

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**  
**Σχολή Χρηματοοικονομικής και Στατιστικής**



**Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ  
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ  
ΠΟΥ ΑΦΟΡΟΥΝ ΤΗΝ ΕΜΦΑΝΙΣΗ  
ΚΑΡΔΙΑΓΓΕΙΑΚΩΝ ΠΑΘΗΣΕΩΝ**

**Ευαγγελία Λιακοπούλου**

**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**

Σεπτέμβριος 2024

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:

- Τριανταφύλλου Ιωάννης (Επιβλέπων, Επίκουρος Καθηγητής)
- Ευαγγελάρας Χαράλαμπος (Αναπληρωτής Καθηγητής)
- Ρακιτζής Αθανάσιος (Επίκουρος Καθηγητής)

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

**UNIVERSITY OF PIRAEUS**  
**School of Finance and Statistics**



**Department of Statistics and Insurance Science**

**POSTGRADUATE PROGRAM IN  
APPLIED STATISTICS**

**STATISTICAL ANALYSIS OF  
CARDIOVASCULAR DISEASES DATA**

By

**Evangelia Liakopoulou**

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

September 2024

*Ένα μεγάλο ευχαριστώ από καρδιάς στην οικογένειά μου, για την αμέριστη στήριξη και πίστη τους σε εμένα, καθώς χωρίς αυτούς τίποτα από όσα έχω καταφέρει δεν θα ήταν εφικτό.*

*Στους γονείς μου,  
Δημήτρη & Γεωργία*

## Περίληψη

Η παρούσα εργασία επικεντρώνεται στη στατιστική ανάλυση των παραγόντων που επηρεάζουν την εμφάνιση καρδιαγγειακών παθήσεων έχοντας ως στόχο την ανάπτυξη προβλεπτικών μοντέλων. Το σύνολο των δεδομένων που χρησιμοποιήθηκε είναι το «*Heart Failure Prediction Dataset*» από το Kaggle το οποίο αποτελείται από 918 παρατηρήσεις με μεταβλητές δημογραφικών χαρακτηριστικών όπως το φύλο και η ηλικία αλλά και ποικίλων ειδικών ιατρικών δεικτών.

Αρχικά, πραγματοποιήθηκε περιγραφική ανάλυση στις μεταβλητές του δείγματος και στη συνέχεια κάποιοι μέθοδοι επαγωγικής στατιστικής συμπερασματολογίας. Συγκεκριμένα, για τις συνεχείς μεταβλητές διεξήχθησαν οι έλεγχοι Shapiro – Wilk για τον έλεγχο της κανονικότητας, και συσχετίσεων. Για τις κατηγορικές οι έλεγχοι  $\chi^2$  για έλεγχο συσχετίσεων, odds ratio, δίνοντας ιδιαίτερη έμφαση στη μεταβλητή Heart Disease που δείχνει τα υγιή και τα νοσούντα από καρδιακή πάθηση άτομα. Επιπλέον, χρησιμοποιήθηκαν οι έλεγχοι Wilcoxon-Mann Whitney και Kruskal-Wallis για τη σύγκριση των ανεξάρτητων ομάδων του δείγματος.

Επιπρόσθετα, εφαρμόστηκαν η μέθοδος παραγοντικής ανάλυσης στις συνεχείς μεταβλητές για την εύρεση κοινών παραγόντων, μοντέλο λογιστικής παλινδρόμηση για την πρόβλεψη της πιθανότητας εμφάνισης καρδιακής πάθησης και σε αυτό, καμπύλη ROC για την αξιολόγηση της διαγνωστικής του ακρίβειας.

Εν κατακλείδι, με την επιλογή κατάλληλων δεικτών και αποτελεσματικών προβλεπτικών μοντέλων, μπορεί να επιτευχθεί ο έγκαιρος εντοπισμός των ατόμων που ανήκουν σε ομάδες αυξημένου κινδύνου για την εμφάνιση καρδιακών παθήσεων.

## **Abstract**

The present study focuses on the statistical analysis of the factors influencing the occurrence of cardiovascular disease with the aim of developing predictive models. The dataset used is the "Heart Failure Prediction Dataset" from Kaggle which consists of 918 observations with variables of demographic characteristics such as gender and age and a variety of specific medical indicators.

Initially, descriptive analysis was performed on the sample variables followed by some methods of inferential statistical inference. In particular, for continuous variables, Shapiro - Wilk tests were conducted to check normality, and correlations. For the categorical ones the  $X^2$  tests for testing correlations, odds ratio, with particular emphasis on the Heart Disease variable showing healthy and heart disease subjects. In addition, Wilcoxon-Mann Whitney and Kruskal-Wallis tests were used to compare the independent sample groups.

In addition, factor analysis method was applied to continuous variables to find common factors, logistic regression model to predict the probability of heart disease and on it, ROC curve to evaluate its diagnostic accuracy.

In conclusion, with the selection of appropriate indicators and effective predictive models, early identification of individuals belonging to groups at increased risk of developing heart disease can be achieved.

# Περιεχόμενα

## Κατάλογος Πινάκων

## Κατάλογος Σχημάτων

## Κατάλογος Συντομογραφιών

<b>1. Εισαγωγή</b>	<b>14</b>
1.1 Ορισμός καρδιαγγειακών παθήσεων	14
1.2 Θνησιμότητα καρδιαγγειακών παθήσεων	14
1.3 Διάκριση καρδιαγγειακών παθήσεων	14
1.3.1 Τύποι και συμπτώματα καρδιακών ασθενειών	15
1.3.2 Τύποι και συμπτώματα εγκεφαλικών αγγειακών προβλημάτων	17
1.3.3 Τύποι και συμπτώματα περιφερικών αγγειακών νοσημάτων	19
1.4 Ιστορική αναδρομή καρδιαγγειακών παθήσεων	21
1.5 Περιγραφή δείγματος	21
<b>2. Έλεγχος και περιγραφική στατιστική ανάλυση</b>	<b>23</b>
2.1 Εισαγωγή	23
2.2 Έλεγχος ελλειπών και διπλότυπων τιμών	23
2.3 Συνεχείς μεταβλητές	24
2.3.1 Εύρεση μη λογικών τιμών	24
2.3.2 Ηλικία	25
2.3.3 Χοληστερίνη	25
2.3.4 Μέγιστος καρδιακός ρυθμός	26
2.3.5 Μέγιστη απόκλιση ηλεκτροκαρδιογραφήματος κατά το τεστ κοπώσεως	27
2.3.6 Αρτηριακή πίεση σε κατάσταση ηρεμίας	28
2.4 Κατηγορικές μεταβλητές	28
2.4.1 Φύλο	28

2.4.2	Στηθάγχη	29
2.4.3	Ζάχαρο νηστείας	30
2.4.4	Ηλεκτροκαρδιογράφημα σε κατάσταση ηρεμίας	31
2.4.5	Πρόκληση στηθάγχης εξαιτίας σωματικής άσκησης	31
2.4.6	Αλλαγή μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης του τεστ κοπώσεως στο ηλεκτροκαρδιογράφημα	32
2.4.7	Ύπαρξη καρδιακής πάθησης	33
<b>3</b>	<b>Επαγωγική στατιστική συμπερασματολογία</b>	<b>33</b>
3.1	Εισαγωγή	33
3.2	Έλεγχος κανονικότητας συνεχών μεταβλητών	34
3.3	Έλεγχος συσχετίσεων μη κανονικά κατανομημένων μεταβλητών	36
3.4	Έλεγχοι ανεξαρτησίας κατηγορικών μεταβλητών	37
3.4.1	Έλεγχος ανεξαρτησίας φύλου και καρδιακής πάθησης	37
3.4.2	Έλεγχος ανεξαρτησίας στηθάγχης και καρδιακής πάθησης	39
3.4.3	Έλεγχος ανεξαρτησίας ζαχάρου νηστείας και καρδιακής πάθησης	40
3.4.4	Έλεγχος ανεξαρτησίας ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας και καρδιακής πάθησης	41
3.4.5	Έλεγχος ανεξαρτησίας στηθάγχης κατά τη διάρκεια άσκησης και καρδιακής πάθησης	43
3.4.6	Έλεγχος ανεξαρτησίας στην αλλαγής του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης της άσκησης ή του στρες και της καρδιακής πάθησης	44
3.4.7	Έλεγχος ανεξαρτησίας μεταξύ των υπόλοιπων μεταβλητών	45
3.5	Συγκρίσεις με χρήση του ελέγχου Wilcoxon  Mann – Whitney	46
3.5.1	Σύγκριση των τιμών των συνεχών μεταβλητών μεταξύ ανδρών και γυναικών	47
3.5.2	Σύγκριση των τιμών των συνεχών μεταβλητών σε πληθυσμούς που ορίζονται βάσει του επιπέδου καρδιακής πάθησης.	48
3.5.3	Σύγκριση των τιμών των συνεχών μεταβλητών σε πληθυσμούς που ορίζονται βάσει του επιπέδου ζαχάρου νηστείας	49



3.5.4	Σύγκριση των τιμών των συνεχών μεταβλητών σε πληθυσμούς που ορίζονται βάσει του επιπέδου της στηθάγχης από άσκηση	51
3.6	Συγκρίσεις με χρήση του ελέγχου Kruskal-Wallis	52
3.6.1	Σύγκριση των τιμών των συνεχών μεταβλητών σε πληθυσμούς που ορίζονται βάσει του επιπέδου του ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας	52
3.6.2	Σύγκριση των τιμών των συνεχών μεταβλητών σε πληθυσμούς που ορίζονται βάσει του επιπέδου της αλλαγής του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης του τεστ κοπώσεως στο ηλεκτροκαρδιογράφημα	55
3.6.3	Σύγκριση των τιμών των συνεχών μεταβλητών σε πληθυσμούς που ορίζονται βάσει του επιπέδου του στηθαγχικού πόνου	58
<b>4</b>	<b>Στατιστική ανάλυση και μοντελοποίηση</b>	<b>61</b>
4.1	Εισαγωγή	61
4.2	Ανάλυση παραγόντων	62
4.2.1	Υποθέσεις ανάλυσης παραγόντων	62
4.2.2	Ανάλυση παραγόντων και καρδιακές παθήσεις	63
4.3	Λογιστική παλινδρόμηση	67
4.4	Καμπύλη ROC	70
<b>5</b>	<b>Συμπεράσματα ανάλυσης</b>	<b>71</b>
	<b>Βιβλιογραφία</b>	<b>73</b>

## Κατάλογος Πινάκων

Πίνακας 1	Έλεγχος ελλειπόν τιμών	24
Πίνακας 2	Έλεγχος διπλότυπων τιμών	24
Πίνακας 3	Περιγραφικά μέτρα αρτηριακής πίεσης σε κατάσταση ηρεμίας και χοληστερίνης	24
Πίνακας 4	Περιγραφικά μέτρα ηλικίας	25
Πίνακας 5	Περιγραφικά μέτρα χοληστερίνης	26
Πίνακας 6	Περιγραφικά μέτρα μέγιστου καρδιακού ρυθμού	26
Πίνακας 7	Περιγραφικά μέτρα μέγιστης απόκλισης του ηλεκτροκαρδιογραφήματος	27
Πίνακας 8	Περιγραφικά μέτρα αρτηριακής πίεσης σε κατάσταση ηρεμίας	28
Πίνακας 9	Συχνότητα φύλου	29
Πίνακας 10	Συχνότητα στηθαγικού πόνου	30
Πίνακας 11	Συχνότητα ζαχάρου νηστείας	30
Πίνακας 12	Συχνότητα τύπων ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας	31
Πίνακας 13	Συχνότητα ύπαρξης στηθάγχης από άσκηση	32
Πίνακας 14	Συχνότητα αλλαγής μοτίβου της κλίση κατά το σημείο μέγιστης έντασης της άσκησης ή στο τεστ κοπώσεως	32
Πίνακας 15	Συχνότητα ύπαρξης καρδιακής πάθησης	33
Πίνακας 16	Test κανονικότητας Shapiro – Wilk	34
Πίνακας 17	Συντελεστής συσχέτισης Spearman	36
Πίνακας 18	Συχνότητα εμφάνισης καρδιακής πάθησης ανά τα φύλα	37
Πίνακας 19	Έλεγχος ανεξαρτησίας $X^2$ μεταξύ φύλου και καρδιακών παθήσεων	38
Πίνακας 20	Συχνότητα εμφάνισης στηθάγχης και καρδιακής πάθησης	39
Πίνακας 21	Έλεγχος ανεξαρτησίας $X^2$ μεταξύ στηθάγχης και καρδιακής πάθησης	39
Πίνακας 22	Συχνότητα εμφάνισης ζαχάρου νηστείας και καρδιακής πάθησης	40
Πίνακας 23	Έλεγχος ανεξαρτησίας $X^2$ μεταξύ ζαχάρου νηστείας και καρδιακής πάθησης	40
Πίνακας 24	Συχνότητα εμφάνισης παλμών στο ηλεκτροκαρδιογράφημα σε κατάσταση ηρεμίας και καρδιακής πάθησης	41

Πίνακας 25	Έλεγχος ανεξαρτησίας $X^2$ μεταξύ παλμών του ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας και καρδιακής πάθησης	42
Πίνακας 26	Συχνότητα εμφάνισης στηθαγχικού πόνου κατά τη διάρκεια της άσκησης και καρδιακής πάθησης.	43
Πίνακας 27	Έλεγχος ανεξαρτησίας $X^2$ μεταξύ στηθαγχικού πόνου κατά τη άσκηση και καρδιακής πάθησης	43
Πίνακας 28	Συχνότητα εμφάνισης αλλαγής του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης της άσκησης ή του στρες και της καρδιακής πάθησης	44
Πίνακας 29	Έλεγχος ανεξαρτησίας $X^2$ μεταξύ της αλλαγής του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης της άσκησης ή του στρες και της καρδιακής πάθησης	44
Πίνακας 30	Έλεγχος ανεξαρτησίας $X^2$ ανά δύο των υπόλοιπων μεταβλητών	45
Πίνακας 31	Συνολικές αναθέσεις συνεχών μεταβλητών ανά φύλο	47
Πίνακας 32	Έλεγχος Mann – Whitney U, Wilcoxon συνεχών μεταβλητών με το φύλο	47
Πίνακας 33	Συνολικές αναθέσεις συνεχών μεταβλητών με ύπαρξη ή όχι καρδιακής πάθησης	48
Πίνακας 34	Έλεγχος Mann – Whitney U, Wilcoxon συνεχών μεταβλητών με την ύπαρξη ή όχι καρδιακής πάθησης	49
Πίνακας 35	Συνολικές αναθέσεις συνεχών μεταβλητών με τα επίπεδα ζαχάρου νηστείας	49
Πίνακας 36	Έλεγχος Mann – Whitney U, Wilcoxon συνεχών μεταβλητών με τα επίπεδα ζαχάρου νηστείας	50
Πίνακας 37	Συνολικές αναθέσεις συνεχών μεταβλητών με στηθάγχη από άσκηση	51
Πίνακας 38	Έλεγχος Mann – Whitney U, Wilcoxon συνεχών μεταβλητών με τα επίπεδα στηθάγχης από άσκηση	51
Πίνακας 39	Έλεγχος Kruskal-Wallis συνεχών μεταβλητών με τα επίπεδα του ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας	52
Πίνακας 40	Μέθοδος bootstrap για προσδιορισμό διαστημάτων εμπιστοσύνης	53
Πίνακας 41	Έλεγχος Kruskal-Wallis συνεχών μεταβλητών με τα επίπεδα της αλλαγής του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης του τεστ κοπώσεως του ηλεκτροκαρδιογραφήματος	55
Πίνακας 42	Μέθοδος bootstrap για προσδιορισμό διαστημάτων εμπιστοσύνης	56
Πίνακας 43	Έλεγχος Kruskal-Wallis συνεχών μεταβλητών με τα επίπεδα του στηθαγχικού πόνου	58
Πίνακας 44	Μέθοδος bootstrap για προσδιορισμό διαστημάτων εμπιστοσύνης	59

Πίνακας 45	Πίνακας συσχετίσεων	63
Πίνακας 46	Έλεγχος καταλληλότητας των μεταβλητών για παραγοντική ανάλυση με τη μέθοδο KMO και Bartlett's Test	64
Πίνακας 47	Έλεγχος καταλληλότητας των μεταβλητών για παραγοντική ανάλυση με τη μέθοδο MSA	64
Πίνακας 48	Πίνακας συνολικής επεξηγηματικής διακύμανσης	65
Πίνακας 49	Κατανομή μεταβλητών στους παράγοντες	66
Πίνακας 50	Έλεγχος προσαρμογής του μοντέλου	67
Πίνακας 51	Έλεγχος καλής προσαρμογής του μοντέλου	68
Πίνακας 52	Έλεγχος ορθής ταξινόμησης	68
Πίνακας 53	Αποτελέσματα λογιστικής παλινδρόμησης	68
Πίνακας 54	Αξιολόγηση διαγνωστικής ικανότητας του μοντέλου	70

## Κατάλογος Σχημάτων

Σχήμα 1	Ιστογράμματα αρτηριακής πίεσης σε κατάσταση ηρεμίας και χοληστερίνης συμπεριλαμβανομένων των μηδενικών τιμών	24
Σχήμα 2	Ιστόγραμμα ηλικίας	25
Σχήμα 3	Ιστόγραμμα χοληστερίνης	26
Σχήμα 4	Ιστόγραμμα μέγιστου καρδιακού ρυθμού	27
Σχήμα 5	Ιστόγραμμα μέγιστης απόκλισης του ηλεκτροκαρδιογραφήματος	27
Σχήμα 6	Ιστόγραμμα αρτηριακής πίεσης σε κατάσταση ηρεμίας	28
Σχήμα 7	Φύλο	29
Σχήμα 8	Στηθαγχικός πόνος	30
Σχήμα 9	Ζάχαρο νηστείας	30
Σχήμα 10	Τύποι ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας	31
Σχήμα 11	Στηθάγχη από άσκηση	32
Σχήμα 12	Αλλαγή μοτίβου κατά το σημείο μέγιστης έντασης της άσκησης ή στο τεστ κοπώσεως	32
Σχήμα 13	Ύπαρξη καρδιακής πάθησης	33
Σχήμα 14	Normal Q – Q Plots	34
Σχήμα 15	Επιλογή αριθμού παραγόντων με βάση τον κανόνα του «αγκώνα»	65
Σχήμα 16	Προβλεπτική ικανότητα μοντέλου με χρήση καμπύλης ROC	71

## Γλωσσάρι

baseline: Είναι το σημείο αναφοράς ή το αρχικό επίπεδο με το οποίο συγκρίνονται τα μελλοντικά δεδομένα.

Δ.Ε: Είναι συντομογραφία για το διάστημα εμπιστοσύνης.



# ΚΕΦΑΛΑΙΟ 1

## Εισαγωγή

### 1.1 Ορισμός καρδιαγγειακών παθήσεων

Οι καρδιαγγειακές παθήσεις είναι μία ομάδα διαταραχών της καρδιάς και των αιμοφόρων αγγείων. Συχνά επηρεάζουν τη λειτουργία της καρδιάς, των αιμοφόρων αγγείων και υποβαθμίζουν την υγεία. Ορισμένες από αυτές είναι η στεφανιαία νόσος (στηθάγχη, οξύ έμφραγμα του μυοκαρδίου, αιφνίδιος καρδιακός θάνατος) <sup>[1]</sup>, η εγκεφαλοαγγειακή νόσος (εγκεφαλικό επεισόδιο, ανεύρισμα, αγγειακές δυσπλασίες) <sup>[2]</sup>, ρευματική καρδιοπάθεια <sup>[3]</sup> και άλλες παθήσεις.

### 1.2 Θνησιμότητα καρδιαγγειακών παθήσεων

Σύμφωνα με τον παγκόσμιο οργανισμό υγείας (ΠΟΥ), οι καρδιαγγειακές παθήσεις είναι η κύρια αιτία θανάτων παγκοσμίως καταγράφοντας 17,9 εκατομμύρια θανάτους το χρόνο. Ακόμα, περισσότεροι από το 80% των θανάτων από καρδιαγγειακά νοσήματα οφείλονται σε καρδιακές προσβολές και εγκεφαλικά επεισόδια καθώς το 33% αυτών συμβαίνουν πρόωρα σε άτομα ηλικίας μικρότερα των 70 ετών.

### 1.3 Διάκριση καρδιαγγειακών παθήσεων

Οι καρδιαγγειακές παθήσεις μπορούν να διακριθούν σε διάφορες κατηγορίες ανάλογα με το ποιο μέρος του καρδιαγγειακού συστήματος επηρεάζουν. Ορισμένες από τις βασικές κατηγορίες περιλαμβάνουν: Καρδιακές παθήσεις, Εγκεφαλικά αγγειακά προβλήματα και Περιφερικά αγγειακά νοσήματα.

### 1.3.1 Τύποι και συμπτώματα καρδιακών ασθενειών

Οι καρδιακές ασθένειες συμπεριλαμβάνουν αρκετούς τύπους καρδιαγγειακών παθήσεων [\[4\]](#). Αν και είναι άρρηκτα συνδεδεμένα με αυτές, διαφοροποιούνται στο γεγονός πως στοχεύουν τις παθήσεις που επηρεάζουν τη δομή και τη λειτουργία της καρδιάς. Οι πιο συνηθισμένοι τύποι αυτών περιλαμβάνουν:

- Στεφανιαία νόσο, που προκαλείται συνήθως από την αθηροσκλήρωση. Έτσι ονομάζεται η διαδικασία κατά την οποία η χοληστερίνη και άλλες ουσίες δημιουργούν την πλάκα η οποία φράσσει τις αρτηρίες και έχει ως αποτέλεσμα τη μείωση της παροχής του αίματος στην καρδιά και στα υπόλοιπα σημεία του σώματος. Αν αυτό πραγματοποιηθεί σε στεφανιαία αρτηρία τότε δημιουργούνται θρόμβοι οι οποίοι μειώνουν ή φράζουν τελείως τη δίοδο προς την καρδιά οδηγώντας έτσι σε καρδιακό επεισόδιο. Η νόσος αυτή κατατάσσεται πρώτη σε συχνότητα εμφάνισης παγκοσμίως [\[5\]](#) με τα κύρια συμπτώματά της να είναι η στηθάγχη και η δυσφορία στο στήθος.
- Καρδιακή προσβολή, γίνεται όταν ένα μέρος του μυ της καρδιάς δεν λαμβάνει επαρκή ποσότητα αίματος με τη στεφανιαία νόσο να είναι η πρώτη στην πρόκληση καρδιακής προσβολής. Μπορεί να προκληθεί κι από σπανιότερες αιτίες όπως είναι οι έντονοι σπασμοί ή η ξαφνική συστολή της στεφανιαίας αρτηρίας που επίσης προκαλούν διακοπή της αιμάτωσης της καρδιάς. Ως συμπτώματα έχει πόνο ή δυσφορία στο στήθος, αδυναμία, ζάλη, λιποθυμία, πόνο ή δυσφορία στο σαγόνι, λαιμό, πλάτη, χέρια ή και ώμους και δύσπνοια.
- Καρδιακή ανεπάρκεια, παρουσιάζεται όταν ο μυς της καρδιάς δεν λειτουργεί όσο καλά θα έπρεπε αδυνατώντας να αντλήσει επαρκή αίμα και οξυγόνο για την κατάλληλη τροφοδοσία των υπόλοιπων οργάνων του σώματος. Τα πιο κοινά συμπτώματα είναι η δύσπνοια κατά τη διάρκεια της καθημερινότητας ή και σε καθιστή θέση, αύξηση βάρους σε συνδυασμό με εφίδρωση στις περιοχές των ποδιών, των αστραγάλων ή και της κοιλιακής χώρας, και τέλος η γενική αίσθηση αδυναμίας ή κόπωσης.
- Κολπική μαρμαρυγή, ανήκει στην ομάδα των καρδιακών αρρυθμιών και προκαλεί προβλήματα στα ηλεκτρικά κυκλώματα της καρδιάς οδηγώντας σε μη φυσιολογικό καρδιακό ρυθμό ή παλμών της καρδιάς. Συγκεκριμένα, ανήκει στην ομάδα των υπερ-



κοιλιακών αρρυθμιών και ανάλογα με τη συχνότητα εμφάνισης διακρίνεται σε τρεις κατηγορίες [\[6\]](#):

- i. Μόνιμη.
- ii. Παροξυσμική, με διάρκεια μικρότερη της μιας εβδομάδας και ρυθμίζεται χωρίς κάποια φαρμακευτική αγωγή ή ιατρική ανάμειξη.
- iii. Εμμένουσα, με διάρκεια μεγαλύτερης της μιας εβδομάδας ή και το πέρας του χρόνου που χρειάζεται όμως κάποια φαρμακευτική αγωγή ή ιατρική ανάμειξη.

