

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ  
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ  
ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΜΕΛΕΤΗ ΤΩΝ ΠΑΡΑΓΟΝΤΩΝ ΠΟΥ ΣΧΕΤΙΖΟΝΤΑΙ  
ΜΕ ΤΟ ΕΠΙΠΕΔΟ ΣΑΚΧΑΡΟΥ ΣΤΟ ΑΙΜΑ**

**Παυλάκης Η. Κωνσταντίνος**

*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

Μάιος 2019

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:

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

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

**UNIVERSITY OF PIRAEUS**



**DEPARTMENT OF STATISTICS  
AND INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN  
APPLIED STATISTICS**

**A STUDY OF THE FACTORS, RELATED TO THE  
LEVEL OF THE BLOOD GLUCOSE**

By

Pavlakis I. Konstantinos

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

May 2019



ΣΤΟΝ ΚΥΡΙΑΚΟ

## Ευχαριστίες

*Στο δύσβατο ακαδημαϊκό δρόμο των μαθηματικών και της στατιστικής, πολλές φορές θα βοηθήσεις και θα βοηθηθείς.*

*Κατά τη διάρκεια της πολυετούς πορείας μου στο Πανεπιστημιακό Ίδρυμα του Πειραιά, είχα την τύχη να έχω εξαιρετικούς συναδέλφους και καθηγητές για συνοδοιπόρους, οι οποίοι μέσα από ευχάριστες ή όχι τόσο στιγμές έχουν εντυπωθεί για πάντα στη μνήμη.*

*Ιδιαίτερη μνεία στον επιβλέπων καθηγητή κ. Γιώργο Τζαβελά με τον οποίο είχα μια άριστη πολύμηνη συχνή επαφή, κερδίζοντας σπάνιες εμπειρίες.*

*Κλείνοντας θα ήθελα να δείξω απερίφραστη ευγνωμοσύνη στην οικογένεια μου για την πίστη και τη στήριξη που μου έδειξαν. Εάν δεν ήταν αυτοί, η πορεία μου θα ήταν αλλιώςτικη.*

*«Η επιτυχία δεν είναι γλυκιά αν δεν έχεις υποφέρει πριν»  
Τσο Αρλάουσкас*

## Περίληψη

Οι σύγχρονες διατροφικές συνήθειες που έχουν αλλάξει άρδην μεταπολεμικά, σε συνδυασμό με την καθιστική και αγχώδη ζωή -ιδιαίτερα στα αστικά κέντρα- είναι υπεύθυνες για την αύξηση της εμφάνισης του σακχαρώδη διαβήτη στον ανεπτυγμένο κόσμο και φυσικά, στην Ελλάδα.

Στην εργασία αυτή, με τη βοήθεια ενός Πανελλαδικού δείγματος μεγέθους 1035 ατόμων εξετάζονται οι δημογραφικοί, σωματομετρικοί και διατροφικοί παράγοντες που σχετίζονται με την εμφάνιση σακχαρώδη διαβήτη. Αφού γίνει περιγραφή του δείγματος που χρησιμοποιήθηκε στην εν λόγω μελέτη, στη συνέχεια εξετάζονται με τη βοήθεια παραμετρικών και μη παραμετρικών μεθόδων οι μεταβλητές που σχετίζονται με το επίπεδο γλυκόζης στο αίμα. Κατόπιν, αφενός με τη βοήθεια του Box-Cox μετασχηματισμού βρίσκεται το βέλτιστο γραμμικό μοντέλο όταν στόχος της μελέτης είναι το επίπεδο σακχάρου στο αίμα, αφετέρου, το βέλτιστο λογιστικό μοντέλο όταν μελετάται η ύπαρξη σακχαρώδους διαβήτη. Ιδιαίτερη έμφαση δίδεται στη μεταβλητή Mediterranean Diet Score η οποία είναι ένας διατροφικός δείκτης που περιγράφει τον βαθμό κατά τον οποίο έχει υιοθετηθεί η μεσογειακή διατροφή από το εξεταζόμενο δείγμα του πληθυσμού.

## **Abstract**

Dietary habits that have changed radically the previous years, along with modern life are responsible for increasing the incidence of diabetes mellitus in the developed world and of course in Greece.

In this study, the dataset derived from a nationwide sample of 1,035 individuals and includes 75 demographic, nutritional, haematological and somatometric variables associated with the incidence of diabetes mellitus. After describing the sample used in this research, the related variables with the glucose blood levels are examined by parametric and non-parametric methods. Next, with the help of the Box-Cox transformation, the optimal linear model is found when the target of the study is the blood sugar level as well as the optimal logistic regression model when examining the presence of diabetes mellitus. Emphasis is given on the variable Mediterranean Nutrition, which is an indicator that describes the adoption strictness of the Mediterranean diet by the sample units in this study.



# Περιεχόμενα

## ΚΕΦΑΛΑΙΟ 1<sup>ο</sup> : ΕΙΣΑΓΩΓΗ

1.1	Εισαγωγή.....	12
1.2	Σακχαρώδης διαβήτης.....	13
1.3	Είδη σακχαρώδη διαβήτη, αιτίες εμφάνισης και τρόποι αντιμετώπισής τους...	14
1.3.1	Προδιαβήτης .....	15
1.3.2	Σακχαρώδης διαβήτης τύπου I.....	15
1.3.3	Σακχαρώδης διαβήτης τύπου II.....	16
1.3.4	Σακχαρώδης διαβήτης κύησης.....	18
1.4	Διαγνωστική διερεύνηση σακχαρώδη διαβήτη.....	19
1.4.1	Συμπτώματα σακχαρώδη διαβήτη.....	19
1.4.2	Διαγνωστικές εξετάσεις σακχαρώδη διαβήτη.....	20
1.5	Επιπλοκές σακχαρώδη διαβήτη.....	22
1.5.1	Χρόνιες επιπλοκές σακχαρώδη διαβήτη.....	22
1.5.2	Οξείες επιπλοκές σακχαρώδη διαβήτη.....	23
1.6	Μεσογειακή διατροφή και σακχαρώδης διαβήτης.....	24
1.6.1	Θετικές και αρνητικές επιδράσεις μεσογειακής διατροφής στο σακχαρώδη διαβήτη. Διαφορές από τη διαβητική δίαιτα.....	25
1.6.2	Διατροφικά στοιχεία μεσογειακής διατροφής.....	26

## ΚΕΦΑΛΑΙΟ 2<sup>ο</sup> : Μεθοδολογία

2.1	Έλεγχος ανεξαρτησίας ποιοτικών δεδομένων.....	27
2.1.1	Πίνακες συνάφειας.....	27
2.1.2	Έλεγχος ανεξαρτησίας $\chi^2$ .....	28
2.2	Έλεγχος τάσης.....	28
2.3	Μέτρα συνάφειας για 2x2 πίνακες συνάφειας.....	29
2.3.1	Μέτρα συνάφειας για μη διατάξιμους πίνακες συνάφειας.....	29
2.3.2	Μέτρα συνάφειας για διατάξιμους πίνακες συνάφειας.....	31
2.4	Σύγκριση μέσων για δίτιμες ποιοτικές μεταβλητές.....	31
2.5	Σύγκριση μέσων για ποιοτικές μεταβλητές με πολλά επίπεδα.....	32
2.6	Έλεγχος ανεξαρτησίας ανάμεσα σε δύο συνεχείς μεταβλητές.....	32
2.7	Γραμμική Παλινδρόμηση.....	33
2.7.1	Βασικές προϋποθέσεις γραμμικού μοντέλου.....	33
2.7.2	Οικογένεια μετασχηματισμών Box – Cox.....	34
2.8	Λογιστική Παλινδρόμηση.....	34

## ΚΕΦΑΛΑΙΟ 3<sup>ο</sup> : ΠΕΡΙΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ ΔΕΙΓΜΑΤΟΣ

3.1	Δημογραφική ανάλυση δείγματος.....	35
3.1.1	Περιγραφική ανάλυση μεταβλητής «Φύλο».....	35
3.1.2	Περιγραφική ανάλυση μεταβλητής «Ηλικία».....	36
3.1.3	Περιγραφική ανάλυση μεταβλητής «Οικογενειακή Κατάσταση»....	37
3.1.4	Περιγραφική ανάλυση μεταβλητής «Εκπαιδευτική Και Οικονομική Κατάσταση».....	38
3.2	Διατροφικά και σωματικά στοιχεία δείγματος.....	40
3.2.1	Περιγραφική ανάλυση μεταβλητής «Κατανάλωση Έτοιμου Φαγητού».....	40
3.2.2	Περιγραφική ανάλυση μεταβλητής «Κατανάλωση Καφέ».....	41
3.2.3	Περιγραφική ανάλυση μεταβλητής «Δείκτης Μάζας Σώματος».....	42
3.2.4	Περιγραφική ανάλυση μεταβλητής «Mediterranean Dietary Group».....	44
3.3	Αθλητική δραστηριότητα και καπνιστικές συνήθειες δείγματος.....	45
3.3.1	Περιγραφική ανάλυση μεταβλητής «Φυσική Δραστηριότητα».....	45
3.3.2	Περιγραφικά στατιστικά μεταβλητής «Καπνιστικές Συνήθειες».....	46
3.4	Κλινικά στοιχεία δείγματος.....	47
3.4.1	Περιγραφικά στατιστικά μεταβλητής «Γλυκόζη Αίματος».....	47
3.4.2	Περιγραφικά στατιστικά μεταβλητής «Σακχαρώδης Διαβήτης».....	49

## ΚΕΦΑΛΑΙΟ 4<sup>ο</sup> : ΣΤΑΤΙΣΤΙΚΟΣ ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ

4.1	Στατιστικός έλεγχος υποθέσεων.....	50
4.2	Στατιστικός έλεγχος δημογραφικών μεταβλητών .....	50
4.2.1	Έλεγχος ανεξαρτησίας μεταβλητής «Φύλο».....	51
4.2.2	Έλεγχος συσχέτισης μεταβλητής «Ηλικία».....	54
4.2.3	Έλεγχος ανεξαρτησίας μεταβλητής «Οικογενειακή Κατάσταση».....	58
4.2.4	Έλεγχος ανεξαρτησίας μεταβλητής «Εκπαιδευτική Και Οικονομική Κατάσταση».....	61
4.3	Στατιστικός έλεγχος διατροφικών και σωματομετρικών μεταβλητών.....	66
4.3.1	Έλεγχος συσχέτισης μεταβλητής «Σωματικού Βάρους».....	66
4.3.2	Έλεγχος συσχέτισης μεταβλητής «Δείκτης Μάζας Σώματος».....	68
4.3.3	Έλεγχος σύγκρισης μέσων όρων της μεταβλητής «Κατανάλωση Έτοιμου Φαγητού».....	72
4.3.4	Έλεγχος ανεξαρτησίας μεταβλητής «Mediterranean Dietary Group».....	77
4.4	Έλεγχος ανεξαρτησίας μεταβλητής «Καπνιστικές Συνήθειες».....	80
4.5	Στατιστικός έλεγχος μεταβλητών κλινικής εικόνας συμμετεχόντων.....	84
4.5.1	Έλεγχος σύγκρισης μέσων όρων της μεταβλητής «Καρδιαγγειακά Επεισόδια».....	84
4.5.2	Έλεγχος συσχέτισης μεταβλητής «Διαστολική Αρτηριακή Πίεση»... ..	87
4.5.3	Έλεγχος ανεξαρτησίας μεταβλητής «υπέρτασης».....	89
4.5.4	Έλεγχος ανεξαρτησίας μεταβλητής «υπερχοληστερολαιμίας».....	93
4.6	Ανακεφαλαίωση και σχόλια.....	96

## ΚΕΦΑΛΑΙΟ 5<sup>ο</sup> : ΠΑΛΙΝΔΡΟΜΗΣΗ

5.1	Τεχνικές παλινδρόμησης.....	97
5.2	Γραμμική παλινδρόμηση .....	98
5.2.1	Το γραμμικό μοντέλο.....	98
5.2.2	Έλεγχος προϋποθέσεων γραμμικού μοντέλου.....	101
5.3	Μετασχηματισμός Box – Cox.....	103
5.3.1	Έλεγχος προϋποθέσεων μετασχηματισμένου γραμμικού μοντέλου	103
5.3.2	Ανάλυση γραμμικής παλινδρόμησης μετασχηματισμένου μοντέλου Box – Cox.....	108
5.4	Ανάλυση γραμμικής παλινδρόμησης με προσαρμοσμένο μόνο το δείκτη Mediterranean Dietary Score	113
5.4.1	Έλεγχος προϋποθέσεων γραμμικού μοντέλου	113
5.4.2	Έλεγχος προϋποθέσεων μετασχηματισμένου γραμμικού μοντέλου Box - cox	115
5.4.3	Ανάλυση γραμμικής παλινδρόμησης μετασχηματισμένου γραμμικού μοντέλου Box - Cox	117
5.5	Λογιστική Παλινδρόμηση.....	119
5.5.1	Ανάλυση μοντέλου λογιστικής παλινδρόμησης χωρίς ερμηνευτικές μεταβλητές.....	119
5.5.2	Ανάλυση μοντέλου λογιστικής παλινδρόμησης με ταυτόχρονη προσαρμογή όλων των ανεξάρτητων μεταβλητών (μέθοδος enter)..	122
5.5.3	Ανάλυση μοντέλου λογιστικής παλινδρόμησης με σταδιακή προσαρμογή των ανεξάρτητων μεταβλητών (μέθοδος Forward Selection-Likelihood Ratio).....	127
5.5.4	Ανάλυση μοντέλου λογιστικής παλινδρόμησης με προσαρμοσμένο το δείκτη Mediterranean Dietary Score, ταυτόχρονη προσαρμογή όλων των ανεξάρτητων μεταβλητών (μέθοδος enter).....	133

## ΚΕΦΑΛΑΙΟ 6<sup>ο</sup> : ΣΥΜΠΕΡΑΣΜΑΤΑ

6.1	Συμπεράσματα.....	137
	Βιβλιογραφία.....	139

# ΚΕΦΑΛΑΙΟ 1<sup>ο</sup>

## ΕΙΣΑΓΩΓΗ

### 1.1 Εισαγωγή

Σακχαρώδη διαβήτη ονομάζουμε μια διαταραχή του μεταβολισμού, η οποία χαρακτηρίζεται από αυξημένο σάκχαρο (γλυκόζη) στο αίμα. Αρχικά, η γλυκόζη είναι η βασική τροφή των ανθρώπινων κυττάρων. Οι περισσότερες τροφές που καταναλώνονται, μετατρέπονται από τον ανθρώπινο οργανισμό σε γλυκόζη. Για να μπορέσει όμως η γλυκόζη να εισέλθει μέσα στα κύτταρα, είναι απαραίτητη η παρουσία της ορμόνης ινσουλίνης. Η ινσουλίνη εκκρίνεται από έναν μεγάλο αδένα που βρίσκεται πίσω από το στομάχι και ονομάζεται πάγκρεας. Όταν λοιπόν το πάγκρεας δεν παράγει καθόλου ινσουλίνη ή παράγει λίγη ποσότητα ή ακόμα και αν αυτή που παράγει δε δρα σωστά, τότε η γλυκόζη δεν μπορεί να απορροφηθεί από τα κύτταρα και επομένως παραμένει στο αίμα. Όταν οι συγκεντρώσεις αυτές είναι υψηλές και ξεπεράσουν τις φυσιολογικές τιμές, τότε εμφανίζεται ο σακχαρώδης διαβήτης.

Οι βασικοί τύποι του σακχαρώδη διαβήτη είναι δύο. Ο τύπος 1, όπου είναι και ο σπανιότερος, εμφανίζεται όταν το πάγκρεας δεν παράγει καθόλου ινσουλίνη και εμφανίζεται κυρίως στη παιδική ηλικία. Στο τύπο 2, όπου εμφανίζεται σε ποσοστό 90-95% των διαβητικών, το πάγκρεας παράγει μικρότερη ποσότητα ινσουλίνης απ' ότι η προβλεπόμενη ή αυτή που παράγει έχει μειωμένη δράση. Ο τελευταίος συνήθως εμφανίζεται στις μεγαλύτερες ηλικίες αν και τα τελευταία χρόνια εμφανίζεται και στα παιδιά. Στη κατηγοριοποίηση θα προσθέσουμε τον προδιαβήτη, όπου ουσιαστικά αποτελεί το κατώφλι του σακχαρώδη διαβήτη τύπου II και ο διαβήτης κύησης, όπου εμφανίζεται στη μητέρα κατά τη διάρκεια της εγκυμοσύνης.

Πέρα από τις ενδεδειγμένες εξετάσεις, όπου μετά από υπόδειξη του θεράποντος ιατρού μπορεί να προχωρήσει ο ασθενής για να επιβεβαιώσει εάν έχει σακχαρώδη διαβήτη ή όχι, υπάρχουν κάποια συμπτώματα που δείχνουν ότι ίσως υπάρχει σχετικό πρόβλημα. Για παράδειγμα ασυνήθιστη δίψα, συχνότερη ούρηση κατά τις βραδινές ώρες, αίσθημα κόπωσης ή και λήθαργου, αίσθημα συνεχιζόμενης πείνας, φαγούρα ή και εξανθήματα στο σώμα, πονοκεφάλους και άλλα.

Η χρόνια υπεργλυκαιμία του αίματος μπορεί να επιφέρει στον ασθενή βλάβες, δυσλειτουργίες και ανεπάρκειες διάφορων οργάνων όπως των νεφρών, των αρτηριών, των οφθαλμών, των νεύρων και της καρδιάς.

Όσον αφορά την αντιμετώπιση του σακχαρώδους διαβήτη, αρχικά θα πρέπει να ενημερώνεται ο ασθενής ποια διατροφή θα πρέπει να ακολουθεί για τον καλύτερο έλεγχο του σακχάρου και φυσικά στη χρήση ινσουλίνης από τον ίδιο, σε περίπτωση που αυτή απαιτείται. Επιπλέον επειδή πρόκειται για μια μακροχρόνια ασθένεια, θα πρέπει να γίνει προσπάθεια για μείωση ή εξάλειψη κάποιων παραγόντων που αυξάνουν τον κίνδυνο εμφάνισης των χρόνιων επιπλοκών του διαβήτη όπως το κάπνισμα. Επιπροσθέτως, ο ασθενής θα πρέπει να ελέγχει την αρτηριακή του πίεση και το επίπεδο της χοληστερόλης του ανά τακτά χρονικά διαστήματα. Από την άλλη μεριά θα πρέπει να γίνει αντιληπτό στον ασθενή ότι υπάρχουν κάποιοι άλλοι παράγοντες οι οποίοι μπορούν να βοηθήσουν στην αντιμετώπιση της ασθένειας όπως η αύξηση της σωματικής άσκησης και γενικότερα η υιοθέτηση ενός υγιεινού τρόπου ζωής.

## 1.2 Σακχαρώδης Διαβήτης

Σακχαρώδη διαβήτη ονομάζουμε την μεταβολική ασθένεια κατά την οποία παρατηρούνται αυξημένες συγκεντρώσεις σακχάρου στο αίμα και διαταραχές στο μεταβολισμό της γλυκόζης.

Οι λόγοι που εμφανίζεται σακχαρώδης διαβήτης είναι κυρίως δύο: είτε λόγω ελαττωμένης έκκρισης ινσουλίνης από τον αδένα του παγκρέατος, είτε επειδή τα κύτταρα του ανθρώπινου σώματος είναι λιγότερο ευαίσθητα στην ορμόνη ινσουλίνη.

Οι κυριότεροι τύποι σακχαρώδους διαβήτη είναι ο διαβήτης τύπου I, τύπου II και αυτός της κύησης.

Μεγάλο ποσοστό των παθόντων δεν γνωρίζει εξ αρχής ότι πάσχει από σακχαρώδη διαβήτη. Η διάγνωση μπορεί να γίνει με κλινική εξέταση από τον θεράποντα ιατρό, αναγνωρίζοντας οπτικά τα συμπτώματα της πάθησης στο σώμα του ασθενή και προχωρώντας στη συνέχεια στις κατάλληλες εξετάσεις.

Η συγκεκριμένη πάθηση θεωρείται μακροχρόνια και υπεύθυνη για μια σειρά από σοβαρές επιπλοκές όπως καρδιαγγειακές ασθένειες, χρόνια νεφρική ανεπάρκεια ή βλάβες στον αμφιβληστροειδή και στα νεφρά.

Ανάλογα τη περίπτωση προτείνονται διαφορετικοί τρόποι αντιμετώπισης της ασθένειας. Για παράδειγμα μπορεί να προταθεί μια συντηρητική αντιμετώπιση όπου ο ασθενής καλείται να ακολουθήσει ένα πρόγραμμα εκγύμνασης, υγιεινής διατροφής ή να διακόψει βλαβερές συνήθειες όπως το κάπνισμα. Φυσικά ο θεράπων ιατρός ενδέχεται να χορηγήσει στον ασθενή ινσουλίνη, την βασική ορμόνη που εκκρίνεται από το πάγκρεας και είναι υπεύθυνη για την δημιουργία σακχαρώδους διαβήτη.



**Εικόνα 1.1:** ο μπλε κύκλος είναι το σύμβολο και η 14<sup>η</sup> Νοεμβρίου έχει ορισθεί ως παγκόσμια ημέρα σακχαρώδη διαβήτη

## 1.3 Είδη Σακχαρώδη Διαβήτη, Αιτίες Εμφάνισης Και Τρόποι Αντιμετώπισής τους

Γενικά υπάρχουν δύο βασικοί τύποι σακχαρώδη διαβήτη, ο διαβήτης τύπου I και τύπου II. Πολλοί επιστήμονες όμως, δέχονται στη κατηγοριοποίηση και τον διαβήτη κύησης ή ακόμα και τον προδιαβήτη. Στο κεφάλαιο αυτό θα αναλύσουμε και θα αναφερθούμε σε όλα τα είδη διαβήτη, ποιες ηλικιακές ομάδες αφορούν ως επί το πλείστον, τους παράγοντες που συμβάλουν στην εμφάνιση τους καθώς και τους προβλεπόμενους τρόπους για την αντιμετώπισή τους.

### 1.3.1 Προδιαβήτης

Όταν τα επίπεδα σακχάρου στο αίμα είναι σαφώς υψηλότερα απ' το φυσιολογικό αλλά όχι τόσο υψηλά ούτως ώστε να θεωρείται διαβήτης τότε αυτή τη περίπτωση την ονομάζουμε προδιαβήτη. Οι περισσότεροι ασθενείς που βρίσκονται στο στάδιο του προδιαβήτη, τελικά εμφανίζουν διαβήτη τύπου II.

Ένας ασθενής ο οποίος έχει διαγνωσθεί με προδιαβήτη, ουσιαστικά βρίσκεται στο κατώφλι της πάθησης και θα πρέπει να εξετάζεται ανά τακτά χρονικά διαστήματα για διαβήτη, κάθε έξι μήνες με ένα χρόνο.

Ουσιαστικά όμως ο προδιαβήτης θα πρέπει να είναι το έναυσμα για λάβει ο ασθενής μέτρα πρόληψης και αντιμετώπισης του σακχαρώδη διαβήτη τύπου II. Ο γιατρός θα προτείνει κάποιες αλλαγές στη καθημερινότητα του ασθενούς οι οποίες θα καθυστερήσουν ή ακόμα και θα αποτρέψουν την εμφάνιση διαβήτη καθώς και των επικίνδυνων επιπλοκών που μπορεί να προκύψουν, όπως οι καρδιακές παθήσεις.

Στο στάδιο του προδιαβήτη ο γιατρός δεν χορηγεί στον ασθενή φαρμακευτική αγωγή. Η καταπολέμηση της πάθησης περιορίζεται κυρίως σε αλλαγή των διατροφικών συνηθειών του πάσχοντα. Επιπλέον ο ασθενής νουθετείται να βάλει τη σωματική άσκηση σε καθημερινή βάση.



**Εικόνα 1.2:** Πρόληψη- Διατροφή- Άσκηση

### 1.3.2 Σακχαρώδης Διαβήτης Τύπου Ι

Ο συγκεκριμένος τύπος εμφανίζεται κυρίως στις παιδικές ηλικίες αλλά μπορεί να προσβάλλει και νεαρούς ενήλικες. Γι' αυτό το λόγο παλαιότερα τον διαβήτη τύπου Ι τον ονόμαζαν *νεανικό διαβήτη* ή *ινσουλινοεξαρτώμενο*.

Το ινσουλινοεξαρτώμενος διαβήτης προκύπτει διότι ο ασθενής είναι απόλυτα εξαρτημένος από την εξωγενή χορήγηση ινσουλίνης ώστε να διατηρούνται τα επίπεδα σακχάρου του αίματος σε φυσιολογικά επίπεδα. Όσοι έχουν σακχαρώδη διαβήτη τύπου Ι χρειάζονται την ορμόνη ινσουλίνη σε καθημερινή βάση και χωρίς αυτήν θεωρείται σίγουρο ότι δεν μπορούν να επιβιώσουν.

Όσον αφορά τις αιτίες εμφάνισης του, το ανθρώπινο ανοσοποιητικό σύστημα αναγνωρίζει λανθασμένα σαν απειλή τα κύτταρα του παγκρέατος που παράγουν ινσουλίνη. Έτσι λοιπόν παράγονται μικρότερες ποσότητες της ορμόνης ή ακόμα χειρότερα, δεν παράγεται καθόλου. Τα κύτταρα του παγκρέατος που δέχονται αυτή την απρόσμενη επίθεση ονομάζονται β-κύτταρα και ο λόγος που συμβαίνει αυτό δεν είναι πλήρως γνωστός ή κατανοητός στους επιστήμονες αφού τις περισσότερες φορές πρόκειται για αυτοάνοση διαδικασία. Η πλειοψηφία των ασθενών με διαβήτη τύπου Ι ανιχνεύονται στο αίμα τους ένα ή και περισσότερα είδη αντισωμάτων τα οποία είναι υπεύθυνα και για άλλες αυτοάνοσες παθήσεις. Στις -ελάχιστες- περιπτώσεις όπου οι ασθενείς με διαβήτη τύπου Ι δεν ανιχνεύονται με αυτοαντισώματα στο αίμα τους, ο τύπος διαβήτη ονομάζεται *ιδιοπαθής σακχαρώδης διαβήτης*.

Ο διαβήτης τύπου Ι εμφανίζεται συνήθως απότομα στον ασθενή και οδηγεί κυρίως στην *ανάπτυξη κέτωσης*<sup>1</sup> ή και διαβητικής κετοξέωσης εάν δεν διαγνωστεί έγκαιρα και δεν αντιμετωπισθεί με εξωγενή χορήγηση ινσουλίνης. Παρόλα αυτά, με καθημερινή θεραπεία, με τακτική παρακολούθηση της πορείας της ασθένειας, με υγιεινή διατροφή και τακτική άσκηση, τα άτομα με διαβήτη τύπου Ι μπορούν να ζήσουν μια φυσιολογική ζωή.



**Εικόνα 1.3:** Ινσουλινοεξαρτώμενος διαβήτης

<sup>1</sup>*Κέτωση*: ονομάζουμε μια φυσιολογική διαδικασία του μεταβολισμού με σκοπό την εύρυθμη λειτουργία του οργανισμού γενικότερα. Σε υγιείς ανθρώπους εμφανίζεται όταν αυτοί ακολουθούν διατροφή χαμηλή σε υδατάνθρακες και ο οργανισμός χρησιμοποιεί τα οξέα λιπαρά για την παραγωγή ενέργειας, με αποτέλεσμα να εκκρίνονται κετόνες κυρίως από την αναπνοή, προκαλώντας χαρακτηριστική δυσσομία. Σε περιπτώσεις όμως ανθρώπων που πάσχουν από σακχαρώδη διαβήτη η διαδικασία αυτή μπορεί να γίνει επικίνδυνη, καθώς συσσωρεύονται κετόνες και αλλάζει η χημική ισορροπία του σώματος.

### 1.3.3 Σακχαρώδης Διαβήτης Τύπου II

Ο διαβήτης τύπου II είναι μια ετερογενής χρόνια νόσος που εμφανίζεται συχνότερα από οποιονδήποτε άλλο τύπο σακχαρώδους διαβήτη, αφού αφορά το 90% έως και 95% του συνόλου του διαβητικού πληθυσμού παγκοσμίως. Ο διαβητικός δε πληθυσμός υπολογίζεται περί τα 415 εκατομμύρια ανθρώπους, οπότε μπορούμε να καταλάβουμε την βαρύτητα της πάθησης από τη συχνότητα που παρατηρείται στο γενικό πληθυσμό. Στην Ελλάδα υπολογίζεται πως περίπου το 6% του συνολικού πληθυσμού πάσχει από κάποιο είδος διαβήτη.

Ο σακχαρώδης διαβήτης τύπου II αφορά κυρίως τους ενήλικες, γι' αυτό το λόγο παλαιότερα αναφερόταν και ως ο «διαβήτης των ενηλίκων». Παρόλα αυτά, τα τελευταία χρόνια έχει παρατηρηθεί ότι εμφανίζεται όλο και συχνότερα σε παιδιά και εφήβους ως επακόλουθο της καλπάζουσας παιδικής παχυσαρκίας. Τα τελευταία χρόνια έχουν διενεργηθεί αρκετές έρευνες που στόχο έχουν να διερευνήσουν σε βάθος τη συγκεκριμένη πάθηση. Οι επιστήμονες έχουν καταλήξει ότι οι παράγοντες που οδηγούν στην ολοένα και αυξανόμενη εμφάνιση της νόσου είναι η καθιστική ζωή, το κάπνισμα, οι νέες διατροφικές συνήθειες και γενικά η αλλαγή κουλτούρας των λαών που εν ολίγη καθορίζεται από τη διαβίωση σε μεγάλα αστικά κέντρα. Γι' αυτό το λόγο απαντάται συχνότερα στον ανεπτυγμένο κόσμο και δυστυχώς συνδέεται με τις κυριότερες αιτίες θανάτου παγκοσμίως. Τα συμπτώματα μπορεί να μη διαφέρουν από αυτά του τύπου I, παρόλα αυτά δεν είναι τόσο οξεία και αυτό έχει σαν αποτέλεσμα την καθυστερημένη διάγνωση της νόσου.

Στον διαβήτη τύπου II, το πάγκρεας είτε παράγει ινσουλίνη και ο οργανισμός γίνεται ανθεκτικός απέναντι στην ορμόνη (*ινσουλινοαντοχή*) είτε παράγει μειωμένες ποσότητες που δεν αρκούν. Σε πρώτη φάση το σύνδρομο χαρακτηρίζεται από ελαττωμένη ευαισθησία του σώματος στην ινσουλίνη αλλά με την πάροδο του χρόνου τα επίπεδα της ινσουλίνης μπορεί να πέσουν τόσο χαμηλά ώστε να μην αρκούν πλέον και να είναι αναποτελεσματικά. Και στις δύο περιπτώσεις τα χαμηλά επίπεδα ινσουλίνης οδηγούν αναπόφευκτα σε υψηλές συγκεντρώσεις γλυκόζης στο αίμα.

Εν αντιθέσει με τον σακχαρώδη διαβήτη τύπου I, οι περισσότεροι ασθενείς δεν εξαρτώνται από την καθημερινή εξωγενή χορήγηση ινσουλίνης για να επιβιώσουν, γι' αυτό το λόγο παλαιότερα οι γιατροί τον αποκαλούσαν *μη ινσουλινοεξαρτώμενο διαβήτη*. Βασικότερος τρόπος αντιμετώπισης της νόσου είναι η υιοθέτηση ενός υγιεινού τρόπου ζωής στο σύνολο της. Έρευνες έχουν δείξει ότι κυριότερος προδιαθεσικός παράγοντας για την ανάπτυξη σακχαρώδη διαβήτη τύπου II είναι η παχυσαρκία, ίσως γιατί τότε παράγονται διάφορες ουσίες από τον λιπώδη ιστό που αυξάνουν την αντοχή των κυττάρων στην ινσουλίνη. Έτσι εξηγείται γιατί οι γιατροί προτείνουν εναλλακτικά πλάνα διατροφής και αυξημένη σωματική δραστηριότητα για τη διατήρηση του σωματικού βάρους. Το διατροφικό πρόγραμμα θα πρέπει να είναι εξατομικευμένο για τον κάθε ασθενή δείχνοντας τη πρέπουσα σημασία στη ποσότητα των τροφών (υπολογισμός πρόσληψης θερμίδων και υδατανθράκων), στους τύπους των τροφών και τον διάστημα μεταξύ των γευμάτων. Από την άλλη μεριά, η μέτρια συστηματική άσκηση συμβάλλει στον έλεγχο των επιπέδων γλυκόζης στο αίμα, βελτιώνει την ευαισθησία των ιστών στην ινσουλίνη, μειώνει τα επίπεδα ινσουλίνης μεταγευματικά και γενικά μειώνει τον κίνδυνο καρδιαγγειακής νόσου αφού βελτιώνει την αρτηριακή πίεση. Ο συνδυασμός άσκησης και διατροφής λοιπόν θεωρείται από τους γιατρούς ως ένας πολύ βασικός παράγοντας για τον γλυκαιμικό έλεγχο στα άτομα που πάσχουν από διαβήτη τύπου II.



Επιπροσθέτως, τα συμπτώματα του διαβήτη τύπου II είναι πιο ήπια σε σχέση με του τύπου I και δεν παρατηρείται συχνά διαβητική κετοξέωση. Παρόλα αυτά επειδή ο κίνδυνος σοβαρών επιπλοκών παραμένει υψηλός, εάν κριθεί απαραίτητο χορηγούνται αντιδιαβητικά φάρμακα, ανάλογα με τις ανάγκες του κάθε ασθενή. Στις μέρες μας, υπάρχει μια τεράστια ποικιλία από θεραπείες στην φαρμακευτική αγορά, είτε πρόκειται για δισκία ή ενέσιμα σκευάσματα.

Τέλος, επιστημονικές έρευνες αναφέρουν ότι άλλοι παράγοντες που προδιαθέτουν τη νόσο είναι η ηλικία, το γονιδιακό περιβάλλον, δηλαδή εάν υπάρχει οικογενειακό ιστορικό με την ασθένεια, ο κακός ύπνος αλλά ακόμα και η εθνικότητα του ασθενούς. Όσον αφορά τον τελευταίο παράγοντα, οι επιστήμονες έχουν σαφείς ενδείξεις ότι κάποιες φυλές όπως οι κάτοικοι των νησιών του ειρηνικού, οι Ασιατοαμερικάνοι, Αφροαμερικάνοι, οι Λατίνοι και οι γηγενής Αμερικάνοι, έχουν προδιάθεση να αναπτύξουν σακχαρώδη διαβήτη.

Ασθενείς με διαβήτη τύπου II δεν θα πρέπει να εφησυχάζονται σε καμία περίπτωση, αφού τις περισσότερες φορές συνυπάρχει και με άλλες ασθένειες όπως η υπέρταση, τα υψηλά επίπεδα σε τριγλυκερίδια κ.α. Στις περιπτώσεις όπου ο διαβήτη τύπου II συνυπάρχει με τις προαναφερθέντες ασθένειες τότε αυτό ονομάζεται *μεταβολικό σύνδρομο*<sup>2</sup>.



**Εικόνα 1.4:** Όσοι πάσχουν από διαβήτη τύπου II, μπορούν να έχουν μια φυσιολογική ζωή με σωστή διατροφή

---

*μεταβολικό σύνδρομο*<sup>2</sup>: οι επιστήμονες ονομάζουν μεταβολικό σύνδρομο μια ομάδα παραγόντων καρδιαγγειακού κινδύνου όπως για παράδειγμα την υπέρταση, την ινσουλινοαντοχή, την παχυσαρκία, την δυσλιπιδαιμία κ.α. Το συγκεκριμένο σύνδρομο αφορά εκατομμύρια ανθρώπους παγκοσμίως και θα πρέπει να λαμβάνεται σοβαρά υπ' όψιν αφού οδηγεί σε αρκετές περιπτώσεις σε εμφάνιση της στεφανιαίας νόσου. Γενικά, η παχυσαρκία στη κοιλιακή χώρα, θεωρείται ιδιαίτερα σημαντικός παράγοντας για ανάπτυξη καρδιακής νόσου. Απαιτείται άμεση αντιμετώπιση του μεταβολικού συνδρόμου και δη της παχυσαρκίας, αφού ασθενείς που είναι εκτεθειμένοι στους παράγοντες που προαναφέρθηκαν, εμφανίζουν συχνά διαβήτη, στεφανιαία νόσο ή παθαίνουν κάποιο αγγειακό εγκεφαλικό επεισόδιο.

### 1.3.4 Διαβήτης Κύησης

Ο σακχαρώδης διαβήτης κύησης είναι η συγκέντρωση υψηλής γλυκόζης στο αίμα μιας εγκυμονούσας γυναίκας, εκδηλώνεται στο 7% περίπου του συνόλου των κυήσεων και θεωρείται εξαιρετικά επικίνδυνος για το έμβρυο αλλά και για την ίδια τη μητέρα.

Όπως φανερώνεται από το όνομά του, ο διαβήτης κύησης εμφανίζεται κατά τη διάρκεια της εγκυμοσύνης αλλά μπορεί να διαγνωστεί μέσω κατάλληλης εξέτασης μόνο όταν η μητέρα βρίσκεται μεταξύ της 24<sup>ης</sup> και της 28<sup>ης</sup> εβδομάδας της κύησης.

Ο σακχαρώδης διαβήτης κύησης μοιάζει αρκετά με αυτόν του τύπου II, με την έννοια ότι χαρακτηρίζεται επίσης από μειωμένη έκκριση ινσουλίνης ή και ελαττωμένη ευαισθησία των κυττάρων στην ινσουλίνη. Επιπλέον, παχύσαρκες μητέρες με καθιστικό τρόπο ζωής είναι πιο πιθανό να αναπτύξουν διαβήτη κύησης. Για να θεωρηθεί ότι ο διαβήτης είναι κύησης και όχι τύπου II, θα πρέπει η γυναίκα να μην έπασχε από την ασθένεια πριν την κύηση αλλά να της εμφανίστηκε για πρώτη φορά κατά τη διάρκεια της εγκυμοσύνης.

Στις περισσότερες περιπτώσεις με διαβήτη κύησης, μετά τη γέννηση του νεογνού, τα επίπεδα γλυκόζης στο αίμα επανέρχονται στο φυσιολογικό αλλά χρειάζεται αυξημένη προσοχή διότι μπορεί να προκαλέσει διάφορες περιγεννητικές επιπλοκές. Τα στοιχεία δείχνουν ότι ένα 30% με 40% των μητέρων θα αναπτύξουν διαβήτη τύπου II κατά τη διάρκεια της υπόλοιπης ζωής τους, οπότε ενδείκνυται οι μητέρες να εξετάζονται για προδιαβήτη ή για διαβήτη τύπου II κάθε τρία χρόνια μετά τον τοκετό. Κλείνοντας, έρευνες έχουν δείξει ότι οι γυναίκες που έφεραν στο κόσμο μωρό που ζύγιζε 4,5 κιλά και πάνω, είναι πολύ πιθανό να εμφανίσουν διαβήτη τύπου II.

Όσον αφορά τους τρόπους αντιμετώπισης του διαβήτη κύησης, λόγω της εγκυμοσύνης προτιμάται η συντηρητική θεραπεία η οποία περιλαμβάνει αλλαγές στο καθημερινό τρόπο ζωής σε συνδυασμό με ένα πλάνο διατροφής. Σε κάποιες άλλες περιπτώσεις όμως οι μητέρες πρέπει να λάβουν ινσουλίνη.

Κλείνοντας τη συζήτηση μας περί των τύπων του σακχαρώδους διαβήτη, θα θέλαμε να αναφέρουμε επιγραμματικά και κάποιες άλλες περιπτώσεις όπου μπορεί να εμφανιστεί διαβήτης όπως λόγω λήψης συγκεκριμένων φαρμάκων, λόγω γενετικά ελαττωματικών β-κυττάρων του παγκρέατος, λόγω γενετικού ελαττώματος της δράσης της ινσουλίνης, λόγω ορμονικών διαταραχών, λόγω ενδοκρινικών νοσημάτων και τέλος λόγω κάποιων συγκεκριμένων παθήσεων του παγκρέατος.



**Εικόνα 1.5:** Διαβήτης κύησης

## 1.4 Διαγνωστική Διερεύνηση Σακχαρώδη Διαβήτη

Σ' αυτό το κεφάλαιο θα συζητήσουμε για την κλινική εικόνα του ασθενή που πάσχει από κάποιο είδος σακχαρώδη διαβήτη, δηλαδή είτε για κάποια κλασικά συμπτώματα του διαβήτη στο σώμα είτε για άλλα κλινικά σημεία που παρουσιάζει. Επίσης θα αναφερθούμε στις εξετάσεις που μπορούν να γίνουν ούτως ώστε να διαγνωστεί ο σακχαρώδης διαβήτης.

### 1.4.1 Συμπτώματα Σακχαρώδη Διαβήτη

Αρχικά, ο προδιαβήτης συνήθως δεν έχει κάποιο σύμπτωμα, παρά μόνο σε λίγες περιπτώσεις συνοδεύεται από κάποια ήπια συμπτώματα όπως η δίψα ή η συχνοουρία. Ένα άλλο ανησυχητικό σημάδι εμφάνισης προδιαβήτη είναι η δημιουργία σκούρων κηλίδων στο δέρμα. Συχνότερα δε εμφανίζονται στο πίσω μέρος του λαιμού, στους αγκώνες, τα γόνατα, τις μασχάλες και γενικά στις αρθρώσεις. Η τελευταία περίπτωση που αναφέραμε ονομάζεται *μελανίζουσα ακάνθωση*.

Όσον αφορά το σακχαρώδη διαβήτη τύπου I, τα συμπτώματα μπορεί να είναι πολύ απότομα με την έννοια ότι συχνά ένα από τα πρώτα συμπτώματα είναι η διαβητική κετοξέωση, η οποία συνοδεύεται από ναυτίες, εμετούς, απώλεια συνείδησης ακόμα και κώμα εάν δεν αντιμετωπιστεί έγκαιρα.

Από την άλλη μεριά, στο σακχαρώδη διαβήτη τύπου II, η νόσος φανερώνεται σταδιακά ενώ αρχίζουν να εμφανίζονται συμπτώματα όπως η κόπωση και η αδυναμία, οι ζαλάδες καθώς και κάποια ευπάθεια σε λοιμώξεις. Σε πολλές περιπτώσεις όμως, η κλασική συμπτωματολογία του διαβήτη τύπου II λείπει παντελώς, ο ασθενής αισθάνεται υγιής και έτσι χάνεται πολύτιμος χρόνος για την καταπολέμηση της νόσου.

Τα συμπτώματα που εμφανίζονται σε κάθε είδος διαβήτη είναι πολλά όπως η πολυουρία που συμβαίνει κυρίως λόγω του ότι υπάρχει πολύ ζάχαρο στο αίμα το οποίο διαφεύγει στα ούρα παρασέρνοντας και νερό μαζί. Επακολούθως, η απώλεια νερού προκαλεί με τη σειρά της αφυδάτωση με αποτέλεσμα τη ξηροστομία και τη πολυδιψία. Επίσης σε μερικές περιπτώσεις παρατηρείται απώλεια βάρους η οποία οφείλεται στο ότι η έλλειψη ινσουλίνης δεν επιτρέπει στα κύτταρα να έχουν αρκετή ενέργεια. Έτσι ο ασθενής μπορεί να μην έχει μειώσει τις μερίδες που έτρωγε ή ακόμα και να τις έχει αυξήσει (πολυφαγία) αλλά το σωματικό του βάρος μειώνεται, χάνοντας κυρίως μυϊκό ιστό και όχι λίπος. Αυτό συμβαίνει σε μια προσπάθεια του οργανισμού να βρει τα θρεπτικά συστατικά που λείπουν από τα κύτταρα. Επιπλέον, ένα ακόμα συχνό σύμπτωμα είναι η θολή όραση λόγω συσσωρευμένης γλυκόζης στο φακό του ματιού που η οποία με τη σειρά της τραβά μαζί της νερό και «φουσκώνει» τον φακό. Άλλα συμπτώματα που μπορεί να συνοδεύονται από κάποιο είδος διαβήτη είναι η φαγούρα, οι κράμπες στα πόδια, οι μυκητιάσεις στα γεννητικά όργανα, οι πληγές αργούν να επουλωθούν, μολύνσεις του δέρματος, εμφάνιση εξανθημάτων ακόμα και συχνό μούδιασμα στα πόδια.

## 1.4.2 Διαγνωστικές Εξετάσεις Σακχαρώδη Διαβήτη

Όταν παρατηρηθούν τα συμπτώματα που περιγράψαμε παραπάνω, τότε η διάγνωση του σακχαρώδη διαβήτη δεν θεωρείται δύσκολη διαδικασία, αφού αρκεί μία μέτρηση των επιπέδων του σακχάρου του αίματος.

Ο Παγκόσμιος Οργανισμός Υγείας (ΠΟΥ) βέβαια συνιστά η διάγνωση του διαβήτη να βασίζεται τα αποτελέσματά της σε τουλάχιστον δύο από τις τρεις δοκιμασίες ελέγχου γλυκόζης αίματος όπως αναφέρονται παρακάτω.

- **Τυχαία μέτρηση γλυκόζης πλάσματος:** μπορεί να πραγματοποιηθεί οποιαδήποτε χρονική στιγμή κατά τη διάρκεια της ημέρας, χωρίς να έχει επιβληθεί κάποιος διατροφικός περιορισμός στον εξεταζόμενο. Τα αποτελέσματα μπορούν να παρουσιάσουν μεγάλες διαφοροποιήσεις ακόμα και όταν έχουν γίνει στο ίδιο άτομο με κάποιες ώρες διαφορά γιατί ουσιαστικά αντικατοπτρίζουν παράγοντες όπως η κατανάλωση τροφής και η σωματική άσκηση. Εάν τα επίπεδα σακχάρου που ελέγχθηκαν με τυχαίο τρόπο υπερβαίνουν τα 200 mg/dL, σε συνδυασμό με τα κλασικά συμπτώματα, τότε η ύπαρξη διαβήτη θεωρείται επιβεβαιωμένη.
- **Γλυκόζη πλάσματος νηστείας:** η διαφορά με την προηγούμενη δοκιμασία είναι ότι το υπό εξέταση δείγμα αίματος λαμβάνεται τουλάχιστον 8 ώρες μετά τη τελευταία λήψη οποιασδήποτε τροφής. Όταν οι τιμές της γλυκόζης νηστείας είναι μικρότερες των 100 mg/dL τότε υποδεικνύουν μια υγιή κατάσταση, ενώ τιμές μεγαλύτερες ή ίσες των 126mg/dL παραπέμπουν σε σακχαρώδη διαβήτη. Βέβαια επειδή οι λοιμώξεις, τα τραύματα σε διαδικασία επούλωσης καθώς και διάφοροι άλλοι στρεσογόνοι παράγοντες, αυξάνουν τα επίπεδα γλυκόζης στο αίμα και εάν οι παρατηρούμενες τιμές της εξέτασης είναι κοντά στα όρια, τότε συνίσταται επανάληψη της δοκιμής.
- **Δοκιμασία ανοχής γλυκόζης από το στόμα:** ουσιαστικά πρόκειται για μια δοκιμασία που καθορίζει τον τρόπο χειρισμού από τον οργανισμό ενός φορτίου γλυκόζης που λαμβάνεται από το στόμα. Η διαδικασία περιλαμβάνει αρχικά μια εξέταση γλυκόζης σε κατάσταση νηστείας και στη συνέχεια χορηγείται στον εξεταζόμενο γλυκόζη 75 γραμμαρίων διαλυμένη σε νερό. Έπειτα από δύο ώρες εξετάζεται η γλυκόζη στο αίμα και εάν η τιμή της βρεθεί μεγαλύτερη των 200mg/dL τότε αυτό αποτελεί κριτήριο για σακχαρώδη διαβήτη.

Έκτος από τις εξετάσεις που παρουσιάστηκαν παραπάνω, υπάρχουν και κάποιες συμπληρωματικές οι οποίες χρησιμοποιούνται από τους γιατρούς κυρίως για την παρακολούθηση της πορείας του σακχαρώδη διαβήτη, παρέχοντας στους ειδικούς εναλλακτικές πληροφορίες όσον αφορά τον έλεγχο του σακχάρου αλλά και άλλων παρεμφερών παθήσεων.

