

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ
ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΕΦΑΡΜΟΓΗ ΤΕΧΝΙΚΩΝ ΕΛΑΤΤΩΣΗΣ ΔΙΑΣΤΑΣΗΣ
ΠΟΛΥΔΙΑΣΤΑΤΩΝ ΔΕΔΟΜΕΝΩΝ ΣΤΗ ΔΗΜΟΓΡΑΦΙΑ**

ΖΟΥΛΙΑ ΕΥΓΕΝΙΑ

Διπλωματική Εργασία

Που υποβλήθηκε στο τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιά ως μέρος των απαιτήσεων
για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης
στην Εφαρμοσμένη Στατιστική

ΠΕΙΡΑΙΑΣ

ΣΕΠΤΕΜΒΡΙΟΣ 2017

UNIVERSITY OF PIRAEYS



DEPARTMENT OF STATISTICS AND INSURANCE SCIENCE

POSTGRADUATE PROGRAM IN APPLIED STATISTICS

***APPLICATION OF DIMENSIONALITY REDUCTION
TECHNIQUES IN DEMOGRAPHY***

By

ZOULIA EVGENIA

MSc Dissertation

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

PIRAEUS, GREECE

SEPTEMBER 2017

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:

- Μάρκος Κούτρας (Επιβλέπων)
- Κλέων Τσίμπος
- Γεωργία Βερροπούλου

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

ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία έχει ως σκοπό να αναδείξει με γραφικό τρόπο τις συσχετίσεις που ενδέχεται να υπάρχουν σε ένα πλήθος κατηγορικών μεταβλητών. Πιο συγκεκριμένα, έχουμε στη διάθεσή μας δημογραφικά χαρακτηριστικά καθώς και στοιχεία για την υγεία ηλικιωμένων άνω των 50 ετών σε διάφορες χώρες της Ευρώπης το έτος 2013. Η μέθοδος που θα χρησιμοποιηθεί ονομάζεται Πολλαπλή Παραγοντική Ανάλυση Αντιστοιχιών και αποτελεί τη γενίκευση της Παραγοντικής Ανάλυσης Αντιστοιχιών που χρησιμοποιείται για την ανάλυση δύο μόνο κατηγορικών μεταβλητών. Μέσα από την παρούσα μελέτη αναδείχθηκε συσχέτιση μεταξύ ομάδων ανθρώπων με συγκεκριμένα δημογραφικά χαρακτηριστικά και την παρουσία ή την απουσία συγκεκριμένων ασθενειών.

Το πρώτο κεφάλαιο έχει ως στόχο να ενημερώσει τον αναγνώστη σχετικά με τους ιδρυτές της μεθόδου και την εξέλιξή της στο χρόνο. Στο δεύτερο κεφάλαιο παρουσιάζεται το θεωρητικό υπόβαθρο της μεθόδου, αρχικά για δύο μεταβλητές και έπειτα γενικεύεται στην περίπτωση των πολλών μεταβλητών. Στο τρίτο κεφάλαιο παρουσιάζουμε το σύνολο των δεδομένων, ενώ στο τέταρτο κεφάλαιο εφαρμόζουμε τη μέθοδο δοκιμάζοντας διάφορες παραλλαγές της ενώ παρουσιάζουμε και συγκρίνουμε τα αποτελέσματα που προκύπτουν. Τέλος, στο πέμπτο κεφάλαιο παραθέτουμε τα γενικά συμπεράσματα όσον αφορά τη μορφή των δεδομένων και τις συσχετίσεις που ανακαλύφθηκαν από τις διάφορες αναλύσεις.

ABSTRACT

The main purpose of the below presented dissertation is to graphically highlight the correlations that may exist in a number of categorical variables. More specifically, we have in our disposal demographic and health data for older people over 50 years of age in different European countries in 2013. The method that will be used is called Multiple Correspondence Analysis and is the generalization of Factorial Correspondence Analysis used for the analysis of only two categorical variables. Through this study, an association was emerged between groups of people with specific demographic characteristics and the presence or absence of specific diseases.

The first chapter aims to inform the reader about the founders of the method and its evolution over time. The second chapter presents the theoretical background of the method, initially for two variables and then generalizes in the multivariate case. In the third chapter we present our data set, while in the fourth chapter we apply the method by testing various models while presenting and comparing the results obtained. Finally, in the fifth chapter we present the general conclusions regarding the structure of the data and the correlations discovered by the various analyses.

ΠΕΡΙΕΧΟΜΕΝΑ ΤΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ

ΠΕΡΙΛΗΨΗ.....	v
ABSTRACT.....	vii
ΠΕΡΙΕΧΟΜΕΝΑ ΤΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ	ix
ΚΕΦΑΛΑΙΟ 1	1
Εισαγωγή και Ιστορική Αναδρομή	1
1.1 Η έννοια της Παραγοντικής Ανάλυσης Αντιστοιχιών.....	1
1.2 Ιστορική Αναδρομή	2
ΚΕΦΑΛΑΙΟ 2	4
Παραγοντική Ανάλυση Αντιστοιχιών	4
2.1 Απλή Ανάλυση Αντιστοιχιών	4
2.2 Πολλαπλή Ανάλυση Αντιστοιχιών	25
ΚΕΦΑΛΑΙΟ 3	30
Δεδομένα.....	30
3.1 Παρουσίαση των μεταβλητών	30
3.2 Περιγραφικά Μέτρα Μεταβλητών.....	31
ΚΕΦΑΛΑΙΟ 4	46
Εφαρμογή της μεθόδου και παρουσίαση των αποτελεσμάτων.....	46
4.1 Εισαγωγή	46
4.2 Πολυμεταβλητή Ανάλυση Αντιστοιχιών στις δημογραφικές μεταβλητές.....	47
4.3 Πολυμεταβλητή Ανάλυση Αντιστοιχιών στις μεταβλητές των ασθενειών.....	57
ΚΕΦΑΛΑΙΟ 5	62
Συμπεράσματα	62

ΑΝΑΦΟΡΕΣ.....	63
Αναφορές ως προς τη χρήση των δεδομένων	63
ΒΙΒΛΙΟΓΡΑΦΙΑ	64
Ελληνικές Βιβλιογραφικές Αναφορές.....	64
Ξένες Βιβλιογραφικές Αναφορές.....	64
ΠΑΡΑΡΤΗΜΑ	66
Π1. Αναλυτικός πίνακας συχνοτήτων ανά ηλικία	66
Π2. Αναλυτικός πίνακας συχνοτήτων ανά έτη εκπαίδευσης.....	67
Π3. Συγκεντρωτικός πίνακας συχνοτήτων ασθενειών.....	68
Π4. Αδράνεια και δείκτες που προκύπτουν από την 1 ^η ανάλυση των δημογραφικών μεταβλητών.....	69
Π5. Αδράνεια και δείκτες που προκύπτουν από την 2 ^η ανάλυση των δημογραφικών μεταβλητών.....	70
Π6. Αδράνεια και δείκτες που προκύπτουν από την 3 ^η ανάλυση των δημογραφικών μεταβλητών.....	70
Π7. Αδράνεια και δείκτες που προκύπτουν από την 4 ^η ανάλυση των δημογραφικών μεταβλητών.....	71
Π8. Αδράνεια και δείκτες που προκύπτουν από την τελική ανάλυση των δημογραφικών μεταβλητών.....	72
Π9. Αδράνεια και δείκτες που προκύπτουν από την ανάλυση των μεταβλητών των ασθενειών	74

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή και Ιστορική Αναδρομή

1.1 Η έννοια της Παραγοντικής Ανάλυσης Αντιστοιχιών

Η Παραγοντική Ανάλυση Αντιστοιχιών (Factorial Correspondence Analysis) χρησιμοποιείται για να αναλύσει τις σχέσεις που υπάρχουν μεταξύ δύο ή περισσότερων κατηγορικών μεταβλητών. Στην περίπτωση που έχουμε 2 μεταβλητές χρησιμοποιούμε την Απλή Παραγοντική Ανάλυση Αντιστοιχιών (CA) ενώ στην περίπτωση των περισσότερων μεταβλητών χρησιμοποιούμε την Πολλαπλή Παραγοντική Ανάλυση Αντιστοιχιών (MCA). Ο όρος παραγοντική προσδίδεται καθώς βασιζόμενοι στις αρχικές μεταβλητές δημιουργούμε ένα μικρό αριθμό «παραγόντων». Η μέθοδος αυτή είναι το αντίστοιχο της Ανάλυσης Κυρίων Συνιστωσών (Principal Component Analysis, PCA), η οποία χρησιμοποιείται στην περίπτωση της ανάλυσης συνεχών μεταβλητών.

Η χρησιμότητα της ανάλυσης αντιστοιχιών εντοπίζεται κυρίως στην περίπτωση όπου θέλουμε να αναλύσουμε δεδομένα με μεγάλο αριθμό κατηγορικών μεταβλητών τις οποίες είναι αδύνατο να αναπαραστήσουμε γραφικά με στόχο να διερευνήσουμε τις σχέσεις μεταξύ τους. Επιπροσθέτως, αποκτούμε το πλεονέκτημα της ταυτόχρονης σύγκρισης όλων των μεταβλητών. Η γενική ιδέα της μεθόδου είναι να ξεκινήσουμε από έναν πολυδιάστατο πίνακα και να τον μετατρέψουμε, χωρίς να χάσουμε σημαντική πληροφορία, σε έναν πίνακα μικρότερων διαστάσεων κατά προτίμηση δισδιάστατο ή τρισδιάστατο, οπότε είναι δυνατό να έχουμε γραφική αναπαράσταση σε ένα διάγραμμα διασποράς. Εξετάζοντας το διάγραμμα διασποράς είναι δυνατό να εντοπίσουμε σχέσεις ή συστάδες (clusters) μεταξύ των κατηγοριών (επιπέδων των κατηγορικών μεταβλητών) και να ερμηνεύσουμε αυτές τις σχέσεις εξετάζοντας τις σχετικές θέσεις των κατηγοριών.

Στη μέθοδο αυτή δεν χρησιμοποιούμε κάποιο μοντέλο πιθανοτήτων στο οποίο προσαρμόζουμε τα δεδομένα, επομένως μπορούμε να ισχυριστούμε ότι είναι μία

περιγραφική ή διερευνητική τεχνική. Το βασικό της πλεονέκτημα είναι ότι δεν χρειάζεται να υποθέσουμε κάποια θεωρητική κατανομή όμως ταυτόχρονα υστερεί σε ακρίβεια ως προς την ερμηνεία των αποτελεσμάτων και επιπλέον δεν είναι δυνατό να έχουμε στατιστική συμπερασματολογία για το σύνολο του πληθυσμού. Η ανάλυση αντιστοιχιών χρησιμοποιείται ευρέως στην κοινωνιολογία, το marketing, την ψυχολογία αλλά και την βιολογία. Στην παρούσα εργασία θα ασχοληθούμε με δημογραφικά δεδομένα.

1.2 Ιστορική Αναδρομή

Η μέθοδος της Ανάλυσης Αντιστοιχιών ξεκίνησε στις αρχές του 20ου αιώνα και οφείλει το όνομά της στη Γαλλική μετάφραση του όρου “Analyses des Correspondences”. Πιο συγκεκριμένα το 1935 ο Hirschfeld (γνωστός και ως H. O. Hartley) ασχολήθηκε για πρώτη φορά με θεωρητικά ζητήματα δίνοντας μία αλγεβρική σύνθεση των συσχετίσεων μεταξύ των γραμμών και των στηλών ενός πίνακα συνάφειας διπλής εισόδου. Θεωρήθηκε ο ιδρυτής της μεθόδου αυτής, ενώ βασίστηκε κυρίως στα έργα του Karl Pearson και του Hotelling που αφορούσαν την Ανάλυση Κυρίων Συνιστωσών.

Στα πρώτα της στάδια, η μέθοδος αναπτυσσόταν στην Ιαπωνία από τον Chikio Hayashi (1950) και σε άλλες χώρες όπως η Γαλλία, η Αγγλία, η Αμερική και η Ολλανδία. Μερικοί από τους στατιστικούς που την εισήγαγαν στις διάφορες χώρες ήταν οι Horst (1935), Fisher (1940), Guttman (1941) και Burt (1950).

Πιο συγκεκριμένα, ο Fisher το 1940 εφάρμοσε μία μέθοδο με τη μορφή της ανάλυσης διακρίσεων (Discriminant Analysis) σε έναν πίνακα συνάφειας διπλής εισόδου που εξέταζε τη συσχέτιση μεταξύ του χρώματος ματιών και μαλλιών μίας ομάδας μαθητών. Η μέθοδος αυτή ονομάστηκε Αμοιβαία Μεσοστάθμιση (Reciprocal Averaging) από τον Horst. Ο Louis Guttman το 1941 εφάρμοσε τη μέθοδο Διπλής ή Βέλτιστης Κλιμακοποίησης (Dual or Optimal Scaling) η οποία στη συνέχεια χαρακτηρίστηκε ως θεμέλιο της Πολλαπλής Ανάλυσης Αντιστοιχιών. Οι δύο προαναφερθείσες μέθοδοι έχουν την ίδια μαθηματική θεωρία με την Ανάλυση Αντιστοιχιών όμως καταλήγουν σε αριθμητικά αποτελέσματα σε αντίθεση με την Ανάλυση Αντιστοιχιών που δίνει γραφικά αποτελέσματα. Μία σημαντική εξέλιξη της μεθόδου Ανάλυσης Αντιστοιχιών

πραγματοποιήθηκε το 1950 από τον Burt, καθώς δημιούργησε τον πίνακα Burt και χρησιμοποιώντας τον στις εφαρμογές που αφορούσαν τη μέθοδο της Πολλαπλής Ανάλυσης Αντιστοιχιών κατάφερε να βελτιώσει να ποσοστά της ερμηνευμένης αδράνειας στο χώρο των μειωμένων διαστάσεων.

Η πρώτη επίσημη εμφάνιση της μεθόδου της Ανάλυσης Αντιστοιχιών, που μπορεί να θεωρηθεί ως η γεωμετρική μορφή των προηγούμενων μεθόδων, έγινε το 1960 από τον Γάλλο Jean-Paul Benzecri του πανεπιστημίου University Pierie-Marie-Curie. Ο Benzecri εστίασε στον τομέα της γλωσσολογίας και το πρώτο σύνολο δεδομένων που ανέλυσε είχε τα σύμφωνα στις γραμμές, της καταλήξεις φωνηέντων στις στήλες και μελέτησε εάν υπάρχουν ενδείξεις ότι επιτρέπεται να χρησιμοποιούνται οι διάφοροι συνδυασμοί μεταξύ τους. Στη Γαλλία η μέθοδος γνώρισε μεγάλη αποδοχή και χρησιμοποιήθηκε ευρέως. Την ίδια περίοδο, στην Ιαπωνία ασχολήθηκαν με τη μέθοδο ο Komazawa και ο Nishisato το 1982 ενώ στην Αγγλία οι Lebart, Morineau, Warwick (1984), ο Greenacre (1984) και ο Gifi (1990).

Σημαντικό εμπόδιο στην ανάπτυξη της μεθόδου αποτέλεσε το γεγονός ότι οι διάφορες σχολές χρησιμοποιούσαν διαφορετικές γλώσσες. Αυτό το πρόβλημα λύθηκε το 1984 όπου ο M. J. Greenacre έγραψε και στη συνέχεια έκδωσε το πρώτο αγγλόφωνο βιβλίο της μεθόδου όπως αυτή είχε αναπτυχθεί στη Γαλλία με τίτλο “Theory and Applications of Correspondence Analysis”.

