό διάγραμμα για την Ύπαρξη Χοληστερόλης



Παρατηρούμε ότι έχουμε ένα αρκετά μεγάλο ποσοστό ατόμων για τα οποία δεν έχουμε κάποια σαφή απάντηση. Επίσης όπως φαίνεται από τις απαντήσεις το 30.8% των ατόμων πάσχουν από υψηλή χοληστερόλη.

ΚΕΦΑΛΑΙΟ 3

Έλεγχος Κατανομών- Συσχετίσεις – Έλεγχος Ανεξαρτησίας

3.1 Εισαγωγή

Στο κεφάλαιο αυτό αρχικά θα παρουσιαστεί ο συντελεστής συσχέτισης του Pearson για κάθε μία από τις συνεχείς μεταβλητές Ολική Χοληστερόλη, LDL Χοληστερόλη, με όλες τις συνεχείς μεταβλητές που διαθέτουμε στο δείγμα μας. Με τον τρόπο αυτό γίνεται μία πρώτη διερεύνηση για το ποιες μεταβλητές σχετίζονται θετικά ή αρνητικά με τις μεταβλητές αυτές. Αυτό θα μας βοηθήσει στο να επιλέξουμε σε επόμενο κεφάλαιο το βέλτιστο μοντέλο για αυτές. Παράλληλα παρουσιάζονται τα αντίστοιχα διαγράμματα διασποράς τα οποία και επιβεβαιώνουν τα συμπεράσματά μας αλλά μας δίνουν και μία οπτική εικόνα.

Επίσης, στο τέλος του κεφαλαίου αυτού, θα παρουσιάσουμε τον έλεγχο ανεξαρτησίας χ^2 – τετράγωνο ανάμεσα στην δίτιμη Ύπαρξη Χοληστερόλης με όλες τις κατηγορικές μεταβλητές. Και εδώ τα αποτελέσματα θα μας δώσουν μία πρώτη εικόνα για το ποιες μεταβλητές εξαρτώνται ή όχι με την Ύπαρξη Χοληστερόλης και έτσι θα μπορέσουμε να διατυπώσουμε τα αρχικά μας συμπεράσματα.

Πριν όμως από τους ανωτέρω ελέγχους παραθέτουμε και τους ελέγχους κανονικότητας για τις μεταβλητές αυτές. Από τον πίνακα που ακολουθεί παρατηρούμε ότι η μηδενική υπόθεση απορρίπτεται για την HDL – Χοληστερόλη και για τα Τριγλυκερίδια. Τα αποτελέσματα μας φαίνονται και γραφικά στα αντίστοιχα διαγράμματα συχνότητας.

Πίνακας 3.1.1

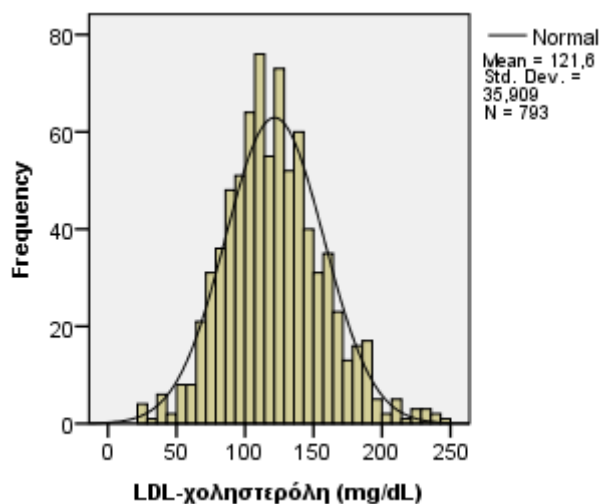
Έλεγχος Κανονικότητας για τις μεταβλητές LDL Χοληστερόλη, HDL – Χοληστερόλη, Συνολική Χοληστερόλη και για τα Τριγλυκερίδια

| Hypothesis Test Summary | | | | |
|-------------------------|-----------------------------------------------------------------------------------------------------------|------------------------------------|------|-----------------------------|
| | Null Hypothesis | Test | Sig. | Decision |
| 1 | The distribution of LDL-χοληστερόλη (mg/dL) is normal with mean 121,6 and standard deviation 35,91. | One-Sample Kolmogorov-Smirnov Test | ,169 | Retain the null hypothesis. |
| 2 | The distribution of HDL-χοληστερόλη (mg/dl) is normal with mean 48,70 and standard deviation 16,04. | One-Sample Kolmogorov-Smirnov Test | ,000 | Reject the null hypothesis. |
| 3 | The distribution of Συνολική Χοληστερίνη (mg/dl) is normal with mean 193,25 and standard deviation 39,75. | One-Sample Kolmogorov-Smirnov Test | ,191 | Retain the null hypothesis. |
| 4 | The distribution of Τριγλυκερίδια (mg/dl) is normal with mean 120,64 and standard deviation 103,14. | One-Sample Kolmogorov-Smirnov Test | ,000 | Reject the null hypothesis. |

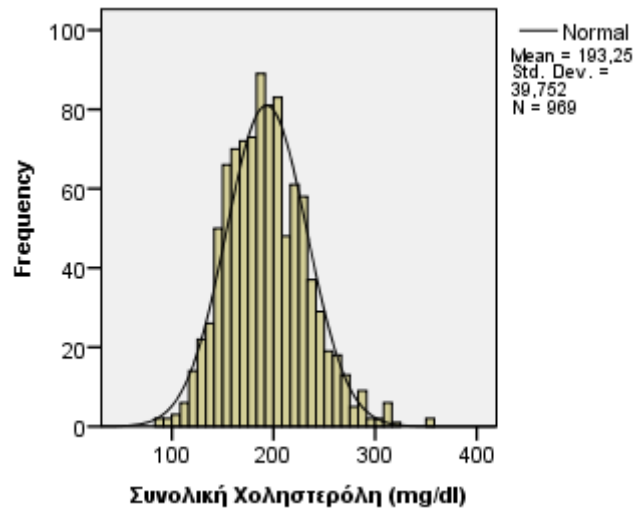
Asymptotic significances are displayed. The significance level is ,05.

Διάγραμμα 3.1.1

Ιστόγραμμα συχνοτήτων για την μεταβλητή LDL Χοληστερόλη



Διάγραμμα 3.1.1
Ιστόγραμμα συχνοτήτων για την μεταβλητή Συνολική Χοληστερόλη



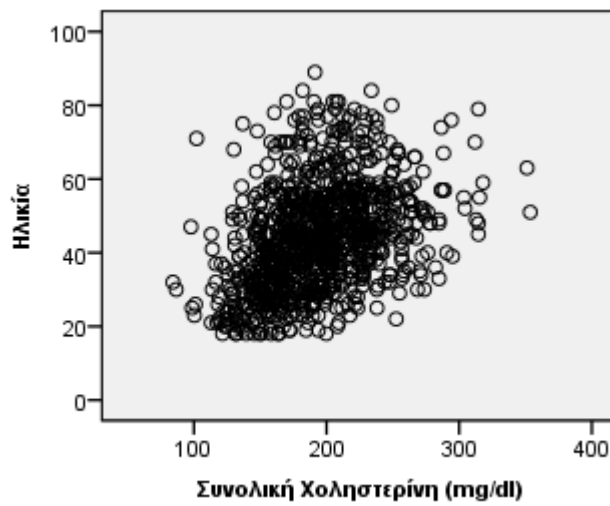
3.2 Συντελεστής συσχέτισης του Pearson για την μεταβλητή Συνολική Χοληστερόλη και τα αντίστοιχα διαγράμματα διασποράς

Θα παρουσιάσουμε τα διαγράμματα για τις μεταβλητές όπου βρέθηκε ότι υπάρχει είτε θετική είτε αρνητική συσχέτιση με την μεταβλητή Ολική Χοληστερόλη και στο τέλος θα παρουσιάσουμε δύο συγκεντρωτικούς πίνακες. Ο πρώτος θα περιλαμβάνει τις μεταβλητές όπου βρέθηκε ότι υπάρχει συσχέτιση, ενώ ο δεύτερος αυτές όπου βρέθηκε ότι δεν υπάρχει συσχέτιση. Όλοι οι έλεγχοι έχουν γίνει σε επίπεδο σημαντικότητας $\alpha=5\%$.

Αρχικά παρατηρούμε ότι υπάρχει θετική συσχέτιση της συνολικής χοληστερόλης με την ηλικία, το βάρος και με τον δείκτη μάζας σώματος, όπως φαίνεται γραφικά.

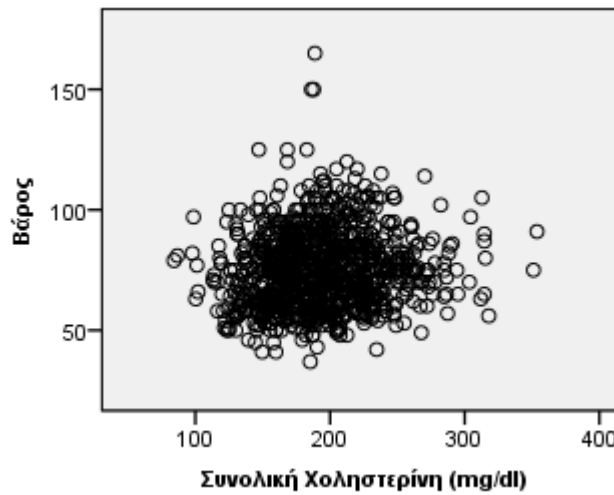
Διάγραμμα 3.2.1

Διάγραμμα διασποράς για τη συνολική χοληστερόλη και την ηλικία



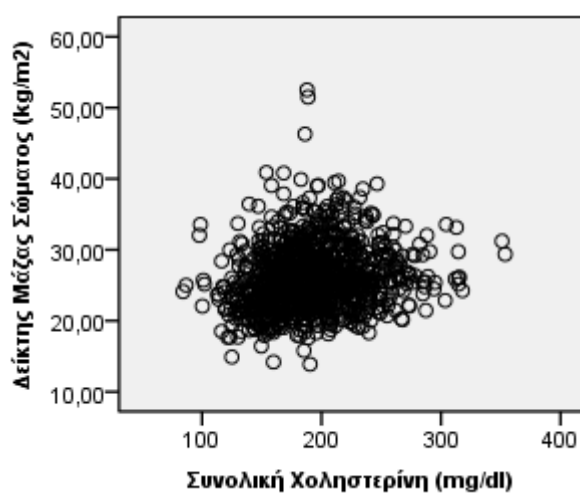
Διάγραμμα 3.2.2

Διάγραμμα διασποράς για τη συνολική χοληστερόλη και το βάρος



Διάγραμμα 3.2.3

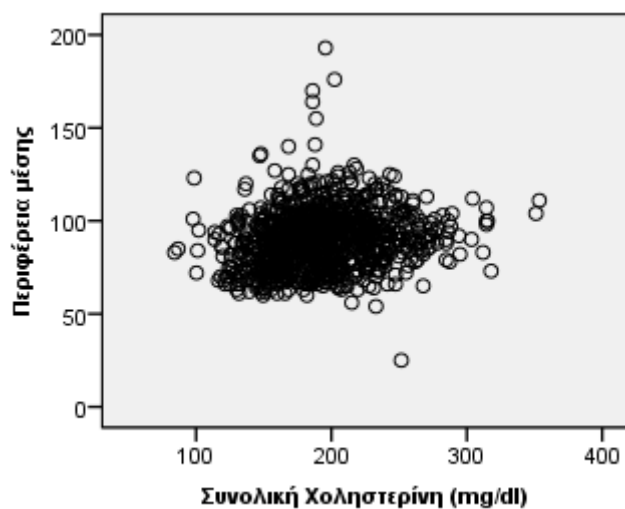
Διάγραμμα διασποράς για τη συνολική χοληστερόλη και τον δείκτη μάζας σώματος



Στη συνέχεια παρατηρούμε ότι η συνολική χοληστερίνη έχει επίσης θετική συσχέτιση με τις μεταβλητές περιφέρεια μέσης και γοφών. Παρουσιάζονται τα αντίστοιχα διαγράμματα.

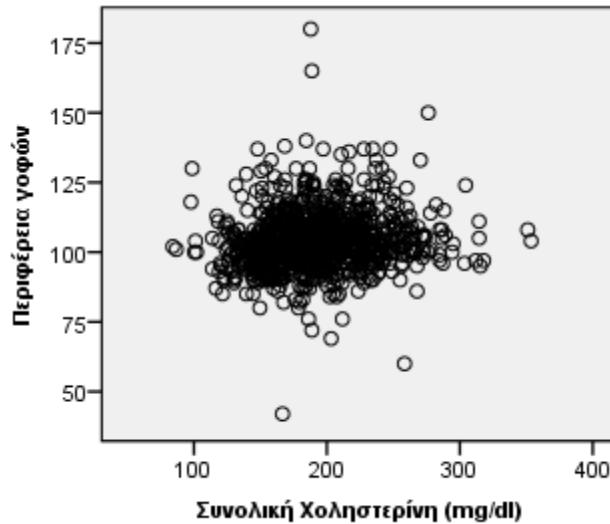
Διάγραμμα 3.2.4

Διάγραμμα διασποράς για τη συνολική χοληστερόλη και την περιφέρεια μέσης



Διάγραμμα 3.2.5

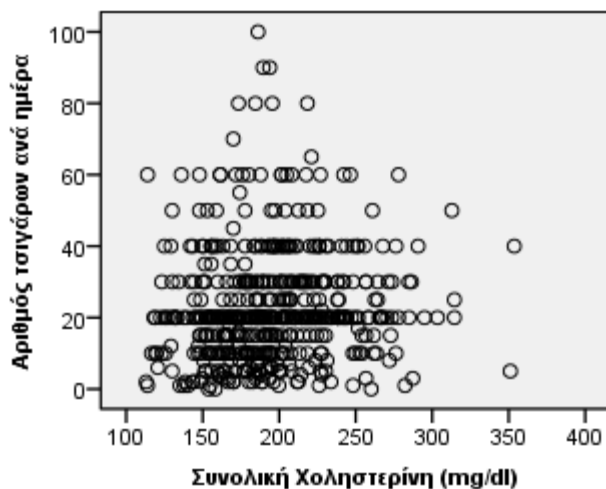
Διάγραμμα διασποράς για τη συνολική χοληστερόλη και την περιφέρεια γοφών



Θετική συσχέτιση υπάρχει και ανάμεσα στις μεταβλητές συνολική χοληστερίνη και αριθμός τσιγάρων ανά ημέρα και αριθμός ετών καπνίσματος, όπως μπορούμε να διακρίνουμε στα αντίστοιχα διαγράμματα.