- **Εξέταση λιπιδαιμικού προφίλ:** οι άνθρωποι που έχουν διαγνωστεί με σακχαρώδη διαβήτη τύπου II διατρέχουν από δύο έως τέσσερις φορές μεγαλύτερο κίνδυνο να εμφανίσουν στεφανιαία νόσο απ' ότι τα υγιή άτομα. Αυτός ο εξαιρετικά αυξημένος κίνδυνος οφείλεται στα επίπεδα λιπιδίων (*δυσλιπιδαιμία*) που βρίσκονται στο αίμα, τα οποία δεν είναι σε φυσιολογικά επίπεδα. Σε τέτοιες περιπτώσεις, τα τριγλυκερίδια, η χοληστερόλη και τα λεγόμενα ελεύθερα λιπαρά οξέα απαντώνται σε μεγάλες συγκεντρώσεις στο αίμα.
- **Αναλυτική εξέταση ούρων:** σε αυτήν την εξέταση ελέγχεται το επίπεδο της γλυκόζης που βρίσκεται στα ούρα μέσω ειδικής ανάλυσης της χημικής σύστασης των ούρων.
- **Μέτρηση κετονών:** η Αμερικανική Διαβητολογική Εταιρία συστήνει τον έλεγχο κετονών στο αίμα έναντι στα ούρα για λόγους αξιοπιστίας. Υπάρχουν αρκετά διαγνωστικά εργαλεία για χρήση από τον ίδιο τον πάσχοντα και τους συγγενείς του όπου παρέχουν αξιόπιστες μετρήσεις των κετονών στο αίμα παράλληλα με τη συγκέντρωση σακχάρου. Έτσι λοιπόν μετρήσεις κάτω των 0,6 mmol/L θεωρούνται φυσιολογικές ενώ από 0,6 έως 1,5 mmol/L υποδεικνύουν ότι ο ασθενής ίσως χρειάζεται επιπλέον δόση ινσουλίνης. Τέλος, μετρήσεις άνω των 1,5 mmol/L συνεπάγονται με άμεσο κίνδυνο του διαβητικού ασθενούς από διαβητική κετοξέωση και συνίσταται η άμεση διακομιδή του στο κοντινότερο κέντρο υγείας.
- **Δείκτης γλυκοζυλιωμένης αιμοσφαιρίνης (HbA1c):** ο συγκεκριμένος δείκτης δε χρησιμοποιείται για τη διάγνωση του σακχαρώδη διαβήτη αλλά για την παρακολούθηση της πορείας του σακχάρου στο αίμα. Πρόκειται για εξέταση που ουσιαστικά αντικατοπτρίζει τη μέση τιμή του σακχάρου στο αίμα κατά τους τρεις τελευταίους μήνες. Τιμές του δείκτη κάτω του 6% θεωρούνται φυσιολογικές.
- **Εξέταση λευκοματινουρίας:** *λευκοματινουρία* σημαίνει απώλεια πρωτεΐνης στα ούρα και είναι χρήσιμη για περιπτώσεις πρώιμης νεφρικής ανεπάρκειας. Η εξέταση αυτή διενεργείται μια φορά το χρόνο και μετρήσεις κάτω των 30 θεωρούνται φυσιολογικές.

- ο **Βυθοσκόπηση**: πρόκειται για μία τακτική εξέταση που διενεργείται τουλάχιστον μια φορά το χρόνο και ανιχνεύει τη διαβητική *αμφιβληστροειδοπάθεια*, έγκαιρα προτού προκύψουν βλάβες, ορισμένες εκ των οποίων είναι μη ιάσιμες όπως η μερική ή ολική τύφλωση.

## 1.5 Επιπλοκές Σακχαρώδη Διαβήτη

Εάν δεν δοθεί η πρέπουσα σημασία στους παράγοντες που ρυθμίζουν τα επίπεδα σακχάρου στο αίμα όπως η παχυσαρκία, το κάπνισμα, η αρτηριακή πίεση, η καθιστική ζωή, η έλλειψη σωματικής άσκησης και η κακή διατροφή, τότε θα επιταχυνθεί η ανάπτυξη επιπλοκών στον ασθενή. Οι επιπλοκές του σακχαρώδη διαβήτη διακρίνονται σε χρόνιες και οξείες.

### 1.5.1 Χρόνιες Επιπλοκές Σακχαρώδη Διαβήτη

Ίσως η πιο επικίνδυνη χρόνια επιπλοκή είναι η διαβητική *μακροαγγειοπάθεια*. Ως επακόλουθο οι γιατροί προειδοποιούν ότι εμφανίζει *αρτηριοσκλήρυνση* βαριάς μορφής και προσβολή των μεσαίων και μεγάλων αρτηριών του σώματος. Εκτός από τις αρτηρίες, η αρτηριοσκλήρυνση είναι υπεύθυνη για εμφάνιση στεφανιαίας νόσου όπου με τη σειρά της είναι υπεύθυνη για στηθάγχη ή και έμφραγμα του μυοκαρδίου. Επίσης, είναι πιθανό να προκληθούν αγγειακά εγκεφαλικά επεισόδια, κυρίως ισχαιμικού τύπου καθώς και εμφάνιση περιφερικής αγγειακής νόσου.

Από την άλλη μεριά υπάρχει καταγεγραμμένη και η διαβητική *μικροαγγειοπάθεια* ως χρόνια επιπλοκή του σακχαρώδη διαβήτη. Η μικροαγγειοπάθεια είναι υπεύθυνη για τη προσβολή των προτριχοειδών και τριχοειδών αγγείων που με τη σειρά τους προκαλούν πολλές παθήσεις. Αρχικά μπορεί να προκληθεί διαβητική *νεφροπάθεια*. Πρόκειται για προσβολή του σπειράματος των νεφρών, του διάμεσου ιστού και των αγγείων, τα οποία οδηγούν σε νεφρική ανεπάρκεια. Μια ακόμα πάθηση είναι η διαβητική νευροπάθεια, η οποία σε κάποιες περιπτώσεις οδηγεί σε *μονονευροπάθεια*, *πολυνευροπάθεια* ή και νευροπάθεια του αυτόνομου νευρικού συστήματος. Σε αυτή την περίπτωση ο ασθενής αντιμετωπίζει απώλεια αισθητικότητας στα κάτω άκρα σε πρώτη φάση και στυτική δυσλειτουργία. Τέλος, μπορεί να προκληθεί διαβητική *αμφιβληστροειδοπάθεια* η οποία είναι υπεύθυνη για τον ακραίο πολλαπλασιασμό των τριχοειδών αγγείων στην επιφάνεια του αμφιβληστροειδή και υπεύθυνη για προσβολή της ωχράς κηλίδας. Η διαβητική αμφιβληστροειδοπάθεια οδηγεί σε απώλεια όρασης. Σε αυτό το σημείο αξίζει να σημειώσουμε ότι ο σακχαρώδης διαβήτης είναι η συχνότερη αιτία τύφλωσης στον ανεπτυγμένο κόσμο.

## 1.5.2 Οξείες Επιπλοκές Σακχαρώδη Διαβήτη

Οι οξείες επιπλοκές του σακχαρώδη διαβήτη χωρίζονται σε τρεις μεγάλες κατηγορίες και θεωρούνται εξαιρετικά επικίνδυνες για την υγεία του διαβητικού.

Αρχικά θα ξεκινήσουμε με την *διαβητική κετοξέωση* και το διαβητικό κώμα. Γενικά, η διαβητική κετοξέωση θεωρείται σοβαρή επιπλοκή και απαιτεί την άμεση διακομιδή του ασθενούς στο κοντινότερο νοσοκομείο. Συναντάται κυρίως σε ασθενείς που έχουν διαβήτη τύπου I και τα πρώτα σημάδια της είναι η εφίδρωση, απόπνοια ακετόνης και η αναπνοή *Kussmaul*<sup>3</sup> συνοδευόμενα από πόνους στη κοιλιακή χώρα. Σε πρώτη φάση παρατηρούνται σημάδια απώλειας συνείδησης, υπνηλίας, λήθαργου μέχρι και κώματος. Σε κάποιες περιπτώσεις έχει παρατηρηθεί υπόταση και κυκλοφορική καταπληξία.

Στη συνέχεια θα συζητήσουμε για το *υπερωσμωτικό μη κετωτικό κώμα*. Αυξημένες συγκεντρώσεις ωσμωτικότητας του πλάσματος, άνω των 350 mOsm/L, καθώς και υψηλές συγκεντρώσεις γλυκόζης στο αίμα, άνω των 600mg/dL, μαζί με απουσία κετοξέωσης, οδηγούν σε υπερωσμωτικό μη κετωτικό κώμα. Προειδοποιητικές ενδείξεις ενδεχομένως είναι η αφυδάτωση και σε ορισμένες περιπτώσεις η απώλεια συνείδησης. Το υπερωσμωτικό μη κετωτικό κώμα προκαλείται πολλές φορές από το έμφραγμα του μυοκαρδίου, τη σηψαιμία, την παγκρεατίτιδα καθώς και από πιθανή αιμορραγία από το γαστρεντερικό σύστημα. Το υπερωσμωτικό μη κετωτικό κώμα εμφανίζεται συχνότερα σε ασθενείς που πάσχουν από διαβήτη τύπου II.

Τέλος θα αναφερθούμε σε επιπλοκές που μπορεί να προκληθούν από τη θεραπεία του σακχαρώδη διαβήτη. Επιπλοκές που μπορεί να έχουν προκληθεί από λανθασμένη δοσολογία φαρμάκων, από υπερβολική σωματική άσκηση ή μειωμένη λήψη συγκεκριμένων τροφών όπως οι υδατάνθρακες. Η επιπλοκή αυτή ονομάζεται υπογλυκαιμία και είναι υπεύθυνη για απώλεια συνείδησης ακόμα και κώμα. Σ' αυτή τη περίπτωση απαιτείται άμεση πρόσληψη γλυκόζης από το στόμα εκτός και εάν ο ασθενής είναι αναισθητός οπότε θα πρέπει να του χορηγηθεί γλυκόζη ενδοφλέβια.



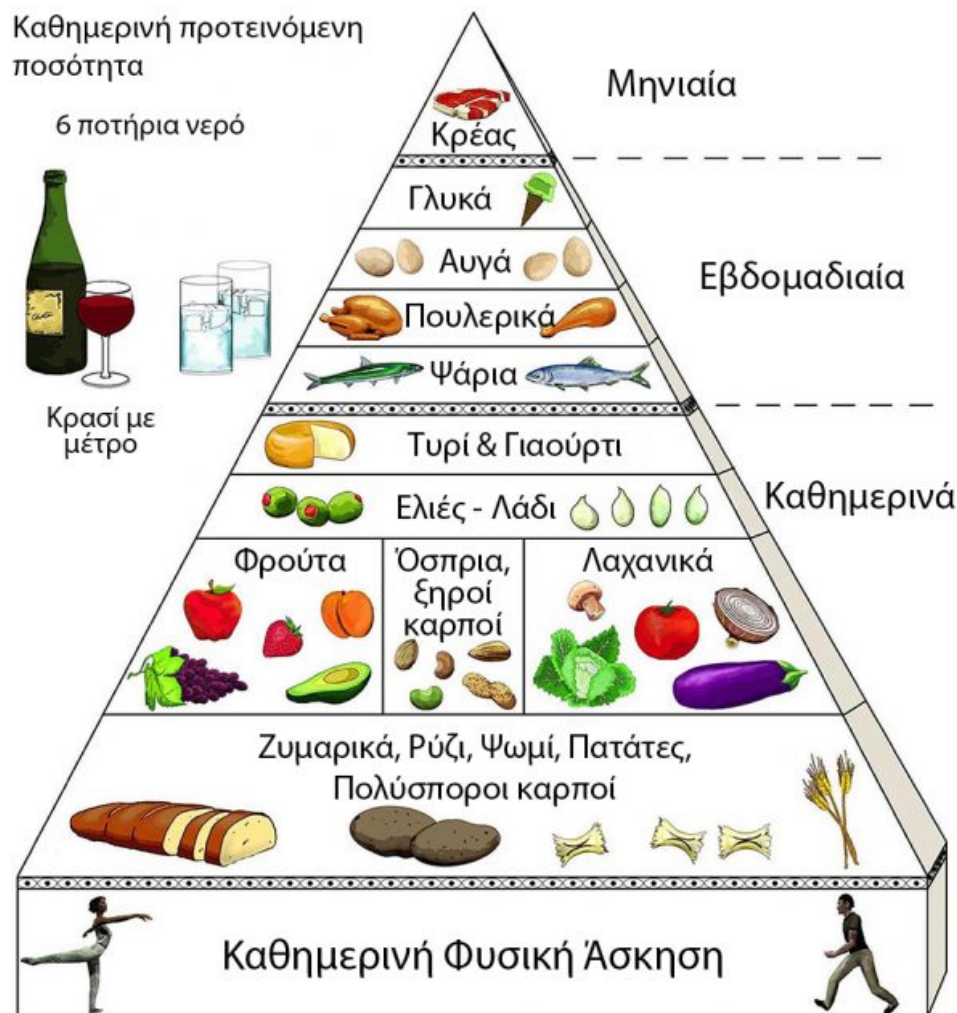
**Εικόνα 1.6:** Επιπλοκές και φαρμακευτική αγωγή για τον σακχαρώδη διαβήτη

*Kussmaul*<sup>3</sup>: πρόκειται για βαθιές αναπνοές, παρατεταμένες που συχνά συνοδεύονται από αναστεναγμού

## 1.6 Μεσογειακή διατροφή και σακχαρώδης διαβήτης

Τα τελευταία πενήντα χρόνια η μεσογειακή διατροφή έχει κεντρίσει το ενδιαφέρον των ειδικών, με αποκορύφωμα τις τελευταίες δύο δεκαετίες όπου έχει μελετηθεί σε πολλές έρευνες. Όλες στο σύνολο τους έδειξαν ότι η διατροφή των λαών της Μεσογείου είναι καλύτερη από την κλασική δίαιτα των Δυτικών χωρών.

Οι κύριοι πυλώνες της μεσογειακής διατροφής είναι η υψηλή κατανάλωση λαχανικών, φρούτων, οσπρίων, ψαριών, ξηρών καρπών, σπόρων και άλλων δημητριακών. Το βασικό λίπος για χρήση στη μαγειρική και τις σαλάτες είναι το ελαιόλαδο. Επίσης, στη μεσογειακή διατροφή καταναλώνεται συχνότερα ψάρι απ' ότι κόκκινο κρέας ή ολόπαχα τυριά και γιαούρτια. Οι λαοί που ζουν στη λεκάνη της Μεσογείου, συνηθίζουν να καταναλώνουν φρέσκα, τοπικά και εποχικά προϊόντα, ενώ η κατανάλωση αλκοόλ αρκείται κυρίως σε κρασί με τα γεύματα. Κλείνοντας, οι Μεσόγειοι χαρακτηρίζονται από τον δραστήριο καθημερινό τρόπο ζωής, λόγω συνθηκών, κουλτούρας και κλίματος.



Εικόνα 1.7: Τροφική πυραμίδα Μεσογειακής Διατροφής



### **1.6.1 Θετικές και αρνητικές επιδράσεις μεσογειακής διατροφής στο σακχαρώδη διαβήτη. Οι διαφορές από τη διαβητική δίαιτα.**

Η μεγαλύτερη σε έκταση έρευνα ξεκίνησε στις αρχές της δεκαετίας του 1960 καλύπτοντας ένα χρονικό διάστημα τριάντα ετών και συνολικά 13.000 άτομα. Διενεργήθηκε από τον Διεθνή Οργανισμό Υγείας (WHO) και το δείγμα ήταν άτομα ηλικίας 40-59 ετών και διέμεναν μόνιμα σε μια από τις Ελλάδα, Ιταλία, Γιουγκοσλαβία, Φινλανδία, Ολλανδία, Ιαπωνία και ΗΠΑ. Τα αποτελέσματα της συγκεκριμένης έρευνας έδειξαν πως οι άνθρωποι που κατοικούν στη Μεσόγειο παρουσιάζουν υψηλότερο προσδόκιμο ζωής και χαμηλότερα επίπεδα θανάτων από καρδιαγγειακές νόσους ή καρκίνους. Τα αποτελέσματα αυτά δε ήταν εντονότερα για τους κατοίκους της Ελλάδας. Σε μία άλλη πρόσφατη έρευνα που έγινε σε συνεργασία της Ιατρικής Σχολής Αθηνών και του Harvard School of Public Health, επιβεβαιώθηκαν τα θετικά της μεσογειακής διατροφής αλλά τονίστηκε και η σημασία του δραστήριου μεσογειακού τρόπου ζωής.

Όμως ποιοι είναι οι μηχανισμοί της μεσογειακής διατροφής που ενισχύουν τον οργανισμό απέναντι στο διαβήτη τύπου II και τη παχυσαρκία; Οι ειδικοί αρχικά υποστηρίζουν ότι η παρουσίας μεγάλης ποσότητας φυτικών ινών οδηγεί σε καθυστερημένη κένωση του στομαχιού. Επιπλέον, η ύπαρξη αντιοξειδωτικών, ασβεστίου, μαγνησίου στα ανθρώπινα μεταβολικά μονοπάτια καθώς και η παρουσία της αδιπνονεκτίνης συμβάλλουν στη πρόληψη της ινσουλινοαντοχής και τη γενικότερη δυσλειτουργία των β-κυττάρων.

Γνωρίζοντας οι ειδικοί ότι τα κύρια γνωρίσματα της μεσογειακής διατροφής είναι από τη μία η μειωμένη κατανάλωση λίπους, κορεσμένων λιπαρών και χοληστερόλης και από την άλλη υψηλή κατανάλωση μονοακόρεστων λιπαρών οξέων και αμυλούχων τροφίμων, κατέληξαν στο γεγονός πως παρότι η συγκεκριμένη δίαιτα είναι ιδιαίτερα ευεργετική για το μεταβολισμό των λιπιδίων, παρόλα αυτά η υψηλή περιεκτικότητα σε υδατάνθρακες δεν επιτρέπει να θεωρηθεί ιδιαίτερα ιδανική για άτομα με διαβήτη ή με υψηλό κίνδυνο εμφάνισης καρδιαγγειακής νόσου.

Με λίγα λόγια, η βασική διαφορά ανάμεσα στη μεσογειακή και τη διαβητική δίαιτα βρίσκεται στη ποσότητα των τροφών που είναι πηγές υδατανθράκων. Πιο συγκεκριμένα στη μεσογειακή διατροφή, τροφές όπως το ρύζι, το ψωμί και οι πατάτες, που είναι πηγές υψηλού γλυκαιμικού δείκτη, συστήνονται σε μεγάλες ποσότητες αφού βρίσκονται στη βάση της διατροφικής της πυραμίδας. Από την άλλη μεριά, στη δίαιτα για το διαβήτη, αυτές οι τροφές επιτρέπονται ώστε να παρέχεται στον οργανισμό η απαραίτητη ενέργεια από τους υδατάνθρακες αλλά σε τέτοιες ποσότητες ώστε να μην εκτοξεύσουν τα επίπεδα γλυκόζης στο αίμα.

## 1.6.2 Διατροφικά στοιχεία μεσογειακής διατροφής

Στη παράγραφο που ακολουθεί, θα αναφερθούμε σε ορισμένα διατροφικά στοιχεία της μεσογειακής διατροφής όπου μπορούν να διαδραματίσουν σημαντικό βοηθητικό ρόλο στη καταπολέμηση του σακχαρώδη διαβήτη.

### Φρούτα και λαχανικά

Ένα κοινό στοιχείο της διαβητικής διαίτας και της μεσογειακής διατροφής είναι η υψηλή πρόσληψη λαχανικών και φρέσκων φρούτων. Οι φυτικές ίνες που περιέχουν είναι πολύ σημαντικές διότι ρυθμίζουν το επίπεδο της γλυκόζης στο αίμα ενώ οι αντιοξειδωτικές ουσίες δρουν κατά των ελευθέρων ριζών και θωρακίζουν τον οργανισμό. Επιπλέον λειτουργούν προστατευτικά είτε απέναντι σε καρδιαγγειακές παθήσεις είτε ακόμα και σε διάφορες μορφές καρκίνου. Κλείνοντας, όσον αφορά τη κατανάλωση φρούτων από ένα διαβητικό ασθενή, θα πρέπει να ελέγχονται ενδελεχώς οι ποσότητες διότι τα φρούτα είναι πλούσια σε σάκχαρα.

### Λιπαρά ψάρια

Το ψάρι κατέχει εξέχουσα θέση στη πυραμίδα της μεσογειακής διατροφής. Το ίδιο ισχύει και στη διαβητική διαίτα αφού συστήνεται η κατανάλωση του τουλάχιστον δύο με τρεις φορές την εβδομάδα, με τάση για ακόμα συχνότερη κατανάλωση τα τελευταία χρόνια. Τα λιπαρά ψάρια όπως το ο σολομός, ο γαύρος, οι σαρδέλες, η τσιπούρα ή το σκουμπρί, είναι πηγές Ω3 λιπαρών οξέων. Τα τελευταία παρέχουν ευεργετική καρδιαγγειακή και αντιφλεγμονώδη δράση ενώ ταυτοχρόνως ρυθμίζουν τα επίπεδα των τριγλυκεριδίων, που συνήθως είναι υψηλά σε ασθενείς διαβήτη τύπου II.

### Ελαιόλαδο

Το ελαιόλαδο σα βασικό λίπος, είναι καλή πηγή μονοακόρεστων λιπαρών οξέων, αντιοξειδωτικών στοιχείων και βιταμίνης E. Το ελαιόλαδο παρέχει στον οργανισμό υψηλή προστασία από καρδιαγγειακές νόσους, μειώνει τα επίπεδα της ολικής, HDL και της LDL χοληστερόλης. Έρευνες έχουν δείξει πώς ακόμα και η βραχυχρόνια κατανάλωση ελαιόλαδου μπορεί να μειώσει πολλούς δείκτες οξειδωτικού στρες των διαβητικών ασθενών.

### Αλκοόλ

Στη μεσογειακή διατροφή συναντάμε σε καθημερινή βάση το αλκοόλ, στη μορφή του κρασιού. Οι ειδικοί επισημαίνουν πως η κατανάλωση αλκοόλης με μέτρο έχει ισχυρή παθολογική προστατευτική δράση του διαβήτη τύπου II αφού αυξάνει την ινσουλινοευαισθησία των κυττάρων. Επιπροσθέτως, η συνετή κατανάλωση αλκοόλης, δηλαδή δύο ποτήρια ημερησίως για τους άνδρες και ένα για τις γυναίκες, είναι πιο ευεργετική για τον οργανισμό από τη παντελή έλλειψή της.

# ΚΕΦΑΛΑΙΟ 2<sup>ο</sup>

## ΜΕΘΟΔΟΛΟΓΙΑ

Η φύση των δεδομένων είναι τέτοια που απαιτείται ιδιαίτερη προσοχή στη διαχείριση και ανάλυση τους. Στη παρούσα μελέτη, θα εστιάσουμε τη προσοχή μας στις διαφοροποιήσεις που προκύπτουν στις μεθόδους ανάλυσης, ανάλογα εάν οι μεταβλητές είναι συνεχείς ή ποιοτικές. Όσον αφορά τις ποιοτικές μεταβλητές, υπάρχει άλλη μια διαφοροποίηση μεταξύ τους εξίσου σημαντική για την εξέλιξη της ανάλυσης, τις κατηγορικές και τις διατακτικές.

Το σετ δεδομένων που έχουμε στη διάθεση μας περιέχει όλα τα είδη μεταβλητών.

### 2.1 Έλεγχος ανεξαρτησίας ποιοτικών δεδομένων

Η σχέση εξάρτησης μεταξύ ποιοτικών μεταβλητών, μπορεί να περιγράψει μέσω πινάκων συνάφειας. Μέσω αυτών μπορούμε να παρουσιάσουμε, να αναλύσουμε, να συγκρίνουμε και να ελέγξουμε εάν υπάρχει σχέση μεταξύ δύο ή περισσότερων ποιοτικών μεταβλητών με τη βοήθεια του  $\chi^2$  ελέγχου ανεξαρτησίας.

#### 2.1.1 Πίνακες συνάφειας

Ο πίνακας συχνοτήτων προκύπτει όταν ταξινομούνται ταυτόχρονα οι τιμές δύο ή περισσότερων ποιοτικών μεταβλητών.

Ο πιο απλός πίνακας συνάφειας είναι ο 2X2 που αποτελείται από δύο ποιοτικές μεταβλητές με δύο κατηγορίες η κάθε μία και συνολικά έχει τέσσερα κελιά. Οι στήλες του πίνακα αποτελούν τις κατηγορίες της μίας μεταβλητής και οι γραμμές της έταιρης. Με  $O_{ij}$  συμβολίζονται οι συχνότητες εμφάνισης των συνδυασμών των  $i$  γραμμών με τις  $j$  στήλες και ονομάζονται παρατηρημένες τιμές (observed values). Επιπλέον εμφανίζονται τα περιθώρια αθροίσματα των στηλών και των γραμμών.

ποιοτική μεταβλητή #1	ποιοτική μεταβλητή #2		
	E	A	Σύνολο
E	$O_{11}$	$O_{12}$	$\Sigma_{1j}$
A	$O_{21}$	$O_{22}$	$\Sigma_{2i}$
Σύνολο	$\Sigma_{i1}$	$\Sigma_{i2}$	n

**Πίνακας 2.1:** Παράδειγμα 2X2 πίνακα συνάφειας

Οι αναμενόμενες τιμές (expected values) συμβολίζονται με  $E_{ij}$  και υπολογίζονται πολλαπλασιάζοντας το άθροισμα μιας στήλης με το άθροισμα της γραμμής και διαιρώντας το με το γενικό σύνολο του δείγματος.

Οι σχετικές συχνότητες υπολογίζονται όταν διαιρούνται οι παρατηρημένες τιμές με το σύνολο του δείγματος. Εάν πάλι χρησιμοποιήσουμε τα περιθώρια αθροίσματα αντί των παρατηρημένων τιμών, τότε ονομάζονται συνολικές σχετικές συχνότητες.

Τέλος, οι σχετικές συχνότητες ανά γραμμή ή στήλη υπολογίζονται όταν διαιρεθεί η παρατηρημένη συχνότητα ενός κελιού με το σύνολο της αντίστοιχης γραμμής ή της στήλης. Πρόκειται ουσιαστικά και δεσμευμένες πιθανότητες.

### 2.1.2 Έλεγχος ανεξαρτησίας $\chi^2$

Σε ιατρικές μελέτες όπως αυτή της παρούσας εργασίας, ενδιαφερόμαστε να ερευνήσουμε εάν υπάρχει σχέση ανάμεσα σε δύο κατηγορικές μεταβλητές. Μέσω του στατιστικού ελέγχου ανεξαρτησίας  $\chi^2$  θα είμαστε σε θέση να απαντήσουμε εάν δύο ποιοτικές μεταβλητές είναι ανεξάρτητες ή όχι.

Εάν ορίσουμε ως περιθώριες πιθανότητες τις  $p_{ij} = O_{ij} / n$ ,  $p_{i1} = O_{i1} / n$  και  $p_{1j} = O_{1j} / n$ , τότε μπορούμε να ορίσουμε και μαθηματικά τις υποθέσεις του ελέγχου  $\chi^2$ . Πιο συγκεκριμένα, η μηδενική υπόθεση της ανεξαρτησίας διατυπώνεται ως  $H_0: p_{ij} = p_{i1} * p_{1j}$ .

Σ' αυτό το σημείο θα πρέπει να αναφερθούμε στο κανόνα ορθής εφαρμογής του ελέγχου  $\chi^2$  ο οποίος υποδεικνύει ότι αρχικά θα πρέπει το μέγεθος δείγματος  $n$  να είναι μεγαλύτερο του τετραπλάσιου του αριθμού των κελιών του πίνακα συνάφειας, δεύτερον να μην υπάρχουν αναμενόμενες συχνότητες μικρότερες της μονάδας και τέλος το ποσοστό των αναμενόμενων συχνοτήτων που είναι μικρότερες του πέντε, να μην υπερβαίνει το 25%.

Στις περιπτώσεις όπου το δείγμα είναι μικρό και δεν μπορούν να καλυφθούν οι παραπάνω προϋποθέσεις, τότε ο ερευνητής μπορεί να καταφύγει στον ακριβή έλεγχο του Fisher.

### 2.2 Έλεγχος τάσης

Στη περίπτωση όπου θέλουμε να μελετήσουμε δύο ποιοτικές μεταβλητές εκ των οποίων τουλάχιστον η μία είναι διατάξιμη, τότε δεν μπορούμε να εξετάσουμε εάν υπάρχει ανεξαρτησία αλλά γραμμική τάση μεταξύ των δύο μεταβλητών. Σε τέτοιες περιπτώσεις εφαρμόζεται ο στατιστικός έλεγχος  $\chi^2$  για τάση.

Η μηδενική υπόθεση ότι δεν υπάρχει γραμμική τάση μπορεί να γραφτεί μαθηματικά ως  $H_0: p_{11} = p_{12} = \dots = p_{1k}$  με  $p_{1i} = a + bx_i$  όπου  $x_i$  είναι τα σκόρ των κατηγοριών της διατάξιμης μεταβλητής. Υπάρχουν συνολικά τρεις διαθέσιμες επιλογές για σκορ. Όταν επιλέγουμε τιμές 1,2, . . . ,  $k$  για τα επίπεδα της διατάξιμης μεταβλητής, τότε αυτά ονομάζονται row scores, όταν η μεταβλητή ταξινόμησης είναι διαστηματική τότε επιλέγουμε τα κέντρα των ομάδων ως σκορ και τέλος μπορούμε να επιλέξουμε το μέσο των τάξεων (midranks) για σκορ.

## 2.3 Μέτρα συνάφειας για 2x2 πίνακες συνάφειας

Η ύπαρξη σχέσης μεταξύ δύο ποιοτικών μεταβλητών ονομάζεται συνάφεια. Για την συνάφεια έχουν ορισθεί κάποια μέτρα τα οποία υποδηλώνουν την ένταση της σχέσης. Τα μέτρα συνάφειας των κατηγορικών και διατάξιμων μεταβλητών διαφέρουν μεταξύ τους.

### 2.3.1 Μέτρα συνάφειας για μη διατάξιμους πίνακες συνάφειας

Παρακάτω θα αναπτύξουμε επιγραμματικά τα κυριότερα μέτρα συνάφειας για μη διατάξιμους πίνακες συνάφειας.

#### Ο Λόγος Συμπληρωματικών Πιθανοτήτων (Odd)

Είναι μια από τις πιο σημαντικές ποσότητες στην ανάλυση των μη διατάξιμων πινάκων συνάφειας. Ουσιαστικά είναι το πηλίκο της πιθανότητας ενός ενδεχομένου προς την πιθανότητα ενός συμπληρωματικού ενδεχομένου.

$$\text{odd} = \frac{P(\text{event})}{1 - P(\text{event})}$$

Οι τιμές που παίρνει ο λόγος συμπληρωματικών πιθανοτήτων είναι μεγαλύτερες ή ίσες του μηδενός.

Τέλος, ο λόγος συμπληρωματικών πιθανοτήτων εκφράζει το πόσο πιθανό είναι να συμβεί το ενδεχόμενο που εξετάζουμε σε σχέση με το να μη συμβεί.

#### Το Πηλίκο Πιθανοτήτων (Odds Ratio)

Είναι ίσως το πιο σημαντικό μέτρο συνάφειας και ορίζεται ως το πηλίκο αυτών που έχουν εκτεθεί σε κάποιο παράγοντα προς αυτούς που δεν εκτέθηκαν.

$$\hat{a} = \frac{O_{11}O_{22}}{O_{12}O_{21}}$$

Ουσιαστικά δεν υπάρχει κάποιος περιορισμός για τις τιμές που παίρνει το πηλίκο πιθανοτήτων.

Το πηλίκο πιθανοτήτων εκφράζει το πόσο πιθανό είναι κάποιος ο οποίος έχει εκτεθεί σε κάποιο κίνδυνο να νοσήσει, σε σχέση με κάποιον που δεν έχει εκτεθεί στο συγκεκριμένο κίνδυνο.

Ο Σχετικός Κίνδυνος (Relative Risk)

Πρόκειται για ένα ακόμα σημαντικό μέτρο συνάφειας το οποίο ορίζεται ως το πηλίκο του κινδύνου με παράγοντας προς τον κίνδυνο χωρίς παράγοντα.

$$\text{Relative risk} = \frac{\frac{p_{11}}{p_{11}+p_{12}}}{\frac{p_{21}}{p_{21}+p_{22}}} = \frac{p_{11}(p_{21}+p_{22})}{p_{21}(p_{11}+p_{12})}$$

Ουσιαστικά δεν υπάρχει κάποιος περιορισμός για τις τιμές που παίρνει ο σχετικός κίνδυνος.

Ο σχετικός κίνδυνος εκφράζει τον αυξημένο κίνδυνο να εμφανιστεί ένα ενδεχόμενο, δεδομένου ότι υπάρχει κάποιος παράγοντας.

Ο Αποδιδόμενος Κίνδυνος (Attribute Risk)

Εάν αφαιρέσουμε από τον κίνδυνο με παράγοντα αυτόν χωρίς παράγοντα τότε παίρνουμε τον αποδιδόμενο κίνδυνο και δεν υπάρχει κάποιος περιορισμός για τις τιμές που μπορεί να πάρει.

Ο Συντελεστής Φ του Pearson

Ορίζεται ως  $\Phi = \sqrt{\frac{\chi^2}{n}}$  και παίρνει τιμές από το μηδέν έως το  $\sqrt{q-1}$ .

Συντελεστής C του Pearson

Ορίζεται ως  $C = \sqrt{\frac{\chi^2}{\chi^2+n}}$  και παίρνει τιμές από το μηδέν έως το  $\sqrt{\frac{q-1}{q}}$ .

Εάν ο συντελεστής πάρει τη τιμή μηδέν τότε σημαίνει ότι οι μεταβλητές είναι μεταξύ τους ανεξάρτητες.

Συντελεστής V του Cramér

Ορίζεται ως  $V = \sqrt{\frac{\chi^2}{n \cdot \min(r-1, c-1)}}$  και παίρνει τιμές από το μηδέν έως το ένα. Εάν ο συντελεστής πάρει τη τιμή μηδέν τότε σημαίνει ότι οι μεταβλητές είναι μεταξύ τους ανεξάρτητες.

### 2.3.2 Μέτρα συνάφειας για διατάξιμους πίνακες συνάφειας

Παρακάτω θα αναπτύξουμε επιγραμματικά τα κυριότερα μέτρα συνάφειας για διατάξιμους πίνακες συνάφειας.

#### Μέτρο $\gamma$ των Goodman και Kruskal

Ορίζεται ως  $\gamma = \frac{n_C - n_D}{n_C + n_D}$  και παίρνει τιμές από το μείον ένα έως το ένα. Εάν το  $\gamma$  βρεθεί ίσο με το μηδέν υποδηλώνει την ανεξαρτησία των υπό μελέτη μεταβλητών.

#### Συντελεστής συσχέτισης του Kendall

Ορίζεται ως  $\tau = \frac{n_C - n_D}{\sqrt{(n^* - n_x)(n^* - n_y)}}$  και παίρνει τις τιμές από το μείον ένα έως το ένα. Εάν το  $\tau$  βρεθεί ίσο με το μηδέν υποδηλώνει την ανεξαρτησία των υπό μελέτη μεταβλητών.

## 2.4 Έλεγχος σύγκρισης μέσων για δίτιμες ποιοτικές μεταβλητές

Στη περίπτωση που θέλουμε να μελετήσουμε τη σχέση δύο μεταβλητών όπου η ανεξάρτητη μεταβλητή είναι δίτιμη μεταβλητή και η εξαρτημένη ή μεταβλητή απόκρισης είναι συνεχής, τότε θα ερευνήσουμε πως κινούνται οι μέσοι της  $Y$  στα δύο επίπεδα της ποιοτικής μεταβλητής.

Ο κατάλληλος παραμετρικός έλεγχος για την εκτίμηση της στατιστικής σημασίας της διαφοράς μεταξύ των δύο μέσων αλλά και τη κατασκευή των αντίστοιχων διαστημάτων εμπιστοσύνης για την διαφορά μεταξύ δύο πληθυσμών, είναι ο  $t$  (studentized  $t$ -test). Η μηδενική υπόθεση  $H_0$  του ελέγχου είναι ότι οι μέσοι είναι ίσοι.

Όπως οι περισσότεροι στατιστικοί έλεγχοι, έτσι και το  $t$ -test έχει ορισμένες προϋποθέσεις για να έχει ισχύει η συμπεραματολογία του. Πιο συγκεκριμένα, προτού προχωρήσουμε στον έλεγχο  $t$ , θα πρέπει να εξετάζονται οι υποθέσεις της κανονικότητας, της ομοσκεδαστικότητας και της ανεξαρτησία των σφαλμάτων.

Όταν μία ή περισσότερες από τις παραπάνω προϋποθέσεις απορρίπτονται, τότε μπορεί να χρησιμοποιηθεί ο αντίστοιχος μη παραμετρικός έλεγχος Sum – Rank του Wilcoxon και ο έλεγχος  $U$  των Mann – Whitney. Οι παραπάνω έλεγχοι είναι ισοδύναμοι στη πράξη γι' αυτό και παρουσιάζονται μαζί από τα στατιστικά πακέτα. Η μηδενική υπόθεση  $H_0$  των ελέγχων είναι ότι οι δύο πληθυσμοί έχουν την ίδια κατανομή ή ισοδύναμα έχουν την ίδια διάμεσο.

## 2.5 Έλεγχος σύγκρισης μέσων για ποιοτικές μεταβλητές πολλών επιπέδων

Βρισκόμενοι στη περίπτωση όπου θέλουμε να μελετήσουμε τη σχέση μεταξύ δύο μεταβλητών όπου η ανεξάρτητη είναι μια ποιοτική κατηγορική ή διατάξιμη μεταβλητή με περισσότερα επίπεδα από δύο και η εξαρτημένη συνεχής, τότε ομοίως θα ερευνήσουμε πως κινούνται οι μέσοι στα επίπεδα της εξαρτημένης μεταβλητής.

Μέσω του ελέγχου ANOVA μπορούμε να αποφανθούμε εάν υπάρχει στατιστική διαφοροποίηση των μέσων μέσα στα επίπεδα της ανεξάρτητης μεταβλητής. Η μηδενική υπόθεση  $H_0$  του ελέγχου είναι ότι οι μέσοι των διαφορετικών επιπέδων είναι ίσοι. Ο έλεγχος ANOVA έχει ακριβώς τις ίδιες προϋποθέσεις με το t-test.

Στη περίπτωση όπου δε καλύπτονται οι προϋποθέσεις της ANOVA, μπορούμε να χρησιμοποιήσουμε τον αντίστοιχο μη παραμετρικό έλεγχο των Kruskal και Wallis οι οποίοι πρότειναν έναν έλεγχο βασισμένο στη βαθμολογία (Rank) των παρατηρήσεων, χωρίς να υποθέσουμε κανονικότητα και ομοσκεδαστικότητα των πληθυσμών. Η μηδενική υπόθεση  $H_0$  του ελέγχου είναι ότι οι δύο πληθυσμοί έχουν την ίδια κατανομή.

## 2.6 Έλεγχος ανεξαρτησίας ανάμεσα σε δύο συνεχείς μεταβλητές

Σ' αυτή τη παράγραφο θα συζητήσουμε τη μέθοδο που θα ακολουθήσουμε όταν θέλουμε να εξετάσουμε την ύπαρξη ανεξαρτησίας μεταξύ δύο συνεχών μεταβλητών.

Ο συντελεστής γραμμικής συσχέτισης  $r$  του Pearson, παίρνει τιμές ανάμεσα στο διάστημα  $[-1,1]$ . Όταν το  $r$  παίρνει τιμές  $-1$  ή  $1$  τότε υποδεικνύεται τέλεια αρνητική ή θετική γραμμική συσχέτιση. Γενικά όσο η τιμή του  $r$  του Pearson τείνει προς το μηδέν, τόσο εξασθενεί η θετική /αρνητική ένταση της γραμμικής συσχέτισης.

Όπως σε όλους τους ελέγχους που έχουμε συζητήσει έως τώρα, υπάρχουν προϋποθέσεις που πρέπει να καλύπτονται για τη χρήση του συντελεστή γραμμικής συσχέτισης του Pearson. Αρχικά θα πρέπει τόσο η εξαρτημένη, όσο και η ανεξάρτητη μεταβλητή να είναι συνεχείς, να είναι γραμμικές μεταξύ τους και να ικανοποιείται η κανονικότητα του δείγματος.

Όταν μία ή περισσότερες από τις παραπάνω προϋποθέσεις δεν επιβεβαιώνονται, τότε ο ερευνητής μπορεί να καταφύγει στον αντίστοιχο έλεγχο συσχετίσεων των Kendall και Spearman. Στο στατιστικό έλεγχο συσχετίσεων των Kendall και Spearman η μηδενική υπόθεση  $H_0$  που εξετάζεται είναι ότι δεν υπάρχει συσχέτιση μεταξύ των εξαρτημένων και ανεξάρτητων μεταβλητών.



## 2.7 Γραμμική παλινδρόμηση

Όταν θέλουμε να μελετήσουμε, μέσω ενός απλού μαθηματικού μοντέλου, τη σχέση μεταξύ δύο συνεχών μεταβλητών, τότε θα πρέπει να αναζητήσουμε μια ευθεία γραμμή της μορφής  $Y = a + \beta x$  η οποία να ταιριάζει καλύτερα στο σύνολο των δεδομένων που έχουμε στη διάθεσή μας. Ουσιαστικά, ψάχνουμε να βρούμε τις κατάλληλες τιμές του σταθερού συντελεστή  $a$  και του συντελεστή της ερμηνευτικής μεταβλητής. Για την εκτίμηση των παραμέτρων  $a$  και  $\beta$  του μοντέλου, χρησιμοποιείται η μέθοδος των ελαχίστων τετραγώνων. Μέσω ενός αποδεκτού μοντέλου γραμμικής παλινδρόμησης μπορούμε να κάνουμε ασφαλείς προβλέψεις για τις τιμές της μεταβλητής απόκρισης για νέες τιμές της εξαρτημένης μεταβλητής. Η μέθοδος αυτή έχει καλή εφαρμογή σε μεγάλη ποικιλία δεδομένων. Στις μέρες μας, η γραμμική παλινδρόμηση έχει συμβάλει στην εξέλιξη της μηχανικής μάθησης. Το μοντέλο μπορεί να εξελιχθεί κατά τον ίδιο τρόπο όταν χρησιμοποιούμε περισσότερες από μια ανεξάρτητες μεταβλητές.

Βασικές παραδοχές της γραμμικής παλινδρόμησης είναι η γραμμικότητα, η κανονικότητα, η ανεξαρτησία και η ομοσκεδαστικότητα. Το συγκεκριμένο γραμμικό μοντέλο δεν μπορεί να εφαρμοστεί όταν η  $Y$  είναι δίτιμη και προκύπτουν τα προβλήματα της μη κανονικότητας και των άνισων διασπορών των σφαλμάτων.

### 2.7.1 Βασικές προϋποθέσεις γραμμικού μοντέλου

#### Έλεγχος κανονικότητας δείγματος

Υπάρχουν τουλάχιστον οκτώ διαθέσιμα κριτήρια για τον έλεγχο κανονικότητας του υπό εξέταση δείγματος. Στο πλαίσιο της συγκεκριμένης έρευνας θα αναφερθούμε σε δύο που είναι πιο αξιόπιστα και αναγνωρίσιμα.

Το γνωστότερο είναι των Kolmogorov – Smirnov, το οποίο χρησιμοποιεί την εμπειρική ή δειγματική συνάρτηση κατανομής. Βασική παραδοχή είναι ότι η παρατηρούμενη δειγματική συνάρτηση κατανομής, θα προσεγγίζει την πραγματική πληθυσμιακή που ερευνάται. Έπειτα, το κριτήριο των Shapiro – Wilk χρησιμοποιείται πολλές φορές παράλληλα από τους ερευνητές με αυτό των Kolmogorov – Smirnov εάν και από πολλούς θεωρείται πιο αξιόπιστο ειδικά όταν αποφασίζει υπέρ της υπόθεσης της κανονικότητας. Η μηδενική υπόθεση και στα δύο κριτήρια είναι ότι το δείγμα προέρχεται από κανονικό πληθυσμό.

#### Έλεγχος τυχαιότητας ή ανεξαρτησίας του δείγματος

Όταν σε ένα δείγμα μεγέθους  $n$  από κάποιο πληθυσμό, θέλουμε να ελέγξουμε εάν οι παρατηρήσεις  $x_i$  αποτελούν τυχαίο δείγμα από κάποια κατανομή  $F$ , τότε χρησιμοποιούμε το Runs Test το οποίο έχει μηδενική υπόθεση ότι το δείγμα είναι τυχαίο.

Έλεγχος ομοσκεδαστικότητας δείγματος

Ο έλεγχος του Levene για την ισότητα των διακυμάνσεων εξετάζει εάν όλες οι τυχαίες μεταβλητές έχουν την ίδια πεπερασμένη διασπορά. Η μηδενική υπόθεση του ελέγχου είναι ότι οι διακυμάνσεις είναι ίσες, δηλαδή ισχύει η ομοσκεδαστικότητα του δείγματος.

**2.7.2 Οικογένεια μετασχηματισμών Box – Cox**

Όπως αναφέραμε και στη προηγούμενη παράγραφο, υπάρχουν προϋποθέσεις για τη χρήση της γραμμικής παλινδρόμησης. Εάν δεν καλύπτονται οι συγκεκριμένες παραδοχές τότε η στατιστική συμπερασματολογία της γραμμικής παλινδρόμησης δεν ισχύει.

Στις περιπτώσεις όπου μας ενδιαφέρει η ερμηνεία του γραμμικού μοντέλου αλλά δεν αποδεχόμαστε την κανονικότητα των καταλοίπων, μπορούμε να ξεπεράσουμε αυτό το εμπόδιο χρησιμοποιώντας την οικογένεια μετασχηματισμών των Box και Cox. Ένας

τέτοιος μετασχηματισμός ορίζεται από τον τύπο: 
$$Y_t(\lambda) = \begin{cases} \frac{y_t^\lambda}{\lambda}, & \lambda \neq 0 \\ \ln y_t, & \lambda = 0 \end{cases}$$

**2.8 Λογιστική παλινδρόμηση**

Η μέθοδος της λογιστικής παλινδρόμησης χρησιμοποιείται για να περιγράψει τη σχέση της πιθανότητας ενός χαρακτηριστικού (μεταβλητή απόκρισης), με έναν ή περισσότερους παράγοντες (ερμηνευτικές μεταβλητές). Η εξαρτημένη μεταβλητή είναι δίτιμη και φανερώνει την παρουσία ή απουσία ενός χαρακτηριστικού ή ασθένειας ενώ οι ανεξάρτητες μεταβλητές μπορεί να είναι συνεχείς ή ποιοτικές. Όταν είναι συνεχείς ονομάζονται συμμεταβλητές, ενώ όταν είναι ποιοτικές, παράγοντες.

Από τη λογιστική συνάρτηση,  $p_i = \frac{e^{b_0 + b_1 x_i}}{1 + e^{b_0 + b_1 x_i}}$ , λογαριθμίζοντας μπορούμε να καταλήξουμε στο μοντέλο λογιστικής παλινδρόμησης  $\ln\left(\frac{p_i}{1-p_i}\right) = b_0 + b_1 x_i$ . Το  $\frac{p_i}{1-p_i}$  είναι ο λόγος των συμπληρωματικών πιθανοτήτων (odds ratio). Η εκτίμηση των παραμέτρων του μοντέλου της λογιστικής παλινδρόμησης γίνεται με τη μέθοδο της μέγιστης πιθανοφάνειας.

Εάν ο λόγος συμπληρωματικών πιθανοτήτων παίρνει τιμές κοντά στη μονάδα ή ισοδύναμα όταν το  $\log(\text{odds ratio})$  παίρνει τιμές κοντά στο μηδέν, τότε η ερμηνευτική μεταβλητή  $x$  δεν έχει ιδιαίτερη προβλεπτική ικανότητα.