Η μέθοδος της Ανάλυσης Αντιστοιχιών μπορεί να εφαρμοσθεί με επιτυχία σε μεγάλα σύνολα δεδομένων και επομένως η χρησιμότητά της ευνοήθηκε με την εμφάνιση και την εξέλιξη των ηλεκτρονικών υπολογιστών. Επίσης γνώρισε μεγάλη εξάπλωση όταν εισάχθηκε σαν επιλογή σε Αμερικάνικα στατιστικά πακέτα όπως το SPSS και το SAS.

ΚΕΦΑΛΑΙΟ 2

Παραγοντική Ανάλυση Αντιστοιχιών

2.1 Απλή Ανάλυση Αντιστοιχιών

Τα δεδομένα που έχουμε στη διάθεσή μας όταν χρησιμοποιούμε την απλή ανάλυση αντιστοιχιών αποτελούνται από δύο μεταβλητές, κάθε μία από τις οποίες αποτελείται από ένα πλήθος κατηγοριών (επιπέδων). Αρχικά δημιουργούμε τον **πίνακα συνάφειας** N που έχει τη μία μεταβλητή στις γραμμές, την άλλη στις στήλες και τα κελιά του περιέχουν τη συχνότητα με την οποία εμφανίζονται ταυτόχρονα οι αντίστοιχες κατηγορίες στις παρατηρήσεις που έχουμε συλλέξει. Πιο συγκεκριμένα, ο πίνακας συνάφειας θα έχει τη μορφή:

	1	2	...	j	...	J	
1	n_{11}	.	.	n_{1j}	.	n_{1J}	n_{1+}
2
...
i	n_{i1}	.	.	n_{ij}	.	.	n_{i+}
...
I	n_{I1}	n_{IJ}	n_{I+}
	n_{+1}	.	.	n_{+j}	.	n_{+J}	N

Πίνακας 1: Ο πίνακας συνάφειας N

όπου με n_{ij} , ($i = 1, 2, \dots, I$ και $j = 1, 2, \dots, J$) συμβολίζουμε τη συχνότητα με την οποία εμφανίζεται η συνδυασμένη κατηγορία (i, j) . Τα σύμβολα n_{i+} και n_{+j} χρησιμοποιούνται για να δηλώσουν τις περιθώριες συχνότητες της κάθε κατηγορίας, δηλαδή $n_{i+} = \sum_j n_{ij}$ και $n_{+j} = \sum_i n_{ij}$. Τέλος, με N θα συμβολίσουμε το άθροισμα όλων των στοιχείων του πίνακα που συμπίπτει και με το συνολικό μέγεθος του δείγματος, δηλαδή το πλήθος των παρατηρήσεων. Επομένως:

$$N = \sum_i \sum_j n_{ij} = \sum_j n_{i+} = \sum_i n_{+j}.$$

Το επόμενο βήμα είναι να δημιουργήσουμε έναν ακόμη πίνακα, τον **πίνακα αντιστοιχιών \mathbf{P}** , διαιρώντας όλα τα κελιά του πίνακα συνάφειας \mathbf{N} με το συνολικό μέγεθος του δείγματος N . Επομένως η αντιστοιχία που υπάρχει μεταξύ των δύο πινάκων είναι:

$$\mathbf{P} = N^{-1}\mathbf{N} = N^{-1}\{n_{ij}\}_{I \times J} = \left\{ \frac{n_{ij}}{N} \right\}_{I \times J} = \{p_{ij}\}_{I \times J}$$

όπου $p_{ij} = n_{ij}/N$.

Το άθροισμα όλων των κελιών του νέου αυτού πίνακα είναι ίσο με τη μονάδα την οποία ονομάζουμε **μάζα** του πίνακα. Οι τιμές των κελιών του πίνακα \mathbf{P} μας δείχνουν με ποιον τρόπο κατανέμεται η συνολική (μοναδιαία) μάζα στα στοιχεία του πίνακα. Ο πίνακας αυτός φαίνεται παρακάτω:

	1	2	...	j	...	J	\overline{mr}
1	p_{11}	·	·	p_{1j}	·	p_{1J}	p_{1+}
2	·	·	·	·	·	·	·
·	·	·	·	·	·	·	·
i	p_{i1}	·	·	p_{ij}	·	·	p_{i+}
·	·	·	·	·	·	·	·
I	p_{I1}	·	·	·	·	p_{IJ}	p_{I+}
\overline{mc}	p_{+1}	·	·	p_{+j}	·	p_{+J}	1

Πίνακας 2: Ο πίνακας αντιστοιχιών \mathbf{P}

Τα στοιχεία p_{ij} ($i = 1, 2, \dots, I$ και $j = 1, 2, \dots, J$) του πίνακα δίνουν το ποσοστό των παρατηρήσεων που ανήκουν στη συνδυασμένη κατηγορία (i, j) . Τα περιθώρια αθροίσματα p_{i+} , p_{+j} αποτελούν τα διανύσματα:

$$\vec{m\bar{r}} = \begin{pmatrix} mr_1 \\ mr_2 \\ \cdot \\ \cdot \\ \cdot \\ mr_I \end{pmatrix} = \begin{pmatrix} p_{1+} \\ p_{2+} \\ \cdot \\ \cdot \\ \cdot \\ p_{I+} \end{pmatrix}$$

και

$$\vec{m\bar{c}} = (mc_1, mc_2, \cdot \cdot \cdot mc_J) = (p_{+1}, p_{+2}, \cdot \cdot \cdot p_{+J})$$

τα οποία μας δείχνουν την κατανομή της μάζας των γραμμών $\vec{m\bar{r}}$ και των στηλών $\vec{m\bar{c}}$ στον πίνακα και λέγονται διανύσματα μαζών των γραμμών και των στηλών αντίστοιχα.

Στο επόμενο βήμα θα δημιουργήσουμε άλλον έναν πίνακα, τον **πίνακα των προφίλ των γραμμών \mathbf{R}** , διαιρώντας το κάθε κελί του πίνακα συνάφειας \mathbf{N} με το περιθώριο άθροισμα της εκάστοτε γραμμής, όπως φαίνεται παρακάτω:

$$\mathbf{R} = \begin{bmatrix} \frac{n_{11}}{n_{1+}} & \frac{n_{12}}{n_{1+}} & \cdot & \cdot & \cdot & \frac{n_{1J}}{n_{1+}} \\ \frac{n_{21}}{n_{2+}} & \frac{n_{22}}{n_{2+}} & \cdot & \cdot & \cdot & \frac{n_{2J}}{n_{2+}} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{n_{I1}}{n_{I+}} & \frac{n_{I2}}{n_{I+}} & \cdot & \cdot & \cdot & \frac{n_{IJ}}{n_{I+}} \\ \frac{n_{I+}}{n_{I+}} & \frac{n_{I+}}{n_{I+}} & \cdot & \cdot & \cdot & \frac{n_{I+}}{n_{I+}} \end{bmatrix} = \begin{bmatrix} \vec{r}_1 \\ \vec{r}_2 \\ \cdot \\ \cdot \\ \cdot \\ \vec{r}_I \end{bmatrix} = \mathbf{D}_{r+}^{-1} \mathbf{N}$$

όπου $\vec{r}_i = (r_{i1}, r_{i2}, \dots, r_{iJ})$ με $r_{ij} = \frac{n_{ij}}{n_{i+}}$, ($i = 1, 2, \dots, I$ και $j = 1, 2, \dots, J$) και \mathbf{D}_{r+} είναι διαγώνιος πίνακας που περιέχει στη διαγώνιο τα περιθώρια αθροίσματα των γραμμών του πίνακα συνάφειας \mathbf{N} . Ο πίνακας σχετικών συχνοτήτων \mathbf{R} μας δείχνει την κατανομή των μαζών ανά γραμμή και καθιστά δυνατή την άμεση σύγκριση μεταξύ των γραμμών του αρχικού πίνακα συνάφειας \mathbf{N} . Στην παραγοντική ανάλυση το διάνυσμα \vec{r}_i που παράγεται από την κάθε γραμμή του πίνακα \mathbf{R} λέγεται προφίλ της i -οστής γραμμής. Το σύνολο των σημείων που καθορίζεται από τα προφίλ των I γραμμών του πίνακα \mathbf{R} , δηλαδή το $\{\vec{r}_1, \vec{r}_2, \dots, \vec{r}_I\}$ ονομάζεται **νέφος των προφίλ των γραμμών**.

Το κέντρο βάρους του νέφους των προφίλ των γραμμών ονομάζεται **σταθμισμένο μέσο προφίλ των γραμμών** και θα συμβολίζεται με $\vec{a\bar{r}}$. Κάθε συντεταγμένη του προκύπτει σαν σταθμισμένος μέσος όρος των συντεταγμένων των προφίλ των γραμμών στην εκάστοτε διάσταση με βάρη ίσα με τις μάζες των γραμμών. Τα προφίλ αποτελούν J -διάστατα διανύσματα που ανήκουν σε ένα χώρο διάστασης $J - 1$, εφόσον το περιθώριο άθροισμα του κάθε προφίλ είναι σταθερό και ίσο με τη μονάδα. Επομένως καταλήγουμε στην έκφραση:

$$\begin{aligned}\vec{a\bar{r}} &= \vec{m\bar{r}}^T \mathbf{R} = (mr_1, \quad mr_2, \quad \dots \quad mr_I) \begin{pmatrix} \vec{r}_1 \\ \vec{r}_2 \\ \vdots \\ \vec{r}_I \end{pmatrix} \\ &= \begin{pmatrix} r_{11}mr_1 + r_{21}mr_2 + \dots + r_{I1}mr_I \\ r_{12}mr_1 + r_{22}mr_2 + \dots + r_{I2}mr_I \\ \vdots \\ r_{1J}mr_1 + r_{2J}mr_2 + \dots + r_{IJ}mr_I \end{pmatrix}^T\end{aligned}$$

Μπορούμε επίσης να αποδείξουμε ότι το κέντρο βάρους του νέφους των γραμμών συμπίπτει με το προφίλ της περιθώριας στήλης του πίνακα συνάφειας \mathbf{N} το οποίο όπως φαίνεται στον Πίνακα 2 μας δίνει επίσης τις μάζες των στηλών του πίνακα \mathbf{R} . Πράγματι, εφόσον $r_{ij} = \frac{n_{ij}}{n_{i+}}$ και $mr_i = p_{i+} = \frac{n_{i+}}{N}$, ($i = 1, 2, \dots, I$ και $j = 1, 2, \dots, J$) η παραπάνω ισότητα μπορεί να γραφεί ως:

$$\begin{aligned}\vec{a\bar{r}} &= \begin{pmatrix} \frac{n_{11}}{n_{1+}} \frac{n_{1+}}{N} + \frac{n_{21}}{n_{2+}} \frac{n_{2+}}{N} + \dots + \frac{n_{I1}}{n_{I+}} \frac{n_{I+}}{N} \\ \frac{n_{12}}{n_{1+}} \frac{n_{1+}}{N} + \frac{n_{22}}{n_{2+}} \frac{n_{2+}}{N} + \dots + \frac{n_{I2}}{n_{I+}} \frac{n_{I+}}{N} \\ \vdots \\ \frac{n_{1J}}{n_{1+}} \frac{n_{1+}}{N} + \frac{n_{2J}}{n_{2+}} \frac{n_{2+}}{N} + \dots + \frac{n_{IJ}}{n_{I+}} \frac{n_{I+}}{N} \end{pmatrix}^T = \begin{pmatrix} \frac{n_{11}}{N} + \frac{n_{21}}{N} + \dots + \frac{n_{I1}}{N} \\ \frac{n_{12}}{N} + \frac{n_{22}}{N} + \dots + \frac{n_{I2}}{N} \\ \vdots \\ \frac{n_{1J}}{N} + \frac{n_{2J}}{N} + \dots + \frac{n_{IJ}}{N} \end{pmatrix}^T = \begin{pmatrix} \frac{n_{+1}}{N} \\ \frac{n_{+2}}{N} \\ \vdots \\ \frac{n_{+J}}{N} \end{pmatrix}^T \\ &= (p_{+1}, \quad p_{+2}, \quad \dots \quad p_{+J}) = \vec{m\bar{c}}\end{aligned}$$

Σε αντιστοιχία με τον πίνακα των προφίλ των γραμμών μπορούμε να ορίσουμε και τον **πίνακα των προφίλ των στηλών** \mathbf{C} , διαιρώντας το κάθε κελί του πίνακα συνάφειας \mathbf{N} με το περιθώριο άθροισμα της εκάστοτε στήλης, όπως φαίνεται παρακάτω:

$$\mathbf{C} = \begin{bmatrix} \frac{n_{11}}{n_{+1}} & \frac{n_{12}}{n_{+2}} & \cdot & \cdot & \cdot & \frac{n_{1J}}{n_{+J}} \\ \frac{n_{21}}{n_{+1}} & \frac{n_{22}}{n_{+2}} & \cdot & \cdot & \cdot & \frac{n_{2J}}{n_{+J}} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{n_{I1}}{n_{+1}} & \frac{n_{I2}}{n_{+2}} & \cdot & \cdot & \cdot & \frac{n_{IJ}}{n_{+J}} \\ \frac{n_{+1}}{n_{+1}} & \frac{n_{+2}}{n_{+2}} & \cdot & \cdot & \cdot & \frac{n_{+J}}{n_{+J}} \end{bmatrix} = [\vec{c}_1, \vec{c}_2, \cdot, \cdot, \cdot, \vec{c}_J] = \mathbf{N}\mathbf{D}_{+c}^{-1}$$

όπου \mathbf{D}_{+c} είναι διαγώνιος πίνακας που περιέχει στη διαγώνιο τα περιθώρια αθροίσματα των στηλών του πίνακα συνάφειας \mathbf{N} . Ο πίνακας \mathbf{C} μας δείχνει την κατανομή των μαζών ανά στήλη. Το διάνυσμα που παράγεται από την κάθε στήλη του πίνακα \mathbf{C} λέγεται προφίλ της στήλης. Το σύνολο των σημείων που καθορίζεται από τα προφίλ των στηλών J ονομάζεται **νέφος των προφίλ των στηλών**. Το κέντρο βάρους του νέφους των προφίλ των στηλών ονομάζεται **σταθμισμένο μέσο προφίλ των στηλών** και θα συμβολίζεται με \vec{c} . Κάθε συντεταγμένη του προκύπτει σαν σταθμισμένος μέσος όρος των συντεταγμένων των προφίλ των στηλών στην εκάστοτε διάσταση με βάρη ίσα με τις μάζες των στηλών. Όμοια με προηγουμένως καταλήγουμε στην έκφραση:

$$\vec{c} = \mathbf{C}\vec{m} = (\vec{c}_1 \quad \vec{c}_2 \quad \cdot \quad \cdot \quad \cdot \quad \vec{c}_J) \begin{pmatrix} mc_1 \\ mc_2 \\ \cdot \\ \cdot \\ \cdot \\ mc_J \end{pmatrix} = \begin{pmatrix} c_{11}mc_1 + c_{12}mc_2 + \dots + c_{1J}mc_J \\ c_{21}mc_1 + c_{22}mc_2 + \dots + c_{2J}mc_J \\ \cdot \\ \cdot \\ \cdot \\ c_{I1}mc_1 + c_{I2}mc_2 + \dots + c_{IJ}mc_J \end{pmatrix}$$

Με όμοιο τρόπο με προηγουμένως μπορούμε να αποδείξουμε ότι το κέντρο βάρους του νέφους των στηλών συμπίπτει με το προφίλ της περιθώριας γραμμής του πίνακα \mathbf{R} .

