

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

**Σχολή Χρηματοοικονομικής και
Στατιστικής**



Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΣΥΓΚΡΙΤΙΚΗ ΜΕΛΕΤΗ ΤΗΣ ΚΑΤΑΝΟΜΗΣ
ΠΟΛΥΜΟΡΦΙΣΜΩΝ ΠΟΥ ΕΜΠΛΕΚΟΝΤΑΙ ΣΤΗΝ
ΑΝΑΠΤΥΞΗ ΠΑΧΥΣΑΡΚΙΑΣ, ΣΤΗΝ ΕΛΛΑΔΑ ΚΑΙ
ΣΕ ΑΛΛΟΥΣ ΠΛΗΘΥΣΜΟΥΣ**

Μαλβίνα Γ. Μιχαλούδη

Διπλωματική Εργασία

*που υποβλήθηκε στο Τμήμα Στατιστικής και
Ασφαλιστικής Επιστήμης του Πανεπιστημίου
Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση
του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην
Εφαρμοσμένη Στατιστική*

Πειραιάς
Σεπτέμβριος 2018

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:

- Αναπλ. Καθ. Πολίτης Κωνσταντίνος (Επιβλέπων)
- Αναπλ. Καθ. Αντζουλάκος Δημήτριος
- Επικ. Καθ. Τζαβελάς Γεώργιος

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
School of Finance and Statistics



Department of Statistics and Insurance Science

**POSTGRADUATE PROGRAM IN
APPLIED STATISTICS**

**COMPARATIVE STUDY FOR THE
DISTRIBUTION OF POLYMORPHISMS LINKED
WITH THE DEVELOPMENT OF OBESITY,
BETWEEN THE GREEK AND OTHER
POPULATIONS**

By

Malvina G. Michaloudi

MSc Dissertation

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

Piraeus, Greece
September 2018

*Στους γονείς μου
Γιώργο και Δώρα*

Ευχαριστίες

Ευχαριστώ εκ βάθους καρδιάς τον επιβλέποντα Καθηγητή μου, κύριο Πολίτη Κωνσταντίνο, ο οποίος με αφοσίωση με καθοδήγησε σ' όλη την διάρκεια της συγγραφικής μου προσπάθειας στη διπλωματική μου εργασία. Ευχαριστώ επίσης την ερευνητική ομάδα Κλινικής Φαρμακολογίας και Φαρμακογονιδιωματικής του Τμήματος Φαρμακευτικής του Εθνικού Καποδιστριακού Πανεπιστημίου Αθηνών, για το δείγμα που μας παραχώρησαν προς μελέτη.

Περίληψη

Το θέμα το οποίο καλούμαστε να μελετήσουμε είναι, ο τρόπος με τον οποίο ορισμένοι πολυμορφισμοί-συγκεκριμένα σημειακοί πολυμορφισμοί, εμπλέκονται στην ανάπτυξη της παχυσαρκίας. Η παχυσαρκία, “φαινόμενο” το οποίο έχει απασχολήσει και απασχολεί πολλούς ερευνητές απ’ όλο τον κόσμο, προκαλεί σοβαρά προβλήματα στην υγεία των παχύσαρκων ατόμων (επιβαρύνοντας και τις οικονομίες των χωρών), κατατάσσεται δε στις πέντε πρώτες αιτίες θανάτου, σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας (ΠΟΥ). Για τη μείωση ή την αποφυγή των προβλημάτων υγείας που προκαλούνται από την παχυσαρκία, θα πρέπει ο κάθε άνθρωπος ξεχωριστά και με αυτοπειθαρχία να φροντίζει για την ποιότητα της ζωής του και κυρίως της διατροφής του, βασικού παράγοντα υγείας.

Όμως, η παχυσαρκία δεν εξαρτάται πάντα από την ποιότητα ζωής και διατροφής. Σε πολλές κλινικές περιπτώσεις σημαντικό ρόλο παίζουν και οι κληρονομικοί παράγοντες, δηλαδή το DNA. Στην παρούσα εργασία μελετάμε τους σημειακούς πολυμορφισμούς, οι οποίοι είναι μεταλλάξεις στην βάση του DNA του κάθε ατόμου. Οι πολυμορφισμοί είναι σημαντικότεροι και χρήσιμοι στους ερευνητές, καθώς τους βοηθούν στον εντοπισμό γονιδίων, τα οποία ενδεχομένως σχετίζονται με κάποια ασθένεια. Το θέμα το οποίο αναπτύσσουμε στην συνέχεια, δηλαδή η σχέση μεταξύ ορισμένων πολυμορφισμών ως προς την ανάπτυξη της παχυσαρκίας, απασχολεί αρκετούς ερευνητές από όλο τον “ανεπτυγμένο” κυρίως κόσμο, όπου και το ως άνω “φαινόμενο” είναι σε έξαρση!

Τα δεδομένα της ανάλυσης ευγενώς μας παραχωρήθηκαν από την Ερευνητική Ομάδα Κλινικής Φαρμακολογίας και Φαρμακογονιδιωματικής της Φαρμακευτικής Σχολής του Πανεπιστημίου Αθηνών. Στην εργασία μας αυτή, αρχικά, μέσω μιας περιγραφικής ανάλυσης, πήραμε μια γενική εικόνα, τόσο για τις ποιοτικές όσο και για ορισμένες συνεχείς μεταβλητές. Κατόπιν, προχωρήσαμε στην ανάλυση, με σκοπό να ερευνήσουμε τυχόν σχέσεις μεταξύ των μεταβλητών, χρησιμοποιώντας μεθόδους της ανάλυσης διακριτών δεδομένων και των γενικευμένων γραμμικών μοντέλων. Τέλος, εφαρμόσαμε τις μεθόδους της ανάλυσης διακριτών δεδομένων και ερευνήσαμε τυχόν σχέσεις μεταξύ περισσότερων μεταβλητών.

Από τις διεθνείς μελέτες παραθέτουμε ορισμένα σχεσιακά αποτελέσματα που μας βοήθησαν να τα συγκρίνουμε με τα δικά μας.

Abstract

The theme that we are going to examine in the present thesis is how some polymorphisms - specifically some single nucleotide polymorphisms- are connected with obesity. Obesity, is a “phenomenon” which has been examined by many researchers all over the world. According to World Health Organization, is one of the most important causes of death. Obesity provokes serious problems in human’s health. To avoid those health problems that obesity provokes, each one of us must have both a better quality of life and better nutrition. But, obesity doesn’t depend only on our quality of life. It’s also a matter of heredity. Single nucleotide polymorphisms are some mutations that happen in our DNA. These polymorphisms are very important and useful for researchers because they locate some genes that are associated with diseases. The subject of how to connect those polymorphisms with obesity has been examined by many researchers.

The data base used in our analysis were collected by the Pharmaceutical School of Athens. Initially a descriptive analysis provided us a general idea of both nominal and continuous variables. In order to test some relations between our variables we used methods of categorical data analysis and methods of generalized linear models. Finally, using methods of categorical data analysis, we wanted to see if there is an association among more variables simultaneously.

The relationship between some polymorphisms and the phenomenon of obesity have been examined by many researchers all over the world. That’s why we presented some of their results and compared them with ours.

Περιεχόμενα

Abstract.....	xi
Κατάλογος Πινάκων.....	xvii
Κατάλογος Διαγραμμάτων	xx
ΚΕΦΑΛΑΙΟ 1	1
ΕΙΣΑΓΩΓΗ.....	1
1.1 Τι είναι Παχυσαρκία;.....	1
1.1.1 Παράγοντες που οδηγούν στην παχυσαρκία	2
1.1.2 Συνέπειες που προκαλούνται στην υγεία	2
1.1.3 Μπορεί η παχυσαρκία να αντιμετωπιστεί;	5
1.1.4 Πώς μετριέται η παχυσαρκία;.....	7
1.1.5 Στατιστικά στοιχεία γύρω από την παχυσαρκία	9
1.2 Τι είναι ο πολυμορφισμός	13
1.2.1 Πώς συνδέονται οι πολυμορφισμοί με την παχυσαρκία.....	14
ΚΕΦΑΛΑΙΟ 2	16
ΠΙΝΑΚΕΣ ΣΥΝΑΦΕΙΑΣ.....	16
2.1 Κατηγορικά δεδομένα.....	16
2.2 Είδη κατηγορικών δεδομένων.....	17
2.3 Πίνακες Συνάφειας.....	17
2.4 Συμβολισμοί ενός πίνακα συνάφειας	20
2.5 Έλεγχος ανεξαρτησίας	23
ΚΕΦΑΛΑΙΟ 3	28
ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ	28
3.1 Βασικές Έννοιες	28
3.2 Κατηγορίες στατιστικών ελέγχων	29
3.3 Στατιστικοί Έλεγχοι Κανονικότητας	30
3.4 Γραφικοί Έλεγχοι Κανονικότητας.....	30
3.5 Μη παραμετρικοί έλεγχοι υποθέσεων	31
3.5.1 Ο έλεγχος U των Mann-Whitney	31
3.5.2 Ο έλεγχος Sum-Rank του Wilcoxon.....	33
3.5.3 Ο έλεγχος H των Kruskal-Wallis	33
ΚΕΦΑΛΑΙΟ 4	36
ΙΣΟΡΡΟΠΙΑ HARDY-WEINBERG.....	36

4.1 Μια σύντομη θεωρητική προσέγγιση	36
4.2 Εξέταση στον πολυμορφισμό ADRB3rs4994	36
4.3 Εξέταση στον πολυμορφισμό ADRB2rs1042713	38
4.4 Εξέταση στον πολυμορφισμό ADRB2rs1042714	40
4.5 Εξέταση στον πολυμορφισμό ADRB1rs1801252	41
4.6 Εξέταση στον πολυμορφισμό ADRB1rs1801253	42
ΚΕΦΑΛΑΙΟ 5	44
ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ.....	44
5.1 Προέλευση και περιγραφή των δεδομένων	44
5.2 Περιγραφική ανάλυση	45
5.2.1 Φύλο των εθελοντών	46
5.2.2 Πολυμορφισμός ADRB3rs4994.....	46
5.2.3 Πολυμορφισμοί ADRB2rs1042713 και ADRB2rs1042714	47
5.2.4 Πολυμορφισμοί ADRB1rs1801252 και ADRB1rs1801253	48
5.2.5 Ο δείκτης μάζας σώματος (BMI).....	48
5.2.6 Μετασηματισμός του δείκτη μάζας σώματος.....	50
5.2.7 Η κατηγοριοποίηση του BMI (ClassificationBMI)	51
5.2.8 Σχέσεις μεταξύ της classificationBMI, των 5 πολυμορφισμών και του φύλου ...	52
ΚΕΦΑΛΑΙΟ 6	57
ΑΠΑΡΑΜΕΤΡΙΚΗ ΑΝΑΛΥΣΗ	57
6.1 Σχέση του δείκτη μάζας σώματος με το φύλο των εθελοντών	57
6.2 Σχέση του δείκτη μάζας σώματος με τους 5 πολυμορφισμούς.....	60
6.2.1 Σχέση του δείκτη μάζας σώματος με τον ADRB3rs4994.....	60
6.2.2 Σχέση του δείκτη μάζας σώματος με τον ADRB2rs1042713.....	62
6.2.3 Σχέση του δείκτη μάζας σώματος με τον ADRB2rs1042714.....	64
6.2.4 Σχέση του δείκτη μάζας σώματος με τον ADRB1rs1801252.....	66
6.2.5 Σχέση του δείκτη μάζας σώματος με τον ADRB1rs1801253.....	67
ΚΕΦΑΛΑΙΟ 7	70
ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ.....	70
7.1 Περιγραφή και προσαρμογή του πρώτου μοντέλου	70
7.2 Περιγραφή και προσαρμογή του δεύτερου μοντέλου.....	75
7.3 Περιγραφή και προσαρμογή του τρίτου μοντέλου	79
ΚΕΦΑΛΑΙΟ 8	82
ΠΙΝΑΚΕΣ ΣΥΝΑΦΕΙΑΣ ΤΡΙΠΛΗΣ ΕΙΣΟΔΟΥ	82
8.1 Προσαρμογή του πρώτου πίνακα τριπλής εισόδου	82
8.2 Προσαρμογή του δεύτερου πίνακα τριπλής εισόδου.....	85

8.3 Προσαρμογή του τρίτου πίνακα τριπλής εισόδου	88
ΚΕΦΑΛΑΙΟ 9	91
ΣΥΓΚΡΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ	91
9.1 Σύνοψη αποτελεσμάτων	91
9.2 Επισκόπηση διεθνούς βιβλιογραφίας	92
9.3 Σύγκριση αποτελεσμάτων	93
9.4 Επίλογος	93
Παράρτημα	94
Βιβλιογραφία	102

Κατάλογος Πινάκων

Π. 1.1: Κατηγορίες διατροφικής κατάστασης με βάση το BMI	5
Π. 2.1: Γενική μορφή ενός πίνακα συνάφειας	15
Π. 2.2: Ειδική περίπτωση ενός 2×2 πίνακα συνάφειας	17
Π. 4.1: Παρατηρούμενος και αναμενόμενος αριθμός γονοτύπων του rs4994	34
Π. 4.2: Παρατηρούμενος και αναμενόμενος αριθμός γονοτύπων του rs1042713	36
Π. 4.3: Παρατηρούμενος και αναμενόμενος αριθμός γονοτύπων του rs1042714	37
Π. 4.4: Παρατηρούμενος και αναμενόμενος αριθμός γονοτύπων του rs1801252	38
Π. 4.5: Παρατηρούμενος και αναμενόμενος αριθμός γονοτύπων του rs1801253	40
Π. 5.1: Πίνακας συχνοτήτων του φύλου των εθελοντών	43
Π. 5.2: Κατανομή του πολυμορφισμού rs4994	44
Π. 5.3: Κατανομή των πολυμορφισμών rs1042713 και rs1042714	45
Π. 5.4: Κατανομή των πολυμορφισμών rs1801252 και rs1801253	45
Π. 5.5: Περιγραφικά μέτρα για τον δείκτη μάζας σώματος των εθελοντών	46
Π. 5.6: Κατανομή της classification bmi	49
Π. 6.1: Έλεγχος Shapiro-Wilk για το δείγμα των ανδρών	54
Π. 6.2: Έλεγχος Shapiro-Wilk για το δείγμα των γυναικών	55
Π. 6.3: Έλεγχος Mann-Whitney	55
Π. 6.4: Μέση τιμή για τον δείκτη μάζας σώματος ανά φύλο	56
Π. 6.5: Έλεγχος Shapiro-Wilk για τον γονότυπο C:T	58
Π. 6.6: Έλεγχος Shapiro-Wilk για τον γονότυπο T:T	58
Π. 6.7: Έλεγχος Mann-Whitney	59
Π. 6.8: Έλεγχος Shapiro-Wilk για τα επίπεδα της μεταβλητής ADRB2rs1042713	59
Π. 6.9: Έλεγχος Kruskal-Wallis	61

Π. 6.10: Έλεγχος Shapiro-Wilk για τα επίπεδα της μεταβλητής ADRB2rs1042714	61
Π. 6.11: Έλεγχος Kruskal-Wallis	62
Π. 6.12: Έλεγχος Shapiro-Wilk για τα επίπεδα της μεταβλητής ADRB1rs1801252	63
Π. 6.13: Έλεγχος Kruskal-Wallis	64
Π. 6.14: Έλεγχος Shapiro-Wilk για τα επίπεδα της μεταβλητής ADRB1rs1801253	64
Π. 6.15: Έλεγχος Kruskal-Wallis	66
Π. 7.1: Πληροφοριακός πίνακας του μοντέλου	67
Π. 7.2: Πληροφορίες των κατηγορικών μεταβλητών του μοντέλου	68
Π. 7.3: Περιγραφικά μέτρα της συνεχούς επεξηγηματικής μεταβλητής	69
Π. 7.4: Κριτήρια καλής προσαρμογής και μέτρα πληροφορίας	69
Π. 7.5: Omnibus Test	70
Π. 7.6: Έλεγχος Wald	70
Π. 7.7: Εκτιμήσεις των παραμέτρων του μοντέλου	71
Π. 7.8: Έλεγχος καλής προσαρμογής Hosmer-Lemeshow	72
Π. 7.9: Περιγραφικά στοιχεία της εξαρτημένης μεταβλητής	72
Π. 7.10: Μέτρα πληροφορίας και κριτήρια καλής προσαρμογής	73
Π. 7.11: Omnibus Test	73
Π. 7.12: Έλεγχοι Wald	74
Π. 7.13: Εκτιμήσεις των παραμέτρων του μοντέλου	75
Π. 7.14: Έλεγχος Hosmer-Lemeshow	76
Π. 7.15: Μέτρα πληροφορίας και κριτήρια καλής προσαρμογής	76
Π. 7.16: Omnibus Test	77

Π. 7.17: Έλεγχοι Wald για τις επεξηγηματικές μεταβλητές	77
Π. 7.18: Εκτιμήσεις των παραμέτρων του τρίτου μοντέλου	78
Π. 7.19: Έλεγχος Hosmer-Lemeshow	78
Π. 8.1: Weightsituation1 ανά φύλο και ηλικιακή ομάδα	80
Π. 8.2: Έλεγχος Mantel-Haenszel	81
Π. 8.3: Εκτίμηση του σχετικού λόγου πιθανοτήτων για τον έλεγχο M-H	81
Π. 8.4: Έλεγχος Breslow and Day	82
Π. 8.5: Weightsituation2 ανά φύλο και ηλικιακή ομάδα	83
Π. 8.6: Έλεγχος Mantel-Haenszel	84
Π. 8.7: Έλεγχος ομοιόμορφης συνάφειας	84
Π. 8.8: Weightsituation3 ανά φύλο και ηλικιακή ομάδα	85
Π. 8.9: Έλεγχος Mantel-Haenszel	86
Π. 8.10: Έλεγχος Breslow-Day	86

Κατάλογος Διαγραμμάτων

Δ. 5.1: Ραβδόγραμμα της ADRB3rs4994	44
Δ. 5.2: Ιστόγραμμα συχνοτήτων για το BMI	47
Δ. 5.3: Ιστόγραμμα της μεταβλητής BMINEW	48
Δ. 5.4: Ραβδόγραμμα της classificationBMI	49
Δ. 6.1: Q-Q plot ανδρών	54
Δ. 6.2: Q-Q plot γυναικών	55
Δ. 6.3: Ιστόγραμμα του δείκτη μάζας σώματος ανά φύλο	57
Δ. 6.4: Q-Q plot C:T	58
Δ. 6.5: Q-Q plot T:T	58
Δ. 6.6: Q-Q plot A:A, G:A, G:G	60
Δ. 6.7: Q-Q plot C:C, G:C, G:G	62
Δ. 6.8: Q-Q plot A:A, G:A	63
Δ. 6.9: Q-Q plot C:C, G:C, G:G	65

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

Σκοπός του κεφαλαίου αυτού είναι να παρουσιάσουμε με σαφήνεια την έννοια της παχυσαρκίας, τους παράγοντες που οδηγούν σε αυτήν, τις συνέπειες που προκαλεί στην υγεία του ανθρώπου, πώς αυτές μπορούν να αντιμετωπιστούν, καθώς και τον τρόπο μέτρησής της. Παραθέτουμε ορισμένα στατιστικά στοιχεία προκειμένου να δείχτουν διαφορές στα ποσοστά παχυσαρκίας τόσο σε ευρωπαϊκό όσο και σε παγκόσμιο επίπεδο. Τέλος, παρουσιάζουμε έναν ορισμό, στοιχεία για τους σημειακούς πολυμορφισμούς καθώς και τον τρόπο που αυτοί συνδέονται με την παχυσαρκία.

1.1 Τι είναι Παχυσαρκία;

Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας, *“ως παχυσαρκία ορίζουμε την υπερβολική και μη φυσιολογική συσσώρευση λίπους η οποία μπορεί να προκαλέσει σοβαρά προβλήματα στην υγεία, γι’ αυτό και χαρακτηρίζεται ως μια από τις βασικές αιτίες θανάτου.”*



(Πηγή : <https://functionalsystem.com>)

Rank	Cause of Death	Percent of Deaths
1	High Blood Pressure	12.8%
2	Tobacco Use	8.7%
3	High Blood Glucose	5.8%
4	Physical Inactivity	5.5%
5	Overweight & Obesity	4.8%
6	High Cholesterol	4.5%
7	Unsafe Sex	4.0%
8	Alcohol Use	3.8%
9	Childhood Underweight	3.8%
10	Indoor Smoke Solid Fuels	3.3%
Source: WHO		

(Πηγή : <http://www.phitamerica.org/Page2954.aspx>)

1.1.1 Παράγοντες που οδηγούν στην παχυσαρκία

Σημαντικότερος παράγοντας στην ανάπτυξη της παχυσαρκίας στον άνθρωπο είναι οι διατροφικές συνήθειες, οι οποίες καθορίζονται από το οικογενειακό περιβάλλον, δηλαδή από τη διατροφική συμπεριφορά μέσα στην οικογένεια ή είναι ευθύνη του ίδιου του ατόμου.

Βασικός εξ ίσου παράγοντας, σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας, είναι η μη ύπαρξη φυσικής δραστηριότητας λόγω του τρόπου ζωής από την αυξανόμενη αστικοποίηση, από τον φόρτο της καθημερινής εργασίας, από τις ατελείωτες ώρες μπροστά στην οθόνη του υπολογιστή και της τηλεόρασης και γενικά από το καθιστικό μοντέλο ζωής στην δουλειά, τη διασκέδαση, το σπίτι κλπ. Έτσι λοιπόν η κακή διατροφική συμπεριφορά ενός ατόμου (π.χ. fast food, ανθυγιεινές τροφές, υπερβολική κατανάλωση τροφών κλπ) και η έλλειψη φυσικής δραστηριότητας έχουν ως αποτέλεσμα να οδηγηθεί το άτομο στην παχυσαρκία, (<https://thehealthlab.gr>) δηλαδή στην αύξηση του σωματικού του βάρους λόγω συσσώρευσης λίπους στο σώμα του.

Στις αναπτυγμένες κυρίως χώρες καταγράφηκαν:

- Έλλειψη φυσικής δραστηριότητας εξαιτίας της αυξημένης καθιστικής ζωής στο περιβάλλον εργασίας, στο σπίτι και γενικά στον τρόπο ζωής
- Αυξημένη κατανάλωση τροφών με υψηλή περιεκτικότητα σε σάκχαρα και λιπαρά. Κατά συνέπεια αυξημένες-περιττές θερμίδες, οι οποίες αποθηκεύονται ως λίπος στο ανθρώπινο σώμα.

Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας, *βασική αιτία της παχυσαρκίας είναι η ενεργειακή ανισορροπία μεταξύ των θερμίδων που προσλαμβάνουμε και των θερμίδων που καταναλώνουμε, δηλαδή έχουμε θετικό ενεργειακό ισοζύγιο και ως γνωστό η περίσσεια των προσλαμβανόμενων θερμίδων αποθηκεύεται στο σώμα μας υπό μορφή λίπους.*

Το θέμα μας όμως αφορά στους σημειακούς πολυμορφισμούς, τους οποίους θα μελετήσουμε και θα σταθούμε σ' αυτό κυρίως, επισημαίνοντας ότι η κληρονομικότητα, το DNA, είναι επίσης σημαντικός παράγοντας στην εμφάνιση της παχυσαρκίας. Εάν οι γονείς είναι παχύσαρκοι πιθανότατα και το παιδί ή τα παιδιά να γίνουν επίσης παχύσαρκα, γεγονός που παρατηρούμε γύρω μας.

1.1.2 Συνέπειες που προκαλούνται στην υγεία

Η συσσώρευση υπερβολικού λίπους στο σώμα του ανθρώπου προκαλεί προβλήματα υγείας όπως:

- Υψηλή αρτηριακή πίεση
- Καρδιαγγειακές νόσους (προβλήματα καρδιοπάθειας και εγκεφαλικά επεισόδια), βασικές αιτίες θανάτου το 2012 (WHO, 2018).
- Διαβήτη τύπου 2
- Μυοσκελετικές διαταραχές (κυρίως οστεοαρθρίτιδα δηλαδή προβλήματα στις αρθρώσεις)
- Ορισμένες μορφές καρκίνου (όπως προστάτη, μαστών, ωοθηκών, ήπατος, νεφρών, παχέος εντέρου και χοληδόχου κύστεως) (WHO, 2018).
- Διαταραχές ψυχικής ευεξίας έως κατάθλιψη και άλλες ψυχικές διαταραχές.

Η απλή και μόνο ύπαρξη σωματικού λίπους επιβαρύνει την καρδιακή λειτουργία διότι κάθε επιπλέον 1Kg εξ αυτού, άνω του κανονικού βάρους σώματος, συνοδεύεται

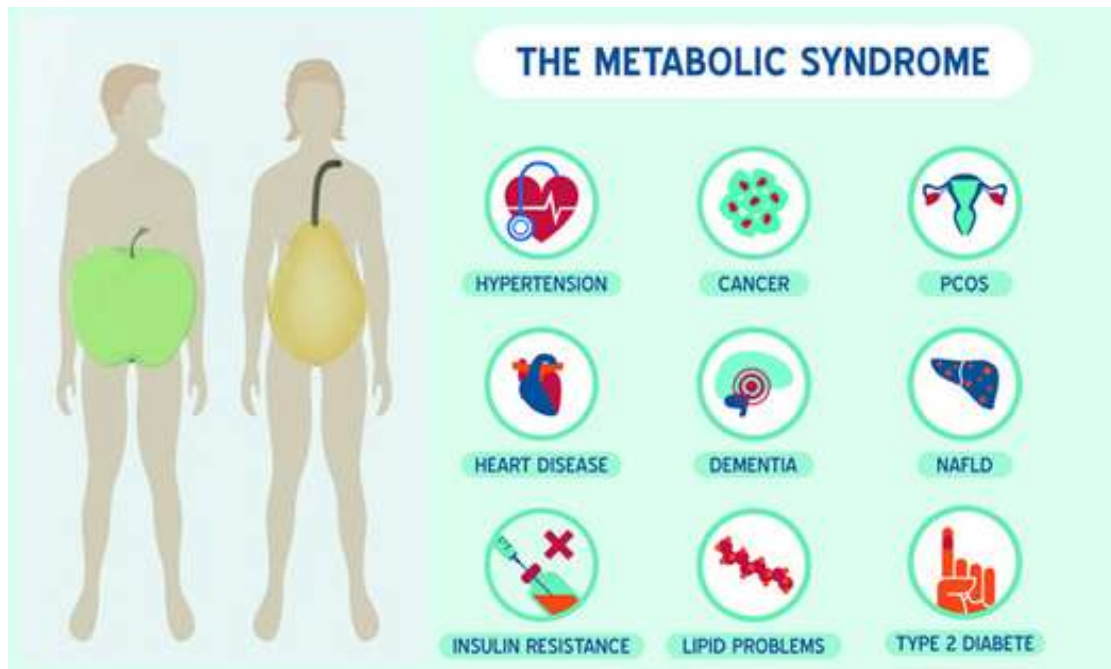
από δίκτυο αιμοφόρων αγγείων μήκους 1,5 Km, και η καρδιά πρέπει να διακινεί συνεχώς και σε μεγαλύτερες αποστάσεις το αίμα.

Ο λιπώδης ιστός στο ανθρώπινο σώμα αποτελεί χρήσιμη ενεργειακή εφεδρεία, όμως η υπέρμετρη συσσώρευση λιπώδους ιστού και μάλιστα περί τα σπλάχνα (περισπλάχνιο λίπος, αυτό είναι το πλέον επιζήμιο) είναι η μάλιστα του καθημερινού τρόπου ζωής, η οποία οδηγεί σε ένα σύμπλεγμα διαταραχών του μεταβολισμού, δηλαδή μια νοσηρή κατάσταση, που αναγνωρίζεται διεθνώς ως “μεταβολικό σύνδρομο” (metabolic syndrome).

Από πολλών ετών ο λιπώδης ιστός δεν θεωρείται περιττή αδρανής ύλη, διότι η επιστημονική έρευνα ανακάλυψε ότι παράγει και διοχετεύει στην κυκλοφορία αίματος, πληθώρα βιολογικά δραστικών μορίων (πρωτεϊνών, οι οποίες συλλήβδην ονομάζονται λιποκίνες) με δράση παρόμοια των ορμονών. Έχουν λοιπόν σημαντικό ρόλο στη ρύθμιση του μεταβολισμού της γλυκόζης και των λιπιδίων του αίματος, σε όργανα όπως ο εγκέφαλος, το ήπαρ, οι σκελετικοί μύες κλπ. και τελικά επηρεάζουν την συστηματική ευαισθησία στην ινσουλίνη (ανάπτυξη αντίστασης στην ινσουλίνη) και τη φλεγμονή και συνεπώς συνδέονται με αυξημένο κίνδυνο σακχαρώδη διαβήτη και καρδιοαγγειακής νόσου. Αν και μεταβολικό σύνδρομο είναι πιθανόν να αναπτύξουν και άτομα που δεν είναι εμφανώς παχύσαρκα, η παχυσαρκία απογειώνει την εκδήλωσή του. Η διάγνωσή του στηρίζεται στην ύπαρξη τουλάχιστον σε τρία από τα ακόλουθα κριτήρια:

- Περίμετρος μέσης >102 cm στους άντρες και >88 cm στις γυναίκες
- Τριγλυκερίδια αίματος >150 mg/dL
- Αρτηριακή Πίεση Αίματος >130/85 mm Hg
- Χαμηλά επίπεδα «καλής» χοληστερόλης (HDL) <40 mg/dL στους άνδρες ή <50 mg/dL στις γυναίκες
- Μέτρηση του σακχάρου στο αίμα μετά από περίοδο νηστείας, γλυκόζη νηστείας >110 mg/dL

Δηλαδή άτομο παχύσαρκο, με αυξημένη αρτηριακή πίεση, αυξημένο σάκχαρο αίματος και αυξημένη χοληστερίνη αίματος έχει τις ιδανικές προϋποθέσεις για να εμφανίσει καρδιοαγγειακές νόσους (έμφραγμα μυοκαρδίου, αγγειακό εγκεφαλικό επεισόδιο, αρτηριοσκληρωτικές βλάβες αγγείων), σακχαρώδη διαβήτη κλπ.



1.1.3 Μπορεί η παχυσαρκία να αντιμετωπιστεί;

Όπως προαναφέρθηκε, η παχυσαρκία ανήκει στις συχνές αιτίες θανάτου, λιγότερο συχνή από το κάπνισμα, ωστόσο επειδή αφορά και στις μικρές ηλικίες, τα παιδιά δηλαδή, πρέπει να ληφθεί σοβαρά υπόψη και να αντιμετωπιστεί εγκαίρως, πριν οδηγηθούν οι άνθρωποι στις προαναφερθείσες και ίσως μη αναστρέψιμες συνέπειές της. Ως εκ τούτου το θέμα αυτό εκτός από την επιμέρους κοινωνική του διάσταση έχει και οικονομική που αφορά στο σύνολο της κοινωνίας μας.

Αναφέρθηκε ήδη ότι η υπερβολική κατανάλωση τροφής καταλήγει σε παχυσαρκία με ολέθριες συνέπειες στην υγεία. Θα ήταν όμως παράλειψη αν δεν αναφερόμαστε συνοπτικά και στην ποιότητα της τροφής και δη στις λιπαρές ουσίες που η κακή τους ποιότητα υποθηκεύει το καρδιαγγειακό μας σύστημα, αυξάνοντας την χοληστερίνη (ή χοληστερόλη) αίματος και μάλιστα το κλάσμα της με το χαμηλό μοριακό βάρος LDL (Low-density lipoprotein) την λεγόμενη και κακή, διότι αυτή είναι υπεύθυνη για την σκλήρυνση των αιμοφόρων αγγείων (αρτηριοσκλήρυνση). Το κλάσμα της με το υψηλό μοριακό βάρος HDL (High-density lipoprotein) λέγεται και καλό διότι δρα αντίθετα της κακής.

Τα λίπη (λιπαρά οξέα) διακρίνονται αναλόγως της χημικής τους δομής σε κορεσμένα όταν φέρουν πλήρη αναλογία ατόμων υδρογόνου, μονοακόρεστα όταν λείπει ένα ζεύγος ατόμων υδρογόνου και πολυακόρεστα όταν λείπουν περισσότερα (ω-

3, ω-6). Τα κορεσμένα είναι κυρίως ζωικής προέλευσης (κρέας, γαλακτοκομικά, βούτυρο) και η λογική κατανάλωσή τους είναι αποδεκτή. Τα μονοακόρεστα (ελαιόλαδο) και δη τα πολυακόρεστα (που βρίσκονται στα λιπαρά ψάρια) είναι τα πλέον επιθυμητά και ένθερμα συστήνονται προς κατανάλωση, καθώς θεωρείται ότι η κατανάλωσή τους επηρεάζει ευνοϊκά την χοληστερίνη του αίματος δηλ., διατηρεί χαμηλά τα επίπεδά της στο αίμα, αλλά επιπλέον αυξάνει τις τιμές της καλής HDL χοληστερίνης. Ξεχωριστή κατηγορία λιπαρών ουσιών αποτελούν τα λεγόμενα τρανς λιπαρά (trans fatty acid). Αυτά είναι ανθρώπινη επινόηση καθώς με χημική παρέμβαση (μερική υδρογόνωση) λιπαρές ουσίες υγρής μορφής, χαμηλής ποιότητας και κόστους μετουσιώνονται σε στέρεη μορφή, τις γνωστές μαργαρίνες. Από φύσεως τους αυτές έχουν υψηλή περιεκτικότητα σε κορεσμένα λιπαρά οξέα και συνεπώς αποτελούν κάκιστη επιλογή για διατροφή.

Σε ατομικό επίπεδο, πρέπει να περιοριστεί η κατανάλωση ενέργειας από σάκχαρα και λίπη κακής ποιότητας (κορεσμένα και τρανς). Η κατανάλωση φρούτων και λαχανικών, οσπρίων (δηλ. τροφές που περιέχουν φυτικές ίνες), ψαριών και ξηρών καρπών (κυρίως καρυδιών), πρέπει να περιλαμβάνονται στο καθημερινό διαιτολόγιο. Δεν πρέπει να παραγκωνίζεται η διατροφή που περιέχει φυτικές ίνες (κατά κύριο λόγο κυτταρίνη), επειδή όσες δεν πέπτονται όταν φθάνουν στο παχύ έντερο με την απορρόφηση νερού διογκώνουν τα κόπρανα και η διέλευση και αποβολή τους είναι γρηγορότερη. Έτσι μαζί με την πιθανολογούμενη προσρόφηση καρκινογόνων ουσιών προστατεύουν το έντερο. Παράλληλα φαίνεται να μειώνουν τα επίπεδα της κακής LDL χοληστερίνης.