Η ερμηνεία των συντελεστών του μοντέλου είναι ίδια με αυτών της γραμμικής παλινδρόμησης. Ο σταθερός όρος  $b_0$  είναι πρακτικά η τιμή του λογαρίθμου του λόγου των συμπληρωματικών πιθανοτήτων όταν η ανεξάρτητη μεταβλητή  $b_1$  είναι μηδέν. Ο συντελεστής  $b_1$  αποτελεί τη κλίση της ευθείας και αντιπροσωπεύει πρακτικά τη μεταβολή που θα επέλθει στο λογάριθμο του λόγου συμπληρωματικών πιθανοτήτων όταν η ερμηνευτική μεταβλητή αυξηθεί κατά μία μονάδα.

# ΚΕΦΑΛΑΙΟ 3<sup>ο</sup>

## ΠΕΡΙΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ ΔΕΙΓΜΑΤΟΣ

### 3.1 Δημογραφική ανάλυση δείγματος

Αφού αναφερθήκαμε στη μεθοδολογία που θα ακολουθήσουμε στην ανάλυση, στο παρόν κεφάλαιο θα παρουσιάσουμε μια περιγραφή του δείγματος. Αυτό θα επιτευχθεί με τη χρήση της περιγραφικής ανάλυσης ορισμένων επιλεγμένων μεταβλητών του δείγματος ώστε να μπορέσουμε να δημιουργήσουμε μια πιο σφαιρική εικόνα για τα δεδομένα.

Θα ξεκινήσουμε από την ανάλυση των δημογραφικών στοιχείων των συμμετεχόντων. Πρόκειται για μια πολύ σημαντική κατηγορία μεταβλητών που απαντάται σχεδόν σε κάθε έρευνα και μπορούμε να αντλήσουμε στοιχεία τα οποία ενδεχομένως να επηρεάζουν αυτό που ερευνάται.

Εκτός από τα δημογραφικά στοιχεία, τα δεδομένα που έχουμε στη διάθεσή μας, κωδικοποιήθηκαν, ομαδοποιήθηκαν και θα χρησιμοποιηθούν ως ανεξάρτητες μεταβλητές στη παρούσα έρευνα. Κύριος στόχος της παρούσας μελέτης είναι να ερευνήσουμε ποιες μεταβλητές από αυτές επηρεάζουν περισσότερο την εμφάνιση σακχαρώδη διαβήτη και ποιες λιγότερο.

#### 3.1.1 Περιγραφική ανάλυση μεταβλητής «Φύλο»

Όπως αναφέραμε και προηγουμένως, το φύλο των συμμετεχόντων καταγράφεται σχεδόν σε κάθε έρευνα. Στην ιατρική βιβλιογραφία δεν αναφέρεται ότι υπάρχουν ενδείξεις ότι ο σακχαρώδης διαβήτης εμφανίζεται συχνότερα σε κάποιο από τα δύο φύλα αλλά η σχέση τους θα ελεγχθεί με τη βοήθεια αυτού του δείγματος. Εκτός από το διαδικασία της ανάλυσης, είναι σημαντικό να γνωρίζουμε τη κατανομή των φύλων μέσα στο δείγμα.

Από τον πίνακα 3.1 που ακολουθεί καταλαβαίνουμε ότι δεν υπάρχει μεροληψία αφού στο σύνολο των 1035 συμμετεχόντων, 510 ήταν οι γυναίκες και 525 οι άνδρες.

Sex of Subjects					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	510	49.3	49.3	49.3
	Male	525	50.7	50.7	100.0
	Total	1035	100.0	100.0	

Πίνακας 3.1: Παρατηρούμενες συχνότητες μεταβλητής φύλου

### 3.1.2 Περιγραφική ανάλυση μεταβλητής «Ηλικία»

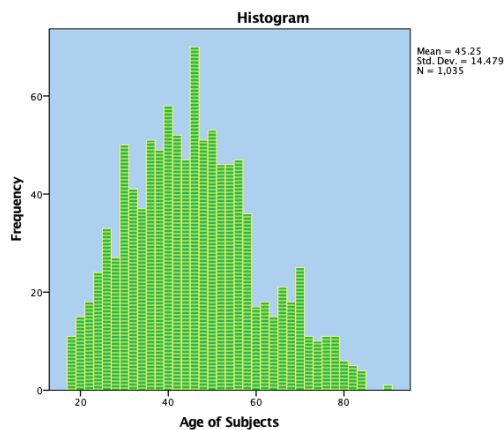
Ένας κομβικός παράγοντας που επηρεάζει την εμφάνιση του σακχαρώδη διαβήτη είναι η ηλικία. Όπως αναφέραμε και στο πρώτο κεφάλαιο, ο διαβήτης τύπου II τείνει να εμφανίζεται συχνότερα όσο μεγαλώνει ηλικιακά το άτομο αν και τα τελευταία χρόνια λόγω κακής διατροφής και του γενικότερου καθιστικού τρόπου ζωής εμφανίζεται σε όλο και νεαρότερα άτομα.

Στο πίνακα 3.2 που ακολουθεί, παρατηρούμε ότι ηλικιακά το δείγμα μας κυμαίνεται από τα 18 ως τα 89 έτη. Επιπλέον, ο μέσος όρος της ηλικίας είναι  $44,25 \pm 14,48$  και η διάμεσος του δείγματος βρέθηκε στα 45 έτη.

Descriptives				
		Statistic	Std. Error	
Age of Subjects	Mean	45.25	.450	
	95% Confidence Interval for Mean	Lower Bound	44.37	
		Upper Bound	46.14	
	Median	45.00		
	Std. Deviation	14.479		
	Minimum	18		
	Maximum	89		
	Range	71		

Πίνακας 3.2: Περιγραφικά μέτρα μεταβλητής ηλικία

Είναι προφανής ο λόγος που επιλέχθηκε για ανάλυση η συγκεκριμένη μεταβλητή αλλά πέρα από αυτό είναι σημαντικό να μελετήσουμε και τη κατανομή των ηλικιών μέσα στο δείγμα και τις συχνότητες εμφάνισής της πάθησης. Έτσι λοιπόν, με τη βοήθεια του διαγράμματος 3.1, παρατηρούμε ότι η συχνότερα εμφανιζόμενη ηλικία στο δείγμα είναι τα 45-46 έτη. Επιπροσθέτως, οι ηλικίες από 26 έως 58 έχουν τις υψηλότερες συχνότητες εμφάνισης σε αντίθεση με τις μικρότερες ή τις μεγαλύτερες.



Γράφημα 3.1: Ιστόγραμμα συχνοτήτων μεταβλητής ηλικία

### 3.1.3 Περιγραφική ανάλυση μεταβλητής «Οικογενειακή Κατάσταση»

Ο λόγος που επιλέχθηκε να ερμηνευθεί η μεταβλητή της οικογενειακής κατάστασης, είναι διότι προφανώς αλλάζουν οι διατροφικές συνήθειες, η ψυχολογική κατάσταση αλλά και γενικότερα ο καθημερινός τρόπος ζωής ενός ατόμου αναλόγως της οικογενειακής κατάστασης του. Και γνωρίζοντας ότι η διατροφή και το επίπεδο σωματικής δραστηριότητας αποτελούν σημαντικούς παράγοντες εμφάνισης ή επιδείνωσης του σακχαρώδη διαβήτη, προχωρούμε σε πρώτη φάση, περιγραφική ανάλυση της μεταβλητής οικογενειακής κατάστασης.

Οι διαθέσιμες κατηγορίες για κάθε έναν από τους συμμετέχοντες ήταν:

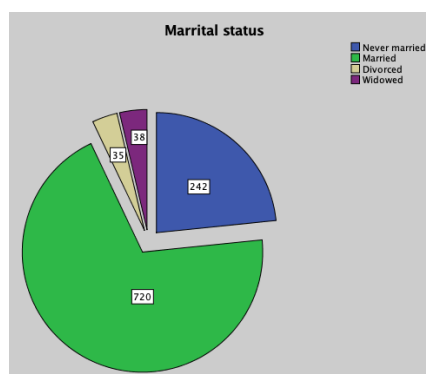
- άγαμος/η
- παντρεμένος/η
- διαζευγμένος/η
- χήρος/α

Έτσι λοιπόν από τον πίνακα 3.3 παρατηρούμε ότι από τους 1035 συμμετέχοντες, 242 απ' αυτούς ήταν άγαμοι, 720 ήταν παντρεμένοι, 35 ήταν διαζευγμένοι και 38 χήροι.

Marrital status					
		Frequenc y	Percent	Valid Percent	Cumulative Percent
Valid	Never married	242	23.4	23.4	23.4
	Married	720	69.6	69.6	92.9
	Divorced	35	3.4	3.4	96.3
	Widowed	38	3.7	3.7	100.0
	Total	1035	100.0	100.0	

Πίνακας 3.3: Παρατηρούμενες συχνότητες μεταβλητής οικογενειακής κατάστασης

Στο διάγραμμα 3.2, παρουσιάζονται οι παρατηρούμενες συχνότητες με τη μορφή ενός διαγράμματος πίτας, για να έχει και μια οπτική άποψη ο αναγνώστης για τη μεταβλητή οικογενειακή κατάσταση.



Γράφημα 3.2: Διάγραμμα πίτας μεταβλητής οικογενειακής κατάστασης

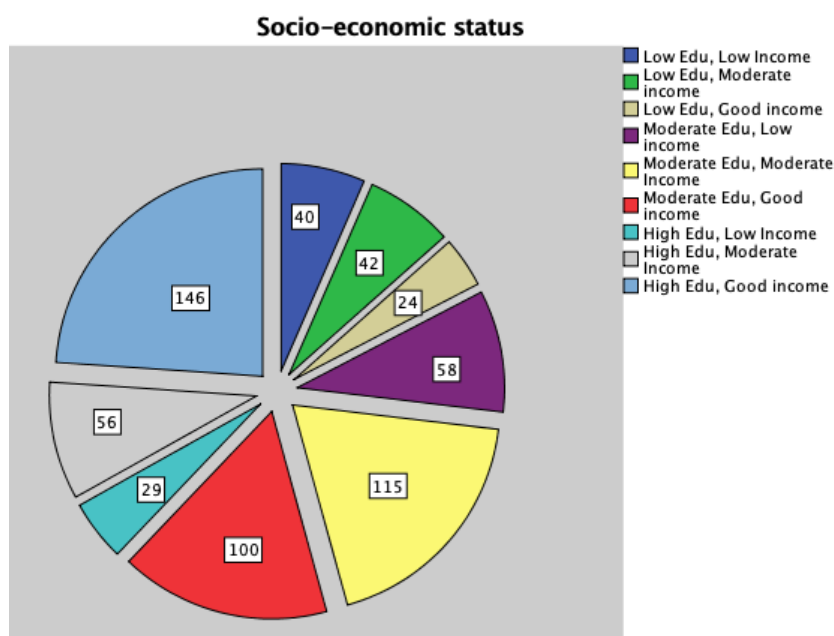
### 3.1.4 Περιγραφική ανάλυση μεταβλητής «Εκπαιδευτική και Οικονομική Κατάσταση»

Τις περισσότερες φορές, στη δημογραφική ανάλυση του ερωτώμενου, ενδιαφερόμαστε να υπάρχουν ερωτήσεις που να μας επιτρέπουν να τον κατατάσσουμε οικονομικά ή μορφωτικά σε κάποια κατηγορία.

Στη μελέτη μας είναι πολύ σημαντικός παράγοντας αν αναλογιστεί κανείς πως ακολουθείται διαφορετικός τρόπος ζωής ανάλογα με το διαθέσιμο εισόδημα. Πιο συγκεκριμένα, ενδεχομένως επηρεάζεται η ψυχολογία του ατόμου σε δύσκολες οικονομικές καταστάσεις. Επιπλέον, μπορεί να ακολουθείται διαφορετική καθημερινή διατροφή, εάν θεωρήσουμε πως οι ποιοτικές τροφές είναι ακριβότερες όπως το ψάρι, το βουβαλίσιο κρέας ή τα βιολογικά λαχανικά. Επιπροσθέτως, η οικονομική κατάσταση πιθανώς διαδραματίζει σοβαρό ρόλο στο κομμάτι της σωματικής άσκησης, με την έννοια πως ένα άτομο με καλή οικονομική κατάσταση μπορεί να επιλέξει ανάμεσα σε αναρίθμητους τρόπους άσκησης, συμβατικής ή μη. Συμβατική γυμναστική θεωρείται το περπάτημα, το ποδήλατο, το γυμναστήριο κ.α. Κλείνοντας, σε αυτό το σημείο αξίζει να αναφέρουμε το ενδεχόμενο κάποιος να αναβάλει ή ακόμα να μην έχει τη δυνατότητα προληπτικών εξετάσεων εάν το οικογενειακό του εισόδημα είναι χαμηλό. Μπορούμε να καταλάβουμε τη σπουδαιότητα αυτού του ενδεχομένου εάν σκεφτούμε πως η πρόληψη και η έγκαιρη διάγνωση των ασθενειών είναι πολύ σημαντική στη σύγχρονη ιατρική.

Στο πίνακα 3.4 παρουσιάζονται οι κατηγορίες που δημιουργήθηκαν για να μπορέσουμε να κατατάξουμε όσο το δυνατόν ακριβέστερα τους συμμετέχοντες στις εκπαιδευτικοοικονομικές κατηγορίες και οι αντίστοιχες συχνότητες εμφάνισής τους. Το εκπαιδευτικό πεδίο θεωρήθηκε καλό να ληφθεί υπ' όψιν διότι σε θεωρητικό επίπεδο όσο περισσότερο μορφωμένος είναι κάποιος, τόσο πιο πιθανό είναι να έχει υψηλότερο εισόδημα.

Κλείνοντας, θα θέλαμε να αναφέρουμε πως 425 από τους 1035 συμμετέχοντες δεν θέλησαν να απαντήσουν στη συγκεκριμένη ερώτηση κατάταξης με οικονομικά και εκπαιδευτικά κριτήρια.



**Γράφημα 3.3:** Διάγραμμα πίτας μεταβλητής εκπαιδευτικής και οικονομικής κατάστασης

Socio-economic status					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Low Edu, Low Income	40	3.9	6.6	6.6
	Low Edu, Moderate income	42	4.1	6.9	13.4
	Low Edu, Good income	24	2.3	3.9	17.4
	Moderate Edu, Low income	58	5.6	9.5	26.9
	Moderate Edu, Moderate Income	115	11.1	18.9	45.7
	Moderate Edu, Good income	100	9.7	16.4	62.1
	High Edu, Low Income	29	2.8	4.8	66.9
	High Edu, Moderate Income	56	5.4	9.2	76.1
	High Edu, Good income	146	14.1	23.9	100.0
	Total	610	58.9	100.0	
Missing	System	425	41.1		
Total		1035	100.0		

Πίνακας 3.4: Παρατηρούμενες συχνότητες μεταβλητής εκπαιδευτικής και οικονομικής κατάστασης

Στο γράφημα 3.3 παρουσιάζονται οι αντίστοιχες συχνότητες εμφάνισης των οικονομικών και εκπαιδευτικών κατηγοριών σε ένα διάγραμμα πίτας, ώστε ο αναγνώστης να έχει και μια οπτική άποψη των δεδομένων.

### 3.2 Διατροφικά Στοιχεία Δείγματος

Σε μια μελέτη για τον σακχαρώδη διαβήτη, μια μεταβολική ασθένεια που κρύβει πολλούς κινδύνους για τον ασθενή, δεν θα μπορούσαν να λείπουν ερωτήσεις σχετικά με τις διατροφικές συνήθειες του συμμετέχοντα.

Όπως αναλύσαμε διεξοδικά στο πρώτο κεφάλαιο, για κάθε τύπο σακχαρώδη διαβήτη, οι γιατροί συστήνουν ένα ισορροπημένο και υγιεινό διατροφικό πλάνο. Στο πλαίσιο αυτό θα προσπαθήσουμε να αναλύσουμε περιγραφικά τις μεταβλητές που αναφέρονται σε διατροφικές συνήθειες των συμμετεχόντων.

#### 3.2.1 Περιγραφική ανάλυση μεταβλητής «Κατανάλωση Έτοιμου Φαγητού»

Στη μελέτη για τον σακχαρώδη διαβήτη προσθέσαμε μια μεταβλητή που αναφέρεται στον αριθμό των μερίδων που τρώει ο συμμετέχοντας σε ταχυφαγεία, ταβέρνες, καντίνες και γενικά σε επιχειρήσεις τέτοιου είδους. Αυτό μπορεί να μας βοηθήσει να κατατάξουμε τον ερωτώμενο ανάλογα με την διατροφή που ακολουθεί, πόσο υγιεινή ή όχι είναι, το επίπεδο των λιπαρών που μπορεί να δέχεται καθώς και το ποιοτικό επίπεδο των τροφών που προτιμά.

Στο πίνακα 3.5 παρουσιάζονται οι συχνότητες που οι συμμετέχοντες τρώνε έτοιμο φαγητό τρώγοντας έξω ή κάνοντας παραγγελία στο σπίτι. Οι πλειοψηφία παρατηρούμε πως μαγειρεύει σε καθημερινή βάση στο σπίτι ή τρώει έτοιμο φαγητό μόνο μια φορά τη βδομάδα.

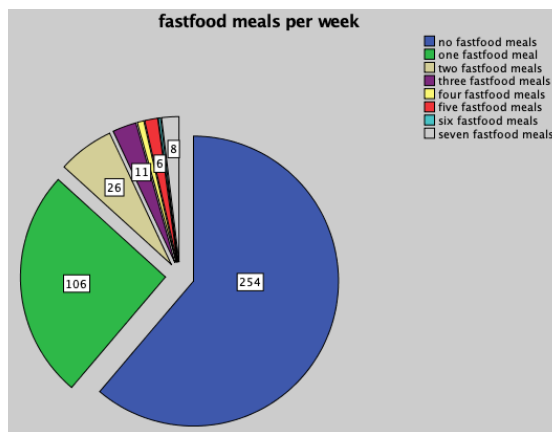
Βέβαια, 620 από τους 1035 συμμετέχοντες δεν απάντησαν σε αυτή την ερώτηση που αφορά τις διατροφικές συνήθειες. Αυτό συνέβη κυρίως επειδή αποφασίστηκε να εισέλθει αυτή η μεταβλητή στην έρευνα καθυστερημένα, οπότε οι πρώτοι συμμετέχοντες δεν την απάντησαν καν, και λιγότερο επειδή δεν ήθελαν να απαντήσουν οι ερωτώμενοι.

fastfood meals per week					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	no fastfood meals	254	24.5	61.2	61.2
	one fastfood meal	106	10.2	25.5	86.7
	two fastfood meals	26	2.5	6.3	93.0
	three fastfood meals	11	1.1	2.7	95.7
	four fastfood meals	3	.3	.7	96.4
	five fastfood meals	6	.6	1.4	97.8
	six fastfood meals	1	.1	.2	98.1
	seven fastfood meals	8	.8	1.9	100.0
	Total	415	40.1	100.0	
Missing	System	620	59.9		
	Total	1035	100.0		

Πίνακας 3.5: Παρατηρούμενες συχνότητες μεταβλητής κατανάλωση έτοιμου φαγητού



Στο γράφημα 3.4 παρουσιάζονται σε ένα διάγραμμα πίτας οι αντίστοιχες συχνότητες εμφάνισης των φορών που συνηθίζει να τρώει έτοιμο φαγητό ο συμμετέχοντας σε μια εβδομάδα, ώστε ο αναγνώστης να έχει και μια οπτική άποψη των δεδομένων.



Γράφημα 3.4: Διάγραμμα πίτας μεταβλητής κατανάλωσης έτοιμου φαγητού

### 3.2.2 Περιγραφική ανάλυση μεταβλητής «Κατανάλωση Καφέ»

Ο λόγος που προστέθηκε αυτή η μεταβλητή στην έρευνα είναι ότι υπάρχουν ισχυρές ενδείξεις ότι η κατανάλωση καφέ μειώνει την πιθανότητα εμφάνισης διαβήτη τύπου II καθώς και ότι ελαττώνει την ινσουλινοαντοχή των κυττάρων των διαβητικών.

Πιο συγκεκριμένα, μια πρόσφατη δανέζικη μελέτη έδειξε ότι η κατανάλωση τριών έως τεσσάρων μερίδων καφέ εσπρέσο, μειώνει τον κίνδυνο εμφάνισης διαβήτη τύπου II κατά 25%, σε σχέση με όσους πίνουν ένα φλιτζάνι ή λιγότερο. Ο λόγος που συμβαίνει αυτό σύμφωνα με την έρευνα είναι λόγω της καφεστόλης που υπάρχει στον καφέ. Η καφεστόλη είναι μια ουσία που βρίσκεται στον καφέ και διαφέρει από την καφεΐνη. Η καφεστόλη όμως μπορεί να εμποδιστεί να περάσει στο φλιτζάνι εάν υπάρχει φίλτρο, γι' αυτό το λόγο κυρίως ο εσπρέσο μπορεί να βοηθήσει ενάντια στο διαβήτη.

Στο ίδιο μήκος κύματος κινήθηκε και μια ολλανδική έρευνα για τον διαβήτη και τον καφέ. Επρόκειτο για μια πολύ μεγάλη έρευνα για τον διαβήτη όπου πήραν μέρος 17.000 άτομα ηλικία 30 έως 60 ετών. Το συμπέρασμα ήταν ότι όσοι κατανάλωναν τουλάχιστον επτά καφέδες την ημέρα είχαν 50% λιγότερο κίνδυνο να προσβληθούν από διαβήτη τύπου II, σε σχέση με αυτούς που έπιναν δύο φλιτζάνια καφέ την ημέρα ή και λιγότερους. Στη συγκεκριμένη έρευνα βρέθηκε ότι εκτός από τη καφεστόλη, υπάρχουν και άλλες ουσίες στον καφέ οι οποίες μπορούν να μειώσουν την αντοχή των κυττάρων στην ινσουλίνη. Τέτοιες ουσίες είναι το κάλλιο, το χλωρογενικό οξύ και το μαγνήσιο.

Στο πίνακα 3.6 παρατηρούμε ότι η πλειοψηφία των συμμετεχόντων, το 85% περίπου, πίνει έστω ένα φλιτζάνι καφέ την ημέρα.

coffee utilization					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No coffee drinking	154	14.9	15.4	15.4
	coffee drinking	849	82.0	84.6	100.0
	Total	1003	96.9	100.0	
Missing	System	32	3.1		
Total		1035	100.0		

Πίνακας 3.6: Παρατηρούμενες συχνότητες μεταβλητής κατανάλωση καφέ

### 3.2.3 Περιγραφική ανάλυση μεταβλητής «Δείκτης Μάζας Σώματος»

Εάν θέλαμε να δημιουργήσουμε μια μεταβλητή η οποία να συμπεριλαμβάνει τις διατροφικές συνήθειες αλλά και τη σωματική άσκηση, τότε η πλέον κατάλληλη μεταβλητή είναι ο Δείκτης Μάζας Σώματος, γνωστός και ως BMI.

Ο Δείκτης Μάζας Σώματος υπολογίζεται διαιρώντας το βάρος (σε κιλά) με το τετράγωνο του ύψους (σε μέτρα). Θεωρείται η πλέον κατάλληλη μεταβλητή για να καταλήξουμε εάν άτομο είναι κανονικό, λιποβαρές ή παχύσαρκο.

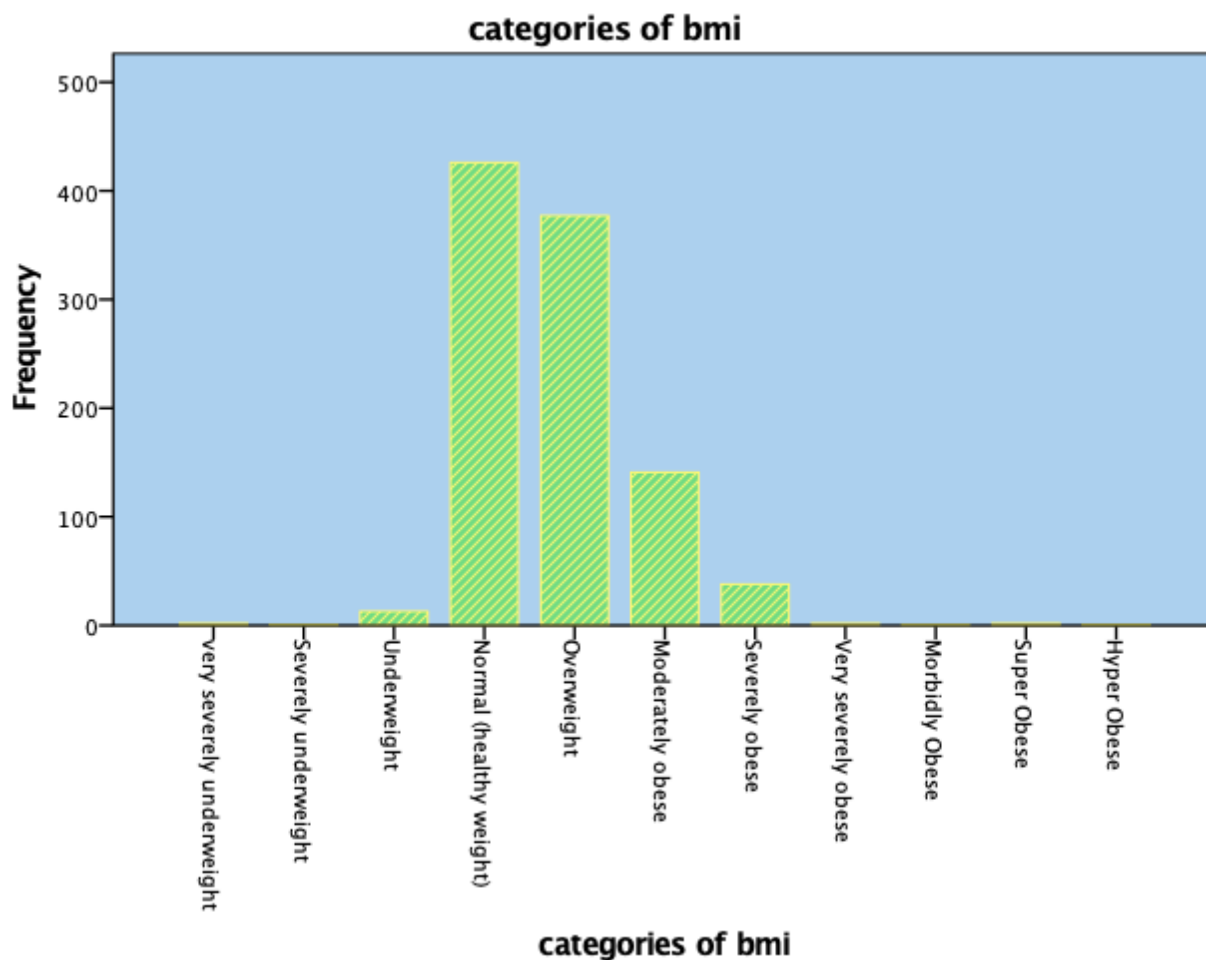
Στο πίνακα 3.7 παρουσιάζονται οι κατηγορίες του Δείκτη Μάζας Σώματος καθώς και πόσοι συμμετέχοντες αντιστοιχούν σε κάθε μια από αυτές. Παρατηρούμε λοιπόν ότι η πλειοψηφία κινείται από φυσιολογικό σωματότυπο έως αρκετά σοβαρή παχυσαρκία.

categories of bmi					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very severely underweight	2	.2	.2	.2
	Severely underweight	1	.1	.1	.3
	Underweight	13	1.3	1.3	1.6
	Normal (healthy weight)	426	41.2	42.4	44.0
	Overweight	377	36.4	37.5	81.6
	Moderately obese	141	13.6	14.0	95.6

Severely obese	38	3.7	3.8	99.4
Very severely obese	2	.2	.2	99.6
Morbidly Obese	1	.1	.1	99.7
Super Obese	2	.2	.2	99.9
Hyper Obese	1	.1	.1	100.0
Total	1004	97.0	100.0	
Missing System	31	3.0		
Total	1035	100.0		

Πίνακας 3.7: Παρατηρούμενες συχνότητες μεταβλητής Δείκτης Μάζας Σώματος

Στο γράφημα 3.5 παρουσιάζονται σε ένα ιστόγραμμα οι αντίστοιχες συχνότητες εμφάνισης των κατηγοριών του Δείκτη Μάζας Σώματος, ώστε ο αναγνώστης να έχει και μια οπτική άποψη των δεδομένων.



Γράφημα 3.5: Ιστόγραμμα συχνότητων μεταβλητής Δείκτη Μάζας Σώματος

### 3.2.4 Περιγραφική ανάλυση μεταβλητής «Med Diet Group»

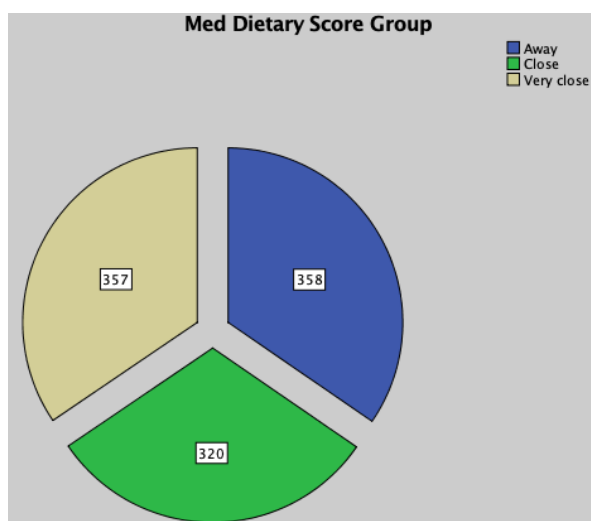
Όπως αναλύσαμε και στο πρώτο κεφάλαιο της παρούσας εργασίας, η μεσογειακή διατροφή μπορεί να μειώσει υπό προϋποθέσεις τον κίνδυνο εμφάνισης διαβήτη τύπου II.

Στο πλαίσιο αυτό, εισήχθη στη μελέτη ο Δείκτης Μεσογειακής Διατροφής, γνωστός και ως Mediterranean Dietary Score. Πρόκειται για μια ποσοτική μεταβλητή που είναι αποτέλεσμα του αθροίσματος των τιμών επιμέρους μεταβλητών. Με βάση τις τιμές που παίρνει η μεταβλητή μπορεί να γίνει καταμερισμός των ατόμων αναλόγως πόσο κοντά ή πόσο μακριά βρίσκεται η διατροφή τους από το Μεσογειακό διατροφικό πρότυπο (Med Dietary Score Group). Ο δείκτης Mediterranean Dietary Score αναπτύχθηκε από τον Παναγιωτάκο, τον Πίτσαβο και τους συνεργάτες του, έχει έντεκα συνιστώσες ομάδες τροφίμων και ο τρόπος βαθμονόμησης γίνεται σε πέντε κατηγορίες. Κάθε συνιστώσα βαθμολογείται από το 0 έως το πέντε.

Στο πίνακα 3.8 παρουσιάζονται τα τρία επίπεδα της μεταβλητής Med Dietary Score Group, η διατροφή του συμμετέχοντα δεν συμπίπτει με τη μεσογειακή διατροφή, η διατροφή του συμμετέχοντα έχει κάποια κοινά με τη μεσογειακή διατροφή και τέλος η διατροφή του συμμετέχοντα συμπίπτει με τη μεσογειακή διατροφή καθώς και πόσοι ερωτώμενοι ανήκουν στην εκάστοτε κατηγορία. Παρατηρούμε ότι οι συμμετέχοντες στη μελέτη είναι σχεδόν μοιρασμένοι ισομερώς στα τρία επίπεδα της μεταβλητής.

Med Dietary Score Group					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Away	358	34.6	34.6	34.6
	Close	320	30.9	30.9	65.5
	Very close	357	34.5	34.5	100.0
	Total	1035	100.0	100.0	

Πίνακας 3.8: Παρατηρούμενες συχνότητες μεταβλητής Mediterranean Dietary Score Group



Γράφημα 3.6: Διάγραμμα πίτας μεταβλητής Mediterranean Dietary Score Group

### 3.3 Αθλητική δραστηριότητα και καπνιστικές συνήθειες δείγματος

Μια αρχαία παροιμία λέει «νοῦς ὑγίης ἐν σώματι ὑγιεῖ». Εἴτε πρόκειται για ασθενή ή όχι, η σωματική άσκηση είναι απαραίτητη στον ανθρώπινο οργανισμό. Τα οφέλη είναι αναρίθμητα. Από μυοσκελετικά, ορμονικά, ψυχολογικά μέχρι αιματολογικά.

Στο σακχαρώδη διαβήτη δύο παράγοντες μπορεί να διαδραματίσουν σοβαρό ρόλο στην εμφάνιση, εξέλιξη και καταπολέμηση της ασθένειας. Η διατροφή, που αναλύθηκε προηγουμένως, και ο υγιεινός καθημερινός τρόπος ζωής.

Με τον όρο υγιεινό καθημερινό τρόπο ζωής οι ειδικοί εννοούν οι πάσχοντες και οι εν δυνάμει ασθενείς, να υιοθετήσουν ένα πιο ενεργητικό καθημερινό προφίλ που να περιλαμβάνει σωματική δραστηριότητα αλλά και αποφυγή βλαβερών συνηθειών όπως το κάπνισμα.

Το κάπνισμα δεν θεωρείται καθαυτό υπεύθυνο για την εμφάνιση οποιαδήποτε μορφής διαβήτη αλλά συνδέεται με αυτόν με πολλούς τρόπους. Από την άλλη μεριά, η σωματική άσκηση μπορεί να μετριάσει την ινσουλινοαντοχή των κυττάρων και να χαμηλώσει τα επίπεδα λίπους στο σώμα που οδηγούν σε διαβήτη τύπου ΙΙ. Επομένως αυτοί οι δύο παράγοντες δεν θα μπορούσαν να λείπουν από μια μελέτη για τον σακχαρώδη διαβήτη.

#### 3.3.1 Περιγραφική ανάλυση μεταβλητής «Φυσική Δραστηριότητα»

Για να μπορέσουμε να διαχωρίσουμε τους συμμετέχοντες ανάλογα με το επίπεδο σωματικής δραστηριότητας που ακολουθούν, τους αφήσαμε να επιλέξουν ανάμεσα από τέσσερις σχετικές κατηγορίες.

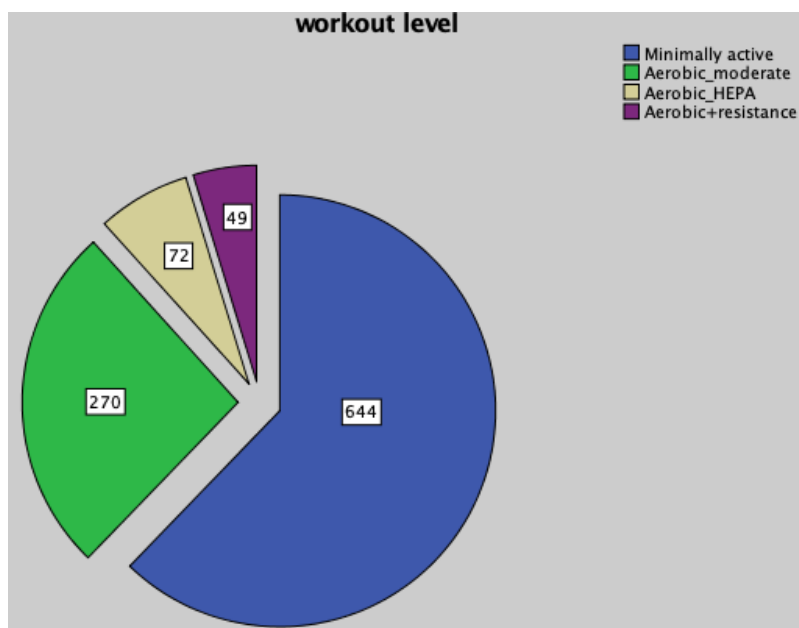
Όλο το δείγμα, και οι χίλιοι συμμετέχοντες, απάντησαν τη συγκεκριμένη ερώτηση, με την πλειοψηφία να θεωρεί ότι ασκείται ελάχιστα ή μέτρια. Μόνο το 12% περίπου θεωρεί ότι η σωματική του καθημερινή δραστηριότητα είναι εντονότερη του μετρίου.

Στο πίνακα 3.9 παρουσιάζονται τα τέσσερα επίπεδα της μεταβλητής φυσικής δραστηριότητας καθώς και πόσοι συμμετέχοντες τις επέλεξαν.

workout level					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Minimally active	644	62.2	62.2	62.2
	Aerobic_moderate	270	26.1	26.1	88.3
	Aerobic_HEPA	72	7.0	7.0	95.3
	Aerobic+resistance	49	4.7	4.7	100.0
	Total	1035	100.0	100.0	

Πίνακας 3.9: Παρατηρούμενες συχνότητες μεταβλητής φυσικής δραστηριότητας

Στο γράφημα 3.7 παρουσιάζονται σε ένα διάγραμμα πίτας οι αντίστοιχες συχνότητες εμφάνισης των συμμετεχόντων στις κατηγορίες της φυσικής δραστηριότητας, ώστε ο αναγνώστης να έχει και μια οπτική άποψη των δεδομένων.



Γράφημα 3.7: Διάγραμμα πίτας μεταβλητής φυσική δραστηριότητα

### 3.3.2 Περιγραφικά στατιστικά μεταβλητής «Καπνιστικές Συνήθειες»

Το κάπνισμα είναι ένας πολύ σοβαρός επιβαρυντικός παράγοντας για τον ανθρώπινο οργανισμό. Συνδέεται άμεσα με αρκετές ασθένειες και επιπλοκές.

Έτσι και στο σακχαρώδη διαβήτη. Το κάπνισμα δεν θεωρείται σε καμία περίπτωση σύμμαχος του διαβητικού ασθενή διότι συνδέεται με αρκετές καρδιαγγειακές παθήσεις και γενικότερα με τον σύγχρονο τρόπο ζωής που επιτάσσει λιγότερη σωματική άσκηση και ποιοτικά χειρότερη διατροφή. Γι' αυτό λοιπόν στη παρούσα μελέτη για τον σακχαρώδη διαβήτη οι συμμετέχοντες κλήθηκαν να απαντήσουν εάν καπνίζουν ή όχι.

Στο πίνακα 3.10 παρατηρούμε ότι περισσότεροι από τους μισούς δεν καπνίζουν.

smoking				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid No smoking	597	57.7	57.7	57.7
smoking	438	42.3	42.3	100.0
Total	1035	100.0	100.0	

Πίνακας 3.10: Παρατηρούμενες συχνότητες μεταβλητής καπνιστικές συνήθειες

### 3.4 Κλινικά στοιχεία δείγματος

Αφού αναλύσαμε τους παράγοντες όπου μπορούν να οδηγήσουν σε διαβήτη καθώς και αυτούς που αποτρέπουν ή καθυστερούν την εμφάνισή του, σειρά έχει να μελετήσουμε ένα σπουδαίο κλινικό παράγοντα όπου να μας υποδεικνύει με σχετική ασφάλεια την ύπαρξη διαβήτη ή όχι. Η κλινική εικόνα ενός ατόμου, δηλαδή τα σωματικά συμπτώματα και τα αποτελέσματα των εξετάσεων είναι τα εργαλεία που έχει ο γιατρός στα χέρια του για να μπορέσει να βοηθήσει τον ασθενή.

Σ' αυτή την ενότητα θα ερευνήσουμε πόσοι συμμετέχοντες στην έρευνα είχαν αναπτύξει σακχαρώδη διαβήτη.

Ένας ασφαλής τρόπος για να επιβεβαιωθεί η ύπαρξη διαβήτη είναι από τα επίπεδα γλυκόζης νηστείας στο ανθρώπινο αίμα.

#### 3.4.1 Περιγραφικά στατιστικά μεταβλητής «Γλυκόζη Αίματος»

Όπως έχουμε αναφέρει και σε προηγούμενο κεφάλαιο, η διάγνωση του σακχαρώδη διαβήτη μπορεί να γίνει με διάφορες εξετάσεις. Σε πολλές περιπτώσεις θα υπάρξουν και ορισμένα σωματικά συμπτώματα, τα οποία σε συνδυασμό με την κλινική εικόνα του ατόμου θα οδηγήσουν τον γιατρό στη διάγνωση. Προφανώς δεν είναι μοναδικό κριτήριο η εξέταση γλυκόζης πλάσματος νηστείας καθώς πολλοί παράγοντες μπορούν να αυξομειώσουν τις τιμές του σακχάρου. Παρ' όλα αυτά, στο παρόν κεφάλαιο θα ερευνήσουμε πως συμπεριφέρονται οι τιμές του σακχάρου στο δείγμα.

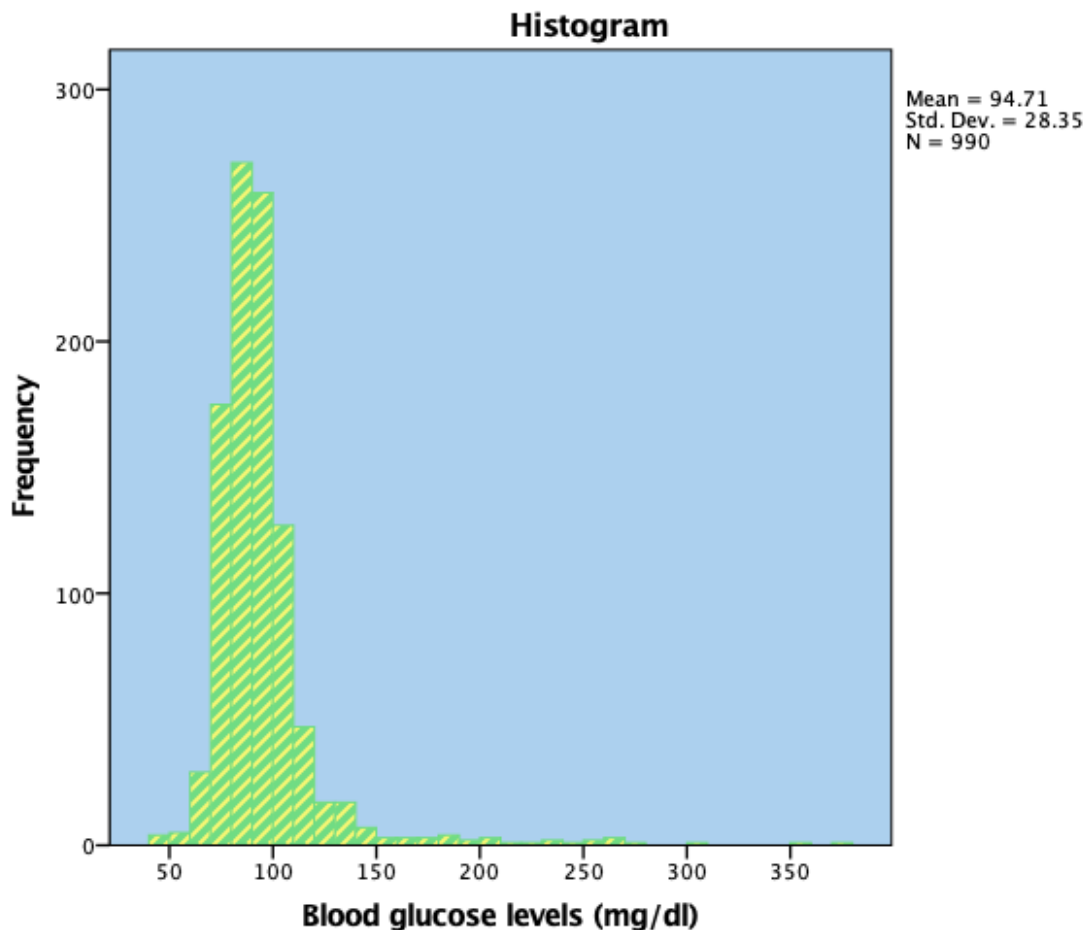
Από τον πίνακα 3.11 πληροφορούμαστε ότι η χαμηλότερη τιμή της γλυκόζης νηστείας στο αίμα των συμμετεχόντων ήταν 40 mg/dL και η υψηλότερη τιμή που παρατηρήθηκε ήταν 371 mg/dL. Επομένως αντιλαμβανόμαστε πως η έρευνα κάλυψε σχεδόν όλο το φάσμα των πιθανών τιμών της γλυκόζης νηστείας εάν αναλογιστούμε πως και οι δύο τιμές θεωρούνται εξαιρετικά επικίνδυνες. Πρόκειται για ακραίες τιμές στα άκρα του διαστήματος των πιθανών τιμών γλυκόζης και απαιτείται ιατρική παρακολούθηση σε κάθε περίπτωση.

Επιπροσθέτως, ο μέσος όρος των τιμών γλυκόζης νηστείας στο αίμα των συμμετεχόντων βρέθηκε  $94,71 \text{ mg/dL} \pm 28,35 \text{ mg/dL}$ . Μια τιμή οριακή για τη διάγνωση του διαβήτη τύπου II, αφού για να διαγνωστεί κάποιος με σακχαρώδη διαβήτη, θα πρέπει τα επίπεδα γλυκόζης να ξεπερνούν τα 126 mg/dL στο αίμα του. Τέλος η διάμεσος του δείγματος υπολογίστηκε στα 90 mg/dL.

Descriptives			Statistic	Std. Error
Blood glucose levels (mg/dl)	Mean		94.71	.901
	95% Confidence Interval for Mean	Lower Bound	92.94	
		Upper Bound	96.48	
	Median		90.00	
	Std. Deviation		28.350	
	Minimum		40	
	Maximum		371	
	Range		331	

Πίνακας 3.11: Περιγραφικά μέτρα μεταβλητής γλυκόζη αίματος

Στο γράφημα 3.8 φαίνεται πως στη πλειοψηφία των συμμετεχόντων στην έρευνα, παρατηρήθηκε γλυκόζη νηστείας στο αίμα τους από 80 mg/dL έως περίπου 120 mg/dL. Πρόκειται για τιμές στο κατώφλι του διαβήτη.



Γράφημα 3.8: Ιστόγραμμα συχνοτήτων μεταβλητής γλυκόζη αίματος



### 3.4.2 Περιγραφικά στατιστικά μεταβλητής «Σακχαρώδης Διαβήτης»

Σ' αυτή τη παράγραφο θα εξετάσουμε την αναλογία των διαβητικών ασθενών και των υγιών περιπτώσεων στο δείγμα.

Η συγκεκριμένη μεταβλητή συμπληρώθηκε έχοντας σαν γνώμονα αρκετές παραμέτρους και όχι μόνο τα επίπεδα σακχάρου στο αίμα. Γι' αυτό το λόγο παρατηρούνται περιπτώσεις όπου ενώ η γλυκόζη νηστείας παίρνει τιμές μικρότερες των 126mg/dL στο αίμα και παρόλα αυτά ο ασθενής είναι διαβητικός.

Από τον πίνακα 3.12 παρατηρούμε ότι η συντριπτική πλειοψηφία των συμμετεχόντων στη μελέτη δεν ήταν επίσημα διαγνωσμένα διαβητικοί τύπου I, τύπου II ή διαβητικές κύησης. Πιθανώς ορισμένοι να ήταν στο στάδιο του προδιαβήτη αφού οι τιμές της γλυκόζης νηστείας ήταν αυξημένη στο αίμα τους. Έτσι λοιπόν μόνο το 10% περίπου του δείγματος πάσχουν από σακχαρώδη διαβήτη.

Diabetes Mellitus					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No diabetes	935	90.3	90.3	90.3
	diabetes	100	9.7	9.7	100.0
	Total	1035	100.0	100.0	

Πίνακας 3.12: Παρατηρούμενες συχνότητες μεταβλητής σακχαρώδη διαβήτη

# ΚΕΦΑΛΑΙΟ 4<sup>ο</sup>

## ΣΤΑΤΙΣΤΙΚΟΣ ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ

---

### 4.1 Στατιστικός έλεγχος υποθέσεων

Στο 4<sup>ο</sup> κεφάλαιο της παρούσας εργασίας, αρχικά θα παρουσιαστεί η διαδικασία με την οποία μπορούν να αναλυθούν δύο ποιοτικές μεταβλητές. Η διαδικασία διαφέρει ανάλογα εάν μια από τις δύο ποιοτικές μεταβλητές είναι διατάξιμη. Επομένως πρώτα θα πρέπει να διευκρινιστεί ποιες μεταβλητές είναι κατηγορικές και ποιες διατάξιμες. Στη πρώτη περίπτωση θα μελετήσουμε μια σχέση εξάρτησης ενώ στη δεύτερη την ύπαρξη γραμμικής τάσης μεταξύ των μεταβλητών.