Τα συχνότερα συμπτώματα είναι ακανόνιστοι καρδιακοί χτύποι, ταχυπαλμίες, αρρυθμίες, ατονία, αυξημένη κόπωση, ζαλάδες, δύσπνοια και πόνοι στο στήθος.

- Καρδιομυοπάθεια [\[7\]](#), είναι ένα σύνολο ασθενειών που προσβάλλουν τη δομή του καρδιακού μυ δυσχεραίνοντας έτσι το έργο του στην άντληση αίματος και τον εφοδιασμό του υπόλοιπου σώματος με αυτό. Μπορεί να προκαλέσει ανικανότητα στη διατήρηση του κανονικού καρδιακού ρυθμού έχοντας ως αποτέλεσμα την καρδιακή ανεπάρκεια ή τις αρρυθμίες. Ακόμα, επηρεάζει την καρδιά ως προς το μέγεθος, το σχήμα και το πάχος της. Η διάκριση γίνεται στους εξής τύπους:

- i. Κληρονομικοί τύποι καρδιομυοπάθειας:

- Υπερτροφική μυοκαρδιοπάθεια [\[8\]](#), όπου μπορεί να προκληθεί από μη φυσιολογικά γονίδια στον καρδιακό μυ και επηρεάζει κυρίως την αριστερή πλευρά των τοιχωμάτων του μυοκαρδίου κάνοντάς το πιο παχύ. Είναι ακόμα πιθανό να προκαλέσει και αρρυθμίες.
- Αρρυθμογόνος μυοκαρδιοπάθεια δεξιάς κοιλίας (ARVC) [\[9\]](#) ή αλλιώς αρρυθμογενής δυσπλασία της δεξιάς κοιλίας, εμφανίζεται στη δεξιά κοιλία και είναι μια γενετική διαταραχή του μυοκαρδίου κατά την οποία δημιουργείται λιπώδης ιστός καταλαμβάνοντας έτσι τον ελεύθερο χώρο στα τοιχωμάτων της δεξιάς κοιλότητας. Επιπλέον, περιορίζει την παραγωγή και διανομή του αίματος στο σώμα.
- Μη συμπαγής (σπογγώδης) μυοκαρδιοπάθεια (LVNC) είναι μία ανατομική διαταραχή με χαρακτηριστικό σπογγώδες σχήμα, το οποίο προκύπτει από την διακοπή της φυσιολογικής διαδικασίας συμπίεσης του μυοκαρδίου κατά την εμβρυϊκή ανάπτυξη [\[10\]](#).

ii. Άλλοι τύποι μυοκαρδιοπάθειας:

- Takotsubo μυοκαρδιοπάθεια ή αλλιώς σύνδρομο της ραγισμένης καρδιάς, που μπορεί να προκληθεί από συναισθηματική και σωματική πίεση.
- Μυοκαρδιοπάθεια της κύησης, η οποία εμφανίζεται από το τελευταίο τρίμηνο της κύησης ή και μετά.
- Περιοριστική μυοκαρδιοπάθεια, που αλλάζει τα ιστολογικά χαρακτηριστικά της καρδιάς καθώς προκαλεί σκλήρυνση του καρδιακού μυ και συχνά οφείλεται στην καρδιακή αμυλοείδωση.

Τα συνηθέστερα συμπτώματα είναι το αίσθημα κόπωσης, ατονίας, ζάλης, ακανόνιστος καρδιακός ρυθμός, σφίξιμο στο στήθος και πρήξιμο στην κοιλιά ή και στα κάτω άκρα.

- Βαλβιδική καρδιοπάθεια [\[11\]](#), αναφέρεται σε οποιαδήποτε δυσλειτουργία ή ανωμαλία μίας ή περισσότερων από τις βαλβίδες της καρδιάς. Η ανθρώπινη καρδιά έχει τέσσερις βαλβίδες, την αορτική, τη μιτροειδή, την πνευμονική και την τριγλώχινα οι οποίες φροντίζουν να οδηγείται το αίμα προς τη σωστή κατεύθυνση, με αφετηρία την καρδιά και κατεύθυνση προς το υπόλοιπο σώμα. Η μη σωστή λειτουργία μπορεί να προκαλέσει καρδιαγγειακά προβλήματα. Η βαλβιδική καρδιακή νόσος μπορεί να υπάρχει κατά τη γέννηση ή να αναπτυχθεί με την πάροδο του χρόνου από άλλες αιτίες. Ως συμπτώματα περιλαμβάνει τη δυσκολία στην αναπνοή, κούραση, ζαλάδα ή και λιποθυμία, πυρετό, απότομη αύξηση βάρους και ακανόνιστο καρδιακό ρυθμό.

### 1.3.2 Τύποι και συμπτώματα εγκεφαλικών αγγειακών προβλημάτων

Στα εγκεφαλικά αγγειακά προβλήματα [\[12\]](#) ανήκουν οι παθήσεις οι οποίες επηρεάζουν τα αγγεία που τροφοδοτούν τον εγκέφαλο, και η μη σωστή λειτουργία αυτών επηρεάζει την αιμάτωσή του. Με τη ρήξη ή και φραγή (με θρόμβο του αίματος) της εγκεφαλικής αρτηρίας προκαλείται βλάβη στον εγκεφαλικό ιστό καθώς διακόπτεται ή μειώνεται η αιμάτωση του εγκεφάλου. Οι εγκεφαλοαγγειακές παθήσεις εμπεριέχουν:

- Εγκεφαλικό επεισόδιο [\[13\]](#), είναι η απότομη διαταραχή της εγκεφαλικής λειτουργίας εξαιτίας της έλλειψης αίματος ή αιμορραγίας και διακρίνεται σε:

- i. Ισχαιμικό, όταν φράσσεται κάποια αρτηρία με θρόμβο και μειώνεται η παροχή αίματος, οξυγόνου και θρεπτικών ουσιών στον εγκέφαλο.
- ii. Παροδικό ισχαιμικό επεισόδιο, στο οποίο συμβαίνει ό,τι και στο ισχαιμικό, με τη διαφορά ότι είναι προσωρινό και η απόφραξη του αγγείου αποκαθίσταται αυτόματα. Συνίσταται όμως ιατρική βοήθεια καθώς το 10% των ανθρώπων που το βιώνουν παρουσιάζουν εγκεφαλικό επεισόδιο μέσα στις επόμενες δύο εβδομάδες, ενώ το 30% των ανθρώπων με σοβαρό εγκεφαλικό είχαν προγενέστερα πάθει παροδικό ισχαιμικό επεισόδιο.
- iii. Αιμορραγικό, με τη ρήξη κάποιου αγγείου του εγκεφάλου με συνέπεια την πρόκληση αιμορραγίας στον εγκέφαλο και στις γύρω περιοχές.

Τα συμπτώματα των εγκεφαλικών επεισοδίων περιλαμβάνουν ζαλη, ναυτία, εμετό, ασυνήθιστα έντονους πονοκεφάλους, σύγχυση, αποπροσανατολισμό, απώλεια μνήμης, μούδιασμα, αδυναμία σε πρόσωπο ή και άκρα συνήθως στη μία πλευρά του σώματος, δυσκολία στο λόγο, στην κατανόηση συζητήσεων, θόλωση ή και απώλεια όρασης και δυσκολία ή αδυναμία στο περπάτημα.

- Στένωση καρωτίδας, είναι η στένεψη ή η έντονη φραγή των καρωτίδων αρτηριών που προκαλείται από την αθηρωσκλήρωση και είναι πιθανό να προκαλέσει παροδικό ισχαιμικό επεισόδιο. Είναι ασυμπτωματική ασθένεια η οποία εντοπίζεται αρχικά από σωματική εξέταση και η διάγνωσή της γίνεται μετά από εξετάσεις.
- Ανεύρυσμα εγκεφάλου, είναι μία εξασθετισμένη περιοχή όπου ένα αιμοφόρο αγγείο του εγκεφάλου αδυνατίζει και προκαλεί εξόγκωμα ή βγαίνει εκτός των τοιχωμάτων μιας αρτηρίας κοντά ή μέσα στον εγκέφαλο. Παρουσιάζεται κυρίως σε ηλικίες μεταξύ 50 και 60 χρόνων, με μεγαλύτερο κίνδυνο να διατρέχει το γυναικείο φύλο, καθώς έχει τριπλάσιες πιθανότητες εμφάνισης σε σχέση με το αντρικό. Στο ανεύρυσμα μπορεί να προκληθεί ρήξη ή και όχι και ανάλογα με αυτό, χαρακτηρίζονται και τα συμπτώματά [\[14\]](#). Συγκεκριμένα:

- i. Αν δεν έχει επέλθει ρήξη και το μέγεθός του είναι μικρό (< 10 χιλιοστά), τότε είναι πιθανό να μην υπάρχουν συμπτώματα.
- ii. Αν είναι μεγάλο και ανάλογα με την περιοχή του εγκεφάλου που βρίσκεται, μπορεί να προκαλέσει πόνο πίσω και πάνω από το μάτι, διαστολή της κόρης του ματιού, αλλαγή ή θόλωση της όρασης, μούδιασμα, αδυναμία και παράλυση των μυών του

μισού προσώπου. Αν επέλθει ρήξη, τότε μπορεί να υπάρξει ευαισθησία στο φως, διπλή όραση, ναυτία, εμετός, πιάσιμο στον αυχένα, κρίσεις, απώλεια συνείδησης και καρδιακό επεισόδιο.

- Αγγειακή δυσπλασία, είναι μια μη φυσιολογική ανάπτυξη των αγγείων. Όταν συμβαίνει στην περιοχή του εγκεφάλου, τα αγγεία που βρίσκονται εκεί δεν έχουν σωστή ανάπτυξη ή λειτουργία. Μία από τις συνηθέστερες αγγειακές δυσπλασίες στον εγκέφαλο είναι η αρτηριοφλεβώδης δυσπλασία η οποία είναι ένα συνοθήλευμα αρτηριών και φλεβών μαζί, με μεγαλύτερη ροή αίματος απο ό,τι συνήθως. Τα συμπτώματα ποικίλουν ανάλογα με τον τύπο και τη θέση της δυσπλασίας στον εγκέφαλο. Ενδεικτικά, κάποια από αυτά είναι σπασμοί, με ή χωρίς απώλεια συνείδησης, πονοκέφαλος, ναυτία, εμετός, ζάλη, επιληψία, προβλήματα στην κίνηση, τη σκέψη, τη μνήμη, τον διάλογο ή και στην όραση.
- Νόσος Μογανογιο [\[15\]](#), είναι μία σπάνια νευρολογική πάθηση κατά την οποία τα αιμοφόρα αγγεία του εγκεφάλου στενεύουν προκαλώντας ελάττωση του ρυθμού αιμάτωσης του εγκεφάλου. Το όνομά της προέρχεται από την Ιαπωνική λέξη «μογιο μογιο» που μεταφράζεται ως «σύννεφο καπνού» και περιγράφει την εικόνα που προκύπτει από την αναπτυσσόμενη διαταραχή των αγγείων του εγκεφάλου. Μπορεί να εμφανιστεί σε ενήλικες και παιδιά με τα συμπτώματα να χωρίζονται σε ενδοκρανιακή αιμορραγία για τους ενήλικες ενώ για τα παιδιά σε παροδικά ισχαιμικά επεισόδια, εγκεφαλικά, σταδιακή μείωση της γνωστικής λειτουργίας, κρίσεις ή ακούσιες κινήσεις των άκρων τους.

### **1.3.3 Τύποι και συμπτώματα περιφερικών αγγειακών νοσημάτων**

Τα περιφερικά αγγειακά νοσήματα [\[16\]](#) είναι μια αργή και προοδευτική διαταραχή των αιμοφόρων αγγείων κατά την οποία προκαλείται αθηροσκλήρωση. Μπορεί να προσβάλλει οποιοδήποτε αιμοφόρο αγγείο εκτός της καρδιάς, επηρεάζοντας αρνητικά τη λειτουργία των οργάνων που αιματώνονται μέσω αυτών, όπως ο εγκέφαλος και πιο συχνά τα πόδια. Οι τύποι των περιφερικών αγγειακών νοσημάτων περιλαμβάνουν:

- Περιφερική αρτηριακή νόσο, κατά την οποία η συσσώρευση πλάκας εντοπίζεται στις αρτηρίες των χεριών και των ποδιών. Έτσι, διαταράσσεται η ομαλή ροή του αίματος σε αυτές τις περιοχές και, σε ακραίες περιπτώσεις, μπορεί να προκαλέσει γάγγραινα και να οδηγήσει στον ακρωτηριασμό μέρος ή και όλου του σημείου (χέρι, πόδι). Ορισμένα

συμπτώματα είναι ο πόνος κατά το περπάτημα, διαφορά θερμοκρασίας ανάμεσα στα πόδια, αποχρωματισμός, εμφάνιση πληγών και παύση της ανάπτυξης των νυχιών και τριχών.

- Εντερικό ισχαιμικό σύνδρομο, μία σπάνια πάθηση κατά την οποία η πλάκα καταλαμβάνει μέρος των αιμοφόρων αγγείων με κατεύθυνση το γαστρεντερικό σύστημα. Τα συμπτώματα διαφέρουν ανάλογα το είδος [\[17\]](#):
  - i. Χρόνια, περιλαμβάνει πονόκοιλο μετά το γεύμα, απώλεια βάρους, έλλειψη όρεξης φαγητού, εμετό, διάρροια και δυσκοιλιότητα.
  - ii. Οξύ, με συχνό εμετό, συσσωρευση υδρογόνου στο αίμα, αύξηση λευκών αιμοσφαιρίων, αίμα στα κόπρανα και απότομη πτώση της πίεσης.
- Νεφρική αρτηριακή νόσο, προκαλείται από την φραγή των νεφρικών αρτηριών καθώς εμποδίζει την απρόσκοπτη μεταφορά του αίματος από την καρδιά στα νεφρά και μπορεί να οδηγήσει σε υπέρταση, χρόνια νεφρική ασθένεια μέχρι και νεφρική ανεπάρκεια. Συμπτώματα της δυσλειτουργίας των νεφρών είναι η αυπνία, κούραση, πονοκέφαλοι, κατάσταση σύγχυσης, δυσκολία στη συγκέντρωση, ναυτία, εμετό, απώλεια όρεξης, πρήξιμο, μυικές κράμπες, δύσπνοια, μεταβολή της υφής του δέρματος σε πιο ξηρό, άγριο ή και του χρώματος σε πιο σκούρες αποχρώσεις, ανεξήγητη απώλεια βάρους και διακοπή ούρησης ή συχνουρία.
- Σύνδρομο ιγνυακής παγίδευσης προκαλείται όταν η ιγνυακή αρτηρία, η οποία είναι η κύρια πηγή μεταφοράς αίματος στο ύψος του ποδιού κάτω από το γόνατο και ο γαστροκνήμιος μυς (μυς στο πίσω και κάτω μέρος της γάμπας), δεν έχει φυσιολογική πορεία. Μειώνει την παροχή του αίματος στο σημείο αυτό και συναντάται σε αθλητές. Κατά την άσκηση, μπορούν να παρατηρηθούν τα εξής συμπτώματα στο πόδι ή στη γάμπα, αδιάκοπος πόνος, κράμπες, κούραση, αίσθημα βάρους, κάψης και μούδιασματος στο πόδι.
- Φαινόμενο Raynaud [\[18\]](#), είναι μία κατάσταση κατά την οποία ελλατώνεται η ροή του αίματος στα δάχτυλα. Ως έναυσμα έχει τις χαμηλές θερμοκρασίες, το άγχος και τις έντονες συναισθηματικές διαταραχές και διακρίνεται σε:
  - i. Ιδιοπαθές, είναι πιο σύνηθες και δεν συνδιάζει άλλα νοσήματα. Ως σύμπτωμα έχει τον ήπιο αποχρωματισμό, το μούδιασμα και το αίσθημα τρυπήματος.
  - ii. Δευτεροπαθές, είναι πιο σπάνιο και συναντάται με άλλα υποκείμενα ή και αυτοάνοσα νοσήματα, φαρμακευτικές αγωγές και παράγοντες που επηρεάζουν τον

τρόπο ζωής. Μπορεί να προκληθούν και πιο σοβαρά συμπτώματα όπως δερματικά έλκη ή και γάγγραινα.

## 1.4 Ιστορική αναδρομή καρδιαγγειακών παθήσεων

Η παλαιότερη καταγεγραμμένη περίπτωση καρδιαγγειακής πάθησης εντοπίστηκε σε μία πριγκίπισσα της αρχαίας Αιγύπτου που έζησε μεταξύ 1580 με 1550 π.Χ. και έπασχε από αθηροσκλήρωση, μαζί με άλλα σημαντικά πρόσωπα της εποχής αυτής. Έπειτα, με τη μελέτη μουμιοποιημένων σωμάτων της αρχαιότητας αποκαλύφθηκε πως οι καρδιαγγειακές παθήσεις ήταν πιο συχνές από ό,τι θεωρούνταν έως τότε [\[19\]](#). Στα τέλη του Μεσαίωνα (13ος-15ος αιώνας), υπήρξε σταδιακή εξέλιξη της ιατρικής γνώσης με έμφαση στη μελέτη του σώματος, τις θεραπείες ασθενειών αλλά και τη σύνδεση αυτών με το καρδιαγγειακό σύστημα [\[20\]](#) κυρίως από Άραβες ιατρούς. Σημαντικές ανακαλύψεις εντοπίζονται κατά την Αναγέννηση (15ος-16ος αιώνας) με τον Leonardo Da Vinci να συμβάλλει σημαντικά στη μελέτη της καρδιάς μέσω λεπτομερών σκίτσων και παρατηρήσεων. Τον 17ο αιώνα, ο William Harvey περιέγραψε για πρώτη φορά την κυκλοφορία του αίματος [\[21\]](#), γεγονός που οδήγησε σε σημαντική πρόοδο στην κατανόηση του καρδιαγγειακού συστήματος. Τέλος, κατά τον 19ο αιώνα και μετά, υπήρξε μεγάλη εξέλιξη της ιατρικής και της επιστημονικής έρευνας γύρω από τις καρδιαγγειακές παθήσεις, με σημαντικές ανακαλύψεις όπως η ηλεκτροκαρδιογραφία (ECG), καθώς και η ανάπτυξη τεχνικών όπως ο καθετηριασμός και η αγγειογραφία.

## 1.5 Περιγραφή δείγματος

Το δείγμα που χρησιμοποιείται στην παρούσα μελέτη προέρχεται από το Kaggle [\[22\]](#) και είναι το “Heart Failure Prediction Dataset” υπό την άδεια under the Open Database License (ODbL) [\[23\]](#). Πρόκειται για έναν συνδυασμό 5 διαφορετικών datasets, τα οποία εμπεριέχουν 12 κοινές μεταβλητές. Συγκεκριμένα, χρησιμοποιήθηκαν τα σύνολα Cleveland, Hungarian, Switzerland, Long Beach VA και Stalog (Heart) Data Set, με 303, 294, 123, 200 και 270 παρατηρήσεις αντίστοιχα.

Οι μεταβλητές που χρησιμοποιούνται με την κωδικοποίησή τους:

- Age: Ηλικία

- Sex: Φύλο (0: Άντρας, 1: Γυναίκα)
- ChestPainType: Στήθαγχη ή πόνος στο στήθος που χωρίζεται στα εξής επίπεδα:
  - i. TA (Typical Angina): Τυπικός πόνος στο στήθος όπου υπάρχει το αίσθημα πίεσης ή βάρους στο θώρακα και δυναμώνει με την άσκηση (0: TA)
  - ii. ATA (Atypical Angina) : Ύπαρξη πόνου διαφορετικού από τον συνηθισμένο όπως ο οξύς πόνος «σουβλιές», πόνος μεγαλύτερης ή μικρότερης διάρκειας που εντοπίζεται και σε διαφορετικά σημεία όπως την πλάτη. (1: ATA)
  - iii. NAP (Non-Anginal Pain): Ύπαρξη πόνου που μοιάζει με τον καρδιακό αλλά οφείλεται σε διαφορετικές αιτίες όπως ο μυοσκελετικός πόνος. (2: NAP)
  - iv. ASY (Asymptomatic): Καθόλου πόνος (3: ASY)
- RestingBP: Αρτηριακή πίεση σε κατάσταση ηρεμίας
- Cholesterol: Χοληστερίνη
- FastingBS: Ζάχαρο νηστείας. Είναι το επίπεδο της γλυκόζης στην κυκλοφορία του αίματος μετά από περίοδο νηστείας. (0: Διαφορετικά, 1: FastingBS > 120 mg/dl)
- RestingECG: Ηλεκτροκαρδιογράφημα σε κατάσταση ηρεμίας με σκοπό την εξέταση του ρυθμού και της ηλεκτρικής δραστηριότητας της καρδιάς και χωρίζεται στα εξής επίπεδα:
  - i. Normal: Κανονικοί παλμοί (0: Normal)
  - ii. ST (Ανωμαλία των κυμάτων ST-T, όπου αναστροφή των κυμάτων T και η ανύψωση ή κατάθλιψη του ST > 0,05 mV, όπου το 0.05 mV δείχνει την απόκλιση από τη baseline). Αν κάτι είναι μεγαλύτερο από αυτή την τιμή τότε θεωρείται μη φυσιολογικό. (1: ST)
  - iii. LVH: Δείχνει πιθανή ή οριστική υπερτροφία της αριστερής κοιλίας με τα κριτήρια του Estes. (2: LVH)
- MaxHR: Ο μέγιστος καρδιακός ρυθμός.
- ExerciseAngina (Exercise-induced angina): Πόνος στο στήθος που προκαλείται από την άσκηση με μονολεκτικές απαντήσεις. (0: Όχι, 1: Ναι)
- Oldpeak: Μετράει τη μέγιστη μεταβολή που εμφανίζεται στο ηλεκτροκαρδιογράφημα κατά τη διαδικασία του τεστ κοπώσεως. Όσο μεγαλύτερη η μεταβολή, τόσο πιο πιθανή η ύπαρξη καρδιαγγειακής πάθησης με κανονικές τιμές να θεωρούνται οι  $\pm 0.5$  από τη βάση.