Η φυσιολογική, ποιοτική, ισορροπημένη διατροφή προσφέρει όλα τα απαραίτητα στοιχεία (πρωτεΐνες, λίπη, υδατάνθρακες) μαζί με βιταμίνες και ιχνοστοιχεία για να δομηθεί και να διατηρηθεί υγιής ο ανθρώπινος οργανισμός. Παράλληλα όμως είναι απολύτως εξ ίσου απαραίτητη η συνεχής φυσική δραστηριότητα (60' την ημέρα για τα παιδιά και 150' την εβδομάδα για τους ενήλικες, δηλαδή μισή ώρα γρήγορο περπάτημα επί πέντε φορές την εβδομάδα), στόχος που πρέπει να θέσει ο καθένας στον εαυτό του για βελτιωμένη και υγιή ζωή (WHO, 2018).

Οι βιομηχανίες τροφίμων έχουν παίξει ρόλο στην ανάπτυξη της παχυσαρκίας και υπονόμηση της υγείας, προωθώντας υλικά όπως ζάχαρα, λιπαρές ουσίες κλπ. στα προϊόντα που διαθέτουν. Παράδειγμα η προσθήκη φοινικέλαιου σχεδόν σε όλα τα βρώσιμα προϊόντα (επειδή είναι φθινό και ουδέτερο στη γεύση). Πολλά από αυτά τα

προϊόντα είναι τόσο κακά ώστε τα εμπλουτίζουν με βιταμίνες και ιχνοστοιχεία για να γίνουν ευυπόληπτα προς διαφήμιση, δηλαδή προσφέρουν σε ένα πακέτο μαζί το πρόβλημα και τη λύση. Με αυστηρότατο έλεγχο από τους νόμους της πολιτείας απαιτείται η διαφοροποίηση στην ποιότητα των υλικών, η μείωση των “κακών” λιπαρών (κορεσμένα-ζωϊκά και τρανς λιπαρά-μαργαρίνες) καθώς και η μείωση της ζάχαρης και του αλατιού από το περιεχόμενο των υπό κατανάλωση τροφίμων. Ο έλεγχος θα πρέπει να εξασφαλίζει ότι οι υγιεινές και θρεπτικές διατροφικές επιλογές είναι διαθέσιμες και προσβάσιμες στους καταναλωτές. Ίσως είναι σκόπιμη η απαγόρευση της εμπορίας τροφίμων με υψηλή περιεκτικότητα σε σάκχαρα, αλάτι και “κακά” λιπαρά, τα οποία προορίζονται για παιδιά και εφήβους. Οι βιομηχανίες τροφίμων θα πρέπει να εξασφαλίζουν την διαθεσιμότητα επιλογών υγιεινής διατροφής και γιατί όχι να ενισχύουν την εφαρμογή καθημερινής φυσικής δραστηριότητας στους εργαζόμενους σ’ αυτές. Κάτι τέτοιο όμως δεν είναι πιθανό να συμβεί, καθώς λειτουργεί αντίθετα στην κερδοφορία!

Συνεπώς, καθώς προτεραιότητα αποτελούν τα παιδιά, η ευρεία ενημέρωση μέσω της σχολικής εκπαίδευσης από μικρή ηλικία, η ενημέρωση και η καθοδήγηση των γονέων στα σχολεία σχετικά με προγράμματα διατροφής των παιδιών τους, είναι μία προτεινόμενη μέθοδος. Θα προτείνουμε επίσης συχνή ενημέρωση από τα ΜΜΕ, με συνεχή προβολή spots, τα οποία να αναφέρονται στην σωστή διατροφή και τη σχέση της με την καλή υγεία, καθώς η παχυσαρκία και τα προβλήματα που προκαλούνται από αυτήν είναι σε υψηλά επίπεδα στη χώρα μας.

1.1.4 Πώς μετριέται η παχυσαρκία;

Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας, ένας ευρέως γνωστός δείκτης μέτρησης της παχυσαρκίας είναι ο Δείκτης Μάζας Σώματος (Body Mass Index-BMI). Ο τρόπος υπολογισμού του δείκτη μάζας σώματος είναι πολύ απλός. Υπολογίζεται διαιρώντας το βάρος του ατόμου υπολογισμένο σε κιλά (kg) με το τετράγωνο του ύψους του υπολογισμένο σε μέτρα (m^2).

$$BMI = \frac{kg}{m^2}$$

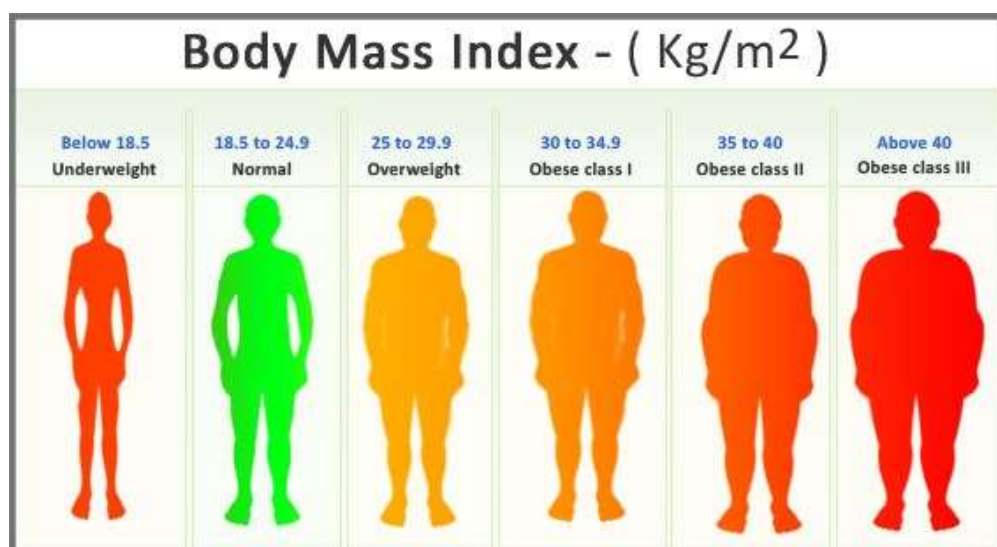
Για παράδειγμα, ένα άτομο βάρους 60 κιλών με ύψος 1,70 θα έχει δείκτη μάζας σώματος $BMI = \frac{60}{1.70^2} = 20,76$.

Με βάση τον δείκτη μάζας σώματος, ο κάθε άνθρωπος κατατάσσεται στην ανάλογη κατηγορία βάρους. Σύμφωνα με τον WHO, οι κατηγορίες βάρους διαχωρίζονται σε 6 και παρουσιάζονται, περιληπτικά, στον παρακάτω πίνακα. Ο Δείκτης Μάζας Σώματος υπολογίζεται με τον ίδιο τρόπο στους άνδρες και τις γυναίκες και όλες τις ηλικίες των ενηλίκων.

Πίνακας 1.1

Κατηγορίες διατροφικής κατάστασης με βάση το BMI

Δείκτης Μάζας Σώματος (BMI)	Διατροφική κατάσταση
Κάτω από 18,5 (<18,5)	Ελλειποβαρής
Από 18,5 έως 24,9 (18.5-24.9)	Κανονικό βάρος
Από 25 έως 29,9 (25-29,9)	Υπέρβαρος
Από 30 έως 34,9 (30-34,9)	Παχύσαρκος κατηγορίας I
Από 35 έως 39,9 (35-39,9)	Παχύσαρκος κατηγορίας II
Πάνω από 40 (>40)	Παχύσαρκος κατηγορίας III



(Πηγή: <http://www.shdoonline.org/blog-5/>)

Κάθε άτομο, με βάση τον δείκτη μάζας σώματός του, κατατάσσεται σε μια κατηγορία:

Στην κατηγορία των ελλειποβαρών (δηλαδή κάτω από 18,5) το άτομο είναι πολύ αδύνατο, που σημαίνει ελλειμματική διατροφή, με πιθανές συνέπειες εμφάνισης αναιμίας, ανεπάρκειας βιταμινών, έλλειψη ενέργειας και μειωμένες άμυνες όσον αφορά το ανοσοποιητικό σύστημα.

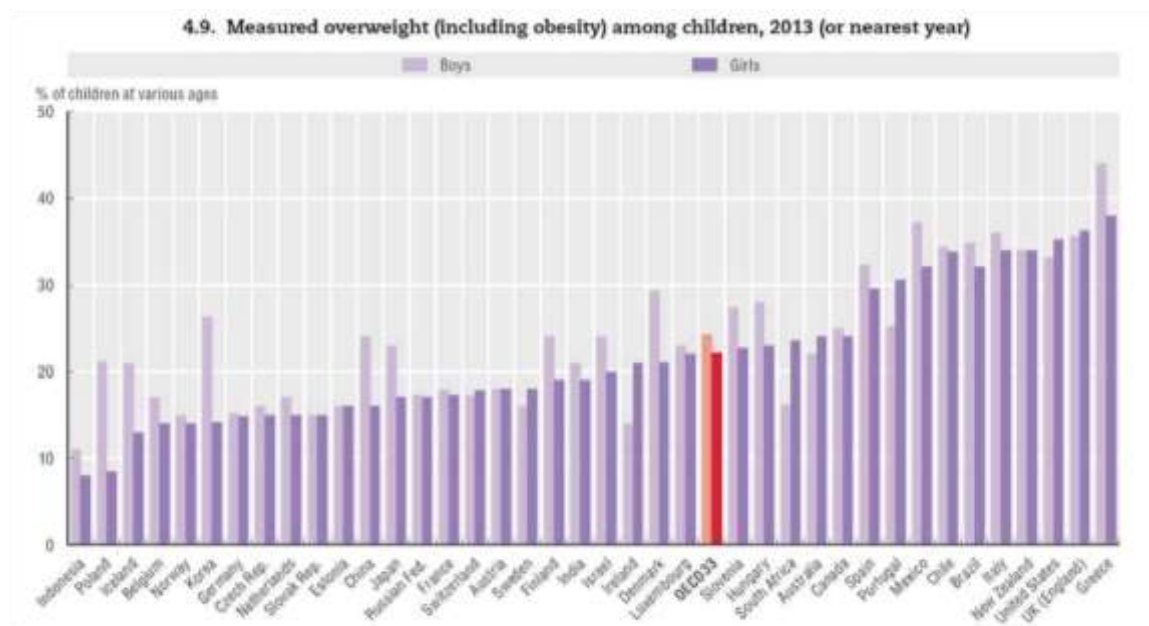
Άτομα με BMI από 18,5 έως 24,9 θεωρούνται κανονικού βάρους.

Από 25 και πάνω τα άτομα κατατάσσονται ως υπέρβαρα και σε αυτά εμφανίζονται κατά κύριο λόγο τα προβλήματα που αναφέρθηκαν στην υποενότητα 1.1.2., όπως υψηλή αρτηριακή πίεση, καρδιαγγειακά νοσήματα, διαβήτης τύπου 2, οστεοαρθρίτιδα, ορισμένοι τύποι καρκίνου κλπ.

1.1.5 Στατιστικά στοιχεία γύρω από την παχυσαρκία

Σύμφωνα με τα τελευταία στοιχεία (2015-2017) του Παγκόσμιου Οργανισμού Υγείας (WHO), στην Ελλάδα και σε άλλες χώρες της νότιας Ευρώπης (Ιταλία, Ισπανία, Μάλτα) παρατηρούνται τα υψηλότερα ποσοστά παιδικής παχυσαρκίας (ποσοστά που κυμαίνονται από το 18% έως το 21%).

Σχήμα 1.1.5.1 (Ποσοστό υπέρβαρων και παχύσαρκων παιδιών για το 2013)



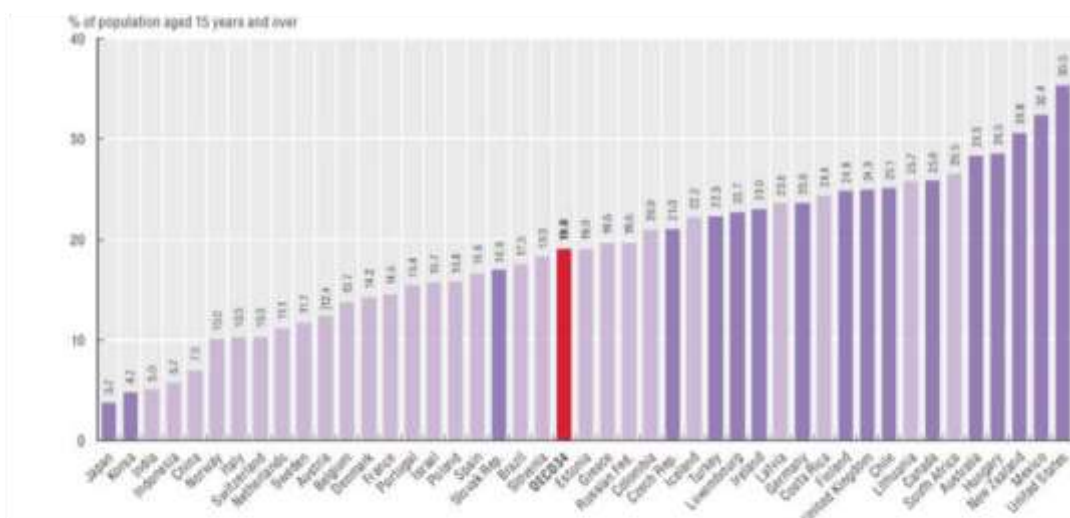
Πηγή: ΟΟΣΑ, 2015

(Πηγή: https://www.dianeosis.org/2016/04/paxysarkia_stin_ellada_tis_krisis/)

Το παραπάνω διάγραμμα του ΟΟΣΑ (2015) παρουσιάζει τα ποσοστά υπέρβαρων και παχύσαρκων παιδιών για το έτος 2013. Παρατηρούμε ότι στην Ελλάδα το ποσοστό των υπέρβαρων και παχύσαρκων αγοριών ανέρχεται σε πάνω από 40% και το ποσοστό

των υπέρβαρων και παχύσαρκων κοριτσιών στο 38-39%. Στην γειτονική Ιταλία τα ποσοστά των υπέρβαρων και παχύσαρκων αγοριών και κοριτσιών δεν παρουσιάζουν μεγάλη διαφορά, περίπου 35% και 33% αντίστοιχα, χαμηλότερα ωστόσο από αυτά της χώρας μας.

Σχήμα 1.1.5.2 (Ποσοστά παχυσαρκίας ενηλίκων για το 2013)

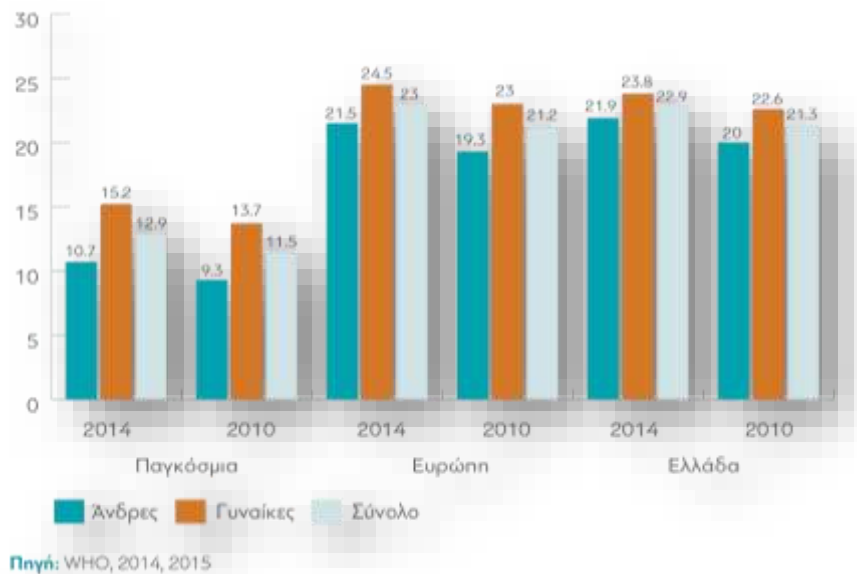


Πηγή: ΟΟΣΑ, 2015

(Πηγή: https://www.dianeosis.org/2016/04/paxysarkia_stin_ellada_tis_krisis/)

Το παραπάνω διάγραμμα δείχνει ποσοστά παχυσαρκίας ενηλίκων-2013. Υψηλά ποσοστά παχυσαρκίας παρατηρούνται στις Ηνωμένες Πολιτείες, το Μεξικό και τη Νέα Ζηλανδία (ποσοστά 35.3%, 32.4%, 30.6% αντίστοιχα). Στην Ελλάδα και σε άλλες χώρες της νότιας Ευρώπης (Ιταλία, Ισπανία) το ποσοστό παχυσαρκίας, στα άτομα ηλικίας άνω των 15 ετών, ανέρχεται στο 19.5%, 10.3% και 16.6% αντίστοιχα. Μικρότερο ποσοστό σε παχύσαρκους ενήλικες (ΟΟΣΑ 2015) εμφανίζεται στην Ιαπωνία, δηλαδή ποσοστό 3.7%, πολύ μικρό συγκριτικά με αυτά των προαναφερθεισών χωρών.

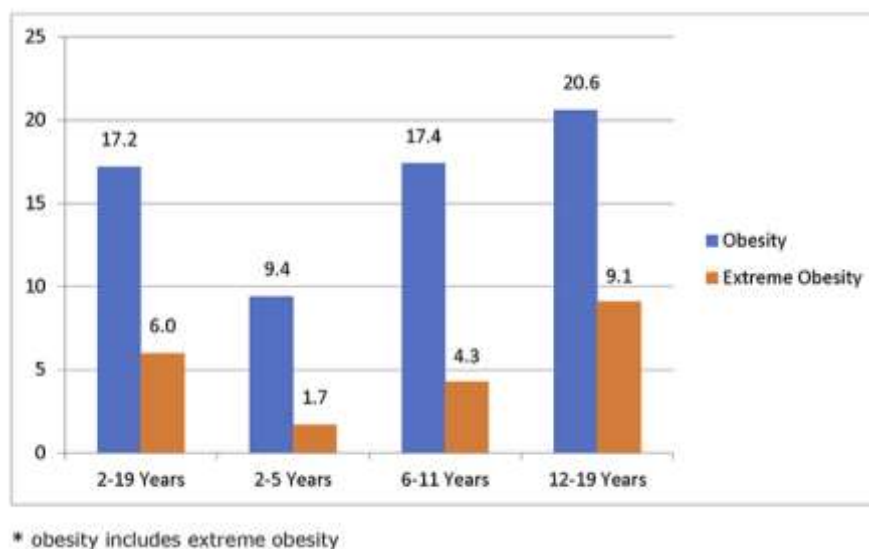
Σχήμα 1.1.5.3 (Ποσοστά παχύσαρκων ενηλίκων σε παγκόσμιο και Ευρωπαϊκό επίπεδο και ανά φύλο)



(Πηγή: https://www.dianeosis.org/2016/04/paxysarkia_stin_ellada_tis_krisis/)

Ενδιαφέρον παρουσιάζει η μελέτη στα ποσοστά των παχύσαρκων ενηλίκων σε παγκόσμιο, ευρωπαϊκό και ελληνικό επίπεδο επί συνόλου πληθυσμού αλλά και ανά φύλο κατά την πενταετία 2010-2014 (*Παγκόσμιος Οργανισμός Υγείας*). Την πενταετία αυτή, στην Ελλάδα, το ποσοστό των παχύσαρκων ενηλίκων **αυξήθηκε** ελαφρά από το 21,3% το 2010 σε 22,9% το 2014. Μικρή αύξηση παρατηρείται στο διάγραμμα ανά φύλο. Για παράδειγμα, το ποσοστό των παχύσαρκων Ελλήνων αντρών το 2010 ήταν 20% και το 2014 κυμαινόταν στο 21,9%. Σε ευρωπαϊκό επίπεδο, κατά την διάρκεια της ως άνω πενταετίας, το ποσοστό των παχύσαρκων ενηλίκων παρουσιάζει αύξηση από 21,2% (2010) σε 23% (2014). Το ίδιο συμβαίνει και ανά φύλο. Για παράδειγμα, το ποσοστό των παχύσαρκων ανδρών παγκοσμίως αυξήθηκε από 19,3% το 2010 σε 21,5% το 2014. Τέλος σε παγκόσμιο επίπεδο, για το σύνολο του πληθυσμού, το ποσοστό των παχύσαρκων ενηλίκων αυξήθηκε από το 11,5% το 2010, σε 12,9% το 2014. Αύξηση παρατηρείται και ανά φύλο.

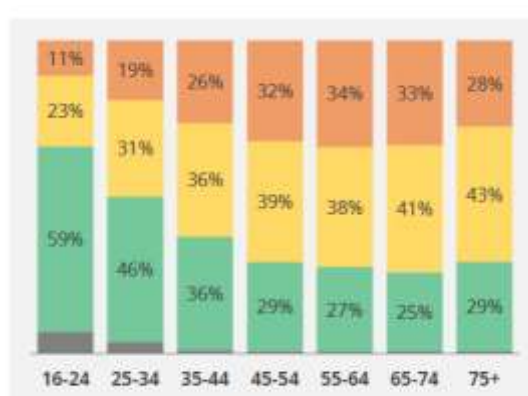
Σχήμα 1.1.5.4 (Ποσοστά παχυσαρκίας νέων στις Ηνωμένες Πολιτείες τα έτη 2013-2014)



(Πηγή : <https://www.niddk.nih.gov/health-information/health-statistics/overweight-obesity>)

Στις Ηνωμένες Πολιτείες το μεγαλύτερο ποσοστό παχυσαρκίας και νοσογόνου παχυσαρκίας των νέων, της τάξης του 20,6% και 9,1% αντίστοιχα, εμφανίζεται σε άτομα ηλικίας 12-19 ετών σε αντίθεση με τα άτομα ηλικίας 2-5 ετών που το ποσοστό παχυσαρκίας ανέρχεται σε 9,4% ενώ το ποσοστό νοσογόνου παχυσαρκίας σε 1,7%.

Σχήμα 1.1.5.5 (Ποσοστά παχυσαρκίας ηλικιών 16-75+ στο Ηνωμένο Βασίλειο)



(Πηγή: Baker, 2018)

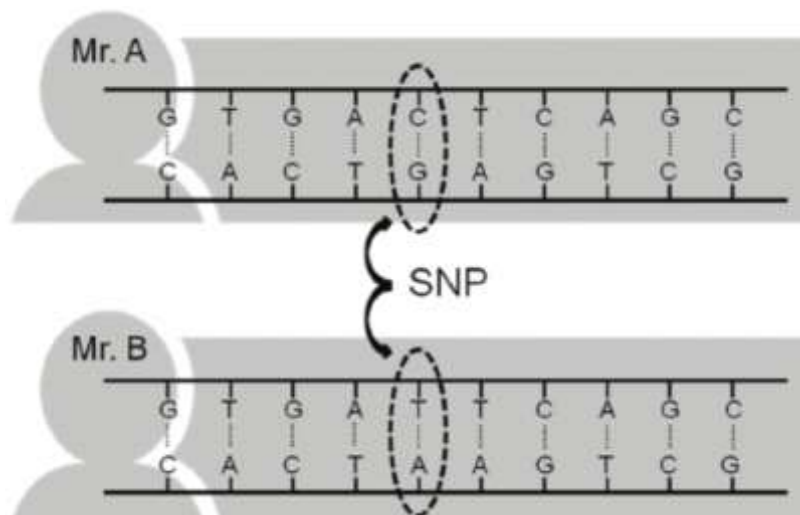
Στην Αγγλία, σύμφωνα με το ως άνω σχήμα, μεγαλύτερα ποσοστά παχυσαρκίας εμφανίζονται από την ηλικία των 45 ετών και πάνω. Στις νεότερες ηλικίες, ηλικίες 16-24, το ποσοστό των παχύσαρκων ατόμων ανέρχεται στο 11%. Όσο αυξάνεται η ηλικία

το ποσοστό παχυσαρκίας παρουσιάζει ανοδική πορεία με εξαίρεση στα άτομα ηλικίας 75 ετών και άνω τα οποία παρουσιάζουν ποσοστό παχυσαρκίας της τάξης του 28%. Στο ίδιο συμπέρασμα, δηλαδή όσο μεγαλύτερη είναι η ηλικία τόσο πιθανότερο είναι να καταταχθεί κάποιος ως υπέρβαρος και παχύσαρκος, καταλήξαμε και εμείς βάσει του δείγματός μας (βλ. κεφάλαιο 7).

1.2 Τι είναι ο πολυμορφισμός

Ο σημειακός ή μονού νουκλεοτιδίου πολυμορφισμός (Single Nucleotide Polymorphism-SNP) είναι το πιο γνωστό είδος γενετικής ποικιλίας μεταξύ των ανθρώπων και η βασική μονάδα στην οποία εκφράζεται κάθε γενετική ποικιλομορφία (Νικολάου, 2015). Κάθε σημειακός πολυμορφισμός αντιπροσωπεύει μια διαφορά σε ένα ενιαίο μπλοκ DNA. Για παράδειγμα, ένας SNP μπορεί να αντικαταστήσει το νουκλεοτίδιο της κυτοσίνης (C) με το νουκλεοτίδιο της θυμίνης (T), σε μια ορισμένη περιοχή του DNA. Οι σημειακοί πολυμορφισμοί εμφανίζονται κανονικά στο DNA του ανθρώπου. Κατά μέσο όρο εμφανίζονται μια φορά σε κάθε 300 νουκλεοτίδια, πράγμα που σημαίνει ότι υπάρχουν 10 εκατομμύρια σημειακοί πολυμορφισμοί στο ανθρώπινο γονιδίωμα. Κάθε ένας σημειακός πολυμορφισμός βρίσκεται στο DNA και “χρησιμοποιείται” από τους επιστήμονες για να εντοπίσουν γονίδια, τα οποία σχετίζονται με μια ασθένεια. Οι περισσότεροι σημειακοί πολυμορφισμοί δεν έχουν καμία επίδραση στην υγεία. Ωστόσο κάποιες από αυτές τις γενετικές διαφορές έχει αποδειχθεί ότι είναι πολύ σημαντικές στην μελέτη της υγείας των ανθρώπων. Ερευνητές έχουν ανακαλύψει σημειακούς πολυμορφισμούς, οι οποίοι βοηθούν στην πρόβλεψη της ανθρώπινης αντίδρασης σε ορισμένα φάρμακα, την ευαισθησία σε παράγοντες όπως οι τοξίνες και τον κίνδυνο της ανάπτυξης συγκεκριμένων ασθενειών. Τέλος, οι σημειακοί πολυμορφισμοί μπορούν να χρησιμοποιηθούν για την παρακολούθηση της κληρονομιάς γονιδίων νόσου μέσα σε οικογένειες. Μελλοντικές μελέτες έχουν ως στόχο να εντοπίσουν SNPs, οι οποίοι σχετίζονται με σύνθετες ασθένειες, όπως για παράδειγμα καρδιοαγγειακές παθήσεις, διαβήτη και καρκίνο. (<https://ghr.nlm.nih.gov>).

Σχήμα 1.5.1 (Το σχήμα του σημειακού πολυμορφισμού)

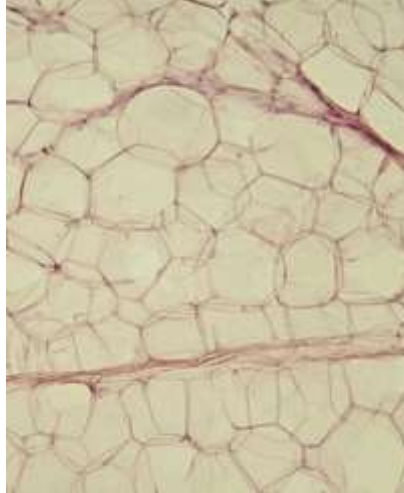


(Πηγή: <https://www.researchgate.net>)

1.2.1 Πώς συνδέονται οι πολυμορφισμοί με την παχυσαρκία

Οι πολυμορφισμοί πράγματι συνδέονται με την παχυσαρκία. Πιο συγκεκριμένα οι πολυμορφισμοί οι οποίοι βρίσκονται στα γονίδια του αδρενεργικού υποδοχέα (ADRB) έχουν εκτεταμένα μελετηθεί για την σχέση που μπορεί να παρουσιάζουν με την παχυσαρκία (Kurokawa, 2008). Στο κεφάλαιο 5, οι σημειακοί πολυμορφισμοί που θα μελετήσουμε είναι οι ADRB3rs4994, ADRB2rs1042713, ADRB2rs1042714, ADRB1rs1801252 και ADRB1rs1801253. Ο πρώτος πολυμορφισμός, ο οποίος μελετήθηκε βρίσκεται στο γονίδιο ADRB3. Το γονίδιο αυτό καλείται αλλιώς και β -3 αδρενεργικός υποδοχέας, είναι μέρος του αδρενεργικού συστήματος, το οποίο ως γνωστόν παίζει σημαντικό ρόλο στον ενεργειακό μεταβολισμό και συνδέεται με την παχυσαρκία (Kurokawa, 2008). Εντοπίζεται στον λιπώδη ιστό (σχήμα 1.5.1.1) και εμπλέκεται στην ρύθμιση της λιπόλυσης. Η λιπόλυση είναι η διαίρεση των λιπιδίων και χρησιμοποιείται για να κινητοποιήσει την αποθηκευμένη ενέργεια (<https://en.wikipedia.org>). Το 1995, μια σειρά εργασιών δημοσιεύτηκαν στο *New England Περιοδικό της Ιατρικής* αποκαλύπτοντας μια πιθανή σχέση μιας μετάλλαξης εντός του ADRB3 προς την ανθρώπινη παχυσαρκία (Rosmond, 2003).

Σχήμα 1.5.1.1 (Λιπώδης ιστός)



(Πηγή: http://health.hellasmagazine.gr/2012/05/blog-post_9.html)

Ο δεύτερος και ο τρίτος πολυμορφισμός που θα μελετήσουμε βρίσκονται στο ίδιο γονίδιο και πιο συγκεκριμένα στο **ADRB2**. Το γονίδιο αυτό καλείται **β-2** αδρενεργικός υποδοχέας και παίζει σημαντικό ρόλο στην ενεργειακή δαπάνη και στην λιπόλυση. Οι τελευταίοι πολυμορφισμοί που μελετάμε βρίσκονται στο γονίδιο **ADRB1**. Το γονίδιο αυτό καλείται **β-1** αδρενεργικός υποδοχέας και βρίσκεται κυρίως στον καρδιακό ιστό.

ΚΕΦΑΛΑΙΟ 2

ΠΙΝΑΚΕΣ ΣΥΝΑΦΕΙΑΣ

Σκοπός του κεφαλαίου αυτού είναι να δοθεί ένας ορισμός για τα κατηγορικά δεδομένα και μετά για τον διαχωρισμό των ειδών τους. Στη συνέχεια, θα δούμε τι είναι ένας πίνακας συνάφειας (Contingency table) και πώς παρουσιάζονται ταυτόχρονα δύο ποιοτικές μεταβλητές, πώς μπορούμε δηλαδή να ελέγξουμε αν υπάρχει σχέση μεταξύ δύο ή περισσότερων ποιοτικών μεταβλητών μέσω του ελέγχου ανεξαρτησίας χ^2 .

2.1 Κατηγορικά δεδομένα

Οι ποιοτικές μεταβλητές, καλούμενες και κατηγορικές μεταβλητές (ή κατηγορικά δεδομένα), είναι μία από τις δύο βασικές κατηγορίες των μεταβλητών στο επιστημονικό πεδίο της έρευνας, της παρατήρησης και βασικότερα του πειράματος. (www.wikipedia.org). Με λίγα λόγια, κατηγορικά δεδομένα ονομάζονται εκείνα που παίρνουν «τιμές» σε ένα σύνολο κατηγοριών. Προέρχονται, δηλαδή, από πραγματοποιήσεις τυχαίων μεταβλητών των οποίων οι τιμές λαμβάνονται από ένα σύνολο που δεν αποτελείται απαραίτητα από αριθμούς. Σε πολλές περιπτώσεις, όμως, αντιστοιχίζουμε τα δεδομένα αυτά με αριθμούς προκειμένου να τα επεξεργαστούμε με τη βοήθεια στατιστικών προγραμμάτων στον Η/Υ μας. Η αντιστοιχία αυτή είναι απλός συμβολισμός, πιο πολύ για διευκόλυνσή μας, και επομένως δεν νοείται κανενός είδους αριθμητική πράξη. Αντιπροσωπευτικά παραδείγματα κατηγορικών δεδομένων είναι τα εξής:

- *Φύλο*: άνδρας, γυναίκα.
- *Θρήσκευμα*: χριστιανός, μουσουλμάνος, βουδιστής, εβραίος κ.α.
- *Πολιτική Τοποθέτηση*: ακροαριστερός, αριστερός, κεντρώος, δεξιός, ακροδεξιός.
- *Κοινωνική Κατάσταση*: παντρεμένος, ανύπαντρος, χήρος κλπ.
- *Επίπεδο Εκπαίδευσης*: υποχρεωτική, τεχνολογική, πανεπιστημιακή.

Για να εξηγήσουμε την παραπάνω αντιστοίχιση των δεδομένων με αριθμούς, ας πάρουμε την μεταβλητή *Φύλο* στην οποία μπορούμε να θέσουμε την κατηγορία *άνδρας* με 0 και την κατηγορία *γυναίκα* με 1.

2.2 Είδη κατηγορικών δεδομένων

Μια Κατηγορική Μεταβλητή μπορεί να χαρακτηριστεί Ονοματική (Nominal) ή Διατακτική (Ordinal). Στις ονοματικές μεταβλητές η σειρά με την οποία εμφανίζονται οι κατηγορίες δεν έχει σημασία. Δηλαδή λαμβάνουν τιμές που δεν ιεραρχούνται. Οι κατηγορίες τους δεν έχουν μια φυσική διάταξη. Ονοματικές μεταβλητές μπορούν να χαρακτηριστούν, βάσει των ως άνω παραδειγμάτων, το φύλο, το θρήσκευμα και η κοινωνική κατάσταση ενός ατόμου. Στις συγκεκριμένες μεταβλητές τα συμπεράσματα που εξάγουμε είναι ανεξάρτητα από τη σειρά των κατηγοριών.