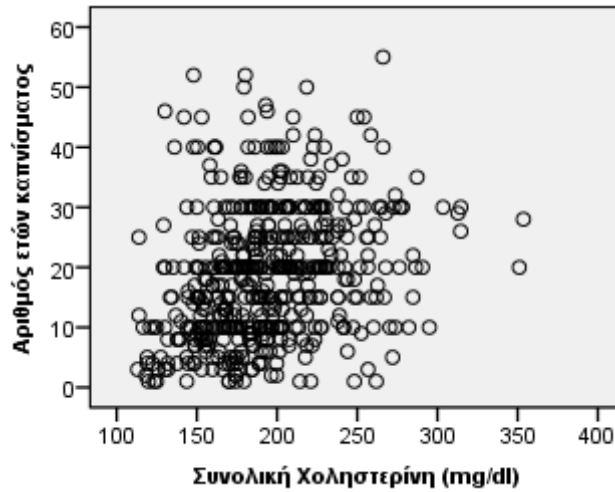
Διάγραμμα 3.2.6

Διάγραμμα διασποράς για τη συνολική χοληστερόλη και τον αριθμό τσιγάρων ανά ημέρα



Διάγραμμα 3.2.7

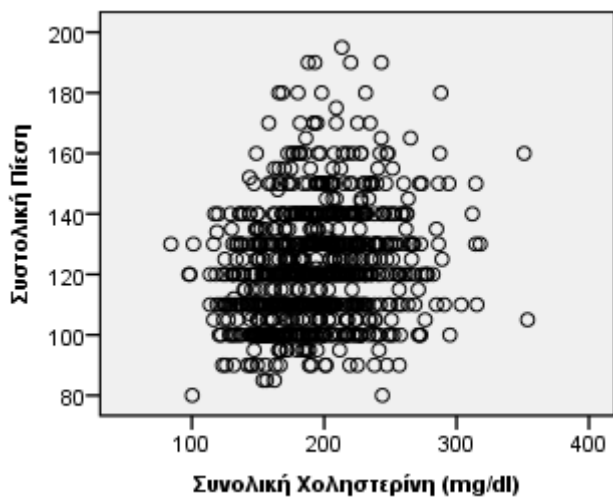
Διάγραμμα διασποράς για τη συνολική χοληστερόλη και τον αριθμό ετών καπνίσματος



Όσον αφορά τη συστολική και τη διαστολική πίεση, υπάρχει θετική συσχέτιση, κάτι το οποίο ισχύει και για το επίπεδο γλυκόζης στο αίμα.

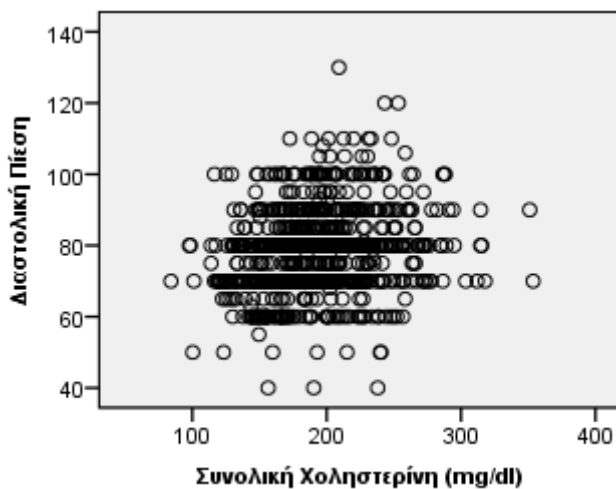
Διάγραμμα 3.2.8

Διάγραμμα διασποράς για τη συνολική χοληστερόλη και τη συστολική πίεση



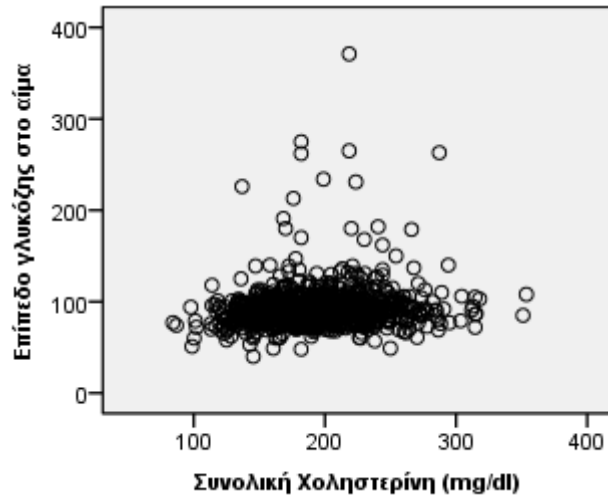
Διάγραμμα 3.2.9

Διάγραμμα διασποράς για τη συνολική χοληστερόλη και τη διαστολική πίεση



Διάγραμμα 3.2.10

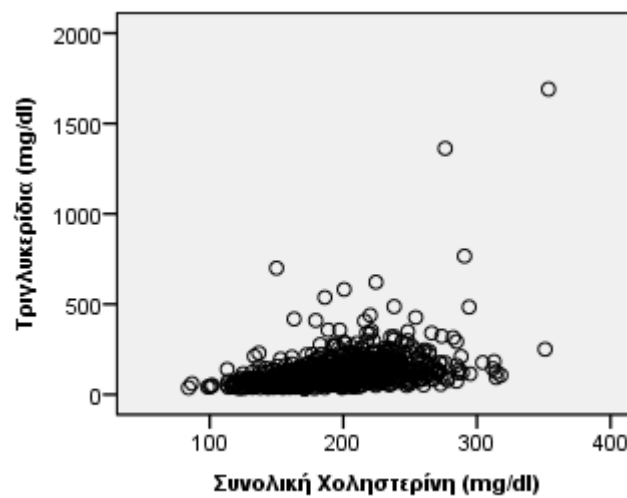
Διάγραμμα διασποράς για τη συνολική χοληστερόλη και το επίπεδο γλυκόζης στο αίμα



Όπως είναι αναμενόμενο, θετική συσχέτιση βρέθηκε και ανάμεσα στην συνολική χοληστερίνη και στα τριγλυκερίδια καθώς και στην LDL χοληστερόλη.

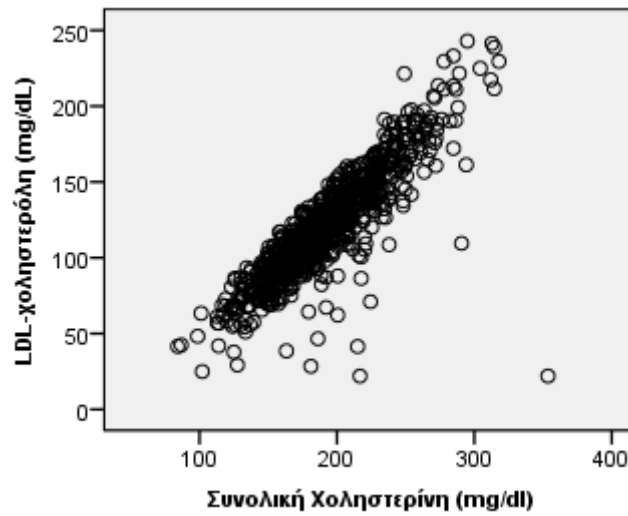
Διάγραμμα 3.2.11

Διάγραμμα διασποράς για τη συνολική χοληστερόλη και τα τριγλυκερίδια



Διάγραμμα 3.2.12

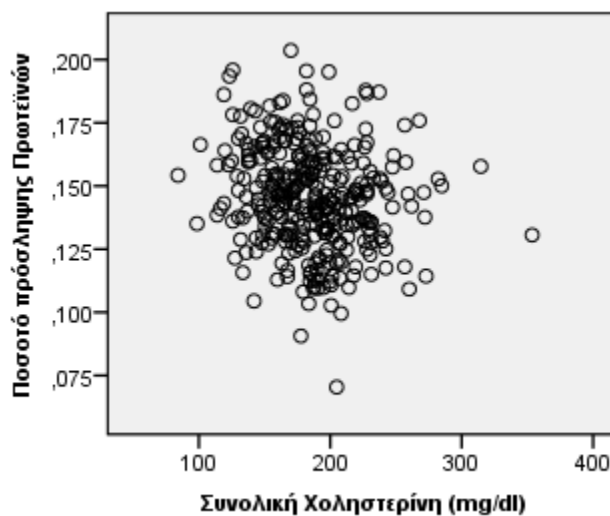
Διάγραμμα διασποράς για τη συνολική χοληστερόλη και την LDL χοληστερόλη



Όσον αφορά το ποσοστό πρόσληψης πρωτεϊνών, κορεσμένων λιπαρών οξέων και το σκορ μεσογειακής διατροφής, βρέθηκε ότι υπάρχει αρνητική συσχέτιση με τη συνολική χοληστερίνη.

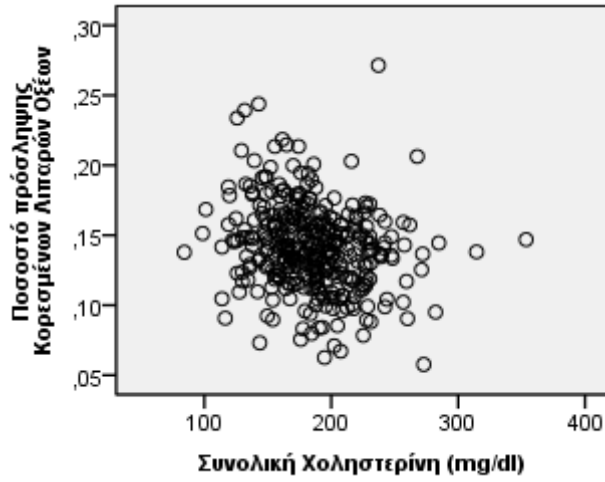
Διάγραμμα 3.2.13

Διάγραμμα διασποράς για τη συνολική χοληστερόλη και το ποσοστό πρόσληψης πρωτεϊνών



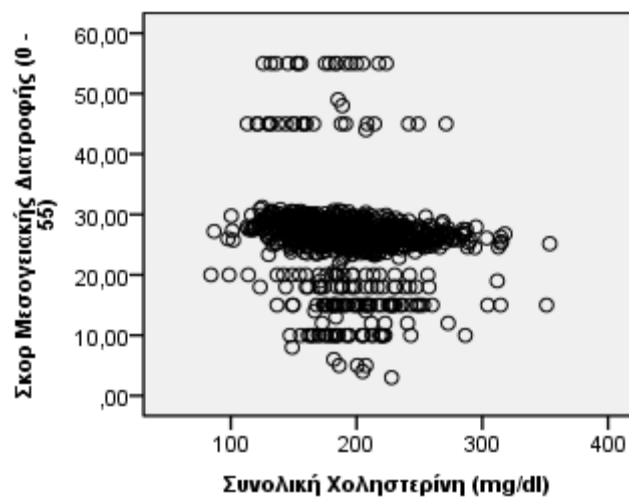
Διάγραμμα 3.2.14

Διάγραμμα διασποράς για τη συνολική χοληστερόλη και το ποσοστό πρόσληψης κορεσμένων λιπαρών οξέων



Διάγραμμα 3.2.15

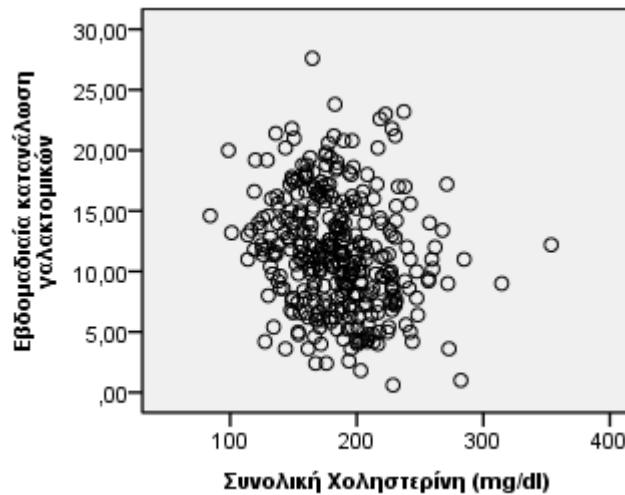
Διάγραμμα διασποράς για τη συνολική χοληστερόλη και το σκορ μεσογειακής διατροφής



Όσον αφορά την εβδομαδιαία κατανάλωση συγκεκριμένων τροφών, βρέθηκε ότι υπάρχει αρνητική συσχέτιση στην κατανάλωση γαλακτοκομικών, γλυκών, αυγών καθώς και στην συνολική κατανάλωση κρέατος.

