

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΕΚΤΙΜΗΣΗ ΤΟΥ ΕΠΙΠΕΔΟΥ ΠΑΧΥΣΑΡΚΙΑΣ ΜΕ ΒΑΣΗ
ΔΗΜΟΓΡΑΦΙΚΑ ΣΤΟΙΧΕΙΑ ΚΑΙ ΔΙΑΤΡΟΦΙΚΕΣ
ΣΥΝΗΘΕΙΕΣ**

ΤΑΡΑΪ ΑΝΤΖΕΛΑ

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

2025

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

Τα μέλη της Επιτροπής ήταν:

- Αναπληρωτής Καθηγητής Τζαβελάς Γεώργιος (Επιβλέπων)
- Καθηγήτρια Βερροπούλου Γεωργία
- Καθηγητής Πολίτης Κωνσταντίνος

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

UNIVERSITY OF PIRAEUS



**DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN
APPLIED STATISTICS**

**OBESITY LEVEL ESTIMATION BASED ON
DEMOGRAPHIC DATA AND DIETARY HABITS**

By
TARAI ANTZELA

MSc Dissertation
submitted to the Department of Statistics and
Insurance Science of the University of Piraeus in
partial fulfilment of the requirements for the degree
of Master of Science in Applied Statistics.

Piraeus, Greece

2025

Στην οικογένειά μου

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου κ. Γεώργιο Τζαβελά για την καθοδήγησή του, το χρόνο και τη διάθεσή του, καθώς και για την εμπιστοσύνη που μου έδειξε.

Θα ήθελα επίσης να ευχαριστήσω από καρδιάς την οικογένειά μου, για την υπομονή τους και τη στήριξή τους όλο αυτό το διάστημα.

ΠΕΡΙΛΗΨΗ

Η παχυσαρκία αποτελεί ένα από τα σημαντικότερα προβλήματα δημόσιας υγείας παγκοσμίως, καθώς συνδέεται με αυξημένο κίνδυνο εμφάνισης καρδιομεταβολικών νοσημάτων, σακχαρώδη διαβήτη τύπου 2 και άλλων σοβαρών επιπλοκών. Εκτός από τις επιπτώσεις στη σωματική υγεία, επηρεάζει αρνητικά και την ψυχοκοινωνική ευεξία, μειώνοντας την ποιότητα ζωής.

Στόχος της παρούσας μελέτης ήταν να διερευνηθούν οι παράγοντες που σχετίζονται με την εμφάνιση και τα επίπεδα παχυσαρκίας, ώστε να κατανοήσουμε καλύτερα τις επιδράσεις των δημογραφικών χαρακτηριστικών, των διατροφικών συνηθειών και της φυσικής δραστηριότητας.

Η έρευνα βασίστηκε σε δεδομένα από 485 άτομα προερχόμενα από το Μεξικό, το Περού και την Κολομβία, τα οποία περιλάμβαναν πληροφορίες για κοινωνικοδημογραφικά στοιχεία, οικογενειακό ιστορικό παχυσαρκίας, συνήθειες διατροφής και καθημερινής δραστηριότητας.

Για την ανάλυση εφαρμόστηκαν περιγραφικά στατιστικά, έλεγχοι συσχέτισης (χ^2 του Pearson, μη παραμετρικός έλεγχος Kruskal–Wallis, έλεγχος κανονικότητας Kolmogorov–Smirnov) καθώς και παλινδρομικά μοντέλα. Συγκεκριμένα, χρησιμοποιήθηκε διωνυμική λογιστική παλινδρόμηση για τη διερεύνηση παραγόντων που σχετίζονται με την πιθανότητα εμφάνισης παχυσαρκίας και πολυωνυμική λογιστική παλινδρόμηση για την ανάλυση των διαφορετικών επιπέδων της.

Τα αποτελέσματα έδειξαν ότι σημαντικό ρόλο στην εμφάνιση παχυσαρκίας παίζει το οικογενειακό ιστορικό, η ηλικία, η ημερήσια κατανάλωση νερού (CH_2O), η εβδομαδιαία φυσική δραστηριότητα (FAF) και σε ορισμένες περιπτώσεις οι διατροφικές συνήθειες όπως τα ενδιάμεσα γεύματα (CAEC). Ιδιαίτερα, το οικογενειακό ιστορικό με παχυσαρκία αυξάνει σημαντικά τις πιθανότητες εμφάνισης παχυσαρκίας, ενώ η αυξημένη φυσική δραστηριότητα και η επαρκής κατανάλωση νερού φαίνεται να λειτουργούν προστατευτικά.

Συμπερασματικά, η παχυσαρκία είναι μια πολυπαραγοντική κατάσταση που επηρεάζεται από το οικογενειακό υπόβαθρο, τη συμπεριφορά και τον τρόπο ζωής, και τα ευρήματα της μελέτης μπορούν να συμβάλουν στον σχεδιασμό στοχευμένων παρεμβάσεων για την πρόληψη και την αντιμετώπισή της.

ABSTRACT

Obesity is one of the major public health challenges worldwide, as it is associated with an increased risk of cardiometabolic diseases, type 2 diabetes and other severe health complications. Beyond its physical consequences, obesity also negatively affects psychosocial well-being and overall quality of life.

The aim of this study was to investigate the factors associated with the occurrence and severity of obesity, focusing on the influence of demographic characteristics, dietary habits and physical activity.

The analysis was based on data from 485 individuals originating from Mexico, Peru and Colombia, including demographic information, family history of obesity, dietary patterns and lifestyle behaviors.

Descriptive statistics, correlation tests (Pearson's χ^2 , Kruskal–Wallis non-parametric test, Kolmogorov–Smirnov normality test) as well as regression models were applied. In particular, binary logistic regression was used to explore the predictors of obesity presence, while multinomial logistic regression was applied to analyze the different obesity levels.

The results indicated that family history of obesity, age, daily water consumption (CH₂O), physical activity per week (FAF), and, in some cases, eating habits such as the frequency of food consumption between meals (CAEC), play an important role. Specifically, a positive family history significantly increases the likelihood of obesity, whereas higher physical activity and adequate water intake appear to be protective factors.

In conclusion, obesity is a multifactorial condition influenced by family background, behavior, and lifestyle, and the findings of this study may contribute to the design of targeted prevention and intervention strategies.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ.....	xv
ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ.....	xix
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ.....	xxi

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

1.1 Εισαγωγή.....	1
1.1.1 Ιστορία.....	1
1.1.2 Τύποι παχυσαρκίας.....	1
1.1.3 Ταξινόμηση παχυσαρκίας.....	1
1.1.4 Κατηγορίες παχυσαρκίας.....	2
1.2 Παιδική παχυσαρκία.....	3
1.3 Αίτια παχυσαρκίας.....	4
1.3.1 Διατροφή.....	4
1.3.2 Έλλειψη γυμναστικής.....	5
1.3.3 Γενετικοί και κοινωνικοί παράγοντες.....	6
1.4 Επιπτώσεις παχυσαρκίας.....	6
1.5 Αντιμετώπιση παχυσαρκίας.....	8
1.6 Παχυσαρκία στην Ελλάδα.....	9
1.7 Μεθοδολογία.....	12
1.7.1 Παρουσίαση μεταβλητών.....	12
1.7.2 Ο έλεγχος χ^2	12
1.7.3 Συντελεστής V του Cramer.....	13
1.7.4 Έλεγχος Κανονικότητας.....	13
1.7.5 Ο έλεγχος Kruskal-Wallis.....	14
1.7.6 Διωνυμική Λογιστική Παλινδρόμηση.....	15
1.7.7 Ανάλυση ROC.....	16
1.7.8 Πολυωνυμική Λογιστική Παλινδρόμηση.....	17
1.7.9 Το κριτήριο AIC.....	18
1.7.10 Δείκτης Nagelkerke R^2	18
1.7.11 Λόγος σχετικών πιθανοτήτων (Odds Ratio – OR).....	19

ΚΕΦΑΛΑΙΟ 2: ΠΕΡΙΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ

2.2 Περιγραφική Ανάλυση	20
2.3 Δημογραφικά χαρακτηριστικά	20
2.3.1 Περιγραφική ανάλυση μεταβλητής «Φύλο».....	20

2.3.2	Περιγραφική ανάλυση μεταβλητής «Ηλικία».....	21
2.3.3	Περιγραφική ανάλυση μεταβλητής «Ύψος».....	22
2.3.4	Περιγραφική ανάλυση μεταβλητής «Βάρος».....	23
2.3.5	Περιγραφική ανάλυση μεταβλητής «Οικογενειακό ιστορικό παχυσαρκίας».....	24
2.4	Διατροφικές Συνήθειες	25
2.4.1	Περιγραφική ανάλυση μεταβλητής «FAVC».....	25
2.4.2	Περιγραφική ανάλυση μεταβλητής «FCVC».....	26
2.4.3	Περιγραφική ανάλυση μεταβλητής «NCP».....	27
2.4.4	Περιγραφική ανάλυση μεταβλητής «CAEC».....	28
2.4.5	Περιγραφική ανάλυση μεταβλητής «CH2O».....	29
2.4.6	Περιγραφική ανάλυση μεταβλητής «SCC».....	30
2.4.7	Περιγραφική ανάλυση μεταβλητής «CALC».....	31
2.5	Φυσική Δραστηριότητα και άλλες συνήθειες.....	32
2.5.1	Περιγραφική ανάλυση μεταβλητής «FAF».....	32
2.5.2	Περιγραφική ανάλυση μεταβλητής «TUE».....	33
2.5.3	Περιγραφική ανάλυση μεταβλητής «MTRANS».....	34
2.5.4	Περιγραφική ανάλυση μεταβλητής «SMOKE».....	35
2.6	Επίπεδο παχυσαρκίας.....	36
2.6.1	Περιγραφική ανάλυση μεταβλητής «NObeyesdad».....	36
2.6.2	Περιγραφική ανάλυση μεταβλητής «NewObesity».....	38
2.7	Δείκτης Μάζας Σώματος.....	39
2.7.1	Περιγραφική ανάλυση «ΔΜΣ».....	39

ΚΕΦΑΛΑΙΟ 3: ΣΥΣΧΕΤΙΣΕΙΣ

3.1	Συσχετίσεις	41
3.2	Συσχετίσεις δημογραφικών στοιχείων με το «Επίπεδο παχυσαρκίας».....	42
3.2.1	Συσχέτιση με τη μεταβλητή «Φύλο».....	42
3.2.2	Συσχέτιση με τη μεταβλητή «Ηλικία».....	43
3.2.3	Συσχέτιση με τη μεταβλητή «Βάρος».....	45
3.2.4	Συσχέτιση με τη μεταβλητή «Ύψος».....	47
3.2.5	Συσχέτιση με τη μεταβλητή «Οικογενειακό ιστορικό παχυσαρκίας».....	49
3.3	Συσχετίσεις διατροφικών συνηθειών με το «Επίπεδο παχυσαρκίας».....	50
3.3.1	Συσχέτιση με τη μεταβλητή «FAVC».....	51
3.3.2	Συσχέτιση με τη μεταβλητή «FCVC».....	52
3.3.3	Συσχέτιση με τη μεταβλητή «NCP».....	53
3.3.4	Συσχέτιση με τη μεταβλητή «CAEC».....	54
3.3.5	Συσχέτιση με τη μεταβλητή «CH2O».....	57
3.3.6	Συσχέτιση με τη μεταβλητή «SCC».....	58
3.3.7	Συσχέτιση με τη μεταβλητή «CALC».....	59
3.4	Συσχετίσεις φυσικής δραστηριότητας και άλλων συνηθειών με το «Επίπεδο παχυσαρκίας».....	60
3.4.1	Συσχέτιση με τη μεταβλητή «FAF».....	61

3.4.2	Συσχέτιση με τη μεταβλητή «TUE».....	62
3.4.3	Συσχέτιση με τη μεταβλητή «MTRANS».....	64
3.4.4	Συσχέτιση με τη μεταβλητή «SMOKE».....	65
3.5	Συσχέτιση BMI με το «Επίπεδο παχυσαρκίας».....	67

ΚΕΦΑΛΑΙΟ 4: ΜΟΝΤΕΛΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

4.1	Στατιστική Ανάλυση.....	70
4.2	Διωνυμική Λογιστική Παλινδρόμηση.....	70
4.2.1	Πρώτη εφαρμογή Διωνυμικής Λογιστικής Παλινδρόμησης.....	71
4.2.2	Δεύτερη εφαρμογή Διωνυμικής Λογιστικής Παλινδρόμησης.....	74
4.3	Πολυωνυμική Λογιστική Παλινδρόμηση.....	79
4.3.1	Πρώτη εφαρμογή Πολυωνυμικής Λογιστικής Παλινδρόμησης.....	79
4.3.2	Ενδιάμεσες εφαρμογές Πολυωνυμικής Λογιστικής Παλινδρόμησης.....	81
4.3.3	Τελευταία εφαρμογή Πολυωνυμικής Λογιστικής Παλινδρόμησης.....	85

ΚΕΦΑΛΑΙΟ 5: ΣΥΖΗΤΗΣΗ

5.1	Εισαγωγή	89
5.2	Αποτελέσματα από περιγραφική ανάλυση	89
5.3	Συσχετίσεις	90
5.4	Αποτελέσματα και συμπεράσματα από τις αναλύσεις.....	90

ΒΙΒΛΙΟΓΡΑΦΙΑ.....92

ΠΑΡΑΡΤΗΜΑ.....95

Κατάλογος Πινάκων

1.1: Πίνακας ταξινόμησης επιπέδων παχυσαρκίας σύμφωνα με το ΔΜΣ.....	2
2.2.1: Στατιστικά στοιχεία για το φύλο.....	20
2.2.2: Στατιστικά στοιχεία για την ηλικία.....	21
2.2.3: Στατιστικά στοιχεία για το ύψος.....	22
2.2.4: Στατιστικά στοιχεία για το βάρος.....	23
2.2.5: Στατιστικά στοιχεία για το οικογενειακό ιστορικό παχυσαρκίας.....	24
2.3.1: Στατιστικά στοιχεία για την κατανάλωση τροφής με πολλές θερμίδες.....	25
2.3.2: Στατιστικά στοιχεία για τη συχνότητα κατανάλωσης λαχανικών.....	26
2.3.3: Στατιστικά στοιχεία για τον αριθμό κυρίων γευμάτων.....	27
2.3.4: Στατιστικά στοιχεία για την κατανάλωση τροφής μεταξύ των γευμάτων.....	28
2.3.5: Στατιστικά για την ποσότητα νερού που καταναλώνεται καθημερινά.....	29
2.3.6: Στατιστικά στοιχεία για την καθημερινή μέτρηση θερμίδων.....	30
2.3.7: Στατιστικά στοιχεία για τη συχνότητα κατανάλωσης αλκοόλ.....	31
2.4.1: Στατιστικά στοιχεία για τη συχνότητα σωματικής άσκησης.....	33
2.4.2: Στατιστικά στοιχεία χρόνου χρήσης τεχνολογικών συσκευών.....	34
2.4.3: Στατιστικά στοιχεία για τον τρόπο μετακίνησης.....	35
2.4.4: Στατιστικά στοιχεία για καπνιστές.....	36
2.5.1: Επίπεδο παχυσαρκίας.....	37
2.5.2: Στατιστικά στοιχεία νέων επιπέδων παχυσαρκίας.....	38
2.6.1(α): Πίνακας ταξινόμησης επιπέδων παχυσαρκίας σύμφωνα με το ΔΜΣ.....	39
2.6.1(β): Στατιστικά στοιχεία για το Δείκτη Μάζας Σώματος.....	39
3.2.1: Έλεγχος ανεξαρτησίας φύλου και επιπέδου παχυσαρκίας.....	42
3.2.2(α): Έλεγχος κανονικότητας Kolmogorov-Smirnov των μεταβλητών «ηλικία» και «επίπεδο παχυσαρκίας».....	43
3.2.2(β): Πίνακας ελέγχου συσχετίσεων Kruskal-Wallis των μεταβλητών «ηλικία» και «επίπεδο παχυσαρκίας»	44
3.2.3(α): Έλεγχος κανονικότητας Kolmogorov-Smirnov των μεταβλητών «βάρος» και «επίπεδο παχυσαρκίας».....	45
3.2.3(β): Πίνακας ελέγχου συσχετίσεων Kruskal-Wallis των μεταβλητών «βάρος» και «επίπεδο παχυσαρκίας»	46

3.2.4(α): Έλεγχος κανονικότητας Kolmogorov-Smirnov των μεταβλητών «ύψος» και «επίπεδο παχυσαρκίας»	47
3.2.4(β): Πίνακας ελέγχου συσχέτισεων Kruskal-Wallis των μεταβλητών «ύψος» και «επίπεδο παχυσαρκίας»	48
3.2.5(α): Έλεγχος οικογενειακού ιστορικού παχυσαρκίας και επιπέδου παχυσαρκίας.....	49
3.2.5(β): Ένταση συσχέτισης οικογενειακού ιστορικού παχυσαρκίας και επιπέδου παχυσαρκίας.....	50
3.3.1: Έλεγχος ανεξαρτησίας κατανάλωσης τροφής με πολλές θερμίδες και επιπέδου παχυσαρκίας.....	51
3.3.2: Έλεγχος ανεξαρτησίας συχνότητας κατανάλωσης λαχανικών μέσα στα γεύματα και επιπέδου παχυσαρκίας.....	52
3.3.3(α): Έλεγχος ανεξαρτησίας ποσότητας καθημερινής κατανάλωσης κύριων γευμάτων και επιπέδου παχυσαρκίας.....	53
3.3.3(β): Ένταση συσχέτισης ποσότητας καθημερινής κατανάλωσης κύριων γευμάτων και επιπέδου παχυσαρκίας.....	54
3.3.4(α): Έλεγχος ανεξαρτησίας συχνότητας κατανάλωσης τροφής μεταξύ των γευμάτων και επιπέδου παχυσαρκίας με μέθοδο Monte Carlo.....	55
3.3.4(β): Ένταση συσχέτισης συχνότητας κατανάλωσης τροφής μεταξύ των γευμάτων και επιπέδου παχυσαρκίας.....	56
3.3.5(α): Έλεγχος ανεξαρτησίας ποσότητας καθημερινής κατανάλωσης νερού και επιπέδου παχυσαρκίας.....	57
3.3.5(β): Ένταση συσχέτισης ποσότητας καθημερινής κατανάλωσης νερού και επιπέδου παχυσαρκίας.....	58
3.3.6: Έλεγχος ανεξαρτησίας καθημερινής μέτρησης θερμίδων και επιπέδου παχυσαρκίας.....	59
3.3.7: Έλεγχος ανεξαρτησίας συχνότητας κατανάλωσης αλκοόλ και επιπέδου παχυσαρκίας με μέθοδο Monte Carlo.....	60
3.4.1(α): Έλεγχος ανεξαρτησίας συχνότητας σωματικής άσκησης και επιπέδου παχυσαρκίας.....	61
3.4.1(β): Ένταση συσχέτισης συχνότητας σωματικής άσκησης και επιπέδου παχυσαρκίας.....	62
3.4.2(α): Έλεγχος ανεξαρτησίας χρόνου χρήσης τεχνολογικών συσκευών και επιπέδου παχυσαρκίας.....	63
3.4.2(β): Ένταση συσχέτισης χρόνου χρήσης τεχνολογικών συσκευών και επιπέδου παχυσαρκίας.....	63
3.4.3: Έλεγχος ανεξαρτησίας τρόπου μετακίνησης και επιπέδου παχυσαρκίας με τη μέθοδο Monte Carlo.....	65

3.4.4(α): Έλεγχος ανεξαρτησίας καπνιστών και επιπέδου παχυσαρκίας με τη μέθοδο Monte Carlo.....	66
3.4.4(β): Ένταση συσχέτισης καπνίσματος και επιπέδου παχυσαρκίας.....	66
3.5(α): Έλεγχος κανονικότητας Kolmogorov-Smirnov των μεταβλητών «BMI» και «επίπεδο παχυσαρκίας»	67
3.5(β): Πίνακας ελέγχου συσχέτισεων Kruskal-Wallis των μεταβλητών «BMI» και «επίπεδο παχυσαρκίας»	68
4.1: Περιγραφή του δείγματος.....	70
4.2: Στατιστικά στοιχεία νέων επιπέδων παχυσαρκίας.....	71
4.2.1(α): Κωδικοποίηση κατηγορικών μεταβλητών.....	72
4.2.1(β): Ακρίβεια πρόβλεψης του πρώτου μοντέλου Λογιστικής Παλινδρόμησης.....	72
4.2.1(γ): Έλεγχος καλής προσαρμογής Hosmer και Lemeshow πρώτου μοντέλου Λογιστικής Παλινδρόμησης.....	73
4.2.1(δ): Μεταβλητές του πρώτου μοντέλου Λογιστικής Παλινδρόμησης.....	73
4.2.1(ε): Εμβადόν κάτω από καμπύλη τελευταίου μοντέλου Λογιστικής Παλινδρόμησης.....	74
4.2.2(α): Ακρίβεια πρόβλεψης του τελικού μοντέλου Λογιστικής Παλινδρόμησης.....	74
4.2.2(β): Έλεγχος καλής προσαρμογής Hosmer και Lemeshow τελικού μοντέλου Λογιστικής Παλινδρόμησης.....	75
4.2.2(γ): Μεταβλητές του τελικού μοντέλου Λογιστικής Παλινδρόμησης.....	75
4.2.2(δ): Εμβადόν κάτω από καμπύλη τελικού μοντέλου Λογιστικής Παλινδρόμησης.....	76
4.2.2(ε): Μεταβλητές των μοντέλων με τη μέθοδο Stepwise	77
4.2.2(στ): Δείκτες προσαρμογής του μοντέλου με τη μέθοδο Stepwise.....	78
4.2.2(ζ): Ακρίβεια πρόβλεψης μοντέλων με τη μέθοδο Stepwise.....	78
4.3.1(α): Σύγκριση αρχικού και τελικού μοντέλου (1^η προσπάθεια).....	79
4.3.1(β): Ακρίβεια πρόβλεψης του πρώτου μοντέλου Πολυωνυμικής Παλινδρόμησης.....	80
4.3.1(γ): Μεταβλητές του πρώτου μοντέλου Πολυωνυμικής Παλινδρόμησης	81
4.3.1(δ): Δείκτες Ψευδο-R² του 1^{ου} μοντέλου Πολυωνυμικής Παλινδρόμησης.....	81
4.3.2(α): Σύγκριση αρχικού και τελικού μοντέλου (2^η προσπάθεια).....	82
4.3.2(β): Ακρίβεια πρόβλεψης του δεύτερου μοντέλου Πολυωνυμικής Παλινδρόμησης.....	82
4.3.2(γ): Μεταβλητές του δεύτερου μοντέλου Πολυωνυμικής Παλινδρόμησης.....	83
4.3.2(δ): Δείκτες Ψευδο-R² του 2^{ου} μοντέλου Πολυωνυμικής Παλινδρόμησης.....	83
4.3.2(ε): Σύγκριση αρχικού και τελικού μοντέλου (3^η προσπάθεια).....	84

4.3.2(στ): Ακρίβεια πρόβλεψης του τρίτου μοντέλου Πολυωνυμικής Παλινδρόμησης.....	84
4.3.2(ζ): Μεταβλητές του τρίτου μοντέλου Πολυωνυμικής Παλινδρόμησης.....	85
4.3.2(η): Δείκτες Ψευδο- R^2 του 3 ^{ου} μοντέλου Πολυωνυμικής Παλινδρόμησης.....	85
4.3.3(α): Σύγκριση αρχικού και τελικού μοντέλου (τελική προσπάθεια).....	85
4.3.3(β): Ακρίβεια πρόβλεψης του τελικού μοντέλου Πολυωνυμικής Παλινδρόμησης	86
4.3.3(γ): Δείκτες Ψευδο- R^2 του τελικού μοντέλου Πολυωνυμικής Παλινδρόμησης	86
4.3.3(δ): Μεταβλητές του τελικού μοντέλου Πολυωνυμικής Παλινδρόμησης.....	87
4.3.3(ε): Εκτίμηση παραμέτρων για το τελικό μοντέλο της Πολυωνυμικής Παλινδρόμησης.....	88

Κατάλογος Σχημάτων

Σχήμα 1.7.7: Παράδειγμα καμπύλης ROC.....	17
Διάγραμμα 2.2.1: Κατανομή φύλου.....	21
Διάγραμμα 2.2.2: Ιστόγραμμα για την κατανομή των ηλικιών.....	22
Διάγραμμα 2.2.3: Ιστόγραμμα κατανομής ύψους.....	23
Διάγραμμα 2.2.4: Ιστόγραμμα κατανομής βάρους.....	24
Διάγραμμα 2.2.5: Κατανομή οικογενειακού ιστορικού παχυσαρκίας.....	25
Διάγραμμα 2.3.1: Κατανομή κατανάλωσης τροφής με πολλές θερμίδες.....	26
Διάγραμμα 2.3.2: Κατανομή συχνότητας κατανάλωσης λαχανικών.....	27
Διάγραμμα 2.3.3: Κατανομή αριθμού κυρίων γευμάτων.....	28
Διάγραμμα 2.3.4: Κατανομή κατανάλωσης τροφής μεταξύ των γευμάτων.....	29
Διάγραμμα 2.3.5: Κατανομή ποσότητας νερού που καταναλώνεται καθημερινά.....	30
Διάγραμμα 2.3.6: Κατανομή καθημερινής μέτρησης θερμίδων.....	31
Διάγραμμα 2.3.7: Κατανομή συχνότητας κατανάλωσης αλκοόλ.....	32
Διάγραμμα 2.4.1: Κατανομή συχνότητας σωματικής άσκησης.....	33
Διάγραμμα 2.4.2: Κατανομή χρόνου χρήσης τεχνολογικών συσκευών.....	34
Διάγραμμα 2.4.3: Κατανομή τρόπου κίνησης.....	35
Διάγραμμα 2.4.4: Κατανομή των καπνιστών.....	36
Διάγραμμα 2.5.1: Κατανομή επιπέδου παχυσαρκίας.....	37
Διάγραμμα 2.5.2: Κατανομή νέων επιπέδων παχυσαρκίας.....	38
Διάγραμμα 2.6.1: Ιστόγραμμα κατανομής Δείκτη Μάζας Σώματος.....	40
Διάγραμμα 3.2.2: Boxplot επιπέδου παχυσαρκίας σε σχέση με την ηλικία.....	44
Διάγραμμα 3.2.3: Boxplot επιπέδου παχυσαρκίας σε σχέση με το βάρος.....	46
Διάγραμμα 3.2.4: Boxplot επιπέδου παχυσαρκίας σε σχέση με το ύψος.....	48
Διάγραμμα 3.2.5: Ραβδόγραμμα συσχέτισης οικογενειακού ιστορικού παχυσαρκίας με το επίπεδο παχυσαρκίας.....	50
Διάγραμμα 3.3.3: Ραβδόγραμμα συσχέτισης της ποσότητας καθημερινής κατανάλωσης κύριων γευμάτων με το επίπεδο παχυσαρκίας.....	54
Διάγραμμα 3.3.4: Ραβδόγραμμα συσχέτισης της συχνότητας κατανάλωσης τροφής μεταξύ των γευμάτων με το επίπεδο παχυσαρκίας.....	56

Διάγραμμα 3.3.5: Ραβδόγραμμα συσχέτισης της ποσότητας καθημερινής κατανάλωσης νερού με το επίπεδο παχυσαρκίας.....	58
Διάγραμμα 3.4.1: Ραβδόγραμμα συσχέτισης συχνότητας σωματικής άσκησης με το επίπεδο παχυσαρκίας.....	62
Διάγραμμα 3.4.2: Ραβδόγραμμα συσχέτισης του χρόνου χρήσης τεχνολογικών συσκευών με το επίπεδο παχυσαρκίας.....	64
Διάγραμμα 3.4.4: Ραβδόγραμμα συσχέτισης του καπνίσματος με το επίπεδο παχυσαρκίας...	67
Διάγραμμα 3.5: Boxplot επιπέδου παχυσαρκίας σε σχέση με το BMI.....	69
Διάγραμμα 4.2: Κατανομή νέων επιπέδων παχυσαρκίας.....	71
Διάγραμμα 4.2.1(α): Καμπύλη ROC δεύτερου μοντέλου Λογιστικής Παλινδρόμησης.....	74
Διάγραμμα 4.2.2(α): Καμπύλη ROC τελευταίου μοντέλου Λογιστικής Παλινδρόμησης.....	76

Κατάλογος Εικόνων

1.3.1: Συγκριτική ημερήσια κατανάλωση θερμίδων.....	5
1.4.1: Επιπτώσεις παχυσαρκίας στην υγεία.....	7
1.4.2: Εκτιμώμενος αριθμός θανάτων παγκοσμίως ανά παράγοντα κινδύνου το 2021.....	8
1.6.1: Εκτιμώμενος αριθμός θανάτων ανά παράγοντα κινδύνου το 2021 στην Ελλάδα.....	9
1.6.2: Ποσοστά παχυσαρκίας ανδρών-γυναικών σε Ελλάδα και Ευρώπη.....	10
1.6.3: Ημερήσια κατανάλωση θερμίδων στην Ελλάδα.....	10
1.6.4: Εξέλιξη παιδικής παχυσαρκίας στην Ελλάδα.....	11
3.2.1: Πίνακας συνάφειας φύλου και επιπέδου παχυσαρκίας.....	42
3.2.2: Q-Q Plot γραφήματα για το επίπεδο παχυσαρκίας σε σχέση με την ηλικία.....	43
3.2.3: Q-Q Plot γραφήματα για το επίπεδο παχυσαρκίας σε σχέση με το βάρος.....	45
3.2.4: Q-Q Plot γραφήματα για το επίπεδο παχυσαρκίας σε σχέση με το ύψος.....	47
3.2.5: Πίνακας συνάφειας οικογενειακού ιστορικού παχυσαρκίας και επιπέδου παχυσαρκίας.....	49
3.3.1: Πίνακας συνάφειας κατανάλωσης τροφής με πολλές θερμίδες και επιπέδου παχυσαρκίας.....	51
3.3.2: Πίνακας συνάφειας συχνότητας κατανάλωσης λαχανικών και επιπέδου παχυσαρκίας.....	52
3.3.3: Πίνακας συνάφειας ποσότητας καθημερινής κατανάλωσης κύριων γευμάτων και επιπέδου παχυσαρκίας.....	53
3.3.4: Πίνακας συνάφειας συχνότητας κατανάλωσης τροφής μεταξύ των γευμάτων και επιπέδου παχυσαρκίας.....	55
3.3.5: Πίνακας συνάφειας ποσότητας καθημερινής κατανάλωσης νερού και επιπέδου παχυσαρκίας.....	57
3.3.6: Πίνακας συνάφειας καθημερινής μέτρησης θερμίδων και επιπέδου παχυσαρκίας.....	59
3.3.7: Πίνακας συνάφειας συχνότητας κατανάλωσης αλκοόλ και επιπέδου παχυσαρκίας.....	60
3.4.1: Πίνακας συνάφειας συχνότητας σωματικής άσκησης και επιπέδου παχυσαρκίας.....	61
3.4.2: Πίνακας συνάφειας χρόνου χρήσης τεχνολογικών συσκευών και επιπέδου παχυσαρκίας.....	63
3.4.3: Πίνακας συνάφειας τρόπου μετακίνησης και επιπέδου παχυσαρκίας.....	65
3.4.4: Πίνακας συνάφειας καπνιστών και επιπέδου παχυσαρκίας.....	66
3.5: Q-Q Plot γραφήματα για το επίπεδο παχυσαρκίας σε σχέση με το BMI.....	68

Κεφάλαιο 1

Εισαγωγή

1.1 Εισαγωγή

Η παχυσαρκία ορίζεται ως το νόσημα που προκαλείται από υπερβολική συγκέντρωση λίπους στο σώμα και αποτελεί ένα από τα μεγαλύτερα διατροφικά προβλήματα της σύγχρονης κοινωνίας. Αναγνωρίστηκε επίσημα ως νόσος το 1948 από τον Παγκόσμιο Οργανισμό Υγείας (ΠΟΥ) ενώ από το 1997 έχει κηρυχθεί ως ασθένεια επιδημικού χαρακτήρα. Ο αριθμός των ανθρώπων που πάσχουν από αυτήν έχει διπλασιαστεί από το 1980. Συγκεκριμένα, από έρευνα του ΠΟΥ το 2022 προκύπτει ότι 2,5 δισεκατομμύρια ενήλικες είναι υπέρβαροι. Από αυτούς, τα 890 εκατομμύρια είναι παχύσαρκοι.

1.1.1 Ιστορία

Η λέξη παχυσαρκία προκύπτει από τη σύνθεση των λέξεων «παχύς» και «σάρκα» και περιγράφει την κατάσταση κατά την οποία υπάρχει υπερβάλλουσα συσσώρευση λίπους στο ανθρώπινο σώμα. Η αρχαία ελληνική ιατρική καθώς και οι αρχαίοι Αιγύπτιοι αναγνώριζαν την παχυσαρκία ως ιατρική διαταραχή. Ο Ιπποκράτης μάλιστα είχε γράψει ότι «η παχυσαρκία δεν είναι απλά μια ασθένεια, αλλά είναι ο προάγγελος άλλων».

Κατά το μεγαλύτερο μέρος της ανθρώπινης ιστορίας, η ανθρωπότητα πάλευε με την έλλειψη τροφίμων, επομένως η παχυσαρκία έχει ιστορικά θεωρηθεί ως δείγμα πλούτου και ευημερίας. Πλέον όμως τα ποσοστά παχυσαρκίας αυξάνονται τόσο στις αναπτυγμένες χώρες όσο και στις αναπτυσσόμενες, με μόνη εξαίρεση την περιοχή της υποσαχάριας Αφρικής. Μάλιστα, υπάρχουν περισσότερα παχύσαρκα άτομα παρά λιποβαρή σε όλο τον κόσμο εκτός της περιοχής της νοτιοανατολικής Ασίας.

1.1.2 Τύποι παχυσαρκίας

Υπάρχουν δύο τύποι παχυσαρκίας, η κεντρική και η περιφερική. Στην κεντρική υπάρχει αυξημένη κατανομή του λίπους στο άνω τμήμα του σώματος ενώ στην περιφερική υπάρχει συσσώρευση λίπους στους μηρούς ή στους γλουτούς ή και συνδυασμό των δύο. Η κεντρική αποτελεί τον πιο επικίνδυνο τύπο παχυσαρκίας καθώς σχετίζεται με υψηλή συγκέντρωση σπλαχνικού λίπους και έχει αποδειχθεί ότι αυξάνει τη νοσηρότητα και τη θνησιμότητα.

1.1.3 Ταξινόμηση παχυσαρκίας

Η ταξινόμηση της παχυσαρκίας παλαιότερα γινόταν με βάση πίνακες των ασφαλιστικών εταιρειών οι οποίοι όριζαν το «ιδανικό» σωματικό βάρος ανάλογα με τη σωματική διάπλαση σε σχέση με το ύψος. Έτσι παχύσαρκος ήταν κάποιος ο οποίος το σωματικό του βάρος ξεπερνούσε το 120% του «ιδανικού» σωματικού βάρους. Αργότερα, ο ΠΟΥ άρχισε να χρησιμοποιεί το Δείκτη Μάζας Σώματος (ΔΜΣ) για την ταξινόμηση. Ο ΔΜΣ ορίζεται ως το βάρος του σώματος σε χιλιόγραμμα (kg) διαιρούμενο από το τετράγωνο του ύψους σε μέτρα.

Έτσι λοιπόν ο ΔΜΣ σχετίζεται με το ποσοστό λίπους στο σώμα και έχει μικρή συσχέτιση με το ύψος. Οπότε ένας άντρας θεωρείται παχύσαρκος αν το ποσοστό του σωματικού του λίπους ξεπερνά το 25%, ενώ μια γυναίκα όταν το ποσοστό αυτό ξεπερνά το 35%.

Η ταξινόμηση ανάλογα με το αποτέλεσμα του ΔΜΣ είναι η εξής (De la Hoz Manotas, A., & Mendoza Palechor, F., 2019)

ΔΜΣ	Ταξινόμηση
<18,5	Λιποβαρής
18,5-24,9	Κανονικό βάρος
25-29,9	Υπέρβαρος
30-34,9	Παχυσαρκία Τύπου Ι
35-39,9	Παχυσαρκία Τύπου ΙΙ
>40	Παχυσαρκία Τύπου ΙΙΙ

Πίνακας 1.1: Πίνακας ταξινόμησης επιπέδων παχυσαρκίας σύμφωνα με το ΔΜΣ

Ένας άλλος τρόπος ταξινόμησης της παχυσαρκίας είναι η Waiste to Hip Ratio, δηλαδή η σχέση της μικρότερης περιμέτρου της μέσης προς τη μεγαλύτερη περίμετρο των ισχίων. Είναι μία μέθοδος που δεν εξαρτάται από το ύψος, συνεπώς θεωρείται απλούστερη από το ΔΜΣ και είναι εξίσου αξιόπιστη. Με τη διαδικασία αυτή γίνεται εκτίμηση της κεντρικής παχυσαρκίας καθώς επίσης και του κινδύνου ανάπτυξης καρδιαγγειακών και άλλων νοσημάτων σχετικών με την παχυσαρκία. Για τις γυναίκες, μια τιμή περιφέρειας μέσης άνω των 88 εκατοστών θεωρείται αυξημένη, ενώ για τους άνδρες η αντίστοιχη τιμή είναι τα 102 εκατοστά.

1.1.4 Κατηγορίες παχυσαρκίας

Με βάση το ΔΜΣ, μπορούμε να περιγράψουμε τις κατηγορίες παχυσαρκίας ως εξής:

- **Λιποβαρής:** Όταν ο ΔΜΣ είναι μικρότερος από 20 kg/m² αυτό σημαίνει ότι ο άνθρωπος είναι πολύ πιο αδύνατος από το φυσιολογικό. Η κατάσταση αυτή δεν σημαίνει ότι είναι καλή και φυσιολογική για την υγεία. Τα λιποβαρή άτομα είναι ευάλωτα σε λοιμώξεις και ασθένειες και επίσης δεν έχουν ενεργειακά αποθέματα για την αντιμετώπιση στρεσογόνων καταστάσεων.
- **Κανονικό βάρος:** Όταν ο ΔΜΣ είναι 20-25 kg/m² αυτό σημαίνει ότι αυτό είναι το καλύτερο επίπεδο υγείας του ανθρώπου.
- **Υπέρβαρος:** Όταν το BMI είναι 25-30 kg/m² σημαίνει ότι το άτομο είναι υπέρβαρο, αυτό είναι το υποκλινικό στάδιο της παχυσαρκίας και οι μηχανισμοί οργανικής βλάβης όσο κι αν δεν είναι εμφανής έχουν αρχίσει να δρουν. Ουσιαστικά είναι η είσοδος στην παχυσαρκία και το καλύτερο στάδιο για προληπτική αγωγή.
- **Παχυσαρκία Τύπου Ι:** Είναι το πρώτο στάδιο της παχυσαρκίας, με ΔΜΣ 30-35 kg/m², όπου έχουν αρχίσει και εκφράζονται οι επιπλοκές και οι δυσκολίες στην καθημερινότητα. Εδώ εμφανίζονται διάφορα συμπτώματα όπως μικρή υπέρταση, αυξημένο σάκχαρο αίματος λίγο πιο πάνω από τα φυσιολογικά, κόπωση χωρίς να έχουμε κάνει κάτι ασυνήθιστο και διάφορα άλλα.

- Παχυσαρκία Τύπου II: Είναι το δεύτερο στάδιο της παχυσαρκίας, με ΔΜΣ 35-40 kg/m² γνωστή και ως κλινική παχυσαρκία. Στο στάδιο αυτό εμφανίζονται περισσότερα συμπτώματα τα οποία δυσκολεύουν την καθημερινότητα του ασθενούς. Επίσης εμφανίζονται διάφορα νοσήματα τα οποία συνδέονται με την παχυσαρκία όπως υπέρταση, διαβήτης, υπνική άπνοια, μυοσκελετικά προβλήματα και άλλα τα οποία χρήζουν φαρμακευτικής αγωγής.
- Παχυσαρκία Τύπου III: Είναι το τρίτο και βαρύτερο στάδιο της παχυσαρκίας, με ΔΜΣ μεγαλύτερο από 40 kg/m². Ονομάζεται και νοσογόνος ή θνησιγόνος παχυσαρκία καθώς επηρεάζει το προσδόκιμο ζωής.

1.2 Παιδική παχυσαρκία

Η παιδική παχυσαρκία αποτελεί εξίσου σημαντικό πρόβλημα όπως η παχυσαρκία στους ενήλικες. Έλαβε διαστάσεις επιδημίας τον 21ο αιώνα, καθώς τα ποσοστά της αυξάνονται και στον ανεπτυγμένο, και στον αναπτυσσόμενο κόσμο. Σύμφωνα με έρευνα του 2022 από τον ΠΟΥ, υπολογίζεται ότι 37 εκατομμύρια παιδιά κάτω των 5 ετών, καθώς και πάνω από 390 εκατομμύρια παιδιά ηλικίας 5-19 ετών, είναι υπέρβαρα. Επιπλέον, η αυξητική τάση του αριθμού των υπέρβαρων ή παχύσαρκων παιδιών και εφήβων ηλικίας 5-19 ετών είναι παρόμοια σε αγόρια και κορίτσια, συγκεκριμένα το 2022 το 19% των κοριτσιών και το 21% των αγοριών είναι υπέρβαρα. Συνολικά, ενώ το 1990 μόνο το 2% των παιδιών και εφήβων ήταν παχύσαρκα, η έρευνα του 2022 έδειξε ότι ο αριθμός αυτός έχει φτάσει στο 8%, δηλαδή από 31 εκατομμύρια έχει αυξηθεί στα 160 εκατομμύρια.

Η παιδική παχυσαρκία κάποτε ήταν πρόβλημα των οικονομικά δυνατών χωρών αλλά πλέον έχει αυξηθεί και σε αναπτυσσόμενες χώρες. Στην ίδια έρευνα τα αποτελέσματα έδειξαν ότι ο μισός πληθυσμός των υπέρβαρων παιδιών κάτω των 5 ετών ζουν στην Ασία ενώ στην Αφρική ο αριθμός των παιδιών αυτών έχει αυξηθεί κατά 23% από το 2000. Ειδικά στην περιοχή της νοτιοανατολικής Ασίας η οποία γενικά είχε πολύ μικρά ποσοστά παχυσαρκίας, έχει παρατηρηθεί ανησυχητική αύξηση των αριθμών τα τελευταία 10-15 χρόνια. Αυτό οφείλεται κατά κύριο λόγο στη στροφή σε πιο ανθυγιεινούς τρόπους ζωής και συγκεκριμένα η ακμάζουσα οικονομική ανάπτυξη και η αστικοποίηση έχουν ως αποτέλεσμα τα παιδιά να μπορούν να τρώνε συνέχεια παχυντικά φαγητά με πολλά λιπαρά αλλά και να μην γυμνάζονται.

Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας, η παχυσαρκία στα παιδιά και τους εφήβους ηλικίας μεταξύ 2 και 19 ετών, ορίζεται με την βοήθεια της ειδικής καμπύλης του ΔΜΣ, σε εκατοστιαίες μονάδες, ανάλογα το φύλο και την ηλικία. Δηλαδή ένα παιδί ή ένας έφηβος με ΔΜΣ από 85 έως 95 εκατοστιαίες μονάδες θεωρείται υπέρβαρο, ενώ αν ο ΔΜΣ είναι πάνω από τις 95 μονάδες τότε θεωρείται παχύσαρκο.

Όπως και στους ενήλικες, έτσι και στα παιδιά υπάρχουν η κεντρική και η περιφερική παχυσαρκία και επηρεάζουν τον οργανισμό με τον ίδιο τρόπο. Έχει παρατηρηθεί ότι το 95% των παχύσαρκων παιδιών υποφέρουν από κεντρικού τύπου παχυσαρκία, που είναι και η πιο επικίνδυνη, ενώ ο αντίστοιχος αριθμός στα υπέρβαρα παιδιά ανέρχεται στο 70%. Ακόμα, το 75% των υπέρβαρων αγοριών παρουσιάζουν κεντρικού τύπου παχυσαρκία ενώ ο αντίστοιχος αριθμός στα κορίτσια ανέρχεται στο 64%.

Σε κάθε κοινωνία ο ρόλος της οικογένειας αλλά και του σχολείου κρίνεται ιδιαίτερα σημαντικός στο θέμα της παχυσαρκίας. Η οικογένεια πρέπει να φροντίζει ώστε τα παιδιά να θρέφονται με σωστό τρόπο και να μην παραγκωνίζονται οι φυσικές δραστηριότητες. Οι γονείς πρέπει να ασχολούνται με την υγεία των παιδιών τους παρά τον περιορισμένο ελεύθερο χρόνο τους ώστε τα παιδιά να καταλάβουν ότι μπορούν να τρώνε ακόμα και τα ανθυγιεινά φαγητά αλλά με μέτρο. Το σχολείο πρέπει και αυτό με τη σειρά του να διδάξει στα παιδιά τις αρετές της καλής διατροφής και να διοργανώνει εκδηλώσεις οι οποίες θα ευαισθητοποιούν το κοινό για το θέμα της παχυσαρκίας. Άλλωστε δεν είναι λίγα τα παχύσαρκα παιδιά που έχουν στιγματιστεί από τους συμμαθητές τους και πιθανότατα έχουν πέσει θύματα εκφοβισμού.

1.3 Αίτια παχυσαρκίας

Υπάρχουν πολλοί λόγοι που οδηγούν ένα άτομο στην παχυσαρκία. Το κύριο αίτιο, το οποίο είναι και από τα μεγαλύτερα προβλήματα της σύγχρονης κοινωνίας, θεωρείται ο συνδυασμός της υπερβολικής πρόσληψης ενέργειας από τροφές και της έλλειψης φυσικής δραστηριότητας. Ένα απλό παράδειγμα είναι το ότι λόγω περιορισμένου ελεύθερου χρόνου, στρεφόμεστε στις εύκολες επιλογές όπως είναι το γρήγορο και εύγευστο φαγητό που μπορούμε να παραγγείλουμε από την άνεση του σπιτιού μας. Αυτές οι επιλογές είναι που οδηγούν σε έναν ανθυγιεινό τρόπο ζωής καθώς, πρώτον, τα φαγητά αυτά είναι πλούσια σε λιπαρά ή υδατάνθρακες και δεύτερον, δεν υπάρχει ανάγκη για κάποια μετακίνηση από πλευράς μας.

1.3.1 Διατροφή

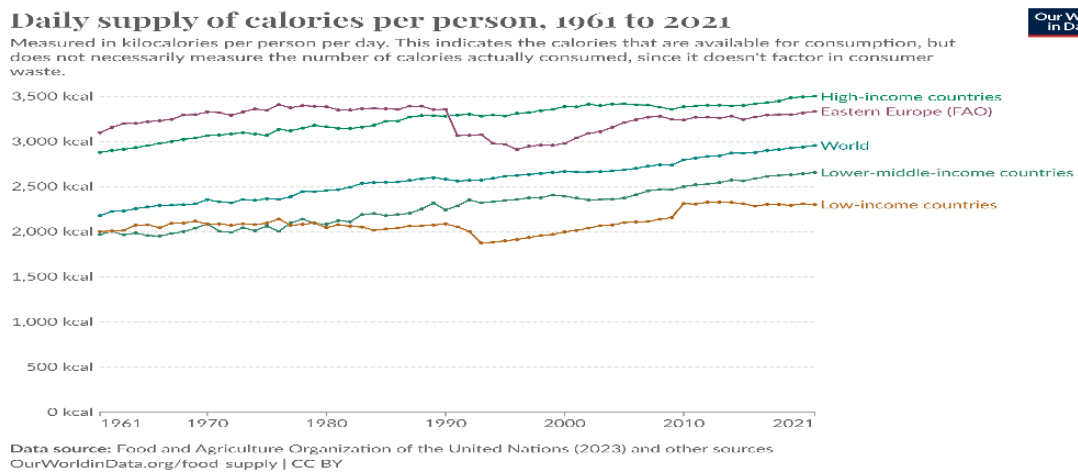
Ο τρόπος διατροφής επηρεάζει με διάφορους τρόπους το σωματικό βάρος. Ο αριθμός των θερμίδων που λαμβάνουμε καθημερινά εξαρτάται από τις διατροφικές μας συνήθειες και αλλάζει συνεχώς με το χρόνο και τον τόπο στον οποίο διαμένουμε. Σε χώρες με υψηλό εισόδημα είναι σύνηθες φαινόμενο η υπερκατανάλωση φαγητού και θερμίδων λόγω της μείωσης των τιμών των τροφίμων σε συνδυασμό με τις διαφημιστικές καμπάνιες των φαστ φουντ. Από τη δεκαετία του 1960 έως τα τέλη της δεκαετίας του 1990 ο μέσος όρος θερμίδων ανά άτομο και ανά ημέρα αυξήθηκε σε όλα τα μέρη του κόσμου εκτός από τις χώρες της Ανατολικής Ευρώπης, όπως φαίνεται στην Εικόνα 1.3.1.