Μία πολύ σημαντική πτυχή της ανάλυσης αντιστοιχιών είναι ο **λόγος συνάφειας** (contingency ratio), ο οποίος δείχνει τη συμμετρία μεταξύ των γραμμών και των στηλών. Δηλώνει ότι η αναλογία του κάθε κελιού του προφίλ των γραμμών προς το σταθμισμένο μέσο προφίλ των γραμμών είναι ίση με την αναλογία του αντίστοιχου κελιού του προφίλ των στηλών προς το σταθμισμένο μέσο προφίλ των στηλών. Γενικότερα, είτε διαλέξουμε να αναλύσουμε τις γραμμές είτε τις στήλες όλοι οι υπολογισμοί καθώς και τα αποτελέσματα που προκύπτουν είναι σε πλήρη αντιστοιχία. Στα πλαίσια της θεωρίας επιλέγουμε να αναφέρουμε την ανάλυση των γραμμών με αναφορά στις στήλες όπου χρειάζεται.

Ωστόσο, δημιουργώντας τον πίνακα των προφίλ αλλοιώνεται η κατανομή των δεδομένων. Η διατήρηση αυτής της πληροφωρίας επιτυγχάνεται χρησιμοποιώντας τις μάζες, καθώς αποτελούν βάρη που μας δείχνουν τη συμμετοχή της κάθε γραμμής, αντιστοίχως στήλης, στο σύνολο του δείγματος. Επομένως, κατά τη γραφική αναπαράσταση, τα προφίλ με τη μεγαλύτερη μάζα συνεισφέρουν περισσότερο στη δημιουργία του σταθμισμένου μέσου προφίλ που είναι το κέντρο βάρους τους¹.

Η Ανάλυση Αντιστοιχιών έχει ως σκοπό να αναγνωρίσει τυχόν συσχετίσεις που υπάρχουν μεταξύ των κατηγοριών των γραμμών ή αντίστοιχα των στηλών. Επομένως χρειάζεται να ορίσουμε ένα μέτρο απόστασης μεταξύ αυτών για να μπορέσουμε να τις συγκρίνουμε. Εφόσον κάθε γραμμή αντιπροσωπεύει ένα σημείο σε έναν πολυδιάστατο χώρο μπορούμε να υπολογίσουμε τις αποστάσεις αυτές. Ένα πολύ σημαντικό μέρος της ανάλυσης αντιστοιχιών είναι η επιλογή του μέτρου το οποίο θα χρησιμοποιήσουμε για να υπολογίσουμε τις αποστάσεις. Σαν πρώτη επιλογή θα σκεφτόμασταν την Ευκλείδεια απόσταση, η οποία όμως έχει ένα πολύ σημαντικό μειονέκτημα. Δίνεται η ίδια βαρύτητα σε όλες τις κατηγορίες ανεξαρτήτως της συχνότητάς τους. Αυτό το πρόβλημα λύνεται εάν χρησιμοποιήσουμε την **απόσταση χ^2** , της οποίας η ονομασία προκύπτει από την ομοιότητα που παρουσιάζει με το στατιστικό χ^2 του Pearson το οποίο χρησιμοποιείται για να ελέγξουμε την υπόθεση της ανεξαρτησίας ή ομοιογένειας σε έναν πίνακα συνάφειας.

¹ Βιβλιογραφική αναφορά (B1) σελ.34

Από τον τρόπο υπολογισμού της μπορούμε να δούμε ότι η απόσταση χ^2 είναι ουσιαστικά μία σταθμισμένη Ευκλείδεια απόσταση όπου τα βάρη που χρησιμοποιούνται στην ανάλυση των γραμμών είναι το αντίστροφο του μέσου προφίλ των γραμμών ή εναλλακτικά των αναμενόμενων τιμών των προφίλ των γραμμών ή ακόμη οι μάζες των στηλών αφού όπως είδαμε αυτά τα δύο μέτρα συμπίπτουν. Επομένως, οι αποστάσεις χ^2 δίνουν μεγαλύτερη βαρύτητα στις αποκλίσεις μεταξύ κατηγοριών που έχουν μικρότερες συχνότητες καθώς ο κάθε όρος διαιρείται με τους αντίστοιχους όρους του μέσου προφίλ.

Είναι πολύ σημαντικό να θυμόμαστε ότι η απόσταση χ^2 έχει νόημα να υπολογιστεί μόνο μεταξύ των κατηγοριών της ίδιας μεταβλητής, δηλαδή είτε μεταξύ των γραμμών είτε μεταξύ των στηλών. Στο γράφημα της Ανάλυσης Αντιστοιχιών φαίνονται αυτές ακριβώς οι αποστάσεις και προφίλ γραμμών με μικρές αποστάσεις θα πρέπει να βρίσκονται σε κοντινές θέσεις στο γράφημα. Τέλος, μπορούμε ακόμη να κάνουμε συγκρίσεις μεταξύ των προφίλ και του μέσου προφίλ για να εντοπίσουμε ομοιότητες και ακόμη να απεικονίσουμε ταυτόχρονα τις γραμμές και τις στήλες σε ένα κοινό διάγραμμα (αυτό προκύπτει από τον ορισμό της απόστασης χ^2).

Ο τύπος υπολογισμού της Ευκλείδειας απόστασης μεταξύ του προφίλ της γραμμής i , που έχουμε συμβολίσει \vec{r}_i και του μέσου προφίλ \vec{ar} είναι:

$$d_{r_i, ar} = \sqrt{\sum_{j=1}^J (r_{ij} - ar_j)^2} = \sqrt{(\vec{r}_i - \vec{ar})^T (\vec{r}_i - \vec{ar})}$$

Αντίστοιχα, ο τύπος υπολογισμού της απόστασης χ^2 είναι:

$$d_{r_i, ar} = \sqrt{\sum_{j=1}^J \frac{(r_{ij} - ar_j)^2}{ar_j}} = \sqrt{(\vec{r}_i - \vec{ar})^T \mathbf{D}_{ar}^{-1} (\vec{r}_i - \vec{ar})}$$

όπου D_{ar} είναι διαγώνιος πίνακας που περιέχει στη διαγώνιο τις συντεταγμένες του μέσου προφίλ των γραμμών.

Ένα σημαντικό προτέρημα των αποστάσεων χ^2 έναντι των Ευκλείδειων αποστάσεων είναι ότι αν σε έναν πίνακα συνάφειας και άρα και σε έναν πίνακα προφίλ γραμμών ή στηλών έχουμε δύο προφίλ που είναι ίδια ή έχουν παρόμοιες κατανομές ή που υπάρχει μία σταθερή αναλογία μεταξύ των τιμών τους τότε μπορούμε να τα ενώσουμε δημιουργώντας ένα νέο προφίλ χωρίς να επηρεαστούν οι αποστάσεις χ^2 . Αυτό συμβαίνει διότι η μάζα του νέου προφίλ που θα δημιουργηθεί θα είναι απλώς το άθροισμα των μαζών των δύο προφίλ που έχουν συγχωνευθεί. Αντίθετα, γίνεται αντιληπτό ότι δεν υπάρχει λόγος να χωρίσουμε μία κατηγορία σε επιμέρους. Λόγω αυτής της ιδιότητας λέμε ότι η απόσταση χ^2 υπακούει την **Αρχή της Ισοδυναμίας των Κατανομών** ή Ισοδυναμίας κατά Κατανομή ή Ισοκατανομής (Principal of Distributional Equivalence). Η ιδιότητα αυτή διατυπώθηκε από τους Lebart, Morineau & Warwick (1984)².

Εφόσον έχουμε υπολογίσει όλες τις αποστάσεις μεταξύ των προφίλ των γραμμών και του μέσου προφίλ των γραμμών θα πρέπει να ορίσουμε ένα μέτρο που να συνοψίζει όλη αυτή την πληροφορία. Ένα τέτοιο μέτρο είναι η **συνολική αδράνεια** ή απλά αδράνεια (που είναι γνωστή και ως Pearson's mean-square contingency) η οποία δίνεται από τον τύπο:

$$A^2 = \sum_{i=1}^I mr_i d_{r_i, ar}^2 = \sum_{i=1}^I mr_i \sum_{j=1}^J \frac{(r_{ij} - ar_j)^2}{ar_j} \quad (1)$$

Η ποσότητα A είναι ένα μέτρο το οποίο μας δείχνει το μέγεθος της μεταβλητότητας που υπάρχει στα προφίλ των γραμμών ή αλλιώς τη διασπορά του νέφους των γραμμών γύρω από το κέντρο βάρους. Η αδράνεια είναι επομένως το αντίστοιχο της διασποράς που υπολογίζουμε στην ανάλυση των κυρίων συνιστωσών για την περίπτωση όπου έχουμε

² Η απόδειξη της ιδιότητας μπορεί να βρεθεί στη βιβλιογραφική αναφορά (A6) σελ. 38

σταθμίσεις. Υψηλές τιμές αδράνειας παρατηρούμε όταν τα προφίλ είναι πολύ απομακρυσμένα από το μέσο προφίλ που συνεπάγεται ότι υπάρχει και μεγάλη διαφοροποίηση μεταξύ των γραμμών. Σε αυτή την περίπτωση έχουμε και υψηλές συσχετίσεις μεταξύ των γραμμών και των στηλών. Ακόμη, εάν όλα τα προφίλ των γραμμών ήταν ίδια τότε θα συνέπιπταν με το μέσο προφίλ και η αδράνεια θα έπαιρνε την τιμή 0. Αντιθέτως, η μέγιστη τιμή της αδράνειας είναι ίση με τη διάσταση του χώρου στον οποίο ανήκουν τα δεδομένα μας. Αυτό συμβαίνει στην περίπτωση που όλα τα μη διαγώνια στοιχεία του πίνακα συνάφειας είναι 0³. Αξίζει να σημειωθεί ότι η αδράνεια είναι ίδια για τις γραμμές και τις στήλες. Από τον τύπο (1) με τον οποίο ορίστηκε η αδράνεια φαίνεται ότι προφίλ παρόμοια με το μέσο προφίλ που βρίσκονται κοντά στο μέσο προφίλ θα έχουν μικρή απόσταση και επομένως δεν θα προσφέρουν σημαντικά στην συνολική αδράνεια.

Αντίστοιχα υπολογίζεται και η αδράνεια στην περίπτωση της ανάλυσης των στηλών και φαίνεται ότι είναι ίση με την αδράνεια όταν εξετάζουμε τις γραμμές. Πράγματι, αναπτύσσοντας την προηγούμενη σχέση, η αδράνεια μπορεί να γραφεί και ως:

$$\begin{aligned}
 A^2 &= \sum_{i=1}^I mr_i \sum_{j=1}^J \frac{(r_{ij} - ar_j)^2}{ar_j} \\
 &= \sum_{i=1}^I \sum_{j=1}^J mr_i \frac{\left(\frac{n_{ij}}{n_{i+}} - mc_j\right)^2}{mc_j} = \sum_{i=1}^I \sum_{j=1}^J \frac{mr_i^2 \left(\frac{p_{ij}}{mr_i} - mc_j\right)^2}{mr_i mc_j} \\
 &= \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - mr_i mc_j)^2}{mr_i mc_j} \tag{2}
 \end{aligned}$$

³ Η απόδειξη της ιδιότητας μπορεί να βρεθεί στη βιβλιογραφική αναφορά (A3) σελ. 406-407

έκφραση η οποία είναι συμμετρική ως προς τις γραμμές και τις στήλες. Επίσης, από την τελευταία σχέση βλέπουμε ότι η αδράνεια συνδέεται με τη στατιστική συνάρτηση χ^2 του Pearson μέσω της έκφρασης:

$$\Lambda^2 = \frac{\chi^2}{N} = \varphi^2$$

όπου με χ^2 συμβολίζουμε την τιμή που λαμβάνει το στατιστικό χ^2 του Pearson και με N το συνολικό μέγεθος του δείγματος. Από αυτή τη σχέση βλέπουμε ακόμη ότι ενώ το στατιστικό χ^2 του Pearson επηρεάζεται από το μέγεθος του δείγματος, δεν ισχύει το ίδιο και για την αδράνεια. Τέλος, η τετραγωνική ρίζα της αδράνειας είναι γνωστή και ως **συντελεστής συνάφειας μέσω τετραγώνου ή φ** (mean square contingency coefficient ή phi coefficient), επομένως μπορούμε να δηλώσουμε την αδράνεια και ως φ^2 .

Η **σχετική αδράνεια** κάθε συνδυασμένης κατηγορίας (i, j) προκύπτει από το πηλίκο του ατομικού χ^2 του Pearson με την συνολική αδράνεια Λ^2 , δηλαδή:

$$\Lambda_{i,j}^2 = \frac{1}{\Lambda^2} \frac{(p_{ij} - mr_i mc_j)^2}{mr_i mc_j}$$

Η αδράνεια της κάθε γραμμής προκύπτει από το άθροισμα της σχετικής αδράνειας των στοιχείων της εκάστοτε γραμμής i , δηλαδή:

$$\Lambda_i^2 = \sum_{j=1}^J \Lambda_{i,j}^2 = \frac{1}{\Lambda^2} \sum_{j=1}^J \frac{(p_{ij} - mr_i mc_j)^2}{mr_i mc_j}$$

και μας πληροφορεί για τη συνεισφορά της στη συνολική αδράνεια.

Τέλος, ο τύπος υπολογισμού του στατιστικού χ^2 του **Pearson** ανάμεσα στη γραμμή i του πίνακα συνάφειας \mathbf{N} που συμβολίζεται με \vec{n}_i και στην αναμενόμενη γραμμή που συμβολίζεται με $\vec{\bar{n}}_i$ είναι⁴:

$$\chi_{i,i}^2 = \sum_{j=1}^J \frac{(n_{ij} - \bar{n}_{ij})^2}{\bar{n}_{ij}} = (\vec{n}_i - \vec{\bar{n}}_i)^T \mathbf{D}_i^{-1} (\vec{n}_i - \vec{\bar{n}}_i)$$

όπου η αναμενόμενη γραμμή περιέχει τις συχνότητες εμφάνισης κάθε συνδυασμένης κατηγορίας κάτω από την υπόθεση της ανεξαρτησίας των μεταβλητών, οι οποίες προκύπτουν από το γινόμενο της περιθώριας γραμμής και της περιθώριας στήλης διαιρεμένο με το συνολικό πλήθος των παρατηρήσεων ως εξής:

$$\bar{n}_{ij} = \frac{n_{i+} n_{j+}}{N}$$

Ακόμη, με D_i συμβολίζουμε τον πίνακα που περιέχει στη διαγώνιο της αναμενόμενες συχνότητες τις γραμμής i , δηλαδή τα στοιχεία \bar{n}_{ij} ($i = 1, 2, \dots, I$).