Αντιθέτως, στις διατακτικές μεταβλητές η σειρά των κατηγοριών είναι σημαντική. Οι κατηγορίες των μεταβλητών αυτών ιεραρχούνται. Το επίπεδο εκπαίδευσης, για παράδειγμα, μπορεί να χαρακτηριστεί ως Διατακτική Μεταβλητή διότι κατηγοριοποιείται σε βασική εκπαίδευση, λύκειο, πανεπιστήμιο, μεταπτυχιακό. Σε συνάφεια με το παράδειγμά μας που αναφέρεται στην εκπαίδευση, Διατακτική μπορεί να θεωρηθεί και η διαγωγή ενός μαθητή καθώς κατηγοριοποιείται σε καλή, κοσμία, κοσμιωτάτη. Ακόμη, Διατακτική Μεταβλητή μπορεί να θεωρηθεί και η κατηγοριοποίηση των τιμών του δείκτη μάζα σώματος (Body Mass Index-BMI) προκειμένου να “τοποθετήσουμε” ένα άτομο στην ανάλογη κατηγορία (φυσιολογικός, υπέρβαρος, παχύσαρκος κλπ.) Στις συγκεκριμένες μεταβλητές επιτρέπεται μόνο αντιστροφή της σειράς.

2.3 Πίνακες Συνάφειας

Γνωρίζουμε ότι για να παρουσιάσουμε μία ποιοτική μεταβλητή χρησιμοποιούμε έναν πίνακα συχνοτήτων. Τι γίνεται όμως αν θέλουμε να παρουσιάσουμε δύο ποιοτικές (κατηγορικές) μεταβλητές μαζί; Στην συγκεκριμένη περίπτωση χρησιμοποιούμε έναν **πίνακα συνάφειας** ο οποίος παριστάνει τις συχνότητες (counts) για όλους τους συνδυασμούς κατηγοριών (επιπέδων) ανάμεσα σε δύο (ή περισσότερες) κατηγορικές μεταβλητές.

Ας υποθέσουμε ότι έχουμε δύο κατηγορικές μεταβλητές X, Y . Έστω I ο αριθμός των κατηγοριών της μεταβλητής X και J ο αριθμός των κατηγοριών της μεταβλητής Y . Ένας πίνακας με I γραμμές για τις κατηγορίες της μεταβλητής X και J στήλες για τις κατηγορίες της μεταβλητής Y , έχει IJ κελιά. Όταν τα κελιά περιέχουν τις συχνότητες για κάθε IJ πιθανό αποτέλεσμα ενός δείγματος, τότε έχουμε έναν διδιάστατο $I \times J$

πίνακα συνάφειας [Agresti, (2007)]. Αν συμβολίσουμε με n_{ij} τη συχνότητα που αντιστοιχεί στην i κατηγορία της X και στην j κατηγορία της Y και με π_{ij} τις από κοινού πιθανότητες των X, Y παίρνουμε τον παρακάτω $I \times J$ πίνακα συνάφειας. Ένας πίνακας συνάφειας, ο οποίος περιγράφει την σχέση μεταξύ δύο κατηγορικών μεταβλητών καλείται πίνακας διπλής εισόδου (two-way table). Ακόμη, στον ακόλουθο πίνακα συνάφειας, παρουσιάζονται και οι από κοινού πιθανότητες π_{ij} των μεταβλητών X και Y . Για κάθε ζευγάρι μεταβλητών X και Y η από κοινού συνάρτηση πιθανότητας (ή από κοινού πιθανότητες) είναι

$$\pi_{ij} = P(X = i \text{ και } Y = j) \text{ (Πολίτης, 2016).}$$

Πίνακας 2.1

Γενική μορφή ενός πίνακα συνάφειας

Μεταβλητή X	Μεταβλητή Y					Σύνολο
	1	2	...	j	J	
1	n_{11} p_{11}	n_{12} p_{12}	...	n_{1j} p_{1j}	n_{1J} p_{1J}	n_{1+} p_{1+}
2	n_{21} p_{21}	n_{22} p_{22}	...	n_{2j} p_{2j}	n_{2J} p_{2J}	n_{2+} p_{2+}
·	·	·	·	·	·	·
·	·	·	·	·	·	·
·	·	·	·	·	·	·
·	·	·	·	·	·	·
i	n_{i1} p_{i1}	n_{i2} p_{i2}	·	n_{ij} p_{ij}	n_{iJ} p_{iJ}	n_{i+} p_{i+}
I	n_{I1} p_{I1}	n_{I2} p_{I2}	·	n_{Ij} p_{Ij}	n_{IJ} p_{IJ}	n_{I+} p_{I+}
Σύνολο	n_{+1} p_{+1}	n_{+2} p_{+2}	·	n_{+j} p_{+j}	n_{+J} p_{+J}	N I

Οι περιθώριες κατανομές ορίζονται με βάση τις περιθώριες πιθανότητες, p_{i+} για την μεταβλητή X και p_{+j} για την μεταβλητή Y (Πολίτης, 2016). Πιο συγκεκριμένα, οι

γενικοί τύποι υπολογισμού των περιθώριων πιθανοτήτων των μεταβλητών X και Y είναι οι εξής:

$$p_{i+} = \sum_{j=1}^J p_{ij} \text{ με } \sum_{i=1}^I p_{i+} = 1 \text{ (γνωστό από την θεωρία πιθανοτήτων)}$$

$$p_{+j} = \sum_{i=1}^I p_{ij} \text{ με } \sum_{j=1}^J p_{+j} = 1$$

Για παράδειγμα, για την πρώτη γραμμή του πρώτου πίνακα συνάφειας (Πίνακας 2.1) προκύπτει ότι $p_{1+} = \sum_{j=1}^J p_{1j} = p_{11} + p_{12} + \dots + p_{1J}$ (το άθροισμα των από κοινού πιθανοτήτων των κελιών (1,1), (1,2) και (1,J) αντίστοιχα) ενώ για την πρώτη στήλη του παραπάνω πίνακα συνάφειας προκύπτει ότι $p_{+1} = \sum_{i=1}^I p_{i1} = p_{11} + p_{21} + \dots + p_{I1}$ (το άθροισμα των από κοινού πιθανοτήτων των κελιών (1,1), (2,1) και (I,1) αντίστοιχα). Ομοίως προκύπτουν και οι από κοινού πιθανότητες των υπολοίπων κελιών του πίνακα.

Η απλούστερη μορφή ενός πίνακα συνάφειας είναι ο 2×2 πίνακας. Ο συγκεκριμένος πίνακας δημιουργείται από δύο ποιοτικές μεταβλητές με δύο κατηγορίες (επίπεδα) η κάθε μια και αποτελείται από 4 κελιά. Εφόσον και οι δύο μεταβλητές που μελετάμε είναι κατηγορικές (ποιοτικές), αναφέρονται συνήθως και ως παράγοντες (και οι τιμές τους ως επίπεδα του αντίστοιχου παράγοντα) ή ως μεταβλητές ταξινόμησης, εφόσον χρησιμοποιούνται για την κατασκευή ενός πίνακα συχνοτήτων. Οι συγκεκριμένοι εμφανίζονται σε ιατρικές μελέτες, στις οποίες το ενδιαφέρον εστιάζεται στην παρουσία ή όχι μιας νόσου-μια δίτιμη ποιοτική μεταβλητή (Πολίτης, 2016). Παράδειγμα ενός 2×2 πίνακα συνάφειας δίνεται στον παρακάτω Πίνακα 2.2. Οι γραμμές του πίνακα αποτελούνται από τις κατηγορίες της μιας ποιοτικής μεταβλητής, ενώ οι στήλες του πίνακα αποτελούνται από τις κατηγορίες της άλλης ποιοτικής μεταβλητής. Με n_{ij} συμβολίζουμε την συχνότητα εμφάνισης του κελιού (i,j) , δηλαδή αυτού που δημιουργείται από την γραμμή i και την στήλη j . Εκτενέστερα, με n_{11} συμβολίζουμε την συχνότητα του κελιού που δημιουργείται από την γραμμή 1 και την στήλη 1 (1,1). Με n_{12} συμβολίζουμε την συχνότητα του κελιού που δημιουργείται από την γραμμή 1 και την στήλη 2 (1,2). Με n_{21} συμβολίζουμε την συχνότητα του κελιού που δημιουργείται από την γραμμή 2 και την στήλη 1 (2,1) και τέλος με n_{22} συμβολίζουμε την συχνότητα του κελιού που δημιουργείται από την γραμμή 2 και την στήλη 2 (2,2).

Πίνακας 2.2

Ειδική Περίπτωση ενός 2×2 πίνακα συνάφειας

Μεταβλητή X	Μεταβλητή Y		Σύνολο
	Επίπεδο 1	Επίπεδο 2	
Επίπεδο 1	n_{11} p_{11}	n_{12} p_{12}	n_{1+} p_{1+}
Επίπεδο 2	n_{21} p_{21}	n_{22} p_{22}	n_{2+} p_{2+}
Σύνολο	n_{+1} p_{+1}	n_{+2} p_{+2}	N I

2.4 Συμβολισμοί ενός πίνακα συνάφειας

Για έναν πίνακα συνάφειας μας δίνεται η δυνατότητα να υπολογίσουμε και άλλες ποσότητες οι οποίες μπορούν να μας περιγράψουν μια κατάσταση.

- Οι προαναφερθείσες συχνότητες n_{ij} , οι οποίες συμβολίζουν την συχνότητα του κελιού που δημιουργείται από την i γραμμή και την j στήλη, ονομάζονται *παρατηρηθείσες τιμές* (observed values). Βάσει αυτών μπορεί να υπολογιστεί το σύνολο των παρατηρημένων τιμών ανά γραμμή και ανά στήλη, n_{i+} και n_{+j} αντίστοιχα. Για παράδειγμα, για την πρώτη γραμμή του παραπάνω πίνακα συνάφειας (Πίνακας 2.2) προκύπτει ότι $n_{1+} = n_{11} + n_{12}$ (το άθροισμα των παρατηρημένων συχνοτήτων των κελιών (1,1) και (1,2) αντίστοιχα) ενώ για την πρώτη στήλη προκύπτει ότι $n_{+1} = n_{11} + n_{21}$ (το άθροισμα των παρατηρημένων συχνοτήτων των κελιών (1,1) και (2,1) αντίστοιχα). Οι τιμές αυτές μας δηλώνουν πόσα άτομα, στο σύνολο του δείγματος, εμφανίζουν το αντίστοιχο χαρακτηριστικό.
- Για έναν πίνακα συνάφειας μπορούν να υπολογιστούν και οι *αναμενόμενες τιμές* (expected values), οι οποίες δηλώνουν τι θα περιμέναμε να παρατηρήσουμε στο δείγμα μας εάν οι μεταβλητές που έχουμε ήταν ανεξάρτητες μεταξύ τους (η έννοια της ανεξαρτησίας θα αναφερθεί εκτενέστερα παρακάτω). Οι τιμές αυτές υπολογίζονται πολλαπλασιάζοντας το σύνολο της γραμμής επί το σύνολο της στήλης και διαιρώντας με το μέγεθος του δείγματος. Δηλαδή $n_{i+} * n_{+j} / N$.

Για την εξήγηση των παρακάτω συμβολισμών ας θεωρήσουμε το επόμενο παράδειγμα:

Έχουμε δύο ποιοτικές μεταβλητές, το φύλο και το έμφραγμα, οι οποίες παρουσιάζονται σε έναν πίνακα συνάφειας. Από μια μελέτη βρέθηκαν τα παρακάτω : 30 άνδρες έπαθαν έμφραγμα, 14 άνδρες δεν έπαθαν έμφραγμα, 17 γυναίκες έπαθαν έμφραγμα και 35 γυναίκες δεν έπαθαν έμφραγμα. Οι τιμές αυτές είναι οι συχνότητες και παρουσιάζονται στον ακόλουθο 2×2 πίνακα συνάφειας :

Φύλο	Έμφραγμα		Σύνολο
	Ναι	Όχι	
Άνδρας	30	14	44
Γυναίκα	17	35	52
Σύνολο	47	49	96

Στο παράδειγμα αυτό οι παρατηρηθείσες συχνότητες είναι : $n_{11} = 30$ και $n_{12} = 14$ για την πρώτη γραμμή του πίνακα. Η τιμή 44, η οποία συμβολίζεται με n_{1+} , ισούται με το άθροισμα των δύο παραπάνω συχνοτήτων. Για την δεύτερη γραμμή του πίνακα έχουμε $n_{21} = 17$ και $n_{22} = 35$. Η τιμή 52, η οποία συμβολίζεται με n_{2+} , ισούται με το άθροισμα των συχνοτήτων της δεύτερης γραμμής του πίνακα. Η τιμή 47, η οποία συμβολίζεται με n_{+1} , ισούται με το άθροισμα των n_{11} και n_{21} . Τέλος η τιμή 49, η οποία συμβολίζεται με n_{+2} , ισούται με το άθροισμα των n_{12} και n_{22} .

- Ακόμη, μπορούμε να υπολογίσουμε τις *σχετικές συχνότητες* που υπολογίζονται εάν διαιρέσουμε την παρατηρημένη συχνότητα κάθε κελιού (n_{ij}) με το σύνολο του δείγματος, δηλαδή n_{ij}/N , και μας δείχνουν πόσο τις εκατό αντιστοιχεί το κάθε κελί στο σύνολο του δείγματος. Για παράδειγμα, το $\frac{30}{96} = 31,25\%$ του δείγματος είναι άνδρες και έπαθαν έμφραγμα, το $\frac{14}{96} = 14,6\%$ του δείγματος είναι άνδρες και δεν έπαθαν έμφραγμα κλπ. Οι παραπάνω συχνότητες μπορούν να υπολογιστούν ξεχωριστά ανά γραμμή και ανά στήλη.
 - Οι *σχετικές συχνότητες ανά γραμμή* υπολογίζονται εάν διαιρέσουμε την παρατηρημένη συχνότητα κάθε κελιού (n_{ij}) με το σύνολο της

αντίστοιχης γραμμής, δηλαδή n_{ij}/n_{i+} , και αντιπροσωπεύουν το ποσοστό κάθε κελιού στο σύνολο της γραμμής. Πρόκειται για δεσμευμένες πιθανότητες οι οποίες γράφονται ως εξής $\pi_{ij}/\pi_{i+} = P(Y = j|X = i)$ (Ηλιόπουλος, 2017). Αυτό δηλώνει πως, δεδομένου ότι εμφανίζεται κάποιο χαρακτηριστικό της ποιοτικής μεταβλητής που βρίσκεται στη γραμμή του πίνακα συνάφειας, παρατηρούμε ότι τόσα άτομα εμφάνισαν το χαρακτηριστικό της ποιοτικής μεταβλητής που βρίσκεται στην στήλη του πίνακα. Οπότε σύμφωνα με το προηγούμενο παράδειγμα, το $\frac{30}{44} = 68,18\%$ των ανδρών έπαθαν έμφραγμα, ενώ το $\frac{14}{44} = 31,81\%$ δεν έπαθαν έμφραγμα κλπ. Εδώ δηλαδή δεσμεύσαμε ως προς το χαρακτηριστικό «φύλο» και καταλήξαμε ότι ένα μεγάλο ποσοστό των ανδρών έπαθαν έμφραγμα.

- Οι σχετικές συχνότητες ανά στήλη υπολογίζονται εάν διαιρέσουμε την παρατηρημένη συχνότητα κάθε κελιού (n_{ij}) με το σύνολο της αντίστοιχης στήλης, δηλαδή n_{ij}/n_{+j} , και αντιπροσωπεύουν το ποσοστό κάθε κελιού στο σύνολο της στήλης. Πρόκειται για δεσμευμένες πιθανότητες οι οποίες γράφονται ως εξής $\pi_{ij}/\pi_{+j} = P(X = i|Y = j)$ (Ηλιόπουλος, 2017). Αυτό σημαίνει πως, δεδομένου ότι παρατηρείται στο υπό εξεταζόμενο δείγμα κάποιο χαρακτηριστικό της ποιοτικής μεταβλητής που βρίσκεται στην στήλη του πίνακα συνάφειας, παρατηρούμε ότι τόσα άτομα εμφάνισαν το χαρακτηριστικό της ποιοτικής μεταβλητής που βρίσκεται στην γραμμή του πίνακα. Οπότε σύμφωνα με το προηγούμενο παράδειγμα, το $\frac{30}{47} = 63,82\%$ αυτών που έπαθαν έμφραγμα είναι άνδρες, ενώ το υπόλοιπο $\frac{17}{47} = 36,17\%$ είναι γυναίκες κλπ. Εδώ δηλαδή δεσμεύσαμε ως προς το χαρακτηριστικό «έμφραγμα» και καταλήξαμε στο συμπέρασμα ότι το μεγαλύτερο ποσοστό αυτών που έπαθαν έμφραγμα είναι άνδρες.

2.5 Έλεγχος ανεξαρτησίας

Μία στατιστική ανάλυση δεν περιορίζεται σχεδόν ποτέ στη μελέτη μίας μεταβλητής, αλλά μας ενδιαφέρει να μελετάμε τις σχέσεις μεταξύ δύο ή περισσότερων μεταβλητών. Η τεχνική που ακολουθείται για την μελέτη των σχέσεων μεταξύ μεταβλητών διαφοροποιείται ανάλογα με το είδος των μεταβλητών που θέλουμε να συγκρίνουμε, δηλαδή ποσοτικές ή ποιοτικές μεταβλητές. Εμείς θα ασχοληθούμε με την μελέτη των σχέσεων μεταξύ δύο *ποιοτικών* μεταβλητών. Είναι δηλαδή οι δύο υπό εξέταση ποιοτικές μεταβλητές ανεξάρτητες ή εξαρτημένες; Όταν μελετάμε τη σχέση μεταξύ δύο ή περισσότερων ποιοτικών μεταβλητών τότε αντί για συσχέτιση χρησιμοποιείται ο όρος συνάφεια (contingency). Γι' αυτό και οι μεταβλητές που μας ενδιαφέρει να μελετήσουμε παρουσιάζονται σε έναν πίνακα συνάφειας (contingency table). Για να ελέγξουμε μια υπόθεση σαν την παραπάνω θα πρέπει να εφαρμόσουμε τον έλεγχο ανεξαρτησίας χ^2 (Chi-Square Test). Ο έλεγχος αυτός διαφοροποιείται ανάλογα με το είδος του πίνακα συνάφειας που έχουμε και όταν το δείγμα μας είναι «μικρό».

1. Πίνακας Συνάφειας $I \times J$ – Έλεγχος Ανεξαρτησίας

Στην περίπτωση που η μία ποιοτική μεταβλητή έχει I επίπεδα (κατηγορίες) και η δεύτερη ποιοτική μεταβλητή έχει J επίπεδα, για να ελέγξουμε αν υπάρχει σχέση μεταξύ των δύο μεταβλητών θα παρουσιάσουμε τα δεδομένα μας σε έναν $I \times J$ πίνακα συνάφειας και στην συνέχεια θα εφαρμόσουμε τον έλεγχο ανεξαρτησίας χ^2 .

Οι πιθανότητες που θα χρησιμοποιήσουμε στην προκειμένη περίπτωση είναι οι εξής :

- Από κοινού πιθανότητες για κάθε ζευγάρι μεταβλητών X και Y (π_{ij}), δηλαδή την πιθανότητα ένα μέλος του πληθυσμού να ανήκει στο κελί (i, j) .
- Τις περιθώριες πιθανότητες π_{i+} και π_{+j} για την μεταβλητή X και Y αντίστοιχα. Δηλαδή την πιθανότητα ένα μέλος του πληθυσμού να ανήκει στο επίπεδο i της μεταβλητής X και την πιθανότητα ένα μέλος του πληθυσμού να ανήκει στο επίπεδο j της μεταβλητής Y αντίστοιχα.

Οι παραπάνω πιθανότητες μπορούν να εκτιμηθούν μέσω των σχέσεων

$$p_{ij} = n_{ij}/N, p_{i+} = n_{i+}/N \text{ και } p_{+j} = n_{+j}/N \text{ αντίστοιχα.}$$

Δηλαδή, για να υπολογίσουμε την εκτίμηση της από κοινού πιθανότητας π_{ij} διαιρούμε την παρατηρημένη τιμή n_{ij} με το μέγεθος του δείγματος N . Αντίστοιχα, για να υπολογίσουμε την εκτίμηση της περιθώριας πιθανότητας π_{i+} διαιρούμε την ποσότητα n_{i+} με το μέγεθος του δείγματος N . Τέλος, για να υπολογίσουμε την εκτίμηση της περιθώριας πιθανότητας π_{+j} διαιρούμε την ποσότητα n_{+j} με το μέγεθος του δείγματος N .

Η μηδενική υπόθεση είναι η υπόθεση την οποία ενδιαφερόμαστε να ελέγξουμε και είναι πάντα η υπόθεση της μη διαφοροποίησης ή της ισότητας. Αντιθέτως, η εναλλακτική είναι η συμπληρωματική υπόθεση. Στη συγκεκριμένη περίπτωση οι υποθέσεις αυτές ορίζονται ως εξής:

H_0 : οι υπό εξέταση μεταβλητές είναι *ανεξάρτητες* μεταξύ τους

Έναντι

H_1 : οι υπό εξέταση μεταβλητές είναι *εξαρτημένες*

Μαθηματικά, η μηδενική υπόθεση (υπόθεση της ανεξαρτησίας) γράφεται ως εξής :

$$H_0: \pi_{ij} = \pi_{i+}\pi_{+j},$$

δηλαδή η από κοινού πιθανότητα του κελιού (i,j) είναι ίση με το γινόμενο των δύο περιθωρίων πιθανοτήτων.

Έστω

- n_{ij} η παρατηρηθείσα συχνότητα για το κελί (i,j) και
- μ_{ij} η αναμενόμενη συχνότητα για το κελί (i,j)

Ο τύπος υπολογισμού του ελέγχου ανεξαρτησίας χ^2 (ή του χ^2 του Pearson) είναι ο εξής :

$$X^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}} = \sum_{i=1}^I \sum_{j=1}^J \frac{(Observed - Expected)^2}{Expected} \quad (2.5.1)$$

Το X^2 είναι ένα στατιστικό μέτρο το οποίο παράγεται λαμβάνοντας υπόψη τις αποκλίσεις των παρατηρούμενων από τις αναμενόμενες συχνότητες και υπολογίζεται για όλα τα κελιά του πίνακα. Μεγάλες αποκλίσεις των αναμενόμενων τιμών από τις παρατηρούμενες τιμές υποδηλώνουν πιθανή εξάρτηση των υπό εξέταση μεταβλητών.

Υπό την μηδενική υπόθεση της ανεξαρτησίας (εάν δηλαδή οι υπό εξέταση μεταβλητές είναι πράγματι ανεξάρτητες μεταξύ τους)

- οι αναμενόμενες συχνότητες μ_{ij} (expected values) υπολογίζονται ως εξής :

$$\mu_{ij} = \frac{n_{i+} * n_{+j}}{N}.$$

- Η στατιστική συνάρτηση ελέγχου X^2 ακολουθεί, υπό την υπόθεση H_0 της ανεξαρτησίας, την κατανομή $\chi^2_{(I-1)(J-1)}$, δηλαδή την κατανομή χ^2 με $(I-1)(J-1)$ βαθμούς ελευθερίας.

Απορρίπτουμε δηλαδή την μηδενική υπόθεση (υπόθεση ανεξαρτησίας) και θεωρούμε ότι οι δύο μεταβλητές δεν είναι ανεξάρτητες μεταξύ τους εάν η τιμή της στατιστικής συνάρτησης X^2 , που υπολογίζεται βάσει της σχέσης 2.5.1, είναι μεγαλύτερη ή ίση από την τιμή $\chi^2_{(I-1)(J-1),\alpha}$ ή όταν η τιμή του p-value είναι μικρότερη του επιπέδου σημαντικότητας α (p-value < α).

Για να εφαρμόσουμε τον έλεγχο ανεξαρτησίας χ^2 πρέπει να τηρούνται κάποιες προϋποθέσεις (ή παραδοχές) οι οποίες είναι οι εξής :

- Θα πρέπει τα δεδομένα μας να προέρχονται από τυχαία δειγματοληψία από τον πληθυσμό
- Οι μεταβλητές που εξετάζονται πρέπει να είναι ποιοτικές
- Οι παρατηρήσεις πρέπει να είναι ανεξάρτητες, κάθε παρατήρηση δηλαδή πρέπει να προέρχεται από διαφορετικό υποκείμενο (ερωτώμενο)
- Θα πρέπει το πολύ το 20% των κελιών του πίνακα να έχει αναμενόμενη συχνότητα κάτω από 5.
- Όλες οι αναμενόμενες συχνότητες να είναι \geq του 1 (Agresti, 2007).

Αν η προτελευταία παραδοχή δεν τηρείται τότε στην περίπτωση των 2×2 πινάκων συνάφειας χρησιμοποιείται ο ακριβής έλεγχος του Fisher, ενώ σε κάθε άλλη περίπτωση χρειάζεται να γίνει συγχώνευση γειτονικών κελιών ή κατηγοριών της μεταβλητής που μας ενδιαφέρει, κατά τέτοιο τρόπο ώστε να εξαλείφεται το παραπάνω πρόβλημα.

2. Πίνακας Συνάφειας 2×2 – Έλεγχος Ανεξαρτησίας

Πολλές φορές τις ιατρικές μελέτες απασχολεί η μελέτη ποιοτικών μεταβλητών που παίρνουν μόνο δύο τιμές (δίτιμες ποιοτικές μεταβλητές), όπως παρουσία ή όχι μιας νόσου, καπνιστής ή όχι, θάνατος ή όχι κλπ. Επιπλέον οι παράγοντες κινδύνου που μελετάμε, ως προς το αν επηρεάζουν την εμφάνιση της νόσου, είναι επίσης ποιοτικές

(πολλές φορές δίτιμες μεταβλητές), όπως άνδρας ή γυναίκα, κατανάλωση ή όχι αλκοόλ κλπ. (Πολίτης, 2016).

Στον συγκεκριμένο πίνακα συνάφειας έχουμε 4 κελιά συχνοτήτων (παρατηρημένων τιμών), ανάλογα με τις τιμές των δύο παραγόντων, A και B. Εδώ θα στηριχθούμε στον Πίνακα 2.2 που δημιουργήσαμε σε παραπάνω ενότητα.

Συνήθως ο Παράγοντας B, ο οποίος τοποθετείται στις στήλες του πίνακα συνάφειας και δηλώνει εάν παρουσιάζεται ή όχι η ασθένεια που μας ενδιαφέρει, είναι η μεταβλητή απόκρισης (εξαρτημένη μεταβλητή) και ο Παράγοντας A είναι μία ερμηνευτική, ή ανεξάρτητη, μεταβλητή.

Αυτό που μας ενδιαφέρει είναι να συγκρίνουμε δύο ομάδες ως προς μια δίτιμη μεταβλητή απόκρισης (Παράγοντας B) ή, διαφορετικά, εάν οι δύο μεταβλητές είναι ανεξάρτητες. Η μηδενική υπόθεση (ή υπόθεση ανεξαρτησίας) είναι ίδια με εκείνη του πίνακα συνάφειας $I \times J$ που είδαμε παραπάνω, δηλαδή

$$H_0: \pi_{ij} = \pi_{i+}\pi_{+j}$$

Ο τύπος υπολογισμού του ελέγχου ανεξαρτησίας χ^2 (ή του χ^2 του Pearson) για έναν 2×2 πίνακα συνάφειας είναι ο εξής :

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}} = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(Observed - Expected)^2}{Expected} \quad (2.5.2)$$

Υπό την μηδενική υπόθεση της ανεξαρτησίας (εάν δηλαδή οι υπό εξέταση μεταβλητές είναι πράγματι ανεξάρτητες), η παραπάνω στατιστική συνάρτηση ελέγχου ακολουθεί την κατανομή χ^2_1 , δηλαδή την κατανομή χ^2 με 1 βαθμό ελευθερίας.

3. Ο ακριβής Έλεγχος του Fisher

Εάν το μέγεθος του δείγματος που εξετάζουμε είναι μικρό (<50), τότε αντί του ελέγχου ανεξαρτησίας χ^2 χρησιμοποιείται ο ακριβής έλεγχος του Fisher (Fisher's Exact Test). Για τον έλεγχο αυτό θα χρησιμοποιήσουμε τις παρατηρούμενες συχνότητες n_{ij} δοθέντων των περιθωρίων αθροισμάτων n_{i+} , n_{+j} . Ο συγκεκριμένος έλεγχος συγκρίνει τον πίνακα του δείγματος μόνο με εκείνους τους πίνακες που έχουν τα ίδια περιθώρια αθροίσματα.

Έστω ότι έχουμε δύο ανεξάρτητες μεταβλητές X_1 και X_2 , οι οποίες ακολουθούν την διωνυμική κατανομή με παραμέτρους n και p_1 και n και p_2 αντίστοιχα. Τα ποσοστά p_1

και p_2 καλούνται ποσοστά επιτυχίας και εμάς μας ενδιαφέρει να ελέγξουμε την υπόθεση της ισότητας των συγκεκριμένων ποσοστών, δηλαδή

$$H_0: p_1 = p_2 = p$$

Υπό την μηδενική υπόθεση ισχύει ότι το άθροισμά των δύο ανεξάρτητων μεταβλητών ακολουθεί την διωνυμική κατανομή με παραμέτρους n και p , δηλαδή $X_1+X_2 \sim B(n, p)$.

Υπό αυτές τις συνθήκες, ισχύει ότι η τυχαία μεταβλητή $(X_1|X_1 + X_2 = x)$ ακολουθεί την υπεργεωμετρική κατανομή $Hg(n, x, n_1/n)$ οπότε

$$P(X_1 = n_{11} | n_{1+}) = \frac{n_{1+}! n_{2+}! n_{+1}! n_{+2}!}{n! n_{11}! n_{12}! n_{21}! n_{22}!} \quad (\text{Σαχλάς, 2016}).$$

ΚΕΦΑΛΑΙΟ 3

ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

Σκοπός του κεφαλαίου αυτού είναι να δοθεί ένας ορισμός για τους μη παραμετρικούς ελέγχους υποθέσεων και για το πότε χρησιμοποιούνται οι συγκεκριμένοι έλεγχοι. Στη συνέχεια, θα αναλύσουμε ξεχωριστά τους μη παραμετρικούς ελέγχους και θα πούμε πότε, ανάλογα με την ανάλυση που επιθυμούμε να κάνουμε, επιλέγουμε να χρησιμοποιήσουμε τον καθένα ξεχωριστά.

3.1 Βασικές Έννοιες

Στους παραμετρικούς ελέγχους υποθέσεων επιθυμούμε να ελέγξουμε αν μία ή περισσότερες παράμετροι ενός πληθυσμού ικανοποιούν μια βασική υπόθεση (η οποία καλείται μηδενική) έναντι μιας δεύτερης (εναλλακτικής) υπόθεσης. Η μηδενική υπόθεση (null hypothesis) είναι η υπόθεση την οποία ενδιαφερόμαστε να ελέγξουμε και είναι πάντα η υπόθεση της μη διαφοροποίησης ή της ισότητας και συμβολίζεται με H_0 . Αντιθέτως, η εναλλακτική υπόθεση (alternative hypothesis) συμβολίζεται με H_1 . Οι παρακάτω έλεγχοι υποθέσεων που θα μας απασχολήσουν είναι δίπλευρης μορφής. Η απόφαση για την απόρριψη ή μη της H_0 γίνεται βάσει πιθανοτήτων γι' αυτόν τον λόγο ορίζουμε επίπεδο εμπιστοσύνης $(1-\alpha)$ για την απόφαση ελέγχου. Το α λέγεται επίπεδο σημαντικότητας (significance level) και είναι η πιθανότητα της απόρριψης της μηδενικής υπόθεσης, δεδομένου ότι ισχύει. Η πιο συνήθης τιμή για το α είναι $\alpha = 5\%$. Η απόφασή μας (απόρριψη ή όχι της μηδενικής υπόθεσης) βασίζεται σε δείγμα το οποίο χρησιμοποιούμε για να εφαρμόσουμε τους ελέγχους μας. Το επίπεδο σημαντικότητας α το συγκρίνουμε με το p-value και αποφασίζουμε αν πρέπει ή όχι να απορρίψουμε την H_0 . Το p-value είναι η πιθανότητα η στατιστική συνάρτηση ελέγχου να πάρει μία τιμή τόσο ακραία ή περισσότερο ακραία από αυτήν που πήρε για το συγκεκριμένο δείγμα, δεδομένου ότι ισχύει η μηδενική υπόθεση. Με λίγα λόγια, είναι το μικρότερο επίπεδο σημαντικότητας στο οποίο η αρχική υπόθεση μπορεί να απορριφθεί. Η στατιστική συνάρτηση ελέγχου (test statistic) είναι απλώς μια συνάρτηση που υπολογίζεται από τα δεδομένα μας. Με βάση τη τιμή που λαμβάνει, αποφασίζουμε αν θα απορρίψουμε ή όχι την αρχική υπόθεση.

- Αν η τιμή p-value είναι μικρότερη από το επίπεδο σημαντικότητας α ($p - value < \alpha$), τότε απορρίπτουμε την H_0 (αποδεχόμαστε την H_1) σε επίπεδο σημαντικότητας α .
- Αν η τιμή p-value είναι μεγαλύτερη από το επίπεδο σημαντικότητας α ($p - value > \alpha$), τότε δεν απορρίπτουμε την H_0 σε επίπεδο σημαντικότητας α .

3.2 Κατηγορίες στατιστικών ελέγχων

Οι στατιστικοί έλεγχοι διακρίνονται στις ακόλουθες δύο κατηγορίες :

- Οι παραμετρικοί έλεγχοι υποθέσεων και
- Οι μη παραμετρικοί έλεγχοι υποθέσεων

Αρχικά εξετάζουμε αν οι παρατηρήσεις μας ακολουθούν κάποια γνωστή κατανομή (π.χ. κανονική κατανομή) και στη συνέχεια αποφασίζουμε εάν θα χρησιμοποιήσουμε παραμετρικό ή μη παραμετρικό έλεγχο.