Αντίθετα, σε πιο αδύναμες οικονομικά χώρες, είναι πολύ σύνηθες να συνυπάρχουν η παχυσαρκία και ο υποσιτισμός ακόμα και στην ίδια περιοχή. Τα παιδιά σε τέτοιες χώρες είναι ευάλωτα καθώς δεν έχουν ούτε την επαρκή ούτε τη σωστή διατροφή που χρειάζονται σε αυτή την ηλικία, δηλαδή είναι εκτεθειμένα σε τροφές με υψηλά λιπαρά, υψηλή ζάχαρη και φτωχές σε θρεπτικά συστατικά αφού τα τρόφιμα που θα πρέπει να τρώνε είναι ή ακριβά ή δυσεύρετα. Μάλιστα, στις αναπτυσσόμενες χώρες, ο ΠΟΥ προβλέπει ότι το πρόβλημα του υποσιτισμού θα αντικατασταθεί με το πρόβλημα του υπερβολικού βάρους και της παχυσαρκίας.

Όπως βλέπουμε στην Εικόνα 1.3.1 μετά το 2000 υπάρχει αυξητική τάση στην ημερήσια κατανάλωση θερμίδων είτε σε δυνατές είτε σε αδύναμες οικονομικά χώρες. Δύο χαρακτηριστικά παραδείγματα είναι οι Ηνωμένες Πολιτείες της Αμερικής και η Κίνα, όπου το 1960 η κατανάλωση θερμίδων ανά άτομο ήταν 3100 και 1400kcal/ημέρα αντίστοιχα και το 2021 έχει φτάσει τις 3900 και 3400kcal/ημέρα αντίστοιχα. Ιδιαίτερα στην Κίνα, η έντονη

αστικοποίηση έχει παίξει σημαντικό ρόλο στη διατροφή των ανθρώπων. Από τη στιγμή που ο μέσος σωματικά ενεργός άνδρας για να διατηρήσει ένα υγιές βάρος χρειάζεται 2500 θερμίδες την ημέρα και αντίστοιχα 2000 η γυναίκα, καταλαβαίνουμε ότι το όριο αυτό έχει ξεπεραστεί κατά πολύ.

Οι προσπάθειες προώθησης της υγιεινής διατροφής δεν είναι αρκετά ισχυρές και αυτό φαίνεται από το γεγονός ότι δεν υπάρχει ο ίδιος αριθμός διαφημίσεων για υγιεινά τρόφιμα όσο για ανθυγιεινά. Σε έρευνα που έγινε στην τηλεόραση της Ταϊλάνδης (Bridget K. et al. 2024) για μια εβδομάδα το Μάιο του 2022 συνολικά ταυτοποιήθηκαν 659 διαφημίσεις τροφίμων, με μέσο όρο 9,2 διαφημίσεις ανθυγιεινών τροφίμων ανά ώρα. Οι πιο συχνές διαφημίσεις αφορούσαν ζαχαρούχα ποτά, τα οποία έχουν συνδεθεί στενά με την παχυσαρκία. Τα υγιεινά τρόφιμα έχουν γίνει πιο κοστοβόρα και, μάλιστα, σε πολλές χώρες είναι δυσεύρετα πλέον. Καθώς πλέον υπάρχουν πολλοί τρόποι διαφήμισης, η εξοικείωση του σύγχρονου ανθρώπου με έναν πιο υγιεινό τρόπο ζωής δεν θα έπρεπε να είναι κοστοβόρα.



Εικόνα 1.3.1: Συγκριτική ημερήσια κατανάλωση θερμίδων

1.3.2 Έλλειψη γυμναστικής

Η παχυσαρκία οφείλεται σε μεγάλο βαθμό στην έλλειψη φυσικής δραστηριότητας. Τα τελευταία χρόνια έχουμε παραμελήσει την εκγύμναση του σώματος, πράγμα που οφείλεται σε μεγάλο βαθμό στη ραγδαία ανάπτυξη της τεχνολογίας. Οι χειρωνακτικές εργασίες έχουν περιοριστεί πολύ λόγω των σύγχρονων μηχανημάτων, οπότε ακόμα και οι εργασίες που δεν είναι γραφείου απαιτούν λιγότερη σωματική άσκηση. Επιπλέον, τρόποι μετακίνησης όπως το ποδήλατο ή τα πόδια έχουν αντικατασταθεί καθώς το αυτοκίνητο έχει γίνει σχεδόν απαραίτητο στη ζωή του σύγχρονου ανθρώπου. Είναι μία πολύ μεγάλη διευκόλυνση και ως προς θέμα χρόνου, αλλά κυρίως ως προς θέμα κόυρασης, όμως περιορίζει σε μεγάλο βαθμό τη σωματική δραστηριότητα.

Έχουμε γίνει επιρρεπείς στο να κάνουμε καθιστική ζωή αφού η τεχνολογία μας προσφέρει τη δυνατότητα να ασχολούμαστε με τις δουλειές μας μέσω των ηλεκτρονικών υπολογιστών, αλλά και να διασκεδάσουμε μέσα στο σπίτι. Τα μικρά παιδιά έχουν αντικαταστήσει τον αθλητισμό και τη γυμναστική με τους ηλεκτρονικούς υπολογιστές και την τηλεόραση και είναι πολύ δύσκολο να κατανοήσουν πόσο βλαβερός είναι για την υγεία τους αυτός ο τρόπος ζωής.

Ιδιαίτερα όμως όσο μεγαλώνουμε, οι ψυχαγωγικές δραστηριότητες γίνονται ολοένα και λιγότερο σωματικά ενεργές.

1.3.3. Γενετικοί και κοινωνικοί παράγοντες

Οι γενετικοί παράγοντες επηρεάζουν την ανάπτυξη της παχυσαρκίας σε ποσοστό που κυμαίνεται από 40% έως 70%. Έρευνες σε οικογένειες έχουν δείξει ότι ο ΔΜΣ και άλλοι δείκτες όπως τα επίπεδα της λεπτίνης, σχετίζονται σημαντικά με τους συγγενείς πρώτου βαθμού. Είναι όμως δύσκολο να διευκρινιστεί σε ποιο βαθμό ευθύνονται οι γενετικοί ή οι περιβαλλοντικοί παράγοντες. Μελέτες που έγιναν σε μονοζυγωτικά δίδυμα, σύμφωνα με τον Sorensen, έδειξαν ότι οι διαφορές στο σωματικό βάρος οφείλονται σε περιβαλλοντικούς παράγοντες, ωστόσο υπάρχουν στοιχεία που δείχνουν ότι υπάρχει γενετική προδιάθεση για την εμφάνιση της παχυσαρκίας. Δηλαδή, εάν οι γονείς είναι παχύσαρκοι, υπάρχει μεγάλη πιθανότητα να γίνει το παιδί παχύσαρκος ενήλικας.

Η εμφάνιση παχυσαρκίας εξαρτάται από γονίδια τα οποία τα κληρονομούμαι από τους γονείς μας. Το πιο χαρακτηριστικό είναι το γονίδιο που σχετίζεται με την παραγωγή της λεπτίνης και ονομάζεται LEP. Αυτό το γονίδιο καθοδηγεί την παραγωγή της ορμόνης λεπτίνης, και τυχόν μεταλλάξεις στο γονίδιο LEP μπορούν να επηρεάσουν τη λειτουργία της ορμόνης, με αποτέλεσμα προβλήματα στη ρύθμιση της όρεξης και του βάρους. Η λεπτίνη είναι η ορμόνη που μειώνει την όρεξη οπότε τα παχύσαρκα άτομα παρουσιάζουν αντίσταση σε αυτήν με αποτέλεσμα να μην μπορούν να ελέγξουν την πείνα τους. Επίσης, ελλείψεις γονιδίων σε διάφορα χρωμοσώματα οδηγούν σε σύνδρομο που σχετίζονται με την παχυσαρκία, όπως το σύνδρομο Prader-Willi και το Bardet-Biedl τα οποία συνδέονται με υπερφαγία και κατά συνέπεια με παχυσαρκία. (Ζουρνατζή Β. et al, 2009 και Rehman A. et al. 2023)

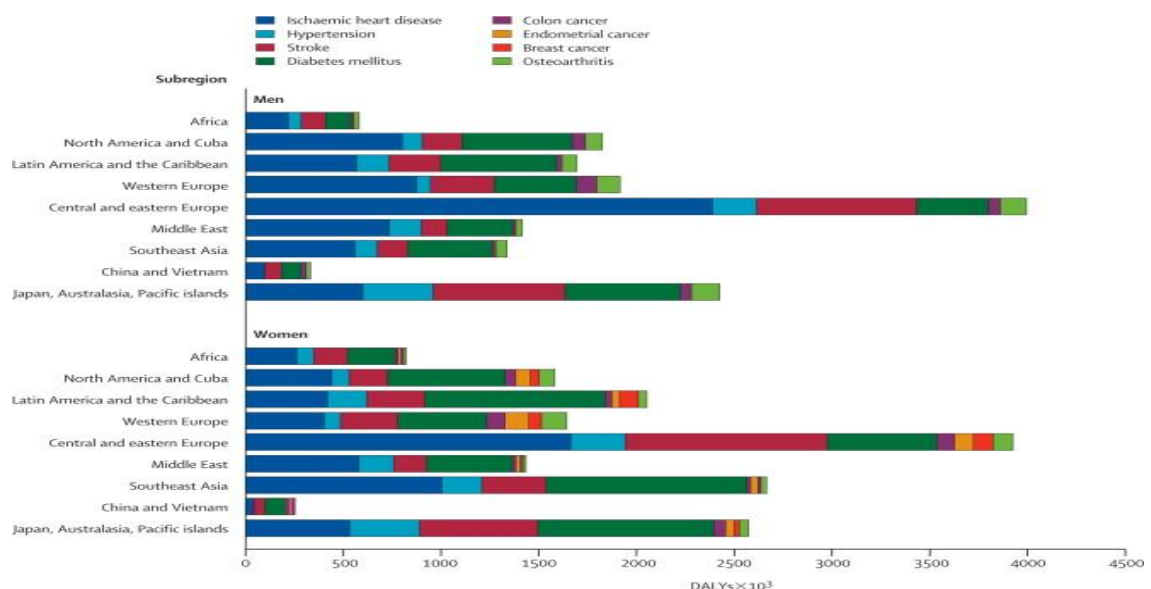
Η σπουδαιότητα των γενετικών παραγόντων και η επίδρασή τους στην ανάπτυξη της παχυσαρκίας είναι αδιαμφισβήτητη, όμως από μόνα τους δεν μπορούν να δικαιολογήσουν την απότομη αύξηση που παρατηρήθηκε παγκοσμίως. Κοινωνικοί και περιβαλλοντικοί παράγοντες όπως το υπερβολικό άγχος ή η έλλειψη ύπνου παίζουν επίσης σημαντικό ρόλο. Όταν είμαστε αγχωμένοι αυξάνονται οι ορμόνες που ελέγχουν την πείνα και είναι πιο πιθανό να τρώμε τροφές με υψηλή περιεκτικότητα σε λιπαρά και ζάχαρη. Ακόμα και φάρμακα που μπορεί να παίρνουμε για τη θεραπεία άλλων παθήσεων μπορεί να συμβάλλουν στην αύξηση βάρους, όπως τα αντικαταθλιπτικά, τα στεροειδή και τα φάρμακα για το διαβήτη.

1.4 Επιπτώσεις παχυσαρκίας

Η παχυσαρκία είναι μία νόσος η οποία σχετίζεται με διάφορες άλλες ασθένειες. Η πάθηση με την οποία συνδέεται περισσότερο είναι ο σακχαρώδης διαβήτης τύπου 2 καθώς το πλεονάζον σωματικό λίπος αποτελεί την κύρια αιτία εμφάνισης διαβήτη κατά 64% στους άνδρες και 77% στις γυναίκες. Αυτό οφείλεται στο ότι όταν αυξάνεται το σωματικό λίπος αλλοιώνει την αντίδραση του σώματος στην ινσουλίνη και συχνά υπάρχει αντίσταση σε αυτήν. Συχνή είναι η εμφάνιση καρδιακών παθήσεων οι οποίες μπορεί να παίρνουν μορφή υψηλής πίεσης ή υψηλές τιμές χοληστερόλης και τριγλυκεριδίων που μπορεί να οδηγήσουν σε στηθάγχη και έμφραγμα του μυοκαρδίου. Ο συνδυασμός παχυσαρκίας, διαβήτη και υψηλής πίεσης ονομάζεται μεταβολικό σύνδρομο. Άλλες ασθένειες που έχουν συνδεθεί με την παχυσαρκία είναι το άσθμα, η οστεοαρθρίτιδα, η υπνική άπνοια, η ηπατίτιδα ή και η κίρρωση

του ήπατος που προκαλούνται από υπερβολική συσσώρευση λίπους στο ήπαρ. Επίσης, υπάρχει αυξημένος κίνδυνος εμφάνισης ορισμένων μορφών καρκίνου όπως καρκίνος της μήτρας, του μαστού, του εντέρου ή του παγκρέατος. Στις γυναίκες μπορεί να προκαλέσει στειρότητα, υπογονιμότητα ή επιπλοκές στην εγκυμοσύνη τους. Τέλος, τα περισσότερα παχύσαρκα παιδιά διατηρούν το πρόβλημα της παχυσαρκίας τους και κατά την ενήλικη ζωή τους.

Στην Εικόνα 1.4.1 παρουσιάζονται λεπτομερείς εκτιμήσεις των ετών κακής υγείας καθώς και των ζών που χάθηκαν μεταξύ 30 και 75 ετών λόγω υπερβολικού βάρους, σε διάφορες περιοχές του κόσμου. Κυριαρχούν τα καρδιαγγειακά νοσήματα και ακολουθούν ο διαβήτης και ορισμένες μορφές καρκίνου, ιδιαίτερα στις γυναίκες. Η σημαντικότητα των ασθενειών είναι υψηλή στην Κεντρική και Ανατολική Ευρώπη καθώς και στις γυναίκες στη Λατινική Αμερική. Οι νοτιοανατολικές ασιατικές χώρες και η Ιαπωνία έχουν κι αυτές παραδόξως υψηλή επιβάρυνση παρόλο που έχουν χαμηλότερα ποσοστά παχυσαρκίας ενώ τα αντίστοιχα ποσοστά σε Κίνα και Βιετνάμ είναι τα χαμηλότερα παγκοσμίως. Αυτό το εύρημα σχετίζεται με τον υψηλότερο απόλυτο κίνδυνο διαβήτη και πιθανώς καρδιαγγειακής νόσου μεταξύ των Ασιατικών, Ισπανόφωνων και ίσως αφρικανικών πληθυσμών, εν μέρει επειδή είναι πιο επιρρεπείς στην κοιλιακή παχυσαρκία με τους υπερβολικούς κινδύνους της. (James W. & Haslam D., 2005)



Εικόνα 1.4.1: Επιπτώσεις παχυσαρκίας στην υγεία

Πέρα όμως από την κρισιμότητα των ασθενειών αυτών, το να ζει κάποιος με την παχυσαρκία ενέχει επιπλοκές και στην καθημερινότητα. Οι δραστηριότητες μπορεί να εμποδίζονται για παράδειγμα λόγω πόνων στις αρθρώσεις ή στην πλάτη, λόγω κόπωσης ή δύσπνοιας. Σημαντικό είναι επίσης και το ψυχολογικό κομμάτι της ασθένειας καθώς πολλές φορές κάποιος παχύσαρκος μπορεί να νιώθει ότι είναι κοινωνικά απομονωμένος ή να έχει χαμηλή αυτοπεποίθηση και αυτοεκτίμηση. Η κατάθλιψη και η εναλλαγή διαθέσεων είναι επίσης συχνά φαινόμενα.

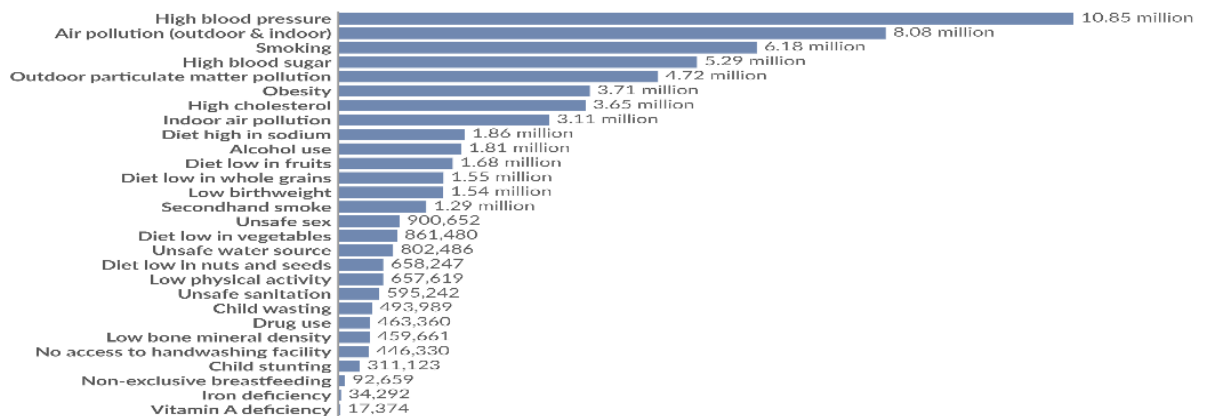
Η παχυσαρκία επηρεάζει και το προσδόκιμο ζωής. Στην Εικόνα 1.4.2 βλέπουμε την ετήσια εκτίμηση του αριθμού θανάτων που οφείλονται σε κάποιους παράγοντες κινδύνου. Η παχυσαρκία βρίσκεται στην έκτη θέση με εκτιμώμενο αριθμό 3,71 εκατομμύρια θανάτους που την καθιστά μία από τις πιο θανατηφόρες ασθένειες. Έρευνες έχουν δείξει ότι η παχυσαρκία

μειώνει το προσδόκιμο ζωής κατά 7 χρόνια στην ηλικία των 40. Η αύξηση του κινδύνου θανάτου με κάθε μονάδα αύξησης του ΔΜΣ μειώνεται προοδευτικά με την ηλικία, αλλά παραμένει σημαντική μέχρι την ηλικιακή ομάδα των 75 ετών και άνω. Αυτό που δεν έχει ακόμη επιβεβαιωθεί, ωστόσο, είναι εάν η σκόπιμη απώλεια βάρους σε παχύσαρκα άτομα παρατείνει τη ζωή καθώς και αν μειώνει τους κινδύνους. Εντούτοις, τα άτομα με πρόσφατα διαγνωσμένο διαβήτη που έχασαν 10 κιλά τον πρώτο χρόνο θεραπείας, βρέθηκε ότι κέρδισαν άλλα 4 χρόνια ζωής.

Deaths by risk factor, World, 2021

The estimated annual number of deaths attributed to each risk factor¹. Estimates come with wide uncertainties, especially for countries with poor vital registration².

Our World
in Data



Data source: IHME, Global Burden of Disease (2024)

OurWorldinData.org/causes-of-death | CC BY

Note: Risk factors¹ are not mutually exclusive. The sum of deaths attributed to each risk factor can exceed the total number of deaths.

Εικόνα 1.4.2: Εκτιμώμενος αριθμός θανάτων παγκοσμίως ανά παράγοντα κινδύνου το 2021

1.5 Αντιμετώπιση παχυσαρκίας

Η καλύτερη θεραπεία για την παχυσαρκία είναι η πρόληψη, που συνδέεται με την ανάπτυξη μιας ισορροπημένης σχέσης με το σώμα μας. Στην περίπτωση όμως που η πρόληψη αυτή δεν έχει επιτευχθεί και ο παχύσαρκος βρίσκεται αντιμέτωπος με τη νοσογόνο αυτή κατάσταση, η θεραπευτική προσέγγιση μπορεί να είναι είτε συντηρητική είτε χειρουργική.

Η πρώτη προσέγγιση θα ήταν καλύτερα να είναι η συντηρητική όπου το παχύσαρκο άτομο καλείται να αλλάξει τον τρόπο ζωής του ξεκινώντας με έναν συνδυασμό ειδικής διατροφής και γυμναστικής. Ακόμη, μπορεί να συμπεριλάβει κάποια φαρμακευτική αγωγή στο πρόγραμμά του όπως για παράδειγμα τα κατασταλτικά ορέξεως. Σε αυτή τη φάση χρειάζεται ψυχολογική υποστήριξη καθώς είναι πολλοί οι παχύσαρκοι που υποτροπιάζουν, παρά την τεράστια προσπάθεια που έχουν καταβάλει και σε μερικές περιπτώσεις παρά τη μεγάλη απώλεια βάρους που έχουν πετύχει. Η συντηρητική θεραπεία απαιτεί μεγάλο ψυχικό σθένος από το παχύσαρκο άτομο και χρειάζεται χρόνο.

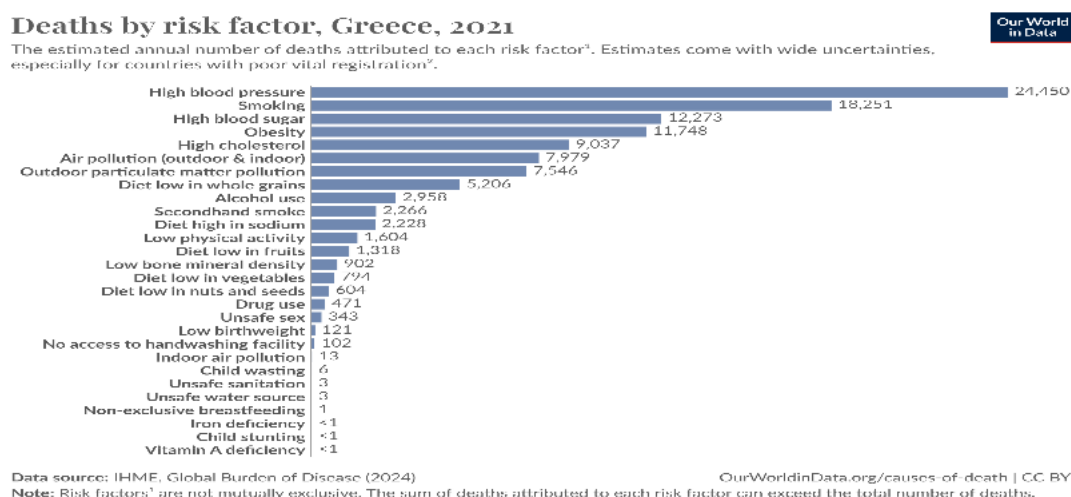
Αν η συντηρητική θεραπεία αποτύχει τότε η επόμενη μέθοδος είναι η χειρουργική θεραπεία η οποία ονομάζεται βαριατρική χειρουργική. Οι επεμβάσεις βαριατρικής χειρουργικής λειτουργούν αλλάζοντας το πεπτικό σύστημα ώστε να ρυθμιστούν οι θερμίδες που μπορούν να καταναλωθούν και να απορροφηθούν (Αθαναϊλίδου Ε). Η μέθοδος αυτή αν και είναι αποτελεσματική, δεν είναι χωρίς κινδύνους και οι ασθενείς μπορεί να αντιμετωπίσουν επιπλοκές τόσο κατά τη διάρκεια όσο και μετά την επέμβαση. Έτσι ασθενείς με ψυχολογικά ή

ψυχιατρικά προβλήματα ή ασθενείς με συνυπάρχουσες ασθένειες όπως κακοήθειες ή βαριά καρδιακά προβλήματα, δεν ενδείκνυται να υποβληθούν σε αυτή τη διαδικασία.

Στα παιδιά η αντιμετώπιση που ακολουθείται συνήθως είναι η παρέμβαση στον τρόπο ζωής και τεχνικές βελτίωσης των διατροφικών τους συνηθειών. Επίσης γίνονται απόπειρες αύξησης των δραστηριοτήτων των παχύσαρκων παιδιών αλλά χωρίς επιτυχία. Η φαρμακευτική αγωγή δεν συνίσταται και μάλιστα στις ΗΠΑ δεν έχει εγκριθεί από την εθνική Υπηρεσία Τροφίμων και Φαρμάκων.

1.6 Παχυσαρκία στην Ελλάδα

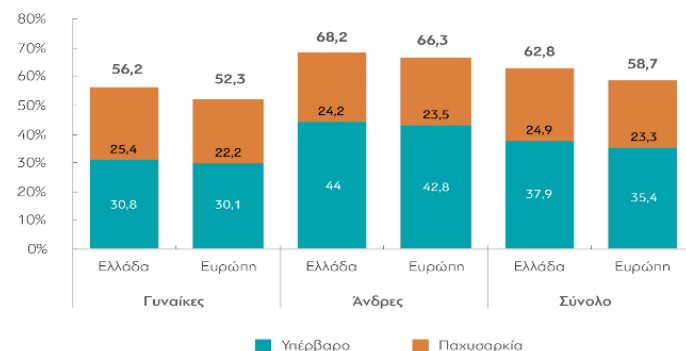
Η Ελλάδα είναι μία από τις χώρες της Ευρώπης που έχει επηρεαστεί πολύ αρνητικά από την παχυσαρκία. Βρίσκεται στις πρώτες θέσεις παχυσαρκίας παιδιών και ενηλίκων ανάμεσα στις ευρωπαϊκές χώρες και μάλιστα βρίσκεται στην πρώτη θέση στην παιδική παχυσαρκία στην Ευρωπαϊκή Ένωση. Όπως φαίνεται στην Εικόνα 1.6.1, η παχυσαρκία στην Ελλάδα είναι το τέταρτο βασικό αίτιο θανάτου με 11.748 θανάτους το έτος 2021.



Εικόνα 1.6.1: Εκτιμώμενος αριθμός θανάτων ανά παράγοντα κινδύνου το 2021 στην Ελλάδα

Ειδικότερα, σύμφωνα με δεδομένα του 2019 όπως φαίνονται στην Εικόνα 1.6.2 το ποσοστό των παχύσαρκων στην Ελλάδα εκτιμήθηκε σε 24,9% (άνδρες 24,2% – γυναίκες 25,4%), πάνω από τον μέσο ευρωπαϊκό όρο 23,3% (άνδρες 23,5% – γυναίκες 22,2%).

Γράφημα. Επιπολασμός υπέρβαρου/παχυσαρκίας σε ενήλικες στην Ελλάδα και την Ευρώπη, σύμφωνα με δεδομένα του ΠΟΥ



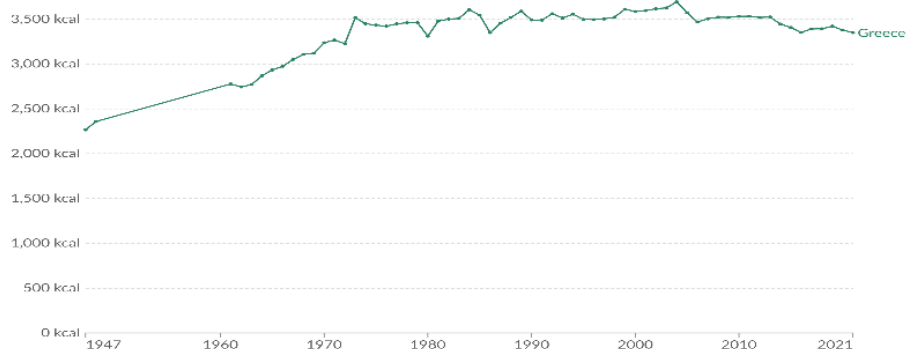
Πηγή: Προσαρμογή από WHO, 2019a.

Εικόνα 1.6.2: Ποσοστά παχυσαρκίας ανδρών-γυναικών σε Ελλάδα και Ευρώπη

Η μεσογειακή διατροφή, η οποία είναι εμπνευσμένη από τις διατροφικές συνήθειες της Ελλάδας και άλλων μεσογειακών χωρών, βοηθά στην πρόληψη αλλά και στην αντιμετώπιση της παχυσαρκίας. Τα τελευταία χρόνια όμως στην Ελλάδα την έχουμε παραμελήσει και έχουμε στραφεί στον πιο δυτικό τρόπο διατροφής, για παράδειγμα τρεις στους τέσσερις Έλληνες ενήλικες τρώνε λιγότερα φρούτα και λαχανικά από τις ενδεδειγμένες ποσότητες (Ανδρούτσος O. et al). Αυτό έχει ως αποτέλεσμα την πρόσληψη πολλών θερμίδων ανά ημέρα και καθώς το 68% των Ελλήνων ενηλίκων δεν γυμνάζονται καθόλου και δεν ασχολούνται με κανένα άθλημα, τότε οδηγούνται στην παχυσαρκία. Στην Εικόνα 1.6.3 βλέπουμε τον αριθμό θερμίδων που προσλαμβάνουν καθημερινά οι άνθρωποι στην Ελλάδα. Παρατηρούμε ότι μέχρι το 2004 υπάρχει μία αυξητική τάση που φτάνει σχεδόν τις 3700 θερμίδες την ημέρα ενώ από τότε μειώνεται σταδιακά και το 2021 φτάνουμε στις 3350 θερμίδες την ημέρα. Συγκεκριμένα την περίοδο 2009-2014, που είναι και η περίοδος της κρίσης στην Ελλάδα, βλέπουμε ότι υπάρχει μια σταθεροποίηση στις 3,5 χιλιάδες θερμίδες και αντίστοιχα εκείνη την περίοδο υπάρχει μια σταθεροποίηση του ποσοστού παχυσαρκίας τόσο στην Ελλάδα όσο και στην Ευρώπη.

Daily supply of calories per person, 1947 to 2021

Measured in kilocalories per person per day. This indicates the calories that are available for consumption, but does not necessarily measure the number of calories actually consumed, since it doesn't factor in consumer waste.



Data source: Food and Agriculture Organization of the United Nations (2023) and other sources
OurWorldinData.org/food_supply | CC BY

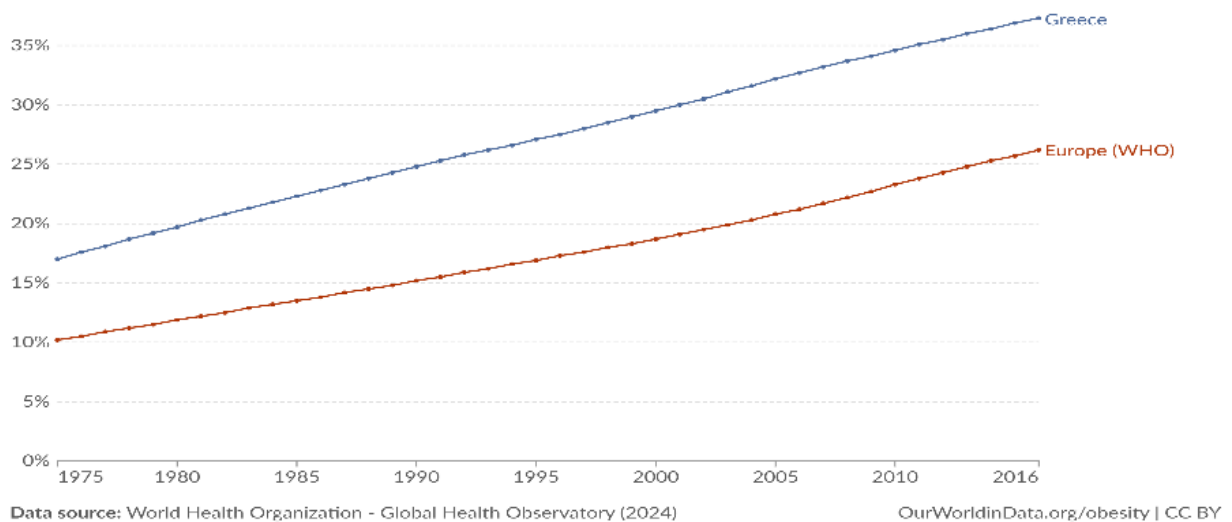
Εικόνα 1.6.3: Ημερήσια κατανάλωση θερμίδων στην Ελλάδα

Πέρα από τους ενήλικες, και τα παιδιά στην Ελλάδα εμφανίζουν τα μεγαλύτερα ποσοστά παχυσαρκίας στην Ευρώπη. Το ποσοστό των παιδιών ηλικίας 4-6 ετών που είναι παχύσαρκα ή υπέρβαρα είναι 20,6%. Στα παιδιά ηλικίας 6-10 ανεβαίνει στο 38,5% και στα παιδιά ηλικίας 10-12 ετών φτάνει το 41,2%. Στην Εικόνα 1.6.4, παρουσιάζεται η εξέλιξη του ποσοστού των παιδιών και εφήβων, ηλικίας 5-19 ετών, που είναι υπέρβαρα ή παχύσαρκα από το 1975 έως το 2016 στην Ελλάδα και στο μέσο όρο της Ευρώπης, σύμφωνα με τα δεδομένα του ΠΟΥ. Βλέπουμε ότι η Ελλάδα έχει υψηλότερα ποσοστά σε όλες τις χρονιές και υπάρχει μία σταθερή αύξηση του ποσοστού και στις δύο περιπτώσεις, όμως η διαφορά διευρύνεται με την πάροδο των ετών, με την Ελλάδα να παρουσιάζει πιο έντονη αύξηση.

Share of children and adolescents who are overweight or obese, 1975 to 2016

Our World in Data

Share of children and adolescents aged 5 to 19 years old that are defined as either overweight or obese. This means their weight-for-height is more than one standard deviations from the median of the World Health Organization (WHO) Child Growth Standards.



Εικόνα 1.6.4: Εξέλιξη παιδικής παχυσαρκίας στην Ελλάδα

Η παιδική και εφηβική παχυσαρκία αυξάνεται σταθερά τόσο στην Ελλάδα όσο και στην Ευρώπη. Η Ελλάδα καταγράφει ιδιαίτερα ανησυχητικά ποσοστά, που την καθιστούν έναν από τους "πρωταθλητές" παχυσαρκίας στην περιοχή. Όπως και οι ενήλικες, έτσι και τα παιδιά δεν ακολουθούν πλέον τη μεσογειακή διατροφή. Η μεγάλη κατανάλωση ενεργειακά πυκνών και πλούσιων σε λιπαρά τροφίμων σε συνδυασμό με τη μικρή κατανάλωση λαχανικών και φρούτων και τη χαμηλή σωματική δραστηριότητα οδηγούν πολλά παιδιά στην παχυσαρκία.

Τα ποσοστά υπέρβαρου και παχυσαρκίας στα παιδιά φαίνεται πως είναι μεγαλύτερα σε επαρχιακές και αγροτικές περιοχές από ό,τι στις πόλεις. Αυτό θα μπορούσε να οφείλεται είτε στην έλλειψη αθλητικών χώρων και αθλητικού εξοπλισμού, είτε στο εκπαιδευτικό επίπεδο των γονέων. Κάτι που βεβαίως δεν είναι μόνο ελληνικό χαρακτηριστικό είναι ότι το 88% των γονέων με παιδιά προσχολικής ηλικίας που είναι υπέρβαρα και το 55,8% των γονέων με παιδιά που είναι παχύσαρκα θεωρούν ότι το παιδί τους έχει φυσιολογικό σωματικό βάρος. Επίσης, έρευνα του 2010 έδειξε ότι τα παιδιά τα οποία φροντίζουν κυρίως οι γιαγιάδες και οι παππούδες έχουν 53% περισσότερες πιθανότητες να γίνουν υπέρβαρα ή παχύσαρκα σε σχέση με αυτά που τα φροντίζουν κυρίως οι γονείς. Άλλοι σημαντικοί παράγοντες φυσικά είναι το είδος των

τροφίμων που υπάρχουν διαθέσιμα στο σπίτι, η έκθεση σε διαφημιστικά μηνύματα από τα ΜΜΕ καθώς και οι επιρροές από το εκπαιδευτικό περιβάλλον και από τους συνομηλίκους.

Το πρόβλημα της παχυσαρκίας είναι ιδιαίτερα εκτεταμένο στην Ελλάδα σε όλες τις ηλικίες, επιβεβαιώνοντας τον χαρακτηρισμό της ως «πανδημία» από τον ΠΟΥ. Το ενθαρρυντικό όμως είναι ότι υπάρχουν πλέον πολλές δράσεις και προγράμματα ενημέρωσης κατά της νόσου που στοχεύουν την πρόληψη και την τελική αντιμετώπιση της παχυσαρκίας.

1.7 Μεθοδολογία

Στην παράγραφο αυτή θα παρουσιάσουμε τη μεθοδολογία που θα ακολουθήσουμε στα επόμενα κεφάλαια προκειμένου να γνωρίσουμε τις μεταβλητές μας αλλά και να εξετάσουμε τη σχέση τους με το επίπεδο παχυσαρκίας.

Για την καταχώρηση, την επεξεργασία και την ανάλυση των αποτελεσμάτων χρησιμοποιήθηκε το στατιστικό πακέτο Statistical Package for Social Sciences (SPSS) ver.22. Ως επίπεδο σημαντικότητας ορίστηκε το $\alpha=5\%$.

1.7.1 Παρουσίαση μεταβλητών

Αρχικά πραγματοποιήθηκε η περιγραφική ανάλυση όλων των μεταβλητών και τα αποτελέσματα παρουσιάστηκαν σε πίνακες. Πιο συγκεκριμένα, για τις κατηγορικές και τις διατεταγμένες μεταβλητές, απαριθμήσαμε πόσοι συμμετέχοντες βρίσκονται σε κάθε κατηγορία (απόλυτες συχνότητες) και υπολογίσαμε το ποσοστό τους επί του συνόλου. Επίσης, κάναμε γραφική απεικόνιση των αποτελεσμάτων, χρησιμοποιώντας ραβδογράμματα και διαγράμματα πίτας που δείχνουν την αριθμητική κατανομή των κατηγοριών. Ενώ οι συνεχείς μεταβλητές παρουσιάστηκαν με τη βοήθεια της μέσης τιμής $\pm$ την τυπική απόκλιση και τη διάμεσο τιμή. Για τη γραφική απεικόνιση των συνεχών μεταβλητών χρησιμοποιήθηκαν ιστογράμματα στα οποία προσαρμόσαμε την καμπύλη κανονικής κατανομής, ώστε να εξεταστεί οπτικά η μορφή της κατανομής τους.

1.7.2 Ο έλεγχος χ^2

Προκειμένου να εξεταστεί η σχέση μεταξύ της μεταβλητής «επίπεδο παχυσαρκίας» και των δημογραφικών στοιχείων, των διατροφικών συνηθειών και της φυσικής δραστηριότητας πραγματοποιήθηκε μονοπαραγοντική ανάλυση.

Ο έλεγχος γίνεται μέσω ενός $r \times c$ πίνακα συνάφειας και πραγματοποιήθηκε με το στατιστικό κριτήριο Pearson's χ^2 . Αυτό που μας ενδιαφέρει είναι να ελέγξουμε αν υπάρχει σχέση μεταξύ των δύο μεταβλητών ταξινόμησης σε γραμμές και στήλες ή εάν οι δύο μεταβλητές είναι ανεξάρτητες.

Για να ελέγξουμε αυτό συγκρίνουμε τις παρατηρούμενες συχνότητες Π κάθε κελιού με τις αναμενόμενες συχνότητες A . Πιο συγκεκριμένα, η αναμενόμενη τιμή κάθε κελιού 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:

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

Ο έλεγχος υποθέσεων που πραγματοποιήθηκε είναι ο ακόλουθος:

Μηδενική υπόθεση

H_0 : η εξεταζόμενη μεταβλητή και η μεταβλητή «επίπεδο παχυσαρκίας» είναι ανεξάρτητες μεταξύ τους

έναντι της εναλλακτικής υπόθεσης

H_1 : η εξεταζόμενη μεταβλητή και η μεταβλητή «επίπεδο παχυσαρκίας» δεν είναι ανεξάρτητες μεταξύ τους.

Η μηδενική υπόθεση του ελέγχου απορρίπτεται όταν $p\text{-value} < 0,05$ και δηλώνει την μη ύπαρξη σχέσης μεταξύ των εξεταζόμενων μεταβλητών.

Προκειμένου να μπορεί να εφαρμοστεί ο έλεγχος ανεξαρτησίας χ^2 και να είναι αξιόπιστος, θα πρέπει να ισχύουν οι παρακάτω προϋποθέσεις:

- i. Καμία από τις αναμενόμενες συχνότητες δεν πρέπει να είναι μικρότερη της μονάδας.
- ii. Το ποσοστό των αναμενόμενων συχνοτήτων οι οποίες είναι μικρότερες του 5 δεν πρέπει να υπερβαίνει το 20%.

Σε περίπτωση που δεν ίσχυε κάποια από τις προϋποθέσεις ορθής εφαρμογής του ελέγχου χ^2 χρησιμοποιήθηκε η προσέγγιση κατά Monte Carlo. (Μπερσίμης Σ, 2022)

1.7.3 Συντελεστής V του Cramer

Στις περιπτώσεις που απορρίπτουμε τη μηδενική υπόθεση του ελέγχου χ^2 , σημαίνει ότι υπάρχει σχέση μεταξύ της ανεξάρτητης μεταβλητής και του επιπέδου παχυσαρκίας. Για να εκτιμήσουμε το μέγεθος της συσχέτισης μεταξύ δύο κατηγορικών μεταβλητών -δηλαδή πόσο ισχυρή είναι η σχέση, θα χρησιμοποιήσουμε το συντελεστή V του Cramer. Ο συντελεστής αυτός χρησιμοποιείται για πίνακες μεγαλύτερους από 2x2 και η ερμηνεία των τιμών του είναι η εξής:

- 0,10–0,20 → μικρή συσχέτιση
- 0,20–0,30 → μέτρια συσχέτιση
- > 0,30 → ισχυρή συσχέτιση

1.7.4 Έλεγχος Κανονικότητας

Για να ελεγχθεί η συσχέτιση ή μη, μιας συνεχούς μεταβλητής με το επίπεδο παχυσαρκίας, πρέπει αρχικά να δούμε ποιον στατιστικό έλεγχο μπορούμε να χρησιμοποιήσουμε στα δεδομένα μας.

Θα ξεκινήσουμε την πορεία μας ελέγχοντας αν οι κατανομές των ποσοτικών μεταβλητών αποκλίνουν σημαντικά από την κανονική κατανομή. Για το λόγο αυτό, χρησιμοποιήθηκε ο έλεγχος Kolmogorov-Smirnov. Η μέθοδος συγκρίνει την αθροιστική κατανομή συχνοτήτων των παρατηρούμενων δεδομένων με την αντίστοιχη θεωρητική κατανομή και υπολογίζει τη μέγιστη απόλυτη διαφορά μεταξύ τους. Ο έλεγχος υποθέσεων που γίνεται είναι ο εξής:

$$\begin{aligned}H_0: & \text{Τα δεδομένα ακολουθούν την κανονική κατανομή.} \\H_1: & \text{Τα δεδομένα δεν ακολουθούν την κανονική κατανομή.}\end{aligned}$$

Σε περιπτώσεις όπου η τιμή p-value είναι μικρότερη από 0,05, απορρίπτεται η μηδενική υπόθεση της κανονικότητας και η μεταβλητή θεωρείται μη κανονικά κατανομημένη. Άρα για να ελέγξουμε αν υπάρχει σχέση ανάμεσα στην ανεξάρτητη και στην εξαρτημένη μεταβλητή, θα πρέπει να χρησιμοποιήσουμε μη παραμετρικούς ελέγχους όπως Kruskal-Wallis, Mann-Whitney U, Spearman κλπ.

Αν όμως η τιμή p-value δεν μας επιτρέπει να απορρίψουμε τη μηδενική υπόθεση της κανονικότητας, τότε θα πρέπει να ελέγξουμε αν τα δεδομένα μας είναι και ομοσκεδαστικά. Ο έλεγχος Levene είναι μια στατιστική μέθοδος που χρησιμοποιείται για να εκτιμήσει αν η διακύμανση μιας συνεχούς μεταβλητής είναι παρόμοια σε διαφορετικές ομάδες. Ελέγχει δηλαδή την ομοσκεδαστικότητα (ισότητα διακυμάνσεων). Η λογική του ελέγχου είναι ότι αν οι διακυμάνσεις είναι παρόμοιες, οι ομάδες μπορούν να συγκριθούν πιο αξιόπιστα, ενώ αν διαφέρουν σημαντικά, μπορεί να επηρεαστεί η εγκυρότητα των συμπερασμάτων. Οι υποθέσεις που ελέγχουμε είναι οι εξής:

$$\begin{aligned}H_0: & \text{Τα δεδομένα είναι ομοσκεδαστικά} \\H_1: & \text{Τα δεδομένα δεν είναι ομοσκεδαστικά}\end{aligned}$$

Σε περιπτώσεις όπου η τιμή p-value είναι μικρότερη από 0,05, απορρίπτεται η μηδενική υπόθεση, δηλαδή η μεταβλητή θεωρείται ότι δεν πληροί την παραδοχή της ομοσκεδαστικότητας. Τότε, πάλι, για να ελέγξουμε αν υπάρχει σχέση ανάμεσα στην ανεξάρτητη και στην εξαρτημένη μεταβλητή, θα πρέπει να χρησιμοποιήσουμε μη παραμετρικούς ελέγχους όπως Welch's ANOVA.

Τέλος, αν η μηδενική υπόθεση δεν απορριφθεί, δηλαδή αποδεχόμαστε ότι η ανεξάρτητη μεταβλητή πληροί την παραδοχή της ομοσκεδαστικότητας, τότε χρησιμοποιούμε παραμετρικούς ελέγχους όπως t-test ή ANOVA. Το πρώτο χρησιμοποιείται όταν θέλουμε να συγκρίνουμε μέσους όρους ανάμεσα σε δύο ομάδες, ενώ το δεύτερο χρησιμοποιείται όταν θέλουμε να συγκρίνουμε μέσους όρους μεταξύ περισσότερων από δύο ομάδων. Συγκεκριμένα, ελέγχει αν υπάρχει τουλάχιστον μία στατιστικά σημαντική διαφορά μεταξύ των μέσων όρων των ομάδων, αλλά δεν μας λέει ποια ομάδα διαφέρει. (Ευαγγελάρας Χ, 2022)

1.7.5 Ο έλεγχος Kruskal-Wallis

Για την σύγκριση μεταξύ της μεταβλητής «επίπεδο παχυσαρκίας» και των συνεχών μεταβλητών χρησιμοποιήθηκε ο μη-παραμετρικός έλεγχος Kruskal-Wallis. Ο λόγος που χρησιμοποιήθηκε ο μη-παραμετρικός έλεγχος είναι γιατί η υπόθεση της κανονικότητας της

εξεταζόμενης μεταβλητής σε κάθε επίπεδο της μεταβλητής «επίπεδο παχυσαρκίας» απορρίφθηκε σύμφωνα με τον έλεγχο Kolmogorov-Smirnov.

Ο έλεγχος Kruskal-Wallis είναι ένας μη παραμετρικός στατιστικός έλεγχος που χρησιμοποιείται για να συγκρίνει τις διαμέσους τριών ή περισσότερων ανεξάρτητων ομάδων, όταν η παραδοχή της κανονικότητας (ή/και της ομοσκεδαστικότητας) δεν πληρείται. Αποτελεί την μη παραμετρική εναλλακτική της μονοπαραγοντικής ANOVA. Βασίζεται στην κατάταξη των δεδομένων και όχι στις πραγματικές τιμές τους και είναι μια επέκταση του τεστ Mann-Whitney U για περισσότερες από δύο ομάδες.

