

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
Σχολή Χρηματοοικονομικής και Στατιστικής



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

Συγκριτική μελέτη επιβίωσης ασθενών με καρδιαγγειακή νόσο με
τον γενικό πληθυσμό, με τη χρήση παραμετρικών μοντέλων
επιβίωσης για την περιοχή της Κέρκυρας

Νικολακάκης Στυλιανός

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, Ελλάδα
Μάρτιος 2025

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

Τα μέλη της Επιτροπής ήταν:

- Αναπληρωτής Καθηγητής Τζαβελάς Γεώργιος (Επιβλέπων)
- Καθηγητής Παναγιωτάκος Δημοσθένης
- Επίκουρος Καθηγητής Ανδρουλάκης Εμμανουήλ

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

UNIVERSITY OF PIRAEUS
School of Finance and Statistics



DEPARTMENT OF STATISTICS AND INSURANCE SCIENCE

**POSTGRADUATE PROGRAM IN
APPLIED STATISTICS**

A comparative study of survival between patients with cardiovascular disease and the general population, from the area of Corfu Island, with the use of parametric survival models.

Nikolakakis Stylianos

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

March 2025

[3]

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω πολύ τον επιβλέποντα καθηγητή μου κύριο Τζαβελά Γεώργιο για την βοήθειά του και για την υπομονή του όλο αυτό το διάστημα.

Ακόμη, θα ήθελα να ευχαριστήσω την οικογένειά μου , τους φίλους μου και την Κάλλια για την συμπαράσταση και την ώθηση όλα αυτά τα χρόνια.

ΠΕΡΙΛΗΨΗ

Παρότι στην πλειοψηφία των μελετών χρησιμοποιείται το μοντέλο αναλογικών κινδύνων (PH) του Cox, οι μελέτες δεν εξετάζουν την καταλληλότητα του μοντέλου. Ο στόχος της συγκεκριμένης μελέτης είναι η σύγκριση διαφόρων παραμετρικών μοντέλων επιβίωσης στον γενικό πληθυσμό. Όλα τα δεδομένα αντλήθηκαν από ένα υπό-δείγμα της μελέτης των 7 χωρών. Η συγκεκριμένη μελέτη εξέτασε την επίδραση διαφόρων παραγόντων στη 40ετή θνησιμότητα (από όλες τις αιτίες και από καρδιαγγειακή νόσο (CVD) ξεχωριστά) σε 529 αγρότες (ηλικία $49,8 \pm 5,7$ έτη) από την Κέρκυρα. Δημιουργήθηκαν πολυπαραγοντικά μοντέλα ανάλυσης επιβίωσης (PH Cox και επιταχυνόμενου χρόνου ζωής (AFT) με κατανομές Exponential, Weibull, Loglogistic, Lognormal). Τα PH Cox μοντέλα δεν είχαν καλή εφαρμογή βάσει των γραφημάτων Cox-Snell και αναδείχθηκαν κατώτερα από τα AFT μοντέλα. Από τα AFT μοντέλα, καλύτερη εφαρμογή είχε το Weibull, βάσει των γραφημάτων Cox-Snell & AIC και στην περίπτωση της ολικής επιβίωσης και της επιβίωσης από CVD. Με βάσει το AFT Weibull πολυπαραγοντικό μοντέλο, η ηλικία μείωνε την σχετική πιθανότητα συνολικής επιβίωσης, όπως και το κάπνισμα, ενώ η σωματική δραστηριότητα την αύξανε. Η ηλικία μείωνε και την επιβίωση από CVD, όπως και η συστολική αρτηριακή πίεση. Το AFT μοντέλο αποτελεί καλή εναλλακτική λύση όταν δεν πληρούνται οι προϋποθέσεις του μοντέλου PH Cox, με πιθανότατα καλύτερη επιλογή τη Weibull κατανομή.

ABSTRACT

Although Cox's proportional hazards (PH) model is widely used in research, proportionality assumption and goodness of fit are not evaluated in most studies. This study aims to compare different parametric survival models in the general population. All data are derived from a subsample of the *Seven Countries Study*. This study examined the effect of various factors on the 40-year mortality (from all causes and from cardiovascular disease (CVD) separately) in 529 farmers (age 49.8 ± 5.7 years) from Corfu Island. Multivariate survival analysis models (PH Cox and accelerated failure time (AFT) with Exponential, Weibull, Loglogistic, Lognormal distributions) were used. PH Cox models did not fit well based on Cox-Snell residual graphs and were found to be inferior to AFT models. Among all AFT models, Weibull distribution exhibited the most favorable properties, based on Cox-Snell graphs & AIC criterion in both overall mortality, and CVD mortality. Based on the AFT Weibull multivariate model, age decreased the relative probability of total survival, as did the smoking, while physical activity increased it. Age also decreased the CVD survival, as did systolic blood pressure. AFT model is a good alternative when the conditions of the PH Cox model are not met, with the Weibull distribution probably being the best choice.

Περιεχόμενα

1. Εισαγωγή	10
Εισαγωγή στην Καρδιαγγειακή Νόσο	10
2. Μεθοδολογία	16
2.1. Περιγραφή της μελέτης	16
2.2. Περιγραφή του δείγματος	16
2.3. Στατιστική ανάλυση	17
2.4. Μοντέλο αναλογικών κινδύνων του Cox.....	17
2.5. Παραμετρικά Μοντέλα επιβίωσης	20
3. Περιγραφική Ανάλυση	25
3.1. Περιγραφή των μεταβλητών	25
3.2. Περιγραφική Ανάλυση Δείγματος	26
3.3. Συσχέτιση θνησιμότητας από CVD με τα χαρακτηριστικά των ασθενών	35
3.4. Γραφήματα Kaplan – Meier για τον θάνατο από CVD	42
4. Μοντέλα ανάλυσης επιβίωσης	44
4.1. Μοντέλα ανάλυσης επιβίωσης για τον θάνατο από CVD.....	44
i. Μοντέλο αναλογικών κινδύνων του Cox.....	44
ii. Μοντέλο AFT	51
4.2. Μοντέλα AFT ανάλυσης επιβίωσης για τον θάνατο από όλες τις αιτίες.....	59
5. Συζήτηση	67
6. Βιβλιογραφία.....	75

Κατάλογος πινάκων

Πίνακας 1. Περιγραφικά μέτρα μεταβλητής ηλικία (N=529).....	26
Πίνακας 2. Θνησιμότητα από όλες τις αιτίες (N=529).....	27
Πίνακας 3. Θνησιμότητα από CVD (N=529).....	27
Πίνακας 4. Σωματική δραστηριότητα συμμετεχόντων (N=529).....	28
Πίνακας 5. Κάπνισμα συμμετεχόντων (N=468).	29
Πίνακας 6. Περιγραφικά μέτρα μεταβλητής συστολική αρτηριακή πίεση (N=529).....	30
Πίνακας 7. Περιγραφικά μέτρα μεταβλητής διαστολική αρτηριακή πίεση (N=529).	31
Πίνακας 8. Περιγραφικά μέτρα μεταβλητής καρδιακός ρυθμός (N=529).	32
Πίνακας 9. Περιγραφικά μέτρα μεταβλητής ολική χοληστερόλη (N=521).....	33
Πίνακας 10. Περιγραφικά μέτρα μεταβλητής δείκτης μάζας σώματος (N=529).....	34
Πίνακας 11. Κατηγοριοποίηση δείκτη μάζας σώματος (N=529).	35
Πίνακας 12. Σύγκριση χαρακτηριστικών μεταξύ θανόντων έναντι μη θανόντων από CVD.	42
Πίνακας 13. Μονοπαραγοντικό και πολυπαραγοντικό μοντέλο αναλογικών κινδύνων του Cox για τον σχετικό κίνδυνο θανάτου από CVD.....	46
Πίνακας 14. Έλεγχος της υπόθεσης των αναλογικών κινδύνων χρησιμοποιώντας τα κατάλοιπα Schoenfeld (συσχέτιση με τον χρόνο επιβίωσης).	50
Πίνακας 15. Μονοπαραγοντικά μοντέλα AFT για την σχετική πιθανότητα επιβίωσης (μη θάνατος από CVD).....	56
Πίνακας 16. Πολυπαραγοντικά μοντέλα AFT για την σχετική πιθανότητα επιβίωσης (μη θάνατος από CVD).....	56
Πίνακας 17. Σύγκριση των AFT μοντέλων βάσει του AIC κριτηρίου.	58
Πίνακας 18. Μονοπαραγοντικά μοντέλα AFT για την σχετική πιθανότητα συνολικής επιβίωσης. .	64
Πίνακας 19. Πολυπαραγοντικά μοντέλα AFT για την σχετική πιθανότητα συνολικής επιβίωσης...	64
Πίνακας 20. Σύγκριση των AFT μοντέλων βάσει του AIC κριτηρίου.	66

ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ

Γράφημα 1. Ιστόγραμμα συχνοτήτων μεταβλητής ηλικία.	26
Γράφημα 2. Διάγραμμα πίτας μεταβλητής θνησιμότητα από όλες τις αιτίες.	27
Γράφημα 3. Διάγραμμα πίτας μεταβλητής θνησιμότητα από CVD.	28
Γράφημα 4. Διάγραμμα πίτας μεταβλητής σωματικής δραστηριότητα συμμετεχόντων.	28
Γράφημα 5. Διάγραμμα πίτας μεταβλητής καπνίσματος.	29
Γράφημα 6. Ιστόγραμμα συχνοτήτων μεταβλητής συστολική αρτηριακή πίεση.	30
Γράφημα 7. Ιστόγραμμα συχνοτήτων μεταβλητής διαστολική αρτηριακή πίεση.	31
Γράφημα 8. Ιστόγραμμα συχνοτήτων μεταβλητής καρδιακός ρυθμός.	32
Γράφημα 9. Ιστόγραμμα συχνοτήτων μεταβλητής ολική χοληστερόλη.	33
Γράφημα 10. Ιστόγραμμα συχνοτήτων μεταβλητής δείκτης μάζας σώματος.	34
Γράφημα 11. Διάγραμμα πίτας μεταβλητής δείκτη μάζας σώματος.	35
Γράφημα 12. Box-Plot της ηλικίας για τους δύο πληθυσμούς (θανόντες και μη θανόντες από CVD).	39
Γράφημα 13. Καμπύλες Kaplan – Meier για το θάνατο από CVD ανά σωματική δραστηριότητα. .	42
Γράφημα 14. Καμπύλες Kaplan – Meier για το θάνατο από CVD ανά καπνιστικές συνήθειες.	43
Γράφημα 15. Καμπύλες Kaplan – Meier για το θάνατο από CVD ανά κατηγορία δείκτη μάζας σώματος.	43
Γράφημα 16. Στικτόγραμμα του χρόνου (άξονας X) σε σχέση με τα κατάλοιπα Schoenfeld (άξονας Y) που προέκυψαν από το πολυπαραγοντικό μοντέλο αναλογικών κινδύνων του Cox για τον σχετικό κίνδυνο θανάτου από CVD (Πίνακας 13) των μεταβλητών ηλικία (A), σωματική δραστηριότητα (B), κάπνισμα (Γ), συστολική αρτηριακή πίεση (Δ), καρδιακός ρυθμός (Ε), ολική χοληστερόλη (ΣΤ) και δείκτης μάζας σώματος (Ζ).	49
Γράφημα 17. Διάγραμμα των καταλοίπων Cox-Snell (άξονας X) έναντι του εκτιμώμενοι αθροιστικού κινδύνου αυτών (Cumulative Hazard of Cox-Snell residuals - άξονας Y).	50
Γράφημα 18. Διαγράμματα των καταλοίπων Cox-Snell (άξονας X) έναντι του εκτιμώμενοι αθροιστικού κινδύνου αυτών (Cumulative Hazard of Cox-Snell residuals - άξονας Y) για τα πολυπαραγοντικά μοντέλα με Exponential, Weibull, Loglogistic και Lognormal κατανομές.	57
Γράφημα 19. Διαγράμματα των καταλοίπων Cox-Snell (άξονας X) έναντι του εκτιμώμενοι αθροιστικού κινδύνου αυτών (Cumulative Hazard of Cox-Snell residuals - άξονας Y) για τα πολυπαραγοντικά μοντέλα συνολικής επιβίωσης με Exponential, Weibull, Loglogistic και Lognormal κατανομές.	65

1. Εισαγωγή

Εισαγωγή στην Καρδιαγγειακή Νόσο

Η καρδιαγγειακή νόσος (ΚΝ) αποτελεί έναν από τους σημαντικότερους υγειονομικούς προβληματισμούς παγκοσμίως, επηρεάζοντας εκατομμύρια ανθρώπους και προκαλώντας υψηλά ποσοστά θνησιμότητας. Σύμφωνα με τα δεδομένα του Παγκόσμιου Οργανισμού Υγείας (Π.Ο.Υ.), η ΚΝ ευθύνεται για περίπου 17,9 εκατομμύρια θανάτους ετησίως, αντιπροσωπεύοντας περίπου το 32% όλων των θανάτων παγκοσμίως (Π.Ο.Υ., 2021). Η σοβαρότητα της κατάστασης αυξάνεται καθώς προβλέπεται περαιτέρω αύξηση της επίπτωσης και των θανάτων που σχετίζονται με ΚΝ τις επόμενες δεκαετίες. Έρευνες δείχνουν ότι η αστικοποίηση, η κακή διατροφή και η καθιστική ζωή συμβάλλουν στην αυξημένη επίπτωση της ΚΝ παγκοσμίως. Η διατροφή με υψηλή περιεκτικότητα σε κορεσμένα λιπαρά, αλάτι και ζάχαρη, σε συνδυασμό με την έλλειψη φυσικής δραστηριότητας, επιδεινώνει την κατάσταση και αυξάνει τον κίνδυνο ανάπτυξης ΚΝ (Mendis et al., 2019). Επιπλέον, οι καρδιαγγειακές παθήσεις δεν επηρεάζουν μόνο τους ηλικιωμένους αλλά και νεότερα άτομα, με την εμφάνιση των παραγόντων κινδύνου να αυξάνεται σταθερά λόγω ανθυγιεινών συνηθειών, όπως το κάπνισμα και η παχυσαρκία.

Η κοινωνικοοικονομική διάσταση της ΚΝ είναι επίσης αξιοσημείωτη. Σύμφωνα με έρευνες, οι χαμηλότερες κοινωνικοοικονομικές τάξεις έχουν αυξημένο κίνδυνο ανάπτυξης καρδιαγγειακών παθήσεων λόγω περιορισμένης πρόσβασης σε ποιοτική ιατρική φροντίδα, ανεπαρκούς ενημέρωσης για υγιεινούς τρόπους ζωής και αυξημένης έκθεσης σε περιβαλλοντικούς κινδύνους (Chi-Young Lee et al., 2021).

Ορισμός και Είδη Καρδιαγγειακής Νόσου

Η καρδιαγγειακή νόσος αναφέρεται σε ένα ευρύ φάσμα παθήσεων που επηρεάζουν την καρδιά και τα αιμοφόρα αγγεία. Οι κύριες μορφές της ΚΝ περιλαμβάνουν:

Στεφανιαία Νόσος (ΣΝ):

Προκαλείται από τη στένωση ή την απόφραξη των στεφανιαίων αρτηριών λόγω αθηρωματικών πλακών. Τα κυριότερα συμπτώματα περιλαμβάνουν στηθάγχη και έμφραγμα του μυοκαρδίου. Η στεφανιαία νόσος επηρεάζει κυρίως τους άνδρες και άτομα μεγαλύτερης ηλικίας, ενώ τα υψηλά επίπεδα LDL χοληστερόλης και η υπέρταση είναι σημαντικοί παράγοντες κινδύνου.

Καρδιακή Ανεπάρκεια:

Κατάσταση κατά την οποία η καρδιά δεν μπορεί να αντλήσει επαρκώς αίμα για να καλύψει τις ανάγκες του οργανισμού. Τα συμπτώματα περιλαμβάνουν δύσπνοια, κατακράτηση υγρών και κόπωση. Η καρδιακή ανεπάρκεια μπορεί να προκύψει από άλλες παθήσεις όπως στεφανιαία νόσος και υπέρταση.

Αγγειακή Εγκεφαλική Νόσος:

Παθήσεις που επηρεάζουν την αιμάτωση του εγκεφάλου, περιλαμβάνοντας εγκεφαλικά επεισόδια (ισχαιμικά και αιμορραγικά) και παροδικά ισχαιμικά επεισόδια. Η αγγειακή εγκεφαλική νόσος συνδέεται στενά με παράγοντες κινδύνου όπως η υπέρταση και η υπερλιπιδαιμία.

Περιφερική Αρτηριακή Νόσος (ΠΑΝ):

Στένωση ή απόφραξη των αρτηριών που τροφοδοτούν τα άκρα και άλλα όργανα, προκαλώντας πόνο, κράμπες και έλκη στα κάτω άκρα. Συνδέεται με το κάπνισμα και τον διαβήτη και είναι συχνότερη σε ηλικιωμένους.

Υπέρταση:

Αν και συχνά θεωρείται παράγοντας κινδύνου, όταν δεν ελέγχεται μπορεί να οδηγήσει σε βλάβες της καρδιάς, των νεφρών και των αιμοφόρων αγγείων. Η υπέρταση είναι πιο συχνή σε άτομα άνω των 50 ετών και συχνά δεν παρουσιάζει συμπτώματα μέχρι να προκαλέσει σοβαρές επιπλοκές (Π.Ο.Υ., 2021).

Επιδημιολογία της Καρδιαγγειακής Νόσου

Η καρδιαγγειακή νόσος είναι η κύρια αιτία θανάτου παγκοσμίως, ευθυνόμενη για περίπου 17,9 εκατομμύρια θανάτους κάθε χρόνο, κατά τον Παγκόσμιο Οργανισμό Υγείας. Σύμφωνα με την Global Burden of Disease Study (Mensah et al., 2023), η ΚΝ παραμένει η κύρια αιτία θανάτου παγκοσμίως, με υψηλότερα ποσοστά θνησιμότητας στις χώρες με χαμηλό και μεσαίο εισόδημα. Η αυξανόμενη επίπτωση της ΚΝ συνδέεται άμεσα με την αστικοποίηση, τη γήρανση του πληθυσμού και την αυξανόμενη επικράτηση των παραγόντων κινδύνου όπως η παχυσαρκία, η υπέρταση και ο διαβήτης (Bloom et al., 2011). Στην Ευρώπη, οι καρδιαγγειακές παθήσεις ευθύνονται για το 45% όλων των θανάτων. Στην Ελλάδα, τα ποσοστά θνησιμότητας από ΚΝ είναι επίσης υψηλά, με τις στατιστικές

μελέτες να δείχνουν ότι περίπου 40.000 άτομα χάνουν τη ζωή τους ετησίως λόγω αυτών των παθήσεων. Στην Κέρκυρα, τα στοιχεία δείχνουν ότι η ΚΝ αποτελεί σημαντική αιτία νοσηρότητας και θνησιμότητας, με αυξημένη συχνότητα εμφάνισης σε άτομα μεγαλύτερης ηλικίας. Είναι αξιοσημείωτο ότι οι άνδρες πλήττονται συχνότερα από τις γυναίκες, αν και η διαφορά αυτή μειώνεται με την ηλικία.

Παράγοντες Κινδύνου

Οι παράγοντες κινδύνου για την ανάπτυξη καρδιαγγειακής νόσου μπορούν να ταξινομηθούν σε σταθεροί και μεταβλητοί. Στους σταθερούς παράγοντες περιλαμβάνονται η ηλικία, το φύλο και η γενετική προδιάθεση. Οι ηλικιωμένοι έχουν μεγαλύτερο κίνδυνο ανάπτυξης ΚΝ, ενώ οι άνδρες πλήττονται συχνότερα από τις γυναίκες. Ωστόσο, μετά την εμμηνόπαυση, ο κίνδυνος για τις γυναίκες αυξάνεται και πλησιάζει αυτόν των ανδρών. Η γενετική προδιάθεση επίσης παίζει σημαντικό ρόλο, καθώς η ύπαρξη οικογενειακού ιστορικού ΚΝ αυξάνει τον κίνδυνο εμφάνισης της νόσου.

Οι μεταβλητοί παράγοντες κινδύνου περιλαμβάνουν την υπέρταση, την υπερλιπιδαιμία, το κάπνισμα, την παχυσαρκία, τον διαβήτη και τον καθιστικό τρόπο ζωής. Η υπέρταση, ή υψηλή αρτηριακή πίεση, αποτελεί έναν από τους πιο σημαντικούς παράγοντες κινδύνου για ΚΝ, καθώς μπορεί να προκαλέσει βλάβες στις αρτηρίες και την καρδιά. Η υπερλιπιδαιμία, ή υψηλά επίπεδα χοληστερόλης στο αίμα, συμβάλλει στη δημιουργία αθηρωματικών πλακών στις αρτηρίες, περιορίζοντας τη ροή του αίματος και αυξάνοντας τον κίνδυνο καρδιακής προσβολής και εγκεφαλικού επεισοδίου. Το κάπνισμα είναι ένας ισχυρός παράγοντας κινδύνου που μπορεί να προκαλέσει βλάβες στα αιμοφόρα αγγεία και να αυξήσει την αρτηριακή πίεση. Η παχυσαρκία και ο διαβήτης επίσης σχετίζονται στενά με την ανάπτυξη ΚΝ, καθώς επηρεάζουν αρνητικά το μεταβολισμό και την καρδιαγγειακή υγεία. Τέλος, ο καθιστικός τρόπος ζωής μειώνει την καρδιαγγειακή αντοχή και αυξάνει τον κίνδυνο ανάπτυξης ΚΝ.

Διάγνωση και Θεραπεία

Η διάγνωση της καρδιαγγειακής νόσου γίνεται μέσω διαφόρων κλινικών και εργαστηριακών εξετάσεων. Η ηλεκτροκαρδιογραφία (ΗΚΓ) είναι μια βασική εξέταση που καταγράφει την ηλεκτρική δραστηριότητα της καρδιάς και μπορεί να ανιχνεύσει ανωμαλίες στον καρδιακό ρυθμό και την αγωγή. Η ηχοκαρδιογραφία χρησιμοποιεί υπερήχους για να απεικονίσει την καρδιά και να αξιολογήσει τη λειτουργία της, ενώ οι αιματολογικές εξετάσεις μπορούν να μετρήσουν τα επίπεδα χοληστερόλης, γλυκόζης και άλλων δεικτών που σχετίζονται με την καρδιαγγειακή υγεία. Η απεικόνιση των στεφανιαίων αρτηριών με αξονική τομογραφία (CT) ή μαγνητική τομογραφία (MRI) μπορεί να

ανιχνεύσει την παρουσία αθηρωματικών πλακών και να αξιολογήσει τον βαθμό στένωσης των αρτηριών.

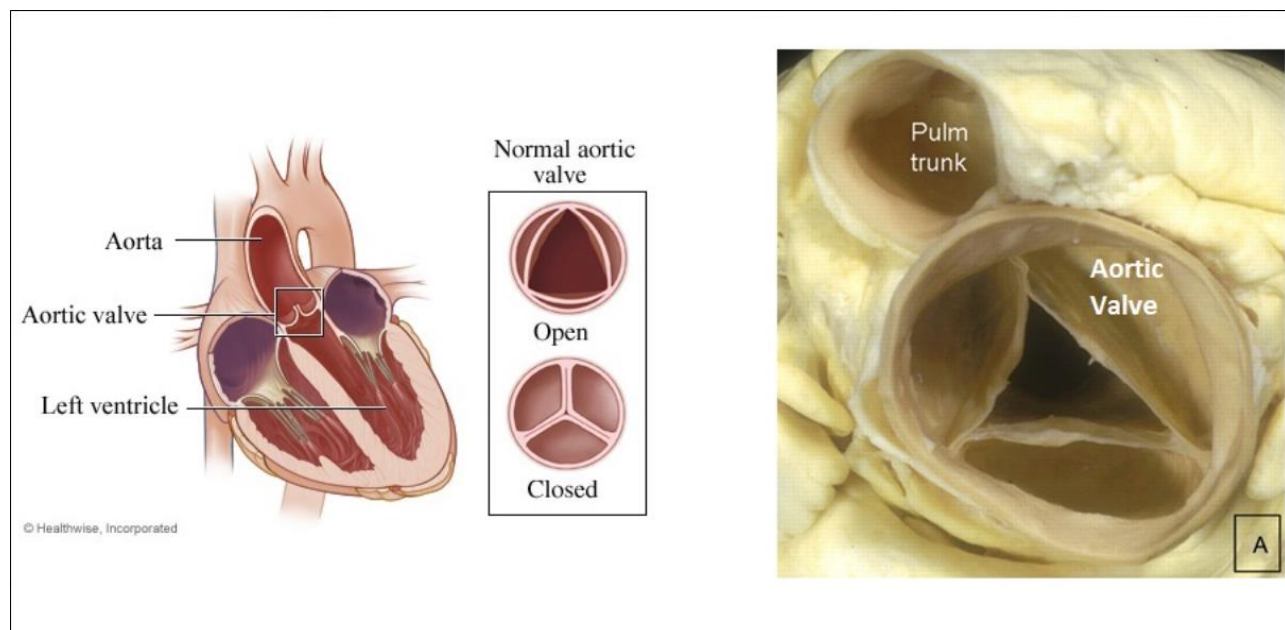
Οι θεραπευτικές επιλογές για την καρδιαγγειακή νόσο περιλαμβάνουν φαρμακευτική αγωγή, αλλαγές στον τρόπο ζωής, επεμβατικές διαδικασίες και χειρουργικές επεμβάσεις. Η φαρμακευτική αγωγή μπορεί να περιλαμβάνει αντιϋπερτασικά, στατίνες για τη μείωση της χοληστερόλης, αντιπηκτικά για την πρόληψη θρόμβων και άλλα φάρμακα που στοχεύουν στη βελτίωση της καρδιακής λειτουργίας. Οι αλλαγές στον τρόπο ζωής, όπως η υιοθέτηση υγιεινής διατροφής, η αύξηση της σωματικής δραστηριότητας, η διακοπή του καπνίσματος και η διαχείριση του στρες, είναι κρίσιμες για την πρόληψη και τη διαχείριση της ΚΝ. Οι επεμβατικές διαδικασίες, όπως η αγγειοπλαστική και η τοποθέτηση στεντ, χρησιμοποιούνται για την άμεση αποκατάσταση της ροής του αίματος στις στενωμένες αρτηρίες. Τέλος, οι χειρουργικές επεμβάσεις, όπως η αορτοστεφανιαία παράκαμψη (bypass), μπορεί να απαιτηθούν για την αντιμετώπιση σοβαρών περιπτώσεων στένωσης ή απόφραξης των αρτηριών.