Στη συνέχεια θα προχωρήσουμε την ανάλυση χρησιμοποιώντας συνεχή μεταβλητή για εξαρτημένη. Η διαδικασία θα διαφοροποιηθεί αναλόγως τι μορφής θα είναι οι ανεξάρτητες μεταβλητές που επιλέχθηκαν για ανάλυση. Επιγραμματικά αναφέρουμε πως στο συνδυασμό συνεχούς εξαρτημένης με δίτιμης ανεξάρτητης μεταβλητής θα περάσουμε σε έλεγχο  $t$ , στη περίπτωση που η εξαρτημένη έχει περισσότερα από δύο επίπεδα θα περάσουμε σε κλασική άνοια, ενώ τέλος στη περίπτωση που και οι δύο μεταβλητές είναι συνεχής τότε θα ερευνήσουμε την ύπαρξη συσχέτισης μεταξύ τους μέσω του ελέγχου του Pearson ή των Kendall – Spearman, αναλόγως εάν αποδεχόμαστε την κανονικότητα της εξαρτημένης μεταβλητής.

### 4.2 Στατιστικός έλεγχος δημογραφικών μεταβλητών

Μπορεί να γίνει κατανοητή η σημαντικότητα της στατιστικής ανάλυσης των δημογραφικών δεικτών, εάν αναλογιστεί κανείς ότι πρόκειται για μεγάλες πληθυσμιακές ομάδες. Θα ήταν εξαιρετικά κομβικής σημασίας εάν για παράδειγμα οι ειδικοί γνώριζαν ότι μια ασθένεια εξαρτάται η εμφάνισή της ή όχι από το φύλο, την ηλικιακή ομάδα ή τι σχέση έχει ενδεχομένως με την οικονομική κατάσταση που βρίσκεται το άτομο.

### 4.2.1 Έλεγχος ανεξαρτησίας μεταβλητής «Φύλο»

Σκοπός της παραγράφου είναι να ερευνήσουμε τη σχέση μεταξύ της δημογραφικής μεταβλητής του φύλου (*sex*) και της μεταβλητής του σακχαρώδη διαβήτη (*diabetes mellitus - DM*). Κύρια επιδίωξη μας είναι να μελετήσουμε τη σχέση εξάρτησης μεταξύ των δύο αυτών μεταβλητών.

Πρόκειται για δύο ποιοτικές μεταβλητές, αφού υποθέτουμε ότι το φύλο μπορεί να έχει δύο καταστάσεις (άνδρας - γυναίκα) και η μεταβλητή του σακχαρώδη διαβήτη μπορεί να είναι είτε θετική είτε αρνητική.

Αρχικά, θα παρουσιάσουμε τη σχέση μεταξύ των ποιοτικών μεταβλητών μέσω του πίνακα συνάφειας όπου μπορούμε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης στα δύο φύλα.

Ενδεικτικά αναφέρουμε πως βρέθηκαν συνολικά 100 άτομα με σακχαρώδη διαβήτη εκ των οποίων οι 30 ήταν γυναίκες και οι 70 άνδρες. Οι υπόλοιποι 935 συμμετέχοντες στην έρευνα χαρακτηρίστηκαν υγιείς. Εκτός από τις συχνότητες, ο πίνακας περιέχει τις συνολικές σχετικές συχνότητες ανά γραμμή και στήλη. Για παράδειγμα, παρατηρούμε πως το 86,7% των ανδρών δεν πάσχει από διαβήτη ενώ από το σύνολο των μη διαβητικών συμμετεχόντων, το 48,7% ήταν άνδρες.

Sex of Subjects * Diabetes Mellitus Crosstabulation					
		Diabetes Mellitus		Total	
		No diabetes	diabetes		
Sex of Subjects	Female	Count	480	30	510
		Expected Count	460.7	49.3	510.0
		% within Sex of Subjects	94.1%	5.9%	100.0%
		% within Diabetes Mellitus	51.3%	30.0%	49.3%
		% of Total	46.4%	2.9%	49.3%
	Male	Count	455	70	525
		Expected Count	474.3	50.7	525.0
		% within Sex of Subjects	86.7%	13.3%	100.0%
		% within Diabetes Mellitus	48.7%	70.0%	50.7%
		% of Total	44.0%	6.8%	50.7%
Total	Count	935	100	1035	
	Expected Count	935.0	100.0	1035.0	
	% within Sex of Subjects	90.3%	9.7%	100.0%	
	% within Diabetes Mellitus	100.0%	100.0%	100.0%	
	% of Total	90.3%	9.7%	100.0%	

Πίνακας 4.1: Πίνακας συνάφειας φύλου - σακχαρώδη διαβήτη, παρατηρούμενες και αναμενόμενες σχετικές συχνότητες

Προτού εφαρμόσουμε τον στατιστικό έλεγχο  $\chi^2$ , θα ελέγξουμε εάν ισχύει ο κανόνας ορθής εφαρμογής του. Παρατηρούμε ότι ισχύουν και οι τρεις προϋποθέσεις του αφού, πρώτον το μέγεθος του δείγματος είναι μεγαλύτερο του τετραπλάσιου του αριθμού των κελιών ( $1035 > 4 \cdot 4$ ), δεύτερον όλες οι αναμενόμενες συχνότητες είναι μεγαλύτερες της μονάδας όπως φαίνεται από τις αναμενόμενες συχνότητες με πράσινο χρώμα του πίνακα 4.1 και τρίτον το ποσοστό των αναμενόμενων συχνοτήτων που είναι μικρότερες του πέντε είναι μικρότερο του 20% με 25%.

Παρά το γεγονός ότι όλα τα p-values που υπολογίστηκαν τείνουν στο μηδέν, εμείς θα πρέπει να λάβουμε υπόψιν μόνο την πρώτη γραμμή του ελέγχου  $\chi^2$  του Pearson, αφού δεν έχουμε μικρό δείγμα ώστε να μας ενδιαφέρει το ακριβές τεστ του Fisher καθώς και το φύλο δεν είναι διατάξιμη μεταβλητή ώστε να μας ενδιαφέρει ο έλεγχος τάσης (linear by linear).

Για να ελεγχθεί η υπόθεση εάν υπάρχει σχέση μεταξύ της εμφάνισης της ασθένειας του σακχαρώδη διαβήτη και του φύλου, θα εφαρμόσουμε τον έλεγχο ανεξαρτησίας  $\chi^2$  του Pearson.

Ο έλεγχος υποθέσεων διαμορφώνεται ως εξής:

$H_0$ : δεν υπάρχει εξάρτηση μεταξύ φύλου και σακχαρώδη διαβήτη

$H_1$ : υπάρχει εξάρτηση μεταξύ φύλου και σακχαρώδη διαβήτη

Από τον πίνακα που ακολουθεί, παρατηρούμε ότι το p-value του ελέγχου  $\chi^2$  του Pearson είναι σχεδόν μηδέν που σημαίνει ότι απορρίπτουμε τη μηδενική υπόθεση σε επίπεδο σημαντικότητας  $\alpha=5\%$ . Ουσιαστικά, σημαίνει ότι υπάρχει σχέση μεταξύ φύλου και σακχαρώδη διαβήτη, δηλαδή το φύλο έχει επίδραση στο εάν θα παρουσιαστεί διαβήτης.

Chi-Square Tests					
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	16.455 <sup>a</sup>	1	.000		
Continuity Correction <sup>b</sup>	15.612	1	.000		
Likelihood Ratio	16.908	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	16.439	1	.000		
N of Valid Cases	1035				

Πίνακας 4.2: Πίνακας αποτελεσμάτων στατιστικών ελέγχων

Μπορούμε να ελέγξουμε την ένταση της συσχέτισης μεταξύ δύο μη διατάξιμων μεταβλητών μέσω αρκετών μέτρων συνάφειας. Στην ανάλυσή μας θα εμπιστευτούμε τον συντελεστή  $\Phi$  του Pearson και τον  $V$  του Cramer. Οι δύο πρώτες γραμμές του παρακάτω πίνακα αναφέρονται σ' αυτούς τους συντελεστές, οι οποίοι είναι ίσοι με 0,126. Αυτό μας οδηγεί στο συμπέρασμα ότι υπάρχει μικρής έντασης συσχέτιση μεταξύ του φύλου και του σακχαρώδη διαβήτη.

Symmetric Measures					
		Value	Asymptotic Standardized Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Nominal by Nominal	Phi	.126			.000
	Cramer's V	.126			.000
	Contingency Coefficient	.125			.000
Interval by Interval	Pearson's R	.126	.029	4.085	.000 <sup>c</sup>
Ordinal by Ordinal	Spearman Correlation	.126	.029	4.085	.000 <sup>c</sup>
N of Valid Cases		1035			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

**Πίνακας 4.3:** Πίνακας μέτρων συνάφειας (Pearson's  $\Phi$  – Cramer's  $V$ ) για μη διατάξιμους πίνακες συνάφειας

Σε συνέχεια της προηγούμενης παραγράφου, θα παρουσιάσουμε κάποια μέτρα συνάφειας επιπλέον, τα οποία έχουν ιδιαίτερα σημαντική ερμηνεία.

Στη πρώτη γραμμή του πίνακα που ακολουθεί, υπολογίζεται το πηλίκο πιθανοτήτων, odds ratio, το οποίο ισούται με 2,462. Αυτό ουσιαστικά σημαίνει ότι η σχετική πιθανότητα να ασθενήσει από σακχαρώδη διαβήτη ένας άνδρας είναι 2,5 φορές περίπου μεγαλύτερη από την αντίστοιχη σχετική πιθανότητα για μια γυναίκα.

Risk Estimate			
	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Sex of Subjects (Female / Male)	2.462	1.575	3.847
For cohort Diabetes Mellitus = No diabetes	1.086	1.043	1.130
For cohort Diabetes Mellitus = diabetes	.441	.293	.665
N of Valid Cases	1035		

Πίνακας 4.4: Πίνακας μέτρων συνάφειας (Odds Ratio – Relative Risk) για μη διατάξιμους πίνακες συνάφειας

#### 4.2.2 Έλεγχος συσχέτισης μεταβλητής «Ηλικία»

Ο χρόνος και η φθορά είναι άρρηκτα συνδεδεμένοι παράγοντες. Απ' αυτόν τον κανόνα δε θα μπορούσε να απουσιάζει ο ανθρώπινος οργανισμός. Ιατρικές μελέτες έχουν δείξει πως περνώντας τα χρόνια, η υγεία του ανθρώπου φθίνει και γίνεται ολοένα πιο ευάλωτη σε ασθένειες, ιούς και αυτοάνοσα νοσήματα.

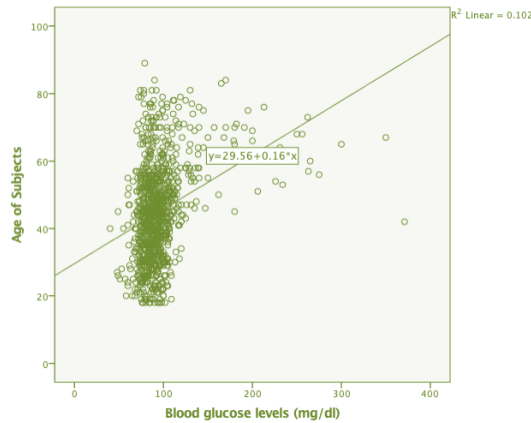
Σκοπός της παραγράφου είναι να ερευνήσουμε την ύπαρξη συσχέτισης μεταξύ της μεταβλητής της ηλικίας (*age*) και της μεταβλητής που φανερώνει τα επίπεδα γλυκόζης στο αίμα (*blood glucose levels- glucose*). Κύρια επιδίωξη μας είναι να μελετήσουμε το επίπεδο της έντασης της συσχέτισης μεταξύ των δύο αυτών μεταβλητών.

Προτού προχωρήσουμε στον έλεγχο συσχέτισης του Pearson, θα πρέπει να ικανοποιούνται κάποιες βασικές προϋποθέσεις. Αρχικά θα πρέπει τόσο η εξαρτημένη, όσο και η ανεξάρτητη μεταβλητή να είναι συνεχείς, να είναι γραμμικές μεταξύ τους και να ικανοποιείται η κανονικότητα του δείγματος.

Αρχικά, πρόκειται για δύο συνεχείς μεταβλητές, αφού τόσο η ηλικία όσο και τα επίπεδα σακχάρου στο αίμα μπορούν να πάρουν θετικές τιμές.

Από το διάγραμμα που ακολουθεί, παρατηρούμε πως μεταξύ των μεταβλητών δεν υπάρχει γραμμική σχέση, γεγονός που επιβεβαιώνεται από τις παρατηρήσεις οι οποίες δεν βρίσκονται σε καμία περίπτωση κοντά στη διαγώνιο.

## Κεφάλαιο 4<sup>ο</sup>: Στατιστικός Έλεγχος Υποθέσεων



**Γράφημα 4.1:** Διάγραμμα διασποράς μεταβλητών ηλικίας και σακχάρου αίματος

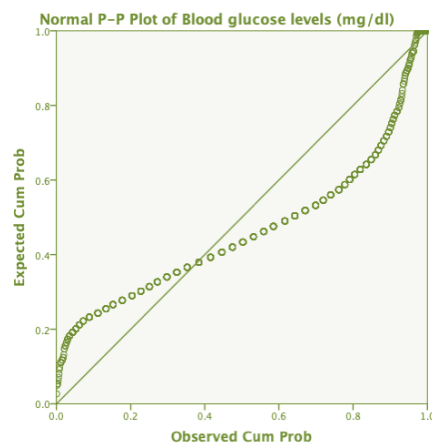
Παρότι δεν ικανοποιείται η πρώτη προϋπόθεση του ελέγχου συσχετίσεων του Pearson, θα προχωρήσουμε και στον έλεγχο των υπόλοιπων προϋποθέσεων.

Αρχικά θα κατασκευάσουμε ένα p-p plot για κάθε μεταβλητή ξεχωριστά ώστε να έχουμε μια πρώτη εικόνα για τη κανονικότητα του δείγματός μας.



**Γράφημα 4.2:** Διάγραμμα normal p-p plot της μεταβλητής «ηλικία»

Φαίνεται να υπάρχει γραφική ένδειξη απόκλισης από τη κανονική κατανομή αφού τα σημεία, ιδιαίτερα της μεταβλητής «επίπεδα σακχάρου στο αίμα», δεν βρίσκονται κοντά στη διχοτόμο της γωνίας των αξόνων.



**Γράφημα 4.3:** Διάγραμμα normal p-p plot της μεταβλητής «επίπεδα σακχάρου αίματος»

Τώρα μπορούμε να προχωρήσουμε στο στατιστικό έλεγχο κανονικότητας των Kolmogorov - Smirnov ο οποίος ελέγχει τις παρακάτω υποθέσεις:

$H_0$ : οι μεταβλητές ακολουθούν την κανονική κατανομή

$H^1$ : οι μεταβλητές δεν ακολουθούν την κανονική κατανομή

Από τον έλεγχο επιβεβαιώνεται αυτό που είχαμε υποθέσει γραφικά παραπάνω, δηλαδή σε επίπεδο σημαντικότητας 5% υπάρχουν ενδείξεις για απόρριψη της μηδενικής υπόθεσης της κανονικότητας των μεταβλητών αφού το p-value είναι σχεδόν μηδέν, το οποίο υπολείπεται κατά πολύ του  $\alpha=0,05$ . Τελικά και οι δύο μεταβλητές δεν ακολουθούν την κανονική κατανομή.

One-Sample Kolmogorov-Smirnov Test			
		Age of Subjects	Blood glucose levels (mg/dl)
N		1035	990
Normal Parameters <sup>a,b</sup>	Mean	45.25	94.71
	Std. Deviation	14.479	28.350
Most Extreme Differences	Absolute	.046	.201
	Positive	.046	.201
	Negative	-.033	-.155
Test Statistic		.046	.201
Asymp. Sig. (2-tailed)		.000 <sup>c</sup>	.000 <sup>c</sup>

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

**Πίνακας 4.5:** Πίνακας ελέγχου κανονικότητας Kolmogorov - Smirnov των μεταβλητών «ηλικία» και «επίπεδα σακχάρου αίματος»

Αφού οι δύο από τις τρεις προϋποθέσεις του ελέγχου δεν εκπληρώθηκαν, θα περάσουμε στον αντίστοιχο μη παραμετρικό έλεγχο συσχετίσεων του Kendall και Spearman.

Στο στατιστικό έλεγχο συσχετίσεων των Kendall και Spearman ελέγχονται οι παρακάτω υποθέσεις:

$H_0$ : δεν υπάρχει συσχέτιση μεταξύ των μεταβλητών ηλικία και επίπεδα σακχάρου στο αίμα



H<sub>1</sub>: υπάρχει συσχέτιση μεταξύ των μεταβλητών ηλικία και επίπεδα σακχάρου στο αίμα

Τα p-values και των δύο ελέγχων βρέθηκαν σχεδόν μηδέν με αποτέλεσμα να απορρίπτουμε την μηδενική υπόθεση, αφού είναι μικρότερα από ένα επίπεδο σημαντικότητας της τάξεως του 5%. Ουσιαστικά υπάρχει συσχέτιση μεταξύ των μεταβλητών «ηλικία» και «επίπεδα σακχάρου στο αίμα».

Correlations				
			Age of Subjects	Blood glucose levels (mg/dl)
Kendall's tau_b	Age of Subjects	Correlation Coefficient	1.000	.196**
		Sig. (2-tailed)	.	.000
		N	1035	990
	Blood glucose levels (mg/dl)	Correlation Coefficient	.196**	1.000
		Sig. (2-tailed)	.000	.
		N	990	990
Spearman's rho	Age of Subjects	Correlation Coefficient	1.000	.280**
		Sig. (2-tailed)	.	.000
		N	1035	990
	Blood glucose levels (mg/dl)	Correlation Coefficient	.280**	1.000
		Sig. (2-tailed)	.000	.
		N	990	990

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

Πίνακας 4.6: Πίνακας ελέγχου συσχετίσεων Kendall - Spearman των μεταβλητών «ηλικία» και «επίπεδα σακχάρου αίματος»

### 4.2.3 Έλεγχος ανεξαρτησίας μεταβλητής «Οικογενειακή κατάσταση»

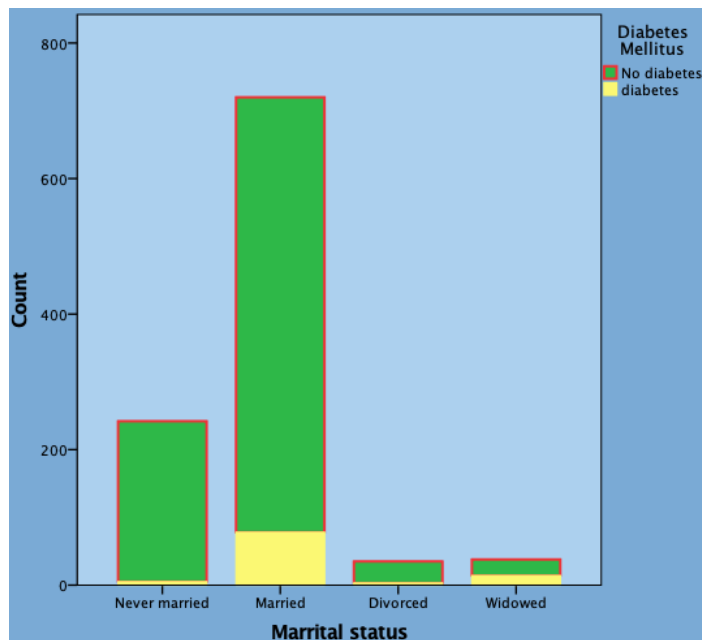
Στις περισσότερες σύγχρονες έρευνες, στο τμήμα των δημογραφικών δεδομένων, υπάρχει ερώτηση που σχετίζεται με την οικογενειακή κατάσταση του συμμετέχοντα. Μπορεί να αποδειχτεί ιδιαίτερα χρήσιμη, ειδικά στις ιατρικές έρευνες. Ο κυριότερος λόγος είναι ότι η καθημερινότητα ενός ατόμου μπορεί να επηρεάζει σχεδόν όλες τις πτυχές των ψυχοσωματικών του διεργασιών. Κάποιοι παράγοντες που συσχετίζονται με την οικογενειακή κατάσταση ενός ατόμου είναι ο ελεύθερος προσωπικός χρόνος, οι διατροφικές συνήθειες, η διαμονή, η σωματική άσκηση ή ακόμα τα επίπεδα στρες και άγχους.

Σκοπός της παραγράφου είναι να ερευνήσουμε τη σχέση μεταξύ της δημογραφικής μεταβλητής της οικογενειακής κατάστασης (*marital status - family*) και της μεταβλητής του σακχαρώδη διαβήτη (*diabetes mellitus - DM*). Κύρια επιδίωξη μας είναι να μελετήσουμε τη σχέση εξάρτησης μεταξύ των δύο αυτών μεταβλητών.

Πρόκειται για δύο ποιοτικές μεταβλητές, αφού η οικογενειακή κατάσταση μπορεί να έχει τέσσερις καταστάσεις (άγαμος/η, παντρεμένος/η, διαζευγμένος/η, χήρος/α) χωρίς να υπάρχει διάταξη μεταξύ τους. Η μεταβλητή του σακχαρώδη διαβήτη μπορεί να είναι είτε θετική είτε αρνητική.

Αρχικά, θα παρουσιάσουμε τη σχέση μεταξύ των ποιοτικών μεταβλητών μέσω του πίνακα συνάφειας όπου μπορούμε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης στα επίπεδα της μεταβλητής της οικογενειακής κατάστασης.

Ενδεικτικά αξίζει να αναφερθεί πως από τους άγαμους συμμετέχοντες μόνο το 2,1% παρουσίασε σακχαρώδη διαβήτη, από τους παντρεμένους το 10,8%, από τους διαζευγμένους το 8,6% και από τους χήρους το 36,8%(!). Παρατηρείται λοιπόν μια αρκετά έντονη διαφοροποίηση ανάμεσα στα επίπεδα της μεταβλητής οικογενειακής κατάστασης και μένει να διαπιστωθεί εάν αυτές οι διαφορές είναι στατιστικά σημαντικές.



Γράφημα 4.4: Ραβδόγραμμα μεταβλητής «οικογενειακή κατάσταση»

Marrital status * Diabetes Mellitus Crosstabulation						
		Diabetes Mellitus		Total		
		No diabetes	diabetes			
Marrital status	Never married	Count	237	5	242	
		Expected Count	218.6	23.4	242.0	
		% within Marrital status	97.9%	2.1%	100.0%	
		% within Diabetes Mellitus	25.3%	5.0%	23.4%	
		% of Total	22.9%	0.5%	23.4%	
		Married	Count	642	78	720
			Expected Count	650.4	69.6	720.0
	% within Marrital status		89.2%	10.8%	100.0%	
	% within Diabetes Mellitus		68.7%	78.0%	69.6%	
	% of Total		62.0%	7.5%	69.6%	
	Divorced		Count	32	3	35
			Expected Count	31.6	3.4	35.0
		% within Marrital status	91.4%	8.6%	100.0%	
		% within Diabetes Mellitus	3.4%	3.0%	3.4%	
		% of Total	3.1%	0.3%	3.4%	
		Widowed	Count	24	14	38
			Expected Count	34.3	3.7	38.0
	% within Marrital status		63.2%	36.8%	100.0%	
	% within Diabetes Mellitus		2.6%	14.0%	3.7%	
	% of Total		2.3%	1.4%	3.7%	
	Total		Count	935	100	1035
Expected Count			935.0	100.0	1035.0	
% within Marrital status		90.3%	9.7%	100.0%		
% within Diabetes Mellitus		100.0%	100.0%	100.0%		
% of Total		90.3%	9.7%	100.0%		

Πίνακας 4.7: Πίνακας συνάφειας οικογενειακής κατάστασης - σακχαρώδη διαβήτη, παρατηρούμενες και αναμενόμενες σχετικές συχνότητες

Για να ελεγχθεί η υπόθεση εάν υπάρχει σχέση μεταξύ της εμφάνισης της ασθένειας του σακχαρώδη διαβήτη και της οικογενειακής κατάστασης ενός ατόμου, θα εφαρμόσουμε τον έλεγχο ανεξαρτησίας  $\chi^2$  του Pearson.

Ο έλεγχος υποθέσεων διαμορφώνεται ως εξής:

$H_0$ : δεν υπάρχει εξάρτηση μεταξύ οικογενειακής κατάστασης και σακχαρώδη διαβήτη  
 $H_1$ : υπάρχει εξάρτηση μεταξύ οικογενειακής κατάστασης και σακχαρώδη διαβήτη

Από τον πίνακα που ακολουθεί, παρατηρούμε ότι το p-value του ελέγχου  $\chi^2$  του Pearson είναι σχεδόν μηδέν που σημαίνει ότι απορρίπτουμε τη μηδενική υπόθεση σε επίπεδο σημαντικότητας  $\alpha=5\%$ . Ουσιαστικά, σημαίνει ότι υπάρχει σχέση μεταξύ οικογενειακής κατάστασης και σακχαρώδη διαβήτη, δηλαδή η οικογενειακή κατάσταση έχει επίδραση στο εάν θα παρουσιαστεί διαβήτης.

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	49.339 <sup>a</sup>	3	.000
Likelihood Ratio	44.282	3	.000
Linear-by-Linear Association	41.473	1	.000
N of Valid Cases	1035		

a. 2 cells (25.0%) have expected count less than 5. The minimum expected count is 3.38.

**Πίνακας 4.8:** Πίνακας αποτελεσμάτων στατιστικών ελέγχων

Μπορούμε να ελέγξουμε την ένταση της συσχέτισης μεταξύ δύο μη διατάξιμων μεταβλητών μέσω αρκετών μέτρων συνάφειας. Στην ανάλυσή μας θα εμπιστευτούμε τον συντελεστή  $\Phi$  του Pearson και τον V του Cramer. Οι δύο πρώτες γραμμές του παρακάτω πίνακα αναφέρονται σ' αυτούς τους συντελεστές, οι οποίοι είναι ίσοι με 0,218. Αυτό μας οδηγεί στο συμπέρασμα ότι υπάρχει μικρής έντασης συσχέτιση μεταξύ της οικογενειακής κατάστασης και του σακχαρώδη διαβήτη.

Symmetric Measures					
		Value	Asymptotic Standardized Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Nominal by Nominal	Phi	.218			.000
	Cramer's V	.218			.000
	Contingency Coefficient	.213			.000
Interval by Interval	Pearson's R	.200	.035	6.570	.000 <sup>c</sup>
Ordinal by Ordinal	Spearman Correlation	.176	.026	5.749	.000 <sup>c</sup>
N of Valid Cases		1035			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

Πίνακας 4.9: Πίνακας μέτρων συνάφειας (Pearson's  $\Phi$  – Cramer's V) για μη διατάξιμους πίνακες συνάφειας

#### 4.2.4 Έλεγχος ανεξαρτησίας μεταβλητής «Εκπαιδευτική και οικονομική κατάσταση»

Οι σχεδιαστές των σύγχρονων μελετών, ενδιαφέρονται για το μορφωτικό επίπεδο και την τρέχουσα οικονομική κατάσταση των συμμετεχόντων, ώστε συχνά ενσωματώνουν ερωτήσεις για να τους κατατάξουν ανάλογα με τις απαντήσεις τους. Σε μια έρευνα σχετικά με το ζάχαρο, θα είχε ενδιαφέρον να ερευνήσουμε εάν υπάρχουν διαφορές ανάμεσα σε άτομα διαφορετικών εκπαιδευτικών βαθμίδων αλλά και οικονομικών καταστάσεων. Εάν αναλογιστεί κανείς πως πρόκειται για μια χρόνια ασθένεια που απαιτεί συχνή ιατρική παρακολούθηση, καλή και ποιοτική διατροφή σε συνδυασμό με τακτική άσκηση, τότε αναμένουμε ενδιαφέροντα αποτελέσματα.

Σκοπός της παραγράφου λοιπόν είναι να ερευνήσουμε τη σχέση μεταξύ της εκπαιδευτικής και οικονομικής κατάστασης (*SocioEconomic Status - SES*) και της μεταβλητής του σακχαρώδη διαβήτη (*diabetes mellitus - DM*). Κύρια επιδίωξη μας είναι να μελετήσουμε την ύπαρξη γραμμικής τάσης μεταξύ των δύο αυτών μεταβλητών.

Πρόκειται για δύο μεταβλητές με διαφορετικά χαρακτηριστικά. Πιο συγκεκριμένα, η εκπαιδευτικό-οικονομική κατάσταση των συμμετεχόντων είναι μια διατάξιμη μεταβλητή αφού μπορεί να έχει εννιά ιεραρχικές καταστάσεις ανάλογα με το μηνιαίο εισόδημα και το επίπεδο της ανώτερης εκπαίδευσης των ατόμων. Αντιθέτως, η μεταβλητή του σακχαρώδη διαβήτη είναι μια κατηγορική μεταβλητή αφού μπορεί να είναι είτε θετική είτε αρνητική.

## Κεφάλαιο 4<sup>ο</sup>: Στατιστικός Έλεγχος Υποθέσεων

Αρχικά, θα παρουσιάσουμε τη σχέση μεταξύ των ποιοτικών μεταβλητών μέσω του πίνακα συνάφειας όπου μπορούμε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης ανάμεσα στα επίπεδα της εκπαιδευτικό-οικονομική μεταβλητής.

Ενδεικτικά αναφέρουμε πως έπασχε από σακχαρώδη διαβήτη το 20% από αυτούς που ανήκουν στη κατηγορία με χαμηλό εισόδημα και εκπαίδευση, το 11,9% από αυτούς που ανήκουν στη κατηγορία με μέτριο εισόδημα και χαμηλή εκπαίδευση και το 16,7% από αυτούς που ανήκουν στη κατηγορία με καλό εισόδημα και χαμηλή εκπαίδευση. Πρόκειται για αρκετά μεγάλα ποσοστά τα οποία πέφτουν όσο πλησιάζουμε προς την κορυφή της ιεραρχικής βαθμίδας. Επιπροσθέτως, έχει ενδιαφέρον να αναφέρουμε πως από το σύνολο των μη διαβητικών περιπτώσεων το 5,6% ανήκει στη πρώτη κατηγορία, το 6,4% στη δεύτερη, το 3,5% στη τρίτη, το 9,9% στη τέταρτη και ούτω καθ' εξής. Παρατηρούνται μειωμένα ποσοστά υγιών ατόμων στις αρχικές βαθμίδες και θα πρέπει να ερευνησουμε εάν αυτές οι διαφορές είναι στατιστικά σημαντικές.

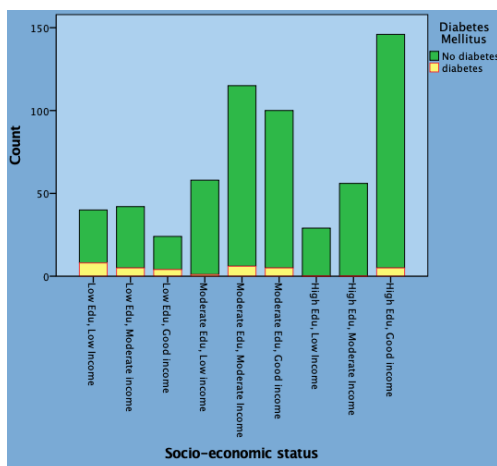
**Socio-economic status \* Diabetes Mellitus Crosstabulation**

			Diabetes Mellitus		
			No diabetes	diabetes	
Socio-economic status	Low Edu, Low Income	Count	32	8	40
		% within Socio-economic status	80.0%	20.0%	100.0%
		% within Diabetes Mellitus	5.6%	23.5%	6.6%
		% of Total	5.2%	1.3%	6.6%
	Low Edu, Moderate income	Count	37	5	42
		% within Socio-economic status	88.1%	11.9%	100.0%
		% within Diabetes Mellitus	6.4%	14.7%	6.9%
		% of Total	6.1%	0.8%	6.9%
	Low Edu, Good income	Count	20	4	24
		% within Socio-economic status	83.3%	16.7%	100.0%
		% within Diabetes Mellitus	3.5%	11.8%	3.9%
		% of Total	3.3%	0.7%	3.9%
Moderate Edu, Low income	Count	57	1	58	
	% within Socio-economic status	98.3%	1.7%	100.0%	
	% within Diabetes Mellitus	9.9%	2.9%	9.5%	
	% of Total	9.3%	0.2%	9.5%	

Κεφάλαιο 4<sup>ο</sup>: Στατιστικός Έλεγχος Υποθέσεων

Moderate Edu, Moderate Income	Count	109	6	115
	% within Socio-economic status	94.8%	5.2%	100.0%
	% within Diabetes Mellitus	18.9%	17.6%	18.9%
	% of Total	17.9%	1.0%	18.9%
Moderate Edu, Good income	Count	95	5	100
	% within Socio-economic status	95.0%	5.0%	100.0%
	% within Diabetes Mellitus	16.5%	14.7%	16.4%
	% of Total	15.6%	0.8%	16.4%
High Edu, Low Income	Count	29	0	29
	% within Socio-economic status	100.0%	0.0%	100.0%
	% within Diabetes Mellitus	5.0%	0.0%	4.8%
	% of Total	4.8%	0.0%	4.8%
High Edu, Moderate Income	Count	56	0	56
	% within Socio-economic status	100.0%	0.0%	100.0%
	% within Diabetes Mellitus	9.7%	0.0%	9.2%
	% of Total	9.2%	0.0%	9.2%
High Edu, Good income	Count	141	5	146
	% within Socio-economic status	96.6%	3.4%	100.0%
	% within Diabetes Mellitus	24.5%	14.7%	23.9%
	% of Total	23.1%	0.8%	23.9%
Total	Count	576	34	610
	% within Socio-economic status	94.4%	5.6%	100.0%
	% within Diabetes Mellitus	100.0%	100.0%	100.0%
	% of Total	94.4%	5.6%	100.0%

Πίνακας 4.10: Πίνακας συνάφειας εκπαιδευτικο-οικογενειακής κατάστασης - σακχαρώδη διαβήτη



**Γράφημα 4.5:** Ραβδόγραμμα μεταβλητής εκπαιδευτικο-οικογενειακής κατάστασης

Για να ελεγχθεί η υπόθεση εάν υπάρχει γραμμική τάση μεταξύ της εμφάνισης της ασθένειας του σακχαρώδη διαβήτη και της εκπαιδευτικο-οικονομικής κατάστασης, θα εφαρμόσουμε τον έλεγχο  $\chi^2$  για τάση.

Ο έλεγχος υποθέσεων διαμορφώνεται ως εξής:

$H_0$ : δεν υπάρχει γραμμική τάση μεταξύ εκπαιδευτικο-οικονομικής κατάστασης και σακχαρώδη διαβήτη

$H_1$ : υπάρχει γραμμική τάση μεταξύ εκπαιδευτικο-οικονομικής κατάστασης και σακχαρώδη διαβήτη

Παρά το γεγονός ότι όλα τα p-values που υπολογίστηκαν τείνουν στο μηδέν, εμείς θα πρέπει να λάβουμε υπόψιν μόνο τη τρίτη γραμμή του ελέγχου  $\chi^2$  (Linear by Linear Association) καθώς μια από τις δύο μεταβλητές είναι διατάξιμη και θα προχωρήσουμε σε έλεγχο τάσης.

Από τον πίνακα που ακολουθεί, παρατηρούμε ότι το p-value του ελέγχου  $\chi^2$  για τάση είναι σχεδόν μηδέν που σημαίνει ότι απορρίπτουμε τη μηδενική υπόθεση σε επίπεδο σημαντικότητας  $\alpha=5\%$ . Ουσιαστικά, σημαίνει ότι υπάρχει ισχυρή ένδειξη τάσης μεταξύ εκπαιδευτικο-οικονομικής κατάστασης και σακχαρώδη διαβήτη, δηλαδή η μεταβλητή της εκπαιδευτικο-οικονομικής κατάστασης έχει επίδραση στο εάν θα παρουσιαστεί διαβήτης.

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	32.649 <sup>a</sup>	8	.000
Likelihood Ratio	29.576	8	.000
Linear-by-Linear Association	17.950	1	.000
N of Valid Cases	610		

**Πίνακας 4.11:** Πίνακας αποτελεσμάτων στατιστικών ελέγχων



Μπορούμε να ελέγξουμε την ένταση της γραμμικής τάσης μεταξύ μιας διατάξιμης και μιας κατηγορικής μεταβλητής μέσω αρκετών μέτρων συνάφειας. Στην ανάλυσή μας θα εμπιστευτούμε τον συντελεστή  $\tau$  του Kendal (-0,134), τον  $\gamma$  των Goodman και Kruskal (-0,435) καθώς και τα κατευθυντήρια μέτρα του αμέσως επόμενου πίνακα (-0,084 & -0,381). Με βάση όλα τα μέτρα συνάφειας που υπολογίστηκαν, υπάρχει ελάχιστη αρνητική έως μέτρια αρνητική τάση, μεταξύ της εκπαιδευτικο-οικονομικής κατάστασης των συμμετεχόντων στην έρευνα και της εμφάνισης σακχαρώδη διαβήτη.

Directional Measures					
		Value	Asymptotic Standardized Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Ordinal by Ordinal	Somers' d Symmetric	-.084	.023	-3.240	.001
	Socio-economic status	-.381	.101	-3.240	.001
	Dependent Diabetes Mellitus	-.047	.015	-3.240	.001

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Πίνακας 4.12: Πίνακας κατευθυντήριων μέτρων συνάφειας για διατάξιμους πίνακες συνάφειας

Symmetric Measures					
		Value	Asymptotic Standardized Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Ordinal by Ordinal	Kendall's tau-b	-.134	.037	-3.240	.001
	Kendall's tau-c	-.080	.025	-3.240	.001
	Gamma	-.435	.110	-3.240	.001
	Spearman Correlation	-.154	.043	-3.835	.000 <sup>c</sup>
Interval by Interval	Pearson's R	-.172	.045	-4.297	.000 <sup>c</sup>
N of Valid Cases		610			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

Πίνακας 4.13: Πίνακας μέτρων συνάφειας (Kendal's  $\tau$  – Kruskal's  $\gamma$ ) για διατάξιμους πίνακες συνάφειας

### 4.3 Έλεγχος ανεξαρτησίας Διατροφικών και σωματομετρικών μεταβλητών

Όπως έχουμε αναφέρει και σε προηγούμενο κεφάλαιο, οι διατροφικές συνήθειες αυτών που πάσχουν από σακχαρώδη διαβήτη, κρίνονται από τους ειδικούς ως κομβικής σημασίας για την επιδείνωση ή όχι της ασθένειας.

#### 4.3.1 Έλεγχος συσχέτισης μεταβλητής «Σωματικού βάρους»

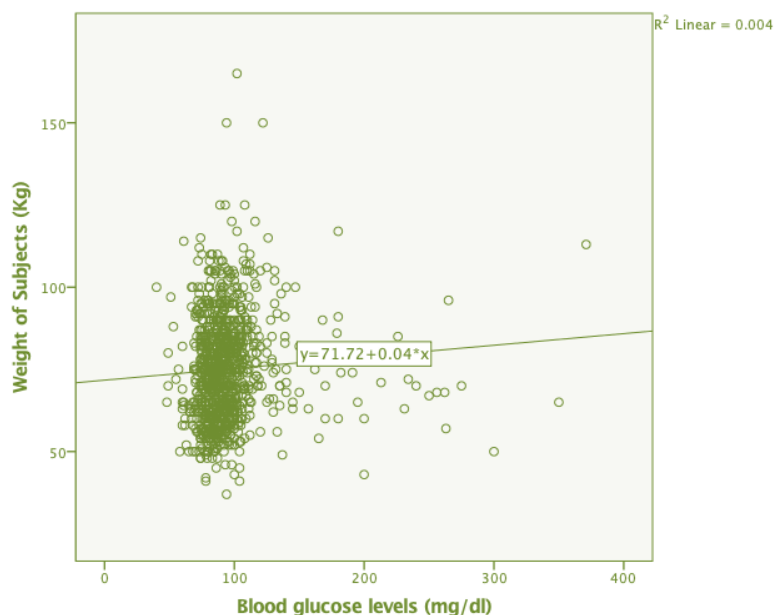
Η διατήρηση του σωματικού βάρους σε φυσιολογικά επίπεδα, είναι μια από τις βασικότερες προϋποθέσεις για την σωματική και πνευματική ευεξία. Η παχυσαρκία είναι συνδεδεμένη με τις περισσότερες καρδιαγγειακές παθήσεις αλλά και τον σακχαρώδη διαβήτη.

Σκοπός της παραγράφου είναι να ερευνήσουμε την ύπαρξη συσχέτισης μεταξύ της μεταβλητής του σωματικού βάρους (*weight of subjects - weight*) και της μεταβλητής που φανερώνει τα επίπεδα γλυκόζης στο αίμα (*blood glucose levels - glucose*). Κύρια επιδίωξη μας είναι να μελετήσουμε το επίπεδο της έντασης της συσχέτισης μεταξύ των δύο αυτών μεταβλητών.

Προτού προχωρήσουμε στον έλεγχο συσχέτισης του Pearson, θα πρέπει να ικανοποιούνται κάποιες βασικές προϋποθέσεις. Αρχικά θα πρέπει τόσο η εξαρτημένη, όσο και η ανεξάρτητη μεταβλητή να είναι συνεχείς, να είναι γραμμικές μεταξύ τους και να ικανοποιείται η κανονικότητα του δείγματος.

Πρόκειται για δύο συνεχείς μεταβλητές, αφού τόσο το σωματικό βάρος όσο και τα επίπεδα σακχάρου στο αίμα μπορούν να πάρουν θετικές τιμές.

Από το διάγραμμα που ακολουθεί, παρατηρούμε πως μεταξύ των μεταβλητών δεν υπάρχει γραμμική σχέση, γεγονός που επιβεβαιώνεται από τις παρατηρήσεις οι οποίες δεν βρίσκονται σε καμία περίπτωση κοντά στο ευθύγραμμο τμήμα.



**Γράφημα 4.6:** Διάγραμμα διασποράς μεταβλητών σωματικού βάρους και σακχάρου αίματος

Παρότι δεν ικανοποιείται η πρώτη προϋπόθεση του ελέγχου συσχετίσεων του Pearson, θα προχωρήσουμε και στον έλεγχο των υπόλοιπων προϋποθέσεων.

Στη συνέχεια θα προχωρήσουμε στο στατιστικό έλεγχο κανονικότητας των Kolmogorov - Smirnov ο οποίος ελέγχει τις παρακάτω υποθέσεις:

H<sub>0</sub>: η μεταβλητή του σωματικού βάρους ακολουθεί τη κανονική κατανομή

H<sub>1</sub>: η μεταβλητή του σωματικού βάρους δεν ακολουθεί τη κανονική κατανομή

Σε επίπεδο σημαντικότητας 5% υπάρχουν ενδείξεις για απόρριψη της μηδενικής υπόθεσης της κανονικότητας της μεταβλητή του σωματικού βάρους αφού το p-value είναι σχεδόν μηδέν, το οποίο υπολείπεται κατά πολύ του  $\alpha=0,05$ . Τελικά η μεταβλητή του σωματικού βάρους δεν ακολουθεί την κανονική κατανομή.

One-Sample Kolmogorov-Smirnov Test		
		Weight of Subjects (Kg)
N		1030
Normal Parameters <sup>a,b</sup>	Mean	75.17
	Std. Deviation	15.968
Most Extreme Differences	Absolute	.064
	Positive	.064
	Negative	-.041
Test Statistic		.064
Asymp. Sig. (2-tailed)		.000 <sup>c</sup>
a. Test distribution is Normal.		
b. Calculated from data.		
c. Lilliefors Significance Correction.		

**Πίνακας 4.14:** Πίνακας ελέγχου κανονικότητας Kolmogorov – Smirnov της μεταβλητής του σωματικού βάρους

Αφού οι δύο από τις τρεις προϋποθέσεις του ελέγχου δεν εκπληρώθηκαν, θα περάσουμε στον αντίστοιχο έλεγχο συσχετίσεων των Kendall και Spearman.

Στο στατιστικό έλεγχο συσχετίσεων των Kendall και Spearman ελέγχονται οι παρακάτω υποθέσεις:

H<sub>0</sub>: δεν υπάρχει συσχέτιση μεταξύ των μεταβλητών σωματικό βάρος και επίπεδα σακχάρου στο αίμα

H<sub>1</sub>: υπάρχει συσχέτιση μεταξύ των μεταβλητών σωματικού βάρους και επίπεδα σακχάρου στο αίμα

Τα p-values και των δύο ελέγχων βρέθηκαν σχεδόν μηδέν με αποτέλεσμα να απορρίπτουμε την μηδενική υπόθεση, αφού είναι μικρότερα από ένα επίπεδο σημαντικότητας της τάξεως του 5%. Ουσιαστικά υπάρχει συσχέτιση μεταξύ των μεταβλητών σωματικό βάρος και επίπεδα σακχάρου στο αίμα.

Correlations				
			Weight of Subjects (Kg)	Blood glucose levels (mg/dl)
Kendall's tau_b	Weight of Subjects (Kg)	Correlation Coefficient	1.000	.096**
		Sig. (2-tailed)	.	.000
	N		1030	986
	Blood glucose levels (mg/dl)	Correlation Coefficient	.096**	1.000
Sig. (2-tailed)		.000	.	
N		986	990	
Spearman's rho	Weight of Subjects (Kg)	Correlation Coefficient	1.000	.140**
		Sig. (2-tailed)	.	.000
	N		1030	986
	Blood glucose levels (mg/dl)	Correlation Coefficient	.140**	1.000
Sig. (2-tailed)		.000	.	
N		986	990	

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

Πίνακας 4.15: Πίνακας ελέγχου συσχέτισεων Kendall - Spearman των μεταβλητών σωματικού βάρους και επίπεδα σακχάρου αίματος

### 4.3.2 Έλεγχος ανεξαρτησίας μεταβλητής «Δείκτης μάζας σώματος»

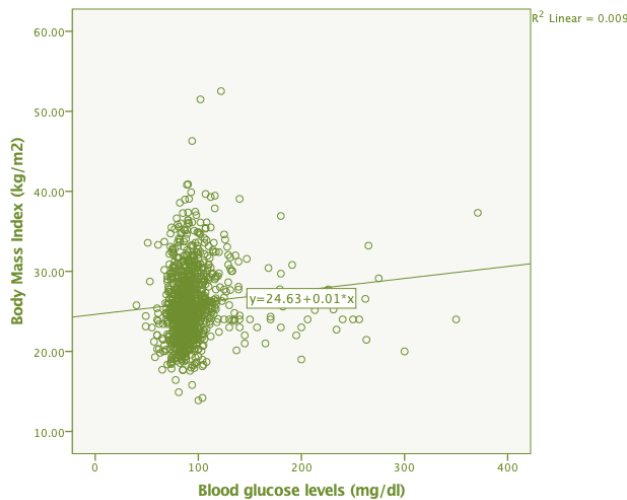
Σε συνέχεια της προηγούμενης παραγράφου, ο δείκτης μάζας σώματος είναι μία εναλλακτική μέθοδος για τη κατανόηση του σωματότυπου ενός ανθρώπου. Πιο συγκεκριμένα, με το βάρος και μόνο μπορεί να παρερμηνευτεί η εικόνα κάποιου, αφού δεν λαμβάνεται υπ' όψιν καθόλου το ύψος του. Αυτό το πρόβλημα λύθηκε με την δημιουργία του δείκτη BMI.

