

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ  
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ  
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΕΛΕΓΧΟΣ ΣΥΜΒΑΤΟΤΗΤΑΣ ΤΩΝ  
ΤΙΜΩΝ ΕΝΕΡΓΕΙΑΚΗΣ ΠΡΟΣΛΗΨΗΣ  
ΔΙΑΤΡΟΦΙΚΩΝ ΔΕΔΟΜΕΝΩΝ**

Αγγελική Π. Πυλαρινού

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  
Νοέμβριος 2004



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ  
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ  
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΕΛΕΓΧΟΣ ΣΥΜΒΑΤΟΤΗΤΑΣ ΤΩΝ  
ΤΙΜΩΝ ΕΝΕΡΓΕΙΑΚΗΣ ΠΡΟΣΛΗΨΗΣ  
ΔΙΑΤΡΟΦΙΚΩΝ ΔΕΔΟΜΕΝΩΝ**

Αγγελική Π. Πυλαρινού

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  
Νοέμβριος 2004

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:

- Πολίτης Κωνσταντίνος (Επιβλέπων)
- Κατέρη Μαρία
- Τσίμπος Κλέων

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**

**ASSESSMENT OF COMPATIBILITY OF  
ENERGY INTAKE MEASUREMENT  
FROM DIETARY DATA**

By

Angeliki P. Pylarinou

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  
November 2004



*Στην οικογένειά μου  
και  
στη φίλη μου Σωτηρία.*





## **Ευχαριστίες**

Ευχαριστώ τον επιβλέποντα καθηγητή μου κ. Πολίτη Κωνσταντίνο για τις πολύτιμες συμβουλές του και την καθοδήγησή του, τον Ορφανό Φίλιππο για την βοήθειά του και την πολύ καλή συνεργασία μας και την κ. Τριχοπούλου Αντωνία για τη διάθεση δείγματος πρωτογενών δεδομένων από την έρευνα του Ευρωπαϊκού Προγράμματος συνεργασίας Ιατρικής και Κοινωνίας (ΕΠΙΚ).



## Περίληψη

Το πρόγραμμα ΕΠΙΚ (Ευρωπαϊκό Πρόγραμμα συνεργασίας Ιατρικής και Κοινωνίας) είναι μια πολυκεντρική επιδημιολογική προοπτική μελέτη (cohort study) που σχεδιάστηκε για να ερευνήσει τη σχέση μεταξύ διατροφής και καρκίνου (άλλα και άλλων χρόνιων παθήσεων) καθώς και τη σχέση μεταξύ διατροφικών και άλλων παραγόντων που αφορούν στον τρόπο ζωής, στα γονίδια κ.λ.π. Στην Ελλάδα, η οποία είναι μία από τις 10 Ευρωπαϊκές χώρες που διεξάγεται το πρόγραμμα και είναι η χώρα με την οποία θα ασχοληθούμε, η μελέτη συντονίζεται από το Εργαστήριο Υγιεινής και Επιδημιολογίας (EYE) της Ιατρικής Σχολής του Πανεπιστημίου Αθηνών. Σκοπός μας είναι η αξιολόγηση της συμβατότητας των αναφερόμενων τιμών ενεργειακής πρόσληψης, που προκύπτουν από δεδομένα διατροφικής ανάκλησης 24ώρου. Η αξιολόγηση αυτή θα γίνει παρουσιάζοντας μια μεθοδολογία για την εύρεση ατόμων των οποίων η ενεργειακή πρόσληψη είναι χαμηλότερη της ελάχιστα απαιτούμενης (υποεκτιμούντες) για τη διατήρηση σταθερού βάρους υπό φυσιολογικές συνθήκες διαβίωσης. Επίσης, θα αναφέρουμε τρόπους εύρεσης της φυσικής δραστηριότητας ενός ατόμου αφού η φυσική δραστηριότητα είναι σημαντικός παράγοντας στον προσδιορισμό των υποεκτιμούντων. Η μέτρηση της φυσικής δραστηριότητας μπορεί να οδηγήσει στον προσδιορισμό της ενέργειας που συνήθως δαπανά ένα άτομο ανά ημέρα ή ώρα, εκφρασμένη είτε σε μονάδες ενέργειας (kcal ή kJ) είτε σε μεταβολικά ισοδύναμα (MET). Η μεθοδολογία που θα αναπτύξουμε θα εφαρμοστεί σε δεδομένα 300 Ελλήνων συμμετεχόντων ενώ η ανάλυσή μας θα γίνει με το στατιστικό πρόγραμμα STATA.



## **Abstract**

EPIC (European Prospective Investigation Into Cancer and Nutrition) is a multi-center prospective epidemiological cohort study investigating the association between diet and cancer (and other chronic diseases) and also the association between dietary and other lifestyle, genetic factors e.t.c. In Greece, which is one of the 10 European countries that participate in this study and is the country that we focus on, the study is coordinated by the Department of Hygiene and Epidemiology of the Medical School of the University of Athens. Our intention is the assessment of compatibility of the reported energy intake data, from data that derive from the “24-hour diet recall” calibration method. This assessment will take place expounding a methodology for the determination of participants that have energy intake lower than the minimum cut-off limits below which a person cannot live a normal life. Besides, we will mention some ways for the definition of physical activity for a participant, because physical activity is a significant factor for the determination of underreporters. The measurement of physical activity can lead to the assessment of habitual energy expenditure per day or hour, that expressed either in energy units (kcal ή kJ) or in metabolic equivalents (MET). The methodology that we will be developed will be applied to data from 300 Greek participants in the EPIC cohort and the statistical package STATA will be used for the statistical analysis.



# Περιεχόμενα

<b>1. Εισαγωγή</b>	<b>1</b>
<b>2. Τυποποίηση της μεθόδου βαθμονόμησης (calibration method) και παρουσίαση του λογισμικού EPIC-SOFT</b>	<b>7</b>
2.1 Εισαγωγή	7
2.2 Ανάπτυξη του EPIC SOFT σε τρία βασικά στάδια	8
2.2.1 Πρώτο στάδιο	9
2.2.1.1 Σφάλματα του ανταποκριτή	9
2.2.1.2 Σφάλματα του συνεντευκτή	10
2.2.1.3 Σφάλματα της μεθόδου	10
2.2.2 Δεύτερο στάδιο	11
2.2.2.1 Μείωση σφάλματος που οφείλεται στη μνήμη	11
2.2.2.2 Αναγνώριση και περιγραφή φαγητών και συνταγών	11
2.2.2.3 Ποσοτικοποίηση φαγητών και συνταγών	12
2.2.2.4 Χειρισμός συνταγών	14
2.2.2.5 Έλεγχος ποιότητας	15
2.2.3 Τρίτο στάδιο	16
<b>3. Διαδικασία για την επιλογή πληθυσμού</b>	<b>19</b>
3.1 Εισαγωγή	19
3.2 Follow-up	20
3.3 Διαδικασίες δειγματοληψίας από πρακτικής άποψης	21
3.4 Χρησιμοποίηση υπολογιστικής θεώρησης (logistics) για το σχεδιασμό βαθμονομημένων υπο-ομάδων του ΕΠΙΚ	25
3.5 Αποκλεισμός και συνυπολογισμός συμμετεχόντων από τα δεδομένα της διατροφικής ανάκλησης 24-ώρου	28
3.6 Ποσοστά συμμετοχής	28
3.7 Μελέτες με βάση το ΕΠΙΚ και το ΕΠΙΚ-Διατροφικής Ανάκλησης 24-ώρου	30

<b>4. Παραγωγή οριακών τιμών (cut-off limits) για αναγνώριση των υποεκτιμώντων (underreporters)</b>	<b>47</b>
4.1 Εισαγωγή	47
4.2 Ορισμός των ελαχίστων οριακών τιμών (cut-off limits)	48
4.3 Λεπτομερής ανάλυση/ παραγωγή του CUT-OFF 2	49
4.3.1 Υπόθεση για ενεργειακή ισορροπία	50
4.3.2 Ανάλυση των συντελεστών του S	51
4.3.3 Προσδιορισμός του κατώτερου ορίου	53
4.3.4 Παράδειγμα-Εφαρμογή (πάνω στο Goldberg cut-off όριο)	55
4.4 Βελτίωση των οριακών τιμών από την εξίσωση της Goldberg	56
4.4.1 Επίπεδο φυσικής δραστηριότητας (PAL)	57
4.4.2 Συντελεστής μεταβλητότητας της ενεργειακής πρόσληψης στις αναφερόμενες τιμές του ίδιου ατόμου $CV_{wEI}$ (within-individual CV for EI)	60
4.4.3 Συντελεστής μεταβλητότητας του δείκτη φυσικής δραστηριότητας $CV_{TP}$ (between-subject CV for PAL)	60
4.4.4 Συντελεστής μεταβλητότητας του ρυθμού βασικού μεταβολισμού $CV_{wB}$ (CV for BMR)	61
4.5 Μεταβολικά ισοδύναμα (MET)	62
<b>5. Στατιστική ανάλυση των δεδομένων</b>	<b>67</b>
5.1 Εισαγωγή	67
5.2 Περιγραφικά χαρακτηριστικά	68
5.3 Υποεκτιμούντες	74
5.3.1 Προσδιορισμός των υποεκτιμώντων	76
5.3.2 Απλή στατιστική συμπερασματολογία	80
5.3.3 Λογιστική παλινδρόμηση	93
5.4 Ανάλυση του δείκτη 24ωρης φυσικής δραστηριότητας όταν εκφράζεται σε MET	98
5.4.1 Περιγραφικά χαρακτηριστικά για το δείκτη 24ωρης φυσικής δραστηριότητας	99



5.4.2	Ανάλυση Διακύμανσης και Ανάλυση Συνδιακύμανσης για τη μεταβλητή MET	101
5.5	Συντελεστές Συσχέτισης	111
5.5.1	Συντελεστές συσχέτισης για τους υποεκτιμούντες	112
5.5.2	Συντελεστές συσχέτισης για τα MET	115
<b>6.</b>	<b>Γενικά συμπεράσματα</b>	<b>119</b>
	<b>Βιβλιογραφία</b>	<b>123</b>



# ΚΕΦΑΛΑΙΟ 1

## Εισαγωγή

Στο πολύπλοκο πεδίο της έρευνας για το ρόλο των παραγόντων και του τρόπου ζωής στην αιτιολογία χρόνιων ασθενειών, η διατροφή είναι, από τα ατομικά χαρακτηριστικά, το πιο δύσκολο να μετρηθεί. Αυτό γίνεται ακόμα πιο δύσκολο όταν ενδιαφερόμαστε για τη διατροφή όχι για μια μικρή περίοδο αλλά όταν ενδιαφερόμαστε για τη διατροφική πρόσληψη μιας μακράς περιόδου. Μετά από δεκαετίες εντατικής έρευνας στο πεδίο της διατροφικής επιδημιολογίας, τα αντιφατικά αποτελέσματα που ήρθαν στο προσκήνιο για τις σχέσεις μεταξύ διατροφής και καρκίνου, οδήγησαν σε μια αναθεώρηση των παραδοσιακών μεθόδων έρευνας καθώς επίσης και στην επεξεργασία νέων μεθοδολογικών σκέψεων και στη μελέτη σχεδιασμών για την έρευνα της σχέσης διατροφής και ασθενειών.

Μέχρι το τέλος της δεκαετίας του 80, ο επικρατών σχεδιασμός για τη μελέτη διατροφής και καρκίνου ήταν αναδρομικού τύπου (μελέτες ασθενών-μαρτύρων, case-control), περιλαμβάνοντας ασθενείς με κλινική διάγνωση καρκίνου (ασθενείς, cases) και άτομα χωρίς καρκίνο αντιπροσωπευτικά του εν δυνάμει πληθυσμού που προήλθε η ομάδα των ασθενών (μάρτυρες, controls). Αυτές οι μελέτες όμως παρουσιάζουν πιθανά μειονεκτήματα<sup>1</sup>. Οι ασθενείς αλλά και τα υπόλοιπα άτομα μπορεί, για παράδειγμα, να αναφέρουν για την περασμένη διατροφή τους διαφορετικά στοιχεία, ιδιαίτερα εάν η ασθένεια έχει οδηγήσει σε συνειδητή ή ασυνείδητη αλλαγή στη διατροφή, εισάγοντας έτσι συστηματική μεροληψία στις διατροφικές μετρήσεις μεταξύ cases και controls και διαστρέφοντας την ανάλυση της σχέσης μεταξύ διατροφής και καρκίνου. Επίσης, οι περιορισμένες αποκλίσεις στις διατροφικές εκθέσεις σε μελέτες που διεξήχθησαν σε ομάδες πληθυσμού με σχετικά ομοιογενή διατροφή, παρουσιάζει ένα επιπλέον ενδιαφέρον και θα μπορούσε να υπολογιστεί για την αναφερόμενη σχέση μεταξύ διατροφής και καρκίνου.

Κατά τη διάρκεια των δεκαετιών του 80 και του 90 μερικές μεγάλες προοπτικές μελέτες (cohort studies) ξεκίνησαν με σκοπό να υπερπηδήσουν τους αρχικούς περιορισμούς που σχετίζονται με case-control studies. Στις προοπτικές μελέτες, τα άτομα επιλέχθηκαν ανάλογα με την έκθεσή τους στον παράγοντα που εξετάζουμε (διατροφή) και παρακολουθούνται για

μία μεγάλη περίοδο κατά τη διάρκεια της οποίας νέα περιστατικά καρκίνου (cases) καταγράφηκαν. Επιπρόσθετα, δύο νέες μελέτες, το Ευρωπαϊκό Πρόγραμμα συνεργασίας Ιατρικής και Κοινωνίας-ΕΠΙΚ (European Prospective Investigation into Cancer and Nutrition-EPIC) και το Πολυεθνικό Cohort στη Χαβάη και το Λος Άντζελες, σχεδιάστηκαν για να αυξήσουν την ετερογένεια τόσο της έκθεσης (διατροφή) όσο και της ασθένειας (π.χ. καρκίνου), με τελικό σκοπό την αύξηση της στατιστικής ισχύος, ώστε να ανιχνεύσουν τη σχέση μεταξύ διατροφής και ασθένειας, εάν αυτή η σχέση είναι υπαρκτή.

Παράλληλα, με τη βελτίωση των σχεδιασμών των μελετών (study designs) και των μεθόδων διατροφικής εκτίμησης, για να μπορέσουμε έτσι να εκτιμήσουμε τη συνήθη ατομική διατροφική πρόσληψη κατά τη διάρκεια των δεκαετιών του 70 και του 80, μεγαλύτερη προσοχή δόθηκε στη μεθοδολογία για αναγνώριση, ποσοτικοποίηση και διόρθωση των σφαλμάτων που σχετίζονται με τις διατροφικές μετρήσεις.

Μία νέα γενιά σύνθετων σχεδιασμών για μελέτες εγκυρότητας ή βαθμονόμησης που χρησιμοποιούν διατροφικές ή βιολογικές μετρήσεις προτάθηκαν για να εκτιμήσουν την αξιοπιστία και να βελτιώσουν τη συγκρισιμότητα των διατροφικών εκθέσεων. Κατά τη διάρκεια των δεκαετιών του 80 και του 90, δύο βιολογικοί δείκτες (biological markers), το άζωτο ούρων (urinary nitrogen) και το διπλά επισημασμένο ύδωρ (doubly-labeled water) για την εκτίμηση του ολικού αζώτου (πρωτεΐνη) και την πρόσληψη ενέργειας, αντίστοιχα, εμφανίστηκαν ως οι πλέον υποσχόμενες ανεξάρτητες μετρήσεις αναφοράς για να χρησιμοποιηθούν σε μελέτες εγκυρότητας ή βαθμονόμησης<sup>1</sup>. Επίσης η χρήση των δύο αυτών βιοδεικτών (biomarkers) συνεισέφερε στο να διαλυθεί ο μύθος, όπου παρέμενε σταθερός μέχρι τα τέλη της δεκαετίας του 80, για την ύπαρξη μιας πρότυπης (gold standard) διατροφικής μεθόδου εκτίμησης. Όλες οι μελέτες εγκυρότητας ή βαθμονόμησης, που διεξήχθησαν μέχρι τώρα, χρησιμοποιώντας βιολογικούς δείκτες έδειξαν με συνέπεια και συμφωνία ότι όλες οι υπάρχουσες διατροφικές μέθοδοι εκτίμησης περικλείουν σφάλματα μέτρησης, αν και αυτά μπορεί να ποικίλλουν στο βαθμό και τη σημασία.

Οι μεγάλες πολυκεντρικές ή πολυεθνικές cohort studies, δημιούργησαν νέα στατιστικά και μεθοδολογικά ζητήματα που σχετίζονται με τη συγκρισιμότητα, στο επίπεδο του φαγητού και των διατροφικών συστατικών, των διατροφικών μετρήσεων που συλλέχθηκαν από πληθυσμούς με διαφορετικές διατροφικές συνήθειες και διαφορετική γεωγραφική και εθνική προέλευση<sup>1</sup>. Το υπάρχον πρόβλημα της εξασφάλισης ακριβών μετρήσεων για ένα άτομο σε μια μεγάλη περίοδο πρόσληψης, που σχετίζεται με όλες τις διατροφικές μεθόδους εκτίμησης,

ενισχύθηκε στις πολυκεντρικές μελέτες, όπου διαφορετικά διατροφικά ερωτηματολόγια χρησιμοποιήθηκαν για να βρουν την ετερογένεια των διατροφικών προτύπων που υπάρχουν στους μελετούμενους πληθυσμούς. Το μέγεθος και η φύση των συστηματικών και τυχαίων σφαλμάτων στις διατροφικές μετρήσεις ενδέχεται να ποικίλλουν στους μελετούμενους πληθυσμούς και να διαστρέψουν την εκτίμηση και την ερμηνεία της γενικής σχέσης μεταξύ διατροφής και ασθένειας όταν όλα τα cohort συνδυάζονται μαζί.

Για να αποφευχθούν ή να ελαχιστοποιηθούν τα σφάλματα μέτρησης και να εξασφαλίσουμε ότι η όποια συστηματική υπέρ ή υποεκτίμηση θα εφαρμοστεί με ίση σπουδαιότητα στους διάφορους μελετούμενους πληθυσμούς, έχει χρησιμοποιηθεί μία τυποποίηση της διατροφικής μεθόδου. Με την τυποποίηση, αναπτύσσουμε μία προσέγγιση που μας διαβεβαιώνει ότι η διαδικασία συλλογής και συνεπώς διαχείρισης και ανάλυσης διατροφικών δεδομένων είναι περίπου η ίδια παντού. Για αυτό το λόγο είναι κρίσιμο να έχουμε κατανοήσει καλά τους τύπους των σφαλμάτων που μπορεί να σχετίζονται με διατροφικές μετρήσεις.

Έτσι, λοιπόν, στην περίπτωση του ΕΠΙΚ, που είναι μια πολυκεντρική επιδημιολογική προοπτική μελέτη προοπτικού σχεδιασμού (cohort study) η οποία διεξήχθη σε 23 κέντρα 10 Ευρωπαϊκών χωρών (Δανία, Γαλλία, Γερμανία, Ελλάδα, Ιταλία, Ισπανία, Ολλανδία, Σουηδία, Αγγλία και Νορβηγία) και την οποία μελέτη θα αναλύσουμε λεπτομερώς παρακάτω, χρησιμοποιήθηκε ως αναφορική μέτρηση βαθμονόμησης (reference calibration measurement), σε όλες τις χώρες, η διατροφική μέθοδος ανάκλησης 24-ώρου (24-hour dietary recall method), και συγκεντρώνει διατροφικά στοιχεία για κάθε συμμετέχοντα για την προηγούμενη μέρα. Η τυποποίηση αυτής της μεθόδου οδήγησε στην δημιουργία ενός λογισμικού, του EPIC-SOFT, το οποίο μπορεί να παρουσιάζει πολλά πλεονεκτήματα αλλά όπως έχουμε πει δεν είναι δυνατόν να μην υπάρχουν ελλείψεις και σε αυτό.

Η δική μας στατιστική ανάλυση θα επικεντρωθεί στην Ελλάδα και συγκεκριμένα σε 300 συμμετέχοντες που επιλέχθηκαν με απλή τυχαία δειγματοληψία από τη βάση δεδομένων του ελληνικού τμήματος του ΕΠΙΚ και θα εξετάσουμε λεπτομερώς κάποια χαρακτηριστικά του δείγματος αυτού. Σκοπός μας είναι η αξιολόγηση της συμβατότητας των αναφερόμενων τιμών ενεργειακής πρόσληψης, που προκύπτουν από δεδομένα διατροφικής ανάκλησης 24-ώρου. Ο όρος «συμβατότητα» ενεργειακής πρόσληψης περιλαμβάνει τόσο άτομα που υποεκτιμούν την ενέργεια όσο και άτομα που την υπερεκτιμούν. Εμείς θα επικεντρωθούμε στην υποεκτίμηση της ενέργειας, αφού το πρόβλημα της υπερεκτίμησης είναι σπανιότερο,

καθώς επίσης και θα αναλύσουμε τρόπους εύρεσης της φυσικής δραστηριότητας ενός ατόμου. Η φυσική δραστηριότητα είναι σημαντικός παράγοντας στον προσδιορισμό αυτών που υποεκτιμούν την ενέργεια που προσλαμβάνουν, γιατί μπορούμε να ορίσουμε με ακρίβεια το επίπεδο φυσικής δραστηριότητας (Physical Activity Level, PAL) το οποίο είναι σημαντική παράμετρος στον προσδιορισμό των υποεκτιμούντων. Η μέτρηση της φυσικής δραστηριότητας μπορεί να οδηγήσει στον προσδιορισμό της ενέργειας που συνήθως δαπανά ένα άτομο ανά ημέρα ή ώρα, εκφρασμένη είτε σε μονάδες ενέργειας (kcal ή kJ) είτε σε μεταβολικά ισοδύναμα (metabolic equivalent, MET)<sup>21</sup>, όπου το ένα MET αντιστοιχεί στην ενέργεια που δαπανά ένα άτομο υπό συνθήκες ανάπαυσης. Τέλος, μπορούμε να ελέγξουμε κατά πόσο τα παχύσαρκα άτομα, υποεκτιμούν σε μεγαλύτερο βαθμό την ενεργειακή πρόσληψη, όπως έχει βρεθεί σε άλλες μελέτες<sup>18,19</sup>, με το δείκτη μάζας σώματος (Body Mass Index, BMI) όπου ορίζεται σαν τον λόγο του βάρους (kg) προς το τετράγωνο του ύψους (m<sup>2</sup>). Ο δείκτης μάζας σώματος αποτελεί έναν ικανοποιητικό δείκτη του βαθμού παχυσαρκίας και όσο μεγαλύτερος είναι αυτός ο δείκτης τόσο μεγαλύτερο είναι και το επίπεδο παχυσαρκίας (εκτός από το δείκτη BMI μπορούμε να ορίσουμε και το δείκτη WHR (waist-to-hip ratio), που είναι ο λόγος της περιφέρειας της μέσης προς την περιφέρεια του γοφού και μας δίνει ένα δείκτη για κεντρική παχυσαρκία<sup>4</sup>).

Αναλυτικά στη συνέχεια της εργασίας:

- στο 2<sup>ο</sup> κεφάλαιο θα αναλύσουμε την τυποποίηση της μεθόδου βαθμονόμησης, την διατροφική ανάκληση 24-ώρου, θα αναφέρουμε τις πιθανές πηγές σφαλμάτων που σχετίζονται με αυτή τη μέθοδο και πώς μπορούν να ελαχιστοποιηθούν αυτά τα σφάλματα, ενώ τέλος θα γίνει μια σύντομη παρουσίαση του λογισμικού EPIC SOFT,
- στο 3<sup>ο</sup> κεφάλαιο θα παρουσιάσουμε λεπτομερώς πώς έγινε η συγκέντρωση και η επιλογή του πληθυσμού (recruitment), ποιες διαδικασίες δειγματοληψίας ακολουθήθηκαν, τη διαχρονική παρακολούθηση (follow-up) και ο σχεδιασμός των βαθμονομημένων υποομάδων στη διατροφική ανάκληση 24-ώρου που χρησιμοποιήθηκε στα κέντρα του ΕΠΙΚ,
- στο 4<sup>ο</sup> κεφάλαιο παρουσιάζουμε λεπτομερώς, σε θεωρητικό επίπεδο, πώς γίνεται η εκτίμηση των δεδομένων πρόσληψης ενέργειας, χρησιμοποιώντας θεμελιώδεις αρχές της ενεργειακής φυσιολογίας, με την εφαρμογή κατώτερων οριακών τιμών (cut-off limits) για την αναγνώριση των υποεκτιμούντων (underreporters) καταναλωτών με βάση τις εξισώσεις των Goldberg και συνεργατών<sup>16</sup>. Επίσης, αναλύουμε μια μέθοδο για τον προσδιορισμό της

φυσικής δραστηριότητας των ατόμων αλλά ορίζουμε και έναν δείκτη 24ωρης φυσικής δραστηριότητας εκφρασμένο σε μεταβολικά ισοδύναμα (MET) και τέλος

• το 5<sup>ο</sup> κεφάλαιο περιλαμβάνει τη στατιστική ανάλυση και παρουσιάζουμε:

- ✓ τα περιγραφικά χαρακτηριστικά του δείγματός μας,
- ✓ τον προσδιορισμό των υποεκτιμούντων, αφού έχουμε αναπτύξει πρώτα μια μεθοδολογία για την εύρεση των ατόμων των οποίων η ενεργειακή πρόσληψη κρίνεται ανακριβής υπό φυσιολογικές συνθήκες διαβίωσης,
- ✓ τη στατιστική συμπερασματολογία για να δούμε κατά πόσο κάποια χαρακτηριστικά διαφέρουν μεταξύ των δύο ομάδων (των υποεκτιμούντων και μη), όπως για παράδειγμα το BMI ή τα MET,
- ✓ αναλύουμε ένα γενικευμένο γραμμικό μοντέλο (logit) έχοντας σαν εξαρτημένη μεταβλητή τους υποεκτιμούντες και ανεξάρτητες τις άλλες μεταβλητές της ανάλυσής μας, δηλαδή την οικογενειακή κατάσταση, την ηλικία, την περιφέρεια, την περιοχή, τα MET, το κάπνισμα, τη μόρφωση, την απασχόληση, το επίπεδο του WHR και το επίπεδο του BMI,
- ✓ επίσης εξετάζουμε ένα μοντέλο ανάλυσης συνδιακύμανσης με εξαρτημένη μεταβλητή τα MET και τέλος
- ✓ δίνουμε τους συντελεστές συσχέτισης (Pearson, Spearman) μεταξύ των υποεκτιμούντων και των MET με τις μεταβλητές που αναφέραμε παραπάνω και βρίσκουμε τους συντελεστές συσχέτισης που είναι στατιστικά σημαντικοί.





## ΚΕΦΑΛΑΙΟ 2

### Τυποποίηση της μεθόδου βαθμονόμησης (calibration method) και παρουσίαση του λογισμικού EPIC SOFT

#### 2.1 Εισαγωγή.

Το πρόγραμμα ΕΠΙΚ (Ευρωπαϊκό Πρόγραμμα συνεργασίας Ιατρικής και Κοινωνίας) είναι μια πολυκεντρική επιδημιολογική προοπτική μελέτη (cohort study) που σχεδιάστηκε για να ερευνήσει τη σχέση μεταξύ διατροφής και καρκίνου και να μας δώσει τη δυνατότητα αξιολόγησης και άλλων χρόνιων παθήσεων. Είναι η μεγαλύτερη πηγή η οποία είναι διαθέσιμη σήμερα παγκοσμίως για προοπτικές έρευνες πάνω στην αιτιολογία του καρκίνου (και άλλων ασθενειών). Το ΕΠΙΚ ξεκίνησε να ερευνά τη σχέση μεταξύ διατροφικών και άλλων παραγόντων που αφορούν στον τρόπο ζωής, στα γονίδια κ.λ.π.

Η μελέτη αυτή περιλαμβάνει 519.978 συμμετέχοντες από τους οποίους 366.521 είναι γυναίκες και 153.457 είναι άνδρες ηλικίας 35-70 χρόνων ενώ, όπως ήδη αναφέραμε, διεξήχθη σε 23 κέντρα 10 Ευρωπαϊκών χωρών (Δανία, Γαλλία, Γερμανία, Ελλάδα, Ιταλία, Ισπανία, Ολλανδία, Σουηδία, Αγγλία και Νορβηγία).

Η συγκέντρωση και η επιλογή (recruitment) του μελετούμενου πληθυσμού αποφασίστηκε από πρακτική και υπολογιστική (logistic) θεώρηση για να έχουμε μεγάλη συμμετοχή και μεγάλης περιόδου follow-up. Οι εγγραφές έγιναν από το 1992 μέχρι το 2000 και συγκεντρώθηκαν πληροφορίες που αφορούν στη διατροφή, σε ανθρωπομετρικούς παράγοντες, στον τρόπο ζωής ενώ συλλέχθηκε και δείγμα αίματος από κάθε συμμετέχοντα. Οι πληροφορίες αυτές, λοιπόν, συγκεντρώθηκαν τόσο μέσω ενός μη διατροφικού ερωτηματολογίου με μεταβλητές πάνω στον τρόπο ζωής όσο και μέσω ενός διατροφικού ερωτηματολογίου ενώ παρουσιάστηκαν ανθρωπομετρικές μετρήσεις και όπως αναφέραμε και πιο πάνω συλλέχθηκαν δείγματα αίματος. Σε κάθε χώρα οι διατροφικές πληροφορίες προήρθαν χρησιμοποιώντας ειδικές διατροφικές μεθόδους (διαφορετικές σε κάθε χώρα), ώστε να μπορέσουμε να εκτιμήσουμε μεγάλης περιόδου πρόσληψη για κάθε άτομο.

Έτσι λοιπόν, μία δεύτερη διατροφική μέθοδος, βασισμένη σε μια μηχανογραφημένη συνέντευξη, που ονομάζεται διατροφική ανάκληση 24-ώρου (24-hour dietary recall),

συγκέντρωσε δεδομένα από ένα υπό-δείγμα του ΕΠΙΚ και χρησιμοποιήθηκε σαν αναφορική μέθοδος βαθμονόμησης (reference calibration measurement) σε όλες τις χώρες. Ο βαθμονομημένος πληθυσμός ορίστηκε σαν ένα τυχαίο δείγμα από κάθε μία από τις ομάδες (cohorts) και έτσι δημιουργήθηκαν υπό-ομάδες όπου η κάθε υπό-ομάδα εκπροσωπείται από υπό δείγματα του 5-12% από τους συμμετέχοντες. Έτσι, συνολικά έχουμε 36.900 μετέχοντες από τη διατροφική ανάκληση 24-ώρου. Στην Ελλάδα, 5.406 από τους συνολικά 28.000 που περιλαμβάνονται στο ΕΠΙΚ, προσεγγίστηκαν και τελικά 2.930 έδωσαν συνέντευξη. Από αυτούς, θα εξετάσουμε λεπτομερώς κάποια χαρακτηριστικά ενός δείγματος 300 συμμετεχόντων.

Στις επιδημιολογικές μελέτες η 24-ωρη διατροφική ανάκληση προσφέρει αρκετά πλεονεκτήματα<sup>1</sup>, όπως για παράδειγμα ότι δεν απαιτεί γνώση ανάγνωσης και γραφής και έτσι μπορούμε να έχουμε ένα υψηλό ποσοστό από άτομα με διαφορετικό μορφωτικό και κοινωνικό-οικονομικό επίπεδο, είναι σχετικά φθηνό και ο χρόνος συνέντευξης είναι μικρός. Φυσικά, στο 24-ωρο ανάκλησης δεν είναι δυνατόν να μην υπάρχουν σφάλματα ενώ η μέθοδος έχει και την τάση να υπερεκτιμά το φαγητό και την εισαγωγή τροφής. Λόγω απουσίας μιας αμερόληπτης αναφορικής διατροφικής μεθόδου (dietary reference method) η τυποποίηση της ανάκλησης 24-ώρου φαίνεται σαν η καλύτερη λύση<sup>1</sup>. Έτσι δημιουργήθηκε ένα λογισμικό (software) πρόγραμμα, το EPIC-SOFT.

## **2.2 Ανάπτυξη του EPIC-SOFT σε τρία βασικά στάδια.**

Η ανάπτυξη του EPIC-SOFT διεξήχθη σε τρία βασικά στάδια<sup>1</sup>:

- Το πρώτο στάδιο ήταν η αναγνώριση των πιθανών πηγών σφαλμάτων που σχετίζονται με τις μετρήσεις της ανάκλησης 24-ώρου.
- Το δεύτερο στάδιο ήταν η ανάπτυξη προσεγγίσεων για να ελαχιστοποιηθεί ή να αποφευχθεί κάθε πιθανή πηγή σφαλμάτων που σχετίζεται με την ανάκληση 24-ώρου.
- Και τέλος, το τρίτο στάδιο ήταν η ανάπτυξη του προγράμματος EPIC-SOFT σε κάθε χώρα.

### **2.2.1 Πρώτο στάδιο.**

Το πρώτο στάδιο, όπως είπαμε, αφορά στην αναγνώριση πιθανών πηγών σφαλμάτων που σχετίζονται με τις μετρήσεις της διατροφικής ανάκλησης 24-ώρου.

Ένας βασικός σκοπός μιας διατροφικής μεθόδου είναι η ελαχιστοποίηση των σφαλμάτων μέτρησης. Η εργασία της τυποποίησης μιας διαιτητικής μέτρησης πρέπει να βασίζεται στη σωστή αναγνώριση και πρόληψη των πιθανών πηγών σφαλμάτων. Οι πιθανές πηγές σφαλμάτων που σχετίζονται με τις μετρήσεις της ανάκλησης 24-ώρου μπορούν να κατηγοριοποιηθούν σε τρεις ομάδες<sup>1</sup>:

1. στα σφάλματα του συνεντευκτή (αυτός που ρωτά),
2. στα σφάλματα του ανταποκριτή (αυτός που απαντά) και
3. στη διατροφική μέθοδο (εδώ είναι η διατροφική ανάκληση 24-ώρου) που χρησιμοποιείται σαν εργαλείο από τον συνεντευκτή για την συλλογή και ακολούθως την ανάλυση πληροφοριών που έχει συγκεντρώσει.

Οι αλληλεπιδράσεις μέσα σε αυτό το τριγωνικό σύστημα (ανταποκριτής, συνεντευκτή, μέθοδος) μπορεί θεωρητικά να επηρεάσει τρεις κύριες συνιστώσες των μετρήσεων του 24-ώρου ανάκλησης:

- αναφορά φαγητού (food reporting),
- αναγνώριση φαγητού (food identification) και
- ποσοτικοποίηση φαγητού (food quantification).

Ανάλογα με τον τύπο του σφάλματος έχουμε υποεκτίμηση ή υπερεκτίμηση της πρόσληψης φαγητού. Το τελικό σφάλμα μέτρησης, «net measurement error», όπως ονομάζεται, είναι αποτέλεσμα της αλληλεπίδρασης αυτών των διαφορετικών πηγών σφαλμάτων κατά τη διάρκεια της τυποποίησης της διατροφικής συνέντευξης.

Αναλύοντας τις τρεις πηγές σφαλμάτων, που αναφέραμε παραπάνω, έχουμε:

#### **2.2.1.1 Σφάλματα του ανταποκριτή.**

Η μνήμη του ανταποκριτή είναι ένα από τα κύρια προβλήματα. Το πρόβλημα μνήμης επηρεάζει την αναφορά του φαγητού με δύο τρόπους: μπορεί να ξεχάσει να αναφέρει φαγητά που κατανάλωσε (“errors of omission”) ή μπορεί να αναφέρει φαγητά που δεν κατανάλωσε (“errors of commission”).

### 2.2.1.2 Σφάλματα του συνεντευκτή.

Ο συνεντευκτής είναι επίσης μια πιθανή πηγή σφαλμάτων<sup>1</sup> λόγω της ανοικτής (open-ended) φύσης της ανάκλησης 24-ώρου και του προβλήματος μνήμης του συνεντευκτή. Αφού δεν γνωρίζει εκ των προτέρων τι θα δηλώσει ο ανταποκριτής, ο συνεντευκτής χρειάζεται να ξέρει όλους τους κανόνες για να αναγνωρίσει σωστά, να περιγράψει, να ποσοτικοποιήσει και να ελέγξει τα χιλιάδες φαγητά και συνταγές που αναφέρονται από τους συμμετέχοντες. Αυτά τα προβλήματα αφορούν κυρίως στον παραδοσιακό τρόπο ανάκλησης 24-ώρου που χρησιμοποιούσε ένα κενό φύλλο χαρτιού για τη συνέντευξη. Η δυσκολία του να αποστηθίσει κάποιος και να χρησιμοποιήσει όλους τους κανόνες κατά τη διάρκεια της συνέντευξης προσθέτει ένα παραπάνω φορτίο στη μνήμη του συνεντευκτή. Υπάρχουν βέβαια και άλλοι παράγοντες που αφορούν στην ανθρώπινη φύση, όπως η διατύπωση της ερώτησης, η μη σωστή χρήση κατάλληλων ερωτήσεων, λεκτικά ή άλλα προβλήματα στις απαντήσεις και η ανικανότητα μιας καλής σχέσης μεταξύ συνεντευκτή και ανταποκριτή, τα οποία μπορεί να γίνουν πιθανές πηγές σφαλμάτων, αν και είναι δύσκολο να καταγραφούν και να ελεγχθούν.

### 2.2.1.3 Σφάλματα της μεθόδου.

Η μέθοδος συλλογής, διαχείρισης και ανάλυσης των δεδομένων με τη μέθοδο της διατροφικής ανάκλησης 24-ώρου μπορεί επίσης να εισάγει τυχαία ή συστηματικά σφάλματα<sup>1</sup>. Τα κύρια βήματα της διαδικασίας και τα πιθανά σφάλματά της φαίνονται παρακάτω:

<b>Κύρια βήματα της μεθόδου 24-HDR</b>	<b>Πιθανά σφάλματα</b>
Συλλογή στοιχείων (συνέντευξη)	Έλλειψη μνήμης Αναγνώριση και περιγραφή φαγητού Ποσοτικοποίηση φαγητού Έλεγχος ποιότητας
Εισαγωγή δεδομένων	Κωδικοποίηση και ταξινόμηση φαγητού Έλεγχος ποιότητας
Υπολογισμός στοιχείων	Υπολογισμός πρόσληψης φαγητού και συνταγών Έλεγχος ποιότητας
Ανάλυση στοιχείων	Στατιστικές μέθοδοι ανάλυσης Συστήματα ταξινόμησης φαγητού Έλεγχος ποιότητας

## **2.2.2 Δεύτερο στάδιο.**

Το δεύτερο στάδιο αφορά στην ανάπτυξη προσεγγίσεων για να ελαχιστοποιηθεί ή να αποφευχθεί κάθε πιθανή πηγή σφαλμάτων που σχετίζεται με τη διατροφική ανάκληση 24-ώρου. Παρακάτω φαίνονται συγκεκριμένες προσεγγίσεις στο EPIC-SOFT για τη μείωση ή αποφυγή πιθανών πηγών σφαλμάτων<sup>1</sup>:

### **2.2.2.1 Μείωση σφάλματος που οφείλεται στη μνήμη.**

Μία σειρά από διαδικασίες συμπεριλήφθηκαν στο EPIC-SOFT για να βελτιώσουν την ανάκληση φαγητού που καταναλώθηκε την προηγούμενη μέρα και για να ελέγξει την πληρότητα και την συνάφεια των απαντήσεων. Η διατροφική ανάκληση σχεδιάστηκε σε δύο κύρια βήματα:

- στο πρώτο βήμα που ονομάστηκε «quick list», ο συμμετέχων ρωτάται να θυμηθεί σε γενικές γραμμές όλα τα φαγητά που έφαγε σε κάθε γεύμα την προηγούμενη μέρα
- στο δεύτερο βήμα, αφού έχει συμπληρωθεί η «quick list», υπάρχουν επιπρόσθετες ερωτήσεις στους μετέχοντες για να περιγράψουν και να ποσοτικοποιήσουν τα φαγητά με λεπτομέρεια.

Η λογική αυτής της προσέγγισης είναι ότι οι μετέχοντες έχουν μικρότερη δυσκολία να θυμηθούν τι έφαγαν εάν δεν ερωτηθούν να περιγράψουν και να ποσοτικοποιήσουν ταυτόχρονα. Στην πρώτη φάση της «quick list» διακρίνουμε 11 περιπτώσεις κατανάλωσης φαγητού κατά τη διάρκεια της μέρας (ξεκινώντας από “πριν το πρωινό” και καταλήγοντας “μετά το βραδινό, κατά τη διάρκεια της νύχτας”), όπου για κάθε περίπτωση στην οθόνη υπάρχουν ερωτήσεις για το μέρος και την ώρα που έγινε η κατανάλωση για να βοηθήσει τον συμμετέχοντα να θυμηθεί τι έκανε την προηγούμενη μέρα και ακολούθως τι κατανάλωσε. Σε κάθε επίπεδο της διαδικασίας, η πληροφορία που εισάγεται από τον συνεντευκτή ελέγχεται συστηματικά για την πληρότητα και για την συνέπειά της (π.χ. έκτροπες τιμές (outlier values), μη πλήρης περιγραφή του φαγητού). Με αυτόν τον τρόπο τα σφάλματα που οφείλονται στη μνήμη ελαχιστοποιούνται.

### **2.2.2.2 Αναγνώριση και περιγραφή φαγητών και συνταγών.**

Λόγω της ανοιχτής φύσης της ανάκλησης 24-ώρου, προκαλούνται συγκεκριμένες δυσκολίες επειδή ο συνεντευξιζόμενος μπορεί να δηλώσει ένα μεγάλο εύρος φαγητών. Η

διαδικασία τυποποίησης για την αναγνώριση και την περιγραφή φαγητών που αναπτύχθηκαν στο EPIC-SOFT βασίζεται σε δύο βασικές αρχές:

- πρώτο, να βεβαιώσει ότι τα φαγητά που πιθανόν αναφέρονται, έχουν εισαχθεί από πριν (pre-entered) στο software και
- δεύτερο, να τυποποιήσει το επίπεδο λεπτομερούς καταγραφής για την περιγραφή αυτών των φαγητών.

Για κάθε χώρα υπάρχουν συγκεκριμένες λίστες φαγητών και συνταγών μέσα στο EPIC-SOFT. Οι λίστες αυτές βασίστηκαν στις πιλοτικές μελέτες του ΕΠΙΚ καθώς και σε δεδομένα άλλων πηγών που έχουν σχέση με την κατανάλωση εθνικών φαγητών. Το EPIC-SOFT δίνει την δυνατότητα να εισάγουμε σαν “προσωρινά ελλείψεις” (“temporarily missing”) κάθε τεμάχιο που δεν βρίσκεται και μετά τη συνέντευξη αποφασίζεται αν η λίστα φαγητών θα αναβαθμιστεί με το νέο φαγητό ή αν ένα υπάρχον φαγητό είναι όμοιο και θα το αντικαταστήσει.

Για να τυποποιήσουμε τις μεθόδους αναγνώρισης και περιγραφής φαγητού, αναπτύχθηκαν δύο συμπληρωματικές περιγραφές φαγητών: “έμμεση περιγραφή” (“implicit description”) και “σαφής περιγραφή” (“explicit description”) που λαμβάνει υπόψη τις ομοιότητες και τις αποκλίσεις των φαγητών που καταναλώνονται μέσα και μεταξύ των χωρών. Η “έμμεση περιγραφή φαγητού” δηλώνει ότι το όνομα του φαγητού είναι αρκετό από μόνο του για να αναγνωρίσει και να περιγράψει το φαγητό π.χ. μήλο. Η “σαφής περιγραφή φαγητού” σημαίνει ότι ο συνεντευξιαζόμενος πρέπει να ερωτηθεί επιπλέον για να αναγνωρίσει και να περιγράψει πως καταναλώθηκε το φαγητό π.χ. διαφορετικές μέθοδοι μαγειρέματος χρησιμοποιούνται για το κρέας, το ψάρι, τα λαχανικά.