Η μηδενική υπόθεση δηλώνει ότι δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των παρατηρούμενων και των αναμενόμενων συχνοτήτων και επομένως οι μεταβλητές μας είναι ανεξάρτητες. Εάν δεν μπορούμε να απορρίψουμε τη μηδενική υπόθεση συμπεραίνουμε ότι η οποιαδήποτε σχέση που μπορεί να παρατηρούμε είναι τυχαία, επομένως ιδανικά θα θέλαμε να την απορρίψουμε εφόσον η μέθοδος εξετάζει τις συσχετίσεις μεταξύ των κατηγοριών (επιπέδων) των μεταβλητών. Για να απορρίψουμε τη μηδενική υπόθεση θέλουμε η τιμή του στατιστικού χ^2 του Pearson να είναι μικρότερη από την θεωρητική τιμή της κατανομής χ^2 με βαθμούς ελευθερίας $(I - 1) \times (J - 1)$. Ωστόσο ακόμη και αν δεν συμβαίνει κάτι τέτοιο η ανάλυσή μας συνεχίζεται κανονικά. Αυτό

⁴ Βιβλιογραφική αναφορά (B1) σελ.32

συμβαίνει διότι σύμφωνα με τον Greenacre (2007)⁵ το παραπάνω στατιστικό δεν θεωρείται τόσο σημαντικό.

Το κύριο μέρος της Ανάλυσης Αντιστοιχιών είναι η εύρεση του βέλτιστου υπόχωρου στον οποίο θα προβληθούν τα αρχικά πολυμεταβλητά σημεία ούτως ώστε να μπορέσουμε να έχουμε μία γραφική απεικόνιση των κατηγοριών των γραμμών και των στηλών. Για να γίνει αυτό θα χρειαστεί να εφαρμόσουμε τη μέθοδο της **παραγοντοποίησης ιδιόμορφων τιμών** (singular value decomposition-συντομογραφία: SVD) στον **πίνακα των τυποποιημένων καταλοίπων \mathbf{S}** . Ο πίνακας αυτός προκύπτει από την αδράνεια μέσω της σχέσης (2) ως εξής:

$$A^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - mr_i mc_j)^2}{mr_i mc_j} = \sum_{i=1}^I \sum_{j=1}^J s_{ij}^2$$

όπου τα στοιχεία s_{ij} συνιστούν τα κελιά του πίνακα \mathbf{S} .

Το κάθε κελί αυτού του πίνακα εμπεριέχει τη διαφορά της πραγματικής σχετικής συχνότητας του πίνακα αντιστοιχιών \mathbf{P} με τη σχετική συχνότητα που υποθέτει το μοντέλο της ανεξαρτησίας των δύο μεταβλητών που μελετάμε διαιρεμένο με τη ρίζα αυτής της θεωρητικής σχετικής συχνότητας. Τα στοιχεία του πίνακα αυτού δείχνουν πόσο η τιμή του κάθε κελιού διαφέρει από αυτήν που θα είχε στην περίπτωση της ανεξαρτησίας⁶. Ακόμη παρατηρούμε ότι το άθροισμα των τετραγώνων των στοιχείων αυτού του πίνακα μας δίνει τη συνολική αδράνεια. Μπορεί γεωμετρικά να έχουμε αναπαραστήσει τα προφίλ, ωστόσο αλγεβρικά είναι απλούστερο να εφαρμόσουμε τη μέθοδο στον πίνακα αντιστοιχιών. Επομένως ο πίνακας \mathbf{S} μπορεί να γραφεί στη μορφή:

$$\mathbf{S} = \left\{ \frac{p_{ij} - mr_i mc_j}{\sqrt{mr_i mc_j}} \right\}_{I \times J} = \mathbf{D}_{mr}^{-1/2} (\mathbf{P} - \overline{mr} \cdot \overline{mc}^T) \mathbf{D}_{mc}^{-1/2}$$

⁵ Βιβλιογραφική αναφορά (B4)

⁶ Βιβλιογραφική αναφορά (A3) σελ.407

όπου p_{ij} είναι οι σχετικές συχνότητες που περιλαμβάνονται στον πίνακα αντιστοιχιών \mathbf{P} , m_{r_i} και m_{c_j} είναι οι μάζες της γραμμής i και της στήλης j και περιλαμβάνονται στα διανύσματα \vec{m}_r και \vec{m}_c αντίστοιχα και \mathbf{D}_{mr} είναι ο διαγώνιος πίνακας με τις μάζες των γραμμών στα στοιχεία της διαγωνίου του και αντίστοιχα ο \mathbf{D}_{mc} με τις μάζες των στηλών. Όπως και προηγουμένως έχουμε ότι $i = 1, 2, \dots, I$ και $j = 1, 2, \dots, J$. Ακόμη, μπορούμε να δούμε ότι η σχέση που ισχύει μεταξύ του πίνακα των καταλοίπων \mathbf{S} και της συνολικής αδράνειας A^2 είναι:

$$A^2 = tr(\mathbf{S}\mathbf{S}^T)$$

όπου το ίχνος (trace) του πίνακα είναι το άθροισμα των διαγώνιων στοιχείων του.

Η εφαρμογή της μεθόδου SVD στον πίνακα \mathbf{S} έχει ως σκοπό να αποσυνθέσει την αδράνεια σε ιδιοτιμές. Το άθροισμα όλων των ιδιοτιμών ισούται με τη συνολική αδράνεια. Κάθε ιδιοτιμή αντιστοιχεί σε μία διάσταση και μας παρέχει πληροφορία σχετικά με την σημαντικότητα αυτής της διάστασης και το ποσοστό της συνολικής αδράνειας που αυτή εξηγεί. Επομένως το πλήθος των ιδιοτιμών ισούται με το πλήθος των διαστάσεων του κοινού υπόχωρου των γραμμών και των στηλών, το οποίο είναι:

$$K = \min\{I, J\} - 1$$

Η μέθοδος SVD θα μας οδηγήσει σε μια έκφραση της μορφής:

$$\mathbf{S} = \mathbf{V}\mathbf{A}\mathbf{W}^T \quad (3)$$

όπου \mathbf{A} διαγώνιος πίνακας τάξης K που περιέχει τις ιδιόμορφες τιμές (singular values) στη διαγώνιο σε φθίνουσα σειρά $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_K > 0$. Ο πίνακας \mathbf{V} είναι ο αριστερά ιδιόμορφος πίνακας (left singular matrix) διάστασης $I \times K$ που περιέχει τα αριστερά ιδιόμορφα διανύσματα (left singular vectors) στις K στήλες του και ο πίνακας \mathbf{W} είναι ο δεξιά ιδιόμορφος πίνακας διάστασης $J \times K$ που περιέχει τα δεξιά ιδιόμορφα διανύσματα

στις K στήλες του. Οι στήλες του πίνακα V είναι οι συντεταγμένες των σημείων των γραμμών και οι στήλες του πίνακα W οι συντεταγμένες των σημείων των στηλών. Οι πίνακες V και W έχουν ορθοκανονικές στήλες, δηλαδή ισχύει ότι $VV^T = WW^T = I$.

Η παραγοντοποίηση ιδιόμορφων τιμών αποτελεί μία γενίκευση της φασματικής ανάλυσης σε μη συμμετρικούς πίνακες. Προκειμένου να υπολογίσουμε τα $\lambda_1, \lambda_2, \dots, \lambda_K$ και τους πίνακες V και W παρατηρούμε ότι από τη σχέση (3) προκύπτουν οι ισότητες:

$$SS^T = V\Lambda^2V^T \text{ και } S^T S = W\Lambda^2W^T$$

οι οποίες υποδεικνύουν ότι οι πίνακες V, W, Λ^2 είναι οι πίνακες που εμφανίζονται στη φασματική ανάλυση του συμμετρικού πίνακα SS^T και $S^T S$. Επομένως, τα δεξιά ιδιόμορφα διανύσματα του πίνακα S ταυτίζονται με τα ιδιοδιανύσματα του πίνακα $S^T S$, τα αριστερά ιδιόμορφα διανύσματά του ταυτίζονται με τα ιδιοδιανύσματα του πίνακα SS^T και οι τετραγωνικές τιμές των ιδιόμορφων τιμών του ταυτίζονται με τις ιδιοτιμές καθενός από τους δύο προηγούμενους πίνακες⁷.

Η Ανάλυση Αντιστοιχιών κάνει χρήση των βασικών ιδιοτήτων της SVD ώστε να μειωθεί η διάσταση των αρχικών δεδομένων και να επιτευχθεί η ερμηνεία τους. Πιο αναλυτικά, κάθε ιδιοτιμή λ_k^2 , $k = 1, 2, \dots, K$ που προκύπτει από την SVD αντιστοιχεί σε έναν άξονα και όλοι οι άξονες μαζί αναπαριστούν πλήρως τα αρχικά δεδομένα. Εφόσον όμως οι ιδιοτιμές υπολογίζονται σε φθίνουσα σειρά σε κάθε επόμενο άξονα που προκύπτει αντιστοιχεί όλο και μικρότερο μέρος αδράνειας. Το ποσοστό της αδράνειας που εξηγεί ο κάθε άξονας βρίσκεται διαιρώντας την ιδιοτιμή που του αντιστοιχεί με το άθροισμα όλων των ιδιοτιμών (το οποίο συμπίπτει με τη συνολική αδράνεια των δεδομένων). Η μείωση των διαστάσεων επιτυγχάνεται αγνοώντας τις διαστάσεις εκείνες που έχουν μικρή αδράνεια και επομένως περιέχουν μικρή πληροφορία. Επομένως αφαιρώντας αυτές τις συγκεκριμένες διαστάσεις καταφέρνουμε να διατηρήσουμε γεωμετρικά την πληροφορία,

⁷ Βιβλιογραφική αναφορά (A3) σελ.412

δηλαδή σημεία γειτονικά στις αρχικές διαστάσεις να συνεχίζουν να είναι γειτονικά και στις νέες μειωμένες διαστάσεις.

Εάν κρατήσουμε μία μόνο διάσταση θα καταλήξουμε να έχουμε έναν άξονα και όλα τα προφίλ θα προβληθούν σε αυτόν τον άξονα τον οποίο ονομάζουμε πρώτο κύριο άξονα (first principal axis). Το μέρος της αδράνειας που αυτός εξηγεί ονομάζεται πρώτη κύρια αδράνεια (first principal inertia). Εάν επιλέξουμε να κρατήσουμε δύο διαστάσεις, οι δύο άξονες που δημιουργούνται συνιστούν ένα επίπεδο. Ο δεύτερος άξονας θα είναι ορθογώνιος στον πρώτο και αντίστοιχα κάθε επόμενος άξονας θα είναι ορθογώνιος σε όλους τους υπόλοιπους που ήδη έχουν χρησιμοποιηθεί. Η πρώτη διάσταση ονομάζεται πρώτος κύριος άξονας, η δεύτερη δεύτερος κύριος άξονας κ.ο.κ. Η αδράνεια που εξηγούν οι δύο πρώτοι άξονες ονομάζεται πρώτη και δεύτερη κύρια αδράνεια αντίστοιχα.

Το επόμενο βήμα είναι να διαλέξουμε το πλήθος των διαστάσεων που θα κρατήσουμε. Δεδομένου ότι ο σκοπός της ανάλυσης αυτής είναι να μας παρέχει τη δυνατότητα να αναπαραστήσουμε γραφικά τα αποτελέσματα το ιδανικό θα ήταν να κρατήσουμε τρεις ή λιγότερες διαστάσεις. Δεν υπάρχει κάποιος σαφής κανόνας που να μας προκαθορίζει τον ιδανικό αριθμό διαστάσεων, όμως έχουν προταθεί τέσσερα κριτήρια που μπορούμε να χρησιμοποιήσουμε για να οδηγηθούμε στην βέλτιστη απόφαση. Ξεκινώντας, το πιο σημαντικό κριτήριο βασίζεται στο **ποσοστό της συνολικής αδράνειας** που εξηγούν, το οποίο και θέλουμε να είναι όσο πιο μεγάλο γίνεται. Σε συνδυασμό με αυτό μπορεί να χρησιμοποιηθεί και το **Scree test**, όπου φτιάχνουμε ένα διάγραμμα κάθε σημείο του οποίου είναι της μορφής (i, λ_i^2) , όπου με λ_i^2 συμβολίζουμε την ιδιοτιμή που αντιστοιχεί στη διάσταση i . Ενώνοντας αυτά τα σημεία δημιουργείται μία φθίνουσα γραμμή. Το σημείο στο οποίο εμφανίζεται η μεγαλύτερη καμπή και έπειτα μειώνεται απότομα η κλίση καθορίζει το πλήθος των διαστάσεων που προτείνεται να κρατήσουμε. Το επόμενο κριτήριο βασίζεται στην **ερμηνεία των τελικών διαστάσεων**. Αν έχουμε βρει κάποια διάσταση, στην οποία δεν μπορούμε να δώσουμε φυσική ερμηνεία τότε το πιθανότερο είναι να μην χρειάζεται. Τέλος, θα μπορούσαμε διαλέγοντας μία **υποομάδα των δεδομένων** μας να καταλήξουμε να κρατήσουμε το πλήθος των διαστάσεων που μοιάζει ταυτόχρονα ιδανικό τόσο για την υποομάδα αυτή όσο και για τα αρχικά δεδομένα.

Έχοντας αποφασίσει για το πλήθος των διαστάσεων που θα κρατήσουμε θα πρέπει ακόμη να ελέγξουμε και κατά πόσο η λύση αυτή είναι σημαντική. Για να το κάνουμε αυτό, μπορούμε αρχικά να υπολογίσουμε το ποσοστό της συνολικής αδράνειας που αυτή εξηγεί, δηλαδή την **ποιότητα** της λύσης. Ο υπολογισμός αυτής βασίζεται σε ένα κλάσμα το οποίο στον αριθμητή έχει το άθροισμα των πρώτων κύριων αδρανειών και στον παρονομαστή τη συνολική αδράνεια των αρχικών δεδομένων. Το κλάσμα αυτό το εκφράζουμε σαν ποσοστό. Έχουμε ήδη αναφέρει ότι μειώνοντας τις διαστάσεις των δεδομένων χάνουμε πληροφορία. Επομένως θα πρέπει να ορίσουμε και ένα μέτρο απώλειας της πληροφορίας που χάνουμε. Αυτό το μέτρο στηρίζεται στην υπολειπόμενη αδράνεια που μένει ανεξήγητη και υπολογίζεται αντικαθιστώντας στον τύπο της ποιότητας τον αριθμητή με αυτήν ακριβώς την ποσότητα. Έτσι καταλήγουμε με ένα ακόμη ποσοστό το οποίο ονομάζεται **σφάλμα** της ανάλυσης.