Σκοπός της παραγράφου είναι να ερευνήσουμε την ύπαρξη συσχέτισης μεταξύ της μεταβλητής του δείκτη μάζας σώματος (*body mass index - BMI*) και της μεταβλητής που φανερώνει τα επίπεδα γλυκόζης στο αίμα (*blood glucose levels - glucose*). Κύρια επιδίωξη μας είναι να μελετήσουμε το επίπεδο της έντασης της συσχέτισης μεταξύ των δύο αυτών μεταβλητών.

Προτού προχωρήσουμε στον έλεγχο συσχέτισης του Pearson, θα πρέπει να ικανοποιούνται κάποιες βασικές προϋποθέσεις. Αρχικά θα πρέπει τόσο η εξαρτημένη, όσο και η ανεξάρτητη μεταβλητή να είναι συνεχείς, να είναι γραμμικές μεταξύ τους και να ικανοποιείται η κανονικότητα του δείγματος.

Πρόκειται για δύο συνεχής μεταβλητές, αφού τόσο ο δείκτης μάζας σώματος όσο και τα επίπεδα σακχάρου στο αίμα μπορούν να πάρουν θετικές τιμές.

Από το διάγραμμα που ακολουθεί, παρατηρούμε πως μεταξύ των μεταβλητών δεν υπάρχει γραμμική σχέση, γεγονός που επιβεβαιώνεται από τις παρατηρήσεις οι οποίες δεν βρίσκονται σε καμία περίπτωση κοντά στο ευθύγραμμο τμήμα.

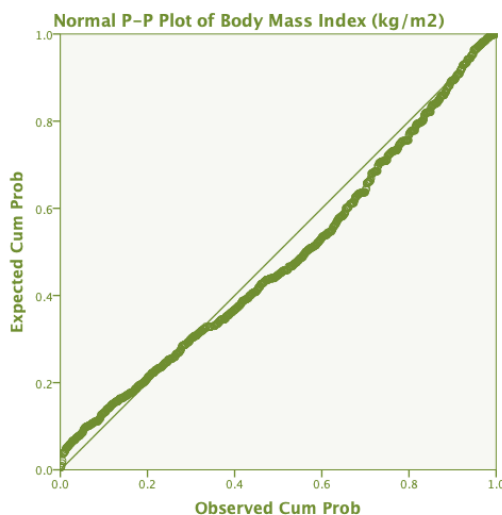


**Γράφημα 4.7:** Διάγραμμα διασποράς μεταβλητών δείκτη μάζας σώματος και σακχάρου αίματος

Παρότι δεν ικανοποιείται η πρώτη προϋπόθεση του ελέγχου συσχετίσεων του Pearson, θα προχωρήσουμε και στον έλεγχο των υπόλοιπων προϋποθέσεων.

Αρχικά θα κατασκευάσουμε ένα p-p plot για τη μεταβλητή του δείκτη μάζας σώματος, ώστε να έχουμε μια πρώτη εικόνα για τη κανονικότητα της μεταβλητής.

Φαίνεται να υπάρχει γραφική ένδειξη απόκλισης από τη κανονική κατανομή αφού τα σημεία, δεν βρίσκονται κοντά στη διχοτόμο της γωνίας των αξόνων.



**Γράφημα 4.8:** Διάγραμμα normal p-p plot της μεταβλητής δείκτης μάζας σώματος

Στη συνέχεια θα προχωρήσουμε στο στατιστικό έλεγχο κανονικότητας των Kolmogorov - Smirnov ο οποίος ελέγχει τις παρακάτω υποθέσεις:

$H_0$ : η μεταβλητή του δείκτη μάζας σώματος ακολουθεί τη κανονική κατανομή  
 $H_1$ : η μεταβλητή του δείκτη μάζας σώματος δεν ακολουθεί τη κανονική κατανομή

Σε επίπεδο σημαντικότητας 5% υπάρχουν ενδείξεις για απόρριψη της μηδενικής υπόθεσης της κανονικότητας της μεταβλητή του δείκτη μάζας σώματος αφού το p-value είναι σχεδόν μηδέν, το οποίο υπολείπεται κατά πολύ του  $\alpha=0,05$ . Τελικά η μεταβλητή του δείκτη μάζας σώματος δεν ακολουθεί την κανονική κατανομή.

One-Sample Kolmogorov-Smirnov Test		
		Body Mass Index (kg/m <sup>2</sup> )
N		1022
Normal Parameters <sup>a,b</sup>	Mean	26.1124
	Std. Deviation	4.75241
Most Extreme Differences	Absolute Positive	.075
	Negative	-.041
Test Statistic		.075
Asymp. Sig. (2-tailed)		.000 <sup>c</sup>

a. Test distribution is Normal.  
 b. Calculated from data.  
 c. Lilliefors Significance Correction.

**Πίνακας 4.16:** Πίνακας ελέγχου κανονικότητας Kolmogorov – Smirnov της μεταβλητής του δείκτη μάζας σώματος

Αφού οι δύο από τις τρεις προϋποθέσεις του ελέγχου δεν εκπληρώθηκαν, θα περάσουμε στον αντίστοιχο έλεγχο συσχετίσεων των Kendall και Spearman.

Στο στατιστικό έλεγχο συσχετίσεων των Kendall και Spearman ελέγχονται οι παρακάτω υποθέσεις:

$H_0$ : δεν υπάρχει συσχέτιση μεταξύ των μεταβλητών δείκτη μάζας σώματος και επίπεδα σακχάρου στο αίμα  
 $H_1$ : υπάρχει συσχέτιση μεταξύ των μεταβλητών δείκτη μάζας σώματος και επίπεδα σακχάρου στο αίμα

Τα p-values και των δύο ελέγχων βρέθηκαν σχεδόν μηδέν με αποτέλεσμα να απορρίπτουμε την μηδενική υπόθεση, αφού είναι μικρότερα από ένα επίπεδο σημαντικότητας της τάξεως του 5%. Ουσιαστικά υπάρχει συσχέτιση μεταξύ των μεταβλητών «δείκτη μάζας σώματος» και «επίπεδα σακχάρου στο αίμα».

Correlations				
			Body Mass Index (kg/m <sup>2</sup> )	Blood glucose levels (mg/dl)
Kendall's tau_b	Body Mass Index (kg/m <sup>2</sup> )	Correlation Coefficient	1.000	.105**
		Sig. (2-tailed)	.	.000
		N	1022	978
	Blood glucose levels (mg/dl)	Correlation Coefficient	.105**	1.000
		Sig. (2-tailed)	.000	.
		N	978	990
Spearman's rho	Body Mass Index (kg/m <sup>2</sup> )	Correlation Coefficient	1.000	.155**
		Sig. (2-tailed)	.	.000
		N	1022	978
	Blood glucose levels (mg/dl)	Correlation Coefficient	.155**	1.000
		Sig. (2-tailed)	.000	.
		N	978	990

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

**Πίνακας 4.17:** Πίνακας ελέγχου συσχετίσεων Kendall - Spearman των μεταβλητών «δείκτης μάζας σώματος» και «επίπεδα σακχάρου αίματος»

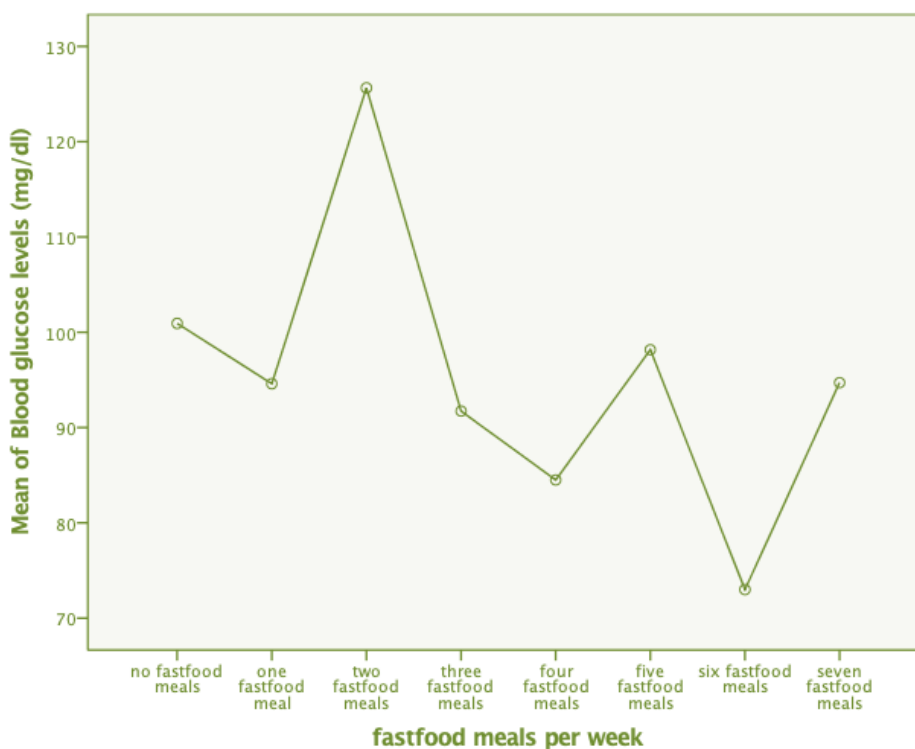
### 4.3.3 Έλεγχος σύγκρισης μέσω όρων της μεταβλητής «Κατανάλωση Έτοιμου Φαγητού»

Σε κάθε μελέτη παρεμφερούς ερευνητικού ενδιαφέροντος, υπάρχει μέριμνα για τουλάχιστον μια μεταβλητή που να αναφέρεται στο γρήγορο φαγητό. Στο μοντέρνο τρόπο ζωής, υπάρχει σοβαρή έλλειψη χρόνου και παράλληλα πληθώρα επιλογών για έτοιμο γεύμα κάθε είδους. Η καθημερινή κατανάλωση έτοιμου φαγητού δεν συνίσταται από τους ειδικούς διότι είναι στην πλειοψηφία τους πλούσια σε λίπη, υδατάνθρακες και σάκχαρα.

Σκοπός της παραγράφου είναι να ερευνήσουμε εάν υπάρχει διαφοροποίηση στα επίπεδα ( μηδέν έως επτά έτοιμα γεύματα ανά εβδομάδα) της μεταβλητής της κατανάλωσης έτοιμου φαγητού (*fastfood meals per week - fastfood*) και της μεταβλητής που φανερώνει τα επίπεδα γλυκόζης στο αίμα (*blood glucose levels - glucose*). Κύρια επιδίωξη μας είναι να αποφανθούμε εάν αυτή η διαφοροποίηση είναι στατιστικά σημαντική.

Πρόκειται για δύο μεταβλητές, διαφορετικής φύσης αφού η εξαρτημένη μεταβλητή του επιπέδου γλυκόζης στο αίμα είναι συνεχής, παίρνοντας θετικές τιμές και η ανεξάρτητη μεταβλητή της κατανάλωσης έτοιμου φαγητού είναι μια διατάξιμη ποιοτική μεταβλητή με επτά επίπεδα.

Αρχικά, θα προσπαθήσουμε να απεικονίσουμε γραφικά πως κινούνται οι μέσοι των επιπέδων της ανεξάρτητης μεταβλητής. Από το line plot που ακολουθεί, παρατηρείται αρκετά έντονη διαφοροποίηση μεταξύ των οκτώ επιπέδων της μεταβλητής κατανάλωσης έτοιμου φαγητού.



Γράφημα 4.9: Line plot μεταβλητών κατανάλωσης έτοιμου φαγητού και σακχάρου αίματος



Προτού προχωρήσουμε στον έλεγχο άπονα, θα εξετάσουμε εάν ισχύουν οι προϋποθέσεις για την εφαρμογή του οι οποίοι είναι η κανονικότητα, η ομοσκεδαστικότητα και η ανεξαρτησία των σφαλμάτων.

Θα ξεκινήσουμε προχωρώντας στο στατιστικό έλεγχο κανονικότητας των Kolmogorov - Smirnov ο οποίος ελέγχει τις παρακάτω υποθέσεις:

$H_0$ : τα κατάλοιπα της μεταβλητής της κατανάλωσης έτοιμου φαγητού ακολουθούν τη κανονική κατανομή

$H_1$ : τα κατάλοιπα της μεταβλητής της κατανάλωσης έτοιμου φαγητού δεν ακολουθούν τη κανονική κατανομή

Σε επίπεδο σημαντικότητας 5% υπάρχουν ενδείξεις για απόρριψη της μηδενικής υπόθεσης της κανονικότητας της κατανάλωσης έτοιμου φαγητού αφού το p-value είναι σχεδόν μηδέν, το οποίο υπολείπεται κατά πολύ του  $\alpha=0,05$ . Τελικά η μεταβλητή της κατανάλωσης έτοιμου φαγητού δεν ακολουθεί την κανονική κατανομή.

Από τη στιγμή που απορρίψαμε την κανονικότητα, δεν υπάρχει λόγος να ελέγξουμε και τις υπόλοιπες δύο προϋποθέσεις του ελέγχου άπονα. Πλέον θα ακολουθήσουμε μη παραμετρική μέθοδο για να εξετάσουμε την διαφοροποίηση των μέσων που μας ενδιαφέρει.

One-Sample Kolmogorov-Smirnov Test		
		Unstandardized Residual
N		968
Normal Parameters <sup>a,b</sup>	Mean	.0000000
	Std. Deviation	23.18538557
	Most Extreme Differences	
	Absolute	.164
	Positive	.164
	Negative	-.122
Test Statistic		.164
Asymp. Sig. (2-tailed)		.000 <sup>c</sup>

a. Test distribution is Normal.  
 b. Calculated from data.  
 c. Lilliefors Significance Correction.

**Πίνακας 4.18:** Πίνακας ελέγχου κανονικότητας Kolmogorov – Smirnov της μεταβλητής της κατανάλωσης έτοιμου φαγητού

Οι Kruskal και Wallis πρότειναν έναν έλεγχο βασισμένο στη βαθμολογία (Rank) των παρατηρήσεων, χωρίς να υποθέσουμε κανονικότητα και ομοσκεδαστικότητα των πληθυσμών. Οι υποθέσεις αυτού του ελέγχου είναι:

$H_0$ : οι δύο πληθυσμοί έχουν την ίδια κατανομή

$H_1$ : οι δύο πληθυσμοί δεν έχουν την ίδια κατανομή

Από τον πίνακα που δίνεται παρακάτω, παρατηρούμε ότι το p-value του ελέγχου είναι 0,224 το οποίο είναι μικρότερο από ένα επίπεδο σημαντικότητας 5%. Αυτό σημαίνει ότι δεν απορρίπτουμε τη μηδενική υπόθεση του ελέγχου. Ουσιαστικά συμπεραίνουμε πως δεν υπάρχει διαφοροποίηση μεταξύ των μέσων στα επίπεδα της μεταβλητής κατανάλωσης έτοιμου φαγητού.

Test Statistics <sup>a,b</sup>	
	Blood glucose levels (mg/dl)
Chi-Square	9.411
df	7
Asymp. Sig.	.224

a. Kruskal Wallis Test  
b. Grouping Variable: fastfood meals per week

Πίνακας 4.19: Πίνακας ελέγχου Kruskal - Wallis

Εάν δεν είχαμε απορρίψει την κανονικότητα θα προχωρούσαμε σε ένα f test. Για λόγους αντιπαράθεσης των αποτελεσμάτων θα παρουσιάσουμε συνοπτικά ένα τέτοιο έλεγχο.

Για να διερευνηθεί και στατιστικά η μεταβολή των μέσων τιμών, θα προχωρήσουμε σε ένα f test. Οι υποθέσεις του ελέγχου είναι οι εξής:

$H_0$ : οι μέσοι των επιπέδων είναι ίσοι

$H_1$ : τουλάχιστον ένας μέσος διαφέρει

Από τον πίνακα άσωνα που ακολουθεί, το p-value του ελέγχου είναι ίσο με 0,032 το οποίο είναι μικρότερο από ένα επίπεδο σημαντικότητας 5%. Επομένως μπορούμε να απορρίψουμε τη μηδενική υπόθεση με σχετική σιγουριά. Επομένως, τουλάχιστον ένας μέσος διαφέρει στατιστικά.

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	21527.608	7	3075.373	2.223	.032
Within Groups	517469.274	374	1383.608		
Total	538996.882	381			

**Πίνακας 4.20:** Πίνακας ανοβα των μεταβλητών «κατανάλωσης έτοιμου φαγητού» και «επίπεδα σακχάρου αίματος»

Πρακτικά, δεν καταλήξαμε στο ίδιο αποτέλεσμα χρησιμοποιώντας παραμετρικό και μη παραμετρικό έλεγχο. Βασική επιδίωξη του συγγραφέα ήταν να τονίσει τη σημασία των προϋποθέσεων κάθε ελέγχου και ότι ο καθένας μπορεί να οδηγήσει σε εντελώς διαφορετικά συμπεράσματα.

Τέλος στο πλαίσιο μια ολοκληρωμένης παρουσίασης παραμετρικού ελέγχου, θα διερευνήσουμε μεταξύ ποιων επιπέδων της μεταβλητής της κατανάλωσης έτοιμου φαγητού, παρατηρείται στατιστικά σημαντική διαφοροποίηση.

Οι υποθέσεις του ελέγχου πολλαπλών συγκρίσεων του Bonferroni είναι οι εξής:

$H_0$ : ο μέσος του  $i$  επιπέδου είναι ίσος με του  $j$

$H_1$ : ο μέσος του  $i$  επιπέδου είναι διάφορος του  $j$

Σύμφωνα με τον πίνακα πολλαπλών συγκρίσεων του Bonferroni που ακολουθεί, παρατηρούμε ότι όλα τα  $p$ -values των ανά δύο ελέγχων, πλησιάζουν τη μονάδα με εξαίρεση τα ζευγάρια κανένα γεύμα – δύο γεύματα και ένα γεύμα – δύο γεύματα όπου τα  $p$ -values είναι 0,047 και 0,006 αντίστοιχα. Αυτό σημαίνει πως μόνο για αυτές τις δύο περιπτώσεις απορρίπτουμε τη μηδενική υπόθεση. Πρακτικά σημαίνει πως παρατηρούνται διαφοροποιήσεις στους μέσους των επιπέδων του σακχάρου στο αίμα μεταξύ των συμμετεχόντων που δεν τρώνε κανένα γεύμα την εβδομάδα με αυτούς που τρώνε δύο γεύματα και τέλος μεταξύ αυτών που τρώνε ένα γεύμα την εβδομάδα με αυτούς που τρώνε δύο γεύματα.

#### Multiple Comparisons

Dependent Variable: Blood glucose levels (mg/dl)

Bonferroni

(I) fastfood meals per week	(J) fastfood meals per week	Mean Difference (I-J)	Std. Error	Sig.
no fastfood meals	one fastfood meal	6.328	4.516	1.000
	two fastfood meals	-24.712*	7.815	.047
	three fastfood meals	9.200	11.460	1.000
	four fastfood meals	16.428	26.379	1.000
	five fastfood meals	2.761	15.358	1.000
	six fastfood meals	26.428	26.379	1.000

Κεφάλαιο 4<sup>ο</sup>: Στατιστικός Έλεγχος Υποθέσεων

	seven fastfood meals	6.213	14.248	1.000
one fastfood meal	two fastfood meals	-31.040	8.350	.006
	three fastfood meals	2.873*	11.831	1.000
	four fastfood meals	10.100	26.542	1.000
	five fastfood meals	-3.567	15.637	1.000
	six fastfood meals	20.100	26.542	1.000
	seven fastfood meals	-.114	14.548	1.000
two fastfood meals	three fastfood meals	33.913	13.440	.337
	four fastfood meals	41.140*	27.298	1.000
	five fastfood meals	27.473*	16.887	1.000
	six fastfood meals	51.140	27.298	1.000
	seven fastfood meals	30.926	15.885	1.000
three fastfood meals	four fastfood meals	7.227	28.555	1.000
	five fastfood meals	-6.439	18.853	1.000
	six fastfood meals	17.227	28.555	1.000
	seven fastfood meals	-2.987	17.961	1.000
four fastfood meals	five fastfood meals	-13.667	30.331	1.000
	six fastfood meals	10.000	37.147	1.000
	seven fastfood meals	-10.214	29.784	1.000
five fastfood meals	six fastfood meals	23.667	30.331	1.000
	seven fastfood meals	3.452	20.667	1.000
	seven fastfood meals	-20.214	29.784	1.000

**Πίνακας 4.21:** Πίνακας πολλαπλών συγκρίσεων Bonferroni των μεταβλητών «κατανάλωσης έτοιμου φαγητού» και «επίπεδα σακχάρου αίματος»

### 4.3.1 Έλεγχος ανεξαρτησίας μεταβλητής «Mediterranean Dietary Group»

Υπάρχει μια συνεχώς αύξουσα τάση στις σύγχρονες οικογένειες του δυτικού κόσμου, να μην μαγειρεύεται καθημερινά σπιτικό φαγητό αλλά να καλύπτονται οι διατροφικές ανάγκες των μελών από διάφορων ειδών καταστήματα εστίασης. Η τάση αυτή γίνεται ολοένα και πιο έντονη στα μεγάλα αστικά κέντρα όπου κυρίως λόγω έλλειψης χρόνου, η μαγειρική δεν είναι προτεραιότητα. Τα αποτελέσματα αυτών των επιλογών είναι εμφανή στο γενικό πληθυσμό όπου επιστήμονες κρούουν το κώδωνα του κινδύνου για την αύξηση της παχυσαρκίας αλλά και διάφορων ασθενειών που συσχετίζονται με την κακή διατροφή. Από τον κανόνα δεν θα μπορούσε να απουσιάζει η μεσογειακή διατροφή που εάν και θεωρείται ότι μπορεί να συμβάλει θετικά στο σακχαρώδη διαβήτη, εντούτοις εγκαταλείπεται στο βωμό του σύγχρονου τρόπου ζωής.

Σκοπός της παραγράφου λοιπόν είναι να ερευνήσουμε τη σχέση μεταξύ της μεταβλητής μεσογειακής διατροφής (*Med Dietary Score - Meddietgroup*) και της μεταβλητής του σακχαρώδη διαβήτη (*diabetes mellitus - DM*). Κύρια επιδίωξη μας είναι να μελετήσουμε την ύπαρξη γραμμικής τάσης μεταξύ των δύο αυτών μεταβλητών.

Πρόκειται για δύο μεταβλητές με διαφορετικά χαρακτηριστικά. Πιο συγκεκριμένα, η μεταβλητή μεσογειακής διατροφής είναι μια διατάξιμη μεταβλητή αφού μπορεί να έχει τρεις ιεραρχικές καταστάσεις ανάλογα με το πόσο «κοντά» βρίσκεται η διατροφή ενός ατόμου στη μεσογειακή διατροφή (μακριά, κοντά, πολύ κοντά). Αντιθέτως, η μεταβλητή του σακχαρώδη διαβήτη είναι μια κατηγορική μεταβλητή αφού μπορεί να είναι είτε θετική είτε αρνητική.

Αρχικά, θα παρουσιάσουμε τη σχέση μεταξύ των ποιοτικών μεταβλητών μέσω του πίνακα συνάφειας όπου μπορούμε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης ανάμεσα στα επίπεδα της εκπαιδευτικό-οικονομική μεταβλητής.

Ενδεικτικά αναφέρουμε πως παρουσίασε σακχαρώδη διαβήτη το 22,9% όσων δεν ακολουθούν τη μεσογειακή διατροφή, το 4,4% όσων η διατροφή πλησιάζει τη μεσογειακή και τέλος το 1,1% όσων ακολουθούν πιστά τη μεσογειακή διατροφή. Όσο πιο πιστός είναι κάποιος στη μεσογειακή διατροφή, τόσο φαίνεται να μειώνονται τα ποσοστά εμφάνισης της ασθένειας. Θα πρέπει να ερευνήσουμε εάν αυτές οι διαφορές είναι στατιστικά σημαντικές.

Med Dietary Score Group * Diabetes Mellitus Crosstabulation					
			Diabetes Mellitus		Total
			No diabetes	diabetes	
Med Dietary Score Group	Away	Count	276	82	358
		% within Med Dietary Score Group	77.1%	22.9%	100.0%
		% within Diabetes Mellitus	29.5%	82.0%	34.6%

	% of Total	26.7%	7.9%	34.6%
Close	Count	306	14	320
	% within Med Dietary Score Group	95.6%	4.4%	100.0%
	% within Diabetes Mellitus	32.7%	14.0%	30.9%
	% of Total	29.6%	1.4%	30.9%
Very close	Count	353	4	357
	% within Med Dietary Score Group	98.9%	1.1%	100.0%
	% within Diabetes Mellitus	37.8%	4.0%	34.5%
	% of Total	34.1%	0.4%	34.5%
Total	Count	935	100	1035
	% within Med Dietary Score Group	90.3%	9.7%	100.0%
	% within Diabetes Mellitus	100.0%	100.0%	100.0%
	% of Total	90.3%	9.7%	100.0%

**Πίνακας 4.22:** Πίνακας συνάφειας μεσογειακής διατροφής - σακχαρώδη διαβήτη

Για να ελεγχθεί η υπόθεση εάν υπάρχει γραμμική τάση μεταξύ της εμφάνισης της ασθένειας του σακχαρώδη διαβήτη και της μεσογειακής διατροφής, θα εφαρμόσουμε τον έλεγχο  $\chi^2$  για τάση.

Ο έλεγχος υποθέσεων διαμορφώνεται ως εξής:

$H_0$ : δεν υπάρχει γραμμική τάση μεταξύ μεσογειακής διατροφής και σακχαρώδη διαβήτη

$H_1$ : υπάρχει γραμμική τάση μεταξύ μεσογειακής διατροφής και σακχαρώδη διαβήτη

Από τον πίνακα που ακολουθεί, παρατηρούμε ότι το p-value του ελέγχου  $\chi^2$  για τάση είναι σχεδόν μηδέν που σημαίνει ότι απορρίπτουμε τη μηδενική υπόθεση σε επίπεδο σημαντικότητας  $\alpha=5\%$ . Ουσιαστικά, σημαίνει ότι υπάρχει ισχυρή ένδειξη μεταξύ μεσογειακής διατροφής και σακχαρώδη διαβήτη, δηλαδή η μεσογειακή διατροφή έχει επίδραση στο εάν θα παρουσιαστεί διαβήτης.

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	112.022 <sup>a</sup>	2	.000
Likelihood Ratio	113.226	2	.000
Linear-by-Linear Association	97.153	1	.000
N of Valid Cases	1035		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 30.92.

**Πίνακας 4.23:** Πίνακας αποτελεσμάτων στατιστικών ελέγχων

Μπορούμε να ελέγξουμε την ένταση της γραμμικής τάσης μεταξύ μιας διατάξιμης και μιας κατηγορικής μεταβλητής μέσω αρκετών μέτρων συνάφειας. Στην ανάλυσή μας θα εμπιστευτούμε τον συντελεστή  $\tau$  του Kendal (-0,289), τον  $\gamma$  των Goodman και Kruskal (-0,810) καθώς και τα κατευθυντήρια μέτρα του αμέσως επόμενου πίνακα (-0,235 & -0,565). Με βάση όλα τα μέτρα συνάφειας που υπολογίστηκαν, υπάρχει λίγη αρνητική έως μέτρια αρνητική τάση, μεταξύ της μεσογειακή διατροφής των συμμετεχόντων στην έρευνα και της εμφάνισης σακχαρώδη διαβήτη.

Directional Measures						
			Value	Asymptotic Standardized Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Ordinal by Ordinal	Somers' d	Symmetric	-.235	.018	-9.408	.000
		Med Dietary Score Group Dependent	-.565	.036	-9.408	.000
		Diabetes Mellitus Dependent	-.148	.016	-9.408	.000

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

**Πίνακας 4.24:** Πίνακας κατευθυντήριων μέτρων συνάφειας για διατάξιμους πίνακες συνάφειας

Symmetric Measures					
		Value	Asymptotic Standardized Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Ordinal by Ordinal	Kendall's tau-b	<b>-.289</b>	.022	-9.408	.000
	Kendall's tau-c	-.197	.021	-9.408	.000
	Gamma	<b>-.810</b>	.045	-9.408	.000
	Spearman Correlation	-.307	.024	-10.352	.000 <sup>c</sup>
Interval by Interval	Pearson's R	-.307	.024	-10.350	.000 <sup>c</sup>
N of Valid Cases		1035			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

**Πίνακας 4.25:** Πίνακας μέτρων συνάφειας (Kendal's  $\tau$  – Kruskal's  $\gamma$ ) για διατάξιμους πίνακες συνάφειας

#### 4.4 Έλεγχος ανεξαρτησίας μεταβλητής «Καπνιστικές συνήθειες»

Το κάπνισμα, ως η πιο βλαβερή συνήθεια της ανθρωπότητας τους τελευταίους αιώνες, συνδέεται αρνητικά σχεδόν με κάθε ανθρώπινη πάθηση ή ασθένεια. Σε προηγούμενο κεφάλαιο είχαμε συζητήσει αναλυτικά πως μπορεί να επηρεάσει και να συμβάλει το κάπνισμα στην εμφάνιση σακχαρώδη διαβήτη τύπου II.

Σκοπός της παραγράφου είναι να ερευνήσουμε τη σχέση μεταξύ της καπνιστικής συνήθειας (*smoking – smoking current*) και της μεταβλητής του σακχαρώδη διαβήτη (*diabetes mellitus - DM*). Κύρια επιδίωξη μας είναι να μελετήσουμε τη σχέση εξάρτησης μεταξύ των δύο αυτών μεταβλητών.

Πρόκειται για δύο ποιοτικές μεταβλητές, αφού το κάπνισμα έχει δυο καταστάσεις, καπνιστής / μη καπνιστής και μεταβλητή του σακχαρώδη διαβήτη μπορεί να είναι είτε θετική είτε αρνητική.

Αρχικά, θα παρουσιάσουμε τη σχέση μεταξύ των ποιοτικών μεταβλητών μέσω του πίνακα συνάφειας όπου μπορούμε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης σ' αυτούς που καπνίζουν καθημερινά και στους μη καπνιστές.

Ενδεικτικά αξίζει να αναφερθεί πως από τους συμμετέχοντες που δεν πάσχουν από σακχαρώδη διαβήτη, το 56,1% δεν καπνίζει συστηματικά ενώ το 43,9% είναι συστηματικοί καπνιστές. Παρατηρείται ποσοστιαία διαφορά της τάξεως του 12,2%, μεταξύ των συμμετεχόντων και μένει να μελετήσουμε στη συνέχεια εάν αυτή η διαφοροποίηση κρίνεται στατιστικά σημαντική.



smoking * Diabetes Mellitus Crosstabulation					
		Diabetes Mellitus		Total	
		No diabetes	diabetes		
smoking	No smoking	Count	525	72	597
		Expected Count	539.3	57.7	597.0
		% within smoking	87.9%	12.1%	100.0%
		% within Diabetes Mellitus	56.1%	72.0%	57.7%
		% of Total	50.7%	7.0%	57.7%
		smoking	Count	410	28
		Expected Count	395.7	42.3	438.0
		% within smoking	93.6%	6.4%	100.0%
		% within Diabetes Mellitus	43.9%	28.0%	42.3%
		% of Total	39.6%	2.7%	42.3%
Total		Count	935	100	1035
		Expected Count	935.0	100.0	1035.0
		% within smoking	90.3%	9.7%	100.0%
		% within Diabetes Mellitus	100.0%	100.0%	100.0%
		% of Total	90.3%	9.7%	100.0%

**Πίνακας 4.26:** Πίνακας συνάφειας καπνιστικής συνήθειας - σακχαρώδη διαβήτη, παρατηρούμενες και αναμενόμενες σχετικές συχνότητες

Για να ελεγχθεί η υπόθεση εάν υπάρχει σχέση μεταξύ της εμφάνισης της ασθένειας του σακχαρώδη διαβήτη και της καπνιστικής συνήθειας, θα εφαρμόσουμε τον έλεγχο ανεξαρτησίας  $\chi^2$  του Pearson.

Ο έλεγχος υποθέσεων διαμορφώνεται ως εξής:

$H_0$ : δεν υπάρχει εξάρτηση μεταξύ της καπνιστικής συνήθειας και σακχαρώδη διαβήτη

$H_1$ : υπάρχει εξάρτηση μεταξύ καπνιστικής συνήθειας και σακχαρώδη διαβήτη

Από τον πίνακα που ακολουθεί, παρατηρούμε ότι το p-value του ελέγχου  $\chi^2$  του Pearson ισούται με 0,002 που σημαίνει ότι απορρίπτουμε τη μηδενική υπόθεση σε επίπεδο σημαντικότητας  $\alpha=5\%$ . Ουσιαστικά, σημαίνει ότι υπάρχει σχέση μεταξύ της καπνιστικής συνήθειας και του σακχαρώδη διαβήτη, δηλαδή το κάπνισμα έχει επίδραση στο εάν θα παρουσιαστεί διαβήτης.

Chi-Square Tests					
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	9.298 <sup>a</sup>	1	.002		
Continuity Correction <sup>b</sup>	8.660	1	.003		
Likelihood Ratio	9.696	1	.002		
Fisher's Exact Test				.003	.001
Linear-by-Linear Association	9.289	1	.002		
N of Valid Cases	1035				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 42.32.

b. Computed only for a 2x2 table

**Πίνακας 4.27:** Πίνακας αποτελεσμάτων στατιστικών ελέγχων

Μπορούμε να ελέγξουμε την ένταση της συσχέτισης μεταξύ δύο μη διατάξιμων μεταβλητών μέσω αρκετών μέτρων συνάφειας. Στην ανάλυσή μας θα εμπιστευτούμε τον συντελεστή  $\Phi$  του Pearson και τον  $V$  του Cramer. Οι δύο πρώτες γραμμές του παρακάτω πίνακα αναφέρονται σ' αυτούς τους συντελεστές, οι οποίοι είναι ίσοι με  $\pm 0,095$ . Αυτό μας οδηγεί στο συμπέρασμα ότι υπάρχει μικρής έντασης συσχέτιση μεταξύ του καπνίσματος και του σακχαρώδη διαβήτη.

Symmetric Measures					
		Value	Asymptotic Standard Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Nominal by Nominal	Phi	-.095			.002
	Cramer's V	.095			.002
	Contingency Coefficient	.094			.002
Interval by Interval	Pearson's R	-.095	.029	-3.060	.002 <sup>c</sup>
Ordinal by Ordinal	Spearman Correlation	-.095	.029	-3.060	.002 <sup>c</sup>
N of Valid Cases		1035			

Πίνακας 4.28: Πίνακας μέτρων συνάφειας (Pearson's  $\Phi$  – Cramer's V) για μη διατάξιμους πίνακες συνάφειας

Σε συνέχεια της προηγούμενης παραγράφου, θα παρουσιάσουμε κάποια μέτρα συνάφειας επιπλέον, τα οποία έχουν ιδιαίτερα σημαντική ερμηνεία.

Στη πρώτη γραμμή του πίνακα που ακολουθεί, υπολογίζεται το πηλίκο πιθανοτήτων, odds ratio, το οποίο ισούται με 0,498. Μια πρώτη ερμηνεία λέει ότι είναι 50,2% ( $1 - 0,498 = 0,502$ ) πιθανότερο να νοσήσει κάποιος από σακχαρώδη διαβήτη που καπνίζει συστηματικά σε σχέση με κάποιον που δεν καπνίζει. Μια δεύτερη προσέγγιση του πηλίκου πιθανοτήτων λέει πως η σχετική πιθανότητα να νοσήσει κάποιος που καπνίζει από σακχαρώδη διαβήτη είναι 2,008 ( $1 / 0,498$ ) φορές μεγαλύτερη από την αντίστοιχη σχετική πιθανότητα να νοσήσει από σακχαρώδη διαβήτη κάποιος δεν καπνίζει.

Επιπλέον στον ίδιο πίνακα δίνεται και η τιμή του σχετικού κινδύνου (relative risk) που ισούται με 0,939. Αυτό ουσιαστικά σημαίνει ότι τα άτομα που δεν καπνίζουν, έχουν 6,1% ( $1 - 0,939 = 0,061$ ) μικρότερο κίνδυνο να νοσήσουν από σακχαρώδη διαβήτη σε σχέση με αυτούς που επιλέγουν να καπνίζουν συστηματικά.

Risk Estimate			
	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for smoking (No smoking / smoking)	.498	.316	.785
For cohort Diabetes Mellitus = No diabetes	.939	.904	.976
For cohort Diabetes Mellitus = diabetes	1.887	1.241	2.868
N of Valid Cases	1035		

Πίνακας 4.29: Πίνακας μέτρων συνάφειας (Odds Ratio – Relative Risk) για μη διατάξιμους πίνακες συνάφειας

## 4.5 Έλεγχος ανεξαρτησίας μεταβλητών κλινικής εικόνας συμμετεχόντων

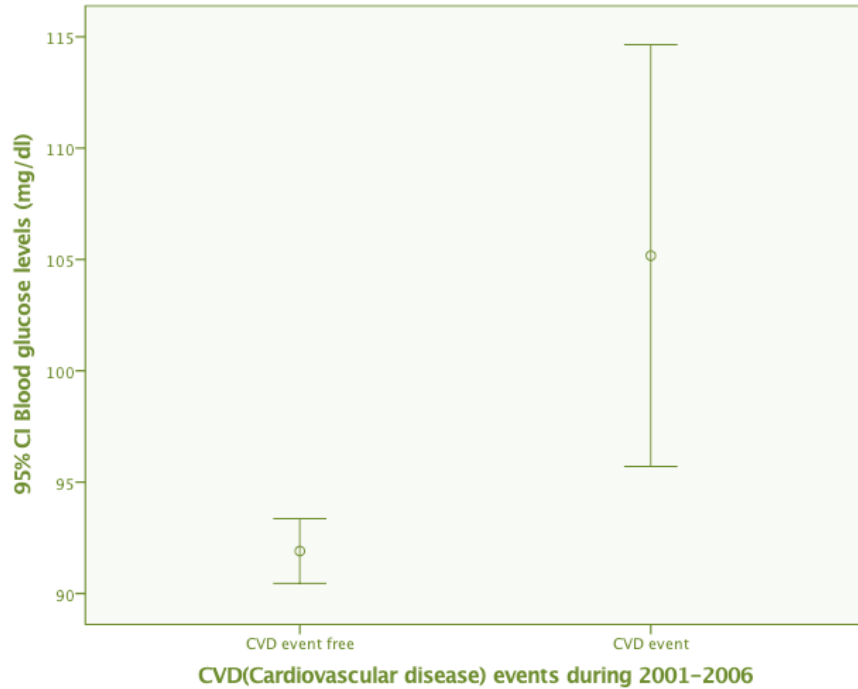
Με την όρο κλινική εικόνα εννοούμε τα συμπτώματα που μπορεί να εμφανίζει ένας ασθενής αλλά και τα αποτελέσματα των εξετάσεων που ενδεχομένως υποβλήθηκε.

### 4.5.1 Έλεγχος σύγκρισης μέσω των όρων της μεταβλητής «Καρδιαγγειακά Επεισόδια»

Η εμφάνιση καρδιαγγειακών επεισοδίων είναι πολύ σοβαρή υπόθεση για την υγεία των ασθενών αφού πολλές φορές οδηγούν ακόμη και στον θάνατο. Οι ειδικοί ισχυρίζονται πως υπάρχει σύνδεση μεταξύ του σακχαρώδη διαβήτη και των καρδιαγγειακών επεισοδίων.

Σκοπός της παραγράφου είναι να ερευνήσουμε εάν διαφοροποιείται ο μέσος του επιπέδου γλυκόζης στο αίμα (*blood glucose levels - glucose*) μεταξύ των συμμετεχόντων που είχαν στο παρελθόν κάποιο καρδιαγγειακό επεισόδιο και μεταξύ αυτών που δεν είχαν (*cardiovascular disease events - CVD*).

Προτού προχωρήσουμε στο στατιστικό έλεγχο, θα πάρουμε μια πρώτη διαγραμματική άποψη για το πως κινούνται οι μέσοι των επιπέδων του σακχάρου στο αίμα. Από το διάγραμμα που ακολουθεί, παρατηρούμε πως υπάρχουν μεγάλες αποκλίσεις μεταξύ των συμμετεχόντων που είχαν στο παρελθόν κάποιο καρδιαγγειακό επεισόδιο και σε αυτούς που δεν είχαν.



**Γράφημα 4.10:** Error bar μεταβλητών κατανάλωσης καρδιαγγειακών επεισοδίων και σακχάρου αίματος

Προτού προχωρήσουμε στον έλεγχο  $t$ , θα εξετάσουμε εάν ισχύουν οι προϋποθέσεις για την εφαρμογή του οι οποίοι είναι η κανονικότητα, η ομοσκεδαστικότητα και η ανεξαρτησία των σφαλμάτων.

Θα ξεκινήσουμε προχωρώντας στο στατιστικό έλεγχο κανονικότητας των Kolmogorov - Smirnov ο οποίος ελέγχει τις παρακάτω υποθέσεις:

- $H_0$ : τα κατάλοιπα της μεταβλητής των καρδιαγγειακών επεισοδίων ακολουθούν τη κανονική κατανομή  
 $H_1$ : τα κατάλοιπα της μεταβλητής των καρδιαγγειακών επεισοδίων δεν ακολουθούν τη κανονική κατανομή

Σε επίπεδο σημαντικότητας 5% υπάρχουν ενδείξεις για απόρριψη της μηδενικής υπόθεσης της κανονικότητας των καρδιαγγειακών επεισοδίων αφού το  $p$ -value είναι σχεδόν μηδέν, το οποίο υπολείπεται κατά πολύ του  $\alpha=0,05$ . Τελικά η μεταβλητή των καρδιαγγειακών επεισοδίων δεν ακολουθεί την κανονική κατανομή.

Από τη στιγμή που απορρίψαμε την κανονικότητα, δεν υπάρχει λόγος να ελέγξουμε και τις υπόλοιπες δύο προϋποθέσεις του ελέγχου  $t$ . Πλέον θα ακολουθήσουμε μη παραμετρική μέθοδο για να εξετάσουμε την διαφοροποίηση των μέσων που μας ενδιαφέρει.

One-Sample Kolmogorov-Smirnov Test		
		Unstandardized Residual
N		968
Normal Parameters <sup>a,b</sup>	Mean	.0000000
	Std. Deviation	23.18538557
Most Extreme Differences	Absolute	.164
	Positive	.164
	Negative	-.122
Test Statistic		.164
Asymp. Sig. (2-tailed)		.000 <sup>c</sup>

a. Test distribution is Normal.  
b. Calculated from data.  
c. Lilliefors Significance Correction.

**Πίνακας 4.30:** Πίνακας ελέγχου κανονικότητας Kolmogorov – Smirnov της μεταβλητής της κατανάλωσης έτοιμου φαγητού

Ο έλεγχος Sum – Rank του Wilcoxon και ο έλεγχος U των Mann – Whitney χρησιμοποιείται αντί του ελέγχου t για δύο ανεξάρτητα δείγματα, όταν οι ελεγχόμενοι πληθυσμοί δεν κατανέμονται κανονικά. Οι παραπάνω έλεγχοι είναι ισοδύναμοι στη πράξη. Οι υποθέσεις του ελέγχου είναι:

$H_0$ : οι δύο πληθυσμοί έχουν την ίδια κατανομή ή ισοδύναμα έχουν την ίδια διάμεσο

$H_1$ : οι δύο πληθυσμοί δεν έχουν την ίδια κατανομή ή ισοδύναμα δεν έχουν την ίδια διάμεσο

Από τον πίνακα που δίνεται παρακάτω, παρατηρούμε ότι το p-value του ελέγχου είναι 0,006 το οποίο είναι μικρότερο από ένα επίπεδο σημαντικότητας 5%. Αυτό σημαίνει ότι απορρίπτουμε τη μηδενική υπόθεση του ελέγχου. Ουσιαστικά συμπεραίνουμε πως υπάρχει διαφοροποίηση μεταξύ των μέσων στα επίπεδα της μεταβλητής εμφάνισης καρδιαγγειακών επεισοδίων.

Test Statistics <sup>a</sup>	
	Blood glucose levels (mg/dl)
Mann-Whitney U	18445.500
Wilcoxon W	438431.500
Z	-2.739
Asymp. Sig. (2-tailed)	.006

**Πίνακας 4.31:** Πίνακας ελέγχου Mann – Whitney - Wilcoxon

## 4.5.2 Έλεγχος ανεξαρτησίας μεταβλητής «Διαστολική Αρτηριακή Πίεση»

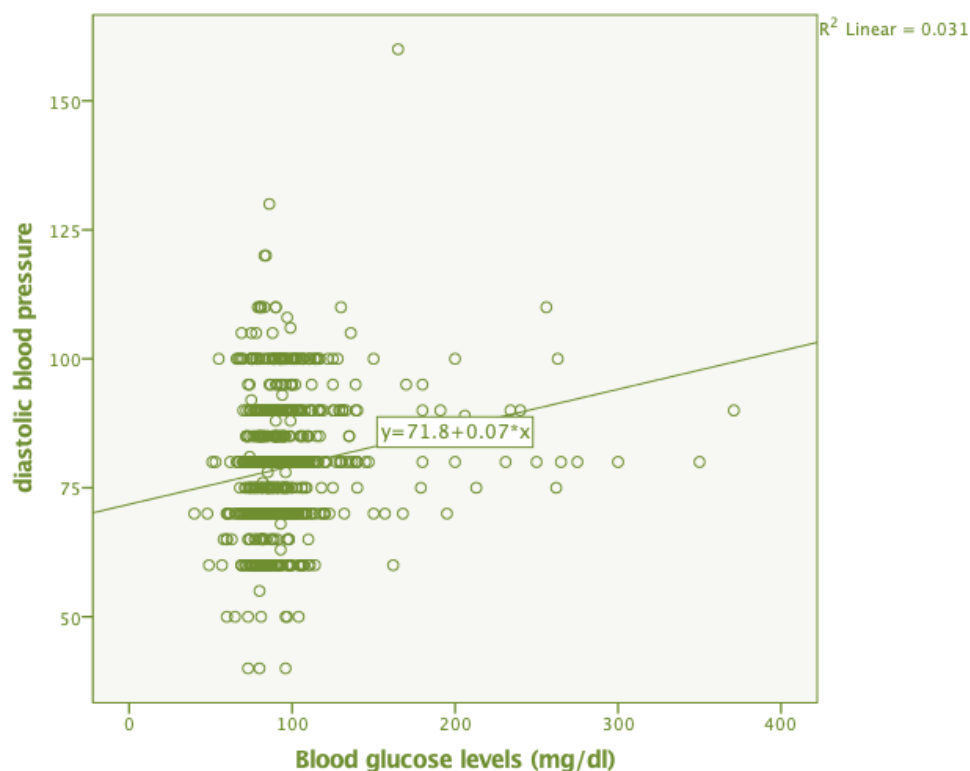
Έρευνες έχουν δείξει πως οι μισοί περίπου περίπου ασθενείς με σακχαρώδη διαβήτη τύπου II, παρουσιάζουν αρτηριακή υπέρταση, ενώ ένα ακόμα 25% περίπου θα εμφανίσει κατά την εξέλιξη της πάθησης.

Σκοπός της παραγράφου είναι να ερευνήσουμε την ύπαρξη συσχέτισης μεταξύ της μεταβλητής της διαστολικής αρτηριακής πίεσης (*diastolic blood pressure- dbp*) και της μεταβλητής που φανερώνει τα επίπεδα γλυκόζης στο αίμα (*blood glucose levels - glucose*). Κύρια επιδίωξη μας είναι να μελετήσουμε το επίπεδο της έντασης της συσχέτισης μεταξύ των δύο αυτών μεταβλητών.

Προτού προχωρήσουμε στον έλεγχο συσχέτισης του Pearson, θα πρέπει να ικανοποιούνται κάποιες βασικές προϋποθέσεις. Αρχικά θα πρέπει τόσο η εξαρτημένη, όσο και η ανεξάρτητη μεταβλητή να είναι συνεχείς, να είναι γραμμικές μεταξύ τους και να ικανοποιείται η κανονικότητα του δείγματος.

Πρόκειται για δύο συνεχής μεταβλητές, αφού τόσο ο δείκτης μάζας σώματος όσο και τα επίπεδα σακχάρου στο αίμα μπορούν να πάρουν θετικές τιμές.