- ST\_Slope: Περιγράφει την αλλαγή του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης του τεστ κοπώσεως στο ηλεκτροκαρδιογράφημα.
  - i. Up (Upsloping): Δείχνει μια θετική ή ανοδική κλίση κατά το τεστ κοπώσεως και περιγράφεται ως λιγότερο ανησυχητική επίπτωση του τεστ.(1: Up)
  - ii. Flat: Αν το ST παραμένει σταθερό, συνίστανται περαιτέρω εξετάσεις, καθώς τα αποτελέσματα μπορεί να διαφέρουν. (2: Flat)
  - iii. Down (Downsloping): Αν η εικόνα στο ST είναι φθίνουσα, τότε θεωρείται πιο σημαντικό εύρημα και μπορεί να είναι ένδειξη καρδιακής ισχαιμίας κατά τη διάρκεια της άσκησης. (3: Down)
- HeartDisease: Αν πάσχει από κάποια καρδιακή ασθένεια ή όχι. (0: Υγιή, 1:Υπαρξη καρδιακής πάθησης)

## ΚΕΦΑΛΑΙΟ 2

### Έλεγχος και περιγραφική στατιστική ανάλυση

#### 2.1 Εισαγωγή

Αρχικά, θα πραγματοποιηθεί έλεγχος των δεδομένων για ελλειπίες και διπλότυπες τιμές. Στη συνέχεια, θα εφαρμοστεί μια πρώτη ανάλυση βάσει περιγραφικών μέτρων, με σκοπό την καλύτερη αποτύπωση του δείγματος χρησιμοποιώντας το στατιστικό πακέτο IBM SPSS Statistics 26.

#### 2.2 Έλεγχος ελλειπών και διπλότυπων τιμών

Από τον έλεγχο που πραγματοποιήθηκε, εντοπίζονται 918 εγγραφές χωρίς ελλειπίες και διπλότυπες τιμές όπως φαίνεται και από τους παρακάτω πίνακες.



Indicator of each last matching case as Primary		
N	Valid	918
	Missing	0

Πίνακας 1: Έλεγχος ελλιπών τιμών.

Indicator of each last matching case as Primary					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Primary Case	918	100,0	100,0	100,0

Πίνακας 2: Έλεγχος διπλότυπων τιμών.

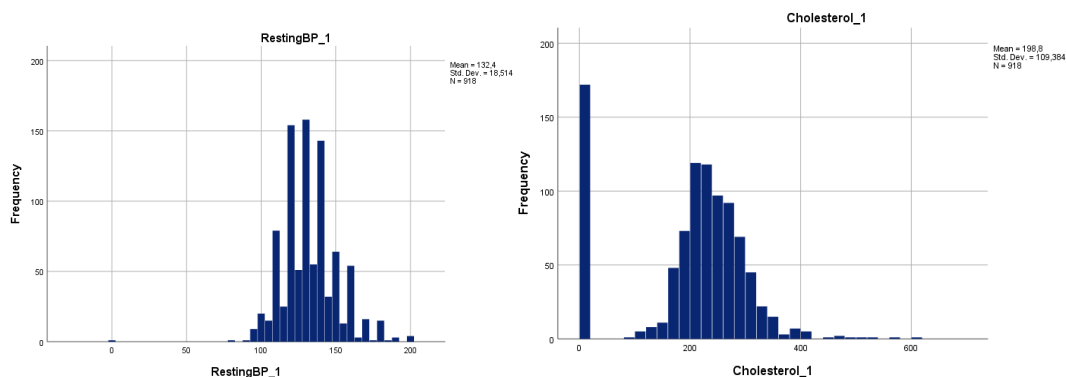
## 2.3 Συνεχείς μεταβλητές

### 2.3.1 Εύρεση μη λογικών τιμών

Ξεκινώντας με την ανάλυση των συνεχών μεταβλητών, από τον παρακάτω πίνακα, παρατηρείται η ύπαρξη μηδενικών τιμών στις μεταβλητές της αρτηριακής πίεσης σε κατάσταση ηρεμίας και της χοληστερίνης. Αυτό πιθανότατα οφείλεται σε εσφαλμένη μέτρηση ή λανθασμένη καταχώριση των δεδομένων, καθώς είναι αδύνατον για ένα άτομο να βρίσκεται εν ζωή χωρίς καρδιακό παλμό και μηδενική τιμή χοληστερίνης.

	Mean	Median	Std. Deviation	Minimum	Maximum
RestingBP_1	132,40	130	18,51	0	200
Cholesterol_1	198,80	223	109,38	0	603

Πίνακας 3: Περιγραφικά μέτρα αρτηριακής πίεσης σε κατάσταση ηρεμίας και χοληστερίνης.



Σχήμα 1: Ιστογράμματα αρτηριακής πίεσης σε κατάσταση ηρεμίας και χοληστερίνης συμπεριλαμβανομένων των μηδενικών τιμών.

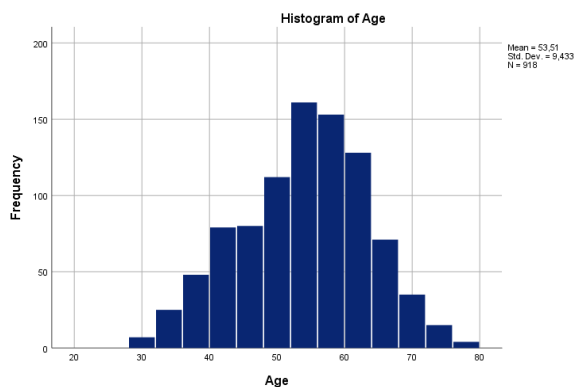
Για το λόγο αυτό, στη συνέχεια της ανάλυσης, θα πραγματοποιηθεί η αντικατάσταση των μηδενικών αυτών τιμών με την αντίστοιχη μέση τιμή των μεταβλητών τους.

### 2.3.2 Ηλικία

Από τον παρακάτω πίνακα φαίνεται πως το ηλικιακό πλαίσιο του δείγματος κυμαίνεται από τα 28 έως και τα 77 έτη, με μέση ηλικία τα 53,5 έτη. Η διάμεσος είναι 54 που υποδηλώνει μία σχεδόν συμμετρική κατανομή της ηλικίας. Η τυπική απόκλιση είναι 9,43 δηλαδή, οι ηλικίες των ατόμων του δείγματος διαφέρουν κατά μέσο όρο 9,43 έτη από τη μέση τιμή. Τέλος, το πηλίκο του συντελεστή προς το τυπικό σφάλμα ασυμμετρίας ισούται με -2,398 (<-2), δείχνοντας έτσι ένα λεπτόκυρτο δείγμα ηλικιών. Αυτό σημαίνει πως οι τιμές συγκεντρώνονται κοντά στην κορυφή.

Variable	Observations	Mean	Median	Std. Deviation	Kurtosis	Std. Error of Kurtosis	Min	Max
Age	918	53,51	54	9,43	-0,386	0,161	28	77

Πίνακας 4: Περιγραφικά μέτρα ηλικίας.



Σχήμα 2: Ιστόγραμμα ηλικίας

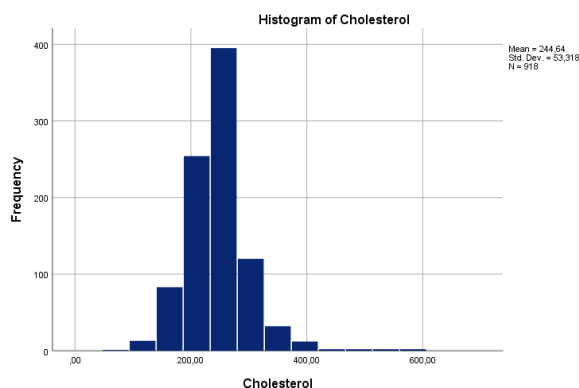
### 2.3.3 Χοληστερίνη

Οι τιμές της χοληστερίνης ξεκινούν από τα 85mm/dl και φτάνουν έως τα 603mm/dl όπως φαίνεται και στον ακόλουθο πίνακα. Η μέση τιμή ταυτίζεται με τη διάμεσο στα 244,64 mm/dl υποδηλώνοντας έτσι μια συμμετρική κατανομή των τιμών της χοληστερίνης. Η τυπική απόκλιση εντοπίζεται στα 53,32 mm/dl όπου κατά μέσο όρο, οι τιμές της απέχουν 53,32 mm/dl από τη μέση

τιμή τους. Ακόμα, το πηλίκο του συντελεστή προς το τυπικό σφάλμα ασυμμετρίας είναι 38,87 ( $>2$ ) παρουσιάζοντας θετική ασυμμετρία δηλαδή, το μέγιστο πλήθος των τιμών της χοληστερίνης εντοπίζεται κάτω από τη μέση τιμή του δείγματος αυτού.

Variable	Observations	Mean	Median	Std. Deviation	Kurtosis	Std. Error of Kurtosis	Min	Max
Cholesterol	918	244,64	244,64	53,32	6,258	0,161	85	603

Πίνακας 5: Περιγραφικά μέτρα χοληστερίνης.



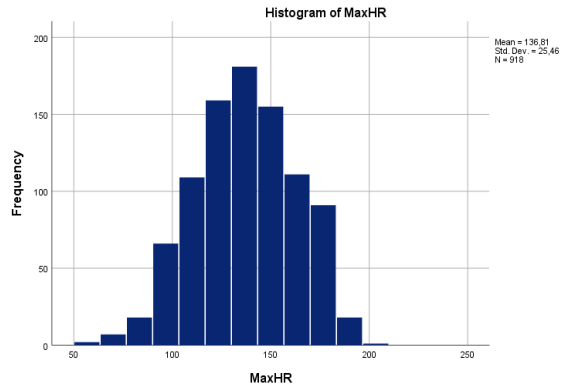
Σχήμα 3: Ιστογράμμα χοληστερίνης.

### 2.3.4 Μέγιστος καρδιακός ρυθμός

Βάσει του επόμενου πίνακα, ο μέγιστος καρδιακός ρυθμός βρίσκεται μεταξύ των 60 και 202 παλμών ανά λεπτό. Η μέση τιμή είναι 136,81 παλμοί ανά λεπτό και η διάμεσος 138 παλμοί ανά λεπτό δείχνοντας έτσι μια σχεδόν συμμετρική κατανομή. Η τυπική απόκλιση είναι 25,46 παλμοί ανά λεπτό, δηλαδή κατά μέσο όρο ο μέγιστος παλμός ανά λεπτό διαφέρει 25,46 παλμούς από τη μέση τιμή αυτών. Το πηλίκο του συντελεστή προς το τυπικό σφάλμα ασυμμετρίας είναι -2,783 ( $<-2$ ) οδηγώντας σε λεπτόκυρτο δείγμα, καθώς μεγάλο πλήθος των τιμών του συγκεντρώνονται κοντά στην κορυφή του δείγματος.

Variable	Observations	Mean	Median	Std. Deviation	Kurtosis	Std. Error of Kurtosis	Min	Max
MaxHR	918	136,81	138	25,46	-0,448	0,161	60	202

Πίνακας 6: Περιγραφικά μέτρα μέγιστου καρδιακού ρυθμού.



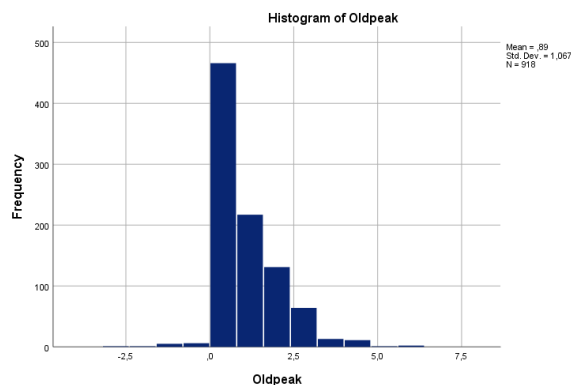
Σχήμα 4: Ιστόγραμμα μέγιστου καρδιακού ρυθμού.

### 2.3.5 Μέγιστη απόκλιση ηλεκτροκαρδιογραφήματος κατά το τεστ κοπώσεως

Η ελάχιστη πτώση από το baseline εντοπίζεται στα -2,6 mm και η μέγιστη αύξηση στα 6,2 mm. Η μέση τιμή αυτών είναι 0,89 και η διασπορά 0,6 μονάδες παρουσιάζοντας ασυμμετρία προς τα δεξιά. Η τυπική απόκλιση είναι 1,07 μονάδες δηλαδή, οι τιμές της μέγιστης απόκλισης του ηλεκτροκαρδιογραφήματος, διαφέρουν κατά μέσο όρο 1,07 mm από τη μέση τιμή τους. Το πηλίκο του συντελεστή προς το τυπικό σφάλμα ασυμμετρίας είναι 7,472 (>2) εμφανίζοντας έτσι θετική ασυμμετρία καθώς το μέγιστο πλήθος των τιμών αυτών βρίσκονται κάτω από τη μέση του τιμή.

Variable	Observations	Mean	Median	Std. Deviation	Kurtosis	Std. Error of Kurtosis	Min	Max
Oldpeak	918	0,89	0,6	1,07	1,203	0,161	-2,6	6,2

Πίνακας 7: Περιγραφικά μέτρα μέγιστης απόκλισης του ηλεκτροκαρδιογραφήματος.



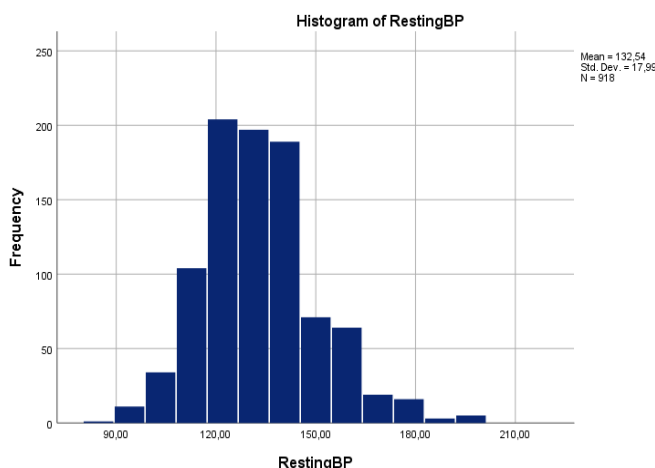
Σχήμα 5: Ιστόγραμμα μέγιστης απόκλισης του ηλεκτροκαρδιογραφήματος.

### 2.3.6 Αρτηριακή πίεση σε κατάσταση ηρεμίας

Η αρτηριακή πίεση σε κατάσταση ηρεμίας όπως φαίνεται από τον ακόλουθο πίνακα, λαμβάνει τιμές από 80 έως 200 mmHg με τη μέση τιμή και διάμεσο αυτών να είναι 132,54 mmHg και 130 mmHg αντίστοιχα. Αυτό υποδηλώνει μια σχεδόν συμμετρική κανομή της αρτηριακής πίεσης σε κατάσταση ηρεμίας. Η τυπική απόκλιση είναι 17.99 mmHg δηλαδή, η μέση αρτηριακή πίεση απέχει κατά μέσο όρο 17,99 mmHg από τη μέση τιμή τους. Επιπρόσθετα, το πηλίκο του συντελεστή προς το τυπικό σφάλμα ασυμμετρίας είναι 4,944 (>2) και παρουσιάζει θετική ασυμμετρία. Δηλαδή, το μέγιστο πλήθος των τιμών της αρτηριακής πίεσης σε κατάσταση ηρεμίας εντοπίζεται κάτω από τη μέση τιμή του δείγματός του.

Variable	Observations	Mean	Median	Std. Deviation	Kurtosis	Std. Error of Kurtosis	Min	Max
RestingBP	918	132,54	130	17,99	0,796	0,161	80	200

Πίνακας 8: Περιγραφικά μέτρα αρτηριακής πίεσης σε κατάσταση ηρεμίας.



Σχήμα 6: Ιστόγραμμα αρτηριακής πίεσης σε κατάσταση ηρεμίας.

## 2.4 Κατηγορικές μεταβλητές

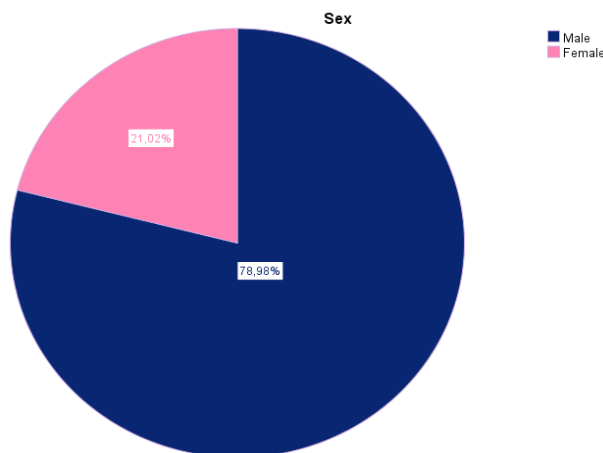
### 2.4.1 Φύλο

Από τον παρακάτω πίνακα παρατηρείται πως από το σύνολο των 918 εγγραφών του δείγματος, οι 725 είναι άντρες, αντιπροσωπεύοντας το 78.98% και οι υπόλοιπες 193 εγγραφές

αντιστοιχούν σε γυναίκες, συμπληρώνοντας το υπόλοιπο 21.02%. Αυτό δείχνει ότι η πλειοψηφία του δείγματος αποτελείται από άντρες.

	Frequency	Valid Percent
Male	725	78,98
Female	193	21,02
Total	918	100,00

Πίνακας 9: Συχνότητα φύλου



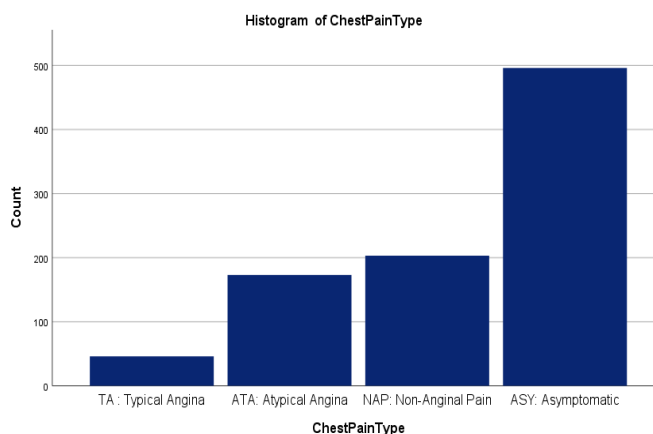
Σχήμα 7: Φύλο

## 2.4.2 Στηθάγχη

Όσον αφορά τον τύπο στηθάγχης (ChestPainType), από τον πίνακα που ακολουθεί παρατηρείται πως από το σύνολο των 918 ατόμων, τα 46 άτομα, το 5,01% των συνολικών καταχωρησεων, εμφάνισαν τυπική στηθάγχη (TA), τα 173, το 18,85%, άτυπη στηθάγχη (ATA), τα 203, το 22,11%, μη στηθαγχικό πόνο (NAP) και τα υπόλοιπα 496, το 54,03% δεν εμφάνισαν πόνο (ASY). Τα δεδομένα δείχνουν ότι η πλειοψηφία των συμμετεχόντων ήταν ασυμπτωματικοί, ενώ το μικρότερο ποσοστό εμφάνισε τυπικό στηθαγχικό πόνο.

	Frequency	Valid Percent
<b>TA : Typical Angina</b>	46	5,01
<b>ATA: Atypical Angina</b>	173	18,85
<b>NAP: Non-Anginal Pain</b>	203	22,11
<b>ASY: Asymptomatic</b>	496	54,03
<b>Total</b>	918	100,00

Πίνακας 10: Συχνότητα στηθαγγχικού πόνου



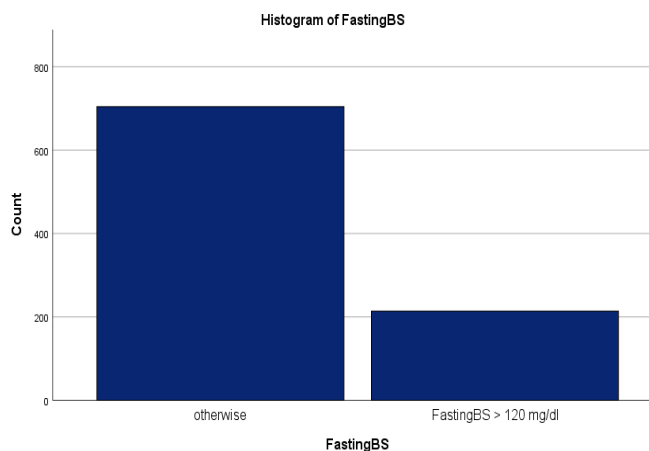
Σχήμα 8: Στήθαγγχικός πόνος

### 2.4.3 Ζάχαρο νηστείας

Από τον κάτω πίνακα παρατηρείται πως από τις 918 εγγραφές, το ζάχαρο νηστείας των 704 ατόμων που αντιπροσωπεύει το 76,69% του συνολικού δείγματος, έχει τιμές από 120mg/dl και κάτω. Τα υπόλοιπα 214 άτομα, το 23,31%, παρουσιάζουν αυξημένη τιμή πάνω από 120mg/dl. Γίνεται αντιληπτό πως το κυρίαρχο σύνολο των συμμετεχόντων έχει χαμηλότερα από 120mg/dl επίπεδα ζαχάρου νηστείας, ενώ υπάρχει κι ένα μικρότερο ποσοστό της τάξης του 23.31% με αυξημένα επίπεδα.

