

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

**Ανάλυση Επιβίωσης από κάθε αιτία και καρδιαγγειακή νόσο,
ανδρών μέσης ηλικίας, στην περιοχή της Κέρκυρας.
(1960-2000)**

ΚΑΡΤΣΑΚΛΗ ΚΩΝΣΤΑΝΤΙΝΑ

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 2022

UNIVERSITY OF PIRAEUS



DEPARTMENT OF STATISTICS AND INSURANCE SCIENCE

POSTGRADUATE PROGRAM IN APPLIED STATISTICS

**Survival analysis by cause of death and cardiovascular disease for
middle age men from the area of Corfu.
(1960-2000)**

By

KARTSAKLI KONSTANTINA

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 2022

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίσθηκε 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)
- Καθηγήτρια Βερροπούλου Γεωργία
- Καθηγητής Χαροκόποιου Πανεπιστημίου Παναγιωτάκος Δημοσθένης

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

ΕΥΧΑΡΙΣΤΙΕΣ

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

Ακόμη, θα ήθελα να ευχαριστήσω τους γονείς μου και την αδερφή μου για την συμπαράσταση και τη στήριξή τους όλους αυτούς τους μήνες της προετοιμασίας.

Κωνσταντίνα Καρτσακλή,

Αθήνα, 2022

ΠΕΡΙΛΗΨΗ

Τα καρδιαγγειακά νοσήματα αφορούν παθήσεις της καρδιάς (του καρδιακού μυός) ή των αιμοφόρων αγγείων. Ωστόσο, όταν χρησιμοποιούμε τον όρο καρδιαγγειακή νόσος αναφερόμαστε συνήθως σε παθήσεις της καρδιάς ή των αιμοφόρων αγγείων που προκαλούνται από το αθήρωμα.

Η καρδιαγγειακή νόσος αποτελεί μια από τις κύριες αιτίες θανάτου, τόσο στον αναπτυγμένο όσο και στον αναπτυσσόμενο κόσμο. Βασικός στόχος της συγκεκριμένης διπλωματικής εργασίας είναι η ανάλυση επιβίωσης από διάφορες αιτίες και καρδιαγγειακά νοσήματα, σε ένα δείγμα 529 ανδρών μέσης ηλικίας, στην περιοχή της Κέρκυρας. Δηλαδή, εξετάζονται οι προγνωστικοί παράγοντες που παίζουν ρόλο στην επιβίωση ή μη των ατόμων από στεφανιαία νόσο ή άλλες αιτίες.

Οι βασικές στατιστικές μέθοδοι χρησιμοποιήθηκαν για την έρευνά μας είναι το μοντέλο αναλογικού κινδύνου του Cox, η πολλαπλή γραμμική παλινδρόμηση καθώς και η λογιστική παλινδρόμηση. Η πολλαπλή γραμμική παλινδρόμηση αποδείχθηκε ανεπαρκής για την μελέτη μας αφού δεν πληρούνταν οι προϋποθέσεις εφαρμογής της. Σύμφωνα όμως με τα άλλα δύο μοντέλα, οι βασικοί προγνωστικοί παράγοντες που παίζουν ρόλο στην επιβίωση ή μη των ατόμων από καρδιαγγειακά νοσήματα είναι η ηλικία εισόδου στην έρευνα καθώς και η συστολική αρτηριακή πίεση ενώ οι βασικοί προγνωστικοί παράγοντες που παίζουν ρόλο στην επιβίωση ή μη των ατόμων από διάφορες αιτίες είναι η ηλικία εισόδου στην έρευνα, το κάπνισμα καθώς και η ολική χοληστερόλη. Τέλος, συγκρίνονται και σχολιάζονται οι συναρτήσεις κινδύνου και επιβίωσης για αυτούς που πέθαναν από καρδιαγγειακά και για αυτούς που πέθαναν από διάφορες αιτίες.

ABSTRACT

Cardiovascular diseases refer to diseases of the heart (heart muscle) or blood vessels. However, when we use the term cardiovascular disease we usually refer to diseases of the heart or blood vessels caused by atheroma. Cardiovascular disease is one of the leading causes of death in both the developed and developing world.

The main objective of this thesis is the analysis of mortality from various causes and cardiovascular diseases, in a sample of 529 middle-aged men, in the area of Corfu. That is, prognostic factors that play a role in the survival or non-survival of individuals from cardiovascular or other causes are examined.

The main statistical methods used for our research are the Cox proportional hazard model, multiple linear regression as well as logistic regression. Multiple linear regression proved insufficient for our study since its conditions were not met. However, according to the other two models, the main prognostic factors that play a role in the survival or non-survival of people from cardiovascular diseases are the age of entry into the research as well as the systolic blood pressure. Also, the main prognostic factors that play a role in the survival or non-survival of individuals from various causes are the age of entry into the research, smoking as well as total cholesterol. Finally, the hazard and survival functions for those who died from cardiovascular and those who died from various causes are compared and commented.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ	10
ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ	14
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ	16

1^ο ΚΕΦΑΛΑΙΟ

1 ΕΙΣΑΓΩΓΗ	17
1.1 ΕΙΣΑΓΩΓΗ	17
1.1.1 ΑΙΤΙΕΣ ΘΑΝΑΤΟΥ ΣΕ ΑΤΟΜΑ ΜΕΣΗΣ ΗΛΙΚΙΑΣ.....	17
1.2 ΚΑΡΔΙΑΓΓΕΙΑΚΑ ΝΟΣΗΜΑΤΑ	18
1.2.1 ΕΙΔΗ ΚΑΡΔΙΑΓΓΕΙΑΚΩΝ ΝΟΣΗΜΑΤΩΝ	19
1.2.2 ΠΑΡΑΓΟΝΤΕΣ ΚΙΝΔΥΝΟΥ ΕΜΦΑΝΙΣΗΣ ΚΑΡΔΙΑΓΓΕΙΑΚΩΝ ΝΟΣΗΜΑΤΩΝ	20
1.2.3 ΜΕΤΡΑ ΠΡΟΛΗΨΗΣ ΚΑΡΔΙΑΓΓΕΙΑΚΩΝ ΝΟΣΗΜΑΤΩΝ	21
1.3 ΔΙΑΤΡΟΦΗ ΚΑΙ ΠΡΟΛΗΨΗ ΚΑΡΔΙΑΓΓΕΙΑΚΩΝ ΠΑΘΗΣΕΩΝ	22
1.3.1 ΜΕΣΟΓΕΙΑΚΗ ΔΙΑΤΡΟΦΗ ΚΑΙ ΥΓΕΙΑ ΤΗΣ ΚΑΡΔΙΑΣ	22
1.3.2 ΣΥΣΤΑΤΙΚΑ ΤΗΣ ΔΙΑΤΡΟΦΗΣ ΣΤΗΝ ΠΡΟΛΗΨΗ ΤΗΣ ΚΑΡΔΙΑΓΓΕΙΑΚΗΣ ΝΟΣΟΥ	23
1.4 ΚΑΡΔΙΑΓΓΕΙΑΚΕΣ ΠΑΘΗΣΕΙΣ ΣΤΗΝ ΕΛΛΑΔΑ	24
1.5 ΜΕΓΑΛΕΣ ΜΕΛΕΤΕΣ ΤΗΣ ΚΑΡΔΙΟΛΟΓΙΚΗΣ ΕΠΙΔΗΜΙΟΛΟΓΙΑΣ	25
1.5.1 Η ΜΕΛΕΤΗ ΤΩΝ ΕΠΤΑ ΧΩΡΩΝ	26
1.6 ΘΑΝΑΤΟΙ ΣΤΗΝ ΚΕΡΚΥΡΑ	26

2^ο ΚΕΦΑΛΑΙΟ

2 ΠΕΡΙΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ	28
2.1 ΕΙΣΑΓΩΓΗ	28
2.1.1 ΠΕΡΙΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ	28
2.1.1.1 ΠΟΣΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ	30
2.1.1.2 ΠΟΙΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ	45

3° ΚΕΦΑΛΑΙΟ

3	ΣΥΣΧΕΤΙΣΕΙΣ.....	50
3.1	ΕΙΣΑΓΩΓΗ	50
3.1.1	ΕΛΕΓΧΟΣ χ^2	50
3.1.2	ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ ΤΟΥ PEARSON	51
3.1.3	ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ ΤΟΥ SPEARMAN	52
3.1.4	ΕΛΕΓΧΟΣ ΓΙΑ ΤΗ ΔΙΑΦΟΡΑ ΜΕΣΩΝ ΔΥΟ ΑΝΕΞΑΡΤΗΤΩΝ ΔΕΙΓΜΑΤΩΝ	53
3.1.5	ΕΛΕΓΧΟΣ ΓΙΑ ΤΟ ΑΝ ΔΥΟ ΑΝΕΞΑΡΤΗΤΑ ΔΕΙΓΜΑΤΑ ΠΡΟΕΡΧΟΝΤΑΙ ΑΠΟ ΤΗΝ ΙΔΙΑ ΚΑΤΑΝΟΜΗ (MANN-WHITNEY U)	54
3.2	ΕΛΕΓΧΟΣ ΣΥΣΧΕΤΙΣΗΣ ΠΟΣΟΣΤΙΚΩΝ ΜΕΤΑΒΛΗΤΩΝ	55
3.3	ΕΛΕΓΧΟΣ ΣΥΣΧΕΤΙΣΗΣ ΠΟΙΟΤΙΚΩΝ ΜΕΤΑΒΛΗΤΩΝ	58
3.4	ΕΛΕΓΧΟΣ ΣΥΣΧΕΤΙΣΗΣ ΠΟΙΟΤΙΚΩΝ ΜΕ ΠΟΣΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ	62
3.5	ΣΥΝΟΠΤΙΚΗ ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΜΕΤΑΒΛΗΤΩΝ ΠΟΥ ΣΥΣΧΕΤΙΖΟΝΤΑΙ ΜΕ ΤΙΣ ΜΕΤΑΒΛΗΤΕΣ “ALLC_MORTALITY” ΚΑΙ “CVD_MORTALITY”	66

4° ΚΕΦΑΛΑΙΟ

4	ΜΟΝΤΕΛΟ ΤΟΥ COX	77
4.1	ΕΙΣΑΓΩΓΗ	77
4.2	ΑΝΑΛΥΣΗ ΕΠΙΒΙΩΣΗΣ	77
4.3	ΣΥΝΑΡΤΗΣΗ ΕΠΙΒΙΩΣΗΣ	77
4.4	ΣΥΝΑΡΤΗΣΗ ΚΙΝΔΥΝΟΥ	78
4.5	ΜΟΝΤΕΛΟ ΑΝΑΛΟΓΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΚΙΝΔΥΝΟΥ	78
4.6	ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΔΙΑΦΟΡΕΣ ΑΙΤΙΕΣ (allc_mortality)	79
4.7	ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΚΑΡΔΙΑΓΓΕΙΑΚΑ ΝΟΣΗΜΑΤΑ (cvd_mortality)	82
4.8	ΣΥΓΚΡΙΣΗ ΤΩΝ ΓΡΑΦΙΚΩΝ ΠΑΡΑΣΤΑΣΕΩΝ ΤΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΕΠΙΒΙΩΣΗΣ ΚΑΙ ΤΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΚΙΝΔΥΝΟΥ ΓΙΑ ΤΟΥΣ ΔΥΟ ΠΛΗΘΥΣΜΟΥΣ ΠΟΥ ΟΡΙΖΟΝΤΑΙ ΑΠΟ ΤΙΣ ΜΕΤΑΒΛΗΤΕΣ ALLC_MORTALITY ΚΑΙ CVD_MORTALITY	86
4.9	ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΓΡΑΦΙΚΩΝ ΠΑΡΑΣΤΑΣΕΩΝ ΤΩΝ ΜΕΣΩΝ ΤΙΜΩΝ ΤΩΝ ΛΟΓΩΝ ΤΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΕΠΙΒΙΩΣΗΣ ΚΑΙ ΤΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΚΙΝΔΥΝΟΥ	89

5° ΚΕΦΑΛΑΙΟ

5	ΠΟΛΛΑΠΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗ	91
5.1	ΕΙΣΑΓΩΓΗ	91
5.2	ΠΟΛΛΑΠΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗ	91
5.3	ΜΕΘΟΔΟΣ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ	92
5.4	ΥΠΟΘΕΣΕΙΣ ΤΟΥ ΠΟΛΛΑΠΛΟΥ ΓΡΑΜΜΙΚΟΥ ΜΟΝΤΕΛΟΥ	92
5.5	ΜΕΘΟΔΟΣ ΤΗΣ STEPWISE REGRESSION	92
5.6	ΜΕΘΟΔΟΣ ENTER REGRESSION	97

6° ΚΕΦΑΛΑΙΟ

6	ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ	102
6.1	ΕΙΣΑΓΩΓΗ	102
6.2	ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ	102
6.3	ΠΟΛΛΑΠΛΟ ΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ	102
6.4	ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΩΝ	103
6.5	ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗ	104
6.5.1	Η ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΚΑΡΔΙΑΓΓΕΙΑΚΑ ΠΡΟΒΛΗΜΑΤΑ ΩΣ ΕΞΑΡΤΗΜΕΝΗ ΜΕΤΑΒΛΗΤΗ ΤΗΣ ΛΟΓΙΣΤΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ	104
6.5.2	Η ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΔΙΑΦΟΡΕΣ ΑΙΤΙΕΣ ΩΣ ΕΞΑΡΤΗΜΕΝΗ ΜΕΤΑΒΛΗΤΗ ΤΗΣ ΛΟΓΙΣΤΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ	107

7° ΚΕΦΑΛΑΙΟ

7	ΣΥΜΠΕΡΑΣΜΑΤΑ	110
7.1	ΕΙΣΑΓΩΓΗ	110
7.2	ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΠΕΡΙΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ	110
7.3	ΣΥΣΧΕΤΙΣΕΙΣ	111
7.4	ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΙΣ ΑΝΑΛΥΣΕΙΣ	112
	ΒΙΒΛΙΟΓΡΑΦΙΑ	114

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

2^ο ΚΕΦΑΛΑΙΟ

ΠΙΝΑΚΑΣ 2.1: ΠΕΡΙΓΡΑΦΙΚΑ ΣΤΑΤΙΣΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΓΙΑ ΤΗΝ ΜΕΤΑΒΛΗΤΗ “ΗΛΙΚΙΑ ΕΙΣΟΔΟΥ ΣΤΗΝ ΕΡΕΥΝΑ”.....	30
ΠΙΝΑΚΑΣ 2.2: ΠΕΡΙΓΡΑΦΙΚΑ ΣΤΑΤΙΣΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΓΙΑ ΤΗΝ ΜΕΤΑΒΛΗΤΗ “ΧΡΟΝΟΣ ΠΑΡΑΜΟΝΗΣ ΣΤΗΝ ΕΡΕΥΝΑ”	31
ΠΙΝΑΚΑΣ 2.3: ΥΠΟΛΟΓΙΣΜΟΣ ΣΥΧΝΟΤΗΤΩΝ ΤΩΝ ΤΙΜΩΝ ΤΗΣ “FU_NEW”	32
ΠΙΝΑΚΑΣ 2.4: ΠΕΡΙΓΡΑΦΙΚΑ ΣΤΑΤΙΣΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΓΙΑ ΤΗΝ ΜΕΤΑΒΛΗΤΗ “ΥΨΟΣ”	33
ΠΙΝΑΚΑΣ 2.5: ΠΕΡΙΓΡΑΦΙΚΑ ΣΤΑΤΙΣΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΓΙΑ ΤΗΝ ΜΕΤΑΒΛΗΤΗ “ΚΙΛΑ”..	34
ΠΙΝΑΚΑΣ 2.6: ΠΕΡΙΓΡΑΦΙΚΑ ΣΤΑΤΙΣΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΓΙΑ ΤΗΝ ΜΕΤΑΒΛΗΤΗ “ΣΥΣΤΟΛΙΚΗ ΑΡΤΗΡΙΑΚΗ ΠΙΕΣΗ”	35
ΠΙΝΑΚΑΣ 2.7: ΠΕΡΙΓΡΑΦΙΚΑ ΣΤΑΤΙΣΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΓΙΑ ΤΗΝ ΜΕΤΑΒΛΗΤΗ “ΔΙΑΣΤΟΛΙΚΗ ΑΡΤΗΡΙΑΚΗ ΠΙΕΣΗ”	36
ΠΙΝΑΚΑΣ 2.8: ΠΕΡΙΓΡΑΦΙΚΑ ΣΤΑΤΙΣΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΓΙΑ ΤΗΝ ΜΕΤΑΒΛΗΤΗ “ΚΑΡΔΙΑΚΟΣ ΡΥΘΜΟΣ”	37
ΠΙΝΑΚΑΣ 2.9: ΠΕΡΙΓΡΑΦΙΚΑ ΣΤΑΤΙΣΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΓΙΑ ΤΗΝ ΜΕΤΑΒΛΗΤΗ “ΟΛΙΚΗ ΧΟΛΗΣΤΕΡΟΛΗ”	38
ΠΙΝΑΚΑΣ 2.10: ΠΕΡΙΓΡΑΦΙΚΑ ΣΤΑΤΙΣΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΓΙΑ ΤΗΝ ΜΕΤΑΒΛΗΤΗ “ΔΕΙΚΤΗΣ ΜΑΖΑΣ ΣΩΜΑΤΟΣ	39
ΠΙΝΑΚΑΣ 2.11: ΠΕΡΙΓΡΑΦΙΚΑ ΣΤΑΤΙΣΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΓΙΑ ΤΗΝ ΜΕΤΑΒΛΗΤΗ “ΗΛΙΚΙΑ ΘΑΝΑΤΟΥ	41
ΠΙΝΑΚΑΣ 2.12: ΠΕΡΙΓΡΑΦΙΚΑ ΣΤΑΤΙΣΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΓΙΑ ΤΗΝ ΜΕΤΑΒΛΗΤΗ “ΗΛΙΚΙΑ ΘΑΝΑΤΟΥ”	42
ΠΙΝΑΚΑΣ 2.13: ΠΙΝΑΚΑΣ ΣΥΧΝΟΤΗΤΩΝ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΔΙΑΦΟΡΕΣ ΑΙΤΙΕΣ”	45
ΠΙΝΑΚΑΣ 2.14: ΠΙΝΑΚΑΣ ΣΥΧΝΟΤΗΤΩΝ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΚΑΡΔΙΑΓΓΕΙΑΚΑ ΝΟΣΗΜΑΤΑ”	46
ΠΙΝΑΚΑΣ 2.15: ΠΙΝΑΚΑΣ ΣΥΧΝΟΤΗΤΩΝ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΚΑΠΝΙΣΜΑ”	47
ΠΙΝΑΚΑΣ 2.16: ΠΙΝΑΚΑΣ ΣΥΧΝΟΤΗΤΩΝ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΦΥΣΙΚΗ ΔΡΑΣΤΗΡΙΟΤΗΤΑ”	48

3° ΚΕΦΑΛΑΙΟ

ΠΙΝΑΚΑΣ 3.1: ΤΙΜΕΣ ΤΟΥ ΔΕΙΚΤΗ ΣΥΣΧΕΤΙΣΗΣ	52
ΠΙΝΑΚΑΣ 3.2: ΠΙΝΑΚΑΣ ΕΛΕΓΧΟΥ ΚΑΝΟΝΙΚΟΤΗΤΑΣ ΓΙΑ ΠΟΣΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ	55
ΠΙΝΑΚΑΣ 3.3: ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΕΩΝ ΓΙΑ ΠΟΣΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ	56
ΠΙΝΑΚΑΣ 3.4: ΠΙΝΑΚΑΣ ΣΥΝΑΦΕΙΑΣ ΓΙΑ ΤΙΣ ΜΕΤΑΒΛΗΤΕΣ ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΔΙΑΦΟΡΕΣ ΑΙΤΙΕΣ – ΚΑΠΝΙΣΜΑ	58
ΠΙΝΑΚΑΣ 3.5: ΠΙΝΑΚΑΣ ΕΛΕΓΧΟΥ χ^2 ΓΙΑ ΤΙΣ ΜΕΤΑΒΛΗΤΕΣ ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΔΙΑΦΟΡΕΣ ΑΙΤΙΕΣ – ΚΑΠΝΙΣΜΑ	58
ΠΙΝΑΚΑΣ 3.6: ΠΙΝΑΚΑΣ ΣΥΝΑΦΕΙΑΣ ΓΙΑ ΤΙΣ ΜΕΤΑΒΛΗΤΕΣ ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΔΙΑΦΟΡΕΣ ΑΙΤΙΕΣ – ΦΥΣΙΚΗ ΔΡΑΣΤΗΡΙΟΤΗΤΑ	59
ΠΙΝΑΚΑΣ 3.7: ΠΙΝΑΚΑΣ ΕΛΕΓΧΟΥ χ^2 ΓΙΑ ΤΙΣ ΜΕΤΑΒΛΗΤΕΣ ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΔΙΑΦΟΡΕΣ ΑΙΤΙΕΣ – ΦΥΣΙΚΗ ΔΡΑΣΤΗΡΙΟΤΗΤΑ	59
ΠΙΝΑΚΑΣ 3.8: ΠΙΝΑΚΑΣ ΣΥΝΑΦΕΙΑΣ ΓΙΑ ΤΙΣ ΜΕΤΑΒΛΗΤΕΣ ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΚΑΡΔΙΑΓΓΕΙΑΚΑ ΝΟΣΗΜΑΤΑ – ΚΑΠΝΙΣΜΑ	60
ΠΙΝΑΚΑΣ 3.9: ΠΙΝΑΚΑΣ ΕΛΕΓΧΟΥ χ^2 ΓΙΑ ΤΙΣ ΜΕΤΑΒΛΗΤΕΣ ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΚΑΡΔΙΑΓΓΕΙΑΚΑ ΝΟΣΗΜΑΤΑ – ΚΑΠΝΙΣΜΑ	60
ΠΙΝΑΚΑΣ 3.10: ΠΙΝΑΚΑΣ ΣΥΝΑΦΕΙΑΣ ΓΙΑ ΤΙΣ ΜΕΤΑΒΛΗΤΕΣ ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΚΑΡΔΙΑΓΓΕΙΑΚΑ ΝΟΣΗΜΑΤΑ – ΦΥΣΙΚΗ ΔΡΑΣΤΗΡΙΟΤΗΤΑ	61
ΠΙΝΑΚΑΣ 3.11: ΠΙΝΑΚΑΣ ΕΛΕΓΧΟΥ χ^2 ΓΙΑ ΤΙΣ ΜΕΤΑΒΛΗΤΕΣ ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΚΑΡΔΙΑΓΓΕΙΑΚΑ ΝΟΣΗΜΑΤΑ – ΦΥΣΙΚΗ ΔΡΑΣΤΗΡΙΟΤΗΤΑ	61
ΠΙΝΑΚΑΣ 3.12: ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΗΣ ΧΡΟΝΟΥ ΠΑΡΑΜΟΝΗΣ ΣΤΗΝ ΕΡΕΥΝΑ ΜΕ ΤΗΝ ΦΥΣΙΚΗ ΔΡΑΣΤΗΡΙΟΤΗΤΑ	62
ΠΙΝΑΚΑΣ 3.13: ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΗΣ ΧΡΟΝΟΥ ΠΑΡΑΜΟΝΗΣ ΣΤΗΝ ΕΡΕΥΝΑ ΜΕ ΤΟ ΚΑΠΝΙΣΜΑ	63
ΠΙΝΑΚΑΣ 3.14: ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΗΣ ΠΟΣΟΤΙΚΩΝ ΜΕΤΑΒΛΗΤΩΝ ΜΕ ΤΗΝ ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΔΙΑΦΟΡΕΣ ΑΙΤΙΕΣ	63
ΠΙΝΑΚΑΣ 3.15: ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΗΣ ΠΟΣΟΤΙΚΩΝ ΜΕΤΑΒΛΗΤΩΝ ΜΕ ΤΗΝ ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΚΑΡΔΙΑΓΓΕΙΑΚΑ ΝΟΣΗΜΑΤΑ	65
ΠΙΝΑΚΑΣ 3.16: ΠΙΝΑΚΑΣ ΠΑΡΟΥΣΙΑΣΗΣ ΠΕΡΙΓΡΑΦΙΚΩΝ ΜΕΤΡΩΝ ΤΩΝ ΜΕΤΑΒΛΗΤΩΝ AGE_ENTR, FU ΚΑΙ SMOKING ΣΥΝΑΡΤΗΣΕΙ ΤΩΝ ΠΛΗΘΥΣΜΩΝ ΠΟΥ ΟΡΙΖΟΝΤΑΙ ΑΠΟ ΤΗΝ ΜΕΤΑΒΛΗΤΗ ALLC_MORTALITY	66

ΠΙΝΑΚΑΣ 3.17: ΠΙΝΑΚΑΣ ΠΑΡΟΥΣΙΑΣΗΣ ΠΕΡΙΓΡΑΦΙΚΩΝ ΜΕΤΡΩΝ ΤΩΝ ΜΕΤΑΒΛΗΤΩΝ AGE_ENTR, SBP, DBP ΚΑΙ FU ΣΥΝΑΡΤΗΣΕΙ ΤΩΝ ΠΛΗΘΥΣΜΩΝ ΠΟΥ ΟΡΙΖΟΝΤΑΙ ΑΠΟ ΤΗΝ ΜΕΤΑΒΛΗΤΗ CVD_MORTALITY	72
---	----

4° ΚΕΦΑΛΑΙΟ

ΠΙΝΑΚΑΣ 4.1: ΠΙΝΑΚΑΣ ΠΕΡΙΓΡΑΦΗΣ ΤΟΥ ΔΕΙΓΜΑΤΟΣ	78
ΠΙΝΑΚΑΣ 4.2: ΠΙΝΑΚΑΣ ΔΟΚΙΜΩΝ ΓΕΝΙΚΩΝ ΣΥΝΤΕΛΕΣΤΩΝ ΜΟΝΤΕΛΟΥ	78
ΠΙΝΑΚΑΣ 4.3: ΠΙΝΑΚΑΣ ΕΞΙΣΩΣΗΣ ΜΕΤΑΒΛΗΤΩΝ	79
ΠΙΝΑΚΑΣ 4.4: ΠΙΝΑΚΑΣ ΠΕΡΙΓΡΑΦΗΣ ΤΟΥ ΔΕΙΓΜΑΤΟΣ	81
ΠΙΝΑΚΑΣ 4.5: ΠΙΝΑΚΑΣ ΔΟΚΙΜΩΝ ΓΕΝΙΚΩΝ ΣΥΝΤΕΛΕΣΤΩΝ ΜΟΝΤΕΛΟΥ	82
ΠΙΝΑΚΑΣ 4.6: ΠΙΝΑΚΑΣ ΕΞΙΣΩΣΗΣ ΜΕΤΑΒΛΗΤΩΝ	82

5° ΚΕΦΑΛΑΙΟ

ΠΙΝΑΚΑΣ 5.1: ΠΙΝΑΚΑΣ ΕΙΣΑΓΩΓΗΣ ΜΕΤΑΒΛΗΤΩΝ ΣΥΜΦΩΝΑ ΜΕ ΤΗ ΜΕΘΟΔΟ STEPWISE	93
ΠΙΝΑΚΑΣ 5.2: ΠΙΝΑΚΑΣ ΣΥΝΤΕΛΕΣΤΩΝ ΤΟΥ ΜΟΝΤΕΛΟΥ ΜΕ ΤΗ ΜΕΘΟΔΟ STEPWISE	94
ΠΙΝΑΚΑΣ 5.3: ΠΙΝΑΚΑΣ ΕΙΣΑΓΩΓΗΣ ΜΕΤΑΒΛΗΤΩΝ ΣΥΜΦΩΝΑ ΜΕ ΤΗ ΜΕΘΟΔΟ STEPWISE	95
ΠΙΝΑΚΑΣ 5.4: ΠΙΝΑΚΑΣ ΣΥΝΤΕΛΕΣΤΩΝ ΤΟΥ ΜΟΝΤΕΛΟΥ ΜΕ ΤΗ ΜΕΘΟΔΟ STEPWISE	96
ΠΙΝΑΚΑΣ 5.5: ΠΙΝΑΚΑΣ ΣΥΝΤΕΛΕΣΤΩΝ ΤΟΥ ΜΟΝΤΕΛΟΥ ΜΕ ΤΗΝ ΜΕΘΟΔΟ ENTER (ΕΠΑΝΑΛΗΨΗ 1Η)	97
ΠΙΝΑΚΑΣ 5.6: ΠΙΝΑΚΑΣ ΣΥΝΤΕΛΕΣΤΩΝ ΤΟΥ ΜΟΝΤΕΛΟΥ ΜΕ ΤΗΝ ΜΕΘΟΔΟ ENTER (ΕΠΑΝΑΛΗΨΗ 2Η)	98
ΠΙΝΑΚΑΣ 5.7: ΠΙΝΑΚΑΣ ΣΥΝΤΕΛΕΣΤΩΝ ΤΟΥ ΜΟΝΤΕΛΟΥ ΜΕ ΤΗΝ ΜΕΘΟΔΟ ENTER (ΕΠΑΝΑΛΗΨΗ 3Η)	98
ΠΙΝΑΚΑΣ 5.8: ΠΙΝΑΚΑΣ ΣΥΝΤΕΛΕΣΤΩΝ ΤΟΥ ΜΟΝΤΕΛΟΥ ΜΕ ΤΗΝ ΜΕΘΟΔΟ ENTER (ΕΠΑΝΑΛΗΨΗ 1Η)	99
ΠΙΝΑΚΑΣ 5.9: ΠΙΝΑΚΑΣ ΣΥΝΤΕΛΕΣΤΩΝ ΤΟΥ ΜΟΝΤΕΛΟΥ ΜΕ ΤΗΝ ΜΕΘΟΔΟ ENTER (ΕΠΑΝΑΛΗΨΗ 2Η)	99

ΠΙΝΑΚΑΣ 5.10: ΠΙΝΑΚΑΣ ΣΥΝΤΕΛΕΣΤΩΝ ΤΟΥ ΜΟΝΤΕΛΟΥ ΜΕ ΤΗΝ ΜΕΘΟΔΟ ENTER (ΕΠΑΝΑΛΗΨΗ 3Η)	100
---	-----

ΠΙΝΑΚΑΣ 5.11: ΠΙΝΑΚΑΣ ΕΛΕΓΧΟΥ ΚΑΝΟΝΙΚΟΤΗΤΑΣ ΤΩΝ ΣΦΑΛΜΑΤΩΝ	100
---	-----

6° ΚΕΦΑΛΑΙΟ

ΠΙΝΑΚΑΣ 6.1: ΠΙΝΑΚΑΣ ΠΕΡΙΓΡΑΦΗΣ ΤΟΥ ΔΕΙΓΜΑΤΟΣ	104
---	-----

ΠΙΝΑΚΑΣ 6.2: ΠΙΝΑΚΑΣ ΕΛΕΓΧΟΥ ΚΑΛΗΣ ΠΡΟΣΑΡΜΟΓΗΣ HOSMER ΚΑΙ LEMESHOW	104
--	-----

ΠΙΝΑΚΑΣ 6.3: ΠΙΝΑΚΑΣ ΑΚΡΙΒΕΙΑΣ ΠΡΟΒΛΕΨΗΣ ΤΟΥ ΜΟΝΤΕΛΟΥ	105
---	-----

ΠΙΝΑΚΑΣ 6.4: ΠΙΝΑΚΑΣ ΤΩΝ ΣΥΝΤΕΛΕΣΤΩΝ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΤΟΥ ΜΟΝΤΕΛΟΥ	105
---	-----

ΠΙΝΑΚΑΣ 6.5: ΠΙΝΑΚΑΣ ΠΕΡΙΓΡΑΦΗΣ ΤΟΥ ΔΕΙΓΜΑΤΟΣ	107
---	-----

ΠΙΝΑΚΑΣ 6.6: ΠΙΝΑΚΑΣ ΕΛΕΓΧΟΥ ΚΑΛΗΣ ΠΡΟΣΑΡΜΟΓΗΣ HOSMER ΚΑΙ LEMESHOW	107
--	-----

ΠΙΝΑΚΑΣ 6.7: ΠΙΝΑΚΑΣ ΑΚΡΙΒΕΙΑΣ ΠΡΟΒΛΕΨΗΣ ΤΟΥ ΜΟΝΤΕΛΟΥ	108
---	-----

ΠΙΝΑΚΑΣ 6.8: ΠΙΝΑΚΑΣ ΤΩΝ ΣΥΝΤΕΛΕΣΤΩΝ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΤΟΥ ΜΟΝΤΕΛΟΥ	108
---	-----

ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ

2° ΚΕΦΑΛΑΙΟ

ΔΙΑΓΡΑΜΜΑ 2.1: ΔΙΑΓΡΑΜΜΑ ΣΥΧΝΟΤΗΤΩΝ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΗΛΙΚΙΑ ΕΙΣΟΔΟΥ ΣΤΗΝ ΕΡΕΥΝΑ”	31
ΔΙΑΓΡΑΜΜΑ 2.2: ΔΙΑΓΡΑΜΜΑ ΣΥΧΝΟΤΗΤΩΝ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΧΡΟΝΟΣ ΠΑΡΑΜΟΝΗΣ ΣΤΗΝ ΕΡΕΥΝΑ”	32
ΔΙΑΓΡΑΜΜΑ 2.3: ΔΙΑΓΡΑΜΜΑ ΣΥΧΝΟΤΗΤΩΝ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΥΨΟΣ”	33
ΔΙΑΓΡΑΜΜΑ 2.4: ΔΙΑΓΡΑΜΜΑ ΣΥΧΝΟΤΗΤΩΝ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΚΙΛΑ”	34
ΔΙΑΓΡΑΜΜΑ 2.5: ΔΙΑΓΡΑΜΜΑ ΣΥΧΝΟΤΗΤΩΝ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΣΥΣΤΟΛΙΚΗ ΑΡΤΗΡΙΑΚΗ ΠΙΕΣΗ”	35
ΔΙΑΓΡΑΜΜΑ 2.6: ΔΙΑΓΡΑΜΜΑ ΣΥΧΝΟΤΗΤΩΝ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΔΙΑΣΤΟΛΙΚΗ ΑΡΤΗΡΙΑΚΗ ΠΙΕΣΗ”	36
ΔΙΑΓΡΑΜΜΑ 2.7: ΔΙΑΓΡΑΜΜΑ ΣΥΧΝΟΤΗΤΩΝ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΚΑΡΔΙΑΚΟΣ ΡΥΘΜΟΣ” .	37
ΔΙΑΓΡΑΜΜΑ 2.8: ΔΙΑΓΡΑΜΜΑ ΣΥΧΝΟΤΗΤΩΝ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΟΛΙΚΗ ΧΟΛΗΣΤΕΡΟΛΗ”	38
ΔΙΑΓΡΑΜΜΑ 2.9: ΔΙΑΓΡΑΜΜΑ ΣΥΧΝΟΤΗΤΩΝ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΔΕΙΚΤΗΣ ΜΑΖΑΣ ΣΩΜΑΤΟΣ”	39
ΔΙΑΓΡΑΜΜΑ 2.10: BOX PLOTS ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ AGE_OF_DEATH ” ΓΙΑ ΔΙΑΦΟΡΕΣ ΑΙΤΙΕΣ ΚΑΙ ΚΑΡΔΙΑΓΓΕΙΑΚΑ ΝΟΣΗΜΑΤΑ	43
ΔΙΑΓΡΑΜΜΑ 2.11: ΡΑΒΔΟΓΡΑΜΜΑ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΔΙΑΦΟΡΕΣ ΑΙΤΙΕΣ”	46
ΔΙΑΓΡΑΜΜΑ 2.12: ΡΑΒΔΟΓΡΑΜΜΑ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΘΝΗΣΙΜΟΤΗΤΑ ΑΠΟ ΚΑΡΔΙΑΓΓΕΙΑΚΑ ΝΟΣΗΜΑΤΑ”	47
ΔΙΑΓΡΑΜΜΑ 2.13: ΡΑΒΔΟΓΡΑΜΜΑ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΚΑΠΝΙΣΜΑ”	48
ΔΙΑΓΡΑΜΜΑ 2.14: ΡΑΒΔΟΓΡΑΜΜΑ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ “ΦΥΣΙΚΗ ΔΡΑΣΤΗΡΙΟΤΗΤΑ”	49