Οι **τυποποιημένες συντεταγμένες** (standardized coordinates) των γραμμών \mathbf{F} και των στηλών \mathbf{G} υπολογίζονται τυποποιώντας τις αρχικές συντεταγμένες που βρίσκονται στους πίνακες \mathbf{V} και \mathbf{W} ως προς τις μάζες των γραμμών και των στηλών αντίστοιχα. Αποτελούν τις συντεταγμένες των σημείων των προφίλ που χρησιμοποιούνται για την γραφική αναπαράσταση της λύσης που προκύπτει και υπολογίζονται μέσω των παρακάτω τύπων:

$$\mathbf{F} = \mathbf{D}_{mr}^{-1/2} \mathbf{V} \text{ και } \mathbf{G} = \mathbf{D}_{mc}^{-1/2} \mathbf{W}$$

Για να πάρουμε τις **κύριες συντεταγμένες** (principal coordinates) χρησιμοποιούμε τους επόμενους τύπους υπολογισμού:

$$\mathbf{F}_P = \mathbf{D}_{mr}^{-1/2} \mathbf{V} \mathbf{\Lambda} \text{ και } \mathbf{G}_P = \mathbf{D}_{mc}^{-1/2} \mathbf{W} \mathbf{\Lambda}$$

Οι πίνακες \mathbf{F}, \mathbf{F}_P είναι διάστασης $I \times K$ όπου οι γραμμές τους είναι οι συντεταγμένες των γραμμών του αρχικού πίνακα και οι στήλες τους αναφέρονται στους κύριους άξονες. Οι πίνακες \mathbf{G}, \mathbf{G}_P είναι διάστασης $J \times K$ όπου οι γραμμές τους είναι οι συντεταγμένες των στηλών του αρχικού πίνακα και οι στήλες τους αναφέρονται στους

κύριους άξονες. Παρατηρούμε σε αυτό το σημείο ότι η μέθοδος ξεκινάει από κατηγορικές μεταβλητές και καταλήγει σε συνεχείς. Αυτό είναι ιδιαίτερα χρήσιμο καθώς μπορούμε να πάρουμε τις νέες αυτές μεταβλητές και να τις αναλύσουμε περαιτέρω με κάποια από τις πολλές διαθέσιμες μεθόδους για συνεχή δεδομένα.

Όπως είχαμε αναφέρει και στην αρχή, υπάρχει πλήρης αντιστοιχία μεταξύ της ανάλυσης που γίνεται για τις γραμμές και αυτής που γίνεται για τις στήλες. Επιπλέον, γνωρίζοντας τις συντεταγμένες, κύριες ή τυποποιημένες των γραμμών μπορούμε να περάσουμε σε αυτές των στηλών και αντίστροφα. Αποτέλεσμα αυτού είναι στις περισσότερες περιπτώσεις να υπολογίζουμε τις συντεταγμένες του μικρότερου πίνακα, για παράδειγμα των γραμμών και μέσω των σχέσεων που συνδέουν τις συντεταγμένες γραμμών και στηλών να υπολογίζουμε και τις συντεταγμένες των στηλών. Οι **σχέσεις μετάβασης** είναι οι παρακάτω:

$$F_P = D_{mr}^{-1} P G_P \Lambda^{-1} \text{ και } G_P = D_{mc}^{-1} P^T F_P \Lambda^{-1}$$

Δύο ακόμη δείκτες που υπολογίζονται στα πλαίσια της ανάλυσης αντιστοιχιών και βοηθούν στην ερμηνεία των αποτελεσμάτων και την απόδοση φυσικής σημασίας στους $k = 1, 2, \dots, K$ παραγοντικούς άξονες (ή διαστάσεις) είναι η συνεισφορά των $i = 1, 2, \dots, I$ σημείων των γραμμών (ή κατηγοριών) στις διαστάσεις ή αλλιώς οι **απόλυτες συνεισφορές CTR** και η συνεισφορά των διαστάσεων στα σημεία των γραμμών ή αλλιώς οι **τετραγωνικές συσχετίσεις COR**. Οι τιμές CTR προκύπτουν από το λόγο της επιμέρους αδράνειας ενός σημείου σε έναν άξονα προς τη συνολική αδράνεια του άξονα και απεικονίζουν το ποσοστό της αδράνειας κάθε μίας διάστασης που εξηγείται από κάθε σημείο. Μας δείχνει τη συνεισφορά του σημείου στον καθορισμό της κατεύθυνσης του εκάστοτε άξονα και η χρησιμότητά τους έγκειται στο γεγονός ότι μας βοηθούν να ερμηνεύσουμε τους άξονες. Οι τιμές COR απεικονίζουν το ποσοστό της αδράνειας ενός σημείου που εξηγείται από την κάθε διάσταση και μας δείχνει την ποιότητα της αναπαράστασης του κάθε σημείου στην εκάστοτε διάσταση. Το άθροισμα των CTR σε όλα τα σημεία της κάθε μεταβλητής ή αλλιώς σε όλες τις κατηγορίες της κάθε μεταβλητής για κάθε άξονα καθώς και το άθροισμα των COR σε όλες τις διαστάσεις για κάθε

κατηγορία ισούται με τη μονάδα. Το άθροισμα των COR ενός σημείου στους επιλεγμένους άξονες εκφράζει την ποιότητα απεικόνισης του σημείου αυτού στον υπόχωρο. Ο τρόπος υπολογισμού των δύο προαναφερθέντων δεικτών για ένα σημείο/γραμμή i είναι⁸:

$$CTR(ik) = \frac{mr_i f_{ik}^2}{\lambda_k^2} \text{ και } COR(ik) = \frac{f_{ik}^2}{d_{r_i,ar}^2} = \cos\omega^2$$

όπου με ω^2 συμβολίζουμε τη γωνία που σχηματίζεται μεταξύ του άξονα k και της ευθείας που ενώνει το σημείο με την αρχή των αξόνων (που συμπίπτει με το κεντροειδές). Όσο μεγαλύτερη είναι η τιμή COR για μία διάσταση τόσο μικρότερη θα είναι και η γωνία ω . Επομένως το σημείο θα έχει την ίδια κατεύθυνση με τον άξονα το οποίο σημαίνει ότι θα εμφανίζει υψηλή συσχέτιση με τον άξονα καθώς το COR εκφράζει συσχέτιση στο τετράγωνο. Τέλος, με f_{ik} συμβολίζουμε την τυποποιημένη συντεταγμένη της i γραμμής στον k παραγοντικό άξονα, με mr_i συμβολίζουμε τη μάζα της i γραμμής και με $d_{r_i,ar}^2$ συμβολίζουμε την απόσταση μεταξύ του προφίλ της γραμμής i , που έχουμε συμβολίσει \vec{r}_i και του μέσου προφίλ που έχουμε συμβολίσει \vec{ar} . Δεν πρέπει να ξεχνάμε ότι η ερμηνεία των αξόνων είναι υποκειμενική.

Όπως αναφέραμε, ο δείκτης CTR προκύπτει από το λόγο της επιμέρους αδράνειας ενός σημείου σε έναν άξονα προς τη συνολική αδράνεια του άξονα. Από τον τύπο υπολογισμού του CTR προκύπτει η επόμενη σχέση για την αδράνεια που ερμηνεύει ο κάθε παραγοντικός άξονας k :

$$\lambda_k^2 = \sum_{i=1}^I mr_i f_{ik}^2, k = 1, 2, \dots, K$$

δηλαδή είναι το άθροισμα των επιμέρους αδρανειών όλων των σημείων σε αυτόν τον άξονα.

⁸ Βιβλιογραφική αναφορά (B3) σελ.16-20

Ένας γραφικός τρόπος να παρατηρήσουμε την ποιότητα της λύσης είναι να συγκρίνουμε τις αποστάσεις χ^2 μεταξύ των αρχικών προφίλ στον οριζόντιο άξονα και αυτών που έχουν βρεθεί μετά την μείωση των διαστάσεων στον κάθετο άξονα. Όσο πιο καλή είναι η αναπαράσταση των δεδομένων στο μειωμένο πλήθος των διαστάσεων που έχουμε διαλέξει τόσο περισσότερα τα σημεία που προβάλλονται θα τείνουν στην ευθεία $y = x$. Εφόσον οι τελικές αποστάσεις προκύπτουν από την προβολή των αρχικών σε έναν χώρο λιγότερων διαστάσεων είναι εύκολο να αντιληφθούμε ότι αυτές θα είναι πάντα μικρότερες από τις αρχικές. Αυτό συμβαίνει καθώς η απόσταση χ^2 αθροίζει μη αρνητικές τιμές και το πλήθος αυτών εξαρτάται από το πλήθος των διαστάσεων.

Το τελευταίο μέρος της Ανάλυσης Αντιστοιχιών είναι η διαγραμματική απεικόνιση των κατηγοριών των μεταβλητών. Αρχικά μπορούμε να προβάλουμε τις κατηγορίες των γραμμών ή των στηλών σε ένα διάγραμμα. Μέσω αυτού μπορούμε να ποσοτικοποιήσουμε τις διαφοροποιήσεις μεταξύ των κατηγοριών, για παράδειγμα να δούμε εάν κάποια κατηγορία είναι πολύ πιο μακριά από τις υπόλοιπες. Ακόμη, μπορούμε να δούμε εάν υπάρχει κάποια φυσική διάταξη στις κατηγορίες. Αυτό που δεν πρέπει να ξεχνάμε είναι ότι χρειάζεται ιδιαίτερη προσοχή στην συμπερασματολογία καθώς όπως έχουμε πει πρόκειται για μία μέθοδο με καθαρά περιγραφικό χαρακτήρα αφού δεν λαμβάνει υπόψη τυχαία σφάλματα και δεν μπορούμε να γενικεύσουμε τα αποτελέσματα που βλέπουμε στο σύνολο του πληθυσμού. Αυτό που λαμβάνουμε από αυτή τη μέθοδο είναι ο εντοπισμός της δομής που υπάρχει στα δεδομένα μας.

Ωστόσο, το πιο σημαντικό διάγραμμα που εξάγεται από αυτή την ανάλυση είναι το **biplot**. Αποτελεί ένα διάγραμμα στο οποίο προβάλλονται ταυτόχρονα οι κατηγορίες τόσο των γραμμών όσο και των στηλών. Το biplot μας βοηθάει να εντοπίσουμε συσχετίσεις μεταξύ των κατηγοριών των γραμμών και των στηλών. Αυτό συμβαίνει εάν αυτές βρίσκονται κοντά η μία στην άλλη. Το γράφημα αυτό μπορεί να απεικονιστεί σε 2 διαφορετικές μορφές, στη μία περίπτωση έχουμε το λεγόμενο **συμμετρικό biplot** ενώ στην άλλη το **μη-συμμετρικό biplot**.

Το συμμετρικό biplot απεικονίζει τόσο τις γραμμές όσο και τις στήλες σε κύριες συντεταγμένες, δηλαδή οι συντεταγμένες των κατηγοριών τους βρίσκονται στην ίδια κλίμακα. Αυτό έχει ως αποτέλεσμα οι κατηγορίες τους να εκτείνονται το ίδιο στο γράφημα. Ακόμη, σημεία με μεγάλη μάζα δεν επηρεάζουν σημαντικά το γράφημα και είναι το συνηθέστερο γράφημα που χρησιμοποιείται. Τα πρόσημα των συντεταγμένων μπορούν να αλλάξουν αντιστρέφοντας μόνο την ερμηνεία των αξόνων και επομένως δεν επηρεάζουν τα συμπεράσματα που προκύπτουν. Αυτό που μας ενδιαφέρει είναι η διαφορά του προσήμου μεταξύ των διαφορετικών σημείων. Αξίζει να σημειωθεί ότι το γράφημα αυτό αποτελείται από 2 διαφορετικά γραφήματα απεικονισμένα το ένα πάνω στο άλλο και επομένως δεν θα πρέπει να ερμηνεύσουμε τις αποστάσεις μεταξύ γραμμών και στηλών ούτε να δημιουργήσουμε συστάδες που να περιέχουν σημεία γραμμών και στηλών ταυτόχρονα. Οι αποστάσεις είναι σωστές μόνο μεταξύ των κατηγοριών της ίδιας μεταβλητής.

Το μη-συμμετρικό biplot χωρίζεται και αυτό σε 2 κατηγορίες, στη μία περίπτωση έχουμε τις γραμμές σε κύριες συντεταγμένες και τις στήλες σε τυποποιημένες συντεταγμένες όπου μιλάμε για το μη-συμμετρικό biplot των γραμμών και στην άλλη περίπτωση έχουμε τις στήλες σε κύριες συντεταγμένες και τις γραμμές σε τυποποιημένες συντεταγμένες και παίρνουμε το μη-συμμετρικό biplot των στηλών. Στο μη-συμμετρικό biplot των γραμμών τα σημεία γραμμών κανονικοποιούνται και επομένως έχουμε σωστές αποστάσεις ανάμεσα στις κατηγορίες των γραμμών και των στηλών. Αυτό έχει σαν αποτέλεσμα την καλύτερη και πιο σωστή απεικόνιση των συσχετίσεων καθώς και την δυνατότητα για εντοπισμό ομαδοποιήσεων. Ακόμη, σε αυτή την περίπτωση τα σημεία των γραμμών γίνονται όσο το δυνατόν πιο ακραία και η απεικόνιση είναι πιο ευκρινής. Αντίστοιχα αποτελέσματα ισχύουν και για τις στήλες.

Τέλος, μπορούμε ακόμη να εισάγουμε **συμπληρωματικά σημεία** στην ανάλυσή μας. Αυτά τα σημεία προστίθενται στον πίνακα σαν γραμμές ή στήλες στον πίνακα και προβάλλονται κανονικά στους παραγοντικούς άξονες, ωστόσο δεν παίζουν ρόλο στην εύρεση των αξόνων καθώς τους ορίζουμε μάζα 0. Βασική προϋπόθεση είναι τα προφίλ τους να είναι συγκρίσιμα με τα υπάρχοντα προφίλ. Για τα σημεία αυτά μπορούμε να

υπολογίσουμε αδράνεια και συσχετίσεις οπότε βλέπουμε ποιος άξονας τα εξηγεί καλύτερα και την ποιότητα της απεικόνισής τους στον υπόχωρο που χρησιμοποιούμε. Οι λόγοι για τους οποίους μπορεί κάποιος να θέλει να εισάγει μεταβλητές ως συμπληρωματικά σημεία είναι:

- Όταν θέλουμε να τα ερμηνεύσουμε πάνω στους παραγοντικούς άξονες αλλά όχι να συνεισφέρουν στη δημιουργία τους επειδή είναι εννοιολογικά διαφορετικά από τα υπόλοιπα σημεία.
- Όταν έχουμε στη διάθεσή μας κάποιο υπερσύνολο ή υποομάδα των δεδομένων. Στην περίπτωση αυτή το θεωρούμε ως συμπληρωματικό σημείο για να μην εισαχθεί η πληροφορία δύο φορές στην ανάλυση.
- Όταν αναλύουμε μία διαχρονική μελέτη και θέλουμε να συμπεριλάβουμε την πληροφορία παρελθοντικών δεδομένων για να παρακολουθήσουμε τις αλλαγές με την πάροδο του χρόνου.

Οι κύριες συντεταγμένες της συμπληρωματικής γραμμής i^{sup} πάνω στον κύριο άξονα s υπολογίζονται βάσει της σχέσης:

$$f_{P_{is}}^{sup} = \sum_{j=1}^J \frac{n_{ij}^{sup}}{n_{i+}^{sup}} g_{js}$$

όπου με g_{is} συμβολίζουμε τις τυποποιημένες συντεταγμένες των στηλών στον κύριο άξονα s . Αντίστοιχα, οι κύριες συντεταγμένες της συμπληρωματικής στήλης j^{sup} πάνω στον κύριο άξονα s υπολογίζονται βάσει της σχέσης:

$$g_{P_{is}}^{sup} = \sum_{j=1}^J \frac{n_{ij}^{sup}}{n_{+j}^{sup}} f_{is}$$

όπου με f_{is} συμβολίζουμε τις τυποποιημένες συντεταγμένες των γραμμών στον κύριο άξονα s .