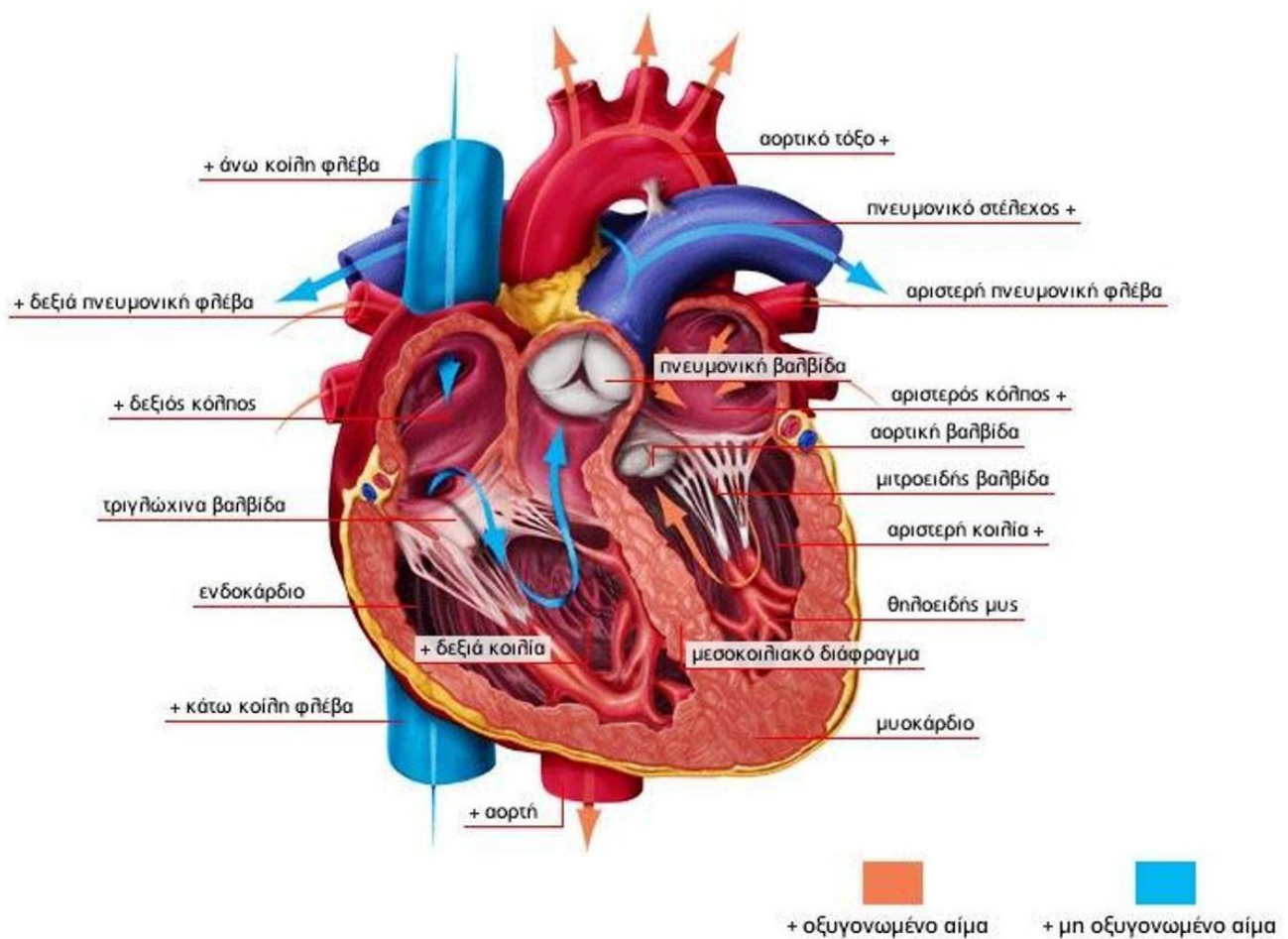
Παραδείγματα Πρόληψης και Επιδημιολογικών Συσχετίσεων

Η μελέτη « *Five Years incidence of cardiovascular disease and its predictors in Greece: The ATTICA study* » αναδεικνύει τη σημασία προληπτικών μεθόδων και πολιτικών που θα έπρεπε να ακολουθηθούν με σκοπό την πρόληψη των ΚΝ στην Ελλάδα ως κύρια αιτία θανάτου. Κύριοι παράγοντες σημειώνονται η υπέρταση, η χοληστερίνη, η παχυσαρκία, η καθιστική ζωή και η κακή διατροφή. Από επιδημιολογική άποψη αναδεικνύει τη σύνδεση των ΚΝ με την ηλικία, το φύλο και τις κοινωνικές συνθήκες. Οι άνδρες και οι ηλικιωμένοι φαίνεται να πλήττονται περισσότερο ενώ σημαντικό ρόλο παίζει και η κοινωνικοοικονομική κατάσταση τους (αγροτικές περιοχές ή και χαμηλότερο εισόδημα). Επιπρόσθετα η μεσογειακή διατροφή η οποία χαρακτηρίζεται από την κατανάλωση φρούτων, λαχανικών, ελαιόλαδου, ολικής αλέσεως προϊόντων, ψαριών και μικρότερης ποσότητας κρέατος, αναδεικνύεται ως μια από τις πιο αποτελεσματικές στρατηγικές πρόληψης των ΚΝ. Η μελέτη δείχνει ότι η υψηλή προσκόλληση στη μεσογειακή διατροφή σχετίζεται με σημαντική μείωση του κινδύνου εμφάνισης καρδιαγγειακών παθήσεων, λόγω των αντιοξειδωτικών ιδιοτήτων των τροφίμων αυτών και της ικανότητάς τους να βελτιώνουν τη χοληστερόλη και τη λειτουργία του καρδιαγγειακού συστήματος. Η υψηλή περιεκτικότητα της μεσογειακής διατροφής σε υγιεινά λιπαρά, όπως το ελαιόλαδο, καθώς και η αύξηση των φυτικών ινών από τα φρούτα και τα λαχανικά, συμβάλλουν στην καλύτερη υγεία της καρδιάς και στη μείωση της φλεγμονής. Πέρα από τη διατροφή, η φυσική δραστηριότητα παίζει επίσης κρίσιμο ρόλο στην πρόληψη των ΚΝ. Η μελέτη επισημαίνει ότι η τακτική άσκηση, ακόμα και ήπιας μορφής, όπως το περπάτημα και η ποδηλασία, βοηθά στη διαχείριση

του βάρους, τη βελτίωση της καρδιαγγειακής λειτουργίας και τη μείωση των επιπέδων της αρτηριακής πίεσης. Επιπλέον, η καταπολέμηση της παχυσαρκίας μέσω της υγιεινής διατροφής και της σωματικής άσκησης αποτελεί βασικό παράγοντα για την πρόληψη των ΚΝ.

Η μελέτη ATTICA (Panagiotakos et al., 2008) αναφέρεται σε μια εκτενή επιδημιολογική έρευνα που πραγματοποιήθηκε στην Ελλάδα για να εκτιμήσει τους παράγοντες κινδύνου που συνδέονται με τις καρδιαγγειακές παθήσεις, καθώς και τις σχέσεις μεταξύ διατροφικών συνηθειών, τρόπου ζωής και καρδιαγγειακής υγείας. Η μελέτη αυτή, που περιλαμβάνει δεδομένα από περισσότερους από 2.500 ενήλικες ηλικίας 18-89 ετών, είναι μία από τις πιο σημαντικές στην ελληνική βιβλιογραφία για τη μελέτη των ΚΝ και της συσχέτισής τους με την τοπική διατροφή και τις συνήθειες. Επιβεβαιώνεται εδώ ότι η συχνότητα εμφάνισης παραγόντων κινδύνου για τις καρδιαγγειακές παθήσεις όπως η χοληστερόλη, η παχυσαρκία, το κάπνισμα, η καθιστική ζωή και η διατροφή, σε ένα αντιπροσωπευτικό δείγμα του πληθυσμού έχουν άμεση συσχέτιση. Τα αποτελέσματα παρέχουν μια ισχυρή βάση για τη δημόσια υγεία στην Ελλάδα, εστιάζοντας στη σημασία της εφαρμογής των υγιεινών διατροφικών στρατηγικών και της ενίσχυσης της φυσικής δραστηριότητας για τη μείωση των καρδιαγγειακών κινδύνων.





Σκοπός και Σημασία της Μελέτης

Ο κύριος σκοπός της μελέτης αυτής είναι να συγκρίνει την επιβίωση των ασθενών με καρδιαγγειακή νόσο στην Κέρκυρα με την επιβίωση του γενικού πληθυσμού, χρησιμοποιώντας παραμετρικά μοντέλα επιβίωσης. Η μελέτη αυτή έχει σημαντική αξία για διάφορους λόγους. Πρώτον, μπορεί να προσφέρει πολύτιμα δεδομένα για την κατανόηση της επίδρασης της ΚΝ στην επιβίωση των ασθενών, επιτρέποντας τη βελτίωση των θεραπευτικών προσεγγίσεων και την ανάπτυξη στοχευμένων παρεμβάσεων. Δεύτερον, τα ευρήματα της μελέτης μπορούν να συμβάλουν στη χάραξη πολιτικών υγείας που στοχεύουν στη μείωση της θνησιμότητας από ΚΝ και στη βελτίωση της καρδιαγγειακής υγείας του πληθυσμού. Τέλος, η χρήση παραμετρικών μοντέλων επιβίωσης μπορεί να προσφέρει ακριβέστερες και πιο αξιόπιστες προβλέψεις, ενισχύοντας την κατανόηση των παραγόντων που επηρεάζουν την επιβίωση και την αποτελεσματικότητα των θεραπευτικών παρεμβάσεων.

2. Μεθοδολογία

2.1. Περιγραφή της μελέτης

Οι μεταβλητές που θα χρησιμοποιήσουμε στην παρούσα μελέτη, αποτελούν ένα μέρος των μεταβλητών που χρησιμοποιήθηκαν στην έρευνα “*Forty-years (1961–2001) of all-cause and coronary heart disease mortality and its determinants: the Corfu cohort from the Seven Countries Study*” που πραγματοποίησε ο κύριος Δημοσθένης Β. Παναγιωτάκος (Καθηγητής Βιοστατιστικής, Μεθόδων Έρευνας και Επιδημιολογίας στο Χαροκόπειο Πανεπιστήμιο Αθηνών, Ελλάδα) και δημοσιεύτηκε το 2002 (Panagiotakos *et al.*, 2003). Η συγκεκριμένη μελέτη εξέτασε την επίδραση διαφόρων παραγόντων κινδύνου στη θνησιμότητα κατά τη διάρκεια μιας περιόδου 40 ετών. Η μελέτη επικεντρώθηκε σε 529 αγρότες μεσήλικες άνδρες από την Κέρκυρα της Ελλάδας, οι οποίοι εγγράφηκαν το 1961 στο πλαίσιο της ευρύτερης μελέτης των επτά χωρών, η οποία διερευνούσε την καρδιαγγειακή νόσο σε διάφορους πληθυσμούς. Στην αρχή της μελέτης συλλέχθηκαν βασικές μετρήσεις, συμπεριλαμβανομένων βασικών παραγόντων καρδιαγγειακού κινδύνου, όπως η ηλικία, η αρτηριακή πίεση (τόσο η συστολική όσο και η διαστολική), τα επίπεδα ολικής χοληστερόλης στον ορό, η κατάσταση καπνίσματος, τα επίπεδα σωματικής δραστηριότητας και ο δείκτης μάζας σώματος. Οι μεταβλητές αυτές καταγράφηκαν προσεκτικά για να εκτιμηθεί η πιθανή επίδρασή τους στη μακροπρόθεσμη θνησιμότητα.

2.2. Περιγραφή του δείγματος

Στη μελέτη των 7 χωρών συμμετείχαν συνολικά 12.763 άνδρες ηλικίας 40-59 ετών από 16 μελέτες κοόρτης σε 7 χώρες: ΗΠΑ, Ιταλία, Ιαπωνία, Ολλανδία, Ελλάδα, πρώην Γιουγκοσλαβία και Φινλανδία. Μεταξύ αυτών, συμμετείχαν 1.215 αγροτικοί άνδρες από την Ελλάδα, εκ των οποίων 529 από Κέρκυρα και 686 από Κρήτη. Οι 529 άνδρες συλλέχθηκαν από έξι χωριά της Κέρκυρας (Άγιος Μάρκος, Σκριπερό, Σοκράκι, Δουκάδες, Γαρδελάδες και Κορακιάνα). Οι περισσότεροι ήταν μικρής κλίμακας, μη μηχανοποιημένοι αγρότες (55%), ενώ άλλοι εργάζονταν ως χειριστές τροφίμων (10%), ψαράδες (6%), μηχανικοί ή εργάτες εργοστασίων (12%), υπάλληλοι (4,9%) ή άνεργοι ή συνταξιούχοι (6,8%). Μετά από 40 έτη, τα επιζώντα μέλη της κοόρτης εξετάστηκαν εκ νέου από ειδικούς του Τμήματος Καρδιολογίας της Ιατρικής Σχολής του Πανεπιστημίου Αθηνών, σύμφωνα με το πρωτόκολλο της μελέτης των επτά χωρών (Panagiotakos *et al.*, 2003). Τα δεδομένα θνησιμότητας συλλέχθηκαν επί τέσσερις δεκαετίες μέσω επισκέψεων σε χωριά για την επιβεβαίωση των θανάτων. Οι αιτίες θανάτου προσδιορίστηκαν από κλινικά αρχεία, νοσοκομειακές εκθέσεις, ευρήματα νεκροψίας (εφόσον υπήρχαν) και πληροφορίες από γιατρούς, συγγενείς και άλλους μάρτυρες. Σε

περιπτώσεις αιφνίδιων ή βίαιων θανάτων που συνέβησαν δημόσια, ζητήθηκαν επίσης πληροφορίες από την αστυνομία.

2.3. Στατιστική ανάλυση

Οι ποσοτικές μεταβλητές (ηλικία, συστολική αρτηριακή πίεση, διαστολική αρτηριακή πίεση, καρδιακός ρυθμός, ολική χοληστερόλη, δείκτης μάζας σώματος) παρουσιάζονται ως μέση τιμή, τυπική απόκλιση και διάμεσος, ενδοτεταρτημοριακό εύρος. Οι κατηγορικές μεταβλητές (θνησιμότητα από όλες τις αιτίες, θνησιμότητα από καρδιαγγειακά νοσήματα (CVD), σωματική δραστηριότητα, κάπνισμα, κατηγοριοποίηση βάσει του δείκτη μάζας σώματος) παρουσιάζονται ως απόλυτη και σχετική συχνότητα. Για τη σύγκριση χαρακτηριστικών μεταξύ θανόντων έναντι μη θανόντων από CVD διεξήχθη ο έλεγχος Student's t-test αν το χαρακτηριστικό ήταν ποσοτική μεταβλητή (όπου ίσχυε η κανονικότητα στις ομάδες, ενώ για τις μεταβλητές που δεν ίσχυε, έγινε ο έλεγχος Mann-Whitney test) και ο έλεγχος χ^2 αν το χαρακτηριστικό ήταν κατηγορική μεταβλητή. Η κανονικότητα ελέγχθηκε γραφικά μέσω των ιστογραμμάτων και των θηκογραμμάτων. Στη συνέχεια σχεδιάστηκαν καμπύλες Kaplan – Meier και διεξήχθησαν έλεγχοι log-rank test για την σύγκριση της ανάλυσης επιβίωσης από CVD ανά σωματική δραστηριότητα, κάπνισμα και κατηγοριοποίηση βάσει του δείκτη μάζας σώματος. Έπειτα σχεδιάστηκαν μονοπαραγοντικά και πολυπαραγοντικά μοντέλα ανάλυσης επιβίωσης (συγκεκριμένα το μοντέλο αναλογικών κινδύνων του Cox) για την εκτίμηση του σχετικού κινδύνου θανάτου από CVD. Επιπλέον, σχεδιάστηκαν μονοπαραγοντικά και πολυπαραγοντικά μοντέλα για την σχετική πιθανότητα επιβίωσης (μη θάνατος από CVD) και συνολικής επιβίωσης (μη θάνατος από οτιδήποτε) και συγκεκριμένα το μοντέλο αναλογικών κινδύνων του Cox και το μοντέλο επιταχυνόμενου χρόνου ζωής (αστοχίας) – AFT μοντέλο, με 4 διαφορετικές κατανομές (Exponential, Weibull, Loglogistic και Lognormal). Η προϋπόθεση της αναλογικότητας των κινδύνων στο μοντέλο αναλογικών κινδύνων του Cox ελέγχθηκε μέσω των καταλοίπων Schoenfeld. Η καλή εφαρμογή και σύγκριση των μοντέλων ελέγχθηκε μέσω διαγραμμάτων των καταλοίπων Cox-Snell (άξονας X) έναντι του εκτιμώμενου αθροιστικού κινδύνου αυτών (Cumulative Hazard of Cox-Snell residuals - άξονας Y) και μέσω του Akaike κριτηρίου πληροφορίας (AIC). Για όλες τις αναλύσεις χρησιμοποιήθηκε το στατιστικό πρόγραμμα SPSS, εκτός από τον σχεδιασμό των AFT μοντέλων, όπου χρησιμοποιήθηκε το στατιστικό πρόγραμμα R.

2.4. Μοντέλο αναλογικών κινδύνων του Cox

Σε τέτοιου είδους μελέτες υπάρχει η ανάγκη να βρούμε την σχέση μεταξύ της μεταβλητής που δηλώνει τον χρόνο ζωής και μιας ή περισσότερων συμμεταβλητών το οποίο το καταφέρνουμε συνήθως μέσω ενός μοντέλου παλινδρόμησης. Οι λόγοι δημιουργίας ενός

τέτοιου μοντέλου είναι ο προσδιορισμός των μεταβλητών που επηρεάζουν περισσότερο την συνάρτηση κινδύνου $h(t)$, η εκτίμηση της ίδιας και κατά συνέπεια της συνάρτησης επιβίωσης. Στην ανάλυση επιβίωσης, η αναλογική παλινδρόμηση (PH) ή μοντέλο αναλογικού κινδύνου (Proportional hazard model) είναι το βασικότερο μοντέλο για τη μοντελοποίηση της συσχέτισης μεταξύ του χρόνου μέχρι το συμβάν με τις συνδιακυμάνσεις ή τις ανεξάρτητες μεταβλητές το οποίο αποτελεί μια μορφή γενικευμένων γραμμικών μοντέλων. Χρησιμοποιείται εκτενώς σήμερα στην ανάλυση λογοκριμένων δεδομένων επιβίωσης που αφορούν βιοϊατρικές εφαρμογές, για την εξακρίβωση των διαφορών στην επιβίωση που οφείλονται στο είδος της θεραπείας και σε προγνωστικούς παράγοντες σε κλινικές δοκιμές.

Συνήθως, στη πράξη εκτός από τη μεταβλητή T που συμβολίζει τον χρόνο επιβίωσης και τη μεταβλητή Δ που συμβολίζει εάν ο χρόνος ζωής είναι πλήρης ή λογοκριμένος, έχουμε και άλλες μεταβλητές $Z_j, j=1,2,\dots,p$, οι οποίες μπορεί να επηρεάζουν τον χρόνο επιβίωσης των ατόμων. Ας υποθέσουμε ότι κάθε άτομο ανήκει σε μια από δυο κατηγορίες.

Για παράδειγμα, μπορεί να έχουμε άνδρες και γυναίκες, καπνιστές και μη καπνιστές, κλπ.

Στο μοντέλο αναλογικού κινδύνου δεχόμαστε ότι ισχύει η σχέση:

$$h_1(t) = c h_2(t) \text{ ή } \frac{h_1(t)}{h_2(t)} = c, \text{ για κάθε } t \geq 0$$

που $h_i(t), i=1,2$ είναι η συνάρτηση κινδύνου της i ομάδας και c είναι ένας θετικός αριθμός που δεν εξαρτάται από το χρόνο t . Το μοντέλο αναλογικού κινδύνου αναφέρει ότι η συνάρτηση κινδύνου της ομάδας 1 είναι c -πλάσια της συνάρτησης κινδύνου της ομάδας 2. Εάν συμβολίζουμε με β τον λογάριθμο του c $\log(c)$ τότε η παραπάνω σχέση παίρνει τη μορφή:

$$h_1(t) = e^\beta h_2(t), \quad \text{για κάθε } t \geq 0$$

Το μοντέλο αναλογικού κινδύνου στην περίπτωση που υπάρχει μια μόνο ανεξάρτητη μεταβλητή δίνεται από τη σχέση:

$$h(t|Z) = h_0(t)e^{\beta z}, \quad \text{για κάθε } t \geq 0,$$

όπου η $h_0(t)$ ονομάζεται συνάρτηση αναφοράς κινδύνου. Η συνάρτηση κινδύνου αναφοράς είναι η συνάρτηση ενός ατόμου όταν η ανεξάρτητη μεταβλητή ισούται με 0. Παρατηρούμε ότι η αναφορική συνάρτηση κινδύνου εξαρτάται μόνο από τον χρόνο και όχι από την ανεξάρτητη μεταβλητή. Αντίθετα, ο όρος $e^{\beta z}$ εξαρτάται μόνο από την ανεξάρτητη μεταβλητή και όχι από τον χρόνο. Όταν η μεταβλητή Z παίρνει την τιμή 1, τότε ισχύει:

$$h(t|Z = 1) = h_0(t)e^{\beta \cdot 1} = h_0(t)e^\beta, \quad \text{για κάθε } t \geq 0,$$

ενώ όταν υπάρχει η τιμή 0 ισχύει:

$$h_0(t|Z = 0) = h_0(t)e^{\beta \cdot 0} = h_0(t), \quad \text{για κάθε } t \geq 0$$

Στην περίπτωση που υπάρχουν περισσότερες από μια ανεξάρτητες μεταβλητές, το μοντέλο αναλογικού κινδύνου γράφεται ως:

$$h(t|Z) = h_0(t)e^{\beta Z} = h_0(t)e^{\sum_{i=1}^p \beta_i z_i}, \quad \text{για κάθε } t \geq 0$$

όπου $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ είναι το διάνυσμα των p παραμέτρων του μοντέλου και $Z = (Z_1, Z_2, \dots, Z_p)'$ είναι το διάνυσμα των p ανεξάρτητων μεταβλητών. Στην περίπτωση αυτή, η συνάρτηση κινδύνου αναφοράς $h_0(t)$ είναι η συνάρτηση κινδύνου ενός ατόμου όταν όλες οι ανεξάρτητες μεταβλητές ισούται με το 0.

Ας θεωρήσουμε τώρα δυο άτομα με τιμές των συμμεταβλητών Z_1 και Z_2 , αντίστοιχα. Ο σχετικός λόγος κινδύνου ισούται με

$$HR = \frac{h(t|Z_1)}{h(t|Z_2)} = \frac{h_0(t)e^{\beta Z_1}}{h_0(t)e^{\beta Z_2}} = \exp\{\beta(Z_1 - Z_2)\},$$

ο οποίος είναι σταθερός και δεν εξαρτάται από τον χρόνο t . Στην πράξη, ο σχετικός λόγος κινδύνου είναι ο κίνδυνος ένα άτομο με χαρακτηριστικά Z_1 να εμφανίσει το υπό μελέτη γεγονός σε σχέση με ένα άτομο με χαρακτηριστικά Z_2 . Από την παραπάνω σχέση, διαπιστώνουμε ότι ο κίνδυνος του ατόμου με χαρακτηριστικά Z_1 ισούται με $\exp\{\beta(Z_1 - Z_2)\}$ φορές τον κίνδυνο ενός ατόμου με χαρακτηριστικά Z_2 . Από την παραπάνω έκφραση του σχετικού λόγου κινδύνου, προκύπτει ότι ο λογάριθμος του ισούται με

$$\log(HR) = \log\left(\frac{h(t|Z_1)}{h(t|Z_2)}\right) = \sum_{i=1}^p \beta_i (Z_{1i} - Z_{2i}).$$

Παρατηρούμε δηλαδή ότι ο λογάριθμος του σχετικού λόγου κινδύνου συνδέεται με τις διαφορές $Z_{1i} - Z_{2i}$, $i=1,2$, με μια γραμμική σχέση.

Το παραπάνω μοντέλο ονομάζεται μοντέλο αναλογικού κινδύνου. Πρόκειται για ένα ημι-παραμετρικό μοντέλο καθώς γίνονται κάποιες υποθέσεις για τον όρο $e^{\beta Z}$ αλλά δεν προκαθορίζεται κάποια μορφή για την αναφορική συνάρτηση κινδύνου $h_0(t)$.

Στην περίπτωση του μοντέλου αναλογικού κινδύνου, η εκτίμηση των παραμέτρων γίνεται με τη μέθοδο μέγιστης πιθανοφάνειας. Διακρίνονται δυο περιπτώσεις. Η πρώτη περίπτωση είναι η απουσία δεσμών (ties) και η δεύτερη είναι η παρουσία δεσμών. Απουσία δεσμών έχουμε όταν σε κάθε χρονική στιγμή συμβαίνει μόνο ένα γεγονός. Αντίθετα, παρουσία δεσμών έχουμε

εάν σε κάποια χρονική στιγμή παρατηρούνται περισσότερα από ένα γεγονότα (Εργαζάκη, 2024).

2.5. Παραμετρικά Μοντέλα επιβίωσης

Έστω ότι $T_1, \dots, T_n; C_1, \dots, C_n$ είναι ανεξάρτητες τυχαίες μεταβλητές. C_i είναι ο χρόνος λογοκρισίας ο οποίος σχετίζεται με τον χρόνο επιβίωσης T_i .

Παρατηρούμε $(Y_1, \delta_1), \dots, (Y_n, \delta_n)$ όπου

$$Y_i = T_i \wedge C_i, \delta_i = I(T_i \leq C_i).$$

Είναι επίσης διαθέσιμες συμμεταβλητές $\underline{x}_1, \dots, \underline{x}_n$; όπου

$$\underline{x}_i = (x_{i1}, \dots, x_{ip}), \quad i = 1, \dots, n$$

οι οποίες σχετίζονται με την εξαρτημένη μεταβλητή T_i .