### **2.2.2.3 Ποσοτικοποίηση φαγητών και συνταγών.**

Η ποσοτικοποίηση στη διατροφική ανάκληση 24-ώρου, επιλέχθηκε να καλύψει τόσο τη μεγάλη ποικιλία τύπων φαγητού όσο και το μεγάλο εύρος στο μέγεθος των μερίδων, συμπεριλαμβανομένων των κυρίων πιάτων άλλα και των συστατικών των συνταγών. Επειδή κάθε συμμετέχων έχει δυσκολία στην εκτίμηση των μερίδων και λόγω της πιθανότητας ότι μπορούν να εισαχθούν σφάλματα όταν χρησιμοποιούνται διαφορετικές μέθοδοι για την ποσοτικοποίηση τους μεγέθους των μερίδων, η εκτίμηση της ποσότητας του φαγητού τυποποιείται στο EPIC-SOFT, βεβαιώνοντας έτσι ότι ίδιες μέθοδοι προσφέρονται σαν επιλογές για την ποσοτικοποίηση φαγητού. Με αυτή την προσέγγιση αυξάνεται η πιθανότητα

ότι, αν υπάρχουν σφάλματα, θα εφαρμοστούν ίσα ώστε να έχουμε εκτιμήσεις συγκρίσιμων φαγητών μέσα και μεταξύ των χωρών. Συνολικά χρησιμοποιούνται πέντε τύποι μεθόδων ποσοτικοποίησης. Για κάθε φαγητό, ο αριθμός και ο τύπος της ποσοτικής μεθόδου, ορίζεται σύμφωνα με 4 παραμέτρους: τύπος φαγητού, φυσική κατάσταση, εάν είναι ωμό ή ψημένο και εάν είναι με ή χωρίς μη φαγώσιμο μέρος. Βασιζόμενοι λοιπόν στο συνδυασμό αυτών των παραμέτρων, η συγκεκριμένη μέθοδος ποσοτικοποίησης που θα χρησιμοποιηθεί παρουσιάζεται στην οθόνη και ο συνεντευκτής επιλέγει την καταλληλότερη επιλογή. Οι πέντε επιλογές που χρησιμοποιούνται για τις μεθόδους ποσοτικοποίησης είναι<sup>1</sup>:

*1) Φωτογραφίες με μονάδες μέτρησης του νοικοκυριού.*

Μονάδες μέτρησης του νοικοκυριού όπως ποτήρια, μπολ, κουτάλια χρησιμοποιούνται για να εκτιμήσουμε την ποσότητα της μερίδας για τα υγρά, τα ποτά και τις ουσίες όπως το αλεύρι ή τη ζάχαρη. Για να διευκολυνθεί η εργασία αποφασίστηκε ότι ήταν πιο πρακτικό να χρησιμοποιούνται φωτογραφίες που δείχνουν τις μονάδες μέτρησης του νοικοκυριού και αφού προσδιορίζεται η μέτρηση ο συνεντευκτής να χρησιμοποιεί ένα χάρακα για να βρει τις ακριβείς διαστάσεις (π.χ. ύψος και διάμετρος ποτηριού).

*2) Τυποποιημένες μονάδες.*

Οι τυποποιημένες μονάδες χρησιμοποιήθηκαν για φαγητά όπως τα φρούτα, τα λαχανικά και άλλα εμπορικά προϊόντα που συσκευάζονται ανά μονάδα. Για τα φρούτα και τα λαχανικά αποφασίστηκε να χρησιμοποιηθούν τρεις τυποποιημένες μονάδες (μικρό, μεσαίο, μεγάλο) για να αποφευχθεί συστηματική υπερεκτίμηση ή υποεκτίμηση.

*3) Φωτογραφίες συγκεκριμένων φαγητών.*

Οι μερίδες των φαγητών και των συνταγών που δεν μπορούν να εκτιμηθούν ούτε από μονάδες μέτρησης του νοικοκυριού ούτε από τυποποιημένες μονάδες εκτιμώνται από φωτογραφίες. Για το ΕΠΙΚ υπάρχει ένα βιβλίο το οποίο περιέχει σετ από φωτογραφίες 140 φαγητών και συνταγών. Κάθε σετ περιέχει 4 έως 6 φωτογραφίες με μερίδες ταξινομημένες κατ' αύξον μέγεθος και με μια διαφορά μεταξύ των φωτογραφιών περίπου 25% για να έχουν οι ερωτώμενοι μια πραγματική οπτική αντίληψη της διαφοράς μεγέθους. Η επιλογή για τη μικρότερη και μεγαλύτερη μερίδα βασίστηκε στα δεδομένα που συλλέχθηκαν κατά την πιλοτική φάση του ΕΠΙΚ. Επιπλέον οι ερωτώμενοι μπορούσαν να χρησιμοποιήσουν κλάσματα ή πολλαπλάσια των μερίδων που έβλεπαν στις φωτογραφίες χωρίς έτσι να αυξηθεί

ο αριθμός των φωτογραφιών (π.χ. ο συμμετέχων μπορεί να δηλώσει ότι κατανάλωσε μόνο το 1/3 από την μερίδα που επέλεξε στη φωτογραφία).

Για τις φωτογραφίες, τα σφάλματα ποσοτικοποίησης του φαγητού μπορούν να αποδοθούν σε πρόβλημα αντίληψης των ερωτώμενων ανεξάρτητα από τα σφάλματα μνήμης. Γι' αυτό το λόγο δόθηκε μεγάλη προσοχή στις τεχνικές μεθόδους που χρησιμοποιήθηκαν για να βγουν οι φωτογραφίες δηλαδή ίδια γωνία, ίδια παρουσίαση του φαγητού για κάθε πιάτο κ.α.

#### *4) Καθορισμός γνωστής ποσότητας.*

Όταν ο ερωτώμενος γνωρίζει ακριβώς την ποσότητα φαγητού που κατανάλωσε, τότε η ποσότητα ή το βάρος μπορούν να εισαχθούν χειρονακτικά σε ένα κενό πεδίο που φαίνεται στην οθόνη. Αυτή η μέθοδος χρησιμοποιείται για να ποσοτικοποιήσει τα συστατικά των συνταγών όπου ο ερωτώμενος γνωρίζει την ποσότητα τους.

#### *5) Υπολογισμός και χρήση του συστήματος για άγνωστες ποσότητες.*

Όταν ο ερωτώμενος δεν μπορεί να δηλώσει το μέγεθος της μερίδας, τότε θεωρείται η ποσότητα σαν ελλιπής (missing). Όμως, έχουν προβλεφθεί προκαθορισμένες μερίδες για μικρές ποσότητες φαγητού (π.χ. λειωμένο τυρί σε πιάτο). Επιπλέον, για κάποιες προσθήκες στα φαγητά (π.χ. λίπη) οι ελλείπουσες ποσότητες αντικαθίστανται από τυποποιημένα ποσά που υπολογίστηκαν αυτόματα βασιζόμενοι στο βάρος του φαγητού που καταναλώθηκε. Επίσης διαφορετικοί αλγόριθμοι αναπτύχθηκαν για να μετατρέψουν τον αρχικό τύπο έκφρασης του φαγητού (όπως εκτιμήθηκε κατά τη διάρκεια της συνέντευξης) στον τελικό τύπο κατανάλωσης (όπως τελικά καταναλώθηκε από τον συμμετέχοντα). Οι μετατροπές από τεμάχια σε βάρος εκφρασμένα σε γραμμάρια, από βάρος ωμού φαγητού σε βάρος μαγειρευμένου και από φαγητό με ή χωρίς μη φαγώσιμο μέρος, υπολογίζονται αυτόματα από το software χρησιμοποιώντας τυποποιημένες μετατροπές για συγκεκριμένα φαγητά.

#### **2.2.2.4 Χειρισμός συνταγών.**

Η διαχείριση των συνταγών είναι κρίσιμο σημείο για την τυποποίηση της διατροφικής ανάκλησης 24-ώρου, αφού τα συστατικά και οι μέθοδοι προετοιμασίας διαφέρουν πολύ ενώ ένα μεγάλο ποσοστό φαγητών προέρχεται από συνταγές. Οι συνταγές προσδιορίζονται σε ένα γκρουπ φαγητών βασιζόμενοι στα κύρια συστατικά. Με αυτή την προσέγγιση έχει βρεθεί<sup>1</sup> ότι υπερεκτιμάμε το γκρουπ των φαγητών που θεωρούνται σαν κύριοι συντελεστές όπως κρέας,



ψάρι κ.α. ενώ υποεκτιμάμε άλλα όπως γαλακτοκομικά προϊόντα, λίπη, λάδι κ.α. (για παράδειγμα τα λίπη και το λάδι διπλασιάζουν τη συνεισφορά ενέργειας όταν τα μεικτά πιάτα αναλυθούν στα συστατικά τους). Στην προσέγγιση που υιοθετήθηκε από το EPIC-SOFT, όλες οι συνταγές αναλύονται στα συστατικά τους έτσι ώστε το πρόβλημα της σύγκρισης, της ταξινόμησης και της διαχείρισης των συνταγών μέσα και μεταξύ των χωρών να μπορέσει να ξεπεραστεί. Για να τυποποιήσουμε την ταξινόμηση σε ομάδες φαγητών, ήταν απαραίτητο να γνωρίζουμε ποιους τύπους προϊόντων θα χειριστούμε σαν συνταγές και ποιους σαν φαγητά (για παράδειγμα ψωμί, κέικ, μπισκότα, σούπες και ποτά θεωρήθηκαν φαγητά).

Η ανάλυση στα συστατικά και ο υπολογισμός αυτών των ποσοτήτων χωριστά, επιτεύχθηκε σε δύο βήματα:

- πρώτον η συνολική ποσότητα της συνταγής όπως καταναλώθηκε εκτιμήθηκε χρησιμοποιώντας μία από τις προκαθορισμένες μεθόδους ποσοτικοποίησης του software (π.χ. φωτογραφίες, τυποποιημένες μονάδες) και
- παράλληλα, κάθε συστατικό προσδιορίστηκε από τον ερωτώμενο στην ωμή μορφή του ενώ στη συνέχεια μετατράπηκε στην τελική του μορφή δηλαδή όπως καταναλώθηκε.

#### **2.2.2.5 Έλεγχος Ποιότητας.**

Ο αντικειμενικός σκοπός του ελέγχου ποιότητας είναι να εμποδίσει και να αναγνωρίσει σφάλματα κατά τη διάρκεια της συνέντευξης και να ελέγξει ότι η διαδικασία που χρησιμοποιείται για τη συλλογή και το χειρισμό των δεδομένων είναι καλά τυποποιημένη μεταξύ των συνεντευκτών. Ένα μηχανογραφημένο πρόγραμμα συνέντευξης διατροφικής ανάκλησης 24-ώρου προσφέρει την τεχνική δυνατότητα να εκτελέσει μία μεγάλη σειρά από συστηματικούς ελέγχους ποιότητας. Το EPIC-SOFT ελέγχει για ελλειπείς πληροφορίες και έκτροπες παρατηρήσεις<sup>1</sup>. Αν και η ανάκληση 24-ώρου δεν είναι αρκετή για να εκτιμήσει ακριβώς την πρόσληψη μεγάλης περιόδου ενέργειας, η πρόσληψη ενέργειας που εκτιμήθηκε από την ανάκληση 24-ώρου συγκρίθηκε με τυποποιημένες απαιτήσεις ενέργειας που βασίστηκαν στην ηλικία, το φύλλο, το βάρος και το ύψος για να διορθωθούν πιθανά σφάλματα που αναγνωρίστηκαν σαν πολύ υψηλά ή χαμηλά συγκρινόμενα με τα τυποποιημένα. Για τις απαιτήσεις ενέργειας θα αναφερθούμε λεπτομερώς παρακάτω.

### 2.2.3 Τρίτο στάδιο.

Το τρίτο και τελευταίο στάδιο αφορά στην ανάπτυξη του προγράμματος EPIC-SOFT σε κάθε χώρα. Η γενική δομή του EPIC-SOFT σχεδιάστηκε για να προκαθορίσει και να τυποποιήσει τα μονοπάτια που ακολουθούνται κατά τη διάρκεια της διατροφικής ανάκλησης 24-ώρου σε όλες τις χώρες. Για το EPIC-SOFT δεν θα δώσουμε μια λεπτομερή περιγραφή<sup>1</sup> αλλά βασιζόμενοι και σε αυτά που έχουμε αναφέρει παραπάνω θα δώσουμε μια γενική εικόνα του προγράμματος:

Αυτή η γενική δομή βασίζεται πρώτα στο διαχωρισμό των «φαγητών» και των «συνταγών», που δηλώνονται από τους συμμετέχοντες. Το «φαγητό» είναι ένα απλό τεμάχιο (item) το οποίο μπορεί να διακριθεί και να ποσοτικοποιηθεί όπως για παράδειγμα φρούτα, λαχανικά, ψάρια, ψωμί, γλυκά. Η «συνταγή» είναι ένα σύνθετο πιάτο όπου όλα τα συστατικά δεν μπορούν να διακριθούν οπτικά και να ποσοτικοποιηθούν χωριστά π.χ. μουςακάς. Για να κάνουμε πιο εύκολη για τους συνεντευκτές τη διάκριση μεταξύ αυτών των δυο<sup>1</sup>, τα περισσότερα από τα «φαγητά» και οι «συνταγές» που καταναλώνονται συνήθως στις διάφορες χώρες του ΕΠΙΚ είναι από πριν περασμένα (pre-entered) σε δυο χωριστές λίστες, συμπεριλαμβάνοντας 1500-2200 φαγητά και 150-350 συνταγές ανάλογα με τη χώρα. Κατά τη διάρκεια της συνέντευξης, κάθε αναφερόμενο τεμάχιο υπάρχει στις λίστες. Ανάλογα με το αν το τεμάχιο είναι «φαγητό» (ή συστατικό συνταγής) ή «συνταγή» ακολουθείται ένα συγκεκριμένο μονοπάτι.

Τα μονοπάτια που ακολουθούμε κάθε φορά μπορούν να παρομοιαστούν σαν διαδοχικές υποδιαιρέσεις ενός ιεραρχικού δέντρου σε διακλαδώσεις<sup>1</sup>, όπου η διακλάδωση κάθε φορά που θα ακολουθήσουμε εξαρτάται από μερικές επιπρόσθετες ερωτήσεις κλειδιά (π.χ. εάν το τεμάχιο είναι διαθέσιμο στις λίστες, εάν η συνταγή είναι γνωστή ή όχι από τον ερωτώμενο, εάν η συνταγή είναι εμπορική ή όχι).

Η διαδικασία συνέντευξης τυποποιείται ανεξάρτητα από το μονοπάτι που ακολουθείται. Για παράδειγμα, τα κύρια βήματα της συνέντευξης (περιγραφή, ποσοτικοποίηση, χρήση κατάλληλων (probing) ερωτήσεων) παραμένουν τα ίδια και με την ίδια ακολουθία για όλα τα μονοπάτια. Επιπλέον, η αρχική αυτόματη έρευνα στην προκαθορισμένη λίστα συνταγών, αναγκάζει τον συνεντευκτή να χρησιμοποιήσει μια τυποποιημένη και από πριν περασμένη συνταγή. Αυτή η προσέγγιση<sup>1</sup> εγγυάται ότι για την ίδια συνταγή θα χρησιμοποιηθεί η ίδια μέθοδος ποσοτικοποίησης του μεγέθους της μερίδας και οι ίδιες κατάλληλες ερωτήσεις. Εάν ο ερωτώμενος γνωρίζει τη συνταγή, η τυποποιημένη συνταγή χρησιμοποιείται σαν βοήθεια

και τα συστατικά που αναφέρονται φτιάχνονται σε λίστα (κατηγοριοποιούνται) και διαδοχικά περιγράφονται και ποσοτικοποιούνται. Εάν ο ερωτώμενος δεν γνωρίζει τη συνταγή, είναι δυνατό να αντικαταστήσουμε μόνο ποιοτικά τα τυποποιημένα συστατικά με συστατικά γνωστά στον συμμετέχοντα (π.χ. αντικατάσταση λίπους που δεν καθορίζεται με λάδι ελιάς). Αυτή η προσέγγιση ελαχιστοποιεί τις αναπόφευκτες διαφορές στην ακρίβεια που αναφέρονται τα φαγητά μεταξύ των συμμετεχόντων που γνωρίζουν τη συνταγή που κατανάλωσαν και σε εκείνους που δεν γνωρίζουν (π.χ. μεταξύ ανδρών και γυναικών).

Όταν ένα «φαγητό» ή μια «συνταγή» λείπει από τις προκαθορισμένες λίστες, τότε προκαθορισμένες επιλογές προτείνονται στο συνεντευκτή για την περιγραφή, ποσοτικοποίηση και τον έλεγχο των νέων τεμαχίων που αναφέρονται. Επιπλέον, στο τέλος από αυτά τα μονοπάτια ένα αρχείο κειμένου (free text file) εμφανίζεται αυτόματα στην οθόνη, για να εισαχθούν όσες περισσότερες σημειώσεις είναι δυνατόν πάνω στα ελλιπή τεμάχια (missing items) που μπορούν να χρησιμοποιηθούν κατόπιν για να αναβαθμίσουν τις ελλιπείς πληροφορίες (missing information) στο αρχείο του EPIC-SOFT σύμφωνα με μια τυποποιημένη διαδικασία.



# ΚΕΦΑΛΑΙΟ 3

## Διαδικασία για την επιλογή του πληθυσμού

### 3.1 Εισαγωγή.

Συγκρίνοντας τις αναδρομικές μελέτες (case-control) με τις πολυκεντρικές μελέτες παρατηρούμε ότι οι μεγάλες πολυκεντρικές προοπτικές μελέτες προσφέρουν πολλά πλεονεκτήματα για την έρευνα της σχέσης μεταξύ διατροφής και άλλων παραγόντων που αφορούν στον τρόπο ζωής, με τον κίνδυνο χρόνιων ασθενειών. Αυτές οι μελέτες σχεδιάστηκαν για την αύξηση της στατιστικής ισχύος ώστε να ανακαλύψουν μία σχέση μεταξύ διατροφής και ασθένειας, συμπεριλαμβάνοντας μεγάλο μελετούμενο πληθυσμό που ποικίλει τόσο στον τύπο των διατροφικών γευμάτων όσο και στα ποσοστά περιστατικών καρκίνου, με αποτέλεσμα να αυξάνεται η ετερογένεια. Ωστόσο, οι πολυκεντρικές μελέτες προκαλούν σχετικά νέα στατιστικά και μεθοδολογικά ζητήματα<sup>2,3</sup> για τη σύγκριση και τη συνολική ανάλυση (pooled analysis) δεδομένων διατροφικής πρόσληψης που συλλέχθηκαν από μεγάλους ετερογενείς πληθυσμούς με ευρείες διαφορές στην κατανάλωση φαγητού, στη γλώσσα και στα κοινωνικό-πολιτιστικά χαρακτηριστικά. Στην πράξη, η έμφυτη δυσκολία εκτίμησης και σύγκρισης συνηθισμένης διατροφικής πρόσληψης των ατόμων, διευρύνονται στις μεγάλες πολυκεντρικές μελέτες, που τα διατροφικά ερωτηματολόγια συχνά διαφέρουν μεταξύ των μελετούμενων πληθυσμών για να συλλεχθούν συγκεκριμένες τοπικές διατροφές. Το μέγεθος και η φύση των συστηματικών και τυχαίων σφαλμάτων, στις μετρήσεις διατροφικής πρόσληψης μπορεί να ποικίλουν μεταξύ των μελετούμενων πληθυσμών και να παραποιούν την εκτίμηση και την ερμηνεία της συνολικής σχέσης μεταξύ διατροφής και ασθένειας όταν όλες οι ομάδες (cohorts) συνδυάζονται.

Για να προσαρμόσουμε για πιθανή συστηματική υπέρ ή υπό-εκτίμηση στις μετρήσεις διατροφικής πρόσληψης και να διορθώσουμε για μείωση μεροληψίας (bias) στις εκτιμήσεις σχετικού κινδύνου, αναπτύχθηκε μια προσέγγιση βαθμονόμησης. Αυτή η προσέγγιση περιλαμβάνει μία επιπρόσθετη διατροφική εκτίμηση μεταξύ των μελετούμενων πληθυσμών ώστε να ξανά εκφράσει ατομικές διατροφικές προσλήψεις σύμφωνα με την ίδια αναφορική κλίμακα (reference scale) και να επιτρέψει την άμεση συγκρισιμότητα των αποτελεσμάτων

μεταξύ των 10 Ευρωπαϊκών χωρών. Μία απλή ανάκληση 24-ώρου συλλέχθηκε, σαν μέθοδος αναφοράς για τη βαθμονόμηση (reference calibration method) των αρχικά συλλεγόμενων πληροφοριών του ΕΠΙΚ, από ένα στρωματοποιημένο τυχαίο δείγμα 36.900 συμμετεχόντων από όλο το ΕΠΙΚ cohort, χρησιμοποιώντας ένα λογισμικό (EPIC-SOFT) ειδικά σχεδιασμένο να τυποποιεί τις διατροφικές μετρήσεις των μελετούμενων πληθυσμών.

Εμείς θα περιγράψουμε το σχεδιασμό των βαθμονομημένων υπό-ομάδων μέσα στο ΕΠΙΚ, που όπως είπαμε απλώνεται σαν ένα δίκτυο από προοπτικές cohort studies που περικλείουν 23 ευρωπαϊκά κέντρα 10 ευρωπαϊκών χωρών. Μεταξύ των πιο σημαντικών χαρακτηριστικών για το ΕΠΙΚ είναι<sup>1</sup>: το μέγεθος, η γεωγραφική κατανομή και η ετερογένεια των διατροφικών προτύπων (dietary patterns) και άλλων παραγόντων τρόπου ζωής και κοινωνικό-πολιτιστικών χαρακτηριστικών των μελετούμενων πληθυσμών. Πληροφορίες για την συνήθη διατροφή, για τον τρόπο ζωής, για περιβαλλοντολογικούς παράγοντες και ανθρωπομετρικά στοιχεία για κάθε άτομο συλλέχθηκαν από βασική εξέταση (baseline examination), όπως δείγμα αίματος<sup>1</sup>.

Οι μελετούμενοι πληθυσμοί στο ΕΠΙΚ δεν επιλέχθηκαν για να εξασφαλίσουν αντιπροσωπευτικά δείγματα. Η συγκέντρωση και η επιλογή (recruitment) αποφασίστηκε από πρακτικής και υπολογιστικής (logistic) θεώρησης για να εξασφαλίσει υψηλή συμμετοχή και μεγάλης περιόδου follow-up από τους μετέχοντες στην μελέτη<sup>3</sup>.

### 3.2 Follow-up.

Έχουμε δύο ειδών follow-up<sup>2</sup>:

- ✓ το follow-up για αλλαγές στις συνθήκες υγείας και στον τρόπο ζωής τους και
- ✓ follow-up για περιστατικά καρκίνου και ολική θνησιμότητα.

Στην πρώτη περίπτωση, μετά την εγγραφή τα μέλη σε τακτά χρονικά διαστήματα, κάθε 3-4 χρόνια, επικοινωνούν για να δώσουν πληροφορίες για διάφορους παράγοντες που αφορούν στον τρόπο ζωής και είναι πιθανό να σχετίζονται με τον κίνδυνο του καρκίνου ενώ μπορούν να αλλάζουν συνεχώς όπως κάπνισμα, αλκοόλ, φυσική άσκηση, βάρος, εμμηνόπαυση κ.α. ενώ μία σειρά από ερωτήσεις προστίθενται για το αν οι συμμετέχοντες υποφέρουν από κάποια ασθένεια. Έτσι ουσιαστικά τα ερωτηματολόγια αυτά που αφορούν στο follow-up χρησιμοποιήθηκαν για να εκσυγχρονιστούν οι πληροφορίες για τη συλλογή στοιχείων.

Στη δεύτερη περίπτωση, το follow-up έχει σκοπό να καταγράψει περιπτώσεις καρκίνου οι οποίες είναι καταγεγραμμένες σε επίσημα βιβλία, από καταγραφές των ασφαλειών ζωής ή από πρόσωπα κοντινής συγγένειας. Από το Μάιο του 2002, το follow-up για τα περιστατικά

καρκίνου οδήγησαν στην αναγνώριση μεγάλου αριθμού ασθενών (1600) που ανέπτυξαν καρκίνο μετά την εγγραφή τους (πνεύμονας, προστάτη, στήθος, έντερο). Το 1996 δημιουργήθηκε μια επιτροπή η οποία προετοίμασε ένα λεπτομερές πρωτόκολλο για την συλλογή κλινικών και παθολογικών δεδομένων που αφορούν στον καρκίνο ενώ παράλληλα σε κάθε κέντρο συλλέχθηκαν στοιχεία για την ολική και ειδική θνησιμότητα βάση των καταγραφών θανάτου.

### **3.3 Διαδικασίες δειγματοληψίας από πρακτικής άποψης.**

Οι μελετούμενοι πληθυσμοί αντιπροσωπεύουν ετερογενή γκρουπ και είναι βασισμένοι στον πληθυσμό (population-based) (Ολλανδία, Ελλάδα, Γερμανία, Σουηδία, Δανία, Νορβηγία, Ισπανία, Ιταλία και ένα μικρό μέρος από την Οξφόρδη της Αγγλίας) ή σε μετέχοντες σε εξέταση του μαστού (breast screening) (Ουτρέχτη της Ολλανδίας, Φλωρεντία της Ιταλίας) ή σε δασκάλους στη Γαλλία. Στην Οξφόρδη, οι περισσότεροι (87%) συγκεντρώθηκαν και επιλέχθηκαν μεταξύ των ατόμων με ενδιαφέρον στην υγεία και θεωρούμενοι σαν “ενδιαφερόμενοι για την υγεία” (“health conscious”). Στα κέντρα της κεντρικής Ιταλίας και στην Ισπανία συγκεντρώθηκαν και επιλέχθηκαν δωρητές αίματος (σε διαφορετικά ποσοστά) ενώ στη Γαλλία, Νορβηγία, στην Ουτρέχτη της Ολλανδίας και στην Νάπολη της Ιταλίας επιλέχθηκαν μόνο γυναίκες<sup>2,3</sup>. Ο παρακάτω πίνακας περιγράφει αναλυτικά τα κέντρα των χωρών, από πού προέρχεται ο πληθυσμός, βάση ποιων κριτηρίων επιλέγεται ο κατάλληλος πληθυσμός αλλά και πώς έγινε η πρώτη επαφή με τους συμμετέχοντες:

Χώρα	Γεωγραφική περιοχή (κέντρα)	Πηγή Πληθυσμού (περιγραφή)	Κριτήρια καταλληλότητας	Αρχική επαφή
<u>Βασικά ΕΠΙΚ cohorts</u>				
<b>Ελλάδα</b>	Ελλάδα: πανεθνικό (nation-wide)	Κατάταξη (recruitment) από τον ολικό πληθυσμό.	Φαινομενικά υγιείς άνδρες και γυναίκες ηλικίας 25-82.	Αυτοπροσώπως και με mail
<b>Ισπανία</b>	Granada: επαρχία	Δωρητές αίματος, ολικός πληθυσμός (από απογραφή, από κέντρα υγείας).	Κάτοικοι: άνδρες ηλικίας 40-64 και γυναίκες ηλικίας 35-64.	Αυτοπροσώπως και με mail
	Murcia: περιφέρεια	Δωρητές αίματος και οι σύντροφοί τους (67%), ολικός πληθυσμός (23%), δημόσιοι υπάλληλοι (5%), εργαζόμενοι σε δύο εταιρίες (3%), μετέχοντες σε μελέτη καρδιαγγειακών κινδύνων (2%).	Κάτοικοι: άνδρες ηλικίας 40-65 και γυναίκες ηλικίας 35-65.	Αυτοπροσώπως και με mail
	Navarra: πόλη Pamplona και περιφέρεια Navarra	Δωρητές αίματος, ολικός πληθυσμός.	Κάτοικοι: άνδρες ηλικίας 40-65 και γυναίκες ηλικίας 35-65.	Mail
	San Sebastian: πόλη και επαρχία Gipuzkoa	Δωρητές αίματος, εργαζόμενοι σε επιλεγμένες επιχειρήσεις (από απογραφές επιλεγμένων δήμων).	Κάτοικοι: άνδρες ηλικίας 40-65 και γυναίκες ηλικίας 35-65.	Αυτοπροσώπως και με mail
	Asturias: περιοχή	Δωρητές αίματος, δημόσιοι υπάλληλοι και ολικός πληθυσμός.	Άνδρες ηλικίας 40-64 και γυναίκες ηλικίας 35-64.	Mail
<b>Ιταλία</b>	Ragusa: επαρχία	Τοπικοί συνεταιρισμοί δωρητών αίματος, τοπικός σύλλογος δασκάλων και άλλες πηγές.	Κάτοικοι: άνδρες ηλικίας 40-65 και γυναίκες ηλικίας 35-65.	Mail
	Florence: επαρχία	Άνδρες και γυναίκες από το γενικό πληθυσμό, μετέχοντες σε εξέταση καρκίνου του μαστού (breast cancer screening).	Κάτοικοι: άνδρες ηλικίας 35-64 και γυναίκες ηλικίας 35-64, χωρίς καρκίνο.	Αυτοπροσώπως και με mail
	Turin: πόλη	Δωρητές αίματος, υπάλληλοι, εθελοντές.	Κάτοικοι: άνδρες ηλικίας 40-74 και γυναίκες ηλικίας 35-74, χωρίς καρκίνο.	Αυτοπροσώπως
	Varese: επαρχία	Εθελοντές από το γενικό πληθυσμό, κυρίως μια επέκταση μιας προοδευτικής μελέτης (ORDET).	Άνδρες ηλικίας 40-65 και γυναίκες ηλικίας 35-65.	Αυτοπροσώπως και με mail



<b>Γαλλία</b>	Γαλλία	Εθνικό (nation-wide) πρόγραμμα ασφάλειας υγείας: δάσκαλοι και υπάλληλοι σε σχολεία κατατάχθηκαν σε μια προοδευτική μελέτη πριν από το ΕΠΙΚ (MGEN).	Γυναίκες ηλικίας 40-65 το 1990 με συναίνεση να αποκτήσουν πληροφορίες από τους μη ανταποκρινόμενους στο MGEN.	Mail
<b>Γερμανία</b>	Heidelberg και περίχωρα	Ολικός πληθυσμός.	Κάτοικοι: άνδρες ηλικίας 40-65 και γυναίκες ηλικίας 35-65, συμπληρωμένα ερωτηματολόγια και εξετάσεις.	Mail
	Potsdam και περίχωρα	Ολικός πληθυσμός.	Κάτοικοι: άνδρες ηλικίας 40-65 και γυναίκες ηλικίας 35-65, συμπληρωμένα ερωτηματολόγια και εξετάσεις.	Mail
<b>Ολλανδία</b>	Bilthoven: Amsterdam, Doetinchem and Maastricht	Πληθυσμός βασισμένος στη στρωματοποίηση ηλικίας και φύλλου του γενικού πληθυσμού.	Κάτοικοι: άνδρες και γυναίκες ηλικίας 20-60 στο Amsterdam και ηλικίας 20-65 στο Doetinchem.	Mail
	Utrecht:περιφέρεια	Πληθυσμός βασισμένος σε μετέχοντες με εξέταση καρκίνου του μαστού (breast cancer screening).	Κάτοικοι: γυναίκες ηλικίας 49-70.	Mail
<b>Αγγλία</b>	Cambridge:Norfolk	Πληθυσμός βασισμένος σε ασθενείς διαφορετικών εξασκούμενων επαγγελματιών (practitioners).	Λίστα από γενικά εξασκούμενα επαγγέλματα (general practitioners): άνδρες και γυναίκες ηλικίας 45-74.	Mail
	Oxford:τοπικές χώρες και "health conscious" από Αγγλία, Ουαλία, Σκωτία και Βόρεια Ιρλανδία	Πληθυσμός βασισμένος στην συνεργασία με εξασκούμενα επαγγέλματα (practitioners), χορτοφάγοι και άλλα "health conscious" άτομα σε συνεργασία με εταιρίες και περιοδικά για χορτοφάγους.	Λίστα από γενικά εξασκούμενα επαγγέλματα (general practitioners): άνδρες και γυναίκες ηλικίας 40-65 και άνδρες και γυναίκες ηλικίας 20+, αλλά με στόχο σε εκείνες με ηλικία 35+.	Mail
<u>Συνεργαζόμενα ΕΠΙΚ cohorts</u>				
<b>Ιταλία</b>	Naples	Γυναίκες εθελόντριες από τον γενικό πληθυσμό.	Γυναίκες ηλικίας 30-69.	Αυτοπροσώπως και με mail

<b>Δανία</b>	Aarhus	Βασισμένοι στον πληθυσμό.	Γεννημένοι στη Δανία: άνδρες και γυναίκες ηλικίας 50-64, χωρίς καρκίνο.	Mail
	Copenhagen	Βασισμένοι στον πληθυσμό.	Γεννημένοι στη Δανία: άνδρες και γυναίκες ηλικίας 50-64, χωρίς καρκίνο.	Mail
<b>Σουηδία</b>	Malmö: πόλη	Βασισμένοι στον πληθυσμό.	Κάτοικοι: άνδρες ηλικίας 50-72 και γυναίκες ηλικίας 46-72.	Mail
	Umea: η χώρα Vasterbotten	Βασισμένοι στον πληθυσμό.	Κάτοικοι: άνδρες και γυναίκες ηλικίας 30, 40, 50 και 60.	Mail
<b>Νορβηγία</b>	Tromsø:εθνικό δείγμα	Βασισμένοι στον πληθυσμό.	Γυναίκες γεννημένες στη Νορβηγία μεταξύ 1943 και 1957.	Mail

Πηγή: Riboli E. et al (2002)<sup>2</sup>  
Slimani N. et al (2002)<sup>3</sup>

Τα μεγέθη επιλέχθηκαν για να εξασφαλίσουν βαθμονόμηση και σε ατομικό και σε πληθυσμιακό επίπεδο ακόμα και όταν αναγνωρίστηκε ότι η ανάκληση 24-ώρου δεν θα είχε πλήρη ανεξάρτητα σφάλματα κατασκευής. Ένα σύνολο από 4000 διατροφικές ανακλήσεις 24-ώρου επιλέχθηκαν από κάθε ομάδα (cohort). Αυτό το μέγεθος του δείγματος επιτεύχθηκε στις περισσότερες χώρες εκτός από την Αγγλία (1117), τη Νορβηγία (1819) και την Ελλάδα (2930)<sup>3</sup>.

Οι διαδικασίες δειγματοληψίας ορίστηκαν σαν να έχουμε μια ίση κατανομή εποχής και ημέρας της συνέντευξης για να ελέγξουμε για πιθανή ημερήσια ή εποχική διακύμανση στην διατροφική κατανάλωση. Αν και στην Ισπανία, για παράδειγμα, τα ποσοστά ανταπόκρισης κατά τη διάρκεια της πιλοτικής φάσης ήταν τόσο υψηλά για τις συνεντεύξεις το Σάββατο όπως και τις υπόλοιπες ημέρες της εβδομάδας, σε άλλες χώρες όπως η Γαλλία και η Ολλανδία εμφανίστηκαν μεγάλα ποσοστά άρνησης για τις συνεντεύξεις του Σαββάτου ενώ σε άλλες χώρες πολλά κέντρα εξέτασης ήταν κλειστά κατά τη διάρκεια του Σαββατοκύριακου. Εξαιτίας αυτών των εμποδίων, θεωρήθηκαν εναλλακτικές μέθοδοι για να προσεγγίσουν και να πάρουν συνέντευξη από τους μετέχοντες κατά τη διάρκεια του Σαββατοκύριακου. Για παράδειγμα, συνεντεύξεις για τις Παρασκευές και τα Σάββατα συλλέγονταν 48 ώρες

αργότερα αντί για την επόμενη ημέρα ενώ προτάθηκαν συνεντεύξεις στο σπίτι για την αύξηση του ποσοστού συμμετοχής<sup>3</sup>.

### **3.4 Χρησιμοποίηση υπολογιστικής θεώρησης (logistics) για το σχεδιασμό βαθμονομημένων υπό-ομάδων του ΕΠΙΚ.**

Η συλλογή στοιχείων για τη μελέτη της βαθμονόμησης του ΕΠΙΚ πήρε πάνω από μία περίοδο 5 χρόνων μεταξύ Μαρτίου 1995 και Ιουνίου 2000. Η πιλοτική φάση ξεκίνησε αρχικά στη Γαλλία, Ισπανία και Ολλανδία για να δοκιμαστεί η πρώτη έκδοση του EPIC-SOFT ενώ στη συνέχεια επεκτάθηκε και στις άλλες χώρες<sup>3</sup>.

Για να έχουμε αντιπροσωπευτικούς βαθμονομημένους υπό-πληθυσμούς, η μελέτη βαθμονόμησης διήρκεσε μέχρι το τέλος της βασικής (baseline) επιλογής του ΕΠΙΚ. Ανάλογα με τη χώρα, χρειάστηκε 10-31 μήνες για να συλλεχθούν τα δείγματα των συνεντεύξεων ώστε να καλύψουν ημερήσια και εποχική διακύμανση. Συνολικά πήρε περισσότερο χρόνο να πραγματοποιηθούν οι συνεντεύξεις σε χώρες όπου: αρκετά κέντρα συμπεριλήφθηκαν σε διαφορετικές γεωγραφικές αποστάσεις (όπως στη Γαλλία, Ιταλία και Ισπανία), επιχειρήθηκε συνολική κάλυψη του πληθυσμού (Ελλάδα) και χρησιμοποιήθηκαν διαφορετικοί σχεδιασμοί και διαφορετικές μέθοδοι συγκέντρωσης και επιλογής. Ο χρόνος που απαιτήθηκε για την συλλογή στοιχείων για τη μελέτη του βαθμονομημένου ΕΠΙΚ δεν ήταν αυστηρά σχετικός με τον συνολικό αριθμό των συνεντεύξεων αφού τουλάχιστον ένας χρόνος χρειάστηκε για να καλυφθούν όλες οι εποχές. Επίσης είναι ενδιαφέρον να σημειωθεί ότι οι βόρειες (σκανδιναβικές) χώρες, που μπήκαν στο ΕΠΙΚ αργότερα, κέρδισαν περισσότερα πλεονεκτήματα<sup>1</sup> ενώ τα EPIC-SOFT προγράμματα ήταν πλήρως ανεπτυγμένα και δοκιμασμένα, που τους επέτρεψαν να τελειώσουν τις συνεντεύξεις πιο γρήγορα από τις άλλες χώρες<sup>3</sup>.

Όποτε ήταν δυνατόν οι συμμετέχοντες συγκεντρώθηκαν και επιλέχθηκαν στη βαθμονομημένη μελέτη αφηνιδιαστικά, όταν ήρθαν για την πρώτη τους βασική εξέταση. Η διατροφική συνέντευξη εκτελέστηκε σε 30-40 λεπτά αμέσως μετά την βασική εξέταση (baseline examination). Ήταν αναμενόμενο ότι αυτή η προσέγγιση συγκέντρωσης και επιλογής θα έδινε μεγαλύτερο ποσοστό συμμετοχής, καθώς οι συμμετέχοντες δεν θα έπρεπε να επιστρέψουν στα κέντρα εξέτασης. Αυτή η μέθοδος χρησιμοποιήθηκε κατά 84-100% στους μετέχοντες στο Παρίσι και στα περίχωρα, Potsdam, Ολλανδία και Δανία και σε μικρότερη έκταση στα άλλα γαλλικά κέντρα, Heidelberg (62%) και Ragusa (42%).

Στα άλλα κέντρα οι συμμετέχοντες είχαν ήδη προσκληθεί για βασική (baseline) εξέταση όταν άρχισε η βαθμονομημένη μελέτη. Οι συμμετέχοντες, που τυχαία επιλέχθηκαν να συμμετέχουν σε αυτή την μελέτη, ξανά ήρθαν σε επαφή είτε με γράμμα (Γαλλία, Αγγλία, Νορβηγία, Σουηδία) ή τηλεφωνικά (Ισπανία, Ιταλία εκτός από τη Ragusa και Ελλάδα). Σε αντίθεση με τη συγκέντρωση και την επιλογή κατά τη βασική εξέταση, στους συμμετέχοντες που ξανά προσκλήθηκαν με γράμμα ή τηλέφωνο ζητήθηκε να επιστρέψουν για να δώσουν επιπλέον πληροφορίες άλλα δεν ενημερώθηκαν για τον τύπο της διατροφικής μεθόδου και την χρονική περίοδο στην οποία θα αναφερθούν (π.χ. προηγούμενη μέρα)<sup>3</sup>. Αυτή η πρόληψη-προφύλαξη πάρθηκε για να αποφευχθούν αλλαγές στις συνήθειες διατροφικές συνήθειες και μεροληψία κατά τη διάρκεια της διατροφικής συνέντευξης (recalled dietary interview). Σύμφωνα με το τι ταιριάζει καλύτερα στους μετέχοντες και στις τοπικές ευκολίες που είναι διαθέσιμες, η πρόσωπο με πρόσωπο ανάκληση 24-ώρου συνέντευξη πραγματοποιήθηκε στο τοπικό κέντρο έρευνας ή στο σπίτι, ειδικά αν οι άνθρωποι ζούσαν μακριά από τα κέντρα έρευνας. Στη Γαλλία η γενική μελέτη διοργανώθηκε από το Παρίσι και οι τοπικές αρχές και οι συνασπισμοί (leagues) καρκίνου έφτιαξαν δωμάτια διαθέσιμα να για να διεξαχθούν οι διατροφικές συνεντεύξεις (π.χ. σχολεία, τοπικοί συνασπισμοί καρκίνου). Στην Ελλάδα που ήταν ιδιαίτερα δύσκολη η συγκέντρωση και η επιλογή (recruitment) τοπικών εθελοντών μακριά από την Αθήνα, χρησιμοποιήθηκαν κινητές μονάδες. Στη Νορβηγία οι συμμετέχοντες συνεντευξιάστηκαν στο σπίτι<sup>1</sup>. Σύμφωνα με τη χώρα και τη μέθοδο στρατολόγησης που χρησιμοποιήθηκε, 1 έως 4 υπενθυμιστικά στέλνονταν όταν οι μετέχοντες δεν απαντούσαν.

Στον παρακάτω πίνακα συνοψίζονται οι μέθοδοι συγκέντρωσης και επιλογής (recruitment) και ο εντοπισμός των διατροφικών συνεντεύξεων στην διατροφική ανάκληση 24-ώρου που πραγματοποιήθηκαν στα κέντρα του ΕΠΙΚ:

	Περίοδος κάλυψης μελέτης	Αριθμός ατόμων	Μέθοδος κατάταξης (recruitment) που χρησιμοποιήθηκε (%)			Αριθμός υπενθυμίσεων	Τύπος υπενθυμίσεων που χρησιμοποιήθηκε (+/-)			Τοποθεσία συνεντεύξεων (%)			
			Γράμμα	Τηλ.	Κέντρο		Κανένα	Γράμμα	Τηλ.	Κέντρο	Σπίτι	Άλλο	Ελλιπές
<b>Ελλάδα</b>	05/97-06/99	5406	-	100	-	2	-	-	+	-	-	100	-
<b>Ισπανία</b>		3741								99.4	0.2	0.2	0.2
<b>Granada</b>	12/95-10/96	722	-	100	-	0	+	-	-	100	-	-	-
<b>Murcia</b>	10/95-09/96	591	2	98	-	0	+	-	-	98.9	0.2	-	0.9
<b>Pamplona</b>	02/96-01/97	850	-	100	-	0	+	-	-	100	-	-	-
<b>San Sebastian</b>	07/95-06/96	806	100	100	-	0	+	-	-	98.1	0.7	0.8	0.4
<b>Asturius</b>	02/96-01/97	772	-	100	-	0	+	-	-	100	-	-	-
<b>Ιταλία</b>		4418								93.7	6.3	-	-
<b>Ragusa</b>	04/96-07/98	348	46	12	42	1	-	-	+	92.0	8.0	-	-
<b>Naples</b>	04/96-03/98	482	-	99	1	0	+	-	-	99.0	1.0	-	-
<b>Florence</b>	05/96-08/98	1155	-	99	1	0	+	-	-	92.0	8.0	-	-
<b>Turin</b>	04/96-03/98	1172	-	99	1	0	+	-	-	87.0	13.0	-	-
<b>Varese</b>	04/96-06/98	1261	-	99	1	0	+	-	-	97.0	3.0	-	-
<b>Γαλλία</b>		6456								29.5	19.9	45.9	4.7
<b>Languedoc/Roussillon</b>	10/95-04/97	869	100	-	-	2-3	-	+	-	-	22.4	73.1	4.5
<b>Aquitaine</b>	10/95-02/97	803	100	-	-	2-3	-	+	-	26.6	16.5	55.1	1.8
<b>Rhone-Alpes</b>	03/95-03/97	623	64	-	36	2	-	+	-	-	33.8	62.5	3.7
<b>Bretagne/Pays-de-Loire</b>	01/96-04/97	665	72	-	28	1-2	-	+	-	0.2	30.1	69.5	0.2
<b>Ile-de France</b>	01/96-07/97	1575	16	-	84	0	-	-	-	23.4	24.5	41.5	10.6
<b>Alsace-Lorraine</b>	10/95-02/97	1343	100	-	-	1-2	-	+	-	84.0	8.0	7.5	0.5
<b>Nord-Pas-de-Calais</b>	10/95-07/97	578	100	-	-	2-3	-	+	-	0.2	19.7	67.5	12.6
<b>Γερμανία</b>		4693								81.1	18.5	0.2	0.2
<b>Heidelberg</b>	06/96-10/98	2349	-	38	62	0	+	-	-	61.7	37.3	0.5	0.5
<b>Potsdam</b>	06/96-04/98	2344	-	-	100	0	+	-	-	98.9	1.1	-	-
<b>Ολλανδία</b>		5642								67.0	31.9	0.1	0.9
<b>Bilthoven</b>	04/95-12/97	3411	-	-	100	0	+	-	-	96.3	2.1	-	1.6
<b>Utrecht</b>	04/95-12/97	2231	-	-	100	0	+	-	-	24.8	74.8	0.3	1.1
<b>Αγγλία</b>		1900								99.4	0.6	-	-
<b>Cambridge</b>	03/97-04/98	905	100	-	-	1-4	-	+	+	98.7	1.3	-	-
<b>Oxford</b>	07/97-07/98	995	100	-	-	0	+	-	-	100	-	-	-
<b>Δανία</b>		4511								96.1	3.9	-	0.1
<b>Copenhagen</b>	07/96-05/97	3268	-	-	100	0	-	-	-	94.6	5.4	-	0.1
<b>Aarhus</b>	08/96-05/97	1243	-	-	100	0	-	-	+	100	-	-	-
<b>Σουηδία</b>		8413								100	-	-	-
<b>Malmo</b>	03/96-07/97	4064	100	-	-	2	-	+	+	100	-	-	-
<b>Umea</b>	03/97-10/98	4349	100	100	-	0	-	-	+	100	-	-	-
<b>Νορβηγία</b>	05/99-06/00	2993	100	-	-	1	-	+	-	-	100	-	-

Πηγή: Slimani N. (2002)<sup>1</sup>  
Slimani N. et al (2002)<sup>3</sup>

- 1.Στις περιπτώσεις του San Sebastian (Ισπανία) και του Umea (Σουηδία) οι μετέχοντες πρώτα ενημερώθηκαν με γράμμα και στην συνέχεια προσεγγίστηκαν τηλεφωνικά για να επιβεβαιώσουν την μέρα και την ώρα της συνέντευξης.
- 2.Στο Heidelberg (Γερμανία) οι μετέχοντες προσεγγίστηκαν στο κέντρο μελέτης όταν ήρθαν για την βασική (baseline) εξέταση. Εάν δεν ήταν δυνατό να κλείσουν ραντεβού αμέσως, οι μετέχοντες ξανά ήρθαν σε επαφή τηλεφωνικά
- 3.Στην Αθήνα, οι συνεντεύξεις διεξήχθησαν στα συντονιστικά κέντρα. Στην υπόλοιπη χώρα, οι συνεντεύξεις διεξήχθησαν σε δημόσια μέρη και περιστασιακά στο σπίτι.