Οι υποθέσεις που ελέγχονται με τη βοήθεια του ελέγχου Kruskal-Wallis είναι:

H_0 : Η κατανομή της ποσοτικής μεταβλητής είναι ίδια σε όλα τα επίπεδα της μεταβλητής «επίπεδο παχυσαρκίας»

H_1 : Η κατανομή της ποσοτικής μεταβλητής δεν είναι ίδια σε όλα τα επίπεδα της μεταβλητής «επίπεδο παχυσαρκίας»

Εάν το p-value του ελέγχου είναι μικρότερο του 0,05 τότε απορρίπτεται η μηδενική υπόθεση. Συνεπώς οι κατανομές ή οι διάμεσοι της μεταβλητής μας, διαφέρουν μεταξύ των επιπέδων της παχυσαρκίας. Αυτό πρακτικά σημαίνει ότι η εξεταζόμενη μεταβλητή και η μεταβλητή «επίπεδο παχυσαρκίας» δεν είναι ανεξάρτητες μεταξύ τους, δηλαδή υπάρχει συσχέτιση μεταξύ τους. Αντίθετα, εάν το p-value είναι μεγαλύτερο του 0,05 τότε δεν απορρίπτεται η μηδενική υπόθεση. Επομένως, η εξεταζόμενη μεταβλητή και η μεταβλητή «επίπεδο παχυσαρκίας» είναι ανεξάρτητες μεταξύ τους, δηλαδή δεν υπάρχει συσχέτιση μεταξύ τους.

1.7.6 Διωνυμική Λογιστική Παλινδρόμηση

Στα δεδομένα μας εφαρμόστηκε Διωνυμική Λογιστική Παλινδρόμηση και παρουσιάστηκε η εκτίμηση των συντελεστών κάθε μοντέλου και ο σχετικός λόγος για κάθε επίπεδο της εξαρτημένης μεταβλητής.

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). Για αυτήν την δίτιμη μεταβλητή εξετάζουμε την πιθανότητα (τα ποσοστά) εμφάνισης των δύο κατηγοριών σε σχέση με τις ανεξάρτητες μεταβλητές - παράγοντες. Επειδή σκοπός είναι να εκτιμηθεί η πιθανότητα εμφάνισης ενός συμβάντος, συνεπάγεται ότι οι τιμές που θα πρέπει να προκύπτουν από το γραμμικό υπόδειγμα περιέχονται στο διάστημα [0,1]. Για τον λόγο αυτό υποθέτουμε ότι η μεταβλητή ακολουθεί διωνυμική κατανομή και ότι η σύνδεση της πιθανότητας εμφάνισης του γεγονότος p_i συνδέεται με το γραμμικό υπόδειγμα μέσω της link function:

$$\log\left(\frac{p_i}{1-p_i}\right) = b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_kx_{ki} + \varepsilon$$

Όπου: $\log\left(\frac{p_i}{1-p_i}\right)$ είναι ο λογάριθμος του λόγου σχετικής πιθανότητας.

Από τα παραπάνω προκύπτει ότι η πιθανότητα της κατηγορίας της εξαρτημένης μεταβλητής θα είναι:

$$p = \frac{e^{\hat{b}_0 + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} + \dots + \hat{b}_k x_{ki}}}{1 + e^{\hat{b}_0 + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} + \dots + \hat{b}_k x_{ki}}}$$

Στο μοντέλο της λογιστικής παλινδρόμησης η εκτίμηση των συντελεστών πραγματοποιείται με τη μέθοδο μέγιστης πιθανοφάνειας (Maximum likelihood Method) αντί της μεθόδου ελαχίστων τετραγώνων.

Όπως και στη γραμμική παλινδρόμηση, ομοίως οι συντελεστές του μοντέλου, οι οποίοι αποτελούν σημειακές εκτιμήσεις, υπόκεινται σε έλεγχο για τη σημαντικότητά τους στο μοντέλο. Δηλαδή η υπόθεση που εξετάζεται είναι η:

$$H_0: b = 0$$

$$H_1: b \neq 0$$

Ο έλεγχος γίνεται σε $(1-\alpha)\%$ επίπεδο σημαντικότητας και η στατιστική συνάρτηση υπό την H_0 είναι η:

$$\frac{\hat{b}_i}{s.e(\hat{b}_i)}$$

η οποία καλείται Wald, ακολουθεί ασυμπτωτικά την κανονική κατανομή, ενώ το τετράγωνο αυτής ακολουθεί την X^2 κατανομή με 1 βαθμό ελευθερίας. Η υπόθεση απορρίπτεται αν η τιμή του ελέγχου είναι μεγαλύτερη από $X^2_{1;\alpha/2}$.

Για κάθε τιμή των ανεξάρτητων μεταβλητών του κάθε συμμετέχοντα στη μελέτη, υπολογίζονται με την βοήθεια των παραπάνω, πιθανότητες οι οποίες εκφράζουν την πιθανότητα το άτομο με τα συγκεκριμένα στοιχεία να ανήκει σε κάθε μία από τις δύο κατηγορίες της εξαρτημένης μεταβλητής, αντίστοιχα. Το άτομο, με τη βοήθεια της Διωνυμικής Λογιστικής Παλινδρόμησης κατατάσσεται στην κατηγορία της εξαρτημένης μεταβλητής για την οποία η πιθανότητα είναι υψηλότερη.

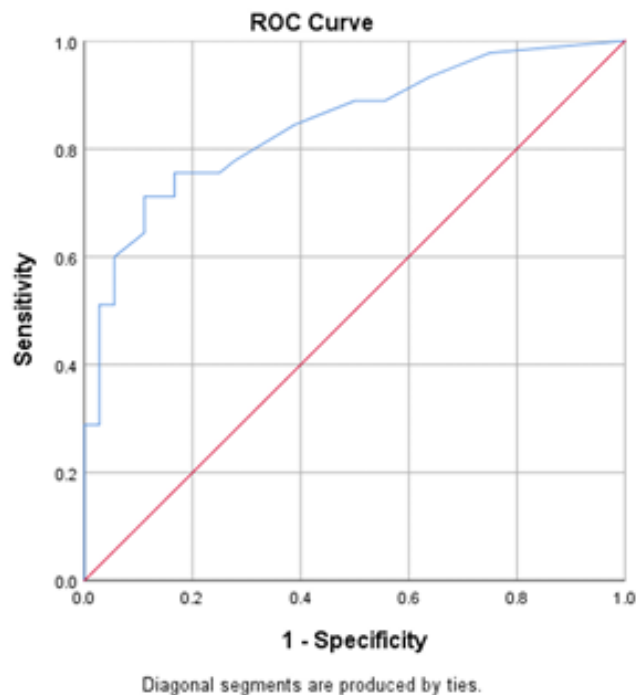
Τέλος, να σημειωθεί ότι στη λογιστική παλινδρόμηση για δίτιμες εξαρτημένες μεταβλητές, ως ανεξάρτητες μεταβλητές μπορούν να χρησιμοποιηθούν είτε ποσοτικά είτε ποιοτικά χαρακτηριστικά.

1.7.7 Ανάλυση ROC

Η ανάλυση ROC (Receiver Operating Characteristic curve) είναι μια μέθοδος αξιολόγησης της απόδοσης ενός μοντέλου ταξινόμησης. Συνήθως, χρησιμοποιείται για μοντέλα που προβλέπουν δυαδικές κατηγορίες, όπως "Θετικό" και "Αρνητικό" σε προβλήματα δυαδικής ταξινόμησης. Επίσης, χρησιμοποιείται προκειμένου να προκύψει μια τιμή αποκοπής που, κατά κάποιο τρόπο θα ελαχιστοποιεί τον αριθμό των ψευδών θετικών και ψευδών αρνητικών αποτελεσμάτων.

Η καμπύλη ROC είναι μία γραφική απεικόνιση της ευαισθησίας έναντι της ποσότητας (1-ειδικότητα), όπου ειδικότητα είναι η αναλογία των αρνητικών περιπτώσεων που διαπίστωσε ο έλεγχος ως προς τον συνολικό αριθμό των πραγματικά υγιών ατόμων. Στον κατακόρυφο άξονα αναπαρίσταται η ευαισθησία, η πιθανότητα δηλαδή ο έλεγχος να δώσει θετικό αποτέλεσμα δοθέντος ότι το άτομο πάσχει από την ασθένεια, ενώ στον οριζόντιο άξονα αναπαρίσταται η

ποσότητα, δηλαδή η πιθανότητα ο έλεγχος να δώσει θετικό αποτέλεσμα δοθέντος ότι το άτομο δεν πάσχει από την ασθένεια. Ένα παράδειγμα καμπύλης ROC δίνεται παρακάτω:



Σχήμα 1.7.7: Παράδειγμα καμπύλης ROC

Γενικότερα μια καλή πρώτη επιλογή για την τιμή αποκοπής είναι αυτή που αντιστοιχεί σε ένα σημείο της καμπύλης ROC που βρίσκεται κοντά στην επάνω αριστερή γωνία του διαγράμματος ROC, δηλαδή το σημείο όπου η ευαισθησία είναι μέγιστη και ταυτόχρονα το ποσοστό των ψευδώς θετικών (1–ειδικότητα) είναι ελάχιστο. Σε διαγνωστικούς ελέγχους που έχουν να κάνουν με την υγεία είναι αρκετά συχνό να δίνεται μεγαλύτερη βαρύτητα στην μεγιστοποίηση της ευαισθησίας από την μεγιστοποίηση της ειδικότητας. Για την εύρεση του βέλτιστου σημείου αποκοπής c μεγιστοποιούμε την συνάρτηση:

$$h(c) = \text{ευαισθησία} + \text{ειδικότητα} - 1$$

η οποία είναι γνωστή ως Δείκτης Youden.

Τέλος, για τη σύγκριση διάφορων διαγνωστικών ελέγχων και την επιλογή τιμών αποκοπής, χρησιμοποιείται το εμβαδόν κάτω από την καμπύλη (AUC) που παίρνει τιμές από 0 έως 1.

1.7.8 Πολυωνυμική Λογιστική Παλινδρόμηση

Στα δεδομένα όπου η εξαρτημένη μεταβλητή έχει περισσότερες από δύο κατηγορίες- στην περίπτωση μας τέσσερεις- εφαρμόζεται η Πολυωνυμική Λογιστική Παλινδρόμηση. Η μέθοδος αυτή αποτελεί γενίκευση της Διωνυμικής Λογιστικής Παλινδρόμησης, αλλά σε αντίθεση με τη Διωνυμική, όπου προβλέπεται η πιθανότητα εμφάνισης ή μη ενός χαρακτηριστικού, στην Πολυωνυμική περίπτωση το ενδιαφέρον εστιάζει στην πρόβλεψη της κατηγορίας στην οποία ανήκει ένα άτομο ανάμεσα σε πολλές δυνατές επιλογές.

Η βασική ιδέα είναι ότι επιλέγεται μία από τις κατηγορίες της εξαρτημένης μεταβλητής ως κατηγορία αναφοράς (reference category), και στη συνέχεια υπολογίζονται λογάριθμοι λόγων

πιθανοτήτων (log-odds) για κάθε άλλη κατηγορία που συγκρίνονται με την κατηγορία αυτή. Η μαθηματική μορφή του μοντέλου είναι:

$$\log\left(\frac{P(Y=j)}{P(Y=ref)}\right) = b_{0j} + b_{1j}x_1 + b_{2j}x_2 + \dots + b_{kj}x_k, \quad j = 1, 2, \dots, m-1$$

όπου $P(Y=j)$ είναι η πιθανότητα να ανήκει η παρατήρηση στην κατηγορία j , $P(Y=ref)$ είναι η πιθανότητα της κατηγορίας αναφοράς, ενώ οι συντελεστές b_{ij} εκφράζουν την επίδραση της ανεξάρτητης μεταβλητής x_i στην πιθανότητα ένταξης στην κατηγορία j σε σχέση με την κατηγορία αναφοράς. Δηλαδή, οι συντελεστές που προκύπτουν δείχνουν αν η παρουσία ή η αύξηση μιας ανεξάρτητης μεταβλητής αυξάνει ή μειώνει την πιθανότητα ένταξης σε μια κατηγορία σε σχέση με την κατηγορία αναφοράς.

Η εκτίμηση των παραμέτρων γίνεται, όπως και στη διωνυμική περίπτωση, με τη μέθοδο της μέγιστης πιθανοφάνειας (Maximum likelihood Method). Η στατιστική σημαντικότητα των συντελεστών ελέγχεται με τον έλεγχο Wald, ενώ η συνολική καταλληλότητα του μοντέλου εξετάζεται μέσω κριτηρίων όπως το -2 Log Likelihood, το AIC και ο ψευδοσυντελεστής προσδιορισμού Nagelkerke R^2 .

Η Πολωνυμική Λογιστική Παλινδρόμηση είναι ιδιαίτερα χρήσιμη όταν η εξαρτημένη μεταβλητή έχει διακριτές και μη διατεταγμένες κατηγορίες (π.χ. τύποι παχυσαρκίας), καθώς επιτρέπει την εκτίμηση του τρόπου με τον οποίο οι ανεξάρτητες μεταβλητές επηρεάζουν την πιθανότητα ένταξης σε κάθε κατηγορία. (Ηλιόπουλος Γ, 2023)

1.7.9 Το κριτήριο AIC

Το Κριτήριο Πληροφορίας του Akaike (AIC), είναι ένα στατιστικό κριτήριο που χρησιμοποιείται για την επιλογή του καλύτερου μοντέλου από μια ομάδα υποψήφιων μοντέλων. Βασίζεται στην ιδέα ότι ένα καλό μοντέλο πρέπει να εξηγεί τα δεδομένα όσο το δυνατόν καλύτερα, αλλά ταυτόχρονα να είναι όσο το δυνατόν απλούστερο (δηλαδή λιγότερες παραμέτρους).

Το AIC υπολογίζει μια τιμή για κάθε μοντέλο, η οποία λαμβάνει υπόψη τόσο την προσαρμογή του μοντέλου στα δεδομένα (μέσω της συνάρτησης πιθανοφάνειας) όσο και τον αριθμό των παραμέτρων του μοντέλου. Ο τύπος για το AIC είναι:

$$AIC = -2\log L + 2k$$

όπου

- $-2\log L$ είναι το deviance (-2 Log Likelihood) και
- k είναι ο αριθμός των παραμέτρων του μοντέλου

Το μοντέλο με το χαμηλότερο AIC θεωρείται το καλύτερο, καθώς αντιπροσωπεύει την καλύτερη ισορροπία μεταξύ προσαρμογής και πολυπλοκότητας.

1.7.10 Δείκτης Nagelkerke R^2

Ο δείκτης Nagelkerke R^2 (ή ψευδο- R^2) αποτελεί μέτρο της προσαρμογής ενός λογιστικού μοντέλου. Σε αντίθεση με το κλασικό R^2 της γραμμικής παλινδρόμησης, που δείχνει το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής που εξηγείται από το μοντέλο, το

Nagelkerke R^2 είναι μια προσαρμοσμένη εκδοχή του Cox & Snell R^2 που παίρνει τιμές στο εύρος $[0,1]$. Έτσι, διευκολύνει την ερμηνεία και επιτρέπει τη σύγκριση διαφορετικών μοντέλων λογιστικής παλινδρόμησης.

Όσο υψηλότερη είναι η τιμή του Nagelkerke R^2 , τόσο μεγαλύτερη είναι η ικανότητα του μοντέλου να εξηγεί την κατανομή της εξαρτημένης μεταβλητής με βάση τις ανεξάρτητες μεταβλητές. Παρότι δεν ερμηνεύεται με τον ίδιο τρόπο όπως το R^2 της γραμμικής παλινδρόμησης, χρησιμοποιείται ευρέως ως ένδειξη του πόσο καλά «ταιριάζει» το μοντέλο στα δεδομένα.

1.7.11 Λόγος σχετικών πιθανοτήτων (Odds Ratio – OR)

Ο λόγος σχετικών πιθανοτήτων (OR) είναι η σχετική πιθανότητα της ομάδας που έχει εκτεθεί σε κάποιον παράγοντα κινδύνου διαιρούμενη με την σχετική πιθανότητα της ομάδας που δεν έχει εκτεθεί.

Ο λόγος σχετικών πιθανοτήτων αποτελεί βασικό δείκτη ερμηνείας στη λογιστική παλινδρόμηση και εκφράζει πόσες φορές αυξάνονται ή μειώνονται οι πιθανότητες εμφάνισης ενός γεγονότος για μια κατηγορία της ανεξάρτητης μεταβλητής σε σύγκριση με την κατηγορία αναφοράς. Συγκεκριμένα, τιμή OR μεγαλύτερη του 1 υποδηλώνει ότι η πιθανότητα εμφάνισης του γεγονότος αυξάνεται, ενώ τιμή μικρότερη του 1 υποδηλώνει μείωση της πιθανότητας. Αν το OR είναι ίσο με 1, τότε η συγκεκριμένη μεταβλητή δεν ασκεί καμία επίδραση στην πιθανότητα εμφάνισης του γεγονότος. Ο δείκτης αυτός επιτρέπει μια πιο διαισθητική κατανόηση της επίδρασης των ανεξάρτητων μεταβλητών στο υπό εξέταση φαινόμενο.

Κεφάλαιο 2

Περιγραφική Ανάλυση

2.1 Περιγραφική Ανάλυση

Στο κεφάλαιο αυτό θα παρουσιάσουμε μια περιγραφή του δείγματος που θα χρησιμοποιήσουμε στην παρούσα εργασία. Τα δεδομένα προέρχονται από μία διαδικτυακή έρευνα για την εκτίμηση των επιπέδων παχυσαρκίας σε άτομα από το Μεξικό, το Περού και την Κολομβία, με ηλικίες μεταξύ 14 και 61 ετών. Η έρευνα ήταν ανώνυμη και οι χρήστες έπρεπε να απαντήσουν σε ερωτήσεις ώστε να αξιολογήσουν τις διατροφικές τους συνήθειες καθώς και τη φυσική τους κατάσταση. Ο αριθμός των εγγραφών που συμμετείχαν στην έρευνα ήταν 485 και οι πληροφορίες υποβλήθηκαν σε επεξεργασία λαμβάνοντας 17 χαρακτηριστικά.

Όλα τα δεδομένα κατηγοριοποιήθηκαν και έτσι δημιουργήθηκε η κατηγορική μεταβλητή NObeyesdad με τιμές: Ανεπαρκές βάρος, Κανονικό βάρος, Υπέρβαρο Επίπεδο I, Υπέρβαρο Επίπεδο II, Παχυσαρκία Τύπου I, Παχυσαρκία Τύπου II και Παχυσαρκία Τύπου III.

Κύριος στόχος της παρούσας μελέτης είναι να ερευνήσουμε ποιες μεταβλητές από αυτές επηρεάζουν περισσότερο την εμφάνιση παχυσαρκίας και ποιες λιγότερο.

2.2 Δημογραφικά χαρακτηριστικά

Τα χαρακτηριστικά που σχετίζονται με τα δημογραφικά στοιχεία είναι: το φύλο (Gender), η ηλικία (Age), το ύψος (Height), το βάρος (Weight) καθώς και το οικογενειακό ιστορικό παχυσαρκίας (family_history_with_overweight).

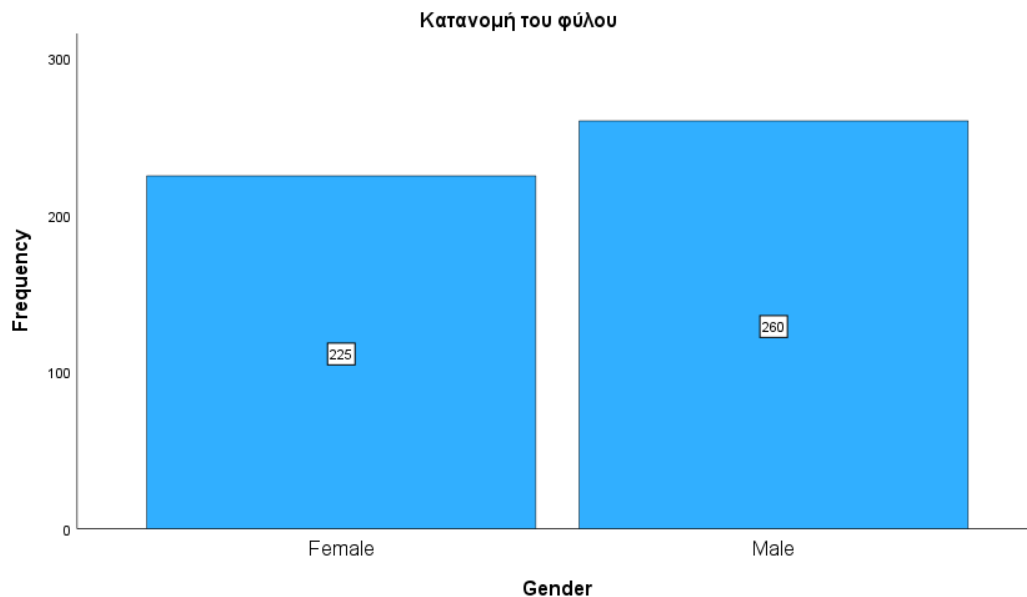
Αυτά τα δημογραφικά χαρακτηριστικά τα μελετάμε διότι τα τελευταία χρόνια υπάρχει μια αυξανόμενη συνειδητοποίηση της ανάγκης να κατανοήσουμε τις διαφορές των φύλων στην παχυσαρκία (Barrea L. et al. 2023). Επίσης, είναι κρίσιμο να μελετήσουμε το οικογενειακό ιστορικό και την ηλικία καθώς το πρώτο μπορεί να υποδεικνύει γενετική προδιάθεση ενώ το δεύτερο μπορεί να επηρεάζει το μεταβολικό ρυθμό και τη συσσώρευση λίπους στο σώμα. Τέλος, το βάρος και το ύψος είναι στενά συνδεδεμένα με την παχυσαρκία, αφού βάσει αυτών υπολογίζουμε το Δείκτη Μάζας Σώματος.

2.2.1 Περιγραφική ανάλυση μεταβλητής «Φύλο»

Είναι σημαντικό να γνωρίζουμε την κατανομή των φύλων μέσα στο δείγμα. Η μεταβλητή φύλο είναι διακριτή και παίρνει τιμές: Άνδρας ή Γυναίκα. Στον Πίνακα 2.2.1 που ακολουθεί καταλαβαίνουμε ότι δεν υπάρχει μεροληψία αφού από τους 485 συμμετέχοντες οι 260 ήταν άνδρες και οι 225 ήταν γυναίκες και αντίστοιχα σε ποσοστά ήταν 53,6% άνδρες και 46,4% γυναίκες.

Gender				
	Frequency	Percent	Valid Percent	Cumulative Percent
Female	225	46,4	46,4	46,4
Male	260	53,6	53,6	100,0
Total	485	100,0	100,0	

Πίνακας 2.2.1: Στατιστικά στοιχεία για το φύλο



Διάγραμμα 2.2.1: Κατανομή φύλου

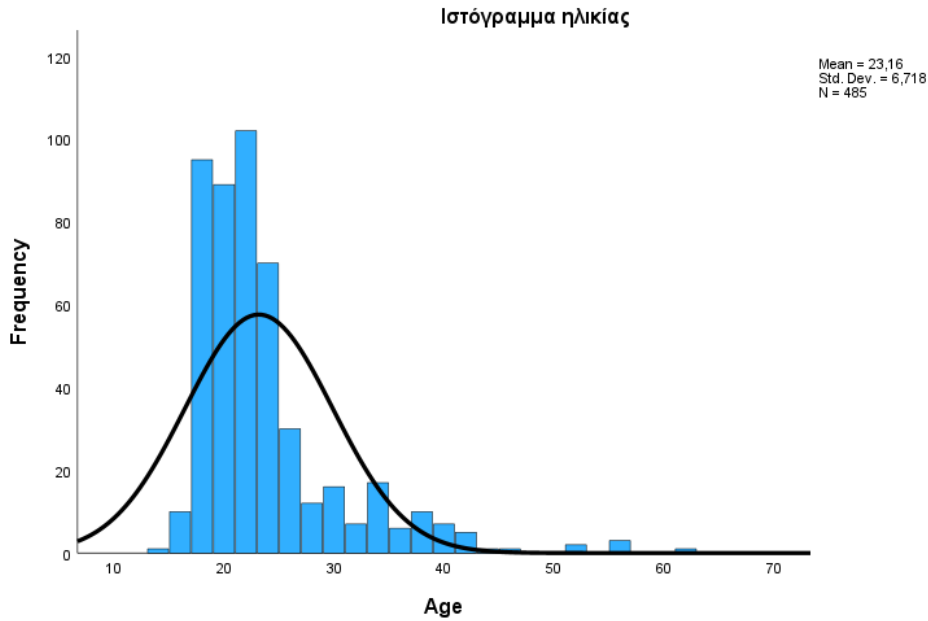
2.2.2 Περιγραφική ανάλυση μεταβλητής «Ηλικία»

Στη συγκεκριμένη μελέτη, η ηλικία είναι μια συνεχής μεταβλητή που οι τιμές της κυμαίνονται μεταξύ 14 και 61 ετών με διάμεσο τα 21 έτη και μέσο όρο ηλικίας τα $23,16 \pm 6,72$ έτη όπου 6,72 περίπου είναι η τυπική απόκλιση.

Statistics		
Age		
Mean		23,16
Median		21,00
Std. Deviation		6,718
Minimum		14
Maximum		61
Percentiles	25	19,00
	50	21,00
	75	24,00

Πίνακας 2.2.2: Στατιστικά στοιχεία για την ηλικία

Στο Διάγραμμα 2.2.2 βλέπουμε ότι η ηλικία δεν ακολουθεί κανονική κατανομή καθώς υπάρχει μια εμφανώς δεξιά ασυμμετρία στο ιστόγραμμα, η οποία υποδηλώνει ότι υπάρχει συσσώρευση στις ηλικίες 18-25 και ότι υπάρχουν λιγότερα άτομα σε μεγαλύτερες ηλικίες.



Διάγραμμα 2.2.2: Ιστόγραμμα για την κατανομή των ηλικιών

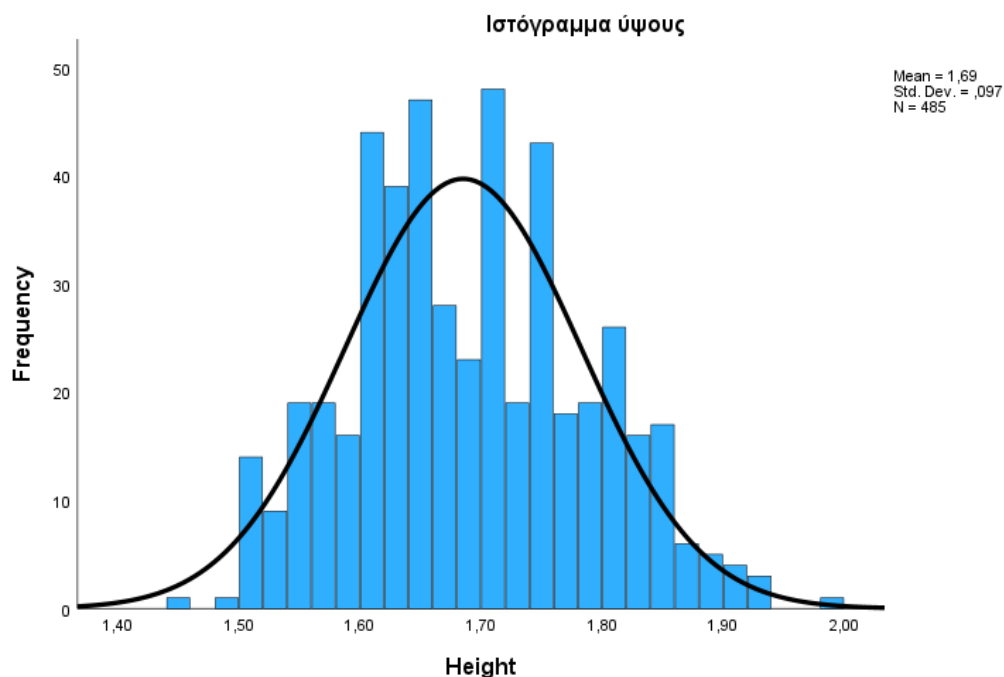
2.2.3 Περιγραφική ανάλυση μεταβλητής «Ύψος»

Το ύψος στη συγκεκριμένη έρευνα είναι μια συνεχής μεταβλητή που έχει υπολογιστεί σε μέτρα(m). Στον Πίνακα 2.2.3 βλέπουμε ότι το ύψος κυμαίνεται από 1,45 m έως 1,98 m. Η διάμεσος και η μέση τιμή είναι περίπου 1,68 m με τυπική απόκλιση περίπου ίση με 0,1.

Statistics		
Height		
Mean		1,6852
Median		1,6800
Std. Deviation		,09745
Minimum		1,45
Maximum		1,98
Percentiles	25	1,6100
	50	1,6800
	75	1,7500

Πίνακας 2.2.3: Στατιστικά στοιχεία για το ύψος

Στο Διάγραμμα 2.2.3 βλέπουμε το ιστόγραμμα του ύψους όπου τα περισσότερα δεδομένα συγκεντρώνονται κοντά στον μέσο όρο, με μικρές αποκλίσεις. Στο ιστόγραμμα φαίνεται η κατανομή να είναι πολύ κοντά στην κανονική, με μία ελαφριά ασυμμετρία προς τα αριστερά, πράγμα που σημαίνει ότι υπάρχουν λίγες τιμές με σχετικά μικρό ύψος. Επίσης, η κατανομή φαίνεται να έχει περισσότερες από μία κορυφές, κάτι που μπορεί να υποδηλώνει την ύπαρξη διαφορετικών υποομάδων μέσα στα δεδομένα και οι οποίες μπορεί να οφείλονται σε χαρακτηριστικά όπως το φύλο ή η ηλικία.



Διάγραμμα 2.2.3: Ιστόγραμμα κατανομής ύψους

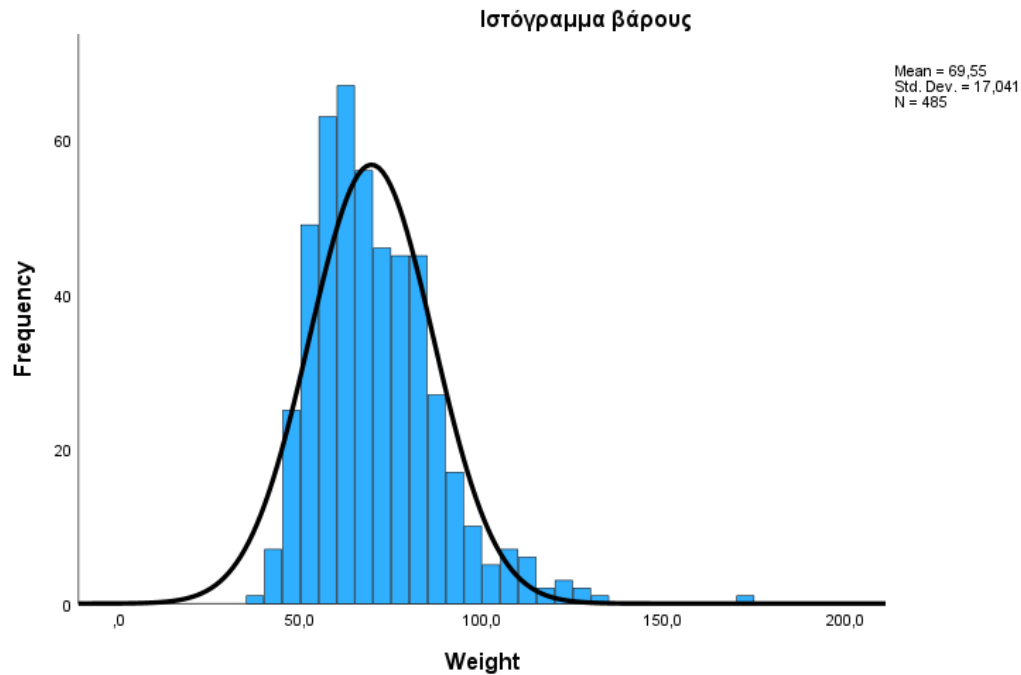
2.2.4 Περιγραφική ανάλυση μεταβλητής «Βάρος»

Το βάρος για την έρευνά μας είναι μια συνεχής μεταβλητή που έχει μετρηθεί σε χιλιόγραμμα(kg). Στον Πίνακα 2.2.4 βλέπουμε ότι το ελάχιστο βάρος είναι 39 κιλά ενώ το μέγιστο είναι 173. Η διάμεσος είναι 67 κιλά ενώ η μέση τιμή 69,5 με τυπική απόκλιση 17 κιλά.

Statistics		
Weight		
Mean		69,552
Median		67,000
Std. Deviation		17,0412
Minimum		39,0
Maximum		173,0
Percentiles	25	57,500
	50	67,000
	75	80,000

Πίνακας 2.2.4: Στατιστικά στοιχεία για το βάρος

Στο Διάγραμμα 2.2.4 βλέπουμε το ιστόγραμμα του βάρους. Φαίνεται ότι υπάρχει δεξιά ασυμμετρία. Η μακριά ουρά στα δεξιά δείχνει ότι υπάρχουν κάποια άτομα με αρκετά μεγαλύτερο βάρος από τον μέσο όρο. Οι περισσότερες παρατηρήσεις βρίσκονται περίπου στο διάστημα 50–75 κιλά, που είναι και κοντά στον μέσο όρο. Ο μέσος όρος επηρεάζεται από κάποιες πολύ μεγάλες τιμές (> 120 κιλά). Αν και το ιστόγραμμα οπτικά πλησιάζει την κανονική κατανομή, η δεξιά ασυμμετρία δείχνει ότι δεν είναι τέλεια κανονική.



Διάγραμμα 2.2.4: Ιστόγραμμα κατανομής βάρους

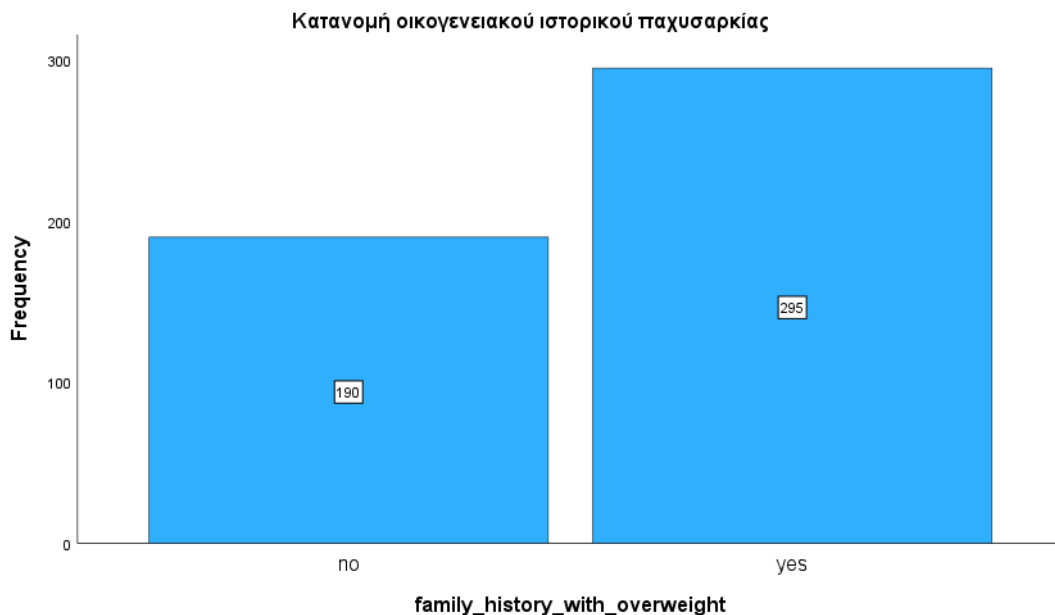
2.2.5 Περιγραφική ανάλυση μεταβλητής «Οικογενειακό ιστορικό παχυσαρκίας»

Η κατανομή των δεδομένων με βάση το οικογενειακό ιστορικό παχυσαρκίας φαίνεται στον Πίνακα 2.2.5 και παρατηρούμε ότι στα δεδομένα μας υπάρχει μια σχετική ισορροπία. Η μεταβλητή αυτή είναι δυαδική και παίρνει τιμές Ναι ή Όχι. Σχεδόν το 61% των δεδομένων μας είχε οικογενειακό ιστορικό παχυσαρκίας ενώ το 39% όχι. Σε αριθμούς, οι 295 είχαν οικογενειακό ιστορικό ενώ οι 190 από τους 485, δεν είχαν.

family_history_with_overweight				
	Frequency	Percent	Valid Percent	Cumulative Percent
no	190	39,2	39,2	39,2
yes	295	60,8	60,8	100,0
Total	485	100,0	100,0	

Πίνακας 2.2.5: Στατιστικά στοιχεία για το οικογενειακό ιστορικό παχυσαρκίας

Στο Διάγραμμα 2.2.5 φαίνεται ότι τα δεδομένα μας έχουν μια σχετική ισορροπία ως προς τη μεταβλητή αυτή, αφού τα άτομα που είχαν οικογενειακό ιστορικό παχυσαρκίας δεν ήταν πολύ περισσότερα από αυτά που δεν είχαν.



Διάγραμμα 2.2.5: Κατανομή οικογενειακού ιστορικού παχυσαρκίας

2.3 Διατροφικές Συνήθειες

Τα χαρακτηριστικά που σχετίζονται με τις διατροφικές συνήθειες είναι: η συχνή κατανάλωση τροφής με πολλές θερμίδες (FAVC), η συχνότητα κατανάλωσης λαχανικών μέσα στα γεύματα (FCVC), ο αριθμός κύριων γευμάτων (NCP), η κατανάλωση τροφής μεταξύ των γευμάτων (CAEC), η ποσότητα νερού που καταναλώνεται καθημερινά (CH20), η καθημερινή μέτρηση θερμίδων (SCC) και η συχνότητα κατανάλωσης αλκοόλ (CALC).

Η διατροφή παίζει σημαντικό ρόλο στη διάπλαση του σώματος, οπότε είναι πολύ χρήσιμο να γνωρίζουμε αν οι συμμετέχοντες πληρούν ή όχι κάποια χαρακτηριστικά και πως αυτά επηρεάζουν στη συνέχεια την ανάπτυξη της παχυσαρκίας.

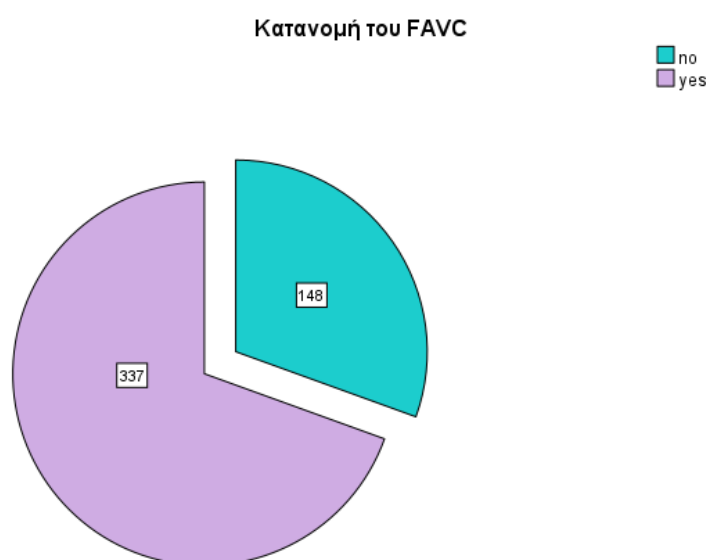
2.3.1 Περιγραφική ανάλυση μεταβλητής «FAVC»

Η μεταβλητή FAVC είναι μια δυαδική μεταβλητή στην οποία οι συμμετέχοντες απάντησαν με Ναι ή Όχι στην ερώτηση αν καταναλώνουν συχνά τροφή με πολλές θερμίδες. Τα δεδομένα δεν είναι ισορροπημένα καθώς έχουμε 337 θετικές απαντήσεις και 148 αρνητικές. Αντίστοιχα σε ποσοστά, το 69,5% των συμμετεχόντων καταναλώνουν συχνά τροφή με πολλές θερμίδες ενώ το 30,5% όχι.

FAVC				
	Frequency	Percent	Valid Percent	Cumulative Percent
no	148	30,5	30,5	30,5
yes	337	69,5	69,5	100,0
Total	485	100,0	100,0	

Πίνακας 2.3.1: Στατιστικά στοιχεία για την κατανάλωση τροφής με πολλές θερμίδες

Στο Διάγραμμα 2.3.1 παρουσιάζονται οι παρατηρούμενες συχνότητες με τη μορφή ενός διαγράμματος πίτας. Σχεδόν το 1/3 του πληθυσμού αποφεύγει τη συχνή κατανάλωση τροφής με πολλές θερμίδες.



Διάγραμμα 2.3.1: Κατανομή κατανάλωσης τροφής με πολλές θερμίδες

2.3.2 Περιγραφική ανάλυση μεταβλητής «FCVC»

Για την ερώτηση «Πόσο συχνά καταναλώνετε λαχανικά στα γεύματά σας;», αρχικά δημιουργήθηκε η κατηγορική μεταβλητή FCVC με πιθανές απαντήσεις:

- Ποτέ
- Μερικές φορές
- Πάντα

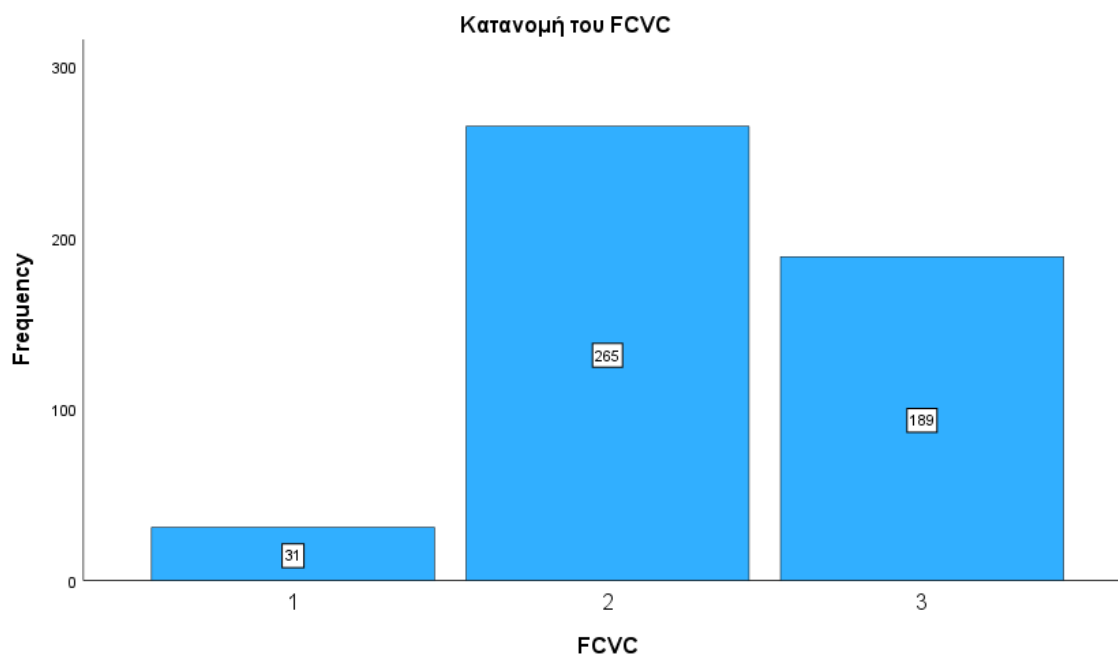
Στη συνέχεια οι απαντήσεις κωδικοποιήθηκαν σε αριθμούς 1, 2 και 3 αντίστοιχα, μετατρέποντας τη μεταβλητή σε ακέραια. Οι 265 από τους 485 απάντησαν ότι καταναλώνουν μερικές φορές λαχανικά ανάμεσα στα γεύματα, οι 189 καταναλώνουν πάντα ενώ μόνο οι 31 δεν καταναλώνουν ποτέ.

FCVC				
	Frequency	Percent	Valid Percent	Cumulative Percent
1	31	6,4	6,4	6,4
2	265	54,6	54,6	61,0
3	189	39,0	39,0	100,0
Total	485	100,0	100,0	

Πίνακας 2.3.2: Στατιστικά στοιχεία για τη συχνότητα κατανάλωσης λαχανικών

Αυτά τα αποτελέσματα, όπως βλέπουμε και στο Διάγραμμα 2.3.2, δηλώνουν ότι οι περισσότερες απαντήσεις ήταν μεταξύ του 2 και του 3. Δηλαδή, οι περισσότεροι

συμμετέχοντες δήλωσαν ότι τρώνε λαχανικά Μερικές φορές ή Πάντα. Συγκεκριμένα βλέπουμε ότι η απάντηση «Μερικές φορές» είχε τη μεγαλύτερη συχνότητα. Από την άλλη, η τιμή 1 έχει τη χαμηλότερη συχνότητα, γεγονός που υποδηλώνει ότι πολύ λίγα άτομα δεν τρώνε καθόλου λαχανικά στα γεύματα.



Διάγραμμα 2.3.2: Κατανομή συχνότητας κατανάλωσης λαχανικών

2.3.3 Περιγραφική ανάλυση μεταβλητής «NCP»

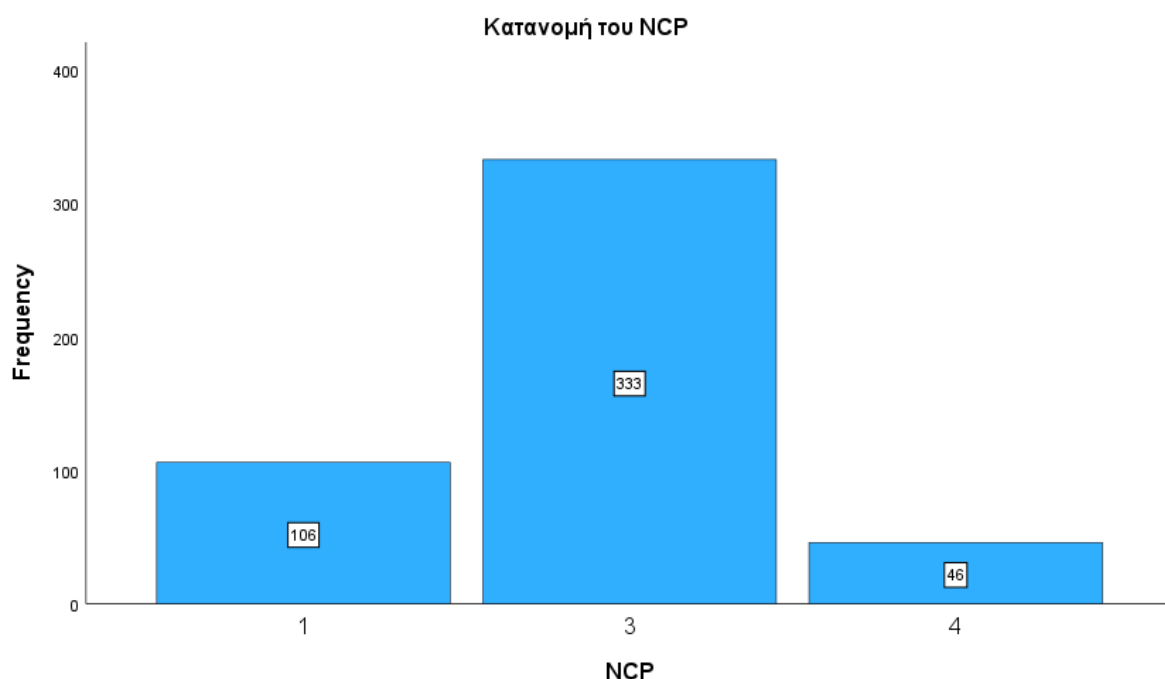
Για την ερώτηση «Πόσα κύρια γεύματα την ημέρα καταναλώνετε;», αρχικά δημιουργήθηκε η κατηγορική μεταβλητή NCP με πιθανές απαντήσεις:

- Μεταξύ ενός και δύο
- Τρία
- Πάνω από τρία

Στη συνέχεια, οι απαντήσεις κωδικοποιήθηκαν ως 1, 3 και 4 αντίστοιχα, μετατρέποντας τη μεταβλητή σε ακέραια. Όπως βλέπουμε στον Πίνακα και στο Διάγραμμα 2.3.3, η απάντηση 3 είναι η πιο συχνή. Δηλαδή, οι περισσότεροι συμμετέχοντες δήλωσαν ότι καταναλώνουν 3 κύρια γεύματα την ημέρα. Λίγοι είναι όμως εκείνοι που καταναλώνουν πάνω από τρία κυρίως γεύματα την ημέρα.