Η συνάρτηση κινδύνου ορίζεται ως

$$\lambda(t; \underline{x}) = \frac{f(t; \underline{x})}{1 - F(t; \underline{x})}.$$

Το αναλογικό μοντέλο συνάρτησης κινδύνου είναι

$$\lambda(t; \underline{x}) = e^{\underline{\beta}'\underline{x}} \lambda_0(t)$$

όπου

$$\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$$

είναι οι συντελεστές παλινδρόμησης.

Τα παραμετρικά μοντέλα επιβίωσης έχουν τα ίδια χαρακτηριστικά με το αναλογικό μοντέλο επιβίωσης του Cox με την διαφορά ότι η συνάρτηση κινδύνου βάσης εκτιμάται μαζί με τυχόν συντελεστές.

Στην παρούσα διπλωματική θα ασχοληθούμε με δυο τύπους παραμετρικών μοντέλων.

- Το παραμετρικό, αναλογικό μοντέλο επιβίωσης
- Το μοντέλο της επιταχυνόμενης (accelerated) συνάρτησης κινδύνου.

Το παραμετρικό μοντέλο επιβίωσης

Το μοντέλο του Cox (PH) περιγράφεται από τη σχέση

$$h(t; x_1, \dots, x_p) = h_0(t) \exp(\beta_1 x_1 + \dots + \beta_p x_p) \quad (1)$$

για το οποίο δεν γίνεται καμιά αναφορά για την κατανομή της τ.μ. T .

Ο λόγος

$$HR = \frac{h(t; x_{i1}, \dots, x_{ip})}{h(t; x_{j1}, \dots, x_{jp})}$$

για δυο μονάδες i και j λέγεται λόγος κινδύνου και είναι ανεξάρτητο το χρόνου.

Δηλαδή αν το μοντέλο προσαρμόζεται καλά, θα πρέπει οι συναρτήσεις κινδύνου για δυο μονάδες να είναι περίπου “παράλληλες” .

Ως εδώ δεν έχει γίνει κάποια υπόθεση για την κατανομή του μοντέλου. Αν θα θέλαμε να υιοθετήσουμε κάποια, θα έπρεπε να λάβουμε υπόψη την ιδιότητα (1) η οποία λέει ότι οι δυο συναρτήσεις κινδύνου h και h_0 είναι ανάλογες.

Έχουμε λοιπόν τον παρακάτω ορισμό.

Ορισμός 1: Αν δυο συναρτήσεις κινδύνου h_0 και h_1 έχουν την ιδιότητα ότι

$$h_1(t) = \psi h_0(t) \text{ για όλα τα } t > 0 \text{ και κάποιο } \psi > 0, \quad (2)$$

τότε λέμε ότι είναι ανάλογες.

Η παραπάνω σχέση μπορεί να γραφεί σε σχέση με τις συναρτήσεις επιβίωσης σαν

$$S_1(t) = \{S_0(t)\}^\psi \text{ για όλα τα } t > 0 \text{ και κάποιο } \psi > 0 \quad (3)$$

Οι κατανομές αυτές λέγονται Lehmann alternatives.

Προφανώς στην εφαρμογή του μοντέλου θέτουμε

$$\psi = \exp(\beta_1 x + \dots + \beta_p x_p).$$

Επομένως κατάλληλες κατανομές είναι αυτές με την ιδιότητα (2) ή (3).

Το PH μοντέλο εστιάζει στον ρυθμό κινδύνου και είναι κατάλληλο όταν υποθέτουμε ότι οι παράγοντες επηρεάζουν τον κίνδυνο συμβάντος τη συγκεκριμένη χρονική στιγμή.

Κατανομές οι οποίες έχουν τις παραπάνω ιδιότητες είναι η Weibull, η κατά τμήματα σταθερή συνάρτηση κινδύνου (pch), η κατανομή Gompertz.

The accelerated failure time model (AFT)

Το AFT μοντέλο καλύτερα περιγράφεται με βάση τη σχέση δυο συναρτήσεων επιβίωσης όταν συγκρίνουμε δυο ομάδες.

- Ομάδα 0 : $P(T \geq t) = S_0(t)$ (ομάδα ελέγχου)
- Ομάδα 1: $P(T \geq t) = S(t) = S_0(\gamma t)$ (πειραματική ομάδα) (4)

Το μοντέλο λέει ότι η ασθένεια επιταχύνει τον χρόνο θανάτου με έναν παράγοντα γ . Αν $\gamma < 1$, τότε η θεραπεία είναι αποτελεσματική, στην αντίθετη περίπτωση είναι αναποτελεσματική. Μια άλλη ερμηνεία είναι ότι η διάμεση διάρκεια ζωής πολλαπλασιάζεται με έναν παράγοντα $1/\gamma$.

Η (4) είναι αυτή που ορίζει το (AFT) μοντέλο.

Ορισμός 2: Ένα μοντέλο επιβίωσης λέγεται *επιταχυνόμενο μοντέλο κινδύνου (AFT)* όταν υπάρχει παράμετρος γ έτσι ώστε η συνάρτηση κινδύνου να γράφεται:

$$h(t; \gamma) = \gamma h(\gamma t; 1) = \gamma h_0(\gamma t) \quad (5)$$

όπου γ είναι ο παράγοντας επιτάχυνσης.

Τα επόμενα είναι ισοδύναμα:

$$h(t; \gamma) = \gamma h(\gamma t; 1) = \gamma h_0(\gamma t)$$

$$S(t; \gamma) = S(\gamma t; 1) = S_0(\gamma t)$$

$$f(t; \gamma) = \gamma f(\gamma t; 1) = \gamma f_0(\gamma t)$$

Παρατηρούμε ότι αν η f έχει παράμετρο κλίμακας λ , δηλαδή

$$f(t; \lambda) = \frac{1}{\lambda} f\left(\frac{1}{\lambda} t; 1\right) \text{ και ισοδύναμα } S(t; \lambda) = S\left(\frac{1}{\lambda} t; 1\right),$$

τότε είναι ένα AFT μοντέλο με συντελεστή επιτάχυνσης $\gamma = \frac{1}{\lambda}$.

Το AFT μοντέλο εστιάζει στην επίδραση των εξηγητικών μεταβλητών στον χρόνο επιβίωσης και είναι κατάλληλο όταν υποθέτουμε ότι οι παράγοντες επιταχύνουν ή επιβραδύνουν τον χρόνο επιβίωσης.

Κατανομές που μπορεί να χρησιμεύσουν σαν AFT μοντέλο είναι η Weibull, η λογαριθμοκανονική (log-normal) , η λογαριθμολογιστική (log-logistic).

Η Weibull είναι η μόνη κατανομή η οποία ικανοποιεί τα κριτήρια και για PH και για AFT μοντέλο.

Για να δημιουργήσουμε το AFT μοντέλο, ξεκινάμε από το γεγονός ότι υπάρχει παράμετρος κλίμακας. Έτσι αν η τυχαία μεταβλητή έχει T συνάρτηση επιβίωσης $S(t)$, τότε η

$$T_\gamma = \frac{T}{\gamma} \quad (5)$$

έχει συνάρτηση επιβίωσης $S(ct)$.

Αν $Y = \log T$ και $Y_\gamma = \log(T_\gamma)$, τότε ισχύει η σχέση (5) η οποία γράφεται σαν

$$Y_\gamma = Y - \log \gamma.$$

Αν $Y = \varepsilon$,

$$Y_\gamma = Y$$

και

$$\log(\gamma) = -\underline{\beta}\underline{x} = -(\beta_1 x + \dots + \beta_p x_p) \quad (6)$$

Η παραπάνω σχέση μπορεί να γραφεί στην γνωστή μορφή

$$Y = \underline{\beta}\underline{x} + \varepsilon,$$

Δηλαδή έχουμε την περίπτωση ενός γραμμικού μοντέλου για τους λογαριθμικούς χρόνους επιβίωσης.

Από την (6) έχουμε ότι ο παράγοντας επιτάχυνσης είναι

$$\gamma = \exp\{-(\beta_1 x_1 + \dots + \beta_p x_p)\}.$$

Ο λόγος

$$\frac{\gamma_i}{\gamma_j}$$

ερμηνεύεται σαν ο παράγοντας επιτάχυνσης της μονάδας i σε σχέση με την μονάδα j .

Για δυο μονάδες οι οποίες διαφέρουν μόνο σε μια μεταβλητή $x_{ik} \neq x_{jk}$ τότε

$$\frac{\gamma_i}{\gamma_j} = e^{-\beta_k(x_{ik}-x_{jk})}$$

Από το οποίο συμπεραίνουμε ότι η τιμή $e^{-\beta_k}$ ερμηνεύεται σαν ο συντελεστής επιτάχυνσης μιας μονάδας όταν αυξάνεται η τιμή της x_k κατά μια μονάδα, όταν όλες οι άλλες μεταβλητές παραμένουν ίδιες.

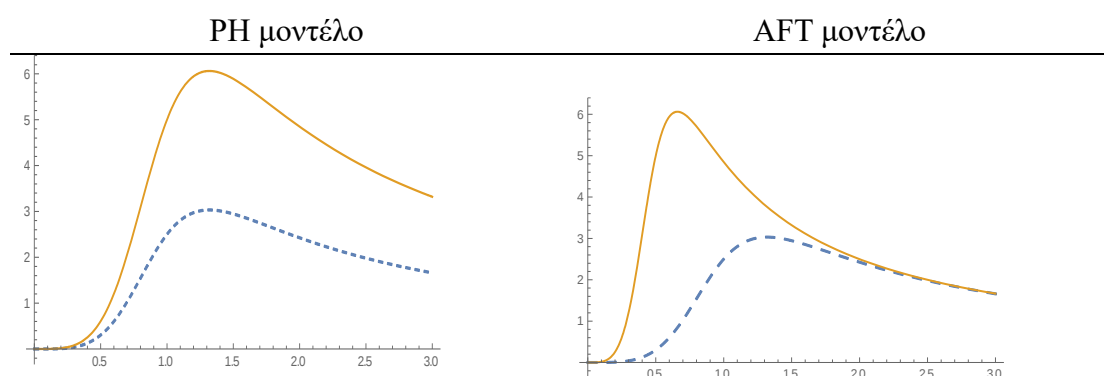
Η επιλογή μεταξύ των δυο αυτών μοντέλων γίνεται με την σύγκριση του AIC κριτηρίου. Επειδή ο αριθμός των παραμέτρων είναι ο ίδιος μπορούμε να συγκρίνουμε τις loglikelihood μέγιστες τιμές.

Για την καλύτερη κατανόηση των δυο εννοιών, κατασκευάσαμε με κατανομή βάσης την loglogistic κατανομή με π.π.

$$f(x; \alpha, \beta) = \frac{\frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1}}{\left(1 + \left(\frac{x}{\alpha}\right)^{\beta}\right)^2}$$

Και συνάρτηση κινδύνου

$$h(t; \alpha, \beta) = \frac{\frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1}}{1 + \left(\frac{t}{\alpha}\right)^{\beta}}$$



Το αριστερό σχήμα το PH μοντέλο. Το δεξί σχήμα το AFT μοντέλο. Η διακεκομμένη γραμμή είναι η h_0 ενώ πλήρης γραμμή η $h(t)$.

Παρατηρούμε ότι στο PH μοντέλο, η συνάρτηση κινδύνου πολλαπλασιάζεται με 2 ενώ στο AFT μοντέλο γίνεται και μια μετατόπιση προς τα αριστερά (Εργαζάκη, 2024).

3. Περιγραφική Ανάλυση

Στη συνέχεια παρουσιάζονται τα αποτελέσματα από τις στατιστικές αναλύσεις που διεξήχθησαν.

3.1.Περιγραφή των μεταβλητών

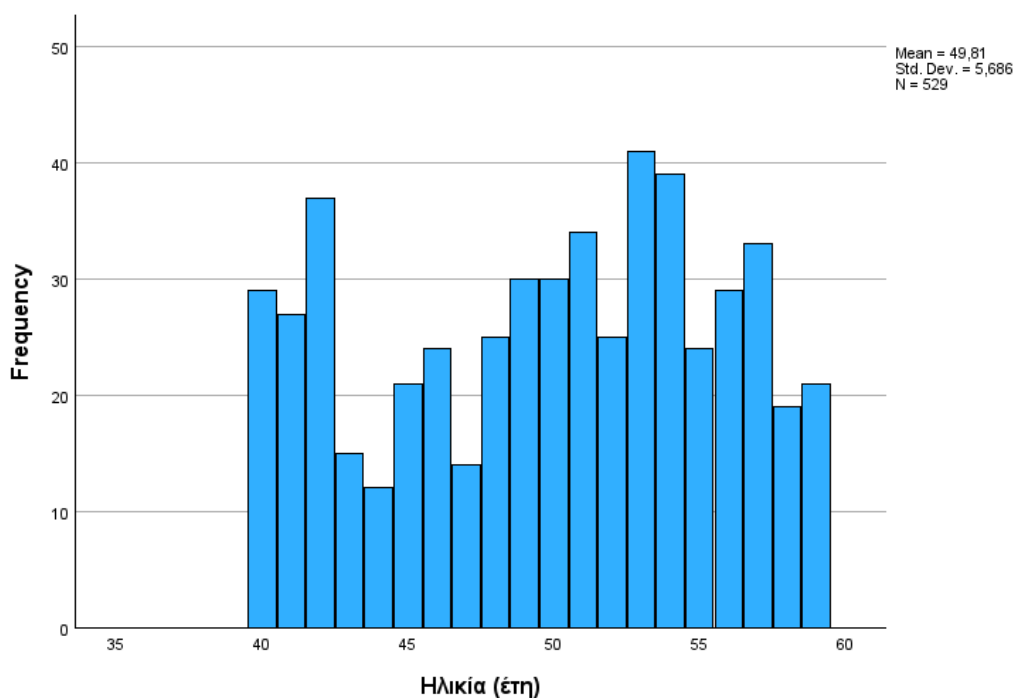
Ως προγνωστικοί παράγοντες χρησιμοποιήθηκαν η ηλικία, η αρτηριακή πίεση, ο καρδιακός ρυθμός, η ολική χοληστερόλη ορού, το κάπνισμα, η φυσική δραστηριότητα και ο δείκτης μάζας σώματος. Τα δεδομένα που θα χρησιμοποιήσουμε αποτελούνται από 529 παρατηρήσεις και 14 στήλες κάθε μία από τις οποίες είναι:

1. Serial : Σειριακός αριθμός
2. age entry : Ηλικία εισόδου στην έρευνα
3. death : Ημερομηνία θανάτου
4. fu (follow up) : Χρόνος παραμονής στην έρευνα
5. alle mortality : Θνησιμότητα από διάφορες αιτίες (συμπεριλαμβάνεται η θνησιμότητα από καρδιαγγειακά νοσήματα).
6. cvd mortality : Θνησιμότητα από καρδιαγγειακά νοσήματα (CVD)
7. physical act : Σωματική δραστηριότητα
8. height : Ύψος
9. weight : Κιλά
10. sbp : Συστολική αρτηριακή πίεση
11. dbp : Διαστολική αρτηριακή πίεση
12. heartrate : Καρδιακός ρυθμός
13. tchol : Ολική χοληστερόλη
14. smoking : Κάπνισμα
15. bmi: Δείκτης μάζας σώματος
16. bmi_group: Κατηγοριοποίηση του δείκτη μάζας σώματος ($\leq 25 \text{ kg/m}^2$ και $> 25 \text{ kg/m}^2$)

3.2.Περιγραφική Ανάλυση Δείγματος

Συνολικά συμπεριλήφθηκαν στη μελέτη 529 συμμετέχοντες ηλικίας 40 έως 59 ετών. Στον Πίνακα 1 περιγράφεται η ηλικία των συμμετεχόντων, ενώ στο Γράφημα 1 το ιστόγραμμα συχνοτήτων της μεταβλητής της ηλικίας. Η μέση τιμή της ηλικίας των συμμετεχόντων κατά την εισαγωγή τους στη μελέτη ήταν $49,81 \pm 5,69$ έτη. Η διάμεσος ήταν 51 έτη και το ενδοτεταρτημοριακό εύρος 45 – 54 έτη.

Πίνακας 1. Περιγραφικά μέτρα μεταβλητής ηλικία (N=529).	
Μέση τιμή	49,81
Τυπική απόκλιση	5,69
Διάμεσος	51
Ενδοτεταρτημοριακό εύρος	45 – 54
Εύρος	40 – 59



Γράφημα 1. Ιστόγραμμα συχνοτήτων μεταβλητής ηλικία.

Στον Πίνακα 2 περιγράφεται η θνησιμότητα από όλες τις αιτίες, ενώ στο Γράφημα 2 το διάγραμμα πίτας της μεταβλητής θνησιμότητα από όλες τις αιτίες. Από το σύνολο των συμμετεχόντων, οι 461 κατέληξαν (87,15% θνησιμότητα), ενώ οι 68 (12,85%) ζουν.

Πίνακας 2. Θνησιμότητα από όλες τις αιτίες (N=529).	
Ζει	68 (12,85%)
Κατέληξε	461 (87,15%)



Γράφημα 2. Διάγραμμα πίτας μεταβλητής θνησιμότητα από όλες τις αιτίες.

Στον Πίνακα 3 περιγράφεται η θνησιμότητα από CVD, ενώ στο Γράφημα 3 το διάγραμμα πίτας της μεταβλητής θνησιμότητα από CVD. Από το σύνολο των συμμετεχόντων, οι 337 κατέληξαν από CVD (63,71% θνησιμότητα από CVD), ενώ οι 192 (36,29%) είτε ζουν είτε κατέληξαν από κάποιο άλλο πρόβλημα υγείας.

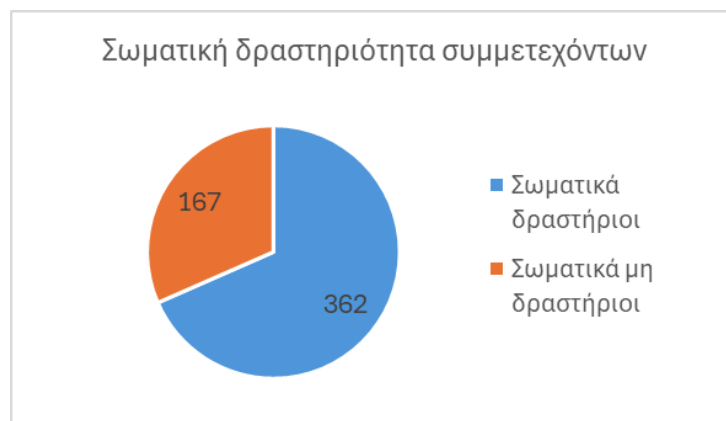
Πίνακας 3. Θνησιμότητα από CVD (N=529).	
Όχι	337 (63,71%)
Ναι	192 (36,29%)



Γράφημα 3. Διάγραμμα πίτας μεταβλητής θνησιμότητα από CVD.

Στον *Πίνακα 4* περιγράφεται η σωματική δραστηριότητα των συμμετεχόντων, ενώ στο *Γράφημα 4* το διάγραμμα πίτας της μεταβλητής σωματική δραστηριότητα. Από το σύνολο των συμμετεχόντων, οι 362 (68,43%) ήταν σωματικά δραστήριοι, ενώ οι 167 (31,57%) δεν ήταν σωματικά δραστήριοι.

Πίνακας 4. Σωματική δραστηριότητα συμμετεχόντων (N=529).	
Σωματικά δραστήριοι	362 (68,43%)
Σωματικά μη δραστήριοι	167 (31,57%)

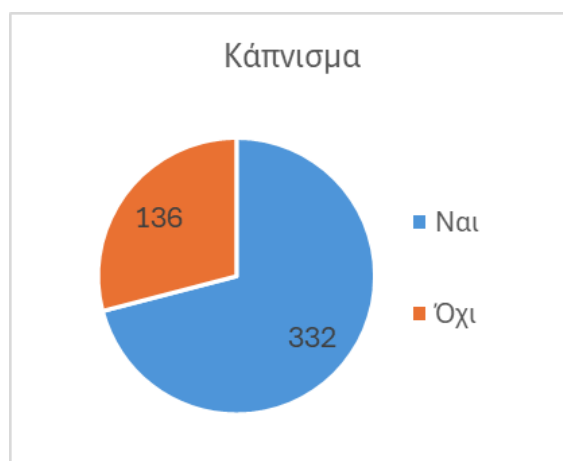


Γράφημα 4. Διάγραμμα πίτας μεταβλητής σωματικής δραστηριότητα συμμετεχόντων.

Στον *Πίνακα 5* περιγράφεται το κάπνισμα των συμμετεχόντων, ενώ στο *Γράφημα 5* το διάγραμμα πίτας της μεταβλητής κάπνισμα. Στη συγκεκριμένη μεταβλητή υπήρχαν 61 ελλείπουσες τιμές. Από το

σύνολο των συμμετεχόντων, οι 332 (70,94%) ήταν σωματικά δραστήριοι, ενώ οι 136 (29,06%) δεν ήταν σωματικά δραστήριοι.

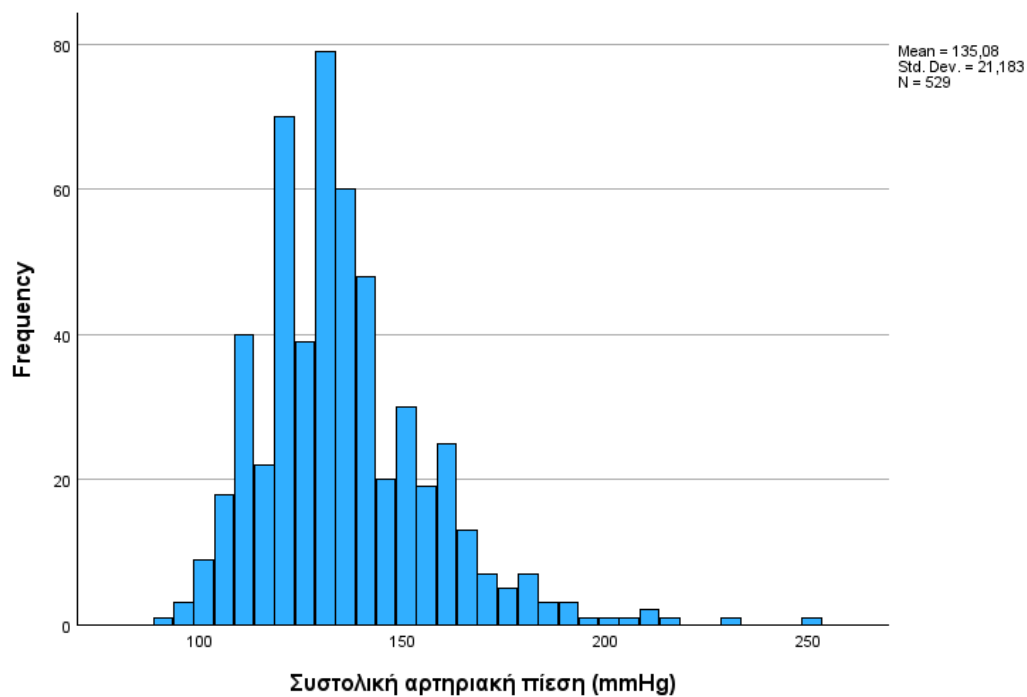
Πίνακας 5. Κάπνισμα συμμετεχόντων (N=468).	
Ναι	332 (70,94%)
Όχι	136 (29,06%)



Γράφημα 5. Διάγραμμα πίτας μεταβλητής καπνίσματος.

Στον Πίνακα 6 περιγράφεται η συστολική αρτηριακή πίεση των συμμετεχόντων, ενώ στο Γράφημα 6 το ιστόγραμμα συχνοτήτων της μεταβλητής της συστολική αρτηριακή πίεση. Η μέση τιμή της συστολικής αρτηριακής πίεσης των συμμετεχόντων κατά την εισαγωγή τους στη μελέτη ήταν $135,1 \pm 21,2$ mmHg. Η διάμεσος ήταν 131 mmHg και το ενδοτεταρτημοριακό εύρος 120 – 146 mmHg, ενώ το συνολικό εύρος 91 – 250 mmHg.

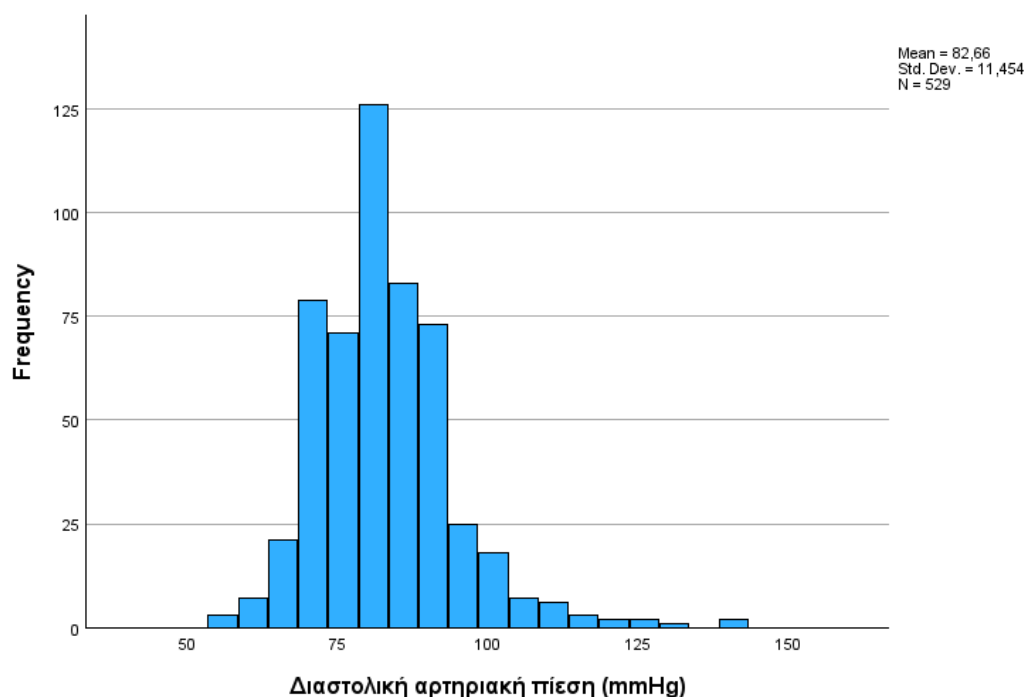
Πίνακας 6. Περιγραφικά μέτρα μεταβλητής συστολική αρτηριακή πίεση (N=529).	
Μέση τιμή	135,1
Τυπική απόκλιση	21,2
Διάμεσος	131
Ενδοτεταρτημοριακό εύρος	120 – 146
Εύρος	91 – 250



Γράφημα 6. Ιστόγραμμα συχνοτήτων μεταβλητής συστολική αρτηριακή πίεση.