Παραμετρικοί έλεγχοι υποθέσεων ονομάζονται οι έλεγχοι των υποθέσεων οι οποίες εκτιμούν μία ή περισσότερες άγνωστες παραμέτρους του πληθυσμού οι οποίοι γίνονται κάτω από την υπόθεση ότι η κατανομή του πληθυσμού είναι γνωστή, συνήθως κανονική, ή τουλάχιστον προσεγγίζει την κανονική. Βασίζονται σε ελεγχοσυναρτήσεις όπου, όπως προαναφέραμε, η κατανομή των δεδομένων μας είναι γνωστή (π.χ. Κανονική, Εκθετική κλπ.) (<https://statistics.scientist.gr>).

Μη παραμετρικοί έλεγχοι υποθέσεων ονομάζονται οι έλεγχοι που χρησιμοποιούμε όταν η κατανομή του πληθυσμού δεν μπορεί να υποτεθεί γνωστή και μπορούν να εφαρμοστούν όταν τα δείγματα είναι μικρά. Δεν επικεντρώνονται στον έλεγχο υποθέσεων αναφορικά με συγκεκριμένες παραμέτρους του πληθυσμού (Ρούσσο, 2007). Στο κεφάλαιο αυτό θα επικεντρωθούμε μόνο στους μη παραμετρικούς ελέγχους υποθέσεων.

Οι μη παραμετρικοί έλεγχοι υποθέσεων, τους οποίους θα αναλύσουμε παρακάτω, είναι οι εξής:

- Ο έλεγχος U των Mann-Whitney
- Ο έλεγχος H των Kruskal-Wallis
- Ο έλεγχος του Wilcoxon

3.3 Στατιστικοί Έλεγχοι Κανονικότητας

Όπως είπαμε παραπάνω, για να αποφασιστεί αν θα χρησιμοποιηθεί παραμετρικός ή μη παραμετρικός έλεγχος εξετάζουμε αρχικά την κανονικότητα των παρατηρήσεών μας τόσο χρησιμοποιώντας στατιστικούς ελέγχους όσο και κάποια γραφήματα. Οι στατιστικοί έλεγχοι που χρησιμοποιούνται είναι αυτοί των Kolmogorov-Smirnov (K-S) και των Shapiro-Wilk (S-W).

Η μηδενική υπόθεση που εξετάζουμε είναι εάν τα δεδομένα μας μπορεί να προέρχονται από την κανονική κατανομή ενώ η εναλλακτική υπόθεση είναι ότι τα δεδομένα μας δεν προέρχονται από την κανονική κατανομή.

Το κριτήριο των K-S είναι ένας μη παραμετρικός έλεγχος που χρησιμοποιείται για να εξετάσει την καλή προσαρμογή ενός τυχαίου δείγματος σε μια δεδομένη κατανομή, και βασίζεται στη διαφορά της εμπειρικής συνάρτησης κατανομής που υπολογίζεται από το δείγμα και της αναμενόμενης συνάρτησης κατανομής υπό την υπόθεση της κανονικότητας, ή κάποιας άλλης κατανομής υπό τη μηδενική υπόθεση. Για να τον εφαρμόσουμε πρέπει το μέγεθος του δείγματος να είναι μεγάλο (50 ή περισσότερο) και οι παράμετροι της αναμενόμενης κατανομής (π.χ. η μέση τιμή και διασπορά μιας κανονικής κατανομής) είναι γνωστοί εκ των προτέρων (Gibbons et al., 2009).

Το κριτήριο των S-W είναι ένας ακόμα πολύ γνωστός μη παραμετρικός έλεγχος για το αν τα δεδομένα μας μπορεί να προέρχονται από κανονικό πληθυσμό. Πολλοί ερευνητές υποστηρίζουν ότι ο συγκεκριμένος έλεγχος ίσως είναι και ο πιο αξιόπιστος, ιδιαίτερα αν αποφασίζει υπέρ της απόρριψης της υπόθεσης της κανονικότητας (Ευαγγελάρας, 2016).

3.4 Γραφικοί Έλεγχοι Κανονικότητας

Ο έλεγχος κανονικότητας επιτυγχάνεται, πέρα από την χρήση κατάλληλων στατιστικών ελέγχων, και με τη χρήση γραφημάτων. Πιο συγκεκριμένα με τη χρήση ιστογραμμάτων ή διαγραμμάτων P-P και Q-Q plots.

Το ιστόγραμμα είναι το πιο συνηθισμένο διάγραμμα περιγραφής ποσοτικών (scale) δεδομένων. Ενδιαφέρον παρουσιάζει και η προσαρμογή της καμπύλης της κανονικής κατανομής, καθώς μας δίνει πληροφορία για το αν τα δεδομένα μας μπορεί να προέρχονται από την κανονική κατανομή (Ευαγγελάρας, 2016).

Το P-P plot είναι ένα γράφημα πιθανότητας το οποίο μας βοηθά να καταλάβουμε εάν τα δεδομένα μας συμπίπτουν ή τείνουν στην ευθεία γραμμή του γραφήματος. Εάν τα σημεία τείνουν προς μία ευθεία γραμμή συνεπάγεται ότι τα δεδομένα ακολουθούν την κατανομή που έχουμε ορίσει.

Τέλος, το Q-Q plot είναι ένα γράφημα πιθανότητας, συχνά τοξοειδούς ή S μορφής, για τη γραφική σύγκριση δύο κατανομών πιθανότητας, απεικονίζοντας τα ποσοστημόρια της μιας σε σχέση με την άλλη. Χρησιμοποιείται για να συγκρίνουμε τα σχήματα των συναρτήσεων κατανομής, παρέχοντας μία γραφική άποψη για το πώς ιδιότητες είναι παρόμοιες ή διαφορετικές στις δύο κατανομές. *Εάν οι κατανομές είναι ίσες, τότε τα σημεία θα βρίσκονται στην ευθεία $y=x$. Εάν οι κατανομές έχουν κάποια σχέση, τότε τα σημεία θα βρίσκονται περίπου σε μία γραμμή* (Χαλικιάς, 2015).

3.5 Μη παραμετρικοί έλεγχοι υποθέσεων

3.5.1 Ο έλεγχος U των Mann-Whitney

Ο συγκεκριμένος έλεγχος εφαρμόζεται για δύο ανεξάρτητα δείγματα στην περίπτωση που οι πληθυσμοί που ελέγχονται δεν κατανέμονται κανονικά. Γενικά εντοπίζει αν τα δύο δείγματα που εξετάζουμε είναι ομογενή ή αν κάποιο διαφέρει από το άλλο. Η βασική μηδενική υπόθεση H_0 που ελέγχεται είναι ότι οι δύο (ανεξάρτητοι) πληθυσμοί προέρχονται από τον ίδιο πληθυσμό (ή έχουν την ίδια κατανομή). Με λίγα λόγια, η μηδενική και η εναλλακτική υπόθεση που καλούμαστε να ελέγξουμε είναι αντίστοιχα :

H_0 : Δεν υπάρχει στατιστικά σημαντική διαφορά της ποσοτικής μεταβλητής ανάμεσα στα δύο επίπεδα της ποιοτικής μεταβλητής

H_1 : Υπάρχει στατιστικά σημαντική διαφορά της ποσοτικής μεταβλητής ανάμεσα στα δύο επίπεδα της ποιοτικής μεταβλητής

Μαθηματικά, αυτές γράφονται ως εξής:

$$H_0: \mu_1 = \mu_2 \text{ έναντι}$$

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

Ας υποθέσουμε ότι έχουμε ένα δείγμα x_1, x_2, \dots, x_n μεγέθους n από έναν πληθυσμό, και ένα δεύτερο δείγμα y_1, y_2, \dots, y_m μεγέθους m από έναν δεύτερο πληθυσμό. Ο έλεγχος U των Mann-Whitney βασίζεται στην σύγκριση κάθε παρατήρησης από το

πρώτο δείγμα με κάθε παρατήρηση από το δεύτερο δείγμα. Ο συνολικός αριθμός συγκρίσεων που θα πραγματοποιήσουμε είναι nm .

Για τον υπολογισμό της τιμής της συνάρτησης U των Mann-Whitney μετράμε τον αριθμό των φορών που μια παρατήρηση x_i από το πρώτο δείγμα είναι μεγαλύτερη από μια παρατήρηση y_j από το δεύτερο δείγμα. Αυτόν τον αριθμό τον συμβολίζουμε με U_X . Παρόμοια, μετράμε τον αριθμό των φορών που μια x_i από το πρώτο δείγμα είναι μικρότερη από μια y_j από το δεύτερο δείγμα και συμβολίζουμε αυτόν τον αριθμό με U_Y . Υπό τη μηδενική υπόθεση περιμένουμε τα U_X και U_Y να είναι περίπου ίσα.

Θέτουμε $U = \min\{U_X, U_Y\}$. Για την απόρριψη ή μη της μηδενικής υπόθεσης βρίσκουμε, με την βοήθεια της εικόνας 3.5.1. του παραρτήματος, την κρίσιμη τιμή U_{nm} . Αν $U < U_{nm}$ τότε απορρίπτουμε την μηδενική υπόθεση σε επίπεδο σημαντικότητας α και καταλήγουμε στο συμπέρασμα ότι υπάρχει στατιστικά σημαντική διαφορά της ποσοτικής μεταβλητής ανάμεσα στα δύο επίπεδα της ποιοτικής μεταβλητής. Όταν ο αριθμός των παρατηρήσεων στο δείγμα είναι τέτοιος ώστε $nm > 20$ τότε ακολουθούμε την κανονική προσέγγιση. Για τους υπολογισμούς θέτουμε

$$\mu_U = \frac{nm}{2} \text{ και } \sigma_U = \sqrt{\frac{nm(n+m+1)}{12}}$$

Παράδειγμα 3.5.1

Έχουμε, για δύο τμήματα, τις αποδόσεις των φοιτητών στις εξετάσεις, οι οποίες φαίνονται στο παρακάτω πλαίσιο.

Πρώτο τμήμα (x_i) : 73, 87, 79, 75, 82, 66, 95, 75, 70
Δεύτερο τμήμα (y_j) : 86, 81, 84, 88, 90, 85, 84, 92, 83, 91, 53, 84

Το πρώτο τμήμα είναι και το πρώτο δείγμα μεγέθους n που ισούται με 9 ενώ το δεύτερο τμήμα είναι το δεύτερο δείγμα μεγέθους m που ισούται με 12. Ο συνολικός αριθμός συγκρίσεων που θα πραγματοποιήσουμε είναι $nm = 9 * 12 = 108$.

Κάνοντας όλες τις συγκρίσεις βρίσκουμε ότι $U_X = 28$ και $U_Y = 80$. Ως τιμή της στατιστικής συνάρτησης U επιλέγουμε $U = \min\{U_X, U_Y\} = 28$. Από τις κρίσιμες τιμές του πίνακα των Mann-Whitney (βλέπε εικόνα Π.Α.1. στο παράρτημα), για μεγέθη δειγμάτων $n = 9$ και $m = 12$ σε επίπεδο σημαντικότητας $\alpha=0.05$, έχουμε κρίσιμη τιμή $U_{nm} = 26$. Επειδή $U > U_{nm}$, δεν απορρίπτουμε την μηδενική υπόθεση και συμπεραίνουμε ότι η απόδοση των φοιτητών δεν διαφέρει στα δύο τμήματα.

3.5.2 Ο έλεγχος Sum-Rank του Wilcoxon

Ο συγκεκριμένος έλεγχος είναι *ισοδύναμος* με τον έλεγχο των Mann-Whitney και εφαρμόζεται για τη σύγκριση των μέσων *δύο* ανεξάρτητων δειγμάτων στην περίπτωση που οι παρατηρήσεις των δειγμάτων που εξετάζουμε *δεν* προέρχονται από την κανονική κατανομή.

Ας υποθέσουμε ότι έχουμε ένα δείγμα x_1, x_2, \dots, x_n μεγέθους n από έναν πληθυσμό, και ένα δεύτερο δείγμα y_1, y_2, \dots, y_m μεγέθους m από έναν δεύτερο πληθυσμό. Ταξινομούμε και βαθμολογούμε όλες τις $n + m$ παρατηρήσεις από την μικρότερη στη μεγαλύτερη, δίνοντας 1 στη μικρότερη, 2 στην αμέσως μεγαλύτερη παρατήρηση κ.ο.κ. (Rank). Το στατιστικό W_n είναι το άθροισμα των βαθμολογιών (Ranks) των παρατηρήσεων του δείγματος το οποίο έχει μικρότερο μέσο βαθμό. Η τιμή W_n συγκρίνεται με την κατάλληλη κρίσιμη τιμή W_{cr} , Εικόνα Π.Α.2. του παραρτήματος, για την απόρριψη ή μη της μηδενικής υπόθεσης. Αν $W_n > W_{cr}$, απορρίπτουμε την μηδενική υπόθεση (Ευαγγελάρας, 2016).

3.5.3 Ο έλεγχος H των Kruskal-Wallis

Σε αντίθεση με τις προηγούμενες περιπτώσεις, ο συγκεκριμένος έλεγχος χρησιμοποιείται όταν εξετάζουμε k ανεξάρτητα δείγματα στην περίπτωση που οι υποθέσεις περί κανονικότητας των k πληθυσμών δεν μπορεί να επαληθευτούν. Η βασική μηδενική υπόθεση H_0 που ελέγχεται είναι ότι οι k (ανεξάρτητοι) πληθυσμοί προέρχονται από τον ίδιο πληθυσμό (ή έχουν την ίδια κατανομή). Η μηδενική και η εναλλακτική υπόθεση που καλούμαστε να ελέγξουμε είναι αντίστοιχα :

H_0 : Δεν υπάρχει στατιστικά σημαντική διαφορά της ποσοτικής μεταβλητής ανάμεσα στα k επίπεδα της ποιοτικής μεταβλητής

H_1 : Υπάρχει στατιστικά σημαντική διαφορά της ποσοτικής μεταβλητής ανάμεσα στα k επίπεδα της ποιοτικής μεταβλητής

Συμβολικά, γράφουμε

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k \text{ έναντι}$$

$$H_1: \text{όχι η } H_0$$

Ας υποθέσουμε ότι έχουμε δείγματα μεγέθους n_i , με $i \in \{1, 2, \dots, k\}$, από k ανεξάρτητους πληθυσμούς. Οι πληθυσμοί από τους οποίους πάρθηκαν τα δείγματα

δηλώνονται ως επίπεδα μιας ποιοτικής μεταβλητής (παράγοντα). Θέλουμε, χωρίς να ισχύει η κανονικότητα και η ομοσκεδαστικότητα των k πληθυσμών, να ελέγξουμε την μηδενική υπόθεση H_0 που αναφέραμε παραπάνω.

Αρχικά, οι $N = n_1 + n_2 + \dots + n_k$ συνολικές παρατηρήσεις που λάβαμε από τους k ανεξάρτητους πληθυσμούς ταξινομούνται από την μικρότερη στη μεγαλύτερη και στην κάθε μια δίνεται ένας βαθμός, που καλείται Rank, ανάλογα με το μέγεθος της κάθε παρατήρησης. Δηλαδή 1 στη μικρότερη, 2 στην αμέσως μεγαλύτερη παρατήρηση κ.ο.κ. Σε περίπτωση που έχουμε ομάδες ίδιων παρατηρήσεων (δεσμούς, ties) η βαθμολογία αναπροσαρμόζεται δίνοντας σε κάθε παρατήρηση της ίδιας ομάδας το μέσο βαθμό που προκύπτει από τους αντίστοιχους αρχικούς βαθμούς. Ύστερα, υπολογίζουμε τις ποσότητες R_i , με $i \in \{1, 2, \dots, k\}$, αθροίζοντας τις τελικές βαθμολογίες των παρατηρήσεων από κάθε δείγμα.

Η τιμή της στατιστικής συνάρτησης H των Kruskal-Wallis δίνεται από τον τύπο

$$H = \frac{\frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)}{1 - \sum \frac{t^3 - t}{N^3 - N}},$$

όπου N είναι το σύνολο των παρατηρήσεων. Η άθροιση στον παρονομαστή γίνεται για κάθε ομάδα ίσων παρατηρήσεων μεγέθους t . Στην περίπτωση που δεν υπάρχουν ίσες παρατηρήσεις, τότε κάθε μια θα αποτελεί μια ομάδα μεγέθους $t=1$ και ο παρονομαστής θα ισούται με την μονάδα. Για την απόρριψη ή όχι της μηδενικής υπόθεσης βρίσκουμε, με τη βοήθεια της εικόνας Π.Α.3 στο παράρτημα, την κρίσιμη τιμή H_{cr} . Αν $H > H_{cr}$, απορρίπτουμε την μηδενική υπόθεση. Εναλλακτικά, για την απόρριψη ή μη της μηδενικής υπόθεσης η στατιστική συνάρτηση H συγκρίνεται με την κρίσιμη τιμή της κατανομής χ^2 με $k-1$ βαθμούς ελευθερίας. Αυτό διότι η H ακολουθεί ασυμπτωτικά την συγκεκριμένη κατανομή.

Παράδειγμα 3.5.3

Στον παρακάτω πίνακα δίνονται οι επιδόσεις κάποιων μαθητών στην Φυσική σε 3 λύκεια της Αττικής. Επιλέγουμε τυχαία 5, 4 και 3 μαθητές από κάθε λύκειο και καταγράφουμε το βαθμό που πήραν στη Φυσική.

Λύκειο Α : 12, 10, 14, 15, 19
 Λύκειο Β : 16, 13, 18, 10
 Λύκειο Γ : 17, 16, 20

Το πρώτο δείγμα είναι το λύκειο Α μεγέθους $n_1 = 5$, το δεύτερο δείγμα είναι το λύκειο Β μεγέθους $n_2 = 4$ και τέλος το τρίτο δείγμα είναι το λύκειο Γ μεγέθους $n_3 = 3$. Τα δείγματα αυτά πάρθηκαν από $k=3$ ανεξάρτητους πληθυσμούς όσα και τα επίπεδα της ποιοτικής μεταβλητής «Λύκειο». Αρχικά, οι $N = 5+4+3 = 12$ συνολικές παρατηρήσεις που λάβαμε από τους 3 ανεξάρτητους πληθυσμούς ταξινομούνται κατά αύξουσα σειρά.

Ταξινομημένες Παρατηρήσεις	10	10	12	13	14	15	16	16	17	18	19	20
Rank	1	2	3	4	5	6	7	8	9	10	11	12
Τελικά Rank	1.5	1.5	3	4	5	6	7.5	7.5	9	10	11	12
Δείγμα	B	A	A	B	A	A	B	Γ	Γ	B	A	Γ

Έτσι έχουμε συγκεντρωτικά : $n_1 = 5$ με $R_1 = 26.5$, $n_2 = 4$ με $R_2 = 23.5$ και $n_3 = 3$ με $R_3 = 29.5$. Επίσης έχουμε 2 ομάδες με $t = 2$ ίσες παρατηρήσεις. Η τιμή της στατιστικής H είναι :

$$H = \frac{\frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)}{1 - \sum \frac{t^3 - t}{N^3 - N}} = \frac{\frac{12}{12 \cdot 13} \left(\frac{26.5^2}{5} + \frac{23.5^2}{4} + \frac{29.5^2}{3} \right) - 3 \cdot 13}{1 - \left(\frac{2^3 - 2}{12^3 - 12} + \frac{2^3 - 2}{12^3 - 12} \right)} = 4.771$$

Από τις κρίσιμες τιμές για $\alpha=0.05$ και μεγέθη δειγμάτων 5,4 και 3 βρίσκουμε $H_{cr} = 5.63$. Επειδή $H < H_{cr}$, δεν απορρίπτουμε τη μηδενική υπόθεση και καταλήγουμε ότι οι επιδόσεις των μαθητών στα 3 λύκεια είναι περίπου ισοδύναμες.

ΚΕΦΑΛΑΙΟ 4

ΙΣΟΡΡΟΠΙΑ HARDY-WEINBERG

Σκοπός αυτού του κεφαλαίου είναι να παρουσιάσουμε την αναγκαιότητα και την χρησιμότητα της συγκεκριμένης ισορροπίας στη γενετική καθώς και να ελέγξουμε, μέσω του στατιστικού ελέγχου χ^2 , εάν οι πληθυσμοί των πολυμορφισμών του δείγματός μας είναι σε ισορροπία Hardy-Weinberg.

4.1 Μια σύντομη θεωρητική προσέγγιση

Ο νόμος των Hardy-Weinberg είναι μια αρχή που δηλώνει ότι η γενετική ποικιλότητα σε έναν πληθυσμό θα παραμείνει σταθερή από τη μια γενιά στην άλλη ελλείψει άλλων εξελικτικών επιδράσεων. Δηλαδή τόσο οι συχνότητες των αλληλόμορφων όσο και των γονότυπων θα παραμένουν σταθερές στις μετέπειτα γενιές (Παπανικολάου κ.α., 2015). Η ισορροπία των Hardy-Weinberg είναι ένας από τους θεμελιώδεις νόμους της γενετικής (Νικολάου κ.α., 2015). Η εξέταση για το αν ένας πληθυσμός βρίσκεται σε ισορροπία Hardy-Weinberg πραγματοποιείται μέσω του ελέγχου καλής προσαρμογής τον οποίο θα παρουσιάσουμε αναλυτικά στις επόμενες ενότητες. Στους ανθρώπινους πληθυσμούς σπάνια δεν ισχύει η συγκεκριμένη ισορροπία. Η απόκλιση από την ισορροπία Hardy-Weinberg είναι ένδειξη μικροεξέλιξης που μπορεί να συμβεί στη διάρκεια μιας γενιάς.

4.2 Εξέταση στον πολυμορφισμό ADRB3rs4994

Από την θεωρία γνωρίζουμε ότι γονότυποι για τον πολυμορφισμό αυτό είναι τρεις : ο C:C, ο C:T και ο T:T (<https://www.snpedia.com>). Από την βιολογία γνωρίζουμε ότι ο συνδυασμός δύο αλληλόμορφων καλείται γονότυπος. Συνεπώς, στον συγκεκριμένο πολυμορφισμό, τα αλληλόμορφα είναι δύο, το C και το T. Για τον συγκεκριμένο πολυμορφισμό έχουμε ότι στα δεδομένα ο γονότυπος C:C εμφανίζεται 1 φορά, ο C:T 47 φορές και ο T:T (που καλείται και φυσικός γονότυπος για τον rs4994) εμφανίζεται 358 φορές. Στην συνέχεια αυτό που μας ενδιαφέρει είναι να υπολογίσουμε τις συχνότητες των δύο προαναφερθέντων αλληλόμορφων. Γι' αυτό λοιπόν συμβολίζουμε με p και q τις συχνότητες στον πληθυσμό των αλληλόμορφων C και T αντίστοιχα και για τον υπολογισμό τους θα πρέπει να λάβουμε υπόψη ότι τα ομόζυγα άτομα έχουν δύο

αντίγραφα του εκάστοτε αλληλόμορφου ενώ τα ετερόζυγα ένα. Αν συμβολίσουμε με N_{CC} τον αριθμό των ομόζυγων ατόμων και N_{CT} τον αριθμό των ετερόζυγων, τότε η συχνότητα p του αλληλόμορφου C θα δίνεται από την σχέση :

$$p = \frac{2 * N_{CC} + N_{CT}}{2N} = \frac{2 * 1 + 47}{2 * 406} = 0,060$$

Η συχνότητα q του αλληλόμορφου T θα δίνεται από τον τύπο $q = 1 - p = 1 - 0,060 = 0,939$. Στηριζόμενοι στις παραπάνω συχνότητες των αλληλόμορφων μπορούμε να υπολογίσουμε τον αναμενόμενο (E) από την ισορροπία H-W αριθμό ατόμων για καθέναν από τους τρεις διαφορετικούς γονότυπους C:C, C:T και T:T όπως φαίνεται παρακάτω :

$$E_{CT} = 2 * p * q * n = 46,043$$

$$E_{CC} = p^2 * n = 1,478$$

$$E_{TT} = q^2 * n = 358,47$$

Τα αποτελέσματα των παραπάνω υπολογισμών συνοψίζονται στον παρακάτω πίνακα :

Πίνακας 4.1

Παρατηρούμενος και αναμενόμενος αριθμός γονοτύπων του rs4994

Γονότυποι	C:C	C:T	T:T
Παρατηρούμενος αριθμός ατόμων (O)	1	47	358
Αναμενόμενος αριθμός ατόμων (E)	1,478	46,043	358,47

Ύστερα ο έλεγχος υπόθεσης που θέλουμε να πραγματοποιήσουμε είναι εάν ο πληθυσμός μας ακολουθεί ή όχι την ισορροπία H-W. Η μηδενική και η εναλλακτική υπόθεση γράφονται αντίστοιχα ως εξής :

H_0 : ο πληθυσμός βρίσκεται σε ισορροπία Hardy-Weinberg.

H_1 : ο πληθυσμός δεν βρίσκεται σε ισορροπία Hardy-Weinberg.

Χρησιμοποιώντας τις τιμές που παρουσιάσαμε στον πίνακα 4.1 μπορούμε να υπολογίσουμε το στατιστικό X^2 και να διεξάγουμε τον έλεγχο καλής προσαρμογής. Έχουμε λοιπόν,

$$X^2 = \sum_{i=1}^n \frac{(O - E)^2}{E} = \frac{(1 - 1,478)^2}{1,478} + \frac{(47 - 46,043)^2}{46,043} + \frac{(358 - 358,47)^2}{358,47} = 0,1753.$$

Από την θεωρία γνωρίζουμε ότι το στατιστικό X^2 , υπό την μηδενική υπόθεση, ακολουθεί την κατανομή χ τετράγωνο με 1 βαθμό ελευθερίας, δηλαδή την χ_1^2 . Η αποδοχή ή η απόρριψη της μηδενικής υπόθεσης μπορεί να πραγματοποιηθεί συγκρίνοντας την τιμή του X^2 που υπολογίσαμε παραπάνω με την κριτική τιμή της κατανομής χ_1^2 σε κατάλληλο επίπεδο σημαντικότητας α . Επιλέγουμε να διεξάγουμε τον έλεγχο σε επίπεδο σημαντικότητας 5%. Ανατρέχοντας στον πίνακα της κατανομής χ^2 με 1 βαθμό ελευθερίας και $\alpha=5\%$ βρίσκουμε ότι $\chi_{1,0.05}^2 = 3,84$. Εφόσον $X^2 < \chi_{1,0.05}^2$ καταλήγουμε στο συμπέρασμα ότι δεν απορρίπτουμε την μηδενική υπόθεση σε επίπεδο σημαντικότητας 5%. Άρα ο πληθυσμός βρίσκεται σε ισορροπία Hardy-Weinberg.

4.3 Εξέταση στον πολυμορφισμό ADRB2rs1042713

Οι γονότυποι του συγκεκριμένου πολυμορφισμού είναι τρεις : ο A:A, ο G:A και ο G:G. Και για αυτόν τον πολυμορφισμό τα αλληλόμορφα είναι δύο, το A και το G. Στον πολυμορφισμό rs1042713 ο γονότυπος A:A εμφανίζεται 62 φορές, ο G:A 199 φορές και ο G:G εμφανίζεται 145 φορές.

Η ονομασία των συχνοτήτων των δύο αλληλόμορφων είναι ίδια με αυτή που αναφέραμε στην προηγούμενη ενότητα. Αυτή τη φορά συμβολίζουμε με N_{GG} τον αριθμό των ομόζυγων ατόμων και με N_{GA} τον αριθμό των ετερόζυγων, η συχνότητα του αλληλόμορφου G υπολογίζεται από την εξίσωση :

$$p = \frac{2 * N_{GG} + N_{GA}}{2N} = \frac{2 * 145 + 199}{2 * 406} = 0,6022$$

Αντιθέτως, η συχνότητα του αλληλόμορφου A υπολογίζεται χρησιμοποιώντας τον τύπο $q = 1 - p = 1 - 0,6022 = 0,3978$. Για ακόμη μια φορά θα υπολογίσουμε μπορούμε να υπολογίσουμε τον αναμενόμενο (E) από την ισορροπία H-W αριθμό

ατόμων για καθέναν από τους τρεις διαφορετικούς γονότυπους A:A,G:A,G:G όπως φαίνεται παρακάτω :

$$E_{AA} = q^2 * n = 64,2474$$

$$E_{GG} = p^2 * n = 147,23$$

$$E_{GA} = 2 * p * q * n = 194,518$$

Τα προαναφερθέντα αποτελέσματα θα τα παρουσιάσουμε περιεκτικά στον ακόλουθο πίνακα :

Πίνακας 4.2

Παρατηρούμενος και αναμενόμενος αριθμός γονοτύπων του rs1042713

Γονότυποι	A:A	G:A	G:G
Παρατηρούμενος αριθμός ατόμων (O)	62	199	145
Αναμενόμενος αριθμός ατόμων (E)	64,24	194,52	147,23

Τόσο η μηδενική όσο και η εναλλακτική υπόθεση είναι οι ίδιες με αυτές που παρουσιάστηκαν στην προηγούμενη ενότητα. Χρησιμοποιώντας τις τιμές του παραπάνω πίνακα θα υπολογίσουμε το στατιστικό χ^2 , όπως φαίνεται παρακάτω :

$$\chi^2 = \sum_{i=1}^n \frac{(O-E)^2}{E} = \frac{(62-64,24)^2}{64,24} + \frac{(199-194,52)^2}{194,52} + \frac{(145-147,23)^2}{147,23} = 0,2157.$$

Τέλος, συγκρίνοντας το στατιστικό χ^2 με την κριτική τιμή της χ τετράγωνο κατανομής με έναν βαθμό ελευθερίας, καταλήγουμε στο συμπέρασμα ότι ο πληθυσμός βρίσκεται σε ισορροπία Hardy-Weinberg.

4.4 Εξέταση στον πολυμορφισμό ADRB2rs1042714

Οι κατηγορίες γονοτύπων του συγκεκριμένου πολυμορφισμού είναι τρεις : ο C:C , ο G:C και ο G:G και τα αλληλόμορφα δύο, το C και το G. Η πρώτη κατηγορία εμφανίζεται 166 φορές, η δεύτερη 188 και η τρίτη εμφανίζεται 52 φορές. Συμβολίζοντας με N_{CC} τον αριθμό των ομόζυγων ατόμων και με N_{GC} τον αριθμό των ετερόζυγων, η συχνότητα του αλληλόμορφου C υπολογίζεται από :

$$p = \frac{2*N_{CC}+N_{GC}}{2N} = \frac{2*166+188}{2*406} = 0,6404.$$

Ενώ η συχνότητα του δεύτερου αλληλόμορφου θα δίνεται από τον τύπο $q = 1 - p = 1 - 0,6404 = 0,3596$. Στη συνέχεια υπολογίζουμε τον αναμενόμενο (E) αριθμό ατόμων για καθέναν από τους τρεις διαφορετικούς γονότυπους G:G,G:C,C:C όπως παρουσιάζονται παρακάτω και όλα τα παραπάνω συνοψίζονται στον ακόλουθο πίνακα:

$$\begin{aligned} E_{GG} &= q^2 * n = 52,500 \\ E_{CC} &= p^2 * n = 166,50 \\ E_{GC} &= 2 * p * q * n = 186,993 \end{aligned}$$

Πίνακας 4.3

Παρατηρούμενος και αναμενόμενος αριθμός γονοτύπων του rs1042714

Γονότυποι	C:C	G:C	G:G
Παρατηρούμενος αριθμός ατόμων (O)	166	188	52
Αναμενόμενος αριθμός ατόμων (E)	166,50	186,99	52,500

Όλες οι παραπάνω ποσότητες θα μας χρησιμεύσουν για να υπολογίσουμε το στατιστικό χ^2 το οποίο δίνεται και υπολογίζεται μέσω της σχέσης :

$$\chi^2 = \sum_{i=1}^n \frac{(O-E)^2}{E} = \frac{(166-166,50)^2}{166,50} + \frac{(188-186,99)^2}{186,99} + \frac{(52-52,500)^2}{52,500} = 0,0117.$$

Για μια ακόμη φορά, εφόσον ισχύει ότι $\chi^2 < \chi_{1,0.05}^2$, συμπεραίνουμε ότι και ο συγκεκριμένος πληθυσμός βρίσκεται σε ισορροπία H-W.

4.5 Εξέταση στον πολυμορφισμό ADRB1rs1801252

Ο πολυμορφισμός rs1801252 του γονιδίου ADRB1 έχει τρεις κατηγορίες γονοτύπων : είναι ο A:A , ο G:A και ο G:G ενώ τα αλληλόμορφα είναι δύο, πιο συγκεκριμένα το A και το G. Για τον πολυμορφισμό rs1801252 ο γονότυπος A:A εμφανίζεται 332 φορές, ο G:A 69 φορές και ο G:G εμφανίζεται μόλις 5 φορές. Αυτό που μας ενδιαφέρει είναι να υπολογίσουμε τις συχνότητες των δύο αλληλόμορφων. Συμβολίζουμε με N_{AA} τον αριθμό των ομόζυγων ατόμων και με N_{GA} τον αριθμό των ετερόζυγων, τότε η συχνότητα του αλληλόμορφου A θα δίνεται από την εξίσωση :

$$p = \frac{2 \cdot N_{AA} + N_{GA}}{2N} = \frac{2 \cdot 332 + 69}{2 \cdot 406} = 0,902.$$

Η συχνότητα του αλληλόμορφου G θα δίνεται από τον τύπο $q = 1 - p = 1 - 0,902 = 0,098$. Εν συνεχεία υπολογίζουμε τον αναμενόμενο (E) αριθμό ατόμων για καθέναν από τους τρεις γονότυπους G:G, G:A, A:A όπως φαίνεται παρακάτω :

$$E_{GG} = q^2 * n = 3,899$$

$$E_{AA} = p^2 * n = 330,323$$

$$E_{GA} = 2 * p * q * n = 71,777$$

Τα αποτελέσματα των παραπάνω υπολογισμών συνοψίζονται στον ακόλουθο πίνακα

Πίνακας 4.4

Παρατηρούμενος και αναμενόμενος αριθμός γονοτύπων του rs1801252

Γονότυποι	A:A	G:A	G:G
Παρατηρούμενος αριθμός ατόμων (O)	332	69	5
Αναμενόμενος αριθμός ατόμων (E)	330,32	71,777	3,899

Το στατιστικό X^2 του ελέγχου καλής προσαρμογής υπολογίζεται μέσω της εξίσωσης :

$$X^2 = \sum_{i=1}^n \frac{(O-E)^2}{E} = \frac{(332-330,32)^2}{330,32} + \frac{(69-71,777)^2}{71,777} + \frac{(5-3,899)^2}{3,899} = 0,4268.$$

Ο συγκεκριμένος πληθυσμός βρίσκεται σε ισορροπία Hardy-Weinberg και αυτό γιατί η τιμή του στατιστικού X^2 είναι μικρότερη από την κριτική τιμή της κατανομής χ -τετράγωνο με 1 β.ε.