NCP				
	Frequency	Percent	Valid Percent	Cumulative Percent
1	106	21,9	21,9	21,9
3	333	68,7	68,7	90,5
4	46	9,5	9,5	100,0
Total	485	100,0	100,0	

Πίνακας 2.3.3: Στατιστικά στοιχεία για τον αριθμό κυρίων γευμάτων



Διάγραμμα 2.3.3: Κατανομή αριθμού κυρίων γευμάτων

2.3.4 Περιγραφική ανάλυση μεταβλητής «CAEC»

Για την ερώτηση αν οι συμμετέχοντες καταναλώνουν φαγητό ανάμεσα στα γεύματα, δημιουργήθηκε η κατηγορική μεταβλητή CAEC με πιθανές απαντήσεις:

- Πάντα
- Συχνά
- Μερικές φορές
- Όχι

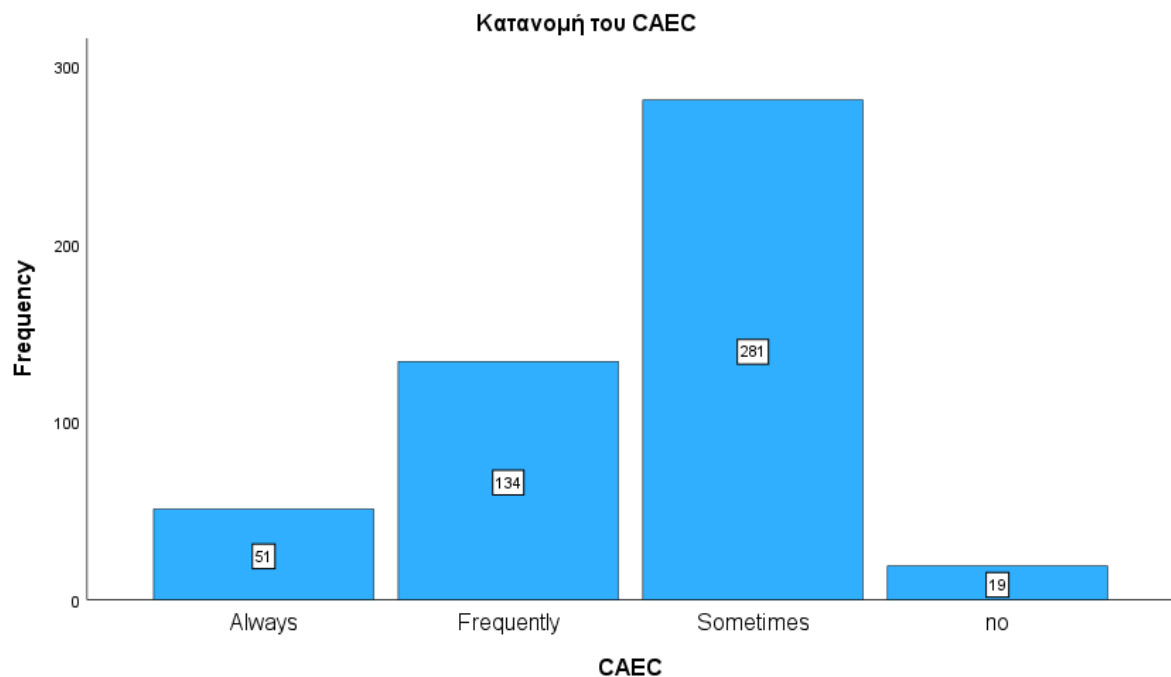
Από τις 485 εγγραφές, οι 51 είπα ότι καταναλώνουν πάντα φαγητό μεταξύ των γευμάτων, οι 134 είπαν συχνά, οι 281 είπαν μερικές φορές και οι 19 είπαν όχι.

CAEC				
	Frequency	Percent	Valid Percent	Cumulative Percent
Always	51	10,5	10,5	10,5
Frequently	134	27,6	27,6	38,1
Sometimes	281	57,9	57,9	100,0
no	19	3,9	3,9	42,1
Total	485	100,0	100,0	

Πίνακας 2.3.4: Στατιστικά στοιχεία για την κατανάλωση τροφής μεταξύ των γευμάτων

Στο Διάγραμμα 2.3.4 απεικονίζονται οι συχνότητες της μεταβλητής CAEC. Παρατηρούμε ότι η πλειοψηφία των ατόμων τρώει φαγητό μεταξύ γευμάτων μερικές φορές ή συχνά, ενώ λιγότερα άτομα το κάνουν πάντα ή ποτέ. Αν η CAEC σχετίζεται με την παχυσαρκία, αυτή η

πληροφορία μπορεί να είναι πολύ χρήσιμη για την ανάλυση και την πρόβλεψη του επιπέδου παχυσαρκίας.



Διάγραμμα 2.3.4: Κατανομή κατανάλωσης τροφής μεταξύ των γευμάτων

2.3.5 Περιγραφική ανάλυση μεταβλητής «CH2O»

Για την ερώτηση «Πόσο νερό καταναλώνετε καθημερινά;», αρχικά δημιουργήθηκε η κατηγορική μεταβλητή CH2O με πιθανές απαντήσεις:

- Λιγότερο από 1 λίτρο
- Μεταξύ 1 και 2 λίτρων
- Περισσότερο από 2 λίτρα

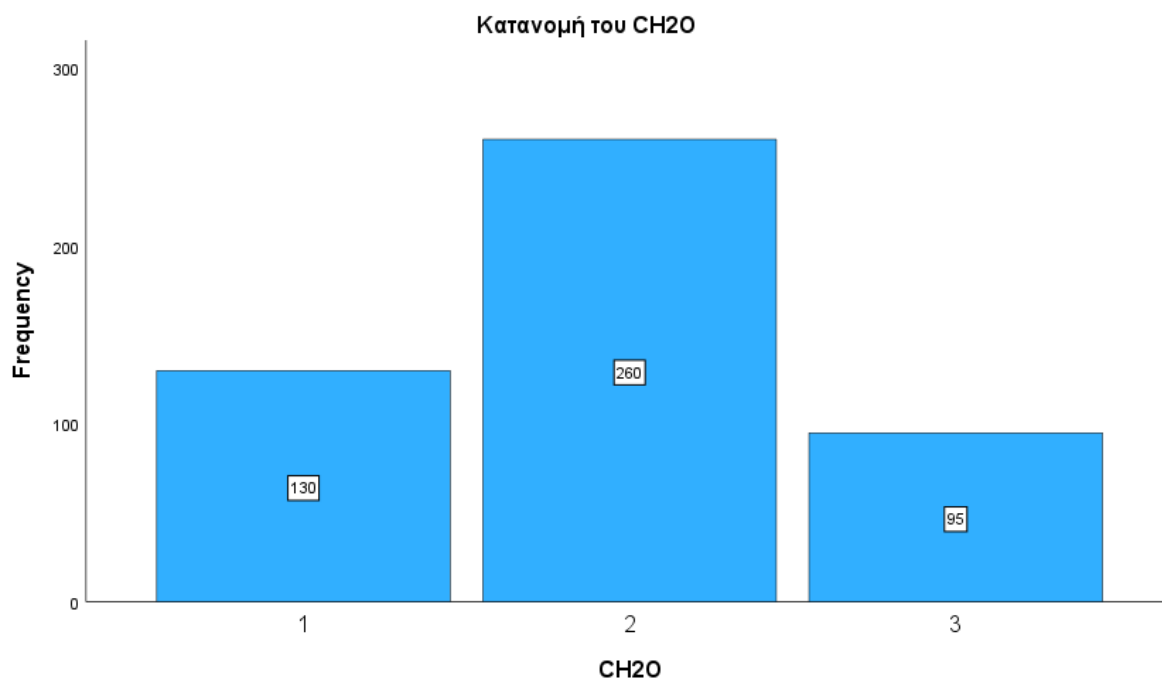
Στη συνέχεια, οι απαντήσεις κωδικοποιήθηκαν σε 1, 2 και 3 αντίστοιχα, μετατρέποντας τη μεταβλητή σε ακέραια.

CH2O				
	Frequency	Percent	Valid Percent	Cumulative Percent
1	130	26,8	26,8	26,8
2	260	53,6	53,6	80,4
3	95	19,6	19,6	100,0
Total	485	100,0	100,0	

Πίνακας 2.3.5: Στατιστικά για την ποσότητα νερού που καταναλώνεται καθημερινά

Όπως παρατηρούμε και στο Διάγραμμα 2.3.5, οι περισσότερες απαντήσεις συγκεντρώνονται στην τιμή 2, δηλαδή οι 260 από τους 485 καταναλώνουν «Μεταξύ 1 και 2 λίτρων» νερού καθημερινά. Η δεύτερη συχνότερη κατηγορία είναι η «Λιγότερο από 1 λίτρο», με 130 άτομα, δηλαδή τα μισά άτομα σε σχέση με την κυρίαρχη κατηγορία. Η τιμή 3 έχει τη

μικρότερη συχνότητα, δηλαδή αυτοί που καταναλώνουν «Περισσότερο από 2 λίτρα» νερό καθημερινά είναι οι λιγότεροι.



Διάγραμμα 2.3.5: Κατανομή ποσότητας νερού που καταναλώνεται καθημερινά

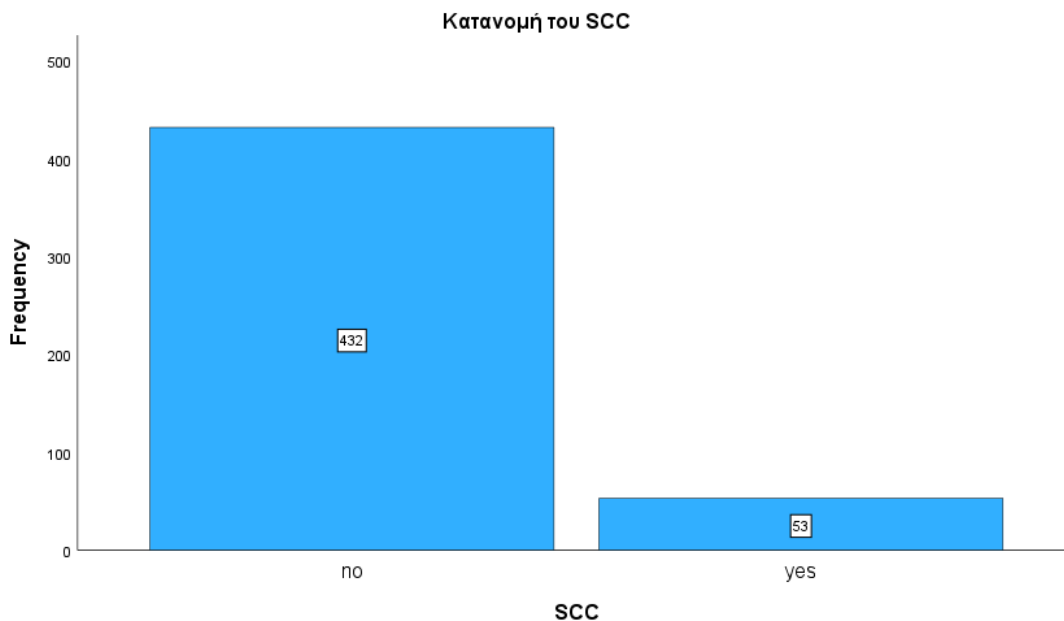
2.3.6 Περιγραφική ανάλυση μεταβλητής «SCC»

Για την ερώτηση αν οι συμμετέχοντες κάνουν καταμέτρηση των θερμίδων που λαμβάνουν καθημερινά δημιουργήθηκε η δυαδική μεταβλητή SCC με πιθανές απαντήσεις: Ναι ή Όχι. Στον Πίνακα 2.3.6 βλέπουμε ότι σχεδόν το 89% των 485 εγγραφών απάντησε ότι δεν τις μετράει ενώ μόνο περίπου το 11% τις μετράει.

SCC				
	Frequency	Percent	Valid Percent	Cumulative Percent
no	432	89,1	89,1	89,1
yes	53	10,9	10,9	100,0
Total	485	100,0	100,0	

Πίνακας 2.3.6: Στατιστικά στοιχεία για την καθημερινή μέτρηση θερμίδων

Στο Διάγραμμα 2.3.6 απεικονίζονται οι παρατηρούμενες συχνότητες της μεταβλητής SCC. Το γράφημα δείχνει μια ισχυρή μη ισόρροπη κατανομή των παρατηρήσεων ανάμεσα στις δύο κατηγορίες. Η πλειοψηφία των ατόμων (432) δεν παρακολουθεί τις θερμίδες, ενώ μόνο 53 άτομα τις παρακολουθούν. Αυτή η απόκλιση στις συχνότητες των δύο κατηγοριών, μπορεί να επηρεάσει την ανάλυση και την πρόβλεψη του επιπέδου παχυσαρκίας.



Διάγραμμα 2.3.6: Κατανομή καθημερινής μέτρησης θερμίδων

2.3.7 Περιγραφική ανάλυση μεταβλητής «CALC»

Για την ερώτηση «Πόσο συχνά καταναλώνετε αλκοόλ;», δημιουργήθηκε η κατηγορική μεταβλητή CALC με πιθανές απαντήσεις:

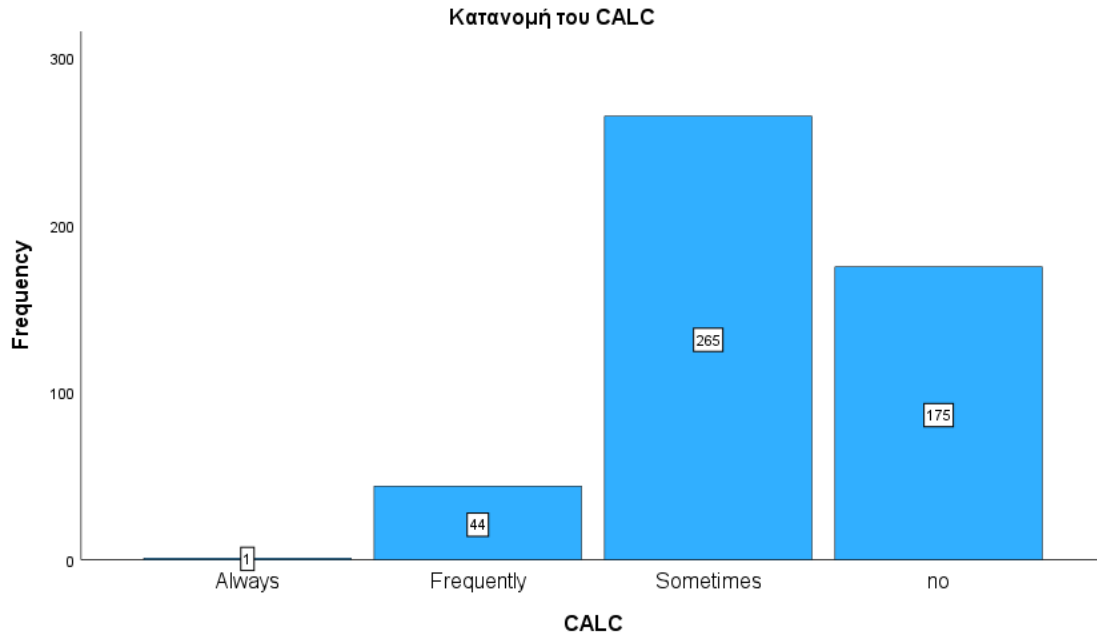
- Πάντα
- Συχνά
- Μερικές φορές
- Όχι

Στον Πίνακα 2.3.7 βλέπουμε ότι μόνο ένα άτομο απάντησε ότι καταναλώνει πάντα αλκοόλ, ενώ το 54,6% απάντησε μερικές φορές, το 36,1% είπε όχι και το 9,1% συχνά.

CALC				
	Frequency	Percent	Valid Percent	Cumulative Percent
Always	1	,2	,2	,2
Frequently	44	9,1	9,1	9,3
Sometimes	265	54,6	54,6	100,0
no	175	36,1	36,1	45,4
Total	485	100,0	100,0	

Πίνακας 2.3.7: Στατιστικά στοιχεία για τη συχνότητα κατανάλωσης αλκοόλ

Στο Διάγραμμα 2.3.7 απεικονίζονται οι παρατηρούμενες συχνότητες της μεταβλητής CALC. Το γράφημα δείχνει ότι η πλειοψηφία των ατόμων ποτέ ή μερικές φορές πίνει αλκοόλ, ενώ λιγότερα άτομα το κάνουν συχνά ή πάντα. Αν η CALC σχετίζεται με την παχυσαρκία, αυτή η πληροφορία μπορεί να είναι πολύ χρήσιμη για την ανάλυση και την πρόβλεψη του επιπέδου παχυσαρκίας.



Διάγραμμα 2.3.7: Κατανομή συχνότητας κατανάλωσης αλκοόλ

2.4 Φυσική Δραστηριότητα και άλλες συνήθειες

Τα χαρακτηριστικά που σχετίζονται με τη φυσική δραστηριότητα είναι: η συχνότητα σωματικής άσκησης (FAF), ο χρόνος χρήσης τεχνολογικών συσκευών (TUE) και ο τρόπος μετακίνησης (MTRANS). Ως άλλες συνήθειες έχουμε τη μεταβλητή SMOKE που δηλώνει αν είναι καπνιστής ή όχι.

Η σωματική διάπλαση επηρεάζεται πολύ από τη σωματική δραστηριότητα, οπότε είναι σημαντικό να μελετήσουμε τη συσχέτιση αυτών των μεταβλητών με την παχυσαρκία. Το κάπνισμα είναι ένας επιπλέον παράγοντας που πρέπει να μελετήσουμε για να δούμε τι σχέση έχει με την παχυσαρκία.

2.4.1 Περιγραφική ανάλυση μεταβλητής «FAF»

Για την ερώτηση «Πόσο συχνά ασκείστε σωματικά την εβδομάδα;», αρχικά δημιουργήθηκε η κατηγορική μεταβλητή FAF με πιθανές απαντήσεις:

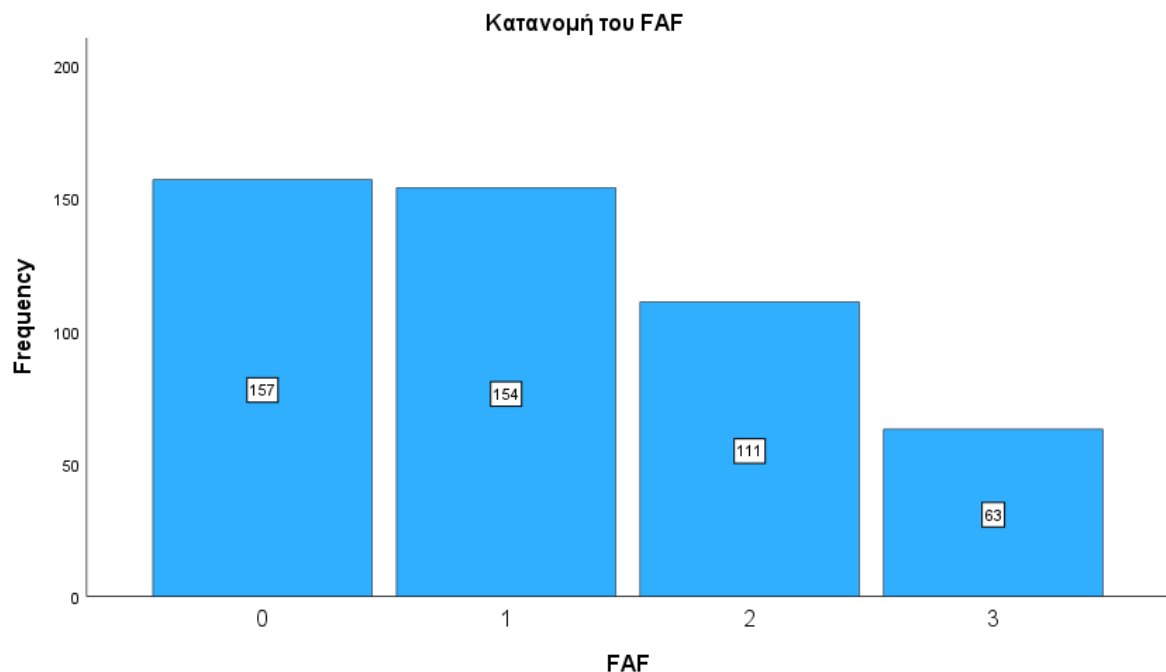
- Καθόλου
- 1 ή 2 ημέρες
- 2 ή 4 ημέρες
- 4 ή 5 ημέρες

Στη συνέχεια, οι απαντήσεις κωδικοποιήθηκαν σε 0, 1, 2 και 3 αντίστοιχα, μετατρέποντας τη μεταβλητή σε ακέραια. Στον Πίνακα 2.4.1 βλέπουμε ότι η πλειοψηφία (157) δεν ασκείται καθόλου, οι 154 ασκούνται 1 ή 2 ημέρες την εβδομάδα, οι 111 ασκούνται 2 ή 4 ημέρες και τέλος μόνο 63 άτομα δήλωσαν ότι ασκούνται 4 ή 5 ημέρες την εβδομάδα.

Στο Διάγραμμα 2.4.1, βλέπουμε ότι η πιο συχνή τιμή είναι το 0. Αυτό σημαίνει ότι η πλειοψηφία των ατόμων στην έρευνα δεν ασκείται καθόλου. Βλέπουμε ότι υπάρχουν άτομα που ασκούν φυσική δραστηριότητα πιο συχνά, αλλά η πλειοψηφία είναι ανάμεσα στις απαντήσεις 0 ή 1, δηλαδή δεν ασκούνται καθόλου ή ασκούνται 1 με 2 ημέρες την εβδομάδα.

FAF				
	Frequency	Percent	Valid Percent	Cumulative Percent
0	157	32,4	32,4	32,4
1	154	31,8	31,8	64,1
2	111	22,9	22,9	87,0
3	63	13,0	13,0	100,0
Total	485	100,0	100,0	

Πίνακας 2.4.1: Στατιστικά στοιχεία για τη συχνότητα σωματικής άσκησης



Διάγραμμα 2.4.1: Κατανομή συχνότητας σωματικής άσκησης

2.4.2 Περιγραφική ανάλυση μεταβλητής «TUE»

Για την ερώτηση «Πόσες ώρες χρησιμοποιείτε τεχνολογικές συσκευές όπως κινητό, βιντεοπαιχνίδια, τηλεόραση, υπολογιστή κτλ.;;» δημιουργήθηκε αρχικά η κατηγορική μεταβλητή TUE, με πιθανές απαντήσεις:

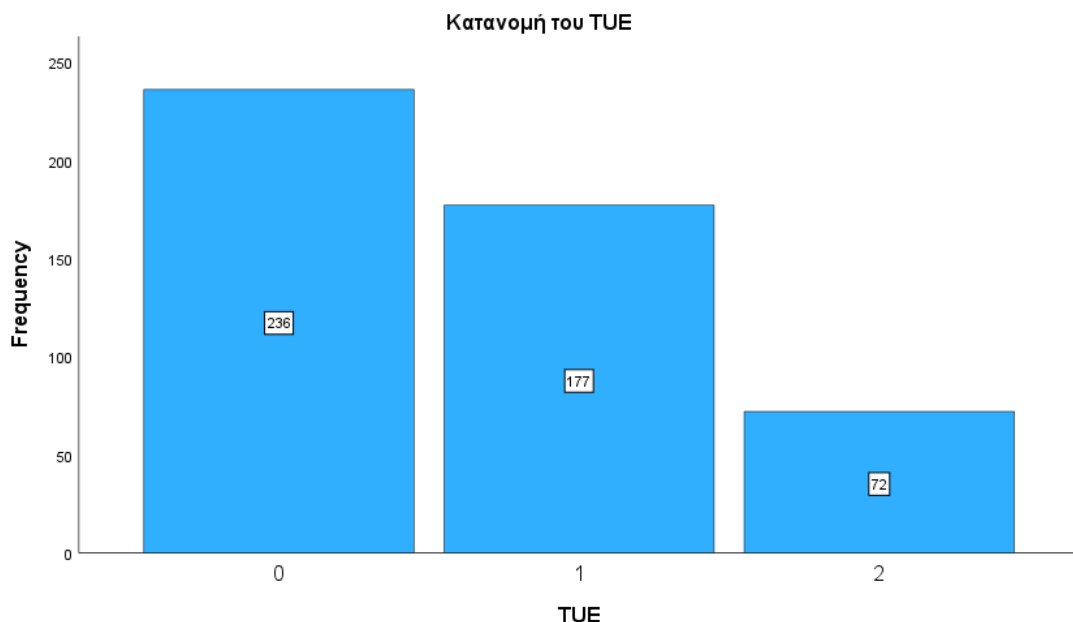
- 0-2 ώρες
- 3-5 ώρες
- Πάνω από 6 ώρες

Στη συνέχεια, οι απαντήσεις κωδικοποιήθηκαν σε 0, 1 και 2 αντίστοιχα και η μεταβλητή έγινε ακέραια. Τη μεταβλητή TUE την κατατάσσουμε στη φυσική δραστηριότητα καθώς όσες περισσότερες ώρες καταναλώνει κανείς στην τεχνολογία, πιθανώς τόσο λιγότερο χρόνο θα έχει για σωματική άσκηση.

TUE				
	Frequency	Percent	Valid Percent	Cumulative Percent
0	236	48,7	48,7	48,7
1	177	36,5	36,5	85,2
2	72	14,8	14,8	100,0
Total	485	100,0	100,0	

Πίνακας 2.4.2: Στατιστικά στοιχεία χρόνου χρήσης τεχνολογικών συσκευών

Στο Διάγραμμα 2.4.2 βλέπουμε ότι η τιμή 0 φαίνεται να είναι η πιο συχνή, δηλαδή 236 άτομα χρησιμοποιούν τεχνολογικά μέσα για 0-2 ώρες τη μέρα. Ακολουθεί η τιμή 1 (2-4 ώρες) με 177 άτομα και τέλος, η τιμή 2 (4-6 ώρες) με 72 άτομα. Αυτό σημαίνει ότι η πλειοψηφία των ατόμων στην έρευνα δεν χρησιμοποιεί τεχνολογικές συσκευές για πολλές ώρες την ημέρα.



Διάγραμμα 2.4.2: Κατανομή χρόνου χρήσης τεχνολογικών συσκευών

2.4.3 Περιγραφική ανάλυση μεταβλητής «MTRANS»

Για την ερώτηση «Ποιο μέσο μεταφοράς χρησιμοποιείτε συνήθως;» δημιουργήθηκε η κατηγορική μεταβλητή MTRANS με πιθανές απαντήσεις:

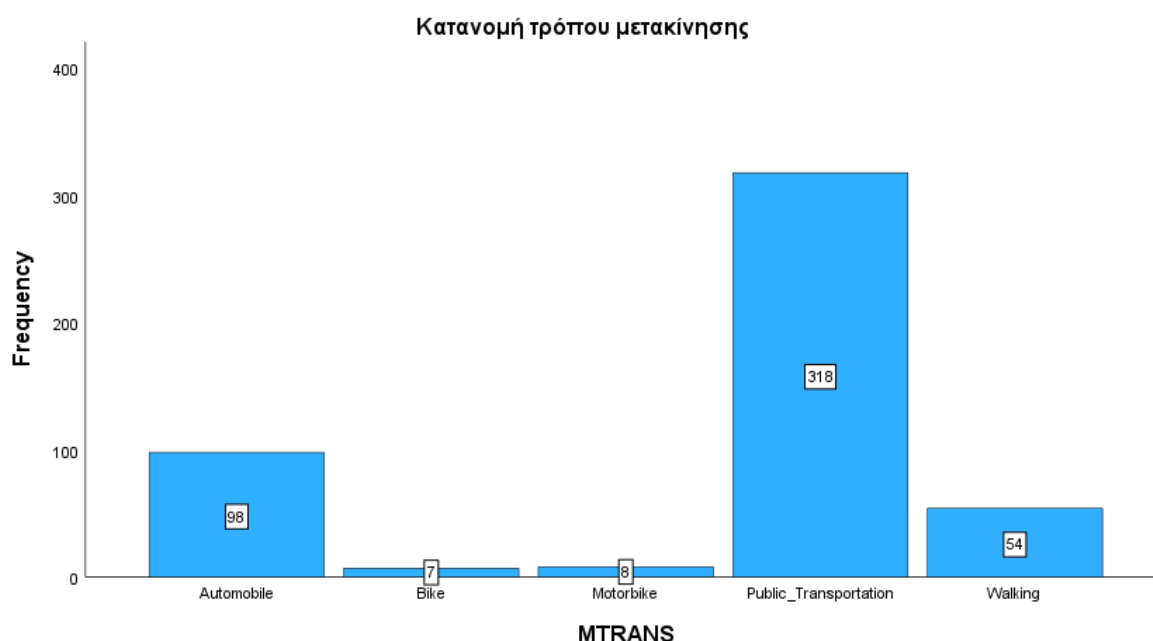
- Αυτοκίνητο
- Ποδήλατο
- Μηχανή
- Δημόσιες συγκοινωνίες
- Περπάτημα

Στον Πίνακα 2.4.3 βλέπουμε τις συχνότητες των απαντήσεων. Οι 318 από τους 485 συμμετέχοντες χρησιμοποιούν δημόσιες συγκοινωνίες, οι 98 αυτοκίνητο, οι 8 μηχανή ενώ μόνο 54 περπατάνε και 7 χρησιμοποιούν ποδήλατο.

MTRANS				
	Frequency	Percent	Valid Percent	Cumulative Percent
Automobile	98	20,2	20,2	20,2
Bike	7	1,4	1,4	21,6
Motorbike	8	1,6	1,6	23,3
Public_Transportation	318	65,6	65,6	88,9
Walking	54	11,1	11,1	100,0
Total	485	100,0	100,0	

Πίνακας 2.4.3: Στατιστικά στοιχεία για τον τρόπο μετακίνησης

Στο Διάγραμμα 2.4.3 απεικονίζονται οι παρατηρούμενες συχνότητες της μεταβλητής MTRANS. Βλέπουμε ότι οι κατηγορίες δεν είναι ισορροπημένες. Το γράφημα δείχνει ότι η πλειοψηφία των ατόμων χρησιμοποιεί δημόσιες συγκοινωνίες, αυτοκίνητο ή περπατά για τη μετακίνησή τους, ενώ λιγότερα άτομα χρησιμοποιούν μηχανή ή ποδήλατο.



Διάγραμμα 2.4.3: Κατανομή τρόπου κίνησης

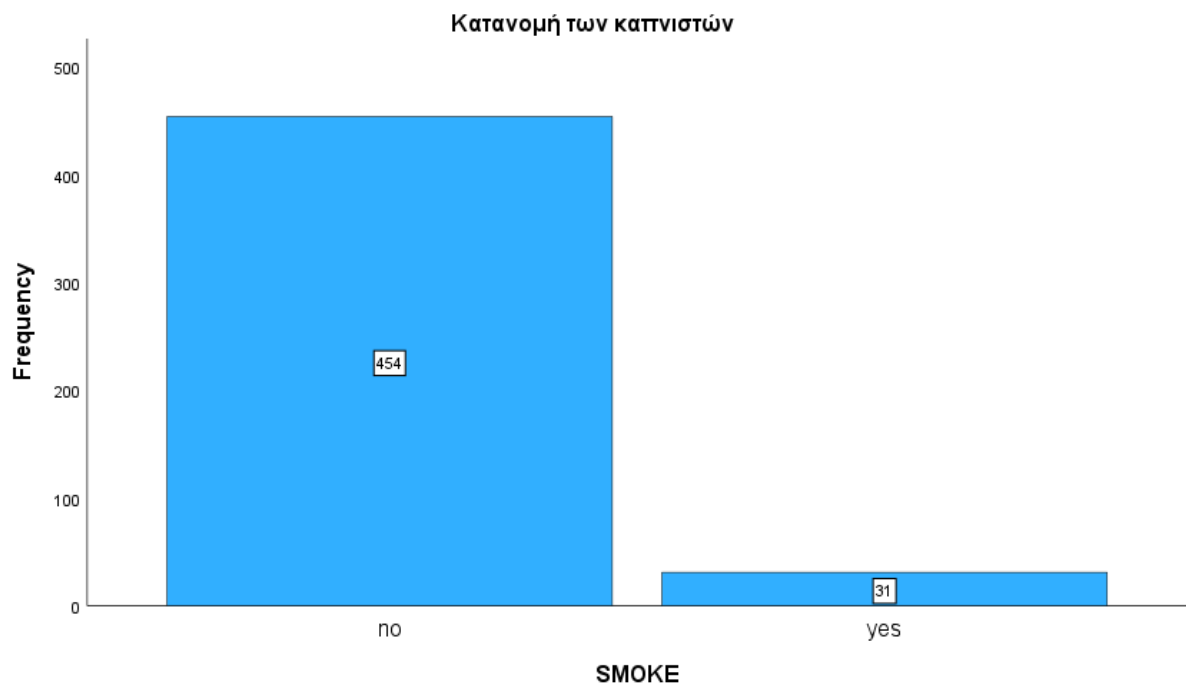
2.4.4 Περιγραφική ανάλυση μεταβλητής «SMOKE»

Για την ερώτηση αν καπνίζουν οι συμμετέχοντες, δημιουργήθηκε η δυαδική μεταβλητή SMOKE με πιθανές απαντήσεις: Ναι ή Όχι. Παρατηρούμε από τον Πίνακα 2.4.4 ότι μόνο το 6,4% ήταν καπνιστές ενώ το 93,6% των συμμετεχόντων δεν ήταν. Αυτά τα ποσοστά δείχνουν ότι υπάρχει ισχυρή ανισοκατανομή μεταξύ των δύο κατηγοριών.

SMOKE				
	Frequency	Percent	Valid Percent	Cumulative Percent
no	454	93,6	93,6	93,6
yes	31	6,4	6,4	100,0
Total	485	100,0	100,0	

Πίνακας 2.4.4: Στατιστικά στοιχεία για καπνιστές

Στο Διάγραμμα 2.4.4 απεικονίζονται οι παρατηρούμενες συχνότητες της μεταβλητής SMOKE. Στο γράφημα φαίνεται ξεκάθαρα ότι η μία κατηγορία υπερτερεί σημαντικά της άλλης. Η πλειοψηφία των ατόμων (454) δεν καπνίζει, ενώ μόνο 31 άτομα καπνίζουν. Αν η μεταβλητή SMOKE σχετίζεται με την παχυσαρκία, τότε η διαφορά στις κατηγορίες μπορεί να επηρεάσει την ικανότητα του μοντέλου να προβλέψει σωστά τους καπνιστές.



Διάγραμμα 2.4.4: Κατανομή των καπνιστών

2.5 Επίπεδο παχυσαρκίας

Αφού αναλύσαμε τους παράγοντες που μπορούν να οδηγήσουν σε παχυσαρκία καθώς και αυτούς που αποτρέπουν ή καθυστερούν την εμφάνισή της, σειρά έχει να αναλύσουμε και τη μεταβλητή που μας δείχνει τα διάφορα επίπεδα παχυσαρκίας του δείγματος.

2.5.1 Περιγραφική ανάλυση μεταβλητής «NObeyesdad»

Η κατηγορική μεταβλητή NObeyesdad ορίζει το επίπεδο της παχυσαρκίας. Παίρνει τις τιμές από το βάρος και το ύψος για να ορίσει το ΔΜΣ και αυτό στη συνέχεια παίρνει τις ονομαστικές τιμές όπως τις ορίσαμε στον Πίνακα 1.1.

		NObeyesdad			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Insufficient_Weight	33	6,8	6,8	6,8
	Normal_Weight	278	57,3	57,3	64,1
	Overweight_Level_I	58	12,0	12,0	76,1
	Overweight_Level_II	56	11,5	11,5	87,6
	Obesity_Type_I	46	9,5	9,5	97,1
	Obesity_Type_II	11	2,3	2,3	99,4
	Obesity_Type_III	3	,6	,6	100,0
	Total	485	100,0	100,0	

Πίνακας 2.5.1: Επίπεδο παχυσαρκίας

Στον Πίνακα 2.5.1 βλέπουμε ότι τα ποσοστά των εγγραφών στις περισσότερες κατηγορίες είναι αρκετά χαμηλά. Τα ποσοστά για τις κατηγορίες «Υπέρβαρο Επίπεδο Ι» και «Υπέρβαρο Επίπεδο ΙΙ» είναι πολύ κοντά (12% και 11,5% αντίστοιχα), γεγονός που υποδηλώνει ότι ένα σημαντικό ποσοστό του πληθυσμού είναι υπέρβαρο.



Διάγραμμα 2.5.1: Κατανομή επιπέδου παχυσαρκίας

Στο Διάγραμμα 2.5.1 απεικονίζονται οι συχνότητες των διάφορων επιπέδων. Οι κατηγορίες «Παχυσαρκία Τύπου Ι», «Παχυσαρκία Τύπου ΙΙ» και «Παχυσαρκία Τύπου ΙΙΙ» φαίνεται να έχουν χαμηλές συχνότητες, γεγονός που υποδηλώνει ότι ένα μικρό ποσοστό του πληθυσμού αντιμετωπίζει προβλήματα παχυσαρκίας. Ειδικά οι δύο τελευταίες κατηγορίες, έχουν τη μικρότερη συχνότητα. Η κατηγορία «Κανονικό βάρος» φαίνεται να έχει την υψηλότερη συχνότητα, που σημαίνει ότι το μεγαλύτερο ποσοστό του πληθυσμού έχει κανονικό βάρος. Τέλος, η κατηγορία «Λιποβαρής» φαίνεται να έχει σχετικά χαμηλή συχνότητα, δηλαδή πολύ λίγα άτομα έχουν ελλιπές βάρος.

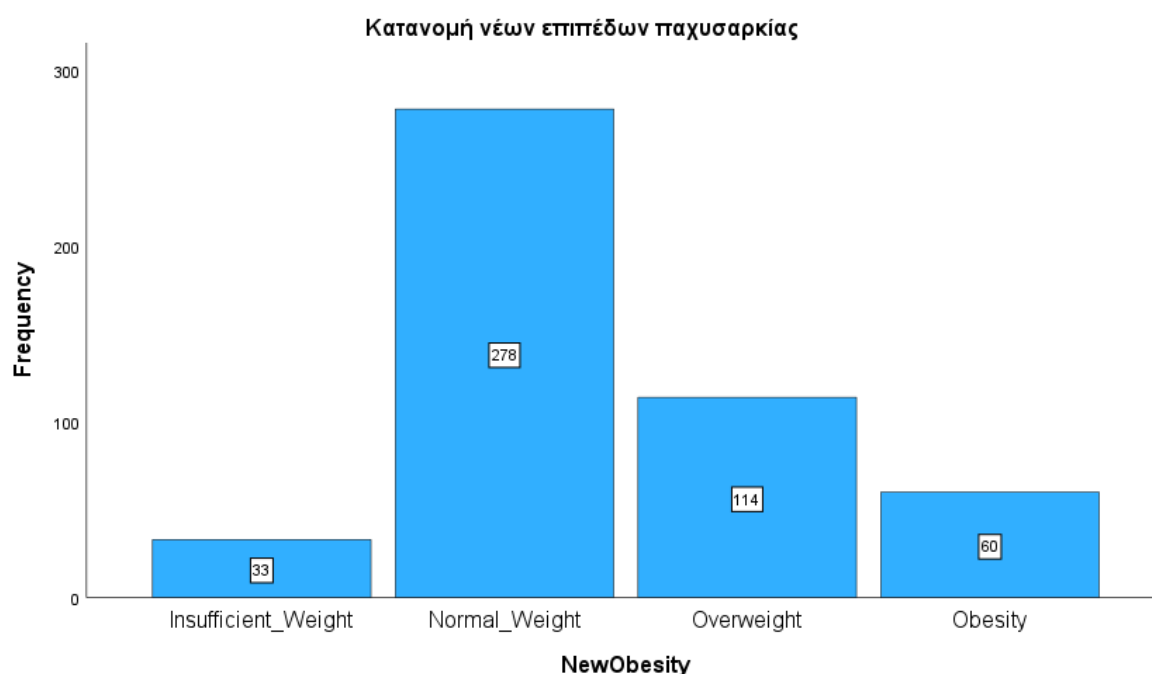
2.5.2 Περιγραφική ανάλυση μεταβλητής «NewObesity»

Για τις ανάγκες της εργασίας, ενώσαμε τα δύο διαφορετικά επίπεδα των υπέρβαρων σε μία ενιαία κατηγορία, την οποία ονομάσαμε «Overweight» και αντίστοιχα τα τρία επίπεδα παχυσαρκίας σε μία, την κατηγορία «Obesity». Παρακάτω βλέπουμε τις συχνότητές τους.

NewObesity					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Insufficient_Weight	33	6,8	6,8	6,8
	Normal_Weight	278	57,3	57,3	64,1
	Overweight	114	23,5	23,5	87,6
	Obesity	60	12,4	12,4	100,0
	Total	485	100,0	100,0	

Πίνακας 2.5.2: Στατιστικά στοιχεία νέων επιπέδων παχυσαρκίας

Στον Πίνακα και στο Διάγραμμα 2.5.2 μπορούμε να δούμε ότι η κατηγορία «Λιποβαρής», με ποσοστό 6,8%, είναι η λιγότερο συχνή, δηλαδή μόνο 33 άτομα από τα 485 έχουν ελλιπές βάρος, ακολουθούμενη από την κατηγορία «Obesity» με ποσοστό 12,4% να είναι η δεύτερη λιγότερο συχνή. Η «Overweight» είναι η δεύτερη πιο συχνή κατηγορία με 23,5%. Αυτά υποδηλώνουν ότι ένα μικρό ποσοστό του πληθυσμού αντιμετωπίζει προβλήματα παχυσαρκίας και λιγότερο από το μισό του πληθυσμού αντιμετωπίζει προβλήματα υπέρβαρου ή παχυσαρκίας. Η κατηγορία «Κανονικό βάρος», με ποσοστό 57,3% είναι η πιο συχνή, που σημαίνει ότι το μεγαλύτερο ποσοστό του πληθυσμού έχει κανονικό βάρος.



Διάγραμμα 2.5.2: Κατανομή νέων επιπέδων παχυσαρκίας

2.6 Δείκτης Μάζας Σώματος

Ο Δείκτης Μάζας Σώματος, όπως έχουμε ήδη αναφέρει, είναι ένα εργαλείο για να μπορούμε να υπολογίσουμε σε ποιο επίπεδο παχυσαρκίας ανήκει κάθε άνθρωπος, γνωρίζοντας απλά το βάρος και το ύψος του. Για να τον υπολογίσουμε χρησιμοποιούμε τον τύπο:

$$\Delta\text{Μ}\Sigma = \frac{\text{Βάρος}}{\text{Υψος}^2}$$

2.6.1 Περιγραφική ανάλυση «ΔΜΣ»

Για τις ανάγκες της εργασίας, φτιάξαμε τη μεταβλητή BMI, η οποία σύμφωνα με την επίσης καινούργια μεταβλητή NewObesity παίρνει τις εξής τιμές:

ΔΜΣ	Ταξινόμηση
<18,5	Λιποβαρής
18,5-24,9	Κανονικό βάρος
25-29,9	Υπέρβαρος
>30	Παχύσαρκος

Πίνακας 2.6.1(α): Πίνακας ταξινόμησης επιπέδων παχυσαρκίας σύμφωνα με το ΔΜΣ

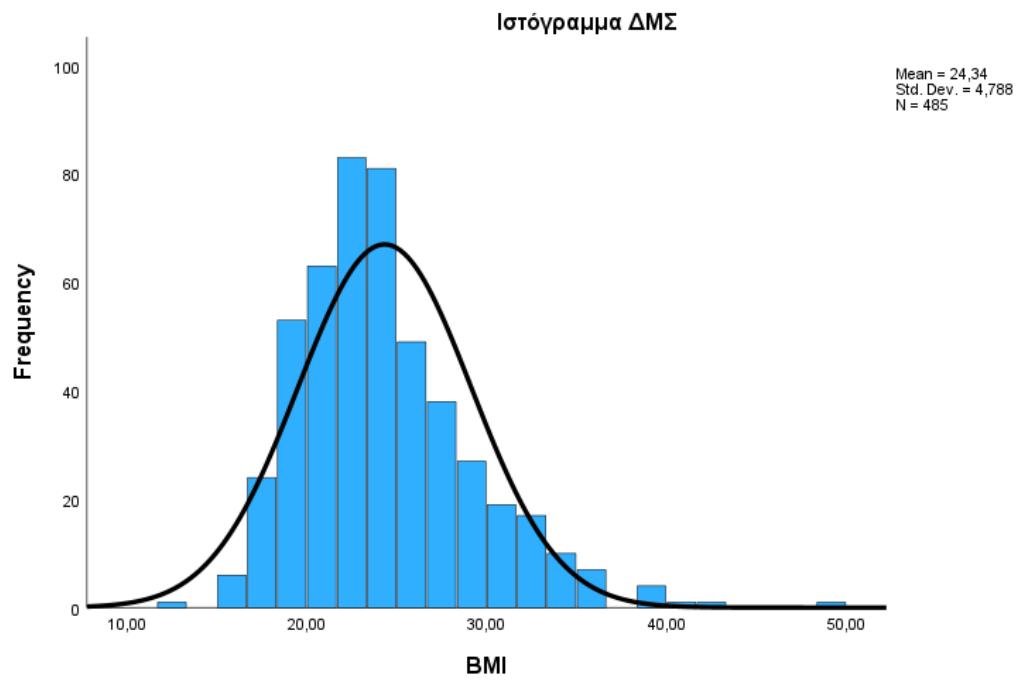
Στη συγκεκριμένη μελέτη, ο ΔΜΣ είναι μια συνεχής μεταβλητή που, σύμφωνα με τον Πίνακα 2.6.1(β), παίρνει τιμές οι οποίες κυμαίνονται μεταξύ 13,29 και 49,47 με διάμεσο περίπου 23,7 και μέσο όρο $24,34 \pm 4,78$ μονάδες που έχουν μετρηθεί σε kg/m^2 .

Statistics

BMI		
Mean		24,3414
Median		23,7118
Std. Deviation		4,78752
Minimum		13,29
Maximum		49,47
Percentiles	25	21,0077
	50	23,7118
	75	26,7432

Πίνακας 2.6.1(β): Στατιστικά στοιχεία για το Δείκτη Μάζας Σώματος

Στο Διάγραμμα 2.6.1 βλέπουμε το ιστόγραμμα του ΔΜΣ. Φαίνεται ότι υπάρχει δεξιά ασυμμετρία. Η μακριά ουρά στα δεξιά δείχνει ότι υπάρχουν κάποια άτομα με αρκετά μεγαλύτερο ΔΜΣ από τον μέσο όρο. Οι περισσότερες παρατηρήσεις βρίσκονται περίπου στο διάστημα 18-28 με κορύφωση γύρω στις τιμές 22-24, που είναι και κοντά στον μέσο όρο. Ο μέσος όρος επηρεάζεται από κάποιες πολύ μεγάλες τιμές (> 38 μονάδες). Το γεγονός ότι ο μέσος ΔΜΣ είναι εντός φυσιολογικών ορίων, αλλά με μεγάλη τυπική απόκλιση, υποδεικνύει ετερογένεια στο σωματικό βάρος του δείγματος. Αν και το ιστόγραμμα οπτικά πλησιάζει την κανονική κατανομή, η δεξιά ασυμμετρία δείχνει ότι δεν είναι τέλεια κανονική.



Διάγραμμα 2.6.1: Ιστόγραμμα κατανομής Δείκτη Μάζας Σώματος

Κεφάλαιο 3

Συσχετίσεις

3.1 Συσχετίσεις

Στο κεφάλαιο αυτό θα ερευνήσουμε αν υπάρχει συσχέτιση μεταξύ των διαθέσιμων χαρακτηριστικών και της μεταβλητής «επίπεδο παχυσαρκίας». Το επίπεδο παχυσαρκίας είναι μια κατηγορική μεταβλητή με τέσσερα επίπεδα. Γνωρίζουμε ότι το δείγμα μας είναι τυχαίο και οι παρατηρήσεις ανεξάρτητες. Παρακάτω θα κάνουμε μια περίληψη της μεθοδολογίας που θα ακολουθήσουμε.