Από το διάγραμμα που ακολουθεί, παρατηρούμε πως μεταξύ των μεταβλητών δεν υπάρχει γραμμική σχέση, γεγονός που επιβεβαιώνεται από τις παρατηρήσεις οι οποίες δεν βρίσκονται σε καμία περίπτωση κοντά στο ευθύγραμμο τμήμα.



**Γράφημα 4.11:** Διάγραμμα διασποράς μεταβλητών διαστολικής αρτηριακής πίεσης και σακχάρου αίματος

Παρότι δεν ικανοποιείται η πρώτη προϋπόθεση του ελέγχου συσχετίσεων του Pearson, θα προχωρήσουμε και στον έλεγχο των υπόλοιπων προϋποθέσεων.

Στη συνέχεια θα προχωρήσουμε στο στατιστικό έλεγχο κανονικότητας των Kolmogorov - Smirnov ο οποίος ελέγχει τις παρακάτω υποθέσεις:

$H_0$ : η μεταβλητή της διαστολικής αρτηριακής πίεσης ακολουθεί τη κανονική κατανομή  
 $H_1$ : η μεταβλητή της διαστολικής αρτηριακής πίεσης δεν ακολουθεί τη κανονική κατανομή

Σε επίπεδο σημαντικότητας 5% υπάρχουν ενδείξεις για απόρριψη της μηδενικής υπόθεσης της κανονικότητας της της διαστολικής αρτηριακής πίεσης αφού το p-value είναι σχεδόν μηδέν, το οποίο υπολείπεται κατά πολύ του  $\alpha=0,05$ . Τελικά η μεταβλητή της διαστολικής αρτηριακής πίεσης δεν ακολουθεί την κανονική κατανομή.

One-Sample Kolmogorov-Smirnov Test		
		diastolic blood pressure
N		956
Normal Parameters <sup>a,b</sup>	Mean	78.93
	Std. Deviation	12.175
Most Extreme Differences	Absolute Positive	.175
	Negative	-.126
Test Statistic		.175
Asymp. Sig. (2-tailed)		.000 <sup>c</sup>
a. Test distribution is Normal.		
b. Calculated from data.		
c. Lilliefors Significance Correction.		

**Πίνακας 4.32:** Πίνακας ελέγχου κανονικότητας Kolmogorov – Smirnov της μεταβλητής διαστολική αρτηριακή πίεση

Αφού οι δύο από τις τρεις προϋποθέσεις του ελέγχου δεν εκπληρώθηκαν, θα περάσουμε στον αντίστοιχο έλεγχο συσχετίσεων των Kendall και Spearman.

Στο στατιστικό έλεγχο συσχετίσεων των Kendall και Spearman ελέγχονται οι παρακάτω υποθέσεις:

$H_0$ : δεν υπάρχει συσχέτιση μεταξύ των μεταβλητών διαστολική αρτηριακή πίεση και επίπεδα σακχάρου στο αίμα

$H_1$ : υπάρχει συσχέτιση μεταξύ των μεταβλητών διαστολική αρτηριακή πίεση και επίπεδα σακχάρου στο αίμα

Τα p-values και των δύο ελέγχων βρέθηκαν σχεδόν μηδέν με αποτέλεσμα να απορρίπτουμε την μηδενική υπόθεση, αφού είναι μικρότερα από ένα επίπεδο σημαντικότητας της τάξεως του 5%. Ουσιαστικά υπάρχει συσχέτιση μεταξύ των μεταβλητών διαστολική αρτηριακή πίεση και επίπεδα σακχάρου στο αίμα.



Correlations				
			diastolic blood pressure	Blood glucose levels (mg/dl)
Kendall's tau_b	diastolic blood pressure	Correlation Coefficient	1.000	.149**
		Sig. (2-tailed)	.	.000
	N		956	918
	Blood glucose levels (mg/dl)	Correlation Coefficient	.149**	1.000
Sig. (2-tailed)		.000	.	
N		918	990	
Spearman's rho	diastolic blood pressure	Correlation Coefficient	1.000	.202**
		Sig. (2-tailed)	.	.000
	N		956	918
	Blood glucose levels (mg/dl)	Correlation Coefficient	.202**	1.000
Sig. (2-tailed)		.000	.	
N		918	990	

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

**Πίνακας 4.33:** Πίνακας ελέγχου συσχετίσεων Kendall - Spearman των μεταβλητών «διαστολική αρτηριακή πίεση» και «επίπεδα σακχάρου αίματος»

#### 4.5.3 Έλεγχος ανεξαρτησίας μεταβλητής «υπέρτασης»

Σακχαρώδης διαβήτης και υπέρταση πάνε παρέα, λένε οι ειδικοί αφού οι διαβητικοί έχουν συνήθως διπλάσια έως τριπλάσια υπέρταση από τους υγιείς. Μάλιστα είναι τόσο στενή η σχέση τους όπου όχι μόνο οι διαβητικοί έχουν υπέρταση αλλά τα άτομα που έχουν υπέρταση θα εμφανίσουν πολύ συχνότερα διαβήτη από κάποιον μη υπερτασικό. Οι επιπλοκές μπορεί να αποδειχθούν ολέθριες γι' αυτό είναι κρίσιμο να διαγνωστεί μια από τις δύο ασθένειες το συντομότερο δυνατό.

Σκοπός της παραγράφου είναι να ερευνήσουμε τη σχέση μεταξύ της υπέρτασης (*hypertension* – *htn*) και της μεταβλητής του σακχαρώδη διαβήτη (*diabetes mellitus* - *DM*). Κύρια επιδίωξη μας είναι να μελετήσουμε τη σχέση εξάρτησης μεταξύ των δύο αυτών μεταβλητών.

Πρόκειται για δύο ποιοτικές μεταβλητές, αφού η υπέρταση έχει δυο καταστάσεις, υπερτασικός / μη υπερτασικός και η μεταβλητή του σακχαρώδη διαβήτη μπορεί να είναι είτε θετική είτε αρνητική.

Αρχικά, θα παρουσιάσουμε τη σχέση μεταξύ των ποιοτικών μεταβλητών μέσω του πίνακα συνάφειας όπου μπορούμε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης σ' αυτούς που έχουν υψηλή υπέρταση.

Ενδεικτικά αξίζει να αναφερθεί πως από τους συμμετέχοντες που δεν πάσχουν από υπέρταση, το 93,9% δεν έχει σακχαρώδη διαβήτη ενώ το 80,7% έχει υπέρταση χωρίς να έχει αναπτύξει διαβήτη παράλληλα. Παρατηρείται σημαντική ποσοστιαία διαφορά μεταξύ των συμμετεχόντων και μένει να μελετήσουμε στη συνέχεια εάν αυτή η διαφοροποίηση κρίνεται στατιστικά σημαντική.

Hypertension * Diabetes Mellitus Crosstabulation					
			Diabetes Mellitus		Total
			No diabetes	diabetes	
Hypertensio n	Normal	Count	630	41	671
		Expected Count	603.0	68.0	671.0
		% within Hypertension	93.9%	6.1%	100.0%
		% within Diabetes Mellitus	72.5%	41.8%	69.4%
		% of Total	65.1%	4.2%	69.4%
	Hypertensiv e	Count	239	57	296
		Expected Count	266.0	30.0	296.0
		% within Hypertension	80.7%	19.3%	100.0%
		% within Diabetes Mellitus	27.5%	58.2%	30.6%
		% of Total	24.7%	5.9%	30.6%
Total	Count	869	98	967	
	Expected Count	869.0	98.0	967.0	
	% within Hypertension	89.9%	10.1%	100.0%	
	% within Diabetes Mellitus	100.0%	100.0%	100.0%	
	% of Total	89.9%	10.1%	100.0%	

**Πίνακας 4.34:** Πίνακας συνάφειας υπέρτασης - σακχαρώδη διαβήτη, παρατηρούμενες και αναμενόμενες σχετικές συχνότητες

Για να ελεγχθεί η υπόθεση εάν υπάρχει σχέση μεταξύ της εμφάνισης της ασθένειας του σακχαρώδη διαβήτη και της υπέρτασης, θα εφαρμόσουμε τον έλεγχο ανεξαρτησίας  $\chi^2$  του Pearson.

Ο έλεγχος υποθέσεων διαμορφώνεται ως εξής:

$H_0$ : δεν υπάρχει εξάρτηση μεταξύ της υπέρτασης και σακχαρώδη διαβήτη

$H_1$ : υπάρχει εξάρτηση μεταξύ υπέρτασης και σακχαρώδη διαβήτη

Από τον πίνακα που ακολουθεί, παρατηρούμε ότι το p-value του ελέγχου  $\chi^2$  του Pearson είναι σχεδόν μηδέν, που σημαίνει ότι απορρίπτουμε τη μηδενική υπόθεση σε επίπεδο σημαντικότητας  $\alpha=5\%$ . Ουσιαστικά, σημαίνει ότι υπάρχει σχέση μεταξύ της υπέρτασης και του σακχαρώδη διαβήτη, δηλαδή η υπέρταση έχει επίδραση στο εάν θα παρουσιαστεί διαβήτης.

Chi-Square Tests					
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	38.977 <sup>a</sup>	1	.000		
Continuity Correction <sup>b</sup>	37.547	1	.000		
Likelihood Ratio	35.720	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	38.937	1	.000		
N of Valid Cases	967				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 30.00.

b. Computed only for a 2x2 table

Πίνακας 4.35: Πίνακας αποτελεσμάτων στατιστικών ελέγχων

Μπορούμε να ελέγξουμε την ένταση της συσχέτισης μεταξύ δύο μη διατάξιμων μεταβλητών μέσω αρκετών μέτρων συνάφειας. Στην ανάλυσή μας θα εμπιστευτούμε τον συντελεστή  $\Phi$  του Pearson και τον  $V$  του Cramer. Οι δύο πρώτες γραμμές του παρακάτω πίνακα αναφέρονται σ' αυτούς τους συντελεστές, οι οποίοι είναι ίσοι με 0,201. Αυτό μας οδηγεί στο συμπέρασμα ότι υπάρχει μικρή θετική συσχέτιση μεταξύ της υπέρτασης και του σακχαρώδη διαβήτη.

Symmetric Measures					
		Value	Asymptotic Standardized Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Nominal by Nominal	Phi	.201			.000
	Cramer's V	.201			.000
	Contingency Coefficient	.197			.000
Interval by Interval	Pearson's R	.201	.035	6.366	.000 <sup>c</sup>
Ordinal by Ordinal	Spearman Correlation	.201	.035	6.366	.000 <sup>c</sup>
N of Valid Cases		967			

**Πίνακας 4.36:** Πίνακας μέτρων συνάφειας (Pearson's Φ – Cramer's V) για μη διατάξιμους πίνακες συνάφειας

Σε συνέχεια της προηγούμενης παραγράφου, θα παρουσιάσουμε κάποια μέτρα συνάφειας επιπλέον, τα οποία έχουν ιδιαίτερα σημαντική ερμηνεία.

Στη πρώτη γραμμή του πίνακα που ακολουθεί, υπολογίζεται το πηλίκο πιθανοτήτων, odds ratio, το οποίο ισούται με 3,665. Αυτό ουσιαστικά σημαίνει ότι η σχετική πιθανότητα να ασθενήσει από σακχαρώδη διαβήτη ένα υπερτασικό άτομο είναι σχεδόν 3,7 φορές μεγαλύτερη από την αντίστοιχη σχετική πιθανότητα για ένα μη υπερτασικό.

Επιπλέον στον ίδιο πίνακα δίνεται και η τιμή του σχετικού κινδύνου (relative risk) που ισούται με 0,317. Αυτό ουσιαστικά σημαίνει ότι οι μη υπερτασικοί έχουν περίπου 68,3% ( $1-0,317=0,683$ ) μικρότερο κίνδυνο να νοσήσουν από σακχαρώδη διαβήτη σε σχέση με τους υπερτασικούς.

Risk Estimate			
	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Hypertension (Normal / Hypertensive)	3.665	2.388	5.623
For cohort Diabetes Mellitus = No diabetes	1.163	1.096	1.233
For cohort Diabetes Mellitus = diabetes	.317	.218	.463
N of Valid Cases	967		

**Πίνακας 4.37:** Πίνακας μέτρων συνάφειας (Odds Ratio – Relative Risk) για μη διατάξιμους πίνακες συνάφειας

#### 4.5.4 Έλεγχος ανεξαρτησίας μεταβλητής υπερχοληστερολαιμίας

Υπερχοληστερολαιμία ονομάζεται η παρουσία μεγάλων συγκεντρώσεων χοληστερόλης στο αίμα. Πρόκειται για μια πολύ σοβαρή πάθηση η οποία μπορεί να οδηγήσει σε έμφραγμα ή εγκεφαλικό. Η υπερχοληστερολαιμία ίσως οφείλεται στη παχυσαρκία, στη κληρονομικότητα ή στον σακχαρώδη διαβήτη.

Σκοπός της παραγράφου είναι να ερευνήσουμε τη σχέση μεταξύ της υπερχοληστερολαιμίας (*hypercholesterolemia* – **hchol**) και της μεταβλητής του σακχαρώδη διαβήτη (*diabetes mellitus* - **DM**). Κύρια επιδίωξη μας είναι να μελετήσουμε τη σχέση εξάρτησης μεταξύ των δύο αυτών μεταβλητών.

Πρόκειται για δύο ποιοτικές μεταβλητές, αφού η υπερχοληστερολαιμία έχει δυο καταστάσεις, ύπαρξη μεγάλων τιμών χοληστερόλης / κανονικές τιμές χοληστερόλης και η μεταβλητή του σακχαρώδη διαβήτη μπορεί να είναι είτε θετική είτε αρνητική.

Αρχικά, θα παρουσιάσουμε τη σχέση μεταξύ των ποιοτικών μεταβλητών μέσω του πίνακα συνάφειας όπου μπορούμε να έχουμε μια πρώτη εικόνα από τις συχνότητες εμφάνισης της πάθησης σ' αυτούς που έχουν υψηλές τιμές χοληστερόλης.

Ενδεικτικά αξίζει να αναφερθεί πως από τους συμμετέχοντες που δεν πάσχουν από υπερχοληστερολαιμία, το 92,5% δεν έχει σακχαρώδη διαβήτη ενώ το 87,3% έχει υπερχοληστερολαιμία χωρίς να έχει αναπτύξει διαβήτη παράλληλα. Παρατηρείται ποσοστιαία διαφορά μεταξύ των συμμετεχόντων και μένει να μελετήσουμε στη συνέχεια εάν αυτή η διαφοροποίηση κρίνεται στατιστικά σημαντική.

**Hypercholesterolemia \* Diabetes Mellitus Crosstabulation**

			Diabetes Mellitus		Total
			No diabetes	diabetes	
Hypercholesterolemia Normal	Count		556	45	601
	Expected Count		542.9	58.1	601.0
	% within Hypercholesterolemia		92.5%	7.5%	100.0%
	% within Diabetes Mellitus		59.5%	45.0%	58.1%
	% of Total		53.7%	4.3%	58.1%
Hypercholesterolemia	Count		379	55	434
	Expected Count		392.1	41.9	434.0
	% within Hypercholesterolemia		87.3%	12.7%	100.0%

## Κεφάλαιο 4<sup>ο</sup>: Στατιστικός Έλεγχος Υποθέσεων

	% within Diabetes Mellitus	40.5%	55.0%	41.9%
	% of Total	36.6%	5.3%	41.9%
Total	Count	935	100	1035
	Expected Count	935.0	100.0	1035.0
	% within Hypercholesterolemia	90.3%	9.7%	100.0%
	% within Diabetes Mellitus	100.0%	100.0%	100.0%
	% of Total	90.3%	9.7%	100.0%

**Πίνακας 4.38:** Πίνακας συνάφειας υπερχοληστερολαιμία - σακχαρώδη διαβήτη, παρατηρούμενες και αναμενόμενες σχετικές συχνότητες

Για να ελεγχθεί η υπόθεση εάν υπάρχει σχέση μεταξύ της εμφάνισης της ασθένειας του σακχαρώδη διαβήτη και της υπέρτασης, θα εφαρμόσουμε τον έλεγχο ανεξαρτησίας  $\chi^2$  του Pearson.

Ο έλεγχος υποθέσεων διαμορφώνεται ως εξής:

$H_0$ : δεν υπάρχει εξάρτηση μεταξύ της υπερχοληστερολαιμίας και σακχαρώδη διαβήτη  
 $H_1$ : υπάρχει εξάρτηση μεταξύ υπερχοληστερολαιμίας και σακχαρώδη διαβήτη

Από τον πίνακα που ακολουθεί, παρατηρούμε ότι το p-value του ελέγχου  $\chi^2$  του Pearson είναι ίσο με 0,005, που σημαίνει ότι απορρίπτουμε τη μηδενική υπόθεση σε επίπεδο σημαντικότητας  $\alpha=5\%$ . Ουσιαστικά, σημαίνει ότι υπάρχει σχέση μεταξύ της υπερχοληστερολαιμίας και του σακχαρώδη διαβήτη, δηλαδή η υπερχοληστερολαιμία έχει επίδραση στο εάν θα παρουσιαστεί διαβήτης.

Chi-Square Tests					
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	7.763 <sup>a</sup>	1	.005		
Continuity Correction <sup>b</sup>	7.180	1	.007		
Likelihood Ratio	7.648	1	.006		
Fisher's Exact Test				.007	.004

Linear-by-Linear Association	7.756	1	.005		
N of Valid Cases	1035				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 41.93.

b. Computed only for a 2x2 table

**Πίνακας 4.39:** Πίνακας αποτελεσμάτων στατιστικών ελέγχων

Μπορούμε να ελέγξουμε την ένταση της συσχέτισης μεταξύ δύο μη διατάξιμων μεταβλητών μέσω αρκετών μέτρων συνάφειας. Στην ανάλυσή μας θα εμπιστευτούμε τον συντελεστή  $\Phi$  του Pearson και τον  $V$  του Cramer. Οι δύο πρώτες γραμμές του παρακάτω πίνακα αναφέρονται σ' αυτούς τους συντελεστές, οι οποίοι είναι ίσοι με 0,087. Αυτό μας οδηγεί στο συμπέρασμα ότι ελάχιστη μικρή θετική συσχέτιση μεταξύ της υπερχοληστερολαιμίας και του σακχαρώδη διαβήτη.

Symmetric Measures					
		Value	Asymptotic Standardized Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Nominal by Nominal	Phi	.087			.005
	Cramer's V	.087			.005
	Contingency Coefficient	.086			.005
Interval by Interval	Pearson's R	.087	.032	2.794	.005 <sup>c</sup>
Ordinal by Ordinal	Spearman Correlation	.087	.032	2.794	.005 <sup>c</sup>
N of Valid Cases		1035			

**Πίνακας 4.40:** Πίνακας μέτρων συνάφειας (Pearson's  $\Phi$  – Cramer's  $V$ ) για μη διατάξιμους πίνακες συνάφειας

Σε συνέχεια της προηγούμενης παραγράφου, θα παρουσιάσουμε κάποια μέτρα συνάφειας επιπλέον, τα οποία έχουν ιδιαίτερα σημαντική ερμηνεία.

Στη πρώτη γραμμή του πίνακα που ακολουθεί, υπολογίζεται το πηλίκο πιθανοτήτων, odds ratio, το οποίο ισούται με 1,793. Αυτό ουσιαστικά σημαίνει ότι η σχετική πιθανότητα να ασθενήσει από σακχαρώδη διαβήτη ένα άτομο με υψηλή χοληστερόλη είναι σχεδόν 1,8 φορές μεγαλύτερη από την αντίστοιχη σχετική πιθανότητα για ένα άτομο που έχει κανονικά επίπεδα χοληστερόλης.

Επιπλέον στον ίδιο πίνακα δίνεται και η τιμή του σχετικού κινδύνου (relative risk) που ισούται με 0,591. Αυτό ουσιαστικά σημαίνει ότι τα άτομα με κανονικά επίπεδα χοληστερόλης έχουν περίπου 40,9% ( $1-0,591=0,409$ ) μικρότερο κίνδυνο να νοσήσουν από σακχαρώδη διαβήτη σε σχέση με τα άτομα που έχουν υψηλές συγκεντρώσεις χοληστερόλης στο αίμα.

Risk Estimate			
	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Hypercholesterolemia (Normal / Hypercholesterolemia)	1.793	1.184	2.715
For cohort Diabetes Mellitus = No diabetes	1.059	1.015	1.105
For cohort Diabetes Mellitus = diabetes	.591	.406	.859
N of Valid Cases	1035		

Πίνακας 4.41: Πίνακας μέτρων συνάφειας (Odds Ratio – Relative Risk) για μη διατάξιμους πίνακες συνάφειας

## 4.6 Ανακεφαλαίωση και σχόλια

Στο 4<sup>ο</sup> κεφάλαιο μελετήσαμε όσες μεταβλητές συσχετίζονται σε κάποιο βαθμό με το σακχαρώδη διαβήτη. Παρόλα αυτά βρέθηκαν και κάποιες οι οποίες δεν φαίνεται από το δείγμα της παρούσας έρευνας να συσχετίζονται. Πιο συγκεκριμένα η μεταβλητή της κατανάλωσης έτοιμου φαγητού με p-value 0,224, η μεταβλητή κατανάλωσης καφέ με p-value ίσο με 0,85, και η μεταβλητή της φυσικής δραστηριότητας των συμμετεχόντων με 0,919. Κλείνοντας θα θέλαμε να αναφέρουμε ότι επειδή με το συγκεκριμένο δείγμα φαίνεται να μην υπάρχει εξάρτηση ή συσχέτιση των παραπάνω μεταβλητών με τον σακχαρώδη διαβήτη, δε πάει να πει αυτομάτως πως δεν έχουν καμία σχέση με την ασθένεια. Υπάρχουν πολλές έρευνες που ισχυρίζονται το αντίθετο και ενδεχομένως με άλλο δείγμα ατόμων να φθάναμε και εμείς στα ίδια συμπεράσματα.



# ΚΕΦΑΛΑΙΟ 5<sup>ο</sup>

## Παλινδρόμηση

---

### 5.1 Τεχνικές παλινδρόμησης

Μια από τις πιο ευρέως διαδομένες στατιστικές τεχνικές μοντελοποίησης και συσχέτισης μιας εξαρτημένης με μια ή περισσότερες ανεξάρτητες μεταβλητές, είναι η παλινδρόμηση.

Επιπροσθέτως, η παλινδρόμηση μπορεί να χρησιμοποιηθεί και ως μια τεχνική εξόρυξης δεδομένων, όπου ουσιαστικά θα αποτελεί ένα μοντέλο προβλέψεων των τιμών μια κατηγορίας για τα «νέα» δεδομένα.

Μέσω της Ανάλυσης Παλινδρόμησης μπορούμε να κατανοήσουμε την όποια μεταβολή της εξαρτημένης μεταβλητής όταν μεταβάλλεται μία από τις ανεξάρτητες, ενώ οι υπόλοιπες ανεξάρτητες παραμένουν σταθερές. Με αυτό το τρόπο μπορεί να εξακριβωθεί η αιτιώδης επίδραση μιας μεταβλητής επάνω σε μια άλλη.

Στη περίπτωση μας θα προσαρμόσουμε ένα μοντέλο γραμμικής παλινδρόμησης για τη συνεχή μεταβλητή απόκρισης που έχουμε στη διάθεσή μας και ένα μοντέλο λογιστική παλινδρόμησης για τη δίτιμη μεταβλητή σακχαρώδη διαβήτη. Στο μοντέλο της λογιστικής παλινδρόμησης θα εισάγουμε ως ανεξάρτητες ερμηνευτικές μεταβλητές όσες βρήκαμε ότι συσχετίζονται με το σακχαρώδη διαβήτη στο κεφάλαιο 4, δηλαδή τη μεταβλητή του φύλου των συμμετεχόντων (*Sex of subjects* – **Sex**), της ηλικίας (*Age of subjects* – **Age**), της οικογενειακής κατάστασης (*Marital status* – **Family**), της κοινωνικοοικονομικής κατάστασης (*Socio-economic status* – **SES**), του σωματικού βάρους (*Weight of subjects (kg)* – **Weight**), του δείκτη μάζας σώματος (*Body Mass Index* - **BMI**), της Μεσογειακής διατροφής (*Mediterranean Dietary Score Group* - **Meddietgroup**), των καπνιστικών συνήθειων (**Smoking\_current**), των καρδιαγγειακών συμβάντων (*Cardiovascular Disease Events during 2001-2006* – **CVD\_Events**), της διαστολικής αρτηριακής πίεσης (*Diastolic Blood Pressure* - **DBP**), της υπέρτασης (*Hypertension* - **HTN**) και της υπερχοληστερολαιμίας (*Hypercholesterolemia* - **hchol**), ενώ σε αυτό της γραμμικής παλινδρόμησης τις συνεχείς μεταβλητές που είναι διαθέσιμες στο σετ δεδομένων.

## 5.2 Γραμμική παλινδρόμηση

Πρόκειται για μια τεχνική μοντελοποίησης και ανάλυσης αριθμητικών δεδομένων, μιας εξαρτημένης και μιας ή περισσότερων ανεξάρτητων μεταβλητών. Ουσιαστικά για να χρησιμοποιηθεί η γραμμική παλινδρόμηση θα πρέπει όλες οι μεταβλητές που είναι διαθέσιμες για ανάλυση να είναι συνεχείς.

Κύριος σκοπός είναι η δημιουργία μιας συνάρτησης συσχέτισης της εξαρτημένης από τις ανεξάρτητες μεταβλητές. Αυτό θα μπορούσε να επιτευχθεί ακόμα και εάν δεν είναι γνωστός εκ των προτέρων ο τρόπος σύνδεσης της εξαρτημένης με τις ανεξάρτητες μεταβλητές. Στη περίπτωση της γραμμικής παλινδρόμησης, το μοντέλο θα πρέπει να είναι ένας γραμμικός συνδυασμός της εξαρτημένης μεταβλητής  $y_i$  με των ανεξάρτητων μεταβλητών.

Στην απλή γραμμική παλινδρόμηση υπάρχει μια ανεξάρτητη μεταβλητή  $x_i$  και δύο παράμετροι  $\beta_0$  και  $\beta_1$ . Τέλος το μοντέλο θα είναι της μορφής  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ , όπου  $\varepsilon_i$  είναι το σφάλμα πρόβλεψης.

Βασικές παραδοχές της γραμμικής παλινδρόμησης είναι η γραμμικότητα, η κανονικότητα, η ανεξαρτησία και η ομοσκεδαστικότητα. Από την άλλη μεριά, το γραμμικό μοντέλο δεν μπορεί να εφαρμοστεί όταν η  $Y$  είναι δίτιμη και προκύπτουν τα προβλήματα της μη κανονικότητας των σφαλμάτων, των άνισων διασπορών των σφαλμάτων και τέλος στο περιορισμό της συνάρτησης απόκρισης.

Στη παρούσα έρευνα, για να αξιοποιήσουμε τη γραμμική παλινδρόμηση, θα χρησιμοποιήσουμε τη συνεχή μεταβλητή που καταγράφει τα επίπεδα γλυκόζης στο αίμα (*blood glucose levels – glucose*), γνωρίζοντας από την ιατρική επιστήμη ότι τιμές της γλυκόζης μεγαλύτερες ή ίσες από 126 mg/dL, υποδεικνύουν την ύπαρξη σακχαρώδη διαβήτη.

### 5.2.1 Το γραμμικό μοντέλο

Προτού ξεκινήσουμε την ανάλυση του γραμμικού μοντέλου οφείλουμε να αναφέρουμε πως έχουμε βάσιμες ενδείξεις μη ικανοποίησης των βασικών προϋποθέσεων του, ήδη από το προηγούμενο κεφάλαιο. Παρόλα αυτά στο πλαίσιο μιας εκπαιδευτικής μελέτης, θα παρουσιάσουμε τα ευρήματα της γραμμικής παλινδρόμησης πριν τους ελέγχους προϋποθέσεων ώστε να έχουμε τη δυνατότητα να προχωρήσουμε σε συγκρίσεις με το μετασχηματισμένο μοντέλο που όπως θα δούμε στη συνέχεια θα προτιμηθεί.

Αρχικά, όπως αναφέραμε και προηγουμένως, επιλέξαμε ως μεταβλητή απόκρισης τα επίπεδα γλυκόζης στο αίμα (*blood glucose levels – glucose*) και ως ανεξάρτητες όλες τις συνεχείς μεταβλητές που ήταν διαθέσιμες στο σετ δεδομένων. Ο λόγος που αποφασίστηκε να εισαχθούν στο μοντέλο όλες οι συνεχείς μεταβλητές και όχι αυτές που γνωρίζουμε ότι συσχετίζονται με την εξαρτημένη, είναι διότι το στατιστικό πακέτο παρέχει τέσσερις διαφορετικές μεθόδους εξαίρεσης μη στατιστικά σημαντικών μεταβλητών από το μοντέλο της γραμμικής παλινδρόμησης. Ενδεικτικά αναφέρουμε τη μέθοδο *Remove* όπου όλες οι μεταβλητές (σε ένα μπλοκ) απομακρύνονται σε ένα μόνο βήμα. Τη μέθοδο *Backward* όπου είναι πρακτικά μια διαδικασία επιλογής στην οποία σε πρώτη φάση εισάγονται στο μοντέλο όλες οι μεταβλητές και στη συνέχεια

απομακρύνονται μία προς μία διαδοχικά. Συνεχίζοντας, υπάρχει διαθέσιμη η μέθοδος *Forward* όπου είναι μια διαδικασία στην οποία οι μεταβλητές εισέρχονται σταδιακά στο μοντέλο και τέλος η μέθοδος *Stepwise*, η οποία και εν τέλει χρησιμοποιήθηκε, είναι μια παραλλαγή της διαδικασίας *forward*, που μας επιτρέπει να επανεξετάσουμε αν κάποια από τις μεταβλητές (εκτός από αυτήν που έχει μόλις εισαχθεί στο μοντέλο), μπορεί να αφαιρεθεί χρησιμοποιώντας τη διαδικασία *backward*.

Ξεκινώντας, παραθέτουμε τον συντελεστή προσδιορισμού του τελικού μοντέλου που επιλέχθηκε με τη μέθοδο *stepwise*. Ο συντελεστής προσδιορισμού  $R^2$  παίρνει τιμές στο διάστημα από μηδέν έως ένα και αποτελεί ένα αξιόπιστο κριτήριο αξιολόγησης της προσαρμογής των δεδομένων στο υπό εξέταση μοντέλο. Όσο η τιμή του συντελεστή προσαρμογής πλησιάζει τη μονάδα τόσο μεγαλύτερη μεταβλητότητα ερμηνεύεται από τις ανεξάρτητες μεταβλητές. Το  $R^2$  παρατηρούμε ότι δεν είναι αρκετά υψηλό μιας και βρέθηκε ίσο με 0,155 αλλά σε αυτό το σημείο αξίζει να αναφέρουμε πως στις αναλύσεις με πραγματικά δεδομένα σπάνια παρατηρούνται υψηλά  $R^2$ .

Model Summary <sup>b</sup>				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.394 <sup>a</sup>	.155	.145	27.710

a. Predictors: (Constant), Smoking years  
b. Dependent Variable: Blood glucose levels (mg/dl)

Πίνακας 5.1: Συνοπτικός πίνακας μοντέλου γραμμικής παλινδρόμησης

Στο πίνακα που ακολουθεί το πακέτο παρουσιάζει αναλυτικά ποιες από τις ανεξάρτητες μεταβλητές εξαιρέθηκαν από το μοντέλο. Ο έλεγχος υποθέσεων που ακολουθήθηκε έχει μηδενική υπόθεση ότι η ανεξάρτητη μεταβλητή δεν είναι στατιστικά σημαντική για την ερμηνεία της απόκρισης και η εναλλακτική υπόθεση ισχυρίζεται ακριβώς το αντίθετο. Πράγματι όλα τα p-values είναι μεγαλύτερα από ένα επίπεδο σημαντικότητας 5%, μη απορρίπτοντας τη μηδενική υπόθεση. Επομένως όλες οι μεταβλητές του παρακάτω πίνακα δεν ελήφθησαν υπόψιν από το μοντέλο της γραμμικής παλινδρόμησης με τη μέθοδο *stepwise*.

Excluded Variables <sup>a</sup>								
Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics			
					Tolerance	VIF	Minimum Tolerance	
1	Age of Subjects	-.118 <sup>b</sup>	-.820	.414	-.090	.490	2.041	.490
	Weight of Subjects (Kg)	.140 <sup>b</sup>	1.374	.173	.149	.954	1.049	.954

Body Mass Index (kg/m <sup>2</sup> )	.189 <sup>b</sup>	1.821	.072	.196	.907	1.103	.907
Waist (cm)	.163 <sup>b</sup>	1.558	.123	.169	.907	1.102	.907
Periphery (cm)	.171 <sup>b</sup>	1.689	.095	.182	.960	1.041	.960
Years of school	-.136 <sup>b</sup>	-1.348	.181	-.146	.985	1.016	.985
Spontaneous Bacterial Peritonitis	.135 <sup>b</sup>	1.234	.221	.134	.839	1.191	.839
diastolic blood pressure	-.054 <sup>b</sup>	-.504	.615	-.055	.898	1.113	.898
Total serum cholesterol (mg/dl)	.108 <sup>b</sup>	1.064	.290	.116	.978	1.022	.978
Triglycerides (mg/dl)	.040 <sup>b</sup>	.385	.701	.042	.928	1.077	.928
HDL-cholesterol (mg/dl)	-.022 <sup>b</sup>	-.213	.832	-.023	.920	1.087	.920
LDL-cholesterol (mg/dL)	.113 <sup>b</sup>	1.112	.269	.121	.976	1.025	.976
Total antioxidant capacity (μmol/L)	-.107 <sup>b</sup>	-1.055	.294	-.115	.983	1.017	.983
C-reactive protein (mg/L)	.077 <sup>b</sup>	.759	.450	.083	.993	1.007	.993
urea	.168 <sup>b</sup>	1.682	.096	.182	.983	1.018	.983

a. Dependent Variable: Blood glucose levels (mg/dl)

b. Predictors in the Model: (Constant), Smoking years

**Πίνακας 5.2:** Πίνακας μεταβλητών που δεν εισήλθαν στο μοντέλο γραμμικής παλινδρόμησης

Παρακάτω δίνεται ο πίνακας με τους συντελεστές που είναι στατιστικά σημαντικοί για το μοντέλο της γραμμικής παλινδρόμησης. Όμοια με προηγουμένως, ο έλεγχος υποθέσεων που ακολουθήθηκε έχει μηδενική υπόθεση ότι η ανεξάρτητη μεταβλητή δεν είναι στατιστικά σημαντική για την ερμηνεία της απόκρισης. Εδώ όμως τα p-values είναι σχεδόν μηδέν οπότε απορρίπτεται η μηδενική υπόθεση και συμπεραίνουμε ότι η σταθερά και τα συνολικά χρόνια που καπνίζει κάποιος, είναι στατιστικά σημαντικοί παράγοντες για το μοντέλο της γραμμικής παλινδρόμησης. Τέλος, ο παράγοντας διόγκωσης διακύμανσης (VIF) βρέθηκε ίσος με τη μονάδα. Από τη θεωρία γνωρίζουμε ότι όταν το VIF παίρνει τιμές μεγαλύτερες του 10, τότε έχουμε πρόβλημα πολυσυγγραμικότητας. Στη περίπτωση μας δεν αντιμετωπίζουμε τέτοιο ενδεχόμενο.

Coefficients <sup>a</sup>											
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	75.162	6.174		12.175	.000					
	Smoking years	1.136	.289	.394	3.924	.000	.394	.394	.394	1.000	1.000

a. Dependent Variable: Blood glucose levels (mg/dl)

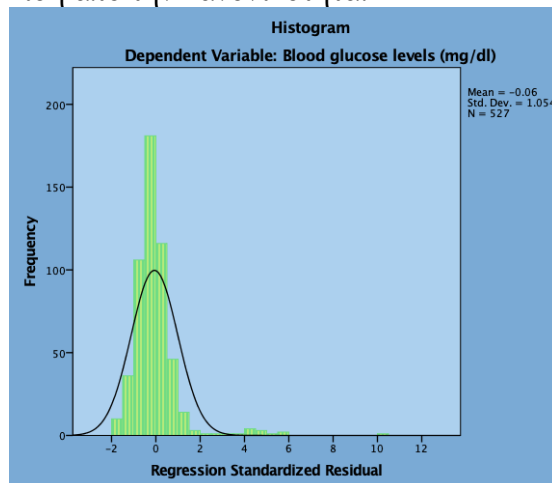
Πίνακας 5.3: Πίνακας συντελεστών μοντέλου γραμμικής παλινδρόμησης

## 5.2.2 Έλεγχος προϋποθέσεων γραμμικού μοντέλου

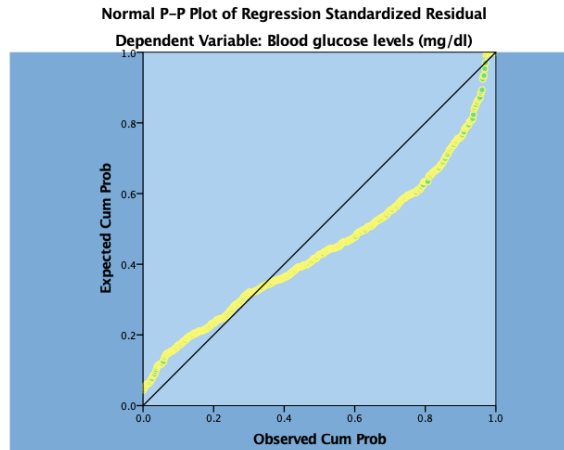
Όπως αναφέραμε στη προηγούμενη παράγραφο, σκόπιμα προχωρήσαμε στον έλεγχο προϋποθέσεων του γραμμικού μοντέλου εκ των υστέρων, διότι γνωρίζαμε εξ αρχής ότι δεν καλύπτονται. Σκοπός μας ήταν να παρουσιάσουμε έστω επιγραμματικά το αρχικό μοντέλο και να έχουμε παράλληλα ένα μέτρο σύγκρισης με ένα επόμενο, το οποίο θα έχει υποστεί κάποιους μετασχηματισμούς.

Οι βασικές προϋποθέσεις που πρέπει να ικανοποιεί ένα γραμμικό μοντέλο είναι η κανονικότητα, η ομοσκεδαστικότητα και η ανεξαρτησία των σφαλμάτων. Στη συνέχεια θα εξετάσουμε τις παραπάνω προϋποθέσεις με τη χρήση γραφημάτων και κατάλληλων στατιστικών ελέγχων.

Αρχικά θα προσπαθήσουμε να πάρουμε μια πρώτη διαγραμματική άποψη σχετικά με το πώς κινούνται τα κατάλοιπα. Από το πρώτο διάγραμμα φαίνεται πως μεγάλο μέρος των καταλοίπων να είναι εντός της «καμπάνας» της κανονικής κατανομής, αλλά μένει να δούμε εάν τα σφάλματα που βρίσκονται εκτός κανονικής κατανομής είναι ικανά στον αριθμό ώστε να μην αποδεχτούμε τη κανονικότητα. Από το επόμενο όμως διάγραμμα τα πράγματα φαίνεται να ξεκαθαρίζουν. Στο λεγόμενο p-p plot διαγράφεται ένα ευθύγραμμο τμήμα στο οποίο θα πρέπει τα κατάλοιπα να προσαρμόζονται για να αποδεχτούμε τη κανονικότητα. Στη περίπτωση μας γίνεται πλέον ξεκάθαρο οπτικά ότι υπάρχει μεγάλη απόκλιση από την κανονικότητα.



Γράφημα 5.1: Ιστογράμμο συχνοτήτων καταλοίπων, με προσαρμοσμένη τη καμπύλη της κανονικής κατανομής



Γράφημα 5.2: P-P plot τυποποιημένων εκτιμώμενων παρατηρήσεων

Για να μπορέσουμε να πάρουμε μια σίγουρη στατιστικά απόφαση για την κανονικότητα των σφαλμάτων θα πρέπει να εξετάσουμε και τον κατάλληλο έλεγχο. Ο μη παραμετρικός έλεγχος των Kolmogorov – Smirnov είναι ο κατάλληλος για να εξετάσει εάν κάθε είδους κατάλοιπα ακολουθούν τη κανονική κατανομή. Η μηδενική υπόθεση είναι ότι τα σφάλματα ακολουθούν τη κανονική κατανομή και η εναλλακτική ισχυρίζεται ακριβώς το αντίθετο. Πράγματι αυτό που είχαμε παρατηρήσει διαγραμματικά επιβεβαιώνεται και στατιστικά, ότι δηλαδή τα σφάλματα δεν ακολουθούν τη κανονική κατανομή. Τα p-values του ελέγχου για τρία διαφορετικά σφάλματα είναι σχεδόν μηδέν, επομένως απορρίπτουμε τη μηδενική υπόθεση ότι υπάρχει κανονικότητα στα κατάλοιπα.

Από τη στιγμή που απορρίφθηκε η κανονικότητα των σφαλμάτων δεν έχει νόημα να προχωρήσουμε στον έλεγχο των υπόλοιπων προϋποθέσεων του αφού πρέπει όλες να ικανοποιούνται ανεξαιρέτως για να αποδεχτούμε την ερμηνεία της γραμμικής παλινδρόμησης.

One-Sample Kolmogorov-Smirnov Test				
		Unstandardized Residual	Standardized Residual	Studentized Residual
N		527	527	527
Normal Parameters <sup>a,b</sup>	Mean	-1.7685210	-.0638235	-.0633424
	Std. Deviation	29.19688911	1.05367515	1.04273257
Most Extreme Differences	Absolute	.158	.158	.156
	Positive	.158	.158	.156
	Negative	-.109	-.109	-.106
Test Statistic		.158	.158	.156
Asymp. Sig. (2-tailed)		.000 <sup>c</sup>	.000 <sup>c</sup>	.000 <sup>c</sup>

Πίνακας 5.4: Πίνακας αποτελεσμάτων μη παραμετρικού ελέγχου των Kolmogorov - Smirnov

## 5.3 Μετασχηματισμός Box – Cox

Χωρίς να γίνονται αποδεκτές οι προϋποθέσεις της γραμμικής παλινδρόμησης, δεν μπορούμε να προχωρήσουμε σε οποιαδήποτε ερμηνεία του μοντέλου καθώς τα συμπεράσματα θα είναι επισφαλή. Σε περιπτώσεις όπως αυτή που αντιμετωπίζουμε πρόβλημα στη κανονικότητα, την ομοσκεδαστικότητα ή την ανεξαρτησία των σφαλμάτων σ' ένα μοντέλο πολλαπλής γραμμικής παλινδρόμησης, τότε μοναδική ευκαιρία να μην απορρίψουμε παντελώς το γραμμικό μοντέλο είναι να καταφύγουμε σε μετασχηματισμό των μεταβλητών ώστε να διορθώσουμε την ανεπιθύμητη απόκλιση.

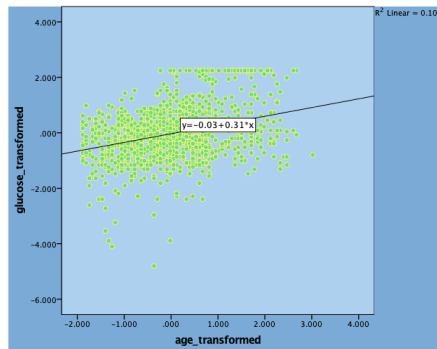
Οι Box και Cox προτείνουν το 1964 μια οικογένεια μετασχηματισμών μέσα από τους οποίους τροποποιείται η μεταβλητή απόκρισης και οι ανεξάρτητες μεταβλητές ώστε να επιτύχουμε την εξάλειψη των προαναφερθέντων αποκλίσεων. Συνήθως μέσω της μεθόδου αυτής επιτυγχάνεται η διόρθωση ούλων των προβλημάτων που μπορεί να αντιμετωπίζουμε στις προϋποθέσεις της γραμμικής παλινδρόμησης.

Το στατιστικό πακέτο που χρησιμοποιήθηκε σε αυτή τη μελέτη παρέχει τη δυνατότητα να διενεργηθούν μετασχηματισμοί Box – Cox μέσω ρουτίνας, ψάχνοντας κάθε φορά το κατάλληλο  $\lambda$ . Οι μετασχηματισμοί εφαρμόζονται στην εξαρτημένη μεταβλητή και στις ανεξάρτητες.

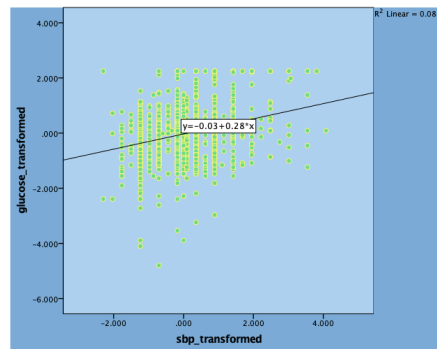
### 5.3.1 Έλεγχος προϋποθέσεων μετασχηματισμένου γραμμικού μοντέλου

Πλέον η ανάλυση και ερμηνεία του γραμμικού μοντέλου θα γίνει στα μετασχηματισμένα δεδομένα τόσο της εξαρτημένης μεταβλητής όσο και των ανεξάρτητων, τα οποία αποθηκεύτηκαν στο στατιστικό πακέτο. Το ίδιο ισχύει και για τους ελέγχους προϋποθέσεων του μοντέλου γραμμικής παλινδρόμησης.

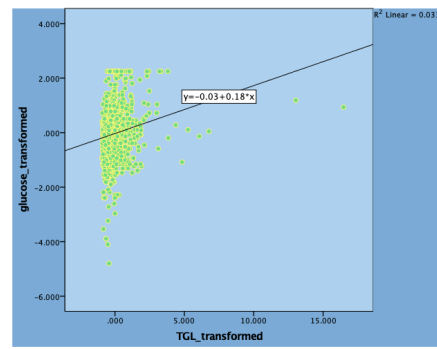
Ένα από τα πρώτα πράγματα που πρέπει να μελετήσει ένας στατιστικός, είναι η καταλληλότητα των δεδομένων που καλείται να αναλύσει όσον αφορά τη γραμμικότητα. Σε αυτό το πλαίσιο, παραθέτουμε τις γραφικές απεικονίσεις της τροποποιημένης εξαρτημένης μεταβλητής του επιπέδου γλυκόζης στο αίμα (*blood glucose levels – glucose*) στο κατακόρυφο άξονα (Y axis) και κάθε μια από τις τροποποιημένες ανεξάρτητες μεταβλητές της ηλικίας (*Age of subjects – Age*), της βακτηριακής περιτονίτιδας (*Spontaneous bacterial peritonitis – sbp*), των τριγλυκεριδίων (*Triglycerides (mg/dl) – TGL*), και των ετών που καπνίζει κάποιος (*years smoking – year smo*), διαδοχικά στον οριζόντιο άξονα (X axis).