3° ΚΕΦΑΛΑΙΟ

ΔΙΑΓΡΑΜΜΑ 3.1: BOX PLOTS ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ AGE_ENTR ΓΙΑ ΑΥΤΟΥΣ ΠΟΥ ΠΕΘΑΝΑΝ ΑΠΟ ΔΙΑΦΟΡΕΣ ΑΙΤΙΕΣ ΚΑΙ ΓΙΑ ΑΥΤΟΥΣ ΠΟΥ ΠΑΡΕΜΕΙΝΑΝ ΖΩΝΤΑΝΟΙ	68
ΔΙΑΓΡΑΜΜΑ 3.2: BOX PLOT ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ AGE_ENTR ΓΙΑ ΟΛΟ ΤΟΝ ΠΛΗΘΥΣΜΟ	68
ΔΙΑΓΡΑΜΜΑ 3.3: BOX PLOT ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ FU ΓΙΑ ΑΥΤΟΥΣ ΠΟΥ ΠΕΘΑΝΑΝ	

ΑΠΟ ΔΙΑΦΟΡΕΣ ΑΙΤΙΕΣ	69
ΔΙΑΓΡΑΜΜΑ 3.4: BOX PLOT ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ FU ΓΙΑ ΟΛΟ ΤΟΝ ΠΛΗΘΥΣΜΟ	69
ΔΙΑΓΡΑΜΜΑ 3.5: ΡΑΒΔΟΓΡΑΜΜΑ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ SMOKING ΓΙΑ ΑΥΤΟΥΣ ΠΟΥ ΠΕΘΑΝΑΝ ΑΠΟ ΔΙΑΦΟΡΕΣ ΑΙΤΙΕΣ (ALLC_MORTALITY=1)	70
ΔΙΑΓΡΑΜΜΑ 3.6: ΡΑΒΔΟΓΡΑΜΜΑ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ SMOKING ΓΙΑ ΑΥΤΟΥΣ ΠΟΥ ΠΑΡΕΜΕΙΝΑΝ ΖΩΝΤΑΝΟΙ (ALLC_MORTALITY=0)	70
ΔΙΑΓΡΑΜΜΑ 3.7: ΡΑΒΔΟΓΡΑΜΜΑ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ SMOKING ΓΙΑ ΟΛΟ ΤΟΝ ΠΛΗΘΥΣΜΟ	71
ΔΙΑΓΡΑΜΜΑ 3.8: BOX PLOTS ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ AGE_ENTR ΓΙΑ ΑΥΤΟΥΣ ΠΟΥ ΠΕΘΑΝΑΝ ΑΠΟ ΚΑΡΔΙΑΓΓΕΙΑΚΑ ΝΟΣΗΜΑΤΑ ΚΑΙ ΓΙΑ ΑΥΤΟΥΣ ΠΟΥ ΔΕΝ ΠΕΘΑΝΑΝ	73
ΔΙΑΓΡΑΜΜΑ 3.9: BOX PLOT ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ AGE_ENTR ΓΙΑ ΟΛΟ ΤΟΝ ΠΛΗΘΥΣΜΟ ...	73
ΔΙΑΓΡΑΜΜΑ 3.10: BOX PLOTS ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ SBP ΓΙΑ ΑΥΤΟΥΣ ΠΟΥ ΠΕΘΑΝΑΝ ΑΠΟ ΚΑΡΔΙΑΓΓΕΙΑΚΑ ΝΟΣΗΜΑΤΑ ΚΑΙ ΓΙΑ ΑΥΤΟΥΣ ΠΟΥ ΔΕΝ ΠΕΘΑΝΑΝ	74
ΔΙΑΓΡΑΜΜΑ 3.11: BOX PLOT ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ SBP ΓΙΑ ΟΛΟ ΤΟΝ ΠΛΗΘΥΣΜΟ	74
ΔΙΑΓΡΑΜΜΑ 3.12: BOX PLOTS ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ DBP ΓΙΑ ΑΥΤΟΥΣ ΠΟΥ ΠΕΘΑΝΑΝ ΑΠΟ ΚΑΡΔΙΑΓΓΕΙΑΚΑ ΝΟΣΗΜΑΤΑ ΚΑΙ ΓΙΑ ΑΥΤΟΥΣ ΠΟΥ ΔΕΝ ΠΕΘΑΝΑΝ	75
ΔΙΑΓΡΑΜΜΑ 3.13: BOX PLOT ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ DBP ΓΙΑ ΟΛΟ ΤΟΝ ΠΛΗΘΥΣΜΟ	75
ΔΙΑΓΡΑΜΜΑ 3.14: BOX PLOTS ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ FU ΓΙΑ ΑΥΤΟΥΣ ΠΟΥ ΠΕΘΑΝΑΝ ΑΠΟ ΚΑΡΔΙΑΓΓΕΙΑΚΑ ΝΟΣΗΜΑΤΑ ΚΑΙ ΓΙΑ ΑΥΤΟΥΣ ΠΟΥ ΔΕΝ ΠΕΘΑΝΑΝ	76
ΔΙΑΓΡΑΜΜΑ 3.15: BOX PLOT ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ FU ΓΙΑ ΟΛΟ ΤΟΝ ΠΛΗΘΥΣΜΟ	76
<u>4° ΚΕΦΑΛΑΙΟ</u>	
ΔΙΑΓΡΑΜΜΑ 4.1: ΔΙΑΓΡΑΜΜΑ ΣΥΝΑΡΤΗΣΕΩΝ ΕΠΙΒΙΩΣΗΣ ΓΙΑ ΤΟΥΣ ΠΛΗΘΥΣΜΟΥΣ ALLC_MORTALITY ΚΑΙ CVD_MORTALITY	87
ΔΙΑΓΡΑΜΜΑ 4.2: ΔΙΑΓΡΑΜΜΑ ΣΥΝΑΡΤΗΣΕΩΝ ΚΙΝΔΥΝΟΥ ΓΙΑ ΤΟΥΣ ΠΛΗΘΥΣΜΟΥΣ ALLC_MORTALITY ΚΑΙ CVD_MORTALITY	88
ΔΙΑΓΡΑΜΜΑ 4.3: ΔΙΑΓΡΑΜΜΑ ΤΩΝ ΜΕΣΩΝ ΤΙΜΩΝ ΤΩΝ ΛΟΓΩΝ ΤΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΕΠΙΒΙΩΣΗΣ	89
ΔΙΑΓΡΑΜΜΑ 4.4: ΔΙΑΓΡΑΜΜΑ ΤΩΝ ΜΕΣΩΝ ΤΙΜΩΝ ΤΩΝ ΛΟΓΩΝ ΤΩΝ ΣΥΝΑΡΤΗΣΕΩΝ ΚΙΝΔΥΝΟΥ	90

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

1° ΚΕΦΑΛΑΙΟ

- ΕΙΚΟΝΑ 1.1 : ΔΙΑΓΡΑΜΜΑΤΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΑΙΤΙΩΝ ΘΑΝΑΤΟΥ ΜΕΣΗΣ ΗΛΙΚΙΑΣ
(ΠΗΓΗ: HANNAH RITCHIE, 06/03/2019, WHAT DO THE PEOPLE OF THE
WORLD DIE FROM?, BBC) 18
- ΕΙΚΟΝΑ 1.2: ΠΥΡΑΜΙΔΑ ΜΕΣΟΓΕΙΑΚΗΣ ΔΙΑΤΡΟΦΗΣ
(ΠΗΓΗ : ΜΑΤΙΝΑ ΠΟΥΛΗ, ΜΕΣΟΓΕΙΑΚΗ ΔΙΑΤΡΟΦΗ: ΤΟ ΜΥΣΤΙΚΟ
ΤΗΣ ΜΑΚΡΟΒΙΟΤΗΤΑΣ, ΔΙΑΤΡΟΦΗ ΚΑΙ ΕΥΕΞΙΑ) 23
- ΕΙΚΟΝΑ 1.3: ΕΙΔΗ ΚΑΡΔΙΑΓΓΕΙΑΚΩΝ ΝΟΣΗΜΑΤΩΝ ΣΕ ΑΝΔΡΕΣ ΚΑΙ ΓΥΝΑΙΚΕΣ.
(ΠΗΓΗ: ΚΑΛΟΓΕΡΟΠΟΥΛΟΥ ΑΙΚΑΤΕΡΙΝΗ, ΔΙΕΡΕΥΝΗΣΗ ΤΟΥ ΡΟΛΟΥ
ΤΗΣ ΚΑΤΑΝΟΜΗΣ Ω6/Ω3 ΛΙΠΑΡΩΝ ΟΞΕΩΝ, ΣΤΗΝ ΠΙΘΑΝΟΤΗΤΑ
ΕΜΦΑΝΙΣΗΣ ΚΑΡΔΙΑΓΓΕΙΑΚΟΥ ΕΠΕΙΣΟΔΙΟΥ, ΥΠΟ ΤΗΝ ΠΑΡΟΥΣΙΑ Η
ΟΧΙ ΚΑΤΑΘΛΙΨΗΣ : ΕΠΙΔΗΜΙΟΛΟΓΙΚΗ ΜΕΛΕΤΗ, ΑΤΤΙΚΗ,
ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ, ΣΕΛ 10) 25
- ΕΙΚΟΝΑ 1.4: ΘΑΝΑΤΟΙ ΣΤΗΝ ΚΕΡΚΥΡΑ ΑΝΑ 100.000 ΚΑΤΟΙΚΟΥΣ ΣΕ ΣΧΕΣΗ ΜΕ ΑΥΤΟΥΣ
ΤΟΥ ΕΘΝΙΚΟΥ ΜΕΣΟΥ ΟΡΟΥ.
(ΠΗΓΗ : ΒΑΝΝΑ ΠΑΝΔΗ-ΑΓΑΘΟΚΛΗ (ΔΡ. ΧΗΜΙΚΟΣ ΕΡΕΥΝΗΤΡΙΑ), 07/2019,
Η ΔΗΜΟΓΡΑΦΙΚΗ ΕΞΕΛΙΞΗ ΣΤΗΝ ΚΕΡΚΥΡΑ ΚΑΙ ΟΙ ΣΥΝΕΠΕΙΕΣ ΤΗΣ) 27

1^ο ΚΕΦΑΛΑΙΟ

ΕΙΣΑΓΩΓΗ

1.1 Εισαγωγή

Στο κεφάλαιο αυτό θα παρουσιάσουμε κάποιες αιτίες θανάτου ανθρώπων μέσης ηλικίας δίνοντας έμφαση στα καρδιαγγειακά νοσήματα, στα είδη τους, στους παράγοντες κινδύνου και στα μέτρα πρόληψής τους. Επίσης, θα καταγράψουμε τον ρόλο που παίζει η μεσογειακή διατροφή στην υγεία της καρδιάς. Τέλος, θα κάνουμε μια αναφορά στις καρδιαγγειακές παθήσεις στην Ελλάδα.

1.1.1 Αιτίες θανάτου σε άτομα μέσης ηλικίας.

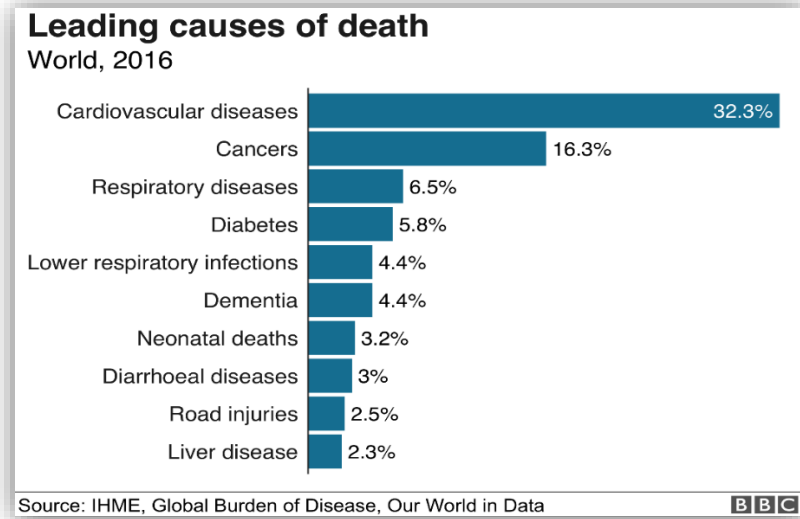
Οι πιο κοινές αιτίες θανάτου από παθήσεις του κυκλοφορικού συστήματος είναι ισχαιμικές καρδιοπάθειες και εγκεφαλοαγγειακές παθήσεις. Οι ισχαιμικές καρδιοπάθειες ήταν η αιτία για 119 θανάτους ανά 100 000 κατοίκους σε όλη την ΕΕ-28 το 2016.

Ο καρκίνος ήταν βασική αιτία θανάτου, με κατά μέσο όρο 259 θανάτους ανά 100 000 κατοίκους σε όλη την ΕΕ-28 το 2016. Στις πιο κοινές μορφές καρκίνου — όλες με τυποποιημένα ποσοστά θνησιμότητας άνω των 10 ανά 100 000 κατοίκους — περιλαμβάνονταν κακοήγη νεοπλάσματα: της τραχείας, των βρόγχων και των πνευμόνων· του παχέος εντέρου, της ορθοσιγμοειδικής διασταύρωσης, του ορθού, του πρωκτού και του πρωκτικού σωλήνα· του μαστού· του παγκρέατος· του προστάτη· του στομάχου· του ήπατος και των χοληφόρων οδών.

Μετά τις παθήσεις του κυκλοφορικού συστήματος και τον καρκίνο, οι παθήσεις του αναπνευστικού συστήματος ήταν η τρίτη συνηθέστερη αιτία θανάτου στην ΕΕ-28, με 83 κατά μέσο όρο θανάτους ανά 100 000 κατοίκους το 2016. Στο πλαίσιο της εν λόγω ομάδας παθήσεων, οι χρόνιες παθήσεις του κατώτερου αναπνευστικού συστήματος ήταν η συνηθέστερη αιτία θνησιμότητας, ακολουθούμενη από άλλες παθήσεις του κατώτερου αναπνευστικού και την πνευμονία.

Οι εξωτερικές αιτίες θανάτου περιλαμβάνουν, μεταξύ άλλων, τους θανάτους λόγω εκούσιου αυτοτραυματισμού (αυτοκτονίες) και λόγω ατυχημάτων στις μεταφορές. Μολονότι η αυτοκτονία δεν συνιστά σημαντική αιτία θανάτου και τα στοιχεία για ορισμένα κράτη μέλη της ΕΕ είναι πιθανόν να είναι πλημμελή, συχνά θεωρείται σημαντικός δείκτης κοινωνικών προβλημάτων. Κατά μέσο όρο, στην ΕΕ-28 σημειώθηκαν 10 θάνατοι ανά 100 000 κατοίκους λόγω αυτοκτονίας το 2016.

Μολονότι ατυχήματα στις μεταφορές συμβαίνουν καθημερινά, η συχνότητα των θανάτων που προκλήθηκαν από ατυχήματα στις μεταφορές στην ΕΕ-28 το 2016 (τυποποιημένο ποσοστό θνησιμότητας 5,6 ανά 100 000 κατοίκους) ήταν μικρότερη από τη συχνότητα των αυτοκτονιών.



Εικόνα 1.1 : Διαγραμματική απεικόνιση αιτιών θανάτου μέσης ηλικίας (Πηγή : Hannah Ritchie, 06/03/2019, What do the people of the world die from?, BBC)

1.2 Καρδιαγγειακά νοσήματα

Τα καρδιαγγειακά νοσήματα αποτελούν έναν γενικό όρο που αντιπροσωπεύει ένα ευρύ φάσμα διαταραχών, οι οποίες προκαλούνται λόγω δημιουργίας αθηρωματικής πλάκας, με αποτέλεσμα τη στένωση των αγγείων και τη μη ομαλή κυκλοφορία του αίματος. Σύμφωνα με τα επίσημα στατιστικά στοιχεία (European Cardiovascular Disease Statistics, 2012), η καρδιαγγειακή νόσος προκαλεί πάνω από 4 εκατομμύρια θανάτους το χρόνο στην Ευρώπη που την καθιστούν την πιο σημαντική αιτία θανάτου στην Ευρωπαϊκή Ένωση. Σχεδόν μία στις πέντε γυναίκες (22%) και ένας στους πέντε άνδρες (20%) πεθαίνει από καρδιαγγειακή νόσο στην Ευρώπη.

1.2.1 Είδη Καρδιαγγειακών νοσημάτων

- Ισχαιμική καρδιοπάθεια - Στεφανιαία νόσος: Προκαλείται από στένωση των αιμοφόρων αγγείων (στεφανιαίων) που αιματώνουν τον καρδιακό μυ, κυρίως εξαιτίας εναπόθεσης λίπους, με αποτέλεσμα τη δημιουργία αθηρωματικών πλακών.
- Αγγειακό εγκεφαλικό επεισόδιο: Πρόκειται για τη διαταραχή της εγκεφαλικής αιματικής κυκλοφορίας. Μπορεί να προκληθεί είτε από διακοπή της αιματικής ροής του εγκεφάλου (ισχαιμικό επεισόδιο), είτε από ρήξη των εγκεφαλικών αγγείων (αιμορραγικό επεισόδιο).
- Ρευματική καρδιοπάθεια: Πρόκειται για την καταστροφή του καρδιακού μυ και των βαλβίδων της καρδιάς, η οποία προκαλείται από άνοση αντίδραση προς τη λοίμωξη από β-αιμολυτικό στρεπτόκοκκο της ομάδας A, η οποία δεν αντιμετωπίστηκε θεραπευτικά.
- Συγγενής Καρδιοπάθεια: Πρόκειται για ανωμαλίες των καρδιαγγειακών δομών π.χ. μεσοκοιλιακά ή μεσοκολπικά ελλείμματα, ανωμαλίες των βαλβίδων, ανωμαλίες των καρδιακών κοιλοτήτων, οι οποίες υπάρχουν από τη γέννηση. Οφείλονται είτε σε γενετικούς παράγοντες για παράδειγμα, σύνδρομο Down, είτε σε περιβαλλοντικούς παράγοντες, όπως για παράδειγμα χρήση αλκοόλ, φαρμάκων (θαλιδομίδη, βαρφαρίνη) από τη μητέρα κατά τη διάρκεια της εγκυμοσύνης, ενδομήτριες λοιμώξεις (ερυθρά), κακή θρέψη της εγκύου (χαμηλή πρόσληψη φυλικού οξέως), συγγενική σχέση εξ' αίματος μεταξύ των γονέων.
- Ανευρύσματα και διαχωρισμός αορτής: Πρόκειται για διάταση και ρήξη της αορτής.
- Εν τω Βάθει Φλεβοθρόμβωση και πνευμονική εμβολή: Πρόκειται για απόφραξη του φλεβικού δικτύου των κάτω άκρων με θρόμβους, οι οποίοι μπορεί να αποσπαστούν και μέσω της κυκλοφορίας να καταλήξουν στον πνεύμονα.
- Περιφερική αρτηριοπάθεια: Πρόκειται για νόσο των περιφερικών αγγείων που αρδεύουν τα άνω και κάτω άκρα.
- Άλλα Καρδιαγγειακά νοσήματα: Αρτηριακή υπέρταση, όγκοι καρδιάς, εγκεφαλικά ανευρύσματα, δυσλειτουργία του καρδιακού μυ όπως για παράδειγμα η καρδιομυοπάθεια και οι βαλβιδοπάθειες.

Αξίζει να σημειωθεί ότι οι συχνότερες μορφές καρδιαγγειακής νόσου είναι η στεφανιαία νόσος και το αγγειακό εγκεφαλικό επεισόδιο.

1.2.2 Παράγοντες κινδύνου εμφάνισης καρδιαγγειακών νοσημάτων

Ανάμεσα στους παράγοντες κινδύνου για την εμφάνιση καρδιαγγειακών νοσημάτων περιλαμβάνονται παράγοντες τους οποίους δεν μπορούμε να αλλάξουμε όπως, το φύλο, η ηλικία και η κληρονομικότητα και άλλοι τους οποίους μπορούμε να τροποποιήσουμε μέσω της διατροφής και του τρόπου ζωής. Οι βασικότεροι παράγοντες κινδύνου στους οποίους μπορούμε να επέμβουμε είναι:

1. Δυσλιπιδαιμία

Η διαταραχή των λιπιδίων στο αίμα, πιο συγκεκριμένα αυξημένα επίπεδα «κακής (LDL)» χοληστερόλης, και τριγλυκεριδίων στο αίμα ή μειωμένα επίπεδα της «καλής» HDL χοληστερόλης, αποτελούν έναν από τους βασικότερους παράγοντες κινδύνου εμφάνισης καρδιαγγειακών νοσημάτων.

2. Κάπνισμα

Το κάπνισμα αποτελεί έναν ισχυρό παράγοντα κινδύνου εμφάνισης καρδιαγγειακών νοσημάτων και συγκεκριμένα εμφράγματος του μυοκαρδίου και αγγειακού εγκεφαλικού επεισοδίου. Σε αντίθεση με ότι μπορεί να πιστεύουμε, ένας καπνιστής διατρέχει μεγαλύτερο κίνδυνο εμφάνισης καρδιαγγειακών νοσημάτων από ότι καρκίνο του πνεύμονα. Το παθητικό κάπνισμα επίσης συγκαταλέγεται στους παράγοντες κινδύνου εμφάνισης καρδιαγγειακών νοσημάτων.

3. Παχυσαρκία & κεντρική κατανομή λίπους

Το αυξημένο σωματικό βάρος καθώς και το αυξημένο σπλαχνικό λίπος αποτελούν παράγοντες κινδύνου εμφάνισης καρδιαγγειακών νοσημάτων και ταυτόχρονα άλλων νοσημάτων όπως, σακχαρώδους διαβήτη και αρτηριακής υπέρτασης, τα οποία με τη σειρά τους αποτελούν ανεξάρτητους παράγοντες κινδύνου καρδιαγγειακών νοσημάτων.

4. Καθιστική ζωή & έλλειψη φυσικής δραστηριότητας

Υπάρχει πληθώρα μελετών οι οποίες υποστηρίζουν τα ευεργετικά αποτελέσματα που έχει η άσκηση ως παράγοντας πρόληψης και θεραπείας των καρδιαγγειακών νοσημάτων. Συγκεκριμένα, σύμφωνα με τον ΠΟΥ, 150 λεπτά ήπιας ή 60 λεπτά έντονης φυσικής δραστηριότητας εβδομαδιαίως μπορεί να μειώσει τον κίνδυνο εμφάνισης καρδιαγγειακών νοσημάτων κατά 30%.

5. Κακή ποιότητα διατροφής

Μία διατροφή πλούσια σε «κακά» λιπαρά, αλάτι και ταυτόχρονα φτωχή σε φρούτα και λαχανικά, αποτελεί παράγοντα κινδύνου εμφάνισης καρδιαγγειακών νοσημάτων.

1.2.3 Μέτρα πρόληψης καρδιαγγειακών νοσημάτων

Εκτιμάται ότι έως και το 90% των καρδιαγγειακών νοσημάτων μπορούν να προληφθούν.

Για τα άτομα που ανήκουν στις ομάδες υψηλού κινδύνου ανάπτυξης καρδιαγγειακής νόσου συνιστάται η φαρμακευτική αγωγή σε συνδυασμό με οδηγίες για την αντιμετώπιση τυχόν ζητημάτων σχετικά με τον τρόπο ζωής των ατόμων. Αυτό συνήθως σημαίνει:

- Φαρμακευτική αγωγή, κατά κανόνα με τη χορήγηση στατίνης, προκειμένου να μειωθεί το επίπεδο της χοληστερόλης, ανεξάρτητα από το σημερινό επίπεδό της. Το επίπεδο της χοληστερόλης θα πρέπει να θεωρηθεί μέρος του συνολικού καρδιαγγειακού κινδύνου. Ο καρδιαγγειακός κίνδυνος σε συνδυασμό με διάφορα επίπεδα χοληστερόλης μπορεί να ποικίλει ανάλογα με το επίπεδο της καλής χοληστερόλης (HDL), καθώς και με άλλους παράγοντες κινδύνου.
- Φαρμακευτική αγωγή, προκειμένου να μειωθεί η αρτηριακή πίεση, σε περίπτωση που είναι υψηλή ή ελαφρώς υψηλή. Περαιτέρω ενθάρρυνση, ανάλογα με την περίπτωση, για την αντιμετώπιση των παραγόντων κινδύνου που σχετίζονται με τον καθημερινό τρόπο ζωής, δηλαδή:
 - Διακοπή του καπνίσματος, σε περίπτωση που πρόκειται για καπνιστές.
 - Υγιεινή διατροφή, συμπεριλαμβανομένης της διατήρησης πρόσληψης άλατος σε 6 γρ. ημερησίως.
 - Διατήρηση του βάρους και της περιφέρειας της μέσης υπό έλεγχο
 - Συστηματική σωματική άσκηση.
 - Μείωση των ποσοτήτων κατανάλωσης αλκοόλ, αν αυτές είναι υπερβολικές.

Τα άτομα που δεν ανήκουν στις ομάδες υψηλού κινδύνου δε σημαίνει ότι δεν διατρέχουν κανένα κίνδυνο εμφάνισης καρδιαγγειακών νοσημάτων, παρά μόνο ότι έχουν λιγότερες πιθανότητες. Η φαρμακευτική αγωγή δεν κρίνεται απαραίτητη, ωστόσο τα άτομα μπορούν να μειώσουν περαιτέρω τις πιθανότητες με τη βελτίωση του τρόπου ζωής τους (όπως περιγράφεται παραπάνω).

1.3 Διατροφή και πρόληψη καρδιαγγειακών παθήσεων

Η διατροφή αποτελεί βασικό τροποποιήσιμο παράγοντα κινδύνου για την πρόληψη της καρδιαγγειακής νόσου. Εκ των επιστημονικών δεδομένων προκύπτει πλέον το συμπέρασμα ότι μεγαλύτερη σημασία έχει η εστίαση κυρίως στην ποιότητα των λιπαρών της διατροφής, παρά στην ποσότητα.

- Αντικατάσταση κορεσμένου λίπους

Πρόσφατες μεταanalύσεις δείχνουν ότι η αντικατάσταση του κορεσμένου λίπους από πολυακόρεστα δύναται να μειώσει τον κίνδυνο καρδιαγγειακών παθήσεων, ενώ αναδεικνύουν τη σημασία της μείωσης των trans λιπαρών.

- Ευεργετικός ρόλος βιταμινών

Υπάρχουν στοιχεία που δείχνουν ενδεχόμενο ευεργετικό ρόλο της βιταμίνης D, των βιταμινών του συμπλέγματος B, των καροτενοειδών, της βιταμίνης E και της βιταμίνης C. Αν και οι μελέτες δείχνουν ότι η χορήγηση αυτών των βιταμινών υπό μορφή συμπληρώματος δεν έχει ουσιαστική επίδραση, η πρόσληψη τους εντός μιας ισορροπημένης διατροφής εμφανίζει σημαντική καρδιοπροστατευτική δράση, συμβάλλοντας στην πρωτογενή πρόληψη της καρδιαγγειακής νόσου.

- Τρόφιμα με προστατευτική δράση

Αναφορικά με τα τρόφιμα, η πιο ισχυρή προστατευτική συσχέτιση έχει δειχθεί για τα φρούτα, τα λαχανικά, τα ψάρια και τα δημητριακά ολικής αλέσεως, ενώ τα επιστημονικά δεδομένα δείχνουν μείωση του κινδύνου καρδιαγγειακής νόσου με μέτρια κατανάλωση αλκοόλ.

1.3.1 Μεσογειακή διατροφή και υγεία της καρδιάς.

Η διατροφή που θεωρείται από τις καλύτερες που έχουν μελετηθεί για την προαγωγή της καρδιαγγειακής υγείας είναι η μεσογειακή διατροφή. Αυτή χαρακτηρίζεται από κατανάλωση ψαριών, μονοακόρεστων εκ του ελαιόλαδου φρούτα, λαχανικά, ξηρούς καρπούς και δημητριακά. Η επίδραση της στη μείωση του κινδύνου καρδιαγγειακών σχετίζεται με τη βελτίωση του λιπιδαιμικού προφίλ, τη μείωση δεικτών φλεγμονής καθώς και τη μείωση των συμβάντων και των θανάτων από καρδιαγγειακή νόσο.



Εικόνα 1.2: Πυραμίδα Μεσογειακής διατροφής (Πηγή : Ματίνα Πούλη, Μεσογειακή Διατροφή: Το μυστικό της μακροβιότητας, Διατροφή και Ευεξία)

1.3.2 Συστατικά της διατροφής στην πρόληψη της καρδιαγγειακής νόσου

- Ω-3 πολυακόρεστα λιπαρά, ψάρι

Αρκετές μετανalύσεις έχουν δείξει δυνητική μείωση της συνολικής καρδιαγγειακής θνησιμότητας από την κατανάλωση ψαριών. Οι Ευρωπαϊκές οδηγίες αναφέρουν τη σημασία της κατανάλωσης ψαριών, τόσο για την πρωτογενή, όσο και για τη δευτερογενή πρόληψη των καρδιαγγειακών παθήσεων. Οι προτεινόμενοι μηχανισμοί σχετίζονται με τη δράση των ω-3 στη βελτίωση του λιπιδαιμικού προφίλ, τη μείωση της πίεσης, την αναστολή της φλεγμονής, την επίσχεση των οξειδωτικών διαδικασιών και την αντιθρομβωτική δράση.

- Μονοακόρεστα λιπαρά, ελαιόλαδο

Οι πρόσφατες οδηγίες αναφέρουν τη σημασία της αντικατάστασης των κορεσμένων λιπαρών με μονοακόρεστα και πολυακόρεστα για την πρωτογενή και τη δευτερογενή πρόληψη των καρδιαγγειακών. Αν και απαιτούνται ακόμη μελέτες προς επιβεβαίωση των δεδομένων, φαίνεται ότι η καρδιοπροστατευτική δράση του ελαιόλαδου οφείλεται αφενός μεν στο ελαϊκό

οξύ, αφετέρου δε στα περιεχόμενα σε αυτό αντιοξειδωτικά συστατικά όπως οι φαινολικές ενώσεις υδροξυτυροσόλη και ελευρωπαίνη.

- Φρούτα και λαχανικά

Μεγάλες επιδημιολογικές μελέτες έχουν καταδείξει τη σημαντική επίδραση της κατανάλωσης φρούτων και λαχανικών στη μείωση του κινδύνου εμφάνισης καρδιαγγειακών παθήσεων. Αυτή φαίνεται να οφείλεται στο χαμηλό θερμιδικό περιεχόμενο των φρούτων και των λαχανικών, στη μεγάλη περιεκτικότητά τους σε μικροθρεπτικά συστατικά και στην υψηλή αντιοξειδωτική τους δράση. Η αντιοξειδωτική τους δράση οφείλεται αφενός μεν στις αντιοξειδωτικές βιταμίνες, αφετέρου δε σε φυτοχημικές ουσίες όπως οι πολυφαινόλες (φλαβονοειδή, όπως κατεχίνες, ανθοκυανίνες κ.α.).

- Δημητριακά, φυτικές ίνες

Οι συστάσεις αναφέρουν ότι μια διατροφή που περιέχει περίπου 25-30g φυτικές ίνες την ημέρα, από διάφορες πηγές όπως τα δημητριακά ολικής αλέσεως, δύναται να μειώσει σημαντικά τον κίνδυνο καρδιαγγειακής νόσου. Οι μελετώμενοι μηχανισμοί δράσεις αφορούν τη μείωση του βάρους, της φλεγμονής και της οξειδωσης, την επίδραση στο μεταβολισμό της γλυκόζης και το λιπιδαιμικό προφίλ.

- Ξηροί καρποί και όσπρια

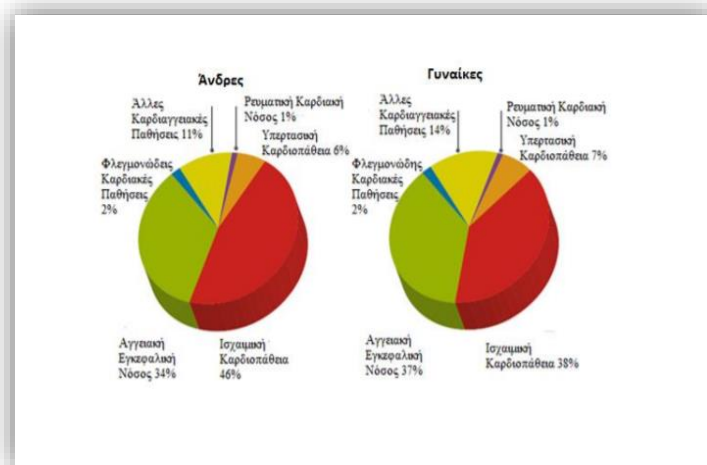
Η κατανάλωση ξηρών καρπών ως κύρια πηγή λίπους μαζί με το ελαιόλαδο φαίνεται ότι έχει καταλυτική επίδραση στη μείωση της καρδιαγγειακής θνησιμότητας. Εξίσου σημαντική επίδραση φαίνεται ότι έχει και η κατανάλωση οσπρίων, ως πολύ καλή πηγή μικροθρεπτικών, αντιοξειδωτικών και φυτικών ινών.

1.4 Καρδιαγγειακές παθήσεις στην Ελλάδα

Στην Ελλάδα ο ρυθμός θνησιμότητας από καρδιαγγειακά συμβάντα είναι κατά μέσο όρο 485 θάνατοι ανά 100.000 άτομα στους άνδρες και 391 θάνατοι ανά 100.000 άτομα στις γυναίκες. Συγκεκριμένα, ο ρυθμός θνησιμότητας από στεφανιαία νόσο ανέρχεται περίπου σε 145 θανάτους ανά 100.000 άτομα στους άνδρες και 68 θανάτους ανά 100.000 άτομα σε γυναίκες.