### **3.5 Αποκλεισμός και συνυπολογισμός συμμετεχόντων από τα δεδομένα της διατροφικής ανάκλησης 24-ώρου.**

Αποκλείστηκαν 358 ( $\approx 1\%$ ) συνεντεύξεις, κυρίως λόγω τεχνικών προβλημάτων του λογισμικού κατά τη διάρκεια της συνέντευξης ή επειδή οι μετέχοντες δεν ήταν κατάλληλα τυχαιοποιημένοι ή αποκλείστηκαν από το ΕΠΙΚ για άλλους λόγους (π.χ. ελλιπή δεδομένα)<sup>3</sup>. Συμμετέχοντες με ηλικία κάτω από 35 και πάνω από 75 (που ήταν παρόντες σε μερικά ΕΠΙΚ cohorts) αποκλείστηκαν από τα δεδομένα πριν τη στατιστική ανάλυση. Αυτό αντιπροσωπεύει ένα σύνολο 945 συμμετεχόντων, κυρίως από το Bithoven (583 νέοι άνθρωποι) και την Ελλάδα (244, κυρίως ηλικιωμένοι).

Επιπλέον 357 ( $\approx 1\%$ ) συμμετέχοντες που αρχικά δεν ήταν στο δείγμα προστέθηκαν στον βαθμονομημένο πληθυσμό. Αυτοί οι συμμετέχοντες περικλείονταν σε άλλες ΕΠΙΚ διαστρωματικές (cross-sectional) μελέτες ή σε προηγούμενες πιλοτικές μελέτες του ΕΠΙΚ.

### **3.6 Ποσοστά συμμετοχής.**

Τα ποσοστά συμμετοχής<sup>1</sup>, στις διαβαθμισμένες (calibration) υπομελέτες, που αποκτήθηκαν στα διαφορετικά διοικητικά κέντρα φαίνονται παρακάτω. Αυτοί οι υπολογισμοί έγιναν πριν τους αποκλεισμούς. Σε επίπεδο χώρας, κυμαίνονται από 91.6% στην Γερμανία μέχρι 54.2% στην Ελλάδα ενώ 7 χώρες έχουν ποσοστό συμμετοχής από 75% και πάνω. Στη Γερμανία, στη Δανία και στην Αγγλία, δεν παρατηρήθηκε καμιά διαφορά στα ποσοστά ανταπόκρισης μεταξύ των κέντρων της ίδιας πόλης. Σε αντίθεση στη Γαλλία, ένα μεγαλύτερο ποσοστό παρατηρήθηκε στο Ile-de-France όπου οι μετέχοντες συνεντευξιάστηκαν αμέσως μετά την βασική (baseline) εξέταση, σε σύγκριση με τα άλλα κέντρα όπου οι μετέχοντες ξανά προσκλήθηκαν για συνέντευξη, μερικές φορές και μετά από 2 χρόνια<sup>1</sup>. Στην Ελλάδα το μικρό ποσοστό συμμετοχής οφείλεται στις υπολογιστικές (logistic) δυσκολίες προσέγγισης των συμμετεχόντων που ζούσαν έξω από την Αθήνα.

	Αριθμός ατόμων που επιλέχθηκαν	Αριθμός ατόμων που συνεντευξιάστηκαν	Ποσοστό συμμετοχής (%)
<b>Ελλάδα</b>	5406	2930	54.2
<b>Ισπανία</b>	3741	3222	86.1
<b>Granada</b>	722	515	71.3
<b>Murcia</b>	591	548	84.1
<b>Pamplona</b>	850	715	91.1
<b>San Sebastian</b>	806	734	86.1
<b>Asturius</b>	772	710	92.0
<b>Ιταλία</b>	4418	3961	89.7
<b>Ragusa</b>	348	306	87.9
<b>Naples</b>	482	403	83.6
<b>Florence</b>	1155	1058	91.6
<b>Turin</b>	1172	1069	91.2
<b>Varese</b>	1261	1125	89.2
<b>Γαλλία</b>	6456	4854	75.2
<b>Languedoc/Roussillon</b>	869	625	72.0
<b>Aquitaine</b>	578	443	76.6
<b>Rhone-Alpes</b>	1575	1018	64.6
<b>Bretagne/Pays-de-Loire</b>	803	635	79.1
<b>Ile-de France</b>	1343	1201	89.4
<b>Alsace-Lorraine</b>	665	480	72.2
<b>Nord-Pas-de-Calais</b>	623	452	72.6
<b>Γερμανία</b>	4693	4299	91.6
<b>Heidelberg</b>	2344	2126	90.7
<b>Potsdam</b>	2349	2173	92.5
<b>Ολλανδία</b>	5642	4585	81.4
<b>Bilthoven</b>	3411	2708	79.4
<b>Utrecht</b>	2231	1877	84.1
<b>Αγγλία</b>	1900	1117	59.0
<b>Cambridge</b>	905	547	60.4
<b>Oxford</b>	995	570	57.3
<b>Δανία</b>	4511	3919	86.9
<b>Copenhagen</b>	3268	2842	87.0
<b>Aarhus</b>	1243	1077	86.6
<b>Σουηδία</b>	8413	6195	73.6
<b>Malmo</b>	4064	3132	77.1
<b>Umea</b>	4349	3063	10.4
<b>Νορβηγία</b>	2993	1819	60.8

Πηγή: Slimani N. (2002)<sup>1</sup>

### 3.7 Μελέτες με βάση το ΕΠΙΚ και το ΕΠΙΚ-Διατροφικής Ανάκλησης 24-ώρου.

Στηριζόμενοι στη επιλογή του πληθυσμού που έχει παρουσιαστεί παραπάνω λεπτομερώς και στα διάφορα χαρακτηριστικά του, έχουν διεξαχθεί πολλές μελέτες. Παρακάτω παραθέτουμε ορισμένες μελέτες που έχουν γίνει με βάση τα δεδομένα του ΕΠΙΚ και της μεθόδου διατροφικής ανάκλησης 24-ώρου:

**(1) “Overweight, obesity and fat distribution in 50-to 64-year-old participants in the European Prospective Investigation into Cancer and Nutrition (EPIC)”<sup>4</sup>**, Haftenberger M, Lahmann PH et.al.

*Επιλογή πληθυσμού:*

Αυτή η μελέτη παρουσιάζει βασικά δεδομένα από το ΕΠΙΚ που συλλέχθηκαν μεταξύ 1992-2000. Η επιλογή του μελετούμενου πληθυσμού για κάθε κέντρο έχει αναλυθεί παραπάνω. Η παρούσα ανάλυση έγινε αυστηρά μεταξύ ηλικιών 50-64 που αντιπροσωπεύουν το κάθε κέντρο και το τελικό σύνολο ήταν 83.178 άνδρες και 163.851 γυναίκες. Όταν έγινε η ανάλυση δεν υπήρχαν ανθρωπομετρικά δεδομένα από τη Νορβηγία.

*Ανάλυση δεδομένων:*

Η στατιστική ανάλυση έγινε στο στατιστικό πακέτο SAS. Όλα τα δεδομένα της ανάλυσης στρωματοποιούνται ως προς το φύλλο. Για κάθε κέντρο παρουσιάζονται ο αριθμητικός μέσος, το τυπικό σφάλμα και το 10<sup>ο</sup>, 50<sup>ο</sup>(διάμεσος) και 90<sup>ο</sup> εκατοστημόριο για το βάρος, το ύψος, το δείκτη μάζας σώματος (body mass index, BMI), την περίμετρο μέσης (waist circumference, WC), την περίμετρο ισχύος (hip circumference, HC) και το λόγο περιμέτρου μέσης προς περίμετρο ισχύος (waist to hip-ratio, WHR, που είναι ο λόγος WC/HC). Τα εκατοστημόρια χρησιμοποιήθηκαν για να περιγράψουν την κατανομή για κάθε μέτρηση του σώματος σε κάθε κέντρο. Επιπλέον εξετάζουμε το σχήμα των κατανομών και την έκταση της ασυμμετρίας των μετρήσεων του σώματος. Η επικράτηση (prevalence) των υπέρβαρων, των παχύσαρκων και των κοιλιακών λιπών που συσσωρεύονται, προσδιορίστηκε εφαρμόζοντας οριακά σημεία που δηλώνουν υψηλό επίπεδο πάχους. [Εδώ πρέπει να προσδιορίσουμε ότι το BMI είναι ο λόγος βάρους (kg) δια το τετράγωνο του ύψους (m) και τα άτομα με βάση το BMI κατατάσσονται στις εξής κατηγορίες<sup>4</sup>: BMI<18.5 kg/m<sup>2</sup> έχουν βάρος λιγότερο του φυσιολογικού, 18.5 kg/m<sup>2</sup>< BMI<24.9 kg/m<sup>2</sup> έχουν φυσιολογικό βάρος, 25 kg/m<sup>2</sup> < BMI<29.9 kg/m<sup>2</sup> θεωρούνται υπέρβαρα και τέλος BMI≥30 kg/m<sup>2</sup> είναι παχύσαρκα. Επίσης οριακά σημεία ορίζονται τόσο για την περίμετρο μέσης (WC) που είναι 102cm για τους



άνδρες και 88cm για τις γυναίκες όσο και για το λόγο περιμέτρου μέσης προς περίμετρο ισχύος (WHR) που είναι 0.95 για τους άνδρες και 0.80 για τις γυναίκες ενώ τέλος οριακά.] Όλα τα στατιστικά μέτρα προσαρμόστηκαν (adjusted) ως προς την ηλικία χρησιμοποιώντας γραμμική παλινδρόμηση με την ηλικία σαν ανεξάρτητη και την ανθρωπομετρική μέτρηση σαν εξαρτημένη. Επίσης βρέθηκε η σχέση μεταξύ BMI και WC ή WHR σαν δείκτη της κατανομής των παχουλών καθώς επίσης και οι συντελεστές συσχέτισης του Pearson τόσο σε ατομικό επίπεδο για κάθε κέντρο όσο και σε επίπεδο των κέντρων. Τέλος περιγράφηκε η σχέση μεταξύ επικράτησης της παχυσαρκίας και επικράτησης μεγάλου WC ή WHR.

#### *Αποτελέσματα:*

Εκτός από την ομάδα των “ενδιαφερόμενων για την υγεία” (“health-conscious”) στην Αγγλία και τα κέντρα της Γαλλίας, το BMI υπερβαίνει το 25kg/m<sup>2</sup>. Η επικράτηση (The prevalence) της παχυσαρκίας (BMI≥30kg/m<sup>2</sup>) ποικίλλει από 8% έως 40% στους άνδρες και από 5% έως 53% στις γυναίκες με υψηλή επικράτηση (>25%) στα κέντρα από την Ισπανία, την Ελλάδα, Ragusa και Νάπολη (Ιταλία) και χαμηλή επικράτηση (<10%) στη Γαλλία και στην ομάδα των “health-conscious”. Η επικράτηση μεγάλου WC ή υψηλού WHR ήταν υψηλή στα κέντρα της Ισπανίας, της Ελλάδας, στις πόλεις Ragusa και Naples (Ιταλία) και μεταξύ των γυναικών στα κέντρα της Γερμανίας και Bilthoven (Ολλανδία). Άρα τα ανθρωπομετρικά δεδομένα διέφεραν μεταξύ του πληθυσμού του ΕΠΙΚ. Αυτά τα δεδομένα παρέχουν μια σταθερή βάση για επιπλέον έρευνα των ανθρωπομετρικών μετρήσεων σε σχέση με τον κίνδυνο χρόνιων ασθενειών και ειδικά καρκίνου.

**(2) “Physical activity of subjects aged 50-64 years involved in the European Prospective Investigation into Cancer and Nutrition (EPIC)”<sup>5</sup>, Haftenberger M, Schuit AJ, Tormo MJ et.al.**

#### *Επιλογή πληθυσμού:*

Ο αριθμός των ατόμων είναι ίδιος με της προηγούμενης μελέτης αφού αναφερόμαστε πάλι σε άτομα ηλικίας 50-64 χρόνων.

#### *Ανάλυση δεδομένων:*

Παρουσιάζονται για τους άνδρες και τις γυναίκες κάθε κέντρου οι κατανομές συχνοτήτων της επαγγελματικής δραστηριότητας (η επαγγελματική δραστηριότητα χωρίζεται σε: μη απασχόληση, καθιστική, όρθια, χειρονακτική και βαριά) και η συμμετοχή τους σε μη επαγγελματική δραστηριότητα καθώς και η διάρκεια (όταν λέμε μη επαγγελματική

δραστηριότητα αναφερόμαστε σε: περπάτημα, ποδηλασία, κηπουρική, σπορ/άσκηση, νοικοκυριό, επισκευές στο σπίτι). Για τη συμμετοχή των ατόμων στις διάφορες μη επαγγελματικές δραστηριότητες παρουσιάζονται ο μέσος, η τυπική απόκλιση και το 10<sup>ο</sup>, 50<sup>ο</sup>(διάμεσος) και 90<sup>ο</sup> εκατοστημόριο του χρόνου που αφιερώνεται σε αυτές τις δραστηριότητες. Όλα τα στατιστικά μέτρα προσαρμόστηκαν ως προς την ηλικία χρησιμοποιώντας γραμμική παλινδρόμηση με την ηλικία σαν ανεξάρτητη και την διάρκεια κάθε δραστηριότητας σαν εξαρτημένη μεταβλητή. Επίσης αναλύθηκε η εποχική διακύμανση της συνολικής διάρκειας της ψυχαγωγικής δραστηριότητας, υπολογίζοντας τη διάμεσο της διαφοράς της συνολική ψυχαγωγικής δραστηριότητας το καλοκαίρι μείον το χειμώνα και υπολογίζοντας το 95% διάστημα εμπιστοσύνης της. Οι διαφορές μεταξύ φυσικής δραστηριότητας το καλοκαίρι και το χειμώνα θεωρούνται σημαντικές όταν το μηδέν δεν περιλαμβάνεται στο δ.ε.

#### *Αποτελέσματα:*

Από την ταξινόμηση της επαγγελματικής δραστηριότητας, πιο συχνά παρατηρούμε την καθιστική και την όρθια σε όλα τα κέντρα. Παρατηρήθηκε μία ευρεία μεταβλητότητα που αφορά στη συμμετοχή σε διάφορους τύπους μη επαγγελματικών δραστηριοτήτων και του χρόνου που αφιερώθηκε σε αυτές σε όλα τα κέντρα. Πάνω από 80% όλων των συμμετεχόντων ασχολούνται με περπάτημα, ενώ λιγότερο από 50% συμμετέχουν στα σπορ. Ο συνολικός χρόνος που αφιέρωσαν στις ψυχαγωγικές δραστηριότητες ήταν υψηλότερος για τους Ολλανδούς μετέχοντες και χαμηλότερος μεταξύ των ανδρών από το Malmo (Σουηδία) και των γυναικών από τη Νάπολη (Ιταλία). Σε όλα τα κέντρα ο συνολικός χρόνος που αφιερώθηκε στις ψυχαγωγικές δραστηριότητες το καλοκαίρι ήταν υψηλότερος από το χειμώνα ενώ οι γυναίκες από τη νότια Ευρώπη ξόδεψαν τον περισσότερο χρόνο στο νοικοκυριό.

### **(3) “Consumption of vegetables, fruit and other plant foods in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohorts from 10 European countries”<sup>6</sup>,**

Agudo A, Slimani N, Ocke MC et.al.

#### *Επιλογή πληθυσμού:*

Τα αρχικά δεδομένα περιλαμβάνουν 39.900 άτομα ενώ μετά από κάποιους αποκλεισμούς λόγω μη πλήρων πληροφοριών ή τεχνικών προβλημάτων και περιορισμό στην ηλικία μεταξύ 35-74 χρόνων, τα άτομα που περιλαμβάνονται στην παρούσα ανάλυση είναι 35.955 (13.031

άνδρες και 22.924 γυναίκες). Αυτή η επιλογή πληθυσμού έχει αναλυθεί λεπτομερώς παραπάνω με τη μέθοδο της διατροφικής ανάκλησης 24-ώρου και είναι ίδια για όλες τις μελέτες παρακάτω.

#### *Ανάλυση δεδομένων:*

Ο ακατέργαστος (crude) μέσος και το τυπικό σφάλμα της πρόσληψης σε γραμμάρια ανά ημέρα, υπολογίστηκε και παρουσιάστηκε για κάθε κέντρο. Όλη η ανάλυση έγινε χωριστά για άνδρες και γυναίκες. Για να βελτιώσουμε τη συγκρισιμότητα μεταξύ των κέντρων, εκτιμήθηκαν μέσοι με τυπικά σφάλματα προσαρμοσμένα ως προς την ηλικία, την εποχή και την ημέρα της εβδομάδας. Αν και οι διαδικασίες δειγματοληψίας σχεδιάστηκαν για να ελέγχουν για εποχική και ανά ημέρα μεταβλητότητα, η κατανομή των μεταβλητών δεν ήταν ομοιογενής μεταξύ των κέντρων. Ομοίως το δείγμα συγκροτείται από άτομα μέσης ηλικίας αλλά η μέση τιμή της ηλικίας ποικίλλει μεταξύ των ομάδων. Κάτω από μία ιδανική δειγματοληψία, όλα τα κέντρα πρέπει να έχουν το 25% των ατόμων συνεντευξιαζόμενων σε κάθε εποχή, το 28.6% (2/7) να ανταποκρίνεται στο Σαβ/κο και το 71.4% στις υπόλοιπες ημέρες. Αφού όμως η δειγματοληψία δεν είναι τέλεια, τα άτομα προσδιορίστηκαν με βάρη για να υπολογίζουν υπέρ- ή υπό- δειγματοληψία. Βάρη υπολογίστηκαν για όλα τα άτομα, για κάθε συνδυασμό ημέρας της εβδομάδας και εποχής, σαν το λόγο μεταξύ της αναμενόμενης συχνότητας, κάτω από ιδανικές συνθήκες και της πραγματικής συχνότητας. Για επιπλέον προσαρμογή ως προς την ηλικία, χρησιμοποιήθηκε η ανάλυση διακύμανσης: ένα σταθμικό μοντέλο παλινδρόμησης προσαρμόστηκε χρησιμοποιώντας τη μεταβλητή που μας ενδιαφέρει σαν εξαρτημένη και τις μεταβλητές κέντρα και ηλικία σαν ανεξάρτητες συμμεταβλητές (covariates) με βάρη όπως ορίστηκαν παραπάνω. Οι σταθμικοί μέσοι (adjusted means) που αναφέρθηκαν, αντιστοιχούν στη μέση κατανάλωση του πληθυσμού με μία ισορροπημένη κατανομή των ατόμων που συνεντευξιάστηκαν για κάθε εποχή και για κάθε ημέρα, με μέση ηλικία 56.8 χρόνων για τους άνδρες και 55.3 για τις γυναίκες. Επίσης για να λάβουμε υπόψη την επίδραση των ανθρωπομετρικών παραγόντων και παραγόντων που αφορούν στον τρόπο ζωής, κάθε μεταβλητή που εκπροσωπεί τέτοιους παράγοντες (BMI, μόρφωση, φυσική δραστηριότητα, κάπνισμα) προστέθηκε στο προηγούμενο μοντέλο. Με την είσοδο κάθε νέας μεταβλητής κοιτάξαμε τις διαφορές στους μέσους κατανάλωσης λαχανικών και φρούτων μετά την προσαρμογή για κάθε παράγοντα.

*Αποτελέσματα:*

Τα κέντρα από τις νότιες χώρες είχαν την υψηλότερη κατανάλωση φρούτων και λαχανικών ενώ η χαμηλότερη πρόσληψη έγινε στην Ολλανδία και τη Σκανδιναβία και για τα δύο φύλλα. Αυτές οι διαφορές ήταν πιο ορατές για τα φρούτα, ιδιαίτερα τα εσπεριδοειδή. Όμως μικρές διαφορές ανέκυσαν για κάποιες υποομάδες των λαχανικών, όπως ρίζες λαχανικών και λάχανα. Οι ανθρωπομετρικές μετρήσεις και οι παράγοντες που αφορούν στον τρόπο ζωής δεν τροποποίησαν ουσιωδώς την πρόσληψη φρούτων και λαχανικών.

**(4) “Food sources of carbohydrates in a European cohort of adults”<sup>7</sup>**, Wirfalt E, McTaggart A, Pala V et.al.

*Επιλογή πληθυσμού:*

Έχει αναφερθεί παραπάνω, στη μελέτη (3).

*Ανάλυση δεδομένων:*

Η ανάλυση διεξάγεται χωριστά για τους άνδρες και τις γυναίκες. Το πακέτο SPSS χρησιμοποιείται για όλες τις στατιστικές αναλύσεις. Οι συμμετέχοντες αναφέρονται από μία ομάδα φαγητών και για κάθε κέντρο περιγράφεται ο ακατέργαστος μέσος. Επειδή από προηγούμενες μελέτες βρέθηκε ότι η κατανάλωση των συμμετεχόντων για κάθε εποχή και ημέρα της εβδομάδας δεν ήταν ίση για κάθε κέντρο, χρησιμοποιήθηκε μια διαδικασία τυποποίησης για να ελέγξει τις εποχικές και ανά ημέρα διαφορές στα κέντρα στην πολυμεταβλητή ανάλυση. Έτσι υπολογίστηκε ο μέσος, που αφορά στην πρόσληψη κάποιων ομάδων φαγητού, και είναι προσαρμοσμένος ως προς την ηλικία για κάθε κέντρο με συγκεκριμένα βάρη για κάθε εποχή και ημέρα της εβδομάδας, χρησιμοποιώντας γενικευμένα γραμμικά μοντέλα (GLM).

*Αποτελέσματα:*

Οι 10 ομάδες φαγητών που συνεισφέρουν περισσότερο υδρογονάνθρακες είναι: ψωμί, φρούτα, γάλα και προϊόντα γάλακτος, τσουρέκια/ κέικ και γλυκά-πίτες, πατάτες, ζάχαρη και μαρμελάδα, ρύζι και μακαρόνια, λαχανικά και όσπρια, τραγανό ψωμί (crispbread) και χυμοί φρούτων και λαχανικών. Η κατανάλωση τόσο φρούτων όσο λαχανικών και όσπριων ήταν υψηλότερη στα νότια κέντρα συγκρινόμενη με τα βόρεια, ενώ η κατανάλωση ελαφριών αναψυκτικών χωρίς αλκοόλ (soft drinks) ήταν υψηλότερη στα βόρεια. Τα κέντρα της Ιταλίας είχαν υψηλή κατανάλωση ρυζιού και μακαρονιών αλλά στα βόρεια κέντρα ήταν πιο υψηλή η κατανάλωση δημητριακών στο πρωινό, πατάτες, τσουρέκια/ κέικ και γλυκά-πίτες. Στη

Σουηδία η χαμηλή κατανάλωση ψωμιού ισορροπήθηκε με μία υψηλή κατανάλωση τραγανού ψωμιού και με τσουρέκια/ κέικ και γλυκά-πίτες. Γενικά οι άνδρες κατανάλωσαν υψηλότερες ποσότητες λαχανικών και όσπριων, ψωμιού, αναψυκτικών χωρίς αλκοόλ (soft drinks), πατάτες, μακαρόνια και ρύζι, δημητριακά στο πρωινό και ζάχαρη και μαρμελάδα από ότι οι γυναίκες, αλλά η κατανάλωση φρούτων ήταν πιο συχνή για τις γυναίκες. Τα φαγητά που είναι πλούσια σε υδρογονάνθρακες και επιλέγονται από τη βόρεια Ευρώπη είναι διαφορετικά από την περιοχή της Μεσογείου.

(Υπογραμμίζουμε ότι: Μέση Ευρώπη ⇒ Γαλλία, Γερμανία, Ολλανδία, Ηνωμένο Βασίλειο

Βόρεια Ευρώπη ⇒ Δανία, Σουηδία, Νορβηγία

Μεσόγειος ⇒ Ελλάδα, Νότια Ισπανία, Νότια Ιταλία )

**(5) “Soy product consumption in 10 European countries: the European Prospective Investigation into Cancer and Nutrition (EPIC) study ”<sup>8</sup>**, Keinan-Boker L, Peeters PHM, Mulligan AA et.al.

*Επιλογή πληθυσμού:*

Έχει αναφερθεί παραπάνω, στη μελέτη (3).

*Ανάλυση δεδομένων:*

Όλα τα αποτελέσματα παρουσιάζονται χωριστά για άνδρες και γυναίκες. Εδώ πρέπει να τονίσουμε και να αναφέρουμε την κατανομή των φαγητών σόγιας σε 7 κατηγορίες: (1) φασόλια και βλαστάρια (sprouts), (2) γαλακτοκομικά προϊόντα (π.χ. γάλα, τυρί, γιαούρτι, παγωτό), (3) δημητριακά προϊόντα (π.χ. ψωμί, μακαρόνια), (4) υποκατάστατα κρέατος (π.χ. σόγια σνίτσελ, χάμπουργκερ), (5) ζυμαρικά, (6) παραδοσιακά φαγητά και (7) μη συγκεκριμένα φαγητά (όταν οι μετέχοντες δεν μπορούν να κάνουν συγκεκριμένο τον τύπο του φαγητού σόγιας που καταναλώθηκε). Η κατανομή των προϊόντων σόγιας περιγράφηκε ανά χώρα. Για να περιγράψουμε τα επίπεδα της κατανάλωσης σόγιας για τις χώρες, εξετάσαμε τις καμπύλες κατανομής και υπολογίστηκαν οι δειγματικοί μέσοι και τα τυπικά τους σφάλματα. Υπολογίστηκαν οι ακατέργαστοι και οι σταθμικοί μέσοι (g/day) της κατανάλωσης για κάθε υποομάδα σόγιας. Χάρη στο χαμηλό αριθμό αναφορών σε πρόσληψη σόγιας, οι ακατέργαστοι και οι σταθμικοί μέσοι της κατανάλωσης παρουσιάστηκαν ανά χώρα και όχι ανά κέντρο, εκτός από το Ηνωμένο Βασίλειο. Οι ακατέργαστοι μέσοι κατανάλωσης προσαρμόστηκαν ως προς την ηλικία χρησιμοποιώντας μοντέλα γραμμικής παλινδρόμησης.

Σταθμικοί μέσοι, ακολούθως, σταθμίστηκαν σύμφωνα με την εποχή και την ημέρα της εβδομάδας, για να εξουδετερωθούν οι εποχικές και ανά ημέρα διακυμάνσεις. Επίσης, περιγράφηκε και η πρόσληψη προϊόντων σόγιας μεταξύ των υποομάδων των συμμετεχόντων με συνήθη υγιεινό τρόπο ζωής (habitual health-conscious lifestyle, HHL), που υποτίθεται ότι καταναλώνουν μεγάλη ποσότητα προϊόντων σόγιας. Περιγράφηκαν δημογραφικά και ανθρωπομετρικά χαρακτηριστικά και διατροφικοί παράγοντες για συμμετέχοντες με HHL και για τους υπόλοιπους συμμετέχοντες.

#### *Αποτελέσματα:*

Αν και η πρόσληψη σόγιας ήταν σχετικά μικρή σε όλες τις χώρες, η υψηλότερη πρόσληψη παρατηρήθηκε στο Ηνωμένο Βασίλειο, χάρη στο μεγάλο αριθμό ατόμων με συνήθη υγιεινό τρόπο ζωής (habitual health-conscious lifestyle, HHL) π.χ. αυτοί που δεν τρώνε κρέας αλλά τρώνε ψάρι, χορτοφάγοι. Η πιο συχνή κατανάλωση προϊόντων σόγιας ήταν τα γαλακτοκομικά προϊόντα στο Ηνωμένο Βασίλειο και τη Γαλλία και τα φασόλια και βλαστοί μεταξύ των χωρών της μέσης Ευρώπης. Και για τα δύο φύλλα η υποομάδα σόγιας που αφορά στα γαλακτοκομικά προϊόντα καταναλώθηκε περισσότερο (1.2g/day για τους άνδρες και 1.9g/day για τις γυναίκες). Συμμετέχοντες με συνήθη υγιεινό τρόπο ζωής, HHL διέφεραν σημαντικά από τους άλλους σε σχέση με δημογραφικούς, ανθρωπομετρικούς και διατροφικούς παράγοντες αφού καταναλώθηκαν μεγαλύτερες ποσότητες από όλα τα προϊόντα σόγιας και τις υποομάδες αυτών.

**(6) “Consumption of added fats and oils in the European Prospective Investigation into Cancer and Nutrition (EPIC) centers across 10 European countries as assessed by 24-hour dietary recalls”<sup>9</sup>**, Linseisen J, Bergstrom E, Gafa L, Gonzalez CA et al.

#### *Επιλογή πληθυσμού:*

Έχει αναφερθεί παραπάνω.

#### *Ανάλυση δεδομένων:*

Ακατέργαστες τιμές πρόσληψης δίνονται σαν αριθμητικός μέσος και σταθμικές τιμές σαν μέσος και τυπικό σφάλμα του μέσου. Στάθμιση μέσα στα κέντρα παρουσιάστηκε για να βελτιώσει για αποκλίσεις από μια “ιδανική” δειγματοληψία της 24ωρου ανάκλησης 24-ώρου (ημέρα της εβδομάδας, εποχικότητα) όπως επίσης και την ηλικία. Για την ημέρα της εβδομάδας δύο ξεχωριστά επίπεδα (Δευτ-Παρ και Σαβ-Κυρ) και για τις εποχές τέσσερα

ξεχωριστά επίπεδα επιλέχθηκαν για τη στάθμιση. Η ηλικία περιλήφθηκε σαν συνεχής μεταβλητή (μοντέλο παλινδρόμησης). Η στάθμιση έγινε χωριστά για άνδρες και γυναίκες. Για να μελετήσουμε τις διαφορές στη συνολική πρόσληψη ενέργειας μεταξύ των ατόμων, υπολογίστηκε για κάθε άτομο το ποσοστό της συνολικής ημερήσιας πρόσληψης ενέργειας, υπό την προϋπόθεση ότι τα λιπίδια προέρχονται από την κατανάλωση πρόσθετων λιπαρών και λαδιού, ακολουθώντας τη ίδια διαδικασία στάθμισης που περιγράφηκε παραπάνω. Παράγοντες που επηρεάζουν σημαντικά την κατανάλωση πρόσθετων λιπαρών και λαδιού, όπως επίσης η πρόσληψη λιπιδίων που προέρχονται από την κατανάλωση πρόσθετων λιπαρών και λαδιού, προσδιορίστηκαν από τους μέσους της ανάλυσης διακύμανσης. Εξετάστηκαν οι παράγοντες κέντρο, ημέρα της εβδομάδας και εποχή της καθορισμένης ανάκλησης, φύλλο, ηλικία, BMI, κάπνισμα, μόρφωση, φυσική δραστηριότητα στη δουλειά και στα σπορ.

#### *Αποτελέσματα:*

Η μέση ημερήσια πρόσληψη πρόσθετων λιπαρών και λαδιού ποικίλλει μεταξύ 16.2g (Varese, Ιταλία) και 41.1g (Malmo, Σουηδία) στις γυναίκες και μεταξύ 24.7g (Ragusa, Ιταλία) και 66g (Potsdam, Γερμανία) στους άνδρες. Η συνολική μέση πρόσληψη λιπιδίων από κατανάλωση πρόσθετων λιπαρών και λαδιού, συμπεριλαμβανομένων αυτών που χρησιμοποιούνται για προετοιμασία σάλτσας, εκτείνονται μεταξύ 18.3 (Νορβηγία) και 37.2g/day (Ελλάδα) στις γυναίκες και 28.4 (Heidelberg, Γερμανία) και 51.2g/day (Ελλάδα) στους άνδρες. Τα Μεσογειακά κέντρα του ΕΠΙΚ με υψηλή κατανάλωση ελαιόλαδου, συνδυάστηκαν με χαμηλή πρόσληψη ζωικών λιπών, σε αντίθεση με τα κέντρα της κεντρικής και βόρειας Ευρώπης όπου καταναλώθηκε λιγότερο φυτικό λάδι, περισσότερα ζωικά λίπη και ένα μεγάλο ποσοστό μαργαρίνης. Η κατανάλωση πρόσθετων λιπαρών και λαδιού που προέρχονται από ζώα ήταν υψηλότερη στη Γερμανία και ακολούθως στη Γαλλία.

**(7) “Meat consumption in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohorts: results from 24-hour dietary recalls ”<sup>10</sup>, Linseisen J, Kesse E, Slimani N et.al.**

#### *Επιλογή πληθυσμού:*

Έχει αναφερθεί παραπάνω.

#### *Ανάλυση δεδομένων:*

Αρχικά πρέπει να υπογραμμιστεί ο καθορισμός των υποομάδων κρέατος: (1) φρέσκο/νωπό κρέας: (i) κόκκινο κρέας (χοιρινό, βοδινό, μοσχαρίσιο, αρνίσιο), (ii) πουλερικά (κοτόπουλο, γαλοπούλα), (iii) άλλο (π.χ. κυνήγι, λαγός, άλογο) και (2) επεξεργασμένο κρέας (π.χ. λουκάνικα, μπέικον, χοιρομέρι)

Ακατέργαστες τιμές πρόσληψης δίνονται σαν αριθμητικοί μέσοι και σταθμισμένες τιμές σαν μέσος και τυπικό σφάλμα του μέσου. Όλα τα συμπεράσματα που δίνονται βασίστηκαν μόνο στις σταθμισμένες τιμές. Στάθμιση μέσα στα κέντρα παρουσιάστηκε για να βελτιώσει για αποκλίσεις από μια “ιδανική” δειγματοληψία της ανάκλησης 24-ώρου (ημέρα της εβδομάδας, εποχικότητα) όπως επίσης και την ηλικία. Επιλέχθηκαν για τη στάθμιση δύο ξεχωριστά επίπεδα για την ημέρα της εβδομάδας (Δευτ-Παρ και Σαβ-Κυρ) και τέσσερα ξεχωριστά επίπεδα για τις εποχές. Η ηλικία περιλήφθηκε σαν συνεχής μεταβλητή. Επιπλέον, η παρουσίαση των αποτελεσμάτων στρωματοποιήθηκε ως προς το φύλλο και το κέντρο. Παράγοντες που επηρεάζουν σημαντικά την κατανάλωση κρέατος, προσδιορίστηκαν από τους μέσους της ανάλυσης διακύμανσης. Ο έλεγχος της στατιστικής σημαντικότητας μεταξύ δύο ομάδων παρουσιάστηκε με το τεστ διαφορών ελαχίστων τετραγώνων (least square differences test, LSD) σε επίπεδο  $\alpha=5\%$ . Εξετάστηκαν οι παράγοντες κέντρο, ημέρα της εβδομάδας και εποχή της καθορισμένης ανάκλησης, φύλλο, ηλικία, BMI, κάπνισμα, μόρφωση, φυσική δραστηριότητα στη δουλειά και στα σπορ.

#### *Αποτελέσματα:*

Η μέση συνολική πρόσληψη κρέατος ήταν χαμηλότερη στα άτομα με υγιεινό τρόπο ζωής (health-conscious) στο Ηνωμένο Βασίλειο (15 και 21g/day στις γυναίκες και στους άνδρες αντίστοιχα) και υψηλότερη στη βόρεια Ισπανία, ειδικά στο San Sebastian (124 και 234g/day αντίστοιχα). Στα κέντρα της νότιας Ισπανίας και στη Νάπολη (Ιταλία), η κατανάλωση κρέατος ήταν ευκρινώς χαμηλότερη από ότι στα βόρεια αυτών των χωρών. Τα κέντρα/χώρες της κεντρικής και βόρειας Ευρώπης έδειξαν μία πανομοιότυπη κατανάλωση κρέατος εκτός από τη Βρετανία και τη Γαλλία. Οι διαφορές στην πρόσληψη των υποομάδων κρέατος ήταν ακόμη υψηλότερες από ότι βρέθηκε για τη συνολική πρόσληψη κρέατος. Με λίγες εξαιρέσεις, τα μεσογειακά κέντρα του ΕΠΙΚ έδειξαν ένα υψηλότερο ποσοστό βοδινού, μοσχαρίσιου και πουλερικών και λιγότερο χοιρινό και επεξεργασμένο κρέας από ότι παρατηρήθηκε στα κέντρα της κεντρικής ή βόρειας Ευρώπης. Η υψηλότερη κατανάλωση



λουκάνικων παρατηρήθηκε στη Γερμανία και ακολούθησαν η Νορβηγία, Σουηδία, Δανία και Ολλανδία.

**(8) “Consumption of dairy products in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort: data from 35955 24-hour dietary recalls in 10 European countries ”<sup>11</sup>, Hjartaker A, Lagiou A, Slimani N et.al.**

*Επιλογή πληθυσμού:*

Έχει αναφερθεί παραπάνω.

*Ανάλυση δεδομένων:*

Αρχικά πρέπει να υπογραμμιστεί ο καθορισμός των υποομάδων γαλακτοκομικών προϊόντων: γάλα, γιαούρτι, τυρί, κρεμώδη επιδόρπια και πουτίγκες βασισμένες σε γάλα, βούτυρο και παγωτό.

Η στατιστική ανάλυση στρωματοποιήθηκε ως προς το φύλλο. Ακατέργαστος και σταθμισμένος μέσος της πρόσληψης γαλακτοκομικών υπολογίστηκε για κάθε κέντρο. Σταθμισμένοι μέσοι υπολογίστηκαν από τους μέσους της ανάλυσης διακύμανσης, χρησιμοποιώντας ένα σταθμισμένο μοντέλο παλινδρόμησης με βάρη στα ατομικά επίπεδα. Βασιζόμενοι στα αποτελέσματα προκαταρκτικών αναλύσεων αποφασίστηκε να συμπεριληφθούν οι ακόλουθες μεταβλητές στα πολλαπλά μοντέλα παλινδρόμησης: ηλικία (συνεχής), ημέρα της εβδομάδας (δύο κατηγορίες: Δευτ-Παρ και Σαβ-Κυρ) και εποχή (4 κατηγορίες). Επίσης παρουσιάστηκαν τα σταθμισμένα μοντέλα παλινδρόμησης, για την συνολική πρόσληψη ενέργειας.

*Αποτελέσματα:*

Μία υψηλή κατανάλωση γαλακτοκομικών προϊόντων παρατηρήθηκε στα περισσότερα κέντρα της Ισπανίας και στο Ηνωμένο Βασίλειο, όπως και στα κέντρα της Ολλανδίας, Σουηδίας και Δανίας. Μία κάπως χαμηλή κατανάλωση αναφέρθηκε στην Ελλάδα και σε κάποια από τα κέντρα της Ιταλίας (Ragusa και Turin). Σε όλα τα κέντρα και για τα δύο φύλλα, το γάλα αποτελεί τη γαλακτοκομική υποομάδα με το μεγαλύτερο ποσοστό κατανάλωσης και ακολουθεί το γιαούρτι και το τυρί. Ακόμα, υπήρξε ένα μεγάλο εύρος στην κατανάλωση διαφόρων γαλακτοκομικών υποομάδων μεταξύ των κέντρων. Τα ισπανικά και νορβηγικά κέντρα γενικά παρουσίασαν υψηλή κατανάλωση γάλατος, τα σουηδικά και τα ολλανδικά παρουσίασαν υψηλή κατανάλωση γιαουρτιού ενώ η υψηλότερη κατανάλωση τυριού αναφέρθηκε στα κέντρα της Γαλλίας. Τα αποτελέσματα αποδεικνύουν ποσοτικές και

ποιοτικές ανομοιοότητες στα γαλακτοκομικά προϊόντα μεταξύ των κέντρων του ΕΠΙΚ. Αυτό προσφέρει ένα δυνατό σημείο για αναλύσεις των σχέσεων μεταξύ κατανάλωσης γαλακτοκομικών και ασθενειών, όπως καρκίνος.

**(9) “Variability of fish consumption within the 10 European countries participating in the European Investigation into Cancer and Nutrition (EPIC) study”<sup>12</sup>, Welch AA, Lund E, Amiano P et.al.**

*Επιλογή πληθυσμού:*

Έχει αναφερθεί παραπάνω.

*Ανάλυση δεδομένων:*

Αρχικά πρέπει να υπογραμμιστεί ο καθορισμός των υποομάδων ψαριού: (1) ψάρι⇒(i) λευκό ψάρι (<4g/100g) και (ii) λιπαρό ψάρι (≥4g/100g), (2) προϊόντα ψαριού⇒(i) προϊόντα λευκού ψαριού, (ii) προϊόντα λιπαρού ψαριού, (iii) προϊόντα από αυγά ψαριού και (3) οστρακοειδή και σαλιγκάρια.

Ακατέργαστοι μέσοι, σταθμισμένοι μέσοι και τυπικά σφάλματα υπολογίστηκαν για τις υποομάδες ψαριών και στρωματοποιήθηκαν ως προς το φύλλο και το κέντρο. Ο σχεδιασμός της μελέτης διήρκεσε πάνω από ένα χρόνο, έτσι κάθε εποχή εκπροσωπείται από το 25% των συμμετεχόντων και κάθε ημέρα της εβδομάδας από το 14.2% των συμμετεχόντων. Όμως πρακτικά αυτό δεν ήταν εντελώς κατορθωτό και για να υπολογίσουμε αυτές τις διακυμάνσεις στη διαδικασία δειγματοληψίας, υπολογίστηκαν σταθμισμένοι μέσοι και τυπικά σφάλματα από την ανάλυση διακύμανσης χρησιμοποιώντας σταθμικούς παράγοντες για την εποχή και την ημέρα της εβδομάδας. Τα βάρη υπολογίστηκαν σαν το λόγο μεταξύ της αναμενόμενης συχνότητας κάτω από ιδανικές συνθήκες και της πραγματικής

συχνότητας. Η ηλικία χρησιμοποιήθηκε σαν συνεχής μεταβλητή. Μία επιπλέον ανάλυση καθόρισε την επίδραση της πρόσληψης ενέργειας, η οποία εισήχθη στο μοντέλο σαν συνεχής μεταβλητή. Η ανά ημέρα μεταβλητότητα στην κατανάλωση ψαριού ερευνήθηκε και μη σταθμισμένοι μέσοι παρουσιάστηκαν για κάθε χώρα. Στο Ηνωμένο Βασίλειο η ομάδα με υγιεινό τρόπο ζωής (health-conscious) εξαιρέθηκε αφού υπήρχαν λιγότερες από 5 καταγραφές για κάποιες ημέρες της εβδομάδας. Υπολογίστηκαν σταθμισμένοι μέσοι χρησιμοποιώντας γενικευμένα γραμμικά μοντέλα (GLM) στο στατιστικό πακέτο SAS και η υπόλοιπη ανάλυση έγινε στο STATA.

#### *Αποτελέσματα:*

Μία εξαπλάσια ή επταπλάσια διακύμανση στην συνολική κατανάλωση ψαριού υπήρξε τόσο στις γυναίκες όσο και στους άνδρες, μεταξύ της χαμηλότερης κατανάλωσης στη Γερμανία και της υψηλότερης στην Ισπανία. Γενικά, το λευκό ψάρι αντιπροσωπεύει το 49% και το 45% της συνολικής πρόσληψης ψαριού στις γυναίκες και στους άνδρες αντίστοιχα, με μεγαλύτερη κατανάλωση στα κέντρα της Ισπανίας και της Ελλάδας και χαμηλότερη στα κέντρα της Γερμανίας και της Ολλανδίας. Η κατανάλωση λιπαρών ψαριών επηρεάστηκε από τα συνολικά ψάρια. Όμως η μεγαλύτερη πρόσληψη πολύ λιπαρών ψαριών ήταν στις παραλιακές περιοχές της βόρειας Ευρώπης (Δανία, Σουηδία και Νορβηγία) και στη Γερμανία. Η κατανάλωση προϊόντων ψαριού ήταν μεγαλύτερη στη βόρεια από ότι στη νότια Ευρώπη, με προϊόντα λευκού ψαριού να επικρατούν στα κέντρα της Γαλλίας, της Ιταλίας, της Ισπανίας, της Ολλανδίας και της Νορβηγίας. Η πρόσληψη προϊόντων από αυγά ψαριού ήταν χαμηλή. Η υψηλότερη κατανάλωση οστρακοειδών βρέθηκε στα κέντρα της Γαλλίας, της Ισπανίας και της Ιταλίας. Η μεγαλύτερη διακύμανση στην κατανάλωση ανά ημέρα της εβδομάδας βρέθηκε στις χώρες με χαμηλότερη πρόσληψη ψαριού.