4.6 Εξέταση στον πολυμορφισμό ADRB1rs1801253

Για τον τελευταίο πολυμορφισμό του δείγματός μας γνωρίζουμε ότι οι γονότυποι είναι τρεις: ο C:C , ο C:G και ο G:G. Τα αλληλόμορφα για τον συγκεκριμένο πολυμορφισμό είναι δύο : το C και το G. Για τον G:G έχουμε συχνότητα 34 ,για τον G:C έχουμε συχνότητα 178 και για τον C:C έχουμε συχνότητα 194. Η συχνότητα του αλληλόμορφου A θα δίνεται από την εξίσωση :

$$p = \frac{2*N_{CC}+N_{GC}}{2N} = \frac{2*194+178}{2*406} = 0,69704.$$

Η συχνότητα του αλληλόμορφου G θα δίνεται από τον τύπο $q = 1 - p = 1 - 0,69704 = 0,30296$

Στηριζόμενοι στις συχνότητες των παραπάνω αλληλόμορφων, μπορούμε να υπολογίσουμε τον αναμενόμενο (E) αριθμό ατόμων για καθέναν από τους γονότυπους όπως φαίνεται παρακάτω:

$$E_{GG} = q^2 * n = 37,265$$

$$E_{CC} = p^2 * n = 197,26$$

$$E_{GC} = 2 * p * q * n = 171,474$$

Ο παρακάτω πίνακας παρουσιάζει εν συντομία όσα βρέθηκαν παραπάνω :

Πίνακας 4.5

Παρατηρούμενος και αναμενόμενος αριθμός γονοτύπων του rs1801253

Γονότυποι	C:C	G:C	G:G
Παρατηρούμενος αριθμός ατόμων (O)	194	178	34
Αναμενόμενος αριθμός ατόμων (E)	197,26	171,474	37,265

Χρησιμοποιώντας αυτές τις τιμές μπορούμε να υπολογίσουμε το χ^2 μέσω του τύπου:

$$\chi^2 = \sum_{i=1}^n \frac{(O-E)^2}{E} = \frac{(194-197,26)^2}{197,26} + \frac{(178-171,474)^2}{171,474} + \frac{(34-37,265)^2}{37,265} = 0,588.$$

Αφού $\chi^2 < \chi_{1,0.05}^2$ δεν απορρίπτουμε την μηδενική υπόθεση. Άρα ο πληθυσμός μπορούμε να πούμε ότι βρίσκεται σε ισορροπία Hardy-Weinberg.

ΚΕΦΑΛΑΙΟ 5

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Σκοπός του κεφαλαίου αυτού είναι αρχικά η παρουσίαση των μεταβλητών που έχουμε στην διάθεσή μας και στην συνέχεια μια αρχική διερευνητική ανάλυση προκειμένου να εξάγουμε κάποια αρχικά αποτελέσματα για τα δεδομένα μας και τη σχέση που μπορεί να έχει η ύπαρξη πολυμορφισμών με την παχυσαρκία.

5.1 Προέλευση και περιγραφή των δεδομένων

Τα δεδομένα μας χορηγήθηκαν από την Φαρμακευτική Σχολή Αθηνών και σε αυτά περιλαμβάνονται οι μετρήσεις που έγιναν σε 406 άτομα και των δύο φύλων. Για τους εθελοντές που συμμετείχαν στην έρευνα έχει καταγραφεί το ύψος τους, το βάρος τους, ο δείκτης μάζας σώματός τους, το φύλο τους καθώς και η ηλικία τους. Στην συνέχεια καταγράφηκαν η χρονολογία γέννησής τους, ο φάκελός τους καθώς και η χρονολογία που πραγματοποιήθηκε η δειγματοληψία. Επίσης καταγράφονται οι πέντε πολυμορφισμοί μέσω των μεταβλητών ADRB3rs4994, ADRB2rs1042713, ADRB2rs1042714, ADRB1rs1801252 και ADRB1rs1801253. Τέλος, δημιουργήθηκε η ποιοτική διατακτική μεταβλητή classificationBMI η οποία κατατάσσει τους εθελοντές του δείγματος, με βάση το δείκτη μάζας σώματός τους, σε κατηγορίες. Για παράδειγμα, όσοι έχουν BMI από 25 μέχρι 30 κατατάσσονται στην κατηγορία των υπέρβαρων (βλ. Πίνακα 1.1).

Πριν προχωρήσουμε στην αρχική ανάλυση θα κάνουμε μια συνοπτική παρουσίαση των μεταβλητών που διαθέτουμε δίνοντας μεγαλύτερη έμφαση στους παραπάνω πέντε πολυμορφισμούς.

- Ο πολυμορφισμός rs4994, γνωστός και ως Trp64Arg ή διαφορετικά ως Arg64, είναι ένας σημειακός πολυμορφισμός ο οποίος βρίσκεται στο γονίδιο ADRB3. Το αλληλόμορφο της κυτοσίνης (C) κωδικοποιεί το αμινοξύ της αργινίνης. Αυτό σημαίνει ότι η ύπαρξη αυτής της βάσης οδηγεί στην έκφραση της αργινίνης. Ο συγκεκριμένος πολυμορφισμός αποτελείται από τρεις κατηγορίες γονοτύπων, τον C:C, τον C:T και τον

T:T. Ο τελευταίος εξ αυτών καλείται φυσικός γονότυπος (<https://www.snpedia.com>).

- Ο πολυμορφισμός rs1042713, γνωστός και ως Arg16, είναι ένας σημειακός πολυμορφισμός ο οποίος βρίσκεται στο γονίδιο του αδρενεργικού β2 υποδοχέα (ADRB2). Το αλληλόμορφο της γουανίνης (G) κωδικοποιεί τη G/ γλυκίνη που είναι το πιο κοινό στους περισσότερους πληθυσμούς ενώ το αλληλόμορφο της αδενίνης κωδικοποιεί την R/ αργινίνη. Ο συγκεκριμένος πολυμορφισμός αποτελείται από τρεις κατηγορίες γονοτύπων, τον G:G, τον G:A και τον A:A. Ο πρώτος εξ αυτών είναι ο φυσικός γονότυπος (<https://www.snpedia.com>).
- Ο πολυμορφισμός rs1042714 είναι ένας σημειακός πολυμορφισμός ο οποίος βρίσκεται και αυτός στο γονίδιο ADRB2, όπως και ο προηγούμενος. Η ύπαρξη της βάσης κυτοσίνης (C) στο γονίδιο κωδικοποιεί τη γλουταμίνη. Ο συγκεκριμένος πολυμορφισμός αποτελείται από τρεις κατηγορίες γονοτύπων, τον C:G, τον C:C και τον G:G. Ο γονότυπος C:C είναι ο φυσικός γονότυπος (<https://www.snpedia.com>).
- Ο σημειακός πολυμορφισμός rs1801252 βρίσκεται στο γονίδιο ADRB1. Ο συγκεκριμένος πολυμορφισμός αποτελείται από τρεις κατηγορίες γονοτύπων, τον A:A, τον A:G και τον G:G. Ο γονότυπος A:A είναι ο φυσικός (<https://www.snpedia.com>).
- Τέλος, ο σημειακός πολυμορφισμός rs1801253 βρίσκεται και αυτός στο γονίδιο ADRB1. Η ύπαρξη της βάσης γουανίνης (G) στο γονίδιο κωδικοποιεί τη γλυκίνη. Ο συγκεκριμένος πολυμορφισμός αποτελείται από τρεις κατηγορίες γονοτύπων, τον C:C, τον C:G και τον G:G. Ο γονότυπος G:G είναι ο φυσικός (<https://www.snpedia.com>).

5.2 Περιγραφική ανάλυση

Στην συγκεκριμένη ενότητα σκοπός μας είναι να προβούμε σε μία συνοπτική παρουσίαση του δείγματός μας, που θα έχει ως αποτέλεσμα την εξαγωγή κάποιων αρχικών συμπερασμάτων. Τα συμπεράσματα αυτά παρουσιάζονται σε μορφή πινάκων αλλά και διαγραμμάτων. Για παράδειγμα, για τις ποιοτικές ή κατηγορικές μεταβλητές παρουσιάζεται η κατανομή τους μέσω ενός πίνακα συχνοτήτων και μέσω ραβδογραμμάτων. Για τις συγκεκριμένες μεταβλητές, πέραν της παρουσίασης της

κατανομής τους, θα γίνουν και κάποιοι έλεγχοι χ^2 ώστε να δούμε αν υπάρχει κάποια σχέση μεταξύ μεταβλητών που πιστεύουμε ότι θα έχει ενδιαφέρον.

5.2.1 Φύλο των εθελοντών

Η πρώτη μεταβλητή που μελετάμε είναι το φύλο των εθελοντών, μια ποιοτική ονομαστική μεταβλητή με δύο επίπεδα. Σύμφωνα με τα δεδομένα, έχουμε ότι οι γυναίκες εθελόντριες, σε απόλυτο αριθμό και σε ποσοστιαία συμμετοχή, είναι περισσότερες από τους άνδρες. Πιο συγκεκριμένα, από τα 406 άτομα του δείγματός μας, τα 256 (σε ποσοστό 63,1%) είναι γυναίκες ενώ τα 150 (σε ποσοστό 36,9%) είναι άνδρες (Πίνακας 5.1).

Πίνακας 5.1

Πίνακας Συχνοτήτων του φύλου των εθελοντών

		Frequency	Valid Percent	Cumulative Percent
Valid	Male	150	36,9	36,9
	Female	256	63,1	100,0
	Total	406	100,0	

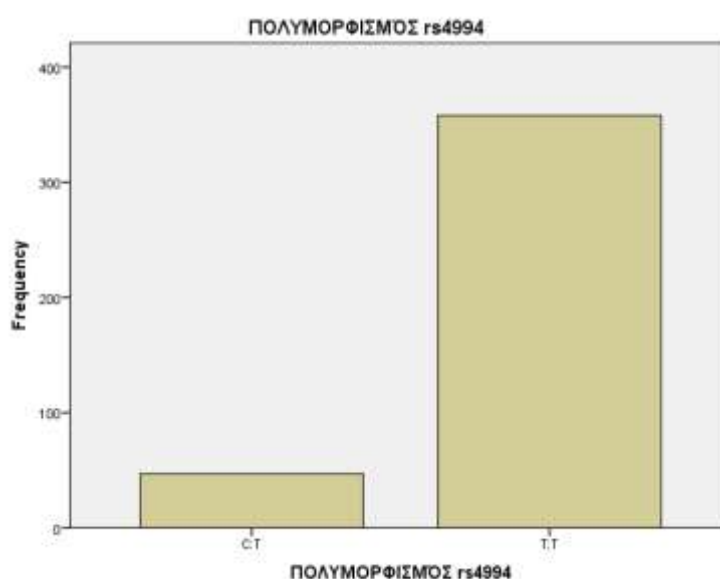
5.2.2 Πολυμορφισμός ADRB3rs4994

Επόμενη και σημαντική μεταβλητή που μελετάμε είναι ο πολυμορφισμός rs4994, μια ποιοτική ονομαστική μεταβλητή με δύο επίπεδα το C:T και το T:T. Σύμφωνα με τα δεδομένα έχουμε ότι η πλειοψηφία των ατόμων έχει τον φυσικό ομόζυγο γονότυπο T:T (Πίνακας 5.2). Να αναφέρουμε ότι έχουμε μία ελλείπουσα παρατήρηση διότι ο C:C γονότυπος εμφανίζεται μόνο μία φορά και γι' αυτό δεν τον θέσαμε ως ξεχωριστή κατηγορία. Η διαφορά στους δύο γονότυπους είναι εμφανής και από το παρακάτω ραβδόγραμμα (Διάγραμμα 5.1).

Πίνακας 5.2

Κατανομή του πολυμορφισμού rs4994

		Frequency	Valid Percent	Cumulative Percent
Valid	C:T	47	11,6	11,6
	T:T	358	88,4	100,0
	Total	405	100,0	
Missing	System	1		
Total		406		



Διάγραμμα 5.1

Ραβδόγραμμα της ADRB3rs4994

5.2.3 Πολυμορφισμοί ADRB2rs1042713 και ADRB2rs1042714

Επόμενες σημαντικές μεταβλητές που μελετάμε είναι οι πολυμορφισμοί rs1042713 και rs1042714, ποιοτικές ονομαστικές μεταβλητές με τρία επίπεδα η καθεμία. Για τον πρώτο πολυμορφισμό έχουμε ότι η πλειοψηφία των ατόμων του δείγματος έχει τον ετερόζυγο γονότυπο G:A ενώ για τον rs1042714 το 46,3% των ατόμων έχει τον ετερόζυγο γονότυπο G:C (Πίνακας 5.3).

Πίνακας 5.3

Κατανομή των πολυμορφισμών rs1042713 και rs1042714

		Count	Column N %
ΠΟΛΥΜΟΡΦΙΣΜΟΣ rs1042713	G:G	145	35,7%
	G:A	199	49,0%
	A:A	62	15,3%
ΠΟΛΥΜΟΡΦΙΣΜΟΣ rs1042714	G:G	52	12,8%
	G:C	188	46,3%
	C:C	166	40,9%

5.2.4 Πολυμορφισμοί ADRB1rs1801252 και ADRB1rs1801253

Οι τελευταίες μεταβλητές πολυμορφισμών που μελετάμε είναι οι rs1801252 και rs1801253, ποιοτικές ονομαστικές μεταβλητές με τρία επίπεδα η καθεμία. Για τον πρώτο πολυμορφισμό έχουμε ότι η πλειοψηφία των ατόμων του δείγματος έχει τον φυσικό ομόζυγο γονότυπο A:A ενώ για τον rs1801253 το 47,8% των ατόμων έχει τον ομόζυγο γονότυπο C:C (Πίνακας 5.4).

Πίνακας 5.4

Κατανομή των πολυμορφισμών rs1801252 και rs1801253

		Count	Column N %
ΠΟΛΥΜΟΡΦΙΣΜΟΣ rs1801252	A:A	332	81,8%
	G:A	69	17,0%
	G:G	5	1,2%
ΠΟΛΥΜΟΡΦΙΣΜΟΣ rs1801253	G:C	178	43,8%
	C:C	194	47,8%
	G:G	34	8,4%

5.2.5 Ο δείκτης μάζας σώματος (BMI)

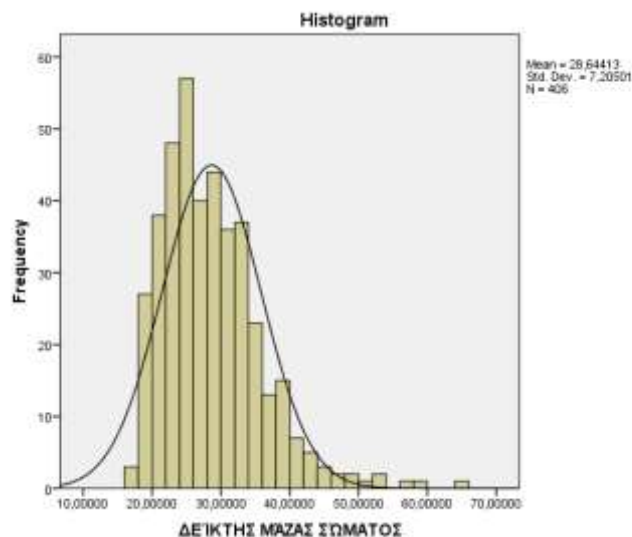
Μία σημαντική μεταβλητή του δείγματός μας είναι η συνεχής μεταβλητή με τις τιμές του δείκτη μάζας σώματος για κάθε ένα άτομο του δείγματος.

Πίνακας 5.5

Περιγραφικά μέτρα για τον δείκτη μάζας σώματος των εθελοντών

Mean	28,644
Median	27,355
Mode	24,489
Skewness	1,248
Std. Error of Skewness	,121
Kurtosis	2,725
Std. Error of Kurtosis	,242
Minimum	16,880
Maximum	65,547

Ο παραπάνω πίνακας παρουσιάζει ορισμένα περιγραφικά μέτρα για την συνεχή μεταβλητή του BMI. Παρατηρούμε ότι ο μέσος δείκτης μάζας σώματος για το δείγμα μας ισούται με 28,6 που σημαίνει ότι ο μέσος άνθρωπος του δείγματος, σύμφωνα με το BMI του, κατατάσσεται στην κατηγορία των υπέρβαρων. Η ελάχιστη και η μέγιστη τιμή του δείκτη μάζας σώματος είναι 16,9 και 65,54 αντίστοιχα. Δηλαδή, όπως θα δούμε και στην παρακάτω ενότητα, το δείγμα μας καλύπτει το φάσμα όλων των κατηγοριών. Ακόμη η επικρατούσα τιμή, δηλαδή η παρατήρηση με την μεγαλύτερη συχνότητα, ισούται με 24,5. Κοιτάζοντας τον συντελεστή ασυμμετρίας (Skewness) και κύρτωσης (Kurtosis) παρατηρούμε ότι η κατανομή που δημιουργείται, βλέπε και διάγραμμα 5.2, είναι πλατύκυρτη και παρουσιάζει θετική ασυμμετρία (Χαλικιάς κ.α., 2015).



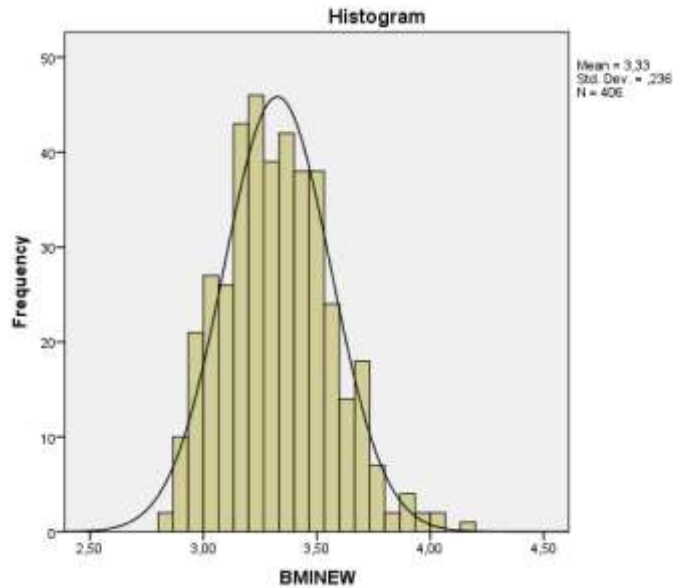
Διάγραμμα 5.2

Ιστόγραμμα συχνοτήτων

Από το παραπάνω ιστόγραμμα συχνοτήτων για τον δείκτη μάζας σώματος φαίνεται ότι η κατανομή αποκλίνει από την κανονική. Τα χαρακτηριστικά της κατανομής παρουσιάστηκαν παραπάνω με την ερμηνεία των περιγραφικών μέτρων (Πίνακας 5.5) και μέσω του ιστογράμματος επιβεβαιώνονται. Παρατηρούμε μάλιστα ότι η κατανομή των δεδομένων είναι πολυκόρυφη.

5.2.6 Μετασχηματισμός του δείκτη μάζας σώματος

Επόμενο βήμα που σκεφτήκαμε ήταν να μετασχηματίσουμε την συνεχή μεταβλητή του δείκτη μάζας σώματος. Γι' αυτό λοιπόν δημιουργήσαμε μια καινούργια μεταβλητή την οποία την ονομάσαμε BMINEW και ισούται με το λογάριθμο της αρχικής μεταβλητής του BMI. Σκοπός αυτού του μετασχηματισμού ήταν να δούμε εάν η κατανομή που θα δημιουργηθεί προσεγγίζει την κανονική.



Διάγραμμα 5.3

Ιστόγραμμα της μεταβλητής BMINEW

Από το παραπάνω διάγραμμα παρατηρούμε ότι η κατανομή, σε σχέση με την προηγούμενη, έχει μετατοπιστεί προς τα δεξιά προσεγγίζοντας αρκετά την κανονική κατανομή (Διάγραμμα 5.3).

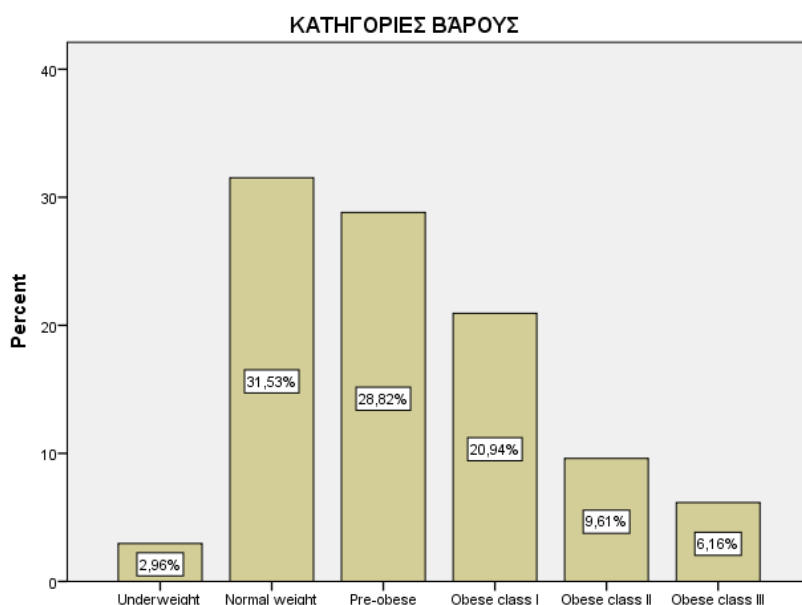
5.2.7 Η κατηγοριοποίηση του BMI (ClassificationBMI)

Μια σημαντική μεταβλητή είναι η classification bmi, μια ποιοτική διατακτική μεταβλητή με έξι κατηγορίες, οι οποίες παρουσιάστηκαν αναλυτικά στο 1^ο κεφάλαιο. Η συγκεκριμένη μεταβλητή είναι κωδικοποίηση του δείκτη μάζας σώματος προκειμένου να κατατάξουμε τα άτομα του δείγματός μας στις ακόλουθες κατηγορίες. Σύμφωνα με τα δεδομένα παρατηρούμε ότι η πλειοψηφία των ατόμων έχει καταταχθεί στην κατηγορία των φυσιολογικών και στην συνέχεια το 28,8% των ατόμων χαρακτηρίζονται ως υπέρβαρα ενώ μόνο το 3% ως ελλειποβαρή (Πίνακας 5.6). Οι πληροφορίες που παίρνουμε για την μεταβλητή αυτή παρουσιάζονται και στο αντίστοιχο ραβδόγραμμα (Διάγραμμα 5.4).

Πίνακας 5.6

Κατανομή της classification bmi

		Frequency	Valid Percent	Cumulative Percent
Valid	Underweight	12	3,0	3,0
	Normal weight	128	31,5	34,5
	Pre-obese	117	28,8	63,3
	Obese class I	85	20,9	84,2
	Obese class II	39	9,6	93,8
	Obese class III	25	6,2	100,0
	Total	406	100,0	



Διάγραμμα 5.4

Ραβδόγραμμα της classificationBMI

5.2.8 Σχέσεις μεταξύ της classificationBMI, των 5 πολυμορφισμών και του φύλου

Ύστερα από μια περιγραφική ανάλυση ορισμένων μεταβλητών είναι ενδιαφέρον να δούμε τι σχέσεις δημιουργούνται ανάμεσα στις παραπάνω ποιοτικές μεταβλητές. Αρχικά εξετάζουμε τις σχέσεις της κατάστασης βάρους και των 5 πολυμορφισμών του δείγματός μας και στην συνέχεια τη σχέση της κατάστασης βάρους με το φύλο.

5.2.8.1 Σχέση μεταξύ classificationBMI και ADRB3rs4994

Ο έλεγχος για τις δύο ποιοτικές μεταβλητές θα γίνει με το chi-square τεστ με το οποίο θα ελέγξουμε την ανεξαρτησία των προαναφερθέντων μεταβλητών. Η μεταβλητή της classificationBMI έχει 6 επίπεδα ενώ η μεταβλητή του πολυμορφισμού rs4994 έχει 2 επίπεδα. Συνεπώς έχουμε έναν 6×2 πίνακα συνάφειας (βλ. πίνακα παραρτήματος Β.1).

Από τον πίνακα συνάφειας μπορούμε να πούμε ότι το 89,4% των παχύσαρκων κατηγορίας I έχει τον φυσικό ομόζυγο γονότυπο T:T και το 91,7% των ελλειποβαρών έχουν τον φυσικό ομόζυγο γονότυπο T:T. Γενικά παρατηρούμε ότι οι περισσότεροι από τα 406 άτομα του δείγματός μας, ανεξαρτήτως κατάστασης βάρους, εμφανίζουν το φυσικό ομόζυγο γονότυπο T:T (σε ποσοστό 88,4%).

Προκειμένου να ελέγξουμε αν υπάρχει σχέση ανάμεσα στο βάρος των ατόμων του δείγματός μας και του πολυμορφισμού rs4994 εφαρμόσαμε τον έλεγχο ανεξαρτησίας χ^2 (Chi-Square Test). Σύμφωνα με τα αποτελέσματα του ελέγχου φαίνεται ότι δεν απορρίπτεται η υπόθεση της ανεξαρτησίας ($p\text{-value} > 0,10$). Άρα συμπεραίνουμε ότι δεν φαίνεται να υπάρχει μια στατιστικά σημαντική συσχέτιση μεταξύ της κατάστασης βάρους και του πολυμορφισμού rs4994.

5.2.8.2 Σχέση μεταξύ classificationBMI και ADRB2rs1042713

Ο έλεγχος ανεξαρτησίας θα πραγματοποιηθεί και εδώ με το chi-square τεστ. Όπως έχει ειπωθεί προηγουμένως, η μεταβλητή της classificationBMI έχει 6 επίπεδα ενώ η μεταβλητή του πολυμορφισμού rs1042713 έχει 3 επίπεδα. Συνεπώς έχουμε έναν 6×3 πίνακα συνάφειας (βλ. πίνακα παραρτήματος Β.2).

Παρατηρούμε ότι το 52% των ατόμων που έχουν καταταχθεί στην παχυσαρκία κατηγορίας III (ή 3^{ου} βαθμού) έχουν το ετερόζυγο γονότυπο G:A και το 28% αυτών έχουν το φυσικό ομόζυγο γονότυπο G:G. Ακόμη το 50,8% των ατόμων που έχουν καταταχθεί ως φυσιολογικοί έχουν το ετερόζυγο γονότυπο G:A και το 37,5% αυτών έχουν το φυσικό ομόζυγο γονότυπο G:G. Ακόμη μπορούμε να πούμε ότι το 48,7% των ατόμων που έχουν καταταχθεί στους υπέρβαρους έχουν το ετερόζυγο γονότυπο G:A και το 32,5% αυτών έχουν το φυσικό ομόζυγο γονότυπο G:G. Γενικά παρατηρούμε ότι οι περισσότεροι από τα 406 άτομα του δείγματός μας, ανεξαρτήτως κατάστασης βάρους, εμφανίζουν το ετερόζυγο γονότυπο G:A (σε ποσοστό 49%).

Αυτό που μας ενδιαφέρει και στην συγκεκριμένη περίπτωση είναι να δούμε αν υπάρχει σχέση ανάμεσα στο βάρος των ατόμων του δείγματός μας και του πολυμορφισμού rs1042713. Εφαρμόζοντας τον έλεγχο ανεξαρτησίας καταλήγουμε στο συμπέρασμα ότι δεν φαίνεται να υπάρχει μια στατιστικά σημαντική συσχέτιση μεταξύ της κατάστασης βάρους και του πολυμορφισμού rs1042713 ($p\text{-value}>0,05$).

5.2.8.3 Σχέση μεταξύ classificationBMI και ADRB2rs1042714

Όπως έχει προαναφερθεί, η μεταβλητή της classificationBMI έχει 6 επίπεδα ενώ η μεταβλητή του πολυμορφισμού rs1042714 έχει 3 επίπεδα. Συνεπώς και εδώ έχουμε έναν 6×3 πίνακα συνάφειας (βλ. πίνακα παραρτήματος Β.3).

Παρατηρούμε ότι το 48% των ατόμων που έχουν καταταχθεί στην παχυσαρκία κατηγορίας III (ή 3^ο βαθμού ή νοσογόνο παχυσαρκία) έχουν το ετερόζυγο γονότυπο G:C και το 36% αυτών έχουν το φυσικό ομόζυγο γονότυπο C:C. Ακόμη το 48,7% των ατόμων που έχουν καταταχθεί στην παχυσαρκία κατηγορίας II (ή 2^ο βαθμού) έχουν το φυσικό ομόζυγο γονότυπο C:C. Γενικά παρατηρούμε ότι οι περισσότεροι από τα 406 άτομα του δείγματός μας, ανεξαρτήτως κατάστασης βάρους, εμφανίζουν το ετερόζυγο γονότυπο G:C (σε ποσοστό 46,3%).

Εφαρμόζοντας για μια ακόμη φορά τον έλεγχο ανεξαρτησίας φαίνεται ότι πάλι δεν απορρίπτεται η υπόθεση της ανεξαρτησίας ($p\text{-value}>0,10$). Άρα συμπεραίνουμε ότι ο πολυμορφισμός rs1042714 του γονιδίου ADRB2 και το βάρος των ατόμων είναι ανεξάρτητα μεταξύ τους.

5.2.8.4 Σχέση μεταξύ classificationBMI και ADRB1rs1801252

Παρατηρούμε ότι το 21,6% των ατόμων που έχουν καταταχθεί ως υπέρβαροι έχουν το ετερόζυγο γονότυπο G:A και το 78,4% αυτών έχουν το φυσικό ομόζυγο γονότυπο A:A. Ακόμη το 89,7% των ατόμων που έχουν καταταχθεί σε αυτούς με παχυσαρκία κατηγορίας II (ή 2^ο βαθμού) έχουν το φυσικό ομόζυγο γονότυπο A:A. Γενικά παρατηρούμε ότι οι περισσότεροι από τα 406 άτομα του δείγματός μας, ανεξαρτήτως κατάστασης βάρους, εμφανίζουν το φυσικό ομόζυγο γονότυπο A:A (σε ποσοστό 82,8%) (βλ. πίνακα παραρτήματος Β.4).

Ο έλεγχος ανεξαρτησίας χ^2 χρησιμοποιείται για να ελέγξουμε αν δύο ποιοτικές μεταβλητές παρουσιάζουν κάποια σχέση μεταξύ τους ή είναι ανεξάρτητες. Και στην συγκεκριμένη περίπτωση, εφόσον $p\text{-value}>0,10$, συμπεραίνουμε ότι ο πολυμορφισμός

rs1801252 του γονιδίου ADRB1 και το βάρος των ατόμων του δείγματός μας είναι ανεξάρτητες μεταβλητές.

5.2.8.5 Σχέση μεταξύ classificationBMI και ADRB1rs1801253

Για τον τελευταίο πολυμορφισμό του δείγματός μας παρατηρούμε ότι το 17,9% των ατόμων που έχουν καταταχθεί στην παχυσαρκία κατηγορίας II (ή 2^{ου} βαθμού) έχουν το φυσικό ομόζυγο γονότυπο G:G και το 43,6% αυτών έχουν το ετερόζυγο γονότυπο G:C. Ακόμη το 53% των ατόμων που έχουν καταταχθεί στους υπέρβαρους έχουν το ομόζυγο γονότυπο C:C (βλ. πίνακα παραρτήματος Β.5).

Η συγκεκριμένη είναι η μοναδική περίπτωση απόρριψης της υπόθεσης της ανεξαρτησίας ($p\text{-value} < 0,10$). Συνεπώς καταλήγουμε στο συμπέρασμα ότι φαίνεται να υπάρχει μια στατιστικά σημαντική συσχέτιση μεταξύ της κατάστασης βάρους και του πολυμορφισμού rs1801253. Αυτό σημαίνει, για παράδειγμα, ότι τα ελλειποβαρή άτομα εμφανίζουν κάθε κατηγορία γονότυπου του συγκεκριμένου πολυμορφισμού σε διαφορετικό ποσοστό από ότι οι παχύσαρκοι κατηγορίας I (βλ. πίνακα παραρτήματος Β.6).

5.2.8.6 Σχέση μεταξύ classificationBMI και φύλου

Στην συγκεκριμένη περίπτωση μας ενδιαφέρει εάν το φύλο των ατόμων του δείγματός μας επηρεάζει το βάρος τους. Με άλλα λόγια, οι άνδρες (male) κατατάσσονται ως φυσιολογικοί, υπέρβαροι, παχύσαρκοι κλπ. σε διαφορετικό ποσοστό από τις γυναίκες (female). Εάν ισχύει αυτό, τότε υπάρχει σχέση μεταξύ της classificationBMI και του φύλου. Η μεταβλητή της classificationBMI έχει 6 επίπεδα ενώ η μεταβλητή του φύλου έχει 2 επίπεδα. Συνεπώς δημιουργείται ένας 6×2 πίνακα συνάφειας (βλ. πίνακα παραρτήματος Β.7).

Παρατηρούμε ότι το 91,7% των ατόμων που έχουν καταταχθεί στην κατηγορία των ελλειποβαρών είναι γυναίκες και το 61,2% των παχύσαρκων κατηγορίας I είναι γυναίκες. Όσον αφορά τους άνδρες, το 48,7% των υπέρβαρων είναι άνδρες ενώ μόνο το 8,3% των ελλειποβαρών είναι άνδρες.

Σύμφωνα με τα αποτελέσματα του ελέγχου, απορρίπτεται η υπόθεση της ανεξαρτησίας ($p\text{-value} < 0,01$) και πράγματι καταλήγουμε στο συμπέρασμα ότι φύλο και κατάσταση βάρους σχετίζονται σημαντικά, δηλαδή οι άνδρες κατατάσσονται ως

φυσιολογικοί, υπέρβαροι, παχύσαρκοι κλπ. σε διαφορετικό ποσοστό από τις γυναίκες (βλ. πίνακα παραρτήματος Β.8).