	Frequency	Valid Percent
<b>otherwise</b>	704	76,69
<b>FastingBS &gt; 120 mg/dl</b>	214	23,31
<b>Total</b>	918	100,00

Πίνακας 11: Συχνότητα ζαχάρου νηστείας



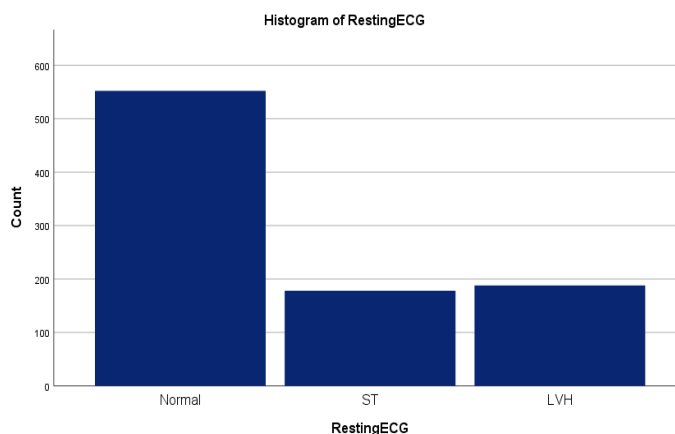
Σχήμα 9: Ζάχαρο νηστείας

## 2.4.4 Ηλεκτροκαρδιογράφημα σε κατάσταση ηρεμίας

Βάσει του πίνακα παρακάτω, από τους 918 συμμετέχοντες στην έρευνα, οι τιμές των ηλεκτροκαρδιογραφήματων σε κατάσταση ηρεμίας (RestingECG) των ατόμων με φυσιολογικές τιμές (Normal) προέρχονται από 552 άτομα, καλύπτοντας το 60,13% του δείγματος. Υπάρχουν επίσης 178 άτομα, το 19,39% με ένδειξη κάποιας ανωμαλίας του ST τμήματος, και 188 άτομα, το 20,48% με παρουσία πιθανής ή μόνιμης υπερτροφίας της αριστερής κοιλίας (LVH). Φαίνεται ότι οι περισσότεροι συμμετέχοντες έχουν φυσιολογικό ηλεκτροκαρδιογράφημα σε κατάσταση ηρεμίας, ενώ ένα σημαντικό ποσοστό εμφανίζει είτε ανωμαλίες του ST τμήματος είτε υπερτροφία της αριστερής κοιλίας.

	Frequency	Valid Percent
Normal	552	60,13
ST	178	19,39
LVH	188	20,48
Total	918	100,00

Πίνακας 12: Συχνότητα τύπων ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας



Σχήμα 10: Τύποι ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας

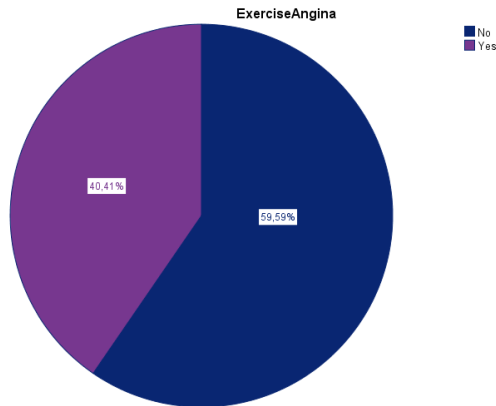
## 2.4.5 Πρόκληση στηθάγχης εξαιτίας σωματικής άσκησης

Σύμφωνα με τον παρακάτω πίνακα, τα αποτελέσματα της ύπαρξης ή όχι στηθαγχικού πόνου από άσκηση (ExerciseAngina) για τους 918 συμμετέχοντες στην έρευνα διαμοιράζονται ως εξής: 371 άτομα, το 40,41% αναφέρουν θετική απάντηση στο βίωμα πόνου, ενώ 547 άτομα, το 59,59% αναφέρουν αρνητική απάντηση. Με αυτόν τον τρόπο, φαίνεται πως η πλειοψηφία των συμμετεχόντων δεν ένιωθε στηθαγχικό πόνο κατά τη διάρκεια της άσκησης.



	Frequency	Valid Percent
<b>Yes</b>	371	40,41
<b>No</b>	547	59,59
<b>Total</b>	918	100,00

Πίνακας 13: Συχνότητα ύπαρξης στηθάγχης από άσκηση



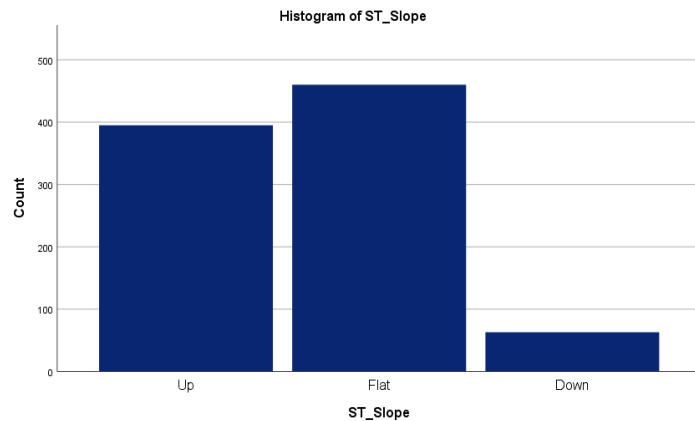
Σχήμα 11: Στηθάγχη από άσκηση

#### 2.4.6 Αλλαγή μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης του τεστ κοπώσεως στο ηλεκτροκαρδιογράφημα

Με βάση τον παρακάτω πίνακα, η αλλαγή του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης του τεστ κοπώσεως του ηλεκτροκαρδιογραφήματος των 918 συμμετεχόντων έδειξε πως τα 395 άτομα, το 43,03% του δείγματος, παρουσιάζουν ανοδική τάση του ST (Up), που συνήθως παραπέμπει σε φυσιολογική καρδιακή λειτουργία. Τα 460 άτομα, το 50,11%, έχουν επίπεδη κλίση του ST (Flat) υποδεικνύοντας πιθανή ισχαιμία ή στεφανιαία νόσο ενώ τα υπόλοιπα 63 άτομα, το 6,86%, έχουν καθοδική κλίση του ST (Down) που τις περισσότερες φορές δείχνει σοβαρή ισχαιμία ή στεφανιαία νόσο.

	Frequency	Valid Percent
<b>Up</b>	395	43,03
<b>Flat</b>	460	50,11
<b>Down</b>	63	6,86
<b>Total</b>	918	100,00

Πίνακας 14: Συχνότητα αλλαγής μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης της άσκησης ή στο τεστ κοπώσεως.



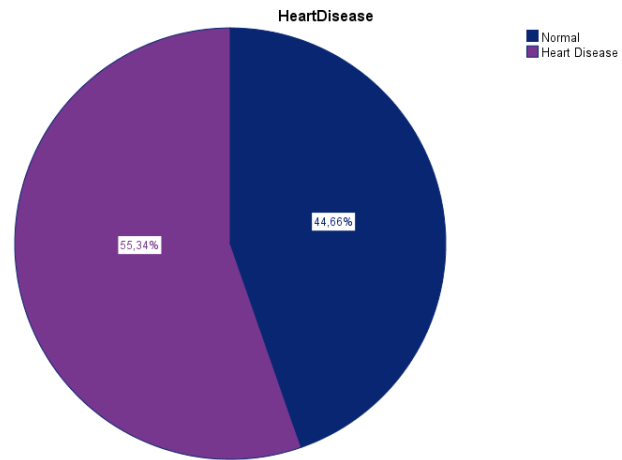
Σχήμα 12: Αλλαγή μοτίβου κατά το σημείο μέγιστης έντασης της άσκησης ή στο τεστ κοπώσεως.

### 2.4.7 Ύπαρξη καρδιακής πάθησης

Από τον ακόλουθο πίνακα φαίνεται πως από τα 918 άτομα, τα 410, το 44,66%, δεν παρουσιάζουν κάποια καρδιακή πάθηση, ενώ τα υπόλοιπα 508 άτομα, το 55,34% έχουν διαγνωστεί. Συνεπώς, η πλειονότητα των συμμετεχόντων πάσχει από κάποια ασθένεια της καρδιάς.

	Frequency	Valid Percent
Normal	410	44,66
Heart Disease	508	55,34
Total	918	100

Πίνακας 15: Συχνότητα ύπαρξης καρδιακής πάθησης.



Σχήμα 13: Ύπαρξη καρδιακής πάθησης.

## ΚΕΦΑΛΑΙΟ 3

### Επαγωγική στατιστική συμπερασματολογία

#### 3.1 Εισαγωγή

Στο κεφάλαιο αυτό θα εξεταστούν οι διάφορες μεταβλητές του δείγματος με σκοπό τη διεξαγωγή σημαντικών πληροφοριών αναφορικά με την επίδρασή τους στην εμφάνιση καρδιακών νόσων. Ως επίπεδο σημαντικότητας ( $\alpha$ ) για όλους τους ελέγχους ορίζεται η τιμή 0,05.

### 3.2 Έλεγχος κανονικότητας συνεχών μεταβλητών

Για την πραγματοποίηση του Shapiro – Wilk test ορίζονται ως μηδενική και εναλλακτική υπόθεση οι παρακάτω προτάσεις:

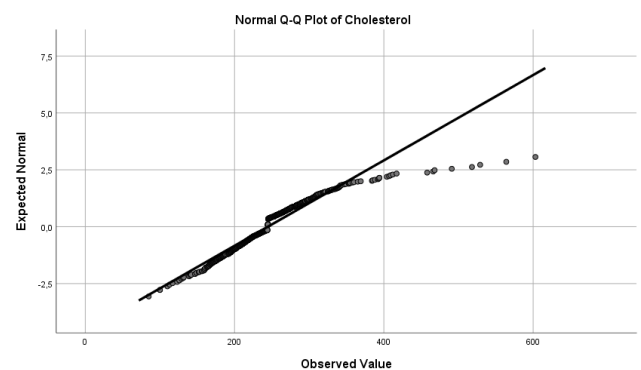
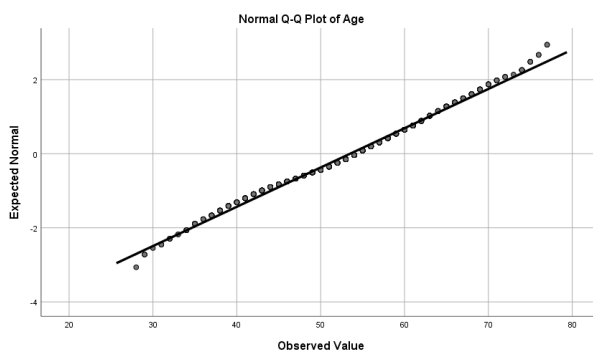
**H<sub>0</sub>:** Η κατανομή των δεδομένων είναι κανονική.

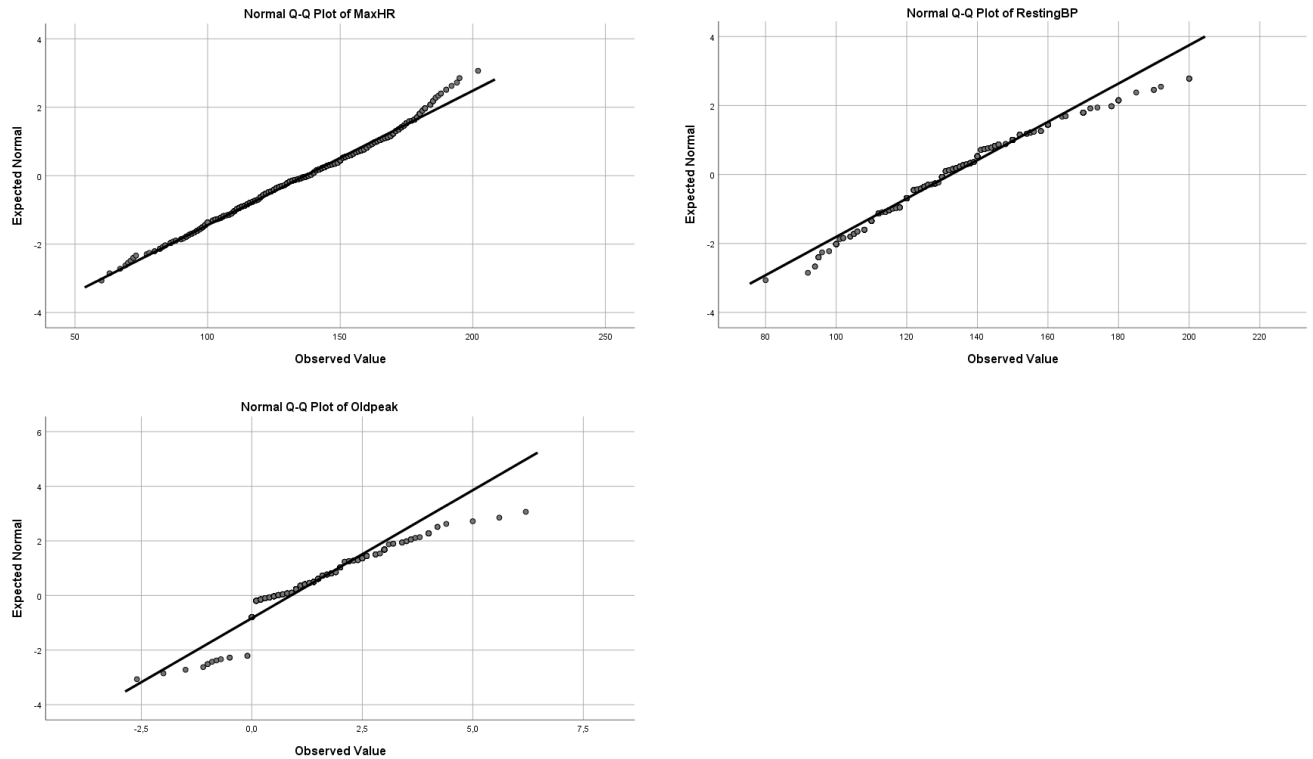
**H<sub>a</sub>:** Η κατανομή των δεδομένων δεν είναι κανονική.

Tests of Normality						
	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Age	,063	918	,000	,991	918	,000
MaxHR	,047	918	,000	,993	918	,000
Oldpeak	,212	918	,000	,860	918	,000
Cholesterol	,137	918	,000	,912	918	,000
RestingBP	,100	918	,000	,971	918	,000

a. Lilliefors Significance Correction

Πίνακας 16: Test κανονικότητας Shapiro – Wilk





Σχήμα 14: Normal Q – Q Plots

Βάσει του παραπάνω πίνακα, όλες οι τιμές pvalue των συνεχών μεταβλητών του Shapiro – Wilk test βρίσκονται πολύ κοντά στην τιμή  $0 < \alpha = 0,05$ . Συνεπώς, σε επίπεδο σημαντικότητας 5% απορρίπτεται η μηδενική υπόθεση  $H_0$  και υποθέτουμε πως τα δεδομένα των μεταβλητών της ηλικίας, χοληστερίνης, μέγιστου καρδιακού ρυθμού, της αρτηριακής πίεσης σε κατάσταση ηρεμίας και της μέγιστης μεταβολής του ηλεκτροκαρδιογραφήματος κατά το τεστ κοπώσεως δεν ακολουθούν την κανονική κατανομή.

Ο ισχυρισμός αυτός ενισχύεται και από τις παραπάνω γραφικές απεικονίσεις των Q – Q Plots των μεταβλητών αυτών, καθώς σε όλες τις περιπτώσεις υπάρχει απόκλιση των σημείων κοντά στην ευθεία γραμμή. Πιο συγκεκριμένα, για τις μεταβλητές Age, MaxHR και RestingBP, τα περισσότερα σημεία αυτών βρίσκονται κοντά στην ευθεία και εντοπίζονται μικρές αποκλίσεις στα άκρα τους. Αντίθετα, για τις υπόλοιπες μεταβλητές Cholesterol και Oldpeak, τα σημεία αποκλίνουν σημαντικά από την ευθεία σε περισσότερα μέρη.

### 3.3 Έλεγχος συσχετίσεων μη κανονικά κατανομημένων μεταβλητών

Για την εξέταση της σχέσης μεταξύ των συνεχών μεταβλητών του δείγματος που δεν ακολουθούν κανονική κατανομή γίνεται χρήση του συντελεστή συσχέτισης Spearman. Ορίζονται ως μηδενική και εναλλακτική υπόθεση τα εξής:

$H_0$ : Δεν υπάρχει συσχέτιση μεταξύ των μεταβλητών ( $\rho = 0$ ).

$H_a$ : Υπάρχει συσχέτιση μεταξύ των μεταβλητών ( $\rho \neq 0$ ).

		Correlations					
		Age	MaxHR	Oldpeak	Cholesterol	RestingBP	
Spearman's rho	Age	Correlation Coefficient	1,000	-,365**	,298**	,086**	,280**
		Sig. (2-tailed)	.	,000	,000	,009	,000
		N	918	918	918	918	918
	MaxHR	Correlation Coefficient	-,365**	1,000	-,205**	-,021	-,106**
		Sig. (2-tailed)	,000	.	,000	,528	,001
		N	918	918	918	918	918
	Oldpeak	Correlation Coefficient	,298**	-,205**	1,000	,072*	,177**
		Sig. (2-tailed)	,000	,000	.	,029	,000
		N	918	918	918	918	918
	Cholesterol	Correlation Coefficient	,086**	-,021	,072*	1,000	,073*
		Sig. (2-tailed)	,009	,528	,029	.	,026
		N	918	918	918	918	918
	RestingBP	Correlation Coefficient	,280**	-,106**	,177**	,073*	1,000
		Sig. (2-tailed)	,000	,001	,000	,026	.
		N	918	918	918	918	918
		**. Correlation is significant at the 0.01 level (2-tailed).					
		*. Correlation is significant at the 0.05 level (2-tailed).					

Πίνακας 17: Συντελεστής συσχέτισης Spearman

Με βάση τον πάνω πίνακα και τις ανά δύο περιπτώσεις που εξετάστηκαν, μόνο το ζεύγος των μεταβλητών Cholesterol – MaxHR, έχει τιμή pvalue = 0,528 >  $\alpha=0,05$ . Επομένως, σε επίπεδο σημαντικότητας 5% δεν υπάρχουν επαρκείς ενδείξεις για την απόρριψη της μηδενικής υπόθεσης  $H_0$ . Αυτό δείχνει πως οι τιμές της χοληστερίνης δεν επηρεάζονται από το μέγιστο καρδιακό ρυθμό.

Αντίθετα, τα υπόλοιπα ζεύγη έχουν τιμές pvalue μικρότερες του  $\alpha = 0,05$  και έτσι, σε επίπεδο σημαντικότητας 5% έχουμε επαρκείς ενδείξεις για την απόρριψη της μηδενικής υπόθεσης  $H_0$  και να θεωρήσουμε πως υπάρχει συσχέτιση μεταξύ αυτών. Οι τιμές του δείκτη κυμαίνονται μακριά από το  $\pm 1$  κι έτσι, η συσχέτιση που υπάρχει είναι ασθενής. Πιο αναλυτικά, φαίνεται πως όσο αυξάνεται η ηλικία ενός ατόμου του δείγματος, αυξάνονται επίσης οι τιμές της μέγιστη απόκλιση του ηλεκτροκαρδιογραφήματός του, οι τιμές της χοληστερίνης του καθώς και η αρτηριακή του πίεση σε κατάσταση ηρεμίας αλλά μειώνεται ο μέγιστος καρδιακός ρυθμός του. Συνεχίζοντας, όσο αυξάνονται οι τιμές του μέγιστου καρδιακού ρυθμού του, μειώνονται και οι τιμές της μέγιστη απόκλιση του ηλεκτροκαρδιογραφήματός του και της αρτηριακής του πίεσης σε κατάσταση ηρεμίας. Επιπρόσθετα, όσο αυξάνεται η μέγιστη απόκλιση του ηλεκτροκαρδιογραφήματός του, αυξάνονται και οι τιμές της χοληστερίνης και της αρτηριακής του πίεσης σε κατάσταση ηρεμίας. Τέλος, όσο αυξάνονται οι τιμές της χοληστερίνης του, αυξάνεται και η αρτηριακή του πίεση σε κατάσταση ηρεμίας.

### 3.4 Έλεγχοι ανεξαρτησίας κατηγορικών μεταβλητών

Για τους παρακάτω ελέγχους ορίζονται οι εξής υποθέσεις:

**$H_0$ :** Οι υπό εξέταση μεταβλητές είναι ανεξάρτητες μεταξύ τους.

**$H_a$ :** Υπάρχει σχέση μεταξύ των υπό μελέτη μεταβλητών.

Για την ύπαρξη συσχέτισης μεταξύ των υπο μελέτη μεταβλητών δύο επιπέδων, θα υπολογιστεί επιπρόσθετα ο λόγος των πιθανοτήτων εμφάνισης του αποτελέσματος υπό την παρουσία και την απουσία του παράγοντα (odds ratio).

#### 3.4.1 Έλεγχος ανεξαρτησίας φύλου και καρδιακής πάθησης

Στον έλεγχο αυτό θα εξεταστεί η ύπαρξη σχέσης μεταξύ του φύλου και της εμφάνισης καρδιακής πάθησης.

Sex * HeartDisease Crosstabulation	
Count	

		HeartDisease		Total
		Normal	Heart Disease	
Sex	Male	267	458	725
	Female	143	50	193
Total		410	508	918

Πίνακας 18: Συχνότητα εμφάνισης καρδιακής πάθησης ανά τα φύλα.

Chi-Square Tests					
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	85,646 <sup>a</sup>	1	,000		
Continuity Correction <sup>b</sup>	84,145	1	,000		
Likelihood Ratio	87,168	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	85,553	1	,000		
N of Valid Cases	918				
a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 86,20.					
b. Computed only for a 2x2 table					

Πίνακας 19: Έλεγχος ανεξαρτησίας  $X^2$  μεταξύ φύλου και καρδιακών παθήσεων.

Από τους πίνακες παραπάνω, γίνεται αντιληπτό πως από τους 725 άντρες του δείγματος, οι 267 δεν έπασχαν από κάποια καρδιακή ασθένεια ενώ οι υπόλοιποι 458 έπασχαν. Αντίστοιχα, από τις 193 γυναίκες, οι 143 ήταν υγιείς και μόλις οι 50 έπασχαν από κάποια καρδιακή πάθηση.

Από τον έλεγχο  $X^2$  που πραγματοποιήθηκε, η τιμή pvalue είναι πολύ κοντά στο  $0 < \alpha = 0,05$  άρα, σε επίπεδο σημαντικότητας 5%, απορρίπτεται η  $H_0$ . Αυτό σημαίνει πως το φύλο μπορεί να σχετίζεται με την πιθανότητα εμφάνισης καρδιακής πάθησης.

Πιο συγκεκριμένα, θα γίνει χρήση του λόγου odds ratio για τη εύρεση της πιθανότητας αυτής.

$$\text{Odds ratio} = (458 \cdot 143) / (267 \cdot 50) = 4,91$$

Συνεπώς, η σχετική πιθανότητα ένας άντρας να πάσχει από κάποια καρδιακή πάθηση είναι περίπου 4,91 φορές μεγαλύτερη από την σχετική πιθανότητα να πάσχει μια γυναίκα. Δηλαδή, οι άνδρες έχουν σημαντικά περισσότερες πιθανότητες να πάσχουν από κάποια καρδιακή πάθηση σε σύγκριση με τις γυναίκες.