Για τις ποιοτικές μεταβλητές, πρέπει να κάνουμε έλεγχο ανεξαρτησίας με το κριτήριο χ^2 του Pearson ή με τη μέθοδο Monte Carlo, αν δεν ικανοποιούνται οι προϋποθέσεις του πρώτου. Για αυτό το λόγο θα κάνουμε τον εξής έλεγχο υποθέσεων:

Μηδενική υπόθεση

H₀: η εξεταζόμενη μεταβλητή και η μεταβλητή «επίπεδο παχυσαρκίας» είναι ανεξάρτητες μεταξύ τους

έναντι της εναλλακτικής υπόθεσης

H₁: η εξεταζόμενη μεταβλητή και η μεταβλητή «επίπεδο παχυσαρκίας» δεν είναι ανεξάρτητες μεταξύ τους.

Όπως αναφέραμε στην Παράγραφο 1.7, η μηδενική υπόθεση του ελέγχου απορρίπτεται όταν $p\text{-value} < 0,05$ και δηλώνει την ύπαρξη σχέσης μεταξύ των εξεταζόμενων μεταβλητών. Στην περίπτωση αυτή, θα υπολογίσουμε το συντελεστή V του Cramer, για να δούμε την ένταση της συσχέτισης.

Για τις ποσοτικές μεταβλητές, πρέπει να κάνουμε έλεγχο κανονικότητας της μεταβλητής για κάθε κατηγορία του επιπέδου παχυσαρκίας, μέσω του στατιστικού ελέγχου Kolmogorov-Smirnov. Οι υποθέσεις που ελέγχονται είναι οι εξής:

H₀: Τα δεδομένα ακολουθούν την κανονική κατανομή.

H₁: Τα δεδομένα δεν ακολουθούν την κανονική κατανομή.

Η μηδενική υπόθεση του ελέγχου δεν απορρίπτεται όταν $p\text{-value} > 0,05$ και δηλώνει ότι τα δεδομένα ακολουθούν την κανονική κατανομή. Σε αυτή την περίπτωση, πρέπει να ελέγξουμε και την ομοσκεδαστικότητα ανάμεσα στις ομάδες. Οι υποθέσεις που ελέγχουμε είναι:

H₀: Τα δεδομένα είναι ομοσκεδαστικά

H₁: Τα δεδομένα δεν είναι ομοσκεδαστικά

Η μηδενική υπόθεση του ελέγχου δεν απορρίπτεται όταν $p\text{-value} > 0,05$ και μπορούμε να εφαρμόσουμε One Way ANOVA για να ελέγξουμε τη συσχέτιση της εκάστοτε μεταβλητής με το «επίπεδο παχυσαρκίας». Στην περίπτωση που η μηδενική υπόθεση απορριφθεί, τότε μπορούμε να κάνουμε Welch's ANOVA.

Τέλος, στην περίπτωση που τα δεδομένα δεν ακολουθούν κανονική κατανομή, θα πρέπει να κάνουμε μη παραμετρικούς ελέγχους, όπως ο έλεγχος Kruskal-Wallis. Οι υποθέσεις που ελέγχουμε σε αυτή την περίπτωση είναι:

H₀: η εξεταζόμενη μεταβλητή και η μεταβλητή «επίπεδο παχυσαρκίας» είναι ανεξάρτητες μεταξύ τους

H₁: η εξεταζόμενη μεταβλητή και η μεταβλητή «επίπεδο παχυσαρκίας» δεν είναι ανεξάρτητες μεταξύ τους.

Όπως και στις ποιοτικές μεταβλητές, η μηδενική υπόθεση του ελέγχου απορρίπτεται όταν $p\text{-value} < 0,05$ και δηλώνει την ύπαρξη σχέσης μεταξύ των εξεταζόμενων μεταβλητών.

3.2 Συσχετίσεις δημογραφικών στοιχείων με το «Επίπεδο παχυσαρκίας»

Σε αυτή την παράγραφο θα ελέγξουμε τη σχέση εξάρτησης που μπορεί να υπάρχει ανάμεσα στα δημογραφικά στοιχεία, όπως παρουσιάστηκαν στο δεύτερο κεφάλαιο, και στη μεταβλητή της παχυσαρκίας.

3.2.1 Συσχέτιση με τη μεταβλητή «Φύλο»

Όπως αναφέραμε στο κεφάλαιο 2, η μεταβλητή φύλο είναι μια κατηγορική μεταβλητή που παίρνει τις τιμές άνδρας ή γυναίκα. Στην Εικόνα 3.2.1 παρουσιάζεται ο πίνακας συνάφειας μεταξύ των δύο ποιοτικών μεταβλητών που μελετάμε, ώστε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης στα δύο φύλα. Συνολικά, 60 συμμετέχοντες πάσχουν από παχυσαρκία, από τους οποίους οι 37 είναι άνδρες και οι 23 γυναίκες.

Επίσης, βλέπουμε ότι ικανοποιούνται οι προϋποθέσεις για τον έλεγχο ανεξαρτησίας χ^2 του Pearson, καθώς όλα τα κελιά έχουν συχνότητα μεγαλύτερη από 1 και όλες οι αναμενόμενες συχνότητες είναι μεγαλύτερες από 5.

Crosstab					
			Gender		Total
			Female	Male	
NewObesity	Insufficient_Weight	Count	17	16	33
		Expected Count	15,3	17,7	33,0
	Normal_Weight	Count	139	139	278
		Expected Count	129,0	149,0	278,0
	Overweight	Count	46	68	114
		Expected Count	52,9	61,1	114,0
	Obesity	Count	23	37	60
		Expected Count	27,8	32,2	60,0
Total	Count	225	260	485	
	Expected Count	225,0	260,0	485,0	

Εικόνα 3.2.1: Πίνακας συνάφειας φύλου και επιπέδου παχυσαρκίας

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson με τιμή 5,043 και $p\text{-value} > 0,05$ όπως φαίνονται στον Πίνακα 3.2.1, δεν απορρίπτει τη μηδενική υπόθεση σε επίπεδο σημαντικότητας 5%. Δηλαδή, το φύλο και το επίπεδο παχυσαρκίας είναι ανεξάρτητα και άρα δεν υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ τους.

Chi-Square Tests		
	Value	Asymptotic Significance (2-sided)
Pearson Chi-Square	5,043	,169
Likelihood Ratio	5,072	,167

Πίνακας 3.2.1: Έλεγχος ανεξαρτησίας φύλου και επιπέδου παχυσαρκίας

3.2.2 Συσχέτιση με τη μεταβλητή «Ηλικία»

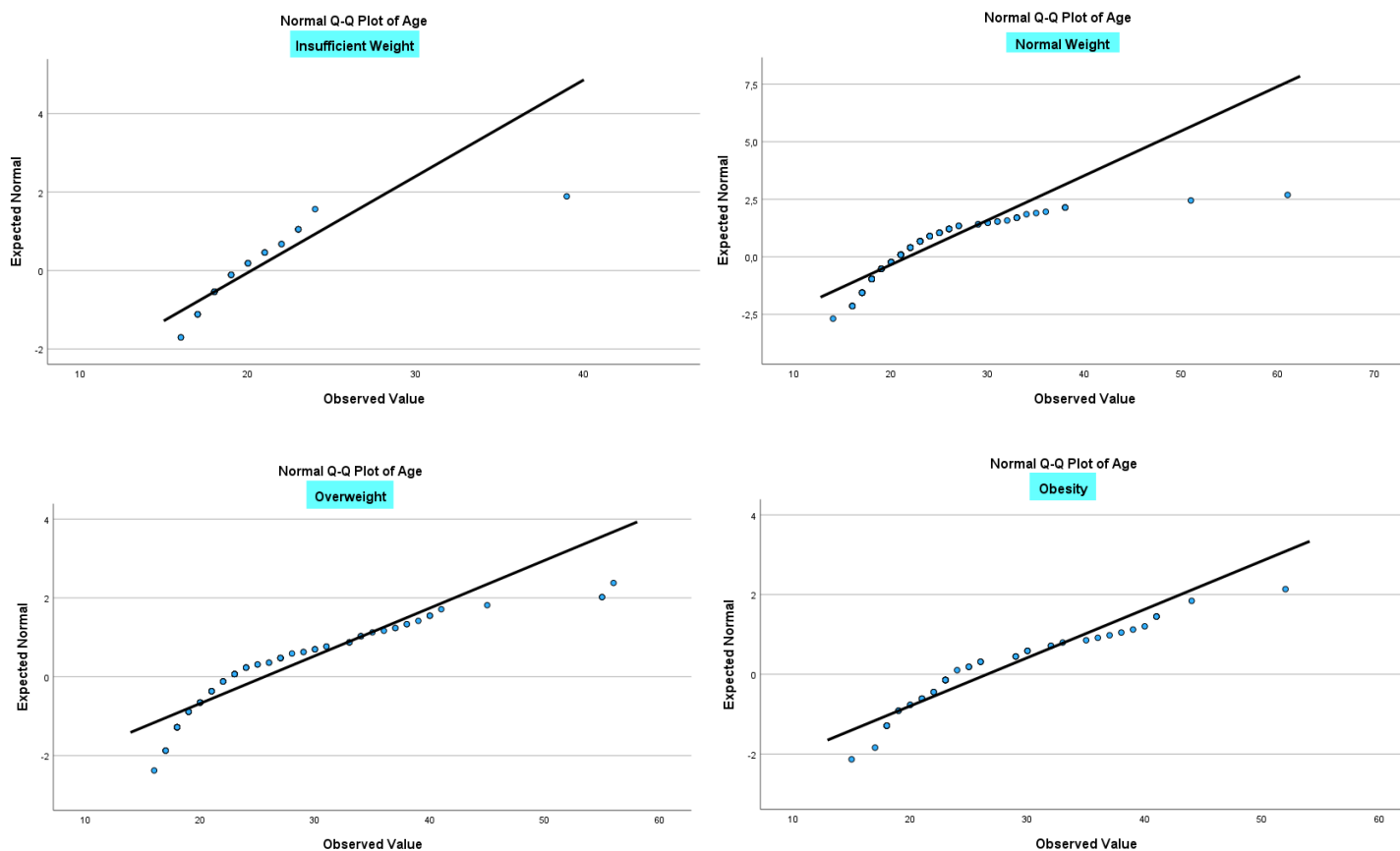
Σκοπός της παραγράφου είναι να ερευνήσουμε την ύπαρξη συσχέτισης μεταξύ της μεταβλητής της ηλικίας και της μεταβλητής που φανερώνει τα επίπεδα παχυσαρκίας. Η ηλικία είναι συνεχής μεταβλητή, οπότε θα κάνουμε πρώτα έλεγχο κανονικότητας. Θα πραγματοποιήσουμε τον έλεγχο Kolmogorov-Smirnov προκειμένου να αποφασίσουμε αν θα χρησιμοποιήσουμε παραμετρικό ή μη-παραμετρικό έλεγχο.

Tests of Normality

		Kolmogorov-Smirnov	
new_obesity		Statistic	p-value
Age	Insufficient_Weight	,186	,005
	Normal_Weight	,205	,000
	Overweight	,194	,000
	Obesity	,201	,000

Πίνακας 3.2.2(α): Έλεγχος κανονικότητας Kolmogorov-Smirnov των μεταβλητών «ηλικία» και «επίπεδο παχυσαρκίας»

Στον παραπάνω πίνακα βλέπουμε ότι η υπόθεση της κανονικότητας απορρίπτεται σε επίπεδο σημαντικότητας $\alpha=5\%$ για κάθε επίπεδο παχυσαρκίας.



Εικόνα 3.2.2: Q-Q Plot γραφήματα για το επίπεδο παχυσαρκίας σε σχέση με την ηλικία

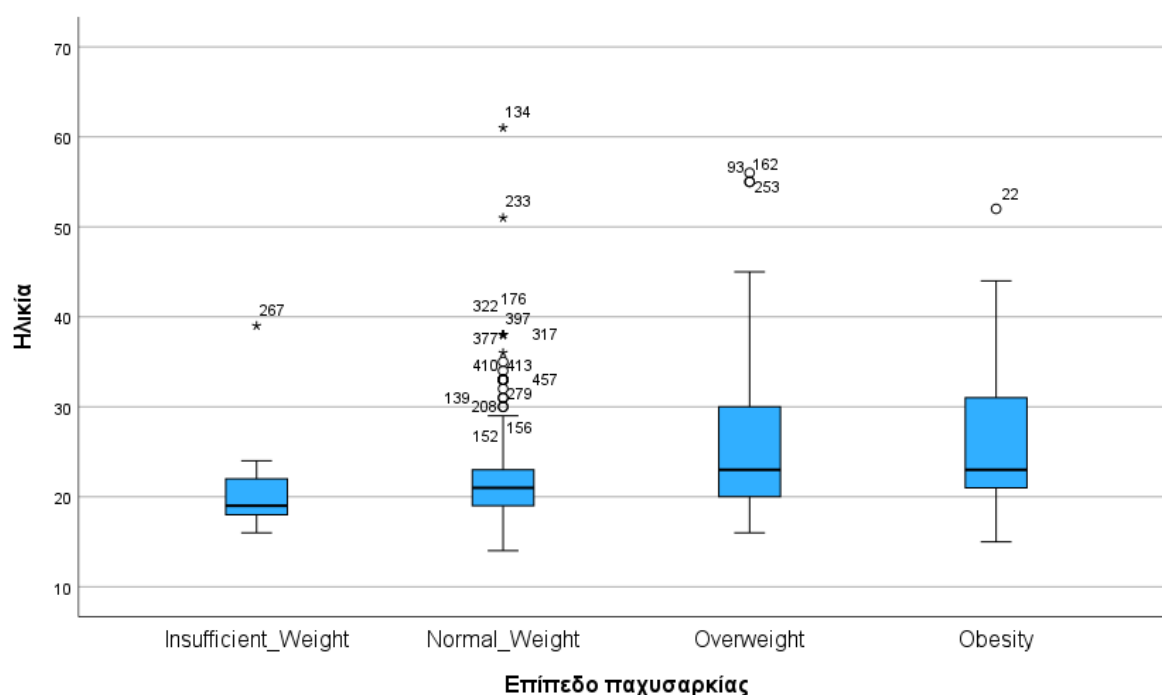
Η απόρριψη της κανονικότητας επιβεβαιώνεται και στα Q-Q Plot της Εικόνας 3.2.2, όπου φαίνεται ότι τα δεδομένα αποκλίνουν από την αναμενόμενη ευθεία της κανονικής κατανομής σε κάθε επίπεδο παχυσαρκίας. Οι αποκλίσεις παρατηρούνται κυρίως στις ακραίες τιμές, στις ουρές δηλαδή της κατανομής φαίνεται να είναι ιδιαίτερα έντονες, ειδικά σε αυτούς με κανονικό βάρος. Σε αυτή την περίπτωση λοιπόν, δεν χρειάζεται να ελέγξουμε την ομοσκεδαστικότητα, αλλά θα συνεχίσουμε με μη-παραμετρικό έλεγχο Kruskal-Wallis.

Test Statistics

	Age
Kruskal-Wallis H	44,912
df	3
Asymp. Sig.	<,001

Πίνακας 3.2.2(β): Πίνακας ελέγχου συσχετίσεων Kruskal-Wallis των μεταβλητών «ηλικία» και «επίπεδο παχυσαρκίας»

Το p-value του ελέγχου, είναι μικρότερο από 0,05 οπότε απορρίπτεται η μηδενική υπόθεση. Δηλαδή, οι κατανομές ή οι διάμεσοι της ηλικίας διαφέρουν μεταξύ των επιπέδων της παχυσαρκίας. Συνεπώς, υπάρχει συσχέτιση μεταξύ των δύο αυτών μεταβλητών.



Διάγραμμα 3.2.2: Boxplot επιπέδου παχυσαρκίας σε σχέση με την ηλικία

Στο Διάγραμμα 3.2.2 παρουσιάζεται η κατανομή της ηλικίας ανά επίπεδο παχυσαρκίας. Οι ομάδες Obesity και Overweight εμφανίζουν μεγαλύτερη διάμεσο και μεγαλύτερη διασπορά ηλικιών σε σχέση με τις ομάδες Normal και Insufficient Weight. Ενισχύεται, δηλαδή, το αποτέλεσμα του ελέγχου Kruskal-Wallis, αφού βλέπουμε την αύξηση της διάμεσης ηλικίας καθώς αυξάνεται το επίπεδο παχυσαρκίας, τεκμηριώνοντας την ύπαρξη συσχέτισης μεταξύ των δύο μεταβλητών.

3.2.3 Συσχέτιση με τη μεταβλητή «Βάρος»

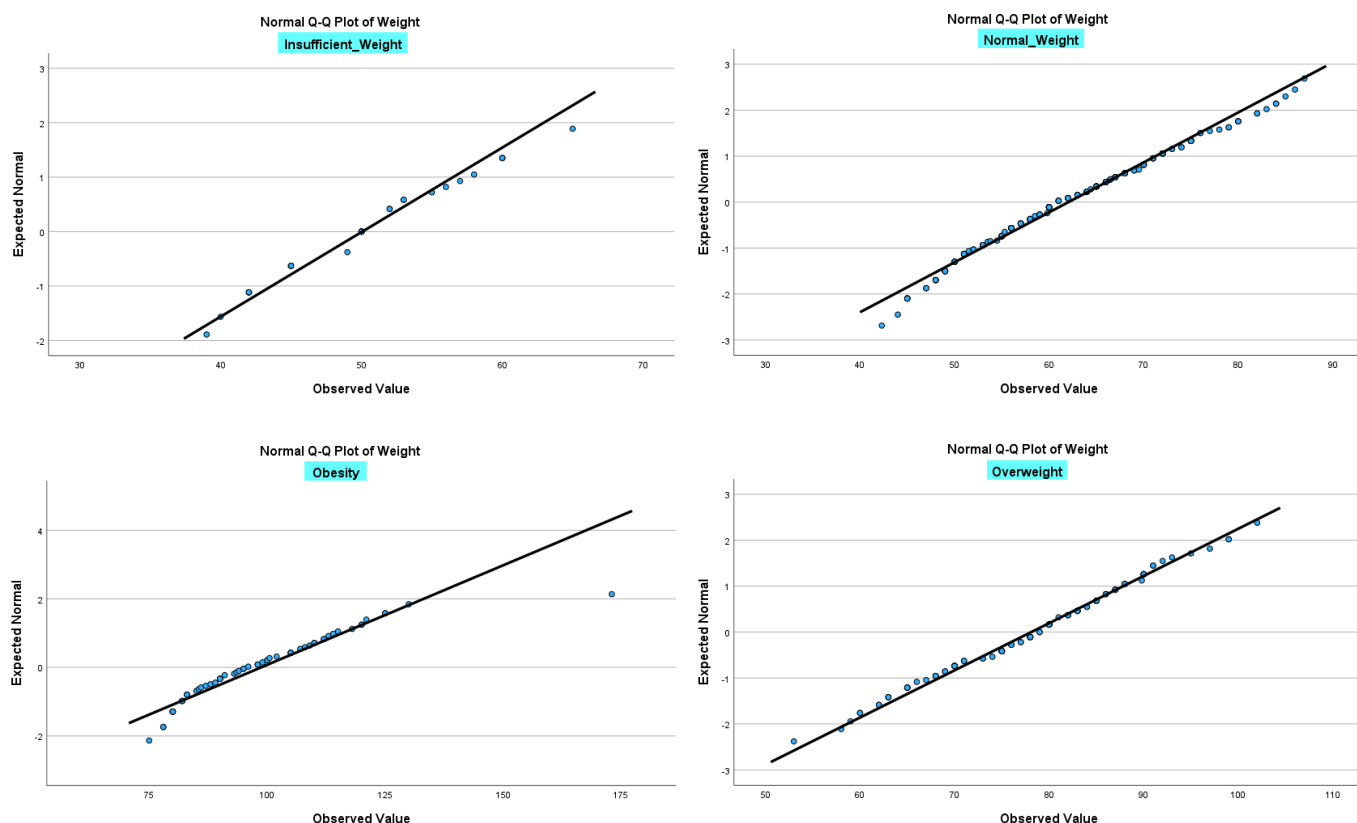
Σκοπός της παραγράφου είναι να ερευνήσουμε την ύπαρξη συσχέτισης μεταξύ της μεταβλητής του βάρους και της μεταβλητής που φανερώνει τα επίπεδα παχυσαρκίας. Το βάρος είναι συνεχής μεταβλητή, οπότε θα κάνουμε πρώτα έλεγχο κανονικότητας. Θα πραγματοποιήσουμε τον έλεγχο Kolmogorov-Smirnov προκειμένου να αποφασίσουμε αν θα χρησιμοποιήσουμε παραμετρικό ή μη-παραμετρικό έλεγχο.

Tests of Normality

		Kolmogorov-Smirnov	
new_obesity		Statistic	p-value
Weight	Insufficient_Weight	,140	,099
	Normal_Weight	,092	,000
	Overweight	,076	,099
	Obesity	,098	,200

Πίνακας 3.2.3(α): Έλεγχος κανονικότητας Kolmogorov-Smirnov των μεταβλητών «βάρος» και «επίπεδο παχυσαρκίας»

Στον παραπάνω πίνακα, βλέπουμε ότι η υπόθεση της κανονικότητας απορρίπτεται για όσους έχουν κανονικό βάρος, σε επίπεδο σημαντικότητας 5%.



Εικόνα 3.2.3: Q-Q Plot γραφήματα για το επίπεδο παχυσαρκίας σε σχέση με το βάρος

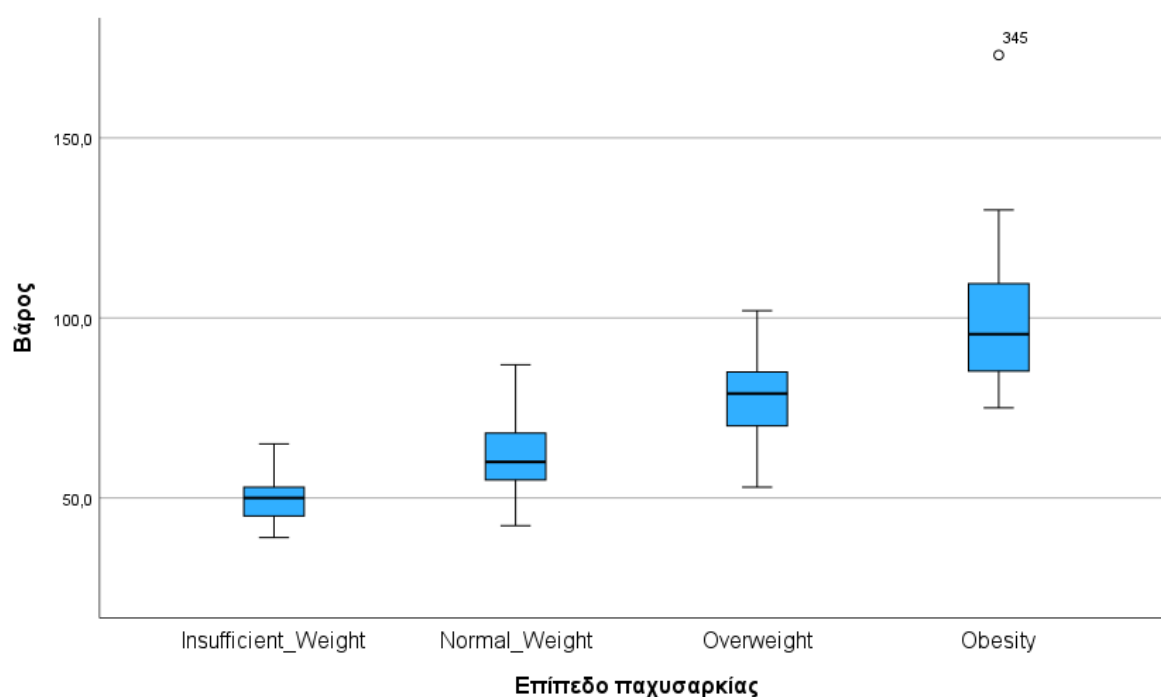
Η απόρριψη της κανονικότητας επιβεβαιώνεται και στο Q-Q Plot της Εικόνας 3.2.3 για τους κανονικούς σε βάρος, όπου φαίνεται ότι τα δεδομένα αποκλίνουν από την αναμενόμενη ευθεία της κανονικής κατανομής. Οι αποκλίσεις παρατηρούνται κυρίως στις ακραίες τιμές, στις ουρές δηλαδή της κατανομής. Οπότε, σε αυτή την περίπτωση, δεν χρειάζεται να ελέγξουμε την ομοσκεδαστικότητα, αλλά θα συνεχίσουμε με μη-παραμετρικό έλεγχο Kruskal-Wallis.

Test Statistics

Weight	
Kruskal-Wallis H	283,367
df	3
Asymp. Sig.	<,001

Πίνακας 3.2.3(β): Πίνακας ελέγχου συσχετίσεων Kruskal-Wallis των μεταβλητών «βάρος» και «επίπεδο παχυσαρκίας»

Το p-value του ελέγχου, είναι μικρότερο από 0,05 οπότε απορρίπτεται η μηδενική υπόθεση. Δηλαδή, οι κατανομές ή οι διάμεσοι του βάρους διαφέρουν μεταξύ των επιπέδων της παχυσαρκίας. Συνεπώς, υπάρχει συσχέτιση μεταξύ των δύο αυτών μεταβλητών.



Διάγραμμα 3.2.3: Boxplot επιπέδου παχυσαρκίας σε σχέση με το βάρος

Στο Διάγραμμα 3.2.3 παρουσιάζεται η κατανομή του βάρους ανά επίπεδο παχυσαρκίας. Οι ομάδες Obesity και Overweight εμφανίζουν μεγαλύτερη διάμεσο και μεγαλύτερη διασπορά βάρους σε σχέση με τις ομάδες Normal και Insufficient Weight. Ενισχύεται, δηλαδή, το αποτέλεσμα του ελέγχου Kruskal-Wallis, αφού βλέπουμε την αύξηση του διάμεσου βάρους καθώς αυξάνεται το επίπεδο παχυσαρκίας, τεκμηριώνοντας την ύπαρξη συσχέτισης μεταξύ των δύο μεταβλητών.

3.2.4 Συσχέτιση με τη μεταβλητή «Ύψος»

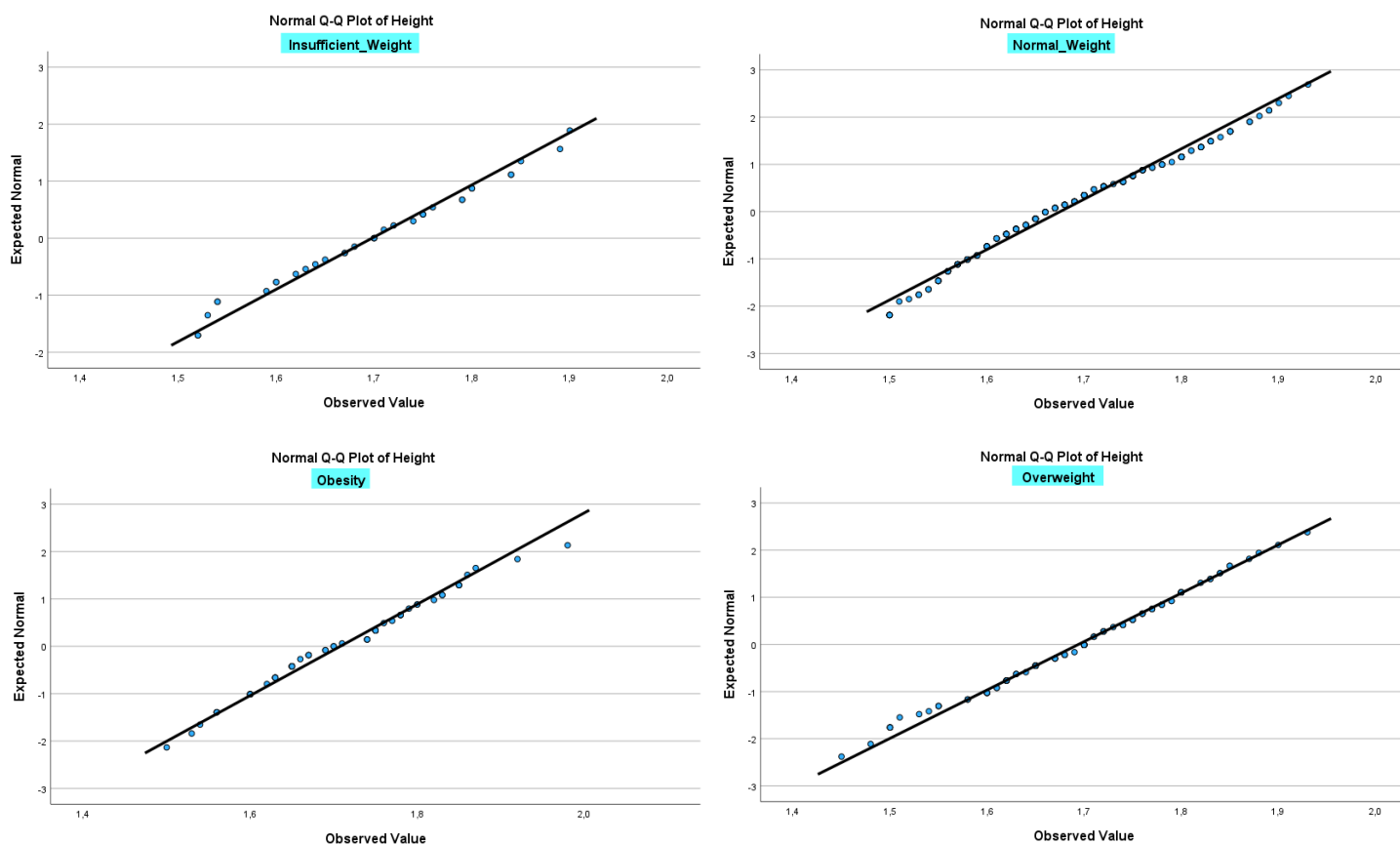
Σκοπός της παραγράφου είναι να ερευνήσουμε την ύπαρξη συσχέτισης μεταξύ της μεταβλητής του ύψους και της μεταβλητής που φανερώνει τα επίπεδα παχυσαρκίας. Το ύψος είναι συνεχής μεταβλητή, οπότε θα κάνουμε πρώτα έλεγχο κανονικότητας. Θα πραγματοποιήσουμε τον έλεγχο Kolmogorov-Smirnov προκειμένου να αποφασίσουμε αν θα χρησιμοποιήσουμε παραμετρικό ή μη-παραμετρικό έλεγχο.

Tests of Normality

		Kolmogorov-Smirnov	
new_obesity		Statistic	p-value
Height	Insufficient_Weight	,078	,200
	Normal_Weight	,082	,000
	Overweight	,085	,040
	Obesity	,095	,200

Πίνακας 3.2.4(α): Έλεγχος κανονικότητας Kolmogorov-Smirnov των μεταβλητών «ύψος» και «επίπεδο παχυσαρκίας»

Στον παραπάνω πίνακα βλέπουμε ότι η υπόθεση της κανονικότητας απορρίπτεται για όσους έχουν κανονικό βάρος και τους υπέρβαρους σε επίπεδο σημαντικότητας 5%.



Εικόνα 3.2.4: Q-Q Plot γραφήματα για το επίπεδο παχυσαρκίας σε σχέση με το ύψος

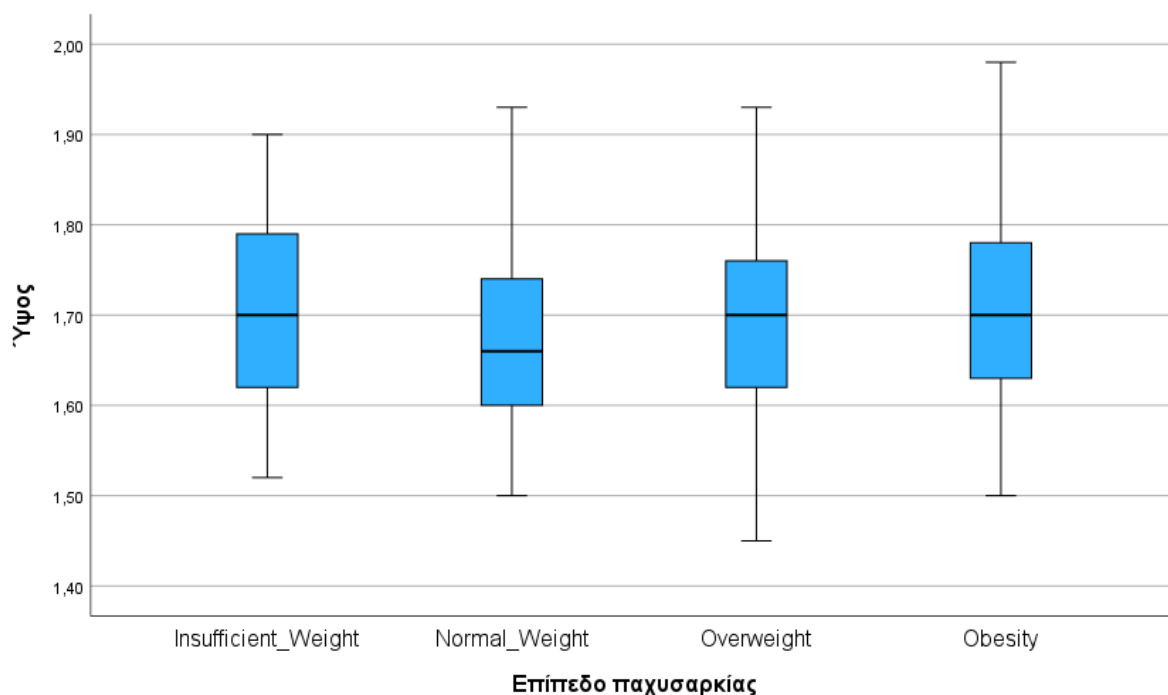
Η απόρριψη της κανονικότητας επιβεβαιώνεται και στο Q-Q Plot της Εικόνας 3.2.4 για τους κανονικούς σε βάρος και τους υπέρβαρους, όπου φαίνεται ότι τα δεδομένα αποκλίνουν από την αναμενόμενη ευθεία της κανονικής κατανομής. Οι αποκλίσεις παρατηρούνται κυρίως στις ακραίες τιμές, στις ουρές δηλαδή της κατανομής. Οπότε, σε αυτή την περίπτωση, δεν χρειάζεται να ελέγξουμε την ομοσκεδαστικότητα, αλλά θα συνεχίσουμε με μη-παραμετρικό έλεγχο Kruskal-Wallis.

Test Statistics

	Height
Kruskal-Wallis H	8,444
df	3
Asymp. Sig.	,038

Πίνακας 3.2.4(β): Πίνακας ελέγχου συσχετίσεων Kruskal-Wallis των μεταβλητών «ύψος» και «επίπεδο παχυσαρκίας»

Στον Πίνακα 3.2.4(β) βλέπουμε ότι $p\text{-value}=0,038$, οπότε απορρίπτεται η μηδενική υπόθεση. Δηλαδή, οι κατανομές ή οι διάμεσοι του ύψους διαφέρουν μεταξύ των επιπέδων της παχυσαρκίας. Συνεπώς, υπάρχει συσχέτιση μεταξύ των δύο αυτών μεταβλητών.



Διάγραμμα 3.2.4: Boxplot επιπέδου παχυσαρκίας σε σχέση με το ύψος

Στο Διάγραμμα 3.2.4 παρουσιάζεται η κατανομή του ύψους ανά επίπεδο παχυσαρκίας. Παρά το γεγονός ότι οι διάμεσοι και τα διατεταρτημοριακά εύρη φαίνονται αρκετά παρόμοια, το αποτέλεσμα του ελέγχου Kruskal-Wallis υποδηλώνει ότι το ύψος ενδέχεται να διαφέρει στατιστικά σημαντικά μεταξύ των επιπέδων παχυσαρκίας. Ωστόσο, η διαφορά φαίνεται να είναι μικρής κλίμακας, όπως φαίνεται και από τη σημαντική επικάλυψη των boxplot.

3.2.5 Συσχέτιση με τη μεταβλητή «Οικογενειακό ιστορικό παχυσαρκίας»

Η μεταβλητή «οικογενειακό ιστορικό παχυσαρκίας» είναι μια κατηγορική μεταβλητή που παίρνει τις τιμές ναι ή όχι. Στην Εικόνα 3.2.5 παρουσιάζεται ο πίνακας συνάφειας μεταξύ των δύο ποιοτικών μεταβλητών που μελετάμε, ώστε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης σε σχέση με το αν είχαν οικογενειακό ιστορικό παχυσαρκίας. Συνολικά, 60 συμμετέχοντες πάσχουν από παχυσαρκία, από τους οποίους οι 52 είχαν οικογενειακό ιστορικό παχυσαρκίας ενώ οι 8 όχι.

Επίσης, βλέπουμε ότι ικανοποιούνται οι προϋποθέσεις για τον έλεγχο ανεξαρτησίας χ^2 του Pearson, καθώς όλα τα κελιά έχουν συχνότητα μεγαλύτερη από 1 και όλες οι αναμενόμενες συχνότητες είναι μεγαλύτερες από 5.

Crosstab					
			family_history_with_overweight		Total
			no	yes	
NewObesity	Insufficient_Weight	Count	18	15	33
		Expected Count	12,9	20,1	33,0
	Normal_Weight	Count	126	152	278
		Expected Count	108,9	169,1	278,0
	Overweight	Count	38	76	114
		Expected Count	44,7	69,3	114,0
	Obesity	Count	8	52	60
		Expected Count	23,5	36,5	60,0
Total		Count	190	295	485
		Expected Count	190,0	295,0	485,0

Εικόνα 3.2.5: Πίνακας συνάφειας οικογενειακού ιστορικού παχυσαρκίας και επιπέδου παχυσαρκίας

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson με τιμή 26,13 και $p\text{-value} < 0,05$ όπως φαίνονται στον Πίνακα 3.2.5(α), απορρίπτει τη μηδενική υπόθεση σε επίπεδο σημαντικότητας 5%. Δηλαδή, το οικογενειακό ιστορικό παχυσαρκίας και το επίπεδο παχυσαρκίας δεν είναι ανεξάρτητα και άρα υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ τους.

Chi-Square Tests		
	Value	Asymptotic Significance (2-sided)
Pearson Chi-Square	26,130	<,001
Likelihood Ratio	28,764	<,001

Πίνακας 3.2.5(α): Έλεγχος οικογενειακού ιστορικού παχυσαρκίας και επιπέδου παχυσαρκίας

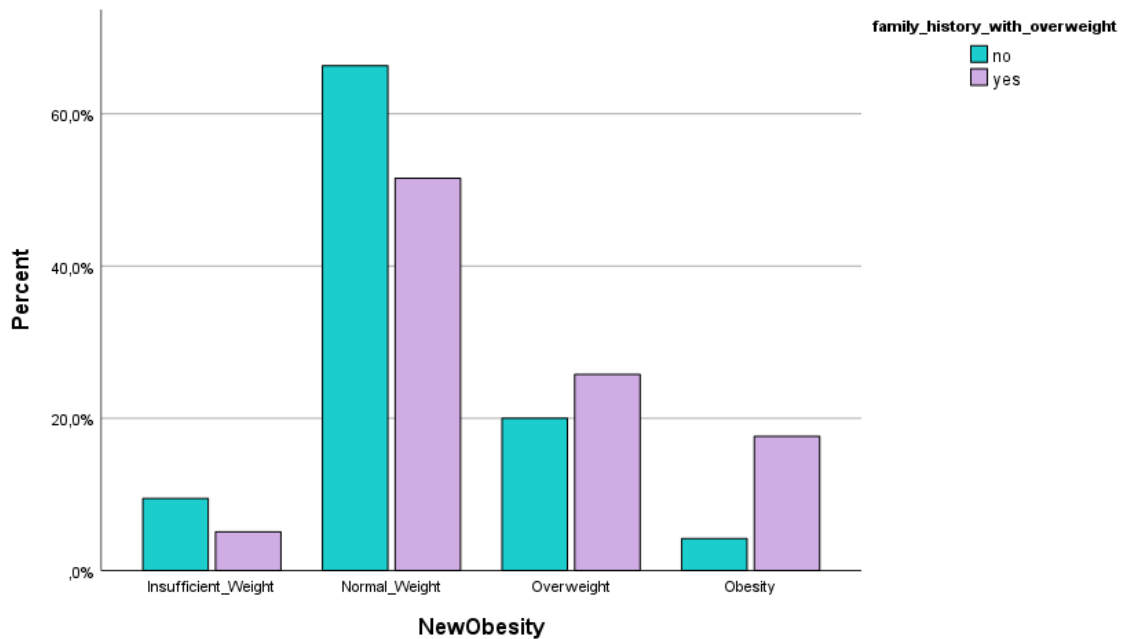
Για να δούμε πόσο ισχυρή είναι η συσχέτιση, θα ελέγξουμε την τιμή V του Cramer. Στον παρακάτω πίνακα βλέπουμε ότι η τιμή αυτή είναι 0,232 που σημαίνει ότι υπάρχει μέτριας έντασης συσχέτιση.

Symmetric Measures

		Value
Nominal by Nominal	Phi	,232
	Cramer's V	,232

Πίνακας 3.2.5(β): Ένταση συσχέτισης οικογενειακού ιστορικού παχυσαρκίας και επιπέδου παχυσαρκίας

Στο ραβδόγραμμα απεικονίζεται η σχέση μεταξύ του ιστορικού παχυσαρκίας στην οικογένεια και του επιπέδου παχυσαρκίας. Παρατηρούμε ότι τα άτομα με ιστορικό παχυσαρκίας στην οικογένεια εμφανίζουν υψηλότερα ποσοστά στις κατηγορίες Overweight και Obesity. Αντίθετα, τα άτομα χωρίς ιστορικό εμφανίζουν υψηλότερα ποσοστά στις κατηγορίες Insufficient και Normal Weight. Αυτό μπορεί να σημαίνει ότι η απουσία οικογενειακού ιστορικού παχυσαρκίας συνδέεται με μεγαλύτερη πιθανότητα διατήρησης φυσιολογικού βάρους, ενώ στην αντίθετη περίπτωση, η γενετική ή κοινή οικογενειακή συμπεριφορά (διατροφή, τρόπος ζωής) μπορεί να αυξάνει τον κίνδυνο εμφάνισης της νόσου.



Διάγραμμα 3.2.5: Ραβδόγραμμα συσχέτισης οικογενειακού ιστορικού παχυσαρκίας με το επίπεδο παχυσαρκίας

3.3 Συσχετίσεις διατροφικών συνηθειών με το «Επίπεδο παχυσαρκίας»

Σε αυτή την παράγραφο θα ελέγξουμε τη σχέση εξάρτησης που μπορεί να υπάρχει ανάμεσα στις διατροφικές συνήθειες, όπως παρουσιάστηκαν στο δεύτερο κεφάλαιο, και στη μεταβλητή της παχυσαρκίας.

3.3.1 Συσχέτιση με τη μεταβλητή «FAVC»

Η μεταβλητή «κατανάλωση τροφής με πολλές θερμίδες» είναι μια κατηγορική μεταβλητή που παίρνει τις τιμές ναι ή όχι. Στην Εικόνα 3.3.1 παρουσιάζεται ο πίνακας συνάφειας μεταξύ των δύο ποιοτικών μεταβλητών που μελετάμε, ώστε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης σε σχέση με την κατανάλωση τροφής με πολλές θερμίδες. Συνολικά, 60 συμμετέχοντες πάσχουν από παχυσαρκία, από τους οποίους οι 44 καταναλώνουν τροφή με πολλές θερμίδες ενώ οι 16 όχι.

Επίσης, βλέπουμε ότι ικανοποιούνται οι προϋποθέσεις για τον έλεγχο ανεξαρτησίας χ^2 του Pearson, καθώς όλα τα κελιά έχουν συχνότητα μεγαλύτερη από 1 και όλες οι αναμενόμενες συχνότητες είναι μεγαλύτερες από 5.

Crosstab					
			FAVC		Total
			no	yes	
NewObesity	Insufficient_Weight	Count	13	20	33
		Expected Count	10,1	22,9	33,0
	Normal_Weight	Count	78	200	278
		Expected Count	84,8	193,2	278,0
	Overweight	Count	41	73	114
		Expected Count	34,8	79,2	114,0
	Obesity	Count	16	44	60
		Expected Count	18,3	41,7	60,0
Total		Count	148	337	485
		Expected Count	148,0	337,0	485,0

Εικόνα 3.3.1: Πίνακας συνάφειας κατανάλωσης τροφής με πολλές θερμίδες και επιπέδου παχυσαρκίας

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson με τιμή 4,035 και p-value=0,258 (μεγαλύτερο από 0,05) όπως φαίνονται στον Πίνακα 3.3.1, δεν απορρίπτει τη μηδενική υπόθεση σε επίπεδο σημαντικότητας 5%. Συνεπώς, η συχνότητα κατανάλωσης τροφής με πολλές θερμίδες και το επίπεδο παχυσαρκίας είναι ανεξάρτητα και άρα δεν υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ τους.

Chi-Square Tests		
	Value	Asymptotic Significance (2-sided)
Pearson Chi-Square	4,035	,258
Likelihood Ratio	3,955	,266

Πίνακας 3.3.1: Έλεγχος ανεξαρτησίας κατανάλωσης τροφής με πολλές θερμίδες και επιπέδου παχυσαρκίας

3.3.2 Συσχέτιση με τη μεταβλητή «FCVC»

Η μεταβλητή «συχνότητα κατανάλωσης λαχανικών μέσα στα γεύματα» είναι μια κατηγορική μεταβλητή. Στην Εικόνα 3.3.2 παρουσιάζεται ο πίνακας συνάφειας μεταξύ των δύο ποιοτικών μεταβλητών που μελετάμε, ώστε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης σε σχέση με τη συχνότητα κατανάλωσης λαχανικών μέσα στα γεύματα. Συνολικά, 60 συμμετέχοντες πάσχουν από παχυσαρκία, από τους οποίους οι 4 δεν καταναλώνουν ποτέ λαχανικά μέσα στα γεύματα, οι 35 μερικές φορές και οι 21 πάντα.

Επίσης, βλέπουμε ότι ικανοποιούνται οι προϋποθέσεις για τον έλεγχο ανεξαρτησίας χ^2 του Pearson, καθώς όλα τα κελιά έχουν συχνότητα μεγαλύτερη από 1 και πάνω από το 80% των κελιών έχουν αναμενόμενη συχνότητα μεγαλύτερη από 5.

Crosstab						
			FCVC			
			1	2	3	Total
NewObesity	Insufficient_Weight	Count	3	11	19	33
		Expected Count	2,1	18,0	12,9	33,0
	Normal_Weight	Count	17	150	111	278
		Expected Count	17,8	151,9	108,3	278,0
	Overweight	Count	7	69	38	114
		Expected Count	7,3	62,3	44,4	114,0
	Obesity	Count	4	35	21	60
		Expected Count	3,8	32,8	23,4	60,0
Total		Count	31	265	189	485
		Expected Count	31,0	265,0	189,0	485,0

Εικόνα 3.3.2: Πίνακας συνάφειας συχνότητας κατανάλωσης λαχανικών και επιπέδου παχυσαρκίας

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson με τιμή 8,235 και p-value=0,221 όπως φαίνονται στον Πίνακα 3.3.2, δεν απορρίπτει τη μηδενική υπόθεση σε επίπεδο σημαντικότητας 5%. Δηλαδή, η συχνότητα κατανάλωσης λαχανικών στα γεύματα και το επίπεδο παχυσαρκίας είναι ανεξάρτητα, οπότε δεν υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ τους,.

Chi-Square Tests		
	Value	Asymptotic Significance (2-sided)
Pearson Chi-Square	8,235 ^a	,221
Likelihood Ratio	8,289	,218

a. 2 cells (16,7%) have expected count less than 5. The minimum expected count is 2,11.