ΚΕΦΑΛΑΙΟ 6

ΑΠΑΡΑΜΕΤΡΙΚΗ ΑΝΑΛΥΣΗ

Αφού πήραμε μια περιγραφική εικόνα για τις μεταβλητές του δείγματος που μελετάμε και αφού μελετήσαμε πιθανές σχέσεις μεταξύ ορισμένων μεταβλητών, τώρα θα περάσουμε στην μη παραμετρική (απαρμετρική) ανάλυση προκειμένου να εξετάσουμε αν ο δείκτης μάζας σώματος επηρεάζεται από το φύλο των εθελοντών καθώς και από τους πέντε πολυμορφισμούς που διαθέτουμε.

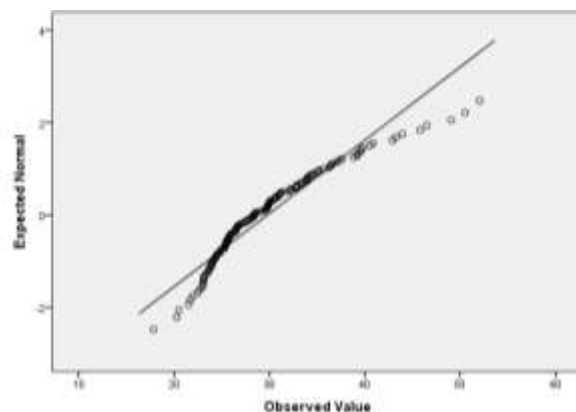
6.1 Σχέση του δείκτη μάζας σώματος με το φύλο των εθελοντών

Αυτό που μας ενδιαφέρει να μελετήσουμε είναι αν διαφέρει ο δείκτης μάζας σώματος στα δύο φύλα ή αν το φύλο των εθελοντών επηρεάζει τις τιμές του δείκτη μάζας σώματος. Αρχικά, πριν προβούμε σε κάποιον έλεγχο, θα ελέγξουμε αν τα δύο δείγματα ως προς το φύλο προέρχονται από την κανονική κατανομή. Πρώτα θα εφαρμόσουμε έλεγχο κανονικότητας στο δείγμα των ανδρών εθελοντών.

Πίνακας 6.1

Έλεγχος Shapiro-Wilk για το δείγμα των ανδρών

Δείγμα	Άνδρες
W	0.919
p-value	0.000



Διάγραμμα 6.1

Q-Q plot ανδρών

Όπως διαπιστώνεται, τόσο διαγραμματικά όσο και μέσω της διενέργειας του αντίστοιχου στατιστικού ελέγχου, το δείγμα των ανδρών εθελοντών δεν προέρχεται από την κανονική κατανομή. Σύμφωνα με τον παραπάνω έλεγχο, παρατηρούμε ότι υπάρχει πολύ ισχυρή στατιστική ένδειξη απόρριψης ότι το προαναφερθέν δείγμα προέρχεται από κανονική κατανομή ($p\text{-value} \cong 0.00$). Από το Q-Q διάγραμμα

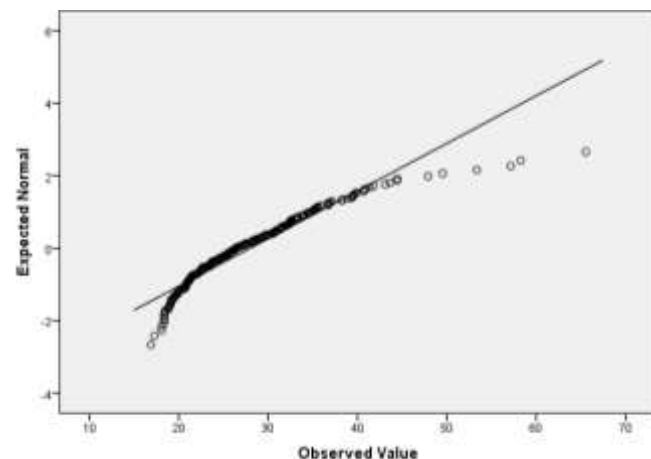
επαληθεύεται το παραπάνω συμπέρασμα. Αυτό συμβαίνει διότι τα σημεία στην πάνω άκρη παρουσιάζουν σημαντικές αποκλίσεις από την ευθεία γραμμή.

Στην συνέχεια, για τις γυναίκες εθελόντριες έχουμε τα εξής:

Πίνακας 6.2

Έλεγχος Shapiro-Wilk για το δείγμα των γυναικών

Δείγμα	Γυναίκες
W	0.917
p-value	0.000



Διάγραμμα 6.2

Q-Q plot γυναικών

Για το δείγμα των γυναικών εθελοντριών παρατηρούμε ότι πάλι υπάρχει εξαιρετικά ισχυρή στατιστική ένδειξη απόρριψης ότι το δείγμα αυτό κατανέμεται κανονικά ($p\text{-value} \cong 0.00$). Στο ίδιο συμπέρασμα καταλήγουμε αν «κοιτάξουμε» και το διάγραμμα πιθανότητας Q-Q (Διάγραμμα 6.2). Συγκεντρωτικά, λοιπόν, καταλήγουμε στο συμπέρασμα ότι τα δύο δείγματα ως προς το φύλο δεν ακολουθούν την κανονική κατανομή. Συνεπώς, προκειμένου να δούμε αν το φύλο των εθελοντών επηρεάζει τον δείκτη μάζα σώματος ή αν υπάρχει στατιστικά σημαντική διαφορά του δείκτη μάζας σώματος στα δύο φύλα, θα στραφούμε σε μη παραμετρικό έλεγχο και πιο συγκεκριμένα στον έλεγχο Mann-Whitney.

Πίνακας 6.3

Έλεγχος Mann-Whitney

W	15489
p-value	0.001

Άρα μπορούμε να απορρίψουμε την υπόθεση ότι δεν υπάρχει στατιστικά σημαντική διαφορά του δείκτη μάζας σώματος στα δύο φύλα ($p\text{-value} < 0.05$). Συνεπώς,

καταλήγουμε στο συμπέρασμα ότι το φύλο των εθελοντών επηρεάζει τον δείκτη μάζας σώματός τους.

Εφόσον βγάλαμε ότι υπάρχει στατιστικά σημαντική διαφορά του δείκτη μάζας σώματος στα δύο φύλα, θέλουμε να «εμβαθύνουμε» και να εξετάσουμε και πώς διαφέρουν. Δηλαδή ποιοι είναι αυτοί με το μεγαλύτερο δείκτη μάζας σώματος, οι άνδρες ή οι γυναίκες; Για να το εξετάσουμε αυτό θα ακολουθήσουμε τις εξής μεθόδους:

- Αρχικά, θα συγκρίνουμε τον μέσο δείκτη μάζας σώματος ανά φύλο.
- Ύστερα, θα κατασκευάσουμε ένα ιστόγραμμα για κάθε φύλο, με τον οριζόντιο άξονα να παρουσιάζει τις τιμές του δείκτη μάζας σώματος. Με αυτόν τον τρόπο θέλουμε να δούμε σε ποιο φύλο οι παχύσαρκοι και οι υπέρβαροι εμφανίζονται περισσότερο.

Σχετικά με τα δύο δείγματα ως προς το φύλο έχουμε:

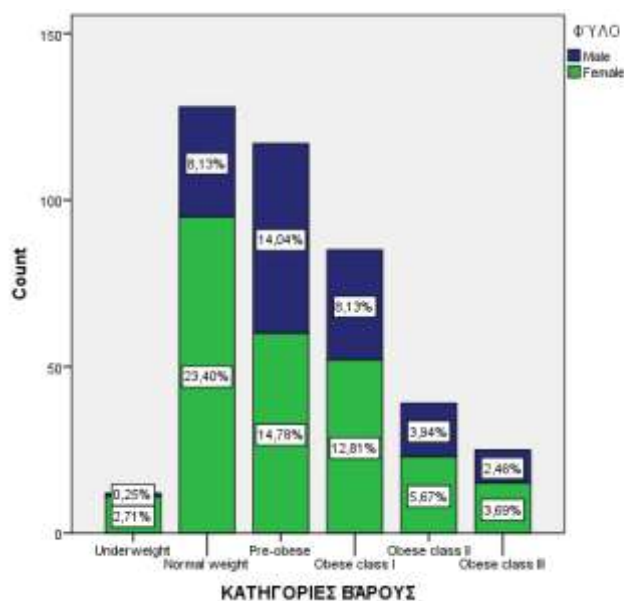
Πίνακας 6.4

Μέση τιμή για τον δείκτη μάζας σώματος ανά φύλο

Φύλο	Μέση τιμή
Άνδρες	29.79
Γυναίκες	27.97

Από τον παραπάνω πίνακα, παρατηρούμε ότι ο μέσος δείκτης μάζας σώματος για τους άνδρες ισούται με 29.79. Δηλαδή, σύμφωνα με τις κατατάξεις του BMI, ο μέσος άνδρας του δείγματός μας κατατάσσεται στην κατηγορία των υπέρβαρων. Στο ίδιο συμπέρασμα καταλήγουμε και για τις γυναίκες με μέσο δείκτη μάζας σώματος 27.97 (βλ. Πίνακα 1.1). Από την παραπάνω περιγραφική «ματιά» καταλήγουμε στο συμπέρασμα ότι το φύλο με το μεγαλύτερο δείκτη μάζας σώματος, για το δείγμα μας, είναι οι άνδρες.

Εν συνεχεία, κατασκευάζουμε το ιστόγραμμα το οποίο στον οριζόντιο άξονα παρουσιάζει τις τιμές, σε κλάσεις, του δείκτη μάζας σώματος και στον κάθετο άξονα τις συχνότητες. Το συγκεκριμένο διάγραμμα δημιουργήθηκε για να δούμε σε ποιο φύλο οι παχύσαρκοι και οι υπέρβαροι εμφανίζονται περισσότερο.



Διάγραμμα 6.3

Ιστόγραμμα του δείκτη μάζας σώματος ανά φύλο

Για τους άνδρες παρατηρούμε ότι το 14,04% κατατάσσονται στην κατηγορία των υπέρβαρων ενώ, για τις γυναίκες, το 14,78% κατατάσσονται στην συγκεκριμένη κατηγορία. Το 12,81% των γυναικών κατατάσσονται στην κατηγορία των παχύσαρκων τάξης I σε αντίθεση με τους άνδρες. Γενικότερα, και στα δύο φύλα, παρατηρείται μεγάλη συγκέντρωση στους «κανονικούς», στους υπέρβαρους και στους παχύσαρκους κατηγορίας I.

6.2 Σχέση του δείκτη μάζας σώματος με τους 5 πολυμορφισμούς

6.2.1 Σχέση του δείκτη μάζας σώματος με τον ADRB3rs4994

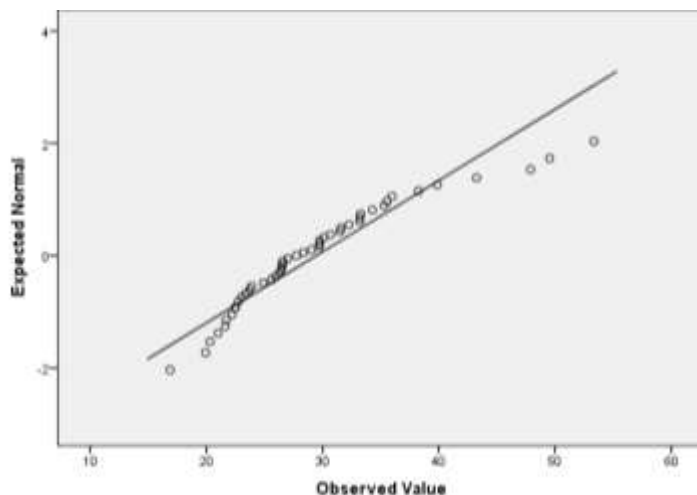
Στην συγκεκριμένη περίπτωση αυτό που καλούμαστε να εξετάσουμε είναι αν υπάρχει στατιστικά σημαντική διαφορά του BMI ανάμεσα στους δύο γονότυπους του συγκεκριμένου πολυμορφισμού. Πριν απαντήσουμε στο ερώτημα αυτό ας εξετάσουμε πρώτα την κανονικότητα στα δύο επίπεδα της μεταβλητής ADRB3rs4994. Αρχικά, για το πρώτο επίπεδό της, τον γονότυπο C:T έχουμε τα ακόλουθα:

Πίνακας 6.5

Έλεγχος Shapiro-Wilk για τον γονότυπο

C:T

Δείγμα	Γονότυπος
	C:T
W	0.916
p-value	0.003



Διάγραμμα 6.4

Q-Q plot C:T

Εφαρμόζοντας τον παραπάνω έλεγχο κανονικότητας για το δείγμα του C:T γονοτύπου καταλήγουμε στο συμπέρασμα ότι το συγκεκριμένο δείγμα δεν κατανέμεται κανονικά ($p\text{-value} < 0.05$). Όσον αφορά το γράφημα Q-Q, παρατηρούμε ότι οι παρατηρήσεις στην πάνω άκρη του δεν «αγκαλιάζουν» την ευθεία γραμμή (Διάγραμμα 6.4).

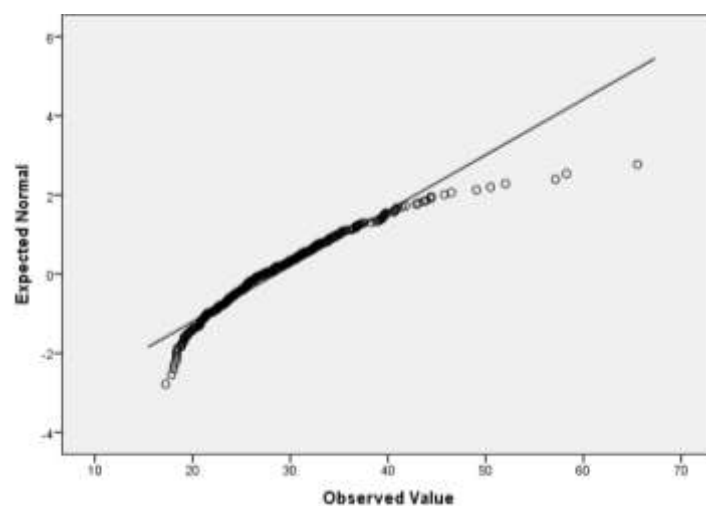
Ύστερα, για τον γονότυπο T:T προκύπτουν τα εξής:

Πίνακας 6.6

Έλεγχος Shapiro-Wilk για τον γονότυπο

T:T

Δείγμα	Γονότυπος
	T:T
W	0.935
p-value	0.000



Διάγραμμα 6.5

Q-Q plot T:T

Για την δεύτερη κατηγορία γονότυπου του πολυμορφισμού ADRB3rs4994 παρατηρούμε ότι απορρίπτεται η υπόθεση της κανονικότητας του συγκεκριμένου

δείγματος ($p\text{-value} \cong 0.00$). Άρα, για να αποφασίσουμε αν τα επίπεδα (γονότυποι) της κατηγορικής μεταβλητής ADRB3rs4994 επηρεάζουν τις τιμές του δείκτη μάζας σώματος, θα προβούμε σε μη παραμετρικό έλεγχο και συγκεκριμένα τον έλεγχο Mann-Whitney.

Πίνακας 6.7

Έλεγχος Mann-Whitney

W	7661
p-value	0.442

Παρατηρούμε ότι πράγματι δεν απορρίπτουμε την υπόθεση περί μη στατιστικά σημαντικής διαφοράς του δείκτη μάζας σώματος στις δύο κατηγορίες γονοτύπων του πολυμορφισμού ADRB3rs4994 ($p\text{-value} > 0.05$).

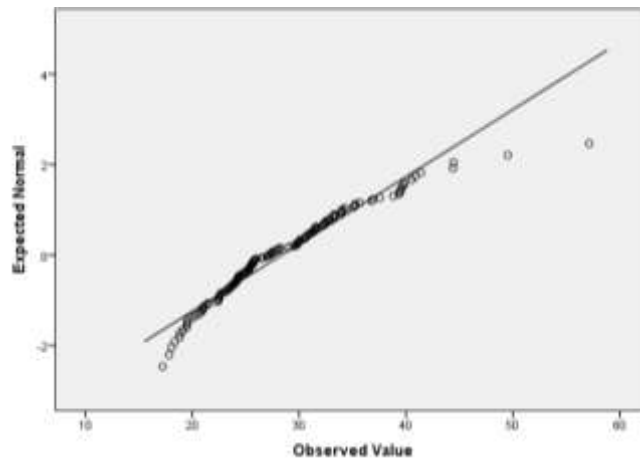
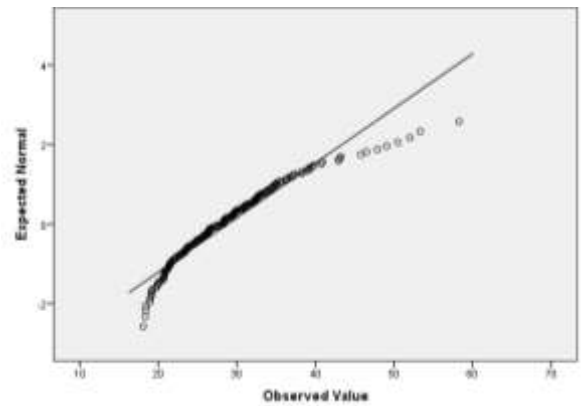
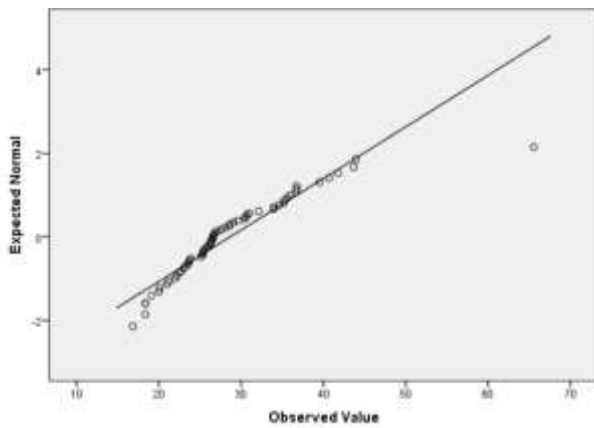
6.2.2 Σχέση του δείκτη μάζας σώματος με τον ADRB2rs1042713

Αυτό που εξετάζουμε είναι αν υπάρχει σημαντική διαφορά του δείκτη μάζας σώματος στις κατηγορίες γονοτύπων του παραπάνω πολυμορφισμού. Ο συγκεκριμένος πολυμορφισμός αποτελείται από τρεις κατηγορίες γονοτύπων. Για κάθε μία από αυτές θα προβούμε σε έλεγχο κανονικότητας. Αυτούς τους ελέγχους θα τους παρουσιάσουμε συγκεντρωτικά σε έναν πίνακα. Για τα τρία δείγματά μας θα κατασκευάσουμε και τα γραφήματα Q-Q.

Πίνακας 6.8

Έλεγχος Shapiro-Wilk για τα επίπεδα της μεταβλητής ADRB2rs1042713

Δείγματα	Γονότυπος A:A	Γονότυπος G:A	Γονότυπος G:G
W	0.875	0.926	0.943
p-value	0.000	0.000	0.000



Διάγραμμα 6.6

Q-Q plot A:A, G:A, G:G

Παρατηρώντας τον παραπάνω συγκεντρωτικό πίνακα παρατηρούμε ότι κανένα δείγμα των τριών γονοτύπων του πολυμορφισμού που εξετάζουμε δεν προέρχεται από την κανονική κατανομή. Στο ίδιο συμπέρασμα καταλήγουμε αν κοιτάξουμε και τα αντίστοιχα διαγράμματα Q-Q. Το πρώτο διάγραμμα αφορά το δείγμα του γονοτύπου A:A, το δεύτερο το δείγμα του G:A και το τρίτο το δείγμα του G:G (Διάγραμμα 6.6). Ίσως θα μπορούσαμε να πούμε ότι, από το πρώτο γράφημα, φαίνεται ότι κατά πλειοψηφία οι παρατηρήσεις «περιτριγυρίζουν» την ευθεία γραμμή. Επίσης και στα άλλα δύο γραφήματα, οι παρατηρήσεις «περιτριγυρίζουν» την ευθεία.

Εφόσον οι υποθέσεις περί κανονικότητας για τα τρία δείγματα δεν επαληθεύτηκαν, καλούμαστε να εφαρμόσουμε έναν μη παραμετρικό έλεγχο, τον έλεγχο Kruskal-Wallis.

Πίνακας 6.9

Έλεγχος Kruskal-Wallis

H	0.154
df	2
p-value	0.926

Άρα δεν μπορούμε να απορρίψουμε την υπόθεση ότι δεν υπάρχει στατιστικά σημαντική διαφορά του δείκτη μάζας σώματος στις τρεις κατηγορίες γονοτύπων ($p\text{-value} > 0.05$). Με λίγα λόγια, οι τρεις κατηγορίες αυτές, A:A, G:A, G:G, δεν διαφέρουν ως προς τις τιμές του BMI.

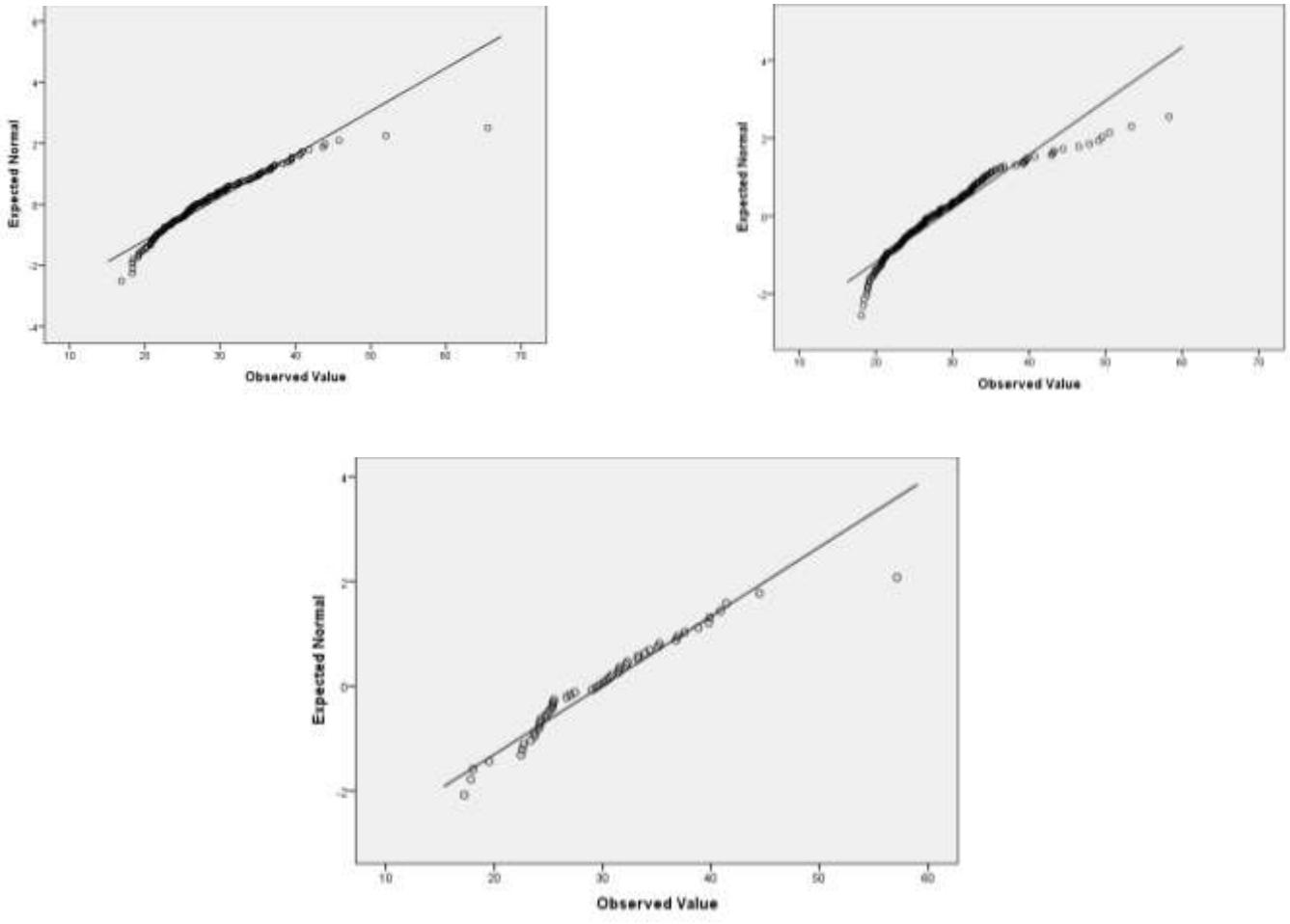
6.2.3 Σχέση του δείκτη μάζας σώματος με τον ADRB2rs1042714

Όπως έχουμε αναφέρει, αυτό που μας ενδιαφέρει είναι ποιες κατηγορίες γονοτύπων επηρεάζουν τον δείκτη μάζας σώματος. Γι' αυτό το λόγο θα εξετάσουμε, μέσω του κατάλληλου ελέγχου, τους γονότυπους του πολυμορφισμού ADRB2rs1042714. Ο συγκεκριμένος αποτελείται από τρεις κατηγορίες γονοτύπων, οπότε για κάθε έναν από αυτούς θα εξετάσουμε την κανονικότητα πριν αποφασίσουμε ποιον έλεγχο θα χρησιμοποιήσουμε.

Πίνακας 6.10

Έλεγχος Shapiro-Wilk για τις κατηγορίες της μεταβλητής ADRB2rs1042714

Δείγματα	Γονότυπος C:C	Γονότυπος G:C	Γονότυπος G:G
W	0.916	0.919	0.943
p-value	0.000	0.000	0.014



Διάγραμμα 6.7

Q-Q plot C:C, G:C, G:G

Και εδώ βλέπουμε ότι υπάρχουν στατιστικές ενδείξεις απόρριψης της κανονικότητας για τα τρία δείγματα γονοτύπων ($p\text{-value} < 0.05$). Συνεπώς, για να δούμε αν ο δείκτης μάζας σώματος διαφέρει στις τρεις κατηγορίες γονοτύπων του πολυμορφισμού που εξετάζουμε, θα εφαρμόσουμε μη παραμετρικό έλεγχο, τον έλεγχο των Kruskal-Wallis.

Πίνακας 6.11

Έλεγχος Kruskal-Wallis

H	2.078
df	2
p-value	0.354

Από τον παραπάνω πίνακα παρατηρούμε ότι η τιμή της στατιστικής συνάρτησης H των Kruskal-Wallis ισούται με 2.0781. Από το p -value του ελέγχου καταλήγουμε στο συμπέρασμα ότι πράγματι ο δείκτης μάζας σώματος δεν διαφέρει στις τρεις κατηγορίες γονοτύπων (p -value>0.05).

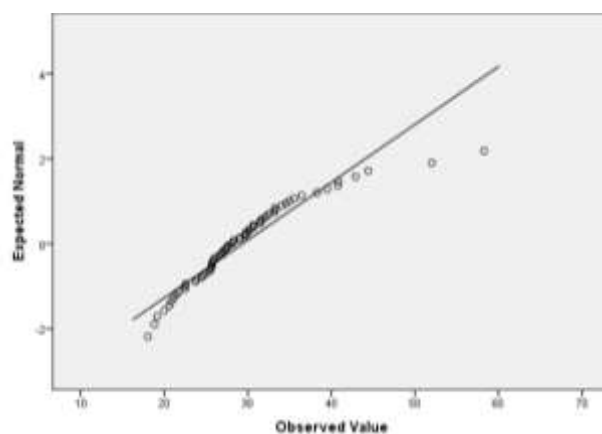
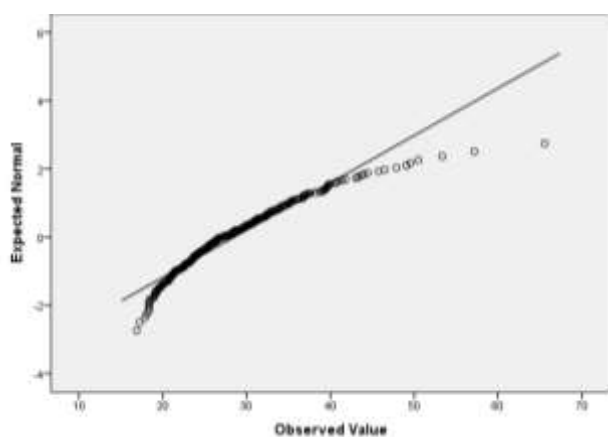
6.2.4 Σχέση του δείκτη μάζας σώματος με τον ADRB1rs1801252

Ο συγκεκριμένος πολυμορφισμός είναι ο προτελευταίος του δείγματος που μελετάμε και αποτελείται από τρεις κατηγορίες γονοτύπων. Αρχικά εξετάζουμε αν τα δείγματα των γονοτύπων ακολουθούν την κανονική κατανομή. Όπως είδαμε στο κεφάλαιο 5, μόνο 5 άτομα του δείγματός μας έχουν τον γονότυπο G:G. Η συχνότητα εμφάνισής του είναι πολύ μικρή οπότε, για τον συγκεκριμένο, δεν έχει νόημα να κάνουμε έλεγχο κανονικότητας.

Πίνακας 6.12

Έλεγχος Shapiro-Wilk για τα επίπεδα της μεταβλητής ADRB1rs1801252

Δείγματα	Γονότυπος A:A	Γονότυπος G:A
W	0.930	0.903
p-value	0.000	0.000



Διάγραμμα 6.8

Q-Q plot A:A, G:A

Ο παραπάνω πίνακας παρουσιάζει τους ελέγχους κανονικότητας για τα δείγματα των δύο γονοτύπων. Κοιτάζοντας τα p -value, συμπεραίνουμε ότι κανένα από αυτά τα

δείγματα δεν κατανέμεται κανονικά ($p\text{-value}<0.05$). Εφόσον δεν ισχύει η υπόθεση της κανονικότητας για τα δείγματα αυτά, θα εφαρμόσουμε μη παραμετρικό έλεγχο και πιο συγκεκριμένα τον έλεγχο των Kruskal-Wallis.

Πίνακας 6.13

Έλεγχος Kruskal-Wallis

H	2.387
df	2
p-value	0.303

Άρα δεν μπορούμε να απορρίψουμε την υπόθεση περί μη στατιστικά σημαντικής διαφοράς του δείκτη μάζας σώματος στις τρεις κατηγορίες γονοτύπων ($p\text{-value}>0.05$). Δηλαδή, ο πολυμορφισμός ADRB1rs1801252 δεν επηρεάζει τις τιμές του δείκτη μάζας σώματος.

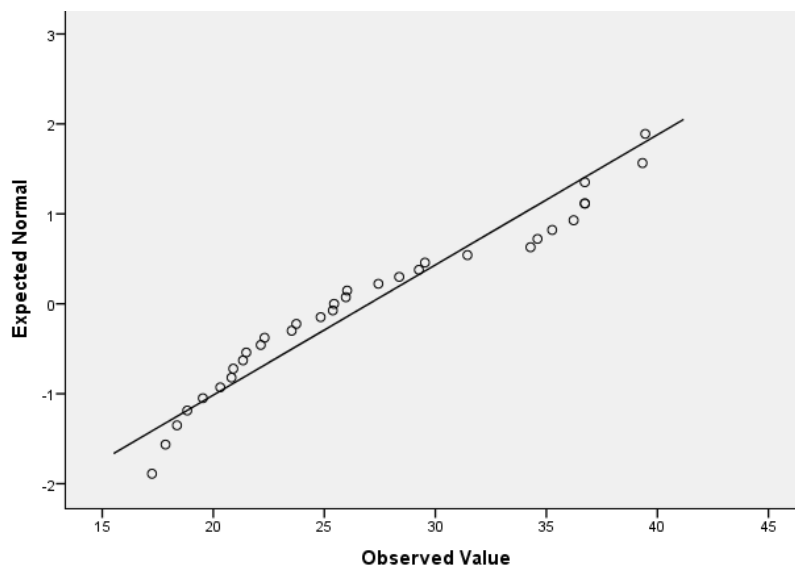
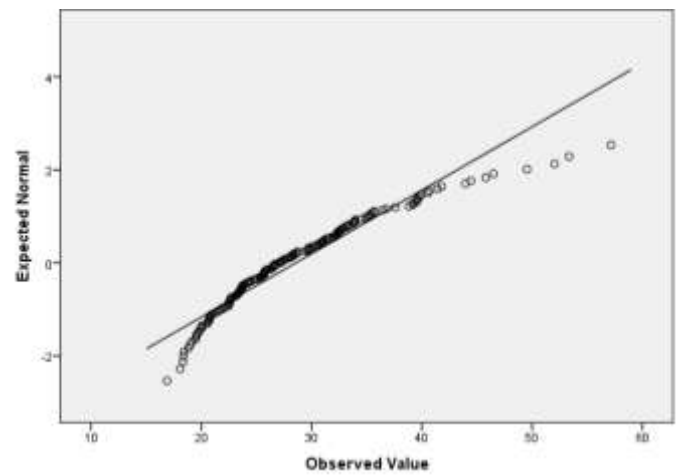
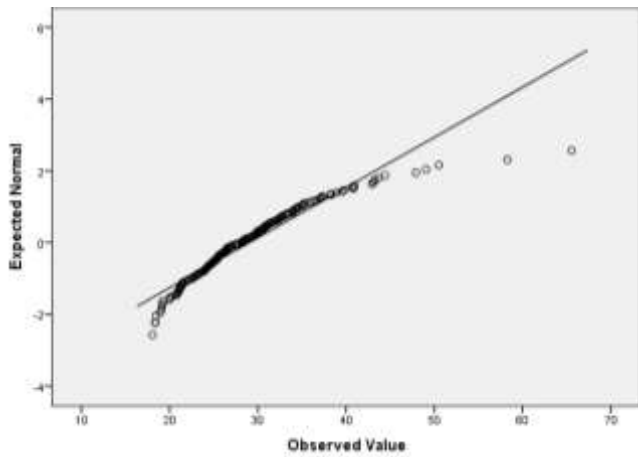
6.2.5 Σχέση του δείκτη μάζας σώματος με τον ADRB1rs1801253

Δουλεύοντας όπως και στις προηγούμενες ενότητες, θα εφαρμόσουμε έλεγχο κανονικότητας στις τρεις κατηγορίες του συγκεκριμένου πολυμορφισμού προκειμένου να καταλήξουμε στον κατάλληλο έλεγχο.