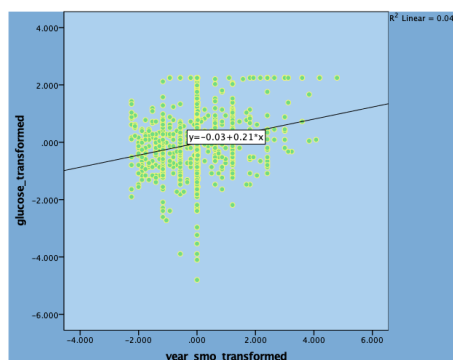
**Γράφημα 5.3:** Διάγραμμα διασποράς των τροποποιημένων μεταβλητών γλυκόζης αίματος και ηλικίας με προσαρμοσμένη την ευθεία ελαχίστων τετραγώνων



**Γράφημα 5.4:** Διάγραμμα διασποράς των τροποποιημένων μεταβλητών γλυκόζης αίματος και βακτηριακής περιτονίτιδας με προσαρμοσμένη την ευθεία ελαχίστων τετραγώνων



**Γράφημα 5.5:** Διάγραμμα διασποράς των τροποποιημένων μεταβλητών γλυκόζης αίματος και τριγλυκεριδίων αίματος με προσαρμοσμένη την ευθεία ελαχίστων τετραγώνων



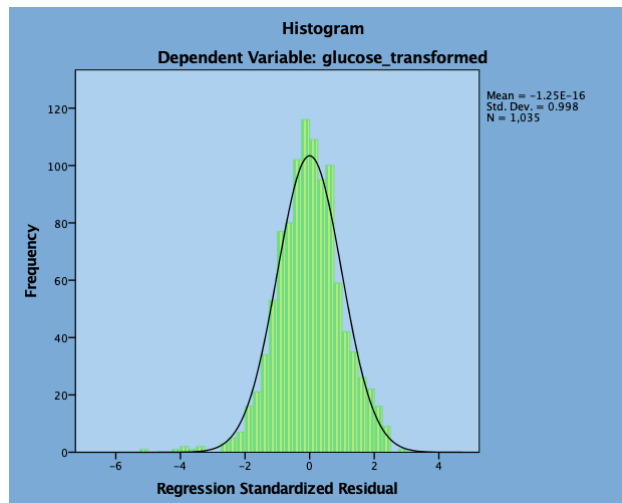
**Γράφημα 5.6:** Διάγραμμα διασποράς των τροποποιημένων μεταβλητών γλυκόζης αίματος και ετών καπνίσματος με προσαρμοσμένη την ευθεία ελαχίστων τετραγώνων



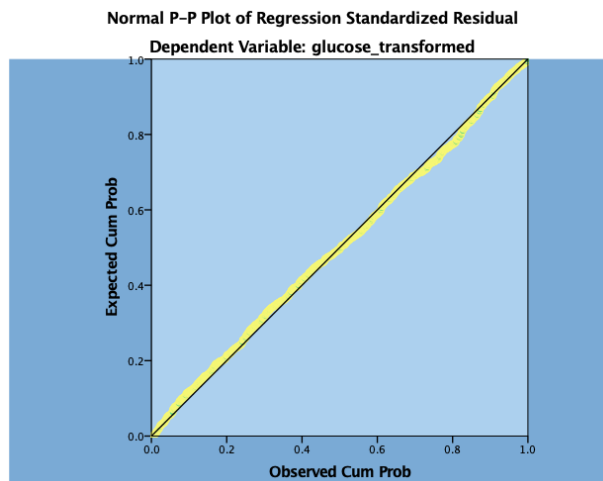
Ερμηνεύοντας τα τέσσερα διαγράμματα μπορούμε να συμπεράνουμε ότι η ηλικία, η βακτηριακή περιτονίτιδα, τα τριγλυκερίδια και και τα έτη που καπνίζει κάποιος φαίνονται να εξαρτώνται γραμμικά από τα επίπεδα γλυκόζης στο αίμα μιας και τα σημεία δεν φαίνεται να έχουν σημαντικές αποστάσεις από την ευθεία ελαχίστων τετραγώνων.

Αφού δεν απορρίπτουμε τη γραμμικότητα των δεδομένων, μπορούμε να ελέγξουμε τις βασικές προϋποθέσεις που πρέπει να ικανοποιεί ένα γραμμικό μοντέλο, οι οποίες είναι η κανονικότητα, η ομοσκεδαστικότητα και η ανεξαρτησία των σφαλμάτων. Στη συνέχεια θα εξετάσουμε τις παραπάνω προϋποθέσεις με τη χρήση γραφημάτων και κατάλληλων στατιστικών ελέγχων.

Ξεκινώντας από την κανονικότητα των σφαλμάτων, από το παρακάτω ιστόγραμμα παρατηρούμε ότι η πλειοψηφία των καταλοίπων βρίσκεται εντός της καμπύλης της κανονικής κατανομής. Επιπλέον, δεν φαίνεται να υπάρχει γραφική ένδειξη απόκλισης από τη κανονική κατανομή, αφού τα σημεία βρίσκονται κοντά στη διχοτόμο της γωνίας των αξόνων. Τα παραπάνω αποτελούν ένδειξη ότι τα σφάλματα ακολουθούν τη κανονική κατανομή.



**Γράφημα 5.7:** Ιστόγραμμα συχνότητας καταλοίπων των τροποποιημένων μεταβλητών, με προσαρμοσμένη τη καμπύλη της κανονικής κατανομής



**Γράφημα 5.8:** P-P plot τυποποιημένων εκτιμώμενων παρατηρήσεων των τροποποιημένων μεταβλητών

Τώρα μπορούμε να κάνουμε τον στατιστικό έλεγχο κανονικότητας σφαλμάτων των Kolmogorov-Smirnov ο οποίος ελέγχει τις παρακάτω υποθέσεις:

$H_0$ : Τα σφάλματα ακολουθούν την κανονική κατανομή

$H_1$ : Τα σφάλματα δεν ακολουθούν την κανονική κατανομή

Πράγματι αυτό που υποθέσαμε γραφικά φαίνεται να ισχύει καθώς τα p-values ναί μεν είναι μεγαλύτερα από ένα επίπεδο σημαντικότητας 5% αλλά βρίσκονται κοντά στο όριο. Δεν απορρίπτεται η μηδενική υπόθεση και με επιφυλάξεις αποδεχόμαστε ότι τα σφάλματα ακολουθούν τη κανονική κατανομή.

One-Sample Kolmogorov-Smirnov Test				
		Unstandardized Residual	Standardized Residual	Studentized Residual
N		1035	1035	1035
Normal Parameters <sup>a,b</sup>	Mean	.0000000	.0000000	-.0000906
	Std. Deviation	.88841178	.99806389	1.00072251
Most Extreme Differences	Absolute	.029	.029	.029
	Positive	.029	.029	.029
	Negative	-.022	-.022	-.022
Test Statistic		.029	.029	.029
Asymp. Sig. (2-tailed)		<b>.061<sup>c</sup></b>	<b>.061<sup>c</sup></b>	<b>.057<sup>c</sup></b>

a. Test distribution is Normal.  
b. Calculated from data.  
c. Lilliefors Significance Correction.

**Πίνακας 5.5:** Πίνακας αποτελεσμάτων μη παραμετρικού ελέγχου των Kolmogorov - Smirnov

Όσον αφορά τον έλεγχο ομοσκεδαστικότητας των σφαλμάτων θα πρέπει πρώτα να χωρίσουμε τα κατάλοιπα σε δύο ισοπληθής ομάδες με κριτήριο τη διάμεσό τους.

Ύστερα μπορούμε να προχωρήσουμε στο στατιστικό έλεγχο του Levene ο οποίος ελέγχει τις ακόλουθες υποθέσεις:

$H_0$ : Τα σφάλματα είναι ομοσκεδαστικά

$H_1$ : Τα σφάλματα δεν είναι ομοσκεδαστικά

Στη περίπτωση της ομοσκεδαστικότητας μπορούμε να απαντήσουμε με σιγουριά πως δεν υπάρχουν ενδείξεις για απόρριψη της μηδενικής υπόθεσης της ομοσκεδαστικότητας των σφαλμάτων αφού το p-value είναι ίσο με 0.199 το οποίο υπολείπεται κατά πολύ του  $\alpha=0,05$  επιπέδου σημαντικότητας.

Independent Samples Test			
		Levene's Test for Equality of Variances	
		F	Sig.
Unstandardized Residual	Equal variances assumed	1.651	.199

Πίνακας 5.6: Πίνακας αποτελεσμάτων ελέγχου ισότητας διακυμάνσεων του Levene

Κλείνοντας με τους ελέγχους προϋποθέσεων της γραμμικής παλινδρόμησης, θα προχωρήσουμε στο στατιστικό έλεγχο των ροών ο οποίος ελέγχει τις ακόλουθες υποθέσεις:

$H_0$ : Τα σφάλματα είναι ανεξάρτητα

$H_1$ : Τα σφάλματα δεν είναι ανεξάρτητα

Μετά τον έλεγχο κανονικότητας, και στην περίπτωση της ανεξαρτησίας είμαστε επιφυλακτικοί. Πιο συγκεκριμένα τα p-values τριών διαφορετικών καταλοίπων είναι μεγαλύτερα από ένα επίπεδο σημαντικότητας 5% αλλά χωρίς ιδιαίτερα μεγάλη διαφορά. Επομένως, δεν απορρίπτουμε τη μηδενική υπόθεση και αποδεχόμαστε επιφυλακτικά ότι τα σφάλματα είναι ανεξάρτητα μεταξύ τους.

Runs Test				
	Data Set	Unstandardized Residual	Standardized Residual	Studentized Residual
Test Value <sup>a</sup>	1.000 <sup>b</sup>	-.01112	-.01249	-.01251
Cases < Test Value	0	517	517	517
Cases >= Test Value	1035	518	518	518
Total Cases	1035	1035	1035	1035
Number of Runs	1 <sup>c</sup>	435	435	435
Z		-5.193	-5.193	-5.193
Asymp. Sig. (2-tailed)		.065	.065	.059

a. Median

b. All values are greater than or less than the cutoff. Runs Test cannot be performed.

c. Only one run occurs. Runs Test cannot be performed.

Πίνακας 5.7: Πίνακας αποτελεσμάτων μη παραμετρικού ελέγχου των ροών

### 5.3.2 Ανάλυση γραμμικής παλινδρόμησης μετασχηματισμένου μοντέλου Box - Cox

Μετά τις απαραίτητες τροποποιήσεις τόσο στην εξαρτημένη μεταβλητή, όσο και στις ανεξάρτητες, γίνανε αποδεκτές οι προϋποθέσεις της γραμμικής παλινδρόμησης. Στο κεφάλαιο αυτό θα ερμηνεύσουμε το γραμμικό μοντέλο.

Ξεκινώντας, παρουσιάζουμε ένα συνοπτικό πίνακα που καταγράφει τα βήματα της μεθόδου stepwise που επιλέχθηκε. Πιο συγκεκριμένα αναφέρει πως στο πρώτο βήμα εισήχθη η τροποποιημένη ανεξάρτητες μεταβλητή της ηλικίας (*Age of subjects* – **Age**), στο δεύτερο της βακτηριακής περιτονίτιδας (*Spontaneous bacterial peritonitis* - **sbp**), στο τρίτο η των τριγλυκεριδίων (*Triglycerides (mg/dl)* - **TGL**), και στο τέταρτο και τελευταίο βήμα η μεταβλητή των ετών που καπνίζει κάποιος (*years smoking* – **year smo**).

Variables Entered/Removed <sup>a</sup>			
Model	Variables Entered	Variables Removed	Method
1	age_transformed		Stepwise (Criteria: Probability-of- F-to-enter <= .050, Probability-of- F-to-remove >= .100).
2	sbp_transformed		Stepwise (Criteria: Probability-of- F-to-enter <= .050, Probability-of- F-to-remove >= .100).
3	TGL_transformed		Stepwise (Criteria: Probability-of- F-to-enter <= .050, Probability-of- F-to-remove >= .100).

4	year_smo_transformed	Stepwise (Criteria: Probability-of- F-to-enter <= .050, Probability-of- F-to-remove >= .100).
a. Dependent Variable: glucose_transformed		

Πίνακας 5.8: Πίνακας βημάτων μεθόδου stepwise

Ο συντελεστής προσδιορισμού  $R^2$  παίρνει τιμές στο διάστημα από μηδέν έως ένα και αποτελεί ένα κριτήριο αξιολόγησης της προσαρμογής των δεδομένων στο υπό εξέταση μοντέλο. Όσο η τιμή του συντελεστή προσαρμογής πλησιάζει τη μονάδα τόσο μεγαλύτερη μεταβλητότητα ερμηνεύεται από τις ανεξάρτητες μεταβλητές. Το  $R^2$  παρατηρούμε ότι δεν είναι αρκετά υψηλό μιας και βρέθηκε ίσο με 0,148 το οποίο ουσιαστικά σημαίνει οι τροποποιημένες ανεξάρτητες μεταβλητή της ηλικίας της βακτηριακής περιτονίτιδας, των τριγλυκεριδίων των ετών που καπνίζει κάποιος ερμηνεύει το 14,8% της συνολικής μεταβλητότητας της γλυκόζης αίματος. Το υπόλοιπο 85,2% της μεταβλητότητας παραμένει ανερμήνευτο λόγω του ότι σχετίζεται με παράγοντες οι οποίοι δεν έχουν ληφθεί υπ' όψιν στο παρών γραμμικό μοντέλο. Σε αυτό το σημείο αξίζει να αναφέρουμε πως στις αναλύσεις με πραγματικά δεδομένα σπάνια παρατηρούνται υψηλά  $R^2$ .

Model Summary <sup>e</sup>				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.327 <sup>a</sup>	.107	.106	.909902
2	.362 <sup>b</sup>	.131	.129	.897802
3	.378 <sup>c</sup>	.143	.140	.892230
4	.384 <sup>d</sup>	.148	.144	.890135

a. Predictors: (Constant), age\_transformed

b. Predictors: (Constant), age\_transformed, sbp\_transformed

c. Predictors: (Constant), age\_transformed, sbp\_transformed, TGL\_transformed

d. Predictors: (Constant), age\_transformed, sbp\_transformed, TGL\_transformed, year\_smo\_transformed

e. Dependent Variable: glucose\_transformed

Πίνακας 5.9: Συνοπτικός πίνακας μοντέλου γραμμικής παλινδρόμησης

Πλέον έχουμε καταλήξει στο μοντέλο της γραμμικής παλινδρόμησης, το οποίο είναι της μορφής  $\hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 * \text{age} + \widehat{\beta}_2 * \text{sbp} + \widehat{\beta}_3 * \text{tgl} + \widehat{\beta}_4 * \text{smoking years} = 0,186 * \text{age} + 162 * \text{sbp} + 0,102 * \text{tgl} + 0,075 * \text{smoking years}$ .

Ενδεικτικά, ο συντελεστής διεύθυνσης  $\widehat{\beta}_1$  της ευθείας  $\hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 \chi$  παριστάνει τη μεταβολή της εξαρτημένης μεταβλητής Y όταν το X μεταβληθεί κατά μία μονάδα. Με άλλα λόγια, όταν το  $\chi$  αυξηθεί κατά μία μονάδα τότε το  $\hat{y}$  αυξάνεται κατά  $\widehat{\beta}_1$  μονάδες όταν το  $\widehat{\beta}_1 > 0$  ή ελαττώνεται κατά  $\widehat{\beta}_1$  μονάδες όταν το  $\widehat{\beta}_1 < 0$ , όταν οι άλλες μεταβλητές παραμένουν σταθερές.

Coefficients <sup>a</sup>										
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
	B	Std. Error	Beta			Zero-order	Partial	Partial	VIF	
1	(Constant)	-.031	.028		-1.106	.269				
	age_transformed	.314	.028	.327	11.109	.000	.327	.327	.327	1.000
2	(Constant)	-.031	.028		-1.121	.263				
	age_transformed	.239	.031	.248	7.660	.000	.327	.232	.222	1.250
	sbp_transformed	.168	.031	.175	5.388	.000	.286	.165	.156	1.250
3	(Constant)	-.031	.028		-1.128	.260				
	age_transformed	.218	.032	.227	6.915	.000	.327	.211	.199	1.291
	sbp_transformed	.163	.031	.170	5.264	.000	.286	.162	.152	1.252
	TGL_transformed	.106	.028	.110	3.732	.000	.183	.115	.108	1.052
4	(Constant)	-.031	.028		-1.130	.259				
	age_transformed	.186	.034	.193	5.459	.000	.327	.168	.157	1.518
	sbp_transformed	.162	.031	.168	5.220	.000	.286	.161	.150	1.253
	TGL_transformed	.102	.028	.106	3.575	.000	.183	.111	.103	1.056
	year_smo_transformed	.075	.031	.078	2.421	.016	.218	.075	.070	1.257

a. Dependent Variable: glucose\_transformed

Πίνακας 5.10: Πίνακας συντελεστών μοντέλου γραμμικής παλινδρόμησης

Στο πίνακα που ακολουθεί το πακέτο παρουσιάζει αναλυτικά ποιες από τις ανεξάρτητες μεταβλητές εξαιρέθηκαν από το μοντέλο. Ο έλεγχος υποθέσεων που ακολουθήθηκε έχει μηδενική υπόθεση ότι η ανεξάρτητη μεταβλητή δεν είναι στατιστικά σημαντική για την ερμηνεία της απόκρισης και η εναλλακτική υπόθεση ισχυρίζεται ακριβώς το αντίθετο. Πράγματι όλα τα p-values είναι μεγαλύτερα από ένα επίπεδο σημαντικότητας 5%, μη απορρίπτοντας τη μηδενική υπόθεση. Επομένως όλες οι μεταβλητές του παρακάτω πίνακα δεν ελήφθησαν υπόψιν από το μοντέλο της γραμμικής παλινδρόμησης με τη μέθοδο stepwise.

Excluded Variables <sup>a</sup>								
Model	Beta	t	Sig.	Partial Correlation	Collinearity Statistics			
					Tolerance	VIF	Minimum Tolerance	
1	weight_transformed	.097 <sup>b</sup>	3.304	.001	.102	.996	1.004	.996
	bmi_transformed	.077 <sup>b</sup>	2.576	.010	.080	.953	1.049	.953
	hip_transformed	.074 <sup>b</sup>	2.515	.012	.078	.982	1.019	.982
	waist_transformed	.108 <sup>b</sup>	3.539	.000	.110	.921	1.086	.921
	educat_transformed	-.056 <sup>b</sup>	-1.693	.091	-.053	.791	1.265	.791
	year_smo_transformed	.090 <sup>b</sup>	2.745	.006	.085	.800	1.251	.800
	sbp_transformed	.175 <sup>b</sup>	5.388	.000	.165	.800	1.250	.800
	dbp_transformed	.110 <sup>b</sup>	3.586	.000	.111	.914	1.094	.914
	TC_transformed	.057 <sup>b</sup>	1.782	.075	.055	.855	1.170	.855
	TGL_transformed	.117 <sup>b</sup>	3.901	.000	.121	.952	1.050	.952
	HDL_C_transformed	-.070 <sup>b</sup>	-	.018	-.074	.999	1.001	.999
			2.374					
	LDL_C_transformed	-.013 <sup>b</sup>	-.440	.660	-.014	.937	1.067	.937
	crp_transformed	.066 <sup>b</sup>	2.246	.025	.070	.994	1.007	.994
	MedDietScore_transformed	-.029 <sup>b</sup>	-.889	.374	-.028	.828	1.207	.828
	ourea_transformed	.050 <sup>b</sup>	1.681	.093	.052	.995	1.005	.995
2	weight_transformed	.059 <sup>c</sup>	1.968	.049	.061	.925	1.081	.743
	bmi_transformed	.043 <sup>c</sup>	1.402	.161	.044	.903	1.108	.758
	hip_transformed	.045 <sup>c</sup>	1.506	.132	.047	.943	1.060	.769
	waist_transformed	.068 <sup>c</sup>	2.154	.031	.067	.849	1.178	.738
	educat_transformed	-.044 <sup>c</sup>	-1.339	.181	-.042	.787	1.271	.675
	year_smo_transformed	.086 <sup>c</sup>	2.645	.008	.082	.799	1.251	.672
	dbp_transformed	.005 <sup>c</sup>	.113	.910	.004	.522	1.914	.457
	TC_transformed	.047 <sup>c</sup>	1.509	.131	.047	.852	1.173	.715
	TGL_transformed	.110 <sup>c</sup>	3.732	.000	.115	.951	1.052	.774
	HDL_C_transformed	-.055 <sup>c</sup>	-1.879	.061	-.058	.989	1.011	.792

LDL_C_transformed	-.018 <sup>c</sup>	-.597	.551	-.019	.936	1.068	.763
crp_transformed	.054 <sup>c</sup>	1.843	.066	.057	.987	1.013	.795
MedDietScore_transformed	-.003 <sup>c</sup>	-.093	.926	-.003	.810	1.235	.715
ourea_transformed	.046 <sup>c</sup>	1.593	.112	.050	.995	1.005	.798
3 weight_transformed	.040 <sup>d</sup>	1.296	.195	.040	.892	1.121	.743
bmi_transformed	.028 <sup>d</sup>	.922	.356	.029	.887	1.128	.758
hip_transformed	.036 <sup>d</sup>	1.209	.227	.038	.937	1.068	.768
waist_transformed	.049 <sup>d</sup>	1.549	.122	.048	.824	1.213	.738
educat_transformed	-.044 <sup>d</sup>	-1.342	.180	-.042	.787	1.271	.657
year_smo_transformed	.078 <sup>d</sup>	2.421	.016	.075	.796	1.257	.659
dbp_transformed	-.007 <sup>d</sup>	-.186	.852	-.006	.519	1.927	.457
TC_transformed	.011 <sup>d</sup>	.341	.733	.011	.765	1.307	.712
HDL_C_transformed	-.033 <sup>d</sup>	-1.127	.260	-.035	.945	1.058	.772
LDL_C_transformed	-.017 <sup>d</sup>	-.563	.573	-.018	.936	1.068	.739
crp_transformed	.049 <sup>d</sup>	1.691	.091	.053	.985	1.015	.774
MedDietScore_transformed	.003 <sup>d</sup>	.105	.916	.003	.807	1.239	.698
ourea_transformed	.045 <sup>d</sup>	1.541	.124	.048	.994	1.006	.772
4 weight_transformed	.036 <sup>e</sup>	1.194	.233	.037	.890	1.123	.651
bmi_transformed	.028 <sup>e</sup>	.909	.363	.028	.887	1.128	.656
hip_transformed	.037 <sup>e</sup>	1.250	.211	.039	.936	1.068	.659
waist_transformed	.046 <sup>e</sup>	1.442	.150	.045	.823	1.216	.654
educat_transformed	-.041 <sup>e</sup>	-1.278	.201	-.040	.786	1.272	.576
dbp_transformed	-.006 <sup>e</sup>	-.161	.872	-.005	.519	1.927	.457
TC_transformed	.011 <sup>e</sup>	.342	.732	.011	.765	1.307	.613
HDL_C_transformed	-.028 <sup>e</sup>	-.936	.349	-.029	.939	1.065	.655
LDL_C_transformed	-.019 <sup>e</sup>	-.625	.532	-.019	.936	1.069	.636
crp_transformed	.047 <sup>e</sup>	1.624	.105	.051	.984	1.016	.659
MedDietScore_transformed	.002 <sup>e</sup>	.055	.956	.002	.807	1.239	.600
ourea_transformed	.044 <sup>e</sup>	1.536	.125	.048	.994	1.006	.658

a. Dependent Variable: glucose\_transformed

b. Predictors in the Model: (Constant), age\_transformed

c. Predictors in the Model: (Constant), age\_transformed, sbp\_transformed

d. Predictors in the Model: (Constant), age\_transformed, sbp\_transformed, TGL\_transformed

e. Predictors in the Model: (Constant), age\_transformed, sbp\_transformed, TGL\_transformed, year\_smo\_transformed

**Πίνακας 5.11:** Πίνακας μεταβλητών που δεν εισήλθαν στο μοντέλο γραμμικής παλινδρόμησης με τη μέθοδο stepwise



## 5.4 Ανάλυση γραμμικής παλινδρόμησης με προσαρμοσμένο μόνο το δείκτη Mediterranean Dietary Score

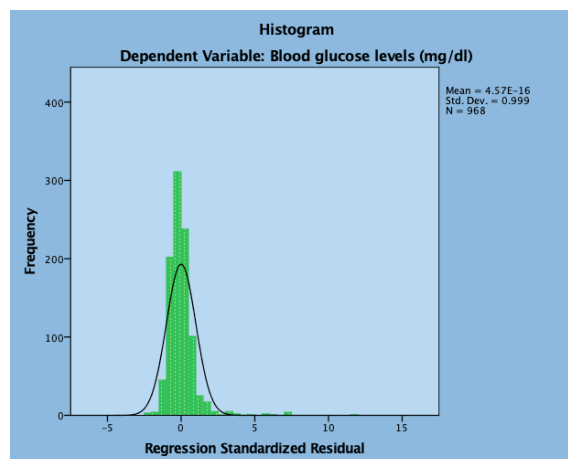
Όπως είδαμε, στο μετασχηματισμένο μοντέλο δε συμμετείχε η μεταβλητή που αναφέρεται στη μεσογειακή διατροφή των συμμετεχόντων. Θέλοντας να μελετήσουμε ένα μοντέλο γραμμικής παλινδρόμησης με τον δείκτη μεσογειακής διατροφής, προσαρμόσαμε ένα νέο μοντέλο χρησιμοποιώντας μόνο αυτή τη μεταβλητή. Η απόφαση αυτή πάρθηκε διότι υπάρχει μεγάλη συσχέτιση μεταξύ των ανεξάρτητων μεταβλητών και το φαινόμενο της πολυσυγγραμμικότητας δεν επιτρέπει ασφαλή συμπερασματολογία.

### 5.4.1 Έλεγχος προϋποθέσεων γραμμικού μοντέλου

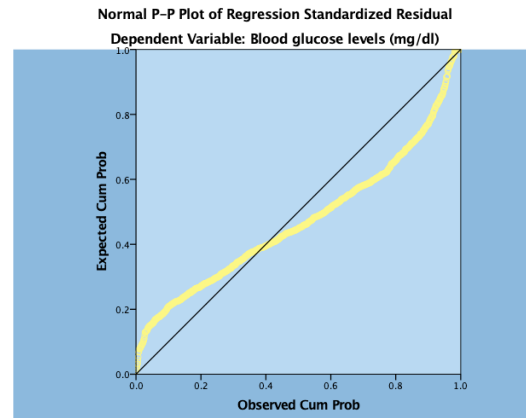
Έχοντας απορρίψει τις βασικές προϋποθέσεις του γραμμικού μοντέλου στις προηγούμενες παραγράφους, εύλογα προβλέπουμε ότι το ίδιο θα συμβεί και σε αυτή τη περίπτωση. Δεν μπορούμε να είμαστε σίγουροι εκ των προτέρων διότι αυτή τη φορά προσαρμόζουμε μόνο μια ανεξάρτητη μεταβλητή στο μοντέλο και τα κατάλοιπα πιθανώς να κινούνται διαφορετικά. Αυτός είναι ο λόγος που θα ελέγξουμε ξανά τα κατάλοιπα εάν πληρούνται οι προϋποθέσεις.

Οι βασικές προϋποθέσεις που πρέπει να ικανοποιεί ένα γραμμικό μοντέλο είναι η κανονικότητα, η ομοσκεδαστικότητα και η ανεξαρτησία των σφαλμάτων. Στη συνέχεια θα εξετάσουμε τις παραπάνω προϋποθέσεις με τη χρήση γραφημάτων και κατάλληλων στατιστικών ελέγχων.

Αρχικά θα προσπαθήσουμε να πάρουμε μια πρώτη διαγραμματική άποψη σχετικά με το πώς κινούνται τα κατάλοιπα. Από το πρώτο διάγραμμα φαίνεται πως μεγάλο μέρος των καταλοίπων να είναι εντός της «καμπάνας» της κανονικής κατανομής, αλλά μένει να δούμε εάν τα σφάλματα που βρίσκονται εκτός κανονικής κατανομής είναι ικανά στον αριθμό ώστε να μην αποδεχτούμε τη κανονικότητα. Από το επόμενο όμως διάγραμμα τα πράγματα φαίνεται να ξεκαθαρίζουν. Στο λεγόμενο p-p plot διαγράφεται ένα ευθύγραμμο τμήμα στο οποίο θα πρέπει τα κατάλοιπα να προσαρμόζονται για να αποδεχτούμε τη κανονικότητα. Στη περίπτωση μας γίνεται πλέον ξεκάθαρο οπτικά ότι υπάρχει μεγάλη απόκλιση από την κανονικότητα.



**Γράφημα 5.9:** Ιστόγραμμα συχνοτήτων καταλοίπων, με προσαρμοσμένη τη καμπύλη της κανονικής κατανομής



**Γράφημα 5.10:** P-P plot τυποποιημένων εκτιμώμενων παρατηρήσεων

Για να μπορέσουμε να πάρουμε μια σίγουρη στατιστικά απόφαση για την κανονικότητα των σφαλμάτων θα πρέπει να εξετάσουμε και τον κατάλληλο έλεγχο. Ο μη παραμετρικός έλεγχος των Kolmogorov – Smirnov είναι ο κατάλληλος για να εξετάσει εάν κάθε είδους κατάλοιπα ακολουθούν τη κανονική κατανομή. Η μηδενική υπόθεση είναι ότι τα σφάλματα ακολουθούν τη κανονική κατανομή και η εναλλακτική ισχυρίζεται ακριβώς το αντίθετο. Πράγματι αυτό που είχαμε παρατηρήσει διαγραμματικά επιβεβαιώνεται και στατιστικά, ότι δηλαδή τα σφάλματα δεν ακολουθούν τη κανονική κατανομή. Τα p-values του ελέγχου για τρία διαφορετικά σφάλματα είναι σχεδόν μηδέν, επομένως απορρίπτουμε τη μηδενική υπόθεση ότι υπάρχει κανονικότητα στα κατάλοιπα.

Από τη στιγμή που απορρίφθηκε η κανονικότητα των σφαλμάτων δεν έχει νόημα να προχωρήσουμε στον έλεγχο των υπόλοιπων προϋποθέσεων του αφού πρέπει όλες να ικανοποιούνται ανεξαιρέτως για να αποδεχτούμε την ερμηνεία της γραμμικής παλινδρόμησης.

#### One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual	Standardized Residual	Studentized Residual
N		968	968	968
Normal Parameters <sup>a,b</sup>	Mean	.0000000	.0000000	.0002067
	Std. Deviation	22.86027564	.99948280	1.00126573
Most Extreme Differences	Absolute	.150	.150	.151
	Positive	.150	.150	.151
	Negative	-.110	-.110	-.110
Test Statistic		.150	.150	.151
Asymp. Sig. (2-tailed)		.000 <sup>c</sup>	.000 <sup>c</sup>	.000 <sup>c</sup>

**Πίνακας 5.12:** Πίνακας αποτελεσμάτων μη παραμετρικού ελέγχου των Kolmogorov - Smirnov

### 5.4.2 Έλεγχος προϋποθέσεων μετασχηματισμένου γραμμικού μοντέλου Box – Cox

Χωρίς να γίνονται αποδεκτές οι προϋποθέσεις της γραμμικής παλινδρόμησης, δεν μπορούμε να προχωρήσουμε σε οποιαδήποτε ερμηνεία του μοντέλου καθώς τα συμπεράσματα θα είναι επισφαλής. Σε περιπτώσεις όπως αυτή που αντιμετωπίζουμε πρόβλημα στη κανονικότητα, την ομοσκεδαστικότητα ή την ανεξαρτησία των σφαλμάτων σ' ένα μοντέλο πολλαπλής γραμμικής παλινδρόμησης, τότε μοναδική ευκαιρία να μην απορρίψουμε παντελώς το γραμμικό μοντέλο είναι να καταφύγουμε σε μετασχηματισμό των μεταβλητών ώστε να διορθώσουμε την ανεπιθύμητη απόκλιση.

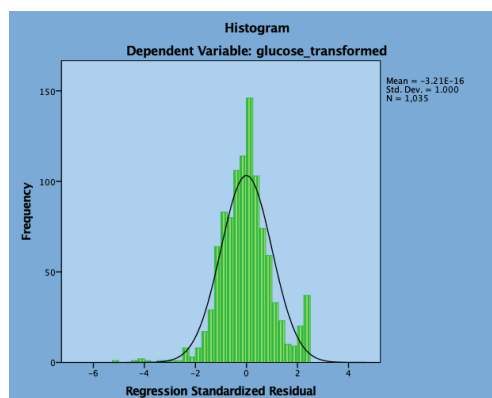
Οι Box και Cox προτείνουν μια οικογένεια μετασχηματισμών μέσα από τους οποίους τροποποιείται η μεταβλητή απόκρισης ώστε να επιτύχουμε την εξάλειψη των προαναφερθέντων αποκλίσεων. Συνήθως μέσω της μεθόδου αυτής επιτυγχάνεται η διόρθωση ούλων των προβλημάτων που μπορεί να αντιμετωπίζουμε στις προϋποθέσεις της γραμμικής παλινδρόμησης.

Το στατιστικό πακέτο που χρησιμοποιήθηκε σε αυτή τη μελέτη παρέχει τη δυνατότητα να διενεργηθούν μετασχηματισμοί Box – Cox μέσω ρουτίνας, ψάχνοντας κάθε φορά το κατάλληλο  $\lambda$ . Επομένως η μεταβλητή απόκρισης μετασχηματίστηκε ακολουθώντας τον τύπο  $Y_i(\lambda) = \frac{y_i^{0,5}}{0,5}$ , σύμφωνα με τον παρακάτω πίνακα.

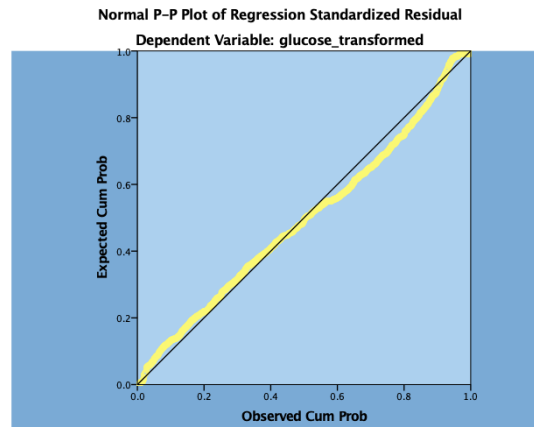
Transformation	Criteria		Lambda
	Mean	SD	
Box-Cox transformation to normality	0.00	1.00	0.50

Πίνακας 5.13: η τιμή του  $\lambda$  στο μετασχηματισμό Box - Cox

Πλέον η ανάλυση και ερμηνεία του γραμμικού μοντέλου θα γίνει στα μετασχηματισμένα δεδομένα τόσο της εξαρτημένης μεταβλητής όσο και των ανεξάρτητων, τα οποία αποθηκεύτηκαν στο στατιστικό πακέτο. Το ίδιο ισχύει και για τους ελέγχους προϋποθέσεων του μοντέλου γραμμικής παλινδρόμησης.



Γράφημα 5.11: Ιστογράμμο συχνοτήτων καταλοίπων της τροποποιημένης εξαρτημένης μεταβλητής, με προσαρμοσμένη τη καμπύλη της κανονικής κατανομής



**Γράφημα 5.12:** P-P plot τυποποιημένων εκτιμώμενων παρατηρήσεων της τροποποιημένης εξαρτημένης μεταβλητής

Τώρα μπορούμε να κάνουμε τον στατιστικό έλεγχο κανονικότητας σφαλμάτων των Kolmogorov-Smirnov ο οποίος ελέγχει τις παρακάτω υποθέσεις:

$H_0$ : Τα σφάλματα ακολουθούν την κανονική κατανομή

$H_1$ : Τα σφάλματα δεν ακολουθούν την κανονική κατανομή

Πράγματι αυτό που υποθέσαμε γραφικά φαίνεται να ισχύει καθώς τα p-values να μεν είναι μεγαλύτερα από ένα επίπεδο σημαντικότητας 5% αλλά βρίσκονται κοντά στο όριο. Δεν απορρίπτεται η μηδενική υπόθεση και με επιφυλάξεις αποδεχόμαστε ότι τα σφάλματα ακολουθούν τη κανονική κατανομή.

One-Sample Kolmogorov-Smirnov Test				
		Unstandardized Residual	Standardized Residual	Studentized Residual
N		1035	1035	1035
Normal Parameters <sup>a,b</sup>	Mean	.0000000	.0000000	.0000900
	Std. Deviation	.94998124	.99951632	1.00051977
Most Extreme Differences	Absolute	.055	.055	.055
	Positive	.055	.055	.055
	Negative	-.036	-.036	-.036
Test Statistic		.055	.055	.055
Asymp. Sig. (2-tailed)		<b>.073<sup>c</sup></b>	<b>.074<sup>c</sup></b>	<b>.061<sup>c</sup></b>

a. Test distribution is Normal.  
 b. Calculated from data.  
 c. Lilliefors Significance Correction.

**Πίνακας 5.14:** Πίνακας αποτελεσμάτων μη παραμετρικού ελέγχου των Kolmogorov - Smirnov

Στη συνέχεια θα περάσουμε στο στατιστικό έλεγχο των ροών ο οποίος ελέγχει τις ακόλουθες υποθέσεις:

$H_0$ : Τα σφάλματα είναι ανεξάρτητα

$H_1$ : Τα σφάλματα δεν είναι ανεξάρτητα

Μετά τον έλεγχο κανονικότητας, και στην περίπτωση της ανεξαρτησίας είμαστε επιφυλακτικοί. Πιο συγκεκριμένα τα p-values τριών διαφορετικών καταλοίπων είναι μεγαλύτερα από ένα επίπεδο σημαντικότητας 5% αλλά χωρίς ιδιαίτερα μεγάλη διαφορά. Επομένως, δεν απορρίπτουμε τη μηδενική υπόθεση και αποδεχόμαστε επιφυλακτικά ότι τα σφάλματα είναι ανεξάρτητα μεταξύ τους.

Runs Test			
	Unstandardized Residual	Standardized Residual	Studentized Residual
Test Value <sup>a</sup>	-.01543	-.01624	-.01625
Cases < Test Value	517	517	517
Cases >= Test Value	518	518	518
Total Cases	1035	1035	1035
Number of Runs	403	403	403
Z	-7.184	-7.184	-7.184
Asymp. Sig. (2-tailed)	.073	.073	.064

a. Median

Πίνακας 5.15: Πίνακας αποτελεσμάτων μη παραμετρικού ελέγχου των ροών

### 5.4.3 Ανάλυση γραμμικής παλινδρόμησης μετασχηματισμένου γραμμικού μοντέλου Box - Cox

Μετά την απαραίτητο μετασχηματισμό στην εξαρτημένη μεταβλητή, γίνανε αποδεκτές οι προϋποθέσεις της γραμμικής παλινδρόμησης. Στο κεφάλαιο αυτό θα ερμηνεύσουμε το γραμμικό μοντέλο.

Ξεκινώντας, παρουσιάζουμε το μοντέλο που θα αναλυθεί. Πιο συγκεκριμένα θα χρησιμοποιήσουμε τη συνεχή μεταβλητή των τιμών του σακχάρου στο αίμα (*diabetes mellitus* – **dm**) με ανεξάρτητη τη μετασχηματισμένη μεταβλητή της μεσογειακής διατροφής (*Mediterranean Diet Score* – **MedDietScor**).

Ο συντελεστής προσδιορισμού  $R^2$  παίρνει τιμές στο διάστημα από μηδέν έως ένα και αποτελεί ένα κριτήριο αξιολόγησης της προσαρμογής των δεδομένων στο υπό εξέταση μοντέλο. Όσο η τιμή του συντελεστή προσαρμογής πλησιάζει τη μονάδα τόσο μεγαλύτερη μεταβλητότητα ερμηνεύεται από τις ανεξάρτητες μεταβλητές. Το  $R^2$  παρατηρούμε ότι δεν είναι αρκετά υψηλό μιας και βρέθηκε ίσο με 0,29 το οποίο ουσιαστικά σημαίνει ότι η τροποποιημένη ανεξάρτητη μεταβλητή της μεσογειακής διατροφής ερμηνεύει το 30% περίπου της συνολικής μεταβλητότητας της γλυκόζης αίματος. Το υπόλοιπο 70% της μεταβλητότητας παραμένει ανερμήνευτο λόγω του ότι σχετίζεται με παράγοντες οι οποίοι δεν έχουν ληφθεί υπ' όψιν στο παρών γραμμικό μοντέλο. Σε αυτό το σημείο αξίζει να αναφέρουμε πως στις αναλύσεις με πραγματικά δεδομένα σπάνια παρατηρούνται υψηλά  $R^2$ .

Model Summary <sup>b</sup>				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.169 <sup>a</sup>	.029	.028	.906826

a. Predictors: (Constant), Med Dietary Score (0 - 55)  
b. Dependent Variable: glucose\_transformed

Πίνακας 5.16: Συνοπτικός πίνακας μοντέλου γραμμικής παλινδρόμησης

Πλέον έχουμε καταλήξει στο μοντέλο της γραμμικής παλινδρόμησης, το οποίο είναι της μορφής  $\hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x = 0,502 - 0,22 * \text{MedDietScore}$

Ενδεικτικά, ο συντελεστής διεύθυνσης  $\widehat{\beta}_1$  της ευθείας  $\hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x$  παριστάνει τη μεταβολή της εξαρτημένης μεταβλητής  $Y$  όταν το  $X$  μεταβληθεί κατά μία μονάδα. Με άλλα λόγια, όταν το  $x$  αυξηθεί κατά μία μονάδα τότε το  $\hat{y}$  αυξάνεται κατά  $\widehat{\beta}_1$  μονάδες όταν το  $\widehat{\beta}_1 > 0$  ή ελαττώνεται κατά  $\widehat{\beta}_1$  μονάδες όταν το  $\widehat{\beta}_1 < 0$ , όταν οι άλλες μεταβλητές παραμένουν σταθερές.

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.502	.111		4.513	.000
	Med Dietary Score (0 - 55)	-.022	.004	-.169	-5.432	.000

a. Dependent Variable: glucose\_transformed

Πίνακας 5.17: Πίνακας συντελεστών μοντέλου γραμμικής παλινδρόμησης

## 5.5 Λογιστική παλινδρόμηση

Πρόκειται για ένα μη γραμμικό λογιστικό μοντέλο, του οποίου τα σφάλματα δεν εντάσσονται στη κανονική κατανομή και η μεταβλητή απόκρισης, δηλαδή η εξαρτημένη μεταβλητή, είναι διακριτή.

Χρησιμοποιείται κυρίως όταν θέλουμε να προβλέψουμε τη παρουσία ή απουσία κάποιου χαρακτηριστικού ή γνωρίσματος. Πρακτικά είναι μια γενίκευση της γραμμικής παλινδρόμησης για την περίπτωση όπου η  $Y$  είναι δίτιμη μεταβλητή και παίρνει τη τιμή 0 όταν απουσιάζει και 1 όταν υπάρχει το χαρακτηριστικό.

Στη λογιστική παλινδρόμηση το μοντέλο είναι της μορφής  $\text{logit}(p)=\log(p/1-p)=\beta_0+\beta_1X_1+\beta_2X_2+\dots+\beta_nX_n$ , όπου  $\beta_0$  εκφράζει τη κλίση της γραμμικής παλινδρόμησης και ισούται με το  $\text{logit}(p)$  όταν οι τιμές όλων των ανεξάρτητων μεταβλητών είναι ίσες με 0, και  $\beta_i$  είναι οι συντελεστές της παλινδρόμησης όπου ο καθένας εκφράζει το μέγεθος της συνεισφοράς στο μοντέλο από τις αντίστοιχες μεταβλητές.

Στη παρούσα μελέτη θα χρησιμοποιήσουμε τη λογιστική παλινδρόμηση με μεταβλητή απόκρισης τη *Diabetes Mellitus* – *dm*, όπου παίρνει τη τιμή 0 όταν ο ερωτώμενος δεν πάσχει από σακχαρώδη διαβήτη και τη τιμή 1 όταν έχει διαγνωστεί από τη συγκεκριμένη πάθηση.

### 5.5.1 Ανάλυση μοντέλου λογιστικής παλινδρόμησης χωρίς ανεξάρτητες μεταβλητές

Εάν το μοντέλο προβλέπει τις περιπτώσεις των ασθενών που δεν πάσχουν από σακχαρώδη διαβήτη τότε αυτή η πρόβλεψη θα είναι σωστή στο  $533/566=0.942$  ή 94,2% των περιπτώσεων. Ποσοστό που κρίνεται αρκετά υψηλό για πρόβλεψη. Από την άλλη μεριά, η πρόβλεψη για τις περιπτώσεις των ασθενών που πάσχουν από σακχαρώδη διαβήτη θα είναι σωστή μόνο στο  $33/566=0,058$  ή 5,8% των περιπτώσεων. Το χαμηλό ποσοστό πρόβλεψης πιθανώς οφείλεται στο γεγονός ότι υπάρχουν διαθέσιμες μόνο 33 περιπτώσεις ασθενών που πάσχουν από σακχαρώδη διαβήτη. Κλείνοντας, το στατιστικό πακέτο ουσιαστικά δίνει απόλυτη ακρίβεια πρόβλεψης για τις περιπτώσεις που είναι διαβητικοί, μηδενική ακρίβεια πρόβλεψης για τους υγιείς και εν τέλει το μοντέλο προβλέπει σωστά το 94,2% το περιπτώσεων.

Classification Table <sup>a,b</sup>					
	Observed		Predicted		
			Diabetes Mellitus		Percentage Correct
			No diabetes	diabetes	
Step	Diabetes	No diabetes	533	0	100.0
0	Mellitus	diabetes	33	0	.0
	Overall Percentage				94.2

a. Constant is included in the model.

b. The cut value is .500

**Πίνακας 5.18:** Πίνακας προβλέψεων και αντιστοίχισης περιπτώσεων αναλόγως εάν πάσχουν ή όχι από σακχαρώδη διαβήτη

Από το πίνακα που ακολουθεί, μπορούμε να ελέγξουμε εάν η σταθερή παράμετρος είναι χρήσιμη για το μοντέλο. Ο έλεγχος Wald, που έχει μηδενική υπόθεση ότι το  $\beta_0=0$  και  $H_1: \beta_0 \neq 0$ , δίνει p-value που είναι σχεδόν μηδέν. Επομένως απορρίπτουμε τη μηδενική υπόθεση. Ουσιαστικά, η σταθερή παράμετρος είναι στατιστικά σημαντική για το μοντέλο και παίρνει τη τιμή -2,782.

Variables in the Equation						
	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	-2.782	.179	240.516	1	.000	.062

Πίνακας 5.19: Πίνακας σταθερής παραμέτρου

Στον επόμενο πίνακα δίνονται οι αντίστοιχοι έλεγχοι  $\chi^2$  για κάθε μια ανεξάρτητη μεταβλητή που προσαρμόσαμε στο μοντέλο της λογιστικής παλινδρόμησης. Η μηδενική υπόθεση του ελέγχου είναι ότι η βι μεταβλητή είναι χρήσιμη στο μοντέλο, έναντι της εναλλακτικής που ισχυρίζεται το αντίθετο. Θα αναλύσουμε σε επόμενη παράγραφο μία μία τις μεταβλητές που εισήγαμε στο μοντέλο, στη παρούσα φάση θα αρκεστούμε στο Overall statistics που έχει p-value σχεδόν μηδέν. Αυτό σημαίνει ότι απορρίπτουμε τη μηδενική υπόθεση και καταλαβαίνουμε ότι ουσιαστικά οι συντελεστές των μεταβλητών που δεν συμπεριελήφθησαν στο μοντέλο είναι στατιστικά σημαντικοί για το μοντέλο της λογιστικής παλινδρόμησης στο σύνολό τους.

Variables not in the Equation			
	Score	df	Sig.
Step 0 Variables age	45.499	1	.000
sex(1)	5.378	1	.020
weight	4.457	1	.035
bmi	7.284	1	.007
family	55.898	3	.000
family(1)	10.784	1	.001
family(2)	.691	1	.406
family(3)	.358	1	.549



SES	30.432	8	.000
SES(1)	13.643	1	.000
SES(2)	3.487	1	.062
SES(3)	6.939	1	.008
SES(4)	1.656	1	.198
SES(5)	.012	1	.913
SES(6)	.042	1	.838
SES(7)	1.755	1	.185
SES(8)	3.696	1	.055
smoking_current(1)	.001	1	.979
htn(1)	15.812	1	.000
dbp	5.976	1	.015
hchol(1)	8.295	1	.004
meddietgroup	29.915	2	.000
meddietgroup(1)	28.906	1	.000
meddietgroup(2)	2.727	1	.099
CVD_events(1)	16.774	1	.000
Overall Statistics	100.287	22	<b>.000</b>

Πίνακας 5.20: Πίνακας ερμηνευτικών παραμέτρων που δεν έχουν συμπεριληφθεί στο μοντέλο

### 5.5.2 Ανάλυση μοντέλου λογιστικής παλινδρόμησης με ταυτόχρονη προσαρμογή όλων των ανεξάρτητων μεταβλητών (μέθοδος enter)

Αφού επιβεβαιώσαμε ότι η σταθερή παράμετρος καθώς και ορισμένες μεταβλητές που συσχετίζονται με τον σακχαρώδη διαβήτη, είναι στατιστικά σημαντικές για το μοντέλο της λογιστικής παλινδρόμησης, είμαστε έτοιμοι να ερευνήσουμε ποιες ακριβώς είναι αυτές.