Πίνακας 3.3.2: Έλεγχος ανεξαρτησίας συχνότητας κατανάλωσης λαχανικών μέσα στα γεύματα και επιπέδου παχυσαρκίας

3.3.3 Συσχέτιση με τη μεταβλητή «NCP»

Η μεταβλητή «ποσότητα καθημερινής κατανάλωσης κύριων γευμάτων» είναι μια κατηγορική μεταβλητή. Στην Εικόνα 3.3.3 παρουσιάζεται ο πίνακας συνάφειας μεταξύ των δύο ποιοτικών μεταβλητών που μελετάμε, ώστε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης ανάλογα με την ποσότητα κύριων γευμάτων που καταναλώνουν οι πάσχοντες καθημερινά. Συνολικά, 60 συμμετέχοντες πάσχουν από παχυσαρκία, από τους οποίους οι 16 καταναλώνουν 1 με 2 κύρια γεύματα την ημέρα, οι 43 καταναλώνουν 3 και ένας καταναλώνει πάνω από 3 γεύματα.

Επίσης, βλέπουμε ότι ικανοποιούνται οι προϋποθέσεις για τον έλεγχο ανεξαρτησίας χ^2 του Pearson, καθώς όλα τα κελιά έχουν συχνότητα μεγαλύτερη από 1 και πάνω από το 80% των κελιών έχουν αναμενόμενη συχνότητα μεγαλύτερη από 5.

Crosstab						
			NCP			
			1	3	4	Total
NewObesity	Insufficient_Weight	Count	5	20	8	33
		Expected Count	7,2	22,7	3,1	33,0
	Normal_Weight	Count	50	199	29	278
		Expected Count	60,8	190,9	26,4	278,0
	Overweight	Count	35	71	8	114
		Expected Count	24,9	78,3	10,8	114,0
	Obesity	Count	16	43	1	60
		Expected Count	13,1	41,2	5,7	60,0
Total		Count	106	333	46	485
		Expected Count	106,0	333,0	46,0	485,0

Εικόνα 3.3.3: Πίνακας συνάφειας ποσότητας καθημερινής κατανάλωσης κύριων γευμάτων και επιπέδου παχυσαρκίας

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson με τιμή 21,152 και p-value=0,002 (μικρότερο από 0,05) όπως φαίνονται στον Πίνακα 3.3.3(α), απορρίπτει τη μηδενική υπόθεση σε επίπεδο σημαντικότητας 5%. Δηλαδή, ο αριθμός των καθημερινών κυρίως γευμάτων και το επίπεδο παχυσαρκίας δεν είναι ανεξάρτητα, οπότε υπάρχει συσχέτιση μεταξύ τους.

Chi-Square Tests

	Value	Asymptotic Significance (2- sided)
Pearson Chi-Square	21,152 ^a	,002
Likelihood Ratio	20,684	,002

a. 1 cell (8,3%) has expected count less than 5. The minimum expected count is 3,13.

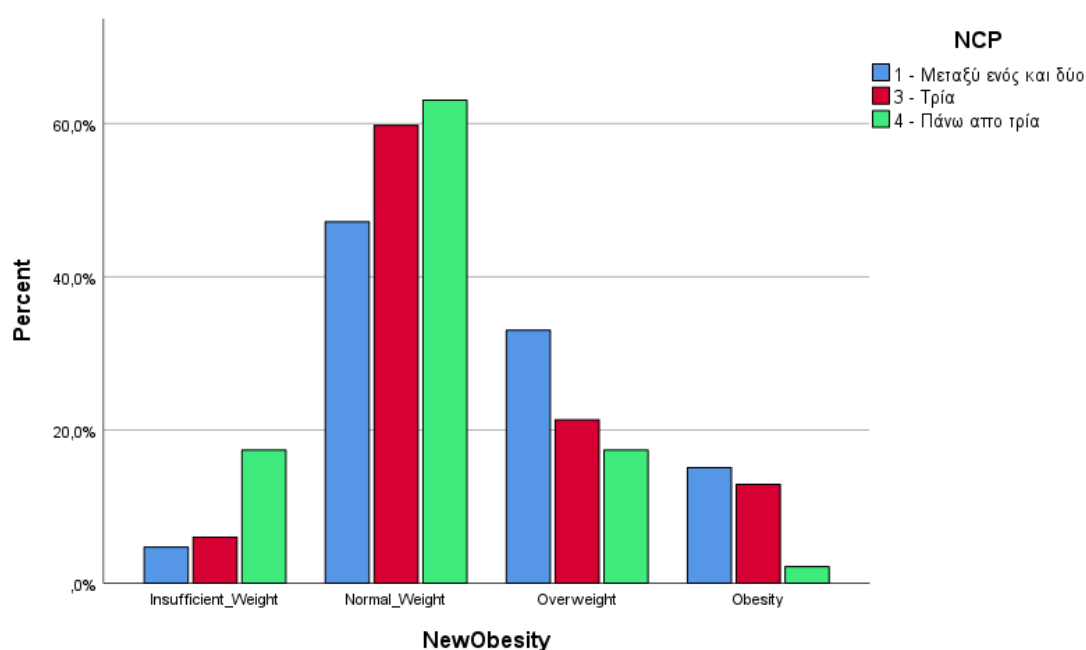
Πίνακας 3.3.3(α): Έλεγχος ανεξαρτησίας ποσότητας καθημερινής κατανάλωσης κύριων γευμάτων και επιπέδου παχυσαρκίας

Για να δούμε πόσο ισχυρή είναι η συσχέτιση, θα ελέγξουμε την τιμή V του Cramer. Στον παρακάτω πίνακα βλέπουμε ότι η τιμή αυτή είναι 0,148 που σημαίνει ότι υπάρχει μικρής έντασης συσχέτιση.

Symmetric Measures		
		Value
Nominal by Nominal	Phi	,209
	Cramer's V	,148

Πίνακας 3.3.3(β): Ένταση συσχέτισης ποσότητας καθημερινής κατανάλωσης κύριων γευμάτων και επιπέδου παχυσαρκίας

Το παρακάτω ραβδόγραμμα εξετάζει τη σχέση μεταξύ του αριθμού των κύριων γευμάτων και της κατανομής του βάρους σε διάφορες κατηγορίες. Βλέπουμε ότι άτομα που καταναλώνουν ένα με δύο γεύματα την ημέρα εμφανίζουν υψηλότερα ποσοστά στις κατηγορίες Overweight ή Obesity, ενώ όσοι καταναλώνουν τρία ή παραπάνω γεύματα την ημέρα συνδέονται περισσότερο με τις κατηγορίες Insufficient και Normal Weight. Αυτή η συσχέτιση των δύο μεταβλητών, μπορεί να σημαίνει ότι εφόσον η κατανάλωση είναι κατανομημένη και ισορροπημένη, υπάρχει μικρότερος κίνδυνος εμφάνισης της νόσου.



Διάγραμμα 3.3.3: Ραβδόγραμμα συσχέτισης της ποσότητας καθημερινής κατανάλωσης κύριων γευμάτων με το επίπεδο παχυσαρκίας

3.3.4 Συσχέτιση με τη μεταβλητή «CAEC»

Η μεταβλητή «συχνότητα κατανάλωσης φαγητού ανάμεσα στα γεύματα» είναι μια κατηγορική μεταβλητή. Στην Εικόνα 3.3.4 παρουσιάζεται ο πίνακας συνάφειας μεταξύ των δύο ποιοτικών μεταβλητών που μελετάμε, ώστε να έχουμε μια πρώτη εικόνα από τις συχνοτήτες εμφάνισης της πάθησης ανάλογα με τη συχνότητα κατανάλωσης φαγητού

ανάμεσα στα γεύματα. Συνολικά, 60 συμμετέχοντες πάσχουν από παχυσαρκία, από τους οποίους οι 8 καταναλώνουν πάντα φαγητό ανάμεσα στα γεύματα, άλλοι 8 συχνά, οι 42 μερικές φορές και οι 2 ποτέ.

Επίσης, βλέπουμε ότι δεν ικανοποιούνται οι προϋποθέσεις για τον έλεγχο ανεξαρτησίας χ^2 του Pearson, καθώς όλα τα κελιά έχουν συχνότητα μεγαλύτερη από 1 αλλά πάνω από 20% των κελιών έχουν συχνότητα μικρότερη από 5. Οπότε, σε αυτή την περίπτωση, θα χρησιμοποιήσουμε τη μέθοδο Monte Carlo.

Crosstab							
			CAEC				Total
			Always	Frequently	Sometimes	no	
NewObesity	Insufficient_Weight	Count	2	16	12	3	33
		Expected Count	3,5	9,1	19,1	1,3	33,0
	Normal_Weight	Count	33	82	153	10	278
		Expected Count	29,2	76,8	161,1	10,9	278,0
	Overweight	Count	8	28	74	4	114
		Expected Count	12,0	31,5	66,0	4,5	114,0
	Obesity	Count	8	8	42	2	60
		Expected Count	6,3	16,6	34,8	2,4	60,0
Total		Count	51	134	281	19	485
		Expected Count	51,0	134,0	281,0	19,0	485,0

Εικόνα 3.3.4: Πίνακας συνάφειας συχνότητας κατανάλωσης τροφής μεταξύ των γευμάτων και επιπέδου παχυσαρκίας

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson ανέδειξε τιμή 21,207 όπως φαίνεται στον Πίνακα 3.3.4(α). Δεδομένου ότι το p-value από τη μέθοδο Monte Carlo είναι μικρότερο από 0,05, απορρίπτεται η μηδενική υπόθεση στο επίπεδο σημαντικότητας 5%. Συνεπώς, η συχνότητα κατανάλωσης φαγητού ανάμεσα στα γεύματα και το επίπεδο παχυσαρκίας δεν είναι ανεξάρτητα, γεγονός που υποδηλώνει την ύπαρξη στατιστικά σημαντικής συσχέτισης μεταξύ τους.

Chi-Square Tests					
	Value	Asymptotic Significance (2- sided)	Monte Carlo Sig. (2-sided)		
			Significance	95% Confidence Interval Lower Bound	Upper Bound
Pearson Chi-Square	21,207 ^a	,012	,013 ^b	,011	,015
Likelihood Ratio	21,219	,012	,016 ^b	,014	,019
Fisher-Freeman-Halton Exact Test	21,037		,009 ^b	,007	,010

a. 4 cells (25,0%) have expected count less than 5. The minimum expected count is 1,29.

b. Based on 10000 sampled tables with starting seed 334431365.

Πίνακας 3.3.4(α): Έλεγχος ανεξαρτησίας συχνότητας κατανάλωσης τροφής μεταξύ των γευμάτων και επιπέδου παχυσαρκίας με μέθοδο Monte Carlo

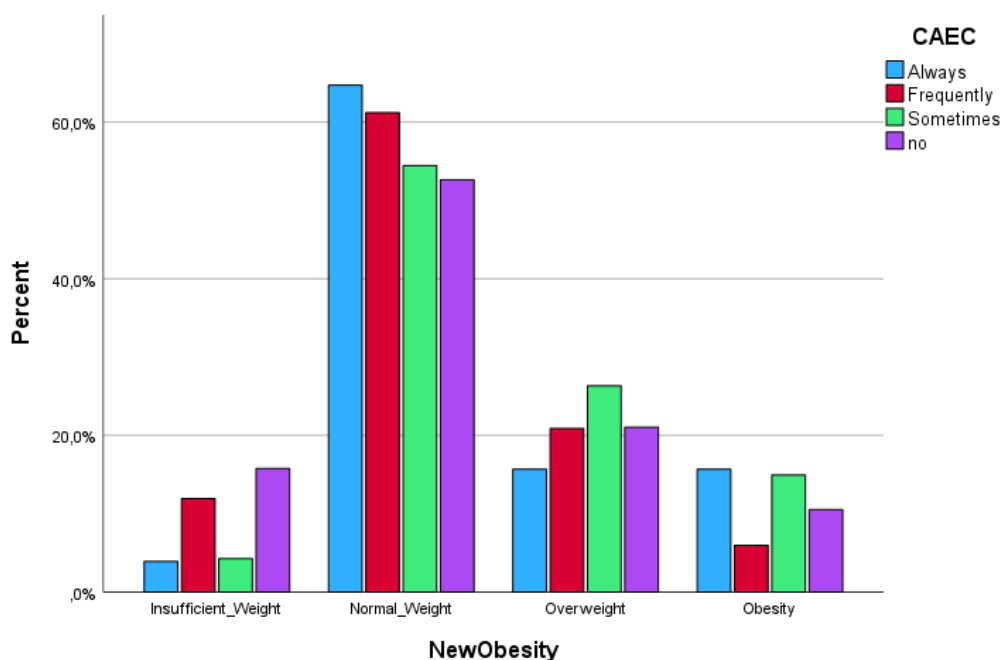
Για να δούμε πόσο ισχυρή είναι η συσχέτιση, θα ελέγξουμε την τιμή V του Cramer. Στον παρακάτω πίνακα βλέπουμε ότι η τιμή αυτή είναι 0,121 που σημαίνει ότι υπάρχει μικρής έντασης συσχέτιση.

Symmetric Measures			
		Value	Approximate Significance
Nominal by Nominal	Phi	,209	,012
	Cramer's V	,121	,012

Πίνακας 3.3.4(β): Ένταση συσχέτισης συχνότητας κατανάλωσης τροφής μεταξύ των γευμάτων και επιπέδου παχυσαρκίας

Το Διάγραμμα 3.3.4 αναλύει τη σχέση μεταξύ των μεταβλητών της συχνότητας κατανάλωσης τροφής ανάμεσα στα γεύματα και του επιπέδου παχυσαρκίας. Παρατηρούμε ότι στην κατηγορία Insufficient Weight οι περισσότεροι άνθρωποι δεν καταναλώνουν ποτέ τροφή ανάμεσα στα γεύματα, ενώ στις κατηγορίες Normal και Obesity η πλειοψηφία καταναλώνει πάντα.

Αυτό πιθανώς να υποδηλώνει ότι η κατανάλωση τροφής ανάμεσα στα γεύματα, γνωστό και ως “snacking”, μπορεί να έχει διαφορετική επίδραση στην υγεία ενός ανθρώπου, ανάλογα με το είδος της τροφής που καταναλώνει ο καθένας και φυσικά και το μέγεθος της μερίδας. Από την άλλη, η έλλειψη κατανάλωσης ενδιάμεσων γευμάτων μπορεί να είναι προστατευτικός παράγοντας όπως για αυτούς που έχουν κανονικό βάρος, αλλά θα μπορούσε να είναι και ζημιογόνος, όπως στην περίπτωση των λιποβαρών.



Διάγραμμα 3.3.4: Ραβδόγραμμα συσχέτισης της συχνότητας κατανάλωσης τροφής μεταξύ των γευμάτων με το επίπεδο παχυσαρκίας

3.3.5 Συσχέτιση με τη μεταβλητή «CH2O»

Η μεταβλητή «ποσότητα καθημερινής κατανάλωσης νερού» είναι κατηγορική. Στην Εικόνα 3.3.5 παρουσιάζεται ο πίνακας συνάφειας μεταξύ των δύο ποιοτικών μεταβλητών που μελετάμε, ώστε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης ανάλογα με την ποσότητα νερού που καταναλώνουν καθημερινά οι πάσχοντες. Συνολικά, 60 συμμετέχοντες πάσχουν από παχυσαρκία, από τους οποίους οι 17 πίνουν λιγότερο από 1 λίτρο την ημέρα, οι 25 πίνουν 1 με 2 λίτρα την ημέρα και οι 18 πίνουν πάνω από 2 λίτρα καθημερινά.

Επίσης, βλέπουμε ότι ικανοποιούνται οι προϋποθέσεις για τον έλεγχο ανεξαρτησίας χ^2 του Pearson, καθώς όλα τα κελιά έχουν συχνότητα μεγαλύτερη από 1 και όλες οι αναμενόμενες συχνότητες είναι μεγαλύτερες από 5.

Crosstab						
			CH2O			
			1	2	3	Total
NewObesity	Insufficient_Weight	Count	10	15	8	33
		Expected Count	8,8	17,7	6,5	33,0
	Normal_Weight	Count	79	160	39	278
		Expected Count	74,5	149,0	54,5	278,0
	Overweight	Count	24	60	30	114
		Expected Count	30,6	61,1	22,3	114,0
	Obesity	Count	17	25	18	60
		Expected Count	16,1	32,2	11,8	60,0
Total		Count	130	260	95	485
		Expected Count	130,0	260,0	95,0	485,0

Εικόνα 3.3.5: Πίνακας συνάφειας ποσότητας καθημερινής κατανάλωσης νερού και επιπέδου παχυσαρκίας

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson με τιμή 15,419 και p-value<0,05 όπως φαίνονται στον Πίνακα 3.3.5(α), απορρίπτει τη μηδενική υπόθεση σε επίπεδο σημαντικότητας 5%. Συνεπώς, η ποσότητα νερού που καταναλώνεται καθημερινά και το επίπεδο παχυσαρκίας δεν είναι ανεξάρτητα. Υπάρχει, δηλαδή, σημαντική συσχέτιση μεταξύ των μεταβλητών.

Chi-Square Tests

	Value	Asymptotic Significance (2- sided)
Pearson Chi-Square	15,419	,017
Likelihood Ratio	15,390	,017

Πίνακας 3.3.5(α): Έλεγχος ανεξαρτησίας ποσότητας καθημερινής κατανάλωσης νερού και επιπέδου παχυσαρκίας

Για να δούμε πόσο ισχυρή είναι η συσχέτιση, θα ελέγξουμε την τιμή V του Cramer. Στον παρακάτω πίνακα βλέπουμε ότι η τιμή αυτή είναι 0,126 που σημαίνει ότι υπάρχει μικρής έντασης συσχέτιση.

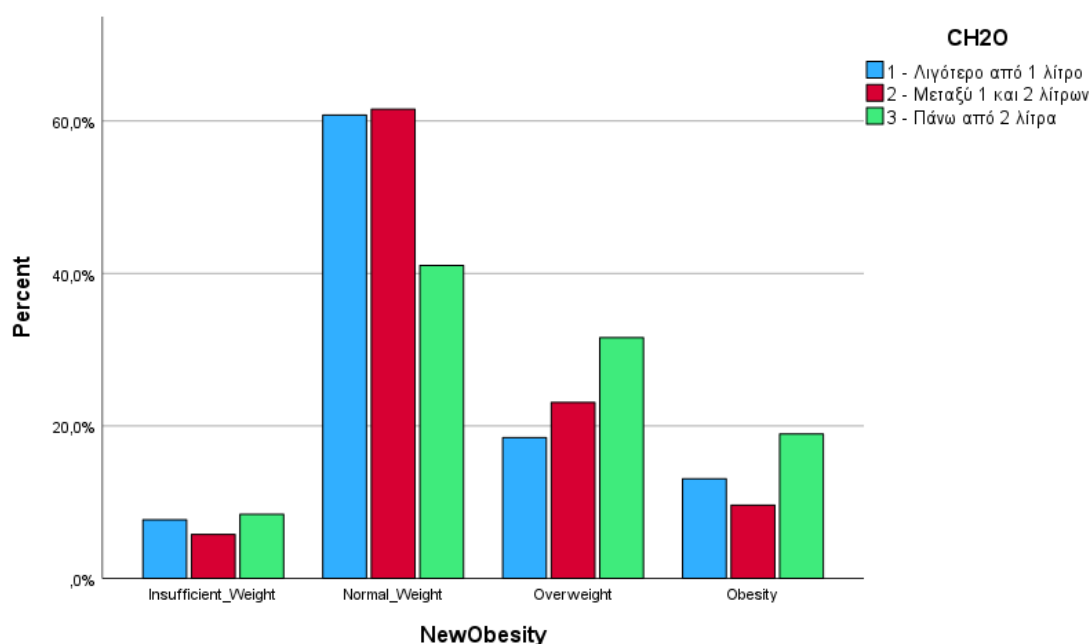
Symmetric Measures

		Value
Nominal by Nominal	Phi	,178
	Cramer's V	,126

Πίνακας 3.3.5(β): Ένταση συσχέτισης ποσότητας καθημερινής κατανάλωσης νερού και επιπέδου παχυσαρκίας

Στο ακόλουθο ραβδόγραμμα απεικονίζονται τα ποσοστά κατανάλωσης νερού ανάλογα με το επίπεδο παχυσαρκίας. Η κατανομή των ποσοστών αναδεικνύει τη διαφοροποίηση στην κατανάλωση νερού ανά επίπεδο παχυσαρκίας, γεγονός που ενισχύει την ένδειξη ύπαρξης συσχέτισης μεταξύ των δύο μεταβλητών.

Παρατηρείται ότι τα άτομα με φυσιολογικό βάρος παρουσιάζουν το υψηλότερο ποσοστό κατανάλωσης νερού μεταξύ 1 και 2 λίτρων ημερησίως (άνω του 60%). Αντιθέτως, στις κατηγορίες «Υπέρβαροι» και «Παχύσαρκοι» αυξάνεται το ποσοστό των ατόμων που καταναλώνουν πάνω από 2 λίτρα νερού την ημέρα. Το γεγονός ότι τα παχύσαρκα άτομα καταναλώνουν περισσότερο νερό δεν σημαίνει ότι η αυξημένη κατανάλωση νερού προκαλεί παχυσαρκία· αντίθετα, είναι πιθανό η αυξημένη κατανάλωση να αποτελεί αντίδραση στην ήδη υπάρχουσα παχυσαρκία, είτε στο πλαίσιο προσπαθειών ελέγχου βάρους είτε λόγω αυξημένων διατροφικών αναγκών.



Διάγραμμα 3.3.5: Ραβδόγραμμα συσχέτισης της ποσότητας καθημερινής κατανάλωσης νερού με το επίπεδο παχυσαρκίας

3.3.6 Συσχέτιση με τη μεταβλητή «SCC»

Η μεταβλητή «καταμέτρηση της καθημερινής λήψης θερμίδων» είναι μια κατηγορική μεταβλητή που παίρνει τις τιμές ναι ή όχι. Στην Εικόνα 3.3.6 παρουσιάζεται ο πίνακας συνάφειας μεταξύ των δύο ποιοτικών μεταβλητών που μελετάμε, ώστε να έχουμε μια πρώτη

εικόνα από τις συχνότητες εμφάνισης της πάθησης ανάλογα με το αν οι πάσχοντες μετρώνε τις θερμίδες που λαμβάνουν καθημερινά με την τροφή τους. Συνολικά, 60 συμμετέχοντες πάσχουν από παχυσαρκία, από τους οποίους οι 57 δεν μετρώνε τις θερμίδες που καταναλώνουν, ενώ μόνο οι 3 τις μετρώνε.

Επίσης, βλέπουμε ότι ικανοποιούνται οι προϋποθέσεις για τον έλεγχο ανεξαρτησίας χ^2 του Pearson, καθώς όλα τα κελιά έχουν συχνότητα μεγαλύτερη από 1 και πάνω από το 80% των κελιών έχουν αναμενόμενη συχνότητα μεγαλύτερη από 5.

Crosstab					
			SCC		Total
			no	yes	
NewObesity	Insufficient_Weight	Count	27	6	33
		Expected Count	29,4	3,6	33,0
	Normal_Weight	Count	249	29	278
		Expected Count	247,6	30,4	278,0
	Overweight	Count	99	15	114
		Expected Count	101,5	12,5	114,0
	Obesity	Count	57	3	60
		Expected Count	53,4	6,6	60,0
Total	Count	432	53	485	
	Expected Count	432,0	53,0	485,0	

Εικόνα 3.3.6: Πίνακας συνάφειας καθημερινής μέτρησης θερμίδων και επιπέδου παχυσαρκίας

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson με τιμή 4,603 και $p\text{-value} > 0,05$ όπως φαίνονται στον Πίνακα 3.3.6, δεν απορρίπτει τη μηδενική υπόθεση σε επίπεδο σημαντικότητας 5%. Δηλαδή, η καταμέτρηση των θερμίδων που λαμβάνονται καθημερινά και το επίπεδο παχυσαρκίας είναι ανεξάρτητα, άρα δεν υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ τους.

Chi-Square Tests		
	Value	Asymptotic Significance (2-sided)
Pearson Chi-Square	4,603 ^a	,203
Likelihood Ratio	4,798	,187

a. 1 cells (12,5%) have expected count less than 5. The minimum expected count is 3,61.

Πίνακας 3.3.6: Έλεγχος ανεξαρτησίας καθημερινής μέτρησης θερμίδων και επιπέδου παχυσαρκίας

3.3.7 Συσχέτιση με τη μεταβλητή «CALC»

Η μεταβλητή «συχνότητα κατανάλωσης αλκοόλ» είναι μια κατηγορική μεταβλητή. Στην Εικόνα 3.3.7 παρουσιάζεται ο πίνακας συνάφειας μεταξύ των δύο ποιοτικών μεταβλητών που μελετάμε, ώστε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης

ανάλογα με τη συχνότητα κατανάλωσης αλκοόλ. Συνολικά, 60 συμμετέχοντες πάσχουν από παχυσαρκία, από τους οποίους οι 21 δεν καταναλώνουν καθόλου αλκοόλ, οι 30 καταναλώνουν μερικές φορές, οι 9 συχνά, ενώ κανένας δεν απάντησε «πάντα».

Επίσης, βλέπουμε ότι δεν ικανοποιούνται οι προϋποθέσεις για τον έλεγχο ανεξαρτησίας χ^2 του Pearson, καθώς δεν έχουν όλα τα κελιά συχνότητα μεγαλύτερη από 1 και πάνω από 20% των κελιών έχουν συχνότητα μικρότερη από 5.

Crosstab							
			CALC				Total
			Always	Frequently	Sometimes	no	
NewObesity	Insufficient_Weight	Count	0	1	18	14	33
		Expected Count	,1	3,0	18,0	11,9	33,0
	Normal_Weight	Count	1	17	155	105	278
		Expected Count	,6	25,2	151,9	100,3	278,0
	Overweight	Count	0	17	62	35	114
		Expected Count	,2	10,3	62,3	41,1	114,0
	Obesity	Count	0	9	30	21	60
		Expected Count	,1	5,4	32,8	21,6	60,0
Total		Count	1	44	265	175	485
		Expected Count	1,0	44,0	265,0	175,0	485,0

Εικόνα 3.3.7: Πίνακας συνάφειας συχνότητας κατανάλωσης αλκοόλ και επιπέδου παχυσαρκίας

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson ανέδειξε τιμή 13,184 όπως φαίνεται στον Πίνακα 3.3.7. Δεδομένου ότι το p-value από τη μέθοδο Monte Carlo είναι μεγαλύτερο από 0,05, δεν απορρίπτεται η μηδενική υπόθεση στο επίπεδο σημαντικότητας 5%. Συνεπώς, η συχνότητα κατανάλωσης αλκοόλ και το επίπεδο παχυσαρκίας είναι ανεξάρτητα, γεγονός που υποδηλώνει ότι δεν υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ τους.

		Chi-Square Tests			
	Value	Asymptotic Significance (2- sided)	Monte Carlo Sig. (2-sided)		
			Significance	95% Confidence Interval	
				Lower Bound	Upper Bound
Pearson Chi-Square	13,184 ^a	,154	,176 ^b	,168	,183
Likelihood Ratio	13,312	,149	,105 ^b	,099	,110
Fisher-Freeman-Halton Exact Test	14,260		,095 ^b	,089	,100

a. 5 cells (31,3%) have expected count less than 5. The minimum expected count is ,07.

b. Based on 10000 sampled tables with starting seed 334431365.

Πίνακας 3.3.7: Έλεγχος ανεξαρτησίας συχνότητας κατανάλωσης αλκοόλ και επιπέδου παχυσαρκίας με μέθοδο Monte Carlo

3.4 Συσχετίσεις φυσικής δραστηριότητας και άλλων συνηθειών με το «Επίπεδο παχυσαρκίας»

Σε αυτή την παράγραφο θα ελέγξουμε τη σχέση εξάρτησης που μπορεί να υπάρχει ανάμεσα στη φυσική δραστηριότητα και άλλες συνήθειες, όπως παρουσιάστηκαν στο δεύτερο κεφάλαιο, με τη μεταβλητή της παχυσαρκίας.

3.4.1 Συσχέτιση με τη μεταβλητή «FAF»

Η μεταβλητή «συχνότητα εβδομαδιαίας σωματικής άσκησης» είναι κατηγορική. Στην Εικόνα 3.4.1 παρουσιάζεται ο πίνακας συνάφειας μεταξύ των δύο ποιοτικών μεταβλητών που μελετάμε, ώστε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης ανάλογα με τη συχνότητα εβδομαδιαίας σωματικής άσκησης. Συνολικά, 60 συμμετέχοντες πάσχουν από παχυσαρκία, από τους οποίους οι 28 δεν ασκούνται καθόλου, οι 15 ασκούνται 1 με 2 ημέρες, οι 12 ασκούνται 2 με 4 ημέρες την εβδομάδα και οι 5 ασκούνται 4 με 5 ημέρες.

Επίσης, βλέπουμε ότι ικανοποιούνται οι προϋποθέσεις για τον έλεγχο ανεξαρτησίας χ^2 του Pearson, καθώς όλα τα κελιά έχουν συχνότητα μεγαλύτερη από 1 και πάνω από το 80% των κελιών έχουν αναμενόμενη συχνότητα μεγαλύτερη από 5.

			Crosstab				
			FAF				
			0	1	2	3	Total
NewObesity	Insufficient_Weight	Count	9	6	14	4	33
		Expected Count	10,7	10,5	7,6	4,3	33,0
	Normal_Weight	Count	77	94	67	40	278
		Expected Count	90,0	88,3	63,6	36,1	278,0
	Overweight	Count	43	39	18	14	114
		Expected Count	36,9	36,2	26,1	14,8	114,0
	Obesity	Count	28	15	12	5	60
		Expected Count	19,4	19,1	13,7	7,8	60,0
Total		Count	157	154	111	63	485
		Expected Count	157,0	154,0	111,0	63,0	485,0

Εικόνα 3.4.1: Πίνακας συνάφειας συχνότητας σωματικής άσκησης και επιπέδου παχυσαρκίας

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson με τιμή 20,194 και $p\text{-value} < 0,05$ όπως φαίνονται στον Πίνακα 3.4.1(α), απορρίπτει τη μηδενική υπόθεση σε επίπεδο σημαντικότητας 5%. Δηλαδή, η συχνότητα εβδομαδιαίας σωματικής άσκησης και το επίπεδο παχυσαρκίας δεν είναι ανεξάρτητα και υπάρχει στατιστικά σημαντική σχέση ανάμεσα στις δύο μεταβλητές.

Chi-Square Tests		
	Value	Asymptotic Significance (2-sided)
Pearson Chi-Square	20,194 ^a	,017
Likelihood Ratio	19,535	,021

a. 1 cells (6,3%) have expected count less than 5. The minimum expected count is 4,29.

Πίνακας 3.4.1(α): Έλεγχος ανεξαρτησίας συχνότητας σωματικής άσκησης και επιπέδου παχυσαρκίας

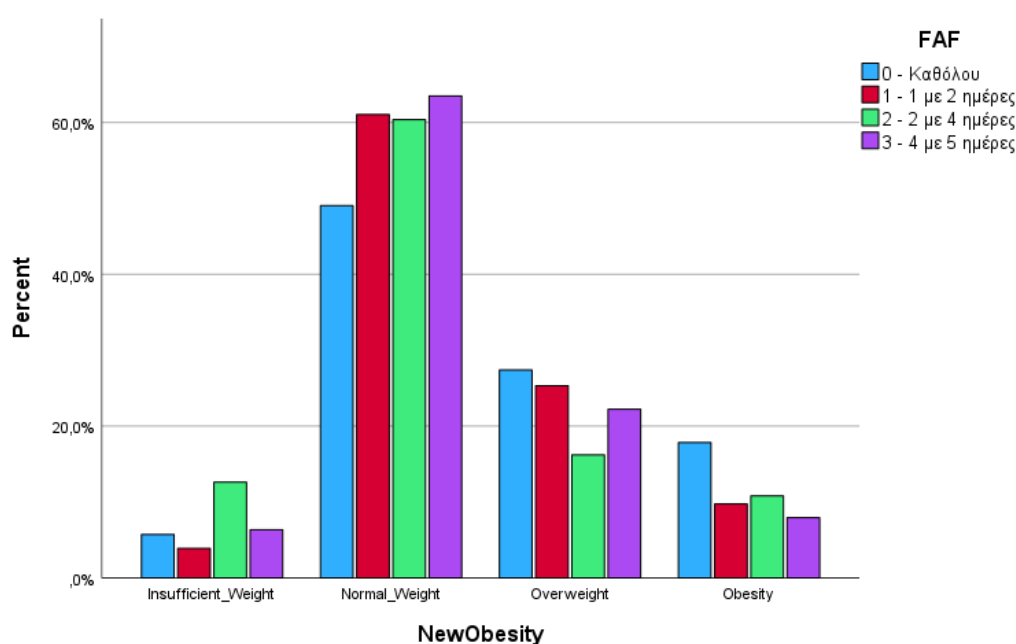
Για να δούμε πόσο ισχυρή είναι η συσχέτιση, θα ελέγξουμε την τιμή V του Cramer. Στον παρακάτω πίνακα βλέπουμε ότι η τιμή αυτή είναι 0,118 που σημαίνει ότι υπάρχει μικρής έντασης συσχέτιση.

Symmetric Measures

		Value
Nominal by Nominal	Phi	,204
	Cramer's V	,118

Πίνακας 3.4.1(β): Ένταση συσχέτισης συχνότητας σωματικής άσκησης και επιπέδου παχυσαρκίας

Στο Διάγραμμα 3.4.1 απεικονίζεται η σχέση μεταξύ της συχνότητας σωματικής άσκησης με το επίπεδο παχυσαρκίας. Παρατηρούμε ότι τα άτομα με φυσιολογικό βάρος παρουσιάζουν υψηλότερα ποσοστά σωματικής δραστηριότητας, ενώ οι υψηλότερες κατηγορίες βάρους συνδέονται με τη μειωμένη σωματική δραστηριότητα. Αυτό υποδηλώνει αρνητική συσχέτιση μεταξύ σωματικής δραστηριότητας και βαθμού παχυσαρκίας, δηλαδή όσο μειώνεται η σωματική δραστηριότητα, τόσο αυξάνεται ο κίνδυνος εμφάνισης της νόσου.



Διάγραμμα 3.4.1: Ραβδόγραμμα συσχέτισης συχνότητας σωματικής άσκησης με το επίπεδο παχυσαρκίας

3.4.2 Συσχέτιση με τη μεταβλητή «TUE»

Η μεταβλητή «χρόνος χρήσης τεχνολογικών συσκευών» είναι μια κατηγορική μεταβλητή. Στην Εικόνα 3.4.2 παρουσιάζεται ο πίνακας συνάφειας μεταξύ των δύο ποιοτικών μεταβλητών που μελετάμε, ώστε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης ανάλογα με το χρόνο που αφιερώνουν οι πάσχοντες σε διάφορα τεχνολογικά μέσα. Συνολικά, 60 συμμετέχοντες πάσχουν από παχυσαρκία, από τους οποίους οι 32 χρησιμοποιούν τεχνολογικές συσκευές από 0 έως 2 ώρες καθημερινά, οι 16 από 3 έως 5 ώρες και οι 12 πάνω από 6 ώρες την ημέρα.

Επίσης, βλέπουμε ότι ικανοποιούνται οι προϋποθέσεις για τον έλεγχο ανεξαρτησίας χ^2 του Pearson, καθώς όλα τα κελιά έχουν συχνότητα μεγαλύτερη από 1 και πάνω από το 80% των κελιών έχουν αναμενόμενη συχνότητα μεγαλύτερη από 5.

Crosstab						
			TUE			Total
			0	1	2	
NewObesity	Insufficient_Weight	Count	13	13	7	33
		Expected Count	16,1	12,0	4,9	33,0
	Normal_Weight	Count	125	118	35	278
		Expected Count	135,3	101,5	41,3	278,0
	Overweight	Count	66	30	18	114
		Expected Count	55,5	41,6	16,9	114,0
	Obesity	Count	32	16	12	60
		Expected Count	29,2	21,9	8,9	60,0
Total		Count	236	177	72	485
		Expected Count	236,0	177,0	72,0	485,0

Εικόνα 3.4.2: Πίνακας συνάφειας χρόνου χρήσης τεχνολογικών συσκευών και επιπέδου παχυσαρκίας

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson με τιμή 14,224 και $p\text{-value} < 0,05$ όπως φαίνονται στον Πίνακα 3.4.2(α), απορρίπτει τη μηδενική υπόθεση σε επίπεδο σημαντικότητας 5%. Δηλαδή, ο χρόνος χρήσης τεχνολογικών συσκευών και το επίπεδο παχυσαρκίας δεν είναι ανεξάρτητα, άρα υπάρχει συσχέτιση μεταξύ τους.

Chi-Square Tests

	Value	Asymptotic Significance (2- sided)
Pearson Chi-Square	14,224 ^a	,027
Likelihood Ratio	14,379	,026

a. 1 cells (8,3%) have expected count less than 5. The minimum expected count is 4,90.

Πίνακας 3.4.2(α): Έλεγχος ανεξαρτησίας χρόνου χρήσης τεχνολογικών συσκευών και επιπέδου παχυσαρκίας

Για να δούμε πόσο ισχυρή είναι η συσχέτιση, θα ελέγξουμε την τιμή V του Cramer. Στον παρακάτω πίνακα βλέπουμε ότι η τιμή αυτή είναι 0,121 που σημαίνει ότι υπάρχει μικρής έντασης συσχέτιση.

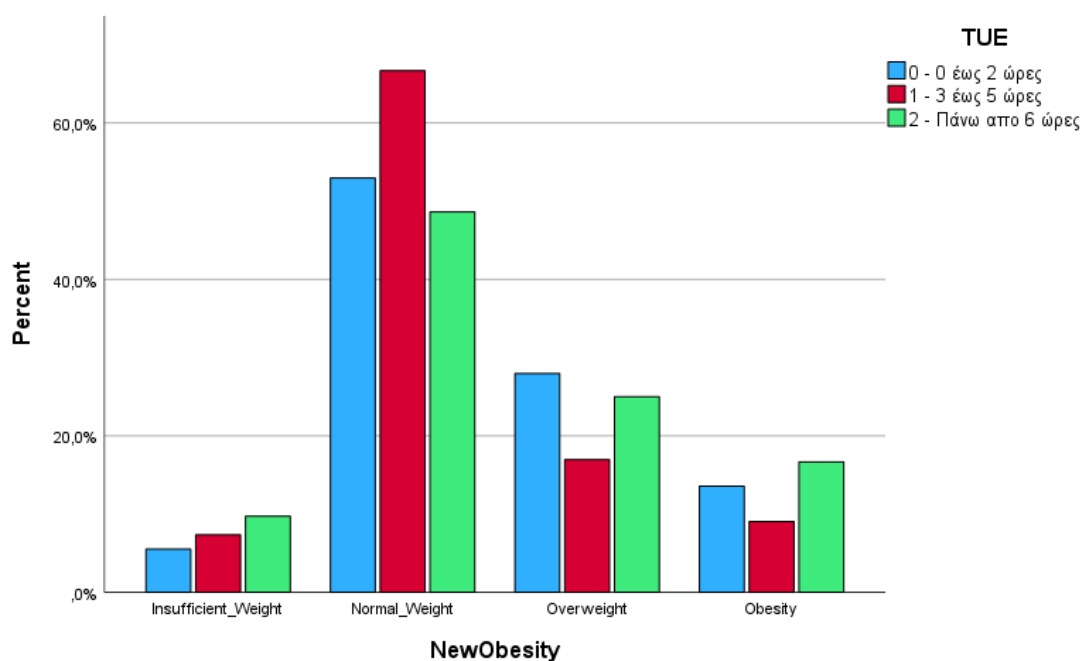
Symmetric Measures

		Value
Nominal by Nominal	Phi	,171
	Cramer's V	,121

Πίνακας 3.4.2(β): Ένταση συσχέτισης χρόνου χρήσης τεχνολογικών συσκευών και επιπέδου παχυσαρκίας

Το παρακάτω ραβδόγραμμα απεικονίζει τη σχέση που υπάρχει ανάμεσα στο χρόνο χρήσης ηλεκτρονικών συσκευών και το επίπεδο παχυσαρκίας. Παρατηρούμε ότι τα άτομα με φυσιολογικό βάρος συγκεντρώνονται κυρίως στην κατηγορία με μέτρια χρήση οθόνης (3–5 ώρες), σε ποσοστό άνω του 65%. Οι αντίστοιχοι δείκτες είναι χαμηλότεροι για τα άτομα που είναι υπέρβαρα ή παχύσαρκα. Συγκεκριμένα, στις κατηγορίες Overweight και Obesity, παρατηρείται αυξημένο ποσοστό ατόμων που περνούν πάνω από 6 ώρες μπροστά σε οθόνη, υποδηλώνοντας μια τάση συσχέτισης μεταξύ αυξημένου χρόνου καθιστικής συμπεριφοράς και υψηλότερου ΔΜΣ. Επιπλέον, όσοι κάνουν χαμηλή χρήση οθόνης (0–2 ώρες) τείνουν να βρίσκονται είτε στην κατηγορία φυσιολογικού βάρους, είτε -σε μικρότερο βαθμό- στις υπέρβαρες ομάδες.

Η κατανομή των ποσοστών υποδεικνύει μια ενδεχόμενη αρνητική επίδραση της πολύωρης έκθεσης σε οθόνες στο σωματικό βάρος. Η συστηματική καθιστική δραστηριότητα που σχετίζεται με την αυξημένη χρήση ηλεκτρονικών συσκευών ενδέχεται να συνεισφέρει στην εμφάνιση υπερβάλλοντος βάρους και παχυσαρκίας.



Διάγραμμα 3.4.2: Ραβδόγραμμα συσχέτισης του χρόνου χρήσης τεχνολογικών συσκευών με το επίπεδο παχυσαρκίας

3.4.3 Συσχέτιση με τη μεταβλητή «MTRANS»

Η μεταβλητή «τρόπος μετακίνησης» είναι μια κατηγορική μεταβλητή. Στην Εικόνα 3.4.3 παρουσιάζεται ο πίνακας συνάφειας μεταξύ των δύο ποιοτικών μεταβλητών που μελετάμε, ώστε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης ανάλογα με το μέσο μεταφοράς που χρησιμοποιούν οι πάσχοντες. Συνολικά, 60 συμμετέχοντες πάσχουν από παχυσαρκία, από τους οποίους οι 19 χρησιμοποιούν αυτοκίνητο, 1 χρησιμοποιεί ποδήλατο, 2 χρησιμοποιούν μηχανή, οι 35 δημόσιες συγκοινωνίες και μόνο 3 περπατάνε.

Επίσης, βλέπουμε ότι δεν ικανοποιούνται οι προϋποθέσεις για τον έλεγχο ανεξαρτησίας χ^2 του Pearson, καθώς δεν έχουν όλα τα κελιά συχνότητα μεγαλύτερη από 1 και πάνω από 20% των κελιών έχουν συχνότητα μικρότερη από 5.

			Crosstab					
			MTRANS					
			Automobile	Bike	Motorbike	Public_Transp ortation	Walking	Total
NewObesity	Insufficient_Weight	Count	3	0	0	24	6	33
		Expected Count	6,7	,5	,5	21,6	3,7	33,0
	Normal_Weight	Count	45	4	4	194	31	278
		Expected Count	56,2	4,0	4,6	182,3	31,0	278,0
	Overweight	Count	31	2	2	65	14	114
		Expected Count	23,0	1,6	1,9	74,7	12,7	114,0
	Obesity	Count	19	1	2	35	3	60
		Expected Count	12,1	,9	1,0	39,3	6,7	60,0
Total	Count	98	7	8	318	54	485	
	Expected Count	98,0	7,0	8,0	318,0	54,0	485,0	

Εικόνα 3.4.3: Πίνακας συνάφειας τρόπου μετακίνησης και επιπέδου παχυσαρκίας

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson ανέδειξε τιμή 19,522 όπως φαίνεται στον Πίνακα 3.4.3. Δεδομένου ότι το p-value από τη μέθοδο Monte Carlo είναι οριακά μεγαλύτερο από 0,05, δεν απορρίπτεται η μηδενική υπόθεση στο επίπεδο σημαντικότητας 5%. Συνεπώς, το μέσο μεταφοράς που χρησιμοποιούν οι συμμετέχοντες και το επίπεδο παχυσαρκίας είναι ανεξάρτητα, γεγονός που υποδηλώνει ότι δεν υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ τους.

Chi-Square Tests					
	Value	Asymptotic Significance (2- sided)	Monte Carlo Sig. (2-sided)		
			Significance	95% Confidence Interval Lower Bound	Upper Bound
Pearson Chi-Square	19,522 ^a	,077	,079 ^b	,074	,084
Likelihood Ratio	20,480	,059	,069 ^b	,064	,074
Fisher-Freeman-Halton Exact Test	19,166		,048 ^b	,043	,052

a. 9 cells (45,0%) have expected count less than 5. The minimum expected count is 0,48.

b. Based on 10000 sampled tables with starting seed 334431365.

Πίνακας 3.4.3: Έλεγχος ανεξαρτησίας τρόπου μετακίνησης και επιπέδου παχυσαρκίας με τη μέθοδο Monte Carlo

3.4.4 Συσχέτιση με τη μεταβλητή «SMOKE»

Η μεταβλητή «κάπνισμα» είναι μια κατηγορική μεταβλητή που παίρνει τους τιμές ναι ή όχι. Στην Εικόνα 3.4.4 παρουσιάζεται ο πίνακας συνάφειας μεταξύ των δύο ποιοτικών μεταβλητών που μελετάμε, ώστε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης ανάλογα με το αν οι πάσχοντες είναι καπνιστές ή όχι. Συνολικά, 60 συμμετέχοντες πάσχουν από παχυσαρκία, από τους οποίους οι 10 καπνίζουν ενώ οι 50 όχι.

Επίσης, βλέπουμε ότι δεν ικανοποιούνται οι προϋποθέσεις για τον έλεγχο ανεξαρτησίας χ^2 του Pearson, καθώς όλα τα κελιά έχουν συχνότητα μεγαλύτερη από 1 αλλά πάνω από 20% των κελιών έχουν συχνότητα μικρότερη από 5.

Crosstab					
			SMOKE		Total
			no	yes	
NewObesity	Insufficient_Weight	Count	32	1	33
		Expected Count	30,9	2,1	33,0
	Normal_Weight	Count	266	12	278
		Expected Count	260,2	17,8	278,0
	Overweight	Count	106	8	114
		Expected Count	106,7	7,3	114,0
	Obesity	Count	50	10	60
		Expected Count	56,2	3,8	60,0
Total	Count	454	31	485	
	Expected Count	454,0	31,0	485,0	

Εικόνα 3.4.4: Πίνακας συνάφειας καπνιστών και επιπέδου παχυσαρκίας

Ο έλεγχος ανεξαρτησίας χ^2 του Pearson ανέδειξε τιμή 13,286 όπως φαίνεται στον Πίνακα 3.4.4(α). Δεδομένου ότι το p-value από τη μέθοδο Monte Carlo είναι μικρότερο από 0,05, απορρίπτεται η μηδενική υπόθεση στο επίπεδο σημαντικότητας 5%. Δηλαδή, το κάπνισμα και το επίπεδο παχυσαρκίας δεν είναι ανεξάρτητα, γεγονός που υποδηλώνει την ύπαρξη στατιστικά σημαντικής συσχέτισης μεταξύ τους.

Chi-Square Tests					
	Value	Asymptotic Significance (2- sided)	Monte Carlo Sig. (2-sided)		
			Significance	95% Confidence Interval Lower Bound	Upper Bound
Pearson Chi-Square	13,286 ^a	,004	,005 ^b	,004	,007
Likelihood Ratio	10,623	,014	,018 ^b	,015	,021
Fisher-Freeman-Halton Exact Test	10,679		,010 ^b	,008	,012

a. 2 cells (25,0%) have expected count less than 5. The minimum expected count is 2,11.

b. Based on 10000 sampled tables with starting seed 334431365.