Στον Πίνακα 7 περιγράφεται η διαστολική αρτηριακή πίεση των συμμετεχόντων, ενώ στο Γράφημα 7 το ιστόγραμμα συχνοτήτων της μεταβλητής της διαστολική αρτηριακή πίεση. Η μέση τιμή της διαστολικής αρτηριακής πίεσης των συμμετεχόντων κατά την εισαγωγή τους στη μελέτη ήταν $82,66 \pm 11,45$ mmHg. Η διάμεσος ήταν 81 mmHg και το ενδοτεταρτημοριακό εύρος 75 – 89 mmHg, ενώ το συνολικό εύρος 56 – 143 mmHg.

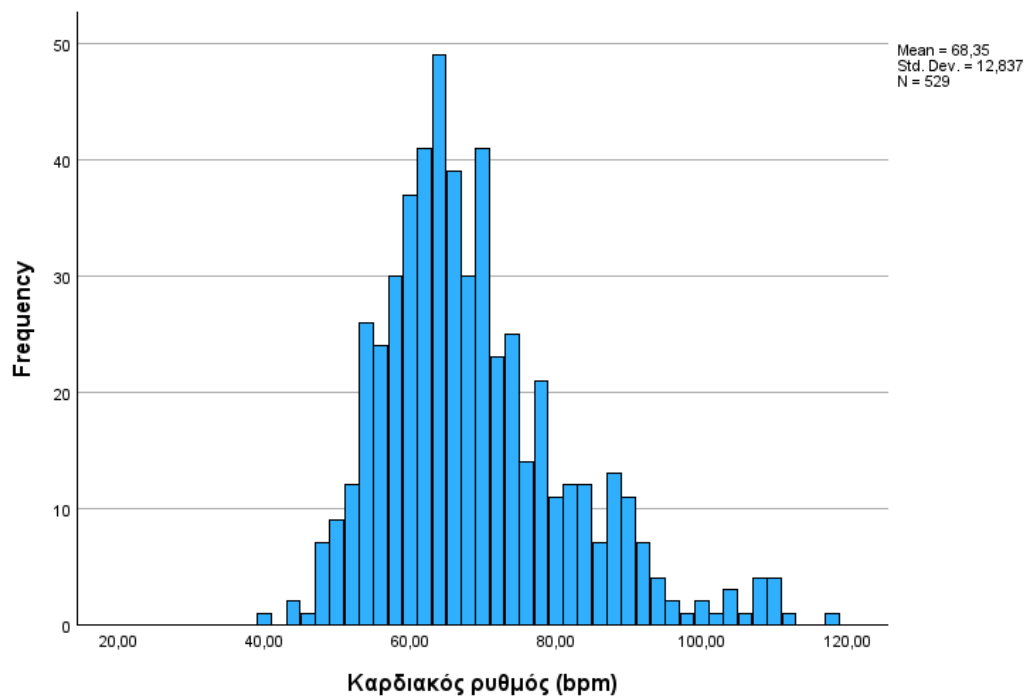
Πίνακας 7. Περιγραφικά μέτρα μεταβλητής διαστολική αρτηριακή πίεση (N=529).	
Μέση τιμή	82,66
Τυπική απόκλιση	11,45
Διάμεσος	81
Ενδοτεταρτημοριακό εύρος	75 – 89
Εύρος	56 – 143



Γράφημα 7. Ιστόγραμμα συχνοτήτων μεταβλητής διαστολική αρτηριακή πίεση.

Στον Πίνακα 8 περιγράφεται ο καρδιακός ρυθμός των συμμετεχόντων, ενώ στο Γράφημα 8 το ιστόγραμμα συχνοτήτων της μεταβλητής καρδιακός ρυθμός. Η μέση τιμή του καρδιακού ρυθμού των συμμετεχόντων κατά την εισαγωγή τους στη μελέτη ήταν $68,35 \pm 12,837$ bpm. Η διάμεσος ήταν 65 bpm και το ενδοτεταρτημοριακό εύρος 60 – 74 bpm, ενώ το συνολικό εύρος 40 – 118 bpm.

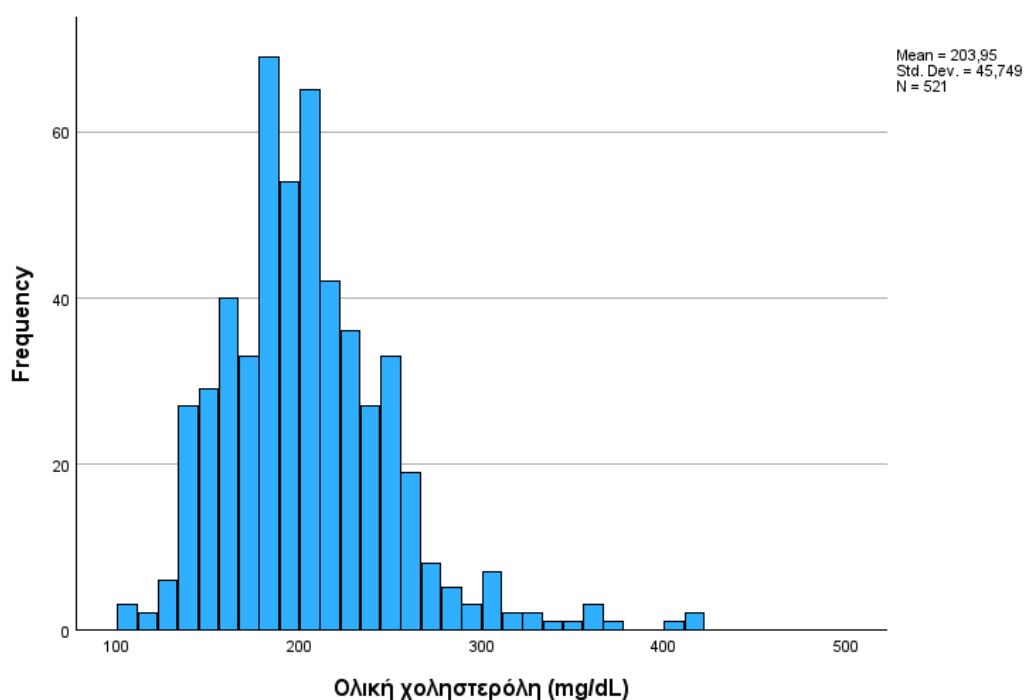
Πίνακας 8. Περιγραφικά μέτρα μεταβλητής καρδιακός ρυθμός (N=529).	
Μέση τιμή	68,35
Τυπική απόκλιση	12,837
Διάμεσος	65
Ενδοτεταρτημοριακό εύρος	60 – 74
Εύρος	40 – 118



Γράφημα 8. Ιστόγραμμα συχνοτήτων μεταβλητής καρδιακός ρυθμός.

Στον Πίνακα 9 περιγράφεται η ολική χοληστερόλη των συμμετεχόντων, ενώ στο Γράφημα 9 το ιστόγραμμα συχνοτήτων της μεταβλητής ολική χοληστερόλη. Η μέση τιμή της ολικής χοληστερόλης των συμμετεχόντων κατά την εισαγωγή τους στη μελέτη ήταν $204 \pm 45,8$ mg/dL. Η διάμεσος ήταν 199 mg/dL και το ενδοτεταρτημοριακό εύρος 174 – 228 mg/dL, ενώ το συνολικό εύρος 101 – 422 mg/dL. Παρατηρήθηκαν 8 ελλείπουσες τιμές για τη συγκεκριμένη μεταβλητή.

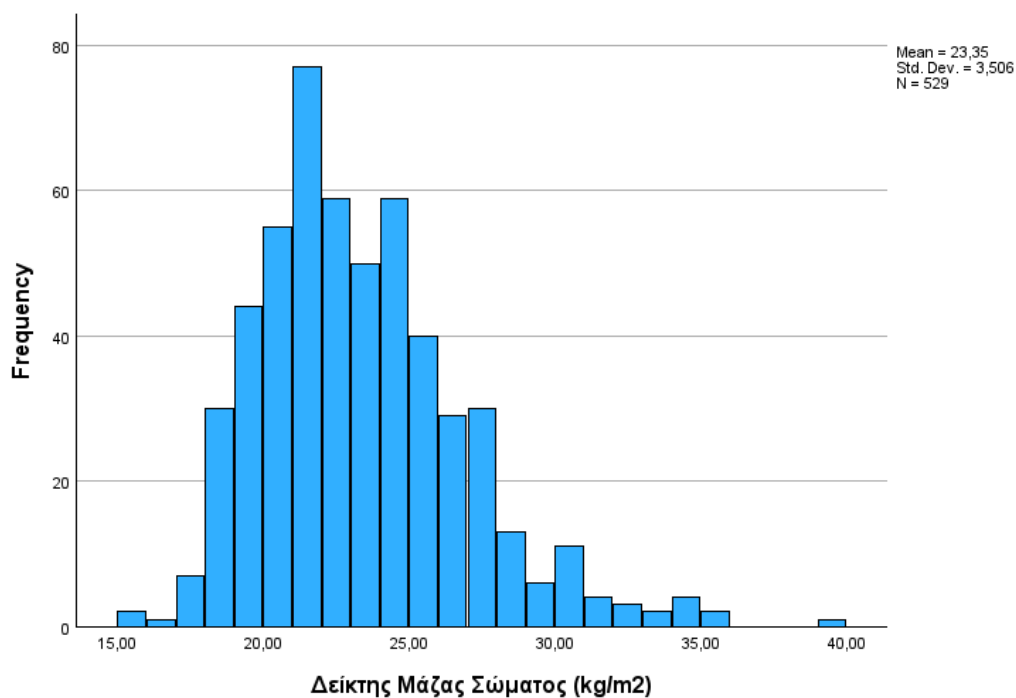
Πίνακας 9. Περιγραφικά μέτρα μεταβλητής ολική χοληστερόλη (N=521).	
Μέση τιμή	204
Τυπική απόκλιση	45,8
Διάμεσος	199
Ενδοτεταρτημοριακό εύρος	174 – 228
Εύρος	101 – 422



Γράφημα 9. Ιστόγραμμα συχνοτήτων μεταβλητής ολική χοληστερόλη.

Στον Πίνακα 10 περιγράφεται ο δείκτης μάζας σώματος των συμμετεχόντων, ενώ στο Γράφημα 10 το ιστόγραμμα συχνοτήτων της μεταβλητής δείκτης μάζας σώματος. Η μέση τιμή του δείκτη μάζας σώματος των συμμετεχόντων κατά την εισαγωγή τους στη μελέτη ήταν $23,4 \pm 3,5 \text{ kg/m}^2$. Η διάμεσος ήταν $22,8 \text{ kg/m}^2$ και το ενδοτεταρτημοριακό εύρος $20,8 - 25,3 \text{ kg/m}^2$, ενώ το συνολικό εύρος $15,8 - 39,1 \text{ kg/m}^2$.

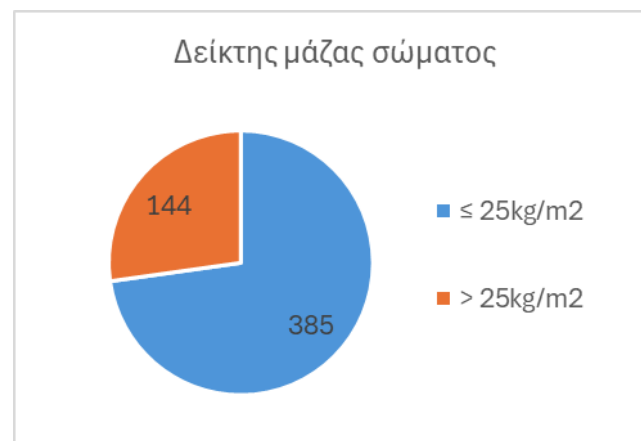
Πίνακας 10. Περιγραφικά μέτρα μεταβλητής δείκτης μάζας σώματος (N=529).	
Μέση τιμή	23,4
Τυπική απόκλιση	3,5
Διάμεσος	22,8
Ενδοτεταρτημοριακό εύρος	20,8 – 25,3
Εύρος	15,8 – 39,1



Γράφημα 10. Ιστόγραμμα συχνοτήτων μεταβλητής δείκτης μάζας σώματος.

Στον Πίνακα 11 περιγράφεται η κατηγοριοποίηση του δείκτη μάζας σώματος, ενώ στο Γράφημα 11 το διάγραμμα πίτας της μεταβλητής κατηγοριοποίηση του δείκτη μάζας σώματος. Από το σύνολο των συμμετεχόντων, οι 385 (72,78%) είχαν δείκτη μάζας σώματος $\leq 25\text{kg/m}^2$, ενώ οι 144 (27,22%) είχαν δείκτη μάζας σώματος $> 25\text{kg/m}^2$.

Πίνακας 11. Κατηγοριοποίηση δείκτη μάζας σώματος (N=529).	
$\leq 25\text{kg/m}^2$	385 (72,78%)
$> 25\text{kg/m}^2$	144 (27,22%)



Γράφημα 11. Διάγραμμα πίτας μεταβλητής δείκτη μάζας σώματος.

3.3.Συσχέτιση θνησιμότητας από CVD με τα χαρακτηριστικά των ασθενών

Στον Πίνακα 12 περιγράφεται η σύγκριση των χαρακτηριστικών μεταξύ θανόντων έναντι μη θανόντων από CVD. Για τις επιμέρους αναλύσεις έχουν χρησιμοποιηθεί οι έλεγχοι Student's t-test, Mann-Whitney test και χ^2 .

Ο έλεγχος **Student's t-test** είναι μια στατιστική μέθοδος που χρησιμοποιείται για να συγκρίνει τις μέσες τιμές δύο ομάδων και να καθορίσει αν οι διαφορές μεταξύ τους είναι στατιστικά σημαντικές. Αναπτύχθηκε από τον William Sealy Gosset, ο οποίος δημοσίευσε υπό το ψευδώνυμο "Student", και αποτελεί ένα από τα πιο διαδεδομένα εργαλεία για τον έλεγχο υποθέσεων στις επιστήμες. Ο t-test χρησιμοποιείται όταν θέλουμε να εξετάσουμε αν οι μέσες τιμές δύο ομάδων διαφέρουν σημαντικά, λαμβάνοντας υπόψη τη διακύμανση των δεδομένων και το μέγεθος του δείγματος (Field, 2018).

Ο τύπος του ελέγχου για δύο ανεξάρτητες ομάδες είναι:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

όπου:

- $\overline{X}_1$ και $\overline{X}_2$ είναι οι μέσες τιμές των δύο ομάδων,
- s_1^2 και s_2^2 είναι οι διακυμάνσεις των δύο ομάδων,
- n_1 και n_2 είναι το μέγεθος δείγματος των δύο ομάδων.

Ο έλεγχος έχει τις παρακάτω υποθέσεις:

- Μηδενική υπόθεση (H_0): η μέση τιμή της ποσοτικής μεταβλητής δεν διαφέρει μεταξύ των 2 ομάδων
- Εναλλακτική υπόθεση (H_1): η μέση τιμή της ποσοτικής μεταβλητής διαφέρει μεταξύ των 2 ομάδων

Η δοκιμασία **Mann-Whitney U test** (επίσης γνωστή ως Wilcoxon rank-sum test) είναι μια μη παραμετρική στατιστική δοκιμασία που χρησιμοποιείται για τη σύγκριση δύο ανεξάρτητων ομάδων όταν τα δεδομένα δεν πληρούν τις προϋποθέσεις που απαιτούνται για την ανεξάρτητη δοκιμασία του Student's t-test. Είναι ιδιαίτερα χρήσιμη για ταξινομημένα δεδομένα ή συνεχή δεδομένα που δεν κατανέμονται κανονικά. Το τεστ αξιολογεί αν η μία ομάδα τείνει να έχει υψηλότερες ή χαμηλότερες τιμές από την άλλη, χωρίς να κάνει υποθέσεις σχετικά με την υποκείμενη κατανομή των δεδομένων (Field, 2018).

Ο τύπος του ελέγχου είναι:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

όπου:

- U_1 και U_2 το στατιστικό κριτήριο U για τις 2 ομάδες,
- n_1 και n_2 ο αριθμός δείγματος για τις 2 ομάδες,
- R_1 και R_2 τα αθροίσματα των τάξεων (ranks) για τις δύο ομάδες

Ο έλεγχος χρησιμοποιείται για να προσδιοριστεί εάν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των κατανομών δύο ανεξάρτητων ομάδων (Field, 2018). Ο έλεγχος έχει τις παρακάτω υποθέσεις:

- Μηδενική υπόθεση (H_0): οι δύο ομάδες προέρχονται από τον ίδιο πληθυσμό
- Εναλλακτική υπόθεση (H_1): οι δύο ομάδες προέρχονται από διαφορετικούς πληθυσμούς

Ο έλεγχος χ^2 είναι μια στατιστική μέθοδος που χρησιμοποιείται για την εξέταση της συσχέτισης μεταξύ κατηγορικών μεταβλητών. Είναι μη παραμετρική δοκιμή, δηλαδή δεν προϋποθέτει συγκεκριμένη κατανομή των δεδομένων, γεγονός που την καθιστά ευέλικτη για την ανάλυση κατηγορικών δεδομένων (Field, 2018). Ο έλεγχος ανεξαρτησίας χ^2 χρησιμοποιείται για να προσδιοριστεί εάν υπάρχει σημαντική συσχέτιση μεταξύ δύο κατηγορικών μεταβλητών. Για παράδειγμα, μπορεί να χρησιμοποιηθεί για να ελεγχθεί αν υπάρχει σχέση μεταξύ του φύλου (άνδρας/γυναίκα) και της προτίμησης για ένα προϊόν (ναι/όχι). Ο έλεγχος συγκρίνει τις παρατηρούμενες συχνότητες σε έναν πίνακα ενδεχομένων με τις συχνότητες που θα αναμένονταν εάν δεν υπήρχε συσχέτιση μεταξύ των μεταβλητών.

Οι υποθέσεις του ελέγχου είναι οι εξής:

- Μηδενική υπόθεση (H_0): Οι δύο μεταβλητές είναι ανεξάρτητες (δεν υπάρχει συσχέτιση).
- Εναλλακτική υπόθεση (H_1): Οι δύο μεταβλητές είναι εξαρτημένες (υπάρχει συσχέτιση).

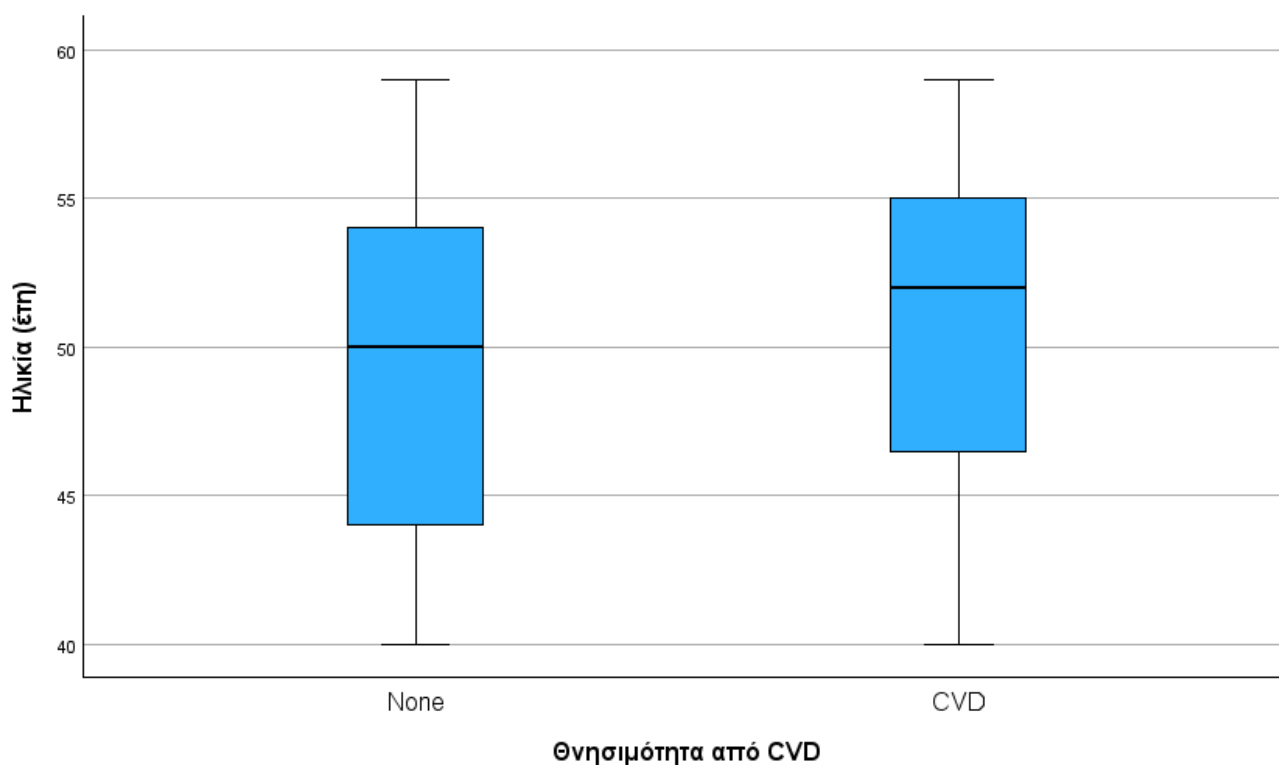
Ο τύπος του ελέγχου είναι:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

όπου:

- O_i οι τιμές που παρατηρήθηκαν
- E_i οι αναμενόμενες τιμές

- Ως προς την **ηλικία εισόδου**, ο έλεγχος που έγινε ήταν ο Mann-Whitney test για τη διερεύνηση διαφοράς της ηλικίας μεταξύ μη θανόντων από CVD και θανόντων από CVD. Οι θανόντες από CVD είχαν υψηλότερη ηλικία σε σχέση με τους μη θανόντες, όπως φαίνεται από το *Γράφημα 12*. Επιπλέον, οι θανόντες από CVD είχαν στατιστικά σημαντικά υψηλότερη διάμεση ηλικία (52 [46,5-55] έτη) σε σχέση με τους μη θανόντες από CVD (50 [44-54] έτη).



Γράφημα 12. Box-Plot της ηλικίας για τους δύο πληθυσμούς (θανόντες και μη θανόντες από CVD).

- Ως προς τη **σωματική δραστηριότητα**, ο έλεγχος που έγινε ήταν ο έλεγχος χ^2 . Ο έλεγχος υποθέσεων διαμορφώνεται ως εξής: H_0 : δεν υπάρχει εξάρτηση μεταξύ σωματικής δραστηριότητας και θνησιμότητας από CVD, H_1 : υπάρχει εξάρτηση μεταξύ σωματικής δραστηριότητας και θνησιμότητας από CVD. Από τον έλεγχο φαίνεται ότι δεν μπορούμε να απορρίψουμε την H_0 , καθώς το $p=0,275$ και συνεπώς δεν υπάρχει εξάρτηση μεταξύ σωματικής δραστηριότητας και θνησιμότητας από CVD. Συγκεκριμένα, 71,4% των θανόντων από CVD και το 66,8% των μη θανόντων από CVD ήταν σωματικά δραστήριοι.

- Ως προς το **κάπνισμα**, ο έλεγχος που έγινε ήταν ο έλεγχος χ^2 . Ο έλεγχος υποθέσεων διαμορφώνεται ως εξής: H_0 : δεν υπάρχει εξάρτηση μεταξύ καπνίσματος και θνησιμότητας από CVD, H_1 : υπάρχει εξάρτηση μεταξύ καπνίσματος και θνησιμότητας από CVD. Από τον έλεγχο φαίνεται ότι δεν μπορούμε να απορρίψουμε την H_0 , καθώς το $p=0,282$ και συνεπώς δεν υπάρχει εξάρτηση μεταξύ καπνίσματος και θνησιμότητας από CVD. Συγκεκριμένα, 67,9% των θανόντων από CVD και το 72,6% των μη θανόντων από CVD κάπνιζαν.
- Ως προς την **συστολική αρτηριακή πίεση**, ο έλεγχος που έγινε ήταν ο Student's t-test για τη διερεύνηση διαφοράς της συστολικής αρτηριακής πίεσης μεταξύ μη θανόντων από CVD και θανόντων από CVD. Ο έλεγχος υποθέσεων διαμορφώνεται ως εξής: H_0 : δεν υπάρχει διαφορά στη μέση συστολική αρτηριακή πίεση θανόντων και μη θανόντων από CVD, H_1 : υπάρχει διαφορά στη μέση συστολική αρτηριακή πίεση θανόντων και μη θανόντων από CVD. Από τον έλεγχο φαίνεται ότι μπορούμε να απορρίψουμε την H_0 , καθώς το $p<0,001$ και συνεπώς υπάρχει διαφορά στη μέση συστολική αρτηριακή πίεση θανόντων και μη θανόντων από CVD. Συγκεκριμένα, οι θανόντες από CVD είχαν στατιστικά σημαντικά υψηλότερη συστολική αρτηριακή πίεση ($139,2 \pm 23,7$ mmHg) σε σχέση με τους μη θανόντες από CVD ($132,7 \pm 19,3$ mmHg).
- Ως προς την **διαστολική αρτηριακή πίεση**, ο έλεγχος που έγινε ήταν ο Student's t-test για τη διερεύνηση διαφοράς της διαστολικής αρτηριακής πίεσης μεταξύ μη θανόντων από CVD και θανόντων από CVD. Ο έλεγχος υποθέσεων διαμορφώνεται ως εξής: H_0 : δεν υπάρχει διαφορά στη μέση διαστολική αρτηριακή πίεση θανόντων και μη θανόντων από CVD, H_1 : υπάρχει διαφορά στη μέση διαστολική αρτηριακή πίεση θανόντων και μη θανόντων από CVD. Από τον έλεγχο φαίνεται ότι μπορούμε να απορρίψουμε την H_0 , καθώς το $p=0,017$ και συνεπώς υπάρχει διαφορά στη μέση διαστολική αρτηριακή πίεση θανόντων και μη θανόντων από CVD. Συγκεκριμένα, οι θανόντες από CVD είχαν στατιστικά σημαντικά υψηλότερη διαστολική αρτηριακή πίεση ($84,2 \pm 13,0$ mmHg) σε σχέση με τους μη θανόντες από CVD ($81,8 \pm 10,4$ mmHg).