Πίνακας 6.14

Έλεγχος Shapiro-Wilk για τις κατηγορίες της μεταβλητής ADRB1rs1801253

Δείγματα	Γονότυπος C:C	Γονότυπος G:C	Γονότυπος G:G
W	0.907	0.927	0.923
p-value	0.000	0.000	0.020



Διάγραμμα 6.9

Q-Q plot C:C, G:C, G:G

Κοιτώντας τον παραπάνω πίνακα, παρατηρούμε ότι και εδώ δεν μπορούμε να πούμε ότι τα δείγματά μας ακολουθούν την κανονική κατανομή ($p\text{-value} < 0.05$). Άρα θα χρησιμοποιήσουμε μη παραμετρικό έλεγχο, τον έλεγχο Kruskal-Wallis, προκειμένου να εξετάσουμε αν οι τρεις συγκεκριμένες κατηγορίες γονοτύπων επηρεάζουν τις τιμές του BMI.

Πίνακας 6.15

Έλεγχος Kruskal-Wallis

H	2.898
df	2
p-value	0.235

Για ακόμα μια φορά δεν μπορούμε να απορρίψουμε την υπόθεση της μη στατιστικά σημαντικής διαφοράς του δείκτη μάζας σώματος στους τρεις γονότυπους. Συνεπώς, οι τρεις γονότυποι δεν επηρεάζουν τις τιμές του BMI.

ΚΕΦΑΛΑΙΟ 7

ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Σκοπός αυτού του κεφαλαίου είναι να αναπτυχθούν και να παρουσιαστούν εκτενώς τρία μοντέλα λογιστικής παλινδρόμησης με διαφορετικά cutoff points για τις εξαρτημένες μεταβλητές. Να μελετήσουμε λίγο «βαθύτερα» τις επεξηγηματικές μεταβλητές σε σχέση με τον δείκτη μάζας σώματος.

7.1 Περιγραφή και προσαρμογή του πρώτου μοντέλου

Στην παρούσα ενότητα θα εφαρμόσουμε ένα μοντέλο λογιστικής παλινδρόμησης (logit) με δίτιμη εξαρτημένη μεταβλητή ή μεταβλητή απόκρισης. Η συγκεκριμένη μεταβλητή που χρησιμοποιήσαμε στο πρώτο μοντέλο ονομάζεται `weightsituation1`, μια ποιοτική μεταβλητή με δύο επίπεδα. “Κατασκευάστηκε” συγχωνεύοντας κάποιες κατηγορίες της `classificationBMI` (την παρουσιάσαμε στο κεφάλαιο 5). Θέσαμε, σαν πρώτη κατηγορία, όσους έχουν $BMI \geq 25$ δηλαδή τους υπέρβαρους και παχύσαρκους ενώ, σαν δεύτερη κατηγορία, όσους έχουν $BMI < 25$ ως αυτούς που το βάρος τους θεωρείται κανονικό. Σαν ανεξάρτητες μεταβλητές χρησιμοποιήθηκαν οι πέντε πολυμορφισμοί του δείγματός μας καθώς και η ηλικία και το φύλο των εθελοντών. Το συγκεκριμένο μοντέλο λογιστικής παλινδρόμησης, όπως γνωρίζουμε από την θεωρία, γράφεται ως εξής :

$$\log\left(\frac{\pi}{1-\pi}\right) \equiv \text{logit}(\pi) = x'\beta \equiv \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7,$$

όπου $x = (1, X_1, X_2, \dots, X_7)'$ το διάνυσμα των επεξηγηματικών μεταβλητών. Η πιθανότητα παρουσίας ενός χαρακτηριστικού συμβολίζεται με π . Επομένως, το logit της πιθανότητας παρουσίας ενός χαρακτηριστικού ισούται με το λογάριθμο της σχετικής πιθανότητας (odds) εμφάνισής του (Agresti, 2007).

Πίνακας 7.1

Πληροφοριακός πίνακας του μοντέλου

	N	Ποσοστά
Συμπεριλήφθηκαν	375	92,4%
Εξαιρούνται	31	7,6%
Σύνολο	406	100%

Για την προσαρμογή του συγκεκριμένου μοντέλου χρησιμοποιήθηκαν 375 παρατηρήσεις (σε ποσοστό 92,4%) ενώ από το μοντέλο μας εξαιρούνται οι 31 (σε ποσοστό 7,6%). Η συνεχής μεταβλητή της ηλικίας παρουσιάζει 30 ελλείπουσες τιμές ενώ η ποιοτική μεταβλητή του πολυμορφισμού rs4994 παρουσιάζει 1 ελλείπουσα παρατήρηση (την κατηγορία C:C). Γι 'αυτό το λόγο εξαιρούνται από το μοντέλο 31 παρατηρήσεις.

Πίνακας 7.2

Πληροφορίες των κατηγορικών μεταβλητών του μοντέλου

			N	Ποσοστά
Μεταβλητή απόκρισης	WEIGHTSITUATION 1	Normal weight	132	35,2%
		Pre-obese & Obese	243	64,8%
		Total	375	100,0%
Παράγοντες	Πολυμορφισμός rs1042713	G:G	132	35,2%
		G:A	187	49,9%
		A:A	56	14,9%
		Total	375	100,0%
	Πολυμορφισμός rs1042714	G:G	45	12,0%
		G:C	177	47,2%
		C:C	153	40,8%
		Total	375	100,0%
	Πολυμορφισμός rs1801252	A:A	308	82,1%
		G:A	64	17,1%
		G:G	3	0,8%
		Total	375	100,0%
	Πολυμορφισμός rs1801253	G:C	166	44,3%
		C:C	182	48,5%
		G:G	27	7,2%
		Total	375	100,0%
	Πολυμορφισμός rs4994	C:T	43	11,5%
		T:T	332	88,5%
		Total	375	100,0%
	Φύλο	Male	140	37,3%
Female		235	62,7%	
Total		375	100,0%	

Πίνακας 7.3

Περιγραφικά μέτρα της συνεχούς επεξηγηματικής μεταβλητής

		N	Minimum	Maximum	Mean	Std. Deviation
Συνεχής	Ηλικία	375	1	89	41,82	13,460

Οι παραπάνω πίνακες μας παρέχουν περιγραφικά στοιχεία για τις ανεξάρτητες ή επεξηγηματικές μεταβλητές του μοντέλου μας. Για παράδειγμα, το 64,8% των δεδομένων μας κατηγοριοποιούνται ως υπέρβαροι και παχύσαρκοι και το 62,7% είναι γυναίκες εθελόντριες. Ακόμη, για τον πολυμορφισμό rs4994, το 88,5% των εθελοντών του δείγματός μας έχουν το φυσικό ομόζυγο γονότυπο T:T (Πίνακας 7.2). Η μέση ηλικία του δείγματός μας είναι περίπου τα 42 έτη με τυπική απόκλιση να αποκλίνει αρκετά από το μέσο όρο, ενώ η ελάχιστη και η μέγιστη ηλικία είναι το 0 και τα 89 έτη αντίστοιχα (Πίνακας 7.3).

Πίνακας 7.4

Κριτήρια καλής προσαρμογής λογιστικού μοντέλου και μέτρα πληροφορίας

	Value	df	Value/df
Deviance	405,906	327	1,241
Scaled Deviance	405,906	327	
Pearson Chi-Square	340,892	327	1,042
Scaled Pearson Chi-Square	340,892	327	
Log Likelihood ^b	-213,638		
Akaike's Information Criterion (AIC)	451,276		
Finite Sample Corrected AIC (AICC)	452,138		
Bayesian Information Criterion (BIC)	498,399		
Consistent AIC (CAIC)	510,399		

Ο Πίνακας 7.4 δίνει τα αποτελέσματα του λογιστικού μοντέλου. Από τη θεωρία γνωρίζουμε ότι αν η τιμή Value/df πλησιάζει στην μονάδα, τότε έχουμε καλή προσαρμογή στα δεδομένα μας πράγμα το οποίο φαίνεται ότι ισχύει στο συγκεκριμένο

μοντέλο καθώς η προαναφερθείσα τιμή ισούται με 1,241 (Πίνακας 7.4). Το κριτήριο πληροφορίας του Akaike στο συγκεκριμένο μοντέλο ισούται με 451,276.

Πίνακας 7.5

Omnibus Test

Likelihood Ratio Chi-Square	df	Sig.
36,239	11	,000

Ο συγκεκριμένος έλεγχος χρησιμοποιείται για να συγκρίνουμε το προσαρμοσμένο μοντέλο με το μοντέλο που περιλαμβάνει μόνο το σταθερό όρο. Η τιμή της στατιστικής συνάρτησης ισούται με 36,239 με 11 βαθμούς ελευθερίας και, σύμφωνα με το p-value του ελέγχου, έχουμε ισχυρή ένδειξη ότι το μοντέλο με τις επτά επεξηγηματικές μεταβλητές (5 πολυμορφισμοί, φύλο και ηλικία) προσαρμόζεται καλύτερα από το μοντέλο με το σταθερό όρο ($p\text{-value} < 0.05$). Ελέγχοντας, ωστόσο, τη σημαντικότητα καθεμιάς από τις επεξηγηματικές μεταβλητές, βρήκαμε ότι κανείς από τους 5 πολυμορφισμούς δεν είναι σημαντικός για την μεταβλητή απόκρισης.

Πίνακας 7.6

Έλεγχος Wald

	Wald Chi-Square	df	Sig.
(Intercept)	1,471	1	,225
Φύλο	12,189	1	,000
Ηλικία	12,574	1	,000

Από τον παραπάνω πίνακα και, σύμφωνα με τις τιμές του p-value του ελέγχου, έχουμε ισχυρή ένδειξη ότι στατιστικά σημαντικές μεταβλητές θεωρούνται το φύλο και η ηλικία των εθελοντών ($p\text{-value} < 0.05$) και μας απομένει να δούμε τις εκτιμήσεις των συντελεστών $\hat{\beta}_i$. Συμπεραίνουμε δηλαδή ότι, τόσο το φύλο όσο και η ηλικία, επηρεάζουν την πιθανότητα να καταταχθεί κάποιος ως υπέρβαρος και παχύσαρκος.

Πίνακας 7.7

Εκτιμήσεις των παραμέτρων του μοντέλου

Parameter	B	Std. Error	Exp(B)	95% Wald Confidence Interval for Exp(B)	
				Lower	Upper
(Intercept)	- 0,983	1,3862	,106	,007	1,597
[Φύλο =1]	,861	,2492	2,387	1,465	3,891
[Φύλο =2]	0 ^a	.	1	.	.
Ηλικία	,032	,0089	1,032	1,014	1,050

Εφόσον, μέσω του πίνακα 7.6, καταλήξαμε στο συμπέρασμα ότι μόνον το φύλο και η ηλικία των εθελοντών θεωρούνται στατιστικά σημαντικές μεταβλητές θα ερμηνεύσουμε *μόνο* για τις συγκεκριμένες εκτιμήσεις των παραμέτρων τους. Αν αυξήσουμε την ηλικία κατά ένα έτος, δεδομένου ότι όλες οι υπόλοιπες εξηγηματικές μεταβλητές παραμένουν σταθερές, τότε ο λογάριθμος του odds θα αυξηθεί κατά 0,032 ή με άλλα λόγια η εκτιμώμενη σχετική πιθανότητα να είναι κάποιος παχύσαρκος/υπέρβαρος θα πολλαπλασιαστεί με $e^{0.032} = 1,032$. Επίσης, όσο μεγαλύτερη είναι η ηλικία τόσο πιθανότερο είναι να καταταχθεί κάποιος ως υπέρβαρος και παχύσαρκος μιας και $\widehat{\beta}_7 = 0,031 > 0$. Ένα 95% διάστημα εμπιστοσύνης του odds ratio είναι το (1.014,1.050).

Όσον αφορά το φύλο των εθελοντών, το log odds του να είναι κάποιος υπέρβαρος και παχύσαρκος είναι κατά 0.861 φορές υψηλότερο στους άνδρες σε σχέση με εκείνο των γυναικών. Δηλαδή, το odds του να είναι κάποιος υπέρβαρος και παχύσαρκος είναι περίπου διπλάσιο μεταξύ των ανδρών συγκρινόμενο με εκείνο των γυναικών.

Προκειμένου να ελέγξουμε αν το μοντέλο μας προσαρμόζεται καλά στα δεδομένα θα εφαρμόσουμε τον έλεγχο καλής προσαρμογής των Hosmer-Lemeshow. Η μηδενική υπόθεση που καλούμαστε να ελέγξουμε μας λέει ότι το μοντέλο προσαρμόζεται καλά στα δεδομένα. Φαίνεται ότι δεν απορρίπτουμε την μηδενική υπόθεση ($p\text{-value} > 0,05$) οπότε καταλήγουμε στο συμπέρασμα ότι το μοντέλο με τις 7 εξηγηματικές μεταβλητές προσαρμόζεται καλά στα δεδομένα.

Πίνακας 7.8

Έλεγχος καλής προσαρμογής Hosmer-Lemeshow

Hosmer and Lemeshow Test			
Step	Chi-Square	df	Sig.
1	10,143	8	,255

7.2 Περιγραφή και προσαρμογή του δεύτερου μοντέλου

Το δεύτερο μοντέλο λογιστικής παλινδρόμησης που χρησιμοποιήσαμε δεν διαφέρει με το προηγούμενο ως προς τις επεξηγηματικές μεταβλητές παρά μόνο ως προς την μεταβλητή απόκρισης. Η δεύτερη εξαρτημένη μεταβλητή που χρησιμοποιήσαμε είναι η *weightsituation2*, μια δίτιμη ποιοτική μεταβλητή. Η πρώτη κατηγορία είναι όσοι έχουν $BMI \geq 30$ τους οποίους τους κατατάξαμε ως παχύσαρκους ενώ η δεύτερη κατηγορία, ως μη παχύσαρκους, όσους έχουν $BMI < 30$. Η εξίσωση του συγκεκριμένου μοντέλου είναι αντίστοιχη με αυτή που παρουσιάσαμε στην προηγούμενη ενότητα. Να αναφέρουμε ότι τόσο οι παρατηρήσεις που συμπεριλήφθηκαν όσο και αυτές που εξαιρούνται από το μοντέλο είναι ίδιες με του μοντέλου της ενότητας 7.1.

Πίνακας 7.9

Περιγραφικά στοιχεία της εξαρτημένης μεταβλητής

			N	Percent
Μεταβλητή απόκρισης	WEIGHTSITUATION2	Non Obese	244	65,1%
		Obese	131	34,9%
		Total	375	100,0%

Όπως προαναφέραμε, στο συγκεκριμένο μοντέλο, το μόνο που διαφέρει είναι η δίτιμη εξαρτημένη μεταβλητή που χρησιμοποιήσαμε. Γι' αυτό το λόγο στον πίνακα 7.9 δεν παρουσιάζονται ξανά οι συχνότητες των παραγόντων του μοντέλου καθώς είναι οι ίδιες με του προηγούμενου μοντέλου. Σύμφωνα με τον παραπάνω πίνακα, το 65,1% του δείγματός μας κατηγοριοποιούνται ως μη παχύσαρκοι. Ακόμη να αναφέρουμε ότι και τα περιγραφικά μέτρα της συνεχούς ανεξάρτητης μεταβλητής παραμένουν τα ίδια και γι' αυτό δεν έχουμε παρουσιάσει τον πίνακα.

Πίνακας 7.10

Μέτρα πληροφορίας και κριτήρια καλής προσαρμογής λογιστικού μοντέλου

	Value	df	Value/df
Deviance	411,012	327	1,257
Scaled Deviance	411,012	327	
Pearson Chi-Square	336,657	327	1,030
Scaled Pearson Chi-Square	336,657	327	
Log Likelihood ^b	-217,577		
Akaike's Information Criterion (AIC)	459,155		
Finite Sample Corrected AIC (AICC)	460,017		
Bayesian Information Criterion (BIC)	506,278		
Consistent AIC (CAIC)	518,278		

Από τον παραπάνω πίνακα, λόγω του ότι ο λόγος Value/df πλησιάζει στην μονάδα, μπορούμε να πούμε ότι και το συγκεκριμένο μοντέλο προσαρμόζεται καλά στα δεδομένα μας. Εδώ ο λόγος Value/df ισούται με 1,257 (Πίνακας 7.10). Ένα ενδιαφέρον και πολύ δημοφιλές μέτρο πληροφορίας είναι το κριτήριο πληροφορίας του Akaike (AIC) το οποίο χρησιμοποιείται για να συγκρίνουμε διαφορετικά μοντέλα. Για το συγκεκριμένο ισούται με 459,155. Γνωρίζουμε ότι ένα «καλό» μοντέλο θεωρείται εκείνο με το μικρότερο AIC συγκρινόμενο με άλλα μοντέλα. Συγκρίνοντας λοιπόν το AIC του μοντέλου που παρουσιάσαμε στην προηγούμενη ενότητα, το οποίο ισούται με 451.276, μπορούμε να αναφέρουμε ότι το μοντέλο με εξαρτημένη μεταβλητή την weightsituation1 είναι «καλύτερο».

Πίνακας 7.11

Omnibus Test

Likelihood Ratio Chi-Square	df	Sig.
24,356	11	,011

Εδώ η τιμή της στατιστικής συνάρτησης του ελέγχου είναι ίση με 24,356 με 11 βαθμούς ελευθερίας. Σύμφωνα με το p-value έχουμε ισχυρή ένδειξη ότι το μοντέλο με τις ανεξάρτητες μεταβλητές προσαρμόζεται καλύτερα από το μοντέλο που περιλαμβάνει μόνο τον σταθερό όρο ($p\text{-value} < 0.05$).

Πίνακας 7.12

Έλεγχοι Wald

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	,000	1	1,000
Πολυμορφισμός rs1042714	6,161	2	,046
Ηλικία	11,069	1	,001

Αυτή τη φορά, κοιτώντας το p-value του ελέγχου Wald, διαπιστώνουμε πως μόνο δύο από τις επεξηγηματικές μεταβλητές είναι στατιστικά σημαντικές, η ηλικία των εθελοντών και ο πολυμορφισμός rs1042714 ($p\text{-value} < 0.05$). Διαπιστώνουμε πως, σε σχέση με το προηγούμενο μοντέλο, το φύλο δεν είναι στατιστικά σημαντικό. Μπορούμε να πούμε δηλαδή ότι, τόσο η ηλικία όσο και ο πολυμορφισμός rs1042714, επηρεάζουν την πιθανότητα να καταταχθεί κάποιος στους παχύσαρκους.

Πίνακας 7.13

Εκτιμήσεις των παραμέτρων του μοντέλου

Parameter	B	Std. Error	Exp(B)	95% Wald Confidence Interval for Exp(B)	
(Intercept)	-2,043	45679,1428	2,580E-11	,000	. ^a
[Πολυμορφισμός rs1042714=1]	0,810	,4461	3,004	1,253	7,201
[Πολυμορφισμός rs1042714=2]	,213	,2806	1,297	,748	2,247
[Πολυμορφισμός rs1042714=3]	0	.	1	.	.
Ηλικία	,028	,0087	1,029	1,012	1,047

Και στην συγκεκριμένη περίπτωση θα ερμηνεύσουμε τις εκτιμήσεις των παραμέτρων μόνο των δύο στατιστικά σημαντικών μεταβλητών, δηλαδή της ηλικίας των εθελοντών και του πολυμορφισμού rs1042714. Επειδή $\widehat{\beta}_7 = 0,028 > 0$, όσο μεγαλύτερη είναι η ηλικία τόσο πιθανότερο είναι να καταταχθεί κάποιος ως παχύσαρκος. Με άλλα λόγια, αν αυξήσουμε την ηλικία κατά ένα έτος τότε ο λογάριθμος του odds θα αυξηθεί κατά 0,028. Αν ένας εθελοντής έχει τον ομόζυγο γονότυπο G:G, τότε ο λογάριθμος της σχετικής πιθανότητας να καταταχθεί κάποιος ως παχύσαρκος έναντι κάποιου που έχει τον ετερόζυγο γονότυπο G:C αυξάνει κατά 0,810. Αυτό με άλλα λόγια σημαίνει ότι η σχετική πιθανότητα να είναι παχύσαρκο το άτομο που έχει τον γονότυπο G:G είναι περίπου 1 φορά μεγαλύτερη από την σχετική πιθανότητα όταν το άτομο έχει τον γονότυπο G:C.

Θέλουμε να ελέγξουμε την καλή προσαρμογή του δεύτερου μοντέλου που εφαρμόσαμε στα δεδομένα. Για να γίνει αυτό θα χρησιμοποιήσουμε τον έλεγχο καλής προσαρμογής Hosmer-Lemeshow. Κοιτώντας το p-value του ελέγχου, συμπεραίνουμε ότι και το συγκεκριμένο μοντέλο προσαρμόζεται καλά στα δεδομένα (p-value>0,05).

Πίνακας 7.14

Έλεγχος Hosmer-Lemeshow

Hosmer and Lemeshow Test			
Step	Chi-Square	df	Sig.
1	5,745	8	,676

7.3 Περιγραφή και προσαρμογή του τρίτου μοντέλου

Το τρίτο και τελευταίο μοντέλο που προσαρμόσαμε αποτελείται ξανά από τις ίδιες επεξηγηματικές μεταβλητές. Το μόνο που αλλάζει και εδώ είναι η μεταβλητή απόκρισης η οποία ονομάζεται *weightsituation3* και είναι μια δίτιμη διατακτική ποιοτική μεταβλητή. Θέσαμε σαν πρώτη κατηγορία όσους έχουν $BMI \geq 35$ ως υπερβολικά παχύσαρκους και τους υπόλοιπους σαν δεύτερη κατηγορία. Στην παρούσα ενότητα θα παρουσιάσουμε τα κριτήρια πληροφορίας, τους ελέγχους Wald καθώς και τις εκτιμήσεις των παραμέτρων του μοντέλου.

Πίνακας 7.15

Μέτρα πληροφορίας και κριτήρια καλής προσαρμογής λογιστικού μοντέλου

	Value	df	Value/df
Deviance	273,591	327	,837
Scaled Deviance	273,591	327	
Pearson Chi-Square	345,463	327	1,056
Scaled Pearson Chi-Square	345,463	327	
Log Likelihood ^b	-142,628		
Akaike's Information Criterion (AIC)	309,256		
Finite Sample Corrected AIC (AICC)	310,118		
Bayesian Information Criterion (BIC)	356,379		
Consistent AIC (CAIC)	368,379		

Από τον παραπάνω πίνακα παρατηρούμε ότι το κριτήριο πληροφορίας του Akaike είναι ίσο με 309,256. Παρατηρώντας, συγκριτικά και με τα δύο προηγούμενα μοντέλα, την τιμή του AIC καταλήγουμε στο συμπέρασμα ότι «καλύτερο» μοντέλο είναι το τρίτο. Ακόμη, η τιμή του λόγου Value/df ισούται με 0,837 πράγμα που σημαίνει ότι, επειδή είναι τιμή που πλησιάζει την μονάδα, και αυτό το μοντέλο προσαρμόζεται καλά στα δεδομένα μας.

Πίνακας 7.16

Omnibus Test

Likelihood Ratio Chi-Square	df	Sig.
10,575	11	,480

Η τιμή της στατιστικής συνάρτησης του ελέγχου ισούται με 10,575 με 11 βαθμούς ελευθερίας. Όμως, σε αντίθεση με τα προηγούμενα μοντέλα που χρησιμοποιήσαμε, το συγκεκριμένο δεν προσαρμόζεται καλύτερα από το μοντέλο που περιλαμβάνει μόνο τον σταθερό όρο ($p\text{-value} > 0.05$).

Πίνακας 7.17

Έλεγχοι Wald για τις επεξηγηματικές μεταβλητές

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	,000	1	1,000
Πολυμορφισμός rs1042714	6,385	2	,041

Αυτή τη φορά διαπιστώνουμε ότι μόνο ο πολυμορφισμός rs1042714 είναι στατιστικά σημαντική επεξηγηματική μεταβλητή ($p\text{-value} < 0.05$). Αυτό σημαίνει ότι ο συγκεκριμένος πολυμορφισμός επηρεάζει την πιθανότητα να καταταχθεί κάποιος στους υπερβολικά παχύσαρκους.

Πίνακας 7.18**Εκτιμήσεις των παραμέτρων του τρίτου μοντέλου**

Parameter	B	Std. Error	Exp(B)	95% Wald Confidence Interval for Exp(B)	
				Lower	Upper
(Intercept)	- 1,595	45486,9256	1,470E-10	,000	. ^a
[Πολυμορφισμός rs1042714=1]	0,391	,6064	3,819	1,164	12,536
[Πολυμορφισμός rs1042714=2]	-,327	,3945	1,001	,462	2,169
[Πολυμορφισμός rs1042714=3]	0	.	1	.	.

Όπως και στις δύο προηγούμενες ενότητες έτσι και τώρα θα ερμηνεύσουμε τις εκτιμήσεις της παραμέτρου μόνο της επεξηγηματικής μεταβλητής που βρέθηκε ότι είναι στατιστικά σημαντική. Αν ένα άτομο έχει τον γονότυπο G:G τότε ο λογάριθμος της σχετικής πιθανότητας να είναι υπερβολικά παχύσαρκο έναντι να μην είναι αυξάνει κατά 0,391.

Τέλος, θα ελέγξουμε ακόμη μια φορά την καλή προσαρμογή του συγκεκριμένου μοντέλου στα δεδομένα χρησιμοποιώντας τον έλεγχο Hosmer-Lemeshow. Από το p-value του ελέγχου, συμπεραίνουμε ότι και αυτό το τελευταίο μοντέλο προσαρμόζεται καλά στα δεδομένα μας.

Πίνακας 7.19**Έλεγχος Hosmer-Lemeshow**

Hosmer and Lemeshow Test			
Step	Chi-Square	df	Sig.
1	3,901	8	,866

ΚΕΦΑΛΑΙΟ 8

ΠΙΝΑΚΕΣ ΣΥΝΑΦΕΙΑΣ ΤΡΙΠΛΗΣ ΕΙΣΟΔΟΥ

Σκοπός του παρόντος κεφαλαίου είναι να μελετήσουμε την συνάφεια μίας επεξηγηματικής μεταβλητής και της μεταβλητής απόκρισης στα επίπεδα μιας τρίτης (επεξηγηματικής) μεταβλητής καθώς και αν υπάρχει σχέση μεταξύ περισσοτέρων ποιοτικών μεταβλητών μέσω του ελέγχου Mantel-Haenszel.

8.1 Προσαρμογή του πρώτου πίνακα τριπλής εισόδου

Για την εφαρμογή πινάκων τριπλής εισόδου θα χρησιμοποιήσουμε τρεις κατηγορικές μεταβλητές. Στην παρούσα ενότητα ως μεταβλητή απόκρισης θα θεωρήσουμε την *weightsituation1* με δύο επίπεδα, η περιγραφή της οποίας παρουσιάστηκε στο κεφάλαιο 7. Η δεύτερη ποιοτική μεταβλητή, με δύο επίπεδα, που θα χρησιμοποιήσουμε είναι το φύλο και η τρίτη είναι η μεταβλητή *classage*. Η διατακτική ποιοτική μεταβλητή *classage* “κατασκευάστηκε” χρησιμοποιώντας την συνεχή μεταβλητή της ηλικίας. Η συγκεκριμένη έχει τέσσερις κατηγορίες. Η πρώτη κατηγορία είναι τα άτομα ηλικίας 20-29, η δεύτερη τα άτομα ηλικίας 30-39, η τρίτη τα άτομα ηλικίας 40-49 και τέλος η τέταρτη κατηγορία είναι τα άτομα ηλικίας 50 ετών και πάνω (Saliba et al., 2014).

Συχνά η συνάφεια μεταξύ μιας επεξηγηματικής μεταβλητής X και της μεταβλητής απόκρισης Y μπορεί να διαφοροποιείται στα επίπεδα μιας τρίτης επεξηγηματικής μεταβλητής Z . Για να μελετηθεί αυτό χρησιμοποιούμε τους πίνακες συνάφειας τριπλής εισόδου (three-way tables). Δηλαδή κατηγοριοποιούμε τις X και Y στα επίπεδα της Z δείχνοντας μας την επίδραση της X στην μεταβλητή απόκρισης Y με βάση τα επίπεδα της Z (Agresti, 2002). Επιθυμούμε δηλαδή να συνδυάσουμε τα δεδομένα από τις 4 διαφορετικές ηλικιακές ομάδες και να δούμε συνολικά τι ισχύει όσον αφορά το αν κάποιος είναι παχύσαρκος ή μη και το φύλο.

Πίνακας 8.1

Weightsituation1 ανά φύλο και ηλικιακή ομάδα

Count					
ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΗΛΙΚΙΑΣ			WEIGHTSITUATION1		Total
			Normal weight	Pre-obese & Obese	
20-29	ΦΥΛΟ	Male	8	11	19
		Female	24	13	37
	Total		32	24	56
30-39	ΦΥΛΟ	Male	8	22	30
		Female	34	31	65
	Total		42	53	95
40-49	ΦΥΛΟ	Male	10	32	42
		Female	21	35	56
	Total		31	67	98
50+	ΦΥΛΟ	Male	7	38	45
		Female	17	51	68
	Total		24	89	113
Total	ΦΥΛΟ	Male	33	103	136
		Female	96	130	226
	Total		129	233	362

Ο παραπάνω πίνακας παρουσιάζει τον $2 \times 2 \times 4$ πίνακα συνάφειας. Μελετάμε την επίδραση του φύλου στο αν κάποιος είναι παχύσαρκος ή μη, αντιμετωπίζοντας την μεταβλητή της κατηγοριοποίησης της ηλικίας ως μεταβλητή ελέγχου. Από τον παραπάνω πίνακα, μπορούμε να πούμε ότι, ανεξαρτήτως φύλου, καθώς αυξάνεται η ηλικία περισσότερα άτομα κατατάσσονται στην κατηγορία των υπέρβαρων και παχύσαρκων. Μπορούμε να παρατηρήσουμε ότι μόνο 24 άτομα ηλικιών 20-29 ετών θεωρούνται ως υπέρβαρα και παχύσαρκα. Η τιμή αυτή ξεκινάει να παρουσιάζει μια άνοδο από τα άτομα ηλικίας 30-39 ετών. Επίσης, μπορούμε να πούμε ότι οι περισσότερες γυναίκες ηλικιών 40-49 και 50+ κατατάσσονται ως υπέρβαρες και παχύσαρκες. Να αναφέρουμε ότι το σύνολο των ατόμων είναι 362 και όχι 406 που είναι το μέγεθος του δείγματός μας. Αυτό συμβαίνει διότι η διατακτική μεταβλητή της ηλικίας έχει 44 ελλείπουσες τιμές.

Αυτό που μας ενδιαφέρει να ελέγξουμε είναι αν υπάρχει σχέση μεταξύ του φύλου και το αν κάποιος κατατάσσεται ως παχύσαρκος ή μη στα διάφορα επίπεδα της μεταβλητής της ηλικίας. Αυτό πραγματοποιείται εφαρμόζοντας τον έλεγχο των

Mantel-Haenszel. Ο συγκεκριμένος έλεγχος εφαρμόζει επαναλαμβανόμενα τεστ ανεξαρτησίας και ελέγχει την μηδενική υπόθεση ότι ο πληθυσμιακός σχετικός λόγος (odds ratio) κάθε υποπίνακα ισούται με 1. Διατυπώνοντας διαφορετικά την μηδενική υπόθεση, μπορούμε να πούμε ότι το αν κάποιος κατατάσσεται ως παχύσαρκος ή μη δεν σχετίζεται με το φύλο, λαμβάνοντας υπόψη τις διαφορετικές ηλικιακές ομάδες.

Πίνακας 8.2

Έλεγχος Mantel-Haenszel

	Chi-Squared	df	Asymptotic Significance (2-sided)
Cochran's	10,919	1	,001
Mantel-Haenszel	10,036	1	,002

Από τον παραπάνω πίνακα παρατηρούμε ότι η στατιστική συνάρτηση χ^2_{MH} ισούται με 10,036 με 1 βαθμό ελευθερίας και με p-value να ισούται με 0,002. Συνεπώς, αφού υπάρχει πολύ ισχυρή στατιστική ένδειξη απόρριψης της μηδενικής υπόθεσης (p-value<0,05), συμπεραίνουμε ότι, υπάρχει σχέση μεταξύ του αν κάποιος κατατάσσεται ως παχύσαρκος ή μη με το φύλο για τις διαφορετικές ηλικιακές ομάδες. Στο ίδιο συμπέρασμα καταλήγουμε και αν κοιτάξουμε τον έλεγχο του Cochran.

Πίνακας 8.3

Εκτίμηση του σχετικού λόγου πιθανοτήτων για τον έλεγχο M-H

Estimate			,445
ln(Estimate)			-,810
Standard Error of ln(Estimate)			,249
Asymptotic Significance (2-sided)			,001
Asymptotic 95% Confidence Interval	Common Odds Ratio	Lower Bound	,273
		Upper Bound	,724
	ln(Common Odds Ratio)	Lower Bound	-1,297
		Upper Bound	-,322

Στον παραπάνω πίνακα παρουσιάζεται ο σχετικός λόγος πιθανοτήτων (odds ratio) για τον έλεγχο Mantel-Haenszel. Ο λόγος αυτός ισούται με 0.445, που σημαίνει ότι η σχετική πιθανότητα ένας άνδρας να καταταχθεί ως κανονικός είναι σχεδόν το 45% της αντίστοιχης σχετικής πιθανότητας των γυναικών. Στο ίδιο συμπέρασμα περί απόρριψης της μηδενικής υπόθεσης καταλήγουμε αν κοιτάξουμε το 95% διάστημα εμπιστοσύνης (0.273,0.724) το οποίο δεν περιλαμβάνει το 1.

Πίνακας 8.4

Έλεγχος Breslow and Day

	Chi-Squared	df	Asymptotic Significance (2-sided)
Breslow-Day	,728	3	,867
Tarone's	,728	3	,867

Ο έλεγχος των Breslow and Day χρησιμοποιείται προκειμένου να ελέγξουμε την ομοιόμορφη συνάφεια, δηλαδή αν οι πληθυσμιακοί σχετικοί λόγοι των τεσσάρων ηλικιακών ομάδων ισούται μεταξύ τους. Η στατιστική συνάρτηση του ελέγχου ισούται με 0.728 με 3 βαθμούς ελευθερίας και το αντίστοιχο p-value ισούται με 0.867. Συνεπώς, καταλήγουμε στο συμπέρασμα ότι οι πληθυσμιακοί σχετικοί λόγοι των τεσσάρων ηλικιακών ομάδων είναι ίσοι μεταξύ τους.