Στην Ελλάδα, ο ρυθμός θνησιμότητας από καρδιαγγειακά νοσήματα μειώθηκε κατά 31% στους άνδρες και κατά 39% στις γυναίκες και από στεφανιαία νόσο κατά 28% στους άνδρες και 39% στις γυναίκες. (Κωνσταντίνος Τσιούφης, 2018, “η καρδιαγγειακή νόσος στην Ελλάδα με αριθμούς”)

Το πρόγραμμα ESC Atlas είναι μια συλλογή καρδιαγγειακών στατιστικών που καταρτίστηκε από το European Heart Agency της ESC και αναπτύχθηκε σε συνεργασία με τις εθνικές καρδιολογικές εταιρείες των χωρών-μελών της ESC, υπό την ηγεσία του τέως προέδρου της ESC, καθηγητή κ. Π. Βάρδα. Σύμφωνα με τον άτλαντα υπάρχουν ενδείξεις ότι παρόλο που η καρδιαγγειακή θνησιμότητα μειώθηκε, τα νεοδιαγνωσθέντα περιστατικά καρδιαγγειακών νοσημάτων συνέχισαν να αυξάνονται μεταξύ των ετών 1990 και 2015 στην πλειοψηφία των χωρών μελών της ESC, συμπεριλαμβανομένης της Ελλάδας, ως αποτέλεσμα της αύξησης του πληθυσμού και της γήρανσης. Ως εκ τούτου, για να διατηρηθεί η πτωτική τάση στην καρδιαγγειακή θνησιμότητα ενθαρρύνεται η εφαρμογή θεραπειών για την αντιμετώπιση παραγόντων κινδύνου όπως η υπέρταση, η δυσλιπιδαιμία και ο διαβήτης.



Εικόνα 1.3: Είδη καρδιαγγειακών νοσημάτων σε άνδρες και γυναίκες. (Πηγή: Καλογεροπούλου Αικατερίνη, Διερεύνηση του ρόλου της κατανομής Ω6/Ω3 λιπαρών οξέων, στην πιθανότητα εμφάνισης καρδιαγγειακού επεισοδίου, υπο την παρουσία ή όχι κατάθλιψης : επιδημιολογική μελέτη, Αττική, Χαροκόπειο Πανεπιστήμιο, σελ 10)

1.5 Μεγάλες μελέτες της καρδιολογικής επιδημιολογίας

Οι πρώτες μελέτες που σήμερα θεωρούνται ορόσημα για την καρδιολογική επιδημιολογία, ήταν οι Framingham Study (United States Public Health Service), και Seven Countries Study (University of Minnesota Laboratory of Physiological Hygiene & Exercise Science, 1978), γνωστές στην ελληνική βιβλιογραφία ως η Μελέτη Framingham και η Μελέτη των Επτά Χωρών, αντίστοιχα. Οι μελέτες αυτές ήταν αυτές που έθεσαν τις βάσεις για την ανίχνευση των κλασικών παραγόντων κινδύνου.

1.5.1 Η μελέτη των επτά χωρών

Ο Keys αποφάσισε να διεξάγει μια μελέτη που θα αποκάλυπτε την τριγωνική σχέση καρδιακών προσβολών, χοληστερίνης και κατανάλωσης κορεσμένου λίπους. Καλλιέργησε λοιπόν την ιδέα μιας μελέτης διεθνούς συνεργασίας η οποία θα αποδείκνυε τη θεωρία του.

Σ' αυτή τη μελέτη συμμετείχαν οι ΗΠΑ, η Ιαπωνία, η Φιλανδία, η Ολλανδία, η Ιταλία, η Γιουγκοσλαβία και η Ελλάδα, χώρες που ο Keys ήξερε καλά γιατί είχε μελετήσει τη διατροφή τους. Ονομάστηκε «Μελέτη των Επτά Χωρών» και έμελλε να παίξει τεράστιο ρόλο στη διάδοση της ιδέας ότι το κορεσμένο λίπος κάνει κακό στην ανθρώπινη υγεία προκαλώντας καρδιακές προσβολές.

Κάθε μια από τις επτά χώρες ανέλαβε την υποχρέωση να θέσει στη διάθεσή του Keys και των συνεργατών του μια ομάδα υγιών ανδρών ηλικίας 40-59 ετών. Συνολικά συγκεντρώθηκαν 12.763 άτομα από 16 περιοχές επτά χωρών με σκοπό να συγκριθεί η θνησιμότητα στη διάρκεια του χρόνου και να εκτιμηθεί ο ρόλος της διατροφής.

Η επιλογή του δείγματος για την Ελλάδα έγινε με το εξής σκεπτικό: οι περιοχές έπρεπε να είναι μακριά από το κέντρο, ώστε να μην έχουν την επίδραση της ζωής του. Επελέγη τελικώς μια αγροτική περιοχή 50 χλμ. έξω από το Ηράκλειο και ένα χρόνο αργότερα μια αγροτική περιοχή 45 χλμ. έξω από την πόλη της Κέρκυρας. Το σκεπτικό σύμφωνα με το οποίο επιλέχθηκε αργότερα και η Κέρκυρα ήταν ότι επειδή η περιοχή δεν είχε βρεθεί υπό τον τουρκικό ζυγό και είχε έλθει από πολύ νωρίς σε επαφή με την Ευρώπη, ίσως είχαν εμφανιστεί αλλαγές σε κάποιες συμπεριφορές. Στην έρευνα συμμετείχε μια ομάδα υγιών ανδρών ηλικίας 40-59 ετών. Τα άτομα αυτά είχαν ετήσια παρακολούθηση σε ό, τι αφορούσε την καταγραφή τυχόν ασθενειών και πενταετή πλήρη επανεξέταση.

1.6 Θάνατοι στην Κέρκυρα

Κατά το χρονικό διάστημα 1940 – 1971 μετανάστευσε από την Κέρκυρα το 17,7% του πληθυσμού της. Επειδή ως μετανάστες έφευγαν τα άτομα νεαρών ηλικιών, με συνέπεια να μειωθούν και οι γεννήσεις, απέμεινε στο νησί ο γερασμένος πληθυσμός. Συνέπεια ήταν να καταγραφούν στο νησί περισσότεροι θάνατοι από καρκίνους, ισχαιμική καρδιοπάθεια και νόσους των αγγείων του εγκεφάλου από ότι στον εθνικό μέσο όρο.

Η πληθυσμιακή εξέλιξη της Κέρκυρας, σύμφωνα με τις απογραφές των ετών 1971-2001, φαίνεται να υπολείπεται πάντα αυτής του εθνικού μέσου όρου της χώρας, ο οποίος από μόνος του θεωρείται μικρός. Πρώτη ένδειξη είναι η σταθερή μείωση του παιδικού πληθυσμού (0-14 ετών), που καταγράφεται με την πάροδο των ετών, από 23,7% το 1971 έως 14,63% το 2001 και

ξεπερνάει αυτήν του εθνικού μέσου όρου. Δεύτερη ένδειξη είναι η αύξηση του ποσοστού των ηλικιών 60-69 ετών από 11,66% σε 11,75% και των ηλικιών 70-79 ετών από 7,28% το 1971 έως 9,91% το 2001. Όλες δε οι τιμές είναι πάνω από τον εθνικό μέσο όρο. Τρίτη απόδειξη είναι η αύξηση, συν τω χρόνω, των θανάτων από τις ομάδες των τριών ασθενειών στις οποίες είναι πιο επιρρεπή τα ηλικιωμένα άτομα: καρκίνους, καρδιοπάθειες και νόσους αγγείων του εγκεφάλου. Στην Εικόνα 1.4 παρατηρούμε ότι το ποσοστό θανάτων από κακοήθη νεοπλασμάτα στην Κέρκυρα για το διάστημα 1971 – 1991 είναι πάντα υψηλότερο από του εθνικού μέσου όρου. Υπάρχει μόνο μια αντιστροφή των ρόλων για τη δεκαετία 2001– 2011 (Πηγή : Βάννα Πανδή-Αγαθοκλή (Δρ. χημικός ερευνήτρια), 07/2019, Η δημογραφική εξέλιξη στην Κέρκυρα και οι συνέπειες της).

Οι θάνατοι από ισχαιμική καρδιοπάθεια και άλλες νόσους καρδιάς είναι υψηλότεροι για τα 30 χρόνια από αυτούς του εθνικού μέσου όρου, μειώνονται όμως κατά το διάστημα 1991-2001, για να υπερβούν πάλι τον εθνικό μέσο όρο τη δεκαετία 2001-2011. Όσον αφορά τους θανάτους από νόσους αγγείων του εγκεφάλου, είναι σταθερά περισσότεροι από αυτούς του εθνικού μέσου όρου για όλο το μελετηθέν χρονικό διάστημα.

Έτος	Κακοήθη νεοπλ.	Ισχαμ. Καρδιοπ.	Νόσοι αγγ. Εγκεφάλου
1956	71,14	140,39	130,91
1956 εθν. μέσος όρος	90,89	114,75	91,97
1961	101,2	154,26	135,59
1961 εθν. μέσος όρος	105,8	110,9	83,68
1971	147,41	206,6	167,86
1971 εθν. μέσος όρος	135,23	149,07	125,03
1981	185,7	251,95	275
1981 εθν. μέσος όρος	167,39	198,36	164,18
1991	211,9	290,9	236
1991 εθν. μέσος όρος	194,62	290,11	183,14
2001	208,8	276,3	226,86
2001 εθν. μέσος όρος	232,22	290,8	169,54
2004	248,31	303	195,28
2011	220,8	464	
2011 εθν. μέσος όρος	250	436,4	

Εικόνα 1.4: Θάνατοι στην Κέρκυρα ανά 100.000 κατοίκους σε σχέση με αυτούς του εθνικού μέσου όρου. (Πηγή : Βάννα Πανδή-Αγαθοκλή (Δρ. χημικός ερευνήτρια), 07/2019, Η δημογραφική εξέλιξη στην Κέρκυρα και οι συνέπειες της)

2^ο ΚΕΦΑΛΑΙΟ

ΠΕΡΙΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ

2.1 Εισαγωγή

Στο κεφάλαιο αυτό θα ασχοληθούμε με την περιγραφική ανάλυση των μεταβλητών μας παρουσιάζοντας μέτρα θέσης, μέτρα διασποράς και πίνακες συχνοτήτων όπου αυτό κρίνεται απαραίτητο.

2.1.1 Περιγραφική Ανάλυση

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.

Αυτή η μελέτη εξέτασε παράγοντες κινδύνου σε σχέση με τη σαρανταετή θνησιμότητα από κάθε αιτία και στεφανιαία νόσο στην κοόρτη της Κέρκυρας. Ο πληθυσμός που μελετήθηκε σε αυτήν την ανάλυση αποτελούνταν από 529 αγρότες, μεσήλικες άνδρες που εγγράφηκαν το 1961.

Ως προγνωστικοί παράγοντες χρησιμοποιήθηκαν η ηλικία, αρτηριακή πίεση, ολική χοληστερόλη ορού, κάπνισμα, φυσική δραστηριότητα, δείκτης μάζας σώματος, πάχος πτυχής δέρματος, ζωτική ικανότητα και ο εξαναγκασμένος εκπνευστικός όγκος.

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) allc mortality : Θνησιμότητα από διάφορες αιτίες (συμπεριλαμβάνεται η θνησιμότητα από καρδιαγγειακά νοσήματα).
- 6) cvd mortality : Θνησιμότητα από καρδιαγγειακά νοσήματα (CVD)
- 7) physical act : Φυσική 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

Για την παρούσα ανάλυση θα υπολογίσουμε μία επιπλέον μεταβλητή, τον δείκτη μάζας σώματος (ΔΜΣ, body mass index (BMI), ή Quetelet index). Ο ορισμός Δείκτης μάζας σώματος είναι μία γενική ιατρική ένδειξη για τον υπολογισμό του βαθμού παχυσαρκίας ενός ατόμου. Λόγω του εύκολου υπολογισμού του είναι ένα ευρέως διαδεδομένο διαγνωστικό εργαλείο των πιθανών προβλημάτων υγείας ενός ατόμου σε σχέση με το βάρος του. Δημιουργήθηκε το 1832 από τον Adolphe Quetelet.

Υπολογίζεται από τον τύπο:

$$\text{BMI} = \frac{\text{Βάρος σε κιλά}}{(\text{Ύψος σε μέτρα})^2}$$

2.1.1.1. Ποσοτικές μεταβλητές

Μια συνοπτική παρουσίαση των δεδομένων ποσοτικών μεταβλητών επιτυγχάνεται με τα περιγραφικά μέτρα, που διακρίνονται σε μέτρα θέσης και μέτρα διασποράς.

Ένα μέτρο θέσης είναι μία αριθμητική τιμή ενδεικτική της θέσης, του σημείου γύρω από το οποίο ένα σύνολο δεδομένων συγκεντρώνεται. Τέτοια είναι η μέση τιμή (μέσος όρος των μετρήσεων), η διάμεσος (η τιμή εκείνη που χωρίζει τα δεδομένα σε δύο ίσα μέρη έτσι ώστε το πλήθος των μετρήσεων που βρίσκονται αριστερά της να είναι ίσο με το πλήθος των μετρήσεων που βρίσκεται δεξιά της) και η επικρατούσα τιμή ή κορυφή (η τιμή με τη μεγαλύτερη συχνότητα).

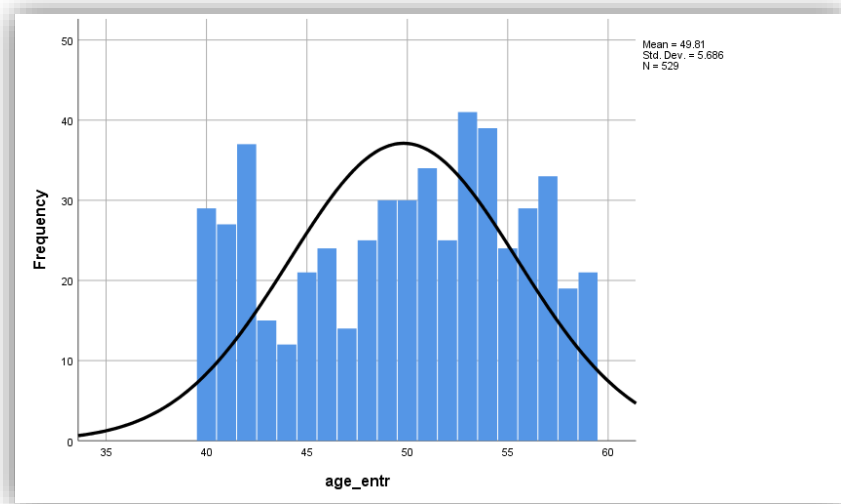
Ένα μέτρο διασποράς είναι μία αριθμητική τιμή ενδεικτική του τρόπου με τον οποίο τα δεδομένα κατανέμονται γύρω από τη μέση τιμή. Τέτοια μέτρα είναι το εύρος (παριστάνει τη διαφορά της ελάχιστης από τη μέγιστη τιμή), η διακύμανση (εκφράζει τη μεταβλητότητα ενός συνόλου αριθμητικών δεδομένων από τη μέση τους τιμή), η τυπική απόκλιση (η θετική τετραγωνική ρίζα της διακύμανσης).

1) Ηλικία εισόδου στην έρευνα (age entry)

Descriptives

		Statistic	Std. Error	
age_entr	Mean	49.81	.247	
	95% Confidence Interval for Mean	Lower Bound	49.33	
		Upper Bound	50.30	
	5% Trimmed Mean	49.86		
	Median	51.00		
	Variance	32.335		
	Std. Deviation	5.686		
	Minimum	40		
	Maximum	59		
	Range	19		
	Interquartile Range	9		
	Skewness	-.217	.106	
	Kurtosis	-1.125	.212	

Πίνακας 2.1: Περιγραφικά στατιστικά αποτελέσματα για την μεταβλητή “ηλικία εισόδου στην έρευνα”



Διάγραμμα 2.1: Διάγραμμα συχνοτήτων της μεταβλητής “ηλικία εισόδου στην έρευνα”

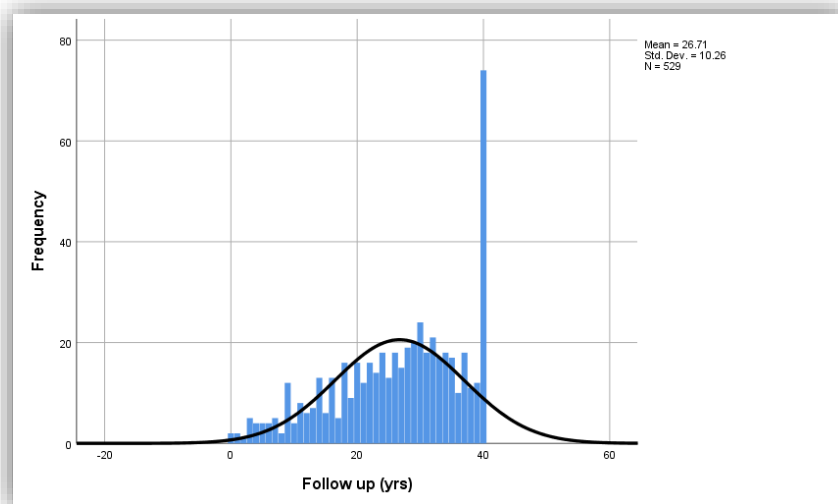
Από τα παραπάνω παρατηρούμε ότι ο μέση ηλικία εισόδου στην έρευνα ήταν 49.81 έτη. Επίσης στο παραπάνω ιστόγραμμα βλέπουμε πως η “ουρά” προς τα αριστερά μας δίνει αρνητικό πρόσημο στο συντελεστή ασυμμετρίας ($\gamma = -0.217 < 0$) και συνεπώς η κατανομή είναι αρνητικά ασύμμετρη.

2) Χρόνος παραμονής στην έρευνα (fu)

Descriptives

		Statistic	Std. Error	
Follow up (yrs)	Mean	26.71	.446	
	95% Confidence Interval for Mean	Lower Bound	25.84	
		Upper Bound	27.59	
	5% Trimmed Mean	27.23		
	Median	28.00		
	Variance	105.277		
	Std. Deviation	10.260		
	Minimum	0		
	Maximum	40		
	Range	40		
	Interquartile Range	15		
	Skewness	-.524	.106	
	Kurtosis	-.578	.212	

Πίνακας 2.2: Περιγραφικά στατιστικά αποτελέσματα για την μεταβλητή “χρόνος παραμονής στην έρευνα”



Διάγραμμα 2.2: Διάγραμμα συχνοτήτων της μεταβλητής “χρόνος παραμονής στην έρευνα”

Από τα παραπάνω παρατηρούμε ότι ο μέσος χρόνος παραμονής στην έρευνα ήταν 26.71 χρόνια. Επίσης στο παραπάνω ιστόγραμμα βλέπουμε πως η “ουρά” προς τα αριστερά μας δίνει αρνητικό πρόσημο στο συντελεστή ασυμμετρίας ($\gamma = -0.524 < 0$) και συνεπώς η κατανομή είναι αρνητικά ασύμμετρη.

Παρατηρούμε ότι η τελευταία στήλη του ιστογράμματος ($fu=40$) είναι μεγαλύτερη από τις υπόλοιπες, αντιστοιχεί σε συχνότητα 74, δηλαδή 74 από τα 529 άτομα ήταν ζωντανά ή πέθαναν την χρονιά που τελείωσε η έρευνα. Ο ακριβής υπολογισμός της συχνότητας έγινε με την δημιουργία της μεταβλητής fu_new η οποία παίρνει την τιμή 1 όταν $fu=40$ διαφορετικά παίρνει την τιμή 0, όπως φαίνεται στον παρακάτω πίνακα.

		fu_new			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	455	86.0	86.0	86.0
	1.00	74	14.0	14.0	100.0
Total		529	100.0	100.0	

Πίνακας 2.3: Υπολογισμός συχνοτήτων των τιμών της “fu_new”

Με βάση τα παραπάνω παρατηρούμε ότι το 14% του δείγματος (74 στα 529) πέθαναν ή ήταν ακόμη ζωντανοί την τελευταία χρονιά της έρευνας. Με δεδομένο ότι η μέση ηλικία εισόδου στην

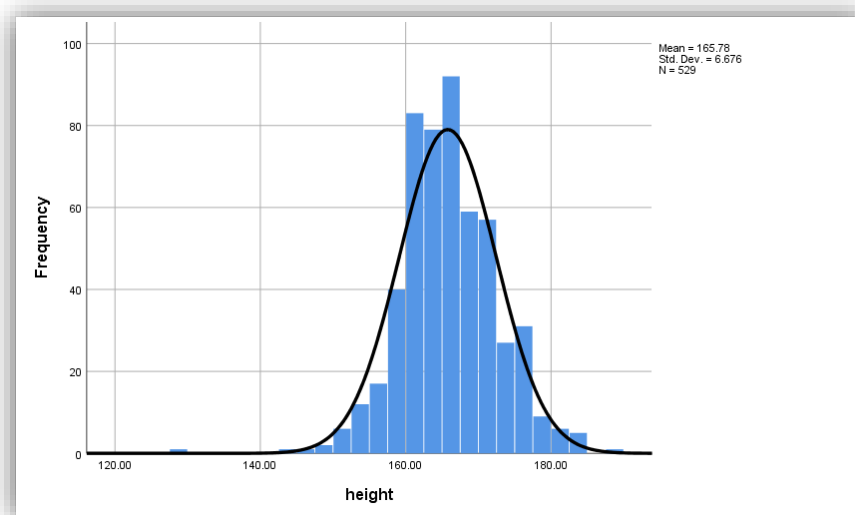
έρευνα ήταν 49.81 έτη και ότι η διάρκεια της ήταν 40 έτη, οι άνθρωποι αυτοί είχαν μεγάλη ηλικία την χρονιά που τελείωσε η έρευνα.

3) Ύψος (height)

Descriptives

		Statistic	Std. Error	
height	Mean	165.7809	.29027	
	95% Confidence Interval for Mean	Lower Bound	165.2107	
		Upper Bound	166.3511	
	5% Trimmed Mean	165.7721		
	Median	165.5000		
	Variance	44.573		
	Std. Deviation	6.67631		
	Minimum	129.50		
	Maximum	189.00		
	Range	59.50		
	Interquartile Range	8.50		
	Skewness	-.099	.106	
	Kurtosis	1.731	.212	

Πίνακας 2.4: Περιγραφικά στατιστικά αποτελέσματα για την μεταβλητή “ύψος”



Διάγραμμα 2.3: Διάγραμμα συχνοτήτων της μεταβλητής “ύψος”

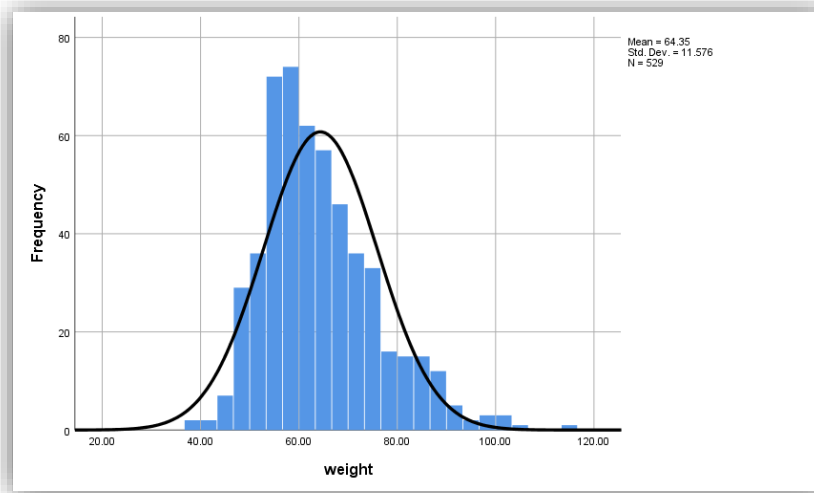
Από τα παραπάνω παρατηρούμε ότι το μέσο ύψος των ανθρώπων που συμμετείχαν στην έρευνα ήταν 165.7809. Επίσης στο παραπάνω ιστόγραμμα βλέπουμε πως η “ουρά” προς τα αριστερά μας δίνει αρνητικό πρόσημο στο συντελεστή ασυμμετρίας ($\gamma = -0.099 < 0$) και συνεπώς η κατανομή είναι αρνητικά ασύμμετρη.

4) Κιλά (weight)

Descriptives

		Statistic	Std. Error	
weight	Mean	64.3544	.50332	
	95% Confidence Interval for Mean	Lower Bound	63.3657	
		Upper Bound	65.3432	
	5% Trimmed Mean	63.7361		
	Median	62.0000		
	Variance	134.012		
	Std. Deviation	11.57636		
	Minimum	37.50		
	Maximum	113.50		
	Range	76.00		
	Interquartile Range	14.00		
	Skewness	.878	.106	
	Kurtosis	.958	.212	

Πίνακας 2.5: Περιγραφικά στατιστικά αποτελέσματα για την μεταβλητή “κιλά”



Διάγραμμα 2.4: Διάγραμμα συχνοτήτων της μεταβλητής “κιλά”

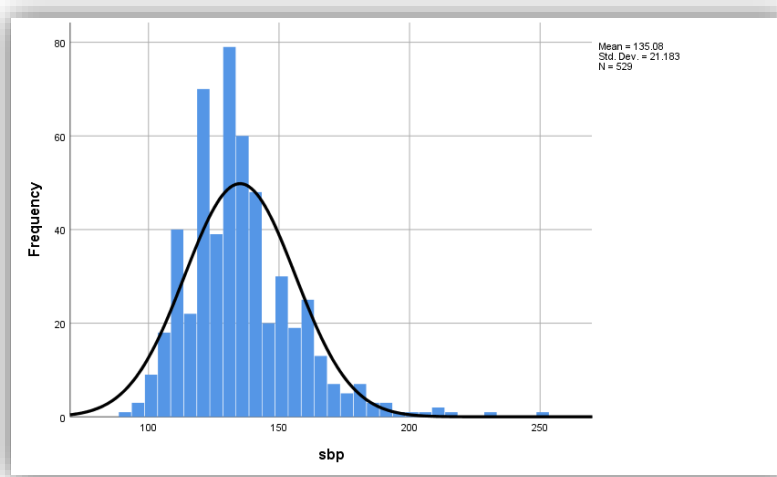
Από τα παραπάνω παρατηρούμε ότι το μέσο βάρος των ανθρώπων που συμμετείχαν στην έρευνα ήταν 64.3544. Επίσης στο παραπάνω ιστόγραμμα βλέπουμε πως η “ουρά” προς τα δεξιά μας δίνει θετικό πρόσημο στο συντελεστή ασυμμετρίας ($\gamma=0.878>0$) και συνεπώς η κατανομή είναι θετικά ασύμμετρη.

5) Συστολική αρτηριακή πίεση (sbp)

Descriptives

		Statistic	Std. Error	
sbp	Mean	135.08	.921	
	95% Confidence Interval for Mean	Lower Bound	133.27	
		Upper Bound	136.89	
	5% Trimmed Mean	133.78		
	Median	131.00		
	Variance	448.700		
	Std. Deviation	21.183		
	Minimum	91		
	Maximum	250		
	Range	159		
	Interquartile Range	26		
	Skewness	1.168	.106	
	Kurtosis	2.777	.212	

Πίνακας 2.6: Περιγραφικά στατιστικά αποτελέσματα για την μεταβλητή “συστολική αρτηριακή πίεση”



Διάγραμμα 2.5: Διάγραμμα συχνοτήτων της μεταβλητής “συστολική αρτηριακή πίεση”

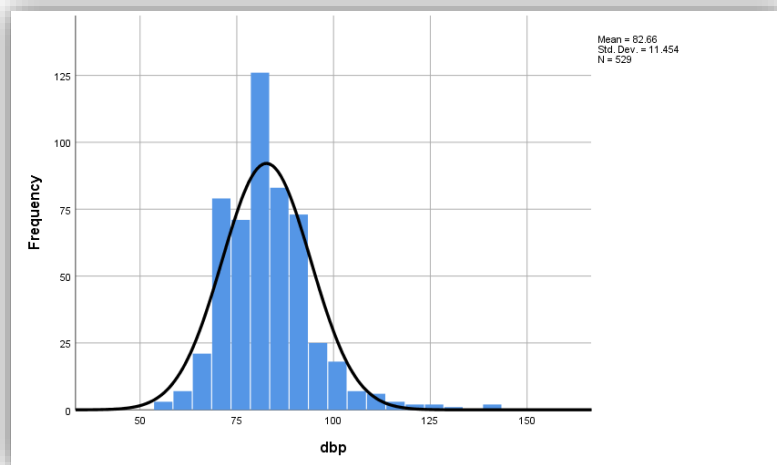
Από τα παραπάνω παρατηρούμε ότι η μέση συστολική αρτηριακή πίεση ήταν 135.08 χιλιοστά της στήλης υδραργύρου (mm Hg). Επίσης στο παραπάνω ιστόγραμμα βλέπουμε πως η “ουρά” προς τα δεξιά μας δίνει θετικό πρόσημο στο συντελεστή ασυμμετρίας ($\gamma=1.168>0$) και συνεπώς η κατανομή είναι θετικά ασύμμετρη.

6) Διαστολική αρτηριακή πίεση (dbp)

Descriptives

		Statistic	Std. Error	
dbp	Mean	82.66	.498	
	95% Confidence Interval for Mean	Lower Bound	81.68	
		Upper Bound	83.64	
	5% Trimmed Mean	82.01		
	Median	81.00		
	Variance	131.191		
	Std. Deviation	11.454		
	Minimum	56		
	Maximum	143		
	Range	87		
	Interquartile Range	14		
	Skewness	1.157	.106	
	Kurtosis	3.392	.212	

Πίνακας 2.7: Περιγραφικά στατιστικά αποτελέσματα για την μεταβλητή “διαστολική αρτηριακή πίεση”



Διάγραμμα 2.6: Διάγραμμα συχνότητων της μεταβλητής “διαστολική αρτηριακή πίεση”

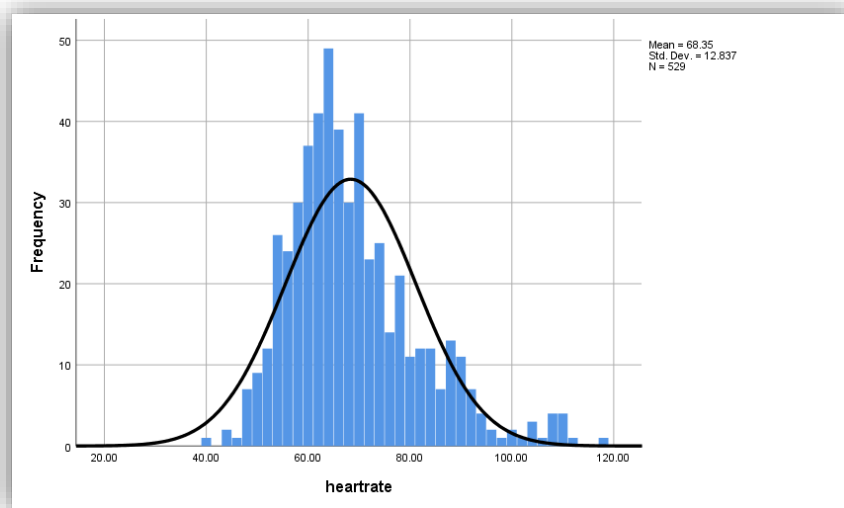
Από τα παραπάνω παρατηρούμε ότι η μέση διαστολική αρτηριακή πίεση ήταν 82.66 χιλιοστά της στήλης υδραργύρου (mm Hg) . Επίσης στο παραπάνω ιστόγραμμα βλέπουμε πως η “ουρά” προς τα δεξιά μας δίνει θετικό πρόσημο στο συντελεστή ασυμμετρίας ($\gamma=1.157>0$) και συνεπώς η κατανομή είναι θετικά ασύμμετρη.

7) Καρδιακός ρυθμός (heartrate)

Descriptives

		Statistic	Std. Error	
heartrate	Mean	68.3516	.55811	
	95% Confidence Interval for Mean	Lower Bound	67.2552	
		Upper Bound	69.4480	
	5% Trimmed Mean	67.5724		
	Median	65.0000		
	Variance	164.778		
	Std. Deviation	12.83657		
	Minimum	40.00		
	Maximum	118.00		
	Range	78.00		
	Interquartile Range	14.50		
	Skewness	.983	.106	
	Kurtosis	1.131	.212	

Πίνακας 2.8: Περιγραφικά στατιστικά αποτελέσματα για την μεταβλητή “καρδιακός ρυθμός”



Διάγραμμα 2.7: Διάγραμμα συχνοτήτων της μεταβλητής “καρδιακός ρυθμός”

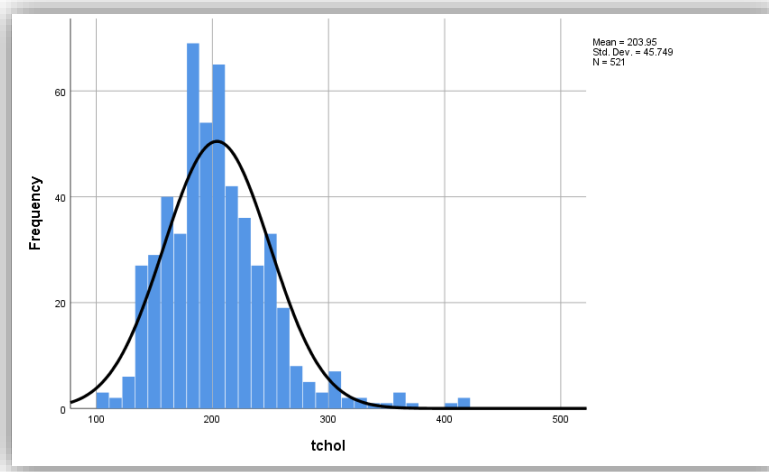
Από τα παραπάνω παρατηρούμε ότι ο μέσος καρδιακός ρυθμός ήταν 68.3516 χτύποι. Επίσης στο παραπάνω ιστόγραμμα βλέπουμε πως η "ουρά" προς τα δεξιά μας δίνει θετικό πρόσημο στο συντελεστή ασυμμετρίας ($\gamma=0.983>0$) και συνεπώς η κατανομή είναι θετικά ασύμμετρη.

8) Ολική χοληστερόλη (tchol)

Descriptives

		Statistic	Std. Error	
tchol	Mean	203.95	2.004	
	95% Confidence Interval for Mean	Lower Bound	200.01	
		Upper Bound	207.89	
	5% Trimmed Mean	201.16		
	Median	199.00		
	Variance	2092.957		
	Std. Deviation	45.749		
	Minimum	101		
	Maximum	422		
	Range	321		
	Interquartile Range	54		
	Skewness	1.106	.107	
	Kurtosis	2.768	.214	

Πίνακας 2.9: Περιγραφικά στατιστικά αποτελέσματα για την μεταβλητή "ολική χοληστερόλη"



Διάγραμμα 2.8: Διάγραμμα συχνότητων της μεταβλητής "ολική χοληστερόλη"

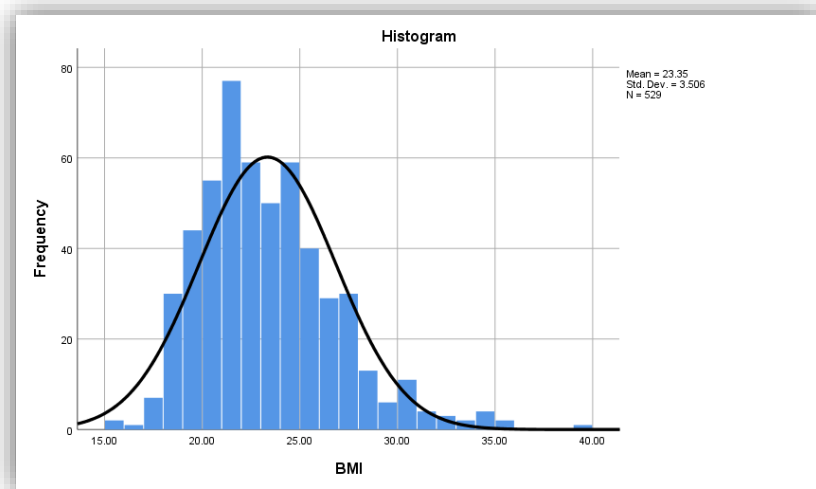
Από τα παραπάνω παρατηρούμε ότι η μέση ολική χοληστερόλη ήταν 203.95 mg/dL. Επίσης στο παραπάνω ιστόγραμμα βλέπουμε πως η “ουρά” προς τα δεξιά μας δίνει θετικό πρόσημο στο συντελεστή ασυμμετρίας ($\gamma=1.106>0$) και συνεπώς η κατανομή είναι θετικά ασύμμετρη.