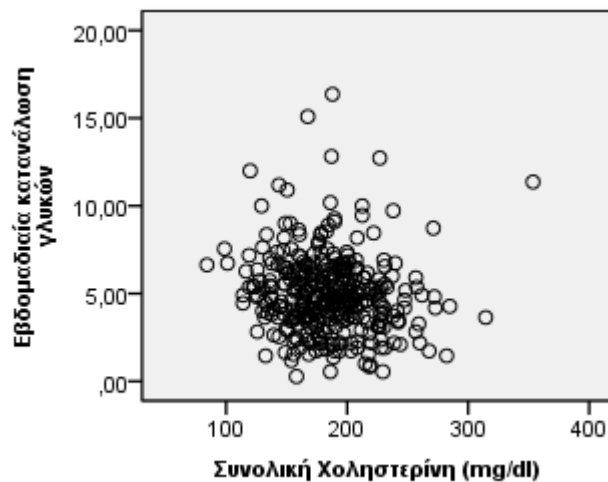
Διάγραμμα 3.2.16

Διάγραμμα διασποράς για τη συνολική χοληστερόλη και την εβδομαδιαία κατανάλωση γαλακτοκομικών



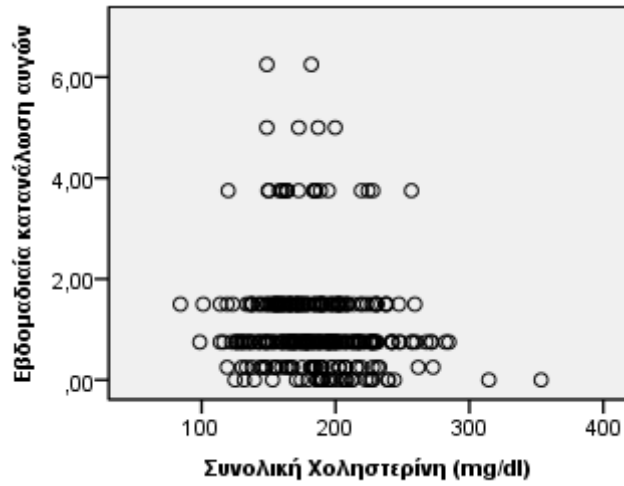
Διάγραμμα 3.2.17

Διάγραμμα διασποράς για τη συνολική χοληστερόλη και την εβδομαδιαία κατανάλωση γλυκών



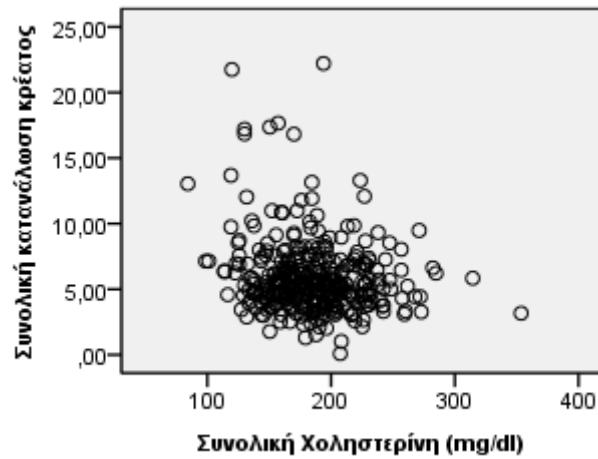
Διάγραμμα 3.2.18

Διάγραμμα διασποράς για τη συνολική χοληστερόλη και την εβδομαδιαία κατανάλωση αυγών



Διάγραμμα 3.2.19

Διάγραμμα διασποράς για τη συνολική χοληστερόλη και τη συνολική κατανάλωση κρέατος



Προκειμένου να παρουσιάσουμε συγκεντρωτικά τα αποτελέσματα μας, παρουσιάζουμε τον ακόλουθο πίνακα. Πιο συγκεκριμένα δίνεται ο πίνακας για τις μεταβλητές εκείνες όπου και υπάρχει συσχέτιση.

Πίνακας 3.2.1

Συγκεντρωτικός πίνακας για την μεταβλητή Συνολική Χοληστερόλη

| Μεταβλητή | Συντελεστής συσχέτισης | |
|--------------------------------------------------|------------------------|---------|
| | Pearson | P-Value |
| Ηλικία | 0,380 | 0,000 |
| Βάρος | 0,106 | 0,001 |
| Δείκτης μάζας σώματος | 0,159 | 0,000 |
| Περιφέρεια μέσης | 0,168 | 0,000 |
| Περιφέρεια γοφών | 0,112 | 0,001 |
| Αριθμός τσιγάρων ανά ημέρα | 0,088 | 0,044 |
| Αριθμός ετών καπνίσματος | 0,270 | 0,000 |
| Συστολική πίεση | 0,210 | 0,000 |
| Διαστολική πίεση | 0,202 | 0,000 |
| Επίπεδο γλυκόζης στο αίμα | 0,121 | 0,000 |
| Τριγλυκερίδια | 0,397 | 0,000 |
| LDL-Χοληστερόλη | 0,866 | 0,000 |
| Ποσοστό πρόσληψης πρωτεϊνών | -0,176 | 0,001 |
| Ποσοστό πρόσληψης κορεσμένων λιπαρών οξέων | -0,215 | 0,000 |

| | | |
|----------------------------------------|--------|-------|
| Σκορ μεσογειακής διατροφής | -0,175 | 0,000 |
| Εβδομαδιαία κατανάλωση γαλακτομικών | -0,201 | 0,000 |
| Εβδομαδιαία κατανάλωση γλυκών | -0,113 | 0,037 |
| Εβδομαδιαία κατανάλωση αυγών | -0,122 | 0,024 |
| Συνολική κατανάλωση κρέατος | -0,186 | 0,001 |

3.3 Συντελεστής συσχέτισης του Pearson για τη μεταβλητή LDL Χοληστερόλη

Όσον αφορά τη μεταβλητή LDL Χοληστερόλη, θα παρουσιάσουμε συνοπτικά σε δύο πίνακες τις μεταβλητές με τις οποίες εμφανίζεται θετική ή αρνητική συσχέτιση και αυτές που εμφανίζεται συσχέτιση.

Πίνακας 3.3.1

Συγκεντρωτικός πίνακας για την μεταβλητή LDL - Χοληστερόλη

| Μεταβλητή | Συντελεστής Συσχέτισης Pearson | P-Value |
|----------------------------|--------------------------------------|---------|
| Ηλικία | 0,297 | 0,000 |
| Βάρος | 0,123 | 0,001 |
| Δείκτης μάζας σώματος | 0,161 | 0,000 |
| Περιφέρεια μέσης | 0,166 | 0,000 |
| Περιφέρεια γοφών | 0,112 | 0,002 |
| Αριθμός τσιγάρων ανά ημέρα | 0,107 | 0,025 |
| Αριθμός ετών καπνίσματος | 0,224 | 0,000 |

| | | |
|-----------------------------------------------|--------|-------|
| Συστολική πίεση | 0,176 | 0,000 |
| Διαστολική πίεση | 0,192 | 0,000 |
| HDL-Χοληστερόλη | -0,215 | 0,000 |
| Πρωτεΐνη C | 0,074 | 0,044 |
| Ποσοστό πρόσληψης κορεσμένων λιπαρών οξέων | -0,151 | 0,005 |
| Σκορ μεσογειακής διατροφής | -0,142 | 0,000 |
| Εβδομαδιαία κατανάλωση γαλακτοκομικών | -0,180 | 0,001 |
| Εβδομαδιαία κατανάλωση γλυκών | -0,164 | 0,002 |
| Συνολική κατανάλωση κρέατος | -0,133 | 0,014 |

Πίνακας 3.3.2

Συγκεντρωτικός πίνακας για την μεταβλητή LDL - Χοληστερόλη

| Μεταβλητή | Συντελεστής Συσχέτισης Pearson | P-Value |
|--------------------------------------|-----------------------------------------------|----------------|
| Ύψος | -0,023 | 0,529 |
| Επίπεδο γλυκόζης στο αίμα | 0,052 | 0,146 |
| Τριγλυκερίδια | 0,053 | 0,134 |
| Συνολική αντιοξειδωτική ικανότητα | -0,008 | 0,917 |
| Κλίμακα κατάθλιψης Zung | -0,018 | 0,768 |
| Κλίμακα άγχους Spielberg | -0,073 | 0,250 |

| | | |
|--------------------------------------------------|--------|-------|
| Πρόσληψη ενέργειας | -0,059 | 0,265 |
| Ποσοστό πρόσληψης υδατανθράκων | 0,005 | 0,932 |
| Ποσοστό πρόσληψης πρωτεϊνών | -0,089 | 0,103 |
| Ποσοστό πρόσληψης λίπους | -0,022 | 0,680 |
| Ποσοστό πρόσληψης μονοακόρεστων λιπαρών οξέων | 0,030 | 0,586 |
| Ποσοστό πρόσληψης πολυακόρεστων λιπαρών οξέων | 0,076 | 0,164 |
| Εβδομαδιαία κατανάλωση ψαριού | -0,012 | 0,830 |
| Εβδομαδιαία κατανάλωση ξηρών καρπών | 0,073 | 0,180 |
| Εβδομαδιαία κατανάλωση λαχανικών | 0,027 | 0,622 |
| Εβδομαδιαία κατανάλωση φρούτων | 0,018 | 0,748 |
| Εβδομαδιαία κατανάλωση πατάτας | -0,095 | 0,082 |
| Εβδομαδιαία κατανάλωση δημητριακών | -0,073 | 0,178 |
| Εβδομαδιαία κατανάλωση κρέατος | -0,004 | 0,945 |
| Εβδομαδιαία κατανάλωση αυγών | -0,094 | 0,084 |
| Εβδομαδιαία κατανάλωση πουλερικών | -0,087 | 0,111 |
| Εβδομαδιαία κατανάλωση αναψυκτικών | -0,09 | 0,864 |

3.4 Έλεγχος ανεξαρτησίας χι – τετράγωνο για τη μεταβλητή Ύπαρξη Χοληστερόλης

Κλείνοντας το κεφάλαιο αυτό, δίνονται δύο πίνακες όπου και δίνονται οι έλεγχοι ανεξαρτησίας ανάμεσα στη μεταβλητή Ύπαρξη Χοληστερόλης και σε όλες τις κατηγορικές μεταβλητές που διαθέτουμε. Στον πρώτο πίνακα παρουσιάζονται οι μεταβλητές για τις οποίες υπάρχει εξάρτηση, ενώ στον δεύτερο αυτές όπου δεν υπάρχει εξάρτηση σύμφωνα με τον έλεγχο χι – τετράγωνο.

Πίνακας 3.6.1

Συγκεντρωτικός πίνακας για τη μεταβλητή Ύπαρξη Χοληστερόλης

| Μεταβλητή | Τιμή Pearson Chi-Square | P-Value |
|----------------------------------------------|-------------------------|---------|
| Φύλο | 8,175 | 0,004 |
| Εκπαίδευση | 8,636 | 0,013 |
| Οικογενειακή κατάσταση | 20,262 | 0,000 |
| Κοινωνικο-οικονομικό επίπεδο | 20,903 | 0,007 |
| Σακχαρώδης Διαβήτης | 18,304 | 0,000 |
| Υπερχοληστερολαιμία | 35,086 | 0,000 |
| Μεσογειακή διατροφή | 28,026 | 0,000 |
| Καρδιαγγειακό επεισόδιο από 2001 έως 2006 | 23,385 | 0,000 |
| Περιστατικό με υψηλή πίεση | 20,021 | 0,000 |

Πίνακας 3.6.2

Συγκεντρωτικός πίνακας για τη μεταβλητή Ύπαρξη Χοληστερόλης

| Μεταβλητή | Τιμή Pearson Chi-Square | P-Value |
|------------------------|-------------------------|---------|
| Επίπεδο δραστηριότητας | 7,037 | 0,071 |
| Καπνιστής | 2,585 | 0,108 |
| Έκθεση στο παθητικό | 0,671 | 0,413 |

| | | |
|--------------------------------|-------|-------|
| κάπνισμα | | |
| Ύπαρξη υπέρτασης | 1,059 | 0,303 |
| Καθημερινή χρήση ελαιόλαδου | 1,062 | 0,303 |
| Καθημερινή χρήση σπορέλαιου | 0,007 | 0,932 |
| Καθημερινή χρήση βουτύρου | 0,648 | 0,421 |
| Καθημερινή χρήση μαργαρίνης | 0,491 | 0,484 |
| Κατανάλωση καφέ | 2,593 | 0,107 |
| Ποσότητα κατανάλωσης καφέ | 3,926 | 0,270 |
| Υψηλό ζάχαρο | 1,028 | 0,311 |

Στη συνέχεια εξετάζουμε αρχικά μέσω του t-test εάν η μέση τιμή της μεταβλητής Συνολικής Χοληστερόλης και LDL Χοληστερόλης διαφέρει ή όχι για όλες τις δίτιμες μεταβλητές που έχουμε στη διάθεσή μας. Κάνουμε τον έλεγχο αυτό για τις μεταβλητές LDL Χοληστερόλη, γιατί όπως προαναφέρθηκε αυτές υπάρχουν στον τύπο υπολογισμού της Ολικής Χοληστερόλης και θεωρούμε χρήσιμο το να βρεθούν τόσο οι παράγοντες που επηρεάζουν την Ολική Χοληστερόλη όσο και οι παράγοντες που επηρεάζουν το βασικό συστατικό της. Πιο συγκεκριμένα ελέγχουμε σε επίπεδο σημαντικότητας $\alpha=5\%$ τις υποθέσεις:

$$H_0 : \mu_1 = \mu_2 \text{ vs } H_1 : \mu_1 \neq \mu_2$$

3.5 T-Test για την μεταβλητή Συνολική Χοληστερόλη

Ακολουθεί ο συγκεντρωτικός πίνακας με τις μεταβλητές για τις οποίες απορρίπτουμε τη μηδενική υπόθεση και συνεπώς η μέση τιμή της Ολικής Χοληστερόλης διαφέρει στις δύο ομάδες ατόμων που δημιουργούνται από αυτές.