**(10) “Patterns of alcohol consumption in 10 European countries participating in the European Prospective Investigation into Cancer and Nutrition (EPIC) project”<sup>13</sup>, Sieri S, Agudo A, Kesse E et.al.**

#### *Επιλογή πληθυσμού:*

Έχει αναφερθεί παραπάνω.

#### *Ανάλυση δεδομένων:*

Αρχικά πρέπει να υπογραμμιστεί ο καθορισμός των υποομάδων των αλκοολούχων ποτών: κρασί, μύρα, οινοπνευματώδη και συνολική πρόσληψη αιθανόλης.

Υπολογίστηκαν ακατέργαστοι μέσοι και σταθμισμένοι μέσοι της κατανάλωσης κάθε κατηγορίας ποτού και της συνολικής αιθανόλης ανά κέντρο, χρησιμοποιώντας μία αποκαλούμενη ανάμεικτη διαδικασία, στην οποία η ανάλυση διακύμανσης έλαβε υπόψη την πιθανή συγχετική επίδραση της ηλικίας και μία σταθμισμένη διαδικασία λαμβάνοντας υπόψη την ανομοιογενή κατανομή των συνεντευξιαζόμενων, μεταξύ των κέντρων, ως προς την ημέρα της εβδομάδας και την εποχή. Όλες οι αναλύσεις έγιναν στο στατιστικό πακέτο STATA, χρησιμοποιώντας αναλυτικά βάρη που υπολογίστηκαν σαν το λόγο μεταξύ του αναμενόμενου και του παρατηρούμενου αριθμού ατόμων σε κάθε κέντρο και του φύλλου σε

σχέση με την κατανομή της εποχής και της ημέρας. Η ηλικία συμπεριλήφθηκε σαν συνεχής μεταβλητή στην ανάλυση διακύμανσης ενώ η ανάλυση παρουσιάστηκε χωριστά για άνδρες και γυναίκες.

*Αποτελέσματα:*

Το φύλλο ήταν ένας καθοριστικός παράγοντας, των τύπων ποτού σε όλες τις χώρες. Η υψηλότερη συνολική κατανάλωση αλκοόλ παρατηρήθηκε στα κέντρα της Ισπανίας (San Sebastian, 41.4g/day) για τους άνδρες και της Δανίας (Copenhagen, 20.9g/day) για τις γυναίκες. Η χαμηλότερη συνολική πρόσληψη αλκοόλ παρατηρήθηκε στα κέντρα της Σουηδίας (Umea, 10.2g/day) για τους άνδρες και στην Ελλάδα (3.4g/day) για τις γυναίκες. Μεταξύ των ανδρών, την κύρια συμβολή στην συνολική πρόσληψη αλκοόλ την είχε το κρασί, για τις Μεσογειακές χώρες και η μπύρα για τα κέντρα της Ολλανδίας, Γερμανίας, Σουηδίας και Δανίας. Στα περισσότερα κέντρα, η κύρια πηγή αλκοόλ για τις γυναίκες ήταν το κρασί εκτός από την Murcia (Ισπανία) που ήταν η μπύρα. Η κατανάλωση αλκοόλ, ιδιαίτερα στις γυναίκες, αυξανόταν αξιοσημείωτα κατά τη διάρκεια του Σαβ/κου σε σχεδόν όλα τα κέντρα. Τα κέντρα της Γερμανίας, Ολλανδίας, Ηνωμένου Βασιλείου (γενικός πληθυσμός) και Δανίας χαρακτηρίστηκαν από το υψηλότερο ποσοστό κατανάλωσης αλκοόλ έξω από την ώρα του γεύματος. Η μεγάλη διακύμανση στους τύπους ποτών μεταξύ των κέντρων του ΕΠΙΚ, δημιουργεί μια ευκαιρία για καλύτερη κατανόηση της σχέσης μεταξύ αλκοόλ και ασθενειών που σχετίζονται με το αλκοόλ.

**(11) “Trends in self-reported past alcoholic beverage consumption and ethanol intake from 1950 to 1995 observed in eight European countries participating in the European Investigation into Cancer and Nutrition (EPIC)”<sup>14</sup>, Klipstein-Grobusch K, Slimani N, Krogh V et.al.**

*Επιλογή πληθυσμού:*

Έχει αναφερθεί παραπάνω.

*Ανάλυση δεδομένων:*

Η ημερομηνία γέννησης των συμμετεχόντων χρησιμοποιήθηκε για να υπολογιστεί ο ημερολογιακός χρόνος όταν οι συμμετέχοντες ήταν 20, 30 ή 40 χρονών και να κατασκευαστούν αντίστοιχες περιόδους: πληροφορίες για την κατανάλωση αλκοολούχων ποτών στην ηλικία των 20 αντιστοιχεί στην χρονική περίοδο 1950-1975, στην ηλικία των 30 αντιστοιχεί στην χρονική περίοδο 1960-1985 και πληροφορίες στην ηλικία των 40

αντιστοιχεί στην χρονική περίοδο 1970-1995. Λόγω του μεγάλου εύρους της ηλικίας (35-75 χρόνια) των συμμετεχόντων της μελέτης και λόγω της διαθεσιμότητας των πληροφοριών που αφορούν στην κατανάλωση αλκοολούχων ποτών στις συγκεκριμένες ηλικίες (20, 30 και 40 χρόνων), είμαστε ικανοί να υπολογίσουμε τη μέση κατανάλωση αλκοολούχων ποτών (glass/week) και την πρόσληψη αιθανόλης (g/day) για τις χρονικές περιόδους 1950-1975 (στην ηλικία των 20), 1960-1985 (στην ηλικία των 30) και 1970-1995 (στην ηλικία των 40) στα μελετούμενα κέντρα. Καθορίστηκε το ποσοστό αυτών που απέχουν από το αλκοόλ. Η συμβολή των αλκοολούχων ποτών στην πρόσληψη αιθανόλης υπολογίστηκε όπως περιγράφηκε παραπάνω. Μία ανάλυση γραμμικής παλινδρόμησης διεξήχθη για να καθοριστεί η γενική τάση της παρελθούσας κατανάλωσης αλκοολούχων ποτών και πρόσληψης αιθανόλης με χρόνους 1950-1975, 1960-1985 και 1970-1995 για τα κέντρα της μελέτης. Στο μοντέλο παλινδρόμησης, ο αριθμός ποτηριών αλκοόλ ανά εβδομάδα και αντίστοιχα γραμμάρια αιθανόλης ανά ημέρα ήταν η εξαρτημένη μεταβλητή και ο ημερολογιακός χρόνος της κατανάλωσης ήταν η ανεξάρτητη μεταβλητή. Όλες οι αναλύσεις παρουσιάστηκαν χωριστά για άνδρες και γυναίκες ενώ έγιναν στο στατιστικό πακέτο SAS.

#### *Αποτελέσματα:*

Γενικά, αυξήσεις στην κατανάλωση μπίρας παρατηρήθηκαν στα περισσότερα κέντρα του ΕΠΙΚ για το 1950-1975, 1960-1985 και 1970-1995. Η γενική τάση στην κατανάλωση κρασιού διέφερε σύμφωνα με τη γεωγραφική τοποθεσία: για τους άνδρες των κέντρων της νότιας Ευρώπης παρατηρήθηκε μία τάση προς τα κάτω με το χρόνο ενώ παρατηρήθηκε μία τάση προς τα πάνω στα κέντρα της κεντρικής/βόρειας Ευρώπης. Για τις γυναίκες, ίσχυαν τα ίδια με τους άνδρες άλλα παρατηρήθηκε μικρότερη τάση. Λόγω του ότι η κατανάλωση κρασιού ήταν ο κύριος παράγοντας συνεισφοράς στην πρόσληψη αιθανόλης τόσο για τους άνδρες όσο και για τις γυναίκες στα περισσότερα κέντρα, η γενική τάση της πρόσληψης αιθανόλης με το χρόνο ήταν πανομοιότυπη με την κατανάλωση κρασιού.

### **(12) “Diversity of dietary patterns observed in the European Prospective Investigation into Cancer and Nutrition (EPIC) project”<sup>15</sup>, Slimani N, Fahey M, Welch AA et.al.**

#### *Επιλογή πληθυσμού:*

Έχει αναφερθεί παραπάνω.

#### *Ανάλυση δεδομένων:*

Στην ανάλυση λήφθηκαν υπόψη όλες οι κύριες ομάδες του ΕΠΚ που κατηγοριοποιούνται καλύπτοντας τα διαφορετικά συστατικά της διατροφής όλων των χωρών εκτός από ανάμεικτα φαγητά (miscellaneous) και σούπες. Τελικά, 22 κύριες υποομάδες θεωρήθηκαν για την ανάλυσή μας. Αυτές οι ομάδες φαγητών ρυθμίστηκαν σύμφωνα με τη φυτική ή ζωική τους προέλευση και το βαθμό επεξεργασίας του φαγητού ώστε να διευκολυνθεί η σύγκριση διατροφικών προτύπων στα κέντρα. Για τη σύγκριση των αλκοολούχων ποτών, που ποικίλουν στον τύπο και την αναλογία αλκοόλ στα κέντρα μελέτης του ΕΠΚ, εκφράσαμε την κατανάλωση αλκοόλ σε γραμμάρια.

Η ανάλυσή μας επικεντρώθηκε στη σύγκριση της μέσης διατροφικής πρόσληψης. Ένα συγκεκριμένο μοντέλο πολλαπλής γραμμικής παλινδρόμησης χρησιμοποιήθηκε για να σταθμιστεί ως προς την ηλικία, την ημέρα της εβδομάδας και την εποχή. Η συνολική πρόσληψη ενέργειας προστέθηκε στο μοντέλο αλλά, λόγω του ότι δεν άλλαξαν τα ολικά διατροφικά πρότυπα στα περισσότερα κέντρα, αυτά τα αποτελέσματα δεν αναφέρθηκαν εδώ. Η ολική μέση κατανάλωση, υπολογίστηκε για τις ίδιες ομάδες φαγητών ως προς το φύλλο, σαν αριθμητικός μέσος. Για να εκφράσουμε τη διακύμανση της μέσης πρόσληψης κάθε κέντρου από το γενικό μέσο του ΕΠΚ, υπολογίστηκε για κάθε ομάδα φαγητού το ποσοστό πρόσληψης φαγητού αναφορικά με το μέσο του ΕΠΚ, ως προς το φύλλο και το κέντρο. Μία πολυδιάστατη γραφική παρουσίαση της πρόσληψης φαγητού χρησιμοποιήθηκε για να δείξει εμφανώς τις διαφορές στα διατροφικά πρότυπα ανά κέντρο και φύλλο.

#### *Αποτελέσματα:*

Αν και παρατηρήθηκαν ευρείες διαφορές στα κέντρα, οι χώρες που συμμετέχουν στο ΕΠΚ χαρακτηρίστηκαν από συγκεκριμένα διατροφικά πρότυπα. Γενικά, η Ιταλία και η Ελλάδα έχουν διατροφικά πρότυπα που χαρακτηρίζονται από κατανάλωση φυτικών φαγητών (plant foods) (εκτός από πατάτες) και μία χαμηλότερη κατανάλωση ζωικών και επεξεργασμένων φαγητών, συγκρινόμενα με τις άλλες χώρες του ΕΠΚ. Η Γαλλία και ιδιαίτερα η Ισπανία έχει μεγαλύτερη ετερογένεια διατροφικών προτύπων, με μία σχετικά υψηλή κατανάλωση τόσο φυτικών όσο και ζωικών προϊόντων. Εκτός από συγκεκριμένα χαρακτηριστικά των ομάδων χορτοφάγων, η ομάδα των ατόμων με υγιεινό τρόπο ζωής (health-conscious) στο Ηνωμένο Βασίλειο μοιράζεται με το γενικό πληθυσμό του Ηνωμένου Βασιλείου μία σχετικά υψηλή κατανάλωση τσαγιού, σάλτσας, κέικ, μη αλκοολούχων ποτών (γυναίκες), μαργαρίνης και βουτύρου. Σε αντίθεση, η διατροφή στις σκανδιναβικές χώρες, στην Ολλανδία, τη Γερμανία

και το Ηνωμένο Βασίλειο (στο γενικό πληθυσμό) ήταν σχετικά υψηλή στις πατάτες και στα ζωικά, επεξεργασμένα και γλυκά φαγητά, με ποσοστά να ποικίλουν μεταξύ των χωρών και των κέντρων. Σε αυτές τις χώρες, η κατανάλωση λαχανικών και φρούτων ήταν παρόμοια ή παρακάτω από τους γενικούς μέσους του ΕΠΙΚ και ήταν χαμηλότερη για τα όσπρια και το φυτικό λάδι. Γενικά, τα διατροφικά πρότυπα ήταν παρόμοια για τους άνδρες και τις γυναίκες, αν και υπήρξαν μεγάλες διαφορές, στο φύλλο, για συγκεκριμένες ομάδες φαγητών.

Η δική μας μελέτη, που παρουσιάζεται στα επόμενα κεφάλαια, αφορά στην ανάλυση και επεξεργασία δεδομένων που προέρχονται από το Ευρωπαϊκό Πρόγραμμα Ιατρικής και Κοινωνίας (ΕΠΙΚ), μόνο για τη χώρα της Ελλάδας. Σκοπός μας είναι να ελέγξουμε τη συμβατότητα ενεργειακής πρόσληψης των δεδομένων της 24ωρης διατροφικής ανάκλησης προσδιορίζοντας τους υποεκτιμούντες (under-reporters). Επίσης, αναλύουμε τρόπους εύρεσης της φυσικής δραστηριότητας ενός ατόμου, αφού η φυσική δραστηριότητα είναι σημαντικός παράγοντας στον προσδιορισμό των υποεκτιμούντων. Η μέτρηση της φυσικής δραστηριότητας μπορεί να οδηγήσει στον προσδιορισμό της ενέργειας που συνήθως δαπανά ένα άτομο ανά ημέρα ή ώρα, εκφρασμένη είτε σε μονάδες ενέργειας (kcal ή kJ) είτε σε μεταβολικά ισοδύναμα (metabolic equivalent, MET)<sup>21</sup>.





## ΚΕΦΑΛΑΙΟ 4

### Παραγωγή οριακών τιμών (cut-off limits) για αναγνώριση των υποεκτιμούντων (underreporters)

#### 4.1 Εισαγωγή.

Πολλές διατροφικές μελέτες εξαρτώνται από τους αξιόπιστους εκτιμητές της κατανάλωσης φαγητού σε ελεύθερους (free-living) ανθρώπους και υπεύθυνοι ερευνητές έκαναν μεγάλες προσπάθειες να επικυρώσουν τις διατροφικές τεχνικές τους. Όμως για αρκετά χρόνια δεν υπήρχε κανένα πρότυπο με το οποίο να συγκρίνουμε τις ποικίλες μεθόδους. Έτσι λοιπόν τα τελευταία χρόνια, λόγω της μεγάλης εμπειρίας των ερευνητών από την αξιολόγηση των αποτελεσμάτων μεγάλων διατροφικών ερευνών, δημιουργήθηκε η ανάγκη για μια κριτική αντιμετώπιση των τιμών της ενεργειακής πρόσληψης, οι οποίες προκύπτουν από τις διαφορετικές μεθόδους καταγραφής της διατροφής. Οι περισσότερες από αυτές τις μελέτες βασίζονται στην παραδοχή ότι η διατροφική πρόσληψη που καταγράφεται είναι αντικειμενική, χωρίς όμως να είναι σε θέση να το αποδείξουν με τη χρήση ανεξάρτητων μεθόδων, ως προς τις οποίες συγκρίνουν τα αποτελέσματά τους.

Έτσι λοιπόν για τη μέτρηση της συνολικής ενεργειακής κατανάλωσης έχουμε δύο μεθόδους<sup>18,19</sup>: τη μέθοδο του διπλά επισημασμένου ύδατος (doubly labelled water, DLW) και τη μέθοδο της ολοσωματικής θερμιδομετρίας (whole body calorimeter). Με την χρησιμοποίηση και στη συνέχεια την καθιέρωση της μεθόδου του διπλά επισημασμένου ύδατος, ξεκίνησαν και οι μελέτες για τον έλεγχο της συμβατότητας των τιμών ενεργειακής πρόσληψης, δηλαδή η σύγκριση της ενεργειακής κατανάλωσης (energy expenditure) με την ενεργειακή πρόσληψη (energy intake). Η μέθοδος αυτή έχει ένα μεγάλο πλεονέκτημα το οποίο την καθιστά και πιο εύχρηστη, συγκρινόμενη με άλλες θερμιδομετρικές μεθόδους, όπως με τη μέθοδο της ολοσωματικής θερμιδομετρίας που χρησιμοποιείται κυρίως για πειραματικές έρευνες. Το πλεονέκτημα<sup>18</sup> αυτό είναι ότι μπορεί να χρησιμοποιηθεί για τον υπολογισμό της ενεργειακής κατανάλωσης σε άτομα που ζουν υπό φυσιολογικές συνθήκες. Από τις πρώτες κιάλας μελέτες που χρησιμοποιήθηκαν μετρήσεις κατανάλωσης ενέργειας, που λήφθηκαν με τη μέθοδο του διπλά επισημασμένου ύδατος, για να διασταυρώσουμε την

εγκυρότητα των εκτιμήσεων πρόσληψης φαγητού, αποδείχθηκε σοβαρή υποεκτίμηση σε παχύσαρκους ενήλικες, σε άτομα που έχουν κάνει μια δίαιτα με επιτυχία άλλα και σε ενήλικες και εφήβους με κανονικό βάρος.

Υπήρξαν και μελέτες στις οποίες βρέθηκε, χρησιμοποιώντας πάντοτε την ίδια μέθοδο, ότι τόσο οι άνδρες όσο και οι γυναίκες υποεκτιμούν την πρόσληψη ενέργειας στον ίδιο βαθμό ενώ οι άνδρες έχουν μεγαλύτερη κατανάλωση ενέργειας από τις γυναίκες με τον ίδιο ρυθμό μεταβολισμού (Basal Metabolic Rate, BMR). Όλα αυτά τα συμπεράσματα, τα οποία αποτελούν πτυχές της ενεργειακής φυσιολογίας αλλά και των μεθόδων προσδιορισμού της διατροφικής πρόσληψης, οδήγησαν τους ερευνητές σε πιο εκτεταμένη χρήση της μεθόδου του διπλά επισημασμένου ύδατος για την αξιολόγηση των αποτελεσμάτων των διατροφικών ερευνών<sup>18</sup>. Όμως η συγκεκριμένη μέθοδος είναι πολύ ακριβή και τεχνικά δύσκολη για να χρησιμοποιηθεί ευρέως, με αποτέλεσμα να μην μπορεί να χρησιμοποιηθεί σε μεγάλες επιδημιολογικές έρευνες. Για αυτό λοιπόν το λόγο, έχουμε τον προσδιορισμό και την εφαρμογή οριακών τιμών (cut-off limits)<sup>16</sup>, όπου δείχνουν τα χαμηλότερα επίπεδα των ενεργειακών αναγκών που μπορούν να έχουν άτομα που ζουν κάτω από φυσιολογικές συνθήκες ώστε να διατηρούν σταθερό το βάρος τους. Επομένως, υπό την προϋπόθεση ότι δεν έχουμε απώλεια βάρους, αναφορές ενεργειακής πρόσληψης χαμηλότερες των οριακών αυτών τιμών θεωρούνται ύποπτες για υποεκτίμηση της πρόσληψης.

#### **4.2 Ορισμός των ελαχίστων οριακών τιμών (cut-off limits).**

Οι Goldberg et al<sup>16</sup>, χρησιμοποίησαν θεμελιώδεις αρχές της ενεργειακής φυσιολογίας για να ορίσουν τα ελάχιστα επίπεδα τιμών για την ενεργειακή πρόσληψη, όπου κάτω από αυτά τα όρια ένα άτομο με δοσμένο φύλλο, ηλικία και βάρος, δεν θα μπορούσε να ζήσει μια φυσιολογική ζωή. Αυτές οι μετρήσεις έχουν προέλθει από τη μέθοδο του διπλά επισημασμένου ύδατος και της ολοσωματικής θερμοδομετρίας σε ένα σύνολο μεγάλου εύρους υγιών ενηλίκων. Τα ελάχιστα όρια (CUT-OFF), τα οποία όπως θα δούμε παρακάτω εξαρτώνται από το μέγεθος του δείγματος και τη διάρκεια των μετρήσεων, αναγνωρίζουν ελάχιστα πιθανά επίπεδα κατανάλωσης ενέργειας εκφρασμένα σαν πολλαπλάσια του βασικού ρυθμού μεταβολισμού ( BMR). Έχουμε δύο ελάχιστα όρια:

### *(1<sup>ο</sup>) CUT-OFF 1*

Ελέγχει εάν η αναφερόμενη πρόσληψη ενέργειας μπορεί να είναι αντιπροσωπευτική για μια συνηθισμένη μεγάλης διάρκειας πρόσληψη (long-term habitual intake). Η τιμή του πρώτου ελαχίστου ορίου είναι 1.35, όπου όπως αναφέραμε παραπάνω θεωρήθηκε ως η ελάχιστη δυνατή τιμή της συνηθισμένης ενεργειακής πρόσληψης ενός υγιούς ενήλικου ατόμου με χαμηλή φυσική δραστηριότητα. Το CUT-OFF 1 αποτελεί την ελάχιστη δυνατή τιμή ενεργειακής πρόσληψης υγιών ατόμων, υπό την προϋπόθεση ότι αυτή είναι αντιπροσωπευτική της συνηθισμένης διατροφής τους, το οποίο μπορεί δύσκολα να διασφαλιστεί ακόμα και όταν έχουμε καταγραφή πρόσληψης φαγητού σε μεγάλες χρονικές περιόδους (π.χ με διατροφικά ιστορικά, ερωτηματολόγια συχνότητας κατανάλωσης τροφίμων). Έτσι μπορούμε να ορίσουμε το CUT-OFF 2.

### *(2<sup>ο</sup>) CUT-OFF 2*

Ελέγχει εάν η αναφερόμενη πρόσληψη ενέργειας είναι ένα αληθοφανές μέτρο της κατανάλωσης φαγητού κατά τη διάρκεια μιας ακριβής περιόδου μέτρησης.

Χρησιμοποιούμε το δεύτερο ελάχιστο όριο για τον προσδιορισμό μιας ελάχιστης τιμής για την ενεργειακή πρόσληψη ακόμα και όταν αυτή δεν είναι αντιπροσωπευτική της συνηθισμένης διατροφής, όπως συμβαίνει στο CUT-OFF 1, ενώ λαμβάνουμε υπόψη την ανά ημέρα διακύμανση της διατροφής κατά την περίοδο που γίνεται η καταγραφή. Το CUT-OFF 2 είναι πιο «φιλελεύθερο» από το CUT-OFF 1 ενώ για τον υπολογισμό του χρειαζόμαστε το μέγεθος του δείγματος και τη διάρκεια των μετρήσεων, δηλαδή τον αριθμό των ημερών που καταγράφεται η πρόσληψη. Τα ελάχιστα όρια μπορούν να χρησιμοποιηθούν για την εκτίμηση δεδομένων ενεργειακής πρόσληψης και τιμές κατώτερες των ελαχίστων ορίων θεωρούνται ανακριβείς.

## **4.3 Λεπτομερής ανάλυση/ παραγωγή του CUT-OFF 2.**

Οι χαμηλές καταγραφές της πρόσληψης ενέργειας μπορούν να προέλθουν από 4 κύριους λόγους<sup>16</sup>:

- (1) μία απλή αποτυχία στο να καταγραφεί οτιδήποτε έχει φαγωθεί,
- (2) συνειδητή ή υποσυνείδητη υποεκτίμηση,
- (3) τροποποίηση των γευμάτων και
- (4) η στατιστική αβεβαιότητα που προκύπτει από το υψηλό επίπεδο της ανά ημέρα διακύμανσης κατά την πρόσληψη φαγητού.

Ο τελευταίος από τους 4 κύριους λόγους δεν απειλεί την γενική αξιοπιστία των μετρήσεων πρόσληψης αφού γεννάει τόσο υπερεκτιμήσεις της πρόσληψης όσο και υποεκτιμήσεις.

Για τον εντοπισμό, λοιπόν, των υποεκτιμούντων είναι απαραίτητη η ανάπτυξη οριακών τιμών. Οι οριακές αυτές τιμές πρέπει να λαμβάνουν υπ' όψιν τη γνωστή διακύμανση της ενεργειακής πρόσληψης, τη μέτρηση ή την πρόβλεψη του βασικού ρυθμού μεταβολισμού (BMR) και τα χαμηλότερα όρια αποδοχής για τις τιμές του επιπέδου φυσικής δραστηριότητας PAL. Οι οριακές αυτές τιμές μπορούν να προέλθουν όπως παρακάτω:

#### 4.3.1 Υπόθεση για ενεργειακή ισορροπία.

Σε μία μεγάλης περιόδου βάση, η πρόσληψη ενέργειας (EI) πρέπει να είναι ίση με την κατανάλωση ενέργειας (EE) σε ενήλικους με σταθερό βάρος. Έτσι:

$$EI = EE \Rightarrow \frac{EI}{EE} = 1 \quad (4.1)$$

Όμως το EE μπορεί να εκφραστεί σαν πολλαπλάσιο του BMR, αφού γνωρίζουμε ότι ισχύει:

$$PAL = \frac{EE}{BMR} \Rightarrow EE = BMR \cdot PAL$$

Αντικαθιστώντας στη σχέση (4.1) έχουμε την ισότητα ενεργειακής ισορροπίας εκφρασμένη διαφορετικά, όπως φαίνεται παρακάτω:

$$\frac{EI}{BMR \cdot PAL} = 1 \quad (4.2)$$

Όμως EI και PAL τείνουν να έχουν μια ασύμμετρη κατανομή η οποία πρέπει να ληφθεί υπόψη από ένα λογαριθμικό μετασχηματισμό. Εάν υποθέσουμε μία λογαριθμοκανονική κατανομή, τότε η διακύμανση του EI/(BMR.PAL) μπορεί να γραφεί:

$$\begin{aligned} \text{Variance}[\log(\frac{EI}{BMR \cdot PAL})] &= \text{var}(\log EI) + \text{var}(\log BMR) + \text{var}(\log PAL) \\ &\quad - 2r_{IB} \sqrt{[\text{var}(\log EI) \cdot \text{var}(\log BMR)]} \\ &\quad - 2r_{IP} \sqrt{[\text{var}(\log EI) \cdot \text{var}(\log PAL)]} \\ &\quad + 2r_{BP} \sqrt{[\text{var}(\log BMR) \cdot \text{var}(\log PAL)]} \end{aligned} \quad (4.3)$$

όπου οι τρεις τελευταίοι όροι ονομάζονται όροι συνδιακύμανσης (covariance terms) και όπου I,B,P αναφέρονται στο EI, BMR και PAL αντίστοιχα. Αλλά,  $\text{Var} = \text{SD}^2$  και  $\text{CV} = \log \text{SD}$ , συνεπώς:

$$\text{var}[\log(\frac{EI}{BMR \cdot PAL})] = \frac{CV_{IW}^2}{k} + CV_B^2 + CV_P^2 + \text{covariance terms}$$

και

$$SD[\log(\frac{EI}{BMR \cdot PAL})] = \sqrt{\frac{CV_{IW}^2}{k} + CV_B^2 + CV_P^2 + \text{covariance terms}} \quad (4.4)$$

όπου IW δηλώνει τη within-individual διακύμανση της ενεργειακής πρόσληψης και k είναι ο αριθμός των ημερών μέτρησης. Παρατηρούμε ότι όταν η διατροφική μέθοδος υποτίθεται ότι εκτιμά συνήθη πρόσληψη με διατροφικά ιστορικά, ερωτηματολόγια συχνότητας κατανάλωσης τροφίμων κ.α τότε  $k=\infty$  και  $CV_{IW}^2/k=0$ . Επίσης παρατηρούμε ότι η between-individual διακύμανση της πρόσληψης ενέργειας δεν εμφανίζεται στη σχέση (4.4) επειδή η πρόσληψη ενέργειας (EI) είναι μία εκτίμηση όχι της πρόσληψης του πληθυσμού αλλά της πρόσληψης ενός δοθέντος ατόμου. Για συντομία, το δεξί μέρος της ισότητας (4.4) θα αναφέρεται σαν S.

Για να εκτιμήσουμε την πιθανότητα ένα δοσμένο αποτέλεσμα να βρίσκεται μέσα στο αποδεκτό εύρος, είναι απαραίτητο να υπολογίσουμε ένα SD score:

$$SD_{score} = \frac{\log(\frac{EI}{BMR \cdot PAL}) - \text{mean}[\log(\frac{EI}{BMR \cdot PAL})]}{\frac{S}{\sqrt{n}}}$$

όπου n είναι ο αριθμός των ατόμων για τους οποίους μετράμε το EI. Άλλα, από την ισότητα (4.2) έχω  $\text{mean}[\log(\frac{EI}{BMR \cdot PAL})] = 0$ , έτσι:

$$SD_{score} = \frac{\log(\frac{EI}{BMR \cdot PAL})}{\frac{S}{\sqrt{n}}} \quad (4.5)$$

#### 4.3.2 Ανάλυση των συντελεστών του S.

Έχοντας ορίσει παραπάνω σαν S το δεξί μέρος της ισότητας (4.4)

(δηλαδή  $S = \sqrt{[\frac{CV_{IW}^2}{k} + CV_B^2 + CV_P^2 + \text{covariance terms}]}$ ) μπορούμε να αναλύσουμε και να περιγράψουμε λεπτομερώς κάθε συντελεστή της ισότητας χωριστά<sup>16,18</sup>:

(i)  $CV_{IW}$  = Συντελεστής μεταβλητότητας της ενεργειακής πρόσληψης στις αναφερόμενες τιμές του ίδιου ατόμου (within-individual CV for EI). Ο Bingham<sup>26</sup> κατέληξε ότι ο μέσος within-individual CV είναι 23%, ύστερα από 17 ξεχωριστές μελέτες σε 15 πληθυσμιακές

ομάδες διαφορετικής κατάστασης (οικονομικής, οικογενειακής κ.α.), διαφορετικής ηλικίας και διαφορετικού βάρους στη Βόρεια Αμερική, στην Αυστραλία και στην Ευρώπη. Οι Nelson et al<sup>27</sup> υπολόγισαν το CV<sub>IW</sub> για 18 ομάδες διαφορετικής ηλικίας και φύλλου σε 7 μελέτες που διεξήχθησαν στην Αγγλία. Τα αποτελέσματα μπορούν να συνοψιστούν όπως παρακάτω: βρέφη=18.7%, παιδιά σχολικής ηλικίας=22.3%, ενήλικες μέσης ηλικίας=22.5% και άτομα μεγαλύτερης ηλικίας=16.5%. Η τιμή, όμως, που υιοθετήθηκε είναι το 23%.

**(ii) CV<sub>B</sub>**= Συντελεστής μεταβλητότητας του ρυθμού βασικού μεταβολισμού, όπως αυτός εκτιμάται από τις εξισώσεις των Schofield et al.(CV for BMR)<sup>28</sup>. Το CV για μετρημένο (measured) και όχι εκτιμημένο BMR είναι κατά μέσο όρο 2.5% (Prentice et al)<sup>29</sup>. Εάν όμως το BMR εκτιμάται από τις εξισώσεις του Schofield<sup>28</sup> τότε το CV το παίρνουμε 8% .

**(iii) CV<sub>P</sub>**= Συντελεστής μεταβλητότητας του δείκτη φυσικής δραστηριότητας, όπως προσδιορίστηκε από τη μελέτη των FAO/WHO/UNU (CV for PAL)<sup>30</sup>. Στο FAO/WHO/UNU αναφέρεται ότι το CV για το PAL είναι 12.5%. Το μέσο CV από μελέτες που χρησιμοποιούν τη μέθοδο του διπλά επισημασμένου ύδατος βρέθηκε 13.7%. Εμείς παίρνουμε CV<sub>P</sub> ίσο με 12.5%.

**(iv) covariance terms** = Γνωρίζω ότι το EI μετριέται, το BMR είτε μετριέται είτε εκτιμάται και το PAL εκτιμάται. Εάν και το EI αλλά και το BMR μετριοούνται, η συσχέτισή τους r<sub>IB</sub> είναι πιθανό να είναι θετική αφού το BMR είναι μία σημαντική συνιστώσα του EE. Έτσι η παρουσία του πρώτου όρου συνδιακύμανσης στην ισότητα (3) τείνει να μειώσει το S. Αντιθέτως, εάν το BMR εκτιμάται από το βάρος του σώματος αυτό είναι μία πληθυσμιακή εκτίμηση όπου εξ' ορισμού είναι ασυσχέτιστο με το EI έτσι ο πρώτος όρος συνδιακύμανσης είναι μηδέν. Ο δεύτερος και τρίτος όρος στους όρους συνδιακύμανσης εμπλέκουν το PAL, για το οποίο υποθέτουμε μία απλή πληθυσμιακή τιμή. Επομένως ξανά εξ' ορισμού και οι δύο όροι συνδιακύμανσης είναι μηδέν. Η επίδραση της υπόθεσης, ότι η συνδιακύμανση μεταξύ EI και BMR, το οποίο μπορεί να μετρηθεί και δεν εκτιμάται, είναι μηδέν ενώ στην πραγματικότητα δεν είναι, θα ήταν να είχαμε το S μεγαλύτερο από ότι θα ήταν στην πραγματικότητα, όπου με τη σειρά του θα έκανε το cut-off χαμηλότερο. Κάθε σφάλμα που προκαλείται από τα παραπάνω, τείνει να κάνει πιο φιλελεύθερα τα απορρέοντα cut-off όρια. Δηλαδή στη συγκεκριμένη περίπτωση τα πραγματικά cut-off όρια θα έπρεπε να είναι

υψηλότερα από αυτά που έχουμε υπολογίσει και θα έπρεπε να ταξινομείται ένας μεγαλύτερος αριθμός αναφερόμενης πρόσληψης σαν μη αποδεκτός.

#### 4.3.3 Προσδιορισμός του κατώτερου ορίου.

Ένα κατώτερο όριο για το SD score ( $SD_{\min}$ ) πρέπει να προσδιοριστεί για να ορίσουμε τα όρια του διαστήματος εμπιστοσύνης. Από τη θεωρία<sup>16</sup>, τα cut-off όρια έχουν παραχθεί με 95% όρια εμπιστοσύνης ( $SD_{\min}=-2$ ) και με 99.7% όρια εμπιστοσύνης ( $SD_{\min}=-3$ ). Αυτά θα αντιπροσωπεύουν τα ελάχιστα όρια κάτω από τα οποία υπάρχει μόνο μία 2.5% ή μία 0.14% πιθανότητα όπου οι παρατηρούμενες τιμές EI/BMR μπορούν να έρθουν κατά τύχη. Η ισότητα (4.5) μπορεί να τροποποιηθεί για να βρούμε τιμές του λόγου EI/BMR που αντιστοιχούν στο  $SD_{\min}$ :

$$SD_{\min} = \frac{\log\left(\frac{EI}{BMR \cdot PAL}\right)}{\frac{S/100}{\sqrt{n}}} \Rightarrow$$

$$\log\left(\frac{EI}{BMR \cdot PAL}\right) = SD_{\min} \times \frac{S/100}{\sqrt{n}} \Rightarrow$$

$$\frac{EI}{BMR \cdot PAL} = \exp\left[SD_{\min} \times \frac{S/100}{\sqrt{n}}\right]$$

Επομένως για να αποφασίσουμε εάν μία δοθείσα τιμή του λόγου  $\frac{EI}{BMR}$  σε  $n$  άτομα είναι αποδεκτή, πρέπει να ικανοποιείται η ακόλουθη σχέση:

$$\frac{EI}{BMR} > PAL \times \exp\left[SD_{\min} \times \frac{S/100}{\sqrt{n}}\right]$$

όπου:

**S**  $\Rightarrow$  το έχουμε αναλύσει λεπτομερώς παραπάνω

**SD<sub>min</sub>** ⇒ -2 ή -3 όπως έχουμε αναφέρει

**n** ⇒ αριθμός ατόμων για τα οποία γίνεται ο έλεγχος (αν ο έλεγχος γίνεται σε ατομικό επίπεδο τότε n=1)

**PAL** ⇒ είναι ο μέσος όρος για το PAL για ένα πληθυσμό και υποθέτουμε ότι η τιμή του είναι 1.55 Η υπόθεση αυτή έγινε στη δημοσίευση των FAO/WHO/UNU<sup>30</sup>. Θα πρέπει να σημειώσουμε ότι η τιμή αυτή είναι χαμηλότερη από το 1.67 που προέρχεται με τη μέθοδο του διπλά επισημασμένου ύδατος. Η χρήση της τιμής 1.55 για το δείκτη φυσικής δραστηριότητας βασίζεται στην παραδοχή ότι τα άτομα που συμμετέχουν στη μελέτη έχουν ελαφρά φυσική δραστηριότητα. Η παραδοχή αυτή χρησιμοποιείται εφόσον δεν υπάρχουν διαθέσιμα δεδομένα για να εκτιμηθεί η συνηθισμένη φυσική δραστηριότητα. Η τιμή ορίστηκε από τους FAO/WHO/UNU και αναφέρεται σε καθιστικό τρόπο ζωής. Παρακάτω φαίνονται οι εκτιμήσεις για τα επίπεδα της φυσικής δραστηριότητας που προέκυψαν από αυτή τη μελέτη, ανάλογα με το φύλλο και το είδος της δραστηριότητάς τους<sup>18</sup>:

**Πίνακας 4.1**

	Δραστηριότητα		
Φύλλο	Ελαφρά	Μέτρια	Βαριά
Άνδρες	1.55	1.78	2.10
Γυναίκες	1.56	1.65	1.82

Παρατηρώ, ότι η τιμή 1.55 που χρησιμοποιούμε και προτάθηκε από τους Goldberg et al.<sup>16</sup>, είναι η ελάχιστη από όλες τις τιμές της φυσικής δραστηριότητας των ατόμων που ακολουθούν φυσιολογικές συνθήκες διαβίωσης. Εάν χρησιμοποιούσαμε μια μεγαλύτερη τιμή για το PAL, για τον προσδιορισμό του CUT-OFF 2, τότε θα είχαμε μεγαλύτερο ποσοστό ατόμων που υποεκτιμούν την πρόσληψη. Ο μοναδικός τρόπος να υπολογίζεται αντικειμενικά η τιμή του CUT-OFF 2, είναι να εκτιμάται για κάθε άτομο χωριστά ανάλογα με τις πραγματικές δραστηριότητές του, πράγμα το οποίο προϋποθέτει την ακριβή μέτρηση της φυσικής δραστηριότητας με μια κατάλληλη μέθοδο. Μία τέτοια μέθοδο θα δούμε στο κεφάλαιο 4.4.1.

**EI** ⇒ είναι η ημερήσια ενεργειακή πρόσληψη η οποία μπορεί να μετρηθεί όπως έχουμε αναφέρει παραπάνω



**BMR** ⇒ είναι, όπως γνωρίζουμε, ο βασικός ρυθμός μεταβολισμού και εκτιμάται από τις εξισώσεις των Schofield et al<sup>28</sup> όπως παρακάτω:

Χρησιμοποιούμε τα σωματομετρικά δεδομένα των συμμετεχόντων, δηλαδή βάρος και ύψος για να μπορέσουμε να βρούμε το βασικό μεταβολισμό τους. Η εκτίμηση, λοιπόν του BMR, γίνεται χρησιμοποιώντας τις εξισώσεις των Schofield et al. Οι εξισώσεις ορίζονται με διαφορετικό τρόπο ανάλογα με το φύλλο και την ομάδα ηλικίας. Έτσι, για παράδειγμα δίνουμε τις παρακάτω εξισώσεις:

*Ηλικία 30-60 ετών*

$$\text{Ανδρες} \quad \text{BMR} = 0.048 \times \text{Βάρος} - 0.001 \times \text{Ύψος} + 3.670$$

$$\text{Γυναίκες} \quad \text{BMR} = 0.034 \times \text{Βάρος} - 0.006 \times \text{Ύψος} + 3.530$$

*Ηλικία ≥60 ετών*

$$\text{Ανδρες} \quad \text{BMR} = 0.038 \times \text{Βάρος} - 4.068 \times \text{Ύψος} + 3.491$$

$$\text{Γυναίκες} \quad \text{BMR} = 0.033 \times \text{Βάρος} + 1.917 \times \text{Ύψος} + 0.074$$

οπού το BMR υπολογίζεται σε MJ/24ωρο, το βάρος σε kg και το ύψος σε m.

#### 4.3.4 Παράδειγμα-Εφαρμογή (πάνω στο Goldberg cut-off όριο).

Για να κατανοήσουμε τα παραπάνω θεωρούμε το παρακάτω παράδειγμα:

Έστω 9 άτομα με «καθιστικό τρόπο ζωής», δηλαδή ελαφρά δραστηριότητα όπου έχουμε μετρήσεις για 7-ημέρες και το  $SD_{\min}$  το θέτουμε  $-2$ .

Εάν το BMR εκτιμάται, τότε:

$$S = \sqrt{\left[ \frac{CV_{TW}^2}{k} + CV_B^2 + CV_P^2 \right]} \Rightarrow S = \sqrt{\left[ \frac{23^2}{7} + 8^2 + 12.5^2 \right]} = 17.2$$

Έτσι η οριακή τιμή είναι:

$$\text{Cut-off value} = PAL \times \exp\left[SD_{\min} \times \frac{S/100}{\sqrt{n}}\right] = 1.55 \times \exp\left[-2 \times \frac{0.172}{\sqrt{9}}\right] = 1.38,$$

εφόσον ο έλεγχος γίνεται σε ομαδικό επίπεδο

$$\text{Cut-off value} = PAL \times \exp\left[SD_{\min} \times \frac{S/100}{\sqrt{n}}\right] = 1.55 \times \exp\left[-2 \times \frac{0.172}{\sqrt{1}}\right] = 1.099,$$

εφόσον ο έλεγχος γίνεται σε ατομικό επίπεδο

Συνεπώς, αν ο μέσος αναφερθείς λόγος  $\frac{EI}{BMR}$  για το σύνολο του δείγματος υπερβαίνει το

1.39 θα μπορούσαμε να κάνουμε μια πρώτη αξιολόγηση της καταγραφόμενης πρόσληψης σε

ομαδικό επίπεδο και θα μπορούσαμε να πούμε ότι οι αναφερόμενες τιμές πρόσληψης του συγκεκριμένου δείγματος δεν είναι ασύμβατες με τις δυνατές τιμές πρόσληψης ενός πληθυσμού με ελαφρά δραστηριότητα.

Σε ατομικό επίπεδο, βρίσκουμε το λόγο  $\frac{EI}{BMR}$  για κάθε άτομο και το συγκρίνουμε με το

1.099, έτσι ώστε να προσδιοριστούν συγκεκριμένα άτομα που αναφέρουν υποεκτιμημένες τιμές πρόσληψης.

#### *ΠΡΟΣΟΧΗ*

(1) Εάν το BMR μετριέται και δεν εκτιμάται, τότε το  $CV_B$ , στον τύπο της διακύμανσης αντί για 8 θα γίνει 2.5.

(2) Εάν τα άτομα είναι πιο δραστήρια τότε αντί για 1.55, η τιμή του PAL αλλάζει σύμφωνα με τον πίνακα που έχουμε δώσει παραπάνω ή σύμφωνα με την αναφορά των FAO/WHO/UNU.

#### **4.4 Βελτίωση των οριακών τιμών από την εξίσωση της Goldberg.**

Ένας μεγάλος αριθμός ερευνών χρησιμοποίησαν το Goldberg cut-off όριο για να αναγνωρίσουν τους υποεκτιμούντες και να εξετάσουν τα χαρακτηριστικά τους ή να εξετάσουν τις επιδράσεις της υποεκτίμησης στα δεδομένα και να εξάγουν κάποια συμπεράσματα. Όμως η τεχνική δεν ήταν απόλυτα κατανοητή ή σωστά εφαρμόσιμη. Έτσι, οι παρερμηνείες που υπήρξαν οδήγησαν την Black<sup>17</sup>, να ξανά ασχοληθεί με το θέμα των υποεκτιμούντων, με σκοπό να ξαναεκφράσει τις αρχές των cut-off ορίων της Goldberg για των προσδιορισμό των υποεκτιμούντων στην πρόσληψη ενέργειας και να αναβαθμίσει τις τιμές που αντικαθιστούμε στην ισότητα για τον υπολογισμό της οριακής τιμής.

Επομένως, προτάθηκαν νέες τιμές για κάθε στοιχείο της εξίσωσης της Goldberg ενώ η τιμή του PAL, που μέχρι πρότινος ήταν 1.55 αντικατοπτρίζοντας έναν ελαφρά ασκούμενο πληθυσμό, προτάθηκε να τροποποιείται κάθε φορά ώστε να αντανάκλα τη γενική εικόνα φυσικής δραστηριότητας του εκάστοτε εξεταζόμενου πληθυσμού. Η προτεινόμενη τιμή για το συντελεστή μεταβλητότητας της ενεργειακής πρόσληψης στις αναφερόμενες τιμές του ίδιου ατόμου παραμένει αμετάβλητη και ίση με 23%. Ο συντελεστής μεταβλητότητας του ρυθμού βασικού μεταβολισμού, για το εκτιμώμενο ή το μετρήσιμο BMR είναι 8.5% και 4% αντίστοιχα (ενώ προηγουμένως είχαμε 8% και 2.5%). Τέλος, για το συντελεστή

μεταβλητότητας του δείκτη φυσικής δραστηριότητας, η προτεινόμενη τιμή είναι 15%. Οι αλλαγές αυτές έχουν σαν επίδραση τη διαπλάτυνση των ορίων εμπιστοσύνης. Παρακάτω, παραθέτουμε αναλυτικά τους παράγοντες που αντικαθιστούμε στην εξίσωση της Goldberg και τις αλλαγές που έχουν υποστεί.