### 3.4.2 Έλεγχος ανεξαρτησίας στηθάγχης και καρδιακής πάθησης

Για τον έλεγχο αυτό, θα διερευνηθεί η ύπαρξη ή όχι κάποιας σχέσης μεταξύ της στηθάγχης και της εμφάνισης καρδιακής πάθησης.

ChestPainType * HeartDisease Crosstabulation				
Count				
		HeartDisease		Total
		Normal	Heart Disease	
ChestPainType	TA : Typical Angina	26	20	46
	ATA: Atypical Angina	149	24	173
	NAP: Non-Anginal Pain	131	72	203
	ASY: Asymptomatic	104	392	496
Total		410	508	918

Πίνακας 20: Συχνότητα εμφάνισης στηθάγχης και καρδιακής πάθησης.

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	268,067 <sup>a</sup>	3	,000
Likelihood Ratio	286,395	3	,000
Linear-by-Linear Association	203,735	1	,000
N of Valid Cases	918		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 20,54.

Πίνακας 21: Έλεγχος ανεξαρτησίας  $\chi^2$  μεταξύ στηθάγχης και καρδιακής πάθησης.



Βάσει των παραπάνω, από τα 46 άτομα που ένιωθαν τυπικό σθηθαγγικό πόνο, τα 26 δεν έπασχαν από κάποια καρδιακή πάθηση ενώ τα υπόλοιπα 20 έπασχαν. Συνεχίζοντας από τα 173 άτομα με συμπτώματα διαφορετικού τύπου πόνου, τα 149 δεν νοσούσαν από καρδιακή πάθηση ενώ τα 24 νοσούσαν. Ακόμα, από τα 203 άτομα που βίωναν πόνο διαφορετικής προέλευσης, τα 104 δεν έπασχαν ενώ τα υπόλοιποι 392 έπασχαν. Τέλος, από τα εναπομείναντα 496 άτομα που δεν ένιωθαν πόνο, οι 410 ήταν υγιείς και οι 508 είχαν κάποια καρδιακή ασθένεια.

Με τον  $X^2$  έλεγχο που εφαρμόστηκε, η pvalue τιμή είναι πολύ κοντά στο  $0 < \alpha = 0,05$  συνεπώς, σε επίπεδο σημαντικότητας 5%, απορρίπτεται η  $H_0$ . Δηλαδή, ο πόνος στο στήθος ενδέχεται να σχετίζεται με την ύπαρξη καρδιακής πάθησης.

### 3.4.3 Έλεγχος ανεξαρτησίας ζαχάρου νηστείας και καρδιακής πάθησης

Στον έλεγχο που θα πραγματοποιηθεί, εξετάζεται αν υπάρχει κάποια σχέση μεταξύ του ζαχαρου νηστείας και της εμφάνισης καρδιακής πάθησης.

FastingBS * HeartDisease Crosstabulation				
Count				
		HeartDisease		Total
		Normal	Heart Disease	
FastingBS	otherwise	366	338	704
	FastingBS > 120 mg/dl	44	170	214
Total		410	508	918

Πίνακας 22: Συχνότητα εμφάνισης ζαχάρου νηστείας και καρδιακής πάθησης.

Chi-Square Tests					
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	65,586 <sup>a</sup>	1	,000		
Continuity Correction <sup>b</sup>	64,321	1	,000		
Likelihood Ratio	69,842	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	65,515	1	,000		

<b>N of Valid Cases</b>	918			
a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 95,58.				
b. Computed only for a 2x2 table				

Πίνακας 23: Έλεγχος ανεξαρτησίας  $X^2$  μεταξύ ζαχάρου νηστείας και καρδιακής πάθησης.

Από τους πίνακες πάνω φαίνεται πως τα 214 άτομα που έχουν τιμή ζαχάρου νηστείας μεγαλύτερη των 120mg/dl διαχωρίζονται σε 44 υγιή και 170 με κάποια καρδιακή πάθηση. Τα υπόλοιπα 704 που έχουν τιμή μικρότερη των 120 mg/ dl κατανέμονται σε 410 χωρίς καρδιακή πάθηση και 508 με καρδιακή πάθηση.

Με βάση τον έλεγχο  $X^2$ , η τιμή pvalue είναι πολύ κοντά στο  $0 < \alpha = 0,05$  άρα, σε επίπεδο σημαντικότητας 5%, απορρίπτεται η  $H_0$ . Αυτό σημαίνει πως το ζάχαρο νηστείας, είναι πιθανό να σχετίζεται με την εμφάνιση καρδιακής πάθησης.

Υπολογίζοντας το odds ratio έχουμε τα εξής.

$$\text{Odds ratio} = (170 \cdot 366) / (44 \cdot 338) = 4,19$$

Η σχετική πιθανότητα ενός ατόμου με τιμές ζαχάρου μεγαλύτερες του 120 mg/dl να εμφανίσει καρδιακή πάθηση είναι σχεδόν τετραπλάσια από την αντίστοιχη σχετική πιθανότητα ενός ατόμου να εμφανίσει, έχοντας τιμές ζαχάρου μικρότερες ή ίσες του 120 mg/dl. Έτσι, τα άτομα με υψηλότερες από 120 md/dl τιμές ζαχάρου, έχουν αρκετά υψηλότερες πιθανότητες να πάσχουν από καρδιακή πάθηση σε σύγκριση με αυτά που έχουν χαμηλότερες τιμές.

### 3.4.4 Έλεγχος ανεξαρτησίας ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας και καρδιακής πάθησης

Στο παρόν, θα εξεταστεί αν υπάρχει σχέση μεταξύ του ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας και της καρδιακής πάθησης.

RestingECG * HeartDisease Crosstabulation			
Count	HeartDisease		Total
	Normal	Heart Disease	

<b>RestingECG</b>	<b>Normal</b>	267	285	552
	<b>ST</b>	61	117	178
	<b>LVH</b>	82	106	188
<b>Total</b>		410	508	918

Πίνακας 24: Συχνότητα εμφάνισης παλμών στο ηλεκτροκαρδιογράφημα σε κατάσταση ηρεμίας και καρδιακής πάθησης.

<b>Chi-Square Tests</b>			
	<b>Value</b>	<b>df</b>	<b>Asymptotic Significance (2-sided)</b>
<b>Pearson Chi-Square</b>	10,931 <sup>a</sup>	2	,004
<b>Likelihood Ratio</b>	11,098	2	,004
<b>Linear-by-Linear Association</b>	3,413	1	,065
<b>N of Valid Cases</b>	918		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 79,50.

Πίνακας 25: Έλεγχος ανεξαρτησίας  $X^2$  μεταξύ παλμών του ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας και καρδιακής πάθησης.

Παρατηρείται ότι από τους 552 συμμετέχοντες με φυσιολογική καρδιακή δραστηριότητα στο ηλεκτροκαρδιογράφημα, οι 267 δεν έπασχαν από κάποια καρδιακή πάθηση ενώ οι 285 έπασχαν. Ομοίως, από τους 178 που είχαν αλλαγές στο ST τμήμα του, οι 61 δεν έπασχαν από καρδιακή πάθηση ενώ οι υπόλοιποι 117 έπασχαν. Τέλος, από τους 188 με υπερτροφία της αριστερής κοιλίας της καρδιάς, οι 82 δεν είχαν καρδιακή πάθηση ενώ οι 106 είχαν.

Από τον  $X^2$  έλεγχο εντοπίζεται τιμή pvalue =0,004 <  $\alpha=0,05$  όπου, σε επίπεδο σημαντικότητας 5%, απορρίπτεται η  $H_0$ . Αυτό σημαίνει πως οι τιμές του ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας μπορεί να σχετίζονται με την πιθανότητα εμφάνισης καρδιακής πάθησης.

### 3.4.5 Έλεγχος ανεξαρτησίας στηθάγχης κατά τη διάρκεια άσκησης και καρδιακής πάθησης

Παρακάτω, εξετάζεται η ύπαρξη σχέσης του στηθαγχικού πόνου κατά τη διάρκεια της άσκησης και της καρδιακής πάθησης.

ExerciseAngina * HeartDisease Crosstabulation				
Count				
		HeartDisease		Total
		Normal	Heart Disease	
ExerciseAngina	No	355	192	547
	Yes	55	316	371
Total		410	508	918

Πίνακας 26: Συχνότητα εμφάνισης στηθαγχικού πόνου κατά τη διάρκεια της άσκησης και καρδιακής πάθησης.

Chi-Square Tests					
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	224,281 <sup>a</sup>	1	,000		
Continuity Correction <sup>b</sup>	222,259	1	,000		
Likelihood Ratio	241,765	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	224,037	1	,000		
N of Valid Cases	918				
a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 165,70.					
b. Computed only for a 2x2 table					

Πίνακας 27: Έλεγχος ανεξαρτησίας  $\chi^2$  μεταξύ στηθαγχικού πόνου κατά την άσκηση και καρδιακής πάθησης.

Από τους παραπάνω πίνακες γίνεται αντιληπτό πως οι 918 συμμετέχοντες κατανέμονται σε 547 που δεν ένιωθαν στηθαγχικό πόνο κατά την άσκηση εκ των οποίων, οι 355 δεν είχαν καρδιακή ασθένεια ενώ οι υπόλοιποι 192 είχαν και στους 371 που ένιωθαν πόνο στο στήθος κατά την άσκηση και από αυτούς, οι 55 δεν έπασχαν από καρδιακή πάθηση αντίθετα με τους υπόλοιπους 316 που έπασχαν.

Από τον  $\chi^2$  έλεγχο που πραγματοποιήθηκε, η τιμή pvalue βρίσκεται πολύ κοντά στο  $0 < \alpha = 0,05$  συνεπώς, σε επίπεδο σημαντικότητας 5% απορρίπτεται η  $H_0$ . Δηλαδή, η ύπαρξη στηθαγχικού πόνο κατά τη διάρκεια της σωματικής άσκησης μπορεί να σχετίζεται με την πιθανότητα ύπαρξης καρδιακής ασθένειας.

Υπολογίζοντας το odds ratio έχουμε τα εξής.

$$\text{Odds ratio} = (355 \times 316) / (192 \times 55) = 10,61$$

Η σχετική πιθανότητα εμφάνισης καρδιακής πάθησης των ατόμων που βιώνουν στηθαγχικό πόνο κατά την άσκηση είναι 10 φορές μεγαλύτερη από τη σχετική πιθανότητα των ατόμων που δεν βιώνουν στηθαγχικό πόνο κατά την άσκηση να πάσχουν. Έτσι, τα άτομα με αίσθημα στηθαγχικού πόνου κατά την άσκηση, έχουν σημαντικά υψηλότερες πιθανότητες να έχουν καρδιακή πάθηση σε σύγκριση με αυτά που δεν νιώθουν πόνο.

### 3.4.6 Έλεγχος ανεξαρτησίας στην αλλαγή του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης της άσκησης ή του στρες και της καρδιακής πάθησης

Γίνεται διερεύνηση για την ύπαρξη σχέσης της αλλαγής του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης της άσκησης ή του στρες και της καρδιακής πάθησης.

ST_Slope * HeartDisease Crosstabulation				
Count				
		HeartDisease		Total
		Normal	Heart Disease	
ST_Slope	Up	317	78	395
	Flat	79	381	460
	Down	14	49	63
Total		410	508	918

Πίνακας 28: Συχνότητα εμφάνισης αλλαγής του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης της άσκησης ή του στρες και της καρδιακής πάθησης.

#### Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
<b>Pearson Chi-Square</b>	355,918 <sup>a</sup>	2	,000
<b>Likelihood Ratio</b>	380,921	2	,000
<b>Linear-by-Linear Association</b>	286,310	1	,000
<b>N of Valid Cases</b>	918		
a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 28,14.			

Πίνακας 29: Έλεγχος ανεξαρτησίας  $X^2$  μεταξύ της αλλαγής του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης της άσκησης ή του στρες και της καρδιακής πάθησης

Παρατηρείται πως η αλλαγή στην κλίση του μοτίβου των 918 ατόμων είναι ανοδική για τα 395 άτομα εκ των οποίων, οι 317 είναι υγείας και οι 78 νοσούν από κάποια καρδιακή πάθηση. Στη συνέχεια, η κλίση είναι σταθερή για τα 460 άτομα όπου από αυτά τα 79 είναι υγεία και τα 381 νοσούν και τέλος, η κλίση είναι καθοδική για τα υπόλοιπα 63 άτομα εξ αυτών τα 14 είναι υγεία και τα 49 πάσχουν.

Βάσει του ελέγχου  $X^2$ , η τιμή pvalue βρίσκεται πολύ κοντά στο  $0 < \alpha = 0,05$  και συνεπώς, σε επίπεδο σημαντικότητας 5% απορρίπτεται η  $H_0$ . Έτσι, υπάρχουν ενδείξεις πως η αλλαγή του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης της άσκησης ή του στρες σχετίζονται με την πιθανότητα ύπαρξης καρδιακής πάθησης.

### 3.4.7 Έλεγχος ανεξαρτησίας μεταξύ των υπόλοιπων μεταβλητών

Ακολουθεί ένας συνοπτικός πίνακας με τους ελέγχους  $X^2$  των υπολοίπων μεταβλητών ανά δύο προκειμένου να αναλυθούν και οι υπόλοιποι συνδιασμοί.

Μεταβλητές ανά δύο	Pvalue
ChestPain*Sex	,000
FastingBS*Sex	,000
RestingECG*Sex	,089
ExerciseAngina * Sex	,000
ST_Slope * Sex	,000
FastingBS * ChestPainType	,000
RestingECG * ChestPainType	0,04
ExerciseAngina * ChestPainType	,000
ST_Slope * ChestPainType	,000

RestingECG * FastingBS	,000
ExerciseAngina * FastingBS	,067
ST_Slope * FastingBS	,000
ExerciseAngina * RestingECG	,005
ST_Slope * RestingECG	,103
ST_Slope * ExerciseAngina	,000

Πίνακας 30: Έλεγχος ανεξαρτησίας  $X^2$  ανά δύο των υπόλοιπων μεταβλητών.

Από τον πίνακα πάνω φαίνεται πως για τις περιπτώσεις μεταξύ ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας και φύλου ( $p$ value=0,089), στηθαγχικού πόνο κατά την άσκηση με ζάχαρο νηστείας ( $p$ value=0,067) και την αλλαγή του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης του τεστ κοπώσεως στο ηλεκτροκαρδιογράφημα με το ηλεκτροκαρδιογράφημα σε κατάσταση ηρεμίας ( $p$ value=0,103), οι τιμές των  $p$ value είναι μεγαλύτερες του επιπέδου σημαντικότητας 5%, συνεπώς δεν υπάρχουν αρκετές ενδείξεις για απόρριψη της  $H_0$  και δεχόμαστε ότι ο συνδυασμός αυτών των ζευγών ανά δύο δεν έχουν κάποια σχέση μεταξύ τους. Δηλαδή, το φύλο του ανθρώπου δεν σχετίζεται με τις τιμές του ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας ούτε οι τιμές ζαχάρου νηστείας έχουν κάποια σχέση με την ύπαρξη στηθαγχικού πόνου ομοίως και οι τιμές του ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας με την αλλαγή στην κλίση του κατά το τεστ κοπώσεως.

Για τις υπόλοιπες όμως μεταβλητές που ελέγχθηκαν, παρατηρούνται τιμές  $p$ value  $< \alpha=0,05$  άρα, σε επίπεδο σημαντικότητας 5% απορρίπτεται η  $H_0$  και δεχόμαστε πως υπάρχει κάποια σχέση μεταξύ αυτών των ανά δύο συνδυασμών.

### 3.5 Συγκρίσεις με χρήση του ελέγχου Wilcoxon| Mann – Whitney

Θα πραγματοποιηθεί ο μη παραμετρικός έλεγχος Wilcoxon, Mann – Whitney μεταξύ των συνεχών και των κατηγορικών μεταβλητών δύο επιπέδων ορίζοντας τις εξής υποθέσεις:

**$H_0$ :** Δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των δύο ομάδων που συγκρίνονται.

**$H_a$ :** Όχι  $H_0$ .

### 3.5.1 Σύγκριση των τιμών των συνεχών μεταβλητών μεταξύ ανδρών και γυναικών

Ranks				
	Sex	N	Mean Rank	Sum of Ranks
Age	Male	725	467,69	339076,00
	Female	193	428,73	82745,00
	Total	918		
MaxHR	Male	725	432,73	313727,50
	Female	193	560,07	108093,50
	Total	918		
Oldpeak	Male	725	475,05	344413,00
	Female	193	401,08	77408,00
	Total	918		
Cholesterol	Male	725	449,47	325866,00
	Female	193	497,18	95955,00
	Total	918		
RestingBP	Male	725	463,31	335901,50
	Female	193	445,18	85919,50
	Total	918		

Πίνακας 31: Συνολικές αναθέσεις συνεχών μεταβλητών ανά φύλο.

Test Statistics <sup>a</sup>					
	Age	MaxHR	Oldpeak	Cholesterol	RestingBP
Mann-Whitney U	64024,000	50552,500	58687,000	62691,000	67198,500
Wilcoxon W	82745,000	313727,500	77408,000	325866,000	85919,500
Z	-1,815	-5,931	-3,564	-2,229	-,848
Asymp. Sig. (2-tailed)	,070	,000	,000	,026	,397

a. Grouping Variable: Sex

Πίνακας 32: Έλεγχος Mann – Whitney U, Wilcoxon συνεχών μεταβλητών με το φύλο.

Βάσει του παραπάνω πίνακα και του ελέγχου Wilcoxon, παρατηρείται πως για τις περιπτώσεις της ηλικίας ( $p$ value = 0,70) και του ζαχάρου νηστείας ( $p$ value=0,397) ανά το φύλο, οι τιμές των  $p$ value τους είναι μεγαλύτερες του επιπέδου σημαντικότητας 5% συνεπώς δεν υπάρχουν αρκετές ενδείξεις για την απόρριψη της  $H_0$ . Δηλαδή, η μέση ηλικία αντρών και



γυναικών και η μέση αρτηριακή πίεση σε κατάσταση ηρεμίας αυτών, δεν έχουν σημαντικές διαφορές μεταξύ τους.

Αντίθετα, ο μέγιστος καρδιακός ρυθμός, η μέγιστη μεταβολή του ηλεκτροκαρδιογραφήματος κατά το τεστ κοπώσεως και η χοληστερίνη έχουν τιμές  $p$ value <  $\alpha=0,05$  άρα, σε επίπεδο σημαντικότητας 5%, απορρίπτεται η  $H_0$ . Έτσι, υπάρχει σημαντική διαφορά ανάμεσα στον μέγιστο καρδιακό ρυθμό, τη μέγιστη μεταβολή του ηλεκτροκαρδιογραφήματος κατά το τεστ κοπώσεως και τη χοληστερίνη μεταξύ των δύο φύλων. Συγκεκριμένα, οι γυναίκες έχουν σημαντικά υψηλότερες τιμές στο μέγιστο καρδιακό ρυθμό και τη χοληστερίνη τους σε σύγκριση με τους άνδρες, οι οποίοι παρουσιάζουν σημαντικά υψηλότερες τιμές στη μέγιστη μεταβολή του ηλεκτροκαρδιογραφήματος κατά το τεστ κοπώσεως.

### 3.5.2 Σύγκριση των τιμών των συνεχών μεταβλητών σε πληθυσμούς που ορίζονται βάσει του επιπέδου καρδιακής πάθησης

Ranks				
	HeartDisease	N	Mean Rank	Sum of Ranks
Age	Normal	410	374,13	153392,50
	Heart Disease	508	528,40	268428,50
	Total	918		
MaxHR	Normal	410	578,89	237344,50
	Heart Disease	508	363,14	184476,50
	Total	918		
Oldpeak	Normal	410	340,05	139419,00
	Heart Disease	508	555,91	282402,00
	Total	918		
Cholesterol	Normal	410	421,94	172995,50
	Heart Disease	508	489,81	248825,50
	Total	918		
RestingBP	Normal	410	425,40	174415,50
	Heart Disease	508	487,02	247405,50
	Total	918		

Πίνακας 33: Συνολικές αναθέσεις συνεχών μεταβλητών με ύπαρξη ή όχι καρδιακής πάθησης.

Test Statistics <sup>a</sup>					
	Age	MaxHR	Oldpeak	Cholesterol	RestingBP
Mann-Whitney U	69137,500	55190,500	55164,000	88740,500	90160,500
Wilcoxon W	153392,500	184476,500	139419,000	172995,500	174415,500
Z	-8,769	-12,259	-12,690	-3,869	-3,513
Asymp. Sig. (2-tailed)	,000	,000	,000	,000	,000

a. Grouping Variable: HeartDisease

Πίνακας 34: Έλεγχος Mann – Whitney U, Wilcoxon συνεχών μεταβλητών με την ύπαρξη ή όχι καρδιακής πάθησης

Από τον πάνω πίνακα παρατηρείται πως όλες οι τιμές pvalue βρίσκονται πολύ κοντά στο 0 συνεπώς, σε επίπεδο σημαντικότητας 5% απορρίπτεται η  $H_0$ . Δηλαδή, υπάρχει σημαντική διαφορά ανάμεσα στα υγιή και στα άτομα με καρδιακή πάθηση, ανάλογα την ηλικία, το μέγιστο καρδιακός ρυθμό, τη μέγιστη μεταβολή του ηλεκτροκαρδιογραφήματος κατά το τεστ κοπώσεως, τις τιμές της χοληστερίνης και της μέσης αρτηριακής τους πίεσης σε κατάσταση ηρεμίας. Αναλυτικά, τα άτομα που πάσχουν από καρδιακές ασθένειες τείνουν να είναι μεγαλύτερης ηλικίας, με χαμηλότερες τιμές του μέγιστου καρδιακού ρυθμού, μεγαλύτερη μεταβολή του ηλεκτροκαρδιογραφήματος κατά το τεστ κοπώσεως, υψηλότερες τιμές χοληστερίνης και μεγαλύτερες τιμές αρτηριακής πίεσης από τα υγιή άτομα.

### 3.5.3 Σύγκριση των τιμών των συνεχών μεταβλητών σε πληθυσμούς που ορίζονται βάσει του επιπέδου ζαχάρου νηστείας

Ranks				
	FastingBS	N	Mean Rank	Sum of Ranks
Age	otherwise	704	430,15	302823,00
	FastingBS > 120 mg/dl	214	556,07	118998,00
	Total	918		
MaxHR	otherwise	704	477,69	336294,00
	FastingBS > 120 mg/dl	214	399,66	85527,00
	Total	918		
Oldpeak	otherwise	704	445,53	313650,50
	FastingBS > 120 mg/dl	214	505,47	108170,50

	<b>Total</b>	918		
<b>Cholesterol</b>	<b>otherwise</b>	704	450,91	317441,00
	<b>FastingBS &gt; 120 mg/dl</b>	214	487,76	104380,00
	<b>Total</b>	918		
<b>RestingBP</b>	<b>otherwise</b>	704	450,05	316833,00
	<b>FastingBS &gt; 120 mg/dl</b>	214	490,60	104988,00
	<b>Total</b>	918		

Πίνακας 35: Συνολικές αναθέσεις συνεχών μεταβλητών με τα επίπεδα ζαχάρου νηστείας.