2.2 Πολλαπλή Ανάλυση Αντιστοιχιών

Στην περίπτωση των 2 μεταβλητών είχαμε έναν πίνακα συνάφειας διπλής εισόδου όπου στις γραμμές είχαμε της κατηγορίες της μίας μεταβλητής και στις στήλες της άλλης. Στην Πολλαπλή Ανάλυση Αντιστοιχιών έχουμε περισσότερες από 2 κατηγορικές μεταβλητές και επομένως τα δεδομένα αναπαρίστανται σε έναν πίνακα συνάφειας πολλαπλής εισόδου. Αξίζει να αναφερθεί ότι η μέθοδος αυτή δεν είναι απλή γενίκευση της Απλής Μεθόδου Αντιστοιχιών καθώς εάν εφαρμόσουμε και τις 2 μεθόδους στις ίδιες 2 μεταβλητές θα πάρουμε διαφορετικά αποτελέσματα. Συνήθως χρησιμοποιείται σε δεδομένα όμοια με αυτά που θα αναλύσουμε και στην παρούσα εργασία, τα οποία έχουν προκύψει μέσα από ένα ερωτηματολόγιο. Όπως και πριν καταλήγουμε σε ένα μικρότερο πλήθος συνεχών μεταβλητών. Μπορεί τα δεδομένα να αναπαρίστανται σε πίνακα πολλαπλής εισόδου, ωστόσο για την εφαρμογή της μεθόδου χρησιμοποιείται άλλη κωδικοποίηση, είτε ο πίνακας δείκτης (disjunctive matrix) είτε ο πίνακας Burt.

Ο πίνακας δείκτης X έχει στις γραμμές τον αύξοντα αριθμό των παρατηρήσεων ή των ατόμων και η κάθε στήλη έχει μία ψευδομεταβλητή που αντιστοιχεί κάθε φορά σε 1 κατηγορία από την κάθε κατηγορική μεταβλητή. Στα κελιά του πίνακα υπάρχει είτε η τιμή 1 που δηλώνει παρουσία του χαρακτηριστικού στο άτομο είτε η τιμή 0 που δηλώνει απουσία αυτού. Μερικά χαρακτηριστικά του πίνακα δείκτη είναι⁹:

- Σε κάθε υποπίνακα ψευδομεταβλητών που δημιουργείται από τις κατηγορίες μίας μεταβλητής θα υπάρχει οπωσδήποτε ένα 1 και οι υπόλοιπες τιμές θα είναι 0.
- Το πλήθος των γραμμών ισούται με το πλήθος των παρατηρήσεων και το πλήθος των στηλών ισούται με το άθροισμα των κατηγοριών όλων των μεταβλητών.
- Το πλήθος των διαφορετικών συνδυασμών που υπάρχουν ανά γραμμή ισούται με το γινόμενο των κατηγοριών των μεταβλητών.

⁹ Βιβλιογραφική αναφορά (A3) σελ.446-447

- Τα περιθώρια αθροίσματα των γραμμών είναι σταθερά και ίσα με το πλήθος των μεταβλητών και των στηλών ίσα με το πλήθος των παρατηρήσεων που παρουσιάζουν το εκάστοτε χαρακτηριστικό.

Ο πίνακας Burt B προκύπτει μέσω του πίνακα δείκτη και πιο συγκεκριμένα είναι ο $B = X^T X$. Πρόκειται για έναν συμμετρικό πίνακα που έχει και στις γραμμές και στις στήλες τις ψευδομεταβλητές που αντιπροσωπεύουν όλες τις κατηγορίες των μεταβλητών μας. Τα κελιά αυτού του πίνακα έχουν το πλήθος των παρατηρήσεων που παρουσιάζουν ταυτόχρονα το χαρακτηριστικό της εκάστοτε στήλης και της εκάστοτε γραμμής. Μερικά χαρακτηριστικά του πίνακα Burt είναι¹⁰:

- Αποτελείται από Q^2 υποπίνακες, όπου Q το πλήθος των μεταβλητών.
- Η διαγώνιος περιέχει p υποπίνακες, έναν για κάθε μεταβλητή. Κάθε ένας από αυτούς είναι διαγώνιος και τα διαγώνια στοιχεία του αποτελούν τις περιθώριες συχνότητες των κατηγοριών της κάθε μεταβλητής.
- Το άθροισμα των στοιχείων της διαγωνίου του κάθε υποπίνακα της διαγωνίου του πίνακα Burt ισούται με το συνολικό πλήθος των παρατηρήσεων N .
- Πάνω από τη διαγώνιο έχουμε $Q(Q - 1)/2$ υποπίνακες, οι οποίοι έχουν ανά δύο όλους τους συνδυασμούς των μεταβλητών και προφανώς οι ίδιοι πίνακες επαναλαμβάνονται και κάτω από την διαγώνιο.
- Έχει καλύτερη γεωμετρική ερμηνεία από τον πίνακα δείκτη και επομένως στις περισσότερες περιπτώσεις δουλεύουμε με αυτόν.
- Εφόσον είναι συμμετρικός η ανάλυση των γραμμών συμπίπτει με αυτή των στηλών και επομένως στα διαγράμματα απεικονίζεται μόνο ένα σετ σημείων.
- Οι τυποποιημένες συντεταγμένες του πίνακα Burt είναι ίδιες με αυτές του πίνακα δείκτη εξ' ορισμού, εφόσον $B = X^T X$. Από αυτό προκύπτει και ότι οι κύριες αδράνειες του πίνακα Burt είναι ίσες με το τετράγωνο αυτών του πίνακα δείκτη.
- Οι κύριες συντεταγμένες προκύπτουν από τον πολλαπλασιασμό των τυποποιημένων συντεταγμένων με την ρίζα των κύριων αδρανειών. Εφόσον οι κύριες αδράνειες

¹⁰ Βιβλιογραφική αναφορά (A3) σελ.448

είναι μικρότερες της μονάδας, υψώνοντας στο τετράγωνο μειώνονται ακόμη περισσότερο με αποτέλεσμα οι κύριες συντεταγμένες του πίνακα Burt να βρίσκονται σε μικρότερη κλίμακα από αυτές του πίνακα δείκτη. Ως αποτέλεσμα, η αδράνεια που εξηγεί η κάθε διάσταση του πίνακα Burt είναι μεγαλύτερη από αυτή του πίνακα δείκτη.

- Όσον αφορά τη γραφική αναπαράσταση σε κύριες συντεταγμένες, λόγω της διαφοράς στην κλίμακα τα σημεία, χρησιμοποιώντας τον πίνακα Burt, είναι πιο «συρρικνωμένα» σε σχέση με του πίνακα δείκτη.

Η Πολλαπλή Ανάλυση Αντιστοιχιών είναι η Απλή Ανάλυση Αντιστοιχιών είτε του πίνακα Burt είτε του πίνακα δείκτη, ωστόσο στα στατιστικά προγράμματα προσφέρεται σαν διαφορετική ρουτίνα.

Εφόσον η μέθοδος στηρίζεται στην ανάλυση των ιδιοτιμών και επομένως στην αδράνεια, το πρώτο βήμα που θα πρέπει να κάνουμε μετά τον υπολογισμό είτε του πίνακα δείκτη είτε του πίνακα Burt θα είναι να υπολογίσουμε την αδράνεια του εκάστοτε πίνακα.

Ξεκινώντας με τον πίνακα δείκτη \mathbf{X} , η αδράνεια του πίνακα αυτού θα είναι ίση με το μέσο όρο των αδρανειών των υποπινάκων που τον συνιστούν. Αυτό συμβαίνει καθώς τα περιθώρια αθροίσματα των γραμμών όλων των υποπινάκων είναι ίσα μεταξύ τους, και η τιμή αυτή φυσικά είναι 1 εφόσον είναι δίτιμοι πίνακες 0-1. Έστω ότι έχουμε Q το πλήθος μεταβλητές και άρα υποπίνακες \mathbf{X}_q όπου $q = 1, 2, \dots, Q$, και έστω ότι κάθε ένας από αυτούς αποτελείται από J_q ψευδομεταβλητές. Επομένως το πλήθος όλων των ψευδομεταβλητών του πίνακα θα ισούται με $J = \sum_{q=1}^Q J_q$. Η ολική αδράνεια του πίνακα δείκτη \mathbf{X} , είναι¹¹:

$$\Lambda^2(\mathbf{X}) = \frac{1}{Q} \sum_{q=1}^Q \Lambda^2(\mathbf{X}_q) = \frac{1}{Q} \sum_{q=1}^Q (J_q - 1) = \frac{J - Q}{Q}$$

¹¹ Βιβλιογραφική αναφορά (B2)

Με βάση την παραπάνω αδράνεια έχει προκύψει και ένα κριτήριο στο οποίο μπορούμε να βασιστούμε για να επιλέξουμε το πλήθος των διαστάσεων που θα κρατήσουμε. Εφόσον η διάσταση του πίνακα \mathbf{X} είναι $J - Q$ η μέση αδράνεια κάθε διάστασης εάν όλες οι διαστάσεις είχαν την ίδια αδράνεια θα ήταν $1/Q$. Μπορούμε να θεωρήσουμε τη μέση αδράνεια ως κατώφλι και να κρατήσουμε τους άξονες που εξηγούν αδράνεια μεγαλύτερη από αυτή την τιμή.

Ο πίνακας Burt \mathbf{B} έχει επίσης ίσα περιθώρια αθροίσματα γραμμών για κάθε ομάδα οριζόντιων πινάκων και ίσα αθροίσματα στηλών για κάθε ομάδα κάθετων πινάκων. Βάσει αυτού η αδράνεια του πίνακα \mathbf{B} ισούται με το μέσο όρο των αδρανειών των υποπινάκων που τον συνιστούν. Η αδράνεια των υποπινάκων της διαγωνίου παρόμοια με τον πίνακα δείκτη είναι ίση με τη διάσταση του χώρου, δηλαδή $J_q - 1$. Ο πίνακας Burt εμφανίζει μικρότερη αδράνεια από τον πίνακα δείκτη, ωστόσο και πάλι λόγω των αδρανειών των διαγώνιων πινάκων του η συνολική αδράνεια του πίνακα παραμένει αρκετά υψηλή. Αυτό οδηγεί σε μικρά ποσοστά ερμηνευμένης αδράνειας από τις πρώτες κύριες διαστάσεις, το οποίο όμως είναι υποεκτίμηση της πραγματικής ποιότητας της αναπαράστασης των δεδομένων.

Μία λύση που έχει προταθεί από τον Benzécri¹² είναι να προσαρμόσουμε τις βασικές αδράνεις που εξηγεί ο κάθε κύριος άξονας $k = 1, 2, \dots, K$ βάσει της σχέσης:

$$\lambda_k^{2*} = \left(\frac{Q}{Q-1}\right)^2 \left(\lambda_k^2 - \frac{1}{Q}\right)^2$$

όπου λ_k^2 είναι η ιδιοτιμή που αντιστοιχεί στον κύριο άξονα k , Q το πλήθος των μεταβλητών και λ_k^{2*} η προσαρμοσμένη ιδιοτιμή που αντιστοιχεί στον κύριο άξονα k . Αυτή η προσαρμογή βασίζεται στο γεγονός ότι δεν λαμβάνουμε υπόψιν τις ιδιοτιμές που είναι μικρότερες της μέσης αδράνειας. Συνεπώς, σε αυτή την περίπτωση το ποσοστό της αδράνειας που ερμηνεύει ο κάθε κύριος άξονας υπολογίζεται:

¹² Βιβλιογραφική αναφορά (B6)

$$\frac{\lambda_k^{2*}}{\sum_{\lambda_k^2 > 1/Q} \lambda_k^{2*}}$$

Μία ακόμη πρόταση έχει γίνει από τον Greenacre¹³, η οποία όμως δίνει μικρότερα ποσοστά ερμηνευμένης αδράνειας και υπολογίζεται:

$$\lambda_k^{2**} = \frac{Q}{Q-1} \left(\sum_{k=1}^K \lambda_k^2 - \frac{J-Q}{Q^2} \right)$$

όπου λ_k^{2**} είναι η προσαρμοσμένη ιδιοτιμή που αντιστοιχεί στον κύριο άξονα k σύμφωνα με τον Greenacre.

Προσαρμόζοντας τις αδράνεις καταφέρνουμε να αντισταθμίσουμε την υποεκτίμηση της ερμηνευμένης αδράνειας που προκύπτει από τους υποπίνακες της διαγωνίου του πίνακα Burt. Όμως, η ποιότητα της ανασύνταξης των αρχικών δεδομένων βασισμένοι στις ιδιοτιμές που είναι μεγαλύτερες από τη μέση αδράνεια ενδέχεται περαιτέρω βελτίωση¹⁴. Γι' αυτό το λόγο έχει αναπτυχθεί μία εναλλακτική μέθοδος, η **Από Κοινού Ανάλυση Αντιστοιχιών** (Joint Correspondence Analysis-συντομογραφία: JCA). Η JCA αγνοεί τους διαγώνιους υποπίνακες, που άλλωστε δεν περιέχουν σημαντική πληροφορία, με αποτέλεσμα να μειώνει σημαντικά την ολική αδράνεια. Ωστόσο, η συγκεκριμένη μέθοδος δεν ενδείκνυται για μεγάλους πίνακες δεδομένων λόγω της πολυπλοκότητας του αλγορίθμου της και επομένως δεν θα χρησιμοποιηθεί στη δική μας ανάλυση.

¹³ Βιβλιογραφική αναφορά (B7)

¹⁴ Βιβλιογραφική αναφορά (B14) σελ. 17

ΚΕΦΑΛΑΙΟ 3

Δεδομένα

3.1 Παρουσίαση των μεταβλητών

Στο παρόν κεφάλαιο θα παρουσιάσουμε αρχικά τις μεταβλητές που απαρτίζουν το σύνολο των δεδομένων που θα αναλύσουμε καθώς και μία βασική περιγραφική εξέταση αυτών. Τα περιγραφικά μέτρα των μεταβλητών έχουν γίνει με το στατιστικό πρόγραμμα SPSS ενώ για την ανάλυση τους χρησιμοποιήσαμε το στατιστικό πρόγραμμα CHIC Analysis¹⁵.

Τα δεδομένα έχουν αντληθεί από την έρευνα του SHARE (Survey of Health, Ageing and Retirement in Europe). Η έρευνα αυτή συλλέγει πληροφορίες από ανθρώπους ηλικίας 50 ετών και άνω μέσω ερωτηματολογίων και περιέχει δεδομένα για την υγεία, την κοινωνικοοικονομική κατάσταση και τα κοινωνικά και οικογενειακά δίκτυα. Είναι μία διαχρονική έρευνα καθώς επαναλαμβάνεται στο χρόνο και έχει ήδη μέχρι στιγμής πραγματοποιηθεί 6 φορές. Στην παρούσα εργασία θα αναλύσουμε τα δεδομένα του 5^{ου} κύματος που ολοκληρώθηκε τον Νοέμβριο του 2013.