- Ως προς τον **καρδιακό ρυθμό**, ο έλεγχος που έγινε ήταν ο Student's t-test για τη διερεύνηση διαφοράς του καρδιακού ρυθμού μεταξύ μη θανόντων από CVD και θανόντων από CVD. Ο έλεγχος υποθέσεων διαμορφώνεται ως εξής: H_0 : δεν υπάρχει διαφορά στο μέσο καρδιακό ρυθμό θανόντων και μη θανόντων από CVD, H_1 : υπάρχει διαφορά στο μέσο καρδιακό ρυθμό θανόντων και μη θανόντων από CVD. Από τον έλεγχο φαίνεται ότι δεν μπορούμε να απορρίψουμε την H_0 , καθώς το $p=0,749$ και συνεπώς δεν υπάρχει διαφορά στο μέσο καρδιακό ρυθμό θανόντων και μη θανόντων από CVD. Συγκεκριμένα, οι θανόντες από CVD είχαν μέσο καρδιακό ρυθμό $68,6 \pm 11,6$ bpm και οι μη θανόντες από CVD $68,2 \pm 13,5$ bpm.
- Ως προς την **χοληστερόλη**, ο έλεγχος που έγινε ήταν ο Student's t-test για τη διερεύνηση διαφοράς της χοληστερόλης μεταξύ μη θανόντων από CVD και θανόντων από CVD. Ο έλεγχος υποθέσεων διαμορφώνεται ως εξής: H_0 : δεν υπάρχει διαφορά στη μέση χοληστερόλη θανόντων και μη θανόντων από CVD, H_1 : υπάρχει διαφορά στη μέση χοληστερόλη θανόντων και μη θανόντων από CVD. Από τον έλεγχο φαίνεται ότι δεν μπορούμε να απορρίψουμε την H_0 , καθώς το $p=0,188$ και συνεπώς δεν υπάρχει διαφορά στη μέση χοληστερόλη θανόντων και μη θανόντων από CVD. Συγκεκριμένα, οι θανόντες από CVD είχαν μέση χοληστερόλη $207,5 \pm 50,1$ mg/dL και οι μη θανόντες από CVD $202,0 \pm 43,1$ mg/dL.
- Ως προς την κατηγοριοποίηση του **δείκτη μάζας σώματος**, ο έλεγχος που έγινε ήταν ο έλεγχος χ^2 . Ο έλεγχος υποθέσεων διαμορφώνεται ως εξής: H_0 : δεν υπάρχει εξάρτηση μεταξύ δείκτη μάζας σώματος και θνησιμότητας από CVD, H_1 : υπάρχει εξάρτηση μεταξύ δείκτη μάζας σώματος και θνησιμότητας από CVD. Από τον έλεγχο φαίνεται ότι δεν μπορούμε να απορρίψουμε την H_0 , καθώς το $p=0,076$ και συνεπώς δεν υπάρχει εξάρτηση μεταξύ δείκτη μάζας σώματος και θνησιμότητας από CVD. Ωστόσο θα μπορούσαμε να χαρακτηρίσουμε το παραπάνω εύρημα οριακά μη στατιστικά σημαντικό. Συγκεκριμένα, οι θανόντες από CVD είχαν οριακά μη σημαντικά υψηλότερο ποσοστό ατόμων με δείκτη μάζας σώματος $>25\text{kg/m}^2$ (31,8%) έναντι των μη θανόντων από CVD (24,6%).

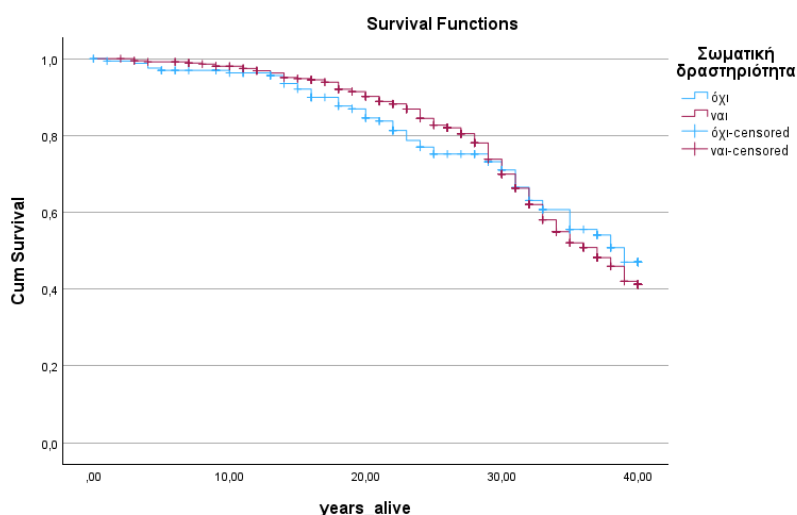
Πίνακας 12. Σύγκριση χαρακτηριστικών μεταξύ θανόντων έναντι μη θανόντων από CVD.

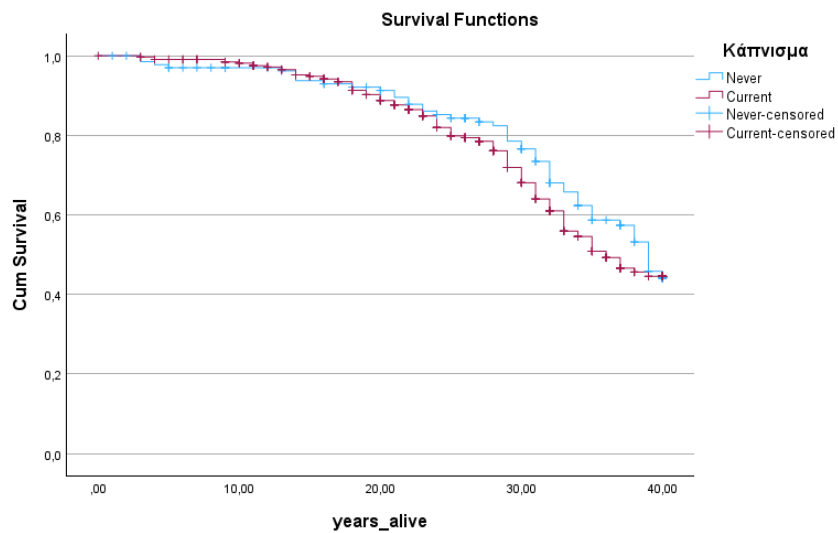
Μεταβλητές	Μη θανόντες από CVD	Θανόντες από CVD	P-value
Ηλικία εισόδου (έτη)	49,2 ± 5,8	50,8 ± 5,5	0,002
Σωματική δραστηριότητα (ναι)	66,8%	71,4%	0,275
Κάπνισμα (ναι)	72,6%	67,9%	0,282
ΣΑΠ (mmHg)	132,7 ± 19,3	139,2 ± 23,7	<0,001
ΔΑΠ (mmHg)	81,8 ± 10,4	84,2 ± 13,0	0,017
KP (bpm)	68,2 ± 13,5	68,6 ± 11,6	0,749
Χοληστερόλη (mg/dL)	202,0 ± 43,1	207,5 ± 50,1	0,188
ΔΜΣ (>25kg/m ²)	24,6%	31,8%	0,076

Συντομογραφίες. Συστολική Αρτηριακή Πίεση (ΣΑΠ), Διαστολική Αρτηριακή Πίεση (ΔΑΠ), Καρδιακός Ρυθμός (ΚΡ), Δείκτης Μάζας Σώματος (ΔΜΣ).

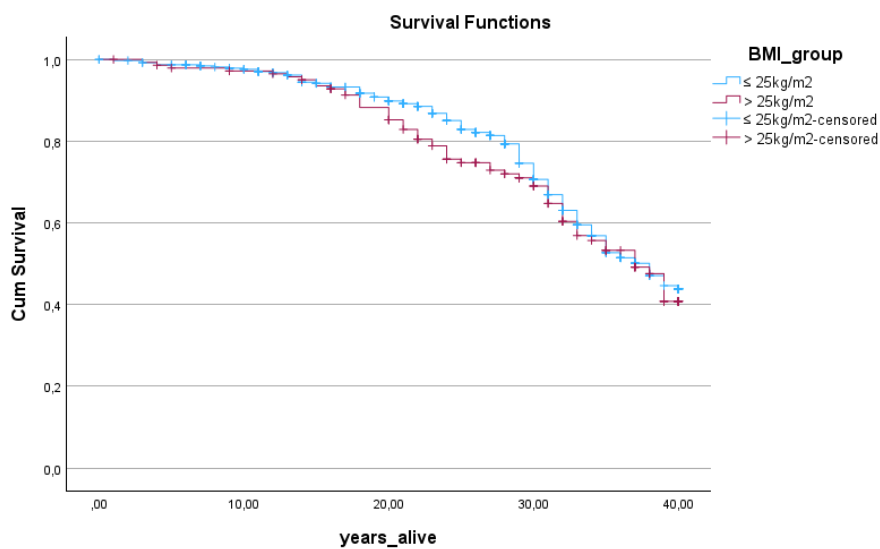
3.4. Γραφήματα Kaplan – Meier για τον θάνατο από CVD

Σύμφωνα με το log-rank test, δεν υπάρχει σημαντική διαφορά στην ανάλυση επιβίωσης με καταληκτικό σημείο τον θάνατο από CVD μεταξύ των σωματικά δραστήριων συμμετεχόντων έναντι των σωματικά μη δραστήριων ($p=0,869$). Το ίδιο συμβαίνει και αναφορικά με το κάπνισμα ($p=0,373$) και τον δείκτη μάζας σώματος ($p=0,431$). Οι καμπύλες K-M παρουσιάζουν επίσης το ίδιο αποτέλεσμα με το log-rank test (Γραφήματα 12 έως 14).

**Γράφημα 13.** Καμπύλες Kaplan – Meier για το θάνατο από CVD ανά σωματική δραστηριότητα.



Γράφημα 14. Καμπύλες Kaplan – Meier για το θάνατο από CVD ανά καπνιστικές συνήθειες.



Γράφημα 15 Καμπύλες Kaplan – Meier για το θάνατο από CVD ανά κατηγορία δείκτη μάζας σώματος.

4. Μοντέλα ανάλυσης επιβίωσης

4.1. Μοντέλα ανάλυσης επιβίωσης για τον θάνατο από CVD

Στη συνέχεια προχωρήσαμε στην κατασκευή μονοπαραγοντικών και πολυπαραγοντικών μοντέλων ανάλυσης επιβίωσης με καταληκτικό σημείο τον θάνατο από CVD για την αξιολόγηση πιθανών παραγόντων κινδύνου για τη θνησιμότητα από CVD.

i. Μοντέλο αναλογικών κινδύνων του Cox

Στην αρχή χρησιμοποιήσαμε το μοντέλο αναλογικών κινδύνων του Cox. Ξεκινήσαμε δημιουργώντας τα μονοπαραγοντικά μοντέλα για να ελέγξουμε όλους τους παράγοντες κινδύνου πριν προχωρήσουμε σε πιο ολοκληρωμένα. Χρησιμοποιήθηκε η μονομεταβλητή παλινδρόμηση αναλογικών κινδύνων Cox για κάθε πιθανό παράγοντα κινδύνου (ηλικία εισόδου, σωματική δραστηριότητα, κάπνισμα, συστολική αρτηριακή πίεση, διαστολική αρτηριακή πίεση, καρδιακός ρυθμός, ολική χοληστερόλη, δείκτης μάζας σώματος). Οι μεταβλητές ηλικία εισόδου, συστολική αρτηριακή πίεση, διαστολική αρτηριακή πίεση, καρδιακός ρυθμός και ολική χοληστερόλη εισήχθησαν στο μοντέλο ως ποσοτικές μεταβλητές. Οι μεταβλητές σωματική δραστηριότητα, κάπνισμα και δείκτης μάζας σώματος εισήχθησαν στο μοντέλο ως κατηγορικές μεταβλητές με 2 κατηγορίες και ομάδα αναφοράς τους σωματικά μη δραστήριους, τους μη καπνιστές και τα άτομα με δείκτη μάζας σώματος $\leq 25\text{kg/m}^2$.

Τα αποτελέσματα παρουσιάζονται στον Πίνακα 13. Από τα αποτελέσματα φαίνεται πως:

- η **ηλικία εισόδου** στη μελέτη αποτέλεσε παράγοντα κινδύνου για θνησιμότητα από CVD και συγκεκριμένα για κάθε 1 έτος παραπάνω ο κίνδυνος αυξανόταν κατά 10% (HR = 1,10, 95% ΔΕ= 1,08, 1,13, $p<0,001$).
- παρομοίως για την **συστολική αρτηριακή πίεση**, για κάθε 1 mmHg παραπάνω ο κίνδυνος να καταλήξει το άτομο από CVD αυξανόταν κατά 1% (HR = 1,01, 95% ΔΕ= 1,01, 1,02, $p<0,001$)
- αναφορικά με την **διαστολική αρτηριακή πίεση**, για κάθε 1 mmHg παραπάνω ο κίνδυνος να καταλήξει το άτομο από CVD αυξανόταν κατά 2% (HR = 1,01, 95% ΔΕ= 1,01, 1,03, $p=0,002$).

- οι υπόλοιπες μεταβλητές δεν συσχετίστηκαν στατιστικά σημαντικά με τον κίνδυνο για θνησιμότητα από CVD (**σωματική δραστηριότητα**: HR = 1,03, 95% ΔΕ= 0,75, 1,40, p=0,871; **κάπνισμα**: HR = 1,16, 95% ΔΕ= 0,83, 1,61, p=0,377; **καρδιακός ρυθμός**: HR = 1,004, 95% ΔΕ= 0,99, 1,02, p=0,431; **ολική χοληστερόλη**: HR = 1,001, 95% ΔΕ= 0,998, 1,004, p=0,578; **δείκτης μάζας σώματος**: HR = 1,13, 95% ΔΕ= 0,83, 1,53, p=0,442).

Στη συνέχεια προχωρήσαμε στην δημιουργία του πολυπαραγοντικού μοντέλου αναλογικών κινδύνων του Cox για τον σχετικό κίνδυνο θανάτου από CVD. Στο μοντέλο εισήχθησαν όλες οι μεταβλητές, εκτός από την διαστολική αρτηριακή πίεση λόγω υψηλής συσχέτισης με την συστολική αρτηριακή πίεση και αποφυγής πολυσυγγραμικότητας στο τελικό πολυπαραγοντικό μοντέλο. Τα αποτελέσματα του μοντέλου παρουσιάζονται στον *Πίνακα 13*.

Η εξίσωση του τελικού μοντέλου ήταν:

$$h_i(t) = h_0(t) \exp(0,098 * age_entry - 0,005 * physical_act_i + 0,33 * smoking_i + 0,01 * sbp + 0,007 * heartrate - 0,003 * tchol + 0,103 * BMI_group)$$

- Η **ηλικία εισόδου** στη μελέτη, όπως και στο μονοπαραγοντικό μοντέλο, αποτέλεσε παράγοντα κινδύνου για θνησιμότητα από CVD και συγκεκριμένα για κάθε 1 έτος παραπάνω ο κίνδυνος αυξανόταν κατά 10% (HR = 1,10, 95% ΔΕ= 1,07, 1,14, p<0,001), ανεξάρτητα από την σωματική δραστηριότητα, το κάπνισμα, τη συστολική αρτηριακή πίεση, τον καρδιακό ρυθμό, την ολική χοληστερόλη και τον δείκτη μάζας σώματος.
- Όμοια, η **συστολική αρτηριακή πίεση** αποτέλεσε παράγοντα κινδύνου για θνησιμότητα από CVD και συγκεκριμένα για κάθε 1 mmHg παραπάνω ο κίνδυνος αυξανόταν κατά 1% (HR = 1,01, 95% ΔΕ= 1,002, 1,02, p=0,012).

- Το **κάπνισμα** συσχετίστηκε οριακά μη σημαντικά με αυξημένο κίνδυνο θανάτου από CVC, με τους συμμετέχοντες όπου κάπνιζαν να εμφανίζουν αυξημένο κίνδυνο κατά 39% έναντι αυτών που δεν κάπνιζαν (HR = 1,39, 95% ΔΕ= 0,99, 1,96, p=0,059).
- Οι υπόλοιπες μεταβλητές δεν συσχετίστηκαν στατιστικά σημαντικά με τον κίνδυνο για θνησιμότητα από CVD (**σωματική δραστηριότητα**: HR = 0,995, 95% ΔΕ= 0,70, 1,43, p=0,979; **καρδιακός ρυθμός**: HR = 1,01, 95% ΔΕ= 0,99, 1,02, p=0,308; **ολική χοληστερόλη**: HR = 0,998, 95% ΔΕ= 0,99, 1,0001, p=0,197; **δείκτης μάζας σώματος**: HR = 1,11, 95% ΔΕ= 0,76, 1,62, p=0,593). Επιπλέον, το AIC ήταν ίσο με 1721,913.

Πίνακας 13. Μονοπαραγοντικό και πολυπαραγοντικό μοντέλο αναλογικών κινδύνων του Cox για τον σχετικό κίνδυνο θανάτου από CVD.

Μεταβλητές	Μονοπαραγοντική ανάλυση			Πολυπαραγοντική ανάλυση		
	HR	95%CI	P-value	HR	95%CI	P-value
Ηλικία εισόδου (έτη)	1,10	(1,08, 1,13)	<0,001	1,10	(1,07, 1,14)	<0,001
Σωματική δραστηριότητα (ναι)	1,03	(0,75, 1,40)	0,871	0,995	(0,70, 1,43)	0,979
Κάπνισμα (ναι)	1,16	(0,83, 1,61)	0,377	1,39	(0,99, 1,96)	0,059
ΣΑΠ (mmHg)	1,01	(1,01,1,02)	<0,001	1,01	(1,002, 1,02)	0,012
ΔΑΠ (mmHg)	1,02	(1,01,1,03)	0,002	-	-	-
KP (bpm)	1,004	(0,99, 1,02)	0,431	1,01	(0,99, 1,02)	0,308
Χοληστερόλη (mg/dL)	1,001	(0,998, 1,004)	0,578	0,998	(0,99, 1,00)	0,197
ΔΜΣ (>25kg/m ²)	1,13	(0,83, 1,53)	0,442	1,11	(0,76, 1,62)	0,593

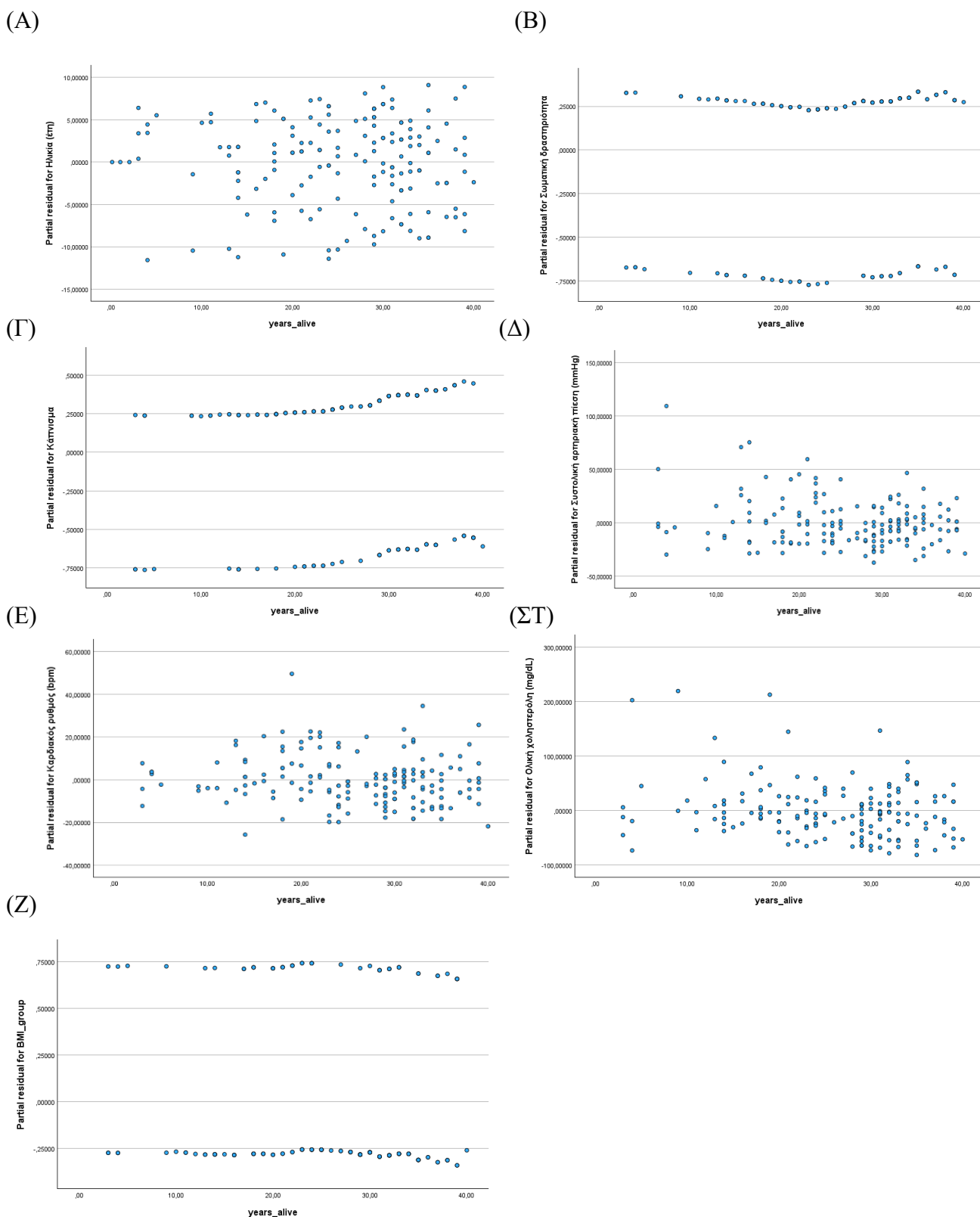
Συντομογραφίες. Λόγος Κινδύνου (HR), Διάστημα Εμπιστοσύνης (CI), Συστολική Αρτηριακή Πίεση (ΣΑΠ), Διαστολική Αρτηριακή Πίεση (ΔΑΠ), Καρδιακός Ρυθμός (KP), Δείκτης Μάζας Σώματος (ΔΜΣ).

Στη συνέχεια αξιολογήθηκε η επάρκεια του μοντέλου αναλογικών κινδύνων Cox (model fit, προϋπόθεση αναλογικότητας κινδύνων, goodness of fit). Σε σχέση με την προϋπόθεση της αναλογικότητας των κινδύνων, έγινε γραφικός έλεγχος μέσω των καταλοίπων Schoenfeld που προέκυψαν από το πολυπαραγοντικό μοντέλο αναλογικών κινδύνων του Cox για τον σχετικό κίνδυνο θανάτου από CVD του Πίνακα 13 (άξονας Y) σε συνάρτηση με τον χρόνο επιβίωσης (άξονας X) για τις μεταβλητές ηλικία (Γράφημα 16Α), σωματική δραστηριότητα (Γράφημα 16Β), κάπνισμα (Γράφημα

16Γ), συστολική αρτηριακή πίεση (Γράφημα 16Δ), καρδιακός ρυθμός (Γράφημα 16Ε), ολική χοληστερόλη (Γράφημα 16ΣΤ) και δείκτης μάζας σώματος (Γράφημα 16Ζ).

- Ως προς την **ηλικία** (Γράφημα 16Α), δεν φαίνεται να υπάρχει κάποια συσχέτιση μεταξύ των καταλοίπων Schoenfeld με τον χρόνο επιβίωσης, καθώς στο αντίστοιχο γράφημα είναι όμοια κατανεμημένα σε όλο το εύρος.
- Αντίστοιχα, ως προς την **σωματική δραστηριότητα** (Γράφημα 16Β), οι δύο ομάδες ατόμων (πάνω και κάτω γραμμές) είναι παράλληλες και δεν μεταβάλλονται με το πέρας του χρόνου, γεγονός που υποδεικνύει ότι πληρείται η προϋπόθεση της αναλογικότητας των κινδύνων.
- Αντίθετα, στο Γράφημα 16Γ, αναφορικά με το **κάπνισμα**, φαίνεται ότι οι δύο γραμμές αυξάνονται με το πέρας του χρόνου, το οποίο αποτελεί ένδειξη ότι η αναλογικότητα των κινδύνων δεν πληρείται για τη μεταβλητή κάπνισμα.
- Ως προς τη **συστολική αρτηριακή πίεση** (Γράφημα 16Δ), φαίνεται να υπάρχει μια πολύ ελαφριά και αρνητική σχέση των καταλοίπων Schoenfeld με το χρόνο, γεγονός που δηλώνει ότι πληρείται η προϋπόθεση αναλογικότητας των κινδύνων.
- Παρόμοια αποτελέσματα παρατηρούμε και για την ολική **χοληστερόλη** (Γράφημα 16ΣΤ), ενώ για τον καρδιακό ρυθμό η σχέση μεταξύ καταλοίπων και χρόνο είναι ακόμη πιο ελαφριά (Γράφημα 15Ε).
- Τέλος, ως προς την κατηγοριοποίηση με βάση τον **δείκτη μάζας σώματος** (Γράφημα 16Ζ), οι δύο ομάδες ατόμων (πάνω και κάτω γραμμές) είναι παράλληλες και δεν μεταβάλλονται με το πέρας του χρόνου, γεγονός που υποδεικνύει ότι πληρείται η προϋπόθεση της αναλογικότητας των κινδύνων.

Έλεγχος της υπόθεσης των αναλογικών κινδύνων χρησιμοποιώντας τα κατάλοιπα Schoenfeld για τις ποσοτικές μεταβλητές έγινε και μέσω του συντελεστή συσχέτισης Spearman's rho με τον χρόνο επιβίωσης (Πίνακας 14). Από τα αποτελέσματα φάνηκε ότι δεν υπήρχε στατιστικά σημαντική σχέση μεταξύ των καταλοίπων Schoenfeld με τον χρόνο επιβίωσης για όλες τις μεταβλητές (ηλικία, συστολική αρτηριακή πίεση, καρδιακός ρυθμός, ολική χοληστερόλη). Παρόλα αυτά, ο συντελεστής συσχέτισης ήταν οριακά μη σημαντικός και για τον καρδιακό ρυθμό ($p=0,059$) και για την ολική χοληστερόλη ($p=0,065$). Επομένως η προϋπόθεση των αναλογικών κινδύνων δεν πληρείται για το κάπνισμα και πιθανώς για τον καρδιακό ρυθμό και για την ολική χοληστερόλη.