9) Δείκτης Μάζας Σώματος (BMI)

Descriptives

		Statistic	Std. Error	
BMI	Mean	23.3458	.15244	
	95% Confidence Interval for Mean	Lower Bound	23.0463	
		Upper Bound	23.6452	
	5% Trimmed Mean	23.1351		
	Median	22.7801		
	Variance	12.292		
	Std. Deviation	3.50606		
	Minimum	15.77		
	Maximum	39.09		
	Range	23.32		
	Interquartile Range	4.45		
	Skewness	.923	.106	
	Kurtosis	1.304	.212	

Πίνακας 2.10: Περιγραφικά στατιστικά αποτελέσματα για την μεταβλητή “Δείκτης Μάζας Σώματος”



Διάγραμμα 2.9: Διάγραμμα συχνοτήτων της μεταβλητής “ Δείκτης Μάζας Σώματος ”

Από τα παραπάνω παρατηρούμε ότι ο μέσος Δείκτης Μάζας Σώματος ήταν 23.3458. Επίσης στο παραπάνω ιστόγραμμα βλέπουμε πως η "ουρά" προς τα δεξιά μας δίνει θετικό πρόσημο στο συντελεστή ασυμμετρίας ($\gamma=0.923>0$) και συνεπώς η κατανομή είναι θετικά ασύμμετρη.

10) Ηλικία θανάτου (age_of_death) για αυτούς που πέθαναν από διάφορες αιτίες και για αυτούς που πέθαναν από καρδιαγγειακά νοσήματα.

Αρκετό ενδιαφέρον παρουσιάζει η εξέταση των περιγραφικών μέτρων μιας μεταβλητής που θα δηλώνει την ηλικία θανάτου των ατόμων που συμμετείχαν στην έρευνα. Συγκεκριμένα, θα είχε ενδιαφέρον η μελέτη της μεταβλητής αυτής, για τα άτομα που πέθαναν από διάφορες αιτίες και για τα άτομα που πέθαναν συγκεκριμένα από καρδιαγγειακά νοσήματα.

Για τον σκοπό αυτό, δημιουργήσαμε μια νέα μεταβλητή, την age_of_death, προσθέτοντας τις μεταβλητές age_entr (ηλικία εισαγωγής ενός ατόμου στην έρευνα) και fu (χρόνος παραμονής ενός ατόμου στην έρευνα).

Αυτή η μεταβλητή περιλαμβάνει και κάποιες λογοκριμένες τιμές. Αυτές οι τιμές προκύπτουν από το γεγονός ότι η τιμή 40 στην μεταβλητή fu είναι λογόκριμένη, αφού στα 40 χρόνια η έρευνα τελείωσε. Δεδομένου ότι η ηλικία εισόδου των ατόμων στην έρευνα δεν είναι ίδια για όλους, οι λογοκριμένες τιμές της age_of_death δεν έχουν μια συγκεκριμένη τιμή.

- Περιγραφικά μέτρα της age_of_death για αυτούς που πέθαναν από διάφορες αιτίες (allc_mortality=1)

Για να μελετήσουμε λοιπόν τα περιγραφικά μέτρα της ηλικίας θανάτου για τα άτομα που πέθαναν από διάφορες αιτίες (allc_mortality=1), εξαιρούμε από το dataset τις παρατηρήσεις που αντιστοιχούν στις τιμές allc_mortality=0. Δηλαδή εξαιρούμε τα άτομα που παρέμειναν ζωντανά ως το τέλος της έρευνας. Αυτό πρακτικά σημαίνει ότι οι λογοκριμένες τιμές της μεταβλητής age_of_death δεν λαμβάνονται υπόψη στην παρακάτω περιγραφική ανάλυση.

Descriptives

		Statistic	Std. Error	
age_of_death	Mean	75.3037	.44458	
	95% Confidence Interval for Mean	Lower Bound	74.4300	
		Upper Bound	76.1773	
	5% Trimmed Mean	75.6392		
	Median	76.0000		
	Variance	91.116		
	Std. Deviation	9.54548		
	Minimum	43.00		
	Maximum	97.00		
	Range	54.00		
	Interquartile Range	13.00		
	Skewness	-.547	.114	
	Kurtosis	.036	.227	

Πίνακας 2.11: Περιγραφικά στατιστικά αποτελέσματα για την μεταβλητή “Ηλικία Θανάτου”

Από τα παραπάνω παρατηρούμε ότι ο μέσος όρος ηλικίας των ατόμων που πέθαναν από οποιαδήποτε αιτία κατά την διάρκεια της έρευνας ήταν 75.3037 χρόνια.

- Περιγραφικά μέτρα της *age_of_death* για αυτούς που πέθαναν από καρδιαγγειακά νοσήματα (*cvd_mortality=1*)

Για να μελετήσουμε τα περιγραφικά μέτρα της ηλικίας θανάτου για τα άτομα που πέθαναν από καρδιαγγειακά νοσήματα (*cvd_mortality=1*), εξαιρούμε από το dataset τις παρατηρήσεις που αντιστοιχούν στις τιμές *cvd_mortality=0*. Δηλαδή εξαιρούμε τα άτομα που δεν πέθαναν από καρδιαγγειακά νοσήματα κατά την διάρκεια της έρευνας.

Descriptives

		Statistic	Std. Error	
age_of_death	Mean	76.3073	.65027	
	95% Confidence Interval for Mean	Lower Bound	75.0247	
		Upper Bound	77.5899	
	5% Trimmed Mean	76.5926		
	Median	78.0000		
	Variance	81.188		
	Std. Deviation	9.01043		
	Minimum	45.00		
	Maximum	96.00		
	Range	51.00		
	Interquartile Range	13.00		
	Skewness	-.542	.175	
	Kurtosis	.167	.349	

Πίνακας 2.12: Περιγραφικά στατιστικά αποτελέσματα για την μεταβλητή “Ηλικία Θανάτου”

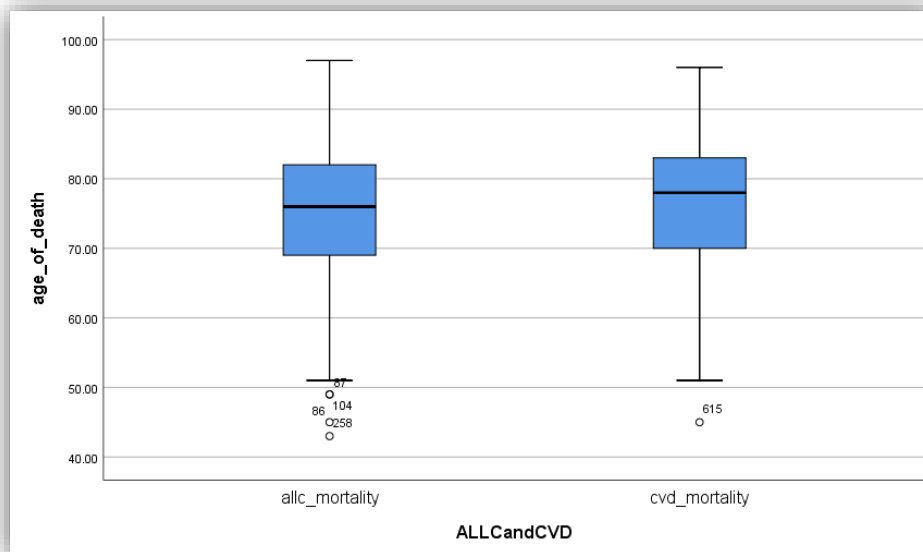
Από τα παραπάνω παρατηρούμε ότι ο μέσος όρος ηλικίας των ατόμων που πέθαναν από καρδιαγγειακά νοσήματα κατά την διάρκεια της έρευνας ήταν 76.3073 χρόνια.

- Παρουσίαση Θηκογραμμάτων της ηλικίας θανάτου από διάφορες αιτίες και από καρδιαγγειακά νοσήματα

Το δείγμα μας αποτελείται από 529 παρατηρήσεις. Προκειμένου να πραγματοποιήσουμε την σύγκριση των θηκογραμμάτων για τους δύο πληθυσμούς που ορίζονται από τις μεταβλητές `allc_mortality=1` και `cvd_mortality=1` προχωρήσαμε στην παρακάτω διαδικασία.

Αρχικά, μετατρέψαμε το δείγμα μας σε 1058 παρατηρήσεις. Αυτό επιτεύχθηκε με την αντιγραφή των 529 παρατηρήσεων κάτω από τις ήδη υπάρχουσες 529 παρατηρήσεις, στο ίδιο dataset. Στη συνέχεια, δημιουργήσαμε δύο νέες μεταβλητές. Η πρώτη κατά σειρά ονομάζεται “mortality” και η δεύτερη “ALLCandCVD”. Για την κατασκευή της μεταβλητής “mortality”, στις πρώτες 529 παρατηρήσεις βάλουμε τις αντίστοιχες τιμές της μεταβλητής “allc_mortality” και στις υπόλοιπες 529 τις τιμές της μεταβλητής “cvd_mortality”. Έτσι, η μεταβλητή “mortality” περιέχει τις τιμές 0 και 1. Η τιμή 1 συμβολίζει τον θάνατο από διάφορες αιτίες καθώς και καρδιαγγειακά νοσήματα και η τιμή 0 συμβολίζει το εάν κάποιος επέζησε από καρδιαγγειακά νοσήματα ή εάν έμεινε ζωντανός μέχρι το τέλος της έρευνας. Για την κατασκευή της μεταβλητής “ALLCandCVD” για τις πρώτες 529 παρατηρήσεις βάλουμε την τιμή 1 και για τις υπόλοιπες 529 την τιμή 2. Με αυτόν τον τρόπο, η μεταβλητή “ALLCandCVD” περιέχει ξεχωριστά τους θανάτους από διάφορες αιτίες, που συμβολίσαμε με 1 και από καρδιαγγειακά νοσήματα, που συμβολίσαμε με 2.

Τέλος, για να μελετήσουμε λοιπόν τα θηκογράμματα της ηλικίας θανάτου για τα άτομα που πέθαναν είτε από διάφορες αιτίες είτε από καρδιαγγειακά (`mortality=1`), εξαιρούμε από το dataset τις παρατηρήσεις που αντιστοιχούν στις τιμές “mortality=0”. Δηλαδή εξαιρούμε τα άτομα που παρέμειναν ζωντανά ως το τέλος της έρευνας ή δεν πέθαναν από καρδιαγγειακά.



Διάγραμμα 2.10: Box plots της μεταβλητής “age_of_death” για διάφορες αιτίες και καρδιαγγειακά νοσήματα

Από τα παραπάνω θηκογράμματα παρατηρούμε ότι, οι διάμεσοι των τιμών της ηλικίας των ατόμων που πέθαναν και για τους δύο πληθυσμούς, είναι αρκετά κοντά με τη διάμεσο της ηλικίας των ατόμων που πέθαναν από όλες τις αιτίες να έχει λίγο μικρότερη τιμή από την διάμεσο της ηλικίας των ατόμων που πέθαναν από καρδιαγγειακά νοσήματα. Επίσης, το εύρος των τιμών της ηλικίας των ατόμων και για τους δύο πληθυσμούς είναι παρόμοιο, με το εύρος των τιμών της ηλικίας των ατόμων που πέθαναν από όλες τις αιτίες να είναι λίγο μεγαλύτερο από αυτό των ατόμων που πέθαναν από καρδιαγγειακά νοσήματα. Τέλος, για την ηλικία των ατόμων που πέθαναν από διάφορες αιτίες οι παρατηρήσεις 87, 104, 86, 258 είναι ακραίες τιμές, δηλαδή 4 άτομα πέθαναν σε μικρή ηλικία συγκριτικά με τους υπόλοιπους, ενώ για την ηλικία των ατόμων που πέθαναν από καρδιαγγειακά υπάρχει μόνο μια ακραία τιμή, δηλαδή ένα άτομο που πέθανε σε μικρή ηλικία σε σχέση με τους υπόλοιπους, που αντιστοιχεί στην παρατήρηση 615.

2.1.1.2. Ποιοτικές μεταβλητές

Η συνοπτική παρουσίαση των δεδομένων μίας ποιοτικής μεταβλητής επιτυγχάνεται

- με τον πίνακα συχνοτήτων των δεδομένων και
- με τις γραφικές τους παραστάσεις (ραβδόγραμμα, κυκλικό διάγραμμα).

Ο πίνακας συχνοτήτων μιας ποιοτικής μεταβλητής προκύπτει από την απαρίθμηση και καταγραφή των δειγματικών τιμών στην αντίστοιχη κατηγορία. Ένας ολοκληρωμένος πίνακας συχνοτήτων μίας ποιοτικής μεταβλητής περιλαμβάνει τη στήλη των Συχνοτήτων (η συχνότητα παριστάνει τον αριθμό των φορών που μία κατηγορία της ποιοτικής μεταβλητής εμφανίζεται στο δείγμα) και τη στήλη των Σχετικών συχνοτήτων (η σχετική συχνότητα παριστάνει το ποσοστό επί τοις εκατό των φορών εμφάνισης μίας τιμής στο δείγμα).

Επιπλέον, μπορεί να συμπεριληφθεί στον πίνακα συχνοτήτων διατάξιμων μόνο ποιοτικών μεταβλητών, η στήλη των Αθροιστικών συχνοτήτων (παριστάνει το πλήθος των τιμών του δείγματος που είναι μικρότερες ή το πολύ ίσες από μία τιμή) .

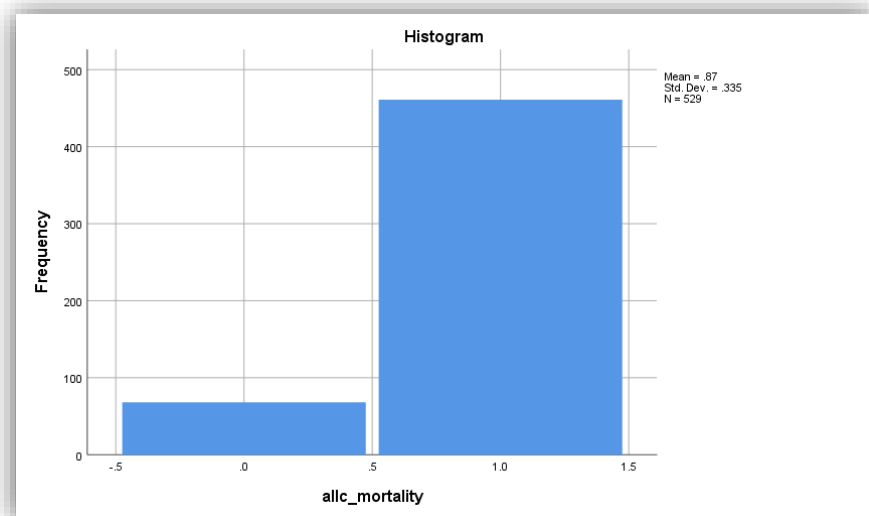
Ένας τρόπος άμεσης κατανόησης των χαρακτηριστικών της κατανομής των συχνοτήτων επιτυγχάνεται με μία ειδική γραφική παράσταση που ονομάζεται ραβδόγραμμα. Στον οριζόντιο άξονα ενός ραβδογράμματος συχνοτήτων (εναλλακτικά ενός ραβδογράμματος σχετικών συχνοτήτων) σημειώνονται οι κατηγορίες στις οποίες τα μέλη του πληθυσμού κατατάσσονται, ενώ στον κατακόρυφο άξονα οι αντίστοιχες συχνότητες (εναλλακτικά οι αντίστοιχες σχετικές συχνότητες).

1) Θνησιμότητα από διάφορες αιτίες (allc mortality)

		allc_mortality			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Alive	68	12.9	12.9	12.9
	Death	461	87.1	87.1	100.0
	Total	529	100.0	100.0	

Πίνακας 2.13: Πίνακας συχνοτήτων της μεταβλητής “θνησιμότητα από διάφορες αιτίες”

Μπορούμε να δούμε ότι το 12.9%, δηλαδή 68 από τους 529 άνδρες του δείγματος, ήταν ζωντανοί στο τέλος της έρευνας, ενώ το 87.1%, δηλαδή 461 από τους 529 άνδρες του δείγματος, πέθαναν κατά την διάρκεια της έρευνας από διάφορες αιτίες θανάτου.



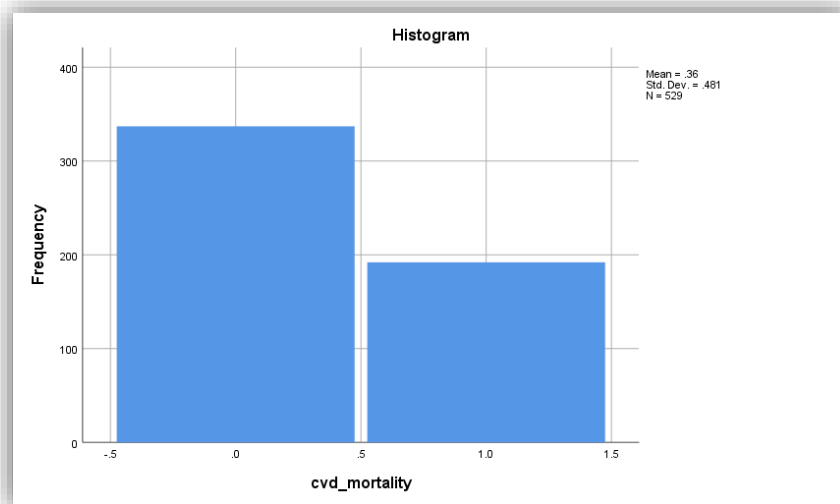
Διάγραμμα 2.11: Ραβδόγραμμα της μεταβλητής “Θνησιμότητα από διάφορες αιτίες”

2) Θνησιμότητα από καρδιαγγειακά νοσήματα (cvd mortality)

		cvd_mortality			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	None	337	63.7	63.7	63.7
	CVD	192	36.3	36.3	100.0
	Total	529	100.0	100.0	

Πίνακας 2.14: Πίνακας συχνότητων της μεταβλητής “Θνησιμότητα από καρδιαγγειακά νοσήματα”

Μπορούμε να δούμε ότι το 63.7%, δηλαδή 337 από τους 529 άνδρες του δείγματος, δεν πέθαναν από καρδιαγγειακά νοσήματα, ενώ το 36.3%, δηλαδή 192 από τους 529 άνδρες του δείγματος, πέθαναν από CVD.



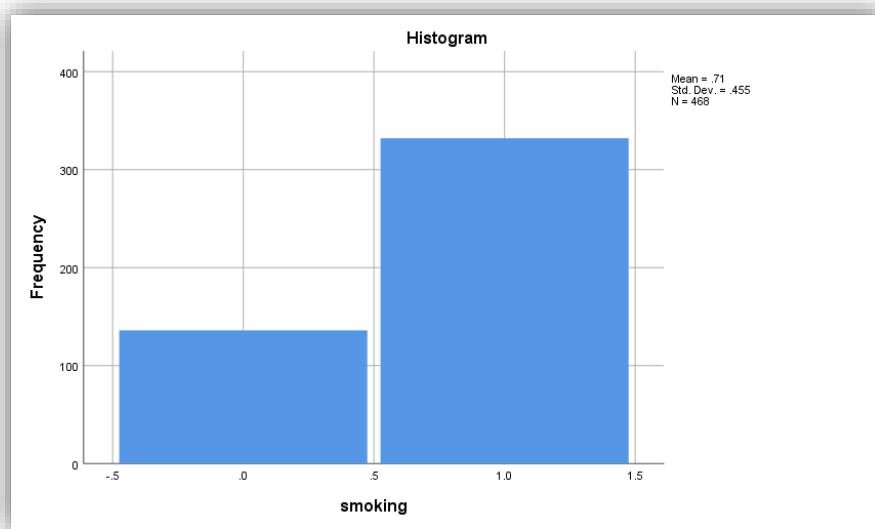
Διάγραμμα 2.12: Ραβδόγραμμα της μεταβλητής “θνησιμότητα από καρδιαγγειακά νοσήματα”

3) Κάπνισμα (smoking)

		smoking			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Never	136	25.7	29.1	29.1
	Current	332	62.8	70.9	100.0
	Total	468	88.5	100.0	
Missing	System	61	11.5		
Total		529	100.0		

Πίνακας 2.15: Πίνακας συχνοτήτων της μεταβλητής “κάπνισμα”

Μπορούμε να δούμε ότι το 25.7%, δηλαδή 136 από τους 529, δεν καπνίζει, ενώ το 62.8%, δηλαδή 332 από τους 529, του δείγματος καπνίζει. Ακόμα, το 11.5%, δηλαδή 61 παρατηρήσεις από τις 529, είναι missing values, δηλαδή δεν ξέρουμε αν καπνίζουν ή όχι.



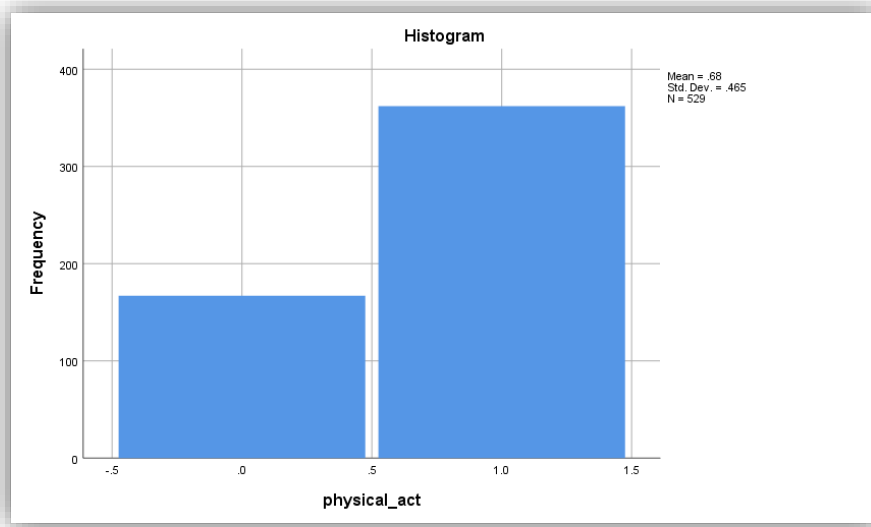
Διάγραμμα 2.13: Ραβδόγραμμα της μεταβλητής “κάπνισμα”

4) Φυσική δραστηριότητα (physical act)

		physical_act			Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	no	167	31.6	31.6	31.6
	yes	362	68.4	68.4	100.0
Total		529	100.0	100.0	

Πίνακας 2.16: Πίνακας συχνοτήτων της μεταβλητής “φυσική δραστηριότητα”

Μπορούμε να δούμε ότι το 31.6%, δηλαδή 167 από τους 529, δεν έχει φυσική δραστηριότητα, ενώ το 68.4%, δηλαδή 362 από τους 529, του δείγματος έχει φυσική δραστηριότητα.



Διάγραμμα 2.14: Ραβδόγραμμα της μεταβλητής “φυσική δραστηριότητα”

3^ο ΚΕΦΑΛΑΙΟ

ΣΥΣΧΕΤΙΣΕΙΣ

3.1 Εισαγωγή

Στο κεφάλαιο αυτό θα ασχοληθούμε με τις συσχετίσεις μεταξύ των μεταβλητών.

3.1.1 Έλεγχος χ^2

Η μέθοδος χ^2 χρησιμοποιείται γενικά για να εκτιμήσουμε αν δύο ή περισσότερα δείγματα τα οποία αποτελούνται από δεδομένα συχνοτήτων διαφέρουν σημαντικά μεταξύ τους. Με άλλα λόγια χρησιμοποιείται κυρίως για την ανάλυση πινάκων συνάφειας με βάση δύο κατηγορικές μεταβλητές.

Όταν δύο μεταβλητές είναι ανεξάρτητες οι αναμενόμενες συχνότητες για τα κελιά του πίνακα συνάφειάς τους, υπολογίζονται ως το γινόμενο των αντίστοιχων περιθωρίων συχνοτήτων κάθε κελιού, διαιρούμενου με το συνολικό πλήθος των παρατηρήσεων του πίνακα.

Γενικά η μηδενική υπόθεση σε ένα πίνακα συνάφειας με r γραμμές και c στήλες είναι ότι δεν υπάρχει σχέση μεταξύ της μεταβλητής <<γραμμή>> και της μεταβλητής <<στήλη>>. Για να ελέγξουμε αυτό συγκρίνουμε τις παρατηρούμενες συχνότητες Π κάθε κελιού με τις αναμενόμενες συχνότητες A που υπολογίζονται υπό την μηδενική υπόθεση. Πιο συγκεκριμένα, αναμενόμενη τιμή κάθε κελιού εφόσον ισχύει η μηδενική υπόθεση δίνεται από τη σχέση:

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

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

Για να ελέγξουμε την μηδενική υπόθεση υπολογίζουμε την τιμή της στατιστικής συνάρτησης χ^2 που συγκρίνει το σύνολο των παρατηρηθέντων συχνοτήτων με το σύνολο των αναμενόμενων συχνοτήτων, ως εξής:

$$\chi^2 = \sum_{i=1}^{rc} \frac{(\Pi - A)^2}{A}$$

όπου:

Π =Observed (παρατηρούμενες συχνότητες)

A =Expected (αναμενόμενες συχνότητες)

Οι προς διερεύνηση υποθέσεις είναι οι ακόλουθες:

H_0 = Οι μεταβλητές X και Y είναι ανεξάρτητες

H_1 = Οι μεταβλητές X και Y δεν είναι ανεξάρτητες

Προϋποθέσεις για να είναι το χ^2 test αξιόπιστο είναι:

- i) Καμία αναμενόμενη συχνότητα δε θα πρέπει να είναι μικρότερη του 1 και
- ii) Το ποσοστό των αναμενόμενων συχνοτήτων που είναι μικρότερες από το 5, δε θα πρέπει να υπερβαίνει το 20%.

Σε έναν πίνακα συνάφειας αν ένα κελί εμφανίσει αναμενόμενη συχνότητα μικρότερη του 5 (επομένως το 25% των αναμενόμενων συχνοτήτων), τότε παραβιάζεται η δεύτερη προϋπόθεση και σε αυτήν την περίπτωση αντί του ελέγχου χ^2 μπορούμε εναλλακτικά να χρησιμοποιήσουμε την ακριβή δοκιμασία του Fisher (Fisher's exact test).

3.1.2 Συντελεστής συσχέτισης του Pearson

Ο συντελεστής συσχέτισης του Pearson χρησιμοποιείται όταν τα δεδομένα μας προέρχονται από την κανονική κατανομή και μας δίνει το βαθμό γραμμικής εξάρτησης δύο ποσοτικών τυχαίων μεταβλητών όπου δίνεται από τη σχέση:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}},$$

Πρόκειται για έναν καθαρό αριθμό μεταξύ του -1 και 1.

Όταν $r = 0$ δεν υπάρχει γραμμική σχέση μεταξύ των X και Y , χωρίς αυτό βέβαια να αποκλείει την ύπαρξη κάποιας σχέσης άλλης μορφής π.χ. εκθετικής.

Όταν $r = +1$ υπάρχει θετική γραμμική εξάρτηση (αύξηση των τιμών της μιας επιφέρει αύξηση στις τιμές της άλλης), ενώ όταν $r = -1$ υπάρχει αρνητική γραμμική εξάρτηση (αύξηση των τιμών της μιας επιφέρει μείωση στις τιμές της άλλης).

Τιμές κοντά στο -1 ή στο 1 υποδηλώνουν αρνητική/ θετική συσχέτιση, αντίστοιχα, ενώ τιμές κοντά στο 0 μη ύπαρξη γραμμικής σχέσης.

$r < 0.3$	Καμία ή πολύ ασθενής συσχέτιση
$0.3 < r < 0.5$	Ασθενής συσχέτιση
$0.5 < r < 0.7$	Μέτρια συσχέτιση
$0.7 < r < 1$	Ισχυρή συσχέτιση
$r = 1$	Απόλυτη συσχέτιση

Πίνακας 3.1: Τιμές του δείκτη συσχέτισης

3.1.3 Ο συντελεστής συσχέτισης του Spearman

Ο συντελεστής συσχέτισης του Spearman χρησιμοποιείται όταν τα δεδομένα μας δεν προέρχονται από την κανονική κατανομή. Συμβολίζεται με r_s και δεν είναι τίποτε άλλο παρά ο συντελεστής συσχέτισης του Pearson όταν αυτός εφαρμόζεται στις τάξεις R_1, \dots, R_n και S_1, \dots, S_n , δηλαδή

$$r_s = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}},$$

όπου $\bar{R} = \frac{\sum_{i=1}^n R_i}{n}$ και $\bar{S} = \frac{\sum_{i=1}^n S_i}{n}$.

Ο συντελεστής συσχέτισης του Spearman έχει αντίστοιχη ερμηνεία με τον συντελεστή γραμμικής συσχέτισης του Pearson.

Βασικό πλεονέκτημα του συντελεστή συσχέτισης του Spearman είναι ότι μπορεί να χρησιμοποιηθεί και για διατάξιμες μεταβλητές, είναι ανθεκτικός στην ύπαρξη ακραίων τιμών, και ως μη παραμετρικός συντελεστής δεν απαιτεί καμία υπόθεση για τους πληθυσμούς. Από

την άλλη μεριά βασικό μειονέκτημα είναι ότι δεν υπολογίζεται από τις πραγματικές τιμές, αλλά από τις τάξεις.

3.1.4 Έλεγχος για τη διαφορά μέσω δύο ανεξάρτητων δειγμάτων.

Έστω ότι θέλουμε να ελέγξουμε την υπόθεση ότι η μέση τιμή μ_1 ενός χαρακτηριστικού κάποιου κανονικού πληθυσμού με άγνωστη διακύμανση είναι ίση με τη μέση τιμή μ_2 ενός άλλου κανονικού πληθυσμού με άγνωστη διακύμανση, έναντι της εναλλακτικής ότι είναι διάφορη, σε δοσμένο επίπεδο σημαντικότητας, δηλαδή

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Με βάση δυο ανεξάρτητα τυχαία δείγματα x_1, x_2, \dots, x_n $x \sim N(\mu_1, \sigma_1^2)$ και y_1, y_2, \dots, y_n $y \sim N(\mu_2, \sigma_2^2)$ από τους δύο αυτούς πληθυσμούς, διακρίνουμε δύο περιπτώσεις:

(α) Οι δυο πληθυσμοί είναι ομοσκεδαστικοί ($\sigma_1^2 = \sigma_2^2$). Κάτω από αυτή την συνθήκη, χρησιμοποιούμε με τη στατιστική συνάρτηση:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

Η οποία υπό την μηδενική υπόθεση ακολουθεί την κατανομή t με $n_1 + n_2 - 2$ βαθμούς ελευθερίας.

(β) οι δύο πληθυσμοί δεν είναι ομοσκεδαστικοί ($\sigma_1^2 \neq \sigma_2^2$). Κάτω από αυτή την συνθήκη, χρησιμοποιούμε με τη στατιστική συνάρτηση:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Η οποία υπό την μηδενική υπόθεση ακολουθεί προσεγγιστικά την κατανομή t βαθμούς ελευθερίας που δίνονται από τον τύπο:

$$df = \frac{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}{\frac{\frac{s_1^2}{n_1}}{n_1 - 1} + \frac{\frac{s_2^2}{n_2}}{n_2 - 1}}$$

3.1.5 Έλεγχος για το αν δύο ανεξάρτητα δείγματα προέρχονται από την ίδια κατανομή (Mann-Whitney U)

Πρόκειται για τον αντίστοιχο μη παραμετρικό έλεγχο της ισότητας δύο μέσων μεταξύ ανεξάρτητων δειγμάτων, τα οποία όμως δεν ακολουθούν κανονική κατανομή. Το κριτήριο Mann-Whitney είναι ένας βαθμολογικός έλεγχος που εξετάζει την υπόθεση ότι το άθροισμα των βαθμών των θέσεων των παρατηρήσεων του ενός δείγματος είναι ίσο με το άθροισμα των βαθμών των θέσεων των παρατηρήσεων του άλλου δείγματος.

$$H_0 : W_A = W_B$$

$$H_1 : W_A \neq W_B$$

Εάν η υπόθεση ότι τα αθροίσματα είναι ίσα δεν απορριφθεί, τότε συνεπάγεται ότι η κατανομή των παρατηρήσεων της μιας ομάδας είναι σχεδόν ίδια με την κατανομή της άλλης και ότι οι παρατηρήσεις εμφανίζονται με τυχαίο τρόπο τόσο στη μια όσο και στην άλλη ομάδα. Συνεπώς, η διαφοροποίηση λόγω της κατηγορικής μεταβλητής δεν σχετίζεται με τη συνεχή μεταβλητή. Η ελεγχουσυνάρτηση βασίζεται στη σχέση:

$$U = \min(U_A, U_B)$$

Όπου

$$U_A = n_A n_B + \frac{n_A(n_A + 1)}{2} - W_A$$

$$U_B = n_A n_B + \frac{n_B(n_B + 1)}{2} - W_B$$

3.2 Έλεγχος συσχέτισης ποσοτικών μεταβλητών

Στην μελέτη μας από εδώ και στο εξής δεν θα χρησιμοποιούνται οι ποσοτικές μεταβλητές height και weight. Στην θέση τους θα χρησιμοποιείται η μεταβλητή BMI της οποίας ο υπολογισμός έγινε με βάση τις άλλες δύο.

Στη συνέχεια, θα κάνουμε έλεγχο κανονικότητας για να αποφασίσουμε ποιον συντελεστή συσχέτισης ποσοτικών μεταβλητών θα εφαρμόσουμε (Pearson αν τα δεδομένα μου ακολουθούν την κανονική κατανομή, διαφορετικά τον Spearman).

- Έλεγχος κανονικότητας

H_0 : Η υπό έλεγχο κατανομή, δε διαφέρει από την κανονική κατανομή. Έναντι της εναλλακτικής

H_1 : Η υπό έλεγχο κατανομή διαφέρει από την κανονική κατανομή

One-Sample Kolmogorov-Smirnov Test

		age_entr	Follow up (yrs)	sbp	dbp	heartrate	tchol	BMI
N		529	529	529	529	529	521	529
Normal Parameters ^{a,b}	Mean	49.81	26.71	135.08	82.66	68.3516	203.95	23.3458
	Std. Deviation	5.686	10.260	21.183	11.454	12.83657	45.749	3.50606
Most Extreme Differences	Absolute	.102	.098	.130	.094	.109	.073	.069
	Positive	.091	.098	.130	.094	.109	.073	.069
	Negative	-.102	-.080	-.052	-.065	-.056	-.046	-.054
Test Statistic		.102	.098	.130	.094	.109	.073	.069
Asymp. Sig. (2-tailed)		.000 ^c	.000 ^c	.000 ^c	.000 ^c	.000 ^c	.000 ^c	.000 ^c

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

Πίνακας 3.2: Πίνακας ελέγχου κανονικότητας για ποσοτικές μεταβλητές

Παρατηρούμε πως σε επίπεδο σημαντικότητας 5% απορρίπτουμε την μηδενική μας υπόθεση για όλες τις μεταβλητές, επομένως δεν ακολουθούν την κανονική κατανομή.