<b>Test Statistics<sup>a</sup></b>					
	<b>Age</b>	<b>MaxHR</b>	<b>Oldpeak</b>	<b>Cholesterol</b>	<b>RestingBP</b>
<b>Mann-Whitney U</b>	54663,000	62522,000	65490,500	69281,000	68673,000
<b>Wilcoxon W</b>	302823,000	85527,000	313650,500	317441,000	316833,000
<b>Z</b>	-6,087	-3,771	-2,997	-1,786	-1,967
<b>Asymp. Sig. (2-tailed)</b>	,000	,000	,003	,074	,049

a. Grouping Variable: FastingBS

Πίνακας 36: Έλεγχος Mann – Whitney U, Wilcoxon συνεχών μεταβλητών με τα επίπεδα ζαχάρου νηστείας.

Με βάση τον έλεγχο Wilcoxon που πραγματοποιήθηκε, η τιμή pvalue της χοληστερίνης ισούται με  $0,074 > \alpha=0,05$ . Άρα σε επίπεδο σημαντικότητας 5% δεν υπάρχουν αρκετές ενδείξεις για την απόρριψη της  $H_0$ . Δηλαδή, δεν υπάρχει σημαντική διαφορά στις τιμές της χοληστερίνης των ατόμων που έχουν τιμές ζαχάρου νηστείας μεγαλύτερες από 120 mg/dl και αυτών με τιμές μικρότερες ή ίσες του 120 mg/dl του ζαχάρου νηστείας.

Στις υπόλοιπες περιπτώσεις, οι τιμές pvalue είναι μικρότερες ή οριακά μικρότερες (pvalueRestigBP=0,049) του επιπέδου σημαντικότητας 5% και έτσι, σε επίπεδο σημαντικότητας 5% απορρίπτεται η  $H_0$ . Δηλαδή, υπάρχουν στατιστικά σημαντικές διαφορές ανάμεσα στην ηλικία, το μέγιστο καρδιακό ρυθμό, τη μέγιστη μεταβολή του ηλεκτροκαρδιογραφήματος κατά το τεστ κοπώσεως, και τη μέση αρτηριακή πίεση σε κατάσταση ηρεμίας μεταξύ των ατόμων που έχουν τιμές ζαχάρου νηστείας  $> 120\text{mg/dl}$  και  $\leq 120\text{ mg/dl}$ . Συγκεκριμένα, τα άτομα με τιμές ζαχάρου μεγαλύτερες του 120 mg/dl είναι μεγαλύτερης ηλικίας, έχουν πιο χαμηλές τιμές μέγιστου καρδιακού ρυθμού, μεγαλύτερη μεταβολή στο ηλεκτροκαρδιογράφημα κατά το τεστ κοπώσεως, υψηλότερες τιμές χοληστερίνης και μεγαλύτερες τιμές αρτηριακής πίεσης από τα άτομα με μικρότερες ή ίσες τιμές του 120 mg/dl.

### 3.5.4 Σύγκριση των τιμών των συνεχών μεταβλητών σε πληθυσμούς που ορίζονται βάσει του επιπέδου της στηθάγχης από άσκηση

Ranks				
	ExerciseAngina	N	Mean Rank	Sum of Ranks
Age	No	547	412,50	225637,50
	Yes	371	528,80	196183,50
	Total	918		
MaxHR	No	547	544,99	298111,50
	Yes	371	333,45	123709,50
	Total	918		
Oldpeak	No	547	365,69	200031,00
	Yes	371	597,82	221790,00
	Total	918		
Cholesterol	No	547	437,64	239388,00
	Yes	371	491,73	182433,00
	Total	918		
RestingBP	No	547	424,15	232010,00
	Yes	371	511,62	189811,00
	Total	918		

Πίνακας 37: Συνολικές αναθέσεις συνεχών μεταβλητών με στηθάγχη από άσκηση.

Test Statistics <sup>a</sup>					
	Age	MaxHR	Oldpeak	Cholesterol	RestingBP
Mann-Whitney U	75759,500	54703,500	50153,000	89510,000	82132,000
Wilcoxon W	225637,500	123709,500	200031,000	239388,000	232010,000
Z	-6,525	-11,865	-13,470	-3,043	-4,923
Asymp. Sig. (2-tailed)	,000	,000	,000	,002	,000

a. Grouping Variable: ExerciseAngina

Πίνακας 38: Έλεγχος Mann – Whitney U, Wilcoxon συνεχών μεταβλητών με τα επίπεδα στηθάγχης από άσκηση.

Από τον έλεγχο Wilcoxon, όλες οι τιμές pvalue < α=0,05 άρα, σε επίπεδο σημαντικότητας 5% απορρίπτεται η H<sub>0</sub>. Έτσι, εντοπίζονται στατιστικά σημαντικές διαφορές στην ηλικία, στο μέγιστο καρδιακό ρυθμό, στη μέγιστη μεταβολή του ηλεκτροκαρδιογραφήματος κατά το τεστ

κοπώσεως, στη χοληστερίνη και στη μέση αρτηριακή πίεση σε κατάσταση ηρεμίας μεταξύ των ατόμων που νιώθουν πόνο στο στήθος κατά τη διάρκεια της άσκησης και αυτών που δεν αισθάνονται πόνο. Αναλυτικότερα, τα άτομα που νιώθουν πόνο στο στήθος κατά τη διάρκεια της άσκησης είναι κατά μέσο όρο μεγαλύτερης ηλικίας, με υψηλότερες τιμές στη μέγιστη μεταβολή του ηλεκτροκαρδιογραφήματός τους, στις τιμές της χοληστερίνης, και στη μέση αρτηριακή πίεση σε κατάσταση ηρεμίας. Αντίθετα, τα άτομα που δεν αισθάνονται πόνο στο στήθος έχουν υψηλότερες τιμές στον μέγιστο καρδιακό ρυθμό τους.

### 3.6 Συγκρίσεις με χρήση του ελέγχου Kruskal-Wallis

Προχωρώντας στην ανάλυση των μεταβλητών με τρία και τέσσερα επίπεδα, θα πραγματοποιηθεί ο μη παραμετρικός έλεγχος Kruskal – Wallis και ορίζονται οι εξής υποθέσεις:

**H<sub>0</sub>:** Η κατανομή των συνεχών μεταβλητών είναι ίδια σε όλα τα επίπεδα της κατηγορικής μεταβλητής.

**H<sub>a</sub>:** Η κατανομή των συνεχών μεταβλητών, διαφέρει σε τουλάχιστον ένα από τα επίπεδα της κατηγορικής.

Στις περιπτώσεις απόρριψης της H<sub>0</sub>, δημιουργούνται μη παραμετρικά διαστήματα εμπιστοσύνης με τη μέθοδο bootstrap σε 1.000 επαναλήψεις για τον εντοπισμό των ζευγών με στατιστικά σημαντικές διαφορές.

#### 3.6.1 Σύγκριση των τιμών των συνεχών μεταβλητών σε πληθυσμούς που ορίζονται βάσει του επιπέδου του ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας

Hypothesis Test Summary			
Null Hypothesis	Test	Sig.	Decision

1	The distribution of Age is the same across categories of RestingECG.	Independent-Samples Kruskal-Wallis Test	,000	Reject the null hypothesis.
2	The distribution of MaxHR is the same across categories of RestingECG.	Independent-Samples Kruskal-Wallis Test	,000	Reject the null hypothesis.
3	The distribution of Oldpeak is the same across categories of RestingECG.	Independent-Samples Kruskal-Wallis Test	,002	Reject the null hypothesis.
4	The distribution of Cholesterol is the same across categories of RestingECG.	Independent-Samples Kruskal-Wallis Test	,013	Reject the null hypothesis.
5	The distribution of RestingBP is the same across categories of RestingECG.	Independent-Samples Kruskal-Wallis Test	,004	Reject the null hypothesis.
Asymptotic significances are displayed. The significance level is ,050.				

Πίνακας 39: Έλεγχος Kruskal-Wallis συνεχών μεταβλητών με τα επίπεδα του ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας.

Bootstrap				
			BCa 95% Confidence Interval	
	RestingECG	Mean	Lower	Upper
Age	Normal	51,74	50,97	52,55
	ST	56,14	54,88	57,45
	L VH	56,22	54,99	57,50
MaxHR	Normal	137,30	135,09	139,57
	ST	128,62	125,08	131,80
	L VH	143,12	139,46	146,62
Oldpeak	Normal	,77	,71	,87
	ST	1,01	,84	1,17
	L VH	1,07	,92	1,25
Cholesterol	Normal	242,80	238,59	247,49
	ST	241,97	235,96	248,51
	L VH	252,56	245,40	259,67
RestingBP	Normal	130,88	129,43	132,43

	ST	135,81	133,13	138,83
	LVH	134,34	131,85	136,89

Πίνακας 40: Μέθοδος bootstrap για προσδιορισμό διαστημάτων εμπιστοσύνης.

Με βάση τον παραπάνω πίνακα, όλες οι τιμές pvalue του ελέγχου Kruskal – Wallis είναι μικρότερες του  $\alpha=0,05$  άρα, σε επίπεδο σημαντικότητας 5% απορρίπτεται η  $H_0$ . Δηλαδή η κατανομή των μεταβλητών της ηλικίας, του μέγιστου καρδιακού ρυθμού, της μέγιστης μεταβολής του ηλεκτροκαρδιογραφήματος κατά τη διαδικασία του τεστ κοπώσεως, της χοληστερίνης και της αρτηριακής πίεσης σε κατάσταση ηρεμίας διαφέρει σε τουλάχιστον ένα από τα τρία επίπεδα του ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας. Πιο συγκεκριμένα, εξετάζοντας ανά δύο το κάθε ζεύγος για όλες τις περιπτώσεις έχουμε:

Ηλικία: Η μέση ηλικία στο επίπεδο Normal είναι 51,74 έτη με  $\Delta E_{Normal}$ : [50,97, 52,55], στο επίπεδο ST είναι 56,14 έτη με  $\Delta E_{ST}$ : [54,88, 57,45] και στο επίπεδο LVH 56,22 έτη με  $\Delta E_{LVH}$ : [54,99, 57,50]. Το ζεύγος ST – LVH παρουσιάζει σχεδόν πλήρη επικάλυψη σε αντίθεση με τα ζεύγη Normal – ST, Normal – LVH, που έχουν στατιστικά σημαντικές διαφορές. Συγκεκριμένα, τα άτομα με φυσιολογικές ενδείξεις στο ηλεκτροκαρδιογράφημα είναι κατά μέσο όρο 4,4 χρόνια μικρότερα από αυτά με ένδειξη κάποιας ανωμαλίας και 4,48 χρόνια μικρότερα από αυτά που υποδεικνύουν παρουσία πιθανής ή μόνιμης υπερτροφίας της αριστερής κοιλίας.

Μέγιστος καρδιακός ρυθμός (MaxHR): Η μέση τιμή του MaxHR για το επίπεδο Normal είναι 137,30 παλμοί με  $\Delta E_{Normal}$ : [135,09, 139,57], για το επίπεδο ST 128,62 παλμοί με  $\Delta E_{ST}$ : [125,08, 131,80] και για το επίπεδο LVH 143,12 παλμοί με  $\Delta E_{LVH}$ : [139,46, 146,62]. Το ζεύγος Normal – LVH παρουσιάζει οριακά επικαλυπτόμενα διαστήματα εμπιστοσύνης ενώ τα ζεύγη Normal – ST, ST – LVH μη επικαλυπτόμενα στα οποία εντοπίζονται οι στατιστικά σημαντικές διαφορές. Συγκεκριμένα, τα άτομα με φυσιολογικές ενδείξεις στο ηλεκτροκαρδιογράφημα έχουν κατά μέσο όρο 8,68 παλμούς περισσότερο από τα αντίστοιχα με ένδειξη κάποιας ανωμαλίας τα οποία με τη σειρά τους έχουν 14,5 παλμούς λιγότερο από αυτά με παρουσία πιθανής ή μόνιμης υπερτροφίας της αριστερής κοιλίας.

Μέγιστη Μεταβολή του ηλεκτροκαρδιογραφήματος κατά τη διαδικασία του τεστ κοπώσεως (Oldpeak): Η μέση τιμή του Oldpeak είναι 0,77 mm για το επίπεδο Normal με  $\Delta E_{Normal}$ : [0,71, 0,87], 1,01 mm για το επίπεδο ST με  $\Delta E_{ST}$ : [0,84, 1,17] και 1,07 mm για το επίπεδο LVH με  $\Delta E_{LVH}$ : [0,92, 1,25]. Τα ζεύγη Normal – ST, ST – LVH έχουν επικαλυπτόμενα διαστήματα

εμπιστοσύνης, αντίθετα με το ζεύγος Normal – LVH στο οποίο εντοπίζεται στατιστική διαφορά. Ειδικά, τα άτομα με φυσιολογικές ενδείξεις στο ηλεκτροκαρδιογράφημα έχουν κατά μέσο όρο 0,3 mm μικρότερη από μεταβολή από αυτά με παρουσία πιθανής ή μόνιμης υπερτροφίας της αριστερής κοιλίας.

Χοληστερίνη: Η μέση τιμή της χοληστερίνης είναι 242,80 mg/dL στο επίπεδο Normal με  $\Delta E_{Normal}$ : [238,59, 247,49], 241,97 mg/dL στο επίπεδο ST με  $\Delta E_{ST}$ : [235,96, 248,51] και 252,56 mg/dL στο επίπεδο LVH με  $\Delta E_{LVH}$ : [245,40, 259,67]. Από αυτά, όλα τα ζεύγη παρουσιάζουν επικαλυπτόμενα διαστήματα εμπιστοσύνης συνεπώς, δεν υπάρχει στατιστική διαφορά σε κανένα από αυτά τα επίπεδα. Δηλαδή, οι τιμές της χοληστερίνης δεν διαφέρουν σημαντικά σε κανένα από τα επίπεδα του ηλεκτροκαρδιογραφήματος σε κατάσταση ηρεμίας.

Αρτηριακή Πίεση σε Κατάσταση Ηρεμίας (RestingBP): Η μέση τιμή της αρτηριακής πίεσης είναι 130,88 mm Hg στο επίπεδο Normal με  $\Delta E_{Normal}$ : [129,43, 132,43], 135,81 mm Hg στο επίπεδο ST με  $\Delta E_{ST}$ : [133,13, 138,83] και 134,34 mm Hg στο επίπεδο LVH με  $\Delta E_{LVH}$ : [131,85, 136,89]. Τα ζεύγη Normal – LVH, ST – LVH παρουσιάζουν επικαλυπτόμενα διαστήματα εμπιστοσύνης σε αντίθεση με το Normal - ST που έχει οριακά μη επικαλυπτόμενα. Έτσι, στατιστικά σημαντική διαφορά εντοπίζεται μόνο στο τελευταίο ζεύγος. Ειδικά, τα άτομα με φυσιολογικές ενδείξεις στο ηλεκτροκαρδιογράφημα έχουν κατά μέσο όρο 3,4 mm Hg χαμηλότερη αρτηριακή πίεση από τα αντίστοιχα με ένδειξη κάποιας ανωμαλίας.

### 3.6.2 Σύγκριση των τιμών των συνεχών μεταβλητών σε πληθυσμούς που ορίζονται βάσει του επιπέδου της αλλαγής του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης του τεστ κοπώσεως στο ηλεκτροκαρδιογράφημα

Hypothesis Test Summary				
	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Age is the same across categories of ST_Slope.	Independent-Samples Kruskal-Wallis Test	,000	Reject the null hypothesis.
2	The distribution of MaxHR is the same across categories of ST_Slope.	Independent-Samples Kruskal-Wallis Test	,000	Reject the null hypothesis.



<b>3</b>	The distribution of Oldpeak is the same across categories of ST_Slope.	Independent-Samples Kruskal-Wallis Test	,000	Reject the null hypothesis.
<b>4</b>	The distribution of Cholesterol is the same across categories of ST_Slope.	Independent-Samples Kruskal-Wallis Test	,027	Reject the null hypothesis.
<b>5</b>	The distribution of RestingBP is the same across categories of ST_Slope.	Independent-Samples Kruskal-Wallis Test	,003	Reject the null hypothesis.
Asymptotic significances are displayed. The significance level is ,050.				

Πίνακας 41: Έλεγχος Kruskal-Wallis συνεχών μεταβλητών με τα επίπεδα της αλλαγής του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης του τεστ κοπώσεως του ηλεκτροκαρδιογραφήματος.

<b>Bootstrap</b>				
			<b>BCa 95% Confidence Interval</b>	
	<b>ST_Slope</b>	<b>Mean</b>	<b>Lower</b>	<b>Upper</b>
<b>Age</b>	Up	50,71	49,74	51,72
	Flat	55,26	54,51	56,02
	Down	58,32	56,08	60,57
<b>MaxHR</b>	Up	148,04	145,69	150,32
	Flat	128,11	125,94	130,49
	Down	129,94	124,14	135,99
<b>Oldpeak</b>	Up	,34	,27	,41
	Flat	1,19	1,09	1,29
	Down	2,15	1,78	2,51
<b>Cholesterol</b>	Up	239,12	234,12	244,56
	Flat	249,61	244,58	254,80
	Down	242,88	232,34	254,89
<b>RestingBP</b>	Up	130,35	128,85	132,08
	Flat	134,52	132,93	136,05
	Down	131,86	126,54	137,43

Πίνακας 42: Μέθοδος bootstrap για προσδιορισμό διαστημάτων εμπιστοσύνης.

Από τον έλεγχο Kruskal – Wallis που πραγματοποιήθηκε, όλες οι περιπτώσεις των συνεχών μεταβλητών που εξετάστηκαν, έχουν τιμές pvalue  $\leq 0,05$  και έτσι, σε επίπεδο σημαντικότητας 5% απορρίπεται η  $H_0$ . Δηλαδή η κατανομή των μεταβλητών της ηλικίας, του μέγιστου καρδιακού ρυθμού, της μέγιστης μεταβολής του ηλεκτροκαρδιογραφήματος κατά το τεστ κοπώσεως, της χοληστερίνης και της αρτηριακής πίεσης σε κατάσταση ηρεμίας διαφέρουν σε τουλάχιστον ένα από τα τρία επίπεδα της αλλαγής του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης του τεστ κοπώσεως στο ηλεκτροκαρδιογράφημα. Πιο συγκεκριμένα:

Ηλικία: Η μέση ηλικία στο επίπεδο Up είναι τα 50,71 έτη με  $\Delta E_{Up}$ : [49,74, 51,72], στο επίπεδο Flat είναι τα 55,26 έτη με  $\Delta E_{Flat}$ : [54,51, 56,02] και στο επίπεδο Down τα 58,32 έτη με  $\Delta E_{Down}$ : [56,08, 60,57]. Σε όλες τις περιπτώσεις, τα διαστήματα εμπιστοσύνης δεν επικαλύπτονται και υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ όλων των ζευγών (Up – Flat, Flat – Down, Up – Down). Συγκεκριμένα, τα άτομα με ανοδική κλίση (Up) είναι κατά μέσο όρο 4,55 χρόνια μικρότερα από αυτά με σταθερή κλίση (Flat) και 7,61 χρόνια μικρότερα από τα αντίστοιχα με καθοδική κλίση (Down). Παρόμοια, τα άτομα με σταθερή κλίση είναι κατά μέσο όρο 3,06 χρόνια μικρότερα από αυτά με καθοδική κλίση.

Μέγιστος καρδιακός ρυθμός (MaxHR): Η μέση τιμή του MaxHR για το επίπεδο Up είναι 148,04 παλμοί με  $\Delta E_{Up}$ : [145,69, 150,32], για το επίπεδο Flat 128,11 παλμοί με  $\Delta E_{Flat}$ : [125,94, 130,49] και για το επίπεδο Down 129,94 παλμοί με  $\Delta E_{Down}$ : [124,14, 135,99]. Τα ζεύγη Up – Flat και Flat – Down παρουσιάζουν επικαλυπτόμενα διαστήματα εμπιστοσύνης σε αντίθεση με το ζεύγος Up – Down. Έτσι, στατιστικά σημαντική διαφορά εντοπίζεται μόνο στα επίπεδα Up και Down. Πιο συγκεκριμένα, τα άτομα με ανοδική κλίση έχουν κατά μέσο όρο 18,1 παλμούς περισσότερο από τα άτομα με καθοδική κλίση.

Μέγιστη Μεταβολή του ηλεκτροκαρδιογραφήματος κατά τη διαδικασία του τεστ κοπώσεως (Oldpeak): Η μέση τιμή του Oldpeak είναι 0,34 mm για το επίπεδο Up με  $\Delta E_{Up}$ : [0,27, 0,41], 1,19 mm για το επίπεδο Flat με  $\Delta E_{Flat}$ : [1,09, 1,29] και 2,15 mm για το επίπεδο Down με  $\Delta E_{Down}$ : [1,78, 2,51]. Κανένα από τα διαστήματα εμπιστοσύνης δεν επικαλύπτονται συνεπώς, υπάρχουν στατιστικά σημαντικές διαφορές σε όλα τα ζεύγη. Ειδικά, τα άτομα με ανοδική κλίση έχουν κατά μέσο όρο 0,85 mm μικρότερη μεταβολή από αυτά με σταθερή κλίση και 1,81 mm

μικρότερη από αυτά με καθοδική κλίση. Ομοίως, τα άτομα με σταθερή κλίση έχουν 0,96 mm μικρότερη μεταβολή από αυτά με καθοδική κλίση.

Χοληστερίνη: Η μέση τιμή της χοληστερίνης είναι 239,12 mg/dL στο επίπεδο Up με  $\Delta E_{Up}$ : [234,12, 244,56], 249,61 mg/dL στο επίπεδο Flat με  $\Delta E_{Flat}$ : [244,58, 254,80] και 242,88 mg/dL στο επίπεδο Down με  $\Delta E_{Down}$ : [232,34, 254,89]. Από αυτά, τα ζεύγη Up – Down και Flat – Down παρουσιάζουν επικαλυπτόμενα διαστήματα εμπιστοσύνης σε αντίθεση με το Up – Flat που έχει οριακά μη επικαλυπτόμενα διαστήματα δείχνοντας έτσι οριακά στατιστική διαφορά. Συγκεκριμένα, τα άτομα με ανοδική κλίση έχουν κατά μέσο όρο 10,49 mg/dL λιγότερη χοληστερίνη από αυτά με σταθερή κλίση.