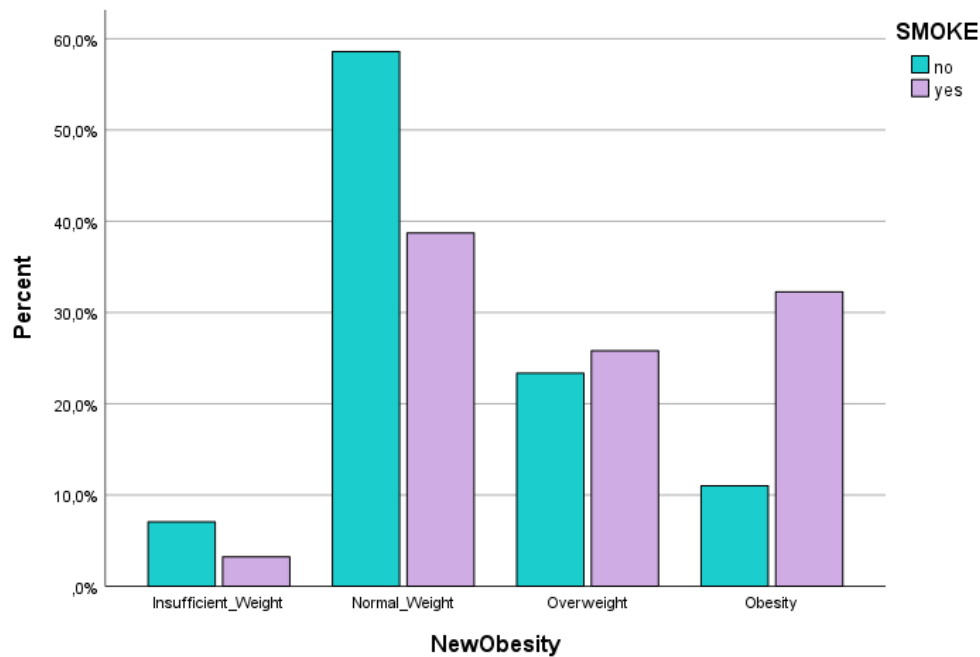
Πίνακας 3.4.4(α): Έλεγχος ανεξαρτησίας καπνιστών και επιπέδου παχυσαρκίας με τη μέθοδο Monte Carlo

Για να δούμε πόσο ισχυρή είναι η συσχέτιση, θα ελέγξουμε την τιμή V του Cramer. Στον παρακάτω πίνακα βλέπουμε ότι η τιμή αυτή είναι 0,166 που σημαίνει ότι υπάρχει μικρής έντασης συσχέτιση.

Symmetric Measures			
		Value	Approximate Significance
Nominal by Nominal	Phi	,166	,004
	Cramer's V	,166	,004

Πίνακας 3.4.4(β): Ένταση συσχέτισης καπνίσματος και επιπέδου παχυσαρκίας

Το επόμενο ραβδόγραμμα αποκαλύπτει μια διακριτή διαφοροποίηση στην κατανομή των καπνιστών και μη καπνιστών ανά επίπεδο παχυσαρκίας.. Συγκεκριμένα, στις κατηγορίες Overweight και Obesity, παρατηρείται αυξημένη αναλογία καπνιστών, σε αντίθεση με τις κατηγορίες Insufficient και Normal Weight, όπου κυριαρχούν οι μη καπνιστές. Το εύρημα αυτό ενδέχεται να υποδηλώνει μια θετική συσχέτιση μεταξύ καπνίσματος και αυξημένου σωματικού βάρους στο συγκεκριμένο δείγμα, παρά τη συνήθη εντύπωση ότι το κάπνισμα συσχετίζεται με μειωμένη όρεξη και συνεπώς με χαμηλότερο βάρος. Σε αυτή την περίπτωση, το κάπνισμα, αντί να λειτουργεί προστατευτικά έναντι της παχυσαρκίας, ενδέχεται να συμβάλλει στην αύξηση του σωματικού βάρους, ιδιαίτερα όταν συνυπάρχει με άλλους παράγοντες κινδύνου.



Διάγραμμα 3.4.4: Ραβδόγραμμα συσχέτισης του καπνίσματος με το επίπεδο παχυσαρκίας

3.5 Συσχέτιση ΔΜΣ με το «Επίπεδο παχυσαρκίας»

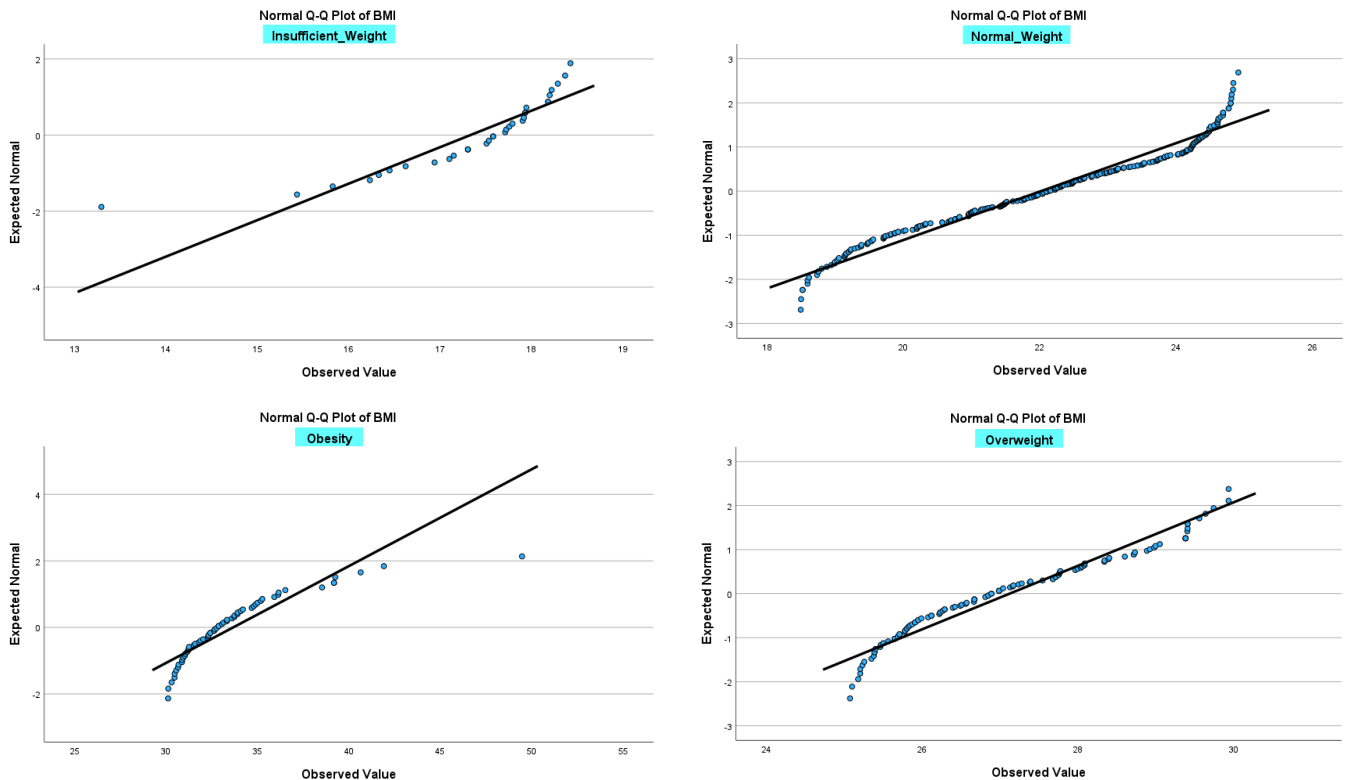
Σκοπός της παραγράφου είναι να ερευνήσουμε την ύπαρξη συσχέτισης μεταξύ της μεταβλητής του ΔΜΣ και της μεταβλητής που φανερώνει τα επίπεδα παχυσαρκίας. Όπως αναφέραμε στο κεφάλαιο 2, ο ΔΜΣ είναι συνεχής μεταβλητή, οπότε θα κάνουμε πρώτα έλεγχο κανονικότητας. Θα πραγματοποιήσουμε τον έλεγχο Kolmogorov-Smirnov προκειμένου να αποφασίσουμε αν θα χρησιμοποιήσουμε παραμετρικό ή μη-παραμετρικό έλεγχο.

Tests of Normality

		Kolmogorov-Smirnov	
new_obesity		Statistics	p-value
BMI	Insufficient_Weight	,186	,005
	Normal_Weight	,072	,002
	Overweight	,095	,014
	Obesity	,155	,001

Πίνακας 3.5(α): Έλεγχος κανονικότητας Kolmogorov-Smirnov των μεταβλητών «BMI» και «επίπεδο παχυσαρκίας»

Στον Πίνακα 3.5(α) βλέπουμε ότι η υπόθεση της κανονικότητας απορρίπτεται σε επίπεδο σημαντικότητας $\alpha=5\%$ για κάθε επίπεδο παχυσαρκίας.



Εικόνα 3.5: Q-Q Plot γραφήματα για το επίπεδο παχυσαρκίας σε σχέση με το BMI

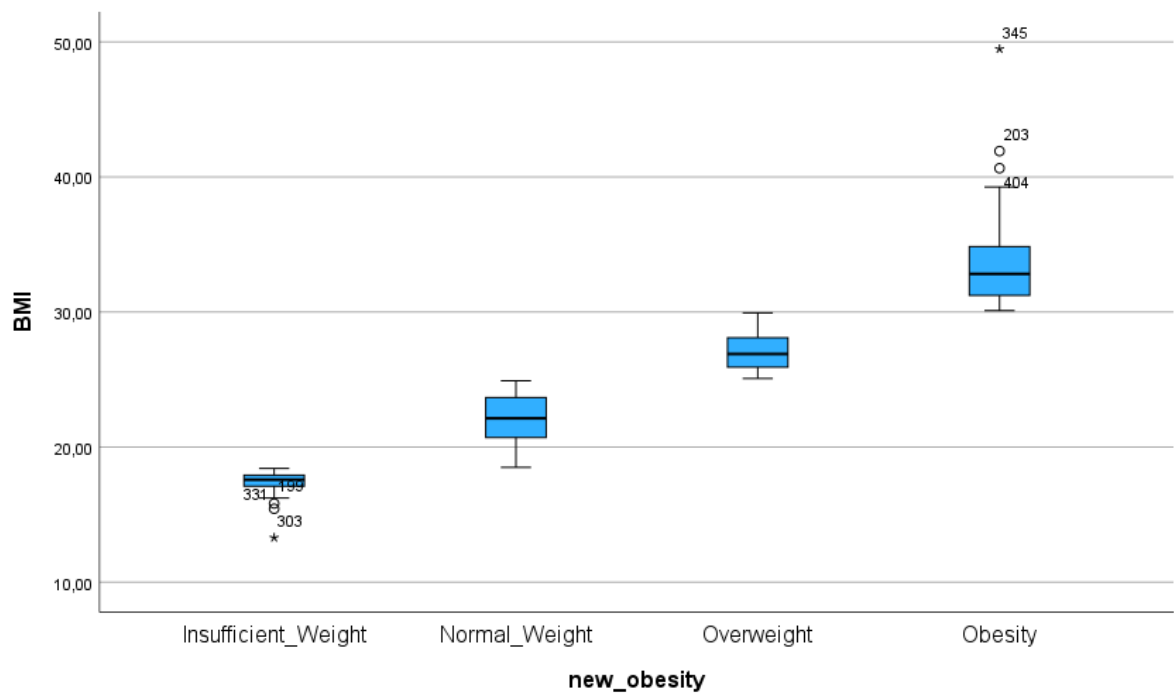
Η απόρριψη της κανονικότητας επιβεβαιώνεται και στα Q-Q Plot της Εικόνας 3.5, όπου φαίνεται ότι τα δεδομένα αποκλίνουν από την αναμενόμενη ευθεία της κανονικής κατανομής σε κάθε επίπεδο παχυσαρκίας. Οι αποκλίσεις παρατηρούνται κυρίως στις ακραίες τιμές, στις ουρές δηλαδή της κατανομής φαίνεται να είναι ιδιαίτερα έντονες, ειδικά στους λιποβαρείς σε αυτούς με κανονικό βάρος. Σε αυτή την περίπτωση λοιπόν, δεν χρειάζεται να ελέγξουμε την ομοσκεδαστικότητα, αλλά θα συνεχίσουμε με μη-παραμετρικό έλεγχο Kruskal-Wallis.

Test Statistics

	BMI
Kruskal-Wallis H	385,502
df	3
Asymp. Sig.	<,001

Πίνακας 3.5(β): Πίνακας ελέγχου συσχετίσεων Kruskal-Wallis των μεταβλητών «BMI» και «επίπεδο παχυσαρκίας»

Το p-value του ελέγχου, είναι μικρότερο από 0,05 οπότε απορρίπτεται η μηδενική υπόθεση. Δηλαδή, οι κατανομές ή οι διάμεσοι του Δείκτη Μάζας Σώματος διαφέρουν μεταξύ των επιπέδων της παχυσαρκίας. Συνεπώς, υπάρχει συσχέτιση μεταξύ των δύο αυτών μεταβλητών.



Διάγραμμα 3.5: Boxplot επιπέδου παχυσαρκίας σε σχέση με το BMI

Στο Διάγραμμα 3.5 παρουσιάζεται η κατανομή του ΔΜΣ ανά επίπεδο παχυσαρκίας. Οι ομάδες Obesity και Overweight εμφανίζουν μεγαλύτερη διάμεσο σε σχέση με τις ομάδες Normal και Insufficient Weight. Ενώ, οι ομάδες Obesity και Normal Weight εμφανίζουν μεγαλύτερη διασπορά στις τιμές που παίρνει ο Δείκτη Μάζας Σώματος. Ενισχύεται, δηλαδή, το αποτέλεσμα του ελέγχου Kruskal-Wallis, αφού βλέπουμε την αύξηση του διάμεσου ΔΜΣ καθώς αυξάνεται το επίπεδο παχυσαρκίας, τεκμηριώνοντας την ύπαρξη συσχέτισης μεταξύ των δύο μεταβλητών.

Κεφάλαιο 4

Μοντέλα Παλινδρόμησης

4.1 Στατιστική Ανάλυση

Σκοπός της παρούσας εργασίας είναι η αναζήτηση και εύρεση ενός κατάλληλου μοντέλου παλινδρόμησης το οποίο θα είναι ικανό να προβλέψει την τιμή της μεταβλητής «επίπεδο παχυσαρκίας». Αυτό θα το επιτύχουμε εξετάζοντας τη σχέση μεταξύ των παραγόντων κινδύνου και της εξεταζόμενης μεταβλητής.

Στη συγκεκριμένη έρευνα δεν υπήρχαν ελλείπουσες τιμές (missing values) και δεν χρειάστηκε να κάνουμε κάποια προεπεξεργασία των δεδομένων. Στον Πίνακα 4.1, παρατηρούμε ότι χρησιμοποιήθηκαν και οι 485 τιμές για τη στατιστική ανάλυση.

Case Processing Summary			
Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	485	100,0
	Missing Cases	0	,0
	Total	485	100,0
Unselected Cases		0	,0
Total		485	100,0

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

Πίνακας 4.1: Περιγραφή του δείγματος

Για να ερμηνεύσουμε τη μεταβλητή-στόχο θα χρησιμοποιήσουμε τις μεταβλητές που σύμφωνα με το κεφάλαιο 3, σχετίζονται με το επίπεδο παχυσαρκίας. Αυτές είναι οι εξής:

- Ηλικία (Age)
- Δείκτης Μάζας Σώματος (BMI)
- Οικογενειακό ιστορικό παχυσαρκίας (family history with overweight)
- Ποσότητα καθημερινής κατανάλωσης κύριων γευμάτων (NCP)
- Συχνότητα κατανάλωσης φαγητού ανάμεσα στα γεύματα (CAEC)
- Ποσότητα καθημερινής κατανάλωσης νερού (CH2O)
- Συχνότητα εβδομαδιαίας σωματικής άσκησης (FAF)
- Χρόνος χρήσης τεχνολογικών συσκευών (TUE) και
- Κάπνισμα (SMOKE).

4.2 Διωνυμική Λογιστική Παλινδρόμηση

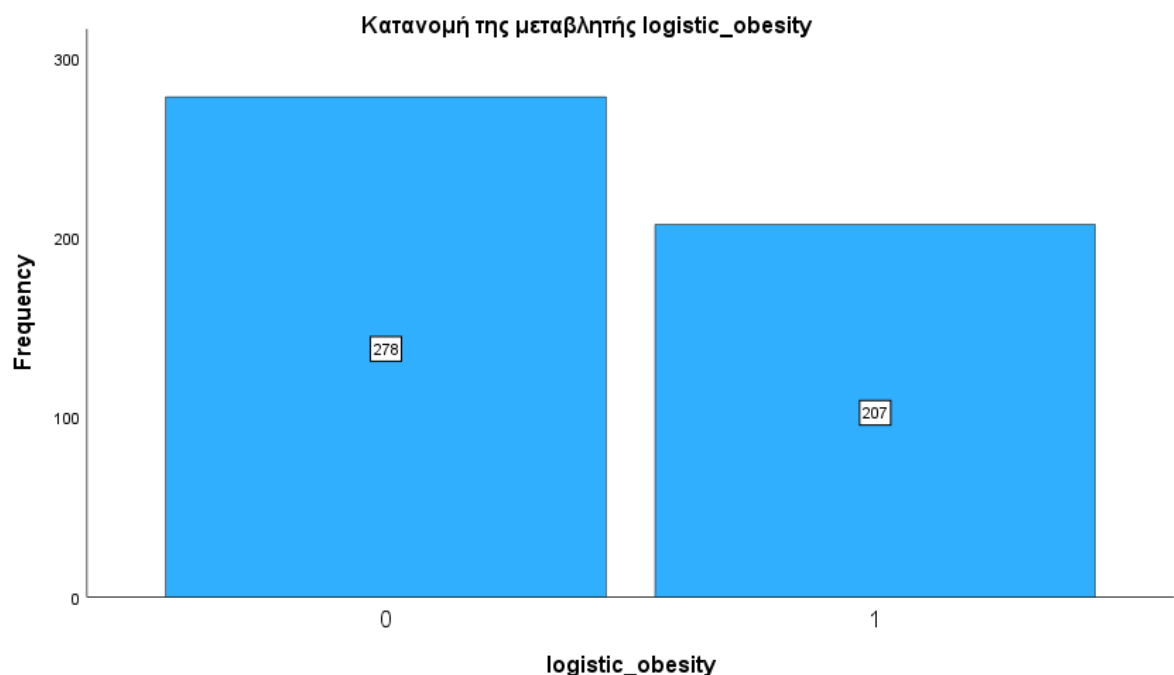
Όπως έχουμε αναφέρει στο πρώτο κεφάλαιο, η Διωνυμική Λογιστική Παλινδρόμηση ή απλά Λογιστική Παλινδρόμηση από εδώ και πέρα, εφαρμόζεται όταν η εξαρτημένη μεταβλητή είναι δίτιμη ερμηνευτική και προσπαθεί να περιγράψει τη σχέση μεταξύ αυτής της μεταβλητής και δύο ή περισσότερων ανεξάρτητων μεταβλητών.

Στη συγκεκριμένη μελέτη, η εξαρτημένη μεταβλητή είναι το επίπεδο παχυσαρκίας «NewObesity» και αποτελείται από τέσσερα επίπεδα όπως τα ορίσαμε στο δεύτερο κεφάλαιο, δηλαδή: «Λιπόσαρκος», «Κανονικό Βάρος», «Υπέρβαρος» και «Παχύσαρκος».

Για να εφαρμόσουμε τη Λογιστική Παλινδρόμηση θα πρέπει το επίπεδο παχυσαρκίας να είναι μια δίτιμη μεταβλητή. Έτσι λοιπόν, θα δημιουργήσουμε τη νέα μεταβλητή «logistic_obesity» στην οποία θα ορίσουμε το επίπεδο «Κανονικό Βάρος» ως 0 -κατηγορία αναφοράς- και τα υπόλοιπα επίπεδα ως 1. Παρακάτω βλέπουμε τις νέες συχνότητες των επιπέδων, όπου αυτοί που έχουν κανονικό βάρος αποτελούν την πλειοψηφία του δείγματος, ενώ όλες οι υπόλοιπες κατηγορίες μαζί αποτελούν το 42,7% των συμμετεχόντων.

logistic_obesity					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	278	57,3	57,3	57,3
	1	207	42,7	42,7	100,0
	Total	485	100,0	100,0	

Πίνακας 4.2: Στατιστικά στοιχεία νέων επιπέδων παχυσαρκίας



Διάγραμμα 4.2: Κατανομή νέων επιπέδων παχυσαρκίας

4.2.1 Πρώτη εφαρμογή Διωνυμικής Λογιστικής Παλινδρόμησης

Για το πρώτο μοντέλο, με τη μέθοδο Enter, θα χρησιμοποιήσουμε όλες τις ανεξάρτητες μεταβλητές που συσχετίζονται με το επίπεδο παχυσαρκίας, εκτός του βάρους, του ύψους και του ΔΜΣ, για να ελέγξουμε ποιες είναι σημαντικές και ποιες όχι, καθώς και αν το μοντέλο έχει καλή προσαρμογή. Τις τρεις αυτές μεταβλητές δεν θα τις χρησιμοποιήσουμε διότι είναι αυτές που καθορίζουν το επίπεδο παχυσαρκίας και συνεπώς θα είναι αρκετές από μόνες τους για να εξηγήσουν το μοντέλο.

Αρχικά, στον Πίνακα 4.2.1(α) βλέπουμε την κωδικοποίηση που χρησιμοποιήθηκε για τις κατηγορικές μεταβλητές: Συχνότητα κατανάλωσης φαγητού ανάμεσα στα γεύματα, Κάπνισμα και Οικογενειακό ιστορικό παχυσαρκίας. Στη μεταβλητή CAEC χρησιμοποιήθηκε το «Μερικές φορές» ως κατηγορία αναφοράς, ενώ στις άλλες δύο το «ναι».

Categorical Variables Codings

			Parameter coding		
		Frequency	(1)	(2)	(3)
CAEC	Always	51	1,000	,000	,000
	Frequent	134	,000	1,000	,000
	no	19	,000	,000	1,000
	Sometime	281	,000	,000	,000
SMOKE	no	454	1,000		
	yes	31	,000		
family_history_with_overweight	no	190	1,000		
	yes	295	,000		

Πίνακας 4.2.1(α): Κωδικοποίηση κατηγορικών μεταβλητών

Στη συνέχεια βλέπουμε τον Πίνακα Ταξινόμησης όπου παρατηρούμε ότι το μοντέλο προβλέπει πως 336 άτομα (224+112) έχουν κανονικό βάρος ενώ οι 149 (54+95) θα ανήκουν στις υπόλοιπες κατηγορίες. Οι 224 είχαν όντως κανονικό βάρος που σημαίνει ότι το μοντέλο έχει υψηλή ειδικότητα (specificity = 80,6%), δηλαδή είναι πολύ καλό στο να αναγνωρίζει τις περιπτώσεις "κανονικού βάρους". Ομοίως, οι 95 ανήκαν όντως σε άλλη κατηγορία βάρους και αποτελούν το 45,9% ($\frac{95}{95+112}$) των πραγματικά θετικών περιπτώσεων. Αυτό δείχνει ότι το μοντέλο ανιχνεύει με μέτρια αξιοπιστία τις περιπτώσεις εκτός του "κανονικού βάρους". Τέλος, το ποσοστό σωστής ταξινόμησης είναι 65,8%.

Classification Table^a

		Predicted		Percentage Correct
Observed		logistic_obesity 0	1	
Step 1	logistic_obesity 0	224	54	80,6
	1	112	95	45,9
Overall Percentage				65,8

a. The cut value is ,500

Πίνακας 4.2.1(β): Ακρίβεια πρόβλεψης του πρώτου μοντέλου Λογιστικής Παλινδρόμησης

Στον Πίνακα 4.2.1(γ) εξετάζουμε αν το μοντέλο έχει καλή προσαρμογή μέσω του ελέγχου Hosmer-Lemeshow. Ο συγκεκριμένος έλεγχος ακολουθεί τις εξής υποθέσεις:

H₀: Το μοντέλο έχει καλή προσαρμογή, έναντι της εναλλακτικής

H₁: Το μοντέλο δεν έχει καλή προσαρμογή

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	5,269	8	,728

Πίνακας 4.2.1(γ): Έλεγχος καλής προσαρμογής Hosmer και Lemeshow πρώτου μοντέλου Λογιστικής Παλινδρόμησης

Με p-value= 0,728, δεν απορρίπτουμε τη μηδενική υπόθεση, οπότε συμπεραίνουμε ότι το μοντέλο έχει καλή προσαρμογή.

Οπότε, τρέχοντας το μοντέλο Διωνυμικής Λογιστικής Παλινδρόμησης, θα παρατηρήσουμε στον Πίνακα 4.2.1(δ) ότι οι μεταβλητές που είναι στατιστικά σημαντικές για το μοντέλο είναι η Ηλικία, το Οικογενειακό ιστορικό παχυσαρκίας, η Ποσότητα καθημερινής κατανάλωσης νερού (CH2O) καθώς και η Συχνότητα εβδομαδιαίας σωματικής άσκησης (FAF).

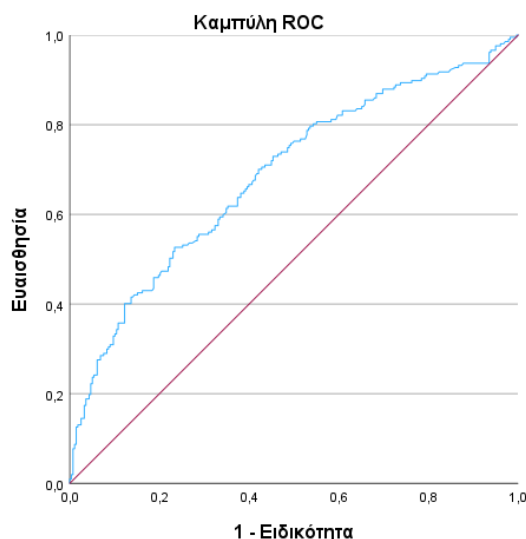
Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Age	,072	,017	18,664	1	<,001	1,075
	family_history_with_overweight(1)	,425	,205	4,296	1	,038	1,530
	NCP	-,164	,108	2,298	1	,130	,849
	CAEC			2,193	3	,533	
	CAEC(1)	,316	,364	,756	1	,385	1,372
	CAEC(2)	,448	,584	,589	1	,443	1,565
	CAEC(3)	,480	,338	2,020	1	,155	1,616
	CH2O	,416	,153	7,435	1	,006	1,516
	FAF	-,216	,100	4,680	1	,031	,806
	TUE	,076	,138	,303	1	,582	1,079
	SMOKE(1)	,454	,413	1,213	1	,271	1,575
	Constant	-2,824	,680	17,251	1	<,001	,059

a. Variable(s) entered on step 1: Age, family_history_with_overweight, NCP, FAF, TUE, CAEC, CH2O, SMOKE.

Πίνακας 4.2.1(δ): Μεταβλητές του πρώτου μοντέλου Λογιστικής Παλινδρόμησης

Παρακάτω παρουσιάζεται το Διάγραμμα ROC για το συγκεκριμένο μοντέλο και θα σχολιαστεί στην επόμενη παράγραφο.



Διάγραμμα 4.2.1(α): Καμπύλη ROC δεύτερου μοντέλου Λογιστικής Παλινδρόμησης

Area Under the Curve	
Test Result Variable(s):	
Area	
	,688

Πίνακας 4.2.1(ε): Εμβαδόν κάτω από καμπύλη δεύτερου μοντέλου Λογιστικής Παλινδρόμησης

4.2.2 Δεύτερη εφαρμογή Διωνυμικής Λογιστικής Παλινδρόμησης

Αφαιρούμε όλες τις μη σημαντικές μεταβλητές: Ποσότητα καθημερινής κατανάλωσης κύριων γευμάτων (NCP), Συχνότητα κατανάλωσης φαγητού ανάμεσα στα γεύματα (CAEC), Χρόνος χρήσης τεχνολογικών συσκευών (TUE) και Κάπνισμα (SMOKE) και ο πίνακας ταξινόμησης είναι ο παρακάτω. Η ακρίβεια πρόβλεψης του μοντέλου παραμένει στο 65,8%, αλλά σε σχέση με το προηγούμενο μοντέλο, η δυνατότητα να προβλέπει σωστά τους ανθρώπους με κανονικό βάρος έχει ανέβει κατά 2,1 μονάδες, ενώ η δυνατότητα να προβλέπει σωστά τους ανθρώπους με μη κανονικό βάρος έχει πέσει στο 43% από 45,9%.

Classification Table^a

Observed		Predicted		Percentage Correct
		logistic_obesity 0	1	
Step 1	logistic_obesity 0	230	48	82,7
	1	118	89	43,0
Overall Percentage				65,8

a. The cut value is ,500

Πίνακας 4.2.2(α): Ακρίβεια πρόβλεψης του τελικού μοντέλου Λογιστικής Παλινδρόμησης

Το p-value του ελέγχου Hosmer-Lemeshow, σύμφωνα με τον Πίνακα 4.2.2(β), πέφτει στο 0,066, που είναι μια οριακή τιμή αλλά εφόσον είναι μεγαλύτερη από 0,05, τότε δεν μπορούμε να απορρίψουμε τη μηδενική υπόθεση, άρα το μοντέλο έχει καλή προσαρμογή.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	14,684	8	,066

Πίνακας 4.2.2(β): Έλεγχος καλής προσαρμογής Hosmer και Lemeshow τελικού μοντέλου Λογιστικής Παλινδρόμησης

Στο μοντέλο στο οποίο καταλήγουμε στον Πίνακα 4.2.2(γ), είναι στατιστικά σημαντικές όλες οι μεταβλητές σε επίπεδο σημαντικότητας 5%.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	Age	,073	,016	20,282	1	<,001	1,076	1,042	1,111
	family_history_with_overweight(1)	,463	,201	5,274	1	,022	1,588	1,070	2,358
	CH2O	,411	,149	7,643	1	,006	1,509	1,127	2,019
	FAF	-,224	,098	5,203	1	,023	,799	,659	,969
	Constant	-2,823	,493	32,746	1	<,001	,059		

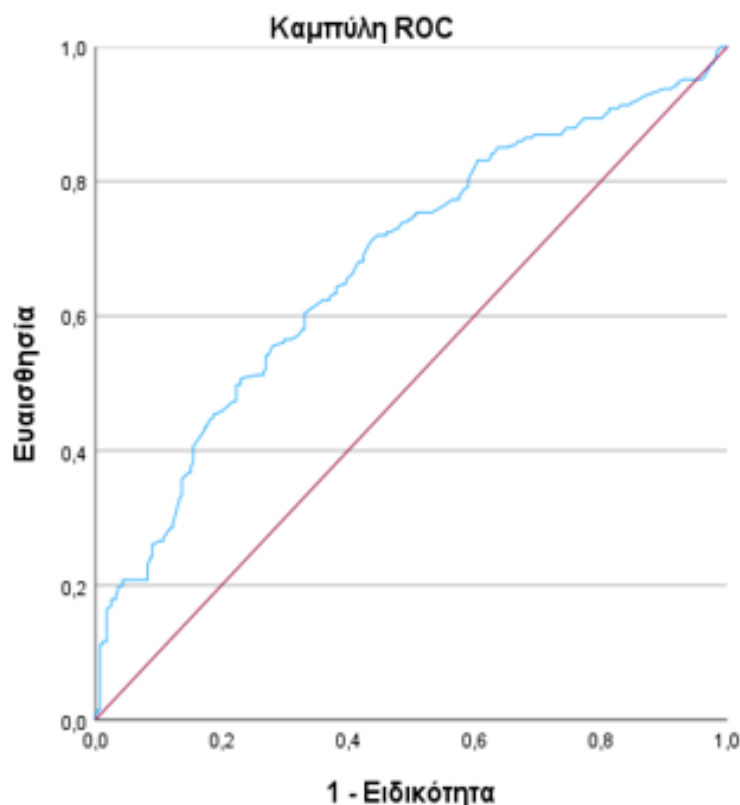
a. Variable(s) entered on step 1: Age, family_history_with_overweight, CH2O, FAF.

Πίνακας 4.2.2(γ): Μεταβλητές του τελικού μοντέλου Λογιστικής Παλινδρόμησης

Στον Πίνακα 4.2.2(γ), οι τιμές B είναι οι εκτιμώμενοι συντελεστές της παλινδρόμησης. Δείχνουν τη μεταβολή στο λογάριθμο των σχετικών πιθανοτήτων για την εμφάνιση παχυσαρκίας, όταν η ανεξάρτητη μεταβλητή αυξάνεται κατά μία μονάδα (για συνεχείς μεταβλητές) ή σε σχέση με την κατηγορία αναφοράς (για κατηγορικές). Οι τιμές Exp(B) είναι το odds ratio (OR), δηλαδή ο λόγος σχετικών πιθανοτήτων και δείχνουν πόσες φορές αυξάνονται ή μειώνονται οι πιθανότητες παχυσαρκίας σε κάθε επίπεδο της μεταβλητής σε σχέση με την κατηγορία αναφοράς. Συγκεκριμένα:

- Για την ηλικία έχουμε $B=0,073$, $\text{Exp}(B)=1,076$. Οπότε το $\text{OR}>1$ σημαίνει ότι κάθε αύξηση ηλικίας κατά 1 έτος αυξάνει τις πιθανότητες κάποιος να μην έχει κανονικό βάρος κατά 7,6% ($1,076 - 1 = 0,076$).
- Για το Οικογενειακό ιστορικό παχυσαρκίας έχουμε $B=0,463$, $\text{Exp}(B)=1,588$. Δηλαδή το $\text{OR}>1$ εδώ σημαίνει ότι άτομα με οικογενειακό ιστορικό έχουν περίπου 1,6 φορές μεγαλύτερες πιθανότητες να μην έχουν κανονικό βάρος σε σχέση με όσους δεν έχουν ιστορικό.
- Για την Ποσότητα καθημερινής κατανάλωσης νερού (CH2O) έχουμε $B=0,411$, $\text{Exp}(B)=1,509$. Εδώ $\text{OR}>1$, δηλαδή η αυξημένη κατανάλωση νερού σχετίζεται με 1,5 φορές μεγαλύτερες πιθανότητες μη κανονικού βάρους (πιθανότατα γιατί παχύσαρκα και υπέρβαρα άτομα καταναλώνουν περισσότερο νερό).
- Τέλος, για την Συχνότητα εβδομαδιαίας σωματικής άσκησης (FAF) έχουμε $B=-0,224$, $\text{Exp}(B)=0,799$. Οπότε εδώ το $\text{OR}<1$, σημαίνει ότι η αυξημένη φυσική δραστηριότητα μειώνει κατά ~20% ($1 - 0,799 = 0,201$) τις πιθανότητες μη κανονικού βάρους.

Η ακρίβεια πρόβλεψης του τελευταίου μοντέλου είναι παρόμοια με του δεύτερου μοντέλου που είχε όλες τις μεταβλητές εκτός της BMI. Το γεγονός βέβαια ότι το τελευταίο μοντέλο έχει λιγότερες μεταβλητές με την ίδια ακρίβεια, λειτουργεί υπέρ του. Ακόμα, σύμφωνα με τον έλεγχο Hosmer-Lemeshow, και τα δύο μοντέλα έχουν καλή προσαρμογή. Επιπλέον, από τις καμπύλες ROC που παρουσιάζονται στη συνέχεια καθώς και τα AUC (Area Under the Curve), μπορούμε να συμπεράνουμε ότι τα μοντέλα δεν έχουν μεγάλη διαφορά. Οι τιμές των AUC (0,677 για το τελευταίο και 0,688 για το δεύτερο) δείχνουν ότι υπάρχει μια σχετικά αποδεκτή διάκριση μεταξύ των κατηγοριών (παχύσαρκοι, υπέρβαροι και λιποβαρείς, από κανονικούς). Οπότε, η επιλογή του μοντέλου Διωνυμικής Λογιστικής Παλινδρόμησης θα εξαρτηθεί από τις στατιστικά σημαντικές μεταβλητές.



Διάγραμμα 4.2.2(α): Καμπύλη ROC τελευταίου μοντέλου Λογιστικής Παλινδρόμησης

Area Under the Curve	
Test Result Variable(s):	
Area	
	,677

Πίνακας 4.2.2(δ): Εμβαδόν κάτω από καμπύλη τελευταίου μοντέλου Λογιστικής Παλινδρόμησης

Σε μια τελική προσπάθεια να δούμε αν το μοντέλο αυτό είναι το καλύτερο από όλα τα δυνατά, θα εφαρμόσουμε άλλη μία φορά Διωνυμική Λογιστική Παλινδρόμηση αλλά με τη μέθοδο Stepwise. Σε αυτή τη διαδικασία, στο μοντέλο εισήχθησαν αρχικά όλες οι ανεξάρτητες μεταβλητές που είχαν θεωρηθεί ως υποψήφιοι προβλεπτικοί παράγοντες (δημογραφικά χαρακτηριστικά, διατροφικές συνήθειες και δείκτες φυσικής δραστηριότητας). Στη συνέχεια, το SPSS μέσω της διαδικασίας Forward: LR αξιολόγησε διαδοχικά τη συνεισφορά κάθε μεταβλητής στην προσαρμογή του μοντέλου και κράτησε μόνο εκείνες που βελτίωναν σημαντικά την προβλεπτική του ικανότητα.

Η μέθοδος stepwise έχει το πλεονέκτημα ότι οδηγεί σε ένα απλούστερο μοντέλο, αποφεύγοντας την υπερπροσαρμογή (overfitting) και λαμβάνοντας υπόψη τις συσχετίσεις μεταξύ των ανεξάρτητων μεταβλητών. Έτσι, το τελικό μοντέλο περιλαμβάνει μόνο τους στατιστικά σημαντικούς παράγοντες, οι οποίοι εξηγούν καλύτερα τη διακύμανση της εξαρτημένης μεταβλητής.

Βλέπουμε στον επόμενο πίνακα ότι το SPSS έκανε 4 βήματα και κράτησε τελικά τις μεταβλητές: Ηλικία, Οικογενειακό ιστορικό παχυσαρκίας, CH2O (καθημερινή κατανάλωση νερού), FAF (συχνότητα φυσικής δραστηριότητας). Δηλαδή, το τελικό μοντέλο είναι το ίδιο στο οποίο καταλήξαμε και με τη μέθοδο Enter και με τους ίδιους συντελεστές.

Variables in the Equation		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Age	,079	,016	24,319	1	<,001	1,083
	Constant	-2,131	,380	31,372	1	<,001	,119
Step 2 ^b	Age	,075	,016	21,948	1	<,001	1,078
	family_history_with_overweight(1)	-,534	,198	7,259	1	,007	,586
	Constant	-1,838	,392	21,929	1	<,001	,159
Step 3 ^c	Age	,076	,016	21,697	1	<,001	1,079
	family_history_with_overweight(1)	-,491	,200	6,027	1	,014	,612
	CH2O	,331	,144	5,305	1	,021	1,392
	Constant	-2,504	,496	25,453	1	<,001	,082
Step 4 ^d	Age	,073	,016	20,282	1	<,001	1,076
	family_history_with_overweight(1)	-,463	,201	5,274	1	,022	,630
	CH2O	,411	,149	7,643	1	,006	1,509
	FAF	-,224	,098	5,203	1	,023	,799
	Constant	-2,360	,501	22,201	1	<,001	,094

a. Variable(s) entered on step 1: Age.

b. Variable(s) entered on step 2: family_history_with_overweight.

c. Variable(s) entered on step 3: CH2O.

d. Variable(s) entered on step 4: FAF.

Πίνακας 4.2.2(ε): Μεταβλητές των μοντέλων με τη μέθοδο Stepwise

Ο Πίνακας 4.2.2(στ) παρουσιάζει τους δείκτες προσαρμογής του μοντέλου σε κάθε βήμα της μεθόδου stepwise. Παρατηρείται ότι η τιμή του -2 Log Likelihood μειώνεται σταδιακά από 633,01 στο πρώτο βήμα σε 614,997 στο τέταρτο, γεγονός που δείχνει βελτίωση της προσαρμογής του μοντέλου με την προσθήκη μεταβλητών. Παράλληλα, ο

δείκτης Nagelkerke R^2 αυξάνεται σταδιακά, με τον τελευταίο να φτάνει το 0,124 στο τελικό μοντέλο. Αυτό σημαίνει ότι περίπου το 12,4% της διακύμανσης της εξαρτημένης μεταβλητής ερμηνεύεται από τις ανεξάρτητες μεταβλητές που παρέμειναν στο μοντέλο.

Model Summary		
Step	-2 Log likelihood	Nagelkerke R Square
1	633,010 ^a	,078
2	625,646 ^a	,097
3	620,282 ^a	,110
4	614,997 ^a	,124

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Πίνακας 4.2.2(στ): Δείκτες προσαρμογής του μοντέλου με τη μέθοδο Stepwise

Τέλος, στον Πίνακα 4.2.2(ζ) παρατηρούμε ότι η ακρίβεια ταξινόμησης για τα άτομα φυσιολογικού βάρους παραμένει υψηλή σε όλα τα βήματα (90,6% στο πρώτο βήμα, 82,7% στο τέταρτο). Αντίθετα, η ικανότητα πρόβλεψης για τα άτομα που ανήκουν στα άλλα επίπεδα παχυσαρκίας, είναι χαμηλότερη, αν και βελτιώνεται ελαφρώς στο τελικό μοντέλο (από 30,0% στο πρώτο βήμα σε 43,0% στο τέταρτο). Συνολικά, στο ποσοστό σωστής ταξινόμησης (Overall Percentage) βλέπουμε ότι υπάρχει μια μικρή αύξηση. Γενικά όμως κυμαίνεται γύρω στο 65%, γεγονός που δείχνει ότι το μοντέλο έχει μέτρια ικανότητα διάκρισης μεταξύ των δύο κατηγοριών.

Classification Table ^a					
Observed			Predicted		Percentage Correct
			logistic_obesity 0	1	
Step 1	logistic_obesity	0	252	26	90,6
		1	145	62	30,0
	Overall Percentage				64,7
Step 2	logistic_obesity	0	246	32	88,5
		1	145	62	30,0
	Overall Percentage				63,5
Step 3	logistic_obesity	0	234	44	84,2
		1	127	80	38,6
	Overall Percentage				64,7
Step 4	logistic_obesity	0	230	48	82,7
		1	118	89	43,0
	Overall Percentage				65,8

a. The cut value is ,500

Πίνακας 4.2.2(ζ): Ακρίβεια πρόβλεψης μοντέλων με τη μέθοδο Stepwise

4.3 Πολυωνυμική Λογιστική Παλινδρόμηση

Η Πολυωνυμική Λογιστική Παλινδρόμηση είναι μια γενίκευση της Διωνυμικής, όπου η εξαρτημένη μεταβλητή έχει περισσότερες από δύο κατηγορίες. Όπως και στη Διωνυμική Λογιστική Παλινδρόμηση, η εξαρτημένη περιγράφει τη σχέση μεταξύ αυτής και δύο ή περισσότερων ανεξάρτητων μεταβλητών.

Στη συγκεκριμένη μελέτη, η εξαρτημένη μεταβλητή είναι το επίπεδο παχυσαρκίας «NewObesity» και αποτελείται από τέσσερα επίπεδα όπως τα ορίσαμε στο δεύτερο κεφάλαιο, δηλαδή: Λιπόσαρκος, Κανονικό Βάρος, Υπέρβαρος και Παχύσαρκος. Για να εφαρμόσουμε την Πολυωνυμική Λογιστική Παλινδρόμηση, από εδώ και πέρα θα αναφέρεται ως Πολυωνυμική Παλινδρόμηση, θα δημιουργήσουμε τη νέα μεταβλητή «polynomial_obesity» στην οποία θα ορίσουμε τα επίπεδα ως εξής: «Κανονικό Βάρος» ως 0, «Λιποβαρής» ως 1, «Υπέρβαρος» ως 2 και «Παχύσαρκος» ως 3. Σε όλα τα μοντέλα που θα ακολουθήσουν, θα χρησιμοποιήσουμε ως κατηγορία αναφοράς το «Κανονικό Βάρος» για να δούμε πόσο αυξάνει ο κίνδυνος εμφάνισης παχυσαρκίας για τις άλλες κατηγορίες, σε σχέση με το φυσιολογικό. Οι συχνότητες των επιπέδων παχυσαρκίας είναι παρόμοιες με τη μεταβλητή «NewObesity» που είχαμε ορίσει στο δεύτερο κεφάλαιο και τα στατιστικά στοιχεία όπως και το διάγραμμα κατανομής, είναι παρόμοια με τον Πίνακα και το Διάγραμμα 2.5.2 αντίστοιχα.

Η κατηγορία «Κανονικό βάρος», με ποσοστό 57,3% είναι η πιο συχνή, η «Overweight» είναι η δεύτερη πιο συχνή με 23,5%, η κατηγορία «Obesity» με ποσοστό 12,4% είναι η τρίτη πιο συχνή και η κατηγορία «Λιποβαρής» με ποσοστό 6,8%, είναι η λιγότερο συχνή.

4.3.1 Πρώτη εφαρμογή Πολυωνυμικής Λογιστικής Παλινδρόμησης

Όπως και στη Διωνυμική Λογιστική Παλινδρόμηση, αρχικά θα χρησιμοποιήσουμε όλες τις ανεξάρτητες μεταβλητές για να ελέγξουμε ποιες είναι σημαντικές και ποιες όχι, καθώς και αν το μοντέλο έχει καλή προσαρμογή. Ο Πίνακας 4.3.1(α), παρουσιάζει τη σύγκριση μεταξύ του τελικού μοντέλου Πολυωνυμικής λογιστικής παλινδρόμησης και του αρχικού μοντέλου που περιλαμβάνει μόνο τη σταθερά. Η τιμή του δείκτη προσαρμογής -2 Log Likelihood μειώθηκε από 1031,208 στο αρχικό μοντέλο σε 879,529 στο τελικό μοντέλο. Η μείωση αυτή είναι στατιστικά σημαντική ($p\text{-value} < 0,001$), γεγονός που υποδηλώνει ότι η εισαγωγή των ανεξάρτητων μεταβλητών βελτιώνει σημαντικά την προσαρμογή του μοντέλου. Συνεπώς, το τελικό μοντέλο περιγράφει καλύτερα τη σχέση μεταξύ των ανεξάρτητων μεταβλητών και της εξαρτημένης μεταβλητής.

Model Fitting Information				
Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1031,208			
Final	879,529	151,680	45	<,001

Πίνακας 4.3.1(α): Σύγκριση αρχικού και τελικού μοντέλου (1^η προσπάθεια)

Παρακάτω βλέπουμε τον πίνακα ταξινόμησης ο οποίος παρουσιάζει τη σύγκριση μεταξύ των παρατηρούμενων και προβλεπόμενων κατηγοριών της εξαρτημένης

μεταβλητής. Το μοντέλο κατηγοριοποίησε σωστά το 61% των περιπτώσεων συνολικά. Η ακρίβεια ταξινόμησης ήταν ιδιαίτερα υψηλή για την κατηγορία «Κανονικό Βάρος» (94,2%), ενώ ήταν πολύ χαμηλή για τις υπόλοιπες κατηγορίες: 3% για την κατηγορία «Λιποβαρής», 18,4% για την κατηγορία «Υπέρβαρος» και 20% για την κατηγορία «Παχύσαρκος». Το αποτέλεσμα αυτό δείχνει ότι το μοντέλο είναι αποτελεσματικό στην πρόβλεψη των ανθρώπων με κανονικό βάρος, αλλά παρουσιάζει χαμηλή ακρίβεια στις υπόλοιπες κατηγορίες.

Classification					
Observed	Predicted				Percent Correct
	0	1	2	3	
0	262	1	11	4	94,2%
1	30	1	1	1	3,0%
2	86	0	21	7	18,4%
3	31	0	17	12	20,0%
Overall Percentage	84,3%	0,4%	10,3%	4,9%	61,0%

Πίνακας 4.3.1(β): Ακρίβεια πρόβλεψης του πρώτου μοντέλου Πολυωνυμικής Παλινδρόμησης

Ο Πίνακας 4.3.1(γ) δείχνει τη συμβολή κάθε ανεξάρτητης μεταβλητής στο μοντέλο μέσω της διαφοράς -2 Log Likelihood, μεταξύ του πλήρους μοντέλου και μειωμένων μοντέλων στα οποία κάθε φορά αφαιρείται μία μεταβλητή. Οι μεταβλητές με στατιστικά σημαντική τιμή (Significance < 0,05) θεωρούνται ότι προσφέρουν ουσιαστική βελτίωση στο μοντέλο.