Πίνακας 3.7.1

Συγκεντρωτικός πίνακας για τη μεταβλητή Συνολική Χοληστερόλη

| Μεταβλητή | Τιμή στατιστικής συνάρτησης T | P-Value |
|----------------------------------------------|----------------------------------|---------|
| Ύπαρξη υπέρτασης | 5,649 | 0,000 |
| Ύπαρξη σακχαρώδη διαβήτη | 2,445 | 0,015 |
| Ύπαρξη υπερχοληστερολαιμίας | 35,555 | 0,000 |
| Καθημερινή χρήση βουτύρου | 2,092 | 0,037 |
| Κατανάλωση καφέ | 4,869 | 0,000 |
| Καρδιαγγειακό επεισόδιο από 2001 έως 2006 | 2,965 | 0,003 |

Πίνακας 3.7.2

Πίνακας μέσω των τιμών και αποκλίσεων ανά κατηγορία

| Μεταβλητή | Κατηγορία | $\mu \pm \sigma$ |
|--------------------------------|-----------------|------------------|
| Ύπαρξη υπέρτασης | Ναι (>14 ή >19) | 204,41±38,203 |
| | Όχι | 188±38,631 |
| Ύπαρξη σακχαρώδη διαβήτη | Ναι | 208,33±37,547 |
| | Όχι | 192,32±39,076 |
| Ύπαρξη υπερχοληστερολαιμίας | Ναι | 228,49±29,358 |
| | Όχι | 167,18±22,441 |
| Καθημερινή χρήση βουτύρου | Ναι | 191,73±39,585 |
| | Όχι | 197,96±39,765 |
| Κατανάλωση καφέ | Ναι | 190,74±38,511 |

| | | |
|----------------------------------------------------|-----|---------------|
| Καρδιαγγειακό επεισόδιο από το 2001 έως το 2006 | Όχι | 208,43±43,104 |
| | Ναι | 209,50±42,263 |
| | Όχι | 192,44±39,378 |

3.6 T-Test για την μεταβλητή LDL Χοληστερόλη

Συγκεντρωτικά τα αποτελέσματα φαίνονται στους δύο πίνακες που ακολουθούν, όπου στον πρώτο πίνακα συγκεντρώνονται οι περιπτώσεις όπου απορρίπτουμε τη μηδενική υπόθεση ενώ στον δεύτερο πίνακα φαίνονται οι περιπτώσεις όπου δεν απορρίπτουμε τη μηδενική υπόθεση.

Πίνακας 3.8.1

Συγκεντρωτικός πίνακας για τη μεταβλητή LDL - Χοληστερόλη

| Μεταβλητή | Τιμή στατιστικής συνάρτησης T | P-Value |
|--------------------------------|----------------------------------|---------|
| Ύπαρξη υπέρτασης | 4,551 | 0,000 |
| Ύπαρξη υπερχοληστερολαιμίας | 23,716 | 0,000 |
| Κατανάλωση καφέ | 4,390 | 0,000 |

Πίνακας 3.8.2

Πίνακας μέσω των τιμών και τυπικών αποκλίσεων ανά κατηγορία

| Μεταβλητή | Κατηγορία | μ±σ |
|------------------|-----------------|---------------|
| Ύπαρξη υπέρτασης | Ναι (>14 ή >19) | 130,89±33,436 |
| | Όχι | 117,93±34,836 |

| | | |
|--------------------------------|-----|---------------|
| Ύπαρξη υπερχοληστερολαιμίας | Ναι | 148,88±30,205 |
| | Όχι | 101,52±22,066 |
| Κατανάλωση καφέ | Ναι | 119,69±33,963 |
| | Όχι | 135,88±37,848 |

Πίνακας 3.8.3

Συγκεντρωτικός πίνακας για τη μεταβλητή LDL - Χοληστερόλη

| Μεταβλητή | Τιμή στατιστικής συνάρτησης T | P-Value |
|----------------------------------------------|----------------------------------|---------|
| Ύπαρξη σακχαρώδη διαβήτη | -0,773 | 0,440 |
| Φύλο | 1,913 | 0,056 |
| Καπνιστής | 0,757 | 0,449 |
| Έκθεση στο παθητικό κάπνισμα | 0,367 | 0,714 |
| Καθημερινή χρήση ελαιόλαδου | -0,995 | 0,320 |
| Καθημερινή χρήση σπορέλαιου | 1,456 | 0,146 |
| Καθημερινή χρήση βουτύρου | 0,965 | 0,335 |
| Καθημερινή χρήση μαργαρίνης | 0,025 | 0,980 |
| Καρδιαγγειακό επεισόδιο από 2001 έως 2006 | 1,417 | 0,157 |
| Περιστατικό με υψηλή πίεση | 1,253 | 0,211 |
| Υψηλό ζάχαρο | 1,218 | 0,224 |

3.7 Μη παραμετρικός έλεγχος για την μεταβλητή Συνολική Χοληστερόλη

Προκειμένου να επιβεβαιώσουμε τα αποτελέσματα μας, παρουσιάζουμε και τον μη παραμετρικό έλεγχο για την μεταβλητή Συνολική Χοληστερόλη στον ακόλουθο πίνακα.

Πίνακας 3.9.1

Συγκεντρωτικός πίνακας για τη μεταβλητή Συνολική Χοληστερόλη

| Κατηγορία | Τιμή στατιστικής συνάρτησης S-W | P - Value |
|------------------------------------|--------------------------------------------|------------------|
| Άνδρες | 0,993 | 0,021 |
| Γυναίκες | 0,982 | 0,000 |
| Καπνιστής | 0,976 | 0,000 |
| Μη καπνιστής | 0,992 | 0,007 |
| Έκθεση στο παθητικό κάπνισμα | 0,989 | 0,001 |
| Μη έκθεση στο παθητικό κάπνισμα | 0,962 | 0,002 |
| Ύπαρξη υπέρτασης (> 14 ή > 9) | 0,984 | 0,004 |
| Μη ύπαρξη υπέρτασης | 0,987 | 0,000 |
| Ύπαρξη σακχαρώδη διαβήτη | 0,988 | 0,717 |
| Μη ύπαρξη σακχαρώδη διαβήτη | 0,986 | 0,000 |
| Καθημερινή χρήση ελαιόλαδου | 0,989 | 0,000 |
| Μη καθημερινή χρήση ελαιόλαδου | 0,971 | 0,137 |
| Καθημερινή χρήση σπορέλαιου | 0,975 | 0,037 |
| Μη καθημερινή χρήση σπορέλαιου | 0,990 | 0,000 |
| Καθημερινή χρήση βουτύρου | 0,988 | 0,000 |

| | | |
|-----------------------------------|-------|-------|
| Μη καθημερινή χρήση βουτύρου | 0,988 | 0,033 |
| Καθημερινή χρήση μαργαρίνης | 0,990 | 0,051 |
| Μη καθημερινή χρήση μαργαρίνης | 0,990 | 0,152 |
| Κατανάλωση καφέ | 0,990 | 0,000 |
| Μη κατανάλωση καφέ | 0,990 | 0,400 |
| Καρδιαγγειακό επεισόδιο | 0,945 | 0,019 |
| Μη καρδιαγγειακό επεισόδιο | 0,990 | 0,000 |
| Περιστατικό με υψηλή πίεση | 0,978 | 0,233 |
| Μη περιστατικό με υψηλή πίεση | 0,988 | 0,000 |
| Υψηλό ζάχαρο (< 120) | 0,954 | 0,129 |
| Μη υψηλό ζάχαρο (≥ 120) | 0,989 | 0,000 |

3.8 Μη παραμετρικός έλεγχος για την μεταβλητή LDL – Χοληστερόλη

Προκειμένου να επιβεβαιώσουμε τα αποτελέσματά μας, παρουσιάζουμε και τον μη παραμετρικό έλεγχο για την μεταβλητή LDL - Χοληστερόλη στον ακόλουθο πίνακα.

Πίνακας 3.10.1

Συγκεντρωτικός πίνακας για τη μεταβλητή LDL - Χοληστερόλη

| Κατηγορία | Τιμή στατιστικής συνάρτησης S-W | P - Value |
|---------------------------------|------------------------------------|-----------|
| Άνδρες | 0,993 | 0,098 |
| Γυναίκες | 0,981 | 0,000 |
| Καπνιστής | 0,981 | 0,054 |
| Μη καπνιστής | 0,994 | 0,000 |
| Έκθεση στο παθητικό κάπνισμα | 0,007 | 0,012 |

| | | |
|---------------------------------|-------|-------|
| Μη έκθεση στο παθητικό κάπνισμα | 0,059 | 0,001 |
| Ύπαρξη υπέρτασης (> 14 ή > 9) | 0,992 | 0,299 |
| Μη ύπαρξη υπέρτασης | 0,990 | 0,002 |
| Ύπαρξη σακχαρώδη διαβήτη | 0,980 | 0,580 |
| Μη ύπαρξη σακχαρώδη διαβήτη | 0,990 | 0,000 |
| Καθημερινή χρήση ελαιόλαδου | 0,990 | 0,000 |
| Μη καθημερινή χρήση ελαιόλαδου | 0,975 | 0,421 |
| Καθημερινή χρήση σπορέλαιου | 0,982 | 0,215 |
| Μη καθημερινή χρήση σπορέλαιου | 0,993 | 0,011 |
| Καθημερινή χρήση βουτύρου | 0,989 | 0,000 |
| Μη καθημερινή χρήση βουτύρου | 0,987 | 0,181 |
| Καθημερινή χρήση μαργαρίνης | 0,994 | 0,181 |
| Μη καθημερινή χρήση μαργαρίνης | 0,988 | 0,712 |
| Κατανάλωση καφέ | 0,993 | 0,003 |
| Μη κατανάλωση καφέ | 0,972 | 0,034 |
| Καρδιαγγειακό επεισόδιο | 0,983 | 0,796 |
| Μη καρδιαγγειακό επεισόδιο | 0,991 | 0,000 |
| Περιστατικό με υψηλή πίεση | 0,973 | 0,208 |
| Μη περιστατικό με υψηλή πίεση | 0,991 | 0,005 |
| Υψηλό ζάχαρο (< 120) | 0,932 | 0,045 |
| Μη υψηλό ζάχαρο (≥ 120) | 0,994 | 0,047 |

ΚΕΦΑΛΑΙΟ 4

Προγνωστικά μοντέλα – Πολλαπλή γραμμική και λογιστική παλινδρόμηση

4.1 Εισαγωγή

Στο κεφάλαιο αυτό, εξετάζουμε μέσω πολλαπλής γραμμικής και λογιστικής παλινδρόμησης, ποιοι δημογραφικοί, σωματομετρικοί και διατροφικοί παράγοντες επηρεάζουν τόσο τη συνολική χοληστερίνη όσο την LDL χοληστερόλη και την HDL χοληστερόλη. Στο τέλος καταλήγουμε στα βέλτιστα μοντέλα για κάθε μία από τις μεταβλητές αυτές. Πιο συγκεκριμένα, έχουμε στη διάθεσή μας δύο συνεχείς μεταβλητές, την Ολική Χοληστερόλη και την LDL Χοληστερόλη, στις οποίες και εφαρμόζουμε το γραμμικό μοντέλο πολλών μεταβλητών καθώς και μία δίτιμη μεταβλητή, την Ύπαρξη Χοληστερόλης, η οποία λαμβάνει την τιμή 0=Ναι εάν η Ολική Χοληστερόλη είναι μεγαλύτερη ή ίση από την τιμή 220 και την τιμή 1=Όχι εάν αυτή είναι μικρότερη του 220, και στην οποία και εφαρμόσαμε λογιστική παλινδρόμηση πολλών μεταβλητών.

4.2 Πολλαπλή Γραμμική Παλινδρόμηση

Μέσω της Πολλαπλής Γραμμικής Παλινδρόμησης, προσπαθούμε να προβλέψουμε μία μεταβλητή απόκρισης μέσω των τιμών $(p-1) \geq 1$ ανεξάρτητων μεταβλητών X_1, X_2, \dots, X_{p-1} . Για το σκοπό αυτό συγκεντρώνουμε δεδομένα που αφορούν τις τιμές $x_{i1}, x_{i2}, \dots, x_{i(p-1)}$ των $(p-1) \geq 1$ ανεξάρτητων μεταβλητών καθώς και την αντίστοιχη τιμή y_i της μεταβλητής απόκρισης για $i=1, 2, \dots, n$ θεωρώντας ότι η τιμή y_i μπορεί να προβλεφθεί μέσω του γραμμικού συνδυασμού:

$$y_i = \beta_0 + \beta_1 \chi_{i1} + \dots + \beta_{p-1} \chi_{i,p-1}$$

Μας ενδιαφέρει να προσδιορίσουμε τις παραμέτρους $\beta_0, \beta_1, \dots, \beta_{p-1}$ έτσι ώστε το άθροισμα των τετραγώνων των αποκλίσεων:

$$\varepsilon_i = y_i - (\beta_0 + \beta_1 \chi_{i1} + \dots + \beta_{p-1} \chi_{i,p-1}) \text{ για } i=1,2,\dots,v$$

να γίνει ελάχιστο.

Στο κανονικό μοντέλο πολλαπλής παλινδρόμησης:

$$Y_i = \beta_0 + \beta_1 \chi_{i1} + \dots + \beta_{p-1} \chi_{i,p-1} + \varepsilon_i \text{ για } i=1,2,\dots,v$$

ισχύουν οι επόμενες συνθήκες:

Σ1: Οι ποσότητες $\beta_0, \beta_1, \dots, \beta_{p-1}$ είναι άγνωστες παράμετροι.