Αρχικά, θα εισάγουμε στο μοντέλο ταυτόχρονα όλες τις μεταβλητές που γνωρίζουμε από προηγούμενο κεφάλαιο ότι συσχετίζονται με τον σακχαρώδη διαβήτη, ανεξάρτητα εάν είναι στατιστικά σημαντικές ή όχι στο μοντέλο (μέθοδος enter).

Στη συνέχεια θα εξετάσουμε τη συνολική προσαρμογή του μοντέλου στα δεδομένα μέσω της πιθανοφάνειας. Ουσιαστικά θα εξετάσουμε τη διαφορά ανάμεσα στις τιμές του -2 Log Likelihood του μοντέλου που περιέχει όλες της ανεξάρτητες μεταβλητές και αυτού με τη σταθερή παράμετρο μόνο. Παρατηρούμε πως το αρχικό -2 Log Likelihood, του μοντέλου με τη σταθερά, βρέθηκε ίσο με 251,615 και του μοντέλου με προσαρμοσμένες όλες τις υπόλοιπες ερμηνευτικές μεταβλητές 174,407.

Iteration History <sup>a,b,c,d</sup>	
Iteration	-2 Log likelihood
Step 1 1	264.810
2	199.452
3	180.995
4	176.002
5	174.836
6	174.554
7	174.461
8	174.427
9	174.414
10	174.410
11	174.408
12	174.407
13	174.407
14	174.407
15	174.407
16	174.407
17	174.407
18	174.407

19	174.407
20	174.407

- a. Method: Enter  
 b. Constant is included in the model.  
 c. Initial -2 Log Likelihood: 251.615  
 d. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

**Πίνακας 5.21:** Πίνακας πιθανοφανειών και επαναλήψεων λογιστικής παλινδρόμησης

Η μεταξύ τους διαφορά είναι της τάξης των 77,208. Για να αποφανθούμε στατιστικά εάν το μοντέλο με τις ανεξάρτητες μεταβλητές, προβλέπει αποτελεσματικότερα από το αρχικό μοντέλο, εάν κάποιος πάσχει από σακχαρώδη διαβήτη, το πακέτο δίνει παράλληλα και ένα έλεγχο wald. Ο τελευταίος έχει μηδενική υπόθεση ότι το αρχικό μοντέλο προσαρμόζεται καλύτερα από το κορεσμένο. Το p-value του ελέγχου βρέθηκε σχεδόν μηδέν, απορρίπτοντας τη μηδενική υπόθεση. Ουσιαστικά συμπεραίνουμε πως το κορεσμένο μοντέλο προσαρμόζεται καλύτερα από το αρχικό και προβλέπει ασφαλέστερα εάν κάποιος πάσχει από σακχαρώδη διαβήτη ή όχι.

Omnibus Tests of Model Coefficients			
	Chi-square	df	Sig.
Step 1 Step	77.208	22	.000
Block	77.208	22	.000
Model	77.208	22	.000

**Πίνακας 5.22:** Πίνακας διαφορών πιθανοφανειών

Υπάρχουν περίπου 10 διαφορετικά μέτρα προσαρμογής του μοντέλου της λογιστικής παλινδρόμησης. Από αυτά άλλα έχουν πολλές ομοιότητες μεταξύ τους και άλλα διαφέρουν σημαντικά. Το πακέτο που επιλέξαμε για ανάλυση υπολογίζει δυο από τα διαθέσιμα, το  $R^2$  των Cox – Snell και το  $R^2$  του Nagelkerke. Και τα δύο παίρνουν τιμές στο διάστημα [0,1] με τη μονάδα να υποδεικνύει άριστη προσαρμογή του μοντέλου στα δεδομένα. Η διαφορά μεταξύ τους είναι ότι ο Nagelkerke διαιρεί αυτό των Cox – Snell με τη μέγιστη τιμή του.

Στη περίπτωση μας παρατηρούμε πως υπάρχει διαφοροποίηση στις τιμές των δύο κριτηρίων, αλλά πρακτικά συμπεραίνουμε πως το μοντέλο δεν είναι ιδιαίτερα ισχυρό. Πιο συγκεκριμένα, το  $R^2$  των Cox – Snell ισχυρίζεται ότι ερμηνεύεται το 12,8% της συνολικής διακύμανσης της εξαρτημένης μεταβλητής και το  $R^2$  του Nagelkerke το 35,5%.

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	174.407 <sup>a</sup>	.128	.355

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

**Πίνακας 5.23:** Πίνακας μέτρων προσαρμογής λογιστικής παλινδρόμησης

Στη συνέχεια παρουσιάζουμε τον πίνακα προβλέψεων αλλά αυτή τη φορά για το μοντέλο που έχει μέσα όλες τις ανεξάρτητες μεταβλητές. Ενδεικτικά, παρατηρούμε ότι το μοντέλο προβλέπει ότι 556 άτομα δεν θα έχουν διαβήτη, πραγματοποιώντας σωστή πρόβλεψη για τους 527, αγγίζοντας το 99% ορθής πρόβλεψης. Από την άλλη μεριά, το μοντέλο προβλέπει σωστά μόνο το 12,1% των περιπτώσεων που πάσχουν από σακχαρώδη διαβήτη και συνολικά το μοντέλο προβλέπει σωστά στο 93,8% των περιπτώσεων, ποσοστό άκρως ικανοποιητικό.

Classification Table <sup>a</sup>				
	Observed	Predicted		
		Diabetes Mellitus		Percentage Correct
		No diabetes	diabetes	
Step 1 Diabetes Mellitus	No diabetes	527	6	98.9
	diabetes	29	4	12.1
Overall Percentage				93.8

a. The cut value is .500

**Πίνακας 5.24:** Πίνακας προβλέψεων και αντιστοίχισης περιπτώσεων αναλόγως εάν πάσχουν ή όχι από σακχαρώδη διαβήτη

Μέσω του πίνακα που ακολουθεί μπορούμε να αποφανθούμε μέσω ελέγχων Wald εάν οι ανεξάρτητες μεταβλητές διαδραματίζουν στατιστικά σημαντικό ρόλο στη διαμόρφωση του μοντέλου. Η μηδενική υπόθεση του ελέγχου είναι ότι  $\beta_i=0$  έναντι  $\beta_i \neq 0$  της εναλλακτικής. Τα p-values δίνονται στην έκτη στήλη του παρακάτω πίνακα και με κόκκινο χρώμα είναι αυτά που μας οδηγούν σε μη απόρριψη της μηδενικής υπόθεσης. Επομένως, τα p-values των μεταβλητών της ηλικίας, της οικογενειακής κατάστασης και της υπέρτασης είναι μικρότερα από ένα επίπεδο σημαντικότητας 5% και επομένως απορρίπτουμε τη μηδενική υπόθεση. Ουσιαστικά αυτές οι μεταβλητές διαδραματίζουν στατιστικά σημαντικό ρόλο στη διαμόρφωση του μοντέλου.

Ένα εναλλακτικός τρόπος να καταλήξουμε στα ίδια συμπεράσματα με παραπάνω αλλά χωρίς τον έλεγχο Wald είναι μέσω των 95% διαστημάτων εμπιστοσύνης της  $\exp(B)$ . Πιο συγκεκριμένα όσα διαστήματα δεν συμπεριλαμβάνουν τη μονάδα, υποδηλώνουν πως οι αντίστοιχες μεταβλητές είναι στατιστικά σημαντικές για το μοντέλο.

Επιπροσθέτως, από τη στήλη B μπορούν να προκύψουν ερμηνείες αντίστοιχες της γραμμικής παλινδρόμησης για τις μεταβλητές που παίζουν στατιστικά σημαντικό ρόλο στο μοντέλο. Ενδεικτικά αναφέρουμε πως το -1,171 της υπέρτασης υποδηλώνει την αντίστοιχη μείωση της εξαρτημένης μεταβλητής του σακχαρώδη διαβήτη που σχετίζεται με αλλαγή μίας μονάδας στην ανεξάρτητη μεταβλητή της υπέρτασης.

Τέλος στη στήλη  $\exp(B)$  αναγράφονται οι λογάριθμοι της στήλης B. Ενδεικτικά αναφέρουμε ότι ο λογάριθμος της σχετικής πιθανότητας να ασθενήσει κάποιος από σακχαρώδη διαβήτη, αυξάνεται κατά 1,059 εάν αυξηθεί η ηλικία του κατά μια μονάδα, θεωρώντας όλες τις άλλες παραμέτρους σταθερές.

Κλείνοντας, η γενική μορφή του μοντέλου θα είναι:  $\text{logit}(p)=\log(p/1-p)=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5$  ενώ το προτεινόμενο μοντέλο λογιστικής παλινδρόμησης για τη μελέτη με τη μέθοδο enter, είναι της μορφής:  $\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = -0,364+(1,059*age)+(0,032*family2)+(0,079*family3)+(0,121*family4)+(0,31*htn2)$ .

Variables in the Equation								
	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Step 1 <sup>a</sup>								
age	.057	.025	5.058	1	.025	1.059	1.007	1.113
sex(1)	-1.102	.684	2.594	1	.107	.332	.087	1.270
weight	.004	.033	.012	1	.914	1.004	.940	1.072
bmi	.022	.112	.037	1	.847	1.022	.821	1.272
family			11.488	3	.009			
family(1)	-3.446	1.340	6.616	1	.010	.032	.002	.440
family(2)	-2.533	.785	10.409	1	.001	.079	.017	.370
family(3)	-2.111	1.196	3.114	1	.078	.121	.012	1.263
SES			5.217	8	.734			
SES(1)	.961	.810	1.408	1	.235	2.615	.534	12.796
SES(2)	.316	.792	.159	1	.690	1.371	.291	6.472
SES(3)	1.249	.829	2.270	1	.132	3.487	.687	17.707
SES(4)	-.950	1.386	.469	1	.493	.387	.026	5.854
SES(5)	.316	.698	.205	1	.651	1.372	.349	5.390
SES(6)	-.064	.708	.008	1	.928	.938	.234	3.755
SES(7)	-16.087	7159.287	.000	1	.998	.000	.000	.
SES(8)	-16.961	4938.587	.000	1	.997	.000	.000	.
smoking_current(1)	-.588	.459	1.643	1	.200	.555	.226	1.365
htn(1)	-1.171	.561	4.356	1	.037	.310	.103	.931
dbp	-.017	.025	.465	1	.495	.983	.937	1.032
hchol(1)	-.438	.423	1.074	1	.300	.645	.282	1.478
meddietgroup			.769	2	.681			
meddietgroup(1)	-.194	1.146	.029	1	.866	.824	.087	7.791
meddietgroup(2)	-.627	1.017	.380	1	.537	.534	.073	3.922
CVD_events(1)	.003	.687	.000	1	.996	1.003	.261	3.857
Constant	-1.011	3.213	.099	1	.753	.364		

a. Variable(s) entered on step 1: age, sex, weight, bmi, family, SES, smoking\_current, htn, dbp, hchol, meddietgroup, CVD\_events.

Πίνακας 5.25: Πίνακας στατιστικής σημαντικότητας ερμηνευτικών μεταβλητών

### 5.5.3 Ανάλυση μοντέλου λογιστικής παλινδρόμησης με σταδιακή προσαρμογή των ανεξάρτητων μεταβλητών (μέθοδος Forward Selection-Likelihood Ratio)

Υπάρχουν αρκετές μέθοδοι για την προσαρμογή των ανεξάρτητων μεταβλητών σε ένα μοντέλο λογιστική παλινδρόμησης. Το στατιστικό πακέτο μέσω του οποίου γίνεται η ανάλυση της παρούσας μελέτης, έχει 7 διαφορετικές μεθόδους. Αφού αναλύθηκε η μέθοδος enter στη προηγούμενη παράγραφο, θα χρησιμοποιήσουμε στη παρούσα μια διαφορετική μέθοδο για να συγκρίνουμε τα αποτελέσματα.

Επιλέχθηκε η μέθοδος forward selection (Likelihood Ratio). Πρόκειται για μια διαδικασία κατά την οποία οι μεταβλητές εισέρχονται σταδιακά στο μοντέλο. Στο πρώτο βήμα της μεθόδου είναι εισαγμένη στο μοντέλο μόνο η σταθερά  $\beta_0$  και ο έλεγχος για την εισαγωγή της επόμενης ανεξάρτητης μεταβλητής βασίζεται στη στατιστική σημαντικότητα της μεταβλητής. Στη συνέχεια πραγματοποιείται έλεγχος για την αφαίρεση όποιας μεταβλητής δεν είναι στατιστικά σημαντική για το μοντέλο. Ο έλεγχος αφαίρεσης μεταβλητών βασίζεται στην πιθανότητα του λόγου πιθανοφανειών (εκτίμηση μεγίστης μερικής πιθανοφάνειας).

Αρχικά θα εξετάσουμε τη συνολική προσαρμογή του μοντέλου στα δεδομένα μέσω της πιθανοφάνειας. Ουσιαστικά θα εξετάσουμε τη διαφορά ανάμεσα στις τιμές του  $-2 \text{ Log Likelihood}$  του μοντέλου που περιέχει τις ανεξάρτητες μεταβλητές μετά από τις κατάλληλες προσθαφαιρέσεις μεταβλητών και αυτού με τη σταθερή παράμετρο μόνο. Παρατηρούμε πως το αρχικό  $-2 \text{ Log Likelihood}$  του μοντέλου με τη σταθερά, βρέθηκε ίσο με 251,615 και του μοντέλου με προσαρμοσμένες τις στατιστικά σημαντικές ερμηνευτικές μεταβλητές 188,133.

Τέλος σε αυτό το σημείο θα θέλαμε να αναφέρουμε πως η μέθοδος forward selection (Likelihood Ratio), κατέληξε στο αποτέλεσμα έπειτα από 4 επαναλήψεις της μεθόδου (έλεγχοι πρόσθεσης και απόρριψης παραμέτρων).

Iteration History <sup>a,b,c,d</sup>		
Iteration	-2 Log likelihood	
Step 1	1	279.124
	2	229.450
	3	220.920
	4	219.319
	5	219.117
	6	219.111
	7	219.111
	8	219.111
Step 2	1	271.797
	2	213.737
	3	200.671
	4	198.344
	5	198.081

	6	198.072
	7	198.072
	8	198.072
Step	1	270.088
3	2	210.007
	3	195.693
	4	193.092
	5	192.810
	6	192.800
	7	192.800
	8	192.800
Step	1	269.266
4	2	207.678
	3	191.724
	4	188.493
	5	188.145
	6	188.133
	7	188.133
	8	<b>188.133</b>

a. Method: Forward Stepwise (Likelihood Ratio)  
 b. Constant is included in the model.  
 c. Initial -2 Log Likelihood: **251.615**  
 d. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.

**Πίνακας 5.26:** Πίνακας πιθανοφανειών και επαναλήψεων λογιστικής παλινδρόμησης

Η μεταξύ τους διαφορά είναι της τάξης των 63,482. Για να αποφανθούμε στατιστικά εάν το μοντέλο με τις ανεξάρτητες μεταβλητές προβλέπει αποτελεσματικότερα από το αρχικό μοντέλο εάν κάποιος πάσχει από σακχαρώδη διαβήτη, το πακέτο δίνει παράλληλα και ένα έλεγχο Wald. Ο τελευταίος έχει μηδενική υπόθεση ότι το αρχικό μοντέλο προσαρμόζεται καλύτερα από το κορεσμένο. Το p-value του ελέγχου βρέθηκε σχεδόν μηδέν, απορρίπτοντας τη μηδενική υπόθεση. Ουσιαστικά συμπεραίνουμε πως το κορεσμένο μοντέλο προσαρμόζεται καλύτερα από το αρχικό και προβλέπει ασφαλέστερα εάν κάποιος πάσχει από σακχαρώδη διαβήτη ή όχι.



Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	32.504	3	.000
	Block	32.504	3	.000
	Model	32.504	3	.000
Step 2	Step	21.039	1	.000
	Block	53.543	4	.000
	Model	53.543	4	.000
Step 3	Step	5.272	1	.022
	Block	58.815	5	.000
	Model	58.815	5	.000
Step 4	Step	4.667	1	.031
	Block	63.482	6	.000
	Model	63.482	6	.000

Πίνακας 5.27: Πίνακας διαφορών πιθανοφανειών

Υπάρχουν περίπου 10 διαφορετικά μέτρα προσαρμογής του μοντέλου της λογιστικής παλινδρόμησης. Από αυτά άλλα έχουν πολλές ομοιότητες μεταξύ τους και άλλα διαφέρουν σημαντικά. Το πακέτο που επιλέξαμε για ανάλυση υπολογίζει δυο από τα διαθέσιμα, το  $R^2$  των Cox – Snell και το  $R^2$  του Nagelkerke. Και τα δύο παίρνουν τιμές στο διάστημα  $[0,1]$  με τη μονάδα να υποδεικνύει άριστη προσαρμογή του μοντέλου στα δεδομένα. Η διαφορά μεταξύ τους είναι ότι ο Nagelkerke διαιρεί αυτό των Cox – Snell με τη μέγιστη τιμή του.

Στη περίπτωση μας παρατηρούμε πως υπάρχει διαφοροποίηση στις τιμές των δύο κριτηρίων, αλλά πρακτικά συμπεραίνουμε πως το μοντέλο δεν είναι ιδιαίτερα ισχυρό. Πιο συγκεκριμένα, το  $R^2$  των Cox – Snell ισχυρίζεται ότι ερμηνεύεται το 10,6% της συνολικής διακύμανσης της εξαρτημένης μεταβλητής και το  $R^2$  του Nagelkerke το 29,6%.

Σ' αυτό το σημείο θα θέλαμε να κάνουμε δύο σχόλια. Αρχικά παρατηρούμε πως επανάληψη με την επανάληψη η τιμή των  $R^2$  αυξάνεται και ότι η ερμηνευμένη συνολική διακύμανση της απόκρισης είναι ελαφρώς περιορισμένη σε σχέση με τη μέθοδο enter του προηγούμενου κεφαλαίου.

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	219.111 <sup>a</sup>	.056	.156
2	198.072 <sup>a</sup>	.090	.252
3	192.800 <sup>a</sup>	.099	.275
4	188.133 <sup>a</sup>	.106	.296

Πίνακας 5.28: Πίνακας μέτρων προσαρμογής λογιστικής παλινδρόμησης

Στη συνέχεια παρουσιάζουμε τον πίνακα προβλέψεων του μοντέλου λογιστικής παλινδρόμησης μετά τη τέταρτη επανάληψη της μεθόδου forward. Ενδεικτικά, παρατηρούμε ότι το μοντέλο προβλέπει ότι 560 άτομα δεν θα έχουν διαβήτη, πραγματοποιώντας σωστή πρόβλεψη για τους 530, φτάνοντας το 99,4% ορθής πρόβλεψης. Από την άλλη μεριά, το μοντέλο προβλέπει σωστά μόνο το 9,1% των περιπτώσεων που πάσχουν από σακχαρώδη διαβήτη και συνολικά το μοντέλο προβλέπει σωστά στο 94,2% των περιπτώσεων, ποσοστό άκρως ικανοποιητικό.

Σε σύγκριση με τη μέθοδο enter, τα ποσοστά ορθής πρόβλεψης είναι οριακά υψηλότερα.

Classification Table <sup>a</sup>					
	Observed		Predicted		
			Diabetes Mellitus		Percentage Correct
			No diabetes	diabetes	
Step 1	Diabetes Mellitus	No diabetes	528	5	99.1
		diabetes	27	6	18.2
	Overall Percentage				94.3
Step 2	Diabetes Mellitus	No diabetes	530	3	99.4
		diabetes	30	3	9.1
	Overall Percentage				94.2
Step 3	Diabetes Mellitus	No diabetes	528	5	99.1
		diabetes	28	5	15.2
	Overall Percentage				94.2
Step 4	Diabetes Mellitus	No diabetes	530	3	99.4
		diabetes	30	3	9.1
	Overall Percentage				94.2

a. The cut value is .500

Πίνακας 5.29: Πίνακας προβλέψεων και αντιστοίχισης περιπτώσεων αναλόγως εάν πάσχουν ή όχι από σακχαρώδη διαβήτη

Μέσω του πίνακα που ακολουθεί μπορούμε να αποφανθούμε μέσω ελέγχων Wald εάν οι ανεξάρτητες μεταβλητές διαδραματίζουν στατιστικά σημαντικό ρόλο στη διαμόρφωση του μοντέλου. Η μηδενική υπόθεση του ελέγχου είναι ότι  $\beta_i=0$  έναντι  $\beta_i \neq 0$  της εναλλακτικής. Τα p-values δίνονται στην έκτη στήλη του παρακάτω πίνακα, με πράσινο χρώμα είναι αυτά που μας οδηγούν σε απόρριψη της μηδενικής υπόθεσης. Επομένως, τα p-values των μεταβλητών της ηλικίας, του φύλου, της οικογενειακής κατάστασης και της υπέρτασης είναι μικρότερα από ένα επίπεδο σημαντικότητας 5% και επομένως απορρίπτουμε τη μηδενική υπόθεση. Ουσιαστικά αυτές οι μεταβλητές διαδραματίζουν στατιστικά σημαντικό ρόλο στη διαμόρφωση του μοντέλου.

Ένα εναλλακτικός τρόπος να καταλήξουμε στα ίδια συμπεράσματα με παραπάνω αλλά χωρίς τον έλεγχο Wald είναι μέσω των 95% διαστημάτων εμπιστοσύνης της  $\exp(B)$ . Πιο συγκεκριμένα όσα διαστήματα δεν συμπεριλαμβάνουν τη μονάδα, υποδηλώνουν πως οι αντίστοιχες μεταβλητές είναι στατιστικά σημαντικές για το μοντέλο.

Επιπροσθέτως, από τη στήλη B μπορούν να προκύψουν ερμηνείες αντίστοιχες της γραμμικής παλινδρόμησης για τις μεταβλητές που παίζουν στατιστικά σημαντικό ρόλο στο μοντέλο. Ενδεικτικά αναφέρουμε πως το -0,859 της υπέρτασης υποδηλώνει την αντίστοιχη μείωση της εξαρτημένης μεταβλητής της υπέρτασης που σχετίζεται με αλλαγή μίας μονάδας στην ανεξάρτητη μεταβλητή του σακχαρώδη διαβήτη.

Τέλος στη στήλη  $\exp(B)$  αναγράφονται οι λογάριθμοι της στήλης B. Ενδεικτικά αναφέρουμε ότι ο λογάριθμος της σχετικής πιθανότητας να ασθενήσει κάποιος από σακχαρώδη διαβήτη, αυξάνεται κατά 1,07 εάν αυξηθεί η ηλικία του κατά μια μονάδα, θεωρώντας όλες τις άλλες παραμέτρους σταθερές.

Κλείνοντας, η γενική μορφή του μοντέλου θα είναι:  $\text{logit}(p)=\log(p/1-p)=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5+\beta_6X_6$  ενώ το προτεινόμενο μοντέλο λογιστικής παλινδρόμησης για τη μελέτη με τη μέθοδο forward, είναι της μορφής:  $\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = -0,055+(1,07*\text{age})+(0,382*\text{sex1})+(0,032*\text{family2})+(0,084*\text{family3})+(0,158*\text{family4})+(0,424*\text{htn2})$

Variables in the Equation								
	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
S			26.846	3	.000			
† family(1)	-	1.172	19.987	1	.000	.005	.001	.053
e	5.239							
P								
1 family(2)	-	.641	19.923	1	.000	.057	.016	.201
a	2.862							
family(3)	-	.956	7.021	1	.008	.079	.012	.517
	2.534							
Constant	.182	.606	.091	1	.763	1.200		
S								
age	.074	.016	20.339	1	.000	1.077	1.043	1.112
† family			10.330	3	.016			
e								
family(1)	-3.142	1.265	6.164	1	.013	.043	.004	.516
p								
family(2)	-	.703	8.698	1	.003	.126	.032	.499
2	2.074							
b								
family(3)	-1.454	1.020	2.032	1	.154	.234	.032	1.724
Constant	-4.410	1.211	13.251	1	.000	.012		
S								
age	.069	.017	16.728	1	.000	1.071	1.037	1.107
† family			9.862	3	.020			
e								
family(1)	-	1.274	5.668	1	.017	.048	.004	.585
p								
3								
c								
family(2)	-	.704	8.709	1	.003	.125	.031	.498
	2.078							
family(3)	-1.590	1.031	2.375	1	.123	.204	.027	1.540
htn(1)	-.917	.401	5.224	1	.022	.400	.182	.877
Constant	-	1.284	8.045	1	.005	.026		
	3.642							
S								
age	.067	.017	15.644	1	.000	1.070	1.035	1.106
† sex(1)	-.963	.473	4.144	1	.042	.382	.151	.965
e								
family			12.190	3	.007			
P								
4								
d								
family(1)	-	1.301	7.007	1	.008	.032	.002	.409
	3.444							
family(2)	-	.751	10.910	1	.001	.084	.019	.365
	2.479							
family(3)	-1.848	1.053	3.077	1	.079	.158	.020	1.242
htn(1)	-.859	.404	4.530	1	.033	.424	.192	.934
Constant	-	1.321	4.837	1	.028	.055		
	2.905							

Πίνακας 5.30: Πίνακας στατιστικής σημαντικότητας ερμηνευτικών μεταβλητών

### 5.5.4 Ανάλυση μοντέλου λογιστικής παλινδρόμησης με προσαρμοσμένο το δείκτη Mediterranean Dietary Score, ταυτόχρονη προσαρμογή όλων των ανεξάρτητων μεταβλητών (μέθοδος enter)

Όπως είδαμε στις προηγούμενες παραγράφους, προσαρμόζοντας τις διαθέσιμες ερμηνευτικές μεταβλητές στο μοντέλο της λογιστικής παλινδρόμησης, είτε με τη μέθοδο enter είτε με τη forward selection-likelihood ratio, το προτεινόμενο μοντέλο δεν εμπεριείχε σε καμία περίπτωση την ανεξάρτητη μεταβλητή του δείκτη μεσογειακής διατροφής (**MedDietScore**). Θέλοντας να μελετήσουμε ένα μοντέλο λογιστικής παλινδρόμησης που περικλείει το δείκτη μεσογειακής διατροφής έπρεπε να «χτίσουμε» από την αρχή ένα νέο μοντέλο.

Η διαδικασία ήταν ιδιαίτερα πολύπλοκη καθώς έπρεπε μέσα από αρκετούς πειραματισμούς και δοκιμές να εντοπίσουμε το μοντέλο που θα συνδύαζε τις στατιστικά σημαντικές μεταβλητές που αναφερθήκαμε προηγουμένως με το δείκτη μεσογειακής διατροφής. Ο λόγος που θεωρήθηκε χρονοβόρα και επίπονη διαδικασία είναι επειδή ορισμένες ανεξάρτητες μεταβλητές είναι ισχυρά συσχετισμένες μεταξύ τους. Αυτό είχε σαν αποτέλεσμα να αλλάζουν άρδην τα αποτελέσματα του μοντέλου όταν προσθέταμε μια νέα ανεξάρτητη μεταβλητή μέσα στο μοντέλο. Επομένως έπρεπε να βρεθεί ο κατάλληλος συνδυασμός στατιστικά σημαντικών μεταβλητών, που να επιτρέπει τη παραμονή και του δείκτη της Μεσογειακής διατροφής.

Εν τέλει καταλήξαμε σε ένα μοντέλο λογιστικής παλινδρόμησης με μεταβλητή απόκρισης τη δίτιμη μεταβλητή που φανερώνει την ύπαρξη σακχαρώδη διαβήτη (*diabetes Mellitus* – **dm**) και με ερμηνευτικές μεταβλητές το δείκτη Μεσογειακής διατροφής (*Mediterranean Diet Score* - **MedDietScore**), την ηλικία (**age**), την οικογενειακή κατάσταση (*marital status* – **family**) και το δείκτη μάζας σώματος (*Body Mass Index* – **bmi**).

Για να αποφανθούμε στατιστικά εάν το μοντέλο με τις ανεξάρτητες μεταβλητές, προβλέπει αποτελεσματικότερα από το αρχικό μοντέλο, εάν κάποιος πάσχει από σακχαρώδη διαβήτη, το πακέτο δίνει παράλληλα και ένα έλεγχο wald. Ο τελευταίος έχει μηδενική υπόθεση ότι το αρχικό μοντέλο προσαρμόζεται καλύτερα από το κορεσμένο. Το p-value του ελέγχου βρέθηκε σχεδόν μηδέν, απορρίπτοντας τη μηδενική υπόθεση. Ουσιαστικά συμπεραίνουμε πως το κορεσμένο μοντέλο προσαρμόζεται καλύτερα από το αρχικό και προβλέπει ασφαλέστερα εάν κάποιος πάσχει από σακχαρώδη διαβήτη ή όχι.

Omnibus Tests of Model Coefficients			
	Chi-square	df	Sig.
Step 1 Step	99.530	6	.000
Block	99.530	6	.000
Model	99.530	6	.000

Πίνακας 5.31: Πίνακας διαφορών πιθανοφανειών

Υπάρχουν περίπου 10 διαφορετικά μέτρα προσαρμογής του μοντέλου της λογιστικής παλινδρόμησης. Από αυτά άλλα έχουν πολλές ομοιότητες μεταξύ τους και άλλα διαφέρουν σημαντικά. Το πακέτο που επιλέξαμε για ανάλυση υπολογίζει δυο από τα διαθέσιμα, το  $R^2$  των Cox – Snell και το  $R^2$  του Nagelkerke. Και τα δύο παίρνουν τιμές στο διάστημα [0,1] με τη μονάδα να υποδεικνύει άριστη προσαρμογή του μοντέλου στα δεδομένα. Η διαφορά μεταξύ τους είναι ότι ο Nagelkerke διαιρεί αυτό των Cox – Snell με τη μέγιστη τιμή του.

Στη περίπτωση μας παρατηρούμε πως υπάρχει διαφοροποίηση στις τιμές των δύο κριτηρίων, αλλά πρακτικά συμπεραίνουμε πως το μοντέλο δεν είναι ιδιαίτερα ισχυρό. Πιο συγκεκριμένα, το  $R^2$  των Cox – Snell ισχυρίζεται ότι ερμηνεύεται το 9,6% της συνολικής διακύμανσης της εξαρτημένης μεταβλητής και το  $R^2$  του Nagelkerke το 24,9%.

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	379.729 <sup>a</sup>	.096	.249

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

**Πίνακας 5.32:** Πίνακας μέτρων προσαρμογής λογιστικής παλινδρόμησης

Στη συνέχεια παρουσιάζουμε τον πίνακα προβλέψεων αλλά αυτή τη φορά για το μοντέλο που έχει μέσα όλες τις ανεξάρτητες μεταβλητές. Ενδεικτικά, παρατηρούμε ότι το μοντέλο προβλέπει ότι 978 άτομα δεν θα έχουν διαβήτη, πραγματοποιώντας σωστή πρόβλεψη για τους 916 αγγίζοντας το 99% ορθής πρόβλεψης. Από την άλλη μεριά, το μοντέλο προβλέπει σωστά μόνο το 6,2% των περιπτώσεων που πάσχουν από σακχαρώδη διαβήτη και συνολικά το μοντέλο προβλέπει σωστά στο 93,2% των περιπτώσεων, ποσοστό άκρως ικανοποιητικό. Σ' αυτό το σημείο θα θέλαμε να αναφέρουμε ότι, ότι τα ποσοστά ορθής ταξινόμησης για το παρόν μοντέλο δε διαφέρουν σε μεγάλο βαθμό από το μοντέλο που είχαμε προσαρμόσει όλες τις ερμηνευτικές μεταβλητές. Αυτό ίσως είναι ένα κριτήριο ώστε να επιλέξουν οι ειδικοί ένα πιο οικονομικό μοντέλο σε ενδεχόμενη μελλοντική έρευνα.

Classification Table <sup>a</sup>				
	Observed	Predicted		
		Diabetes Mellitus		Percentage Correct
		No diabetes	diabetes	
Step 1	Diabetes Mellitus	No diabetes	6	99.3
		diabetes	4	6.2
	Overall Percentage			93.2

a. The cut value is .500

**Πίνακας 5.33:** Πίνακας προβλέψεων και αντιστοίχισης περιπτώσεων αναλόγως εάν πάσχουν ή όχι από σακχαρώδη διαβήτη

Μέσω του πίνακα που ακολουθεί μπορούμε να αποφανθούμε μέσω ελέγχων Wald εάν οι ανεξάρτητες μεταβλητές διαδραματίζουν στατιστικά σημαντικό ρόλο στη διαμόρφωση του μοντέλου. Η μηδενική υπόθεση του ελέγχου είναι ότι  $\beta_i=0$  έναντι  $\beta_i \neq 0$  της εναλλακτικής. Τα p-values δίνονται στη πέμπτη στήλη του παρακάτω πίνακα και με πράσινο χρώμα είναι αυτά που μας οδηγούν σε απόρριψη της μηδενικής υπόθεσης. Επομένως, τα p-values των μεταβλητών της ηλικίας, της οικογενειακής κατάστασης, του δείκτη μάζας σώματος και του δείκτη Μεσογειακής διατροφής είναι μικρότερα από ένα επίπεδο σημαντικότητας 10% και επομένως απορρίπτουμε τη μηδενική υπόθεση. Ουσιαστικά αυτές οι μεταβλητές διαδραματίζουν στατιστικά σημαντικό ρόλο στη διαμόρφωση του μοντέλου.

Ένα εναλλακτικός τρόπος να καταλήξουμε στα ίδια συμπεράσματα με παραπάνω αλλά χωρίς τον έλεγχο Wald είναι μέσω των 95% διαστημάτων εμπιστοσύνης της  $\exp(B)$ . Πιο συγκεκριμένα όσα διαστήματα δεν συμπεριλαμβάνουν τη μονάδα, υποδηλώνουν πως οι αντίστοιχες μεταβλητές είναι στατιστικά σημαντικές για το μοντέλο.

Επιπροσθέτως, από τη στήλη B μπορούν να προκύψουν ερμηνείες αντίστοιχες της γραμμικής παλινδρόμησης για τις μεταβλητές που παίζουν στατιστικά σημαντικό ρόλο στο μοντέλο. Ενδεικτικά αναφέρουμε πως το -0,042 του δείκτη Μεσογειακής διατροφής, υποδηλώνει την αντίστοιχη μείωση της εξαρτημένης μεταβλητής του σακχαρώδη διαβήτη που σχετίζεται με αλλαγή μίας μονάδας στην ανεξάρτητη μεταβλητή του δείκτη Μεσογειακής διατροφής.

Τέλος στη στήλη exp (B) αναγράφονται οι λογάριθμοι της στήλης B. Ενδεικτικά αναφέρουμε ότι ο λογάριθμος της σχετικής πιθανότητας να ασθενήσει κάποιος από σακχαρώδη διαβήτη, αυξάνεται κατά 1,074 εάν αυξηθεί η ηλικία του κατά μια μονάδα, θεωρώντας όλες τις άλλες παραμέτρους σταθερές.

Κλείνοντας, η γενική μορφή του μοντέλου θα είναι:  $\text{logit}(p)=\log(p/1-p)=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5$  ενώ το προτεινόμενο μοντέλο λογιστικής παλινδρόμησης για τη μελέτη με τη μέθοδο enter, είναι της μορφής:  $\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = 0,002+(1,074*\text{age})+(0,467*\text{family2})+(0,407*\text{family3})+(0,777*\text{family4})+(1,062*\text{bmi})+(0,96*\text{MedDietScore})$

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Step 1 <sup>a</sup> age	.071	.013	31.160	1	.000	1.074	1.047	1.101
family			4.972	3	.174			
family(1)	-.762	.727	1.097	1	.295	.467	.112	1.942
family(2)	-.899	.439	4.193	1	.041	.407	.172	.962
family(3)	-.252	.760	.110	1	.740	.777	.175	3.446
bmi	.060	.028	4.755	1	.029	1.062	1.006	1.122
MedDietScore	-.041	.024	2.841	1	.092	.960	.915	1.007
Constant	-6.246	1.575	15.731	1	.000	.002		

a. Variable(s) entered on step 1: age, family, bmi, MedDietScore.

Πίνακας 5.34: Πίνακας στατιστικής σημαντικότητας ερμηνευτικών μεταβλητών



# ΚΕΦΑΛΑΙΟ 6<sup>ο</sup>

## ΣΥΜΠΕΡΑΣΜΑΤΑ

---

### 6.1 Συμπεράσματα

Σακχαρώδης διαβήτης ονομάζεται η διαταραχή του μεταβολισμού ή οποία συνοδεύεται από υψηλά επίπεδα σακχάρου στο αίμα. Διακρίνονται δύο κατηγορίες διαβήτη, ο τύπου I και II. Ο τύπου I συναντάται σπάνια και είναι αποτέλεσμα κακής λειτουργίας ζωτικών οργάνων όπως το πάγκρεας. Συνήθως ο ασθενής δεν ευθύνεται για την εμφάνιση του αλλά θα πρέπει να είναι ιδιαίτερα προσεκτικός με την εξέλιξη και την γενικότερη παρακολούθηση της πάθησης. Από την άλλη μεριά ο σακχαρώδης διαβήτης τύπου II απασχολεί το 90% με 95% των διαβητικών παγκοσμίως και οφείλεται κατά κύριο λόγο στη γενικότερη ζωή του ασθενή. Είμαστε σε θέση πλέον να γνωρίζουμε πως ο σύγχρονος τρόπος ζωής ωθεί τους ανθρώπους στους κόλπους της ασθένειας. Ο περιορισμένος ελεύθερος χρόνος σε συνδυασμό με τις προσταγές του σύγχρονου δυτικού lifestyle έχει σοβαρό αντίκτυπο στην υγεία του πληθυσμού του ανεπτυγμένου κόσμου. Η καθημερινή οικιακή μαγειρική διαδικασία βρίσκεται μειούμενη χρόνο με το χρόνο και τη θέση της παίρνουν ταχυφαγεία, εστιατόρια, ταβέρνες και καταστήματα εστίασης και αποστολής φαγητού κατ' οίκων. Επιπλέον, η σωματική δραστηριότητα των σύγχρονων ανθρώπων βρίσκεται σε διαρκή κάθοδο αφού πλέον οι μετακινήσεις είναι εξαιρετικά άμεσες και η γυμναστική έχει περάσει σε δεύτερη μοίρα. Εάν συνυπολογίσει κανείς τα κληρονομικά γονίδια και το γενικότερο τοξικό αστικό περιβάλλον, τότε μπορεί να καταλάβει γιατί ο σακχαρώδης διαβήτης τείνει να θεωρηθεί μάστιγα.

Η παρούσα μελέτη έγινε με βάση τα σεν δεδομένων δύο ερευνών, οι οποίες διεξήχθησαν από το 2001 έως το 2002 σε όλη την Ελλάδα. Το δείγμα ήταν 1035 ατόμων. Από αυτούς οι 100 είχαν διαγνωστεί επίσημα ότι πάσχουν από κάποιο είδος σακχαρώδη διαβήτη. Όσον αφορά τη κατανομή των δύο φύλων, φροντίστηκε να είναι σχεδόν ίδια μέσα στο δείγμα ώστε να μην υπάρχει μεροληψία. Από εκεί και πέρα, υπήρξε μέριμνα ώστε να συμπεριληφθούν και να μελετηθούν στο δείγμα διάφορες δημογραφικές μεταβλητές που ενδεχομένως επηρεάζουν σε θετικό ή αρνητικό βαθμό το σακχαρώδη διαβήτη, όπως η ηλικία, η οικογενειακή κατάσταση και το εκπαιδευτικό-οικονομικό επίπεδο των συμμετεχόντων. Σε μια έρευνα για μια μεταβολική ασθένεια όπως ο σακχαρώδης διαβήτης, δε θα μπορούσαν να απουσιάζουν μεταβλητές άμεσα συσχετισμένες με τη καθημερινή διατροφή των συμμετεχόντων. Μας ενδιέφερε να ενημερωθούμε σχετικά με το πόσα γεύματα έτοιμου φαγητού καταναλώνουν οι ερωτώμενοι σε μια εβδομάδα, επιπλέον καταλήξαμε ότι οι περισσότεροι συμμετέχοντες (85%) έπιναν τουλάχιστον ένα καφέ τη μέρα, η πλειοψηφία ανήκει σωματικά στους κανονικούς και παχύσαρκους με βάση το δείκτη μάζας σώματος και ότι το δείγμα είναι σχετικά ισομοιρασμένο σε αυτούς που ακολουθούν πιστά τη μεσογειακή διατροφή, αυτούς που βρίσκονται κοντά σε αυτήν

και τέλος σ' αυτούς που η διατροφή τους δεν έχει καμία σχέση με τη μεσογειακή. Τέλος υπήρξε μέριμνα για να έχουμε εικόνα όσων καπνίζουν συστηματικά (42%) και του επιπέδου φυσικής σωματικής δραστηριότητας των ερωτώμενων. Τα συμπεράσματα έδειξαν ότι οι περισσότεροι δεν αθλούνται καθημερινά (62%), ενώ μόνο το 5% των συμμετεχόντων απάντησε ότι κάνει καθημερινές απαιτητικές προπονήσεις.

Θέλοντας να ερευνήσουμε ποιες από τις διαθέσιμες μεταβλητές του δείγματος είχαν σχέση με τον σακχαρώδη διαβήτη, προχωρήσαμε σε διάφορους στατιστικούς ελέγχους. Ο στατιστικός έπρεπε πρώτα να βεβαιωθεί ότι οι προϋποθέσεις των επιλεγμένων στατιστικών μεθόδων ευσταθούν και στη συνέχεια να περάσει στην ερμηνεία των αποτελεσμάτων. Αναλυτικότερα, υπήρχαν δύο μεταβλητές οι οποίες μπορούσαν να χρησιμοποιηθούν εξίσου αποτελεσματικά σα μεταβλητή απόκρισης χωρίς να παραπέμπουν σε διαφορετικά συμπεράσματα. Η πρώτη ήταν μια δίτιμη μεταβλητή η οποία φανέρωνε την ύπαρξη σακχαρώδη διαβήτη ή όχι και η δεύτερη ήταν μια συνεχή μεταβλητή η οποία φανέρωνε τα επίπεδα γλυκόζης στο αίμα. Έχοντας δύο είδη δεδομένων για εξαρτημένες μεταβλητές υπήρχαν αρκετές επιλογές στη φαρέτρα του στατιστικού, ανάλογα και με τη φύση των ερμηνευτικών μεταβλητών.

Επιγραμματικά, βρέθηκαν ότι οι μεταβλητές του φύλου των συμμετεχόντων, η ηλικία, η οικογενειακή κατάσταση, η οικονομική-εκπαιδευτική κατάσταση, το βάρος, ο δείκτης μάζας σώματος, η κατανάλωση έτοιμων γευμάτων, η μεσογειακή διατροφή, οι καπνιστικές συνήθειες, τα καρδιαγγειακά επεισόδια, η διαστολική αρτηριακή πίεση, η υπέρταση και η υπερχοληστερολαιμία συσχετίζονται θετικά ή αρνητικά, ισχυρά ή όχι με το σακχαρώδη διαβήτη.

Επιπλέον χρησιμοποιήθηκαν οι τεχνικές της γραμμικής και λογιστικής παλινδρόμησης, αφού το επέτρεψε η φύση των μεταβλητών απόκρισης. Η γραμμικότητα και η κανονικότητα των καταλοίπων δεν επιβεβαιώθηκε σε κανένα σημείο της έρευνας, γι' αυτό το λόγο προχωρήσαμε σε μετασχηματισμό Box – Cox για να έχει στατιστική ισχύ η ερμηνεία των αποτελεσμάτων της γραμμικής παλινδρόμησης. Όσον αφορά το τελικό μοντέλο που επιλέχθηκε για το μοντέλο της γραμμικής παλινδρόμησης με τη μέθοδο stepwise, χρησιμοποίησε ως ερμηνευτικές μεταβλητές αυτές τις ηλικίας, της βακτηριακής περιτονίτιδας, των τριγλυκεριδίων και των ετών καπνίσματος. Από την άλλη μεριά, το τελικό μοντέλο που επιλέχθηκε για το μοντέλο της λογιστικής παλινδρόμησης με τη μέθοδο forward selection, χρησιμοποίησε ως ερμηνευτικές μεταβλητές αυτές της ηλικίας, του φύλου, της οικογενειακής κατάστασης και της υπέρτασης. Τέλος, αποφασίστηκε να «χτιστεί» από την αρχή ένα μοντέλο λογιστικής παλινδρόμησης στο οποίο να εμπερικλείεται η ερμηνευτική μεταβλητή της Μεσογειακής διατροφής. Σ' αυτό το μοντέλο προσαρμόστηκαν επίσης η ηλικία, η οικογενειακή κατάσταση και ο δείκτης μάζας σώματος. Ουσιαστικά παρατηρήσαμε ότι μόνο το φύλο βρίσκεται μέσα και στα τρία τελικά μοντέλα ως ερμηνευτική μεταβλητή.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

---

### Ελληνική Βιβλιογραφία

1. Μ. Κούτρας - Χ. Ευαγγελάρας (2010), *Ανάλυση Παλινδρόμησης*, εκδόσεις Αθ. Σταμούλης
2. Κ. Τσίμπος – Φ. Γεωργιακώδης (1999), *Περιγραφική και Διερευνητική Στατιστική Ανάλυση Δεδομένων*, εκδόσεις Αθ. Σταμούλης
3. Χ. Γναρδέλλης (2009), *Ανάλυση Δεδομένων με το SPSS Statistics 17.0*, εκδόσεις Παπαζήση
4. Γ. Τζαβελάς (2018), *Βιοστατιστική και Στατιστικές Μέθοδοι στην Επιδημιολογία*, πανεπιστημιακές σημειώσεις
5. Κ. Πολίτης (2018), *Πίνακες Συνάφειας*, πανεπιστημιακές σημειώσεις
6. Α. Σαχλάς – Σ. Μπερσίμης (2016), *Εφαρμοσμένη Στατιστική με Έμφαση στις Επιστήμες Υγείας*, εκδόσεις Α. Τζιόλα
7. Α. Σαχλάς – Σ. Μπερσίμης (2016), *Εφαρμοσμένη Στατιστική με χρήση του IBM SPSS Statistics: Με έμφαση στις επιστήμες υγείας*, εκδόσεις Α. Τζιόλα
8. Α. Σαχλάς – Σ. Μπερσίμης (2017), *Ανάλυση Ιατρικών Δεδομένων*, πανεπιστημιακές σημειώσεις
9. Γ. Ηλιόπουλος (2017), *Γενικευμένα Γραμμικά Μοντέλα*, πανεπιστημιακές σημειώσεις
10. Μ. Κούτρας (2012), *Πολυμεταβλητή Ανάλυση*, πανεπιστημιακές σημειώσεις
11. Ελληνική Διαβητολογική Εταιρεία (2004), *Οδηγός Διατροφής για τη Ρύθμιση του Διαβήτη*, Ιατρικές εκδόσεις Ζήτα

## Ξένη Βιβλιογραφία

1. M. Pagano – K. Gauvreau (2000), Principles of Biostatistics, Thomson Publishing Company
2. International Diabetes Federation (2014), IDF Diabetes Atlas, IDF publishing
3. American Diabetes Association (2003), Report of the Expert Committee of the Diagnosis and Classification of Diabetes Mellitus, ADS publishing
4. Amylin Pharmaceuticals (2003), Diabetes a Multihormonal Disease
5. The Canadian Diabetes Association, Guidelines for the Nutritional Management of Diabetes Mellitus in the New Millennium
6. K. Peterson – G. Shulman (2002), Pathogenesis of Skeletal Muscle Insulin Resistance in Type 2 Diabetes Mellitus

## Διαδικτυακές Ιστοσελίδες

1. [www.onmed.gr](http://www.onmed.gr)
2. [www.medlabnews.gr](http://www.medlabnews.gr)
3. [www.pharmaserve.gr](http://www.pharmaserve.gr)
4. [www.lillydiabetes.com](http://www.lillydiabetes.com)
5. [www.wikipedia.com](http://www.wikipedia.com)
6. [www.sugarfree.gr](http://www.sugarfree.gr)
7. [www.ede.gr](http://www.ede.gr)
8. [www.ngda.gr](http://www.ngda.gr)
9. [www.emedip.gr](http://www.emedip.gr)