Με βάση τα παραπάνω συμπεραίνουμε ότι θα πρέπει να χρησιμοποιήσουμε τον συντελεστή συσχέτισης του Spearman για να βρούμε ποιες μεταβλητές και πόσο ισχυρά συνδέονται μεταξύ τους.

H_0 = Δεν θα υπάρξει στατιστικά σημαντική σχέση μεταξύ των μεταβλητών
 H_1 = Θα υπάρξει στατιστικά σημαντική σχέση μεταξύ των μεταβλητών

- Συντελεστής συσχέτισης Spearman

		Correlations							
			Follow up (yrs)	age_entr	sbp	dbp	heartrate	tchol	BMI
Spearman's rho	Follow up (yrs)	Correlation Coefficient	1.000	-.410**	-.098*	-.103*	-.079	.075	.087*
		Sig. (2-tailed)	.	.000	.024	.018	.069	.088	.044
		N	529	529	529	529	529	521	529
	age_entr	Correlation Coefficient	-.410**	1.000	.134**	.074	-.072	.017	-.058
		Sig. (2-tailed)	.000	.	.002	.088	.098	.700	.182
		N	529	529	529	529	529	521	529
	sbp	Correlation Coefficient	-.098*	.134**	1.000	.714**	.270**	.183**	.268**
		Sig. (2-tailed)	.024	.002	.	.000	.000	.000	.000
		N	529	529	529	529	529	521	529
	dbp	Correlation Coefficient	-.103*	.074	.714**	1.000	.240**	.195**	.288**
		Sig. (2-tailed)	.018	.088	.000	.	.000	.000	.000
		N	529	529	529	529	529	521	529
	heartrate	Correlation Coefficient	-.079	-.072	.270**	.240**	1.000	.161**	.199**
		Sig. (2-tailed)	.069	.098	.000	.000	.	.000	.000
		N	529	529	529	529	529	521	529
	tchol	Correlation Coefficient	.075	.017	.183**	.195**	.161**	1.000	.296**
		Sig. (2-tailed)	.088	.700	.000	.000	.000	.	.000
		N	521	521	521	521	521	521	521
	BMI	Correlation Coefficient	.087*	-.058	.268**	.288**	.199**	.296**	1.000
		Sig. (2-tailed)	.044	.182	.000	.000	.000	.000	.
		N	529	529	529	529	529	521	529

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Πίνακας 3.3: Πίνακας συσχετίσεων για ποσοτικές μεταβλητές

Συσχετίσεις με τον χρόνο παραμονής στην έρευνα (fu)

- Για τον χρόνο παραμονής (fu) στην έρευνα και για την ηλικία εισαγωγής (age_entr) βλέπουμε πως το $p\text{-value} = 0 < 0.05$. Επομένως, υπάρχει στατιστικά σημαντική σχέση μεταξύ τους. Ο συντελεστής συσχέτισης του Spearman είναι -0.410 . Αυτό σημαίνει ότι οι μεταβλητές fu και age_entr έχουν ασθενή αρνητική συσχέτιση.
- Για τον χρόνο παραμονής στην έρευνα (fu) και την συστολική αρτηριακή πίεση (sbr) βλέπουμε πως το $p\text{-value} = 0.024 < 0.05$. Επομένως, υπάρχει στατιστικά σημαντική σχέση μεταξύ τους. Ο συντελεστής συσχέτισης του Spearman είναι -0.098 . Αυτό σημαίνει ότι οι μεταβλητές fu και sbr έχουν πολύ ασθενή αρνητική συσχέτιση.
- Για τον χρόνο παραμονής στην έρευνα (fu) και την διαστολική αρτηριακή πίεση (dbr) βλέπουμε πως το $p\text{-value} = 0.018 < 0.05$. Επομένως, υπάρχει στατιστικά σημαντική σχέση μεταξύ τους. Ο συντελεστής συσχέτισης του είναι -0.103 . Αυτό σημαίνει ότι οι μεταβλητές fu και dbr έχουν πολύ ασθενή αρνητική συσχέτιση.
- Για τον χρόνο παραμονής στην έρευνα (fu) και τον χτύπο της καρδιάς (heartrate) βλέπουμε πως το $p\text{-value} = 0.069 > 0.05$. Επομένως, δεν υπάρχει στατιστικά σημαντική σχέση μεταξύ τους.
- Για τον χρόνο παραμονής στην έρευνα (fu) και την ολική χοληστερόλη (tchol) βλέπουμε πως το $p\text{-value} = 0.088 > 0.05$. Επομένως, δεν υπάρχει στατιστικά σημαντική σχέση μεταξύ τους.
- Για τον χρόνο παραμονής στην έρευνα (fu) και τον δείκτη μάζας σώματος (BMI) βλέπουμε πως το $p\text{-value} = 0.044 < 0.05$. Επομένως, υπάρχει στατιστικά σημαντική σχέση μεταξύ τους. Ο συντελεστής συσχέτισης του Spearman είναι 0.087 . Αυτό σημαίνει ότι οι μεταβλητές fu και BMI έχουν πολύ ασθενή θετική συσχέτιση.

3.3 Έλεγχος συσχέτισης ποιοτικών μεταβλητών

Θα κάνουμε έλεγχο της συσχέτισης των ποιοτικών μεταβλητών κάτω από την υπόθεση

H₀: Οι μεταβλητές δεν είναι συσχετισμένες. Έναντι της εναλλακτικής

H₁: Οι μεταβλητές είναι συσχετισμένες

- Συσχέτιση θνησιμότητας από διάφορες αιτίες με κάπνισμα

allc_mortality * smoking Crosstabulation

		smoking		Total	
		Never	Current		
allc_mortality	Alive	Count	25	37	62
		Expected Count	18.0	44.0	62.0
	Death	Count	111	295	406
		Expected Count	118.0	288.0	406.0
Total	Count	136	332	468	
	Expected Count	136.0	332.0	468.0	

Πίνακας 3.4: Πίνακας συνάφειας για τις μεταβλητές θνησιμότητα από διάφορες αιτίες - κάπνισμα

Chi-Square Tests

	Value	df	Asymptotic Significance (2- sided)	Exact Sig. (2- sided)	Exact Sig. (1- sided)
Pearson Chi-Square	4.398 ^a	1	.036		
Continuity Correction ^b	3.790	1	.052		
Likelihood Ratio	4.174	1	.041		
Fisher's Exact Test				.050	.028
Linear-by-Linear Association	4.388	1	.036		
N of Valid Cases	468				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 18.02.

b. Computed only for a 2x2 table

Πίνακας 3.5: Πίνακας ελέγχου χ^2 για τις μεταβλητές θνησιμότητα από διάφορες αιτίες - κάπνισμα

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson, για το αν η θνησιμότητα από διάφορες αιτίες συσχετίζεται με το κάπνισμα, δεν ήταν στατιστικά σημαντικός σε επίπεδο σημαντικότητας 5% δείχνοντας ότι οι μεταβλητές συσχετίζονται αφού $\chi^2=4.398$ (p -value=0.036).

- Συσχέτιση θνησιμότητας από διάφορες αιτίες με φυσική δραστηριότητα

allc_mortality * physical_act Crosstabulation

		physical_act		Total	
		no	yes		
allc_mortality	Alive	Count	21	47	68
		Expected Count	21.5	46.5	68.0
	Death	Count	146	315	461
		Expected Count	145.5	315.5	461.0
Total		Count	167	362	529
		Expected Count	167.0	362.0	529.0

Πίνακας 3.6: Πίνακας συνάφειας για τις μεταβλητές θνησιμότητα από διάφορες αιτίες – φυσική δραστηριότητα

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.017 ^a	1	.896		
Continuity Correction ^b	.000	1	1.000		
Likelihood Ratio	.017	1	.896		
Fisher's Exact Test				1.000	.509
Linear-by-Linear Association	.017	1	.896		
N of Valid Cases	529				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 21.47.

b. Computed only for a 2x2 table

Πίνακας 3.7: Πίνακας ελέγχου χ^2 για τις μεταβλητές θνησιμότητα από διάφορες αιτίες – φυσική δραστηριότητα

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson, για το αν η θνησιμότητα από διάφορες αιτίες συσχετίζεται με την φυσική δραστηριότητα, ήταν στατιστικά σημαντικός σε επίπεδο σημαντικότητας 5% δείχνοντας ότι οι μεταβλητές δεν συσχετίζονται αφού $\chi^2=0.017$ (p -value=0.896).

- Συσχέτιση θνησιμότητας από καρδιαγγειακά νοσήματα με κάπνισμα

cvd_mortality * smoking Crosstabulation

		smoking		Total	
		Never	Current		
cvd_mortality	None	Count	83	220	303
		Expected Count	88.1	214.9	303.0
	CVD	Count	53	112	165
		Expected Count	47.9	117.1	165.0
Total		Count	136	332	468
		Expected Count	136.0	332.0	468.0

Πίνακας 3.8: Πίνακας συνάφειας για τις μεταβλητές θνησιμότητα από καρδιαγγειακά νοσήματα – κάπνισμα

Chi-Square Tests

	Value	df	Asymptotic Significance (2- sided)	Exact Sig. (2- sided)	Exact Sig. (1- sided)
Pearson Chi-Square	1.159 ^a	1	.282		
Continuity Correction ^b	.941	1	.332		
Likelihood Ratio	1.149	1	.284		
Fisher's Exact Test				.288	.166
Linear-by-Linear Association	1.156	1	.282		
N of Valid Cases	468				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 47.95.

b. Computed only for a 2x2 table

Πίνακας 3.9: Πίνακας ελέγχου χ^2 για τις μεταβλητές θνησιμότητα από καρδιαγγειακά νοσήματα - κάπνισμα

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson, για το αν η θνησιμότητα από καρδιαγγειακά νοσήματα συσχετίζεται με το κάπνισμα, ήταν στατιστικά σημαντικός σε επίπεδο σημαντικότητας 5% δείχνοντας ότι οι μεταβλητές δεν συσχετίζονται αφού $\chi^2=1.159$ (p -value=0.282).

- Συσχέτιση θνησιμότητας από καρδιαγγειακά νοσήματα με φυσική δραστηριότητα

cvd_mortality * physical_act Crosstabulation

		physical_act		Total	
		no	yes		
cvd_mortality	None	Count	112	225	337
		Expected Count	106.4	230.6	337.0
	CVD	Count	55	137	192
		Expected Count	60.6	131.4	192.0
Total		Count	167	362	529
		Expected Count	167.0	362.0	529.0

Πίνακας 3.10: Πίνακας συνάφειας για τις μεταβλητές θνησιμότητα από καρδιαγγειακά νοσήματα – φυσική δραστηριότητα

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.192 ^a	1	.275		
Continuity Correction ^b	.989	1	.320		
Likelihood Ratio	1.202	1	.273		
Fisher's Exact Test				.286	.160
Linear-by-Linear Association	1.190	1	.275		
N of Valid Cases	529				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 60.61.

b. Computed only for a 2x2 table

Πίνακας 3.11: Πίνακας ελέγχου χ^2 για τις μεταβλητές θνησιμότητα από καρδιαγγειακά νοσήματα – φυσική δραστηριότητα

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson, για το αν η θνησιμότητα από καρδιαγγειακά συσχετίζεται με την φυσική δραστηριότητα, ήταν στατιστικά σημαντικός σε επίπεδο σημαντικότητας 5% δείχνοντας ότι οι μεταβλητές δεν συσχετίζονται αφού $\chi^2=1.192$ (p -value=0.275).

3.4 Έλεγχος συσχέτισης ποιοτικών με ποσοτικές μεταβλητές

Σύμφωνα με την παράγραφο 3.1.6 το δείγμα μας δεν ακολουθεί την κανονική κατανομή. Επομένως, για την συσχέτιση ποιοτικών και ποσοτικών μεταβλητών θα χρησιμοποιήσουμε μη παραμετρικούς ελέγχους υποθέσεων.

Μη παραμετρικός έλεγχος (Έλεγχος U των Mann-Whitney για δύο ανεξάρτητα δείγματα)

H₀: Τα δύο ανεξάρτητα δείγματα έχουν την ίδια κατανομή. Έναντι της εναλλακτικής

H₁: Τα δύο ανεξάρτητα δείγματα δεν έχουν την ίδια κατανομή.

- Έλεγχος συσχέτισης φυσικής δραστηριότητας (physical_act) με χρόνο παραμονής στην έρευνα (fu).

	Follow up (yrs)
Mann-Whitney U	26463.000
Wilcoxon W	40491.000
Z	-2.308
Asymp. Sig. (2-tailed)	.021

a. Grouping Variable: physical_act

Πίνακας 3.12: Πίνακας συσχέτισης χρόνου παραμονής στην έρευνα με την φυσική δραστηριότητα

Για την μεταβλητή fu έχουμε πως το p-value του ελέγχου είναι 0.021. Επομένως σε επίπεδο σημαντικότητας 5% απορρίπτουμε την μηδενική υπόθεση και συμπεραίνουμε ότι τα δύο δείγματα δεν έχουν την ίδια κατανομή. Άρα οι μεταβλητές fu και physical_act συσχετίζονται.

- Έλεγχος συσχέτισης καπνίσματος (smoking) με χρόνο παραμονής στην έρευνα (fu).

Test Statistics^a

	Follow up (yrs)
Mann-Whitney U	18337.000
Wilcoxon W	73615.000
Z	-3.197
Asymp. Sig. (2-tailed)	.001

a. Grouping Variable: smoking

Πίνακας 3.13: Πίνακας συσχέτισης χρόνου παραμονής στην έρευνα με το κάπνισμα

Για την μεταβλητή fu έχουμε πως το p-value του ελέγχου είναι 0.001. Επομένως σε επίπεδο σημαντικότητας 5% απορρίπτουμε την μηδενική υπόθεση και συμπεραίνουμε ότι τα δύο δείγματα δεν έχουν την ίδια κατανομή. Άρα οι μεταβλητές fu και smoking συσχετίζονται.

- Έλεγχος συσχέτισης ποσοτικών μεταβλητών με την θνησιμότητα από διάφορες αιτίες (allc_mortality).

Test Statistics^a

	age_entr	Follow up (yrs)	sbp	dbp	heartrate	tchol	BMI
Mann-Whitney U	6898.500	204.000	13468.500	14420.500	15431.000	15314.500	14797.000
Wilcoxon W	9244.500	106695.000	15814.500	16766.500	121922.000	118145.500	121288.000
Z	-7.470	-13.170	-1.876	-1.067	-.207	-.076	-.745
Asymp. Sig. (2-tailed)	.000	.000	.061	.286	.836	.940	.456

a. Grouping Variable: allc_mortality

Πίνακας 3.14: Πίνακας συσχέτισης ποσοτικών μεταβλητών με την θνησιμότητα από διάφορες αιτίες

Για την μεταβλητή age_entr έχουμε πως το p-value του ελέγχου είναι 0. Επομένως σε επίπεδο σημαντικότητας 5% απορρίπτουμε την μηδενική υπόθεση και συμπεραίνουμε ότι τα δύο δείγματα δεν έχουν την ίδια κατανομή. Άρα οι μεταβλητές age_entr και allc_mortality συσχετίζονται.

Για την μεταβλητή fu έχουμε πως το p-value του ελέγχου είναι 0. Επομένως σε επίπεδο σημαντικότητας 5% απορρίπτουμε την μηδενική υπόθεση και συμπεραίνουμε ότι τα δύο δείγματα δεν έχουν την ίδια κατανομή. Άρα οι μεταβλητές fu και alic_mortality συσχετίζονται.

Για την μεταβλητή sbr έχουμε πως το p-value του ελέγχου είναι 0.061. Επομένως σε επίπεδο σημαντικότητας 5% αποδέχομαι την μηδενική υπόθεση και συμπεραίνουμε ότι τα δύο δείγματα έχουν την ίδια κατανομή. Άρα οι μεταβλητές sbr και allc_mortality δεν συσχετίζονται.

Για την μεταβλητή dbr έχουμε πως το p-value του ελέγχου είναι 0.286. Επομένως σε επίπεδο σημαντικότητας 5% αποδέχομαι την μηδενική υπόθεση και συμπεραίνουμε ότι τα δύο δείγματα έχουν την ίδια κατανομή. Άρα οι μεταβλητές dbr και allc_mortality δεν συσχετίζονται.

Για την μεταβλητή heartrate έχουμε πως το p-value του ελέγχου είναι 0.836. Επομένως σε επίπεδο σημαντικότητας 5% αποδέχομαι την μηδενική υπόθεση και συμπεραίνουμε ότι τα δύο δείγματα έχουν την ίδια κατανομή. Άρα οι μεταβλητές heartrate και allc_mortality δεν συσχετίζονται.

Για την μεταβλητή tchol έχουμε πως το p-value του ελέγχου είναι 0.940. Επομένως σε επίπεδο σημαντικότητας 5% αποδέχομαι την μηδενική υπόθεση και συμπεραίνουμε ότι τα δύο δείγματα έχουν την ίδια κατανομή. Άρα οι μεταβλητές tchol και allc_mortality δεν συσχετίζονται.

Για την μεταβλητή BMI έχουμε πως το p-value του ελέγχου είναι 0.456. Επομένως σε επίπεδο σημαντικότητας 5% αποδέχομαι την μηδενική υπόθεση και συμπεραίνουμε ότι τα δύο δείγματα έχουν την ίδια κατανομή. Άρα οι μεταβλητές BMI και allc_mortality δεν συσχετίζονται.

- Έλεγχος συσχέτισης ποσοτικών μεταβλητών με την θνησιμότητα από καρδιαγγειακά νοσήματα (cvd_mortality).

Test Statistics ^a							
	age_entr	Follow up (yrs)	sbp	dbp	heartrate	tchol	BMI
Mann-Whitney U	27223.500	27690.000	27686.000	28957.500	30131.000	29515.000	30132.000
Wilcoxon W	84176.500	46218.000	84639.000	85910.500	87084.000	86131.000	87085.000
Z	-3.039	-2.763	-2.762	-2.011	-1.314	-.952	-1.313
Asymp. Sig. (2-tailed)	.002	.006	.006	.044	.189	.341	.189

a. Grouping Variable: cvd_mortality

Πίνακας 3.15: Πίνακας συσχέτισης ποσοτικών μεταβλητών με την θνησιμότητα από καρδιαγγειακά νοσήματα

Για την μεταβλητή age_entr έχουμε πως το p-value του ελέγχου είναι 0.002. Επομένως σε επίπεδο σημαντικότητας 5% απορρίπτουμε την μηδενική υπόθεση και συμπεραίνουμε ότι τα δύο δείγματα δεν έχουν την ίδια κατανομή. Άρα οι μεταβλητές age_entr και cvd_mortality συσχετίζονται.

Για την μεταβλητή fu έχουμε πως το p-value του ελέγχου είναι 0.006. Επομένως σε επίπεδο σημαντικότητας 5% απορρίπτουμε την μηδενική υπόθεση και συμπεραίνουμε ότι τα δύο δείγματα δεν έχουν την ίδια κατανομή. Άρα οι μεταβλητές fu και cvd_mortality συσχετίζονται.

Για την μεταβλητή sbp έχουμε πως το p-value του ελέγχου είναι 0.006. Επομένως σε επίπεδο σημαντικότητας 5% απορρίπτουμε την μηδενική υπόθεση και συμπεραίνουμε ότι τα δύο δείγματα δεν έχουν την ίδια κατανομή. Άρα οι μεταβλητές sbp και cvd_mortality συσχετίζονται.

Για την μεταβλητή dbp έχουμε πως το p-value του ελέγχου είναι 0.044. Επομένως σε επίπεδο σημαντικότητας 5% απορρίπτουμε την μηδενική υπόθεση και συμπεραίνουμε ότι τα δύο δείγματα δεν έχουν την ίδια κατανομή. Άρα οι μεταβλητές dbp και cvd_mortality συσχετίζονται.

Για την μεταβλητή heartrate έχουμε πως το p-value του ελέγχου είναι 0.189. Επομένως σε επίπεδο σημαντικότητας 5% αποδέχομαι την μηδενική υπόθεση και συμπεραίνουμε ότι τα δύο δείγματα έχουν την ίδια κατανομή. Άρα οι μεταβλητές heartrate και cvd_mortality δεν συσχετίζονται.

Για την μεταβλητή tchol έχουμε πως το p-value του ελέγχου είναι 0.341. Επομένως σε επίπεδο σημαντικότητας 5% αποδέχομαι την μηδενική υπόθεση και συμπεραίνουμε ότι τα δύο δείγματα έχουν την ίδια κατανομή. Άρα οι μεταβλητές tchol και cvd_mortality δεν συσχετίζονται.

Για την μεταβλητή BMI έχουμε πως το p-value του ελέγχου είναι 0.189. Επομένως σε επίπεδο σημαντικότητας 5% αποδέχομαι την μηδενική υπόθεση και συμπεραίνουμε ότι τα δύο δείγματα έχουν την ίδια κατανομή. Άρα οι μεταβλητές BMI και cvd_mortality δεν συσχετίζονται.

3.5 Συνοπτική παρουσίαση των μεταβλητών που συσχετίζονται με τις μεταβλητές “allc_mortality” και “cvd_mortality”

- **Μεταβλητή allc_mortality (θνησιμότητα από διάφορες αιτίες)**

Σύμφωνα με τα παραπάνω είδαμε πως η μεταβλητή allc_mortality συσχετίζεται με τις μεταβλητές age_entr, fu και smoking.

Παρακάτω βλέπουμε μία σύντομη παρουσίαση των περιγραφικών μέτρων των μεταβλητών age_entr, fu και smoking συναρτήσει των πληθυσμών που ορίζονται από την μεταβλητή allc_mortality.

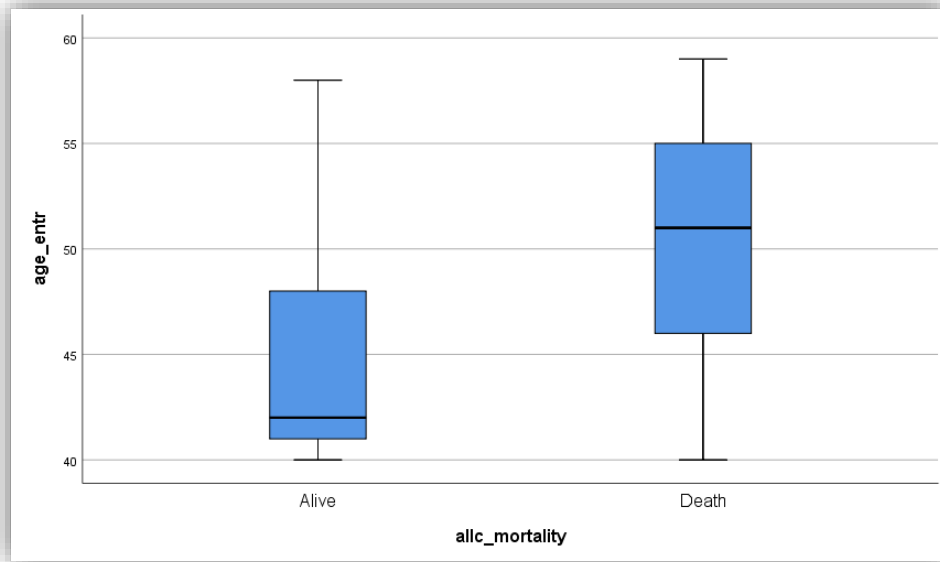
	age_entr (ηλικία εισόδου στην έρευνα)		fu (Χρόνος παραμονής στην έρευνα)		smoking (κάπνισμα)	
	Μέση τιμή	Τυπική απόκλιση	Μέση τιμή	Τυπική απόκλιση	Ποσοστό καπνιστών	Ποσοστό μη καπνιστών
allc_mortality =0 (alive)	44.79	5.083	40	0	54.4%	36.8%
allc_mortality =1 (death)	50.55	5.393	24.75	9.534	64%	24.1%
allc_mortality =(0,1) (both)	49.81	5.686	26.71	10.260	62.8%	25.7%

Πίνακας 3.16: Πίνακας παρουσίασης περιγραφικών μέτρων των μεταβλητών age_entr, fu και smoking συναρτήσει των πληθυσμών που ορίζονται από την μεταβλητή allc_mortality.

Συμπερασματικά, σύμφωνα με τον παραπάνω πίνακα, η μέση ηλικία εισαγωγής στην έρευνα των ατόμων που παρέμειναν ζωντανοί ως το τέλος της είναι 44.79 και τυπική απόκλιση ισούται με 5.083. Ο μέσος χρόνος παραμονής στην έρευνα των ατόμων αυτών είναι 40 χρόνια και η τυπική απόκλιση 0. Αυτό εξηγείται από το γεγονός ότι όποιος παρέμεινε ζωντανός από όλες τις αιτίες, έμεινε στην έρευνα για όλη την διάρκεια της. Τέλος, το ποσοστό των καπνιστών, που παρέμειναν ζωντανοί, είναι 54.4% ενώ των μη καπνιστών είναι 36.8%. Παρατηρούμε πως το άθροισμα των ποσοστών των καπνιστών και των μη καπνιστών ισούται με 91.2% των ατόμων που παρέμειναν ζωντανοί. Αυτό οφείλεται στο γεγονός ότι η μεταβλητή smoking περιέχει missing values.

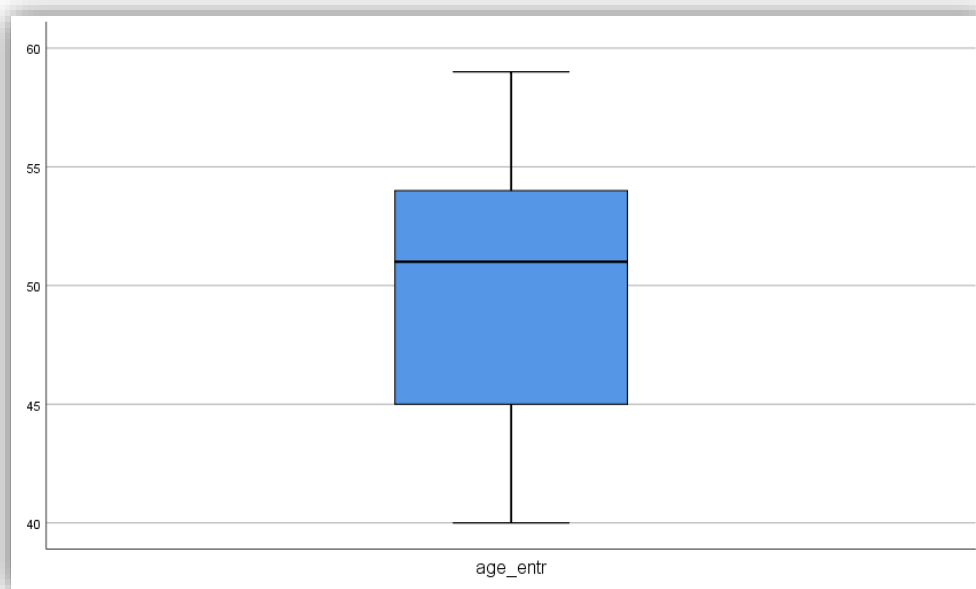
Επιπλέον, η μέση ηλικία εισαγωγής στην έρευνα των ατόμων που πέθαναν από διάφορες αιτίες είναι 50.55 και τυπική απόκλιση ισούται με 5.393. Ο μέσος χρόνος παραμονής στην έρευνα των ατόμων αυτών είναι 24.75 χρόνια και η τυπική απόκλιση 9.534. Τέλος, το ποσοστό των καπνιστών, που πέθαναν από διάφορες αιτίες, είναι 64% ενώ των μη καπνιστών είναι 24.1%. Παρατηρούμε πως το άθροισμα των ποσοστών των καπνιστών και των μη καπνιστών ισούται με 88.1% των ατόμων που πέθαναν. Αυτό οφείλεται στο γεγονός ότι η μεταβλητή smoking περιέχει missing values.

- Παρουσίαση θηκογραμμάτων της μεταβλητής `age_entr` για αυτούς που πέθαναν από διάφορες αιτίες και για αυτούς που παρέμειναν ζωντανοί



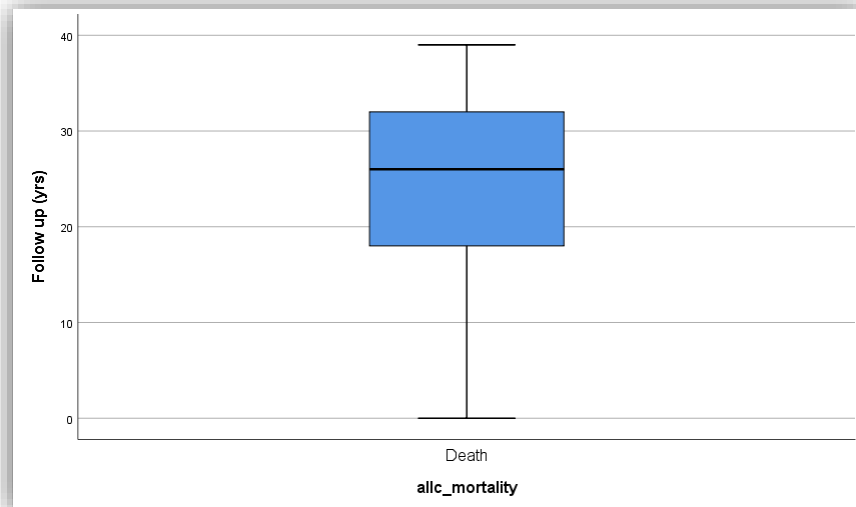
Διάγραμμα 3.1: box plots της μεταβλητής `age_entr` για αυτούς που πέθαναν από διάφορες αιτίες και για αυτούς που παρέμειναν ζωντανοί

- Παρουσίαση θηκογράμματος της μεταβλητής `age_entr` για όλο τον πληθυσμό



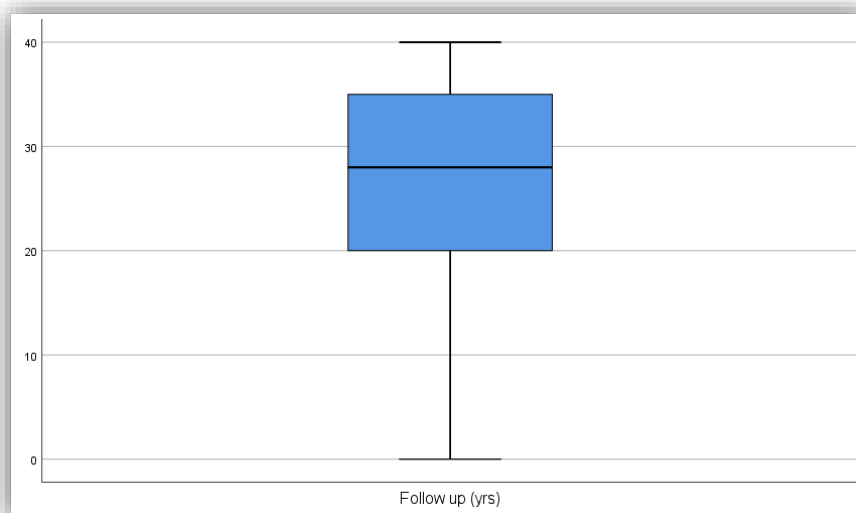
Διάγραμμα 3.2: box plot της μεταβλητής `age_entr` για όλο τον πληθυσμό

- Παρουσίαση θηκογραμμάτων της μεταβλητής fu για αυτούς που πέθαναν από διάφορες αιτίες



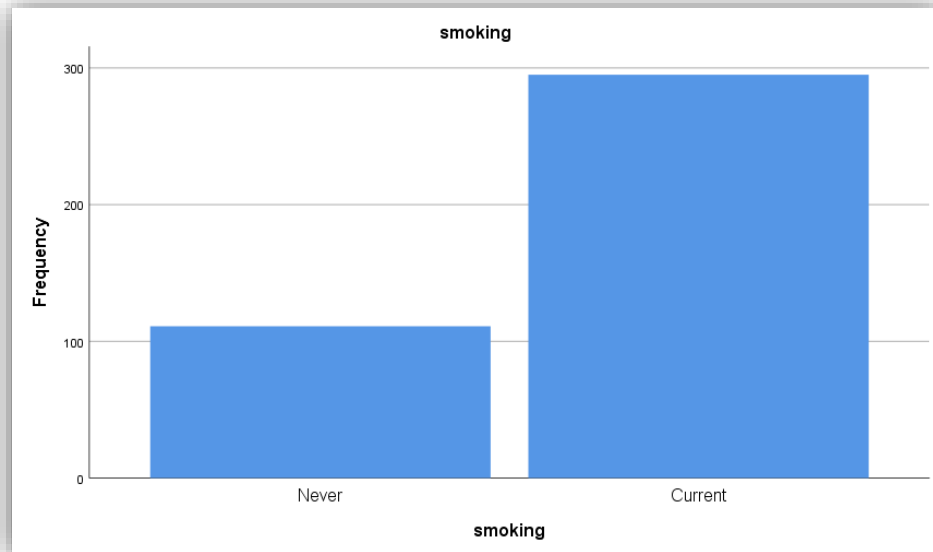
Διάγραμμα 3.3: box plot της μεταβλητής fu για αυτούς που πέθαναν από διάφορες αιτίες

- Παρουσίαση θηκογράμματος της μεταβλητής fu για όλο τον πληθυσμό



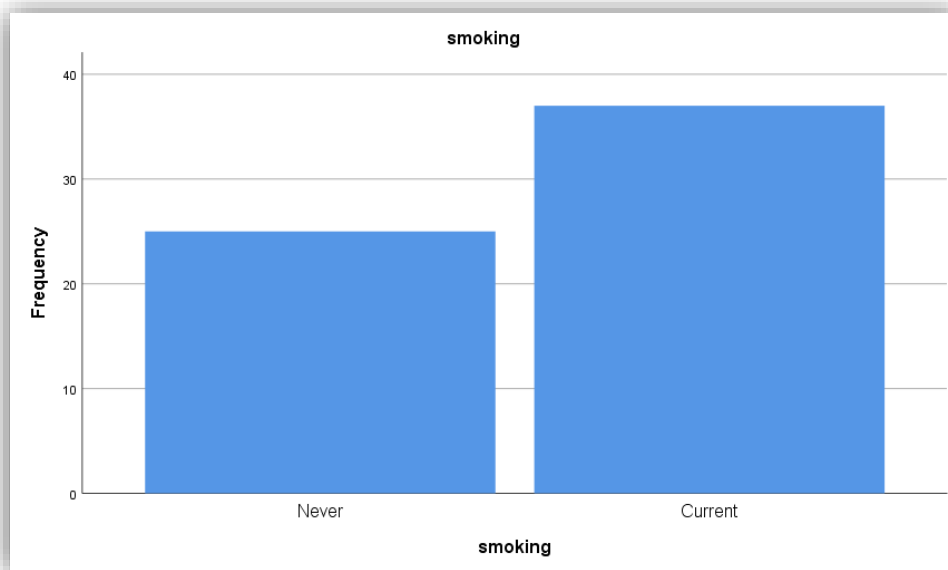
Διάγραμμα 3.4: box plot της μεταβλητής fu για όλο τον πληθυσμό

- Παρουσίαση ραβδόγραμματος της μεταβλητής smoking για αυτούς που πέθαναν από διάφορες αιτίες (allc_mortality=1)



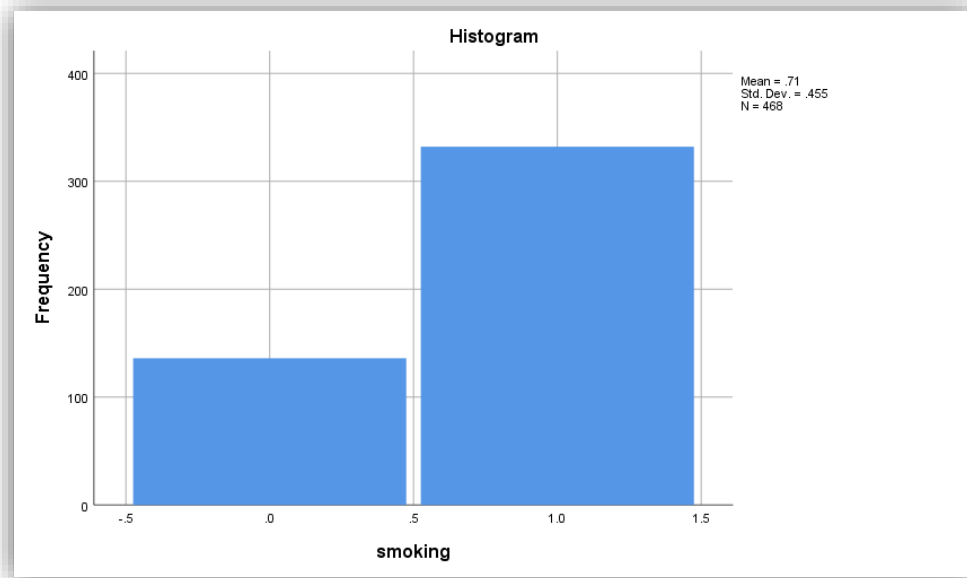
Διάγραμμα 3.5: ραβδόγραμμα της μεταβλητής smoking για αυτούς που πέθαναν από διάφορες αιτίες (allc_mortality=1)

- Παρουσίαση ραβδόγραμματος της μεταβλητής smoking για αυτούς που παρέμειναν ζωντανοί (allc_mortality=0)



Διάγραμμα 3.6: ραβδόγραμμα της μεταβλητής smoking για αυτούς που παρέμειναν ζωντανοί (allc_mortality=0)

- Παρουσίαση ραβδογράμματος της μεταβλητής smoking για όλο τον πληθυσμό



Διάγραμμα 3.7: ραβδόγραμμα της μεταβλητής smoking για όλο τον πληθυσμό

- **Μεταβλητή `cvd_mortality` (θνησιμότητα από καρδιαγγειακά)**

Σύμφωνα με τα παραπάνω είδαμε πως η μεταβλητή `cvd_mortality` συσχετίζεται με τις μεταβλητές `age_entr`, `sbp`, `dbp` και `fu`.