Σ2: Τα $\chi_{i1}, \dots, \chi_{i,p-1}$ είναι γνωστοί αριθμοί.

Σ3: Το Y_i είναι η τιμή της εξαρτημένης μεταβλητής (ή μεταβλητής απόκρισης) κατά την i επανάληψη του πειράματος και είναι τυχαία μεταβλητή.

Σ4: Τα ε_i είναι τυχαία σφάλματα τα οποία ακολουθούν την Κανονική κατανομή με μέση τιμή 0 και διασπορά σ^2 .

Σ5: Τα σφάλματα ε_i και ε_j που αντιστοιχούν σε διαφορετικές επαναλήψεις του πειράματος ($i \neq j$) θεωρούνται ασυσχέτιστα.

Οι συντελεστές β_i εκφράζουν την μεταβολή που θα πραγματοποιηθεί στην μεταβλητή απόκρισης, εάν η αντίστοιχη μεταβλητή X_i αυξηθεί κατά μία μονάδα, ενώ οι υπόλοιπες μεταβλητές παραμείνουν σταθερές.

4.2.1 Πολυσυγγραμμικότητα (Multicollinearity)

Το φαινόμενο της πολυσυγγραμμικότητας εμφανίζεται όταν υπάρχει υψηλή συσχέτιση μεταξύ κάποιων ανεξάρτητων μεταβλητών. Οι συνέπειες που έχει αυτή στα προσαρμοσμένα μοντέλα, είναι οι ακόλουθες:

1. Οι προβλέψεις των τιμών της μεταβλητής απόκρισης, Y , με το πλήρες μοντέλο είναι περίπου ίδιες με τις προβλέψεις που προκύπτουν όταν αφαιρέσουμε μία ή περισσότερες εξαρτημένες μεταβλητές.
2. Το άθροισμα τετραγώνων των υπολοίπων του πλήρους μοντέλου δεν διαφέρει σημαντικά από το άθροισμα τετραγώνων των υπολοίπων του μοντέλου που προκύπτει όταν αφαιρέσουμε μία ή περισσότερες εξαρτημένες μεταβλητές.
3. Τα πρόσθετα αθροίσματα τετραγώνων που αντιστοιχούν στην εισαγωγή κάποιας ανεξάρτητης μεταβλητής έχουν μικρές τιμές σε σχέση με την μεταβλητότητα που έμεινε ανερμήνευτη μετά τη διαμόρφωση ενός μη πλήρους μοντέλου.
4. Τα τυπικά σφάλματα των ελαχίστων εκτιμητριών τετραγώνων για το πλήρες μοντέλο είναι πολύ μεγαλύτερα από τα τυπικά σφάλματα των εκτιμητριών των ιδίων παραμέτρων σε ένα μη πλήρες μοντέλο.
5. Τα διαστήματα εμπιστοσύνης παραμέτρων για το πλήρες μοντέλο είναι πολύ πλατύτερα από τα αντίστοιχα δείγματα εμπιστοσύνης για τις ίδιες παραμέτρους σε ένα μη πλήρες μοντέλο.
6. Όταν από ένα μοντέλο πολλαπλής παλινδρόμησης αφαιρεθεί μία ανεξάρτητη μεταβλητή οποία είναι υψηλά συσχετισμένη με κάποια ή κάποιες άλλες, τότε μεταβάλλεται σημαντικά ο συντελεστής της μεταβλητής ή των μεταβλητών που είναι υψηλά συσχετισμένος με αυτήν που αφαιρέθηκε.

Όταν διαπιστωθεί ότι για ένα σύνολο δεδομένων συμβαίνει κάτι από τα παραπάνω, θα πρέπει να είμαστε ιδιαίτερα προσεκτικοί στην ερμηνεία των αποτελεσμάτων των ελέγχων υποθέσεων, και να εξετάσουμε σε ποιες μεταβλητές εντοπίζεται το πρόβλημα της πολυσυγγραμμικότητας, εάν είναι δυνατόν, και να τις αφαιρέσουμε από το μοντέλο. Ένας δείκτης ο οποίος αποτελεί διαγνωστικό κριτήριο για την ύπαρξη πολυσυγγραμμικότητας, είναι ο παράγοντας διόγκωσης διακύμανσης (Variance Inflation Factor) και υπολογίζεται ως εξής:

$$VIF_k = \frac{1}{1-R_k^2} \text{ για } k=1,2,\dots,p-1$$

Όπου R_k^2 είναι ο συντελεστής προσδιορισμού του μοντέλου που χρησιμοποιεί ως εξαρτημένη τη μεταβλητή X_k και ως ανεξάρτητες τις υπόλοιπες $p-2$ ανεξάρτητες μεταβλητές $X_j, j \neq k$. Αν λάβουμε τιμή κοντά στη μονάδα, τότε η αντίστοιχη μεταβλητή X_k δεν έχει πρόβλημα πολυσυγγραμμικότητας σε σχέση με τις υπόλοιπες, ενώ εάν λάβουμε τιμή μεγαλύτερη του δέκα τότε εμφανίζει. .

4.3 Βέλτιστα μοντέλα

Στην παρούσα εργασία, κρίθηκε σκόπιμο εκτός από το να βρεθεί το βέλτιστο μοντέλο για την Συνολική Χοληστερίνη, να βρεθεί επίσης και το βέλτιστο μοντέλο για τις τρεις μεταβλητές που συμβάλλουν στον τύπο υπολογισμού της. Έτσι, παρακάτω παρουσιάζουμε τα μοντέλα και για τις μεταβλητές LDL Χοληστερόλη, Τριγλυκερίδια και HDL Χοληστερόλη. Με αυτόν τον τρόπο μπορούμε να βρούμε όλους τους παράγοντες που επηρεάζουν τόσο την Συνολική Χοληστερίνη όσο και των τριών ποσοτήτων που αυτή υπολογίζεται, έχοντας έτσι μία πιο πλήρη εικόνα. . Οι μέθοδοι που χρησιμοποιήσαμε είναι η Stepwise Regression, η Forward Selection και η Backward Elimination

4.3.1 Βέλτιστο μοντέλο για την Συνολική Χοληστερόλη

Αφού δοκιμάσαμε αρκετά μοντέλα, καταλήξαμε στο γεγονός ότι σύμφωνα με τρεις διαφορετικές μεθόδους, οι μεταβλητές που ερμηνεύουν καλύτερα την συνολική χοληστερόλη είναι η LDL χοληστερόλη, τα τριγλυκερίδια και ο δείκτης μάζας σώματος. Τελικά λαμβάνουμε τα ακόλουθα:

Πίνακας 4.3.1.1
Βέλτιστο μοντέλο για τη Συνολική Χοληστερόλη

| Coefficients ^a | | | | | | |
|---------------------------|-----------------------------------------------|------------|----------------------|--------|--------|------|
| Model | Unstandardized Coefficients | | Standardized | t | Sig. | |
| | B | Std. Error | Coefficients Beta | | | |
| | (Constant) | 72,203 | 2,918 | 24,745 | ,000 | |
| 1 | Δείκτης Μάζας Σώματος (kg/m ²) | -,424 | ,105 | -,051 | -4,055 | ,000 |
| | LDL-χοληστερόλη (mg/dL) | ,935 | ,013 | ,853 | 70,091 | ,000 |

| | | | | | |
|-----------------------|------|------|------|--------|------|
| Τριγλυκερίδια (mg/dl) | ,157 | ,005 | ,383 | 30,961 | ,000 |
|-----------------------|------|------|------|--------|------|

a. Dependent Variable: Συνολική Χοληστερίνη (mg/dl)

Το μοντέλο μας είναι το ακόλουθο:

$$\text{Συνολική Χοληστερόλη} = 72,203 - 0,424 * \text{Δείκτης Μάζας Σώματος} + 0,935 * \text{LDL Χοληστερόλη} + 0,157 * \text{Τριγλυκερίδια}$$

Συνεπώς εάν η συνολική χοληστερόλη αυξηθεί κατά μία μονάδα, ο δείκτης μάζας σώματος θα αυξηθεί κατά 0,424 μονάδες, δεδομένου ότι η LDL χοληστερόλη και τα τριγλυκερίδια παραμένουν σταθερά. Θα αυξηθεί κατά 0,935 μονάδες εάν ο δείκτης μάζας σώματος και τα τριγλυκερίδια παραμένουν σταθερά, και τέλος θα αυξηθεί κατά 0,157 μονάδες εάν μείνουν σταθερές οι υπόλοιπες.

Πίνακας 4.3.1.2

F – Test για το βέλτιστο μοντέλο

| ANOVA ^a | | | | | | |
|--------------------|------------|----------------|-----|-------------|----------|-------------------|
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| | Regression | 1064479,141 | 3 | 354826,380 | 2053,537 | ,000 ^b |
| 1 | Residual | 134429,026 | 778 | 172,788 | | |
| | Total | 1198908,167 | 781 | | | |

a. Dependent Variable: Συνολική Χοληστερόλη (mg/dl)

b. Predictors: (Constant), Τριγλυκερίδια (mg/dl), LDL-χοληστερόλη (mg/dL), Δείκτης Μάζας Σώματος (kg/m²)

Πίνακας 4.3.1.3

Πίνακας ορθής πρόβλεψης

| Υπαρξη χοληστερίνης * sinoliki_cut Crosstabulation | | |
|----------------------------------------------------|--------------|-------|
| | sinoliki_cut | Total |
| | ,00 | 1,00 |

| | | | | | |
|---------------------|-------------|------------|-------|-------|--------|
| | Ναι (>=220) | Count | 255 | 53 | 308 |
| Υπαρξη χοληστερίνης | | % of Total | 64,9% | 13,5% | 78,4% |
| | Όχι (<220) | Count | 60 | 25 | 85 |
| | | % of Total | 15,3% | 6,4% | 21,6% |
| Total | | Count | 315 | 78 | 393 |
| | | % of Total | 80,2% | 19,8% | 100,0% |

4.3.2 Βέλτιστο μοντέλο για την LDL Χοληστερόλη

Όσον αφορά την LDL Χοληστερόλη, επιλέξαμε ως βέλτιστα μοντέλα δύο. Πιο συγκεκριμένα το ένα περιλαμβάνει τον δείκτη μάζας σώματος και τον αριθμό ετών καπνίσματος, ενώ το δεύτερο περιλαμβάνει επίσης και την συστολική πίεση. Ακολουθούν οι αντίστοιχοι πίνακες.

Πίνακας 4.3.2.1
Βέλτιστο μοντέλο για τη LDL - Χοληστερόλη

| Coefficients ^a | | | | | | |
|---------------------------|--------------------------------------------|------------|---------------------------|------|-------|------|
| Model | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | |
| | B | Std. Error | | | | |
| | (Constant) | 78,244 | 10,094 | | 7,752 | ,000 |
| 1 | Δείκτης Μάζας Σώματος (kg/m ²) | 1,173 | ,387 | ,146 | 3,029 | ,003 |
| | Αριθμός ετών καπνίσματος | ,674 | ,163 | ,200 | 4,141 | ,000 |

a. Dependent Variable: LDL-χοληστερόλη (mg/dL)

Το πρώτο μοντέλο που προκύπτει είναι το ακόλουθο:

$$\text{LDL Χοληστερόλη} = 78,244 + 1,173 * \text{Δείκτης Μάζας Σώματος} + 0,674 * \text{Αριθμός ετών καπνίσματος}$$

Σύμφωνα με αυτό το μοντέλο, εάν η LDL Χοληστερόλη αυξηθεί κατά μία μονάδα ο Δείκτης Μάζας Σώματος θα αυξηθεί κατά 1,173 μονάδες, δεδομένου ότι ο αριθμός ετών καπνίσματος θα παραμείνει σταθερός. Επίσης εάν LDL Χοληστερόλη αυξηθεί κατά μία μονάδα, ο αριθμός ετών καπνίσματος θα αυξηθεί κατά 0,674 μονάδες εάν ο Δείκτης Μάζας Σώματος παραμείνει σταθερός.

Πίνακας 4.3.2.2

F – Test για το βέλτιστο μοντέλο

| ANOVA ^a | | | | | | |
|--------------------|------------|----------------|-----|-------------|--------|-------------------|
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| | Regression | 39885,147 | 2 | 19942,574 | 16,400 | ,000 ^b |
| 1 | Residual | 505865,620 | 416 | 1216,023 | | |
| | Total | 545750,768 | 418 | | | |

a. Dependent Variable: LDL-χοληστερόλη (mg/dL)

b. Predictors: (Constant), Αριθμός ετών καπνίσματος, Δείκτης Μάζας Σώματος (kg/m²)

Πίνακας 4.3.2.3

Πίνακας ορθής πρόβλεψης

| Bcl * LDL_cut1 Crosstabulation | | | | | |
|--------------------------------|------------|------------|-------|--------|-------|
| | | LDL_cut1 | | Total | |
| | | ,00 | 1,00 | | |
| Bcl | low | Count | 107 | 110 | 217 |
| | | % of Total | 13,5% | 13,9% | 27,4% |
| | higt | Count | 312 | 264 | 576 |
| | | % of Total | 39,3% | 33,3% | 72,6% |
| Total | Count | 419 | 374 | 793 | |
| | % of Total | 52,8% | 47,2% | 100,0% | |

Πίνακας 4.3.2.4
Βέλτιστο μοντέλο για τη LDL - Χοληστερόλη

| Coefficients ^a | | | | | | |
|---------------------------|-----------------------------------------------|------------|----------------------|------|-------|------|
| Model | Unstandardized Coefficients | | Standardized | t | Sig. | |
| | B | Std. Error | Coefficients Beta | | | |
| | (Constant) | 70,786 | 12,884 | | 5,494 | ,000 |
| 1 | Δείκτης Μάζας Σώματος (kg/m ²) | ,825 | ,419 | ,106 | 1,968 | ,050 |
| | Αριθμός ετών καπνίσματος | ,585 | ,170 | ,177 | 3,441 | ,001 |
| | Συστολική Πίεση | ,149 | ,105 | ,078 | 1,426 | ,155 |

a. Dependent Variable: LDL-χοληστερόλη (mg/dL)

Το δεύτερο μοντέλο που βρέθηκε είναι το ακόλουθο:

$$\text{LDL Χοληστερόλη} = 70,786 + 0,825 \cdot \text{Δείκτης Μάζας Σώματος} + 0,585 \cdot \text{Αριθμός ετών καπνίσματος} + 0,149 \cdot \text{Συστολική πίεση}$$

Σύμφωνα το δεύτερο μοντέλο, εάν αυξηθεί η LDL Χοληστερόλη κατά μία μονάδα, ο Δείκτης μάζας σώματος θα αυξηθεί κατά 0,825 μονάδες, δεδομένου ότι ο αριθμός ετών καπνίσματος και η συστολική πίεση παραμένουν σταθερές. Αντίστοιχα, θα αυξηθεί κατά 0,585 μονάδες ο αριθμός ετών καπνίσματος ενώ η συστολική πίεση κατά 0,149.