Από το σύνολο των διαθέσιμων μεταβλητών έχουμε επιλέξει να χρησιμοποιήσουμε 19 μεταβλητές που χωρίζονται σε δύο βασικές κατηγορίες, 6 δημογραφικές και 13 που περιέχουν πληροφορία σχετικά με τον εάν κάποιο άτομο είχε παρουσιάσει στο παρελθόν ή έπασχε κατά τη στιγμή της συλλογής των δεδομένων από κάποια ασθένεια. Ακόμη, από το σύνολο των δεδομένων έχουν αφαιρεθεί τα άτομα στα οποία υπήρχαν ελλείψεις τιμές (missing values) και το σύνολο των δεδομένων μας περιλαμβάνει τελικώς 60.153 άτομα. Στον παρακάτω πίνακα μπορούμε να δούμε τις μεταβλητές αυτές καθώς και την περιγραφή της κάθε μίας:

¹⁵ Βιβλιογραφική αναφορά (A4)

ΜΕΤΑΒΛΗΤΕΣ	ΑΓΓΛΙΚΗ ΠΕΡΙΓΡΑΦΗ	ΕΛΛΗΝΙΚΗ ΠΕΡΙΓΡΑΦΗ
MERGEID	Person identifier	Αναγνωριστικό προσώπου
COUNTRY	Country	Χώρα
GENDER	Gender	Φύλο
AGE	Age	Ηλικία
YEDU	Years of education	Έτη εκπαίδευσης
MSTAT	Marital status	Οικογενειακή κατάσταση
FDISTRESS	Household able to make ends meet	Οικονομική κατάσταση νοικοκυριού
HEART_ATT	Heart attack	Καρδιακή προσβολή
H_BLOODPR	High blood pressure or hypertension	Υψηλή αρτηριακή πίεση ή υπέρταση
H_CHOL	High blood cholesterol	Υψηλή χοληστερόλη αίματος
STROKE	Stroke	Εγκεφαλική συμφόρηση
DIAB	Diabetes or high blood sugar	Διαβήτης ή υψηλό σάκχαρο στο αίμα
CR_LUNG	Chronic lung disease	Χρόνια πνευμονική νόσο
CANCER	Cancer	Καρκίνος
STOM_ULCHER	Stomach or duodenal ulcer, peptic ulcer	Στομαχικό ή δωδεκαδακτυλικό έλκος, πεπτικό έλκος
CATAR	Cataracts	Καταρράκτης
EMOT_DIS	Other affective/emotional disorders	Άλλες συγκινησιακές/συναισθηματικές διαταραχές
RH_ARTH	Rheumatoid arthritis	Ρευματοειδής αρθρίτιδα
OSTEOAR	Osteoarthritis/other rheumatism	Οστεοαρθρίτιδα / άλλοι ρευματισμοί
EURODCAT	EURO-D caseness	Κατάθλιψη

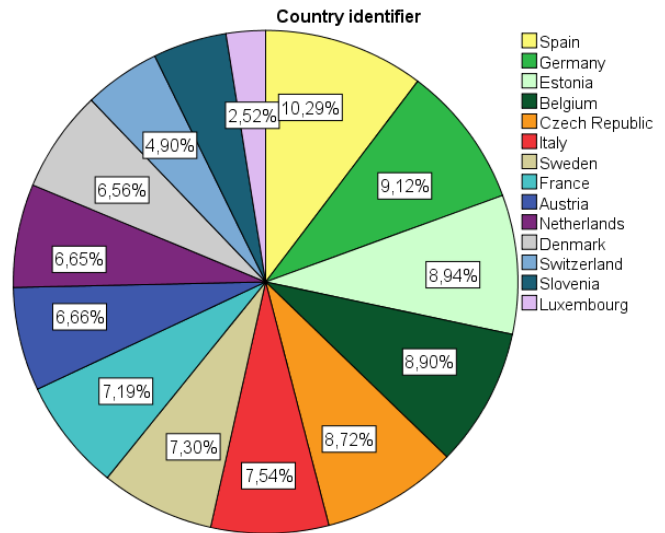
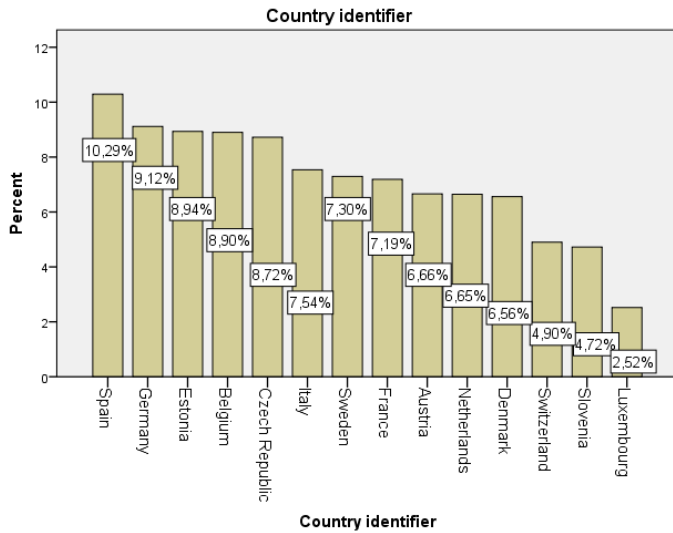
3.2 Περιγραφικά Μέτρα Μεταβλητών

Οι δημογραφικές μεταβλητές που έχουμε επιλέξει είναι η Χώρα, το Φύλο, η Ηλικία, τα Έτη εκπαίδευσης, η Οικογενειακή κατάσταση καθώς και η Οικονομική κατάσταση του νοικοκυριού. Στο παρόν σημείο παρατίθενται οι πίνακες συχνοτήτων των μεταβλητών καθώς και τα σχετικά στατιστικά διαγράμματα (ραβδογράμματα, κυκλικά διαγράμματα, ιστογράμματα κτλ.). Πιο αναλυτικά έχουμε:

- **Χώρα (COUNTRY):** Η μεταβλητή αυτή περιλαμβάνει τη χώρα στην οποία μένει το άτομο που ερωτάται. Στον παρακάτω πίνακα μπορούμε να δούμε πως κατανέμονται τα άτομα στις χώρες σε φθίνουσα σειρά συχνότητας εμφάνισης:

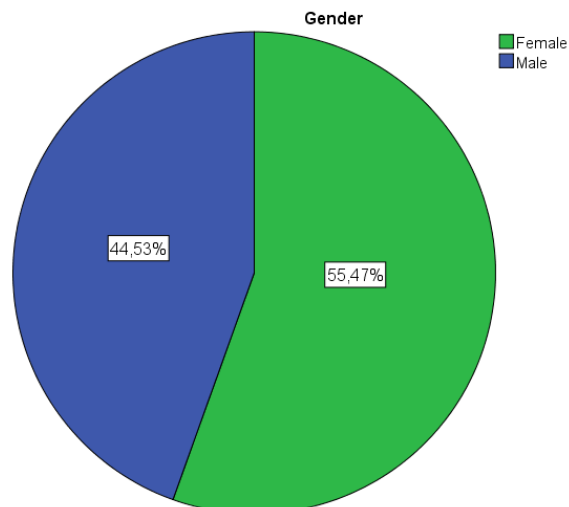
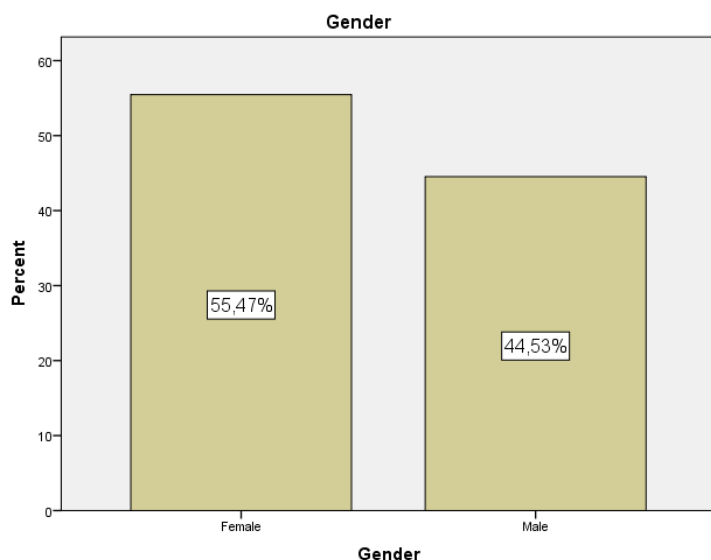
Country	Code	Frequency	Percent	Cumulative Percent
Spain	Spain	6192	10,29%	10,29%
Germany	Germany	5484	9,12%	19,41%
Estonia	Estonia	5375	8,94%	28,35%
Belgium	Belgium	5352	8,90%	37,24%
Czech Republic	Czech	5247	8,72%	45,97%
Italy	Italy	4533	7,54%	53,50%
Sweden	Sweden	4389	7,30%	60,80%
France	France	4326	7,19%	67,99%
Austria	Austria	4007	6,66%	74,65%
Netherlands	Netherl	3998	6,65%	81,30%
Denmark	Denmark	3946	6,56%	87,86%
Switzerland	Swiss	2947	4,90%	92,76%
Slovenia	Slovenia	2842	4,72%	97,48%
Luxembourg	Luxem	1515	2,52%	100,00%
Total		60153	100,00%	

Από τον πίνακα αυτόν καθώς και από το αντίστοιχο ραβδόγραμμα και κυκλικό διάγραμμα που ακολουθούν παρατηρούμε ότι το μεγαλύτερο μέρος του δείγματος προέρχεται από την Ισπανία με ποσοστό 10,29% και το μικρότερο μέρος στο Λουξεμβούργο με ποσοστό 2,52%. Ωστόσο σε γενικές γραμμές όλες οι χώρες εκπροσωπούνται αρκετά ικανοποιητικά στο δείγμα.



- Φύλο (GENDER):** Η μεταβλητή αυτή μας παρέχει την πληροφορία πως κατανέμονται τα δύο φύλα στα δεδομένα μας. Οι γυναίκες αποτελούν το 55,47% του δείγματος ενώ οι άντρες το υπόλοιπο 44,53%. Επομένως το δείγμα μας αποτελείται από περισσότερες γυναίκες σε σχέση με τους άντρες όμως η διαφορά αυτή είναι μικρή και συμπεραίνουμε ότι και τα δύο φύλα εκπροσωπούνται εξίσου καλά στο δείγμα. Οι ακριβείς συχνότητες (απόλυτες και σχετικές) με τις οποίες αυτά εμφανίζονται δίνονται παρακάτω μαζί με το ραβδόγραμμα και το κυκλικό διάγραμμα.

Gender	Frequency	Percent	Cumulative Percent
Female	33368	55,47%	55,47%
Male	26785	44,53%	100,00%
Total	60153	100,00%	



- Ηλικία (AGE):** Στον παρακάτω πίνακα μπορούμε να δούμε μερικά στατιστικά στοιχεία για να καταλάβουμε πως διαμορφώνονται οι ηλικίες των ατόμων στο δείγμα.

Statistics of Age		
Mean		66,82
Std. Deviation		9,837
Median		66
Mode		65
Range		54
Minimum		50
Maximum		104
Skewness		,396
Kurtosis		-,585
Percentiles	25	59
	50	66
	75	74

Το δείγμα επομένως αποτελείται από ανθρώπους ηλικίας 50 έως 104 ετών με μέση ηλικία τα 66,82 έτη και τυπική απόκλιση 9,837 έτη. Η πιο συχνά εμφανιζόμενη ηλικία είναι τα 65 έτη ενώ η διάμεση ηλικία τα 66 έτη. Το 25% των ατόμων ήταν στη διάρκεια της διεξαγωγής της έρευνας 50 έως 59 ετών, το επόμενο 25% απαρτίζεται από τις ηλικίες 60 έως 66 ετών, το 75% του δείγματος είναι έως 75

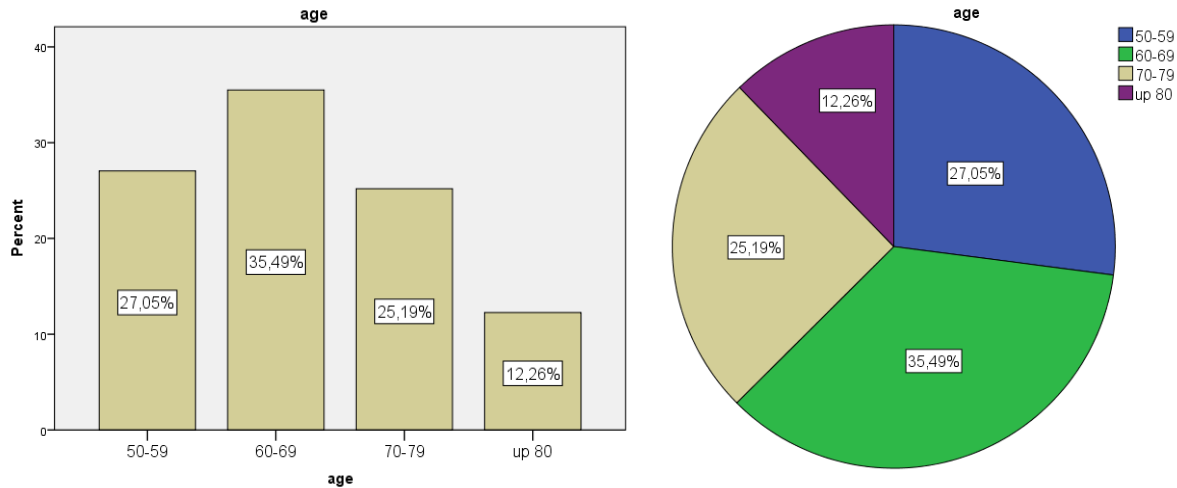
ετών ενώ τα άτομα άνω των 75 ετών απαρτίζουν το τελευταίο 25% του δείγματος. Τέλος η θετική λοξότητα με τιμή 0,396 μας δείχνει ότι τα δεδομένα μας έχουν δεξιά λοξότητα, δηλαδή το μεγαλύτερο πλήθος των ατόμων συναντάται σε ηλικίες μικρότερες της μέσης ηλικίας ενώ η αρνητική της κύρτωσης μας δείχνει ότι η κατανομή μας είναι πλατύκυρτη που σημαίνει ότι το δείγμα μας έχει μικρότερη συγκέντρωση στη μέση τιμή από μία κανονική κατανομή και οι ουρές της κατανομής μας είναι πιο ψηλές. Αυτό συμβαίνει κυρίως επειδή όπως βλέπουμε και στο ιστόγραμμα δεν υπάρχει καμία παρατήρηση μικρότερη των 50 ετών. Για την καλύτερη κατανόηση των ηλικιών του δείγματος παρατίθεται παρακάτω το ιστόγραμμα ενώ ο αναλυτικός πίνακας συχνοτήτων δίνεται στο Παράρτημα (Π1).