Παρακάτω βλέπουμε μία σύντομη παρουσίαση των περιγραφικών μέτρων των μεταβλητών `age_entr`, `sbp`, `dbp` και `fu` συναρτήσει των πληθυσμών που ορίζονται από την μεταβλητή `cvd_mortality`.

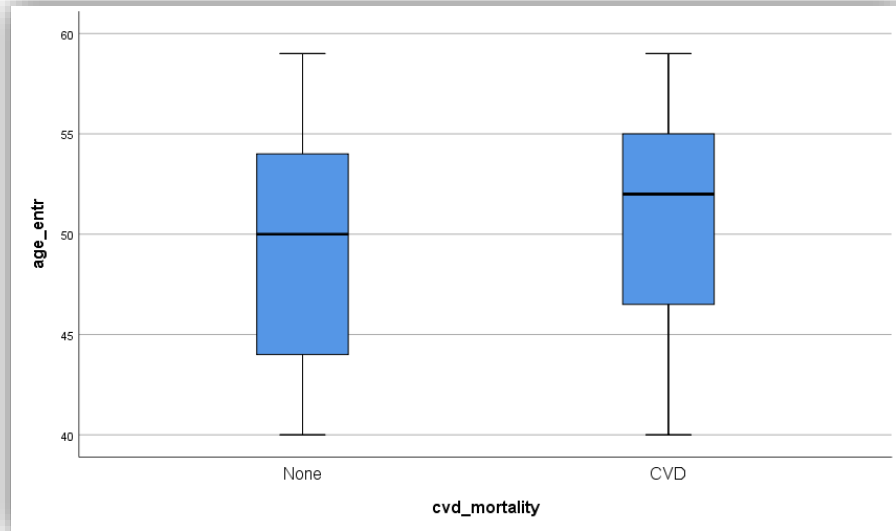
	age_entr (ηλικία εισόδου στην έρευνα)		sbp (Συστολική αρτηριακή πίεση)		dbp (Διαστολική αρτηριακή πίεση)		fu (Χρόνος παραμονής στην έρευνα)	
	Μέση τιμή	Τυπική απόκλιση	Μέση τιμή	Τυπική απόκλιση	Μέση τιμή	Τυπική απόκλιση	Μέση τιμή	Τυπική απόκλιση
<code>cvd_mortality =0</code> (None)	49.23	5.748	132.71	19.270	81.76	10.387	27.41	10.834
<code>cvd_mortality =1</code> (CVD)	50.82	5.445	139.24	23.661	84.23	12.999	25.48	9.064
<code>cvd_mortality =(0,1)</code> (both)	49.81	5.686	135.08	21.183	82.66	11.454	26.71	10.260

Πίνακας 3.17: Πίνακας παρουσίασης περιγραφικών μέτρων των μεταβλητών `age_entr`, `sbp`, `dbp` και `fu` συναρτήσει των πληθυσμών που ορίζονται από την μεταβλητή `cvd_mortality`.

Συμπερασματικά, σύμφωνα με τον παραπάνω πίνακα, η μέση ηλικία εισαγωγής στην έρευνα των ατόμων που δεν πέθαναν από καρδιαγγειακά νοσήματα είναι 49.23 και τυπική απόκλιση ισούται με 5.748. Η μέση συστολική αρτηριακή πίεση αυτών των ατόμων είναι 132.71 και η τυπική απόκλιση 19.270. Η μέση διαστολική αρτηριακή πίεση είναι 81.76 και η τυπική απόκλιση 10.387. Τέλος, ο μέσος χρόνος παραμονής στην έρευνα αυτών των ατόμων είναι 27.41 χρόνια και η τυπική απόκλιση 10.834.

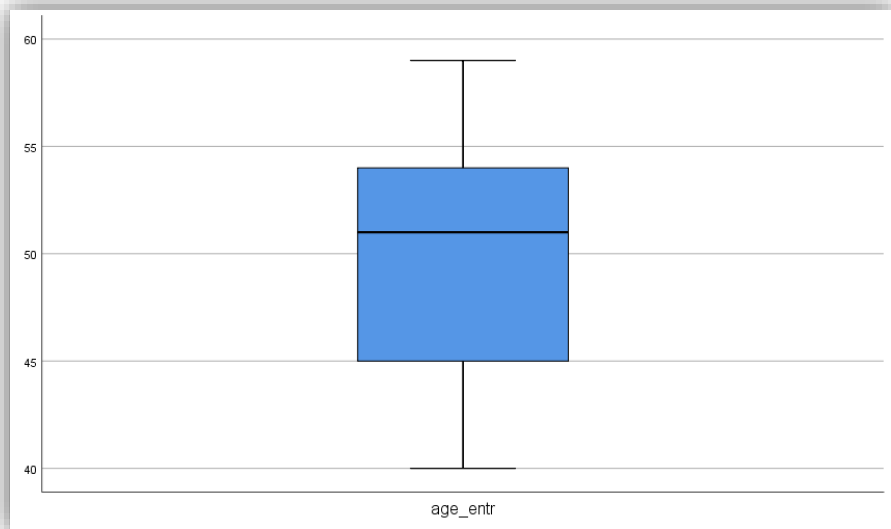
Επίσης, σύμφωνα με τον παραπάνω πίνακα, η μέση ηλικία εισαγωγής στην έρευνα των ατόμων που πέθαναν από καρδιαγγειακά νοσήματα είναι 50.82 και τυπική απόκλιση ισούται με 5.445. Η μέση συστολική αρτηριακή πίεση αυτών των ατόμων είναι 139.24 και η τυπική απόκλιση 23.661. Η μέση διαστολική αρτηριακή πίεση είναι 84.23 και η τυπική απόκλιση 12.999. Τέλος, ο μέσος χρόνος παραμονής αυτών των ατόμων στην έρευνα είναι 25.48 χρόνια και η τυπική απόκλιση 9.064.

- Παρουσίαση θηκογραμμάτων της μεταβλητής `age_entr` για αυτούς που πέθαναν από καρδιαγγειακά νοσήματα και για αυτούς που δεν πέθαναν



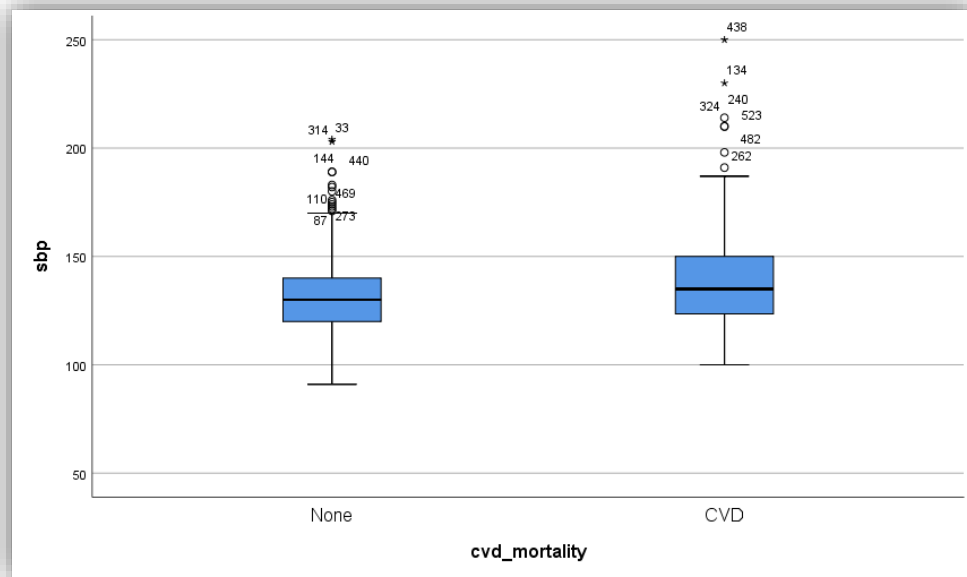
Διάγραμμα 3.8: box plots της μεταβλητής `age_entr` για αυτούς που πέθαναν από καρδιαγγειακά νοσήματα και για αυτούς που δεν πέθαναν

- Παρουσίαση θηκογράμματος της μεταβλητής `age_entr` για όλο τον πληθυσμό



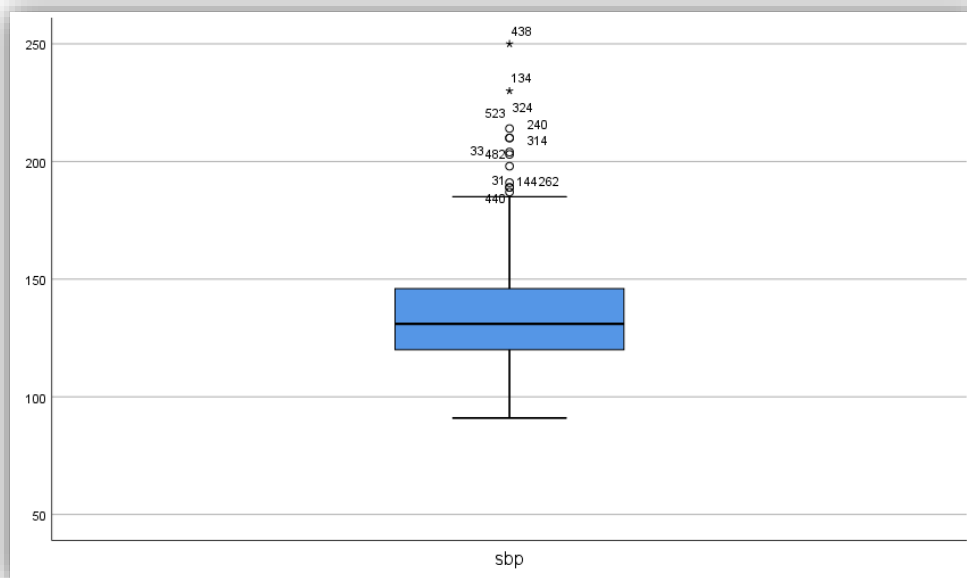
Διάγραμμα 3.9: box plot της μεταβλητής `age_entr` για όλο τον πληθυσμό

- Παρουσίαση θηκογραμμάτων της μεταβλητής sbp για αυτούς που πέθαναν από καρδιαγγειακά νοσήματα και για αυτούς που δεν πέθαναν



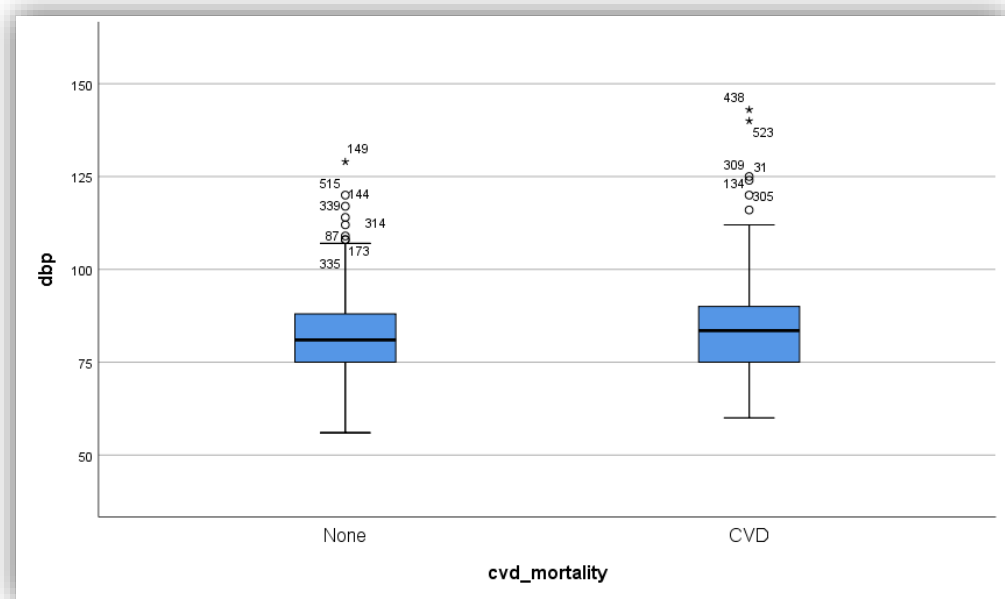
Διάγραμμα 3.10: box plots της μεταβλητής sbp για αυτούς που πέθαναν από καρδιαγγειακά νοσήματα και για αυτούς που δεν πέθαναν

- Παρουσίαση θηκογράμματος της μεταβλητής sbp για όλο τον πληθυσμό



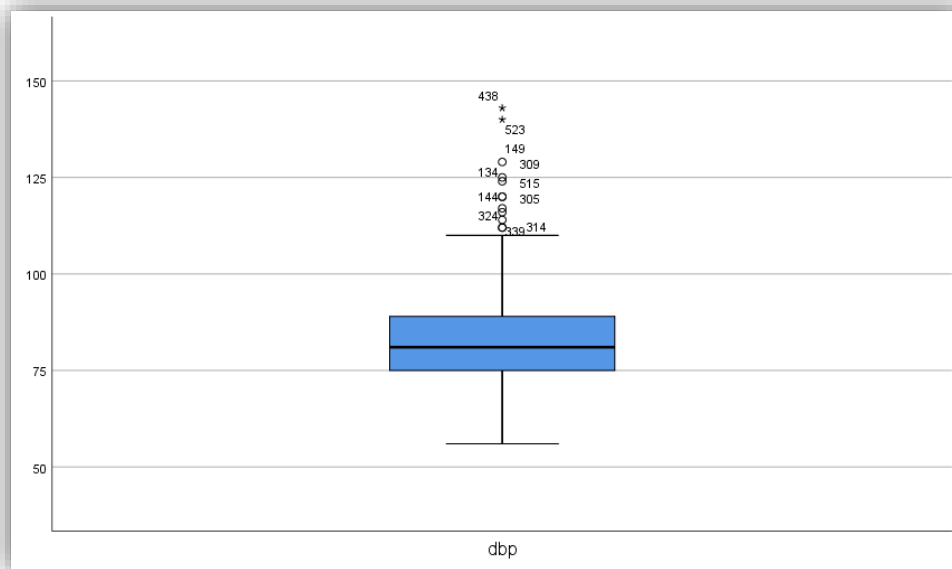
Διάγραμμα 3.11: box plot της μεταβλητής sbp για όλο τον πληθυσμό

- Παρουσίαση θηκογραμμμάτων της μεταβλητής dbp για αυτούς που πέθαναν από καρδιαγγειακά νοσήματα και για αυτούς που δεν πέθαναν



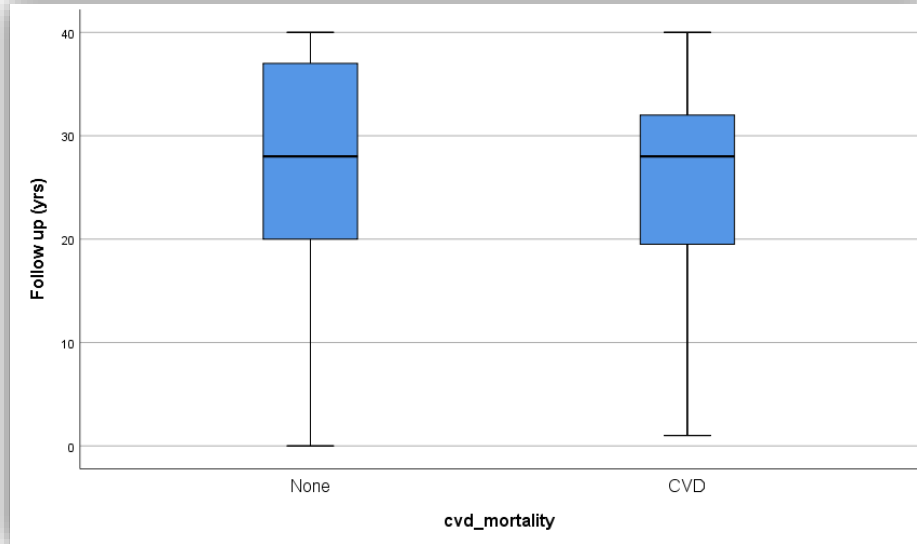
Διάγραμμα 3.12: box plots της μεταβλητής dbp για αυτούς που πέθαναν από καρδιαγγειακά νοσήματα και για αυτούς που δεν πέθαναν

- Παρουσίαση θηκογράμματος της μεταβλητής dbp για όλο τον πληθυσμό



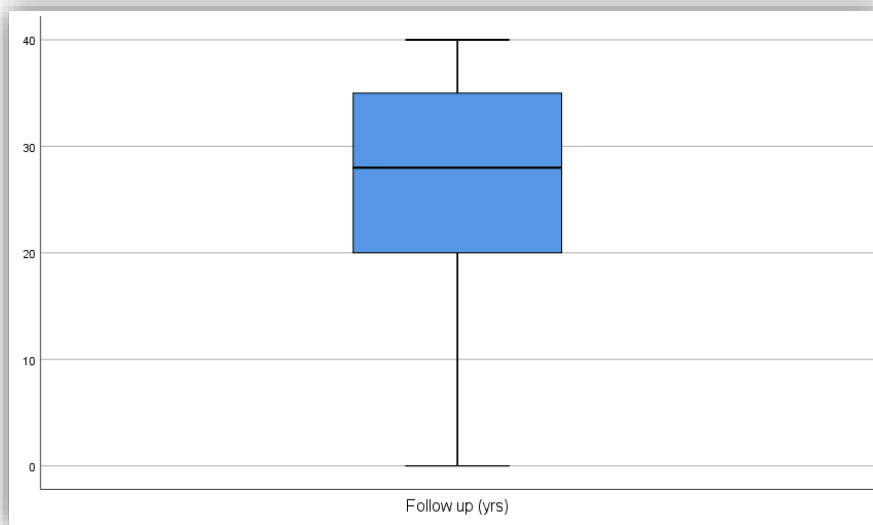
Διάγραμμα 3.13: box plot της μεταβλητής dbp για όλο τον πληθυσμό

- Παρουσίαση θηκογραμμάτων της μεταβλητής fu για αυτούς που πέθαναν από καρδιαγγειακά νοσήματα και για αυτούς που δεν πέθαναν



Διάγραμμα 3.14: box plots της μεταβλητής fu για αυτούς που πέθαναν από καρδιαγγειακά νοσήματα και για αυτούς που δεν πέθαναν

- Παρουσίαση θηκογράμματος της μεταβλητής fu για όλο τον πληθυσμό



Διάγραμμα 3.15: box plot της μεταβλητής fu για όλο τον πληθυσμό

4^ο ΚΕΦΑΛΑΙΟ

ΜΟΝΤΕΛΟ ΤΟΥ COX

4.1 Εισαγωγή

Στο κεφάλαιο αυτό θα ασχοληθούμε με την ανάλυση επιβίωσης (survival analysis) των δεδομένων μας, κάνοντας εφαρμογή του μοντέλου ανάλογων συναρτήσεων κινδύνου του Cox.

4.2 Ανάλυση Επιβίωσης

Η ανάλυση επιβίωσης (survival analysis) αναφέρεται στην ανάλυση δεδομένων που αφορούν στο χρόνο που μεσολαβεί μέχρι κάποιο συγκεκριμένο συμβάν. Αρχικά η ανάλυση αναφερόταν στο χρόνο μεταξύ της θεραπείας μέχρι τον θάνατο και για αυτό το λόγο πήρε και το συγκεκριμένο όνομα. Η ανάλυση επιβίωσης όμως μπορεί να εφαρμοστεί σε αρκετές περιπτώσεις, όπως για παράδειγμα στη μηχανολογία, για την ανάλυση του χρόνου μέχρι την εμπλοκή ενός μηχανήματος ή τη γεωργία, για την ανάλυση του χρόνου μέχρι την βγάλει καρπό ένα δέντρο. Στην περίπτωση της μηχανολογίας η ανάλυση αναφέρεται και ως θεωρία αξιοπιστίας (reliability theory). Ο χρόνος επιβίωσης χρίζει ειδικής μεταχείρισης για το λόγο ότι είναι περιορισμένος στο να είναι πάντα θετικός, και γιατί τα δεδομένα περιέχουν λογοκριμένες (censored) παρατηρήσεις. Τα λογοκριμένα δεδομένα είναι αυτά για τα οποία δεν είναι γνωστός ο χρόνος που συμβαίνει το γεγονός. Το μόνο που μπορεί να λεχθεί είναι ότι ο χρόνος επιβίωσής τους είναι μεγαλύτερος από την τιμή που έχει καταγραφεί. Στην ανάλυση επιβίωσης είναι πολύ σημαντικές δύο συναρτήσεις οι οποίες περιγράφουν την κατανομή του χρόνου επιβίωσης: η συνάρτηση επιβίωσης και η συνάρτηση κινδύνου.

4.3 Συνάρτηση Επιβίωσης

Συμβολίζοντας το χρόνο επιβίωσης με T , η συνάρτηση επιβίωσης (survival function) $S(t)$ ορίζεται ως η πιθανότητα επιβίωσης ενός ατόμου πέραν τη χρονική στιγμή t και δίνεται από τη σχέση :

$$S(t) = P(T > t) = 1 - F(t)$$

Η συνάρτηση επιβίωσης είναι μη αρνητική και μη αύξουσα συνάρτηση του t με $S(0) = 1$ και $S(\infty) = 0$. Η γραφική παράσταση της $S(t)$ συναρτήσεως του t είναι γνωστή ως καμπύλη επιβίωσης και είναι πολύ σημαντική στην ανάλυση δεδομένων χρόνου επιβίωσης.

4.4 Συνάρτηση Κινδύνου

Η συνάρτηση κινδύνου, $h(t)$, ορίζεται ως η πιθανότητα αποβίωσης (ή πραγμάτωσης του γεγονότος που εξετάζεται) τη χρονική στιγμή t , δεδομένου ότι το άτομο έχει επιβιώσει μέχρι τη χρονική στιγμή t . Δηλαδή

$$h(t) = \lim_{s \rightarrow 0} \frac{P(t \leq T \leq t+s \mid T \geq t)}{s}$$

Η συνάρτηση κινδύνου δίνει ένα μέτρο του πόσο πιθανό είναι ένα άτομο να αποβιώσει ως συνάρτηση της ηλικίας του ατόμου.

4.5 Μοντέλο ανάλογων συναρτήσεων κινδύνου

Στην ανάλυση επιβίωσης παίζει μεγάλο ρόλο η εξεύρεση παραγόντων οι οποίοι να σχετίζονται με το χρόνο επιβίωσης. Αυτοί οι παράγοντες θα πρέπει να συμπεριληφθούν στο μοντέλο που θα χρησιμοποιηθεί για τη σχετική ανάλυση των δεδομένων. Αφού η συνάρτηση κινδύνου είναι μη αρνητική, ο λογάριθμος της μπορεί να εκφραστεί ως γραμμική συνάρτηση επεξηγηματικών μεταβλητών :

$$\ln h(t) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

Το μοντέλο αυτό όμως είναι πολύ περιοριστικό αφού υποθέτει ότι η συνάρτηση κινδύνου δεν εξαρτάται από το χρόνο. Υπάρχουν διάφορες μέθοδοι με τις οποίες το μοντέλο θα μπορούσε να υιοθετήσει την εξάρτηση του χρόνου, με την πιο γνωστή να είναι το μοντέλο ανάλογων συναρτήσεων κινδύνου (Cox 1972). Το μοντέλο αυτό δίνεται από

$$\ln h(t) = \alpha(t) + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

όπου $\alpha(t)$ είναι οποιαδήποτε συνάρτηση του χρόνου. Ο όρος "ανάλογων συναρτήσεων κινδύνου" είναι λόγω του γεγονότος ότι για οποιαδήποτε άτομα για οποιοδήποτε σημείο του χρόνου, ο λόγος των συναρτήσεων κινδύνου είναι σταθερός. Εξαιτίας του ότι η συνάρτηση κινδύνου $\alpha(t)$ δεν είναι ανάγκη να οριστεί εξ ολοκλήρου, το μοντέλο ανάλογων συναρτήσεων κινδύνου θεωρείται ως ημιπαραμετρικό. Ο Cox εισηγήθηκε μια μέθοδο δεσμευμένης πιθανοφάνειας για εκτίμηση των παραμέτρων. Το σημαντικό στοιχείο αυτής της μεθόδου είναι οι εκτιμήσεις εξαρτώνται από τη σειρά με την οποία συμβαίνει το γεγονός και όχι τον ακριβή χρόνο.

4.6 Θνησιμότητα από διάφορες αιτίες (allc_mortality)

Εφαρμόζουμε την μέθοδο Cox θεωρώντας ότι η μεταβλητή fu είναι ο χρόνος μέχρι να συμβεί το γεγονός του θανάτου, η μεταβλητή allc_mortality θα είναι το status με την τιμή 0 να είναι λογοκριμένη και την 1 full. Οι μεταβλητές age_entr, physical_act, sbp, dbp, heartrate, tchol, smoking και BMI θα είναι οι ανεξάρτητες μεταβλητές.

Case Processing Summary

		N	Percent
Cases available in analysis	Event ^a	402	76.0%
	Censored	62	11.7%
	Total	464	87.7%
Cases dropped	Cases with missing values	65	12.3%
	Cases with negative time	0	0.0%
	Censored cases before the earliest event in a stratum	0	0.0%
	Total	65	12.3%
Total		529	100.0%

a. Dependent Variable: Follow up (yrs)

Πίνακας 4.1: Πίνακας περιγραφής του δείγματος

Οι 402 από τις συνολικά 464 περιπτώσεις έχουν εμφανίσει το ενδεχόμενο (θάνατος) ενώ οι υπόλοιπες 62 είναι λογοκριμένοι χρόνοι επιβίωσης. Υπάρχουν 65 missing values, κανένας αρνητικός χρόνος και καμιά λογοκριμένη περίπτωση πριν τον πρώτο θάνατο σε κάποιο από τα στρώματα.

Omnibus Tests of Model Coefficients^a

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
4289.425	119.808	8	.000	118.811	8	.000	118.811	8	.000

a. Beginning Block Number 1. Method = Enter

Πίνακας 4.2: Πίνακας δοκιμών γενικών συντελεστών μοντέλου

Επειδή η μεταβολή στην αλλαγή του chi-square είναι σημαντική ($p < 0,001$), το μοντέλο προσαρμόζεται καλά. Δηλαδή, τουλάχιστον μια από τις ερμηνευτικές μεταβλητές επηρεάζει σημαντικά τον χρόνο επιβίωσης.

Variables in the Equation								
	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
age_entr	.091	.010	89.743	1	.000	1.095	1.075	1.116
physical_act	-.219	.114	3.703	1	.054	.803	.642	1.004
sbp	.004	.004	.774	1	.379	1.004	.996	1.011
dbp	.005	.008	.377	1	.539	1.005	.990	1.020
heartrate	.007	.004	2.526	1	.112	1.007	.998	1.016
tchol	-.003	.001	4.636	1	.031	.997	.995	1.000
smoking	.475	.116	16.661	1	.000	1.608	1.280	2.019
BMI	-.018	.016	1.293	1	.256	.982	.951	1.013

Πίνακας 4.3: Πίνακας εξίσωσης μεταβλητών

- Για την μεταβλητή age_entr η στατιστική συνάρτηση του ελέγχου Wald ισούται με 89.743 και το p-value με 0. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% απορρίπτουμε την μηδενική υπόθεση ότι το B της μεταβλητής ισούται με το 0 (ισοδύναμα ότι το Exp(B) ισούται με το 1). Αυτό σημαίνει ότι το B ισούται με 0.091.
- Για την μεταβλητή physical_act η στατιστική συνάρτηση του ελέγχου Wald ισούται με 3.703 και το p-value με 0.054. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% αποδεχόμαστε την μηδενική υπόθεση ότι το B της μεταβλητής ισούται με το 0 (ισοδύναμα ότι το Exp(B) ισούται με το 1)
- Για την μεταβλητή sbp η στατιστική συνάρτηση του ελέγχου Wald ισούται με 0.774 και το p-value με 0.379. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% αποδεχόμαστε την μηδενική υπόθεση ότι το B της μεταβλητής ισούται με το 0 (ισοδύναμα ότι το Exp(B) ισούται με το 1)
- Για την μεταβλητή dbp η στατιστική συνάρτηση του ελέγχου Wald ισούται με 0.377 και το p-value με 0.539. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% αποδεχόμαστε την μηδενική υπόθεση ότι το B της μεταβλητής ισούται με το 0 (ισοδύναμα ότι το Exp(B) ισούται με το 1).

- Για την μεταβλητή *heartrate* η στατιστική συνάρτηση του ελέγχου Wald ισούται με 2.526 και το p-value με 0.112. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% αποδεχόμαστε την μηδενική υπόθεση ότι το B της μεταβλητής ισούται με το 0 (ισοδύναμα ότι το $\text{Exp}(B)$ ισούται με το 1).
- Για την μεταβλητή *tchol* η στατιστική συνάρτηση του ελέγχου Wald ισούται με 4.636 και το p-value με 0.031. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% απορρίπτουμε την μηδενική υπόθεση ότι το B της μεταβλητής ισούται με το 0 (ισοδύναμα ότι το $\text{Exp}(B)$ ισούται με το 1). Αυτό σημαίνει ότι το B ισούται με -0.003.
- Για την μεταβλητή *smoking* η στατιστική συνάρτηση του ελέγχου Wald ισούται με 16.661 και το p-value με 0. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% απορρίπτουμε την μηδενική υπόθεση ότι το B της μεταβλητής ισούται με το 0 (ισοδύναμα ότι το $\text{Exp}(B)$ ισούται με το 1). Αυτό σημαίνει ότι το B ισούται με 0.475.
- Για την μεταβλητή *BMI* η στατιστική συνάρτηση του ελέγχου Wald ισούται με 1.293 και το p-value με 0.256. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% αποδεχόμαστε την μηδενική υπόθεση ότι το B της μεταβλητής ισούται με το 0 (ισοδύναμα ότι το $\text{Exp}(B)$ ισούται με το 1).

Με βάση τα παραπάνω έχω πως το μοντέλο ανάλογων συναρτήσεων κινδύνου του Cox δίνεται από την σχέση :

$$h(t/Z) = h_0(t) \exp(0.091age_entr - 0.003tchol + 0.475smoking) , \text{ για κάθε } t \geq 0$$

Από το παραπάνω μοντέλο συμπεραίνουμε ότι:

- Για αύξηση μίας μονάδας της ηλικίας που κάποιος μπήκε στην έρευνα (age_entr), η συνάρτηση κινδύνου πολλαπλασιάζεται με τον εκθετικό συντελεστή $e^{0.091}=1.095$.
- Για αύξηση μίας μονάδας της ολικής χοληστερόλης (tchol), η συνάρτηση κινδύνου πολλαπλασιάζεται με τον εκθετικό συντελεστή $e^{-0.003}=0.997$.
- Όταν κάποιος είναι καπνιστής (smoking=1), η συνάρτηση κινδύνου πολλαπλασιάζεται με τον εκθετικό συντελεστή $e^{0.475}=1.608$.

4.7 Θνησιμότητα από καρδιαγγειακά νοσήματα (cvd_mortality)

Εφαρμόζουμε την μέθοδο Cox θεωρώντας ότι η μεταβλητή fu είναι ο χρόνος μέχρι να συμβεί το γεγονός του θανάτου, η μεταβλητή cvd_mortality θα είναι το status με την τιμή 0 να είναι λογοκριμένη και την 1 full. Οι μεταβλητές age_entr, physical_act, sbp, dbp, heartrate, tchol, smoking και BMI θα είναι οι ανεξάρτητες μεταβλητές.

Case Processing Summary

		N	Percent
Cases available in analysis	Event ^a	162	30.6%
	Censored	299	56.5%
	Total	461	87.1%
Cases dropped	Cases with missing values	65	12.3%
	Cases with negative time	0	0.0%
	Censored cases before the earliest event in a stratum	3	0.6%
	Total	68	12.9%
Total		529	100.0%

a. Dependent Variable: Follow up (yrs)

Πίνακας 4.4: Πίνακας περιγραφής του δείγματος

Οι 162 από τις συνολικά 461 περιπτώσεις έχουν εμφανίσει το ενδεχόμενο (θάνατος) ενώ οι υπόλοιπες 299 είναι λογοκριμένοι χρόνοι επιβίωσης. Υπάρχουν 65 missing values, κανένας

αρνητικός χρόνος και 3 λογοκριμένες περιπτώσεις πριν τον πρώτο θάνατο σε κάποιο από τα στρώματα.