Αρτηριακή Πίεση σε Κατάσταση Ηρεμίας (RestingBP): Η μέση τιμή της αρτηριακής πίεσης είναι 130,35 mm Hg στο επίπεδο Up με  $\Delta E_{Up}$ : [128,85, 132,08], 134,52 mm Hg στο επίπεδο Flat με  $\Delta E_{Flat}$ : [132,94, 136,05] και 131,86 mm Hg στο επίπεδο Down με  $\Delta E_{Down}$ : [126,54, 137,43]. Τα ζεύγη Up – Down και Flat – Down παρουσιάζουν επικαλυπτόμενα διαστήματα εμπιστοσύνης αντίθετα με το Up – Flat που έχει οριακά μη επικαλυπτόμενα, έχοντας έτσι μόνο στο τελευταίο ζεύγος στατιστικά σημαντική διαφορά. Ειδικά, τα άτομα με ανοδική κλίση έχουν κατά μέσο όρο 4,14 mm Hg χαμηλότερη αρτηριακή πίεση από αυτά με σταθερή κλίση.

### 3.6.3 Σύγκριση των τιμών των συνεχών μεταβλητών σε πληθυσμούς που ορίζονται βάσει του επιπέδου του στηθαγχικού πόνου

Hypothesis Test Summary				
	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Age is the same across categories of ChestPainType.	Independent-Samples Kruskal-Wallis Test	,000	Reject the null hypothesis.
2	The distribution of MaxHR is the same across categories of ChestPainType.	Independent-Samples Kruskal-Wallis Test	,000	Reject the null hypothesis.
3	The distribution of Oldpeak is the same across categories of ChestPainType.	Independent-Samples Kruskal-Wallis Test	,000	Reject the null hypothesis.
4	The distribution of Cholesterol is the same across categories of ChestPainType.	Independent-Samples Kruskal-Wallis Test	,017	Reject the null hypothesis.

<b>5</b>	The distribution of RestingBP is the same across categories of ChestPainType.	Independent-Samples Kruskal-Wallis Test	<b>,061</b>	Retain the null hypothesis.
Asymptotic significances are displayed. The significance level is ,050.				

Πίνακας 43: Έλεγχος Kruskal-Wallis συνεχών μεταβλητών με τα επίπεδα του σθηθαγγικού πόνου.

<b>Bootstrap</b>				
			<b>BCa 95% Confidence Interval</b>	
	<b>ChestPainType</b>	<b>Mean</b>	<b>Lower</b>	<b>Upper</b>
<b>Age</b>	TA : Typical Angina	54,83	51,47	57,93
	ATA: Atypical Angina	49,24	47,94	50,54
	NAP: Non-Anginal Pain	53,31	51,99	54,66
	ASY: Asymptomatic	54,96	54,09	55,84
<b>MaxHR</b>	TA : Typical Angina	147,89	141,42	154,46
	ATA: Atypical Angina	150,21	146,34	154,00
	NAP: Non-Anginal Pain	143,24	139,62	146,65
	ASY: Asymptomatic	128,48	126,44	130,56
<b>Oldpeak</b>	TA : Typical Angina	1,04	,77	1,31
	ATA: Atypical Angina	,31	,22	,40
	NAP: Non-Anginal Pain	,68	,55	,79
	ASY: Asymptomatic	1,16	1,06	1,27
<b>Cholesterol</b>	TA : Typical Angina	233,66	221,70	245,60
	ATA: Atypical Angina	242,95	235,63	250,62
	NAP: Non-Anginal Pain	238,41	231,23	245,72
	ASY: Asymptomatic	248,79	243,97	253,44

Πίνακας 44: Μέθοδος bootstrap για προσδιορισμό διαστημάτων εμπιστοσύνης.

Βάσει του ελέγχου Kruskal – Wallis που εφαρμόστηκε παραπάνω, η μεταβλητή της αρτηριακής πίεσης σε κατάσταση ηρεμίας λαμβάνει τιμή  $p_{value} = 0,061 > \alpha = 0,05$  συνεπώς, σε επίπεδο σημαντικότητας 5% δεν έχουμε αρκετές ενδείξεις να απορρίψουμε τη μηδενική υπόθεση και έτσι, η κατανομή της μεταβλητής αυτής είναι ίδια και στα

τέσσερα επίπεδα, τυπικός πόνος, ασυνήθιστος, πόνος διαφορετικής προέλευσης και καθόλου πόνος, του στηθαγχικού πόνου.

Για τις υπόλοιπες υπό εξέταση μεταβλητές, οι τιμές  $p\text{value} < \alpha=0,05$  και σε επίπεδο σημαντικότητας 5% απορρίπτεται η  $H_0$ . Δηλαδή, η κατανομή της ηλικίας, του μέγιστου καρδιακού ρυθμού, της μέγιστης μεταβολής του ηλεκτροκαρδιογραφήματος κατά το τεστ κοπώσεως, και της χοληστερίνης διαφέρουν σε τουλάχιστον ένα από τα τέσσερα επίπεδα του στηθαγχικού πόνου.

Ηλικία: Η μέση ηλικία στο επίπεδο TA είναι τα 54,83 έτη με  $\Delta E_{TA} = [51,47, 57,93]$ , στο επίπεδο ATA, είναι τα 49,24 έτη με  $\Delta E_{ATA} = [47,94, 50,54]$ , στο NAP τα 53,31 έτη με  $\Delta E_{NAP} = [51,99, 54,66]$  και στο ASY, τα 54,96 έτη με  $\Delta E_{ASY} = [54,09, 55,84]$ . Τα μόνα ζεύγη με μη επικαλυπτόμενα διαστήματα είναι τα ATA – NAP και ATA – ASY στα οποία μόνο αυτά υπάρχουν στατιστικά σημαντικές διαφορές. Δηλαδή, τα άτομα που βιώνουν στηθαγχικό πόνο διαφορετικό από τον συνηθισμένο είναι κατά μέσο όρο 4,07 χρόνια μικρότερα σε ηλικία από τα αντίστοιχα με αίσθημα πόνου που μοιάζει με τον καρδιακό αλλά οφείλεται σε άλλους παράγοντες και 5,72 χρόνια μικρότερα από αυτά που είναι ασυμπτωματικά.

Μέγιστος καρδιακός ρυθμός (MaxHR): Οι μέσοι παλμοί του MaxHR για το επίπεδο TA, είναι 147,89 με  $\Delta E_{TA} = [141,42, 154,46]$ , για το ATA είναι 150,21 με  $\Delta E_{ATA} = [146,34, 154]$ , για το NAP 143,24 με  $\Delta E_{NAP} = [139,62, 146,65]$  και για το ASY είναι 128,48 με  $\Delta E_{ASY} = [126,44, 130,56]$ . Από τα ζεύγη, μόνο τα TA – ASY, ATA – ASY και NAP - ASY έχουν μη επικαλυπτόμενα διαστήματα και συνεπώς μόνο σε αυτά εντοπίζεται στατιστικά σημαντική διαφορά. Δηλαδή, τα άτομα με συνηθισμένο στηθαγχικό πόνο έχουν κατά μέσο όρο 19,41 παλμούς περισσότερους από αυτά χωρίς συμπτώματα. Τα άτομα που βιώνουν στηθαγχικό πόνο διαφορετικό από τον συνηθισμένο έχουν κατά μέσο όρο 21,73 παλμούς περισσότερους από τα ασυμπτωματικά και τέλος τα άτομα που αισθάνονται πόνο όμοιο με τον καρδιακό αλλά οφείλεται σε άλλους παράγοντες έχουν κατά μέσο όρο 14,76 παλμούς περισσότερους από τα άτομα χωρίς συμπτώματα.

Μέγιστη Μεταβολή του ηλεκτροκαρδιογραφήματος κατά τη διαδικασία του τεστ κοπώσεως (Oldpeak): Η μέση τιμή του Oldpeak στο επίπεδο TA είναι 1,04 με  $\Delta E_{TA} = [0,77, 1,31]$ , στο ATA είναι 0,31 με  $\Delta E_{ATA} = [0,22, 0,40]$ , στο NAP είναι 0,68 με  $\Delta E_{NAP} = [0,55, 0,79]$  και στο επίπεδο ASY 1,16 με  $\Delta E_{ASY} = [1,06, 1,27]$ . Τα ζεύγη με μη επικαλυπτόμενα διαστήματα που έχουν στατιστικά σημαντικές διαφορές είναι τα TA – ATA, ATA – NAP, ATA – ASY και NAP – ASY. Συγκεκριμένα, τα άτομα με συνηθισμένο στηθαγχικό πόνο έχουν κατά μέσο όρο 0,73 mm

μικρότερη μεταβολή στο ηλεκτροκαρδιογράφημα από αυτά που βιώνουν στηθαγχικό πόνο διαφορετικό από τον συνηθισμένο τα οποία όμως έχουν κατά μέσο όρο 0,37 mm μικρότερη μεταβολή στο ηλεκτροκαρδιογράφημα από αυτά που αισθάνονται πόνο παρόμοιο με τον καρδιακό αλλά οφείλεται σε άλλους παράγοντες και 0,85 mm από τα ασυμπτωματικά. Τέλος, τα άτομα με αίσθημα πόνο που μοιάζει με τον καρδιακό αλλά οφείλεται σε άλλους παράγοντες έχουν κατά μέσο όρο 0,48 mm μικρότερη μεταβολή του ηλεκτροκαρδιογραφήματος.

Χοληστερίνη: Η μέση τιμή της χοληστερίνης είναι 233,66 mg/dL στο επίπεδο TA με  $\Delta E_{TA} = [221,70, 245,60]$ , στο επίπεδο ATA με μέση τιμή 242,95 mg/dL και  $\Delta E_{ATA} = [235,63, 250,62]$ , στο NAP με μέση τιμή 238,41 mg/dL με  $\Delta E_{NAP} = [231,23, 245,72]$  και στο ASY με μέση τιμή 248,79 με  $\Delta E_{ASY} = [243,97, 253,44]$ . Όλα τα ζεύγη έχουν επικαλυπτόμενα διαστήματα συνεπώς δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των επιπέδων. Δηλαδή, οι τιμές της χοληστερίνης δεν διαφέρουν σημαντικά σε κανένα από τα επίπεδα του στηθαγχικού πόνου.

## ΚΕΦΑΛΑΙΟ 4

### Στατιστική Ανάλυση και Μοντελοποίηση

#### 4.1 Εισαγωγή

Στο κεφάλαιο αυτό θα αναλυθούν οι διάφορες μεταβλητές του δείγματος με σκοπό την κατανόηση της επίδρασής τους στην πρόβλεψη καρδιακών παθήσεων. Συγκεκριμένα, θα εφαρμοστεί ανάλυση παραγόντων για την εύρεση παραγόντων που επηρεάζουν στις καρδιακές παθήσεις, μοντέλο πολλαπλής λογιστικής παλινδρόμηση που θα δείξει ποιες μεταβλητές μπορούν να προβλέψουν την εμφάνιση των καρδιακών παθήσεων και καμπύλες ROC για την αξιολόγηση της ικανότητας του μοντέλου να προβλέπει σωστά την πιθανότητα εμφάνισης καρδιακής πάθησης.

## 4.2 Ανάλυση παραγόντων

Για την εφαρμογή της παραγοντικής ανάλυσης επιλέχθηκαν όλες οι συνεχείς μεταβλητές: ηλικία (Age), αρτηριακή πίεση σε κατάσταση ηρεμίας (RestingBP), μέγιστη μεταβολή του ηλεκτροκαρδιογραφήματος κατά το τεστ κοπώσεως (Oldpeak), χοληστερίνη (Cholesterol) και ο μέγιστος καρδιακός ρυθμός (MaxHR) καθώς υπήρχε συσχέτιση μεταξύ τους (3.3 έλεγχος Spearman's). Πριν συνεχίσουμε, θα ελεγχθεί η ύπαρξη επαρκών συσχετίσεων ανάμεσα στις μεταβλητές αυτές κάνοντας χρήση των παρακάτω στατιστικών δεικτών:

### 4.2.1 Υποθέσεις ανάλυσης παραγόντων

#### Kaiser-Meyer-Olkin (KMO)

Είναι ένα μέτρο που χρησιμοποιείται για την αξιολόγηση της καταλληλότητας των δεδομένων στην ανάλυση παραγόντων, συγκρίνοντας τα μεγέθη των συντελεστών συσχέτισης με τους μερικούς συντελεστές συσχέτισης.

$$KMO = \frac{\sum_{i=1}^{\rho} \sum_{j \neq i} \rho_{ij}^2}{\sum_{i=1}^{\rho} \sum_{j \neq i} \rho_{ij}^2 + \sum_{i=1}^{\rho} \sum_{j \neq i} a_{ij}^2}$$

Όπου,

- Οι  $\rho_{ij}$ ,  $a_{ij}$ : συντελεστές συσχέτισης και μερικής συσχέτισης των  $X_i X_j$
- $KMO > 0,5$  για την καταλληλότητα των μεταβλητών της παραγοντικής ανάλυσης. Παίρνοντας τιμές κοντά στο 0,8 γίνεται λήψη ικανοποιητικών αποτελεσμάτων στην ανάλυση.

#### Measure of sampling adequacy (MSA)

Όπως και το μέτρο KMO, έτσι κι αυτό, ως στόχο του έχει την αξιολόγηση των μεταβλητών της ανάλυσης παραγόντων. Διαφοροποιείται στο ότι υπολογίζει την κάθε μεταβλητή ξεχωριστά και όχι σαν σύνολο όπως το KMO.

$$MSA_i = \frac{\sum_{j=1}^{\rho} \rho}{\sum_{j=1}^{\rho} \rho_{ij}^2 + \sum_{j=1}^{\rho} a_{ij}^2}, i = 1, 2, \dots, \rho$$

Όπου,

- Όταν η τιμή του δείκτη MSA βρίσκεται κοντά στη μονάδα, τότε η κάθε υπό εξέταση μεταβλητή είναι κατάλληλη για ανάλυση παραγόντων<sup>[24]</sup> (Κούτρας Μ. 2021).

#### 4.2.2 Ανάλυση παραγόντων και καρδιακές παθήσεις

Θα πραγματοποιηθεί προσπάθεια εύρεσης κοινών παραγόντων των επιλεγμένων μεταβλητών, με σκοπό την καλύτερη κατανόηση αυτών και διεξαγωγή συμπερασμάτων.

Correlation Matrixa						
		Age	MaxHR	Oldpeak	RestingBP	Cholesterol
Correlation	Age	1,000	-,382	,259	,263	,053
	MaxHR	-,382	1,000	-,161	-,110	-,017
	Oldpeak	,259	-,161	1,000	,174	,053
	RestingBP	,263	-,110	,174	1,000	,083
	Cholesterol	,053	-,017	,053	,083	1,000
Sig. (1-tailed)	Age		,000	,000	,000	,053
	MaxHR	,000		,000	,000	,301
	Oldpeak	,000	,000		,000	,054
	RestingBP	,000	,000	,000		,006
	Cholesterol	,053	,301	,054	,006	
a. Determinant = ,722						

Πίνακας 45: Πίνακας συσχετίσεων

Από τον παραπάνω πίνακα συσχετίσεων φαίνεται πως τα ζεύγη των μεταβλητών Cholesterol - Age, Cholesterol – MaxHR και Cholesterol – Oldpeak έχουν τιμές pvalue > 0,05. Άρα σε επίπεδο σημαντικότητας 5% δεν υπάρχουν επαρκείς ενδείξεις για την απόρριψη της H<sub>0</sub> (H<sub>0</sub>: Δεν υπάρχει συσχέτιση μεταξύ των μεταβλητών έναντι H<sub>a</sub>: Not H<sub>0</sub>). Συνεπώς, οι τιμές της χοληστερίνης δεν σχετίζονται με την ηλικία, το μέγιστο καρδιακό ρυθμό και τη μέγιστη μεταβολή του ηλεκτροκαρδιογραφήματος κατά το τεστ κοπώσεως. Τα υπόλοιπα ζεύγη



παρουσιάζουν τιμές p-value κοντά στο 0, επομένως, στο ίδιο επίπεδο σημαντικότητας, απορρίπτεται η  $H_0$ . Αυτό σημαίνει ότι υπάρχουν στατιστικά σημαντικές σχέσεις μεταξύ αυτών

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,619
Bartlett's Test of Sphericity	Approx. Chi-Square	297,884
	df	10
	Sig.	,000

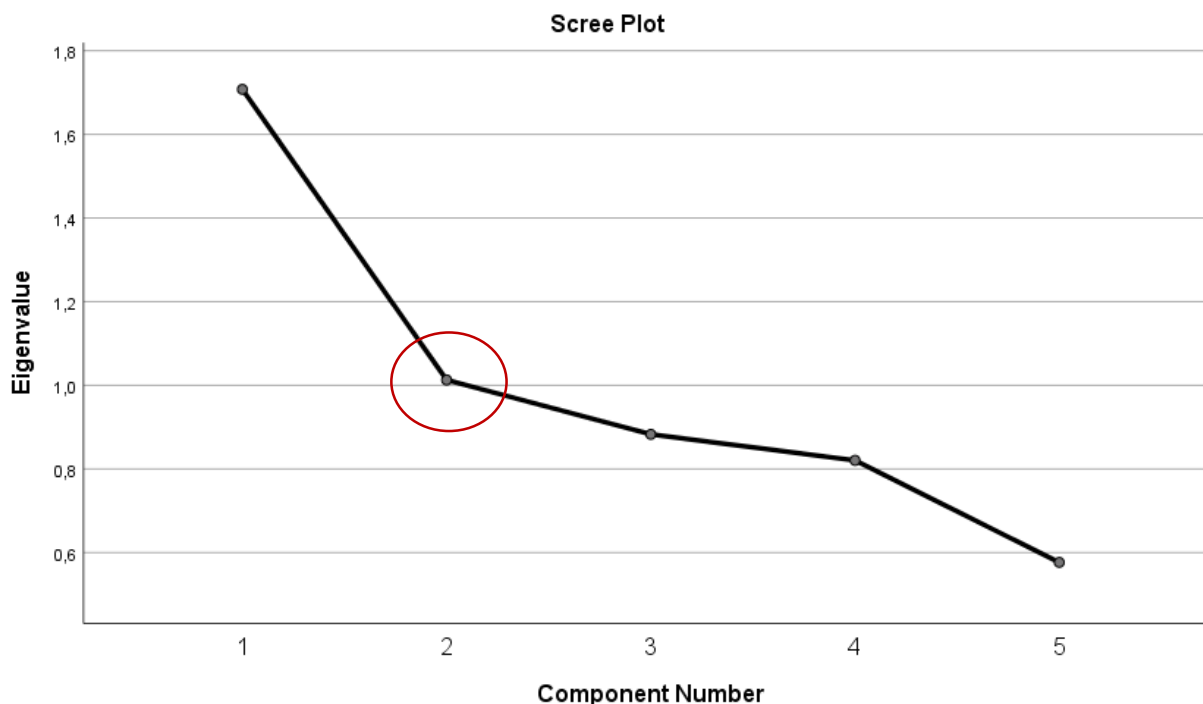
Πίνακας 46: Έλεγχος καταλληλότητας των μεταβλητών για παραγοντική ανάλυση με τη μέθοδο KMO and Bartlett's Test.

Βάσει του πάνω πίνακα, η τιμή του δείκτη  $KMO=0,619 > 0,5$ . Έτσι, τα δεδομένα είναι κατάλληλα για την εφαρμογή της παραγοντικής ανάλυσης. Συνεχίζοντας, η τιμή pvalue του ελέγχου Bartlett's Test of Sphericity είναι πολύ κοντά στο  $0 < \alpha = 0,05$ . Άρα, σε επίπεδο σημαντικότητας 5% απορρίπτεται η  $H_0$  (όπου  $H_0$  ο πίνακας συσχετίσεων είναι μοναδιαίος έναντι  $H_a: \text{Not } H_0$ ) δηλαδή, υπάρχει εξάρτηση μεταξύ κάποιων μεταβλητών.

Anti-image Matrices						
		Age	MaxHR	Oldpeak	RestingBP	Cholesterol
Anti-image Covariance	Age	,776	,283	-,156	-,178	-,022
	MaxHR	,283	,850	,061	,002	-,006
	Oldpeak	-,156	,061	,916	-,102	-,032
	RestingBP	-,178	,002	-,102	,915	-,064
	Cholesterol	-,022	-,006	-,032	-,064	,991
Anti-image Correlation	Age	,586 <sup>a</sup>	,348	-,185	-,212	-,025
	MaxHR	,348	,593 <sup>a</sup>	,069	,003	-,006
	Oldpeak	-,185	,069	,707 <sup>a</sup>	-,111	-,033
	RestingBP	-,212	,003	-,111	,658 <sup>a</sup>	-,068
	Cholesterol	-,025	-,006	-,033	-,068	,670 <sup>a</sup>
a. Measures of Sampling Adequacy(MSA)						

Πίνακας 47: Έλεγχος καταλληλότητας των μεταβλητών για παραγοντική ανάλυση με τη μέθοδο MSA.

Οι τιμές της διαγωνίου του επάνω πίνακα είναι όλες μεγαλύτερες του 0,5 και είναι κατάλληλες για ανάλυση. Επίσης, οι τιμές εκτός αυτής είναι αρκετά μικρές.



Σχήμα 15: Επιλογή αριθμού παραγόντων με βάση τον κανόνα του «αγκώνα».

Από το επάνω γράφημα παρατηρείται η ύπαρξη δύο παραγόντων σύμφωνα με τον εμπειρικό κανόνα του αγκώνα. Ο κανόνας αυτός αναφέρει πως η επιλογή των συνιστωσών γίνεται πριν την ομαλοποίηση της πτώσης, δηλαδή στο σημείο που σχηματίζεται ένας αγκώνας. Επιπλέον, αυτό επαληθεύεται και από το κριτήριο του Kaiser, καθώς η ιδιοτιμή  $\lambda_3$  είναι  $< 1$  ( $\lambda_3 = 0,883$ ).

#### Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings <sup>a</sup>
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	1,708	34,151	34,151	1,708	34,151	34,151	1,694
2	1,013	20,256	54,407	1,013	20,256	54,407	1,060
3	,883	17,655	72,062				
4	,820	16,410	88,472				
5	,576	11,528	100,000				

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

Πίνακας 48: Πίνακας συνολικής επεξηγηματικής διακύμανσης.

Από τον παραπάνω πίνακα, εντοπίζεται η ύπαρξη δύο παραγόντων. Ο πρώτος παράγοντας μόνος του επεξηγεί το 34,15% της συνολικής μεταβλητότητας, ενώ μαζί με τον δεύτερο παράγοντα, επεξηγούν το 54,41% της συνολικής μεταβλητότητας.

Component Matrixa		
	Component	
	1	2
Age	,781	-,150
MaxHR	-,645	,350
Oldpeak	,583	,043
RestingB	,555	,278
Cholesterol	,184	,888
Extraction Method: Principal Component Analysis.		

Πίνακας 49: Κατανομή μεταβλητών στους παράγοντες.