Πίνακας 4.3.2.5
F – Test για το βέλτιστο μοντέλο

| ANOVA ^a | | | | | | |
|--------------------|------------|----------------|-----|-------------|-------|-------------------|
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| | Regression | 34148,884 | 3 | 11382,961 | 9,827 | ,000 ^b |
| 1 | Residual | 447132,339 | 386 | 1158,374 | | |
| | Total | 481281,223 | 389 | | | |

a. Dependent Variable: LDL-χοληστερόλη (mg/dL)

b. Predictors: (Constant), Συστολική Πίεση, Αριθμός ετών καπνίσματος, Δείκτης Μάζας Σώματος (kg/m²)

Πίνακας 4.3.2.6
Πίνακας ορθής πρόβλεψης

| Bcl * LDL_cut Crosstabulation | | | | | |
|-------------------------------|------------|------------|---------|--------|-------|
| | | | LDL_cut | | Total |
| | | | ,00 | 1,00 | |
| Bcl | low | Count | 97 | 120 | 217 |
| | | % of Total | 12,2% | 15,1% | 27,4% |
| | higt | Count | 293 | 283 | 576 |
| | | % of Total | 36,9% | 35,7% | 72,6% |
| Total | Count | 390 | 403 | 793 | |
| | % of Total | 49,2% | 50,8% | 100,0% | |

4.4 Πολλαπλή Λογιστική Παλινδρόμηση

Η λογιστική παλινδρόμηση (logistic regression) χρησιμοποιείται για να περιγράψει τη σχέση της πιθανότητας ενός χαρακτηριστικού με διάφορους παράγοντες. Πιο συγκεκριμένα η πολλαπλή λογιστική παλινδρόμηση προσπαθεί να περιγράψει τη σχέση μεταξύ μίας δίτιμης εξαρτημένης μεταβλητής και δύο ή περισσότερων ανεξάρτητων μεταβλητών. Οι ανεξάρτητες μεταβλητές μπορεί να είναι είτε συνεχείς, και τότε ονομάζονται συμμεταβλητές (covariates), είτε κατηγορικές, και τότε ονομάζονται παράγοντες (factors). Το μοντέλο πολλαπλής λογιστικής παλινδρόμησης είναι το ακόλουθο:

$$\log\left(\frac{p_i}{1-p_i}\right) = b_0 + b_1X_{1i} + b_2X_{2i} + \dots + b_pX_{pi} + \varepsilon$$

Όπου:

η σταθερά b_0 είναι η τιμή του λογαρίθμου του λόγου συμπληρωματικών πιθανοτήτων ($\log\left(\frac{p_i}{1-p_i}\right)$) όταν όλες οι ανεξάρτητες μεταβλητές πάρουν την τιμή μηδέν.

Ο συντελεστής b_1 παριστάνει τη μεταβολή που επιφέρει στον λογάριθμο συμπληρωματικών πιθανοτήτων η αύξηση της ανεξάρτητης μεταβλητής X_1 κατά μία μονάδα όταν όλες οι άλλες ανεξάρτητες μεταβλητές παραμείνουν σταθερές. Οι υπόλοιποι συντελεστές ερμηνεύονται με παρόμοιο τρόπο.

Η εκτίμηση των παραμέτρων του μοντέλου γίνονται με τη μέθοδο της μέγιστης πιθανοφάνειας.

4.4.1 Βέλτιστο μοντέλο για τη Συνολική Χοληστερόλη

Πίνακας 4.4.1

Βέλτιστο μοντέλο για τη Συνολική Χοληστερόλη

| | | Variables in the Equation | | | | | |
|---------------------|--------------|---------------------------|-------|-------|----|------|--------|
| | | B | S.E. | Wald | df | Sig. | Exp(B) |
| Step 1 ^a | sbp | ,007 | ,012 | ,320 | 1 | ,572 | 1,007 |
| | year_smo | ,049 | ,017 | 8,567 | 1 | ,003 | 1,050 |
| | MedDietScore | -,014 | ,035 | ,170 | 1 | ,680 | ,986 |
| | Constant | -2,724 | 1,964 | 1,924 | 1 | ,165 | ,066 |

a. Variable(s) entered on step 1: sbp, year_smo, MedDietScore.

Το βέλτιστο μοντέλο είναι το ακόλουθο:

$$\text{Συνολική Χοληστερόλη} = -2,724 + 0,007 * \text{Συστολική πίεση} + 0,049 * \text{Αριθμός ετών καπνίσματος} - 0,014 * \text{Δείκτης μάζας σώματος}$$

Πίνακας 4.4.2

Έλεγχος χι – τετράγωνο του μοντέλου

| Omnibus Tests of Model Coefficients | | | | |
|-------------------------------------|-------|------------|----|------|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 11,901 | 3 | ,008 |
| | Block | 11,901 | 3 | ,008 |
| | Model | 11,901 | 3 | ,008 |

Πίνακας 4.4.3

Πίνακας ορθής πρόβλεψης

| Classification Table ^a | | | |
|-----------------------------------|---------------------|------------|--|
| Observed | Predicted | | |
| | Ύπαρξη χοληστερίνης | Percentage | |
| | | | |

| | | Nai (>=220) | Όχι (<220) | Correct | |
|--------------------|---------------------|-------------|------------|---------|------|
| Step 1 | Υπαρξη χοληστερίνης | Nai (>=220) | 148 | 2 | 98,7 |
| | | Όχι (<220) | 36 | 3 | 7,7 |
| Overall Percentage | | | | | 79,9 |

a. The cut value is ,500

4.4.2 Βέλτιστο μοντέλο για την LDL - Χοληστερόλη

Πίνακας 4.5.1

Βέλτιστο μοντέλο για την LDL - Χοληστερόλη

| | | Variables in the Equation | | | | | |
|---------------------|--------------|---------------------------|-------|-------|----|------|--------|
| | | B | S.E. | Wald | df | Sig. | Exp(B) |
| Step 1 ^a | sbp | ,023 | ,008 | 8,281 | 1 | ,004 | 1,023 |
| | year_smo | ,030 | ,013 | 5,256 | 1 | ,022 | 1,031 |
| | MedDietScore | -,022 | ,018 | 1,413 | 1 | ,235 | ,978 |
| | Constant | -1,500 | 1,153 | 1,692 | 1 | ,193 | ,223 |

a. Variable(s) entered on step 1: sbp, year_smo, MedDietScore.

Το βέλτιστο μοντέλο είναι το ακόλουθο:

$$\text{LDL - Χοληστερόλη} = -1,500 + 0,023 * \text{Συστολική πίεση} + 0,030 * \text{Αριθμός ετών καπνίσματος} - 0,022 * \text{Δείκτης μάζας σώματος}$$

Το 0,023 δείχνει κατά πόσο θα μεταβληθεί ο λογάριθμος συμπληρωματικών πιθανοτήτων εάν αυξηθεί η Συστολική πίεση κατά μία μονάδα και ο αριθμός ετών καπνίσματος και ο Δείκτης μάζας σώματος παραμείνουν σταθεροί. Ομοίως για το 0,030 εφόσον αυξηθεί ο αριθμός ετών καπνίσματος κατά μία μονάδα ενώ οι άλλες δύο ανεξάρτητες μεταβλητές παραμείνουν σταθερές. Και τέλος ο λογάριθμος των συμπληρωματικών πιθανοτήτων θα μειωθεί κατά 0,022 εάν αυξηθεί ο Δείκτης μάζας σώματος κατά μία μονάδα και οι άλλες δύο ανεξάρτητες μεταβλητές θ παραμείνουν σταθερές.

Πίνακας 4.5.2

Έλεγχος χι – τετράγωνο του μοντέλου

| Omnibus Tests of Model Coefficients | | | | |
|--------------------------------------------|-------|------------|----|------|
| | | Chi-square | df | Sig. |
| | Step | 27,485 | 3 | ,000 |
| Step 1 | Block | 27,485 | 3 | ,000 |
| | Model | 27,485 | 3 | ,000 |

Πίνακας 4.5.3

Πίνακας ορθής πρόβλεψης

| Classification Table^a | | | | | |
|-----------------------------------------|-----|------|-----------|------|------------|
| Observed | | | Predicted | | |
| | | | Bcl | | Percentage |
| | | | low | higt | Correct |
| Step 1 | Bcl | low | 5 | 92 | 5,2 |
| | | higt | 1 | 292 | 99,7 |
| Overall Percentage | | | | | 76,2 |

a. The cut value is ,500

4.5 Διαχωριστική ανάλυση

Η διαχωριστική ανάλυση χωρίζεται σε δύο τεχνικές, την περιγραφική διαχωριστική ανάλυση (Descriptive Discriminant Analysis – DDA) και την προβλεπτική διαχωριστική ανάλυση (Predicted Discriminant Analysis – PDA). Στην περιγραφική είναι γνωστό σε ποια ομάδα ανήκει κάθε περίπτωση μελέτης και καθορίζονται ποιες μεταβλητές κάνουν τις διαφορές μεταξύ δύο ή περισσότερων ομάδων. Στην προβλεπτική στόχος είναι η ταξινόμηση

μίας νέας περίπτωσης σε μία από τις ομάδες, έχει κοινά με την ανάλυση παλινδρόμησης αλλά εδώ η εξαρτημένη είναι κατηγορική. Επίσης δημιουργείται ένα σύνολο κανόνων όσων και των κατηγορικών ταξινόμησης και αποτελείται από εξισώσεις γραμμικού συνδυασμού των ανεξάρτητων μεταβλητών όπου και δημιουργούνται βάρη για κάθε μεταβλητή. Τα βάρη αυτά σχετίζονται με τις κατηγορικές ομάδες.

Έστω τώρα ότι έχουμε k υπό – πληθυσμούς – ομάδες, τις $\Pi_1, \Pi_2, \dots, \Pi_k$ με $k \geq 2$ και για κάθε ομάδα έχουμε μία κατανομή την $f_k(k)$. Η διαχωριστική συνάρτηση του Fisher είναι η εξής:

$$U(x) = L' \underline{x} - \frac{1}{2} L' (\mu_1 + \mu_2)$$

Η κύρια ιδέα ήταν ο μετασχηματισμός των πολλών μεταβλητών \underline{x} σε μεταβλητές παρατηρήσεις y τέτοιες ώστε να προέρχονται από πληθυσμούς Π_1 και Π_2 όσο το δυνατόν χωρισμένες περισσότερο. Για λόγους ευκολίας πρότεινε τους γραμμικούς συνδυασμούς του \underline{x} . Δεν υποθέτει ότι οι πληθυσμοί είναι κανονικοί αλλά υποθέτει ότι οι πίνακες συνδιακύμανσης είναι ίσοι. Η μετατροπή των πολλών μεταβλητών σε μονοδιάστατα σκορ γίνεται μέσω της παραπάνω διαχωριστικής συνάρτησης. Τα σκορ των δύο ομάδων πρέπει να είναι απομακρυσμένα αρκετά έτσι ώστε να γίνει ο διαχωρισμός και η ταξινόμηση των δύο ομάδων.

Έστω τα σκορ $y_{11}, y_{12}, \dots, y_{1n_1}$ για τον πρώτο πληθυσμό και $y_{21}, y_{22}, \dots, y_{2n_2}$ για τον δεύτερο. Το μέτρο απόστασης των δύο πληθυσμών – ομάδων που χρησιμοποιείται είναι το εξής:

$$D = \text{seperation} = \frac{|\bar{y}_1 - \bar{y}_2|}{s_y}$$

$$\text{Όπου: } s_y^2 = \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2}$$

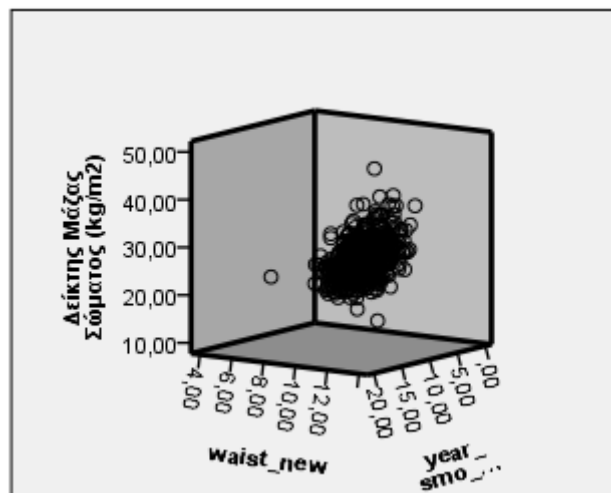
Σκοπός είναι η μεγιστοποίηση του D δείχνοντας έτσι ότι τα σκορ θα διαφοροποιούνται μεταξύ τους πιο πολύ.