Omnibus Tests of Model Coefficients^a

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
1707.497	59.087	8	.000	57.224	8	.000	57.224	8	.000

a. Beginning Block Number 1. Method = Enter

Πίνακας 4.5: Πίνακας δοκιμών γενικών συντελεστών μοντέλου

Επειδή η μεταβολή στην αλλαγή του chi-square είναι σημαντική ($p < 0,001$), το μοντέλο προσαρμόζεται καλά. Δηλαδή, τουλάχιστον μια από τις ερμηνευτικές μεταβλητές επηρεάζει σημαντικά τον χρόνο επιβίωσης.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
age_entr	.096	.015	40.233	1	.000	1.101	1.069	1.134
physical_act	-.031	.184	.028	1	.866	.969	.675	1.391
sbp	.014	.006	5.623	1	.018	1.015	1.003	1.027
dbp	-.010	.012	.693	1	.405	.990	.967	1.014
heartrate	.007	.007	1.197	1	.274	1.007	.994	1.021
tchol	-.002	.002	1.342	1	.247	.998	.994	1.002
smoking	.320	.176	3.295	1	.069	1.377	.975	1.944
BMI	.006	.025	.068	1	.795	1.006	.959	1.057

Πίνακας 4.6: Πίνακας εξίσωσης μεταβλητών

- Για την μεταβλητή age_entr η στατιστική συνάρτηση του ελέγχου Wald ισούται με 40.233 και το p-value με 0. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% απορρίπτουμε την μηδενική υπόθεση ότι το B της μεταβλητής ισούται με το 0 (ισοδύναμα ότι το Exp(B) ισούται με το 1). Αυτό σημαίνει ότι το B ισούται με 0.096.
- Για την μεταβλητή physical_act η στατιστική συνάρτηση του ελέγχου Wald ισούται με 0.028 και το p-value με 0.866. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5%

αποδεχόμαστε την μηδενική υπόθεση ότι το B της μεταβλητής ισούται με το 0 (ισοδύναμα ότι το $\text{Exp}(B)$ ισούται με το 1).

- Για την μεταβλητή sbr η στατιστική συνάρτηση του ελέγχου Wald ισούται με 5.623 και το p -value με 0.018. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% απορρίπτουμε την μηδενική υπόθεση ότι το B της μεταβλητής ισούται με το 0 (ισοδύναμα ότι το $\text{Exp}(B)$ ισούται με το 1). Αυτό σημαίνει ότι το B ισούται με 0.014.
- Για την μεταβλητή dbp η στατιστική συνάρτηση του ελέγχου Wald ισούται με 0.693 και το p -value με 0.405. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% αποδεχόμαστε την μηδενική υπόθεση ότι το B της μεταβλητής ισούται με το 0 (ισοδύναμα ότι το $\text{Exp}(B)$ ισούται με το 1).
- Για την μεταβλητή $heartrate$ η στατιστική συνάρτηση του ελέγχου Wald ισούται με 1.197 και το p -value με 0.274. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% αποδεχόμαστε την μηδενική υπόθεση ότι το B της μεταβλητής ισούται με το 0 (ισοδύναμα ότι το $\text{Exp}(B)$ ισούται με το 1).
- Για την μεταβλητή $tchol$ η στατιστική συνάρτηση του ελέγχου Wald ισούται με 1.342 < και το p -value με 0.247. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% αποδεχόμαστε την μηδενική υπόθεση ότι το B της μεταβλητής ισούται με το 0 (ισοδύναμα ότι το $\text{Exp}(B)$ ισούται με το 1).
- Για την μεταβλητή $smoking$ η στατιστική συνάρτηση του ελέγχου Wald ισούται με 3.295 και το p -value με 0.069. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% αποδεχόμαστε την μηδενική υπόθεση ότι το B της μεταβλητής ισούται με το 0 (ισοδύναμα ότι το $\text{Exp}(B)$ ισούται με το 1).
- Για την μεταβλητή BMI η στατιστική συνάρτηση του ελέγχου Wald ισούται με 0.068 < και το p -value με 0.795. Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% αποδεχόμαστε την μηδενική υπόθεση ότι το B της μεταβλητής ισούται με το 0 (ισοδύναμα ότι το $\text{Exp}(B)$ ισούται με το 1).

Με βάση τα παραπάνω έχω πως το μοντέλο ανάλογων συναρτήσεων κινδύνου του Cox δίνεται από την σχέση :

$$h(t/Z) = h_0(t) \exp(0.096age_entr + 0.014sbp) , \text{ για κάθε } t \geq 0$$

Από το παραπάνω μοντέλο συμπεραίνουμε ότι:

- Για αύξηση μίας μονάδας της ηλικίας που κάποιος μπήκε στην έρευνα (age_entr), η συνάρτηση κινδύνου πολλαπλασιάζεται με τον εκθετικό συντελεστή $e^{0.096}=1.101$.
- Για αύξηση μίας μονάδας της συστολικής αρτηριακής πίεσης (sbp), η συνάρτηση κινδύνου πολλαπλασιάζεται με τον εκθετικό συντελεστή $e^{0.014}=1.015$.

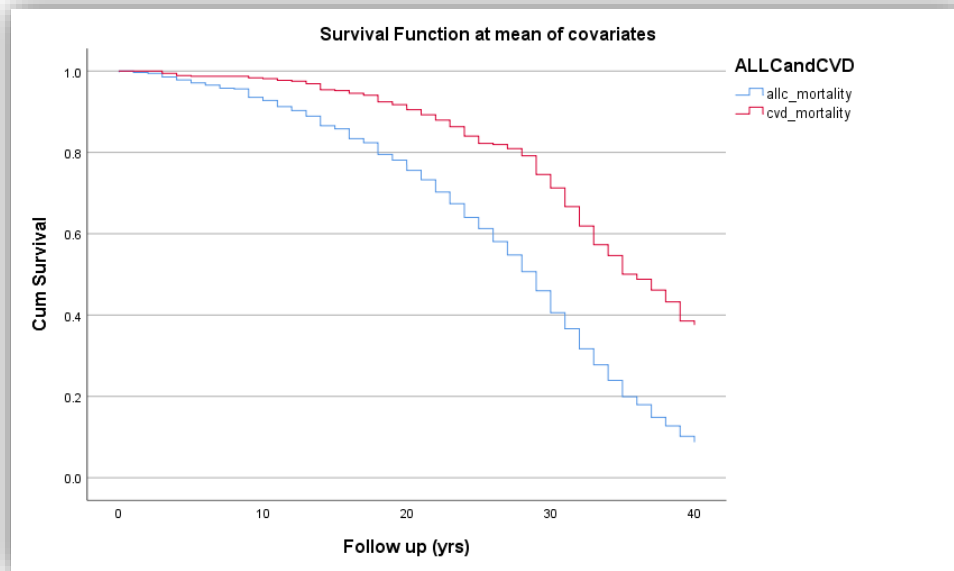
4.8 Σύγκριση των γραφικών παραστάσεων των συναρτήσεων επιβίωσης και των συναρτήσεων κινδύνου για τους δύο πληθυσμούς που ορίζονται από τις μεταβλητές `allc_mortality` και `cvd_mortality`.

Το δείγμα μας αποτελείται από 529 παρατηρήσεις. Προκειμένου να πραγματοποιήσουμε την σύγκριση των γραφικών παραστάσεων της συνάρτησης επιβίωσης και της συνάρτησης κινδύνου για τους δύο πληθυσμούς που ορίζονται από τις μεταβλητές `allc_mortality` και `cvd_mortality` προχωρήσαμε στην παρακάτω διαδικασία.

Αρχικά, μετατρέψαμε το δείγμα μας σε 1058 παρατηρήσεις. Αυτό επιτεύχθηκε με την αντιγραφή των 529 παρατηρήσεων κάτω από τις ήδη υπάρχουσες 529 παρατηρήσεις, στο ίδιο dataset. Στη συνέχεια, δημιουργήσαμε δύο νέες μεταβλητές. Η πρώτη κατά σειρά ονομάζεται “mortality” και η δεύτερη “ALLCandCVD”. Για την κατασκευή της μεταβλητής “mortality”, στις πρώτες 529 παρατηρήσεις βάλαμε τις αντίστοιχες τιμές της μεταβλητής “allc_mortality” και στις υπόλοιπες 529 τις τιμές της μεταβλητής “cvd_mortality”. Έτσι, η μεταβλητή “mortality” περιέχει τις τιμές 0 και 1. Η τιμή 1 συμβολίζει τον θάνατο από διάφορες αιτίες καθώς και καρδιαγγειακά νοσήματα και η τιμή 0 συμβολίζει το εάν κάποιος επέζησε από καρδιαγγειακά νοσήματα ή εάν έμεινε ζωντανός μέχρι το τέλος της έρευνας. Για την κατασκευή της μεταβλητής “ALLCandCVD” για τις πρώτες 529 παρατηρήσεις βάλαμε την τιμή 1 και για τις υπόλοιπες 529 την τιμή 2. Με αυτόν τον τρόπο, η μεταβλητή “ALLCandCVD” περιέχει ξεχωριστά τους θανάτους από διάφορες αιτίες, που συμβολίσαμε με 1 και από καρδιαγγειακά νοσήματα, που συμβολίσαμε με 2.

Συμπερασματικά, έχοντας πλέον δημιουργήσει αυτές τις δύο μεταβλητές, κατασκευάζουμε τις γραφικές παραστάσεις συνάρτησης επιβίωσης και συνάρτησης κινδύνου, βάσει του μοντέλου αναλογικού κινδύνου του Cox, όπως φαίνεται παρακάτω.

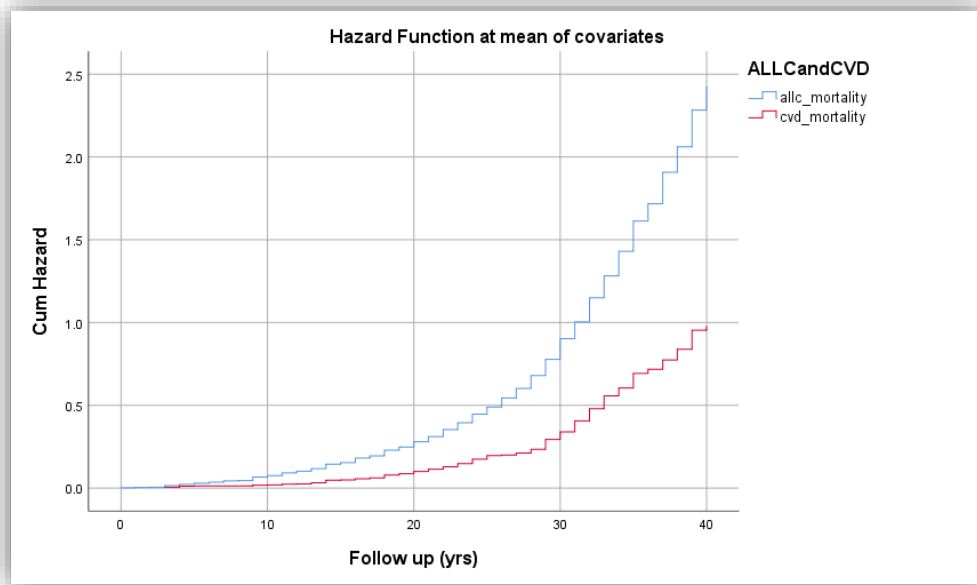
- Παρουσίαση συναρτήσεων επιβίωσης για τους πληθυσμούς `allc_mortality` και `cvd_mortality`



Διάγραμμα 4.1: Διάγραμμα συναρτήσεων επιβίωσης για τους πληθυσμούς `allc_mortality` και `cvd_mortality`.

Από το παραπάνω διάγραμμα παρατηρούμε ότι οι δύο γραφικές παραστάσεις είναι φθίνουσες. Αυτό σημαίνει ότι όσο αυξάνει ο χρόνος παραμονής ενός ατόμου στην έρευνα, η πιθανότητα επιβίωσης από καρδιαγγειακά νοσήματα και από διάφορες αιτίες μειώνεται. Επίσης, παρατηρούμε ότι η γραφική παράσταση της συνάρτησης επιβίωσης του πληθυσμού που ορίζεται από την μεταβλητή “`cvd_mortality`” βρίσκεται πάνω από την γραφική παράσταση της συνάρτησης επιβίωσης του πληθυσμού που ορίζεται από την μεταβλητή “`allc_mortality`”. Αυτό εξηγεί το γεγονός ότι η πιθανότητα επιβίωσης ενός ατόμου από καρδιαγγειακά νοσήματα είναι μεγαλύτερη από την πιθανότητα επιβίωσης ενός ατόμου από διάφορες αιτίες. Τέλος, βλέπουμε πως όσο αυξάνεται ο χρόνος παραμονής ενός ατόμου στην έρευνα, οι γραφικές παραστάσεις των συναρτήσεων επιβίωσης αποκτούν μεγαλύτερη απόσταση μεταξύ τους. Αυτό πρακτικά σημαίνει ότι όσο αυξάνεται ο χρόνος παραμονής ενός ατόμου στην έρευνα τόσο πιο πιθανή είναι η επιβίωση από τα καρδιαγγειακά νοσήματα σε σχέση με την επιβίωση από διάφορες αιτίες.

- Παρουσίαση συναρτήσεων κινδύνου για τους πληθυσμούς `allc_mortality` και `cvd_mortality`



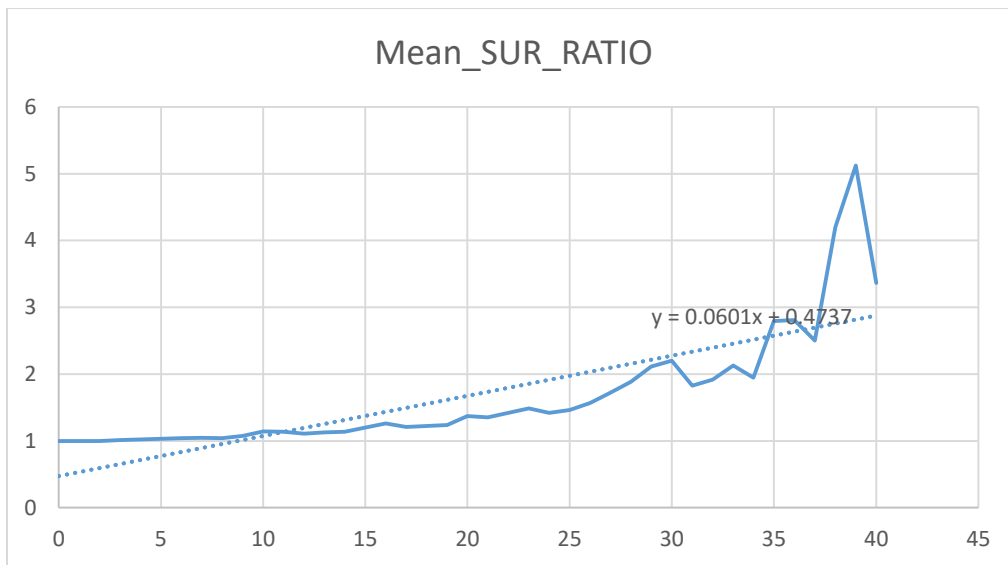
Διάγραμμα 4.2: Διάγραμμα συναρτήσεων κινδύνου για τους πληθυσμούς `allc_mortality` και `cvd_mortality`

Από το παραπάνω διάγραμμα παρατηρούμε ότι οι δύο γραφικές παραστάσεις είναι αύξουσες. Αυτό σημαίνει ότι όσο αυξάνει ο χρόνος παραμονής ενός ατόμου στην έρευνα, η πιθανότητα αποβίωσης από καρδιαγγειακά νοσήματα και από διάφορες αιτίες αυξάνεται. Επίσης, παρατηρούμε ότι η γραφική παράσταση της συνάρτησης κινδύνου του πληθυσμού που ορίζεται από την μεταβλητή “`cvd_mortality`” βρίσκεται κάτω από την γραφική παράσταση της συνάρτησης επιβίωσης του πληθυσμού που ορίζεται από την μεταβλητή “`allc_mortality`”. Αυτό εξηγεί το γεγονός ότι η πιθανότητα αποβίωσης ενός ατόμου από καρδιαγγειακά νοσήματα είναι μικρότερη από την πιθανότητα αποβίωσης ενός ατόμου από διάφορες αιτίες. Τέλος, βλέπουμε πως όσο αυξάνεται ο χρόνος παραμονής ενός ατόμου στην έρευνα, οι γραφικές παραστάσεις των συναρτήσεων κινδύνου αποκτούν μεγαλύτερη απόσταση μεταξύ τους. Αυτό πρακτικά σημαίνει ότι όσο αυξάνεται ο χρόνος παραμονής ενός ατόμου στην έρευνα τόσο πιο πιθανή είναι η αποβίωση από διάφορες αιτίες σε σχέση με την αποβίωση από τα καρδιαγγειακά νοσήματα.

4.9 Παρουσίαση των γραφικών παραστάσεων των μέσων τιμών των λόγων των συναρτήσεων επιβίωσης και των συναρτήσεων κινδύνου.

Μέσω της διαδικασίας του Cox Regression στο SPSS με status το allc_mortality υπολογίζονται οι τιμές της συνάρτησης επιβίωσης (SUR_allc) και της συνάρτησης κινδύνου (HAZ_allc). Αντίστοιχα, από την παραπάνω διαδικασία με status το cvd_mortality υπολογίζονται οι τιμές της συνάρτησης επιβίωσης (SUR_cvd) και της συνάρτησης κινδύνου (HAZ_cvd). Προκειμένου να κατασκευάσουμε τις ζητούμενες γραφικές παραστάσεις, υπολογίσαμε το πηλίκο των συναρτήσεων επιβίωσης (SUR_cvd/ SUR_allc) το οποίο ονομάσαμε SUR_RATIO. Αντίστοιχα, υπολογίσαμε το πηλίκο των συναρτήσεων κινδύνου (HAZ_cvd/ HAZ_allc) το οποίο ονομάσαμε HAZ_RATIO. Κατόπιν, σε ένα φύλλο excel περάσαμε τις στήλες fu, SUR_RATIO και HAZ_RATIO. Χρησιμοποιώντας την γλώσσα προγραμματισμού R, δημιουργήσαμε από το παραπάνω excel ένα νέο, με στήλες τις unique τιμές της fu και τις μέσες τιμές των λόγων που αντιστοιχούν σε κάθε τιμή της fu (Mean_SUR_RATIO, Mean_HAZ_RATIO). Με βάση τις νέες στήλες που παράξαμε και σύμφωνα με τις επιλογές του excel φτιάξαμε τις παρακάτω γραφικές παραστάσεις των μέσων όρων των λόγων των συναρτήσεων επιβίωσης και των μέσων όρων των λόγων των συναρτήσεων κινδύνου.

- **Γραφική παράσταση των μέσων τιμών των λόγων των συναρτήσεων επιβίωσης**



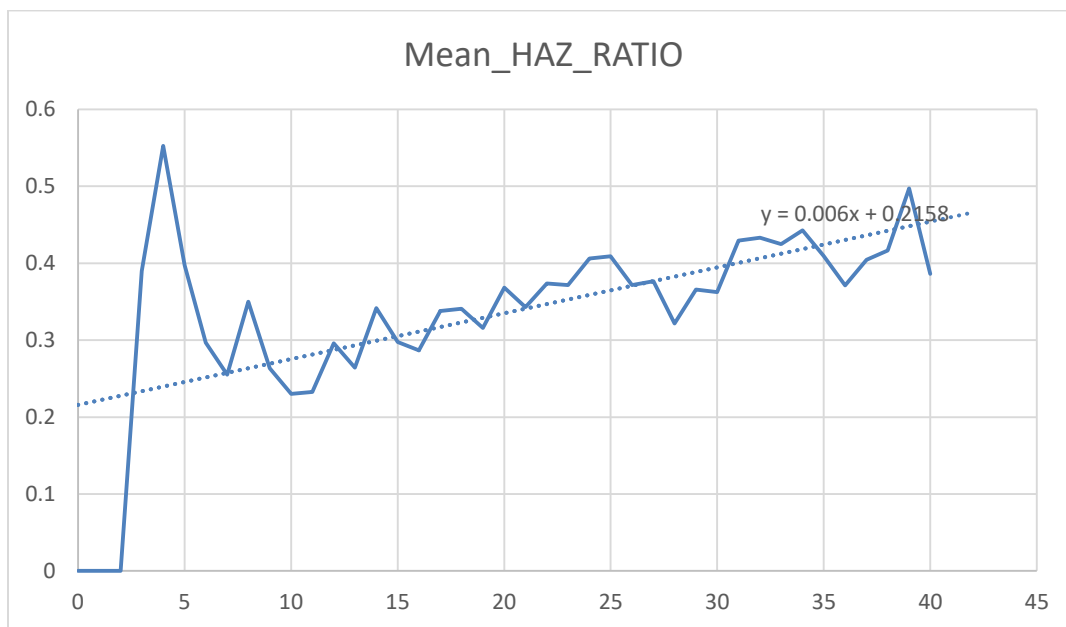
Διάγραμμα 4.3: Διάγραμμα των μέσων τιμών των λόγων των συναρτήσεων επιβίωσης

Από το παραπάνω διάγραμμα και με την βοήθεια του excel βλέπουμε ότι η κατάλληλη καμπύλη τάσης για την καμπύλη μας είναι η ευθεία

$$\text{Mean_SUR_RATIO} = 0.0601 * fu + 0.4737$$

η οποία είναι αύξουσα. Αυτό πρακτικά σημαίνει ότι η καμπύλη των μέσων τιμών των λόγων των συναρτήσεων επιβίωσης έχει αυξητική γραμμική τάση. Δηλαδή ο λόγος των μέσων τιμών των συναρτήσεων επιβίωσης αυξάνεται σταθερά με την πιθανότητα επιβίωσης ενός ατόμου από καρδιαγγειακά νοσήματα να αυξάνει πιο γρήγορα και να είναι μεγαλύτερη από την πιθανότητα επιβίωσης ενός ατόμου από διάφορες αιτίες.

- **Γραφική παράσταση των μέσων τιμών των λόγων των συναρτήσεων κινδύνου**



Διάγραμμα 4.4: Διάγραμμα των μέσων τιμών των λόγων των συναρτήσεων κινδύνου

Από το παραπάνω διάγραμμα και με την βοήθεια του excel βλέπουμε ότι η κατάλληλη καμπύλη τάσης για την καμπύλη μας είναι η ευθεία

$$\text{Mean_HAZ_RATIO} = 0.006 * fu + 0.2158$$

η οποία είναι αύξουσα. Αυτό πρακτικά σημαίνει ότι η καμπύλη των μέσων τιμών των λόγων των συναρτήσεων κινδύνου έχει αυξητική γραμμική τάση. Δηλαδή ο λόγος των μέσων τιμών των συναρτήσεων κινδύνου αυξάνεται σταθερά με την πιθανότητα αποβίωσης ενός ατόμου από καρδιαγγειακά νοσήματα να αυξάνει πιο γρήγορα αλλά να είναι πάντα μικρότερη από την πιθανότητα αποβίωσης ενός ατόμου από διάφορες αιτίες.

5^ο ΚΕΦΑΛΑΙΟ

ΠΟΛΛΑΠΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

5.1 Εισαγωγή

Στο κεφάλαιο αυτό θα δούμε την μέθοδο της παλινδρόμησης η οποία χρησιμεύει στο να αναπτύξουμε σχέση μεταξύ της εξαρτημένης μεταβλητής (fu) και των ανεξάρτητων μεταβλητών $smoking$, $physical_act$, age_entr , sbr , dbp και BMI για τους πληθυσμούς που ορίζονται από τις σχέσεις $cvd_mortality=1$ και $allc_mortality=1$. Δηλαδή για τον πληθυσμό που έχει πεθάνει από καρδιαγγειακά νοσήματα και για τον πληθυσμό που έχει πεθάνει από διάφορες αιτίες.

Η επιλογή των ανεξάρτητων μεταβλητών έγινε με βάση το κεφάλαιο 3 όπου εξετάζεται η σχέση των μεταβλητών μεταξύ τους. Στο κεφάλαιο αυτό βλέπουμε ότι η fu εξαρτάται από τις μεταβλητές $smoking$, $physical_act$, age_entr , sbr , dbp και BMI .

5.2 Πολλαπλή Παλινδρόμηση

Σε αρκετά προβλήματα η μεταβλητή απόκρισης Y μπορεί να θεωρηθεί ότι επηρεάζεται από περισσότερες από μια ερμηνευτικές μεταβλητές, έστω X_1, X_2, \dots, X_{p-1} . Μπορούμε να χρησιμοποιήσουμε ένα γραμμικό μοντέλο για να διερευνήσουμε την εξάρτηση της Y από τις X_1, X_2, \dots, X_{p-1} . Το μοντέλο αυτό, το οποίο αποτελεί γενίκευση του μοντέλου $Y = b_0 + b_1X + \varepsilon$, θα έχει τη μορφή:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_{p-1}X_{p-1} + \varepsilon,$$

όπου τα «σφάλματα» $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ θεωρούνται ανεξάρτητες τ.μ. από την $N(0,1)$ ενώ οι ερμηνευτικές μεταβλητές X_1, X_2, \dots, X_{p-1} , όπως και στο απλό γραμμικό μοντέλο, δεν θεωρούνται τυχαίες. Το παραπάνω μοντέλο γράφεται με τη βοήθεια πινάκων στην απλούστερη μορφή:

$$Y = Xb + \varepsilon$$

Όπου

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{11} & \dots & X_{1,p-1} \\ 1 & X_{21} & & X_{2,p-1} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \dots & X_{n,p-1} \end{bmatrix}, \quad b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Ο πίνακας X καλείται και πίνακας σχεδιασμού. Στο πολλαπλό μοντέλο θεωρούμε ότι τα σημεία δεν βρίσκονται «κοντά» σε μία ευθεία αλλά «κοντά» σε ένα (υπερ)επίπεδο στις p διαστάσεις.

5.3 Μέθοδος ελαχίστων τετραγώνων

Για την εκτίμηση των συντελεστών εφαρμόζουμε τη μέθοδο των ελαχίστων τετραγώνων

$$\begin{aligned}SSE(\beta) &= \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left[Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 X_{i3} + \varepsilon_i \dots + \hat{\beta}_{p-1} X_{i,p-1}) \right]^2 \\ &= (Y - X\hat{\beta})^T (Y - X\hat{\beta})\end{aligned}$$

Παίρνοντας τις πρώτες μερικές παραγώγους προκύπτει ένα σύστημα με p εξισώσεις και p αγνώστους. Ισοδύναμα:

$$X\hat{\beta} = Y \Rightarrow X^T X\beta = X^T Y \Rightarrow \hat{\beta} = (X^T X)^{-1} X^T Y$$

5.4 Υποθέσεις του πολλαπλού γραμμικού μοντέλου

1. *Ανεξαρτησία*: Οι παρατηρήσεις Y_i πρέπει να είναι στατιστικά ανεξάρτητες μεταξύ τους (όταν λαμβάνονται πολλαπλές μετρήσεις από το ίδιο υποκείμενο τότε θα πρέπει να χρησιμοποιηθούν άλλες τεχνικές αναλύσεις).

2. *Ομοσκεδαστικότητα*: Η διασπορά των Y_i είναι ίδια για κάθε συνδυασμούς X_1, X_2, \dots, X_{p-1} με:

$$\sigma_{Y/X_1, X_2, \dots, X_{p-1}}^2 \equiv \sigma^2 \text{ ή αλλιώς } \sigma_{Y/X_1, X_2, \dots, X_{p-1}}^2 \equiv \sigma^2$$

3. *Κανονικότητα*: Για κάθε συνδυασμό X_1, X_2, \dots, X_{p-1} η μεταβλητή Y κατανέμεται κανονικά:

$$Y_i \sim N(\mu_{Y|X_1, \dots, X_{p-1}}, \sigma^2) \text{ ή αλλιώς } \varepsilon_i \sim N(0, \sigma^2)$$

5.5 Η Μέθοδος Stepwise Regression

Σύμφωνα με τη μέθοδο Stepwise Regression, το βέλτιστο γραμμικό μοντέλο αναπτύσσεται με την προσθήκη μιας ανεξάρτητης μεταβλητής κάθε φορά (θεωρείται ότι είναι σημαντική για την πρόβλεψη της εξαρτημένης) αλλά και τον παράλληλο έλεγχο κατά πόσο το διαμορφωμένο σε κάθε φάση μοντέλο χρειάζεται όλες τις εισαχθείσες μεταβλητές.

Η εισαγωγή της κάθε μεταβλητής γίνεται εφόσον ικανοποιείται κάποιο κριτήριο επιλογής της για εισαγωγή, ενώ αντίστοιχο κριτήριο εφαρμόζεται για να αποφασισθεί αν κάποια ήδη εισηγμένη μεταβλητή θα πρέπει να αφαιρεθεί από το μοντέλο (γιατί κάποιες άλλες μεταβλητές μπορούν να καλύψουν την «απουσία» της).

Το κριτήριο που χρησιμοποιούμε είναι :

$$F_i^* = \frac{MSR(X_i)}{MSE(X_i)} > F_{1,n-p}(a)$$

και αν ισχύει, εισάγουμε στο μοντέλο μας την μεταβλητή που έχει το μεγαλύτερο F_i^* .

- Η μέθοδος Stepwise για τον πληθυσμό ατόμων που πέθαναν από καρδιαγγειακά νοσήματα (cvd_mortality=1 (CVD))

Variables Entered/Removed ^{a,b}			
Model	Variables Entered	Variables Removed	Method
1	age_entr	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	dbp	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	smoking	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: Follow up (yrs)

b. Models are based only on cases for which cvd_mortality = CVD

Πίνακας 5.1: Πίνακας εισαγωγής μεταβλητών σύμφωνα με τη μέθοδο stepwise

Από τον παραπάνω πίνακα βλέπουμε πως, με την μέθοδο stepwise, οι μεταβλητές age_entr, dbp και smoking εισαχθήκανε στο μοντέλο.

Coefficients^{a,b}

Model		Unstandardized Coefficients		Standardized	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	48.206	6.299		7.654	.000
	age_entr	-.443	.124	-.270	-3.585	.000
2	(Constant)	58.676	7.004		8.378	.000
	age_entr	-.380	.122	-.232	-3.111	.002
	dbp	-.165	.053	-.231	-3.104	.002
3	(Constant)	63.252	7.130		8.872	.000
	age_entr	-.377	.120	-.230	-3.130	.002
	dbp	-.194	.054	-.272	-3.617	.000
	smoking	-3.520	1.404	-.186	-2.507	.013

a. Dependent Variable: Follow up (yrs)

b. Selecting only cases for which cvd_mortality = CVD

Πίνακας 5.2: Πίνακας συντελεστών του μοντέλου με τη μέθοδο stepwise

Σύμφωνα με τον πίνακα Coefficients το μοντέλο που προκύπτει έχει την μορφή:

$$Fu = 63.252 - 0.377 * age_entr - 0.194 * dbp - 3.520 * smoking$$

- Η μέθοδος Stepwise για τον πληθυσμό ατόμων που πέθαναν από διάφορες αιτίες (allc_mortality= 1 (Death))

Variables Entered/Removed^{a,b}

Model	Variables Entered	Variables Removed	Method
1	age_entr		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	physical_act		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	smoking		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
4	dbp		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
5	BMI		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: Follow up (yrs)

b. Models are based only on cases for which allc_mortality = Death

Πίνακας 5.3: Πίνακας εισαγωγής μεταβλητών σύμφωνα με τη μέθοδο stepwise

Από τον παραπάνω πίνακα βλέπουμε πως, με την μέθοδο stepwise, όλες οι μεταβλητές εκτός από την sbp εισαχθήκανε στο μοντέλο.

Coefficients^{a,b}

Model		Unstandardized Coefficients		Standardized	t	Sig.
		B	Std. Error	Coefficients Beta		
1	(Constant)	46.861	4.339		10.801	.000
	age_entr	-.439	.086	-.248	-5.135	.000
2	(Constant)	44.675	4.389		10.179	.000
	age_entr	-.431	.085	-.243	-5.077	.000
	physical_act	2.576	.989	.125	2.605	.010
3	(Constant)	46.267	4.432		10.439	.000
	age_entr	-.434	.085	-.245	-5.128	.000
	physical_act	2.737	.987	.133	2.772	.006
	smoking	-2.178	1.018	-.102	-2.141	.033
4	(Constant)	53.023	5.544		9.563	.000
	age_entr	-.424	.084	-.239	-5.020	.000
	physical_act	2.412	.997	.117	2.420	.016
	smoking	-2.475	1.024	-.116	-2.416	.016
	dbp	-.083	.041	-.098	-2.015	.045
5	(Constant)	46.787	6.130		7.633	.000
	age_entr	-.414	.084	-.233	-4.926	.000
	physical_act	2.717	1.000	.132	2.717	.007
	smoking	-2.046	1.035	-.096	-1.977	.049
	dbp	-.111	.043	-.131	-2.594	.010
	BMI	.324	.139	.119	2.329	.020

a. Dependent Variable: Follow up (yrs)

b. Selecting only cases for which allc_mortality = Death

Πίνακας 5.4: Πίνακας συντελεστών του μοντέλου με τη μέθοδο stepwise

Σύμφωνα με τον πίνακα Coefficients το μοντέλο που προκύπτει έχει την μορφή:

$$Fu=46.787-0.414*age_entr+2.717*physical_act-2.046*smoking-0.111*dbp+0.324*BMI$$

5.6 Η Μέθοδος Enter Regression

Με αυτή τη μέθοδο εισάγονται στο μοντέλο όλες οι μεταβλητές ανεξάρτητα από το αν είναι στατιστικά σημαντικές σε σχέση με την εξαρτημένη μεταβλητή.

- Η μέθοδος Enter για τον πληθυσμό ατόμων που πέθαναν διάφορες αιτίες (allc_mortality= 1 (Death))

- Επανάληψη 1^η

		Coefficients ^{a,b}				
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	46.782	6.141		7.619	.000
	age_entr	-.414	.085	-.234	-4.888	.000
	physical_act	2.717	1.001	.132	2.714	.007
	smoking	-2.045	1.037	-.096	-1.972	.049
	sbp	.001	.035	.002	.026	.979
	dbp	-.112	.065	-.133	-1.724	.086
	BMI	.324	.139	.119	2.324	.021

a. Dependent Variable: Follow up (yrs)

b. Selecting only cases for which allc_mortality = Death

- Πίνακας 5.5: Πίνακας συντελεστών του μοντέλου με την μέθοδο enter (Επανάληψη 1η)

Σύμφωνα με τον παραπάνω πίνακα Coefficients οι μεταβλητές sbp και dbp έχουν αντίστοιχα p-value 0.979 και 0.086. Οι τιμές αυτές είναι μεγαλύτερες από 0.05, επομένως τις αφαιρούμε και ξανατρέχουμε το μοντέλο μας.

ο Επανάληψη 2^η

Coefficients^{a,b}

Model		Unstandardized Coefficients		Standardized	t	Sig.
		B	Std. Error	Coefficients Beta		
1	(Constant)	40.412	5.656		7.146	.000
	age_entr	-.429	.084	-.242	-5.084	.000
	physical_act	3.022	1.000	.146	3.022	.003
	smoking	-1.814	1.039	-.085	-1.746	.082
	BMI	.224	.135	.082	1.661	.098

a. Dependent Variable: Follow up (yrs)

b. Selecting only cases for which allc_mortality = Death

Πίνακας 5.6: Πίνακας συντελεστών του μοντέλου με την μέθοδο enter (Επανάληψη 2η)

Σύμφωνα με τον παραπάνω πίνακα Coefficients οι μεταβλητές smoking και BMI έχουν αντίστοιχα p-value 0.082 και 0.098. Οι τιμές αυτές είναι μεγαλύτερες από 0.05, επομένως τις αφαιρούμε και ξανατρέχουμε το μοντέλο μας.