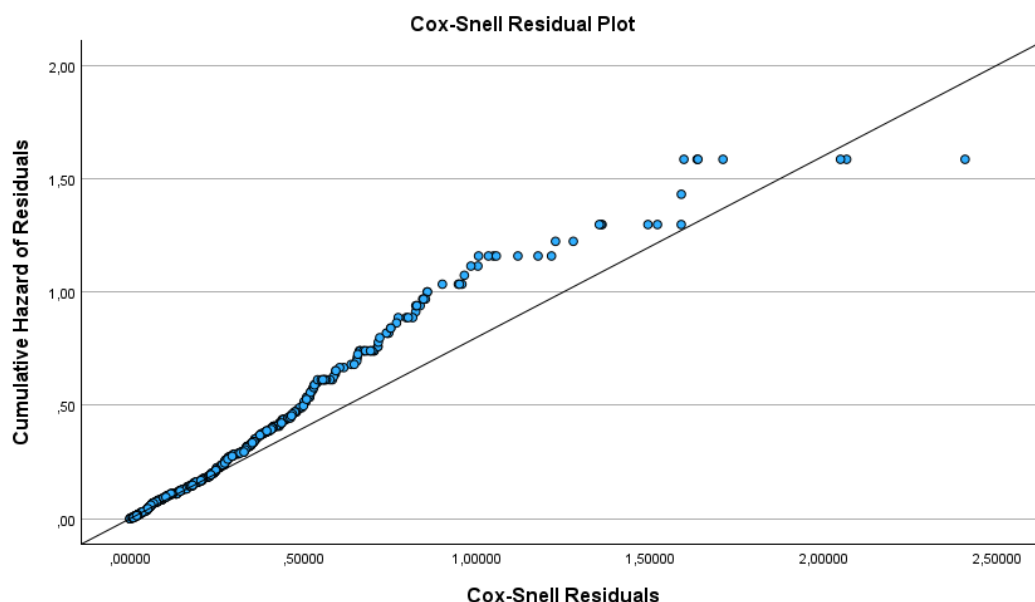
Γράφημα 16. Στικτόγραμμα του χρόνου (άξονας X) σε σχέση με τα κατάλοιπα Schoenfeld (άξονας Y) που προέκυψαν από το πολυπαραγοντικό μοντέλο αναλογικών κινδύνων του Cox για τον σχετικό κίνδυνο θανάτου από CVD (Πίνακας 13) των μεταβλητών ηλικία (A), σωματική δραστηριότητα (B), κάπνισμα (Γ), συστολική αρτηριακή πίεση (Δ), καρδιακός ρυθμός (Ε), ολική χοληστερόλη (ΣΤ) και δείκτης μάζας σώματος (Ζ).

Πίνακας 14. Έλεγχος της υπόθεσης των αναλογικών κινδύνων χρησιμοποιώντας τα κατάλοιπα Schoenfeld (συσχέτιση με τον χρόνο επιβίωσης).

Schoenfeld residuals	Spearman's rho	P-value
Ηλικία	-0,059	0,454
ΣΑΠ	-0,051	0,517
KP	-0,149	0,059
Χοληστερόλη	-0,145	0,065

Συντομογραφίες. Συστολική Αρτηριακή Πίεση (ΣΑΠ), Καρδιακός Ρυθμός (KP).

Στη συνέχεια αξιολογήθηκε η καλή προσαρμογή του μοντέλου μέσω του διαγράμματος καταλοίπων Cox-Snell. Αναλυτικότερα, παρουσιάζεται ένα διάγραμμα των καταλοίπων Cox-Snell (άξονας X) έναντι του εκτιμώμενου αθροιστικού κινδύνου αυτών (Cumulative Hazard of Cox-Snell residuals - άξονας Y) (Γράφημα 16). Από το Γράφημα παρατηρούμε ότι υπάρχει σημαντική απόκλιση από την ευθεία γραμμή, γεγονός που υποδηλώνει κακή προσαρμογή του μοντέλου. Επομένως, φαίνεται πως το μοντέλο αναλογικών κινδύνων του Cox δεν εφαρμόζει καλά και υστερεί αξιοπιστίας.



Γράφημα 17. Διάγραμμα των καταλοίπων Cox-Snell (άξονας X) έναντι του εκτιμώμενου αθροιστικού κινδύνου αυτών (Cumulative Hazard of Cox-Snell residuals - άξονας Y).

ii. Μοντέλο AFT

Καθώς το μοντέλο αναλογικών κινδύνων του Cox δεν εφαρμόζει καλά και υστερεί αξιοπιστίας, όπως είδαμε από την προηγούμενη ενότητα θα προχωρήσουμε στην εφαρμογή εναλλακτικών μοντέλων ανάλυσης επιβίωσης που πιθανώς να είναι καταλληλότερα και πιο αξιόπιστα για τη συγκεκριμένη ανάλυση. Το μοντέλο AFT είναι μια άλλη εναλλακτική λύση του μοντέλου Cox όταν παραβιάζεται η υπόθεση της αναλογικότητάς των κινδύνων. Οι κατανομές για τη δημιουργία των διαφόρων μοντέλων που χρησιμοποιήθηκαν στις συγκεκριμένες αναλύσεις ήταν η εκθετική (exponential), η Weibull, η Loglogistic και η Lognormal. Για κάθε είδος μοντέλου, προσαρμόσαμε τόσο το μονοπαραγοντικό όσο και το πολυπαραγοντικό μοντέλο AFT. Αναφορικά με τα πολυπαραγοντικά μοντέλα, εισήχθησαν όλες οι μεταβλητές, εκτός από την διαστολική αρτηριακή πίεση λόγω υψηλής συσχέτισης με την συστολική αρτηριακή πίεση και αποφυγής πολυσυγγραμικότητας.

Τα μονοπαραγοντικά μοντέλα AFT για την σχετική πιθανότητα επιβίωσης (μη θάνατος από CVD) παρουσιάζονται στον Πίνακα 15:

- Η **ηλικία** φάνηκε να αποτελεί στατιστικά σημαντικό αρνητικό προγνωστικό παράγοντα σε όλα τα μοντέλα, όπου για κάθε έτος παραπάνω η πιθανότητα επιβίωσης μειωνόταν κατά:
 - 5,7% στο Exponential μοντέλο ($e^{\beta} = 0,943$, 95% ΔΕ=0,919, 0,968 , $p<0,001$)
 - 3,1% στο Weibull μοντέλο ($e^{\beta} = 0,969$, 95% ΔΕ= 0,96, 0,977, $p<0,001$)
 - 3,4% στο Loglogistic μοντέλο ($e^{\beta} = 0,966$, 95% ΔΕ= 0,956, 0,975, $p<0,001$)
 - 4,2% στο Lognormal μοντέλο ($e^{\beta} = 0,958$, 95% ΔΕ= 0,946, 0,969, $p<0,001$)
- Αντίθετα, η **σωματική δραστηριότητα** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικό παράγοντα σε κανένα μοντέλο:
 - Exponential μοντέλο: $e^{\beta} = 0,964$, 95% ΔΕ= 0,705, 1,318, $p=0,818$
 - Weibull μοντέλο: $e^{\beta} = 0,995$, 95% ΔΕ= 0,886, 1,119, $p=0,937$

- Loglogistic μοντέλο: $e^{\beta} = 1,027$, 95% ΔΕ= 0,901, 1,171, $p=0,689$
 - Lognormal μοντέλο: $e^{\beta} = 1,087$, 95% ΔΕ= 0,923, 1,279, $p=0,317$).
- Ομοίως, το **κάπνισμα** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικό παράγοντα σε κανένα μοντέλο:
 - Exponential μοντέλο: $e^{\beta} = 1,047$, 95% ΔΕ= 0,755, 1,452, $p=0,782$
 - Weibull μοντέλο: $e^{\beta} = 0,955$, 95% ΔΕ= 0,85, 1,073, $p=0,434$
 - Loglogistic μοντέλο: $e^{\beta} = 0,945$, 95% ΔΕ= 0,829, 1,076, $p=0,394$
 - Lognormal μοντέλο: $e^{\beta} = 0,994$, 95% ΔΕ= 0,844, 1,171, $p=0,943$
- Η **συστολική αρτηριακή πίεση** φάνηκε να αποτελεί στατιστικά σημαντικό αρνητικό προγνωστικό παράγοντα σε όλα τα μοντέλα, όπου για κάθε 1mmHg αυξημένης συστολικής αρτηριακής πίεσης η πιθανότητα επιβίωσης μειωνόταν κατά:
 - 1,2% στο Exponential μοντέλο ($e^{\beta} = 0,988$, 95% ΔΕ=0,982, 0,994, $p<0,001$),
 - 0,5% στο Weibull μοντέλο ($e^{\beta} = 0,995$, 95% ΔΕ= 0,992, 0,997, $p<0,001$),
 - 0,7% στο Loglogistic μοντέλο ($e^{\beta} = 0,993$, 95% ΔΕ= 0,99, 0,996, $p<0,001$) και
 - 0,9% στο Lognormal μοντέλο ($e^{\beta} = 0,991$, 95% ΔΕ= 0,987, 0,994, $p<0,001$).
- Παρόμοια αποτελέσματα υπήρχαν και ως προς την **διαστολική αρτηριακή πίεση**, σε όλα τα μοντέλα.
- Επιπλέον, ο **καρδιακός ρυθμός** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικό παράγοντα σε κανένα μοντέλο:
 - Exponential μοντέλο: $e^{\beta} = 0,996$, 95% ΔΕ= 0,985, 1,007, $p=0,495$
 - Weibull μοντέλο: $e^{\beta} = 0,998$, 95% ΔΕ= 0,994, 1,002, $p=0,423$

- Loglogistic μοντέλο: $e^{\beta} = 0,997$, 95% ΔE= 0,993, 1,002, p=0,298
 - Lognormal μοντέλο: $e^{\beta} = 0,997$, 95% ΔE= 0,991, 1,003, p=0,293
- Παρομοίως, η **ολική χοληστερόλη** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικό παράγοντα σε κανένα μοντέλο:
 - Exponential μοντέλο: $e^{\beta} = 0,999$, 95% ΔE= 0,996, 1,002, p=0,452
 - Weibull μοντέλο: $e^{\beta} = 0,999$, 95% ΔE= 0,998, 1,001, p=0,538
 - Loglogistic μοντέλο: $e^{\beta} = 0,999$, 95% ΔE= 0,998, 1,001, p=0,326
 - Lognormal μοντέλο: $e^{\beta} = 0,999$, 95% ΔE= 0,998, 1,001, p=0,312
- Όμοια, η κατηγοριοποίηση του **δείκτη μάζας σώματος** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικό παράγοντα σε κανένα μοντέλο:
 - Exponential μοντέλο: $e^{\beta} = 0,843$, 95% ΔE= 0,623, 1,142, p=0,271
 - Weibull μοντέλο: $e^{\beta} = 0,953$, 95% ΔE= 0,85, 1,067, p=0,404
 - Loglogistic μοντέλο: $e^{\beta} = 0,935$, 95% ΔE= 0,822, 1,063, p=0,303
 - Lognormal μοντέλο: $e^{\beta} = 0,92$, 95% ΔE= 0,781, 1,084, p=0,312).

Τα πολυπαραγοντικά μοντέλα AFT για την σχετική πιθανότητα επιβίωσης (μη θάνατος από CVD) παρουσιάζονται στον *Πίνακα 16*.

- Η **ηλικία** στα πολυπαραγοντικά μοντέλα παρέμεινε στατιστικά σημαντικός αρνητικός προγνωστικός παράγοντας, σε όλα τα μοντέλα, όπου για κάθε έτος παραπάνω η πιθανότητα επιβίωσης μειωνόταν κατά:
 - 5,3% στο Exponential μοντέλο ($e^{\beta} = 0,947$, 95% ΔE=0,921, 0,975 , p<0,001),
 - 3% στο Weibull μοντέλο ($e^{\beta} = 0,970$, 95% ΔE= 0,962, 0,979, p<0,001),

- 3,1% στο Loglogistic μοντέλο ($e^{\beta} = 0,969$, 95% ΔΕ= 0,959, 0,979, $p < 0,001$) και
- 3,5% στο Lognormal μοντέλο ($e^{\beta} = 0,965$, 95% ΔΕ= 0,952, 0,977, $p < 0,001$),

ανεξάρτητα από την σωματική δραστηριότητα, το κάπνισμα, τη συστολική αρτηριακή πίεση, τον καρδιακό ρυθμό, την ολική χοληστερόλη και τον δείκτη μάζας σώματος.

- Η **σωματική δραστηριότητα** παρέμεινε μη στατιστικά σημαντικός προγνωστικός παράγοντας σε κανένα πολυπαραγοντικό μοντέλο:
 - Exponential μοντέλο: $e^{\beta} = 0,932$, 95% ΔΕ= 0,653, 1,331, $p = 0,701$
 - Weibull μοντέλο: $e^{\beta} = 1,01$, 95% ΔΕ= 0,899, 1,134, $p = 0,865$
 - Loglogistic μοντέλο: $e^{\beta} = 1,019$, 95% ΔΕ= 0,896, 1,160, $p = 0,775$
 - Lognormal μοντέλο: $e^{\beta} = 1,025$, 95% ΔΕ= 0,874, 1,204, $p = 0,758$
- Το **κάπνισμα** παρέμεινε μη στατιστικά σημαντικός προγνωστικός παράγοντας στο Exponential μοντέλο ($e^{\beta} = 0,981$, 95% ΔΕ= 0,700, 1,374, $p = 0,913$). Αντίθετα, στο Weibull μοντέλο φάνηκε ότι το κάπνισμα ήταν οριακά μη σημαντικός και αρνητικός προγνωστικός παράγοντας, όπου οι καπνιστές είχαν χαμηλότερη πιθανότητα επιβίωσης κατά 8,8%, με το $p\text{-value} < 0,1$ ($e^{\beta} = 0,912$, 95% ΔΕ= 0,817, 1,017, $p = 0,098$). Παρόμοια, στο Loglogistic μοντέλο οι καπνιστές είχαν χαμηλότερη πιθανότητα επιβίωσης κατά 10,2%, με το $p\text{-value} < 0,1$ ($e^{\beta} = 0,898$, 95% ΔΕ= 0,795, 1,013, $p = 0,081$). Ωστόσο, στο Lognormal μοντέλο το κάπνισμα αποτέλεσε μη στατιστικά σημαντικό και αρνητικό προγνωστικό παράγοντας ($e^{\beta} = 0,940$, 95% ΔΕ= 0,805, 1,096, $p = 0,430$).
- Η **συστολική αρτηριακή πίεση** παρέμεινε στατιστικά σημαντικός αρνητικός προγνωστικός παράγοντας σε όλα τα μοντέλα (εκτός του Exponential), όπου για κάθε 1mmHg αυξημένης συστολικής αρτηριακής πίεσης η πιθανότητα επιβίωσης μειωνόταν κατά:
 - 0,8% στο Exponential μοντέλο ($e^{\beta} = 0,992$, 95% ΔΕ= 0,984, 1,000, $p = 0,06$),
 - 0,3% στο Weibull μοντέλο ($e^{\beta} = 0,997$, 95% ΔΕ= 0,994, 0,999, $p = 0,012$),

- 0,4% στο Loglogistic μοντέλο ($e^{\beta} = 0,996$, 95% ΔΕ= 0,992, 0,999, $p=0,005$) και
 - 0,5% στο Lognormal μοντέλο ($e^{\beta} = 0,995$, 95% ΔΕ= 0,991, 0,999, $p=0,006$).
- Ο **καρδιακός ρυθμός** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικό παράγοντα σε κανένα μοντέλο:
 - Exponential μοντέλο: $e^{\beta} = 0,997$, 95% ΔΕ= 0,983, 1,010, $p=0,627$
 - Weibull μοντέλο: $e^{\beta} = 0,998$, 95% ΔΕ= 0,994, 1,002, $p=0,348$
 - Loglogistic μοντέλο: $e^{\beta} = 0,997$, 95% ΔΕ= 0,992, 1,002, $p=0,215$
 - Lognormal μοντέλο: $e^{\beta} = 0,997$, 95% ΔΕ= 0,991, 1,004, $p=0,405$).
- Παρομοίως, η **ολική χοληστερόλη** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικό παράγοντα σε κανένα μοντέλο:
 - Exponential μοντέλο: $e^{\beta} = 1,001$, 95% ΔΕ= 0,997, 1,004, $p=0,765$
 - Weibull μοντέλο: $e^{\beta} = 1,001$, 95% ΔΕ= 1,000, 1,002, $p=0,247$
 - Loglogistic μοντέλο: $e^{\beta} = 1,000$, 95% ΔΕ= 0,999, 1,002, $p=0,627$
 - Lognormal μοντέλο: $e^{\beta} = 0,999$, 95% ΔΕ= 0,998, 1,001, $p=0,593$).
- Όμοια, η κατηγοριοποίηση του **δείκτη μάζας σώματος** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικό παράγοντα σε κανένα μοντέλο:
 - Exponential μοντέλο: $e^{\beta} = 0,914$, 95% ΔΕ= 0,627, 1,332, $p=0,640$
 - Weibull μοντέλο: $e^{\beta} = 0,970$, 95% ΔΕ= 0,859, 1,096, $p=0,629$
 - Loglogistic μοντέλο: $e^{\beta} = 0,990$, 95% ΔΕ= 0,864, 1,135, $p=0,890$
 - Lognormal μοντέλο: $e^{\beta} = 0,998$, 95% ΔΕ= 0,884, 1,182, $p=0,982$).

Πίνακας 15. Μονοπαραγοντικά μοντέλα AFT για την σχετική πιθανότητα επιβίωσης (μη θάνατος από CVD).

	Exponential				Weibull				Loglogistic				Lognormal			
	β	e ^β	95%CI	p	β	e ^β	95%CI	p	β	e ^β	95%CI	p	β	e ^β	95%CI	p
Ηλικία εισόδου (έτη)	-0,059	0,943	(0,919, 0,968)	<0,001	-0,032	0,969	(0,96, 0,977)	<0,001	-0,035	0,966	(0,956, 0,975)	<0,001	-0,043	0,958	(0,946, 0,969)	<0,001
Σωματική δραστηριότητα (ναι)	-0,037	0,964	(0,705, 1,318)	0,818	-0,005	0,995	(0,886, 1,119)	0,937	0,027	1,027	(0,901, 1,171)	0,689	0,083	1,087	(0,923, 1,279)	0,317
Κάπνισμα (ναι)	0,046	1,047	(0,755, 1,452)	0,782	-0,046	0,955	(0,85, 1,073)	0,434	-0,057	0,945	(0,829, 1,076)	0,394	-0,006	0,994	(0,844, 1,171)	0,943
ΣΑΠ (mmHg)	-0,012	0,988	(0,982, 0,994)	<0,001	-0,005	0,995	(0,992, 0,997)	<0,001	-0,007	0,993	(0,99, 0,996)	<0,001	-0,009	0,991	(0,987, 0,994)	<0,001
ΔΑΠ (mmHg)	-0,016	0,984	(0,972, 0,996)	0,007	-0,007	0,993	(0,988, 0,997)	0,001	-0,009	0,991	(0,985, 0,996)	<0,001	-0,014	0,986	(0,98, 0,992)	<0,001
KP (bpm)	-0,004	0,996	(0,985, 1,007)	0,495	-0,002	0,998	(0,994, 1,002)	0,423	-0,003	0,997	(0,993, 1,002)	0,298	-0,003	0,997	(0,991, 1,003)	0,293
Χοληστερόλη (mg/dL)	-0,001	0,999	(0,996, 1,002)	0,452	-	0,999	(0,998, 1,001)	0,538	-0,001	0,999	(0,998, 1,001)	0,326	-0,001	0,999	(0,998, 1,001)	0,312
ΔΜΣ (>25kg/m ²)	-0,171	0,843	(0,623, 1,142)	0,271	0,0004	0,953	(0,85, 1,067)	0,404	-0,067	0,935	(0,822, 1,063)	0,303	-0,083	0,92	(0,781, 1,084)	0,321

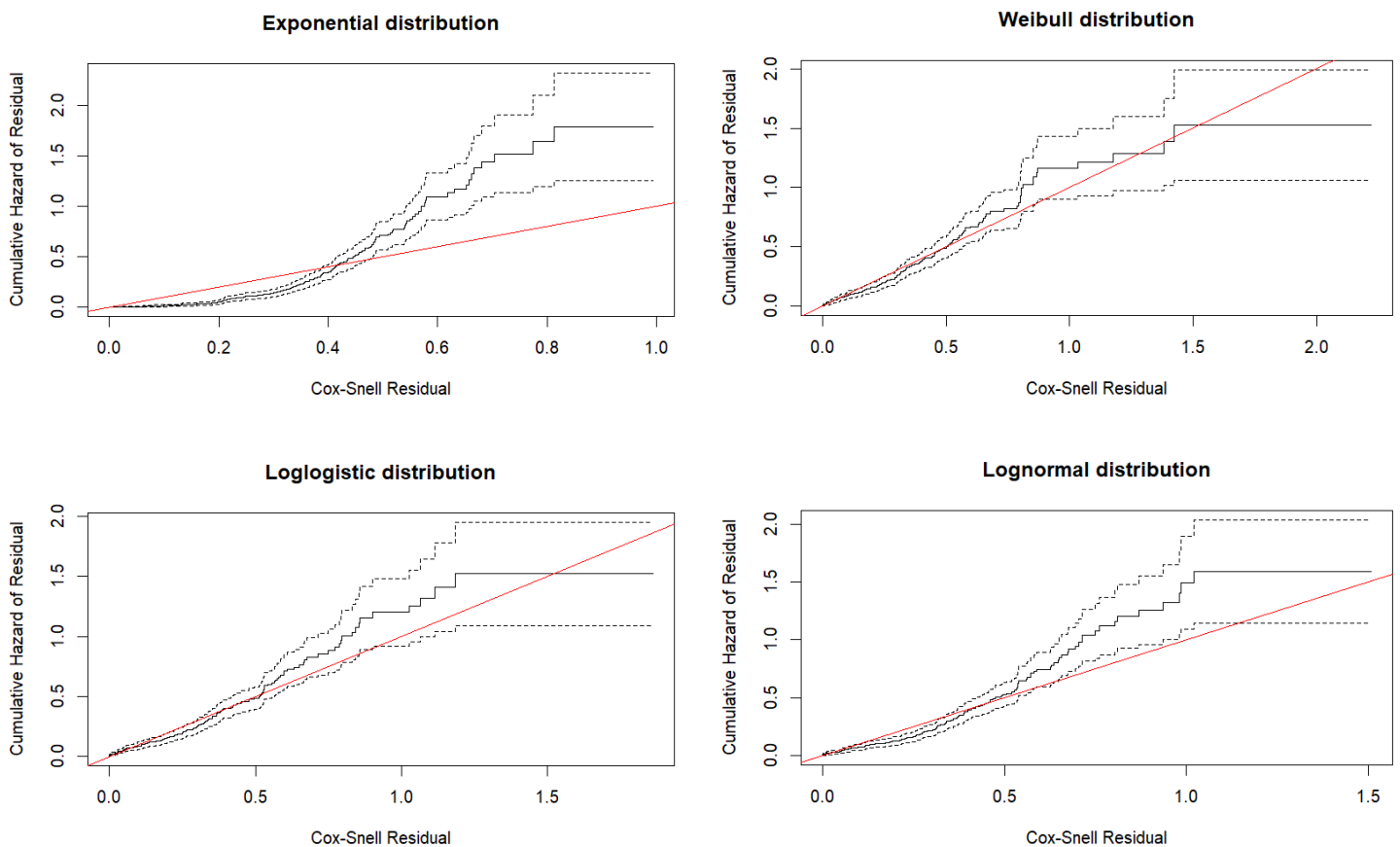
Συντομογραφίες. Διάστημα Εμπιστοσύνης (CI), Συστολική Αρτηριακή Πίεση (ΣΑΠ), Καρδιακός Ρυθμός (ΚΡ), Δείκτης Μάζας Σώματος (ΔΜΣ).

Πίνακας 16. Πολυπαραγοντικά μοντέλα AFT για την σχετική πιθανότητα επιβίωσης (μη θάνατος από CVD).

	Exponential				Weibull				Loglogistic				Lognormal			
	β	e ^β	95%CI	p	β	e ^β	95%CI	p	β	e ^β	95%CI	p	β	e ^β	95%CI	P
Ηλικία εισόδου (έτη)	-0,054	0,947	(0,921, 0,975)	<0,001	-0,030	0,970	(0,962, 0,979)	<0,001	-0,032	0,969	(0,959, 0,979)	<0,001	-0,036	0,965	(0,952, 0,977)	<0,001
Σωματική δραστηριότητα (ναι)	-0,070	0,932	(0,653, 1,331)	0,701	0,010	1,010	(0,899, 1,134)	0,865	0,019	1,019	(0,896, 1,160)	0,775	0,025	1,025	(0,874, 1,204)	0,758
Κάπνισμα (ναι)	-0,019	0,981	(0,700, 1,374)	0,913	-0,092	0,912	(0,817, 1,017)	0,098	-0,108	0,898	(0,795, 1,013)	0,081	-0,062	0,940	(0,805, 1,096)	0,430
ΣΑΠ (mmHg)	-0,008	0,992	(0,984, 1,000)	0,060	-0,003	0,997	(0,994, 0,999)	0,012	-0,004	0,996	(0,992, 0,999)	0,005	-0,005	0,995	(0,991, 0,999)	0,006
KP (bpm)	-0,003	0,997	(0,983, 1,010)	0,627	-0,002	0,998	(0,994, 1,002)	0,348	-0,003	0,997	(0,992, 1,002)	0,215	-0,003	0,997	(0,991, 1,004)	0,405
Χοληστερόλη (mg/dL)	0,001	1,001	(0,997, 1,004)	0,765	0,001	1,001	(1,000, 1,002)	0,247	0,0003	1,000	(0,999, 1,002)	0,627	-0,0004	0,999	(0,998, 1,001)	0,593
ΔΜΣ (>25kg/m ²)	-0,090	0,914	(0,627, 1,332)	0,640	-0,030	0,970	(0,859, 1,096)	0,629	-0,010	0,990	(0,864, 1,135)	0,890	-0,002	0,998	(0,844, 1,182)	0,982

Συντομογραφίες. Διάστημα Εμπιστοσύνης (CI), Συστολική Αρτηριακή Πίεση (ΣΑΠ), Καρδιακός Ρυθμός (ΚΡ), Δείκτης Μάζας Σώματος (ΔΜΣ)..