4.6 Διαχωριστική ανάλυση για τη LDL – χοληστερόλη

Παρατηρούμε ότι η από κοινού κατανομή των τριών μεταβλητών φαίνεται να είναι η πολυμεταβλητή κανονική κατανομή.

Διάγραμμα 4.6.1

Διάγραμμα διασποράς τριών διαστάσεων για τις μεταβλητές Δείκτης Μάζας Σώματος, Περιφέρεια Μέσης και Αριθμός Ετών Καπνίσματος



Παρατηρούμε ότι οι διακυμάνσεις είναι οι ίδιες.

Πίνακας 4.6.1
Box Test

| Test Results | |
|--------------|------------|
| Box's M | 13,307 |
| Approx. | 2,192 |
| F | |
| df1 | 6 |
| df2 | 216028,129 |
| Sig. | ,051 |

Tests null hypothesis of equal
population covariance matrices.

Στο πίνακα που ακολουθεί δίνεται ο έλεγχος της ισότητας των μέσων τιμών στις δύο ομάδες για κάθε μεταβλητή χωριστά και παρατηρούμε ότι η υπόθεση απορρίπτεται.

Πίνακας 4.6.2
Έλεγχος της ισότητας των μέσων τιμών

| Tests of Equality of Group Means | | | | | |
|-----------------------------------------|---------------|--------|-----|-----|------|
| | Wilks' Lambda | F | df1 | df2 | Sig. |
| Δείκτης Μάζας Σώματος (kg/m2) | ,974 | 10,668 | 1 | 400 | ,001 |
| waist_new | ,958 | 17,749 | 1 | 400 | ,000 |
| year_smo_new | ,957 | 17,899 | 1 | 400 | ,000 |

Πίνακας 4.6.3

| Classification Results^{a,c} | | | | | |
|---------------------------------------------|-----------------|------|----------------------------|------|-------|
| | | Bcl | Predicted Group Membership | | Total |
| | | | ,00 | 1,00 | |
| Original | Count | ,00 | 64 | 37 | 101 |
| | | 1,00 | 112 | 189 | 301 |
| | Ungrouped cases | | 35 | 62 | 97 |
| | % | ,00 | 63,4 | 36,6 | 100,0 |
| | | 1,00 | 37,2 | 62,8 | 100,0 |
| Cross-validated ^b | Ungrouped cases | | 36,1 | 63,9 | 100,0 |
| | Count | ,00 | 63 | 38 | 101 |
| | | 1,00 | 114 | 187 | 301 |

| | | | | |
|---|------|------|------|-------|
| | ,00 | 62,4 | 37,6 | 100,0 |
| % | 1,00 | 37,9 | 62,1 | 100,0 |

- a. 62,9% of original grouped cases correctly classified.
- b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.
- c. 62,2% of cross-validated grouped cases correctly classified.

Η μέση τιμή που βρίσκουμε από τον επόμενο πίνακα είναι η -0,1505 και είναι πιο κοντά στη χαμηλή στάθμη της κωδικοποιημένης μεταβλητής για την LDL Χοληστερόλη.

Πίνακας 4.6.4
Έλεγχος του Fisher

| Functions at Group Centroids | |
|--------------------------------------------------------------------------------------|----------|
| Bcl | Function |
| | 1 |
| low | -,453 |
| higt | ,152 |
| Unstandardized canonical discriminant functions evaluated at group means | |

Η διαχωριστική εξίσωση είναι η ακόλουθη:

$$D=0.093*\Delta\acute{\epsilon}\iota\kappa\tau\eta\varsigma\ \text{Μ}\acute{\alpha}\zeta\alpha\varsigma\ \text{Σ}\acute{\omega}\mu\alpha\tau\omicron\varsigma+0,547*\text{waist_new}+0.626*\text{year_smo_new}$$

Πίνακας 4.6.5
Διαχωριστική Συνάρτηση

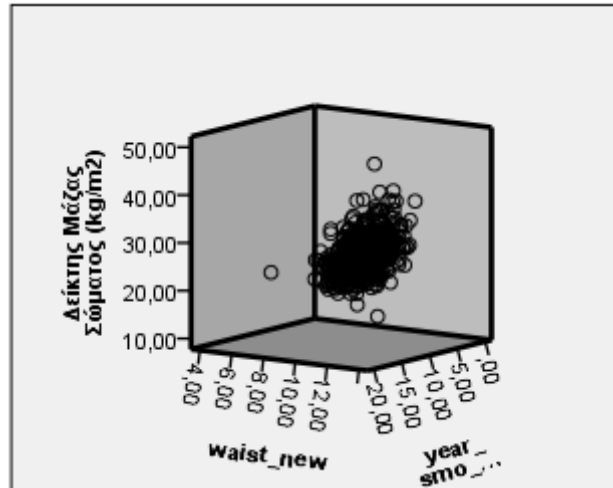
| Standardized Canonical Discriminant Function Coefficients | |
|--------------------------------------------------------------|----------|
| | Function |
| | 1 |
| Δείκτης Μάζας Σώματος (kg/m2) | ,093 |
| waist_new | ,547 |

4.7 Διαχωριστική ανάλυση για τη Συνολική Χοληστερόλη

Παρατηρούμε ότι η από κοινού κατανομή των τριών μεταβλητών φαίνεται να είναι η πολυμεταβλητή κανονική κατανομή.

Διάγραμμα 4.6.1

Διάγραμμα διασποράς τριών διαστάσεων για τις μεταβλητές Δείκτης Μάζας Σώματος, Ύπαρξη Χοληστερόλης και Αριθμός Ετών Καπνίσματος



Παρατηρούμε ότι οι διακυμάνσεις είναι οι ίδιες.

Πίνακας 4.7.1
Box Test

| Test Results | |
|--------------|-----------|
| Box's M | 8,269 |
| Approx. | 1,338 |
| df1 | 6 |
| F | 29549,737 |
| df2 | 29549,737 |
| Sig. | ,236 |

Tests null hypothesis of equal population covariance matrices.

Πίνακας 4.7.2
Έλεγχος της ισότητας των μέσων τιμών

| Tests of Equality of Group Means | | | | | |
|----------------------------------|---------------|-------|-----|-----|------|
| | Wilks' Lambda | F | df1 | df2 | Sig. |
| Δείκτης Μάζας Σώματος (kg/m2) | ,997 | ,568 | 1 | 197 | ,452 |
| waist_new | ,977 | 4,681 | 1 | 197 | ,032 |
| year_smo_new | ,955 | 9,312 | 1 | 197 | ,003 |

Πίνακας 4.7.3

| Classification Results ^{a,c} | | | | | |
|---------------------------------------|-------|---------------------|----------------------------|------------|-------|
| | | Ύπαρξη χοληστερίνης | Predicted Group Membership | | Total |
| | | | Ναι (>=220) | Όχι (<220) | |
| Original | Count | Ναι (>=220) | 102 | 57 | 159 |
| | | Όχι (<220) | 15 | 25 | 40 |
| | | Ungrouped cases | 123 | 177 | 300 |
| | % | Ναι (>=220) | 64,2 | 35,8 | 100,0 |
| | | Όχι (<220) | 37,5 | 62,5 | 100,0 |
| | | Ungrouped cases | 41,0 | 59,0 | 100,0 |
| Cross-validated ^b | Count | Ναι (>=220) | 101 | 58 | 159 |
| | | Όχι (<220) | 15 | 25 | 40 |
| | | Ungrouped cases | 123 | 177 | 300 |
| | % | Ναι (>=220) | 63,5 | 36,5 | 100,0 |
| | | Όχι (<220) | 37,5 | 62,5 | 100,0 |
| | | Ungrouped cases | 41,0 | 59,0 | 100,0 |

a. 63,8% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 63,3% of cross-validated grouped cases correctly classified.

Από τον πίνακα που ακολουθεί βρίσκουμε ότι η μέση τιμή είναι η 0,186 και είναι πιο κοντά στα άτομα που δεν έχουν χοληστερόλη.

Πίνακας 4.7.4

Έλεγχος του Fisher

| Functions at Group Centroids | |
|--------------------------------------------------------------------------|----------|
| Ύπαρξη χοληστερίνης | Function |
| | 1 |
| Ναι (>=220) | -,125 |
| Όχι (<220) | ,497 |
| Unstandardized canonical discriminant functions evaluated at group means | |

Η διαχωριστική εξίσωση είναι η ακόλουθη:

$$D = -0,531 * \text{Δείκτης Μάζας Σώματος} + 0,798 * \text{waist_new} + 0,718 * \text{year_smo_new}$$

Πίνακας 4.7.5

Διαχωριστική Συνάρτηση

| Standardized Canonical Discriminant Function Coefficients | |
|----------------------------------------------------------------------|---------------|
| | Function 1 |
| Δείκτης Μάζας Σώματος (kg/m ²) | -,531 |
| waist_new | ,798 |
| year_smo_new | ,718 |

ΚΕΦΑΛΑΙΟ 5

Ανάλυση Καμπύλης ROC (Receiver Operating Curve)

5.1 Εισαγωγή

Οι καμπύλες ROC χρησιμοποιούνται όταν θέλουμε να εφαρμόσουμε κάποιον διαγνωστικό έλεγχο για την εμφάνιση ή όχι κάποιας ασθένειας και έχουμε στην διάθεσή μας συνεχή δεδομένα. Η διαδικασία που εφαρμόζουμε είναι η ακόλουθη. Θεωρούμε ότι υπάρχουν δύο ομάδες ατόμων, μία φυσιολογική (αρνητική) και μία άλλη θετική. Ο διαγνωστικός

έλεγχος δεν είναι τέλειος και συνεπώς θα αναγνωρίσει κάποιες αλλά όχι όλες τις θετικές περιπτώσεις. Ο αληθινά θετικός ρυθμός, γνωστός και ως ευαισθησία, είναι η αναλογία των θετικών περιπτώσεων που διαπίστωσε ο διαγνωστικός έλεγχος ως προς τον συνολικό αριθμό των ατόμων που είναι γνωστό ότι πάσχουν από την ασθένεια. Αντιθέτως, ο αληθινά αρνητικός ρυθμός, γνωστός και ως ειδικότητα, είναι η αναλογία των αρνητικών περιπτώσεων που διαπίστωσε ο διαγνωστικός έλεγχος ως προς τον συνολικό αριθμό των ατόμων που είναι γνωστό ότι δεν πάσχουν από την ασθένεια.

Η ανάλυση ROC χρησιμοποιείται για να βρεθεί μία τιμή αποκοπής που θα ελαχιστοποιεί τον αριθμό των ψευδώς θετικών και αρνητικών αποτελεσμάτων ή ομοίως θα μεγιστοποιεί την ευαισθησία και την ειδικότητα. Πιο συγκεκριμένα, η καμπύλη ROC είναι μία γραφική παράσταση η οποία στον κατακόρυφο άξονα αναπαριστά την ευαισθησία και στον οριζόντιο άξονα την ποσότητα (1-ειδικότητα). Μία καλή πρώτη επιλογή για την τιμή αποκοπής του διαγνωστικού ελέγχου να αντιστοιχεί σε ένα σημείο που βρίσκεται κοντά στην επάνω αριστερή γωνία του διαγράμματος. Έτσι μεγιστοποιείται η ευαισθησία και παράλληλα και η ειδικότητα. Σε ορισμένους όμως ελέγχους είναι σημαντικό να μην παραληφθεί η ανίχνευση μιας αρνητικής περίπτωσης, και συνεπώς το βέλτιστο σημείο αποκοπής θα μετακινηθεί προς την επάνω δεξιά γωνία, όπου και μεγιστοποιείται η ειδικότητα. Για την εύρεση του βέλτιστου σημείου αποκοπής, μεγιστοποιούμε την συνάρτηση: $h(c) = \text{ευαισθησία} + \text{ειδικότητα} - 1$, η οποία είναι γνωστή ως Δείκτης του Youden.

Το εμβαδόν κάτω από την καμπύλη (Area Under Curve) χρησιμοποιείται για την σύγκριση διαφορετικών διαγνωστικών ελέγχων. Όσο μεγαλύτερο είναι αυτό, τόσο πιο αξιόπιστος είναι ο έλεγχος. Το εμβαδόν αυτό είναι ισοδύναμο με την πιθανότητα ότι ο έλεγχος θα ταξινομήσει ένα τυχαία επιλεγμένο άτομο από την ομάδα των ασθενών υψηλότερα από ότι ένα τυχαία επιλεγμένο άτομο από την ομάδα των υγιών. Αυτό ισοδυναμεί με το Wilcoxon rank test και αποτελεί μία σημαντική στατιστική ιδιότητα των καμπυλών ROC.

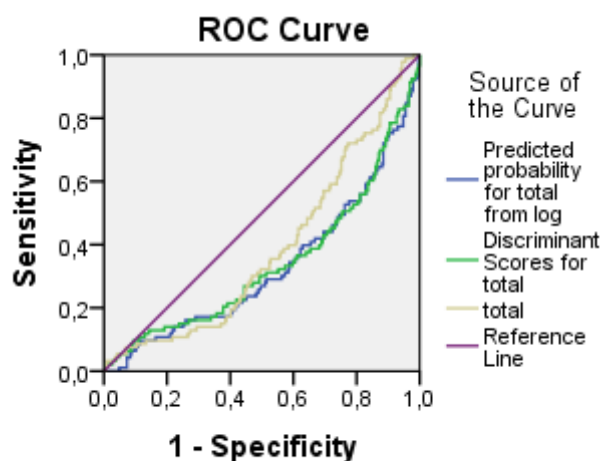
(Στατιστικές Μέθοδοι Στις Κοινωνικές Επιστήμες – Σημειώσεις)

5.2 Ανάλυση Καμπύλης ROC

Εφαρμόζουμε την ανάλυση της καμπύλης ROC για την μεταβλητή Συνολική Χοληστερόλη και παρουσιάζουμε το διάγραμμα που ακολουθεί.