8.2 Προσαρμογή του δεύτερου πίνακα τριπλής εισόδου

Στην παρούσα ενότητα θα χρησιμοποιήσουμε τις ίδιες μεταβλητές που χρησιμοποιήσαμε στην προηγούμενη ενότητα, δηλαδή το φύλο και τις τέσσερις διαφορετικές ηλικιακές ομάδες. Το μόνο που θα αλλάξει είναι η μεταβλητή απόκρισης. Ως μεταβλητή απόκρισης θα χρησιμοποιηθεί η weightsituation2, η περιγραφή της οποίας παρουσιάστηκε στο προηγούμενο κεφάλαιο.

Πίνακας 8.5

Weightsituation2 ανά φύλο και ηλικιακή ομάδα

Count					
ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΗΛΙΚΙΑΣ			WEIGHTSITUATION2		Total
			Non Obese	Obese	
20-29	ΦΥΛΟ	Male	14	5	19
		Female	32	5	37
	Total		46	10	56
30-39	ΦΥΛΟ	Male	20	10	30
		Female	48	17	65
	Total		68	27	95
40-49	ΦΥΛΟ	Male	28	14	42
		Female	35	21	56
	Total		63	35	98
50+	ΦΥΛΟ	Male	22	23	45
		Female	38	30	68
	Total		60	53	113
Total	ΦΥΛΟ	Male	84	52	136
		Female	153	73	226
	Total		237	125	362

Ο παραπάνω πίνακας παρουσιάζει τους 4 2×2 πίνακες συνάφειας που αντιστοιχούν στις τέσσερις διαφορετικές ηλικιακές ομάδες. Παρατηρούμε ότι, ανεξαρτήτως φύλου, τα περισσότερα άτομα και στις 4 ηλικιακές κλάσεις κατατάσσονται ως μη παχύσαρκα. Για παράδειγμα, 60 άτομα ηλικίας 50 και πάνω κατατάσσονται ως μη παχύσαρκα ενώ 53 άτομα ίδιας ηλικίας κατατάσσονται ως παχύσαρκα. Λαμβάνοντας υπόψη και το φύλο, παρατηρούμε ότι στις νέες ηλικίες (20-29 ετών) οι περισσότερες γυναίκες κατατάσσονται ως κανονικές ενώ μόνο 5 ως παχύσαρκες.

Στόχος μας είναι να ελέγξουμε εάν η μεταβλητή απόκρισης Y και η επεξηγηματική μεταβλητή X είναι υπό συνθήκη ανεξάρτητα δοθέντος του επιπέδου του τρίτου χαρακτηριστικού Z (Σαχλάς, 2016). Στην περίπτωση μας επεξηγηματική μεταβλητή X θεωρείται το φύλο και το τρίτο χαρακτηριστικό Z θεωρούνται οι διαφορετικές ηλικιακές ομάδες ενώ μεταβλητή απόκρισης Y η διατακτική μεταβλητή weightsituation2. Αυτό θα ελεγχθεί χρησιμοποιώντας τον Mantel-Haenszel έλεγχο.

Πίνακας 8.6

Έλεγχος Mantel-Haenszel

	Chi-Squared	df	Asymptotic Significance (2-sided)
Cochran's	,865	1	,352
Mantel-Haenszel	,654	1	,419

Η τιμή της στατιστικής συνάρτησης του Mantel-Haenszel ισούται με 0,865 με 1 βαθμό ελευθερίας και το αντίστοιχο p-value ισούται με 0,352. Συνεπώς, μπορούμε να πούμε ότι, στην συγκεκριμένη περίπτωση, δεν φαίνεται να υπάρχει σχέση μεταξύ του αν κάποιος κατατάσσεται ως παχύσαρκος ή μη με το φύλο για τις διαφορετικές ηλικιακές ομάδες ($p\text{-value}>0,05$).

Πίνακας 8.7

Έλεγχος ομοιόμορφης συνάφειας

	Chi-Squared	df	Asymptotic Significance (2-sided)
Breslow-Day	1,725	3	,631
Tarone's	1,724	3	,632

Η τιμή της στατιστικής συνάρτησης του ελέγχου ομοιόμορφης συνάφειας Breslow and Day ισούται με 1,725 με 3 βαθμούς ελευθερίας ενώ το p-value του ελέγχου ισούται με 0,631. Συνεπώς, φαίνεται ότι δεν απορρίπτουμε την μηδενική υπόθεση περί ισότητας των πληθυσμιακών σχετικών λόγων των τεσσάρων ηλικιακών ομάδων ($p\text{-value}>0,05$). Δηλαδή τα odds ratio των 4 ηλικιακών ομάδων δεν διαφέρουν. Η εκτίμηση του κοινού πληθυσμιακού σχετικού λόγου ισούται με 0,807, πράγμα που σημαίνει ότι η σχετική πιθανότητα του να μην είναι κάποιος παχύσαρκος είναι σχεδόν 0,8 φορές μεγαλύτερη στους άνδρες από ότι στις γυναίκες. Επειδή το 95% δ.ε για τον

σχετικό λόγο (0.513,1.271) περιλαμβάνει το 1, ορθώς δεν απορρίψαμε την μηδενική υπόθεση προηγουμένως.

8.3 Προσαρμογή του τρίτου πίνακα τριπλής εισόδου

Στην συγκεκριμένη ενότητα ως επεξηγηματική μεταβλητή X θα χρησιμοποιήσουμε την δίτιμη μεταβλητή του φύλου ενώ ως τρίτο χαρακτηριστικό Z θα χρησιμοποιηθούν οι τέσσερις ηλικιακές κλάσεις. Η μεταβλητή απόκρισης Y θα είναι η weightsituation3 η οποία έχει παρουσιαστεί εκτενώς στο κεφάλαιο 7. Σκοπός μας είναι να κατασκευάσουμε έναν πίνακα τριπλής εισόδου καθώς και να μελετήσουμε την σχέση του αν κάποιος είναι υπερβολικά παχύσαρκος ή όχι με το φύλο στα διαφορετικά επίπεδα ηλικιακών ομάδων.

Πίνακας 8.8

Weightsituation3 ανά φύλο και ηλικιακή ομάδα

Count					
ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΗΛΙΚΙΑΣ			WEIGHTSITUATI ON3		Total
			Non Obese	Extreme Obese	
20-29	ΦΥΛΟ	Male	17	2	19
		Female	34	3	37
	Total		51	5	56
30-39	ΦΥΛΟ	Male	25	5	30
		Female	55	10	65
	Total		80	15	95
40-49	ΦΥΛΟ	Male	37	5	42
		Female	47	9	56
	Total		84	14	98
50+	ΦΥΛΟ	Male	36	9	45
		Female	58	10	68
	Total		94	19	113
Total	ΦΥΛΟ	Male	115	21	136
		Female	194	32	226
	Total		309	53	362

Ο παραπάνω πίνακας αποτελείται από τέσσερις 2×2 πίνακες συνάφειας που αντιστοιχούν στις 4 διαφορετικές ηλικιακές κλάσεις. Ανεξαρτήτως φύλου, παρατηρούμε ότι σε κάθε μια ηλικιακή ομάδα τα περισσότερα άτομα του δείγματός

μας κατατάσσονται ως μη παχύσαρκα. Για παράδειγμα, 94 άτομα ηλικίας 50 και πάνω κατατάσσονται ως μη παχύσαρκα. Στις νεότερες ηλικίες (20-29 ετών) 51 άτομα κατατάσσονται ως μη παχύσαρκα ενώ μόνο 5 ως υπερβολικά παχύσαρκα.

Πίνακας 8.9

Έλεγχος Mantel-Haenszel

	Chi-Squared	df	Asymptotic Significance (2-sided)
Cochran's	,096	1	,757
Mantel-Haenszel	,024	1	,878

Η τιμή της στατιστικής συνάρτησης του ελέγχου Mantel-Haenszel ισούται με 0,024 με 1 βαθμό ελευθερίας ενώ το αντίστοιχο p-value ισούται με 0,878. Παρατηρούμε ότι, και στην συγκεκριμένη περίπτωση, δεν υπάρχει σχέση μεταξύ του αν κάποιος είναι υπερβολικά παχύσαρκος ή όχι με το φύλο στα διαφορετικά επίπεδα ηλικιακών ομάδων.

Πίνακας 8.10

Έλεγχος Breslow-Day

	Chi-Squared	df	Asymptotic Significance (2-sided)
Breslow-Day	,893	3	,827
Tarone's	,893	3	,827

Η τιμή της στατιστικής συνάρτησης του ελέγχου ομοιόμορφης συνάφειας ισούται με 0,893 με 3 βαθμούς ελευθερίας ενώ το p-value του ελέγχου ισούται με 0,827. Συνεπώς, φαίνεται ότι δεν απορρίπτουμε την μηδενική υπόθεση περί ισότητας των πληθυσμιακών σχετικών λόγων των τεσσάρων ηλικιακών ομάδων ($p\text{-value} > 0,05$). Η εκτίμηση του σχετικού λόγου πιθανοτήτων ισούται με 0,910. Αυτό σημαίνει ότι η σχετική πιθανότητα ένας άνδρας να καταταχθεί ως κανονικός είναι σχεδόν το 91% της

αντίστοιχης σχετικής πιθανότητας των γυναικών. Το 95% δ.ε για τον σχετικό λόγο είναι το (0.500, 1.657).

ΚΕΦΑΛΑΙΟ 9

ΣΥΓΚΡΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Σκοπός του κεφαλαίου αυτού είναι να γίνει αρχικά μια σύνοψη των αποτελεσμάτων της ανάλυσής μας, στην συνέχεια επισκόπηση της βιβλιογραφίας που υπάρχει πάνω στο θέμα των πολυμορφισμών και της παχυσαρκίας και τέλος σύγκριση αποτελεσμάτων της δικής μας στατιστικής ανάλυσης με αποτελέσματα άλλων μελετών που έχουν διεξαχθεί για το συγκεκριμένο θέμα.

9.1 Σύνοψη αποτελεσμάτων

Στην ενότητα αυτή θα παρουσιάσουμε συνοπτικά τα αποτελέσματα της στατιστικής ανάλυσής μας. Στο κεφάλαιο 5 πραγματοποιήθηκε αρχικά μια περιγραφική ανάλυση στις ποιοτικές μεταβλητές του δείγματός μας (φύλο, πολυμορφισμοί ADRB3rs4994, ADRB2rs1042713, ADRB2rs1042714, ADRB1rs1801252 και ADRB1rs1801253 και classificationbmi). Στην συνέχεια ελέγξαμε κατά πόσο οι πολυμορφισμοί καθώς και το φύλο συσχετίζονται με την ποιοτική διατακτική μεταβλητή classificationbmi. Από την ανάλυσή μας, μέσω του ελέγχου ανεξαρτησίας, στατιστικά σημαντική συσχέτιση παρατηρείται με τον πολυμορφισμό ADRB1rs1801253 καθώς και με το φύλο των εθελοντών. Τρίτο κομμάτι της στατιστικής ανάλυσής μας ήταν η εφαρμογή μεθόδων απαραμετρικής ανάλυσης προκειμένου να εξετάσουμε εάν ο δείκτης μάζας σώματος επηρεάζεται από το φύλο των εθελοντών καθώς και από τους πέντε πολυμορφισμούς του δείγματός μας. Προτού αποφασίσουμε ποιες μεθόδους ανάλυσης θα χρησιμοποιήσουμε πρώτα ελέγξαμε, μέσω του ελέγχου Shapiro-Wilk καθώς και μέσω Q-Q διαγραμμάτων, αν τα δείγματα που διαθέτουμε προέρχονται από την κανονική κατανομή. Σε όλες τις περιπτώσεις τα δείγματα απέκλιναν από την κανονική κατανομή. Από την ανάλυση μας, μέσω του ελέγχου Mann-Whitney, καταλήξαμε στο συμπέρασμα ότι μόνο το φύλο των εθελοντών επηρεάζει τον δείκτη μάζας σώματος. Συγκεκριμένα, το φύλο με το μεγαλύτερο δείκτη μάζας σώματος, για το δείγμα μας, είναι οι άνδρες. Στο κεφάλαιο 7 εφαρμόσαμε τρία μοντέλα λογιστικής παλινδρόμησης με ανεξάρτητες μεταβλητές τους πέντε πολυμορφισμούς του δείγματός μας, το φύλο και την ηλικία των εθελοντών με διαφορετική μεταβλητή απόκρισης για το καθένα. Μέσω του ελέγχου Hosmer-Lemeshow καταλήξαμε στο συμπέρασμα ότι και τα τρία

μοντέλα που εφαρμόσαμε προσαρμόζονται καλά στα δεδομένα. Στο πρώτο μοντέλο λογιστικής παλινδρόμησης στατιστικά σημαντικές μεταβλητές ήταν μόνο το φύλο και η ηλικία των εθελοντών καταλήγοντας στο συμπέρασμα ότι όσο μεγαλύτερη είναι η ηλικία τόσο πιθανότερο είναι να καταταχθεί κάποιος ως υπέρβαρος και παχύσαρκος. Στο δεύτερο μοντέλο στατιστικά σημαντικές μεταβλητές κρίθηκαν ο πολυμορφισμός rs1042714 και η ηλικία ενώ στο τρίτο και τελευταίο μοντέλο μόνο ο πολυμορφισμός rs1042714.

9.2 Επισκόπηση διεθνούς βιβλιογραφίας

Όσον αφορά την διεθνή βιβλιογραφία μπορούμε να παρατηρήσουμε ότι ορισμένα αποτελέσματα συμπίπτουν μεταξύ τους. Αρχικά, όλα τα δείγματα της διεθνούς βιβλιογραφίας που κοιτάξαμε βρίσκονται σε ισορροπία Hardy-Weinberg (Kurokawa et al. 2008, Takeuchi et al. 2012). Σχετικά με την συσχέτιση του δείκτη μάζας σώματος με τον πολυμορφισμό rs4994 του γονιδίου ADRB3 έρευνες έχουν δείξει ότι δεν παρατηρείται συσχέτιση μεταξύ των δύο αυτών μεταβλητών (Takeuchi et al. 2012, Tahara et al. 2010, Garcia et al. 2016). Αντίθετα άλλη έρευνα που διεξήχθη έδειξε ότι ο παραπάνω πολυμορφισμός παρουσιάζει στατιστικά σημαντική σχέση με τον δείκτη μάζας σώματος. Στην συγκεκριμένη έρευνα μάλιστα κατέληξαν στο συμπέρασμα ότι ο πολυμορφισμός ADRB3 Trp64 παρουσιάζει σχέση με το BMI στην Ανατολική Ασία και όχι στην Ευρώπη (Kurokawa et al., 2008). Όσον αφορά τον πολυμορφισμό rs1801253 του γονιδίου ADRB1 έρευνα έδειξε ότι υπάρχει στατιστικά σημαντική σχέση του πολυμορφισμού αυτού με τον δείκτη μάζας σώματος (Garcia et al. 2016, Chou et al., 2012). Επιπλέον, εφαρμόζοντας πολλαπλή γραμμική παλινδρόμηση, η συνεχής μεταβλητή της ηλικίας βρέθηκε στατιστικά σημαντική σε αντίθεση με το φύλο (Takeuchi et al. 2012), αλλά και η μεταβλητή του φύλου βρέθηκε στατιστικά σημαντική (Garcia et al., 2016). Τα δείγματα που χρησιμοποιήθηκαν στις συγκεκριμένες έρευνες είναι μεγάλα και σύμφωνα με την έρευνα των Kurokawa et al χρειάζονται μεγάλα μεγέθη δειγμάτων προκειμένου να εκτιμηθούν και να ερμηνευτούν αξιόπιστα σχέσεις γονιδίων. Οι μέθοδοι που χρησιμοποιήθηκαν στις προαναφερθείσες έρευνες ήταν έλεγχοι ανεξαρτησίας χ^2 , έλεγχοι Mann-Whitney και Kruskal-Wallis καθώς και πολλαπλή γραμμική παλινδρόμηση.

9.3 Σύγκριση αποτελεσμάτων

Το δείγμα που χρησιμοποιήσαμε στην δική μας στατιστική ανάλυση ήταν μικρό (406 άτομα) σε σχέση με το μέγεθος του δείγματος που χρησιμοποιήθηκε στις έρευνες που αναφέραμε προηγουμένως. Συγκρίνοντας τα αποτελέσματά μας με εκείνα των άλλων ερευνών παρατηρούμε ότι και στο δικό μας δείγμα ο δείκτης μάζας σώματος και ο πολυμορφισμός rs1801253 παρουσιάζουν μια στατιστικά σημαντική σχέση. Επιπλέον, βλέπουμε να συμπίπτουν τα αποτελέσματά μας με εκείνα των άλλων μελετών όσον αφορά τη σχέση του δείκτη μάζας σώματος με τον πολυμορφισμό rs4994. Όσον αφορά την συνεχή μεταβλητή της ηλικίας καθώς και για το φύλο τα αποτελέσματά μας συμφωνούν με την έρευνα των Takeuchi et al. και Garcia et al. αντίστοιχα.

9.4 Επίλογος

Παρατηρήσαμε ότι τα αποτελέσματα που προέκυψαν από την στατιστική ανάλυσή μας συμφωνούν με τα αποτελέσματα των ερευνών που διεξήχθησαν στο εξωτερικό. Οι περισσότεροι πολυμορφισμοί του δείγματός μας δεν βρέθηκαν στατιστικά σημαντικοί. Τα αποτελέσματα της ανάλυσής μας και κάθε ανάλυσης που πραγματεύεται το συγκεκριμένο θέμα μπορούν να μας ωθήσουν στο να αναρωτηθούμε, και να ερευνήσουμε στο μέλλον έχοντας δείγμα μεγαλύτερης κλίμακας, το τι συμβαίνει με τους υπόλοιπους πολυμορφισμούς.

Παράρτημα

Παράρτημα Α: Κρίσιμες τιμές

Στο κεφάλαιο 3 αναφερθήκαμε στους μη παραμετρικούς ελέγχους υποθέσεων. Στις ακόλουθες εικόνες εμφανίζονται οι κρίσιμες τιμές των τριών ελέγχων που αναφερθήκαμε.

Κρίσιμες τιμές U_{α} -του ελέγχου Mann - Whitney																				
	m																			
N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	0	0	0	0	1	1	1	1	1	2	2	2	2
3	-	-	-	-	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4	-	-	-	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5	-	-	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6	-	-	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7	-	-	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	-	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	-	0	2	4	7	10	12	15	17	21	23	26	28	31	34	37	39	42	45	48
10	-	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	-	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	-	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	-	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	-	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15	-	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	-	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	-	2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18	-	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	-	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	-	2	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

$\alpha=0.05$

Εικόνα Π.Α.1.: Οι κρίσιμες τιμές του ελέγχου Mann-Whitney ($\alpha=0.05$)

Κρίσιμες τιμές π_{α} του ελέγχου Sum Rank του Wilcoxon $n \geq m$															
	m														
n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	10	-	-	-	-	-	-	-	-	-	-	-
5	-	-	6	11	17	-	-	-	-	-	-	-	-	-	-
6	-	-	7	12	18	26	-	-	-	-	-	-	-	-	-
7	-	-	7	13	20	27	36	-	-	-	-	-	-	-	-
8	-	3	8	14	21	29	38	49	-	-	-	-	-	-	-
9	-	3	8	14	22	31	40	51	62	-	-	-	-	-	-
10	-	3	9	15	23	32	42	53	65	78	-	-	-	-	-
11	-	3	9	16	24	34	44	55	68	81	96	-	-	-	-
12	-	4	10	17	26	35	46	58	71	84	99	115	-	-	-
13	-	4	10	18	27	37	48	60	73	88	103	119	136	-	-
14	-	4	11	19	28	38	50	62	76	91	106	123	141	160	-
15	-	4	11	20	29	40	52	65	79	94	110	127	145	164	184

Εικόνα Π.Α.2.: Οι κρίσιμες τιμές του ελέγχου Sum-Rank του Wilcoxon ($\alpha=0.05$)

Μεγέθη δειγμάτων			H_{cr}	Μεγέθη δειγμάτων			H_{cr}
n_1	n_2	n_3		n_1	n_2	n_3	
3	2	2	4.68	5	3	1	4.92
3	3	3	5.6	5	3	2	5.2
4	2	2	5.15	5	3	3	5.35
4	3	1	5.21	5	4	1	4.92
4	3	2	5.42	5	4	2	5.27
4	3	3	5.73	5	4	3	5.63
4	4	1	4.92	5	4	4	5.62
4	4	2	5.35	5	5	1	5.01
4	4	3	5.58	5	5	2	5.26
4	4	4	5.68	5	5	3	5.64
5	2	1	4.95	5	5	4	5.64
5	2	2	5.08	5	5	5	5.72

$\alpha=0.05$

Εικόνα Π.Α.3 : Οι κρίσιμες τιμές του ελέγχου Kruskal-Wallis ($\alpha=0.05$)

Παράρτημα Β: Πίνακες Συνάφειας

Στο κεφάλαιο 5 παρουσιάσαμε την στατιστική ανάλυση των δεδομένων μας εφαρμόζοντας, για να εξετάσουμε την σχέση των πολυμορφισμών με την παχυσαρκία, πίνακες συνάφειας. Στους ακόλουθους πίνακες παρουσιάζουμε τους πίνακες συνάφειας καθώς και ορισμένους ελέγχους ανεξαρτησίας.

Πίνακας Π.Β.1

Πίνακας συνάφειας: classificationBMI-rs4994

		ΠΟΛΥΜΟΡΦΙΣΜΟΣ		Total	
		rs4994			
		C:T	T:T		
ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	Underweight	Count	1	11	12
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	8,3%	91,7%	100,0%
	Normal weight	Count	14	113	127
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	11,0%	89,0%	100,0%
	Pre-obese	Count	14	103	117
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	12,0%	88,0%	100,0%
	Obese class I	Count	9	76	85

		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	10,6%	89,4%	100,0%
	Obese class II	Count	5	34	39
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	12,8%	87,2%	100,0%
	Obese class III	Count	4	21	25
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	16,0%	84,0%	100,0%
Total	Count		47	358	405
	% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ		11,6%	88,4%	100,0%

Πίνακας Π.Β.2

Πίνακας συνάφειας: classificationBMI-rs1042713

			ΠΟΛΥΜΟΡΦΙΣΜΟΣ rs1042713			Total
			G:G	G:A	A:A	
ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	Underweight	Count	4	4	4	12
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	33,3%	33,3%	33,3%	100,0%
	Normal weight	Count	48	65	15	128
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	37,5%	50,8%	11,7%	100,0%
	Pre-obese	Count	38	57	22	117
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	32,5%	48,7%	18,8%	100,0%
	Obese class I	Count	34	43	8	85
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	40,0%	50,6%	9,4%	100,0%
	Obese class II	Count	14	17	8	39

		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	35,9%	43,6%	20,5%	100,0%
	Obese class III	Count	7	13	5	25
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	28,0%	52,0%	20,0%	100,0%
Total		Count	145	199	62	406
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	35,7%	49,0%	15,3%	100,0%

Πίνακας Π.Β.3

Πίνακας συνάφειας: classificationBMI-rs1042714

			ΠΟΛΥΜΟΡΦΙΣΜΟΣ rs1042714			Total
			G:G	G:C	C:C	
ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	Underweight	Count	3	3	6	12
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	25,0%	25,0%	50,0%	100,0%
	Normal weight	Count	13	64	51	128
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	10,2%	50,0%	39,8%	100,0%
	Pre-obese	Count	11	53	53	117
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	9,4%	45,3%	45,3%	100,0%
	Obese class I	Count	13	44	28	85
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	15,3%	51,8%	32,9%	100,0%
	Obese class II	Count	8	12	19	39
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	20,5%	30,8%	48,7%	100,0%
		Count	4	12	9	25

	Obese class III	% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	16,0%	48,0%	36,0%	100,0%
Total	Count		52	188	166	406
	% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ		12,8%	46,3%	40,9%	100,0%

Πίνακας Π.Β.4

Πίνακας συνάφειας: classificationBMI-rs1801252

			ΠΟΛΥΜΟΡΦΙΣΜΟΣ rs1801252			Total	
			A:A	G:A	G:G		
ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	Underweight	Count	11	1	0	12	
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	91,7%	8,3%	0,0%	100,0%	
	Normal weight	Count	109	16	3	128	
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	85,2%	12,5%	2,3%	100,0%	
	Pre-obese	Count	91	25	1	117	
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	77,8%	21,4%	0,9%	100,0%	
	Obese class I	Count	67	17	1	85	
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	78,8%	20,0%	1,2%	100,0%	
	Obese class II	Count	35	4	0	39	
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	89,7%	10,3%	0,0%	100,0%	
	Obese class III	Count	19	6	0	25	
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	76,0%	24,0%	0,0%	100,0%	
	Total		Count	332	69	5	406
			% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	81,8%	17,0%	1,2%	100,0%

Πίνακας Π.Β.5

Πίνακας συνάφειας: classificationBMI-rs1801253

			ΠΟΛΥΜΟΡΦΙΣΜΟΣ rs1801253			Total	
			G:C	C:C	G:G		
ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	Underweight	Count	5	4	3	12	
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	41,7%	33,3%	25,0%	100,0%	
	Normal weight	Count	60	55	13	128	
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	46,9%	43,0%	10,2%	100,0%	
	Pre-obese	Count	47	62	8	117	
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	40,2%	53,0%	6,8%	100,0%	
	Obese class I	Count	37	45	3	85	
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	43,5%	52,9%	3,5%	100,0%	
	Obese class II	Count	17	15	7	39	
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	43,6%	38,5%	17,9%	100,0%	
	Obese class III	Count	12	13	0	25	
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	48,0%	52,0%	0,0%	100,0%	
	Total		Count	178	194	34	406
			% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	43,8%	47,8%	8,4%	100,0%

Πίνακας Π.Β.6

Έλεγχος Chi-Square

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	17,348 ^a	10	,067
Likelihood Ratio	17,746	10	,059
Linear-by-Linear Association	,291	1	,589
N of Valid Cases	406		

a. 3 cells (16,7%) have expected count less than 5. The minimum expected count is 1,00.

Πίνακας Π.Β.7

Πίνακας συνάφειας: classificationBMI-gender

			ΦΥΛΟ		Total
			Male	Female	
ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	Underweight	Count	1	11	12
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	8,3%	91,7%	100,0%
	Normal weight	Count	33	95	128
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	25,8%	74,2%	100,0%
	Pre-obese	Count	57	60	117
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	48,7%	51,3%	100,0%
	Obese class I	Count	33	52	85
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	38,8%	61,2%	100,0%
	Obese class II	Count	16	23	39
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	41,0%	59,0%	100,0%

	Obese class III	Count	10	15	25
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	40,0%	60,0%	100,0%
Total		Count	150	256	406
		% within ΚΑΤΗΓΟΡΙΕΣ ΒΑΡΟΥΣ	36,9%	63,1%	100,0%

Πίνακας Π.Β.8

Έλεγχος Chi-Square

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	18,533 ^a	5	,002
Likelihood Ratio	19,715	5	,001
Linear-by-Linear Association	5,469	1	,019
N of Valid Cases	406		
a. 1 cells (8,3%) have expected count less than 5. The minimum expected count is 4,43.			

Βιβλιογραφία

Ελληνική

- Ευαγγελάρας Χ. (2016), *Ανάλυση δεδομένων με τη χρήση στατιστικών πακέτων*, σημειώσεις Π.Μ.Σ Εφαρμοσμένης Στατιστικής
- Ηλιόπουλος Γ. (2017), *Ανάλυση Διακριτών Δεδομένων*, σημειώσεις Π.Μ.Σ Εφαρμοσμένης Στατιστικής
- Ηλιόπουλος Γ. (2017), *Γενικευμένα Γραμμικά Μοντέλα*, σημειώσεις Π.Μ.Σ Εφαρμοσμένης Στατιστικής
- Νικολάου, Χ., Χουβαρδάς, Π., (2015), *Ανάλυση της Γενετικής Ποικιλομορφίας* [Κεφάλαιο 10], Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, Αθήνα.
- Παπανικολάου, Γ., Παλαιολόγου, Δ., Κατσαρέλη, Ε., Κατσίλα, Θ., Τσαρουχά, Χ., Τζέτη, Μ., Λιλάκος, Κ., Δούκισσας, Λ. (2015). *Εργαστηριακές ασκήσεις γενετικής του ανθρώπου*. [ηλεκτρ. βιβλ.] Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών.
- Πολίτης Κ. (2016), *Βιοστατιστική και στατιστικές μέθοδοι στην επιδημιολογία*, σημειώσεις Π.Μ.Σ Εφαρμοσμένης Στατιστικής
- Σαχλάς Α. (2016), *Ανάλυση Διακριτών Δεδομένων*, εργαστηριακές σημειώσεις Π.Μ.Σ Εφαρμοσμένης Στατιστικής
- Σαχλάς Α. (2016), *Βιοστατιστική και στατιστικές μέθοδοι στην επιδημιολογία*, εργαστηριακές σημειώσεις Π.Μ.Σ Εφαρμοσμένης Στατιστικής
- Σπύρογλου Σ., (2007), *Ο λιπώδης ιστός ως ενδοκρινές όργανο, λιποκύτταρο και μεταβολικό σύνδρομο*. Διπλωματική εργασία, πανεπιστήμιο Πατρών, σχολή επιστημών υγείας – τμήμα ιατρικής, πρόγραμμα μεταπτυχιακών σπουδών στις εφαρμογές των βασικών ιατρικών επιστημών
- Χαλικιάς, Χ., Μανωλέσσου, Α., Λάλου, Π., (2015), *Μεθοδολογία Έρευνας και Εισαγωγή στη Στατιστική Ανάλυση Δεδομένων με το IBM SPSS STATISTICS*, Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, Αθήνα.

Ξένα

Agresti, A. (2007), *An introduction to categorical data analysis second edition*, Wiley

Aradillas-Garcia, Celia, et al. Obesity is associated with the Arg389Gly ADRB1 but not with the Trp64Arg ADRB3 polymorphism in children from San Luis Potosí and León, México. *Journal of biomedical research*, 2017, 31.1: 40.

Baker, C. (2018), *Obesity Statistics*, *House of Commons Library*

Chou, Yi-Chun, et al. Association of adrenergic receptor gene polymorphisms with adolescent obesity in Taiwan. *Pediatrics international*, 2012, 54.1: 111-116.

Gibbons RD., Bhaumik DK., Aryal S. (2009), *Statistical Methods for Groundwater Monitoring second edition*, Wiley

Herrera, Blanca M.; Lindgren, Cecilia M. The genetics of obesity. *Current diabetes reports*, 2010, 10.6: 498-505.

Kurokawa, Naoyuki, et al. Association of BMI with the β 3-adrenergic receptor gene polymorphism in Japanese: meta-analysis. *Obesity Research*, 2001, 9.12: 741-745.

Kurokawa, Naoyuki, et al. The ADRB3 Trp64Arg variant and BMI: a meta-analysis of 44 833 individuals. *International journal of obesity*, 2008, 32.8: 1240.

Pratt J., Gibbons JD. (1981), *Concepts of Nonparametric Theory*, Springer-Verlag

Rosmond, R. Association studies of genetic polymorphisms in central obesity: a critical review. *International journal of obesity*, 2003, 27.10: 1141.

Saliba, Louise F., et al. Obesity-related gene ADRB2, ADRB3 and GHRL polymorphisms and the response to a weight loss diet intervention in adult women. *Genetics and molecular Biology*, 2014, 37.1: 15-22.

Szendrei, Barbara, et al. Influence of ADRB2 Gln27Glu and ADRB3 Trp64Arg polymorphisms on body weight and body composition changes after a controlled weight-loss intervention. *Applied Physiology, Nutrition, and Metabolism*, 2015, 41.3: 307-314.

Takeuchi, Shouhei, et al. ADRB3 polymorphism associated with BMI gain in Japanese men. *Experimental diabetes research*, 2012, 2012.

Διαδικτυακοί Τόποι

<http://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight>

<https://functionalmedsystem.com/el/%CF%80%CE%B1%CF%87%CF%85%CF%83%CE%B1%CF%81%CE%BA%CE%AF%CE%B1-%CE%BA%CE%B1%CE%B9-%CF%80%CE%B5%CF%81%CE%B9%CF%84%CF%84%CE%AC-%CE%BA%CE%B9%CE%BB%CE%AC/>

<https://thehealthlab.gr/ygeia/%CF%84%CE%B9-%CE%B5%CE%AF%CE%BD%CE%B1%CE%B9-%CE%B7-%CF%80%CE%B1%CF%87%CF%85%CF%83%CE%B1%CF%81%CE%BA%CE%AF%CE%B1>

<https://coolweb.gr/bmi-ti-einai/>

<https://dfordiet.wordpress.com/2015/06/23/%CF%85%CF%80%CE%BF%CE%BB%CE%BF%CE%B3%CE%B9%CF%83%CE%BC%CF%8C%CF%82-%CE%B4%CE%B5%CE%AF%CE%BA%CF%84%CE%B7-%CE%BC%CE%AC%CE%B6%CE%B1%CF%82-%CF%83%CF%8E%CE%BC%CE%B1%CF%84%CE%BF%CF%82/>

<https://www.dianeosis.org/2016/04/paxysarkia stin ellada tis krisis/>

<https://www.niddk.nih.gov/health-information/health-statistics/overweight-obesity>

<https://ghr.nlm.nih.gov/primer/genomicresearch/snp>

https://www.researchgate.net/figure/The-schema-of-Single-nucleotide-polymorphism-SNP_fig1_221912832

<https://en.wikipedia.org/wiki/Lipolysis>

http://health.hellasmagazine.gr/2012/05/blog-post_9.html

https://el.wikipedia.org/wiki/%CE%A0%CE%BF%CE%B9%CE%BF%CF%84%CE%B9%CE%BA%CE%AE_%CE%BC%CE%B5%CF%84%CE%B1%CE%B2%CE%BB%CE%B7%CF%84%CE%AE

<https://statistics.scientist.gr/5.pdf>

<http://users.uoa.gr/~roussosp/>

<https://www.snpedia.com/index.php/Rs4994>

<https://www.snpedia.com/index.php/Rs1042713>

<https://www.snpedia.com/index.php/Rs1042714>

<https://www.snpedia.com/index.php/Rs1801252>

<https://www.snpedia.com/index.php/Rs1801253>

<http://www.restore.ac.uk/srme/www/fac/soc/wie/research-new/srme/modules/mod4/12/index.html>

<http://core.ecu.edu/psyc/wuenschk/MV/Multreg/Logistic-SPSS.PDF>