#### **4.4.1 Επίπεδο φυσικής δραστηριότητας (PAL).**

Όπως έχουμε αναφέρει, η τιμή 1.55 που χρησιμοποιείται για το PAL, είναι η τιμή του WHO για “ελαφριά” δραστηριότητα. Οι λόγοι της επιλογής αυτής της τιμής ήταν αρκετοί, όπως<sup>17</sup>: η κατανάλωση ενέργειας ή η φυσική δραστηριότητα ήταν άγνωστη, επιλέγοντας μια υψηλή τιμή ήταν πιθανό να μεγαλοποιήσουμε το μέγεθος της υποεκτίμησης, οι υποδείξεις του WHO για τις απαιτήσεις ενέργειας ήταν γνωστές και ευρέως χρησιμοποιούμενες και τέλος οι μέθοδοι της θερμιδομετρίας και του διπλά επισημασμένου ύδατος επιβεβαίωσαν την τιμή του 1.55 για την ελαφρά δραστηριότητα σαν πιθανή ελάχιστη απαίτηση ενέργειας για κανονικά δραστήριο άλλα καθιστικό πληθυσμό. Μία διαδοχική ανασκόπηση από 74 μελέτες, με τη μέθοδο του διπλά επισημασμένου ύδατος, έδειξαν ότι το 1.55 ήταν μια συντηρητική τιμή και πολλές ομάδες έχουν υψηλότερη κατανάλωση ενέργειας. Άλλες τιμές για το PAL, για άτομα πιο δραστήρια δίνονται στην αναφορά των FAO/WHO/UNU<sup>30</sup>.

Η χρήση της τιμής 1.55 για το δείκτη φυσικής δραστηριότητας βασίζεται στην παραδοχή ότι τα άτομα που συμμετέχουν σε κάθε μελέτη έχουν ελαφριά φυσική δραστηριότητα (παρακάτω θα αναφέρουμε λεπτομερώς πως προσδιορίζεται η φυσική δραστηριότητα). Αυτή η παραδοχή είναι αναπόφευκτη σε περίπτωση που δεν έχουμε διαθέσιμα δεδομένα για να εκτιμηθεί η συνηθισμένη φυσική δραστηριότητά τους. Η τιμή λοιπόν αυτή προσδιορίστηκε ως ενδεικτική για φυσική δραστηριότητα που αντιστοιχεί σε καθιστικό τρόπο ζωής. Οι εκτιμήσεις που προέκυψαν από τη μελέτη αυτή για τα επίπεδα φυσικής δραστηριότητας ανδρών και γυναικών ανάλογα με το είδος της δραστηριότητάς τους τα έχουμε αναφέρει και παραπάνω και φαίνονται στον Πίνακα 4.1

Σκοπός τώρα είναι να βρούμε μία μέθοδο ώστε να μπορέσουμε να προσδιορίσουμε πότε μια δραστηριότητα είναι ελαφρά, πότε μέτρια και πότε υψηλή. Η μέθοδος που χρησιμοποιούμε περιγράφεται παρακάτω<sup>20</sup>:

Στα ερωτηματολόγια του ΕΠΙΚ υπάρχουν ερωτήσεις που αφορούν στη φυσική δραστηριότητα των συμμετεχόντων τόσο επαγγελματικά όσο και ψυχαγωγικά. Οι ερωτήσεις είναι οι εξής:

*1<sup>η</sup>) (αφορά στην επαγγελματική δραστηριότητα)*

Σε ποια από τις ακόλουθες κατηγορίες θα κατατάσσατε τη δουλειά σας:

1. Καθιστική απασχόληση (π.χ. δουλειά γραφείου)
2. Όρθια απασχόληση (π.χ. βοηθός μαγαζιού, κομμωτής, φύλακας)
3. Χειρονακτική εργασία (π.χ. υδραυλικός, μαραγκός)
4. Βαριά χειρονακτική εργασία (π.χ. οικοδόμος, ανθρακωρύχος)

*2<sup>η</sup>) (αφορά στην ψυχαγωγική δραστηριότητα)*

Κατά τη διάρκεια του περασμένου έτους πόσες ώρες αφιερώσατε κατά μέσο όρο, ανά εβδομάδα σε καθεμιά από τις παρακάτω δραστηριότητες:

- 2.1 Περπάτημα (συμπεριλαμβάνεται το περπάτημα για τη δουλειά, τα ψώνια, την ψυχαγωγία) το καλοκαίρι.
- 2.2 Περπάτημα (συμπεριλαμβάνεται το περπάτημα για τη δουλειά, τα ψώνια, την ψυχαγωγία) το χειμώνα.
- 2.3 Ποδηλασία (συμπεριλαμβάνεται ποδηλασία προς τη δουλειά, ψυχαγωγία) το καλοκαίρι.
- 2.4 Ποδηλασία (συμπεριλαμβάνεται ποδηλασία προς τη δουλειά, ψυχαγωγία) το χειμώνα.
- 2.5 Κηπουρική το καλοκαίρι.
- 2.6 Κηπουρική το χειμώνα.
- 2.7 Μαστορέματα στο σπίτι.
- 2.8 Φυσική άσκηση, όπως για τη διατήρηση φόρμας, αεροβική γυμναστική, κολύμπι, τρέξιμο, τένις κ.λ.π το καλοκαίρι.
- 2.9 Φυσική άσκηση, όπως για τη διατήρηση φόρμας, αεροβική γυμναστική, κολύμπι, τρέξιμο, τένις κ.λ.π το χειμώνα.
- 2.10 Οικιακά, όπως καθάρισμα, πλύσιμο, μαγείρεμα, ντάντεμα παιδιών.

Με βάση λοιπόν, την επαγγελματική δραστηριότητα και την ψυχαγωγική δραστηριότητα μπορούμε να βρούμε σε ποιο επίπεδο φυσικής δραστηριότητας ανήκει το κάθε άτομο. Αποφασίσαμε τα άτομα που δεν ανέφεραν επαγγελματική δραστηριότητα να ταξινομηθούν στην ομάδα με καθιστική απασχόληση.

Επικεντρώσαμε την προσοχή μας μόνο στην ποδηλασία και την φυσική άσκηση επειδή γενικά η φυσική δραστηριότητα με μεγαλύτερη ένταση αναφέρθηκε με μεγαλύτερη ακρίβεια στις συγκεκριμένες ερωτήσεις του ερωτηματολογίου. Έτσι, λοιπόν έχουμε δύο μεταβλητές όπου η μία περιλαμβάνει το άθροισμα της ποδηλασίας και της φυσικής άσκησης και η άλλη

το άθροισμα των άλλων δραστηριοτήτων που ονομάζεται δραστηριότητα χαμηλής έντασης και δεν μας ενδιαφέρει στον υπολογισμό μας<sup>20</sup>.

Η μεταβλητή που αφορά στην ποδηλασία και την φυσική άσκηση χωρίζεται σε τέσσερις κατηγορίες: 0h/day (ή 0h/week), από 0 μέχρι 0.5h/day (ή (0,3.5]h/week), από 0.5 έως 1h/day (ή (3.5,7]h/week) και περισσότερο από 1h/day (ή (7,168]h/week).

Ο τρόπος που κατηγοριοποιούνται οι συμμετέχοντες ως προς τη δραστηριότητα τους με βάση την επαγγελματική και την ψυχαγωγική δραστηριότητα (ποδηλασία και φυσική δραστηριότητα) είναι ο εξής:

**Πίνακας 4.2**

<b>Επίπεδο δραστηριότητας</b>	<b>Περιγραφή</b>
<b><i>Ελαφριά δραστηριότητα</i></b>	Καθιστική εργασία και 0h/week ψυχαγωγική δραστηριότητα ή Καθιστική εργασία και (0,3.5]h/week ψυχαγωγική δραστηριότητα ή Όρθια εργασία και 0h/week ψυχαγωγική δραστηριότητα
<b><i>Μέτρια δραστηριότητα</i></b>	Καθιστική εργασία και (3.5,7]h/week ψυχαγωγική δραστηριότητα ή Όρθια εργασία και (0,3.5]h/week ψυχαγωγική δραστηριότητα ή Χειρονακτική εργασία και 0h/week ψυχαγωγική δραστηριότητα
<b><i>Βαριά δραστηριότητα</i></b>	Καθιστική εργασία και (7,168]h/week ψυχαγωγική δραστηριότητα ή Όρθια εργασία και (3.5,168]h/week ψυχαγωγική δραστηριότητα ή Χειρονακτική εργασία και τουλάχιστον κάποια ψυχαγωγική δραστηριότητα ή Βαριά χειρονακτική εργασία

Πηγή: Wareham J Nicholas et al (2003)<sup>20</sup>

#### 4.4.2 Συντελεστής μεταβλητότητας της ενεργειακής πρόσληψης στις αναφερόμενες τιμές του ίδιου ατόμου $CV_{wEI}$ (within-individual CV for EI).

Η within- subject ανά ημέρα διακύμανση στην πρόσληψη φαγητού των ατόμων είναι μεγάλη. Το  $CV_w$  από μελέτες ανασκόπησης<sup>17</sup> που έχουν γίνει από τους Bingham<sup>26</sup>, Nelson και άλλους συνεργάτες<sup>27</sup> εκτείνεται από 14% έως 45% με ένα συνολικό μέσο (pooled mean) 23%. Αυτό το 23% θεωρήθηκε σαν μια κατάλληλη μέση τιμή για να αντικατασταθεί στην εξίσωση Gordberg. Όμως ένα συγκεκριμένο  $CV_{wEI}$  μπορεί να υπολογιστεί για κάθε δοσμένη μελέτη, σύμφωνα με τον τύπο:

$$\text{Pooled mean } CV_w = \sqrt{\frac{\sum (CV_i^2)}{n}}$$

όπου  $CV_i$  είναι το CV υπολογισμένο για κάθε άτομο από τον αριθμό των ημερών της διατροφικής εκτίμησης που είναι διαθέσιμος για κάθε άτομο και  $n$  είναι ο αριθμός των ατόμων.

#### 4.4.3 Συντελεστής μεταβλητότητας του δείκτη φυσικής δραστηριότητας $CV_{IP}$ (between-subject CV for PAL).

Όπως γνωρίζουμε το  $CV_{IP}$  είναι ο συντελεστής της συνολικής μεταβλητότητας της απαιτούμενης-καταναλωμένης ενέργειας υπολογισμένη από το μέσο και την τυπική απόκλιση της μελέτης. Το  $CV_{IP}$  αν και αναφέρεται σαν “between-subject” μεταβλητότητα περιλαμβάνει within ( $CV_w$ ) και between ( $CV_b$ ) subject μεταβλητότητα άλλα και μεθοδολογικά σφάλματα. Η τιμή του 12.5% που ορίστηκε στο αρχικό άρθρο<sup>16</sup>, πάρθηκε από την αναφορά των FAO/WHO/UNU<sup>30</sup> για τις απαιτήσεις ενέργειας και πρωτεϊνών, άλλα μόνο μία μελέτη αναφέρθηκε για να υποστηρίξει αυτή την τιμή.

Χρησιμοποιώντας συσσωρευμένα δεδομένα από μελέτες που χρησιμοποιούν τη μέθοδο του διπλά επισημασμένου ύδατος, βλέπουμε ότι υπάρχει ένα μεγάλο εύρος τιμών για το PAL, από 1.2 έως πάνω από 2 για όλες τις ηλικίες. Έτσι το  $CV_{IP}$  από αυτά τα δεδομένα κυμαίνεται από 9.5% έως 23.8% για κάθε με διαφορετική ηλικία και φύλλο, με συνολικό μέσο 15.4%. Εξετάζοντας όμως μελέτες με επαναλαμβανόμενες μετρήσεις με τη μέθοδο του διπλά επισημασμένου ύδατος, βρέθηκε ότι οι τιμές ήταν λίγο χαμηλότερες και μία τιμή γύρω στο 15% προτάθηκε σαν κατάλληλη για την αντικατάστασή του στην εξίσωση της Goldberg (μπορούν να χρησιμοποιηθούν και άλλες τιμές εάν κριθούν κατάλληλες για κάποιες συγκεκριμένες μελέτες)<sup>17</sup>.

#### 4.4.4 Συντελεστής μεταβλητότητας του ρυθμού βασικού μεταβολισμού $CV_{WB}$ (CV for BMR).

Αφού η πρόσληψη και η κατανάλωση ενέργειας εκφράζονται σαν πολλαπλάσια του BMR, η between-subject μεταβλητότητα παίρνεται υπολογισμένη και μόνο η within-subject μεταβλητότητα χρειάζεται να θεωρηθεί στον υπολογισμό των ορίων εμπιστοσύνης. Αυτό περιλαμβάνει και το σφάλμα μετρήσεων αλλά και τη μεταβλητότητα στις επαναλαμβανόμενες μετρήσεις του BMR<sup>17</sup>.

Η κλασική μέτρηση του BMR γίνεται τουλάχιστον 12 ώρες μετά το τελευταίο γεύμα, μετά από ένα ολονύχτιο ύπνο στο μέρος της μέτρησης, εντελώς σε ανάπαυση και σε θερμοκρασία 23°C. Όμως, η πράξη υπαγορεύει ότι αυτές οι συνθήκες δύσκολα συναντώνται, οπότε σε αυτή την περίπτωση η μέτρηση ονομάζεται RMR (Resting Metabolic Rate)<sup>17</sup>. Είτε το άτομο πρέπει να περπατάει από το μέρος του ύπνου στο μέρος που γίνεται η μέτρηση (In-patient RMR) είτε να κοιμάται στο σπίτι και να ταξιδεύει γρήγορα στο μέρος των μετρήσεων το πρωί (Out-patient RMR), όπου και στις δυο περιπτώσεις η μέτρηση γίνεται μετά από μια περίοδο παύσης, συνήθως 30-60 λεπτά. Για τους Out-patient RMR, οι ερευνητές δεν έχουν τον έλεγχο πάνω στο μέγεθος του τελευταίου γεύματος, τη διάρκεια της ανάπαυσης ή της δραστηριότητας το προηγούμενο 24ωρο, όπου όλα αυτά επηρεάζουν το RMR. Άλλοι παράγοντες που μπορεί να επηρεάσουν το RMR και που διαφέρουν μεταξύ των μελετών είναι: η φύση των εφοδίων που χρησιμοποιήθηκαν (μάσκα προσώπου, κουκούλα για οξυγόνο), η εξοικείωση του ατόμου με τα εφόδια και τις διαδικασίες, η διάρκεια της δραστηριότητας το πρωινό της μέτρησης ή η περίοδος ανάπαυσης πριν τη μέτρηση, όπου όλα αυτά μπορεί να αυξήσουν τη within-subject μεταβλητότητα πάνω από αυτή που αποκτάται κάτω από επιθυμητές συνθήκες τυποποίησης του κλασικού BMR.

Έτσι, λοιπόν, κάτω από επιθυμητές συνθήκες τυποποίησης το  $CV_{WB}$ , για μετρήσιμο BMR, είναι περίπου κατά μέσο όρο 2.5%. Αυτή η τιμή προήρθε από 279 άτομα που μετρήθηκαν με τη μέθοδο της ολοσωματικής θερμιδομετρίας, σε διαδοχικές νύχτες όπου η προηγούμενη διατροφή και δραστηριότητα ήταν ίδια. Όμως στην κοινότητα των μελετών, κάποια άτομα πάρθηκαν σαν δείγμα από τον πληθυσμό και η μεταβλητότητα οφείλεται στην προηγούμενη διατροφή και δραστηριότητα, στις φυσικές διακυμάνσεις του βάρους, στον εμμηνορροϊκό κύκλο και στο μεθοδολογικό σφάλμα. Τα αποτελέσματα από 11 μελέτες που ερεύνησαν ειδικά τη within-subject μεταβλητότητα του BMR, και η πλειοψηφία διεξήχθη σε νέα άτομα, έδωσαν μέσο  $CV_{WB}$  ίσο με 3.9%. Τα αποτελέσματα από 12 μελέτες που αποκτήθηκαν με τη

μέθοδο του διπλά επισημασμένου ύδατος με επαναλαμβανόμενες μετρήσεις, έδωσε τιμή ίση με 5.5%, άλλα ο μέσος, από μη περιορισμένης δραστηριότητας άτομα, ήταν γύρω στο 4.7%. Ο μέσος όλων αυτών των μελετών υπολογίστηκε σε 4.2%. Επομένως μία τιμή γύρω στο 4% προτάθηκε σαν κατάλληλη μέση τιμή για να αντικαταστήσουμε το  $CV_{WB}$  στην εξίσωση της Goldberg, ενώ μπορούν να χρησιμοποιηθούν και άλλες τιμές εάν κριθούν καταλληλότερες για κάποια μελέτη.

Όταν δεν είναι διαθέσιμο το μετρήσιμο BMR ή RMR, αυτό μπορεί να εκτιμηθεί από το βάρος και το ύψος ή μόνο από το βάρος<sup>17</sup>. Πολλές εξισώσεις είναι διαθέσιμες και κάθε μία δίνει ένα ελαφρώς διαφορετικό αποτέλεσμα. Οι εξισώσεις στις οποίες βασίζεται ο μεγαλύτερος αριθμός δεδομένων και χρησιμοποιούνται ευρέως είναι αυτές του Schofield<sup>28</sup>. Η τιμή για το  $CV_{WB}$  που χρησιμοποιήθηκε στο αρχικό άρθρο ήταν 8%<sup>16</sup>. Όμως συγκεκριμένες τιμές για διαφορετικές ομάδες ηλικίας-φύλλου είναι διαθέσιμες από τις εξισώσεις του Schofield. Έτσι μια τιμή του 8.5% προτάθηκε σαν κατάλληλη μέση τιμή για να αντικαταστήσουμε το  $CV_{WB}$  στην εξίσωση της Goldberg, ενώ μπορούν να χρησιμοποιηθούν και άλλες τιμές εάν κριθούν καταλληλότερες για κάποια μελέτη.

#### **4.5 Μεταβολικά ισοδύναμα (MET).**

Ο προσδιορισμός της φυσικής δραστηριότητας είναι σημαντικός όχι μόνο για τον προσδιορισμό των υποεκτιμούντων στην περίπτωση μας άλλα και για άλλα σημαντικά προβλήματα. Ενδείξεις από πολλές επιδημιολογικές μελέτες στηρίζουν την άποψη ότι η έλλειψη φυσικής δραστηριότητας συσχετίζεται θετικά με τον κίνδυνο εμφάνισης χρόνιων νοσημάτων π.χ. στεφανιαία νόσο, οστεοπόρωση, καρκίνο του παχέος εντέρου.

Τα τελευταία χρόνια έχουν προταθεί μια σειρά από μεθόδους για τη μέτρηση της φυσικής δραστηριότητας<sup>21</sup>, από τις οποίες οι πιο ακριβείς θεωρούνται οι μέθοδοι του διπλά επισημασμένου ύδατος (doubly labeled water) και της άμεσης και έμμεσης θερμιδομετρίας (direct and indirect calorimetry). Όμως λόγω του μεγάλου κόστους και των τεχνικών δυσκολιών, στο να χρησιμοποιηθούν αυτές οι μέθοδοι ως διαδικασίες ρουτίνας, έχουμε περιορισμό στη χρήση τους σε μεγάλες επιδημιολογικές μελέτες. Επομένως, σαν εναλλακτική λύση έχουν βρεθεί τα τελευταία χρόνια οι μέθοδοι πεδίου (field methods), όπως: η ταξινόμηση των επαγγελματιών με βάση τη σωματική δραστηριότητα που απαιτείται κατά την εκτέλεσή τους (job classification), τα ερωτηματολόγια καταγραφής επαγγελματικών δραστηριοτήτων και ελεύθερου χρόνου που αφορούν συνήθως το ένα έτος (occupational and



leisure time physical activity questionnaires), τα αναμνηστικά ερωτηματολόγια μικρής χρονικής περιόδου αναφοράς που αφορούν συνήθως ένα 24ωρο (recall questionnaires) και τα ημερολόγια που αφορούν μέχρι μία εβδομάδα (diaries)<sup>21</sup>. Οι μέθοδοι πεδίου, λοιπόν, λόγω του ότι έχουν χαμηλό κόστος και λόγω της απλότητας στη χρήση τους είναι οι πιο κατάλληλες για τον προσδιορισμό ατόμων που ζουν υπό φυσιολογικές συνθήκες διαβίωσης. Από αυτές τις μεθόδους τα ερωτηματολόγια, είτε μικρής είτε μεγάλης περιόδου αναφοράς, είναι η μέθοδος που χρησιμοποιείται ευρέως στη μεγάλη πλειοψηφία των επιδημιολογικών μελετών. Όσον αφορά στις άλλες δύο μεθόδους παρατηρώ τα παρακάτω: η μέθοδος των ημερολογίων δεν χρησιμοποιείται συχνά γιατί απαιτεί τη συμμετοχή στην έρευνα ατόμων με μεγάλη υπευθυνότητα, για να διασφαλίζεται η ακρίβεια κατά τη συγγραφή του ημερολογίου ενώ η μέθοδος της ταξινόμησης των επαγγελμάτων δεν χρησιμοποιείται ευρέως γιατί απευθύνεται σε ομάδες του πληθυσμού που είναι επαγγελματικά ενεργές ενώ δεν λαμβάνει υπόψη της τη σωματική δραστηριότητα των συμμετεχόντων κατά των ελεύθερο χρόνο τους.

Τα ερωτηματολόγια είναι σχετικά ανέξοδη μέθοδος προσδιορισμού της φυσικής δραστηριότητας. Το γεγονός ότι μπορούν να συμπληρώνονται από τους ίδιους τους ερωτώμενους τα καθιστά ως την πλέον κατάλληλη μέθοδο μέτρησης της φυσικής δραστηριότητας σε μεγάλες πληθυσμιακές ομάδες. Τις τελευταίες δεκαετίες αναπτύχθηκαν ερωτηματολόγια φυσικής δραστηριότητας διαφορετικού βαθμού λεπτομέρειας και έκτασης, τα οποία περιλαμβάνουν δραστηριότητες διαφόρων τύπων και διαφορετικής περιόδου αναφοράς ανάλογα με τους σκοπούς της εκάστοτε έρευνας. Στην ενότητα 4.4 αναπτύξαμε το ερωτηματολόγιο που χρησιμοποιήσαμε στο ΕΠΙΚ, για τον προσδιορισμό της φυσικής δραστηριότητας των συμμετεχόντων.

Η μέτρηση της φυσικής δραστηριότητας μέσω ενός ερωτηματολογίου μπορεί να οδηγήσει στον προσδιορισμό της ενέργειας που συνήθως δαπανά ένα άτομο ανά ημέρα ή ώρα, εκφρασμένη είτε σε μονάδες ενέργειας (kcal ή kJ) είτε σε μεταβολικά ισοδύναμα. Το μεταβολικό ισοδύναμο (metabolic equivalent) ή MET, αντιστοιχεί στην ενέργεια που δαπανά ένα άτομο υπό συνθήκες ανάπαυσης<sup>21</sup>.

Για κάθε μια από τις δραστηριότητες που αναφέρει ο συμμετέχων ότι πραγματοποιεί, ορίζουμε το μεταβολικό ισοδύναμό της, ο λόγος δηλαδή του ρυθμού του μεταβολισμού κατά τη συγκεκριμένη δραστηριότητα, προς το ρυθμό του μεταβολισμού υπό συνθήκες ανάπαυσης. Οι τιμές των μεταβολικών ισοδυνάμων (MET values) για κάθε δραστηριότητα

προέκυψαν από δημοσιευμένους πίνακες όπως ενδεικτικά φαίνεται παρακάτω ένας τέτοιος πίνακας:

<b>Δραστηριότητες</b>	<b>MET</b>
<b>Εργασία</b>	
Καθιστική απασχόληση	1.5
Όρθια απασχόληση	2
Χειρονακτική εργασία	3
Βαριά χειρονακτική εργασία	5
<b>Εργασία στο σπίτι (οικιακά)</b>	
Ελαφριά εργασία	2
Μέση εργασία	2.5
Βαριά εργασία	4
Ύπνος	0.9
Ξεκούραση(διάβασμα, TV, γράψιμο κ.λ.π.)	1
Χρήση λεωφορείου, τρόλεϊ, τρένου	1.2
Οδήγηση αυτοκινήτου	1
Οδήγηση μηχανής	2.5
Κηπουρική	3.5
Ποδηλασία (για ψυχαγωγία ή για δουλειά)	4
<b>Αθλήματα</b>	
Βόλεϊ	4
Κολύμπι	4
Μπάσκετ	6
Ποδόσφαιρο	7
Τρέξιμο	7
Γυμναστική	4
Ορειβασία	8
Χορός	4.5
Ποδήλατο	7
Τένις	7
Κωπηλασία	7

Σκι χειμερινό	7
Σκι θερινό	6
Ιστιοσανίδα	3

Πηγή:Γναρδέλλης Χ. κ.α. (1999)<sup>21</sup>

Μια δραστηριότητα με μεταβολικό ισοδύναμο 2-MET απαιτεί διπλάσια δαπάνη ενέργειας από αυτή που δαπανάται σε συνθήκες ανάπαυσης. Πρακτικά, ο ορισμός των μεταβολικών ισοδυνάμων χαρακτηρίζει τις διάφορες δραστηριότητες από πλευράς έντασης λαμβάνοντας ως μονάδα μέτρησης της έντασης (1 MET) την ενέργεια που καταναλώνει ένα άτομο όταν αναπαύεται.

Σε περίπτωση που οι δαπάνες ενέργειας μιας δραστηριότητας είναι σκόπιμο να εκφραστούν σε θερμίδες, που καταναλώνονται από ένα άτομο συγκεκριμένου βάρους για ένα συγκεκριμένο χρονικό διάστημα, τότε θα πρέπει το μεταβολικό ισοδύναμο της συγκεκριμένης δραστηριότητας να πολλαπλασιαστεί επί το βάρος του ατόμου, επί τη χρονική διάρκεια άσκησης της δραστηριότητας. Για παράδειγμα μια δραστηριότητα 4 METs (ποδηλασία) ασκούμενη από ένα άτομο 60 κιλών επί 40 λεπτά, δημιουργεί κατανάλωση ενέργειας:

$4 \text{ MET} \times 60 \text{ kg} \times (40 \text{ min}/60 \text{ min}) = 160\text{kcal}$ , ενώ η κατανάλωση ενέργειας για την ίδια δραστηριότητα από ένα άτομο 80 κιλών είναι:  $4 \text{ MET} \times 80 \text{ kg} \times (40 \text{ min}/60 \text{ min}) = 213\text{kcal}$ .

Οι ημερήσιοι χρόνοι των διαφόρων δραστηριοτήτων που αναφέρθηκαν και οι οποίες καλύπτουν το τυπικό 24ωρο κάθε ατόμου, πολλαπλασιάστηκαν επί τις τιμές των μεταβολικών ισοδυνάμων τους και στη συνέχεια αθροίστηκαν. Με τον τρόπο αυτό για κάθε άτομο υπολογίστηκε ένας δείκτης 24ωρης φυσικής δραστηριότητας (MET-hours-per-day score). Ο δείκτης αυτός ορίζει τις θερμίδες που δαπανά κάθε άτομο ανά κιλό βάρους του, κατά τη διάρκεια ενός τυπικού 24ώρου. Έτσι συμβολίζοντας με  $h_i$  τις ώρες άσκησης ανά ημέρα μιας δραστηριότητας  $i$ , με μεταβολικό ισοδύναμο  $MET_i$ , ο δείκτης 24ωρης φυσικής δραστηριότητας ενός ατόμου για ένα σύνολο  $k$  δραστηριοτήτων την ημέρα ορίζεται ως εξής:

Δείκτης 24ωρης φυσικής δραστηριότητας =  $\sum (MET_i \times h_i)$  όπου  $i=1,2,\dots,k$ .



# ΚΕΦΑΛΑΙΟ 5

## Στατιστική ανάλυση των δεδομένων

### 5.1 Εισαγωγή

Όπως έχουμε αναφέρει, η εργασία αυτή αφορά στην ανάλυση και επεξεργασία δεδομένων που προέρχονται από το Ευρωπαϊκό Πρόγραμμα συνεργασίας Ιατρικής και Κοινωνίας (ΕΠΙΚ). Η βάση δεδομένων που θα χρησιμοποιήσουμε αφορά σε 300 Έλληνες που έχουν συμπληρώσει το διατροφικό ερωτηματολόγιο. Για τα άτομα αυτά έγινε προσδιορισμός:

- (i) των σωματομετρικών δεδομένων (π.χ. βάρος, ύψος) από τα οποία εκτιμήθηκε ο ρυθμός του βασικού τους μεταβολισμού,
- (ii) των δημογραφικών χαρακτηριστικών τους (π.χ. φύλλο, ηλικία) και
- (iii) των κοινωνικό-οικονομικών (π.χ. μορφωτικό επίπεδο) χαρακτηριστικών.

Επίσης, περιλαμβάνονται στοιχεία για την ημερήσια πρόσληψη ενέργειας και θρεπτικών συστατικών που προκύπτουν με την μέθοδο της διατροφικής ανάκλησης 24-ώρου και για τη μέση φυσική δραστηριότητα. Ο λόγος της ενεργειακής πρόσληψης προς το ρυθμό του βασικού μεταβολισμού (EI/BMR) αποτέλεσε το κριτήριο για την αξιολόγηση των αναφερόμενων τιμών πρόσληψης και την ταυτοποίηση των υποεκτιμούντων βάσει των οριακών τιμών (cut-off points) που προκύπτουν από τις εξισώσεις των Goldberg και συνεργατών<sup>16</sup>. Η στατιστική ανάλυση θα γίνει στο στατιστικό πακέτο STATA<sup>23</sup> και θα περιλαμβάνει μεταξύ άλλων και τα παρακάτω:

- (1) Ανάπτυξη ενός γενικευμένου γραμμικού μοντέλου χρησιμοποιώντας σαν ανεξάρτητες μεταβλητές το φύλλο, την ηλικία, το BMR, το επίπεδο μόρφωσης των συμμετεχόντων κ.α και σαν εξαρτημένη την κατάσταση των υποεκτιμούντων (ναι/όχι). Για την συγκεκριμένη περίπτωση θα χρησιμοποιήσουμε ένα μοντέλο λογιστικής παλινδρόμησης (logistic regression model) όπου μας δίνει και τη δυνατότητα πρόβλεψης.
- (2) Συγκρίσεις κάποιων χαρακτηριστικών π.χ. φύλλο, BMI στις δύο ομάδες, υποεκτιμούντων και μη, μέσω t-test.
- (3) Συντελεστές συσχέτισης του Pearson, όταν έχουμε συνεχείς μεταβλητές και του Kendall ή Spearman για κατηγορικές μεταβλητές κ.τ.λ.

## 5.2 Περιγραφικά χαρακτηριστικά.

Η επιλογή των ατόμων που συμμετέχουν στην έρευνα για την εκτίμηση των υποεκτιμούντων είναι εκείνοι που δήλωσαν ότι δεν ακολούθησαν κάποια συγκεκριμένη διαίτα, γεγονός που διασφαλίζει, σε μεγάλη έκταση, την απαίτηση για σταθερότητα του βάρους την περίοδο του ελέγχου.

Το υπό μελέτη δείγμα αποτελείται από 162 (54%) γυναίκες και 138 (46%) άνδρες. Παρακάτω παρουσιάζουμε μία σειρά από περιγραφικά χαρακτηριστικά τόσο στο επίπεδο των δύο φύλλων όσο και συνολικά για ολόκληρο το δείγμα μας:

### ❖ Ηλικία και ανθρωπομετρικά χαρακτηριστικά

Παρατηρώ ότι το δείγμα μας αποτελείται κυρίως από μεσήλικες. Η μέση ηλικία για τους άνδρες είναι 59.5 χρόνια, για τις γυναίκες 58 χρόνια ενώ για το σύνολο των συμμετεχόντων 59 χρόνια. Αναλύοντας τα ανθρωπομετρικά χαρακτηριστικά παρατηρούμε ότι το μέσο ύψος για τις γυναίκες είναι 155.5cm ενώ για τους άνδρες 170cm ενώ το βάρος είναι 81kg για τους άνδρες και 69.6kg για τις γυναίκες. Ο δείκτης μάζας σώματος (BMI), που είναι ο λόγος του βάρους προς το τετράγωνο του ύψους, για τους άνδρες και τις γυναίκες είναι 28.11 και 28.83 αντίστοιχα. Η περιφέρεια του γοφού και της μέσης είναι 103 και 99cm για τους άνδρες αντίστοιχα και 107 και 87cm για τις γυναίκες αντίστοιχα. Τέλος, όπως ορίσαμε το BMI μπορούμε να ορίσουμε το δείκτη WHR (waist-to-hip ratio), που είναι ο λόγος της περιμέτρου της μέσης προς την περίμετρο του ισχύος και είναι 0.96 για τους άνδρες και 0.81 για τις γυναίκες. Όλα αυτά που αναφέραμε παραπάνω συνοψίζονται στον πίνακα:

	Άνδρες (n=138)		Γυναίκες (n=162)		Σύνολο (n=300)	
	Mean	SD	Mean	SD	Mean	SD
<b>Ηλικία</b>	59.50	11.94	58.24	10.19	58.82	11.03
<b>Ύψος</b>	169.60	7.55	155.50	6.34	162.00	9.86
<b>Βάρος</b>	80.82	11.95	69.62	11.44	74.77	12.93
<b>BMI</b>	28.11	3.84	28.83	4.53	28.50	4.24

<b>Γοφός</b>	103.10	6.77	106.65	8.72	105.00	8.10
<b>Μέση</b>	98.90	10.65	87.50	11.99	92.72	12.71
<b>WHR</b>	0.96	0.07	0.82	0.08	0.88	0.10

Παρατηρούμε ότι το μέσο BMI είναι μεγαλύτερο στις γυναίκες από τους άνδρες γιατί έχουμε μέση ηλικία περίπου 59 χρόνια και σύμφωνα με προηγούμενες μελέτες έχει βρεθεί ότι οι άνδρες εμφανίζουν μεγαλύτερο BMI από τις γυναίκες μέχρι τα 44 έτη ενώ σε μεγαλύτερες ηλικίες η σχέση αυτή αντιστρέφεται, πράγμα το οποίο συμβαίνει και στα δικά μας δεδομένα:

```

-----
      |
      |          age5
      | sex | 24-34  35-44  45-54  55-64  65+
-----+-----
  male | 24.32372  27.7377  27.62419  28.00134  28.80581
  female | 24.74517  26.53553  27.76705  29.63301  29.99627
-----

```

Από την άλλη πλευρά το WHR είναι πάντοτε μεγαλύτερο στους άνδρες από τις γυναίκες σε όλες τις ηλικιακές ομάδες:

```

-----
      |
      |          age5
      | sex | 24-34  35-44  45-54  55-64  65+
-----+-----
  male | .8636551  .9249911  .929674  .9739661  .9756337
  female | .8163466  .7612352  .8051113  .8115311  .8612463
-----

```

#### ❖ Οικογενειακή κατάσταση

Βλέπουμε από τον παρακάτω πίνακα ότι το μεγαλύτερο ποσοστό είναι παντρεμένοι, με τους άνδρες να έχουν πιο υψηλό ποσοστό (91%) έναντι των γυναικών (80%). Αξιοσημείωτο είναι το ποσοστό των γυναικών που είναι χήρες (12%) και είναι πολύ μεγαλύτερο από το ποσοστό των ανδρών που είναι χήροι, το οποίο είναι μόλις 1.5%:

	<b>Άνδρες (n=138)</b>	<b>Γυναίκες (n=162)</b>	<b>Σύνολο (n=300)</b>
	<b>Ποσοστό (%)</b>	<b>Ποσοστό (%)</b>	<b>Ποσοστό (%)</b>
<b>Οικογενειακή κατάσταση</b>			
<i>Ελεύθερος-η</i>	5.0	4.0	5.0
<i>Παντρεμένος-η</i>	91.0	80.0	85.0
<i>Χωρισμένος-η</i>	2.5	4.0	3.0
<i>Χήρος-α</i>	1.5	12.0	7.0

❖ *Επίπεδο μόρφωσης*

Το επίπεδο μόρφωσης χωρίζεται σε 3 κατηγορίες: στη χαμηλή (που είναι μέχρι 6 χρόνια), τη μέση (που είναι από 7-14 χρόνια) και την ανώτερη (που είναι από 15 χρόνια και πάνω). Παρατηρούμε ότι το ποσοστό των γυναικών με χαμηλή μόρφωση (62%) είναι αρκετά υψηλό όπως και το αντίστοιχο ποσοστό των ανδρών το οποίο είναι λίγο χαμηλότερο από τις γυναίκες (52%). Στη μέση εκπαίδευση υπάρχει μια ισορροπία στα ποσοστά ανδρών και γυναικών (23%) ενώ το ποσοστό των ανδρών που έχουν υψηλότερη μόρφωση (25%) είναι μεγαλύτερο από αυτό των γυναικών (15%). Η κατανομή των συμμετεχόντων στις 3 κατηγορίες φαίνεται παρακάτω:

	<b>Άνδρες (n=138)</b>	<b>Γυναίκες (n=162)</b>	<b>Σύνολο (n=300)</b>
	<b>Ποσοστό (%)</b>	<b>Ποσοστό (%)</b>	<b>Ποσοστό (%)</b>
<b>Εκπαίδευση</b>			
<i>Χαμηλή, ≥ 6 χρόνια</i>	52	62	57
<i>Μέση, 7-14 χρόνια</i>	23	23	23
<i>Ανώτερη, 15+χρόνια</i>	25	15	20



❖ *Κατάσταση καπνίσματος*

Παρατηρούμε ότι ο αριθμός των γυναικών που δεν έχουν καπνίσει ποτέ (80%) είναι πολύ υψηλός σε σχέση με τον αντίστοιχο αριθμό των ανδρών (30.2%). Επίσης το μεγαλύτερο ποσοστό δεν κάπνισαν ποτέ, ήταν πρώην καπνιστές ή καπνίζουν μέχρι 10 τσιγάρα και μόνο το 6% των γυναικών καπνίζουν πάνω από 10 τσιγάρα ενώ το αντίστοιχο ποσοστό των ανδρών είναι σχεδόν τριπλάσιο:

	<b>Άνδρες (n=138)</b>	<b>Γυναίκες (n=162)</b>	<b>Σύνολο (n=300)</b>
	<b>Ποσοστό (%)</b>	<b>Ποσοστό (%)</b>	<b>Ποσοστό (%)</b>
<b>Κατάσταση καπνίσματος</b>			
<i>Ποτέ</i>	30	80	58.0
<i>Πρώην καπνιστής</i>	40	5	21
<i>Νυν καπνιστής</i>	30	15	22
<b>Αριθμός τσιγάρων</b>			
<i>Κανένα</i>	30	80	58
<i>Πρώην ή έως 10 τσιγάρα</i>	51	14	31
<i>11-20</i>	7	4	5
<i>21-30</i>	2	1	1
<i>31-40</i>	6	1	3
<i>41+</i>	4	-	2

Από τον παραπάνω πίνακα παρατηρούμε ότι 11% των ανδρών καπνίζουν μέχρι 10 τσιγάρα (51%-40%) ενώ το αντίστοιχο ποσοστό των γυναικών είναι 9% (14%-5%).

❖ *Απασχόληση*

Βλέπουμε ότι ένα πολύ μεγάλο ποσοστό γυναικών δεν εργάζονται ή κάνουν καθιστική εργασία (83%) ενώ το αντίστοιχο ποσοστό ανδρών είναι σαφώς μικρότερο αλλά και πάλι πολύ υψηλό (63%). Επίσης, παρατηρούμε ότι δεν υπάρχει καμία γυναίκα που να κάνει βαριά χειρονακτική εργασία ενώ το ποσοστό για τους άνδρες είναι μόλις 2%:

	Άνδρες (n=138)	Γυναίκες (n=162)	Σύνολο (n=300)
	Ποσοστό (%)	Ποσοστό (%)	Ποσοστό (%)
<b>Απασχόληση</b>			
<i>Ανεργοί/ Καθιστική εργασία</i>	63	83	74
<i>Ορθια εργασία</i>	17	10	13
<i>Χειρονακτική εργασία</i>	18	7	12
<i>Βαριά χειρονακτική εργασία</i>	2	-	1

❖ *Περιοχή-Περιφέρεια*

Από τον παρακάτω πίνακα βλέπουμε ότι το μεγαλύτερο ποσοστό των συμμετεχόντων ζει στην Αττική (34%), όπου το ποσοστό αυτό είναι αρκετά αυξημένο στους άνδρες (43%) και χαμηλότερο στις γυναίκες (27%) ενώ το ποσοστό των γυναικών που ζουν στην Νότια Ελλάδα είναι διπλάσιο από αυτό των ανδρών. Τέλος, παρατηρούμε ότι ένα πολύ μεγάλο ποσοστό ανδρών ζει σε αστικές περιοχές (72%) ενώ το ποσοστό των γυναικών που ζει τόσο σε αγροτικές όσο και σε αστικές περιοχές είναι σχεδόν μοιρασμένο:

	Άνδρες (n=138)	Γυναίκες (n=162)	Σύνολο (n=300)
	Ποσοστό (%)	Ποσοστό (%)	Ποσοστό (%)
<b>Περιφέρεια</b>			
<i>Αττική</i>	43	27	34
<i>Κεντρική Ελλάδα</i>	22	16	19
<i>Βόρεια Ελλάδα</i>	15	24	20
<i>Νότια Ελλάδα</i>	9	18	14
<i>Νησιά Αιγαίου</i>	11	15	13

Περιοχή			
Αστική περιοχή	72	57	64
Αγροτική περιοχή	28	43	36

❖ *MET*

Ο δείκτης 24ωρης φυσικής δραστηριότητας δεν παρουσιάζει σημαντική διαφορά μεταξύ ανδρών και γυναικών αφού η τιμή του κυμαίνεται γύρω στο 35:

	Άνδρες (n=138)		Γυναίκες (n=162)		Σύνολο (n=300)	
	Mean	SD	Mean	SD	Mean	SD
<b>MET</b>	34.91	6.3	35.13	3.64	35.03	5.02

❖ *Διατροφικά συστατικά*

Παρατηρούμε ότι για όλα τα διατροφικά συστατικά οι μέσες τιμές των ανδρών είναι αρκετά μεγαλύτερες από αυτές των γυναικών όπως έχει παρατηρηθεί και σε άλλες μελέτες<sup>19</sup> :

Διατροφικά συστατικά (g)	Άνδρες (n=138)		Γυναίκες (n=162)		Σύνολο (n=300)	
	Mean	SD	Mean	SD	Mean	SD
<b>Πρωτεΐνες</b>	79.98	37.62	56.33	29.97	67.61	35.66
<b>Υδατάνθρακες</b>	201.59	110.77	159.09	75.08	178.64	95.44
<b>Λιπίδια</b>	103.41	47.80	76.64	39.82	88.95	45.6
<b>Αλκοόλες</b>	23.18	41.64	3.26	8.93	12.42	30.60

❖ *Ενέργεια και BMR*

Η ημερήσια θερμιδική πρόσληψη για τους άνδρες είναι 2583 kcal και για τις γυναίκες 2130 kcal. Με τη βοήθεια των γραμμικών υποδειγμάτων των Schofield et al εκτιμήθηκε το BMR για κάθε ένα από τα 300 άτομα με βάση το φύλλο, την ηλικία, το βάρος και το ύψος. Οι μέσες τιμές του BMR για τους άνδρες και τις γυναίκες είναι 1666 kcal και 1374 kcal αντίστοιχα. Επιπλέον για κάθε άτομο υπολογίστηκε ο λόγος της ενεργειακής πρόσληψης προς το ρυθμό μεταβολισμού, τον οποίο λόγο θα τον συγκρίνουμε με το cut-off σημείο της Goldberg:

Πίνακας 5.1

	Άνδρες (n=138)		Γυναίκες (n=162)		Σύνολο (n=300)	
	Mean	SD	Mean	SD	Mean	SD
<b>Ενέργεια (kcal)</b>	2582.53	340.94	2130.42	198.76	2338.39	354.23
<b>BMR</b>	1666.15	219.94	1374.48	128.24	1508.65	228.52
<b>EI/BMR</b>	1.30	0.51	1.13	0.50	1.21	0.51

### 5.3 Υποεκτιμούντες.

Για να υπολογίσουμε τους υποεκτιμούντες στηρίζομαστε στις εξισώσεις της Goldberg. Εκεί που πρέπει να εστιάσουμε είναι στον υπολογισμό του επιπέδου φυσικής δραστηριότητας. Όπως αναφέραμε παραπάνω, τα σωματομετρικά δεδομένα των συμμετεχόντων και ειδικότερα το βάρος και το ύψος, χρησιμοποιήθηκαν για τον προσδιορισμό του βασικού μεταβολισμού τους. Οι τιμές, λοιπόν, του BMR που προέκυψαν, χρησιμοποιήθηκαν για τον υπολογισμό του λόγου EI/BMR.

Για τον υπολογισμό του cut-off ορίου, λαμβάνουμε υπόψη τόσο την ανά ημέρα διακύμανση της διατροφής κάθε ατόμου (intra-individual variation) όσο και τη διακύμανση των εκτιμούντων τιμών του BMR και PAL (Goldberg et al).