Στη συνέχεια αξιολογήθηκε η καλή προσαρμογή των μοντέλων μέσω του διαγράμματος καταλοίπων Cox-Snell. Αναλυτικότερα, παρουσιάζεται ένα διάγραμμα των καταλοίπων Cox-Snell (άξονας X) έναντι του εκτιμώμενου αθροιστικού κινδύνου αυτών (Cumulative Hazard of Cox-Snell residuals - άξονας Y) για κάθε πολυπαραγοντικό μοντέλο (Γράφημα 17). Από το Γράφημα παρατηρούμε ότι το πολυπαραγοντικό μοντέλο με Exponential κατανομή είχε την χειρότερη προσαρμογή, καθώς υπήρχε τεράστια απόκλιση από την ευθεία κόκκινη γραμμή. Παρόμοια, αλλά ελαφρώς καλύτερη προσαρμογή έδειξε να έχει το πολυπαραγοντικό μοντέλο με Lognormal κατανομή, όπου και πάλι στο γράφημα υπήρχε σημαντική απόκλιση από την ευθεία κόκκινη γραμμή, αλλά μικρότερη σε σχέση με το μοντέλο με Exponential κατανομή. Το πολυπαραγοντικό μοντέλο με Loglogistic κατανομή είχε εμφανώς καλύτερη προσαρμογή από τα παραπάνω δύο μοντέλα, με το διάγραμμα καταλοίπων να είναι αρκετά κοντά στην ευθεία κόκκινη γραμμή. Ωστόσο, την καλύτερη προσαρμογή την είχε το πολυπαραγοντικό μοντέλο με Weibull κατανομή, με το διάγραμμα καταλοίπων να είναι πολύ κοντά στην ευθεία κόκκινη γραμμή σε όλο σχεδόν το διάγραμμα.



Γράφημα 18. Διαγράμματα των καταλοίπων Cox-Snell (άξονας X) έναντι του εκτιμώμενου αθροιστικού κινδύνου αυτών (Cumulative Hazard of Cox-Snell residuals - άξονας Y) για τα πολυπαραγοντικά μοντέλα με Exponential, Weibull, Loglogistic και Lognormal κατανομές.

Στη συνέχεια συγκρίθηκαν όλα τα πολυπαραγοντικά μοντέλα με βάσει το κριτήριο AIC, των οποίων τα αποτελέσματα παρουσιάζονται και στον *Πίνακα 19*. Για το συγκεκριμένο κριτήριο, καλύτερο μοντέλο θεωρείται αυτό με την χαμηλότερη τιμή AIC. Όπως φαίνεται από τα αποτελέσματα, το καλύτερο μοντέλο ήταν το πολυπαραγοντικό AFT μοντέλο με Weibull κατανομή, με τιμή AIC = 519,082, στη συνέχεια ακολουθούσε το πολυπαραγοντικό AFT μοντέλο με Loglogistic κατανομή, με τιμή AIC = 528,1676, το πολυπαραγοντικό AFT μοντέλο με Lognormal κατανομή, με τιμή AIC = 566,5112, το πολυπαραγοντικό AFT μοντέλο με Exponential κατανομή, με τιμή AIC = 704,3146 και τέλος με την υψηλότερη τιμή το πολυπαραγοντικό μοντέλο αναλογικών κινδύνων του Cox, με τιμή AIC = 1721,913.

Πίνακας 17. Σύγκριση των AFT μοντέλων βάσει του AIC κριτηρίου.		
	Παράμετροι	AIC
Exponential	1	704,3146
Weibull	2	519,082
Loglogistic	2	528,1676
Lognormal	2	566,5112

4.2. Μοντέλα AFT ανάλυσης επιβίωσης για τον θάνατο από όλες τις αιτίες

Καθώς το μοντέλο αναλογικών κινδύνων του Cox δεν εφαρμόζει καλά και υστερεί αξιοπιστίας, όπως είδαμε από την προηγούμενη ενότητα θα επιλέξουμε να εφαρμόσουμε τα μοντέλα AFT για την σχετική πιθανότητα συνολικής επιβίωσης (θάνατος από όλες τις αιτίες) με τις κατανομές Exponential, Weibull, Loglogistic και Lognormal. Αναφορικά με τα πολυπαραγοντικά μοντέλα, εισήχθησαν όλες οι μεταβλητές, εκτός από την διαστολική αρτηριακή πίεση λόγω υψηλής συσχέτισης με την συστολική αρτηριακή πίεση και αποφυγής πολυσυγγραμικότητας.

Τα μονοπαραγοντικά μοντέλα AFT για την σχετική πιθανότητα συνολικής επιβίωσης (μη θάνατος) παρουσιάζονται στον Πίνακα 18:

- Η **ηλικία** φάνηκε να αποτελεί στατιστικά σημαντικό αρνητικό προγνωστικό παράγοντα σε όλα τα μοντέλα, όπου για κάθε έτος παραπάνω η πιθανότητα επιβίωσης μειωνόταν κατά:
 - 4,9% στο Exponential μοντέλο ($e^{\beta} = 0,951$, 95% ΔΕ=0,935, 0,967, $p<0,001$)
 - 3,1% στο Weibull μοντέλο ($e^{\beta} = 0,969$, 95% ΔΕ= 0,964, 0,976, $p<0,001$)
 - 3,4% στο Loglogistic μοντέλο ($e^{\beta} = 0,966$, 95% ΔΕ= 0,958, 0,973, $p<0,001$)
 - 4,3% στο Lognormal μοντέλο ($e^{\beta} = 0,957$, 95% ΔΕ= 0,946, 0,968, $p<0,001$)
- Η **σωματική δραστηριότητα** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικό παράγοντα στα μοντέλα με κατανομή Exponential ($e^{\beta} = 1,113$, 95% ΔΕ= 0,915, 1,354, $p=0,286$) και Weibull ($e^{\beta} = 1,057$, 95% ΔΕ= 0,973, 1,146, $p=0,187$), ενώ αποτέλεσε στατιστικά σημαντικό και θετικό προγνωστικό παράγοντα επιβίωσης στα υπόλοιπα μοντέλα, με τα σωματικά δραστήρια άτομα να έχουν υψηλότερη πιθανότητα επιβίωσης κατά 14,6% σε σχέση με τη μη σωματικά δραστήρια στο Loglogistic μοντέλο ($e^{\beta} = 1,146$, 95% ΔΕ= 1,033, 1,271, $p=0,010$) και υψηλότερη πιθανότητα επιβίωσης κατά 25,4% στο Lognormal μοντέλο ($e^{\beta} = 1,254$, 95% ΔΕ= 1,097, 1,432, $p=0,001$).

- Ομοίως, το **κάπνισμα** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικό παράγοντα στα μοντέλα με κατανομή Exponential ($e^{\beta} = 0,833$, 95% ΔΕ= 0,670, 1,036, $p=0,100$) και Lognormal ($e^{\beta} = 0,908$, 95% ΔΕ= 0,791, 1,044, $p=0,175$). Αντίθετα, αποτελούσε αρνητικό προγνωστικό παράγοντα επιβίωσης στα υπόλοιπα μοντέλα, με τα άτομα που καπνίζουν να έχουν χαμηλότερη πιθανότητα επιβίωσης κατά 12,5% σε σχέση με τους μη καπνίζοντες στο Weibull μοντέλο ($e^{\beta} = 0,875$, 95% ΔΕ= 0,8, 0,955, $p=0,003$) και κατά 13,5% στο Loglogistic μοντέλο ($e^{\beta} = 0,865$, 95% ΔΕ= 0,776, 0,965, $p=0,009$).
- Η **συστολική αρτηριακή πίεση** φάνηκε να αποτελεί στατιστικά σημαντικό αρνητικό προγνωστικό παράγοντα σε όλα τα μοντέλα (εκτός από το Exponential), όπου για κάθε 1mmHg αυξημένης συστολικής αρτηριακής πίεσης η πιθανότητα επιβίωσης μειωνόταν κατά:
 - 0,4% στο Exponential μοντέλο ($e^{\beta} = 0,996$, 95% ΔΕ=0,991, 1,000, $p=0,066$),
 - 0,6% στο Weibull μοντέλο ($e^{\beta} = 0,994$, 95% ΔΕ= 0,996, 0,999, $p=0,006$),
 - 0,4% στο Loglogistic μοντέλο ($e^{\beta} = 0,996$, 95% ΔΕ= 0,994, 0,999, $p=0,002$) και
 - 0,4% στο Lognormal μοντέλο ($e^{\beta} = 0,996$, 95% ΔΕ= 0,993, 0,999, $p=0,004$).
- Παρόμοια αποτελέσματα υπήρχαν και ως προς την **διαστολική αρτηριακή πίεση**, σε όλα τα μοντέλα (εκτός από Exponential).
- Επιπλέον, ο **καρδιακός ρυθμός** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικό παράγοντα σε κανένα μοντέλο:
 - Exponential μοντέλο: $e^{\beta} = 0,999$, 95% ΔΕ= 0,992, 1,006, $p=0,784$
 - Weibull μοντέλο: $e^{\beta} = 0,999$, 95% ΔΕ= 0,996, 1,002, $p=0,636$
 - Loglogistic μοντέλο: $e^{\beta} = 0,997$, 95% ΔΕ= 0,993, 1,001, $p=0,115$

- Lognormal μοντέλο: $e^{\beta} = 0,995$, 95% ΔE= 0,991, 1,00, p=0,069
- Παρομοίως, η **ολική χοληστερόλη** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικό παράγοντα σε κανένα μοντέλο (εκτός από το Lognormal):
 - Exponential μοντέλο: $e^{\beta} = 1,001$, 95% ΔE= 0,998, 1,003, p=0,618
 - Weibull μοντέλο: $e^{\beta} = 1,001$, 95% ΔE= 0,999, 1,001, p=0,491
 - Loglogistic μοντέλο: $e^{\beta} = 1,001$, 95% ΔE= 0,999, 1,002, p=0,311
 - Lognormal μοντέλο: $e^{\beta} = 1,001$, 95% ΔE= 1,00, 1,003, p=0,041
- Όμοια, η κατηγοριοποίηση του **δείκτη μάζας σώματος** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικό παράγοντα σε κανένα μοντέλο:
 - Exponential μοντέλο: $e^{\beta} = 1,044$, 95% ΔE= 0,85, 1,281, p=0,681
 - Weibull μοντέλο: $e^{\beta} = 1,033$, 95% ΔE= 0,949, 1,124, p=0,452
 - Loglogistic μοντέλο: $e^{\beta} = 1,057$, 95% ΔE= 0,951, 1,172, p=0,308
 - Lognormal μοντέλο: $e^{\beta} = 1,082$, 95% ΔE= 0,941, 1,246, p=0,267

Τα πολυπαραγοντικά μοντέλα AFT για την σχετική πιθανότητα επιβίωσης (μη θάνατος από CVD) παρουσιάζονται στον *Πίνακα 19*.

- Η **ηλικία** στα πολυπαραγοντικά μοντέλα παρέμεινε στατιστικά σημαντικός αρνητικός προγνωστικός παράγοντας, σε όλα τα μοντέλα, όπου για κάθε έτος παραπάνω η πιθανότητα επιβίωσης μειωνόταν κατά:
 - 5% στο Exponential μοντέλο ($e^{\beta} = 0,95$, 95% ΔE=0,933, 0,968, p<0,001),
 - 3,1% στο Weibull μοντέλο ($e^{\beta} = 0,969$, 95% ΔE= 0,963, 0,976, p<0,001),
 - 3,4% στο Loglogistic μοντέλο ($e^{\beta} = 0,966$, 95% ΔE= 0,958, 0,974, p<0,001) και

- 3,8% στο Lognormal μοντέλο ($e^{\beta} = 0,962$, 95% ΔΕ= 0,951, 0,972, $p < 0,001$),

ανεξάρτητα από την σωματική δραστηριότητα, το κάπνισμα, τη συστολική αρτηριακή πίεση, τον καρδιακό ρυθμό, την ολική χοληστερόλη και τον δείκτη μάζας σώματος.

- Η **σωματική δραστηριότητα** αποτέλεσε στατιστικά σημαντικός θετικός προγνωστικός παράγοντας, σε όλα τα μοντέλα (εκτός του Exponential), όπου για κάθε έτος παραπάνω η πιθανότητα επιβίωσης μειωνόταν κατά:

- 14,6% στο Exponential μοντέλο ($e^{\beta} = 1,146$, 95% ΔΕ= 0,919, 1,43, $p = 0,227$),
- 9% στο Weibull μοντέλο ($e^{\beta} = 1,09$, 95% ΔΕ= 1,003, 1,184, $p = 0,042$),
- 15,6% στο Loglogistic μοντέλο ($e^{\beta} = 1,156$, 95% ΔΕ= 1,044, 1,281, $p = 0,005$) και
- 20,4% στο Lognormal μοντέλο ($e^{\beta} = 1,204$, 95% ΔΕ= 1,055, 1,374, $p = 0,006$),

- Το **κάπνισμα** παρέμεινε μη στατιστικά σημαντικός προγνωστικός παράγοντας στο Exponential μοντέλο ($e^{\beta} = 0,81$, 95% ΔΕ= 0,648, 1,013, $p = 0,065$). Αντίθετα, στο Weibull μοντέλο φάνηκε ότι το κάπνισμα ήταν σημαντικός και αρνητικός προγνωστικός παράγοντας, όπου οι καπνιστές είχαν χαμηλότερη πιθανότητα επιβίωσης κατά 14,9% ($e^{\beta} = 0,851$, 95% ΔΕ= 0,783, 0,925, $p < 0,001$). Παρόμοια, στο Loglogistic μοντέλο οι καπνιστές είχαν χαμηλότερη πιθανότητα επιβίωσης κατά 14,6% ($e^{\beta} = 0,854$, 95% ΔΕ= 0,771, 0,946, $p = 0,002$). Ωστόσο, στο Lognormal μοντέλο το κάπνισμα αποτέλεσε μη στατιστικά σημαντικό και αρνητικό προγνωστικό παράγοντας ($e^{\beta} = 0,895$, 95% ΔΕ= 0,783, 1,022, $p = 0,103$).

- Η **συστολική αρτηριακή πίεση** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικός παράγοντας σε κανένα μοντέλο:

- Exponential μοντέλο: $e^{\beta} = 0,998$, 95% ΔΕ= 0,992, 1,003, $p = 0,370$
- Weibull μοντέλο: $e^{\beta} = 0,998$, 95% ΔΕ= 0,996, 1,00, $p = 0,089$
- Loglogistic μοντέλο: $e^{\beta} = 0,998$, 95% ΔΕ= 0,996, 1,001, $p = 0,141$

Lognormal μοντέλο: $e^{\beta} = 0,998$, 95% ΔE= 0,995, 1,001, p=0,266

- Ο **καρδιακός ρυθμός** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικό παράγοντα σε όλα τα μοντέλα ($p > 0,05$), εκτός του Loglogistic, όπου για κάθε bpm παραπάνω η πιθανότητα επιβίωσης μειωνόταν κατά 0,6% ($e^{\beta} = 0,994$, 95% ΔE= 0,99, 0,999, p=0,008).
- Η **ολική χοληστερόλη** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικό παράγοντα σε κανένα μοντέλο:
 - Exponential μοντέλο: $e^{\beta} = 1,001$, 95% ΔE= 0,999, 1,003, p=0,398
 - Weibull μοντέλο: $e^{\beta} = 1,001$, 95% ΔE= 1,000, 1,002, p=0,05
 - Loglogistic μοντέλο: $e^{\beta} = 1,001$, 95% ΔE= 1,000, 1,002, p=0,156
 - Lognormal μοντέλο: $e^{\beta} = 1,001$, 95% ΔE= 1,000, 1,003, p=0,057
- Όμοια, η κατηγοριοποίηση του **δείκτη μάζας σώματος** δεν αποτέλεσε στατιστικά σημαντικό προγνωστικό παράγοντα σε κανένα μοντέλο:
 - Exponential μοντέλο: $e^{\beta} = 1,046$, 95% ΔE= 0,817, 1,34, p=0,719
 - Weibull μοντέλο: $e^{\beta} = 1,022$, 95% ΔE= 0,932, 1,121, p=0,641
 - Loglogistic μοντέλο: $e^{\beta} = 1,089$, 95% ΔE= 0,973, 1,219, p=0,138
 - Lognormal μοντέλο: $e^{\beta} = 1,083$, 95% ΔE= 0,884, 1,182, p=0,982

Πίνακας 18. Μονοπαραγοντικά μοντέλα AFT για την σχετική πιθανότητα συνολικής επιβίωσης.

	Exponential				Weibull				Loglogistic				Lognormal			
	β	e ^β	95%CI	p	β	e ^β	95%CI	p	β	e ^β	95%CI	p	β	e ^β	95%CI	p
Ηλικία εισόδου (έτη)	-0,05	0,951	(0,935, 0,967)	<0,001	-0,031	0,969	(0,964, 0,976)	<0,001	-0,035	0,966	(0,958, 0,973)	<0,001	-0,044	0,957	(0,946, 0,968)	<0,001
Σωματική δραστηριότητα (ναι)	0,107	1,113	(0,915, 1,354)	0,286	0,055	1,057	(0,973, 1,146)	0,187	0,136	1,146	(1,033, 1,271)	0,010	0,226	1,254	(1,097, 1,432)	0,001
Κάπνισμα (ναι)	-0,183	0,833	(0,670, 1,036)	0,100	-0,134	0,875	(0,8, 0,955)	0,003	-0,145	0,865	(0,776, 0,965)	0,009	-0,096	0,908	(0,791, 1,044)	0,175
ΣΑΠ (mmHg)	-0,004	0,996	(0,991, 1,000)	0,066	-0,003	0,997	(0,996, 0,999)	0,006	-0,004	0,996	(0,994, 0,999)	0,002	-0,004	0,996	(0,993, 0,999)	0,004
ΔΑΠ (mmHg)	-0,006	0,994	(0,986, 1,003)	0,177	-0,004	0,996	(0,993, 1,00)	0,040	-0,005	0,995	(0,991, 0,999)	0,013	-0,008	0,992	(0,987, 0,998)	0,005
KP (bpm)	-0,001	0,999	(0,992, 1,006)	0,784	-0,001	0,999	(0,996, 1,002)	0,636	-0,003	0,997	(0,993, 1,001)	0,115	-0,005	0,995	(0,991, 1,00)	0,069
Χοληστερόλη (mg/dL)	0,001	1,001	(0,998, 1,003)	0,618	0,001	1,001	(0,999, 1,001)	0,491	0,001	1,001	(0,999, 1,002)	0,311	0,001	1,001	(1,000, 1,003)	0,041
ΔΜΣ (>25kg/m²)	0,043	1,044	(0,85, 1,281)	0,681	0,032	1,033	(0,949, 1,124)	0,452	0,055	1,057	(0,951, 1,172)	0,308	0,079	1,082	(0,941, 1,246)	0,267

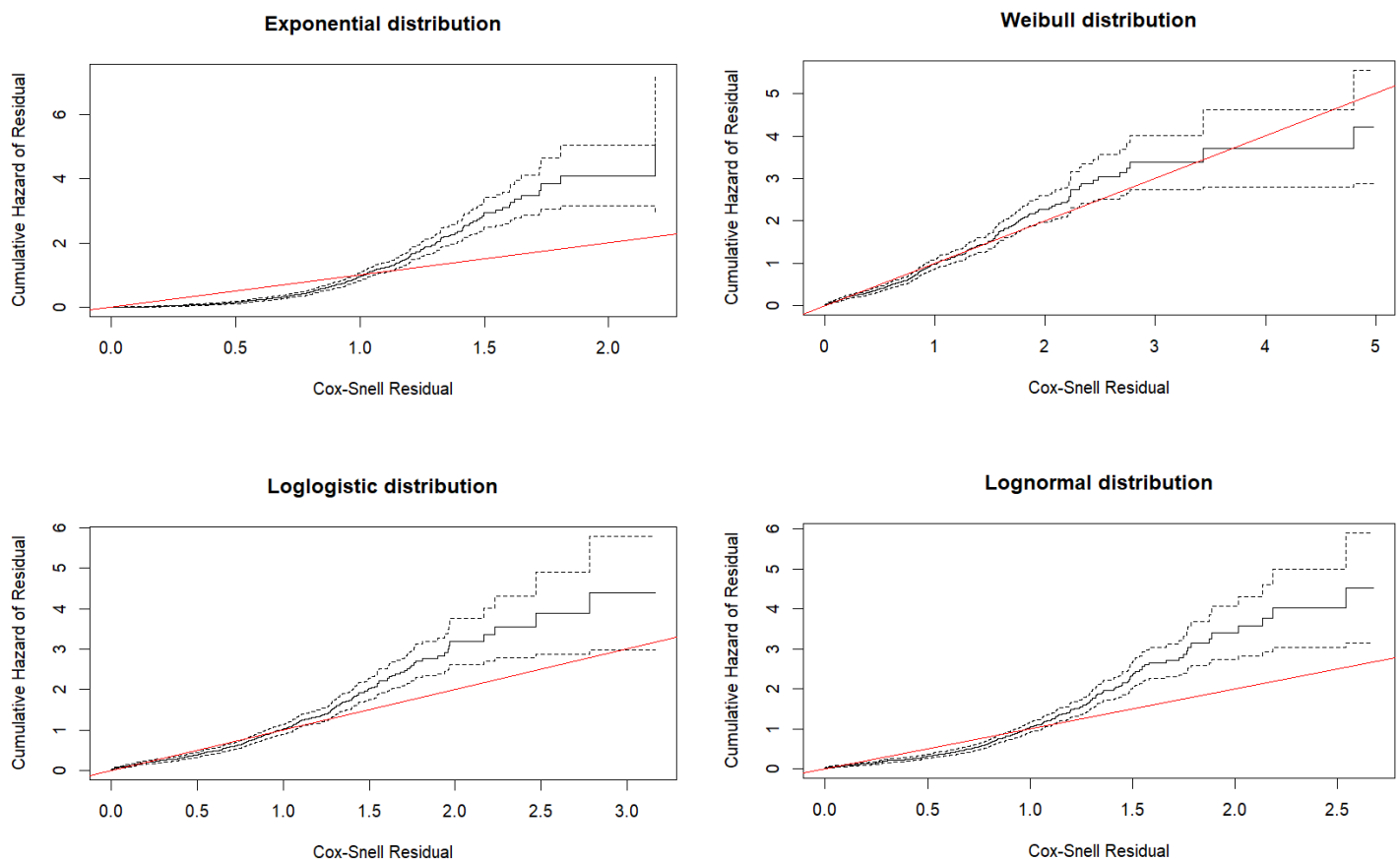
Συντομογραφίες. Διάστημα Εμπιστοσύνης (CI), Συστολική Αρτηριακή Πίεση (ΣΑΠ), Καρδιακός Ρυθμός (ΚΡ), Δείκτης Μάζας Σώματος (ΔΜΣ).

Πίνακας 19. Πολυπαραγοντικά μοντέλα AFT για την σχετική πιθανότητα συνολικής επιβίωσης.

	Exponential				Weibull				Loglogistic				Lognormal			
	β	e ^β	95%CI	p	β	e ^β	95%CI	p	β	e ^β	95%CI	p	β	e ^β	95%CI	P
Ηλικία εισόδου (έτη)	-0,051	0,95	(0,933, 0,968)	<0,001	-0,031	0,969	(0,963, 0,976)	<0,001	-0,035	0,966	(0,958, 0,974)	<0,001	-0,039	0,962	(0,951, 0,972)	<0,001
Σωματική δραστηριότητα (ναι)	0,136	1,146	(0,919, 1,43)	0,227	0,086	1,09	(1,003, 1,184)	0,042	0,145	1,156	(1,044, 1,281)	0,005	0,186	1,204	(1,055, 1,374)	0,006
Κάπνισμα (ναι)	-0,211	0,81	(0,648, 1,013)	0,065	-0,161	0,851	(0,783, 0,925)	<0,001	-0,158	0,854	(0,771, 0,946)	0,002	-0,111	0,895	(0,783, 1,022)	0,103
ΣΑΠ (mmHg)	-0,002	0,998	(0,992, 1,003)	0,370	-0,002	0,998	(0,996, 1,00)	0,089	-0,002	0,998	(0,996, 1,001)	0,141	-0,002	0,998	(0,995, 1,001)	0,266
KP (bpm)	-0,004	0,996	(0,988, 1,005)	0,391	-0,003	0,997	(0,994, 1,001)	0,112	-0,006	0,994	(0,99, 0,999)	0,008	-0,005	0,995	(0,99, 1,00)	0,065
Χοληστερόλη (mg/dL)	0,001	1,001	(0,999, 1,003)	0,398	0,001	1,001	(1,00, 1,002)	0,050	0,001	1,001	(1,00, 1,002)	0,156	0,001	1,001	(1,00, 1,003)	0,057
ΔΜΣ (>25kg/m²)	0,045	1,046	(0,817, 1,34)	0,719	0,022	1,022	(0,932, 1,121)	0,641	0,085	1,089	(0,973, 1,219)	0,138	0,08	1,083	(0,935, 1,254)	0,287

Συντομογραφίες. Διάστημα Εμπιστοσύνης (CI), Συστολική Αρτηριακή Πίεση (ΣΑΠ), Καρδιακός Ρυθμός (ΚΡ), Δείκτης Μάζας Σώματος (ΔΜΣ).

Στη συνέχεια αξιολογήθηκε η καλή προσαρμογή των μοντέλων μέσω του διαγράμματος καταλοίπων Cox-Snell. Αναλυτικότερα, παρουσιάζεται ένα διάγραμμα των καταλοίπων Cox-Snell (άξονας X) έναντι του εκτιμώμενου αθροιστικού κινδύνου αυτών (Cumulative Hazard of Cox-Snell residuals - άξονας Y) για κάθε πολυπαραγοντικό μοντέλο συνολικής επιβίωσης (Γράφημα 18). Από το Γράφημα παρατηρούμε ότι το πολυπαραγοντικό μοντέλο με Exponential κατανομή είχε την χειρότερη προσαρμογή, καθώς υπήρχε τεράστια απόκλιση από την ευθεία κόκκινη γραμμή. Παρόμοια, αλλά ελαφρώς καλύτερη προσαρμογή έδειξε να έχει τα πολυπαραγοντικά μοντέλα με Lognormal και Loglogistic κατανομή, όπου και πάλι στο γράφημα υπήρχε σημαντική απόκλιση από την ευθεία κόκκινη γραμμή. Την καλύτερη προσαρμογή με διαφορά την είχε το πολυπαραγοντικό μοντέλο με Weibull κατανομή, με το διάγραμμα καταλοίπων να είναι πολύ κοντά στην ευθεία κόκκινη γραμμή σε όλο σχεδόν το διάγραμμα.