ο Επανάληψη 3^η

Coefficients^{a,b}

Model		Unstandardized Coefficients		Standardized	t	Sig.
		B	Std. Error	Coefficients Beta		
1	(Constant)	47.385	4.075		11.627	.000
	age_entr	-.484	.079	-.274	-6.161	.000
	physical_act	2.717	.910	.133	2.984	.003

a. Dependent Variable: Follow up (yrs)

b. Selecting only cases for which allc_mortality = Death

Πίνακας 5.7: Πίνακας συντελεστών του μοντέλου με την μέθοδο enter (Επανάληψη 3η)

Σύμφωνα με τον πίνακα Coefficients το μοντέλο που προκύπτει έχει την μορφή:

$$Fu=47.385 +2.717*physical_act-0.484*age_entr$$

- Η μέθοδος Enter για τον πληθυσμό ατόμων που πέθαναν από καρδιαγγειακά νοσήματα (cvd_mortality=1 (CVD))

- Επανάληψη 1^η

		Coefficients ^{a,b}				
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	61.521	8.346		7.371	.000
	age_entr	-.375	.121	-.228	-3.089	.002
	physical_act	.826	1.541	.042	.536	.593
	smoking	-3.507	1.430	-.185	-2.452	.015
	sbp	-.018	.054	-.045	-.330	.742
	dbp	-.161	.102	-.226	-1.577	.117
	BMI	.033	.188	.014	.175	.861

a. Dependent Variable: Follow up (yrs)

b. Selecting only cases for which cvd_mortality = CVD

Πίνακας 5.8: Πίνακας συντελεστών του μοντέλου με την μέθοδο enter (Επανάληψη 1η)

Σύμφωνα με τον παραπάνω πίνακα Coefficients οι μεταβλητές physical_act, sbp, dbp και BMI έχουν αντίστοιχα p-value 0.593, 0.742, 0.117 και 0.861. Οι τιμές αυτές είναι μεγαλύτερες από 0.05, επομένως τις αφαιρούμε και ξανατρέχουμε το μοντέλο μας.

- Επανάληψη 2^η

		Coefficients ^{a,b}				
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	50.114	6.360		7.880	.000
	age_entr	-.448	.123	-.273	-3.646	.000
	smoking	-2.436	1.422	-.128	-1.713	.089

a. Dependent Variable: Follow up (yrs)

b. Selecting only cases for which cvd_mortality = CVD

Πίνακας 5.9: Πίνακας συντελεστών του μοντέλου με την μέθοδο enter (Επανάληψη 2η)

Σύμφωνα με τον παραπάνω πίνακα Coefficients η μεταβλητή smoking έχει p-value = 0.089 > 0.05, επομένως την αφαιρούμε και ξανατρέχουμε το μοντέλο μας.

- ο Επανάληψη 3^η

Coefficients^{a,b}

Model		Unstandardized Coefficients		Standardized	t	Sig.
		B	Std. Error	Coefficients Beta		
1	(Constant)	51.727	5.869		8.814	.000
	age_entr	-.516	.115	-.310	-4.497	.000

a. Dependent Variable: Follow up (yrs)

b. Selecting only cases for which cvd_mortality = CVD

Πίνακας 5.10: Πίνακας συντελεστών του μοντέλου με την μέθοδο enter (Επανάληψη 3η)

Σύμφωνα με τον πίνακα Coefficients το μοντέλο που προκύπτει έχει την μορφή:

$$Fu=51.727-0.516*age_entr$$

Έλεγχος κανονικότητας των σφαλμάτων

H₀ : Τα σφάλματα ακολουθούν την κανονική κατανομή, έναντι της εναλλακτικής

H₁: Τα σφάλματα δεν ακολουθούν την κανονική κατανομή

One-Sample Kolmogorov-Smirnov Test

		allc_step_res	cvd_step_res	allc_entr_res	cvd_entr_res
N		468	468	529	529
Normal Parameters ^{a,b}	Mean	.1813949	.0737551	.1760183	.0816898
	Std. Deviation	1.04041400	1.16606512	1.04288829	1.10257781
Most Extreme Differences	Absolute	.080	.084	.069	.083
	Positive	.052	.050	.045	.049
	Negative	-.080	-.084	-.069	-.083
Test Statistic		.080	.084	.069	.083
Asymp. Sig. (2-tailed)		.000 ^c	.000 ^c	.000 ^c	.000 ^c

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

Πίνακας 5.11: Πίνακας ελέγχου κανονικότητας των σφαλμάτων

Από τον παραπάνω πίνακα βλέπουμε πως τα residuals και για τις τέσσερις περιπτώσεις γραμμικής παλινδρόμησης δεν ακολουθούν την κανονική κατανομή σε επίπεδο σημαντικότητας 5%. Επομένως δεν μπορούμε να συνεχίσουμε σε στατιστική συμπερασματολογία.

Εάν η υπόθεση της κανονικότητας των σφαλμάτων του μοντέλου μας δεν γίνεται δεκτή τότε μπορούμε να καταφύγουμε σε ένα μετασχηματισμό των τιμών της εξαρτημένης μεταβλητής έτσι ώστε να επιτευχθεί η κανονικότητα. Ένας τέτοιος μετασχηματισμός είναι ο μετασχηματισμός Box – Cox. Σε αυτή τη περίπτωση όμως η ερμηνεία των συσχετίσεων θα ήταν δύσκολη. Επομένως, καταλήγουμε στο συμπέρασμα πως η γραμμική παλινδρόμηση δεν είναι κατάλληλη μέθοδος για σαφή στατιστικά συμπεράσματα στο δείγμα μας.

6^ο ΚΕΦΑΛΑΙΟ

ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

6.1 Εισαγωγή

Στο κεφάλαιο αυτό θα δούμε την μέθοδο της λογιστικής παλινδρόμησης η οποία χρησιμεύει στο να αναπτύξουμε σχέση μιας δίτιμης ανεξάρτητης τυχαίας μεταβλητής και συνεχών η διακριτών ανεξάρτητων μεταβλητών. Ουσιαστικά η μέθοδος αυτή γενικεύει τα γραμμικά μοντέλα, έτσι ώστε η εξαρτημένη μμεταβλητή να ακολουθεί την εκθετική οικογένεια κατανομών.

Συγκεκριμένα, θα αναπτύξουμε σχέση μεταξύ της εξαρτημένης μεταβλητής `allc_mortality` και των ανεξάρτητων μεταβλητών `smoking` και `age_entr` καθώς και για την εξαρτημένη μεταβλητή `cvd_mortality` και των ανεξάρτητων μεταβλητών `age_entr`, `sbr` και `dbp`. Η επιλογή των ανεξάρτητων μεταβλητών έγινε με βάση το κεφάλαιο 3 όπου εξετάζεται η σχέση των μεταβλητών μεταξύ τους. Στο κεφάλαιο αυτό βλέπουμε ότι η `allc_mortality` εξαρτάται από τις μεταβλητές `smoking` και `age_entr` ενώ η `cvd_mortality` εξαρτάται από τις `age_entr`, `sbr` και `dbp`.

6.2 Λογιστική Παλινδρόμηση

Το λογιστικό μοντέλο είναι ένα μη γραμμικό μοντέλο στο οποίο όμως τα σφάλματα δεν ακολουθούν κανονική κατανομή και η μεταβλητή απόκρισης είναι διακριτή. Η λογιστική παλινδρόμηση χρησιμοποιείται σε περιπτώσεις στις οποίες επιθυμούμε να προβλέψουμε την απουσία ή την παρουσία ενός χαρακτηριστικού, ή ενός συμβάντος. Είναι μια γενίκευση της απλής γραμμικής παλινδρόμησης για την περίπτωση όπου η εξαρτημένη μεταβλητή (Y) είναι δίτιμη (δηλαδή παίρνει την τιμή 0 όταν απουσιάζει το χαρακτηριστικό ή την τιμή 1 όταν υπάρχει το χαρακτηριστικό).

6.3 Πολλαπλό λογιστικό μοντέλο

Το πολλαπλό λογιστικό μοντέλο είναι:

$$Y_i = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p-1} X_{p-1}}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p-1} X_{p-1}}} + \varepsilon_i = E(Y_i) + \varepsilon_i$$

για να διευκολυνθούμε με τις πράξεις αλλά και για την απλοποίηση των σχέσεων θα χρησιμοποιήσουμε πίνακες και διανύσματα:

$$\beta_{p \times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \quad X_{p \times 1} = \begin{bmatrix} 1 \\ X_1 \\ \vdots \\ X_{p-1} \end{bmatrix} \quad X_i_{p \times 1} = \begin{bmatrix} 1 \\ X_{i1} \\ X_{i2} \\ \vdots \\ X_{i,p-1} \end{bmatrix}$$

Με τη βοήθεια των πιο πάνω και με τα Y_i να είναι ανεξάρτητες μεταβλητές μπορούμε να γράψουμε την αναμενόμενη λογιστική ως εξής:

$$E(Y_i) = \pi_i = \frac{e^{\beta' X_i}}{1 + e^{\beta' X_i}}$$

6.4 Εκτίμηση παραμέτρων

Η εκτίμηση παραμέτρων θα γίνει με τη βοήθεια της μεθόδου μέγιστης πιθανοφάνειας. Η συνάρτηση πιθανοφάνειας είναι:

$$\ln L(\beta) = \sum_{i=1}^n Y_i (\beta' X_i) - \sum_{i=1}^n \ln(1 + e^{\beta' X_i})$$

Για να βρούμε τους εκτιμητές θα πρέπει να χρησιμοποιήσουμε κάποια αριθμητική μέθοδο η οποία θα μας δίνει τις τιμές των $\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}$ που θα μεγιστοποιούν τη παραπάνω σχέση.

6.5 Λογιστική Παλινδρόμηση

6.5.1 Η θνησιμότητα από καρδιαγγειακά νοσήματα ως εξαρτημένη μεταβλητή της λογιστικής παλινδρόμησης

- Πίνακας Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	529	100.0
	Missing Cases	0	.0
	Total	529	100.0
Unselected Cases		0	.0
Total		529	100.0

a. If weight is in effect, see classification table for the total number of cases.

Πίνακας 6.1: Πίνακας περιγραφής του δείγματος

Από τον παραπάνω πίνακα παρατηρούμε ότι χρησιμοποιήθηκαν και οι 529 τιμές, αφού δεν έχουμε missing values.

- Έλεγχος καλής προσαρμογής των Hosmer και Lemeshow

H₀: Το μοντέλο έχει καλή προσαρμογή. Έναντι της εναλλακτικής

H₁: Το μοντέλο δεν έχει καλή προσαρμογή

Step	Chi-square	df	Sig.
1	4.821	8	.777

Πίνακας 6.2: Πίνακας ελέγχου καλής προσαρμογής Hosmer και Lemeshow

Παρατηρούμε πως $p\text{-value}=0.777 > 0.05$. Επομένως σε επίπεδο σημαντικότητας 5% αποδεχόμαστε την μηδενική μας υπόθεση και συμπεραίνουμε ότι το μοντέλο μας έχει καλή προσαρμογή.

- **Ακρίβεια πρόβλεψης του μοντέλου**

Classification Table^a

	Observed	Predicted		Percentage Correct	
		cvd_mortality None	CVD		
Step 1	cvd_mortality	None	317	20	94.1
		CVD	173	19	9.9
	Overall Percentage				63.5

a. The cut value is .500

Πίνακας 6.3: Πίνακας ακρίβειας πρόβλεψης του μοντέλου

Από τον παραπάνω πίνακα παρατηρούμε ότι το μοντέλο προβλέπει πως 490 (317+173) άτομα δεν έχουν πεθάνει από καρδιαγγειακά νοσήματα και 39 (20+19) άτομα έχουν πεθάνει από καρδιαγγειακά.

Επίσης, 317 άτομα όντως δεν έχουν πεθάνει από καρδιαγγειακά και 19 άτομα έχουν πεθάνει από καρδιαγγειακά νοσήματα.

Τέλος, σύμφωνα με τα παραπάνω, το ποσοστό σωστής ταξινόμησης είναι 63.5.

- **Πίνακας των συντελεστών παλινδρόμησης του μοντέλου**

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)		
							Lower	Upper	
Step 1 ^a									
	age_entr	.043	.017	6.792	1	.009	1.044	1.011	1.079
	sbp	.014	.007	4.129	1	.042	1.014	1.000	1.027
	dbp	-.002	.012	.022	1	.883	.998	.975	1.022
	Constant	-4.438	1.043	18.085	1	.000	.012		

a. Variable(s) entered on step 1: age_entr, sbp, dbp.

Πίνακας 6.4: Πίνακας των συντελεστών παλινδρόμησης του μοντέλου

Από τον παραπάνω πίνακα βλέπουμε ότι:

- Για αύξηση μίας μονάδας της ηλικίας που κάποιος μπήκε στην έρευνα (age_entr), ο σχετικός λόγος πιθανοτήτων θετικού αποτελέσματος (cvd_mortality=1) πολλαπλασιάζεται με τον εκθετικό συντελεστή $e^{0.043}=1.044 > 1$. Αυτό σημαίνει πως η πιθανότητα επιβίωσης από καρδιαγγειακά νοσήματα αυξάνεται και αντίστοιχα μειώνεται η πιθανότητα αποβίωσης.

- Για αύξηση μίας μονάδας της συστολικής αρτηριακής πίεσης (sbp), ο σχετικός λόγος πιθανοτήτων θετικού αποτελέσματος (cvd_mortality=1) πολλαπλασιάζεται με τον εκθετικό συντελεστή $e^{0.014}=1.014 > 1$. Αυτό σημαίνει πως η πιθανότητα επιβίωσης από καρδιαγγειακά νοσήματα αυξάνεται και αντίστοιχα μειώνεται η πιθανότητα αποβίωσης.
- Για αύξηση μίας μονάδας της διαστολικής αρτηριακής πίεσης (dbp), ο σχετικός λόγος πιθανοτήτων θετικού αποτελέσματος (cvd_mortality=1) πολλαπλασιάζεται με τον εκθετικό συντελεστή $e^{-0.002}=0.998 < 1$. Αυτό σημαίνει πως η πιθανότητα επιβίωσης από καρδιαγγειακά νοσήματα μειώνεται και αντίστοιχα αυξάνεται η πιθανότητα αποβίωσης.

Το μοντέλο που προσαρμόστηκε είναι το:

$$\ln \left(\frac{p}{1-p} \right) = -4.438 - 0.002 * dbp + 0.014 * sbp + 0.043 * age_entr$$

Ωστόσο, από τον παραπάνω πίνακα βλέπουμε πως μόνο ο σταθερός όρος καθώς και η συστολική αρτηριακή πίεση προκύπτουν στατιστικά σημαντικοί για την πιθανότητα αποβίωσης (p-value=0 < 0.05) σύμφωνα με τον έλεγχο Wald.

Σύμφωνα με τα παραπάνω, το μοντέλο που τελικά προκύπτει είναι το :

$$\ln \left(\frac{p}{1-p} \right) = -4.438 + 0.014 * sbp + 0.043 * age_entr$$

6.5.2 Η θνησιμότητα από διάφορες αιτίες ως εξαρτημένη μεταβλητή της λογιστικής παλινδρόμησης

- Πίνακας Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	468	88.5
	Missing Cases	61	11.5
	Total	529	100.0
Unselected Cases		0	.0
Total		529	100.0

a. If weight is in effect, see classification table for the total number of cases.

Πίνακας 6.5: Πίνακας περιγραφής του δείγματος

Από τον παραπάνω πίνακα παρατηρούμε ότι χρησιμοποιήθηκαν και οι 468 τιμές από τις 529, αφού έχουμε 61 missing values.

- Έλεγχος καλής προσαρμογής των Hosmer και Lemeshow

H₀: Το μοντέλο έχει καλή προσαρμογή. Έναντι της εναλλακτικής

H₁: Το μοντέλο δεν έχει καλή προσαρμογή

Step	Chi-square	df	Sig.
1	5.274	7	.627

Πίνακας 6.6: Πίνακας ελέγχου καλής προσαρμογής Hosmer και Lemeshow

Παρατηρούμε πως $p\text{-value}=0.627 > 0.05$. Επομένως σε επίπεδο σημαντικότητας 5% αποδεχόμαστε την μηδενική μας υπόθεση και συμπεραίνουμε ότι το μοντέλο μας έχει καλή προσαρμογή.

- **Ακρίβεια πρόβλεψης του μοντέλου**

Classification Table^a

	Observed	Predicted		Percentage Correct	
		allc_mortality Alive	Death		
Step 1	allc_mortality	Alive	5	57	8.1
		Death	3	403	99.3
Overall Percentage					87.2

a. The cut value is .500

Πίνακας 6.7: Πίνακας ακρίβειας πρόβλεψης του μοντέλου

Από τον παραπάνω πίνακα παρατηρούμε ότι το μοντέλο προβλέπει πως 8 (5+3) άτομα δεν έχουν πεθάνει από διάφορες αιτίες και 460 (57+403) άτομα έχουν πεθάνει.

Επίσης, 5 άτομα όντως δεν πέθαναν από διάφορες αιτίες και 403 άτομα έχουν πεθάνει.

Τέλος, σύμφωνα με τα παραπάνω, το ποσοστό σωστής ταξινόμησης είναι 87.2.

- **Πίνακας των συντελεστών παλινδρόμησης του μοντέλου**

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Step 1 ^a								
smoking(1)	-.644	.309	4.356	1	.037	.525	.287	.961
age_entr	.217	.031	47.643	1	.000	1.242	1.168	1.321
Constant	-8.149	1.425	32.711	1	.000	.000		

a. Variable(s) entered on step 1: smoking, age_entr.

Πίνακας 6.8: Πίνακας των συντελεστών παλινδρόμησης του μοντέλου

Από τον παραπάνω πίνακα βλέπουμε ότι:

- Όταν κάποιος είναι καπνιστής (smoking=1) ο λόγος σχετικών πιθανοτήτων θετικού αποτελέσματος (allc_mortality=1) πολλαπλασιάζεται με τον εκθετικό συντελεστή $e^{-0.644}=0.525 < 1$. Αυτό σημαίνει πως η πιθανότητα επιβίωσης από διάφορες αιτίες μειώνεται και αντίστοιχα αυξάνεται η πιθανότητα αποβίωσης.
- Για αύξηση μίας μονάδας της ηλικίας που κάποιος μπήκε στην έρευνα (age_entr), ο σχετικός λόγος πιθανοτήτων θετικού αποτελέσματος (allc_mortality=1) πολλαπλασιάζεται με τον εκθετικό συντελεστή $e^{0.217}=1.242 > 1$. Αυτό σημαίνει πως η πιθανότητα επιβίωσης από διάφορες αιτίες αυξάνεται και αντίστοιχα μειώνεται η πιθανότητα αποβίωσης.

Το μοντέλο που προσαρμόστηκε είναι το:

$$\ln \left(\frac{p}{1-p} \right) = -8.149 - 0.644 * \text{smoking} + 0.217 * \text{age_entr}$$

Από τον παραπάνω πίνακα βλέπουμε πώς ο σταθερός όρος, το κάπνισμα καθώς και η ηλικία εισόδου στην έρευνα προκύπτουν στατιστικά σημαντικοί για την πιθανότητα αποβίωσης ($p\text{-value} < 0.05$) σύμφωνα με τον έλεγχο Wald.

7^ο ΚΕΦΑΛΑΙΟ

ΣΥΜΠΕΡΑΣΜΑΤΑ

7.1 Εισαγωγή

Σε αυτό το κεφάλαιο θα παρουσιάσουμε συνοπτικά τα αποτελέσματα που έχουν προκύψει σε όλα τα προηγούμενα κεφάλαια μετά από την ανάλυση. Θα επικεντρωθούμε στους παράγοντες κινδύνου που σχετίζονται με τη θνησιμότητα 40 ετών από κάθε αιτία και στεφανιαία νόσο.

7.2 Αποτελέσματα από περιγραφική ανάλυση

Αρχικά, θα παρουσιάσουμε κάποια από τα σημαντικότερα αποτελέσματα που προέκυψαν μετά την περιγραφική ανάλυση του δείγματος μας. Σύμφωνα με όσα προηγήθηκαν παρατηρούμε ότι η μέση ηλικία εισόδου στην έρευνα ήταν 49.81 έτη και ο μέσος χρόνος παραμονής ήταν 26.71 χρόνια. Το μέσο ύψος των ανθρώπων που συμμετείχαν στην έρευνα ήταν 165.7809 cm και το μέσο βάρος ήταν 64.3544 kg. Αντίστοιχα, ο μέσος δείκτης μάζας σώματος ήταν 23.3458. Σχετικά με την αρτηριακή πίεση βλέπουμε πως η μέση συστολική αρτηριακή πίεση ήταν 135.08 χιλιοστά της στήλης υδραργύρου (mmHg) ενώ η μέση διαστολική αρτηριακή πίεση ήταν 82.66 χιλιοστά της στήλης υδραργύρου (mmHg). Παρατηρούμε ακόμα ότι ο μέσος καρδιακός ρυθμός ήταν 68.3516 χτύποι το λεπτό καθώς και η μέση ολική χοληστερόλη ήταν 203.95 mg/dL.

Επίσης, μπορούμε να δούμε ότι το 25.7%, δηλαδή 136 από τους 529, δεν κάπνιζε, ενώ το 62.8%, δηλαδή 332 από τους 529, του δείγματος κάπνιζε. Ακόμα, το 11.5%, δηλαδή 61 παρατηρήσεις από τις 529, ήταν missing values, δηλαδή δεν ξέρουμε αν κάπνιζαν ή όχι. Το 31.6%, δηλαδή 167 από τους 529, δεν είχε φυσική δραστηριότητα, ενώ το 68.4%, δηλαδή 362 από τους 529, του δείγματος είχε φυσική δραστηριότητα. Παρατηρούμε επίσης, πως το 12.9%, δηλαδή 68 από τους 529 άνδρες του δείγματος, ήταν ζωντανοί στο τέλος της έρευνας, ενώ το 87.1%, δηλαδή 461 από τους 529 άνδρες του δείγματος, πέθαναν κατά την διάρκεια της έρευνας από διάφορες αιτίες θανάτου και πως το 63.7%, δηλαδή 337 από τους 529 άνδρες του δείγματος, δεν πέθαναν από καρδιαγγειακά νοσήματα, ενώ το 36.3%, δηλαδή 192 από τους 529 άνδρες του δείγματος, πέθαναν από CVD.

Επιπλέον, για τους δύο πληθυσμούς που πέθαναν ή επέζησαν από διάφορες αιτίες, παρατηρήσαμε πως τα άτομα που απεβίωσαν είχαν μεγαλύτερη μέση ηλικία εισαγωγής στην έρευνα από αυτά που παρέμειναν ζωντανά ως το τέλος της. Συγκεκριμένα, η μέση ηλικία εισαγωγής των ατόμων που απεβίωσαν ήταν 50.55 χρόνια ενώ αυτών που επέζησαν ήταν 47.79. Επιπροσθέτως, ο μέσος χρόνος παραμονής στην έρευνα των ατόμων που πέθαναν από

διάφορες αιτίες ήταν 24.75 χρόνια ενώ αυτών που επέζησαν ήταν 40 χρόνια, όπως είναι αναμενόμενο, αφού η έρευνα διήρκεσε 40 χρόνια. Τέλος, από τα άτομα που απεβίωσαν από διάφορες αιτίες το 64% ήταν καπνιστές και το 24.1% ήταν μη καπνιστές, ενώ από τα άτομα που επέζησαν ως το τέλος της έρευνας το 54.4% ήταν καπνιστές και το 36.8% ήταν μη καπνιστές. Μπορούμε να δούμε λοιπόν πως τα άτομα που απεβίωσαν από διάφορες αιτίες είχαν μεγαλύτερο ποσοστό καπνιστών από τα άτομα που επέζησαν.

Για τους δύο πληθυσμούς που πέθαναν ή δεν πέθαναν από καρδιαγγειακά νοσήματα, παρατηρήσαμε πως τα άτομα που απεβίωσαν είχαν μεγαλύτερη μέση ηλικία εισαγωγής στην έρευνα από αυτά που δεν πέθαναν. Συγκεκριμένα, η μέση ηλικία εισαγωγής των ατόμων που απεβίωσαν ήταν 50.82 χρόνια ενώ αυτών που δεν πέθαναν ήταν 49.23. Επίσης, η μέση συστολική αρτηριακή πίεση των ατόμων που πέθαναν ήταν μεγαλύτερη από αυτή των ατόμων που δεν πέθαναν. Συγκεκριμένα, η μέση συστολική αρτηριακή πίεση των ατόμων που απεβίωσαν ήταν 139.24 χιλιοστά της στήλης υδραργύρου (mmHg) ενώ αυτών που δεν πέθαναν ήταν 132.71 χιλιοστά της στήλης υδραργύρου (mmHg). Ακόμα, η μέση διαστολική αρτηριακή πίεση των ατόμων που πέθαναν ήταν μεγαλύτερη από αυτή των ατόμων που δεν πέθαναν. Συγκεκριμένα, η μέση διαστολική αρτηριακή πίεση των ατόμων που απεβίωσαν από καρδιαγγειακά ήταν 84.23 χιλιοστά της στήλης υδραργύρου (mmHg) ενώ αυτών που δεν πέθαναν ήταν 81.76 χιλιοστά της στήλης υδραργύρου (mmHg). Τέλος, ο μέσος χρόνος παραμονής στην έρευνα των ατόμων που πέθαναν από καρδιαγγειακά ήταν μικρότερος από τον μέσο χρόνο παραμονής στην έρευνα των ατόμων που δεν πέθαναν από αυτή τη νόσο. Συγκεκριμένα, ο μέσος χρόνος παραμονής στην έρευνα των ατόμων που απεβίωσαν από καρδιαγγειακά ήταν 25.48 χρόνια ενώ αυτών που δεν πέθαναν ήταν 27.41 χρόνια.

7.3 Συσχετίσεις

Σύμφωνα με τη μελέτη που πραγματοποιήσαμε συμπεραίνουμε ότι οι παράγοντες που φαίνεται να επηρεάζουν τον χρόνο παραμονής ενός ατόμου στην έρευνα είναι το κάπνισμα, η φυσική άσκηση, η ηλικία εισαγωγής στην έρευνα, η συστολική και διαστολική αρτηριακή πίεση καθώς και ο δείκτης μάζας σώματος.

Επιπροσθέτως, οι παράγοντες που φαίνεται να επηρεάζουν τον θάνατο ενός ατόμου από στεφανιαία νόσο είναι η ηλικία εισαγωγής στην έρευνα καθώς και η συστολική και διαστολική αρτηριακή πίεση.

Τέλος, οι παράγοντες που φαίνεται να επηρεάζουν τον θάνατο ενός ατόμου από όλες τις αιτίες είναι η ηλικία εισαγωγής στην έρευνα καθώς και το κάπνισμα.

7.4 Αποτελέσματα από τις αναλύσεις

Στη συνέχεια, στα δεδομένα μας προσαρμόσαμε 3 διαφορετικά μοντέλα για να εξετάσουμε ποιοι παράγοντες επηρεάζουν την επιβίωση των ατόμων που συμμετείχαν στην έρευνα.

- **Συμπεράσματα από την προσαρμογή του μοντέλου του μοντέλου του Cox**

- Για την ανάλυση επιβίωσης των ατόμων από διάφορες αιτίες καταλήξαμε στο συμπέρασμα ότι οι μεταβλητές `age_entr`, `tchol` και `smoking` είναι στατιστικά σημαντικές. Δηλαδή, βασικό ρόλο στην επιβίωση των ασθενών από διάφορες αιτίες παίζουν η ηλικία εισαγωγής του ατόμου στην έρευνα (`age_entr`), η ολική χοληστερόλη (`tchol`) καθώς και το εάν το άτομο αυτό καπνίζει ή όχι (`smoking`).
- Για την ανάλυση επιβίωσης των ατόμων από καρδιαγγειακά νοσήματα καταλήξαμε στο συμπέρασμα ότι οι μεταβλητές `age_entr` καθώς και η `sbr` είναι στατιστικά σημαντικές. Δηλαδή, βασικό ρόλο στην επιβίωση των ασθενών από καρδιαγγειακά νοσήματα παίζουν η ηλικία εισαγωγής του ατόμου στην έρευνα (`age_entr`) καθώς και η συστολική αρτηριακή πίεση (`sbr`).
- Σύμφωνα με τις γραφικές παραστάσεις των συναρτήσεων επιβίωσης για τους πληθυσμούς που πέθαναν από διάφορες αιτίες και από καρδιαγγειακά νοσήματα φαίνεται ότι όσο αυξάνει ο χρόνος παραμονής ενός ατόμου στην έρευνα, η πιθανότητα επιβίωσης από καρδιαγγειακά νοσήματα και από διάφορες αιτίες μειώνεται. Επιπλέον, η πιθανότητα επιβίωσης ενός ατόμου από καρδιαγγειακά νοσήματα είναι μεγαλύτερη από την πιθανότητα επιβίωσης ενός ατόμου από διάφορες αιτίες. Τέλος, παρατηρούμε πως όσο αυξάνεται ο χρόνος παραμονής ενός ατόμου στην έρευνα τόσο πιο πιθανή είναι η επιβίωση από τα καρδιαγγειακά νοσήματα σε σχέση με την επιβίωση από διάφορες αιτίες.
- Σύμφωνα με τις γραφικές παραστάσεις των συναρτήσεων κινδύνου για τους πληθυσμούς που πέθαναν από διάφορες αιτίες και από καρδιαγγειακά νοσήματα φαίνεται ότι όσο αυξάνει ο χρόνος παραμονής ενός ατόμου στην έρευνα, η πιθανότητα αποβίωσης από καρδιαγγειακά νοσήματα και από διάφορες αιτίες αυξάνεται. Επιπλέον, η πιθανότητα αποβίωσης ενός ατόμου από καρδιαγγειακά νοσήματα είναι μικρότερη από την πιθανότητα αποβίωσης ενός ατόμου από διάφορες αιτίες. Τέλος, παρατηρούμε πως όσο αυξάνεται ο χρόνος παραμονής ενός ατόμου στην έρευνα τόσο πιο πιθανή είναι η αποβίωση από διάφορες αιτίες σε σχέση με την αποβίωση από τα καρδιαγγειακά νοσήματα.

- **Συμπεράσματα από την προσαρμογή του μοντέλου γραμμικής παλινδρόμησης**
 - Για την πολλαπλή γραμμική παλινδρόμηση, κάνοντας τους ελέγχους προϋποθέσεων για την προσαρμογή του μοντέλου, παρατηρήσαμε ότι τα σφάλματα, για όλες τις περιπτώσεις, δεν ακολουθούν κανονική κατανομή και έτσι δεν μπορέσαμε να συνεχίσουμε σε στατιστική συμπερασματολογία. Επομένως, καταλήγουμε στο συμπέρασμα πως η γραμμική παλινδρόμηση δεν είναι κατάλληλη μέθοδος για σαφή στατιστικά συμπεράσματα στο δείγμα μας.

- **Συμπεράσματα από την προσαρμογή του μοντέλου λογιστικής παλινδρόμησης**
 - Για την ανάλυση επιβίωσης των ατόμων από καρδιαγγειακά νοσήματα οι μεταβλητές `dbp`, `sbr` και `age_entr` μπήκαν στο μοντέλο μας, όμως σύμφωνα με τον έλεγχο Wald είδαμε πως ο σταθερός όρος καθώς, η ηλικία εισόδου στην έρευνα (`age_entr`) και η συστολική αρτηριακή πίεση (`sbr`) αποδείχθηκαν στατιστικά σημαντικές για την πιθανότητα επιβίωσης. Δηλαδή, βασικό ρόλο στην επιβίωση των ασθενών από καρδιαγγειακά νοσήματα παίζει η ηλικία εισαγωγής του ατόμου στην έρευνα και η συστολική αρτηριακή πίεση.

 - Για την ανάλυση επιβίωσης των ατόμων από διάφορες αιτίες οι μεταβλητές `smoking` καθώς και η `age_entr` μπήκαν στο μοντέλο μας. Σύμφωνα με τον έλεγχο Wald είδαμε πως ο σταθερός όρος, η ηλικία εισόδου στην έρευνα (`age_entr`) και το κάπνισμα (`smoking`) προκύπτουν στατιστικά σημαντικοί για την πιθανότητα επιβίωσης. Δηλαδή, βασικό ρόλο στην επιβίωση των ασθενών από διάφορες αιτίες παίζει η ηλικία εισαγωγής του ατόμου στην έρευνα και το κάπνισμα.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ξενόγλωσση Βιβλιογραφία

1. Panagiotakos D., Chrysohoou C., Pitsavos C., Menotti A., Dontas A., Skoumas J., Stefan C., 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 - Journal – Elsevier, International journal of Cardiology, ELSEVIER
2. European Statistics Explained (https://ec.europa.eu/eurostat/statisticsexplained/index.php?title=Main_Page)
3. Ritchie H., 06/03/2019, “What do the people of the world die from?“, BBC

Ελληνική Βιβλιογραφία

4. Εμβαλωτής Α., Κατσής Α., Σιδερίδης Γ., Ιωάννινα 2006, “Στατιστική μεθοδολογία εκπαιδευτικής έρευνας”
5. Καλογεροπούλου Α., 2013, “Διερεύνηση του ρόλου της κατανομής Ω6/Ω3 λιπαρών οξέων, στην πιθανότητα εμφάνισης καρδιαγγειακού επεισοδίου, υπο την παρουσία ή όχι κατάθλιψης : επιδημιολογική μελέτη”, Αττική, Χαροκόπειο Πανεπιστήμιο
6. Μπούτσικας Μ., 2004, “Πολλαπλή Γραμμική Παλινδρόμηση (Multiple Linear Regression)”, Τμήμα Στατιστικής & Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς
7. Ξενή Μ., 2012, “Λογιστική Παλινδρόμηση & Διαχωριστική Ανάλυση”, τμήμα Μαθηματικών πανεπιστήμιο Πατρών, διατμηματικό μεταπτυχιακό πρόγραμμα σπουδών «Μαθηματικά των υπολογιστών και των αποφάσεων»
8. Πανδή-Αγαθοκλή Β. (Δρ. χημικός ερευνήτρια), 07/2019, “Η δημογραφική εξέλιξη στην Κέρκυρα και οι συνέπειες της”
9. Παπαδόπουλος Γ., “Στατιστικός Έλεγχος Υποθέσεων”, Εργαστήριο Μαθηματικών και Στατιστικής (<https://www.aua.gr/gpapadopoulos/files/hypoth-tests12e.pdf>)

- 10.** Πούλη Μ., Μεσογειακή Διατροφή: “Το μυστικό της μακροβιότητας”, Διατροφή και Ευεξία
- 11.** Τροχοπούλου Β., Πειραιάς 2009, “Μέθοδοι επιλογής βέλτιστου συνόλου ανεξάρτητων μεταβλητών σε μοντέλα γραμμικής παλινδρόμησης”, τμήμα Στατιστικής και Ασφαλιστικής επιστήμης, μεταπτυχιακό πρόγραμμα σπουδών «Εφαρμοσμένη Στατιστική»
- 12.** Τσιούφης Κ., 2018, “η καρδιαγγειακή νόσος στην Ελλάδα με αριθμούς”
- 13.** Φωκιανός Κ. , Χαραλάμπους Χ. (<http://www.mas.ucy.ac.cy/~fokianos/GreekRbook/survival.pdf>)
- 14.** Ανώνυμος, 17/06/2013, “Η μελέτη των επτά χωρών”, HealthyLiving