Συγκεκριμένα, οι μεταβλητές Ηλικία (Age), Οικογενειακό ιστορικό παχυσαρκίας (family history with overweight), Συχνότητα κατανάλωσης τροφής ανάμεσα στα γεύματα (CAEC), Ποσότητα καθημερινής κατανάλωσης νερού (CH2O) και Συχνότητα εβδομαδιαίας σωματικής άσκησης (FAF) βρέθηκαν στατιστικά σημαντικές, υποδηλώνοντας ότι συμβάλλουν σημαντικά στην πρόβλεψη της εξαρτημένης μεταβλητής. Αντίθετα, οι μεταβλητές Ποσότητα καθημερινής κατανάλωσης κύριων γευμάτων (NCP), Χρόνος χρήσης τεχνολογικών συσκευών (TUE) και Κάπνισμα (SMOKE), δεν παρουσίασαν στατιστικά σημαντική επίδραση ($p\text{-value} > 0,05$), γεγονός που υποδηλώνει ότι η αφαίρεσή τους δεν οδηγεί σε σημαντική επιδείνωση της προσαρμογής του μοντέλου.

Likelihood Ratio Tests				
Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	879,529	,000	0	.
Age	913,901	34,372	3	<,001
family_history_with_overweight	896,197	16,668	3	<,001
NCP	889,396	9,867	6	,130
CAEC	899,493	19,964	9	,018
CH20	897,262	17,733	6	,007
FAF	895,756	16,228	9	,062
TUE	887,236	7,708	6	,260
SMOKE	883,705	4,176	3	,243

Πίνακας 4.3.1(γ): Μεταβλητές του πρώτου μοντέλου Πολυωνυμικής Παλινδρόμησης

Ο Πίνακας 4.3.1(δ) δείχνει διάφορους δείκτες που προσεγγίζουν το ποσοστό της διακύμανσης που εξηγείται από το μοντέλο. Ο δείκτης Nagelkerke R^2 , ο οποίος προσαρμόζει τη τιμή του Cox and Snell ώστε να φτάνει θεωρητικά το 1, έχει τιμή 0,302, γεγονός που υποδηλώνει ότι το μοντέλο εξηγεί περίπου 30,2% της διακύμανσης της εξαρτημένης μεταβλητής. Η τιμή αυτή θεωρείται ικανοποιητική για την προσαρμογή του μοντέλου, ειδικά σε κοινωνικές και βιοϊατρικές επιστήμες όπου τα φαινόμενα είναι πολυπαραγοντικά.

Pseudo R-Square	
Cox and Snell	,269
Nagelkerke	,302
McFadden	,142

Πίνακας 4.3.1(δ): Δείκτες Ψευδο- R^2 του 1^{ου} μοντέλου Πολυωνυμικής Παλινδρόμησης

4.3.2 Ενδιάμεσες εφαρμογές Πολυωνυμικής Λογιστικής Παλινδρόμησης

Σε αυτή την ενότητα παρουσιάζονται συνοπτικά τα ενδιάμεσα στάδια που οδήγησαν στη διαμόρφωση του τελικού μοντέλου. Αρχικά, εφαρμόστηκε διαδικασία βηματικής αφαίρεσης, κατά την οποία απομακρύνονταν διαδοχικά οι μεταβλητές που δεν παρουσίαζαν στατιστική σημαντικότητα ($p\text{-value} > 0,05$). Στόχος ήταν να διαπιστωθεί κατά πόσο η αφαίρεση κάποιας μη σημαντικής μεταβλητής θα καθιστούσε άλλες μεταβλητές σημαντικές ή θα βελτίωνε τη συνολική απόδοση του μοντέλου. Στο πρώτο αυτό στάδιο, αφαιρέθηκε η μεταβλητή Χρόνος χρήσης τεχνολογικών συσκευών (TUE), η οποία δεν εμφάνιζε στατιστική σημαντικότητα.

- **Αφαίρεση μεταβλητής TUE**

Στον Πίνακα 4.3.2(α), παρατηρούμε ότι η τιμή του δείκτη προσαρμογής -2 Log Likelihood μειώθηκε από 1001,965 στο αρχικό μοντέλο σε 857,993 στο τελικό μοντέλο. Η μείωση αυτή είναι στατιστικά σημαντική ($p\text{-value} < 0,001$), γεγονός που υποδηλώνει ότι η εισαγωγή των ανεξάρτητων μεταβλητών βελτιώνει σημαντικά την προσαρμογή του μοντέλου. Συνεπώς, το τελικό μοντέλο περιγράφει καλύτερα τη σχέση μεταξύ των ανεξάρτητων μεταβλητών και της εξαρτημένης.

Model Fitting Information				
Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1001,965			
Final	857,993	143,972	39	<,001

Πίνακας 4.3.2(α): Σύγκριση αρχικού και τελικού μοντέλου (2^η προσπάθεια)

Στον παρακάτω πίνακα ταξινόμησης, βλέπουμε ότι το μοντέλο κατηγοριοποίησε σωστά το 61,2% των περιπτώσεων συνολικά, δηλαδή η προβλεπτική του ικανότητα αυξήθηκε. Η ακρίβεια ταξινόμησης για τις κατηγορίες «Κανονικό Βάρος», «Υπέρβαρος» και «Παχύσαρκος» παρέμεινε σταθερή στο 94,2%, 18,4% και 20% αντίστοιχα, ενώ αυξήθηκε για την κατηγορία «Λιποβαρής» κατά 3,1 μονάδες.

Classification					
Observed	Predicted				Percent Correct
	0	1	2	3	
0	262	0	10	6	94,2%
1	29	2	1	1	6,1%
2	84	0	21	9	18,4%
3	35	0	13	12	20,0%
Overall Percentage	84,5%	0,4%	9,3%	5,8%	61,2%

Πίνακας 4.3.2(β): Ακρίβεια πρόβλεψης του δεύτερου μοντέλου Πολυωνυμικής Παλινδρόμησης

Στον Πίνακα 4.3.2(γ) βλέπουμε τη συμβολή κάθε ανεξάρτητης μεταβλητής στο τρίτο μοντέλο Πολυωνυμικής Παλινδρόμησης. Οι μεταβλητές που είναι στατιστικά σημαντικές για το μοντέλο είναι η Ηλικία, το Οικογενειακό ιστορικό παχυσαρκίας, η Ποσότητα καθημερινής κατανάλωσης νερού (CH₂O) και η Συχνότητα κατανάλωσης φαγητού ανάμεσα στα γεύματα (CAEC) ενώ η Συχνότητα εβδομαδιαίας σωματικής άσκησης (FAF) είναι οριακά σημαντική. Οπότε, τη συγκεκριμένη μεταβλητή δεν θα την αφαιρέσουμε προς το παρόν για να δούμε τι θα συμβεί όταν θα αφαιρέσουμε τη μεταβλητή Κάπνισμα (SMOKE) η οποία δεν είναι στατιστικά σημαντική.

Likelihood Ratio Tests				
Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	857,993	,000	0	.
Age	895,972	37,979	3	<,001
family_history_with_overweight	874,909	16,916	3	<,001
NCP	868,379	10,386	6	,109
CAEC	877,097	19,104	9	,024
CH2O	875,912	17,920	6	,006
FAF	875,159	17,167	9	,046
SMOKE	862,332	4,339	3	,227

Πίνακας 4.3.2(γ): Μεταβλητές του δεύτερου μοντέλου Πολυωνυμικής Παλινδρόμησης

Στον Πίνακα 4.3.2(δ), ο δείκτης Nagelkerke R^2 έχει τιμή 0,289, γεγονός που υποδηλώνει ότι το μοντέλο εξηγεί περίπου 28,9% της διακύμανσης της εξαρτημένης μεταβλητής. Η τιμή αυτή θεωρείται ικανοποιητική, δηλαδή το μοντέλο έχει καλή προσαρμογή.

Pseudo R-Square	
Cox and Snell	,257
Nagelkerke	,289
McFadden	,135

Πίνακας 4.3.2(δ): Δείκτες Ψευδο- R^2 του 2^{ου} μοντέλου Πολυωνυμικής Παλινδρόμησης

- **Αφαίρεση μεταβλητής SMOKE**

Στον Πίνακα 4.3.3(ε), το μοντέλο αξιολογήθηκε ως προς την καταλληλότητά του χρησιμοποιώντας τον αρνητικό διπλάσιο λογάριθμο πιθανοφάνειας (-2 Log Likelihood), ο οποίος μειώνεται όσο το μοντέλο ταιριάζει καλύτερα στα δεδομένα. Εδώ, η τιμή αυτή μειώθηκε από 996,42 στο αρχικό μοντέλο σε 856,787 στο τελικό. Η μείωση αυτή είναι στατιστικά σημαντική ($p\text{-value} < 0,001$), γεγονός που υποδηλώνει ότι η εισαγωγή των ανεξάρτητων μεταβλητών βελτιώνει σημαντικά την προσαρμογή του μοντέλου. Συνεπώς, το τελικό μοντέλο περιγράφει καλύτερα τη σχέση μεταξύ των ανεξάρτητων μεταβλητών και της εξαρτημένης μεταβλητής.

Model Fitting Information				
Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	996,420			
Final	856,787	139,633	36	<,001

Πίνακας 4.3.2(ε): Σύγκριση αρχικού και τελικού μοντέλου (3^η προσπάθεια)

Στον παρακάτω πίνακα ταξινόμησης, βλέπουμε ότι το μοντέλο κατηγοριοποίησε σωστά το 60,6% των περιπτώσεων συνολικά, δηλαδή η προβλεπτική του ικανότητα μειώθηκε σε σχέση με τα προηγούμενα μοντέλα. Η ακρίβεια ταξινόμησης μειώθηκε για τις κατηγορίες «Κανονικό Βάρος» και «Παχύσαρκος» στα 93,9% και 13,3% αντίστοιχα, ενώ παρέμεινε σταθερή για την κατηγορία «Λιποβαρής» σε σχέση με το προηγούμενο μοντέλο και αυξήθηκε για την κατηγορία «Υπέρβαρος» κατά 2 μονάδες περίπου.

Classification					
Observed	Predicted				Percent Correct
	0	1	2	3	
0	261	0	10	7	93,9%
1	29	2	1	1	6,1%
2	85	0	23	6	20,2%
3	37	0	15	8	13,3%
Overall Percentage	84,9%	0,4%	10,1%	4,5%	60,6%

Πίνακας 4.3.2(στ): Ακρίβεια πρόβλεψης του τρίτου μοντέλου Πολυωνυμικής Παλινδρόμησης

Στον Πίνακα 4.3.2(ζ) βλέπουμε τη συμβολή κάθε ανεξάρτητης μεταβλητής στην τέταρτη προσπάθεια να χτίσουμε ένα μοντέλο Πολυωνυμικής Παλινδρόμησης. Οι μεταβλητές που είναι στατιστικά σημαντικές για το μοντέλο είναι η Ηλικία, το Οικογενειακό ιστορικό παχυσαρκίας, η Ποσότητα καθημερινής κατανάλωσης νερού (CH2O) και η Συχνότητα κατανάλωσης φαγητού ανάμεσα στα γεύματα (CAEC) ενώ η Συχνότητα εβδομαδιαίας σωματικής άσκησης (FAF) είναι οριακά σημαντική. Οπότε, τη συγκεκριμένη μεταβλητή δεν θα την αφαιρέσουμε προς το παρόν για να δούμε τι θα συμβεί όταν θα αφαιρέσουμε τη μεταβλητή Ποσότητα καθημερινής κατανάλωσης κύριων γευμάτων (NCP) η οποία δεν είναι στατιστικά σημαντική.

Likelihood Ratio Tests				
Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	856,787	,000	0	.
Age	897,090	40,304	3	<,001
family_history_with_overweight	876,019	19,232	3	<,001
NCP	866,923	10,136	6	,119
CAEC	875,787	19,001	9	,025
CH2O	874,942	18,155	6	,006
FAF	873,940	17,153	9	,046

Πίνακας 4.3.2(ζ): Μεταβλητές του τρίτου μοντέλου Πολυωνυμικής Παλινδρόμησης

Τέλος, στον Πίνακα 4.3.2(η), η τιμή του δείκτη Nagelkerke R^2 έχει πέσει στο 0,281, γεγονός που υποδηλώνει ότι το μοντέλο εξηγεί περίπου 28,1% της διακύμανσης της εξαρτημένης μεταβλητής. Η τιμή αυτή, παρόλο που είναι πιο κάτω σε σχέση με των προηγούμενων μοντέλων, θεωρείται ικανοποιητική, δηλαδή το μοντέλο έχει καλή προσαρμογή.

Pseudo R-Square	
Cox and Snell	,250
Nagelkerke	,281
McFadden	,131

Πίνακας 4.3.2(η): Δείκτες Ψευδο- R^2 του 3^{ου} μοντέλου Πολυωνυμικής Παλινδρόμησης

4.3.3 Τελευταία εφαρμογή Πολυωνυμικής Λογιστικής Παλινδρόμησης

Για το τελευταίο μοντέλο, θα αφαιρέσουμε τη μεταβλητή NCP και παρατηρούμε ότι η τιμή του δείκτη προσαρμογής -2 Log Likelihood μειώθηκε από 944,263 στο αρχικό μοντέλο σε 814,767 στο τελικό μοντέλο. Η μείωση αυτή είναι στατιστικά σημαντική (p -value < 0,001), γεγονός που υποδηλώνει ότι η εισαγωγή των ανεξάρτητων μεταβλητών βελτιώνει σημαντικά την προσαρμογή του μοντέλου. Συνεπώς, το τελικό μοντέλο περιγράφει καλύτερα τη σχέση μεταξύ των ανεξάρτητων μεταβλητών και της εξαρτημένης μεταβλητής.

Model Fitting Information				
Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	944,263			
Final	814,767	129,497	30	<,001

Πίνακας 4.3.3(α): Σύγκριση αρχικού και τελικού μοντέλου (τελική προσπάθεια)

Στον παρακάτω πίνακα ταξινόμησης, βλέπουμε ότι το μοντέλο κατηγοριοποίησε σωστά το 59,4% των περιπτώσεων συνολικά, δηλαδή η προβλεπτική του ικανότητα μειώθηκε λίγο σε σχέση με το προηγούμενο μοντέλο. Η ακρίβεια ταξινόμησης για την κατηγορία «Κανονικό Βάρος» αυξήθηκε και έφτασε στο επίπεδο που ήταν στο πρώτο μοντέλο, ενώ για τους λιποβαρείς έμεινε σταθερή σε σχέση με το πρώτο μοντέλο, αλλά μειώθηκε κατά τρεις μονάδες σε σχέση με το προηγούμενο. Για τις κατηγορίες «Υπέρβαρος» και «Παχύσαρκος», η ακρίβεια μειώθηκε σε σχέση και με το προηγούμενο και με το αρχικό μοντέλο. Το αποτέλεσμα αυτό δείχνει ότι το μοντέλο είναι αποτελεσματικό στην πρόβλεψη των ανθρώπων με κανονικό βάρος, αλλά παρουσιάζει χαμηλή ακρίβεια στις υπόλοιπες κατηγορίες.

Classification					
Observed	Predicted				Percent Correct
	0	1	2	3	
0	262	0	10	6	94,2%
1	30	1	1	1	3,0%
2	90	0	19	5	16,7%
3	40	0	14	6	10,0%
Overall Percentage	87,0%	0,2%	9,1%	3,7%	59,4%

Πίνακας 4.3.3(β): Ακρίβεια πρόβλεψης του τελικού μοντέλου Πολυωνυμικής Παλινδρόμησης

Στον Πίνακα 4.3.3(γ), η τιμή του δείκτη Nagelkerke R^2 έχει πέσει στο 0,263, γεγονός που υποδηλώνει ότι το μοντέλο εξηγεί περίπου το 26,3% της διακύμανσης της εξαρτημένης μεταβλητής. Η τιμή αυτή, παρόλο που είναι πιο κάτω σε σχέση με των προηγούμενων μοντέλων, θεωρείται ικανοποιητική, δηλαδή το μοντέλο έχει μέτρια προς καλή προσαρμογή.

Pseudo R-Square	
Cox and Snell	,234
Nagelkerke	,263
McFadden	,121

Πίνακας 4.3.3(γ): Δείκτες Ψευδο- R^2 του τελικού μοντέλου Πολυωνυμικής Παλινδρόμησης

Τέλος, ο Πίνακας 4.3.3(δ) δείχνει τη συμβολή κάθε ανεξάρτητης μεταβλητής στην τελική προσπάθεια να χτίσουμε ένα μοντέλο Πολυωνυμικής Παλινδρόμησης, όπου έχουμε αφαιρέσει όλες τις μη στατιστικά σημαντικές μεταβλητές. Παρατηρούμε ότι όλες οι μεταβλητές είναι στατιστικά σημαντικές (Significance < 0,05).

Likelihood Ratio Tests				
Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	814,767	,000	0	.
Age	856,941	42,174	3	<,001
family_history_with_overweight	836,665	21,898	3	<,001
CAEC	835,944	21,177	9	,012
CH2O	831,570	16,804	6	,010
FAF	833,876	19,109	9	,024

Πίνακας 4.3.3(δ): Μεταβλητές του τελικού μοντέλου Πολυωνυμικής Παλινδρόμησης

Στη συνέχεια, για κάθε μία από τις κατηγορίες «Λιποβαρής», «Υπέρβαρος» και «Παχύσαρκος» ελέγξαμε τις παραμέτρους (B) για κάθε ανεξάρτητη μεταβλητή, σε σχέση με την κατηγορία αναφοράς «Κανονικό Βάρος». Στον Πίνακα 4.3.3(ε), έχουμε τη σύγκριση των παχύσαρκων με αυτούς που είχαν κανονικό βάρος. Οι τιμές B δείχνουν τη μεταβολή στο λογάριθμο των πιθανοτήτων για την εμφάνιση παχυσαρκίας σε σχέση με τη μεταβλητή αναφοράς, ενώ οι τιμές $\text{Exp}(B)$ αντιστοιχούν στον σχετικό λόγο πιθανοτήτων για κάθε επίπεδο της εκάστοτε μεταβλητής. Για παράδειγμα, για τη μεταβλητή Age, με $p\text{-value} < 0,001$ η τιμή $B = 0,098$ και $\text{Exp}(B) = 1,103$ σημαίνει ότι για κάθε 1 χρόνο αύξησης στην ηλικία, οι πιθανότητες εμφάνισης παχυσαρκίας αυξάνονται περίπου κατά 10,3% ($1,103 - 1 = 0,103$). Για τη μεταβλητή family history with overweight όπου η κατηγορία αναφοράς αποτελείται από αυτούς που είχαν οικογενειακό ιστορικό παχυσαρκίας, βλέπουμε τις τιμές $B = -1,623$, $p\text{-value} < 0,001$, $\text{Exp}(B) = 0,197$. Αυτό σημαίνει ότι άτομα χωρίς οικογενειακό ιστορικό παχυσαρκίας έχουν σημαντικά μικρότερες πιθανότητες να είναι παχύσαρκα (περίπου 80% λιγότερες πιθανότητες) σε σχέση με άτομα που έχουν οικογενειακό ιστορικό. Αντίθετα, η μεταβλητή που αφορά τη σωματική δραστηριότητα (FAF) δεν ανέδειξε στατιστικά σημαντική επίδραση σε όλα τα επίπεδα, αν και η πλήρης απουσία άσκησης (FAF=0) έδειξε αυξημένο σχετικό κίνδυνο εμφάνισης παχυσαρκίας ($\text{OR} \sim 3,4$, $p\text{-value} = 0,033$) σε σχέση με όσους ασκούνται 4 με 5 φορές την εβδομάδα. Όσον αφορά τη μεταβλητή της ποσότητας κατανάλωσης νερού (CH2O) - η οποία είναι στατιστικά σημαντική με $p\text{-value} = 0,02$, για το επίπεδο κατανάλωσης «1 με 2 λίτρα την ημέρα», έχουμε $\text{Exp}(B) = 0,298$ (άρα $\text{OR} = 0,30$ περίπου). Δηλαδή, η μέτρια κατανάλωση νερού μειώνει την πιθανότητα εμφάνισης παχυσαρκίας κατά 70% σε σχέση με την υψηλή. Τέλος, η μεταβλητή CAEC, για το επίπεδο «Συχνά» με $\text{Exp}(B) = 0,314$ (δηλαδή $\text{OR} = 0,31$ περίπου) σημαίνει ότι η συχνή κατανάλωση φαγητού ανάμεσα στα γεύματα μειώνει την πιθανότητα να ανήκει κανείς στην κατηγορία «Παχύσαρκος» κατά 68,6% ($1 - 0,314$) σε σχέση με την κατηγορία «Μερικές φορές».

Parameter Estimates									
polynomial_obesity ^a		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B)	
								Lower Bound	Upper Bound
3	Intercept	-2,974	,749	15,776	1	<,001			
	Age	,098	,022	19,808	1	<,001	1,103	1,056	1,151
	[family_history_with_overweight=no]	-1,623	,413	15,433	1	<,001	,197	,088	,443
	[family_history_with_overweight=yes]	0	.	.	0	.	.	.	.
	[CAEC=Always]	-,238	,473	,253	1	,615	,788	,312	1,992
	[CAEC=Frequently]	-1,157	,434	7,110	1	,008	,314	,134	,736
	[CAEC=no]	-,326	,853	,146	1	,702	,722	,136	3,840
	[CAEC=Sometimes]	0	.	.	0	.	.	.	.
	[CH2O=1]	-,832	,450	3,424	1	,064	,435	,180	1,051
	[CH2O=2]	-1,211	,395	9,388	1	,002	,298	,137	,646
	[CH2O=3]	0	.	.	0	.	.	.	.
	[FAF=0]	1,230	,577	4,546	1	,033	3,421	1,104	10,598
	[FAF=1]	,442	,593	,554	1	,457	1,555	,486	4,972
	[FAF=2]	,468	,612	,585	1	,444	1,597	,481	5,300
	[FAF=3]	0	.	.	0	.	.	.	.

a. The reference category is: 0.

Πίνακας 4.3.3(ε): Εκτίμηση παραμέτρων για το τελικό μοντέλο της Πολυωνυμικής Παλινδρόμησης

Κεφάλαιο 5

Συζήτηση

5.1 Εισαγωγή

Η παχυσαρκία αποτελεί ένα από τα σημαντικότερα προβλήματα δημόσιας υγείας παγκοσμίως, με αυξανόμενες επιπτώσεις σε σωματικό, ψυχολογικό και κοινωνικό επίπεδο. Στο πλαίσιο της παρούσας εργασίας, αξιοποιήθηκε μέρος των δεδομένων της έρευνας “Estimation of obesity levels based on eating habits and physical condition” που διεξήχθη στο Μεξικό, το Περού και την Κολομβία και αφορούν άτομα με διαφορετικό σωματικό βάρος (λιποβαρείς, φυσιολογικού βάρους, υπέρβαροι και παχύσαρκοι). Στόχος είναι η κατανόηση των παραγόντων που σχετίζονται με την παχυσαρκία δίνοντας ιδιαίτερη έμφαση στην επίδραση δημογραφικών, διατροφικών και συμπεριφορικών χαρακτηριστικών. Η ανάλυση επικεντρώθηκε τόσο σε περιγραφικά και συσχετιστικά στατιστικά, όσο και στη διερεύνηση ενός διωνυμικού και ενός πολυωνυμικού λογιστικού μοντέλου παλινδρόμησης, με σκοπό την εξαγωγή συμπερασμάτων σχετικά με την πρόβλεψη της παχυσαρκίας.

5.2 Αποτελέσματα από περιγραφική ανάλυση

Αρχικά, θα παρουσιάσουμε κάποια από τα σημαντικότερα αποτελέσματα που προέκυψαν μετά την περιγραφική ανάλυση του δείγματος μας. Σύμφωνα με το δεύτερο κεφάλαιο, για τα δημογραφικά χαρακτηριστικά παρατηρήσαμε ότι η κατανομή του φύλου στο δείγμα ήταν ισορροπημένη, ενώ η πλειονότητα των συμμετεχόντων κατηγοριοποιήθηκε στο επίπεδο «Κανονικό Βάρος», με τις κατηγορίες «Παχύσαρκος», «Υπέρβαρος» και «Λιποβαρής» να συγκεντρώνουν μικρότερα ποσοστά. Η μέση ηλικία ήταν περίπου 23 έτη, με τυπική απόκλιση 6,7 ετών. Το μέσο ύψος των ανθρώπων που συμμετείχαν στην έρευνα ήταν 1,68 m και το μέσο βάρος ήταν 69,5 kg με τυπική απόκλιση 17 κιλά. Αντίστοιχα, ο μέσος Δείκτης Μάζας Σώματος ήταν 24,34. Ακόμη, ένα μεγάλο ποσοστό του δείγματος είχε οικογενειακό ιστορικό παχυσαρκίας.

Από την περιγραφική διερεύνηση των μεταβλητών που αφορούν τις διατροφικές συνήθειες αλλά και τη φυσική δραστηριότητα, προέκυψε ότι οι συμμετέχοντες με φυσιολογικό βάρος παρουσίαζαν, κατά μέσο όρο, υψηλότερη εβδομαδιαία σωματική δραστηριότητα (FAF) και μεγαλύτερη κατανάλωση νερού (CH2O). Αντίθετα, οι συμμετέχοντες που ανήκαν στις κατηγορίες παχυσαρκίας εμφάνιζαν αυξημένα ποσοστά θετικού οικογενειακού ιστορικού παχυσαρκίας και συχνότερη κατανάλωση φαγητού εκτός κύριων γευμάτων (CAEC). Παρόλα αυτά όμως, άτομα που καταναλώνουν ένα με δύο γεύματα την ημέρα (NCP) εμφανίζουν και αυτά, υψηλότερα ποσοστά στις κατηγορίες παχυσαρκίας. Επίσης, η πλειοψηφία των παχύσαρκων δεν κάνουν καθημερινή καταμέτρηση των θερμίδων που καταναλώνουν (SCC), ενώ καταναλώνουν τροφή με πολλές θερμίδες (FAVC) και μερικές φορές καταναλώνουν λαχανικά μέσα στα γεύματα (FCVC) και αλκοόλ (CALC). Τέλος, το μεγαλύτερο ποσοστό των παχύσαρκων δεν αφιερώνουν πολύ χρόνο σε τεχνολογικά μέσα (TUE), δεν καπνίζουν (SMOKE) και χρησιμοποιούν δημόσιες συγκοινωνίες (MTRANS).

Τα boxplots ανέδειξαν διαφορές κυρίως στο βάρος (Weight) και την ηλικία (Age) και μικρότερες διαφορές στο ύψος (Height) μεταξύ των επιπέδων παχυσαρκίας, υποδηλώνοντας την ανάγκη για περαιτέρω επαγωγική διερεύνηση των σχέσεων αυτών. Από την ανάλυση της μεταβλητής της ηλικίας προέκυψαν στατιστικά σημαντικές διαφορές μεταξύ των ομάδων παχυσαρκίας (Kruskal-Wallis p-value < 0,05), με την τάση να δείχνει ότι τα υψηλότερα επίπεδα παχυσαρκίας σχετίζονται με μεγαλύτερες ηλικίες.

5.3 Συσχετίσεις

Στη συγκεκριμένη εργασία χρησιμοποιήθηκαν οι έλεγχοι ανεξαρτησίας χ^2 του Pearson, καθώς και οι έλεγχοι Kolmogorov-Smirnov και Kruskal-Wallis για συνεχείς μεταβλητές, οι οποίοι έδειξαν στατιστικά σημαντικές συσχετίσεις μεταξύ πολλών παραγόντων και του επιπέδου παχυσαρκίας. Συγκεκριμένα:

- Τα δημογραφικά χαρακτηριστικά που παρουσίασαν στατιστικά σημαντική συσχέτιση με το επίπεδο παχυσαρκίας είναι η ηλικία και το οικογενειακό ιστορικό παχυσαρκίας. Η ένταση της σχέσης, όπως προκύπτει από τον δείκτη Cramer's V (0,232), χαρακτηρίζεται ως μέτρια. Το βάρος και το ύψος προφανώς συσχετίζονται με το επίπεδο παχυσαρκίας, αφού από αυτά προκύπτει ο ΔΜΣ.
- Για τις διατροφικές συνήθειες, σημαντική σχέση διαπιστώθηκε με τις μεταβλητές Ποσότητα καθημερινής κατανάλωσης κύριων γευμάτων (NCP) με μικρής έντασης σχέση με την παχυσαρκία (Cramer's V 0,148), Κατανάλωση φαγητού εκτός γευμάτων (CAEC) με μικρής έντασης σχέση (Cramer's V 0,121) και Ποσότητα καθημερινής κατανάλωσης νερού (CH20) με μικρής έντασης σχέση (Cramer's V 0,126).
- Για τη φυσική δραστηριότητα, οι μεταβλητές που παρουσίασαν στατιστικά σημαντική σχέση με το επίπεδο παχυσαρκίας είναι η Συχνότητα εβδομαδιαίας σωματικής άσκησης (FAF), ο Χρόνος χρήσης τεχνολογικών συσκευών (TUE) και το Κάπνισμα. Όλες οι μεταβλητές παρουσίασαν μικρής έντασης συσχέτιση με την εξαρτημένη μεταβλητή.

5.4 Αποτελέσματα και συμπεράσματα από τις αναλύσεις

Στη συνέχεια, προσαρμόσαμε δύο διαφορετικά μοντέλα στα δεδομένα μας. Το ένα ήταν μοντέλο Διωνυμικής Λογιστικής Παλινδρόμησης, με εξαρτημένη μεταβλητή την κατηγοριοποιημένη μεταβλητή παχυσαρκίας (0=Κανονικό Βάρος, 1=υπόλοιπες κατηγορίες). Τρία μοντέλα συγκρίθηκαν:

- Το πρώτο μοντέλο παρουσίασε τιμή AIC=630,187, με στατιστικά μη σημαντικές μεταβλητές: Ποσότητα καθημερινής κατανάλωσης κύριων γευμάτων (NCP), Συχνότητα κατανάλωσης φαγητού ανάμεσα στα γεύματα (CAEC), Χρόνος χρήσης τεχνολογικών συσκευών (TUE) και Κάπνισμα (SMOKE). Ο έλεγχος Hosmer-Lemeshow έδωσε p=0,728 που υποδηλώνει καλή προσαρμογή. Η ακρίβεια ταξινόμησης ήταν 65,8% και η τιμή AUC=0,688.
- Στο δεύτερο μοντέλο, όλες οι μεταβλητές ήταν στατιστικά σημαντικές και παρουσίασε τιμή AIC=624,187. Ο έλεγχος Hosmer-Lemeshow έδωσε p=0,066, που υποδηλώνει καλή προσαρμογή. Η ακρίβεια ταξινόμησης ήταν ίδια με του προηγούμενου μοντέλου και η τιμή AUC=0,677. Καθώς στο μοντέλο αυτό έμειναν

μόνο οι μεταβλητές: Ηλικία, Οικογενειακό ιστορικό παχυσαρκίας, Ποσότητα καθημερινής κατανάλωσης νερού (CH2O) και Συχνότητα εβδομαδιαίας σωματικής άσκησης (FAF) και η τιμή AIC είναι μικρότερη από του προηγούμενου μοντέλου, αυτό θα είναι το τελικό μας μοντέλο:

$$\log\left(\frac{p}{1-p}\right) = -2,36 + 0,73 * Age - 0,463 * FH + 0,411 * CH2O - 0,224 * FAF,$$

όπου FH είναι η μεταβλητή του Οικογενειακού ιστορικού παχυσαρκίας.

Στη συνέχεια, εφαρμόστηκε Πολυωνυμική Λογιστική Παλινδρόμηση στην εξαρτημένη μεταβλητή με τέσσερις κατηγορίες (0=Κανονικό Βάρος, 1=Λιποβαρής, 2=Υπέρβαρος, 3=Παχύσαρκος). Ομοίως και σε αυτή την περίπτωση συγκρίναμε τρία μοντέλα. Συγκεκριμένα:

- Το πρώτο μοντέλο είχε τιμή AIC=897,529, με στατιστικά μη σημαντικές μεταβλητές: Ποσότητα καθημερινής κατανάλωσης κύριων γευμάτων (NCP), Χρόνος χρήσης τεχνολογικών συσκευών (TUE) και Κάπνισμα (SMOKE). Ο δείκτης Ψευδό- R² του Nagelkerke είχε τιμή 0,302 που υποδηλώνει καλή προσαρμογή. Η ακρίβεια ταξινόμησης ήταν 61%.
- Το τελευταίο μοντέλο εμφάνισε ικανοποιητική προσαρμογή (Ψευδό- R² Nagelkerke = 0,263). Οι μεταβλητές που ήταν στατιστικά σημαντικές είναι: Ηλικία, Οικογενειακό ιστορικό παχυσαρκίας, Συχνότητα κατανάλωσης φαγητού ανάμεσα στα γεύματα (CAEC), Ποσότητα καθημερινής κατανάλωσης νερού (CH2O) και Συχνότητα εβδομαδιαίας σωματικής άσκησης (FAF). Οι κατηγορικές μεταβλητές μετατράπηκαν σε dummy variables και η εκτίμηση των παραμέτρων έγινε με μέθοδο μέγιστης πιθανοφάνειας. Η ακρίβεια πρόβλεψης του μοντέλου έχει πέσει σε σχέση με το προηγούμενο, στο 59,4%, αλλά η τιμή AIC είναι μικρότερη (826,767) και δεδομένου ότι έχουμε λιγότερες μεταβλητές, αυτό θα είναι το τελικό μοντέλο Πολυωνυμικής Λογιστικής Παλινδρόμησης. Η εξίσωση του μοντέλου για την κατηγορία «Παχύσαρκος» έναντι της κατηγορίας αναφοράς («Κανονικό Βάρος») έχει τη μορφή:

$$\log\left(\frac{P(\text{Obesity})}{P(\text{Normal})}\right) = \beta_0 + \beta_1 * Age + \beta_2 * FH + \beta_3 * CAEC + \beta_4 * CH2O + \beta_5 * FAF$$

όπου FH είναι η μεταβλητή του Οικογενειακού ιστορικού παχυσαρκίας. Εξαιτίας της κατηγορικής φύσης όλων των μεταβλητών, η εξίσωση περιλαμβάνει έναν ξεχωριστό συντελεστή (β_i) για κάθε επίπεδο των μεταβλητών αυτών. Επομένως, δεν παρουσιάζεται ως μία ενιαία εξίσωση, αλλά ερμηνεύονται τα σημαντικά αποτελέσματα με βάση τις εκτιμήσεις των συντελεστών που παρουσιάσαμε στον Πίνακα 4.3.4(ε).

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική Βιβλιογραφία

Αθαναϊλίδου, Ε. (2024) *Διατροφικές ανεπάρκειες και επιπλοκές μετά από βαριατρική επέμβαση*. [Διπλωματική εργασία, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών]. Διαθέσιμο στο: <https://pergamos.lib.uoa.gr/uoa/dl/object/3390450/file.pdf>

Ανδρούτσος, Ο., Μαλακού, Ε., Μανιός, Ι., Μαυρογιάννη, Χ., & Μοσχώνης, Γ. (2022). *Παχυσαρκία και συνοδά καρδιομεταβολικά νοσήματα: Αιτίες – Συνέπειες – Λύσεις*. Διαθέσιμο στο: https://www.dianeosis.org/wp-content/uploads/2022/02/obesity_final11022022.pdf

Ευαγγελάρας, Χ. (2022). *Ανάλυση δεδομένων με τη χρήση στατιστικών πακέτων: Σημειώσεις για το SPSS V19*. Σημειώσεις.

Ζουρνατζή, Β., Καρκανάκη, Α., Κατσίκης, Η., Πανίδης, Δ., Φλωράκης, Δ., & Χατζηδημητρίου, Δ. (2009). *Παχυσαρκία Ι: Ορισμός, ταξινόμηση, αιτιολογία, παθοφυσιολογία*. Διαθέσιμο στο: https://www.iatrikionline.gr/ELL_M_4_2009/3.pdf

Ηλιόπουλος, Γ., & Πολίτης, Κ. (2023). *Γενικευμένα Γραμμικά Μοντέλα*. Σημειώσεις μαθήματος «Γενικευμένα Γραμμικά Μοντέλα».

Λαδά, Φ., Μαυροματάκη, Μ., & Νικολακάκη, Δ. (2017). *Παιδική παχυσαρκία*. Διαθέσιμο στο: https://apothesis.lib.hmu.gr/bitstream/handle/20.500.12688/8276/LadaFoteini_MavromatakiMikaela_NikolakakiDimitra2017.pdf?sequence=5&isAllowed=y

Μιχαλούδη, Μ. Γ. (2018). *Συγκριτική μελέτη της κατανομής πολυμορφισμών που εμπλέκονται στην ανάπτυξη παχυσαρκίας, στην Ελλάδα και σε άλλους πληθυσμούς*. [Διπλωματική εργασία, Πανεπιστήμιο Πειραιώς]. Διαθέσιμο στο: <https://dione.lib.unipi.gr/xmlui/handle/unipi/11445>

Μπερσίμης, Σ. (2022). *Βιοστατιστική και στατιστικές μέθοδοι στην επιδημιολογία*. Πανεπιστημιακές σημειώσεις μαθήματος «Βιοστατιστική και στατιστικές μέθοδοι στην επιδημιολογία».

Ξενόγλωσση Βιβλιογραφία

Barrea, L., Colao, A., Muscogiuri, G., Savastano, S., Verde, L., & Vetrani, C. (2023). *Obesity: a gender-view*. Διαθέσιμο στο: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10859324/>

Bridget, K., Nongnuch, J., & Sirinya, P. (2024). *A comparative analysis of unhealthy food and beverage television advertising to children in Thailand, between 2014 and 2022*. Globalization and Health. Διαθέσιμο στο: <https://globalizationandhealth.biomedcentral.com/articles/10.1186/s12992-023-01007-7>

Chooi, Y. C., Ding, C., & Magkos, F. (2018). *The epidemiology of obesity*. Διαθέσιμο στο: <https://pubmed.ncbi.nlm.nih.gov/30253139/>

Crowder, M., & Hand, D. (1996). *Practical Longitudinal Data Analysis*. Chapman and Hall, Mandras (India).

De la Hoz Manotas, A., & Mendoza Palechor, F. (2019). *Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico*. Universidad de la Costa, CUC, Colombia.

Glidden, D. V., McCulloch, C. E., Shiboski, S. C., & Vittinghoff, E. (2012). *Regression Methods in Biostatistics* (2nd Ed.). Springer Verlag.

James, W. P. T., & Haslam, D. W. (2005). *Obesity*. The Lancet. Διαθέσιμο στο: <https://www.sciencedirect.com/science/article/pii/S0140673605674831#cesec20>

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data* (2nd Ed.). Wiley Interscience.

Manna, S., & Jewkes, A. M. (2015). *Understanding early childhood obesity risks: An empirical study using fuzzy signatures*. Διαθέσιμο στο: https://www.researchgate.net/publication/286591578_Understanding_early_childhood_obesity_risks_An_empirical_study_using_fuzzy_signatures

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models* (2nd Ed.). Chapman and Hall, New York.

Rehman, A., Said, M. S., & Tirthani, E. (2023). *Genetics and Obesity*. In StatPearls. Διαθέσιμο στο: <https://www.ncbi.nlm.nih.gov/books/NBK573068/>

Sorensen, T. I. (1995). *The genetics of obesity*. Metabolism, 44(4), 4–6. Διαθέσιμο στο: <https://pubmed.ncbi.nlm.nih.gov/7674914/>

Ηλεκτρονικές Πηγές

Dianeosis. (2022). *Η παχυσαρκία στην Ελλάδα και οι συνέπειές της*. Διαθέσιμο στο: <https://www.dianeosis.org/2022/02/i-paxysarkia-stin-ellada-kai-oi-synpeies-tis/>

eSurgery. *Παχυσαρκία – Τι είναι;*. Διαθέσιμο στο: https://esurgery.gr/paxysarkia_ti_einai/

eSurgery. *Τύποι παχυσαρκίας*. Διαθέσιμο στο: https://www.esurgery.gr/%CF%84%CF%8D%CF%80%CE%BF%CE%B9_%CF%80%CE%B1%CF%87%CF%85%CF%83%CE%B1%CF%81%CE%BA%CE%AF%CE%B1%CF%82/

Health.com. (2023). *Overweight vs. obese: What's the difference?*. Διαθέσιμο στο: <https://www.health.com/overweight-vs-obese-11708497>

Ιακώβος Θεοδοσίου. *Κατηγορίες παχυσαρκίας*. Διαθέσιμο στο: <https://iakonos-theodosiou.com/pachysarkia/katigories-pachysarkias/>

Kianous Stavros. *Παχυσαρκία: Τρόποι αντιμετώπισης*. Διαθέσιμο στο: <https://kianous-stavros.gr/%CF%80%CE%B1%CF%87%CF%85%CF%83%CE%B1%CF%81%CE%BA%CE%AF%CE%B1-%CF%84%CF%81%CF%8C%CF%80%CE%BF%CE%B9-%CE%B1%CE%BD%CF%84%CE%B9%CE%BC%CE%B5%CF%84%CF%8E%CF%80%CE%B9%CF%83%CE%B7%CF%82/>

Kosmoiatniki. *Παιδική παχυσαρκία*. Διαθέσιμο στο: <https://kosmoiatniki.com/paidiki-pachysarkia/>

Liberal.gr. *Πάνω από τον μέσο όρο στην Ευρώπη τα ποσοστά παχύσαρκων και υπέρβαρων στην Ελλάδα*. Διαθέσιμο στο: <https://www.liberal.gr/ygeia/poy-pano-apo-meso-oro-stin-eyropi-ta-pososta-pahysarkon-kai-yperbaron-stin-ellada>

MyJuniper. (2023). *Obese vs. overweight*. Διαθέσιμο στο: <https://www.myjuniper.com/blog/obese-vs-overweight>

NHS. *Obesity – Causes*. Διαθέσιμο στο: <https://www.nhs.uk/conditions/obesity/causes/>

NumberAnalytics. *How to interpret Nagelkerke R² metrics in regression models*. Διαθέσιμο στο: <https://www.numberanalytics.com/blog/how-to-interpret-nagelkerke-r-squared-metrics-in-regression-models>

Our World in Data. (2021). *Daily per capita caloric supply*. Διαθέσιμο στο: <https://ourworldindata.org/grapher/daily-per-capita-caloric-supply?time=earliest..2021&country=>

Our World in Data. (2021). *Number of deaths by risk factor*. Διαθέσιμο στο: <https://ourworldindata.org/grapher/number-of-deaths-by-risk-factor>

World Health Organization (WHO). (2023). *Obesity*. Διαθέσιμο στο: https://www.who.int/health-topics/obesity#tab=tab_1

World Health Organization (WHO). (2023). *Obesity and overweight*. Διαθέσιμο στο: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>

World Health Organization (WHO) – SEARO. (2023). *Obesity*. Διαθέσιμο στο: <https://www.who.int/southeastasia/health-topics/obesity>

ΠΑΡΑΡΤΗΜΑ

Ερωτήσεις που τέθηκαν στους συμμετέχοντες και οι πιθανές απαντήσεις:

Table 1
Questions of the survey used for initial recollection of information.

Questions	Possible Answers
¿What is your gender?	• Female • Male
¿what is your age?	Numeric value
¿what is your height?	Numeric value in meters
¿what is your weight?	Numeric value in kilograms
¿Has a family member suffered or suffers from overweight?	• Yes • No
¿Do you eat high caloric food frequently?	• Yes • No
¿Do you usually eat vegetables in your meals?	• Never • Sometimes • Always
¿How many main meals do you have daily?	• Between 1 y 2 • Three • More than three
¿Do you eat any food between meals?	• No • Sometimes • Frequently • Always
¿Do you smoke?	• Yes • No
¿How much water do you drink daily?	• Less than a liter • Between 1 and 2 L • More than 2 L
¿Do you monitor the calories you eat daily?	• Yes • No
¿How often do you have physical activity?	• I do not have • 1 or 2 days • 2 or 4 days • 4 or 5 days
¿How much time do you use technological devices such as cell phone, videogames, television, computer and others?	• 0–2 hours • 3–5 hours • More than 5 hours
¿how often do you drink alcohol?	• I do not drink • Sometimes • Frequently • Always
¿Which transportation do you usually use?	• Automobile • Motorbike • Bike • Public Transportation • Walking

Ολοκληρωμένος Πίνακας Εκτίμησης παραμέτρων για το τελικό μοντέλο της Πολυωνυμικής Παλινδρόμησης, για κάθε μία από τις κατηγορίες «Λιποβαρής», «Υπέρβαρος» και «Παχύσαρκος», σε σχέση με την κατηγορία αναφοράς «Κανονικό Βάρος»:

Parameter Estimates									
polynomial_obesity ^a		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B)	
								Lower Bound	Upper Bound
1	Intercept	,048	1,397	,001	1	,973			
	Age	-,111	,060	3,445	1	,063	,895	,795	1,006
	[family_history_with_overweight=no]	,335	,388	,743	1	,389	1,398	,653	2,993
	[family_history_with_overweight=yes]	0 ^b	.	.	0	.	.	.	.
	[CAEC=Always]	-,303	,801	,143	1	,705	,738	,154	3,549
	[CAEC=Frequently]	,946	,419	5,107	1	,024	2,576	1,134	5,855
	[CAEC=no]	1,191	,766	2,418	1	,120	3,291	,733	14,768
	[CAEC=Sometimes]	0 ^b	.	.	0	.	.	.	.
	[CH2O=1]	-,553	,574	,929	1	,335	,575	,187	1,771
	[CH2O=2]	-,832	,499	2,776	1	,096	,435	,164	1,158
	[CH2O=3]	0 ^b	.	.	0	.	.	.	.
	[FAF=0]	,325	,659	,243	1	,622	1,384	,380	5,042
	[FAF=1]	-,592	,692	,732	1	,392	,553	,142	2,148
	[FAF=2]	,658	,618	1,135	1	,287	1,932	,575	6,485
	[FAF=3]	0 ^b	.	.	0	.	.	.	.
2	Intercept	-2,163	,583	13,762	1	<,001			
	Age	,088	,019	22,333	1	<,001	1,092	1,053	1,133
	[family_history_with_overweight=no]	-,406	,247	2,707	1	,100	,666	,411	1,081
	[family_history_with_overweight=yes]	0 ^b	.	.	0	.	.	.	.
	[CAEC=Always]	-,814	,444	3,366	1	,067	,443	,186	1,057
	[CAEC=Frequently]	-,342	,276	1,527	1	,216	,711	,413	1,221
	[CAEC=no]	-,358	,642	,311	1	,577	,699	,198	2,462
	[CAEC=Sometimes]	0 ^b	.	.	0	.	.	.	.
	[CH2O=1]	-1,101	,370	8,848	1	,003	,333	,161	,687
	[CH2O=2]	-,821	,307	7,149	1	,007	,440	,241	,803
	[CH2O=3]	0 ^b	.	.	0	.	.	.	.
	[FAF=0]	,653	,395	2,736	1	,098	1,921	,886	4,164
	[FAF=1]	,310	,389	,635	1	,425	1,363	,636	2,921
	[FAF=2]	-,187	,431	,187	1	,665	,830	,356	1,932
	[FAF=3]	0 ^b	.	.	0	.	.	.	.
3	Intercept	-2,974	,749	15,776	1	<,001			
	Age	,098	,022	19,808	1	<,001	1,103	1,056	1,151
	[family_history_with_overweight=no]	-1,623	,413	15,433	1	<,001	,197	,088	,443
	[family_history_with_overweight=yes]	0 ^b	.	.	0	.	.	.	.
	[CAEC=Always]	-,238	,473	,253	1	,615	,788	,312	1,992
	[CAEC=Frequently]	-1,157	,434	7,110	1	,008	,314	,134	,736
	[CAEC=no]	-,326	,853	,146	1	,702	,722	,136	3,840
	[CAEC=Sometimes]	0 ^b	.	.	0	.	.	.	.
	[CH2O=1]	-,832	,450	3,424	1	,064	,435	,180	1,051
	[CH2O=2]	-1,211	,395	9,388	1	,002	,298	,137	,646
	[CH2O=3]	0 ^b	.	.	0	.	.	.	.
	[FAF=0]	1,230	,577	4,546	1	,033	3,421	1,104	10,598
	[FAF=1]	,442	,593	,554	1	,457	1,555	,486	4,972
	[FAF=2]	,468	,612	,585	1	,444	1,597	,481	5,300
	[FAF=3]	0 ^b	.	.	0	.	.	.	.

a. The reference category is: 0.

b. This parameter is set to zero because it is redundant.