Γράφημα 19. Διαγράμματα των καταλοίπων Cox-Snell (άξονας X) έναντι του εκτιμώμενου αθροιστικού κινδύνου αυτών (Cumulative Hazard of Cox-Snell residuals - άξονας Y) για τα πολυπαραγοντικά μοντέλα συνολικής επιβίωσης με Exponential, Weibull, Loglogistic και Lognormal κατανομές.

Στη συνέχεια συγκρίθηκαν όλα τα πολυπαραγοντικά AFT μοντέλα συνολικής επιβίωσης με βάσει το κριτήριο AIC, των οποίων τα αποτελέσματα παρουσιάζονται και στον *Πίνακα 20*. Όπως φαίνεται από τα αποτελέσματα, το καλύτερο μοντέλο ήταν το πολυπαραγοντικό AFT μοντέλο με Weibull κατανομή, με τιμή $AIC = 3173,374$, στη συνέχεια ακολουθούσε το πολυπαραγοντικό AFT μοντέλο με Loglogistic κατανομή, με τιμή $AIC = 3253,97$, το πολυπαραγοντικό AFT μοντέλο με Lognormal κατανομή, με τιμή $AIC = 3376,857$ και τέλος το πολυπαραγοντικό AFT μοντέλο με Exponential κατανομή, με τιμή $AIC = 3538,55$.

Πίνακας 20. Σύγκριση των AFT μοντέλων βάσει του AIC κριτηρίου.		
	Παράμετροι	AIC
Exponential	1	3538,55
Weibull	2	3173,374
Loglogistic	2	3253,97
Lognormal	2	3376,857

5. Συζήτηση

Τα ευρήματα της μελέτης ανέδειξαν την επιβαρυντική επίδραση της ηλικίας, της συστολικής και διαστολικής πίεσης και πιθανώς του καπνίσματος στην 40ετή θνησιμότητα από CVD, γεγονός όπου φάνηκε να ισχυρίζεται και το αρχικό άρθρο (ως αναφορά το κάπνισμα και την ηλικία) (Panagiotakos *et al.*, 2003). Επιπλέον, από τα AFT μοντέλα για την σχετική πιθανότητα συνολικής επιβίωσης (μη θάνατος), φάνηκε ότι η ηλικία εισόδου αποτελούσε σημαντικό και αρνητικό προγνωστικό παράγοντα για τη συνολική επιβίωση, όπως και το κάπνισμα. Αντίθετα, η σωματική δραστηριότητα αποτελούσε σημαντικό και θετικό προγνωστικό παράγοντα για τη συνολική επιβίωση.

Στη συγκεκριμένη ανάλυση συγκρίθηκαν διάφορα μοντέλα ανάλυσης επιβίωσης για την εύρεση του βέλτιστου από αυτά. Μεταξύ αυτών, χρησιμοποιήθηκαν καμπύλες Kaplan – Meier, μονοπαραγοντικά και πολυπαραγοντικά μοντέλα αναλογικών κινδύνων του Cox και μονοπαραγοντικά και πολυπαραγοντικά AFT μοντέλα με τη χρήση Exponential, Weibull, Loglogistic και Lognormal κατανομών. Από αυτά, φάνηκε ότι τα AFT μοντέλα υπερτερούσαν έναντι των μοντέλων αναλογικών κινδύνων του Cox. Επιπλέον, μεταξύ των AFT μοντέλων, είτε στην περίπτωση της επιβίωσης από CVD (μη θάνατος από CVD) είτε στην περίπτωση της συνολικής επιβίωσης (μη θάνατος), φάνηκε ότι το μοντέλο με την βέλτιστη εφαρμογή με διαφορά ήταν αυτό με Weibull κατανομή, είτε με το κριτήριο των γραφημάτων των Cox Snell καταλοίπων είτε με το κριτήριο AIC. Στη συνέχεια ακολουθούσε και στις δύο περιπτώσεις το Loglogistic μοντέλο, το Lognormal μοντέλο και τελευταίο το Exponential μοντέλο.

Παρότι στην πλειοψηφία των μελετών χρησιμοποιείται το μοντέλο αναλογικών κινδύνων του Cox, στη μεγάλη πλειοψηφία τους οι μελέτες δεν εξετάζουν αν πληρείται η προϋπόθεση της αναλογικότητας των κινδύνων και η καλή εφαρμογή του τελικού μοντέλου που παρουσιάζεται (Altman *et al.*, 1995). Το μοντέλο αναλογικών κινδύνων του Cox θεωρείται αναξιόπιστο ότι δεν πληρείται η προϋπόθεση της αναλογικότητας των κινδύνων, ενώ σε αυτή την περίπτωση θα πρέπει να χρησιμοποιούνται εναλλακτικά μοντέλα (Lalanne and Mesbah, 2016), όπως και στην περίπτωση της συγκεκριμένης μελέτης, στην οποία τα μοντέλα αναλογικών κινδύνων του Cox δεν έκαναν καλή εφαρμογή. Το AFT μοντέλο είναι ένα είδος παραμετρικού μοντέλου ανάλυσης επιβίωσης όπου αποτελεί εναλλακτική στο κλασσικό μοντέλο αναλογικών κινδύνων του Cox στην περίπτωση που δεν πληρούνται οι προϋποθέσεις του και δεν κάνει καλή εφαρμογή (Wei, 1992; Swindell, 2009). Επιπλέον, τα AFT μοντέλα οδηγούν σε περισσότερα αποδοτικές εκτιμήσεις παραμέτρων από το μοντέλο Cox κάτω από ορισμένες περιστάσεις,

όπως στο παράδειγμα κακής εφαρμογής του μοντέλου αναλογικών κινδύνων Cox (Oakes, 1977; Cox, 2018).

Το μοντέλο των αναλογικών κινδύνων του Cox χρησιμοποιείται ευρέως και έχει μελετηθεί εκτενώς, καθιστώντας το ένα τυπικό εργαλείο στην ανάλυση επιβίωσης. Ένα από τα κύρια πλεονεκτήματά του είναι ότι δεν απαιτεί υποθέσεις σχετικά με την υποκείμενη κατανομή των χρόνων επιβίωσης, επιτρέποντας την εφαρμογή του σε ένα ευρύ φάσμα συνόλων δεδομένων. Επιπλέον, επιτρέπει την εκτίμηση καμπυλών επιβίωσης με ταυτόχρονη προσαρμογή για επεξηγηματικές μεταβλητές, παρέχοντας πληροφορίες για το πώς διάφοροι παράγοντες επηρεάζουν την πιθανότητα επιβίωσης με την πάροδο του χρόνου. Ένα άλλο σημαντικό πλεονέκτημα του μοντέλου είναι η ικανότητά του να ενσωματώνει χρονικά εξαρτώμενες συμμεταβλητές, γεγονός που το καθιστά ιδιαίτερα ευέλικτο στο χειρισμό μεταβαλλόμενων παραγόντων κινδύνου. Παρά τα πλεονεκτήματά του, το μοντέλο έχει ορισμένους περιορισμούς. Ένα από τα σημαντικότερα μειονεκτήματα είναι ότι απαιτεί να ισχύει η υπόθεση των αναλογικών κινδύνων, δηλαδή ότι ο λόγος των κινδύνων μεταξύ των ομάδων παραμένει σταθερός με την πάροδο του χρόνου. Εάν η υπόθεση αυτή παραβιάζεται, το μοντέλο μπορεί να παράγει μεροληπτικές εκτιμήσεις και να οδηγήσει σε εσφαλμένα συμπεράσματα. Επιπλέον, το μοντέλο Cox εκφράζει το μέγεθος της επίδρασης των συμμεταβλητών σε όρους λόγων κινδύνου, οι οποίοι μπορεί να είναι λιγότερο διαισθητικοί και κλινικά σχετικοί σε σύγκριση με άλλα μέτρα, όπως ο χρόνος επιβίωσης ή οι λόγοι χρόνου. Αυτό καθιστά την ερμηνεία πιο δύσκολη, ιδίως για τους κλινικούς ιατρούς οι οποίοι μπορεί να θεωρούν ευκολότερο να κατανοήσουν τα αποτελέσματα που εκφράζονται σε όρους χρόνου παρά σε όρους ποσοστών κινδύνου (Qi, 2009; Swindell, 2009; Faruk, 2018).

Το μοντέλο AFT παρέχει έναν πιο άμεσο και κατατοπιστικό τρόπο ανάλυσης των δεδομένων επιβίωσης. Σε αντίθεση με το μοντέλο αναλογικών κινδύνων του Cox, το οποίο επικεντρώνεται στις αναλογίες κινδύνου, το μοντέλο AFT επιτρέπει την εκτίμηση των προβλεπόμενων συναρτήσεων κινδύνου, των συναρτήσεων επιβίωσης, των μέσων χρόνων επιβίωσης και των αναλογιών χρόνου. Αυτές οι πρόσθετες εκροές προσφέρουν μια πιο ολοκληρωμένη εικόνα των τάσεων επιβίωσης και της επίδρασης διαφόρων παραγόντων. Ένα άλλο βασικό πλεονέκτημα του μοντέλου AFT είναι ότι περιγράφει την επίδραση των μεταβλητών σε όρους ενός συντελεστή επιτάχυνσης ή ενός χρονικού λόγου, ο οποίος δείχνει πόσο πιο γρήγορα ή πιο αργά συμβαίνει ένα συμβάν σε μια ομάδα σε σύγκριση με μια άλλη. Αυτή η ερμηνεία είναι συχνά πιο ουσιαστική και σχετική για τους κλινικούς ιατρούς, καθώς σχετίζεται άμεσα με την

πρόγνωση των ασθενών και τα αποτελέσματα της θεραπείας. παρά τα πλεονεκτήματά του, το μοντέλο AFT είναι σχετικά άγνωστο και χρησιμοποιείται σπάνια σε σύγκριση με το μοντέλο Cox PH. Ένας σημαντικός περιορισμός είναι ότι απαιτεί να ισχύει η υπόθεση AFT, δηλαδή ότι η σχέση μεταξύ των συνδιαμεταβλητών και του χρόνου επιβίωσης ακολουθεί μια συγκεκριμένη λειτουργική μορφή. Εάν η υπόθεση αυτή παραβιάζεται, το μοντέλο ενδέχεται να μην παρέχει αξιόπιστα αποτελέσματα. Επιπλέον, το μοντέλο AFT απαιτεί τον προσδιορισμό μιας κατανομής για τον χρόνο επιβίωσης (π.χ. Weibull, Lognormal, Exponential), αλλά η επιλογή της κατάλληλης κατανομής μπορεί να αποτελέσει πρόκληση, ιδίως όταν η πραγματική κατανομή είναι άγνωστη (Qi, 2009; Swindell, 2009; Faruk, 2018).

Συμπερασματικά, το AFT μοντέλο αποτέλεσε καλύτερη εναλλακτική λύση από το μοντέλο αναλογικότητας των κινδύνων του Cox. Επιπλέον, από το σύνολο των AFT μοντέλων, την καλύτερη εφαρμογή φαίνεται να την έχει αυτό με Weibull κατανομή. Στις διαχρονικές μελέτες, όπου οι αναλύσεις που θα πρέπει να γίνουν θα ανήκουν στην ομπρέλα της ανάλυσης επιβίωσης, το τελικό μοντέλο θα πρέπει να επιλέγεται με μεγάλη προσοχή, καθώς ένα «κακό» μοντέλο που δεν εφαρμόζει καλά μπορεί να οδηγήσει σε αναξιόπιστα, ακόμη και τελείως διαφορετικά αποτελέσματα από τα πραγματικά.

Παράρτημα Α:

Κώδικας R

```
library(haven)
Corfu_data_cox_final_survANDhaz <- read_sav("Corfu_data_cox_final_survANDhaz.sav")
data = Corfu_data_cox_final_survANDhaz
rm(Corfu_data_cox_final_survANDhaz)

library(survival)
library(flexsurv)
#remove irrelevant columns
library(dplyr)
data = data %>% select(-allc_step_res)
data = data %>% select(- cvd_step_res, - allc_enter_res, - cvd_enter_res,
  - SUR_allc, - HAZ_allc, - SUR_cvd, - HAZ_cvd, - SUR_RATIO, - HAZ_RATIO)
data = data %>% select(- HAZ_1, - PR1_1, - PR2_1, - PR3_1,
  - PR4_1, - PR5_1, - PR6_1, - PR7_1, - SUR_1, - CumHaz)

#Running the Cox PH model
cox_model = coxph(Surv(years_alive, cvd_mortality) ~ age_entr + physical_act +
  smoking + sbp + heartrate + tchol + BMI_group, data = data,
  ties = "breslow")
summary(cox_model)

#all analyses and graphs were done in SPSS regarding Cox PH model

#Running the AFT models for CVD mortality
#Exponential distribution
aft_model_exp <- survreg(Surv(years_alive, cvd_mortality) ~ age_entr + physical_act +
  smoking + sbp + heartrate + tchol + BMI_group,
  data = data, dist = "exponential")

#Weibull distribution
aft_model_w <- survreg(Surv(years_alive, cvd_mortality) ~ age_entr + physical_act +
  smoking + sbp + heartrate + tchol + BMI_group,
  data = data, dist = "weibull")

#Loglogistic distribution
aft_model_ll <- survreg(Surv(years_alive, cvd_mortality) ~ age_entr + physical_act +
  smoking + sbp + heartrate + tchol + BMI_group,
  data = data, dist = "loglogistic")

#Lognormal distribution
aft_model_ln <- survreg(Surv(years_alive, cvd_mortality) ~ age_entr + physical_act +
  smoking + sbp + heartrate + tchol + BMI_group,
  data = data, dist = "lognormal")

#Summary of each model
```

```

summary(aft_model_exp)
summary(aft_model_w)
summary(aft_model_ll)
summary(aft_model_ln)

#AIC of each model
AIC(aft_model_exp)
AIC(aft_model_w)
AIC(aft_model_ll)
AIC(aft_model_ln)

#Creating the Cox-Snell residuals plot for each model
#Exponential distribution
fitg.exp <- flexsurvreg(formula = Surv(years_alive, cvd_mortality) ~ age_entr +
                        physical_act + smoking + sbp + heartrate + tchol +
                        BMI_group, data = data, dist = "exponential")
cs.exp <- coxsnell_flexsurvreg(fitg.exp)
## Model appears to fit well, with some small sample noise
surv.exp <- survfit(Surv(cs.exp$est, cs.exp$status) ~ 1)
plot(surv.exp, fun="cumhaz",
     xlab = "Cox-Snell Residual",
     ylab = "Cumulative Hazard of Residual",
     main = "Exponential distribution")
abline(0, 1, col="red")

#Weibull distribution
fitg.w <- flexsurvreg(formula = Surv(years_alive, cvd_mortality) ~ age_entr +
                      physical_act + smoking + sbp + heartrate + tchol +
                      BMI_group, data = data, dist = "Weibull")
cs.w <- coxsnell_flexsurvreg(fitg.w)
## Model appears to fit well, with some small sample noise
surv.w <- survfit(Surv(cs.w$est, cs.w$status) ~ 1)
plot(surv.w, fun="cumhaz",
     xlab = "Cox-Snell Residual",
     ylab = "Cumulative Hazard of Residual",
     main = "Weibull distribution")
abline(0, 1, col="red")

#Loglogistic distribution
fitg.ll <- flexsurvreg(formula = Surv(years_alive, cvd_mortality) ~ age_entr +
                      physical_act + smoking + sbp + heartrate + tchol +
                      BMI_group, data = data, dist = "llogis")
cs.ll <- coxsnell_flexsurvreg(fitg.ll)
## Model appears to fit well, with some small sample noise

```

```

surv.ll <- survfit(Surv(cs.ll$est, cs.ll$status) ~ 1)
plot(surv.ll, fun="cumhaz",
     xlab = "Cox-Snell Residual",
     ylab = "Cumulative Hazard of Residual",
     main = "Loglogistic distribution")
abline(0, 1, col="red")

```

```

#Lognormal distribution
fitg.ln <- flexsurvreg(formula = Surv(years_alive, cvd_mortality) ~ age_entr +
                      physical_act + smoking + sbp + heartrate + tchol +
                      BMI_group, data = data, dist = "lnorm")
cs.ln <- coxsnell_flexsurvreg(fitg.ln)
## Model appears to fit well, with some small sample noise
surv.ln <- survfit(Surv(cs.ln$est, cs.ln$status) ~ 1)
plot(surv.ln, fun="cumhaz",
     xlab = "Cox-Snell Residual",
     ylab = "Cumulative Hazard of Residual",
     main = "Lognormal distribution")
abline(0, 1, col="red")

```

```

#Running the AFT models for all cause mortality
#Exponential distribution
aft_model_exp <- survreg(Surv(years_alive, allc_mortality) ~ age_entr + physical_act +
  smoking + sbp + heartrate + tchol + BMI_group,
  data = data, dist = "exponential")

#Weibull distribution
aft_model_w <- survreg(Surv(years_alive, allc_mortality) ~ age_entr + physical_act +
  smoking + sbp + heartrate + tchol + BMI_group,
  data = data, dist = "weibull")

#Loglogistic distribution
aft_model_ll <- survreg(Surv(years_alive, allc_mortality) ~ age_entr + physical_act +
  smoking + sbp + heartrate + tchol + BMI_group,
  data = data, dist = "loglogistic")

#Lognormal distribution
aft_model_ln <- survreg(Surv(years_alive, allc_mortality) ~ age_entr + physical_act +
  smoking + sbp + heartrate + tchol + BMI_group,
  data = data, dist = "lognormal")

#Summary of each model
summary(aft_model_exp)
summary(aft_model_w)
summary(aft_model_ll)
summary(aft_model_ln)

#AIC of each model
AIC(aft_model_exp)
AIC(aft_model_w)
AIC(aft_model_ll)
AIC(aft_model_ln)

#Creating the Cox-Snell residuals plot for each model
#Exponential distribution
fitg.exp2 <- flexsurvreg(formula = Surv(years_alive, allc_mortality) ~ age_entr +
  physical_act + smoking + sbp + heartrate + tchol +
  BMI_group, data = data, dist = "exponential")
cs.exp <- coxsnell_flexsurvreg(fitg.exp2)
## Model appears to fit well, with some small sample noise
surv.exp <- survfit(Surv(cs.exp$est, cs.exp$status) ~ 1)
plot(surv.exp, fun="cumhaz",
  xlab = "Cox-Snell Residual",
  ylab = "Cumulative Hazard of Residual",
  main = "Exponential distribution")
abline(0, 1, col="red")

```

```

#Weibull distribution
fitg.w2 <- flexsurvreg(formula = Surv(years_alive, allc_mortality) ~ age_entr +
  physical_act + smoking + sbp + heartrate + tchol +
  BMI_group, data = data, dist = "Weibull")
cs.w <- coxsnell_flexsurvreg(fitg.w2)
## Model appears to fit well, with some small sample noise
surv.w <- survfit(Surv(cs.w$est, cs.w$status) ~ 1)
plot(surv.w, fun="cumhaz",
  xlab = "Cox-Snell Residual",
  ylab = "Cumulative Hazard of Residual",
  main = "Weibull distribution")
abline(0, 1, col="red")

```

```

#Loglogistic distribution
fitg.ll2 <- flexsurvreg(formula = Surv(years_alive, allc_mortality) ~ age_entr +
  physical_act + smoking + sbp + heartrate + tchol +
  BMI_group, data = data, dist = "llogis")
cs.ll <- coxsnell_flexsurvreg(fitg.ll2)
## Model appears to fit well, with some small sample noise
surv.ll <- survfit(Surv(cs.ll$est, cs.ll$status) ~ 1)
plot(surv.ll, fun="cumhaz",
  xlab = "Cox-Snell Residual",
  ylab = "Cumulative Hazard of Residual",
  main = "Loglogistic distribution")
abline(0, 1, col="red")

```

```

#Lognormal distribution
fitg.ln2 <- flexsurvreg(formula = Surv(years_alive, allc_mortality) ~ age_entr +
  physical_act + smoking + sbp + heartrate + tchol +
  BMI_group, data = data, dist = "lnorm")
cs.ln <- coxsnell_flexsurvreg(fitg.ln2)
## Model appears to fit well, with some small sample noise
surv.ln <- survfit(Surv(cs.ln$est, cs.ln$status) ~ 1)
plot(surv.ln, fun="cumhaz",
  xlab = "Cox-Snell Residual",
  ylab = "Cumulative Hazard of Residual",
  main = "Lognormal distribution")
abline(0, 1, col="red")

```

6. Βιβλιογραφία

Lee, C. Y., & Im, E. O. (2021). Socioeconomic Disparities in Cardiovascular Health in South Korea: A Systematic Review. *The Journal of cardiovascular nursing*, 36(1), 8–22.

<https://doi.org/10.1097/JCN.0000000000000624>

Poulimeneas, D. *et al.* (2020) ‘Exploring the relationship between the Mediterranean diet and weight loss maintenance: the MedWeight study’, *British Journal of Nutrition*, 124(8), pp. 874–880. doi:10.1017/S0007114520001798.

Hooper L. (2007). Primary prevention of CVD: diet and weight loss. *BMJ clinical evidence*, 2007, 0219. Available at: <https://pmc.ncbi.nlm.nih.gov/articles/PMC2943801/>

Bloom, D. E., *et al.* (2011). The Global Economic Burden of Noncommunicable Diseases. Geneva: World Economic Forum. available at: <https://www.weforum.org/publications/global-economic-burden-non-communicable-diseases/>

Mendis, S., *et al.* (2019). *Cardiovascular Disease: Risk Factors and Prevention*. WHO Report. pp. 38-49. Available at: <https://DOI:10.4038/ajim.v1i1.30>

Mensah, G, Fuster, V, Murray, C. *et al.* Global Burden of Cardiovascular Diseases and Risks, 1990-2022. *JACC*. 2023 Dec, 82 (25) 2350–2473. Available at: <https://doi.org/10.1016/j.jacc.2023.11.007>

Altman, D.G. *et al.* (1995) ‘Review of survival analyses published in cancer journals’, *British Journal of Cancer*, 72(2), pp. 511–518. Available at: <https://doi.org/10.1038/bjc.1995.364>.

Cox, D.R. (2018) *Analysis of Survival Data*. New York: Chapman and Hall/CRC. Available at: <https://doi.org/10.1201/9781315137438>.

Faruk, A. (2018) ‘The comparison of proportional hazards and accelerated failure time models in analyzing the first birth interval survival data’, *Journal of Physics: Conference Series*, 974(1), p. 012008. Available at: <https://doi.org/10.1088/1742-6596/974/1/012008>.

Field, A. (2018) *Discovering statistics using IBM® SPSS® statistics*. North American edition, 5th edition. Edited by J. Seaman. Los Angeles London New Delhi [und 3 weitere]: SAGE.

FHS (2025) *Framingham Heart Study*. Available at: <https://www.framinghamheartstudy.org/> (Accessed: 7 March 2025).

Finegold, J.A., Asaria, P. and Francis, D.P. (2013) ‘Mortality from ischaemic heart disease by country, region, and age: Statistics from World Health Organisation and United Nations’, *International Journal of Cardiology*, 168(2), pp. 934–945. Available at: <https://doi.org/10.1016/j.ijcard.2012.10.046>.

Lalanne, C. and Mesbah, M. (2016) ‘5 - Survival Data Analysis’, in C. Lalanne and M. Mesbah (eds) *Biostatistics and Computer-based Analysis of Health Data using Stata*. Elsevier, pp. 101–115. Available at: <https://doi.org/10.1016/B978-1-78548-142-0.50005-5>.

Michas, G., Karvelas, G. and Trikas, A. (2019) ‘Cardiovascular disease in Greece; the latest evidence on risk factors’, *Hellenic Journal of Cardiology*, 60(5), pp. 271–275. Available at: <https://doi.org/10.1016/j.hjc.2018.09.006>.

Oakes, D. (1977) ‘The Asymptotic Information in Censored Survival Data’, *Biometrika*, 64(3), pp. 441–448. Available at: <https://doi.org/10.2307/2345319>.

Panagiotakos, D.B. *et al.* (2003) ‘Forty-years (1961-2001) of all-cause and coronary heart disease mortality and its determinants: the Corfu cohort from the Seven Countries Study’, *International Journal of Cardiology*, 90(1), pp. 73–79. Available at: [https://doi.org/10.1016/s0167-5273\(02\)00526-0](https://doi.org/10.1016/s0167-5273(02)00526-0).

Qi, J. (2009) *Comparison of Proportional Hazards and Accelerated Failure Time Models*. Master. University of Saskatchewan. Available at: <https://core.ac.uk/download/pdf/226159159.pdf>.

Swindell, W.R. (2009) ‘ACCELERATED FAILURE TIME MODELS PROVIDE A USEFUL STATISTICAL FRAMEWORK FOR AGING RESEARCH’, *Experimental gerontology*, 44(3), pp. 190–200. Available at: <https://doi.org/10.1016/j.exger.2008.10.005>.

Wei, L.J. (1992) ‘The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis’, *Statistics in Medicine*, 11(14–15), pp. 1871–1879. Available at: <https://doi.org/10.1002/sim.4780111409>.

Zhao, D. *et al.* (2024) ‘Impact of Aging on Cardiovascular Diseases: From Chronological Observation to Biological Insights: JACC Family Series’, *JACC: Asia*, 4(5), pp. 345–358. Available at: <https://doi.org/10.1016/j.jacasi.2024.02.002>.

Εργαζάκη, Α. (2024) *Παραμετρικά μοντέλα επιβίωσης ασθενών για διάφορα είδη καρκίνου*. Master Thesis. Πανεπιστήμιο Πειραιώς. Available at: <https://dione.lib.unipi.gr/xmlui/handle/unipi/16797> (Accessed: 11 March 2025).

NIH (2025) *Framingham Heart Study (FHS) | NHLBI, NIH*. Available at: <https://www.nhlbi.nih.gov/science/framingham-heart-study-fhs> (Accessed: 7 March 2025).