Ο πίνακας επάνω, δείχνει τη συνεισφορά των μεταβλητών στους δύο παράγοντες. Παρατηρούμε πως οι μεταβλητές Age, MaxHR, Oldpeak και RestingBP έχουν υψηλά φορτία στον 1<sup>ο</sup> παράγοντα ενώ η μεταβλητή Cholesterol στον 2<sup>ο</sup>.

Ο 1<sup>ος</sup> παράγοντας φαίνεται να σχετίζεται με την ηλικία, τον μέγιστο καρδιακό ρυθμό, τη μέγιστη μεταβολή που εμφανίζεται στο ηλεκτροκαρδιογράφημα κατά τη διαδικασία του τεστ κοπώσεως και την αρτηριακή πίεση σε κατάσταση ηρεμίας. Έτσι, ο παράγοντας αυτός θα μπορούσε να περιγράψει τη μεταβολή της καρδιακής κατάστασης κατά τη φυσιολογική γήρανση. Αντίστοιχα, ο 2<sup>ος</sup> παράγοντας σχετίζεται με τη χοληστερίνη άρα θα μπορούσε να περιγράψει τη χοληστερόλη και τα λιπίδια.

### 4.3 Λογιστική Παλινδρόμηση

Με σκοπό την πρόβλεψη των τιμών μιας δίτιμης εξαρτημένης μεταβλητής και περισσότερων από δύο ανεξάρτητων μεταβλητών, θα γίνει χρήση της πολλαπλής λογιστικής παλινδρόμησης που είναι της μορφής:

$$\text{Log} \left( \frac{p_i}{1 - p_i} \right) = b_0 + b_1 X_{1i} + b_2 X_{2i} + \dots + b_k X_{ki} + \varepsilon$$

Στην ανάλυση αυτή, η δίτιμη μεταβλητή HeartDisease ορίζεται ως η εξαρτημένη, και διαχωρίζει τα άτομα με καρδιακή ασθένεια από τα υγιή <sup>[25]</sup> (Σαχλάς Α., Μπερσίμης Σ., 2018). Επιπλέον, η μεταβλητή της ηλικίας χωρίζεται σε δύο υποομάδες, η πρώτη περιλαμβάνει τα άτομα ηλικίας 28 έως 54 ετών, ενώ η δεύτερη τα άτομα ηλικίας 55 έως 77 ετών. Ο διαχωρισμός βασίζεται στη διάμεσο της ηλικίας, που είναι τα 54 έτη, που μπορεί να θεωρηθεί μια κρίσιμη ηλικιακή μετάβαση. Ως κατηγορία αναφοράς ορίζεται η τελευταία τιμή του κάθε επιπέδου.

Στη συνέχεια, για τον έλεγχο της στατιστικής σημαντικότητας του μοντέλου θα γίνει χρήση του ελέγχου καλής προσαρμογής Hosmer and Lemeshow με τις εξής υποθέσεις:

H<sub>0</sub>: Το μοντέλο έχει καλή προσαρμογή

H<sub>a</sub>: Το μοντέλο δεν έχει καλή προσαρμογή

Το μοντέλο που επιλέχθηκε να παρουσιαστεί αποτελείται από τις μεταβλητές Sex, ChestPainType, FastingBS, ExerciseAngina, Oldpeak, ST\_Slope και Age\_grouped. Χρησιμοποιώντας απλή γραμμική παλινδρόμηση, ελέγχθηκαν οι ανεξάρτητες μεταβλητές για πολυσυγγραμμικότητα μέσω του δείκτη VIF. Σε όλες τις περιπτώσεις, οι τιμές VIF των ανεξάρτητων μεταβλητών βρέθηκαν σημαντικά μικρότερες από το 10 και αρκετά κοντά στη μονάδα.

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	609,459a	,509	,681

Πίνακας 50: Έλεγχος προσαρμογής του μοντέλου.

Σύμφωνα με τον παραπάνω πίνακα, το επιλεγμένο μοντέλο προσαρμόζεται ικανοποιητικά στα δεδομένα. Συγκεκριμένα, με βάση τις τιμές των δεικτών Cox & Snell R Square (0,509) και Nagelkerke R Square (0,681), το μοντέλο εξηγεί το 50,9% έως 68,1% της συνολικής μεταβλητότητας.

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	8,144	8	,420

Πίνακας 51: Έλεγχος καλής προσαρμογής του μοντέλου.

Από τον επάνω πίνακα, παρατηρείται πως το  $pvalue=0,420 > 0,05$ . Άρα, σε επίπεδο σημαντικότητας 5% δεν υπάρχουν επαρκείς ενδείξεις για την απόρριψη της  $H_0$ . Δηλαδή, το μοντέλο έχει καλή προσαρμογή στα δεδομένα.

Classification Tablea					
		Predicted			
		HeartDisease		Percentage Correct	
Observed	Normal	Heart Disease	Overall Percentage		
	Step 1	HeartDisease		Normal	340
		Heart Disease	54	454	89,4
					86,5

Πίνακας 52: Έλεγχος ορθής ταξινόμησης.

Σύμφωνα με τον παραπάνω πίνακα, το ποσοστό σωστής ταξινόμησης του μοντέλου για τα υγιή άτομα είναι 82,9%, για τα άτομα με καρδιακή πάθηση είναι 89,4%, ενώ το συνολικό ποσοστό ορθής ταξινόμησης ανέρχεται στο 86,5%, γεγονός που θεωρείται αρκετά ικανοποιητικό.

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1a	Sex(Male)	1,601	,273	34,334	1	,000	4,956
	ChestPainType			66,205	3	,000	
	ChestPainType(TA)	-1,576	,426	13,706	1	,000	,207
	ChestPainType(ATA)	-1,990	,323	38,058	1	,000	,137
	ChestPainType(NAP)	-1,663	,256	42,312	1	,000	,190

<b>FastingBS(0)</b>	-1,351	,264	26,161	1	,000	,259
<b>ExerciseAngina(YES)</b>	-,941	,234	16,207	1	,000	,390
<b>Oldpeak</b>	,334	,113	8,704	1	,003	1,397
<b>ST_Slope</b>			108,983	2	,000	
<b>ST_Slope(1)</b>	-1,139	,442	6,657	1	,010	,320
<b>ST_Slope(2)</b>	1,298	,425	9,351	1	,002	3,663
<b>Age_grouped(1)</b>	-,566	,211	7,178	1	,007	,568
<b>Constant</b>	1,321	,560	5,562	1	,018	3,748

a. Variable(s) entered on step 1: Sex, ChestPainType, FastingBS, ExerciseAngina, Oldpeak, ST\_Slope, Age\_grouped.

Πίνακας 53: Αποτελέσματα λογιστικής παλινδρόμησης.

Με βάση τον παραπάνω πίνακα, γίνεται αντιληπτό πως όλες οι μεταβλητές που έχουν συμπεριληφθεί στο μοντέλο είναι στατιστικά σημαντικές καθώς έχουν  $p$ value  $< 0,05$ . Συνεπώς, σε επίπεδο σημαντικότητας 5% απορρίπτεται η  $H_0$  και οι μεταβλητές έχουν κάποια αξία στην πρόβλεψη καρδιακών παθήσεων. Αναλυτικά:

Φύλο: ο σχετικός λόγος πιθανοτήτων των ατόμων του αντρικού φύλου είναι 4,956. Άρα, έχουν 395,6% μεγαλύτερη σχετική πιθανότητα να εμφανίσουν καρδιακή πάθηση από τα άτομα του γυναικείου φύλου.

Στηθαγγικός πόνος: ο σχετικός λόγος πιθανοτήτων των ατόμων με συνηθισμένο στηθαγγικό πόνο είναι 0,207 δηλαδή, έχουν 79,3% μικρότερη σχετική πιθανότητα να εμφανίσουν καρδιακή πάθηση από τα ασυμπτωματικά άτομα. Ομοίως, ο σχετικός λόγος πιθανοτήτων των ατόμων με διαφορετικό από τον συνηθισμένο στηθαγγικό πόνο είναι 0,137, και των ατόμων με παρόμοιο καρδιακό πόνο που οφείλεται όμως σε διαφορετικούς παράγοντες είναι 0,190. Έτσι, έχουν 86,3% και 81% αντιστοίχως μικρότερη σχετική πιθανότητα να εμφανίσουν καρδιακή πάθηση από τα ασυμπτωματικά άτομα.

Ζάχαρο νηστείας: ο σχετικός λόγος πιθανοτήτων των ατόμων με τιμή ζαχάρου νηστείας μικρότερη ή ίση του 120 mg/dl είναι 0,259. Συνεπώς, έχουν 74,1% μικρότερη σχετική πιθανότητα να εμφανίσουν καρδιακή πάθηση σε σχέση με τα άτομα με υψηλότερες τιμές ζαχάρου.

Πρόκληση στηθαγγικού πόνου από άσκηση: ο σχετικός λόγος πιθανοτήτων των ατόμων που δεν βιώνουν στηθαγγικό πόνο κατά τη διάρκεια της άσκησης είναι 0,390. Αυτό σημαίνει ότι έχουν 61% μικρότερη σχετική πιθανότητα εμφάνισης καρδιακής πάθησης από τα άτομα που βιώνουν στηθαγγικό πόνο κατά την άσκηση..

Μέγιστη μεταβολή του ηλεκτροκαρδιογραφήματος κατά τη διαδικασία του τεστ κοπώσεως: ο σχετικός λόγος πιθανοτήτων των ατόμων που κάνουν το τεστ κοπώσεως είναι 1,397. Άρα, για κάθε μονάδα απόκλισης της μέγιστης μεταβολής του ηλεκτροκαρδιογραφήματος, αυξάνεται κατά 39,7% η σχετική πιθανότητα εμφάνισης καρδιακής πάθησης.

Αλλαγή του μοτίβου της κλίσης κατά το σημείο μέγιστης έντασης του τεστ κοπώσεως στο ηλεκτροκαρδιογράφημα: ο σχετικός λόγος πιθανοτήτων των ατόμων με αλλαγή σε ανοδική μεταβολή είναι 0,320 και χωρίς αλλαγή είναι 3,663. Έτσι, τα άτομα με ανοδική μεταβολή στο μοτίβο, έχουν 68% μικρότερη σχετική πιθανότητα εμφάνισης καρδιακή πάθησης από αυτά με καθοδική μεταβολή στην κλίση τους και αυτά με καμία αλλαγή στην κλίση του μοτίβου τους έχουν 266,3% μεγαλύτερη σχετική πιθανότητα από τα άτομα με καθοδική μεταβολή.

Ηλικιακή ομάδα (Age grouped): ο σχετικός λόγος πιθανοτήτων των ατόμων που ανήκουν στην ηλικιακή ομάδα των 28 έως 54 χρόνων είναι 0,568. Έχουν δηλαδή 43,2% μικρότερη σχετική πιθανότητα να εμφανίσουν καρδιακή πάθηση σε σχέση με τα άτομα που ανήκουν στην ηλικιακή ομάδα των 55 έως 77 χρόνων.

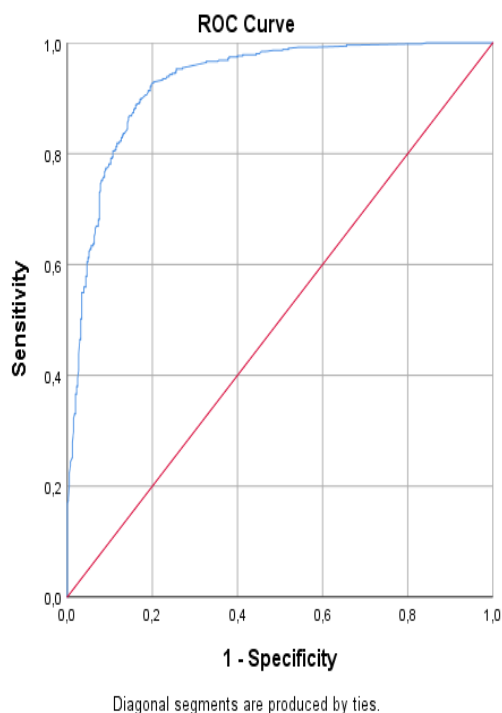
#### 4.4 Καμπύλη ROC

Στο παραπάνω μοντέλο πρόβλεψης, θα εφαρμοστεί η καμπύλη ROC προκειμένου να αξιολογηθεί η ακρίβεια του μοντέλου με χρήση της τιμής AUC (Area Under Curve). Η τιμή AUC κυμαίνεται από 0 έως 1 και πρέπει να είναι μεγαλύτερη από 0,5 για να υποδηλώνει ότι οι σωστές εκτιμήσεις υπερτερούν των λανθασμένων. Για τιμές AUC μεγαλύτερες του 0,8, λαμβάνουμε αρκετά καλές προβλέψεις.<sup>[26]</sup> (Σαχλάς Α., Μπερσίμης Σ., 2018).

Area Under the Curve
Test Result Variable(s): Predicted probability
Area
,929
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Πίνακας 54: Αξιολόγηση διαγνωστικής ικανότητας του μοντέλου.

Με βάση τον πάνω πίνακα, η τιμή AUC είναι 0,929, δηλαδή το μοντέλο έχει αρκετά καλή προβλεπτική ικανότητα στην διάκριση μεταξύ υγιών ατόμων και ατόμων με καρδιακή πάθηση.



Σχήμα 16: Προβλεπτική ικανότητα μοντέλου με χρήση καμπύλης ROC.

Από το πάνω σχήμα φαίνεται πως η καμπύλη βρίσκεται πολύ κοντά στην πάνω αριστερή γωνία υποδηλώνοντας έτσι μια πολύ καλή διαγνωστική αξία του μοντέλου.

## 5. Συμπεράσματα ανάλυσης

Στο κεφάλαιο αυτό θα γίνει μια σύνοψη των κύριων αποτελεσμάτων της ανάλυσης που πραγματοποιήθηκε προκειμένου να διαπιστωθεί ποιες είναι οι μεταβλητές οι οποίες μπορούν να επηρεάσουν την εμφάνιση καρδιακών παθήσεων.

Αρχικά, το φύλο είναι ένας αρκετά σημαντικός παράγοντας που σχετίζεται με τις καρδιακές παθήσεις. Όπως διαπιστώσαμε, οι άντρες παρουσιάζουν υψηλότερη συχνότητα εμφάνισης καρδιακών παθήσεων από τις γυναίκες. Το ίδιο ισχύει και για τα άτομα



μεγαλύτερης ηλικίας 55 χρόνων και πάνω, που όσο μεγαλώνουν, παρατηρείται αύξηση των καρδιακών προβλημάτων τους σε σύγκριση με τα άτομα νεότερης ηλικίας. Συνεχίζοντας με τη μελέτη ορισμένων ιατρικών δεικτών όπως είναι το ζάχαρο νηστείας, βλέπουμε πως τα άτομα με υψηλότερες τιμές από 120mg/dl διατρέχουν μεγαλύτερο κίνδυνο από τα υπόλοιπα. Επιπλέον, υπάρχουν και συγκεκριμένες εξετάσεις που σχετίζονται με τον εντοπισμό των καρδιακών νοσημάτων όπως είναι το τεστ κοπώσεως και άλλες μετρήσεις που γίνονται στο ηλεκτροκαρδιογράφημα και μπορούν να δώσουν πληροφορίες για την υγεία της καρδιάς. Για παράδειγμα, κατά το τεστ κοπώσεως που πραγματοποιείται με χρήση ηλεκτροκαρδιογραφήματος, όσο μεγαλύτερη είναι η μεταβολή που μετράται, τόσο μεγαλώνει και η πιθανότητα της εμφάνισης κάποιας καρδιακής πάθησης. Επιπρόσθετα, υπάρχουν κι άλλοι σημαντικοί παράγοντες που είναι άρρηκτα συνδεδεμένοι με τις καρδιακές παθήσεις όπως είναι ο πόνος στο στήθος. Γενικότερα, ο πόνος στο στήθος μπορεί να εμφανίζεται μετά από σωματική άσκηση ή και σε κατάσταση ηρεμίας. Είναι κάτι που δεν πρέπει να αμελείται, καθώς από τη μελέτη αυτή καταλήξαμε στο λογικό συμπέρασμα πως τα άτομα που βιώνουν στηθόπονο έχουν μεγαλύτερες πιθανότητες για ασθένειες που σχετίζονται με την καρδιά.

Συνοψίζοντας, σύμφωνα με τα αποτελέσματα της έρευνας, οι μεταβλητές δημογραφικών χαρακτηριστικών και ορισμένοι ειδικοί ιατρικοί δείκτες παίζουν σημαντικό ρόλο στην πρόβλεψη καρδιακών ασθενειών. Συμπερασματικά, είναι καίριας σημασίας η σωστή ενημέρωση των ατόμων σχετικά με τους παράγοντες κινδύνου ώστε να λαμβάνουν επαρκή μέτρα για τη βελτίωση της υγείας τους όπως είναι η μεταστροφή σε έναν πιο υγιινό τρόπο ζωής, η τακτική παρακολούθηση των ιατρικών εξετάσεων αλλά και η εσπευσμένη αναζήτηση συμβουλής από εμπειρογώμονες ιατρούς σε περίπτωση ανησυχητικών συμπτωμάτων.

# ΒΙΒΛΙΟΓΡΑΦΙΑ

## Ελληνική

1. Λεμπιδάκης Δ. , Λουλακάκης Μ. , Χαμηλός Μ. , Σκαλίδης Ε. (2021). *Στεφανιαία νόσος: Ένας κατά συρροή δολοφόνος ζει δίπλα(;) μας*, Ελληνική Καρδιολογική Εταιρεία. <https://www.hcs.gr/astheneis-koino/stefaniaia-nosos/>
6. Ελληνική Καρδιολογική Εταιρεία, (2022). *ΚΟΛΠΙΚΗ ΜΑΡΜΑΡΥΓΗ* από <https://www.hcs.gr/astheneis-koino/pliforisi/kolpiki-marmarygi/>
7. ΕΚΠΑ, *Έλεγχος Μυοκαρδιοπαθειών*, από <https://ekkan.gr/elegchos-myokardiopatheion/>
10. Κωστάκου Π. Μ., Κουρής Ν. Θ., Τρύφου Ε. Σ., Δαμάσκος Δ. Σ. , Ολύμπιος Χ. Δ. (2017). Μη Συμπαγές Μυοκάρδιο: Ένας Ξεχωριστός Φαινότυπος με Γενετική Ετερογένεια; Νεώτερα Δεδομένα, Ελληνική Καρδιολογική Επιθεώρηση, 322-329. από <https://www.hcs.gr/wp-content/uploads/2023/08/66-1-194-1-10-20170712.pdf>
12. Ελληνική Δημοκρατία Υπουργείο Υγείας, (2024). *Παγκόσμια Ημέρα για τα Αγγειακά Εγκεφαλικά Επεισόδια*, από <https://www.moh.gov.gr/articles/news/1442-pagkosmia-hmera-gia-ta-aggeiaka-kardiaka-epeisodia>
24. Κούτρας Μ., (2021). *ΠΜΣ «Εφαρμοσμένη Στατιστική» Εφαρμοσμένη Πολυμεταβλητή Ανάλυση – Ανάλυση παραγόντων»*
25. Σαχλάς, Α. & Μπερσίμης (2018). «Εφαρμοσμένη Στατιστική με Έμφαση στις Επιστήμες Υγείας, Εκδόσεις» Τζιόλα, Αθήνα.

## Ξένη

2. American Association of Neurosurgical Surgeons από <https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Cerebrovascular-Disease>
3. World Health Organization (2020). *Rheumatic heart disease* από <https://www.who.int/news-room/fact-sheets/detail/rheumatic-heart-disease>
4. CDC, (2024). *About Heart Disease*, από [https://www.cdc.gov/heart-disease/about/?CDC\\_AAref\\_Val=https://www.cdc.gov/heartdisease/about.htm](https://www.cdc.gov/heart-disease/about/?CDC_AAref_Val=https://www.cdc.gov/heartdisease/about.htm)
5. British Heart Foundation, (2024). *Global Heart & Circulatory Diseases Factsheet*, από <https://www.bhf.org.uk/-/media/files/for-professionals/research/heart-statistics/bhf-cvd-statistics-global->

[factsheet.pdf?rev=f323972183254ca0a1043683a9707a01&hash=5AA21565EEE5D85691D37157B31E4AAA](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5501035/)

8. McFarland J. (2024), *Hypertrophic Cardiomyopathy (HCM)*, American Heart Association από <https://www.heart.org/en/health-topics/cardiomyopathy/what-is-cardiomyopathy-in-adults/hypertrophic-cardiomyopathy>
9. Shah S. N. , Kishore Umapathi K., Rout P., Horenstein M. S., Oliver T. I., (2024). *Arrhythmogenic Right Ventricular Cardiomyopathy*, National Library of Medicine από <https://www.ncbi.nlm.nih.gov/books/NBK470378/>
11. CDC, (2024). *About Valvular Heart Disease*, από [https://www.cdc.gov/heart-disease/about/valvular-heart-disease.html?CDC\\_AAref\\_Val=https://www.cdc.gov/heartdisease/valvular\\_disease.htm](https://www.cdc.gov/heart-disease/about/valvular-heart-disease.html?CDC_AAref_Val=https://www.cdc.gov/heartdisease/valvular_disease.htm)
13. American Association of Neurological Surgeons από <https://www.aans.org/patients/conditions-treatments/cerebrovascular-disease/>
14. Office of Communications and Public Liaison, National Institute of Neurological Disorders and Stroke, National Institutes of Health Department of Health and Human Services (2018). *Cerebral Aneurysms*, Maryland 20892-2540, NIH Publication No. 18-NS-5505 από <https://catalog.ninds.nih.gov/sites/default/files/publications/cerebral-aneurysms.pdf>
15. Cleveland Clinic (2024). *Moyamoya Disease*, από <https://my.clevelandclinic.org/health/diseases/17244-moyamoya-disease>
16. Cleveland Clinic (2022). *Vascular Disease (Vasculopathy)* από <https://my.clevelandclinic.org/health/diseases/17604-vascular-disease>
17. UC Davis Vascular Center, *Intestinal ischemic syndrome* από [https://health.ucdavis.edu/vascular/diseases/intestinal\\_ischemic\\_syndromes.html](https://health.ucdavis.edu/vascular/diseases/intestinal_ischemic_syndromes.html)
18. Cleveland Clinic (2022). *Raynaud's Syndrome*, από <https://my.clevelandclinic.org/health/diseases/9849-raynauds-phenomenon>
19. Jekowsky A. (2011). *Earliest Case Of Coronary Artery Disease Found In Egyptian Princess*, American college of cardiology από <https://www.acc.org/about-acc/press-releases/2011/04/05/16/22/mummies>
20. Ranhel A. S. Mesquita E. T. (2016), *The Middle Ages Contributions to Cardiovascular Medicine* από <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5062728/>
21. Hajar R. (2017). *Coronary Heart Disease: From Mummies to 21st Century*, από <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5501035/>
22. Kaggle (2021). *Heart Failure Prediction Dataset* από <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>
23. Άδεια χρήσης από <https://opendatacommons.org/licenses/odbl/1-0/>