Διάγραμμα 5.2.1

Ανάλυση της καμπύλης ROC για την μεταβλητή Συνολική Χοληστερόλη



Diagonal segments are produced by ties.

Το εμβαδόν κάτω από την καμπύλη είναι ίσο με 32,9% και βρίσκεται εντός των ορίων. Από τον παρακάτω πίνακα συγκρίνουμε την αποτελεσματικότητα των τριών μεθόδων από το εμβαδόν. Όσο πιο μικρό τόσο καλύτερα.

| Test Result Variable(s) | Area Under the Curve | | | | |
|------------------------------------------|----------------------|-------------------------|------------------------------|------------------------------------|-------------|
| | Area | Std. Error ^a | Asymptotic Sig. ^b | Asymptotic 95% Confidence Interval | |
| | | | | Lower Bound | Upper Bound |
| Predicted probability for total from log | ,329 | ,034 | ,000 | ,263 | ,395 |
| Discriminant Scores for total | ,340 | ,035 | ,000 | ,273 | ,408 |
| total | ,388 | ,033 | ,001 | ,324 | ,451 |

The test result variable(s): Predicted probability for total from log has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

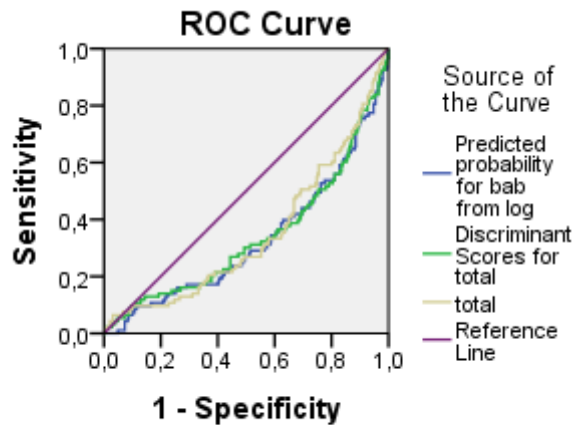
a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Εφαρμόζουμε την ανάλυση της καμπύλης ROC για την μεταβλητή LDL Χοληστερόλη και παρουσιάζουμε το διάγραμμα που ακολουθεί.

Διάγραμμα 5.2.2

Ανάλυση της καμπύλης ROC για την μεταβλητή LDL - Χοληστερόλη



Diagonal segments are produced by ties.

Το εμβαδόν κάτω από την καμπύλη είναι ίσο με 32,9% και βρίσκεται εντός των ορίων.

| Test Result Variable(s) | Area Under the Curve | | | | |
|----------------------------------------|----------------------|-------------------------|------------------------------|------------------------------------|-------------|
| | Area | Std. Error ^a | Asymptotic Sig. ^b | Asymptotic 95% Confidence Interval | |
| | | | | Lower Bound | Upper Bound |
| Predicted probability for bab from log | ,329 | ,034 | ,000 | ,263 | ,395 |
| Discriminant Scores for total | ,340 | ,035 | ,000 | ,273 | ,408 |
| total | ,348 | ,033 | ,000 | ,283 | ,413 |

The test result variable(s): Predicted probability for bab from log has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

ΚΕΦΑΛΑΙΟ 6

Συμπεράσματα

Στο δείγμα που διαθέτουμε, το 50,2% είναι άνδρες ενώ το 49,8% γυναίκες. Ακόμη το 24,1% είναι ανύπαντροι, το 68,8% παντρεμένοι, ενώ το 3,5% διαζευγμένοι. Επίσης το 24,1% ε

είναι ανύπαντροι και το 3,6% είναι χήροι. Χοληστερόλη έχουν το 78,4% ενώ το 21,6% δεν έχουν.

Το μέσο βάρος κυμαίνεται στα 75,58 κιλά, το μέσο ύψος στα 169,45 cm, ενώ ο δείκτης μάζας σώματος στις 26,229 μονάδες. Η περιφέρεια μέσης στις 90,25 μονάδες ενώ η περιφέρεια γοφών στις 104,22 μονάδες. Η LDL χοληστερόλη στις 48,70 μονάδες ενώ η HDL στις 121,60 μονάδες. Η μέση συνολική χοληστερόλη είναι 193,25 μονάδες.

Βρέθηκε επίσης ότι η συνολική χοληστερόλη και η LDL χοληστερόλη ακολουθούν την Κανονική Κατανομή.

Πιο συγκεκριμένα, βρέθηκε ότι η συνολική χοληστερόλη παρουσιάζει θετική συσχέτιση με τις μεταβλητές ηλικία, βάρος, δείκτης μάζας σώματος, περιφέρεια μέσης, περιφέρεια γοφών, αριθμός τσιγάρων ανά ημέρα, αριθμός ετών καπνίσματος, συστολική και διαστολική πίεση, επίπεδο γλυκόζης στο αίμα και προφανώς με τα τριγλυκερίδια.

Η συνολική χοληστερόλη εμφανίζει αρνητική συσχέτιση με τις μεταβλητές ποσοστό πρόσληψης πρωτεϊνών, ποσοστό πρόσληψης κορεσμένων λιπαρών οξέων, σκορ μεσογειακής διατροφής, κατανάλωση γαλακτοκομικών, γλυκών, αυγών και τη συνολική κατανάλωση κρέατος.

Η LDL χοληστερόλη παρουσιάζει θετική συσχέτιση με την ηλικία, το βάρος, τον δείκτη μάζας σώματος, την περιφέρεια μέσης, την περιφέρεια γοφών, τον αριθμό τσιγάρων ανά ημέρα, τον αριθμό ετών καπνίσματος, τη συστολική πίεση, τη διαστολική πίεση, την HDL χοληστερόλη, την πρωτεΐνη C, το ποσοστό πρόσληψης κορεσμένων λιπαρών οξέων, το σκορ μεσογειακής διατροφής, την εβδομαδιαία κατανάλωση γαλακτοκομικών και γλυκών και τέλος την συνολική κατανάλωση κρέατος.

Βρέθηκε ότι όσον αφορά την συνολική χοληστερόλη απορρίπτουμε τη μηδενική υπόθεση για το φύλο, για τους καπνιστές και για όσους είναι εκτεθειμένοι στο παθητικό κάπνισμα. Για όσους πάσχουν από υπέρταση και έχουν στο ιστορικό τους κάποιο καρδιαγγειακό επεισόδιο. Για όσους καταναλώνουν καθημερινά ελαιόλαδο, βούτυρο και καφές. Για την LDL χοληστερόλη βρέθηκε ότι απορρίπτουμε τη μηδενική υπόθεση για τις ίδιες μεταβλητές.

Στη πολλαπλή γραμμική παλινδρόμηση για τη συνολική χοληστερόλη βρέθηκε ότι την επηρεάζουν ο δείκτης μάζας σώματος, η LDL χοληστερόλη και τα τριγλυκερίδια, ενώ για την LDL χοληστερόλη έχουμε δύο μοντέλα. Το πρώτο περιλαμβάνει τον δείκτη μάζας σώματος και τον αριθμό ετών καπνίσματος, ενώ το δεύτερο περιλαμβάνει και τη συστολική πίεση. Το ποσοστό πρόβλεψης για το μοντέλο της συνολικής χοληστερόλης είναι ίσο με 71,3%, ενώ για

το πρώτο μοντέλο της LDL χοληστερόλης είναι ίσο με 47,9% και για το δεύτερο είναι ίσο με 46,8%.

Στη πολλαπλή λογιστική παλινδρόμηση βρέθηκε ότι η συστολική πίεση, ο δείκτης μάζας σώματος και ο αριθμός ετών καπνίσματος επηρεάζουν τόσο τη συνολική όσο και την LDL χοληστερόλη. Το ποσοστό ορθής πρόβλεψης για το μοντέλο της συνολικής χοληστερόλης είναι ίσο με 79,89% ενώ για το μοντέλο της LDL χοληστερόλης είναι ίσο με 76,15%.

Όσον αφορά τη διαχωριστή ανάλυση, παρατηρήσαμε ότι και για την LDL χοληστερόλη και για τη συνολική χοληστερόλη, το μοντέλο είναι το ίδιο. Πιο συγκεκριμένα τις επηρεάζουν η περιφέρεια μέσης, ο δείκτης μάζας σώματος και ο αριθμός ετών καπνίσματος. Αρχικά βέβαια έγινε ο έλεγχος της πολυμεταβλητής κανονικής κατανομής, όπου και βρέθηκε ότι ισχύει, καθώς και ο έλεγχος ισότητας των διακυμάνσεων. Το ποσοστό ορθής πρόβλεψης για το μοντέλο της LDL χοληστερόλης είναι ίσο με 63,4% ενώ για το μοντέλο της συνολικής χοληστερόλης είναι ίσο με 64,2%.

Τέλος εφαρμόσαμε την ανάλυση της καμπύλης ROC για τη συνολική χοληστερόλη και την LDL χοληστερόλη και βρήκαμε ότι το εμβαδόν κάτω από την καμπύλη ισούται με 32,9% και βρίσκεται εντός των ορίων. Από τις τρεις μεθόδους καλύτερη βρέθηκε η πρώτη που δίνει και το παραπάνω εμβαδόν για τις δύο μεταβλητές μας. Παρατηρούμε ότι είναι ακριβώς το ίδιο και για την Συνολική Χοληστερόλη και για την LDL – Χοληστερόλη.

Τα συμπεράσματα μας συμπίπτουν με την διεθνή βιβλιογραφία και με έρευνες όπως αυτή του Διατροφολόγου κ. Μάριου Δημόπουλου (2007) και του κ. Χρήστου Πίτσαβου (2009) αναπληρωτή καθηγητή Καρδιολογίας του Πανεπιστημίου Αθηνών και του κ. Παναγιωτάκου Δημοσθένη (2004).

ΕΛΛΗΝΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

- 1) Δημοσθένης Β. Παναγιωτάκος, (2006) Μεθοδολογία της Έρευνας και της Ανάλυσης δεδομένων για τις επιστήμες της Υγείας; Αθήνα

- 2) Ηλιόπουλος Γιώργος (2006) Βασικές Μέθοδοι Εκτίμησης Παραμέτρων με σημείο και με διάστημα; Εκδόσεις Αθ. Σταμούλης; Αθήνα
- 3) Κλέων Τσίμπος, Φώτης Γεωργιακώδης, (1999) Περιγραφική και Διερευνητική Στατιστική, Ανάλυση δεδομένων, Μονοδιάστατη Ανάλυση; Εκδόσεις Αθ. Σταμούλης; Αθήνα
- 4) Μιχαήλ Παπαδάκης, Κλέων Τσίμπος, (2004) Δημογραφική Ανάλυση, Αρχές – Μέθοδοι – Υποδείγματα; Εκδόσεις Αθ. Σταμούλης; Αθήνα
- 5) Π. Οικονόμου, Χ. Καρώνη,(2010) Στατιστικά Μοντέλα Παλινδρόμησης; Εκδόσεις Συμεων; Αθήνα

ΣΗΜΕΙΩΣΕΙΣ - ΕΡΓΑΣΙΕΣ

- 1) Αλέξανδρος Β. Μήτρου (2007) Διαχωριστική ανάλυση με εφαρμογές στην εκπαίδευση, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Τμήμα Μαθηματικό, Πρόγραμμα Μεταπτυχιακών Σπουδών στην Στατιστική Και Επιχειρησιακή Έρευνα
- 2) Ανδρέας Π. Λιάπης (2013) Κλινικά και δημογραφικά δεδομένα ασθενών με επιληψία που προσέρχονται στα κέντρα κατάταξης του στρατού ξηράς στην Ελλάδα, Πανεπιστήμιο Πειραιώς, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πρόγραμμα Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική.
- 3) Γεώργιος Τζαβελάς (2012) Κλινικές Δοκιμές, Πανεπιστήμιο Πειραιώς, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πρόγραμμα Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική.
- 4) Ευαγγελάρας Χρήστος (2011) Ανάλυση δεδομένων με τη χρήση στατιστικών πακέτων: Σημειώσεις για το SPSS, Πανεπιστήμιο Πειραιώς, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πρόγραμμα Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική.
- 5) Κούτρας Μάρκος (2012), Σημειώσεις για το μάθημα Ανάλυση παλινδρόμησης και ανάλυση διακύμανσης, Πανεπιστήμιο Πειραιώς, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πρόγραμμα Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική.

ΞΕΝΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

- 1) Altman. DG, (1991) Practical Statistics for Medical Research. London: Chapman and Hall: CRC
- 2) Cox DR ,(1972) Regression models and life tables. J Roy Static Soc B
- 3) Dimopoulos N, Piperi C, Salonicioti A, et al. (2007) Characterization of the lipid profile in dementia and depression in the elderly. J Geriatr Psychiatry Neurol, 20:138–144.
- 4) Neil H, Timm, (2002) Aplied Multivariate Analysis Springer Verlang New York
- 5) Panagiotakos DB1, Pitsavos C, Chrysohoou C, Skoumas J, Stefanadis C. (2004) Status and management of blood lipids in Greek adults and their relation to socio-demographic, lifestyle and dietary factors: the ATTICA Study. Blood lipids distribution in Greece.
- 6) Pitsavos Christos, (2009) Beyond LDL Cholesterol: The role of LDL particles and HDL. Hospital Chronicles, Vol 4, No 3: 102-104.
- 7) Subhash Sharma, (1999) Reading and Undrestanding Multivariate Statistics.
- 8) (www.doctor4all.gr)
- 9) (www.medidiet.gr)