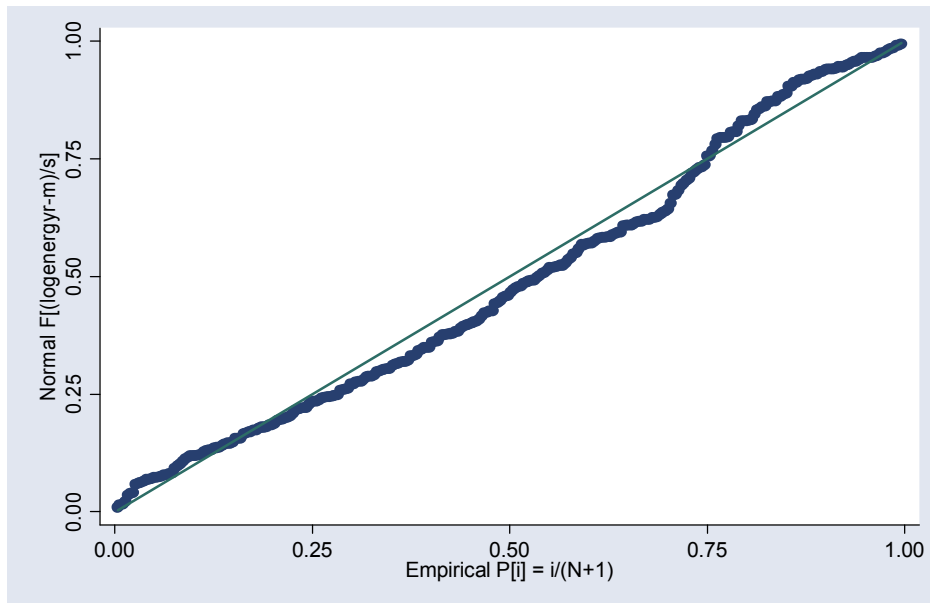
Για να ορίσουμε το cut-off όριο έχουμε υποθέσει<sup>16</sup> ότι η ενεργειακή πρόσληψη EI και ο λόγος EI/BMR ακολουθούν λογαριθμοκανονική κατανομή. Αρκεί, λοιπόν, να αποδείξουμε ότι ο λογάριθμός τους ακολουθεί κανονική κατανομή:

*(i) έλεγχος για την πρόσληψη ενέργειας EI*

Θα ελέγξουμε αν ο λογάριθμος της EI ακολουθεί κανονική κατανομή με το Kolmogorov-Smirnov test και με ένα διάγραμμα P-P Plot:

One-sample Kolmogorov-Smirnov test against theoretical distribution  
 normprob(( logenergyr-7.746181)/0.1477314)

Smaller group	D	P-value	Corrected
logenergyr:	0.0631	0.092	
Cumulative:	-0.0514	0.205	
Combined K-S:	0.0631	<b>0.183</b>	0.164



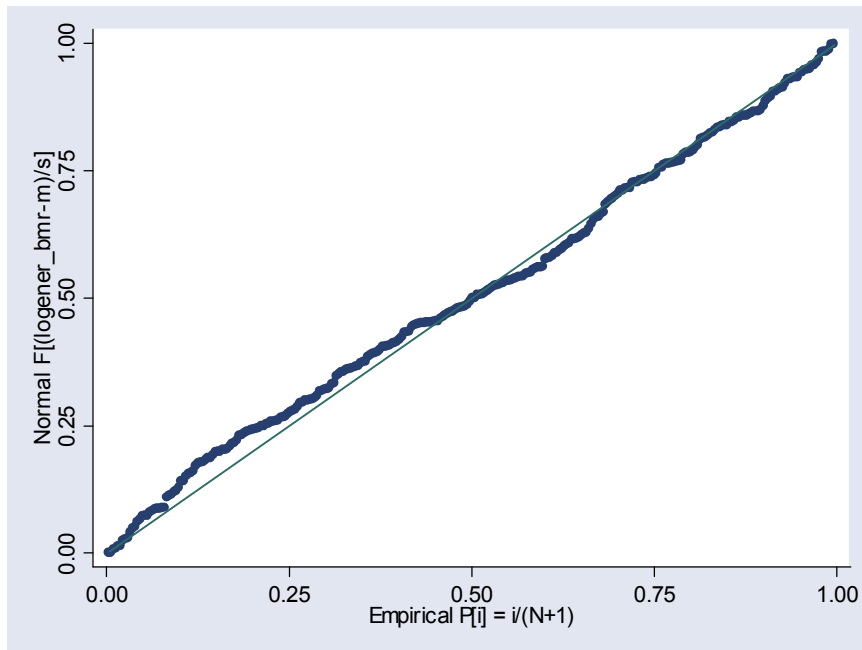
Παρατηρούμε ότι το p-value στο παραπάνω τεστ είναι μεγάλο (=0.183) που σημαίνει ότι δεν απορρίπτουμε την μηδενική υπόθεση, δηλαδή ο λογάριθμος της ενεργειακής πρόσληψης ακολουθεί κανονική κατανομή άρα η ενεργειακή πρόσληψη ακολουθεί λογαριθμοκανονική. Κοιτάζοντας το διάγραμμα παρατηρούμε ότι υπάρχει καλή προσαρμογή των σημείων μου στη γραμμή που σημαίνει ότι καταλήγουμε στο ίδιο συμπέρασμα με παραπάνω.

(ii) έλεγχος για το λόγο  $\frac{EI}{BMR}$

Θα ελέγξουμε αν ο λογάριθμος του λόγου EI/BMR ακολουθεί κανονική κατανομή με το Kolmogorov-Smirnov test και με ένα διάγραμμα P-P Plot:

One-sample Kolmogorov-Smirnov test against theoretical distribution  
 normprob(( logener\_bmr-0.1069604)/0.4265584)

Smaller group	D	P-value	Corrected
logener_bmr:	0.0397	0.388	
Cumulative:	-0.0519	0.198	
Combined K-S:	0.0519	<b>0.394</b>	0.365



Παρατηρούμε ότι το p-value στο παραπάνω τεστ είναι μεγάλο (=0.394) που σημαίνει ότι δεν απορρίπτουμε την μηδενική υπόθεση δηλαδή ο λογάριθμος του λόγου EI/BMR ακολουθεί κανονική κατανομή άρα ο λόγος  $\frac{EI}{BMR}$  ακολουθεί λογαριθμοκανονική. Κοιτάζοντας το διάγραμμα παρατηρούμε ότι υπάρχει καλή προσαρμογή των σημείων μου στη γραμμή, δηλαδή ο λογάριθμος του λόγου EI/BMR ακολουθεί κανονική κατανομή.

### 5.3.1 Προσδιορισμός των υποεκτιμούντων.

Από τη στιγμή που έχω αποδείξει ότι η πρόσληψη ενέργειας (EI) και ο λόγος EI/BMR ακολουθούν λογαριθμοκανονική κατανομή, σημαίνει ότι μπορούμε να χρησιμοποιήσουμε το cut-off όριο της Goldberg<sup>16</sup> που γνωρίζω ότι δίνεται από τον τύπο:

$$PAL \times \exp\left[SD_{\min} \times \frac{S/100}{\sqrt{n}}\right]$$

όπου:

$SD_{\min} = -2$  αφού το cut-off όριο τοποθετείται στο κατώτερο όριο του 95% διαστήματος εμπιστοσύνης της αναμενόμενης μέσης τιμής του PAL

$$\text{και } S = \sqrt{\left[\frac{CV_{IW}^2}{k} + CV_B^2 + CV_P^2\right]} \Rightarrow S = \sqrt{\left[\frac{23^2}{1} + 8.5^2 + 15^2\right]} = 28.744565$$

Δηλαδή το cut-off όριο δίνεται από τον τύπο:

$$PAL \times \exp\left[-2 \times \frac{0.28744565}{\sqrt{n}}\right] \quad (5.1)$$

Θα διακρίνουμε δύο μεθόδους για τον προσδιορισμό του επιπέδου φυσικής δραστηριότητας:

➤ *1<sup>η</sup> μέθοδος*

Σε αυτή την περίπτωση παίρνουμε το PAL ίσο με 1.55, που αντιστοιχεί σε ένα πληθυσμό με καθιστικό τρόπο ζωής.

- Σε ομαδικό επίπεδο έχουμε  $n=300$  και από την σχέση (5.1) συνεπάγεται:

$$1.55 \times \exp\left[-2 \times \frac{0.28744565}{\sqrt{300}}\right] = 1.50$$

Από τον Πίνακα 5.1 έχω βρει ότι η μέση τιμή του λόγου EI/BMR για το σύνολο του δείγματος είναι 1.21, δηλαδή χαμηλότερη από το cut-off όριο, το οποίο προσδιορίστηκε για το συγκεκριμένο δείγμα 1.50. Οι τιμές του ίδιου λόγου παραμένουν μικρότερες του cut-off ορίου και για καθένα από τα δύο φύλλα (1.30 για τους άνδρες και 1.13 για τις γυναίκες). Η πρώτη αυτή επισήμανση αποτελεί και μια πρώτη αξιολόγηση της καταγραφόμενης πρόσληψης σε ομαδικό επίπεδο. Θα μπορούσαμε επομένως να πούμε ότι οι αναφερόμενες τιμές πρόσληψης είναι «ασύμβατες» με τις δυνατές τιμές πρόσληψης του συγκεκριμένου πληθυσμού με ελαφρά δραστηριότητα. Αυτό όμως που μας ενδιαφέρει δεν είναι να πάρουμε μια γενική εικόνα του πληθυσμού μας αλλά να βρούμε ακριβώς ποιοι και πόσοι είναι οι υποεκτιμούντες. Επομένως εξετάζουμε σε ατομικό επίπεδο:

- Σε ατομικό επίπεδο έχουμε  $n=1$  και από την σχέση (5.1) συνεπάγεται:

$$1.55 \times \exp\left[-2 \times \frac{0.28744565}{\sqrt{1}}\right] = 0.87$$

Συγκρίνοντας, λοιπόν, αυτή την τιμή του cut-off ορίου με το λόγο EI/BMR που έχουμε βρει για κάθε άτομο, βρίσκουμε ότι το ποσοστό των ατόμων στο σύνολο του δείγματος που αναφέρουν υποεκτιμημένες τιμές πρόσληψης είναι 26%. Μεταξύ των δύο φύλλων έχουμε 18.84% άνδρες και 32.10% γυναίκες που υποεκτιμούν τις τιμές πρόσληψης:

Πίνακας 5.2

	Άνδρες	Γυναίκες	Σύνολο
<b>Ποσοστά υποεκτίμησης</b>	18.84%	32.10%	26%

➤ 2<sup>η</sup> μέθοδος

Η χρήση της τιμής 1.55 για το δείκτη φυσικής δραστηριότητας, όπως έχουμε επισημάνει επανειλημμένως, βασίζεται στην παραδοχή ότι τα άτομα που συμμετέχουν σε κάθε μελέτη έχουν ελαφριά φυσική δραστηριότητα. Στην ενότητα 4.4.1 έχουμε ορίσει λεπτομερώς μία μέθοδο<sup>20</sup> ώστε να μπορέσουμε να προσδιορίσουμε πότε μια δραστηριότητα είναι ελαφρά, πότε μέτρια και πότε υψηλή. Στηριζόμενοι λοιπόν σε αυτή τη μέθοδο και τον Πίνακα 4.2 κατατάσσουμε τους 300 συμμετέχοντες στις 3 κατηγορίες του επιπέδου δραστηριότητας. Από όλους τους συμμετέχοντες, 222 έχουν κατηγοριοποιηθεί με ελαφριά δραστηριότητα (†), 55 έχουν κατηγοριοποιηθεί με μέτρια δραστηριότητα (‡) και 22 με υψηλή δραστηριότητα (§). (έχουμε μία ελλείπουσα τιμή)

Πίνακας 5.3

Ψυχαγωγική δραστηριότητα (ποδηλ.+φυσική άσκηση)	Επαγγελματική δραστηριότητα (Απασχόληση)			
	Καθιστική	Όρθια	Χειρονακτική	Βαριά Χειρονακτική
0	137†	18†	27‡	2§
0-3.5	67†	15‡	9§	-§
3.5-7	13‡	5§	-§	-§
>7	4§	1§	-§	1§

Έτσι, λοιπόν από τη σχέση (5.1) και δίνοντας κάθε φορά την τιμή του δείκτη φυσικής δραστηριότητας που πρέπει σε σχέση με το επίπεδο φυσικής δραστηριότητας και το φύλλο (βλέπε Πίνακα 4.1), βρίσκουμε το αντίστοιχο κάθε φορά cut-off όριο :



Άνδρες ⇒

$1.55 * \exp(-2 * (0.28744565 / \sqrt{1})) = 0.87228736$  → Ελαφριά δραστηριότητα

$1.78 * \exp(-2 * (0.28744565 / \sqrt{1})) = 1.0017235$  → Μέτρια δραστηριότητα

$2.1 * \exp(-2 * (0.28744565 / \sqrt{1})) = 1.1818087$  → Βαριά δραστηριότητα

Γυναίκες ⇒

$1.56 * \exp(-2 * (0.28744565 / \sqrt{1})) = 0.87791502$  → Ελαφριά δραστηριότητα

$1.64 * \exp(-2 * (0.28744565 / \sqrt{1})) = 0.9229363$  → Μέτρια δραστηριότητα

$1.82 * \exp(-2 * (0.28744565 / \sqrt{1})) = 1.0242342$  → Βαριά δραστηριότητα

Συγκρίνοντας, λοιπόν, την αντίστοιχη κάθε φορά τιμή του cut-off ορίου με το λόγο EI/BMR που έχουμε βρει για κάθε άτομο, βρίσκουμε ότι το ποσοστό των ατόμων στο σύνολο του δείγματος που αναφέρουν υποεκτιμημένες τιμές πρόσληψης είναι 29%. Μεταξύ των δύο φύλλων έχουμε 23.91% άνδρες και 33.95% γυναίκες που υποεκτιμούν τις τιμές πρόσληψης. Δηλαδή με αυτή τη μέθοδο παρατηρώ μια αύξηση στα ποσοστά πράγμα αναμενόμενο αφού το 1.55 που χρησιμοποιήθηκε στην πρώτη περίπτωση είναι η ελάχιστη τιμή που θα μπορούσε να χρησιμοποιηθεί για τη φυσική δραστηριότητα. Είναι προφανές ότι στη δεύτερη περίπτωση που χρησιμοποιούμε μεγαλύτερες τιμές για το δείκτη φυσικής δραστηριότητας το ποσοστό των ατόμων που προκύπτει να υποεκτιμούν την πρόσληψη είναι μεγαλύτερο.

**Πίνακας 5.4**

	<b>Άνδρες</b>	<b>Γυναίκες</b>	<b>Σύνολο</b>
<b>Ποσοστά υποεκτίμησης</b>	23.91%	33.95%	29%

Από τα αποτελέσματα δημοσιευμένων μελετών εγκυρότητας<sup>18</sup> προκύπτει ότι τα ποσοστά υποεκτίμησης κυμαίνονται μεταξύ 19% έως 50%, πράγμα το οποίο συμβαίνει και στη μελέτη μας.

Θέλοντας να συγκρίνουμε τις δύο αυτές μεθόδους παρουσιάζουμε τον παρακάτω πίνακα:

under_reporters (2 <sup>η</sup> περίπτωση)	under_reporters (1 <sup>η</sup> περίπτωση)		Total
	0	1	
0	212	0	212
1	10	78	88
Total	222	78	300

Δηλαδή μόνο 10 άτομα βρέθηκαν να υποεκτιμούν την πρόσληψη με τη δεύτερη μέθοδο ενώ με την πρώτη μέθοδο δεν ήταν υποεκτιμούντες. Αν βρούμε και το συντελεστή συμφωνίας (**kappa του Cohen**) παρατηρούμε ότι υπάρχει μεγάλη συμφωνία μεταξύ των δύο μεθόδων:

Agreement	Expected Agreement	Kappa	Std. Err.	Z	Prob>Z
<b>96.67%</b>	59.92%	0.9168	0.0575	15.94	0.0000

Αφού παρατηρούμε ότι δεν υπάρχει σημαντική διαφορά στις δύο μεθόδους, για την περαιτέρω ανάλυσή μας, στα επόμενα κεφάλαια θα χρησιμοποιήσουμε τη δεύτερη μέθοδο προσδιορισμού των υποεκτιμούντων.

### 5.3.2 Απλή στατιστική συμπερασματολογία.

Με τη στατιστική συμπερασματολογία βγάζουμε συμπεράσματα για κάποιες μεταβλητές που μας ενδιαφέρει να εξετάσουμε για να έχουμε μια γενική εικόνα, χρησιμοποιώντας είτε Εκτιμητική είτε Ελέγχους Στατιστικών Υποθέσεων Από τη στιγμή, λοιπόν, που έχουμε βρει τους υποεκτιμούντες θέλουμε να ελέγξουμε ποιες μεταβλητές διαφέρουν σημαντικά στις δύο ομάδες: των υποεκτιμούντων και μη. Αυτό θα γίνει με την εφαρμογή t-tests για τις συνεχείς μεταβλητές που ακολουθούν κανονική κατανομή ή μη παραμετρικών στατιστικών τεστ, για μεταβλητές που δεν ακολουθούν κανονική κατανομή και exact test του Fisher για τις κατηγορικές.

Έτσι παρουσιάζοντας μία-μία τις μεταβλητές που μας ενδιαφέρουν αναλυτικά, έχουμε:

#### ❖ BMI

Θέλουμε να ελέγξουμε αν ο δείκτης μάζας σώματος διαφέρει μεταξύ των δύο ομάδων, των υποεκτιμούντων και των μη. Θα κάνουμε ένα t-test, ελέγχοντας αρχικά τις υποθέσεις της κανονικότητας και των ίσων διακυμάνσεων:

*Κανονικότητα:*

Ελέγχοντας λοιπόν την κανονικότητα για το BMI με τον έλεγχο Kolmogorov-Smirnov για ένα δείγμα, παρατηρούμε ότι η υπόθεση της κανονικότητας δεν απορρίπτεται αφού έχω p-value ίσο με 0.357:

```
One-sample Kolmogorov-Smirnov test against theoretical distribution
normprob(( rbmi-28.49967)/4.237014)
```

Smaller group	D	P-value	Corrected
rbmi:	0.0535	0.180	
Cumulative:	-0.0362	0.455	
Combined K-S:	0.0535	<b>0.357</b>	0.330

*Έλεγχος ισότητας διακυμάνσεων:*

Στη συνέχεια κάνουμε έναν έλεγχο ισότητας διακυμάνσεων στις δύο ομάδες των υποεκτιμούντων και μη ως προς το BMI, όπου παρατηρούμε ότι  $p=0.4549$  δηλαδή δεν υπάρχει σημαντική διαφορά στις διακυμάνσεις:

```
Variance ratio test
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	212	27.97474	.2801788	4.079465	27.42243	28.52705
1	88	29.76429	.4652378	4.364317	28.83958	30.689
combined	300	28.49967	.2446241	4.237014	28.01827	28.98108

Ho: sd(0) = sd(1)

F(211,87) observed = F\_obs = 0.874  
F(211,87) lower tail = F\_L = F\_obs = 0.874  
F(211,87) upper tail = F\_U = 1/F\_obs = 1.145

Ha: sd(0) < sd(1)      Ha: sd(0) != sd(1)      Ha: sd(0) > sd(1)  
P < F\_obs = 0.2178      **P < F\_L + P > F\_U = 0.4549**      P > F\_obs = 0.7822

Επομένως αφού πληρούνται οι παραπάνω προϋποθέσεις μπορούμε να εφαρμόσουμε ένα t-test για να δούμε αν υπάρχει σημαντική διαφορά στο BMI στις δύο ομάδες:

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	212	27.97474	.2801788	4.079465	27.42243	28.52705
1	88	29.76429	.4652378	4.364317	28.83958	30.689
combined	300	28.49967	.2446241	4.237014	28.01827	28.98108
diff		-1.789551	.5281154		-2.828859	-.7502428

Degrees of freedom: 298

Ho: mean(0) - mean(1) = diff = 0

Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
t = -3.3886	t = -3.3886	t = -3.3886
P < t = 0.0004	<b>P &gt;  t  = 0.0008</b>	P > t = 0.9996

Επειδή το p-value είναι πολύ μικρό (=0.0008) σημαίνει ότι δεν αποδεχόμαστε την  $H_0$ , άρα υπάρχει σημαντική διαφορά στο BMI στις δύο ομάδες των υποεκτιμούντων και μη και μάλιστα η ομάδα των υποεκτιμούντων έχει μεγαλύτερο BMI, όπως είναι αναμενόμενο, αφού το BMI είναι δείκτης παχυσαρκίας και γνωρίζουμε από προηγούμενες μελέτες<sup>18,19</sup>, ότι τα παχύσαρκα άτομα υποεκτιμούν σε μεγαλύτερο βαθμό από ότι τα άτομα με φυσιολογικό βάρος.

Για να αποδείξουμε ποιοι υποεκτιμούν περισσότερο την πρόσληψη ενέργειας, και στη δική μας μελέτη, μπορούμε να διακρίνουμε τα άτομα με βάση το δείκτη μάζας του σώματός τους σε 4 κατηγορίες, σύμφωνα με τις οδηγίες του Παγκόσμιου Οργανισμού Υγείας (World Health Organization)<sup>4</sup>:

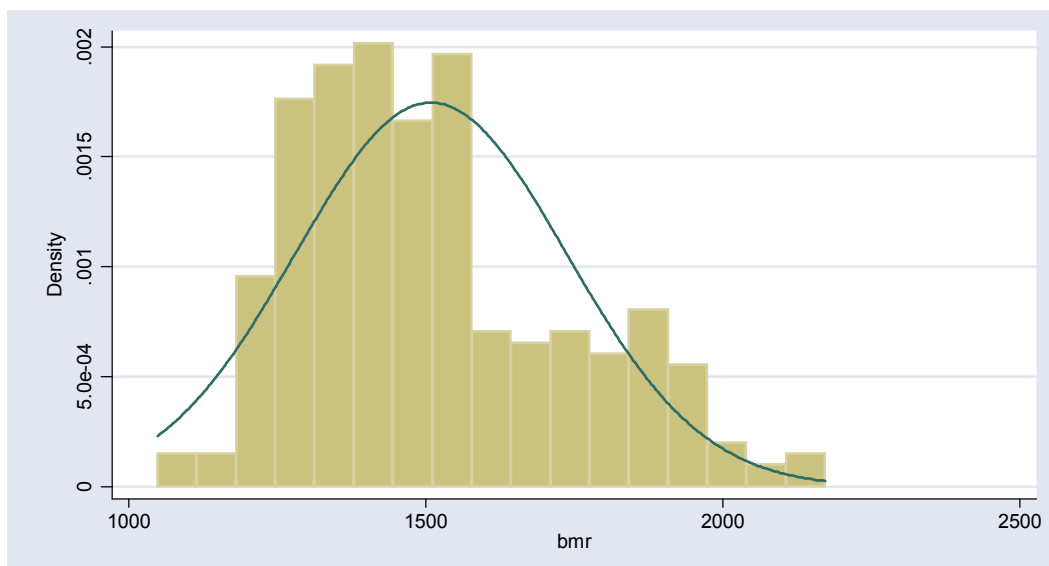
- (1) άτομα με  $BMI < 18.5 \text{ kg/m}^2$  έχουν βάρος λιγότερο του φυσιολογικού,
- (2) άτομα με  $18.5 \text{ kg/m}^2 < BMI < 24.99 \text{ kg/m}^2$  έχουν φυσιολογικό βάρος,
- (3) άτομα με  $25 \text{ kg/m}^2 < BMI < 29.99 \text{ kg/m}^2$  θεωρούνται υπέρβαρα και τέλος
- (4) άτομα με  $BMI \geq 30 \text{ kg/m}^2$  είναι παχύσαρκα.

Κάνοντας αυτή τη διάκριση εντοπίζουμε σημαντικές διαφορές στην υποεκτίμηση της πρόσληψης. Παρατηρούμε ότι τα υπέρβαρα άτομα ανέφεραν τιμές πρόσληψης μικρότερες του cut-off ορίου σε μεγαλύτερο ποσοστό από ότι τα άτομα με φυσιολογικό βάρος, ενώ τα παχύσαρκα σε σχεδόν διπλάσιο ποσοστό από ότι τα άτομα με φυσιολογικό βάρος:

bmi_cat	under_reporters		Total
	0	1	
2	56	15	71
	78.87	21.13	100.00
3	89	30	119
	74.79	25.21	100.00
4	67	43	110
	60.91	39.09	100.00
Total	212	88	300
	70.67	29.33	100.00

#### ❖ BMR

Θέλουμε να ελέγξουμε αν το BMR διαφέρει μεταξύ των δύο ομάδων, των υποεκτιμούντων και των μη. Για να κάνουμε t-test, ελέγχουμε αρχικά την υπόθεση της κανονικότητας με τον έλεγχο Kolmogorov-Smirnov για ένα δείγμα, όπου λόγω του ότι έχω  $p\text{-value}=10^{-3}$  σημαίνει ότι η μεταβλητή BMR δεν ακολουθεί κανονική κατανομή. Δίνοντας το παρακάτω ιστόγραμμα μπορούμε να διαπιστώσουμε και οπτικά την ασυμμετρία της κατανομής:



Επομένως θα εφαρμόσουμε ένα μη παραμετρικό στατιστικό τεστ, το τεστ των Wilcoxon-Mann-Whitney:

```

Two-sample Wilcoxon rank-sum (Mann-Whitney)
test

under_repos |      obs      rank sum      expected
-----+-----
           0 |      212      31931.5      31906
           1 |       88      13218.5      13244
-----+-----
      combined |      300      45150      45150

unadjusted variance      467954.67
adjustment for ties           -1.25
-----
adjusted variance      467953.42

Ho: bmr (under_~s==0) = bmr (under_~s==1)
      z =      0.037
      Prob > |z| =      0.9703

```

Παρατηρούμε ότι δεν απορρίπτουμε την  $H_0$ , επειδή  $p\text{-value}=0.9703$ , άρα δεν υπάρχει σημαντική διαφορά στο BMR στις δύο ομάδες των υποεκτιμούντων και μη.

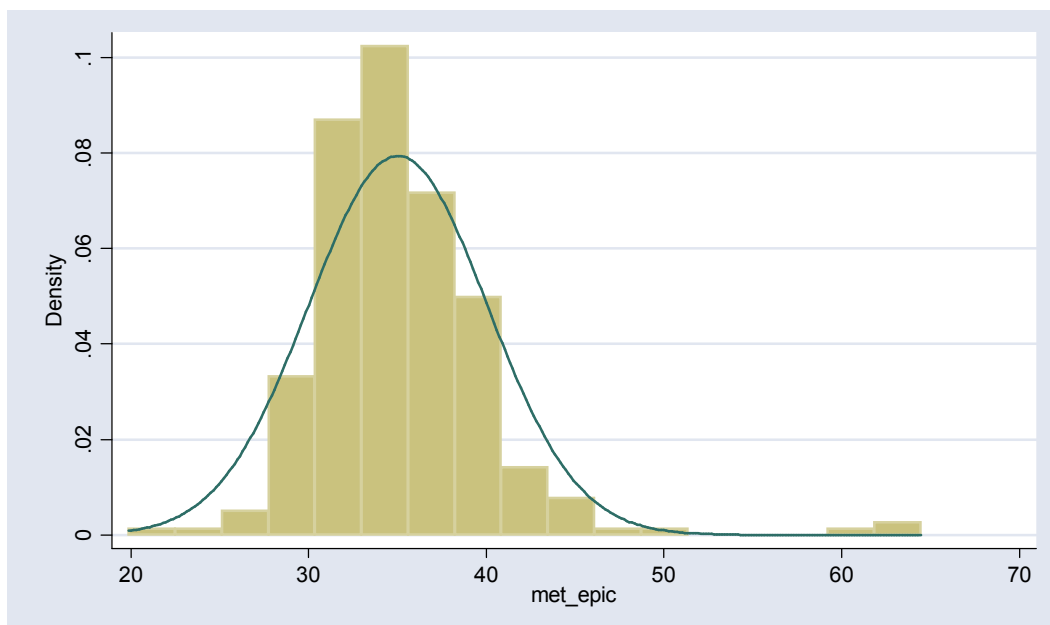
#### ❖ *ΕΝΕΡΓΕΙΑ*

Ομοίως με το BMR θέλουμε να ελέγξουμε αν η ενέργεια διαφέρει μεταξύ των δύο ομάδων. Ελέγχοντας την υπόθεση της κανονικότητας με τον έλεγχο Kolmogorov-Smirnov για ένα δείγμα, βρίσκουμε ότι  $p\text{-value}=0.012$  που σημαίνει ότι η ενέργεια δεν ακολουθεί κανονική κατανομή. Έτσι εφαρμόζοντας το μη παραμετρικό στατιστικό τεστ των Wilcoxon-Mann-Whitney βρίσκουμε  $p\text{-value}=0.9703$ , άρα δεν υπάρχει σημαντική διαφορά στην ενέργεια στις δύο ομάδες των υποεκτιμούντων και μη.

#### ❖ *MET*

Θέλοντας να ελέγξουμε αν ο δείκτης 24ωρης δραστηριότητας (εκφρασμένος σε MET) διαφέρει μεταξύ των δύο ομάδων, εφαρμόζουμε το μη παραμετρικό στατιστικό τεστ των Wilcoxon-Mann-Whitney, αφού ελέγχοντας την υπόθεση της κανονικότητας με το One-sample Kolmogorov-Smirnov test, βρίσκω ότι  $p\text{-value}=0.014$  που σημαίνει ότι υπάρχει

ισχυρή ένδειξη πως η μεταβλητή μου δεν ακολουθεί κανονική κατανομή όπως φαίνεται και από το παρακάτω ιστόγραμμα:



Έτσι από το μη παραμετρικό αυτό τεστ έχω  $p\text{-value}=0.7363$ , άρα δεν υπάρχει σημαντική διαφορά στα MET στις δύο ομάδες.

#### ❖ ΦΥΛΛΟ

Θέλουμε να συγκρίνω αν τα δύο φύλλα υποεκτιμούν στον ίδιο βαθμό την πρόσληψη. Κάνοντας ένα Exact test του Fisher, επειδή έχουμε κατηγορικές μεταβλητές, παρατηρούμε ότι σε επίπεδο σημαντικότητας 0.05 δεν απορρίπτω την  $H_0$  δηλαδή δεν υπάρχει διαφορά στο φύλλο στις δύο ομάδες αφού έχω  $p\text{-value} = 0.075$ . Βλέπουμε ότι σε αντίθεση με το δείκτη μάζας σώματος, όπου είναι ξεκάθαρη η διαφορά της υποεκτίμησης στα άτομα με φυσιολογικό βάρος, στα υπέρβαρα και τα παχύσαρκα (πράγμα το οποίο έχει αποδειχθεί και από μια σειρά πειραματικών μελετών), η διαφορά υποεκτίμησης στα δύο φύλλα δεν είναι ξεκάθαρη αν και σε επίπεδο 5% η υποεκτίμηση της πρόσληψης δεν διαφοροποιήθηκε σημαντικά στα δύο φύλλα (γεγονός που εντοπίστηκε και κατά την αξιολόγηση 37 διατροφικών ερευνών στη μελέτη Black et al).

sex	under_reporters		Total
	0	1	
male	105	33	138
	76.09	23.91	100.00
female	107	55	162
	66.05	33.95	100.00
Total	212	88	300
	70.67	29.33	100.00
Pearson chi2(1) = 3.6220			Pr = 0.057
Fisher's exact =			<b>0.075</b>
1-sided Fisher's exact =			0.037

#### ❖ WHR

Αρχικά διακρίνουμε τα άτομα με βάση το λόγο περιμέτρου μέσης προς περίμετρο ισχύος σε δύο κατηγορίες ανά φύλλο, όπου μας δίνει ένα δείκτη για κεντρική παχυσαρκία<sup>4</sup>: για τους άνδρες το οριακό σημείο είναι το 0.95 και για τις γυναίκες το 0.80. Έτσι κατηγοριοποιούμε: με 1⇒τους άνδρες με  $WHR \leq 0.95$ , με 2⇒τους άνδρες με  $WHR > 0.95$ , με 3⇒τις γυναίκες με  $WHR \leq 0.80$  και με 4⇒τις γυναίκες με  $WHR > 0.80$ . Κάνοντας αυτή τη διάκριση παρατηρούμε την κατανομή των συμμετεχόντων σε αυτές τις κατηγορίες (έχουμε και 4 ελλείπουσες τιμές):

WHR_cat	under_reporters		Total
	0	1	
1	48	11	59
	81.36	18.64	100.00
2	55	21	76
	72.37	27.63	100.00
3	51	15	66
	77.27	22.73	100.00
4	55	40	95
	57.89	42.11	100.00
Total	209	87	296
	70.61	29.39	100.00
Pearson chi2(3) = 12.2089			Pr = 0.007
Fisher's exact =			<b>0.008</b>

Για να δούμε αν υπάρχει διαφορά στην υποεκτίμηση της πρόσληψης ενέργειας, σε αυτές τις 4 υποομάδες κάνουμε έναν ακριβή έλεγχο του Fisher επειδή έχουμε κατηγορικές μεταβλητές και παρατηρούμε ότι δεν αποδεχόμαστε την  $H_0$ , δηλαδή υπάρχει διαφορά στην υποεκτίμηση στις ομάδες του WHR, αφού έχω  $p\text{-value} = 0.008$ .



Μάλιστα παρατηρούμε ότι τα άτομα με παχυσαρκία υποεκτιμούν σε μεγαλύτερο ποσοστό από τα άτομα με φυσιολογικό βάρος (27.63% για τους άνδρες και 42.11% για τις γυναίκες).

❖ *HAIKIA*

Αρχικά παρατηρούμε την κατανομή των υποεκτιμούντων και μη, στις διάφορες ηλικιακές ομάδες:

age5	under_reporters		Total
	0	1	
24-34	5	2	7
	71.43	28.57	100.00
35-44	28	5	33
	84.85	15.15	100.00
45-54	55	14	69
	79.71	20.29	100.00
55-64	54	33	87
	62.07	37.93	100.00
65+	70	34	104
	67.31	32.69	100.00
Total	212	88	300
	70.67	29.33	100.00

Pearson chi2(4) = 9.5947 Pr = 0.048  
Fisher's exact = **0.041**

Από τον πίνακα γίνεται εμφανές ότι η υποεκτίμηση της πρόσληψης είναι εντονότερη στις μεγαλύτερες ηλικιακά ομάδες με πιο έντονη την ηλικία 55-64 με ποσοστό 38% και ακολουθεί η ηλικιακή ομάδα 65+ με ποσοστό 33%. Το μικρότερο ποσοστό (15%) αντιστοιχεί στην ηλικιακή ομάδα 35-44

Θέλοντας να δούμε αν οι ηλικιακές ομάδες υποεκτιμούν στον ίδιο βαθμό την πρόσληψη κάνω ένα Exact test του Fisher και παρατηρούμε ότι απορρίπτουμε την  $H_0$  δηλαδή υπάρχει διαφορά στην υποεκτίμηση στις ηλικιακές ομάδες αφού έχω  $p\text{-value} = 0.041$ .

❖ *ΟΙΚΟΓΕΝΕΙΑΚΗ ΚΑΤΑΣΤΑΣΗ*

Αρχικά παρατηρούμε την κατανομή των υποεκτιμούντων και μη, στις 4 ομάδες της οικογενειακής κατάστασης (έχουμε μία ελλείπουσα τιμή):

mar_stat	under_reporters		Total
	0	1	
single	11	3	14
	78.57	21.43	100.00
married/living togeth	179	75	254
	70.47	29.53	100.00
divorced/seperated	9	1	10
	90.00	10.00	100.00
widowed	12	9	21
	57.14	42.86	100.00
Total	211	88	299
	70.57	29.43	100.00
Pearson chi2(3) = 4.0733 Pr = 0.254			
Fisher's exact = <b>0.286</b>			
Fisher's exact = 0.286			

Θέλοντας να δούμε αν οι ηλικιακές ομάδες υποεκτιμούν στον ίδιο βαθμό την πρόσληψη κάνουμε ένα Exact test του Fisher και παρατηρούμε ότι δεν απορρίπτουμε την  $H_0$  δηλαδή δεν υπάρχει διαφορά στην υποεκτίμηση στις 4 ομάδες της οικογενειακής κατάστασης αφού έχω  $p\text{-value}=0.286$ .

❖ *ΕΠΙΠΕΔΟ ΜΟΡΦΩΣΗΣ*

Αρχικά παρατηρούμε την κατανομή των υποεκτιμούντων και μη, στα τέσσερα επίπεδα μόρφωσης (έχουμε δύο ελλείπουσες τιμές):

edu_epic	under_reporters		Total
	0	1	
Low, <=6 years	110	61	171
	64.33	35.67	100.00
Middle, 7-14 years	55	13	68
	80.88	19.12	100.00
Upper, 15+ years	45	14	59
	76.27	23.73	100.00
Total	210	88	298
	70.47	29.53	100.00
Pearson chi2(2) = 7.5973 Pr = 0.022			
Fisher's exact = 0.023			

Βλέπουμε ότι το μεγαλύτερο ποσοστό υποεκτίμησης (36%) παρατηρείται στους συμμετέχοντες με μόρφωση μικρότερη των έξι χρόνων.

Θέλοντας να δούμε αν οι τα άτομα με χαμηλή, μεσαία και ανώτερη μόρφωση υποεκτιμούν στον ίδιο βαθμό την πρόσληψη κάνω ένα Exact test του Fisher και παρατηρούμε ότι απορρίπτουμε την  $H_0$ , δηλαδή υπάρχει διαφορά στην υποεκτίμηση στις διάφορες κατηγορίες μόρφωσης αφού έχω  $p\text{-value} = 0.023$ .

#### ❖ ΚΑΤΑΣΤΑΣΗ ΚΑΠΝΙΣΜΑΤΟΣ

Κατατάσσουμε τους υποεκτιμούντες και μη, στις τρεις κατηγορίες: σε αυτούς που δεν έχουν καπνίσει ποτέ, στους πρώην καπνιστές και στους νυν (έχουμε δύο ελλείπουσες τιμές):

smo_epic	under_reporters		Total
	0	1	
Never smoked	113	56	169
	66.86	33.14	100.00
Past smokers	45	19	64
	70.31	29.69	100.00
Current smokers	52	13	65
	80.00	20.00	100.00
Total	210	88	298
	70.47	29.53	100.00
Pearson chi2(2) = 3.8936 Pr = 0.143			
Fisher's exact = 0.140			

Θέλοντας να δουμε αν τα άτομα των τριών αυτών κατηγοριών υποεκτιμούν στον ίδιο βαθμό την πρόσληψη κάνουμε ένα Exact test του Fisher επειδή έχουμε κατηγορικές μεταβλητές και παρατηρούμε ότι δεν απορρίπτουμε την  $H_0$  δηλαδή δεν υπάρχει διαφορά στην υποεκτίμηση στις διάφορες κατηγορίες καπνίσματος αφού έχω  $p\text{-value} = 0.140$ .

❖ *ΑΠΑΣΧΟΛΗΣΗ*

Αρχικά παρατηρούμε την κατανομή των υποεκτιμούντων και μη, στις 4 κατηγορίες της απασχόλησης (έχουμε μία ελλείπουσα τιμή):

occupy	under_reporters		Total
	0	1	
no/sedentary work	157	64	221
	71.04	28.96	100.00
standing work	26	13	39
	66.67	33.33	100.00
manual work	26	10	36
	72.22	27.78	100.00
heavy manual work	2	1	3
	66.67	33.33	100.00
Total	211	88	299
	70.57	29.43	100.00
Pearson chi2(3) = 0.3790 Pr = 0.945			
Fisher's exact = <b>0.917</b>			

Θέλοντας να δούμε αν τα άτομα των τεσσάρων αυτών κατηγοριών υποεκτιμούν στον ίδιο βαθμό την πρόσληψη κάνω ένα Exact test του Fisher και παρατηρούμε ότι δεν απορρίπτουμε την  $H_0$  δηλαδή δεν υπάρχει διαφορά στην υποεκτίμηση στις τέσσερις κατηγορίες απασχόλησης αφού έχω  $p\text{-value} = 0.917$ .

❖ ΠΕΡΙΦΕΡΕΙΑ

Στην αρχή βλέπουμε την κατανομή των υποεκτιμούντων και μη, στις περιφέρειες της Αττικής, της κεντρικής Ελλάδας, της Βόρειας Ελλάδας, της Νότιας Ελλάδας και στα νησιά του Αιγαίου (έχουμε δύο ελλείπουσες τιμές):

prf_epic	under_reporters		Total
	0	1	
<b>Attica</b>	72	31	103
	69.90	30.10	100.00
<b>Central Greece</b>	40	16	56
	71.43	28.57	100.00
<b>North Greece</b>	35	24	59
	59.32	40.68	100.00
<b>South Greece</b>	31	10	41
	75.61	24.39	100.00
<b>Aegean Islands</b>	32	7	39
	82.05	17.95	100.00
Total	210	88	298
	70.47	29.53	100.00
Pearson chi2(4) = 6.5983			Pr = 0.159
Fisher's exact =			<b>0.170</b>

Θέλοντας να δούμε αν οι περιφέρειες υποεκτιμούν στον ίδιο βαθμό την πρόσληψη κάνουμε ένα Exact test του Fisher επειδή έχουμε κατηγορικές μεταβλητές και παρατηρούμε ότι δεν απορρίπτουμε την  $H_0$  δηλαδή δεν υπάρχει διαφορά στην υποεκτίμηση στις περιφέρειες αφού έχω  $p\text{-value} = 0.170$ .

❖ ΠΕΡΙΟΧΗ

Αρχικά παρατηρούμε την κατανομή των υποεκτιμούντων και μη, στις αστικές και αγροτικές περιοχές (έχουμε δύο ελλείπουσες τιμές):

urb_epic	under_reporters		Total
	0	1	
Urban	132	59	191
	69.11	30.89	100.00
Rural	78	29	107
	72.90	27.10	100.00
Total	210	88	298
	70.47	29.53	100.00
Pearson chi2(1) =			0.4727 Pr = 0.492
Fisher's exact =			<b>0.511</b>
1-sided Fisher's exact =			0.291

Θέλοντας να δούμε αν οι περιοχές αυτές υποεκτιμούν στον ίδιο βαθμό την πρόσληψη κάνουμε ένα Exact test του Fisher επειδή έχουμε κατηγορικές μεταβλητές και παρατηρούμε ότι δεν απορρίπτουμε την  $H_0$  δηλαδή δεν υπάρχει διαφορά στην υποεκτίμηση στις αγροτικές και αστικές περιοχές αφού έχω  $p\text{-value} = 0.511$ .

#### ❖ ΠΡΟΣΛΗΨΗ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΙΑΤΡΟΦΙΚΩΝ ΣΥΣΤΑΤΙΚΩΝ

Στον παρακάτω πίνακα αναφέρονται οι τιμές της μέσης ημερήσιας πρόσληψης για τις πρωτεΐνες, τους υδατάνθρακες τα λιπίδια και τις αλκοόλες, τόσο για τους άνδρες όσο και για τις γυναίκες. Οι μέσες αυτές τιμές προσδιορίζονται πριν και μετά των αποκλεισμό των υποεκτιμούντων.

	Σύνολο		Αποκλεισμός υποεκτιμούντων		Υποεκτιμούντες	
	Άνδρες	Γυναίκες	Άνδρες	Γυναίκες	Άνδρες	Γυναίκες
<b>Διατροφικά συστατικά</b>						
<b>Πρωτεΐνες</b>	79.98	56.33	89.72	67.47	48.97	34.65
<b>Υδατάνθρακες</b>	201.59	159.09	223.25	184.77	132.7	109.14
<b>Λιπίδια</b>	103.40	76.64	116.46	93.37	61.88	44.1
<b>Αλκοόλες</b>	23.18	3.26	27.72	4.41	8.73	1.02

Παρατηρούμε ότι όλες οι τιμές παρουσιάζουν αύξηση 10-20% όταν αφαιρέσουμε τα δεδομένα από τους υποεκτιμούντες. Αυτό φαίνεται και αν εφαρμόσουμε, για κάθε ένα από τα διατροφικά χαρακτηριστικά χωριστά, το μη παραμετρικό στατιστικό τεστ των Wilcoxon-Mann-Whitney οπότε βρίσκουμε και για τις πρωτεΐνες και για τους υδατάνθρακες άλλα και τα λιπίδια και τις αλκοόλες  $p\text{-value}=10^{-3}$ , που σημαίνει ότι υπάρχει σημαντική διαφορά στα διατροφικά χαρακτηριστικά για τις δύο ομάδες των υποεκτιμούντων και μη (εφαρμόζουμε μη παραμετρικά τεστ γιατί κάνοντας έλεγχο κανονικότητας με τον έλεγχο Kolmogorov-Smirnov, έχουμε πολύ μικρά  $p\text{-value}\approx 0.012$  που σημαίνει ότι δεν ακολουθούν κανονική κατανομή).

### 5.3.3 Λογιστική παλινδρόμηση.

Θέλουμε να βρούμε ένα μοντέλο ώστε να μπορέσουμε να προβλέψουμε την πιθανότητα ένα άτομο να ανήκει στην ομάδα των υποεκτιμούντων ή μη. Δηλαδή θέλουμε να βρούμε ένα κατάλληλο μοντέλο έχοντας την μεταβλητή των υποεκτιμούντων (κατηγορική) σαν εξαρτημένη και τις άλλες μεταβλητές της ανάλυσης μου σαν ανεξάρτητες.

Επειδή η εξαρτημένη μου μεταβλητή είναι δίτιμη, μπορούμε να εφαρμόσουμε ένα μοντέλο λογιστικής παλινδρόμησης (logit) και να ελέγξουμε ποιες μεταβλητές είναι σημαντικές και ποιες όχι (logistic regression). Σε ένα τέτοιο μοντέλο, η ανερμήνευτη μεταβλητότητα των τιμών της εξαρτημένης μεταβλητής μετριέται από την απόκλιση των καταλοιπών (residual deviance). Συγκρίνοντας κάθε φορά τη μεταβολή αυτής της απόκλισης (deviance residual), που προκύπτει από την αφαίρεση μιας μεταβλητής, με τις τιμές της κατανομής  $\chi^2$  με τους αντίστοιχους βαθμούς ελευθερίας, βρίσκουμε αν η συγκεκριμένη μεταβλητή είναι σημαντική ή όχι. Το deviance residual μας δείχνει τη συμβολή κάθε μεταβλητής στην συνολική μεταβλητότητα των  $X_i$ . Όσο μεγαλύτερο είναι αυτό, τόσο μεγαλύτερη είναι η συμβολή της μεταβλητής και θα πρέπει να παραμείνει στο μοντέλο μου.

Έτσι από τον τύπο του logit μοντέλου θα έχω:

$$\text{logit}(\pi(i)) = \log(\pi(i)/1-\pi(i)) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots$$

όπου για συγκεκριμένες τιμές των  $\beta_0, \beta_1, \beta_2, \beta_3, \dots$  που φαίνονται από τον πίνακα των coefficients και για συγκεκριμένες τιμές των  $X_{1i}, X_{2i}, X_{3i}, \dots$  μπορούμε να εκτιμήσουμε (προβλέψουμε) την πιθανότητα,  $\pi(i)$ , να ανήκει ένα άτομο στην ομάδα των υποεκτιμούντων.

Οι συντελεστές  $\beta_0, \beta_1, \beta_2, \beta_3, \dots$  μας δείχνουν πως επηρεάζει κάθε μεταβλητή  $X_{1i}, X_{2i}, X_{3i}, \dots$  την πιθανότητα  $\pi(i)$ . (Ο λόγος  $\pi(i)/1-\pi(i)$  είναι γνωστός ως σχετική πιθανότητα, odds).

Αρχικά δημιουργούμε ένα μοντέλο με εξαρτημένη μεταβλητή τους υποεκτιμούντες και ανεξάρτητες μεταβλητές την οικογενειακή κατάσταση, την ηλικία, την περιφέρεια, την περιοχή, τα MET, το κάπνισμα, τη μόρφωση, την απασχόληση, το επίπεδο του WHR και το επίπεδο του BMI και εφαρμόζουμε ένα μοντέλο λογιστικής παλινδρόμησης (εδώ πρέπει να σημειώσουμε ότι επειδή έχουμε 5 ελλείπουσες τιμές αφαιρούμε αυτά τα άτομα από την ανάλυσή μας).

Από τον πίνακα ANOVA που προκύπτει, βλέπουμε ποιες μεταβλητές δεν είναι σημαντικές, συγκρίνοντας το deviance residual της κάθε μεταβλητής με το  $\chi^2$  με τους αντίστοιχους βαθμούς ελευθερίας. Αν το deviance residual δεν είναι μεγαλύτερο από το  $\chi^2$  με τους αντίστοιχους βαθμούς ελευθερίας, σημαίνει ότι η μεταβλητή μου δεν είναι σημαντική και την αφαιρώ. Εφαρμόζοντας Stepwise Regression (και μάλιστα ξεκινώντας με το πλήρες μοντέλο), καταλήγουμε σε ένα μοντέλο με σημαντικές μεταβλητές την ηλικία και το επίπεδο του BMI (σε επίπεδο σημαντικότητας 10%). Ελέγχοντας για το αν η αλληλεπίδραση των δύο αυτών μεταβλητών είναι σημαντική παρατηρώ πως δεν είναι και επομένως έχω:

	DF	DEVIANCE RES.	DF	RESID.DEVIANCE
NULL			294	357.8316
age5	4	9.45227	290	348.3793
bmi cat	2	6.28748	288	342.0919

Για να δούμε αν το μοντέλο μας είναι ικανοποιητικό συγκρίνω το residual deviance του μοντέλου μας με το με το  $\chi^2_{n-p}$  όπου  $n$ = ο αριθμός των παρατηρήσεων και  $p$ =οι παράμετροι μαζί με τον σταθερό όρο. Αν το residual deviance του μοντέλου μας είναι μεγαλύτερο από το  $\chi^2$  με τους αντίστοιχους βαθμούς ελευθερίας, σημαίνει ότι το μοντέλο μας δεν είναι ικανοποιητικό. Για το συγκεκριμένο μοντέλο έχω βρει residual deviance=342.1 το οποίο είναι μεγαλύτερο από το  $\chi^2_{0.1,288}=319.15$ , άρα το μοντέλο μας δεν παρουσιάζει καλή προσαρμογή. Αυτό προκύπτει και αν βρούμε το p-value του μοντέλου μας που είναι πολύ μεγάλο, δηλαδή 0.98 που σημαίνει ότι το μοντέλο μας δεν είναι καλό.

Έτσι, λοιπόν, δημιουργούμε ένα νέο μοντέλο με τις ίδιες ανεξάρτητες μεταβλητές αλλά προσθέτοντας και τα διατροφικά συστατικά, δηλαδή εξαρτημένη μεταβλητή τους υποεκτιμούντες και ανεξάρτητες μεταβλητές την οικογενειακή κατάσταση, τις πρωτεΐνες,



τους υδατάνθρακες, τα λιπίδια, τις αλκοόλες, την ηλικία, την περιφέρεια, την περιοχή, τα MET, το κάπνισμα, τη μόρφωση, την απασχόληση, το επίπεδο του WHR και το επίπεδο του BMI και εφαρμόζουμε λογιστική παλινδρόμηση:

	<b>DF</b>	<b>DEVIANCE RES.</b>	<b>DF</b>	<b>RESID.DEVIANCE</b>
<i>NULL</i>			294	357.8316
<i>mar_stat</i>	3	4.381092	291	353.4505
<i>prot</i>	1	112.5446	290	240.9059
<i>carb</i>	1	35.0307	289	205.8752
<i>lipi</i>	1	62.6656	288	143.2096
<i>alco</i>	1	1.3261	287	141.8835
<i>age5</i>	4	1.5143	283	140.3691
<i>prf_epic</i>	4	6.9606	279	133.4086
<i>urb_epic</i>	1	1.6369	278	131.7716
<i>met_epic</i>	1	1.6278	277	130.1438
<i>smo_epic</i>	2	2.7885	275	127.3553
<i>edu_epic</i>	2	1.694	273	125.6613
<i>occupy</i>	3	11.0303	270	114.631
<i>bmi_cat</i>	2	17.8314	268	96.7995
<i>whr_cat</i>	3	28.0384	265	68.7611

Από τον παραπάνω πίνακα, παρατηρούμε ποιες μεταβλητές δεν είναι σημαντικές, και με Stepwise Regression, όπως παραπάνω, καταλήγουμε σε ένα μοντέλο με σημαντικές μεταβλητές τις πρωτεΐνες, τους υδατάνθρακες, τα λιπίδια, την απασχόληση, το επίπεδο του BMI και το επίπεδο του WHR. Ελέγχοντας και τις αλληλεπιδράσεις των έξι αυτών μεταβλητών παρατηρούμε ότι καμία δεν είναι στατιστικά σημαντική. Έτσι το τελικό μας μοντέλο αποτελείται από τις παρακάτω μεταβλητές:

	<b>DF</b>	<b>DEVIANCE RES.</b>	<b>DF</b>	<b>RESID.DEVIANCE</b>
<i>NULL</i>			294	357.8316
<i>prot</i>	1	115.4284	293	242.4032
<i>carb</i>	1	34.5132	292	207.89
<i>lipi</i>	1	61.0832	291	146.8068
<i>occupy</i>	3	11.2229	288	135.5839
<i>bmi_cat</i>	2	7.7156	286	127.8683
<i>whr_cat</i>	3	17.4302	283	110.4381

Δηλαδή :

για τη μεταβλητή πρωτεΐνες έχω deviance residual=115.4284 το οποίο είναι μεγαλύτερο από το  $\chi^2_{0.05,1}=3.841$

για τη μεταβλητή υδατάνθρακες έχω deviance residual =34.5132 το οποίο είναι μεγαλύτερο από το  $\chi^2_{0.05,1}=3.841$

για τη μεταβλητή λιπίδια έχω deviance residual =61.0832 το οποίο είναι μεγαλύτερο από το  $\chi^2_{0.05,1}=3.841$

για τη μεταβλητή απασχόληση έχω deviance residual =11.2229 το οποίο είναι μεγαλύτερο από το  $\chi^2_{0.05,3}=7.815$

για τη μεταβλητή που δείχνει το επίπεδο του BMI έχω deviance residual =7.7156 το οποίο είναι μεγαλύτερο από το  $\chi^2_{0.05,2}=5.991$

για τη μεταβλητή που δείχνει το επίπεδο του WHR έχω deviance residual =17.4302 το οποίο είναι μεγαλύτερο από το  $\chi^2_{0.05,3}=7.815$

Για το συγκεκριμένο μοντέλο έχω βρει residual deviance=110.4381 το οποίο είναι μικρότερο από το  $\chi^2_{0.05,283}=323.2359$ , άρα το μοντέλο μας παρουσιάζει καλή προσαρμογή, το οποίο προκύπτει και από το p-value του μοντέλου μας που είναι πολύ μικρό, δηλαδή p-value  $\approx 0$  που σημαίνει ότι το μοντέλο μας είναι καλό. Κάνοντας έλεγχο για την κανονικότητα των Pearson Residuals με τον έλεγχο Kolmogorov-Smirnov για ένα δείγμα, βρίσκω p-value= $10^{-3}$ , δηλαδή απορρίπτεται η υπόθεση της κανονικότητας των Pearson Residuals.

Από τον πίνακα των coefficients φαίνονται οι συντελεστές  $\beta_i$  :

	Value	Std. Error	t value
<i>Intercept</i>	19.18989509	7.44797757	2.5765243
<i>prot</i>	-0.05607821	0.01606617	-3.4904528
<i>carb</i>	-0.03880293	0.00748658	-5.1829973
<i>lipi</i>	-0.10980544	0.01855843	-5.9167429
<i>Occupy manual work</i>	-1.81866903	6.87514346	-0.2645282
<i>Occupy no/sedentary work</i>	-4.18527182	6.86335467	-0.6097997
<i>Occupy standing work</i>	-2.34194274	6.86899708	-0.3409439
<i>bmi.cat3</i>	0.31469388	0.63311429	0.497057
<i>bmi.cat4</i>	1.85004963	0.74252107	2.4915786
<i>whr.cat2</i>	0.24903185	0.78944387	0.3154522
<i>whr.cat3</i>	-2.33962724	0.84395188	-2.7722283
<i>whr.cat4</i>	-2.07730291	0.81041331	-2.5632636

Επομένως το τελικό μου μοντέλο γράφεται:

$$\text{logit}(\pi) = \log\left(\frac{P}{1-P}\right) = 19.19 - 0.056X_1 - 0.039X_2 - 0.11X_3 - 1.82X_4 - 4.19X_5 - 2.34X_6 + 0.31X_7 + 1.85X_8 \\ + 0.25X_9 - 2.34X_{10} - 2.08X_{11}$$

όπου

- $X_1$  είναι οι πρωτεΐνες
- $X_2$  είναι οι υδατάνθρακες
- $X_3$  είναι τα λιπίδια
- $X_4$  είναι η χειρονακτική εργασία
- $X_5$  είναι η καθιστική εργασία
- $X_6$  είναι η όρθια εργασία
- $X_7$  είναι το επίπεδο 3 του BMI
- $X_8$  είναι το επίπεδο 4 του BMI
- $X_9$  είναι το επίπεδο 2 του WHR
- $X_{10}$  είναι το επίπεδο 3 του WHR
- $X_{11}$  είναι το επίπεδο 4 του WHR

### **Παράδειγμα:**

Για να καταλάβουμε τη χρησιμότητα του παραπάνω μοντέλου δίνουμε το παρακάτω παράδειγμα:

Θέλουμε να εκτιμήσουμε την πιθανότητα, ένας άνδρας που έχει καταναλώσει 63gr πρωτεΐνες, 152gr υδατάνθρακες, 80gr λιπίδια, κάνει όρθια εργασία, έχει φυσιολογικό βάρος ( $18.5 \leq \text{BMI} < 25 \text{ kg/m}^2$ ) και έχει το λόγο  $\text{WHR} = 0.89$ , να ανήκει στους υποεκτιμούντες:

$$\text{logit}(\pi) = \log\left(\frac{P}{1-P}\right) = 19.19 - 0.056*63 - 0.039*152 - 0.11*80 - 1.82*0 - 4.19*0 - 2.34*1 + 0.31*0 \\ + 1.85*0 + 0.25*0 - 2.34*0 - 2.08*0 = -1.406 \Rightarrow \\ \pi/1-\pi = 0.245 \Rightarrow \\ \hat{p} = 0.197 \text{ ή } 19.7\%$$

Ο λόγος  $\frac{P}{1-P}$  (odds) είναι η σχετική πιθανότητα ένα άτομο με τα παραπάνω χαρακτηριστικά και το οποίο έχει φυσιολογικό βάρος, να ανήκει στους υποεκτιμούντες.

Στη συνέχεια να υπολογιστεί η ίδια πιθανότητα για έναν υπέρβαρο άνδρα ( $25 \leq \text{BMI} < 30$   $\text{kg/m}^2$ ):

$$\begin{aligned} \text{logit}(\pi_1) &= \log\left(\frac{p_1}{1-p_1}\right) = 19.19 - 0.056 * 63 - 0.039 * 152 - 0.11 * 80 - 1.82 * 0 - 4.19 * 0 - 2.34 * 1 + 0.31 * 1 \\ &\quad + 1.85 * 0 + 0.25 * 0 - 2.34 * 0 - 2.08 * 0 = -1.096 \Rightarrow \\ \pi_1 / 1 - \pi_1 &= 0.334 \Rightarrow \\ \hat{p}_1 &= 0.25 \text{ ή } 25\% \end{aligned}$$

Ο λόγος  $\frac{p_1}{1-p_1}$  (odds) είναι η σχετική πιθανότητα ένα άτομο με τα παραπάνω χαρακτηριστικά και το οποίο είναι υπέρβαρο, να ανήκει στους υποεκτιμούντες.

Για να βρω το odds ratio, διαιρώ τα δύο odds και έχω:

$$\frac{p_1}{1-p_1} / \frac{p}{1-p} = 0.334 / 0.245 = 1.4$$

Δηλαδή ένα άτομο υπέρβαρο έχει 1.4 φορές μεγαλύτερη σχετική πιθανότητα να έχει υποεκτιμήσει την ενεργειακή πρόσληψη από ότι ένα άτομο με φυσιολογικό βάρος.

#### 5.4 Ανάλυση του δείκτη 24ωρης δραστηριότητας όταν εκφράζεται σε MET.

Όπως έχουμε αναφέρει παραπάνω, λεπτομερώς, για κάθε μια από τις δραστηριότητες που αναφέρει ο συμμετέχων ότι πραγματοποιεί, ορίζουμε το μεταβολικό ισοδύναμό της, ο λόγος δηλαδή του ρυθμού του μεταβολισμού κατά τη συγκεκριμένη δραστηριότητα, προς το ρυθμό του μεταβολισμού υπό συνθήκες ανάπαυσης. Οι τιμές των μεταβολικών ισοδυνάμων (MET values) για κάθε δραστηριότητα προκύπτουν από δημοσιευμένους πίνακες.

Η μέτρηση της φυσικής δραστηριότητας μέσω ενός ερωτηματολογίου μπορεί να οδηγήσει στον προσδιορισμό της ενέργειας που συνήθως δαπανά ένα άτομο ανά ημέρα ή ώρα, εκφρασμένη είτε σε μονάδες ενέργειας (kcal ή kJ) είτε σε μεταβολικά ισοδύναμα. Το μεταβολικό ισοδύναμο (metabolic equivalent) ή MET, αντιστοιχεί στην ενέργεια που δαπανά ένα άτομο υπό συνθήκες ανάπαυσης<sup>21</sup>.

Για κάθε άτομο υπολογίστηκε ένας δείκτης 24ωρης φυσικής δραστηριότητας (MET-hours-per-day score). Ο δείκτης αυτός ορίζει τις θερμίδες που δαπανά κάθε άτομο ανά κιλό βάρους του, κατά τη διάρκεια ενός τυπικού 24ώρου. Έτσι συμβολίζοντας με  $h_i$  τις ώρες άσκησης ανά

ημέρα μιας δραστηριότητας  $i$ , με μεταβολικό ισοδύναμο  $MET_i$ , ο δείκτης 24ωρης φυσικής δραστηριότητας ενός ατόμου για ένα σύνολο  $k$  δραστηριοτήτων την ημέρα ορίζεται ως εξής:

$$\text{Δείκτης 24ωρης φυσικής δραστηριότητας} = \sum (MET_i \times h_i) \text{ όπου } i=1,2,\dots,k.$$

#### 5.4.1 Περιγραφικά χαρακτηριστικά για το δείκτη 24ωρης φυσικής δραστηριότητας.

Ο παρακάτω πίνακας δίνει τις μέσες τιμές και τις τυπικές αποκλίσεις του δείκτη 24ωρης φυσικής δραστηριότητας (ημερήσιες δαπάνες ενέργειας εκφρασμένες σε μεταβολικά ισοδύναμα-MET) χωριστά για τους άνδρες και τις γυναίκες άλλα και το σύνολο των συμμετεχόντων:

	Άνδρες		Γυναίκες		Σύνολο	
	Μέση τιμή	Τυπική Απόκλιση	Μέση τιμή	Τυπική Απόκλιση	Μέση τιμή	Τυπική Απόκλιση
<b>Ηλικία</b>						
24-34	38	3.3	32.8	3.7	35.8	4.2
35-44	34.8	5	36.3	3.1	35.5	4.2
45-54	37	4	35.9	2.8	36.2	3.3
55-64	35.4	8.9	35.4	4	35.5	6.6
65+	33.4	4.9	34	3.8	33.7	4.4
<b>Εκπαίδευση</b>						
Χαμηλή	35.1	7.6	35	4.1	35	5.8
Μέση	34.7	5.6	35.5	2.4	35	4.2
Ανώτερη	34.6	3.4	35.3	3.2	34.9	3.3
<b>Περιφέρεια</b>						
Αττική	35.33	5.81	35.89	2.76	35.57	4.75
Κεντρική Ελλάδα	32.36	5.18	34.96	2.73	33.57	4.38
Βόρεια Ελλάδα	37.44	9.73	35.91	4.48	36.43	6.69
Νότια Ελλάδα	35.03	4.40	33.78	4.24	34.15	4.27
Νησιά Αιγαίου	34.85	4.45	34.30	3.20	34.51	3.68

<b>Περιοχή</b>						
<i>Αστική</i>	34.90	6.10	34.96	3.20	34.93	4.90
<i>Γεωργική</i>	34.94	9.88	35.36	4.17	35.21	5.27
<b>BMI</b>						
<i>Φυσιολογικό</i>	35.5	7.3	34.7	3.3	35.1	5.7
<i>Παχύσαρκοι</i>	34.4	6	34.8	3.6	34.6	4.9
<i>Υπέρβαροι</i>	35	5.8	35.7	3.8	35.4	4.7
<b>WHR</b>						
<i>&lt;=0.95</i>	34.99	7.37	-	-	34.99	7.37
<i>&gt;0.95</i>	34.84	5.44	-	-	34.84	5.44
<i>&lt;=0.80</i>	-	-	35.44	3.10	35.44	3.10
<i>&gt;0.80</i>	-	-	34.85	3.95	34.85	3.95
<b>Υποεκτιμούντες</b>						
<i>Όχι</i>	34.6	6.3	35.2	3.4	34.9	5
<i>Ναι</i>	35.7	6.5	34.9	4	35.2	5.1
<b>Κάπνισμα</b>						
<i>Ποτέ</i>	35.35	7.82	35.07	3.77	35.14	5.03
<i>Παλιός καπνιστής</i>	35.82	5.98	34.79	5.01	35.67	5.83
<i>Τωρινός καπνιστής</i>	33.20	4.57	35.55	2.22	34.10	3.99
<b>Οικογενειακή κατάσταση</b>						
<i>Ελεύθερος-η</i>	33.01	7.19	33.78	2.66	33.40	5.22
<i>Παντρεμένος-η</i>	35.09	6.37	35.46	3.79	35.28	5.21
<i>Χωρισμένος-η</i>	33.53	1.87	35.14	1.38	34.66	1.63
<i>Χήρος-α</i>	32.04	0.30	33.38	2.95	33.25	1.83

<b>Απασχόληση</b>						
<i>Όχι/Καθιστική</i>	31.9	3.7	34.3	3.1	33.4	3.5
<i>Όρθια</i>	36.5	2.8	37	2.4	36.7	2.6
<i>Χειρονακτική</i>	40.3	2.9	42.3	2.9	40.9	3
<i>Βαριά Χειρονακτική</i>	62.6	2.8	-	-	62.6	2.8

Από τον πίνακα αυτό, προκύπτει ότι η μέση ημερήσια δαπάνη ενέργειας για το σύνολο των συμμετεχόντων είναι 35.03 MET (35.13 MET για τις γυναίκες και 34.91 MET για τους άνδρες) δηλαδή για ένα άτομο 75 κιλών η ημερήσια θερμιδική δαπάνη είναι  $75 \times 35.03 = 2627$  kcal. Οι ενεργειακές δαπάνες ανδρών και γυναικών εξειδικεύονται στον ίδιο πίνακα κατά ηλικιακή ομάδα, εκπαίδευση, περιφέρεια, περιοχή, δείκτη μάζας σώματος, WHR, υποεκτιμούντες, κάπνισμα, οικογενειακή κατάσταση και απασχόληση. Από την ταξινόμηση των μέσων ημερήσιων ενεργειακών δαπανών των δύο φύλλων ως προς τις παραπάνω μεταβλητές εμφανίζονται έντονες διαφοροποιήσεις μεταξύ αυτών που ασκούν καθιστική εργασία και χειρονακτική (κατά 8.4 MET στους άνδρες και 7.5 MET για τις γυναίκες) αλλά και αυτών που ασκούν βαριά χειρονακτική εργασία (περίπου 30 MET). Όσον αφορά στις άλλες μεταβλητές εμφανίζεται μεικτή εικόνα ως προς την κατανάλωση των MET. Φτιάχνοντας ένα μοντέλο ανάλυση συνδιακύμανσης με εξαρτημένη μεταβλητή το δείκτη 24ωρης φυσικής δραστηριότητας (σε μεταβολικά ισοδύναμα, MET) και ανεξάρτητες τις υπόλοιπες μεταβλητές μπορούμε να βγάλουμε κάποια χρήσιμα συμπεράσματα για τις παραπάνω μεταβλητές.

#### **5.4.2 Ανάλυση Διακύμανσης και Ανάλυση Συνδιακύμανσης για τη μεταβλητή MET.**

Όταν λέμε Ανάλυση Διασποράς ή Ανάλυση Διακύμανσης εννοούμε μια κατηγορία προβλημάτων πολλαπλής γραμμικής παλινδρόμησης όπου οι ανεξάρτητες μεταβλητές είναι κατηγορικές. Στα προβλήματα ανάλυσης διακύμανσης έχουμε μία εξαρτημένη μεταβλητή, την Y, και κάποιους παράγοντες (factors) που είναι οι ανεξάρτητες μεταβλητές. Κάθε παράγοντας μπορεί να εμφανίζει δύο ή περισσότερα επίπεδα. Όταν έχουμε ένα παράγοντα με κ-επίπεδα τότε έχουμε Ανάλυση Διακύμανσης με έναν παράγοντα (one-way ANOVA), όταν

έχουμε δύο παράγοντες έχουμε Ανάλυση Διακύμανσης με δύο παράγοντες (two-way ANOVA) κ.τ.λ.

Χαρακτηριστικά παρουσιάζουμε παρακάτω τη μορφή ενός μοντέλου ANOVA κατά δύο παράγοντες:

$$Y_{ijk} = \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

όπου

$\alpha_i$ : η επίδραση που έχει στις τιμές της Y το επίπεδο i του παράγοντα  $X_1$

$\beta_j$ : η επίδραση που έχει στις τιμές της Y το επίπεδο j του παράγοντα  $X_2$

$(\alpha\beta)_{ij}$ : η αλληλεπίδραση ανάμεσα στο επίπεδο i του παράγοντα  $X_1$  και στο επίπεδο j του παράγοντα  $X_2$

$\varepsilon_{ijk}$ : τυχαίο σφάλμα για το επίπεδο i του  $X_1$  και το επίπεδο j του  $X_2$

Το μοντέλο ANOVA μπορεί να γραφεί ισοδύναμα με τη χρήση πολλαπλής παλινδρόμησης ως εξής:  $Y = X \cdot \beta + \varepsilon$

Για να είναι κατάλληλο ένα μοντέλο ANOVA πρέπει να πληρούνται κάποιες βασικές υποθέσεις, οι οποίες είναι:

- τα σφάλματα πρέπει να ακολουθούν κανονική κατανομή  $\varepsilon_{ijk} \sim N(0, \sigma^2)$
- η διακύμανση των σφαλμάτων πρέπει να είναι σταθερή και
- δεν πρέπει να υπάρχει εξάρτηση μεταξύ των σφαλμάτων

Μια άλλη κατηγορία προβλημάτων γραμμικής παλινδρόμησης όπου οι ανεξάρτητες μεταβλητές είναι και κατηγορικές και συνεχείς, είναι η Ανάλυση Συνδιακύμανσης (ANCOVA). Στην Ανάλυση Συνδιακύμανσης ισχύει ότι και στην Ανάλυση Διακύμανσης και για να είναι κατάλληλο ένα μοντέλο ANCOVA πρέπει να πληρούνται οι ίδιες βασικές υποθέσεις.

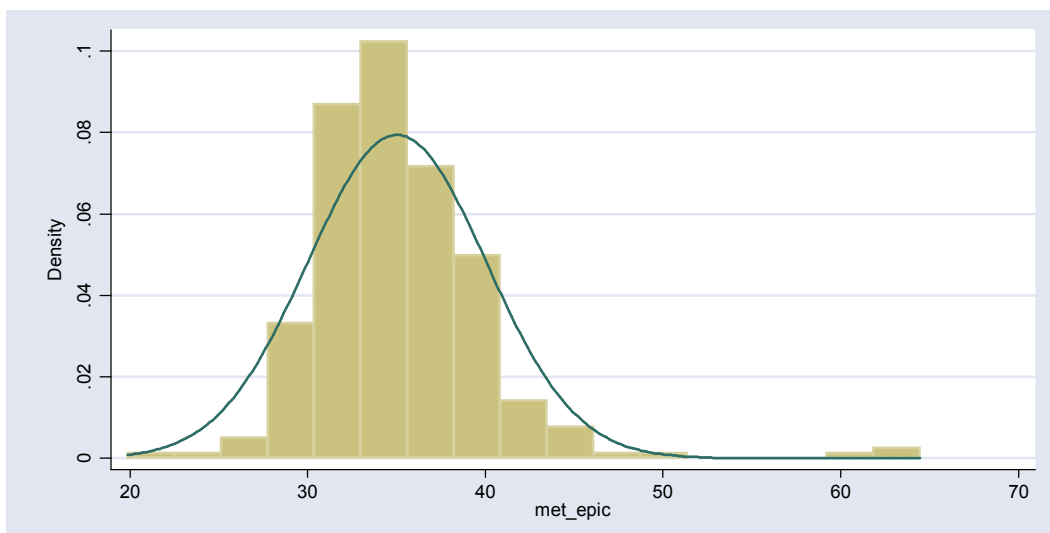
Πηγαίνοντας λοιπόν στα δικά μας δεδομένα επειδή η ανεξάρτητες μεταβλητές που έχουμε δεν είναι μόνο κατηγορικές αλλά και συνεχείς, θα εφαρμόσουμε Ανάλυση Συνδιακύμανσης (ANCOVA). Αρχικά ελέγχουμε αν η μεταβλητή μου MET ακολουθεί κανονική κατανομή με το τεστ Kolmogorov-Smirnov για ένα δείγμα:

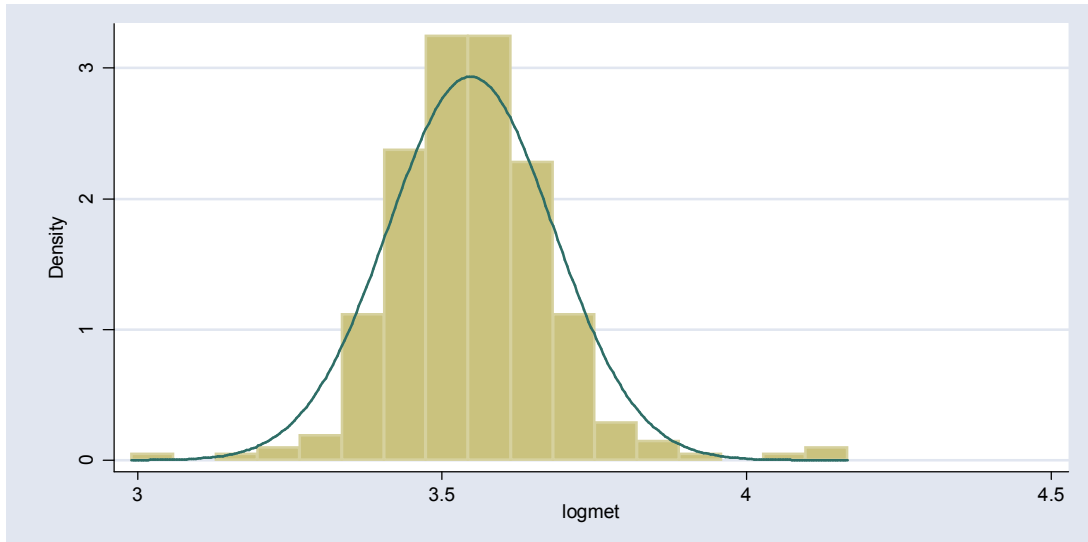


```
One-sample Kolmogorov-Smirnov test against theoretical distribution
normprob(( met_epic-35.0299)/5.025178)
```

Smaller group	D	P-value	Corrected
met_epic:	0.0912	0.007	
Cumulative:	-0.0687	0.060	
Combined K-S:	0.0912	<b>0.014</b>	0.012

Παρατηρούμε ότι  $p\text{-value} = 0.014$  δηλαδή δεν δεχόμαστε την  $H_0$  επομένως η μεταβλητή μας δεν ακολουθεί κανονική κατανομή και δεν μπορούμε να εφαρμόσουμε Ανάλυση Διακύμανσης ή Ανάλυση Συνδιακύμανσης. Λογαριθμίζοντας την μεταβλητή MET και ελέγχοντας τη νέα μεταβλητή μου, που την ονομάζω  $\log(\text{MET})$ , για κανονικότητα με το τεστ Kolmogorov-Smirnov βρίσκω  $p\text{-value} = 0.214$ , δηλαδή δεν απορρίπτουμε την  $H_0$  και επομένως η μεταβλητή μου ακολουθεί κανονική κατανομή. Παρουσιάζοντας παρακάτω το ιστόγραμμα της μεταβλητής MET και της  $\log(\text{MET})$  φαίνεται καθαρά ότι η δεύτερη ακολουθεί κανονική κατανομή ενώ η πρώτη παρουσιάζει δεξιά ασυμμετρία:





Δημιουργούμε ένα μοντέλο με εξαρτημένη μεταβλητή τα  $\log(\text{MET})$  και ανεξάρτητες μεταβλητές το φύλλο, το BMR, την πρόσληψη ενέργειας, το ύψος, το βάρος, την ηλικία, τις αλκοόλες, τις πρωτεΐνες, τους υδατάνθρακες, τα λιπίδια, την περιφέρεια, την περιοχή, το κάπνισμα, τη μόρφωση, την απασχόληση, την οικογενειακή κατάσταση, τους underreporters, το WHR και το BMI και εφαρμόζουμε ένα μοντέλο ANCOVA (Analysis of Covariance) γιατί οι ανεξάρτητες μεταβλητές είναι και κατηγορικές και συνεχείς:

Εφαρμόζοντας Stepwise Regression (ξεκινώντας από το πλήρες μοντέλο) καταλήγουμε σε ένα μοντέλο με σημαντικές μεταβλητές το φύλλο, την περιφέρεια και την απασχόληση ενώ οι αλληλεπιδράσεις των τριών αυτών μεταβλητών δεν είναι στατιστικά σημαντικές όπως φαίνεται από τον παρακάτω πίνακα ANOVA:

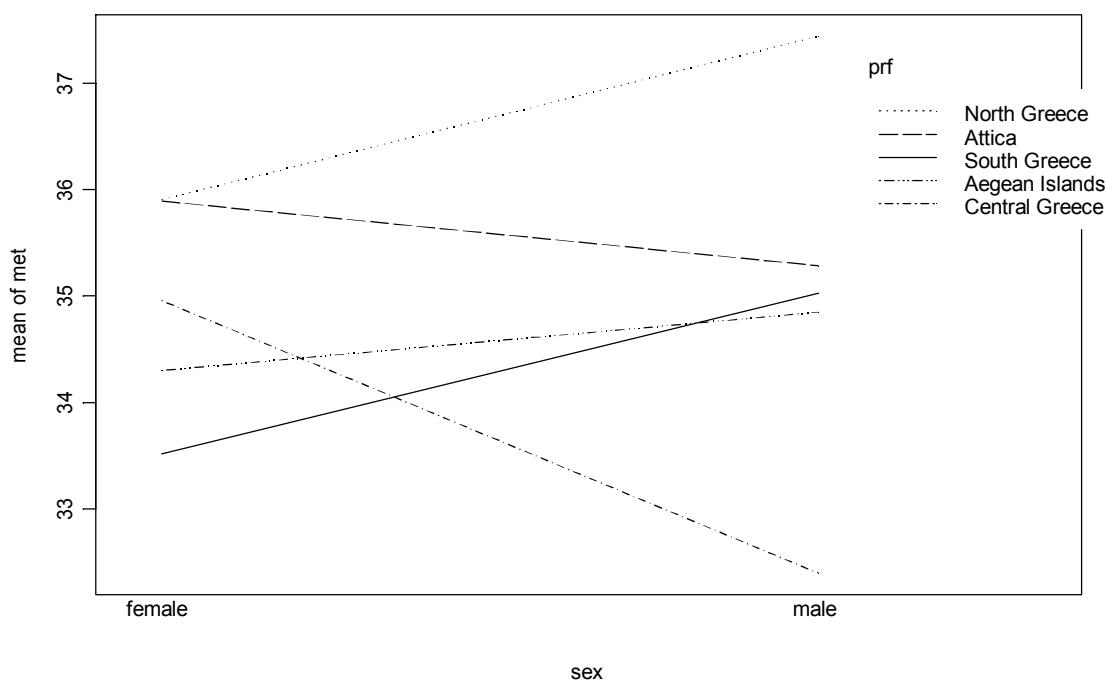
	DF	Sum of Square	Mean Sq	F Value	Pr(F)
<i>occupy</i>	3	1.736796	0.5789319	67.22700	0
<i>sex</i>	1	0.287742	0.2877416	33.41327	0
<i>prf_epic</i>	4	0.178814	0.0447035	5.19108	0.0004809
<i>sex*prf_epic</i>	4	0.050903	0.0127257	1.47773	0.2091438
<i>sex*occupy</i>	2	0.033355	0.0166773	1.93661	0.1461751
<i>prf_epic*occupy</i>	8	0.058266	0.0072833	0.84575	0.5631093
<i>Residuals</i>	272	2.342355	0.0086116		

Για να πάρουμε μια πρώτη ιδέα για τον αν υπάρχουν αλληλεπιδράσεις μεταξύ των μεταβλητών μου, μπορώ να φτιάξω τρεις πίνακες 2x2, για κάθε ζευγάρι αλληλεπίδρασης

όπου θα φαίνονται τα MET, σε κάθε συνδυασμό κατηγοριών, καθώς επίσης μπορούμε να βρούμε και τα διαγράμματα αλληλεπίδρασης (Interaction Plots):

**Φύλλο-Περιοχή:**

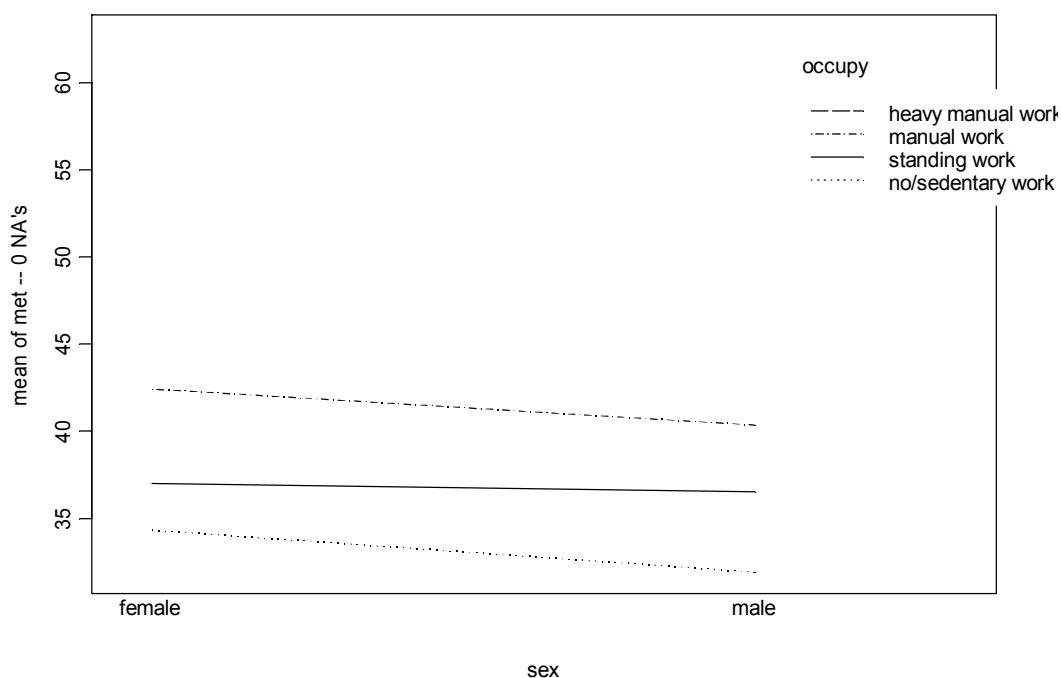
sex	prf_epic				
	Attica	Central Greece	North Greece	South Greece	Aegean Islands
male	35.33493	32.36154	37.44394	35.02776	34.84769
female	35.89234	34.9571	35.90671	33.7812	34.3005



Παρατηρούμε ότι μεταξύ φύλλου και περιοχής υπάρχει ισχυρή ένδειξη αλληλεπίδρασης. Αξιοσημείωτο είναι ότι τόσο οι άνδρες όσο και οι γυναίκες στη Βόρεια Ελλάδα έχουν τον πιο υψηλό δείκτη 24ωρης δραστηριότητας (εκφρασμένος σε MET). Επίσης, στην Κεντρική Ελλάδα και την Αττική οι γυναίκες έχουν μεγαλύτερο δείκτη 24ωρης δραστηριότητας από ότι οι άνδρες. Συμπερασματικά μπορούμε να πούμε ότι υπάρχει ισχυρή ένδειξη αλληλεπίδρασης μεταξύ των δύο αυτών μεταβλητών αλλά η αλληλεπίδραση αυτή δεν είναι στατιστικά σημαντική.

### Φύλλο-Απασχόληση:

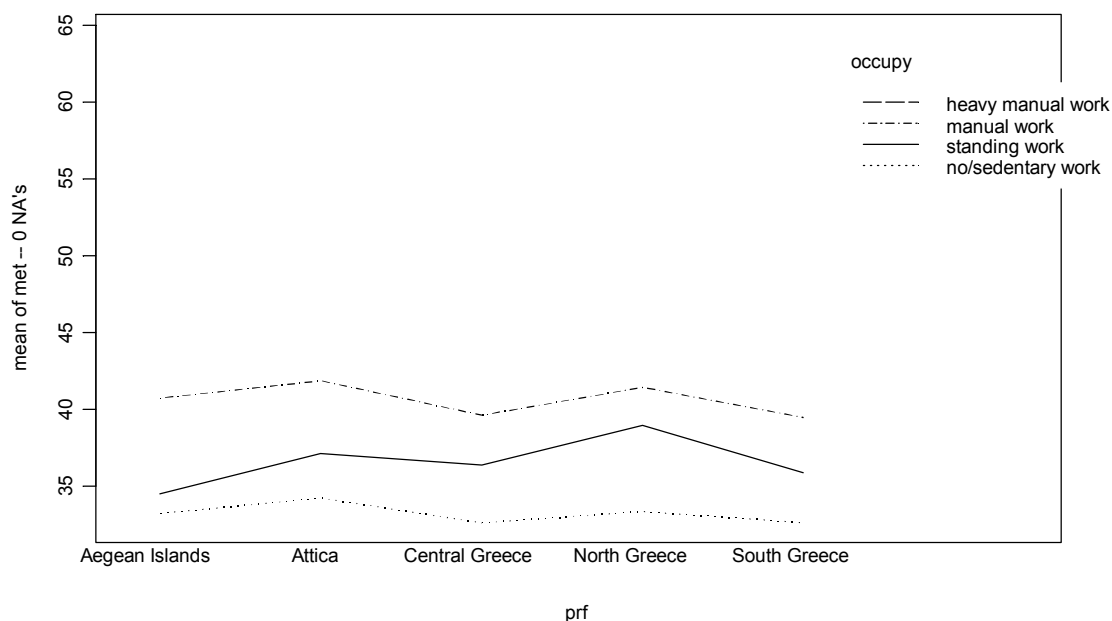
sex	occupy			
	no/sedentary work	standing work	manual work	heavy manual work
male	31.91454	36.52958	40.2672	62.64603
female	34.32182	37.02931	42.31472	



Εδώ είναι φανερό ότι δεν υπάρχει αλληλεπίδραση, ενώ έχουμε μόνο τρεις γραμμές στο διάγραμμα αλληλεπίδρασης, ενώ τα επίπεδα της απασχόλησης είναι τέσσερα, γιατί δεν υπάρχει καμία γυναίκα που να ανήκει στην τέταρτη κατηγορία, δηλαδή να κάνει βαριά χειρονακτική εργασία. Επίσης παρατηρούμε ότι όσο πιο βαριά είναι η εργασία τόσο αυξάνει ο δείκτης 24ωρης δραστηριότητας (το οποίο είναι λογικό αφού καταναλώνονται περισσότερα MET) ενώ σε όλα τα επαγγέλματα οι γυναίκες έχουν μεγαλύτερη κατανάλωση σε MET.

### Περιοχή-Απασχόληση:

prf_epic	no/sedentary work	occupy standing work	manual work	occupy heavy manual work
Attica	34.22063	37.14759	41.46089	64.4524
Central Greece	32.58076	36.37117	39.6123	
North Greece	33.34554	38.94666	41.43333	61.74285
South Greece	32.62274	35.89243	39.7623	
Aegean Islands	33.22541	34.50178	40.73055	



Φαίνεται καθαρά ότι μεταξύ περιοχής και απασχόλησης δεν υπάρχει αλληλεπίδραση, ενώ έχουμε τρεις γραμμές στο διάγραμμα αλληλεπίδρασης, αν και τα επίπεδα της απασχόλησης είναι τέσσερα, γιατί δεν υπάρχει κανένα άτομο στην Κεντρική και Νότια Ελλάδα καθώς και στα νησιά του Αιγαίου που να κάνουν βαριά χειρονακτική εργασία.

Έτσι το τελικό μου μοντέλο φαίνεται παρακάτω:

	DF	Sum of Square	Mean Sq	F Value	Pr(F)
<i>occupy</i>	3	1.736796	0.5789319	66.63285	0
<i>sex</i>	1	0.287742	0.2877416	33.11796	0
<i>prf_epic</i>	4	0.178814	0.0447035	5.14520	0.0005125
<i>Residuals</i>	286	2.484878	0.0086884		

Παρατηρούμε ότι η απασχόληση έχει τη μεγαλύτερη επίδραση από τις άλλες μεταβλητές και αυτό φαίνεται τόσο από τη στήλη Sum of Square, όπου το άθροισμα τετραγώνων για την απασχόληση είναι πολύ μεγαλύτερο από αυτό των άλλων δύο μεταβλητών όσο και από το p-value που είναι μηδέν. Μάλιστα όσο πιο δύσκολη είναι η απασχόληση, δηλαδή από καθιστική πάμε σε όρθια και στη συνέχεια σε χειρονακτική και βαριά χειρονακτική τόσο αυξάνει ο δείκτης 24ωρης δραστηριότητας που εκφράζεται σε MET. Μάλιστα τα άτομα που κάνουν βαριά χειρονακτική εργασία, έχουν σχεδόν διπλάσια MET από τα άτομα που κάνουν τις υπόλοιπες εργασίες.

Από τον πίνακα των coefficients φαίνονται οι συντελεστές  $\beta_i$  :

#### Estimated Coefficients

	Value
<i>Intercept</i>	4.147999
<i>Occupy Manual work</i>	-0.3988772
<i>Occupy No/Sedentary work</i>	-0.6311622
<i>Occupy Standing work</i>	-0.5178425
<i>sex</i>	-0.07361196
<i>prf_epic Attica</i>	0.04588183
<i>prf_epic Central Greece</i>	-0.00395869
<i>prf_epic North Greece</i>	0.009667313
<i>prf_epic South Greece</i>	-0.02276145

Δηλαδή το μοντέλο μου γράφεται:

$$\log Y = 4.14 - 0.4X_1 - 0.63X_2 - 0.52X_3 - 0.074X_4 + 0.046X_5 - 0.004X_6 + 0.0097X_7 - 0.023X_8$$

όπου,

$X_1$  είναι η μεταβλητή που αφορά στην χειρονακτική εργασία, με τιμή 1 και 0 αν ένα άτομο κάνει χειρονακτική εργασία ή όχι, αντίστοιχα

$X_2$  είναι η μεταβλητή που αφορά στην καθιστική εργασία, με τιμή 1 και 0 αν ένα άτομο κάνει καθιστική εργασία ή όχι, αντίστοιχα

$X_3$  είναι η μεταβλητή που αφορά στην όρθια εργασία, με τιμή 1 και 0 αν ένα άτομο κάνει όρθια εργασία ή όχι, αντίστοιχα

$X_4$  είναι η μεταβλητή φύλλο, με τιμή 0 αν είναι γυναίκα και 1 αν είναι άνδρας

$X_5$  είναι η μεταβλητή που αφορά στην περιφέρεια της Αττικής, με τιμή 1 και 0 αν ένα άτομο ζει στην Αττική ή όχι, αντίστοιχα

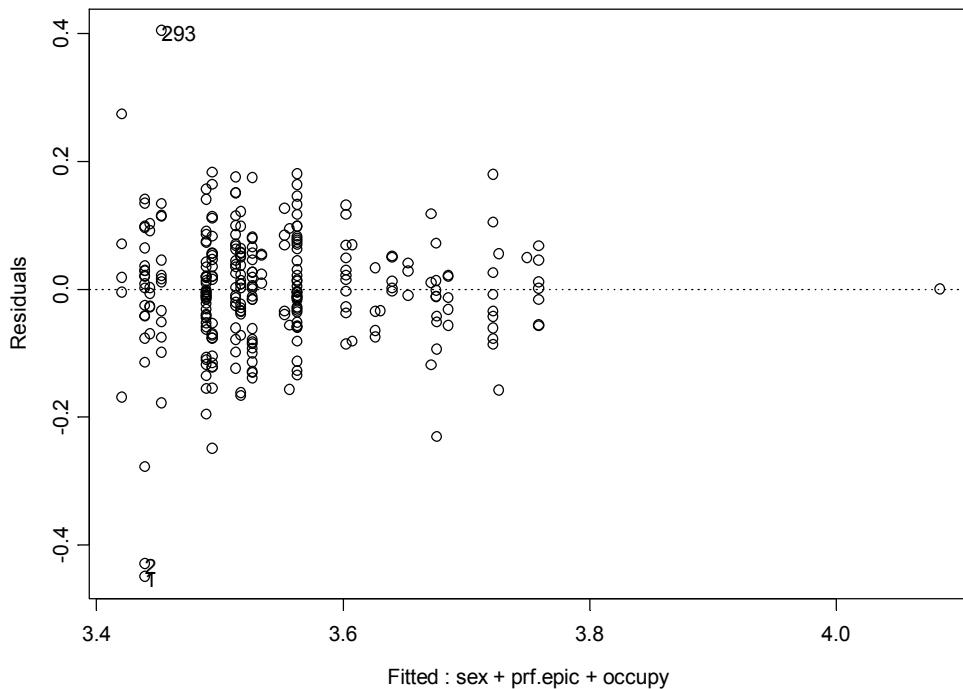
$X_6$  είναι η μεταβλητή που αφορά στην περιφέρεια της Κεντρικής Ελλάδας, με τιμή 1 και 0 αν ένα άτομο ζει στην Κεντρική Ελλάδα ή όχι, αντίστοιχα

$X_7$  είναι η μεταβλητή που αφορά στην περιφέρεια της Βόρειας Ελλάδας, με τιμή 1 και 0 αν ένα άτομο ζει στη Βόρεια Ελλάδα εργασία ή όχι, αντίστοιχα

$X_8$  είναι η μεταβλητή που αφορά στην περιφέρεια της Νότιας Ελλάδας, με τιμή 1 και 0 αν ένα άτομο ζει στη Νότια Ελλάδα εργασία ή όχι, αντίστοιχα

Για να ελέγξουμε αν το μοντέλο είναι κατάλληλο για χρήση ώστε να βρίσκουμε τις αναμενόμενες τιμές της  $Y$  για τις διάφορες τιμές της  $X_i$  θα ελέγξουμε τη συμπεριφορά των καταλοίπων, δηλαδή θα πρέπει να πληρούνται οι προϋποθέσεις που έχουμε αναφέρει παραπάνω:

- Ο έλεγχος κανονικότητας των σφαλμάτων δίνει μεγάλο p-value=0.170, συνεπώς δεν απορρίπτουμε την υπόθεση της κανονικότητας.
- Ο έλεγχος ροών δείχνει πως υπάρχει ένδειξη για μη ανεξαρτησία των σφαλμάτων, αφού το αντίστοιχο p-value είναι 0
- Τέλος, για να ελέγξουμε αν η διακύμανση είναι σταθερή κάνουμε ένα scatter plot:



Μπορούμε να πούμε ότι η διακύμανση των σφαλμάτων είναι σταθερή και δεν αυξάνεται ούτε ελαττώνεται ανάλογα με το  $Y$ .

Επίσης, ο συντελεστής προσδιορισμού είναι:

$$R^2 = \text{άθροισμα τετραγώνων που οφείλεται στην παλινδρόμηση/ολική μεταβλητότητα} = \text{SSR/SST} = 2.203352/4.68823 = 0.47,$$

δηλαδή το 47% από τη μεταβλητότητα των  $Y$  ερμηνεύεται από την παλινδρόμηση ενώ το υπόλοιπο 53% οφείλεται στο τυχαίο σφάλμα.

Γενικά, μπορούμε να πούμε ότι το μοντέλο μας δεν είναι πολύ καλό για να κάνουμε ακριβείς εκτιμήσεις για τον δείκτη 24ωρης δραστηριότητας. Εάν θέλουμε να χρησιμοποιήσουμε αυτό το μοντέλο τότε μπορούμε να καταλάβουμε τη χρησιμότητά του δίνοντας το παρακάτω παράδειγμα:



### **Παράδειγμα:**

Θέλουμε να εκτιμήσουμε τον δείκτη 24ωρης δραστηριότητας (που είναι εκφρασμένος σε MET) ενός άνδρα, που ζει στην Αττική και κάνει καθιστική εργασία:

$$\log \hat{Y} = 4.14 - 0.4 * 0 - 0.63 * 1 - 0.52 * 0 - 0.074 * 1 + 0.046 * 1 - 0.004 * 0 + 0.0097 * 0 - 0.023 * 0 \Rightarrow$$

$$\log \hat{Y} = 3.482 \Rightarrow \hat{Y} = 32.52$$

Στη συνέχεια να εκτιμηθεί ο δείκτης 24ωρης δραστηριότητας μιας γυναίκας που ζει στην Αττική και κάνει καθιστική εργασία:

$$\log \hat{Y} = 4.14 - 0.4 * 0 - 0.63 * 1 - 0.52 * 0 - 0.074 * 0 + 0.046 * 1 - 0.004 * 0 + 0.0097 * 0 - 0.023 * 0 \Rightarrow$$

$$\log \hat{Y} = 3.556 \Rightarrow \hat{Y} = 35.02$$

### **5.5 Συντελεστές Συσχέτισης.**

Ο συντελεστής συσχέτισης, γενικά σε θεωρητικό επίπεδο, είναι ένα μέτρο συσχέτισης μεταξύ δύο μεταβλητών που μας δείχνει το βαθμό γραμμικής εξάρτησης μεταξύ δύο τυχαίων μεταβλητών. Είναι ένας καθαρός αριθμός και παίρνει τιμές μεταξύ  $-1$  και  $1$  και είναι ένα μέτρο του βαθμού της γραμμικής συσχέτισης των δύο συγκεκριμένων μεταβλητών. Όταν ο συντελεστής συσχέτισης είναι  $0$  τότε δεν υπάρχει γραμμική σχέση μεταξύ των δύο μεταβλητών (μπορεί οι δύο μεταβλητές να σχετίζονται διαφορετικά π.χ καμπυλόγραμμα) και οι τυχαίες μεταβλητές λέγονται ασυσχέτιστες. Όταν ο συντελεστής συσχέτισης ισούται με  $+1$  ή  $-1$  τότε λέμε ότι οι τ.μ συσχετίζονται γραμμικά με τέλεια σχέση.

Επειδή ο θεωρητικός συντελεστής συσχέτισης είναι δύσκολο να υπολογιστεί στην πράξη για αυτό εκτιμάμε τους παρακάτω συντελεστές, ανάλογα με το αν έχουμε συνεχείς ή κατηγορικές μεταβλητές<sup>24</sup>:

*Συντελεστής συσχέτισης του Pearson (r).* Όταν θέλουμε να βρούμε τη γραμμική συσχέτιση μεταξύ δύο συνεχών μεταβλητών χρησιμοποιούμε το συντελεστή συσχέτισης του Pearson, ο οποίος είναι ένας καθαρός αριθμός και παίρνει τιμές από  $-1$  μέχρι  $1$ . Όταν  $r=0$  δεν υπάρχει γραμμική σχέση μεταξύ των  $X$  και  $Y$ , όταν  $r=+1$  έχουμε θετική γραμμική εξάρτηση και όταν  $r=-1$  έχουμε αρνητική γραμμική εξάρτηση μεταξύ των τ.μ.  $X$  και  $Y$

Συντελεστής συσχέτισης του Spearman ( $r_s$ ). Όταν θέλουμε να βρούμε τη γραμμική συσχέτιση δύο κατηγορικών μεταβλητών ή μιας συνεχούς και μιας κατηγορικής μεταβλητής τότε χρησιμοποιούμε το συντελεστή συσχέτισης του Spearman. Ο συντελεστής ( $r_s$ ) παίρνει τιμές από  $-1$  μέχρι  $1$  και έχει ανάλογες ιδιότητες με το συντελεστή συσχέτισης του Pearson (εκτός του συντελεστή  $r_s$  υπάρχει και ο συντελεστής συσχέτισης  $r$  του Kendall).

Όταν υπολογίζουμε κάποιον από τους παραπάνω συντελεστές συσχέτισης και βρίσκουμε τιμή διάφορη του μηδενός τότε έχουμε σοβαρές ενδείξεις ότι και ο θεωρητικός συντελεστής συσχέτισης  $\rho$  του πληθυσμού είναι διάφορος του μηδενός και επομένως οι τυχαίες μεταβλητές  $X$  και  $Y$  είναι συσχετιζόμενες. Στις περιπτώσεις αυτές καλό είναι να ελέγχεται και στατιστικά με ένα κριτήριο ελέγχου συντελεστή συσχέτισης:

Για τον έλεγχο της  $H_0: \rho=0$  έναντι των εναλλακτικών  $H_a: \rho>0$ ,  $H_a: \rho<0$ ,  $H_a: \rho \neq 0$  χρησιμοποιείται η στατιστική συνάρτηση:

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

με κρίσιμες περιοχές αντίστοιχα

$$t \geq t_{n-2, \alpha} \quad , \quad t \leq -t_{n-2, \alpha} \quad , \quad t \geq t_{n-2, \alpha/2}$$

### 5.5.1 Συντελεστές συσχέτισης για τους υποεκτιμούντες.

Θέλουμε να βρούμε όλες τις συσχετίσεις της μεταβλητής των υποεκτιμούντων με όλες τις μεταβλητές μας και να δούμε ποιες συσχετίσεις είναι σημαντικές. Επειδή η μεταβλητή των υποεκτιμούντων είναι κατηγορική (και μάλιστα δίτιμη) δεν μπορούμε να εφαρμόσουμε το συντελεστή συσχέτισης του Pearson αλλά μπορούμε να εφαρμόσουμε τους συντελεστές συσχέτισης του Spearman ή του Kendal (όπου στην συγκεκριμένη περίπτωση που έχω δίτιμη μεταβλητή οι δύο συντελεστές ταυτίζονται). Εμείς στον παρακάτω πίνακα θα δώσουμε το συντελεστή συσχέτισης του Spearman των υποεκτιμούντων με κάποιες μεταβλητές και θα δώσουμε και τα αντίστοιχα  $p$ -value για να δούμε ποιες συσχετίσεις είναι στατιστικά σημαντικές:

Μεταβλητές	Συντελεστής συσχέτισης του Spearman	P-value
Μόρφωση	-0.1415	0.0145
Περιφέρεια	-0.0495	0.3946
Φύλλο	0.1099	0.0573
BMI	0.1631	0.0046
Ηλικία	0.1262	0.0289
Οικογενειακή κατάσταση	0.0423	0.4659
BMR	-0.0022	0.9703
EI	-0.0021	0.9709
WHR	0.1750	0.0025
Ύψος	-0.1066	0.0652
Βάρος	0.0930	0.1078
Πρωτεΐνες	-0.5688	0
Υδατάνθρακες	-0.4985	0
Λιπίδια	-0.6469	0
Αλκοόλες	-0.2724	0
Περιοχή	-0.0398	0.4934
MET	0.0195	0.7369
Κάπνισμα	-0.1063	0.0670
Απασχόληση	0.0133	0.8194

Από τον παραπάνω πίνακα παρατηρούμε ότι οι συντελεστές συσχέτισης του Spearman των υποεκτιμούντων με τη μόρφωση, το BMI, την ηλικία, το WHR και τα διατροφικά συστατικά (πρωτεΐνες, λιπίδια, υδατάνθρακες, αλκοόλες) είναι στατιστικά σημαντικοί σε επίπεδο σημαντικότητας 5% ενώ οι συντελεστές συσχέτισης του Spearman των υποεκτιμούντων με το φύλλο, το ύψος και το κάπνισμα είναι στατιστικά σημαντικοί σε επίπεδο σημαντικότητας 10%.

Ερμηνεύοντας τους συντελεστές συσχέτισης καταλήγουμε στα παρακάτω συμπεράσματα:

- Μόρφωση-Υποεκτιμούντες ( $r_s=-0.1415$ )

Το αρνητικό πρόσημο σημαίνει ότι όσο αυξάνει το επίπεδο της μόρφωσης, δηλαδή από μεσαία πηγαίνει σε ανώτερη, τόσο μικρότερη είναι η πιθανότητα ένα άτομο να ανήκει στους υποεκτιμούντες.

- Φύλλο-Υποεκτιμούντες ( $r_s=0.1099$ )

Οι γυναίκες είναι πιο πιθανό να ανήκουν στους υποεκτιμούντες.

- BMI-Υποεκτιμούντες ( $r_s=0.1631$ )

Όσο αυξάνει το επίπεδο του BMI, δηλαδή από φυσιολογικό βάρος πηγαίνει σε παχύσαρκους, τόσο μεγαλύτερη είναι η πιθανότητα ένα άτομο να ανήκει στους υποεκτιμούντες.

- Ηλικία-Υποεκτιμούντες ( $r_s=0.1262$ )

Όσο αυξάνει το επίπεδο της ηλικίας, τόσο μεγαλύτερη είναι η πιθανότητα ένα άτομο να ανήκει στους υποεκτιμούντες.

- WHR-Υποεκτιμούντες ( $r_s=0.1750$ )

Όσο αυξάνει το WHR, τόσο μεγαλύτερη είναι η πιθανότητα ένα άτομο να ανήκει στους υποεκτιμούντες.

- Ύψος-Υποεκτιμούντες ( $r_s=-0.1066$ )

Όσο μειώνεται το ύψος, τόσο μεγαλύτερη είναι η πιθανότητα ένα άτομο να ανήκει στους υποεκτιμούντες.

- Πρωτεΐνες-Υποεκτιμούντες ( $r_s=-0.5688$ )

Όσο μειώνεται η ποσότητα πρωτεϊνών που καταναλώνεται από ένα άτομο, τόσο μεγαλύτερη είναι η πιθανότητα ένα άτομο να ανήκει στους υποεκτιμούντες.

- Υδατάνθρακες-Υποεκτιμούντες ( $r_s=-0.4985$ )

Όσο μειώνεται η ποσότητα υδατανθράκων που καταναλώνεται από ένα άτομο, τόσο μεγαλύτερη είναι η πιθανότητα ένα άτομο να ανήκει στους υποεκτιμούντες.

- Λιπίδια-Υποεκτιμούντες ( $r_s=-0.6469$ )

Όσο μειώνεται η ποσότητα λιπιδίων που καταναλώνεται από ένα άτομο, τόσο μεγαλύτερη είναι η πιθανότητα ένα άτομο να ανήκει στους υποεκτιμούντες.

- Αλκοόλες-Υποεκτιμούντες ( $r_s=-0.2724$ )

Όσο μειώνεται η ποσότητα αλκοολών που καταναλώνεται από ένα άτομο, τόσο μεγαλύτερη είναι η πιθανότητα ένα άτομο να ανήκει στους υποεκτιμούντες.

- Κάπνισμα-Υποεκτιμούντες ( $r_s=-0.1063$ )

Όσο αυξάνει το επίπεδο καπνίσματος, και όταν λέμε ότι αυξάνει το επίπεδο καπνίσματος πρέπει να αναφέρουμε ότι έχουμε ορίσει 1⇒ποτέ καπνιστής, 2⇒παλιός καπνιστής, 3⇒τωρινός καπνιστής, τόσο μικρότερη είναι η πιθανότητα ένας συμμετέχων να ανήκει στην ομάδα των υποεκτιμούντων.

### 5.5.2 Συντελεστές συσχέτισης για τα MET.

Θέλουμε να βρούμε όλες τις συσχετίσεις της μεταβλητής των MET με όλες τις μεταβλητές μας και να δούμε ποιες συσχετίσεις είναι σημαντικές. Επειδή η μεταβλητή MET είναι συνεχής εφαρμόζουμε το συντελεστή συσχέτισης του Pearson όταν θέλω τη συσχέτιση των MET με συνεχείς μεταβλητές, ενώ όταν θέλω τη συσχέτιση των MET με κατηγορικές μεταβλητές εφαρμόζουμε τους συντελεστές συσχέτισης του Spearman ή του Kendall. Εμείς στους παρακάτω πίνακες θα δώσουμε τους συντελεστές συσχέτισης του Pearson και του Spearman των MET με κάποιες μεταβλητές και θα δώσουμε και τα αντίστοιχα p-value για να δούμε ποιες συσχετίσεις είναι στατιστικά σημαντικές:

Μεταβλητές	Συντελεστής συσχέτισης του Pearson	P-value
BMR	0.0467	0.480
EI	0.0467	0.480
Ύψος	0.0203	0.793
Βάρος	0.0141	0.883
Γοφοί	-0.039	0.508
Μέση	-0.078	0.179
Πρωτεΐνες	0.0589	0.324

Υδατάνθρακες	0.0627	0.273
Λιπίδια	0.0058	0.960
Αλκοόλες	0.0075	0.945

Μεταβλητές	Συντελεστής συσχέτισης του Spearman	P-value
Φύλλο	0.0943	0.1041
Ηλικία	-0.2875	0
BMI	0.0289	0.6194
WHR	0.0531	0.3632
Περιφέρεια	-0.0670	0.2490
Περιοχή	0.0429	0.4603
Κάπνισμα	-0.0348	0.5491
Εκπαίδευση	0.0617	0.2883
Απασχόληση	0.5963	0
Οικογενειακή κατάσταση	-0.0515	0.3758
Υποεκτιμούντες	0.0195	0.7369

Από τον παραπάνω πίνακα παρατηρούμε ότι οι συντελεστές συσχέτισης του Spearman των υποεκτιμούντων με την απασχόληση και την ηλικία είναι στατιστικά σημαντικοί σε επίπεδο σημαντικότητας 5% .

Ερμηνεύοντας τους συντελεστές συσχέτισης καταλήγουμε στα παρακάτω συμπεράσματα:

- Απασχόληση-MET ( $r_s=0.5963$ )

Όσο αυξάνει το επίπεδο της απασχόλησης, δηλαδή από καθιστική πηγαίνει σε βαριά χειρονακτική, τόσο μεγαλύτερη είναι η κατανάλωση των MET.

□ Ηλικία-MET ( $r_s=-0.2875$ )

Όσο αυξάνει το επίπεδο της ηλικίας, τόσο λιγότερη είναι η κατανάλωση των MET.





# ΚΕΦΑΛΑΙΟ 6

## Γενικά Συμπεράσματα

Το πρόγραμμα ΕΠΙΚ (Ευρωπαϊκό Πρόγραμμα συνεργασίας Ιατρικής και Κοινωνίας), όπως έχουμε αναφέρει είναι μια πολυκεντρική επιδημιολογική προοπτική μελέτη (cohort study) που σχεδιάστηκε για να ερευνήσει τη σχέση μεταξύ διατροφής και καρκίνου άλλα ερευνά και τη σχέση μεταξύ διατροφικών και άλλων παραγόντων που αφορούν στον τρόπο ζωής, στα γονίδια κ.λ.π. Οι διατροφικές πληροφορίες προήρθαν χρησιμοποιώντας ερωτηματολόγια συχνότητας τροφίμων διαφορετικά σε κάθε χώρα, ώστε να μπορέσουμε να εκτιμήσουμε μεγάλης περιόδου πρόσληψη για κάθε άτομο. Για αυτό το λόγο, μία δεύτερη διατροφική μέθοδος, βασισμένη σε μια προτυποποιημένη συνέντευξη, που επιτρέπει τη συλλογή στοιχείων ανάκλησης 24-ώρου (24-hour dietary recall), συγκέντρωσε δεδομένα από ένα υπό-δείγμα του ΕΠΙΚ και χρησιμοποιήθηκε σαν μέθοδος αναφοράς για τη βαθμονόμηση των αρχικά συλλεγόμενων πληροφοριών.. Λόγω απουσίας μιας αμερόληπτης αναφορικής διατροφικής μεθόδου η τυποποίηση της ανάκλησης 24-ώρου φάνηκε σαν η καλύτερη λύση. Έτσι δημιουργήθηκε ένα λογισμικό (software) πρόγραμμα, το EPIC- SOFT. Στην Ελλάδα, η οποία είναι μία από τις 10 χώρες που συμμετέχουν στο ΕΠΙΚ και είναι η χώρα από την οποία θα επιλέξουμε το δείγμα της ανάλυσης μας, 5.406 συμμετέχοντες προσεγγίστηκαν και τελικά 2930 έδωσαν συνέντευξη.

Οι μελετούμενοι πληθυσμοί στο ΕΠΙΚ δεν επιλέχθηκαν για να εξασφαλίσουν αντιπροσωπευτικά δείγματα. Αυτό συνεπάγεται και από τη φύση των προοπτικών μελετών όπου δεν μας ενδιαφέρει η αντιπροσωπευτικότητα. Η συγκέντρωση και η επιλογή (recruitment) αποφασίστηκε από πρακτικής και υπολογιστικής (logistic) θεώρησης για να εξασφαλίσει υψηλή συμμετοχή και μεγάλης περιόδου follow-up από τους μετέχοντες στην μελέτη. Στηριζόμενοι στην επιλογή του πληθυσμού που έχει παρουσιαστεί παραπάνω λεπτομερώς και στα διάφορα χαρακτηριστικά του, έχουν διεξαχθεί πολλές ερευνητικές μελέτες. Η δική μας μελέτη, που παρουσιάστηκε στα προηγούμενα κεφάλαια, είχε σκοπό τον έλεγχο της συμβατότητας ενεργειακής πρόσληψης των δεδομένων της 24ωρης διατροφικής ανάκλησης προσδιορίζοντας τους υποεκτιμούντες (under-reporters) καθώς επίσης και την

ανάλυση τρόπων εύρεσης της φυσικής δραστηριότητας ενός ατόμου, αφού η φυσική δραστηριότητα είναι σημαντικός παράγοντας στον προσδιορισμό των υποεκτιμούντων. Η μέτρηση της φυσικής δραστηριότητας οδήγησε στον προσδιορισμό της ενέργειας που συνήθως δαπανά ένα άτομο ανά ημέρα ή ώρα, εκφρασμένη είτε σε μονάδες ενέργειας (kcal ή kJ) είτε σε μεταβολικά ισοδύναμα (metabolic equivalent, MET).

Έτσι, λοιπόν, η υποεκτίμηση της ενεργειακής πρόσληψης, που προσδιορίστηκε στα άτομα της έρευνας (26% στην πρώτη περίπτωση και 29% στη δεύτερη) θεωρείται αναμενόμενη σε διατροφικές έρευνες, κατά τον προσδιορισμό της διατροφικής πρόσληψης. Αυτό μπορούμε να το διαπιστώσουμε από διατροφικές επιδημιολογικές έρευνες που έχουν γίνει σε άτομα που ζουν υπό φυσιολογικές συνθήκες διαβίωσης. Σε μελέτη «μετα-ανάλυσης» των Black et al<sup>18</sup> έγινε αξιολόγηση της ενεργειακής πρόσληψης, η οποία προσδιορίστηκε σε 37 δημοσιευμένες διατροφικές μελέτες που έγιναν στην Ευρώπη και την Αμερική από το 1975 και μετά. Από τα αποτελέσματα της εργασίας αυτής βρέθηκε ότι το 68% των διατροφικών ερευνών υποεκτιμούν την ενεργειακή πρόσληψη. Η αξιολόγηση της ενεργειακής πρόσληψης, που προσδιορίστηκε στις μελέτες αυτές, έγινε συγκρίνοντας κάθε φορά το λόγο EI/BMR με το αντίστοιχο CUT-OFF όριο κάθε μελέτης το οποίο, όπως έχουμε αναφέρει, αντιστοιχεί σε «καθιστικό τρόπο ζωής» (PAL=1.55).

Η χρήση όμως του 1.55 για το δείκτη φυσικής δραστηριότητας βασίζεται στην παραδοχή ότι τα άτομα που συμμετέχουν σε κάθε μελέτη έχουν ελαφριά φυσική δραστηριότητα, παραδοχή που είναι αναπόφευκτη, εφόσον δεν υπάρχουν διαθέσιμα δεδομένα για να εκτιμηθεί η «συνηθισμένη» φυσική δραστηριότητά τους. Η τιμή του 1.55 είναι η ελάχιστη που θα μπορούσε να χρησιμοποιηθεί για την ελαφρά δραστηριότητα ατόμων που ακολουθούν φυσιολογικές συνθήκες διαβίωσης. Ο μοναδικός τρόπος να υπολογίζεται αντικειμενικά η τιμή του ελαχίστου ορίου, είναι να εκτιμάται για κάθε άτομο χωριστά ανάλογα με τις πραγματικές δραστηριότητες του. Μια τέτοια περίπτωση, αναφέραμε λεπτομερώς, όπου με βάση την επαγγελματική δραστηριότητα και την ψυχαγωγική δραστηριότητα, που προσδιορίζεται από ένα ερωτηματολόγιο, μπορούμε να βρούμε σε ποιο επίπεδο φυσικής δραστηριότητας ανήκει το κάθε άτομο (ελαφριά, μέτρια ή βαριά δραστηριότητα). Επικεντρώσαμε την προσοχή μας μόνο στην ποδηλασία και την φυσική άσκηση. Σε αυτή την περίπτωση, το ποσοστό των ατόμων που βρέθηκε να υποεκτιμούν την πρόσληψη στη μελέτη μας είναι 29%, μεγαλύτερο από αυτό που βρέθηκε αν θεωρήσουμε ότι οι συμμετέχοντες έχουν ελαφριά δραστηριότητα (26%).

Από τα αποτελέσματα δημοσιευμένων μελετών εγκυρότητας<sup>18</sup> προκύπτει ότι τα ποσοστά υποεκτίμησης κυμαίνονται μεταξύ 19% έως 50%. Η υποεκτίμηση της πρόσληψης, που προέκυψε κατά τη σύγκριση του λόγου EI/BMR με τα ελάχιστα όρια, δεν διαφοροποιήθηκε σημαντικά μεταξύ των δύο φύλων, γεγονός που εντοπίστηκε και κατά την αξιολόγηση των 37 διατροφικών ερευνών στην μελέτη Black et al. Αντίθετα, εντοπίστηκαν σημαντικές διαφορές στην υποεκτίμηση της πρόσληψης μεταξύ υπέρβαρων, παχύσαρκων και ατόμων με φυσιολογικό βάρος. Τα παχύσαρκα και υπέρβαρα άτομα υποεκτιμούν σε μεγαλύτερο ποσοστό την πρόσληψη ενέργειας από ότι τα άτομα με φυσιολογικό βάρος, όπως προκύπτει από μια σειρά πειραματικών μελετών αξιολόγησης της ενεργειακής πρόσληψης με τη μέθοδο του διπλά επισημασμένου ύδατος<sup>18</sup> αλλά και όπως βρέθηκε και από τη δικιά μας μελέτη. Μερικά χρήσιμα συμπεράσματα από τη μελέτη μας, εκτός από αυτά που αναφέραμε παραπάνω, συνοψίζονται παρακάτω:

- ο λόγος της περιμέτρου μέσης προς την περίμετρο ισχύος (WHR), είναι σημαντικός παράγοντας για την πρόβλεψη της υποεκτίμησης, το οποίο είναι αναμενόμενο από τη στιγμή που ο λόγος αυτός αποτελεί έναν δείκτη για την παχυσαρκία.
- οι άνθρωποι μεγαλύτερης ηλικίας υποεκτιμούν σε μεγαλύτερο βαθμό την πρόσληψη ενέργειας όπως και αυτοί που έχουν στοιχειώδη μορφωτική εκπαίδευση.
- τέλος, η πρόσληψη θρεπτικών συστατικών (πρωτεΐνες, λιπίδια, υδατάνθρακες), όπως είναι αναμενόμενο, μειώνεται κατά πολύ στην ομάδα των υποεκτιμούντων ενώ αποτελούν και σημαντικούς παράγοντες στην εκτίμηση της πιθανότητας για να ανήκει ένα άτομο στην ομάδα των υποεκτιμούντων.

Όπως έχουμε επισημάνει παραπάνω, για κάθε μια από τις δραστηριότητες που αναφέρει ο συμμετέχων ότι πραγματοποιεί, ορίζουμε το μεταβολικό ισοδύναμό της. Οι τιμές των μεταβολικών ισοδυνάμων για κάθε δραστηριότητα προκύπτουν από δημοσιευμένους πίνακες. Για κάθε άτομο έχουμε ένα δείκτη 24ωρης φυσικής δραστηριότητας, που ορίζει τις θερμίδες που δαπανά κάθε άτομο ανά κιλό βάρους του, κατά τη διάρκεια ενός τυπικού 24ώρου.

Από τους συντελεστές συσχέτισης άλλα και από το μοντέλο ANCOVA που φτιάξαμε βρήκαμε ότι η σημασία της απασχόλησης σαν προβλεπτικός παράγοντας για τα MET είναι πολύ μεγάλη ενώ οι διαφοροποιήσεις του δείκτη 24ωρης δραστηριότητας σε σχέση με την ηλικία και την απασχόληση είναι προς την αναμενόμενη κατεύθυνση. Έτσι, όσον αφορά στην ηλικία εμφανίζεται ελάττωση της φυσικής δραστηριότητας στις μεγαλύτερες ηλικίες, κυρίως μετά την ηλικία των 65 χρόνων ενώ ως προς την απασχόληση παρατηρούμε ότι όσο πιο

βαριά εργασία κάνει ένα άτομο (π.χ. χειρονακτική εργασία) τόσο περισσότερα MET καταναλώνει, και τόσο πιο αυξημένος είναι ο δείκτης δραστηριότητας.

Συνοψίζοντας, μπορούμε να επισημάνουμε ότι η αξιολόγηση της εγκυρότητας των τιμών πρόσληψης επικεντρώθηκε αποκλειστικά στην κατεύθυνση της υποεκτίμησης, ενώ παραμένει αδιευκρίνιστο το πιθανό εύρος της υπερεκτίμησης, που πιθανώς να συνυπάρχει. Μια ενδεχόμενη προσπάθεια προς αυτή την κατεύθυνση, θα έπρεπε να λάβει υπόψη το επίπεδο φυσικής δραστηριότητας κάθε ατόμου, για να προσδιοριστούν τα αντίστοιχα αποδεκτά «ανώτερα όρια» της διατροφικής πρόσληψης. Εάν, πρέπει να διερευνηθεί σε μια μελέτη και η περίπτωση ύπαρξης υπερεκτίμησης της πρόσληψης αυτό γίνεται προσδιορίζοντας αρχικά απαραίτητως το εύρος της υποεκτίμησης<sup>18</sup>. Σε κάθε περίπτωση πάντως η αναγκαιότητα καθαρίσματος (cleaning) των δεδομένων που συλλέγονται σε μια διατροφική έρευνα είναι απαραίτητη.

## BIBΛΙΟΓΡΑΦΙΑ:

1. Slimani Nadia (2002): 24-hour dietary recalls as reference calibration measurements in EPIC: from statistical theory to epidemiological application. PhD-thesis Wageningen University.
2. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, Charrondiere UR, Hemon B, Casagrande C, Vignat J, Overvad K, Tjonneland A, Clavel-Chapelon F, Thiebaut A, Wahrendorf J, Boeing H, Trichopoulos D, Trichopoulou A, Vineis P, Palli D, Bueno-de-Mesquita HB, Peeters PHM, Lund E, Engeset D, Gonzalez CA, Barricate A, Berglund G, Hallmans G, Day NE, Key TJ, Kaaks R and Saracci R (2002): European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutrition*. **5(6B)**, 1113-1124.
3. Slimani N, Kaaks R, Ferrari P, Casagrande C, Clavel-Chapelon F, Lotze G, Kroke A, Trichopoulos D, Trichopoulou A, Lauria C, Bellegotti M, Ocke MC, Peeters PHM, Engeset D, Lund E, Agudo A, Larranaga N, Mattisson I, Andren C, Johanson I, Davey G, Welch AA, Overvad K, Tjonneland A, WA van Staveren, Saracci R and Riboli E (2002): European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study: rationale, design and population characteristics. *Public Health Nutrition*. **5(6B)**, 1125-1145.
4. Haftenberger M, Lahmann PH, Panico S, Gonzalez CA, Seidell JC, Boeing H, Giurdanella MC, Krogh V, Bueno-de-Mesquita HB, Peeters PHM, Skeie G, Hjartaker A, Rodriguez M, Quiros JR, Berglund G, Janlert U, Khaw KT, Spencer EA, Overvad K, Tjonneland A, Clavel-Chapelon F, Tehard B, Miller AB, Klipstein-Grobusch K, Benetou V, Kiriazi G, Riboli E and Slimani N (2002): Overweight, obesity and fat distribution in 50-to 64-year-old participants in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Public Health Nutrition*. **5(6B)**, 1147-1162.
5. Haftenberger M, Schuit AJ, Tormo MJ, Boeing H, Wareham N, Bueno-de-Mesquita HB, Kulme M, Hjartaker A, Chirlaque MD, Ardanaz E, Andren C, Lindahl B, Peeters PHM, Allen NE, Overvad K, Tjonneland A, Clavel-Chapelon F, Linseisen J, Bergmann MM, Trichopoulou A, Lagiou P, Salvini S, Panico S, Riboli E, Ferrari P and Slimani N (2002): Physical activity of subjects aged 50-64 years involved in the European Prospective

- Investigation into Cancer and Nutrition (EPIC). *Public Health Nutrition*. **5(6B)**, 1163-1177.
6. Agudo A, Slimani N, Ocke MC, Naska A, Miller AB, Kroke A, Bamia C, Karalis D, Vineis P, Palli D, Bueno-de-Mesquita HB, Peeters PHM, Engeset D, Hjartaker A, Navarro C, Martinez Garcia C, Wallstrom P, Zhang JX, Welch AA, Spencer E, Stripp C, Overvad K, Clavel-Chapelon F, Casagrande C and Riboli E (2002): Consumption of vegetables, fruit and other plant foods in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohorts from 10 European countries. *Public Health Nutrition*. **5(6B)**, 1179-1196.
  7. Wirfalt E, McTaggart A, Pala V, Gullberg B, Frasca G, Panico S, Bueno-de-Mesquita HB, Peeters PHM, Engeset D, Skeie G, Chirilaque MD, Amiano P, Lundin E, Mulligan A, Spencer EA, Overvad K, Tjonneland A, Clavel-Chapelon F, Linseisen J, Nothlings U, Polychronopoulos E, Georga K, Charrondiere UR and Slimani N (2002): Food sources of carbohydrates in a European cohort of adults. *Public Health Nutrition*. **5(6B)**, 1197-1215.
  8. Keinan-Boker L, Peeters PHM, Mulligan AA, Navarro C, Slimani N and the EPIC Study Group on Soy Consumption: Mattisson I, Lundin E, Mc Taggart A, Allen NE, Overvad K, Tjonneland A, Clavel-Chapelon F, Linseisen J, Haftenberger M, Lagiou P, Kalapothaki V, Evangelista A, Frasca G, Bueno-de-Mesquita HB, YT van der Schouw, Engeset D, Skeie G, Tormo MJ, Ardanaz E, Charrondiere UR and Riboli E (2002): Soy product consumption in 10 European countries: the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Public Health Nutrition*. **5(6B)**, 1217-1226.
  9. Linseisen J, Bergstrom E, Gafa L, Gonzalez CA, Thiebaut A, Trichopoulou A, Tumino R, Navarro Sanchez C, Martinez Garcia C, Mattisson I, Nilsson S, Welch A, Spencer EA, Overvad K, Tjonneland A, Clavel-Chapelon F, Kesse E, Miller AB, Schulz M, Botsi K, Naska A, Sieri S, Sacerdote C, Ocke MC, Peeters PHM, Skeie G, Engeset D, Charrondiere UR and Slimmani N (2002): Consumption of added fats and oils in the European Prospective Investigation into Cancer and Nutrition (EPIC) centers across 10 European countries as assessed by 24-hour dietary recalls. *Public Health Nutrition*. **5(6B)**, 1227-1242.
  10. Linseisen J, Kesse E, Slimani N, Bueno-de-Mesquita HB, Ocke MC, Skeie G, Kulme M, Dorronsoro Iraeta M, Morote Gomez P, Janzon L, Stattin P, Welch AA, Spencer EA, Overvad K, Tjonneland A, Clavel-Chapelon F, Miller AB, Klipstein-Grobusch K, Lagiou

- P, Kalapothaki V, Masala G, Giurdanella MC, Norat T and Riboli E (2002): Meat consumption in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohorts: results from 24-hour dietary recalls. *Public Health Nutrition*. **5(6B)**, 1243-1258.
11. Hjartaker A, Lagiou A, Slimani N, Lund E, Dhirlaque MD, Vasilopoulou E, Zavitsanos X, Berrino F, Sacerdote C, Ocke MC, Peeters PHM, Engeset D, Skeie G, Aller A, Amiano P, Berglund G, Nilsson S, McTaggart A, Spencer EA, Overvad K, Tjønneland A, Clavel-Chapelon F, Linseisen J, Schulz M, Hemon B and Riboli E (2002): Consumption of dairy products in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort: data from 35955 24-hour dietary recalls in 10 European countries. *Public Health Nutrition*. **5(6B)**, 1259-1271.
12. Welch AA, Lund E, Amiano P, Dorronsoro M, Brustad M, Kulme M, Rodriguez M, Lasheras C, Janzon L, Jansson J, Luben R, Spencer EA, Overvad K, Tjønneland A, Clavel-Chapelon F, Linseisen J, Klipstein-Grobusch K, Benetou V, Zavitsanos X, Tumino R, Galasso R, Bueno-de-Mesquita HB, Ocke MC, Charrondiere UR and Slimani N (2002): Variability of fish consumption within the 10 European countries participating in the European Investigation into Cancer and Nutrition (EPIC) study. *Public Health Nutrition*. **5(6B)**, 1273-1285.
13. Sieri S, Agudo A, Kesse E, Klipstein-Grobusch K, San-Jose B, Welch AA, Krogh V, Luben R, Allen N, Overvad K, Tjønneland A, Clavel-Chapelon F, Thiebaut A, Miller AB, Boeing H, Kolyva M, Saiena C, Celentano E, Ocke MC, Peeters PHM, Brustad M, Kulme M, Dorronsoro M, Fernandez Feito A, Mattisson I, Weinehall L, Riboli E and Slimani N (2002): Patterns of alcohol consumption in 10 European countries participating in the European Prospective Investigation into Cancer and Nutrition (EPIC) project. *Public Health Nutrition*. **5(6B)**, 1287-1296.
14. Klipstein-Grobusch K, Slimani N, Krogh V, Keil U, Boeing H, Overvad K, Tjønneland A, Clavel-Chapelon F, Thiebaut A, Linseisen J, Schulze MB, Lagiou P, Papadimitrou A, Saieva C, Veglia F, Bueno-de-Mesquita HB, Peeters PHM, Kumle M, Brustad M, Martinez Garcia C, Barricarte A, Berglund G, Weinehall L, Mulligan A, Allen N, Ferrari P and Riboli E (2002): Trends in self-reported past alcoholic beverage consumption and ethanol intake from 1950 to 1995 observed in eight European countries participating in the European Investigation into Cancer and Nutrition (EPIC). *Public Health Nutrition*. **5(6B)**, 1297-1310.

- 15.Slimani N, Fahey M, Welch AA, Wirfalt E, Stripp C, Bergstrom E, Linseisen, Schulze MB, Bamia C, Chloptsios Y, Veglia F, Panico S, Bueno-de-Mesquita HB, Ocke MC, Brustad M, Lund E, Gonzalez CA, Barcos A, Berglund G, winkvist A, Mulligan A, Appleby P, Overvad K, Tjonneland A, Clavel-Chapelon F, Kesse E, Ferrari P, WA Van Staveren and Riboli E (2002): Diversity of dietary patterns observed in the European Prospective Investigation into Cancer and Nutrition (EPIC) project. *Public Health Nutrition*. **5(6B)**, 1311-1328.
- 16.Goldberg G.R, Black A.E, Jebb S.A, Cole T.J, Murgatroyd P.R, Coward W.A and Prentice A.M. (1991): Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. *European Journal of Clinical Nutrition*. **45**, 569-581.
- 17.Black Alison E. (2000): Critical evaluation of energy intake using the Goldberg cut-off for EI:BMR. A practical guide to its calculation, use and limitations. *International Journal of Obesity and Related Metabolic Disorders*. **24(9)**, 1119-1130
- 18.Γναρδέλλης Χ, Μπουλού Χ, Δανιά Ρ, Τριχοπούλου Α. (1998): Στάθμιση της εγκυρότητας διατροφικών δεδομένων από την έρευνα ΕΠΙΚ μέσω του δείκτη φυσικής δραστηριότητας των συμμετεχόντων. *Ιατρική*. **73(6)**, 497-503.
- 19.Gnardellis Charalambos, Boulou Chariklia and Trichopoulou Antonia. (1998): Magnitude, determinants and impact of under-reporting of energy intake in a cohort study in Greece. *Public Health Nutrition*. **1(2)**, 131-137.
- 20.Wareham J Nicholas, Jakes W Rupert, Rennie L Kirsten, Schuit Jantine, Mitchell Jo, Hennings Susie and Day E Nicholas. (2003):Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Public Health Nutrition*. **6(4)**, 407-413.
- 21.Γναρδέλλης Χ, Λάγιου Α, Χλόπτσιος Ι, Μπενέτου Β, Τριχοπούλου Α. (1999) Εκτίμηση φυσικής δραστηριότητας σε επιδημιολογικές μελέτες-Η ελληνική εμπειρία στο πλαίσιο του προγράμματος ΕΠΙΚ. *Ιατρική*. **76(6)**, 551-558.
- 22.Black AE. (2000):The sensitivity and specificity of the Goldberg cut-off for EI:BMR for indentifying diet reports of poor validity. *European Journal of Clinical Nutrition*. **54**, 395-404.



23. Rabe-Hesketh Sophia, Everitt Brian (2000): A handbook of Statistical Analyses using Stata. Chapman & Hall/CRC, London.
24. Παπαϊωάννου Τ, Φερεντίνος Κ. (2000): Ιατρική Στατιστική και Στοιχεία Βιομαθηματικών. Εκδόσεις Αθ. Σταμούλη, Αθήνα.
25. Agresti Alan. (2002): Categorical Data Analysis. Wiley, Canada-New Jersey.
26. Bingham SA (1987): The dietary assessment of individuals; methods, accuracy, new techniques and recommendations. *Nutrition Abstracts and Reviews*. **57**, 705-742.
27. Nelson M, Black AE, Morris JA & Cole TJ (1989): Between- and within-subject variation in nutrient intake from infancy to old age: estimating the numbers of days required to rank dietary intakes with desired precision. *American Journal of Clinical Nutrition*. **50**, 155-167.
28. Schofield WN, Schofield C & James WPT (1985): Basal metabolic rate. *Human Nutrition. Clinical Nutrition*. **39C (Suppl.1)**, 1-96.
29. Prentice AM, Goldberg GR, Davies HL, Murgatroyd PR & Scott W (1989): Energy-sparing adaptations in human pregnancy assessed by whole-body calorimetry. *British Journal of Nutrition*. **62**, 5-22.
30. FAO/WHO/UNU (1985): Report of a joint expert consultation. Energy and Protein Requirements. *WHO Tech. Rep. Series No. 724*. WHO: Geneva