Εφόσον η Πολλαπλή Ανάλυση Αντιστοιχιών διαχειρίζεται αποκλειστικά κατηγορικά δεδομένα θα πρέπει η μεταβλητή ηλικία να χωριστεί σε κατηγορίες. Ο τρόπος με τον οποίο θα δημιουργηθούν μπορεί να θεωρηθεί υποκειμενικός ωστόσο η απόφασή μας έχει στηριχθεί τόσο στο ιστόγραμμα όσο και στα αποτελέσματα που μας έχει δώσει η μέθοδος με διάφορους διαχωρισμούς που έχουν δοκιμαστεί. Ο βέλτιστος διαχωρισμός στον οποίο καταλήξαμε είναι:

Age	Frequency	Percent	Cumulative Percent
50-59	16271	27,05%	27,05%
60-69	21351	35,49%	62,54%
70-79	15155	25,19%	87,74%
up 80	7376	12,26%	100,00%
Total	60153	100,00%	

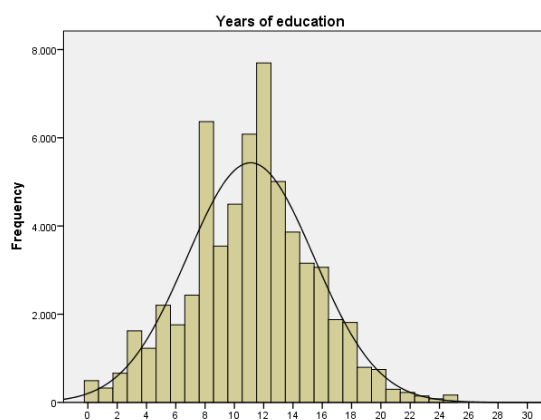
Βάσει αυτής της κατηγοριοποίησης οι ομάδες ηλικιών έχουν χωριστεί σε όσο το δυνατόν ισοπληθείς ομάδες με εξαίρεση τους άνω των 80 ετών. Όπως φαίνεται και από το ραβδόγραμμα και το κυκλικό διάγραμμα που ακολουθούν, το 27,05% του πληθυσμού είναι ηλικίας 50-59 ετών, το 35,49% ηλικίας 60-69 ετών, το 25,19% ηλικίας 70-79 ετών και ένα ποσοστό ίσο με 12,26% αποτελείται από ανθρώπους άνω των 80 ετών.



- **Έτη εκπαίδευσης (YEDU):** Κάποια βασικά στατιστικά στοιχεία για τον τρόπο με τον οποίο διαμορφώνονται τα έτη εκπαίδευσης στο δείγμα είναι:

Statistics of Years of Education	
Mean	11,10
Std. Deviation	4,33
Median	11
Mode	12
Range	25
Minimum	0
Maximum	25
Skewness	,014
Kurtosis	,094
Percentiles	
	25
	50
	75
	8
	11
	14

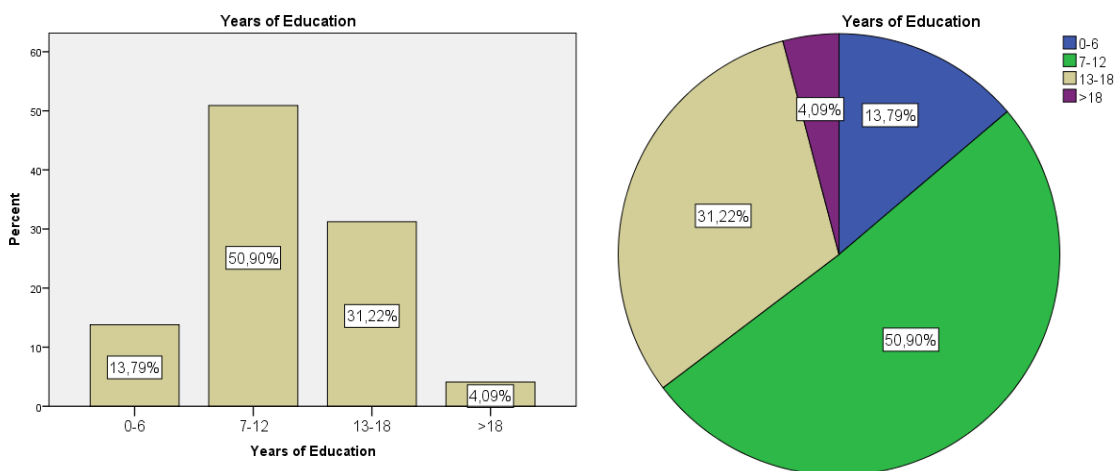
Η μέση τιμή των ετών εκπαίδευσης στο δείγμα διαμορφώνεται περίπου στα 11 έτη με τυπική απόκλιση 4,33 έτη. Η πιο συχνά παρατηρούμενη τιμή είναι τα 11 έτη ενώ ο διάμεσος είναι 12 έτη. Το δείγμα μας περιλαμβάνει από άτομα χωρίς καθόλου εκπαίδευση έως και άτομα με 25 έτη εκπαίδευσης ενώ η θετική λοξότητα και η θετική συμμετρία μας δείχνουν ότι έχουμε μία ελαφρώς λεπτόκυρτη κατανομή με λίγη δεξιά λοξότητα η οποία οφείλετε κυρίως στο γεγονός ότι υπάρχουν πολύ λιγότερες παρατηρήσεις με περισσότερα από 20 έτη εκπαίδευσης σε σχέση με λιγότερα από 5 έτη. Από το ιστόγραμμα παρατηρούμε ότι μεγάλο μέγεθος του δείγματος συγκεντρώνεται στα 8, 11 και 12 έτη. Τέλος, το 25% του δείγματος έχει συμπληρώσει τα 8 έτη εκπαίδευσης, το 50% συμπληρώνεται στα 11 έτη ενώ ένα ποσοστό 25% έχουν πάνω από 14 έτη.



Ο αναλυτικός πίνακας συχνοτήτων βρίσκεται στο Παράρτημα (Π2) και όπως και στην μεταβλητή ηλικία έτσι και εδώ έχει μετατραπεί η μεταβλητή σε κατηγορική με τις εξής κλάσεις:

Years of Education	Code	Frequency	Percent	Cumulative Percent
0-6	Basic	8298	13,79%	13,79%
7-12	High school	30615	50,90%	64,69%
13-18	Uni	18778	31,22%	95,91%
>18	Advanced	2462	4,09%	100,00%
Total		60153	100,00%	

Εφόσον η έρευνα αφορά περισσότερες από μία χώρες με διαφορετική διάρθρωση των επιπέδων εκπαίδευσης η κατηγοριοποίηση που έχει γίνει μπορεί να θεωρηθεί υποκειμενική. Ωστόσο, μπορούμε να πούμε ότι η πρώτη κλάση εκφράζει ανθρώπους με βασική εκπαίδευση, η δεύτερη με μέση εκπαίδευση, η τρίτη με ανώτερη και η τέταρτη με ανώτατη εκπαίδευση. Από τον παραπάνω πίνακα καθώς και από το ραβδόγραμμα και το κυκλικό διάγραμμα που δίνονται παρακάτω φαίνεται ότι το 50,9% του δείγματος έχει 7-12 έτη εκπαίδευσης, ακολουθεί το 31,22% με 13-18 έτη, το 13,79% έχει έως 6 έτη εκπαίδευσης ενώ μόλις το 4,09% έχει παραπάνω από 18 έτη εκπαίδευσης.



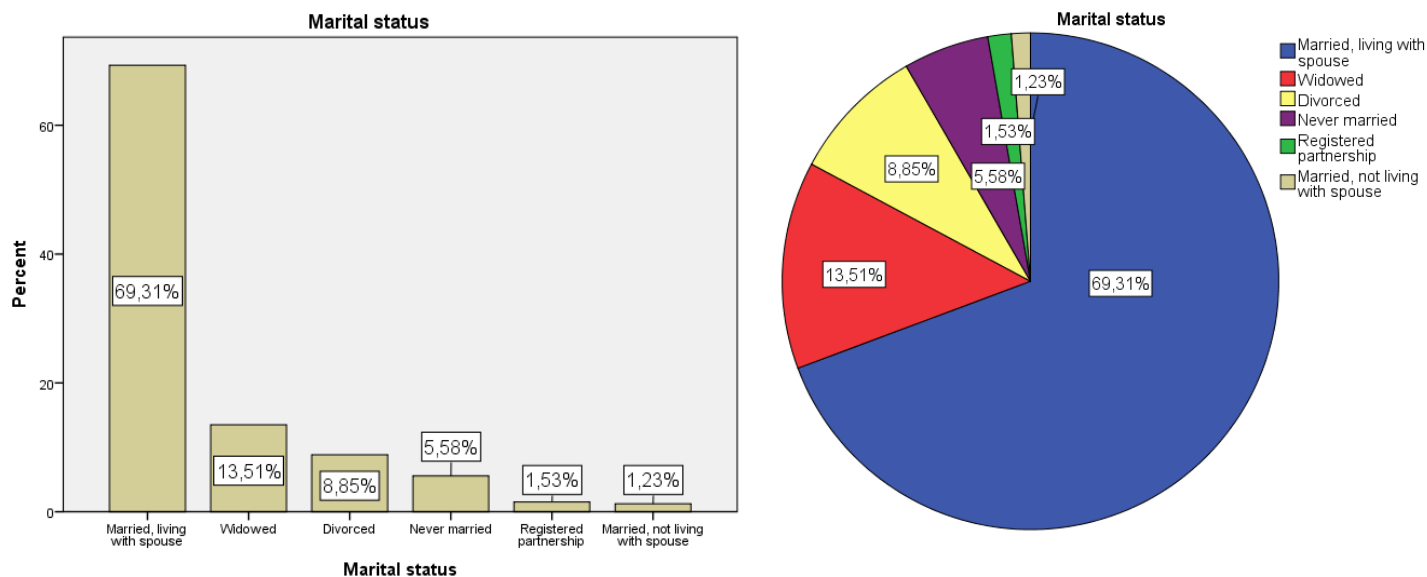
- **Οικογενειακή κατάσταση (MSTAT):** Στην παρούσα μεταβλητή περιέχεται η πληροφορία ως προς το εάν το άτομο ανήκει σε μία από τις παρακάτω κατηγορίες:

ΑΓΓΛΙΚΗ ΠΕΡΙΓΡΑΦΗ	ΕΛΛΗΝΙΚΗ ΠΕΡΙΓΡΑΦΗ
Married, living with spouse	Παντρεμένος/η, συζεί με τον/την σύζυγο
Registered partnership	Έχει εγγεγραμμένο/η σύντροφο
Married, not living with spouse	Παντρεμένος/η, δεν ζει με τον/την σύζυγο
Never married	Δεν έχει παντρευτεί ποτέ
Divorced	Διαζευγμένος/η
Widowed	Χήρος/α

Οι συχνότητες με τις οποίες εμφανίζονται οι παραπάνω κατηγορίες στο δείγμα έχουν ως εξής:

Marital status	Code	Frequency	Percent	Cumulative Percent
Married, living with spouse	Married with	41690	69,31%	69,31%
Registered partnership	Partner	919	1,53%	70,83%
Married, not living with spouse	Marr. without	740	1,23%	72,06%
Never married	N. Married	3356	5,58%	77,64%
Divorced	Divorced	5321	8,85%	86,49%
Widowed	Widowed	8127	13,51%	100,00%
Total		60153	100,00%	

Παρατηρούμε ότι το 69,31% των ανθρώπων είναι παντρεμένοι και ζουν μαζί τον/την σύζυγό τους ενώ το αμέσως επόμενο ποσοστό είναι το 13,51% που αφορά τους χήρους/ες. Το ποσοστό των διαζευγμένων είναι μόλις στο 8,85%, λογικό εάν σκεφτούμε ότι τη χρονική περίοδο της συλλογής των δεδομένων τα άτομα που ερωτήθηκαν ήταν ήδη σε ηλικία άνω των 50 ετών και επομένως ανήκουν σε μία γενιά που γνωρίζουμε ότι δεν είχε υψηλά ποσοστά διαζυγίων σε αντίθεση με τις νεότερες γενιές. Το ραβδόγραμμα και το κυκλικό διάγραμμα της μεταβλητής εμφανίζονται παρακάτω:

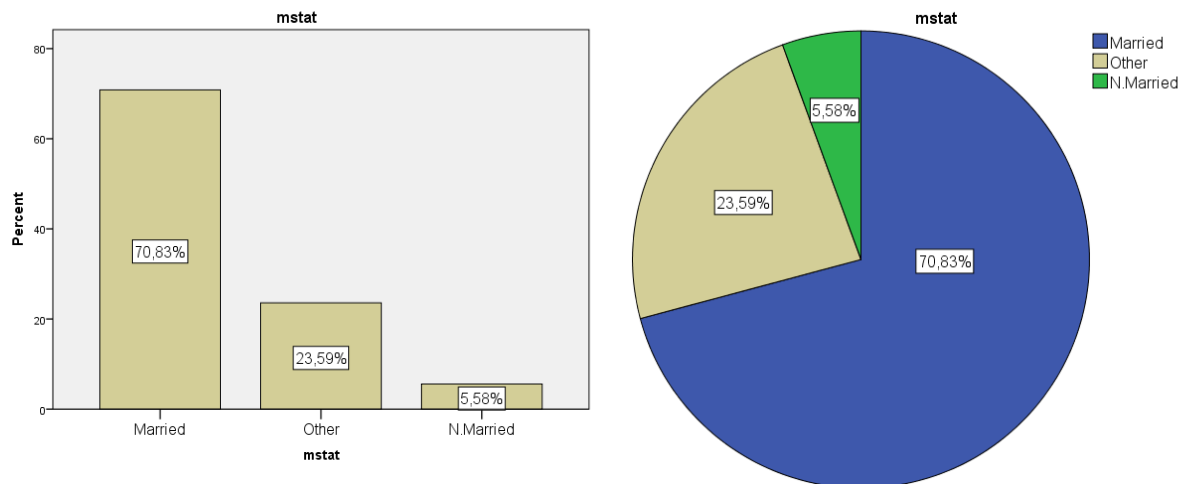


Παρατηρούμε ότι έχουμε κατηγορίες με πολύ μικρές συχνότητες εμφάνισης. Ταυτόχρονα, βάσει της ανάλυσης που έχει γίνει στο επόμενο κεφάλαιο και της ομοιότητας ως προς τον τρόπο ζωής των ανθρώπων, μπορούμε να συγχωνεύσουμε

κάποιες κατηγορίες. Πιο συγκεκριμένα, έχουμε θεωρήσει τους παντρεμένους που συζούν με τους συζύγους τους και όσους έχουν εγγεγραμμένο σύντροφο ως παντρεμένους, έχουμε κρατήσει ακέραια την κατηγορία των ανθρώπων που δεν έχουν παντρευτεί ποτέ και έχουμε θεωρήσει ως άλλη οικογενειακή κατάσταση τους παντρεμένους που δεν ζουν με τον σύζυγό τους, τους διαζευγμένους και τους χήρους. Οι νέες συχνότητες εμφάνισης είναι:

Marital status	Frequency	Percent	Cumulative Percent
Married	42609	70,83%	70,83%
N. Married	3356	5,58%	76,41%
Other	14188	23,59%	100,00%
Total	60153	100,00%	

Από τον παραπάνω πίνακα συχνότητας εμφάνισης καθώς και από το παρακάτω ραβδόγραμμα και κυκλικό διάγραμμα παρατηρούμε ότι το μεγαλύτερο ποσοστό (70,83%) των ανθρώπων ανήκει στην κατηγορία των παντρεμένων, το 23,59% είναι παντρεμένοι που μένουν μόνοι τους, διαζευγμένοι ή χήροι και τέλος με το μικρότερο ποσοστό εμφάνισης (5,58%) έχουμε τους ανθρώπους που δεν έχουν παντρευτεί ποτέ.



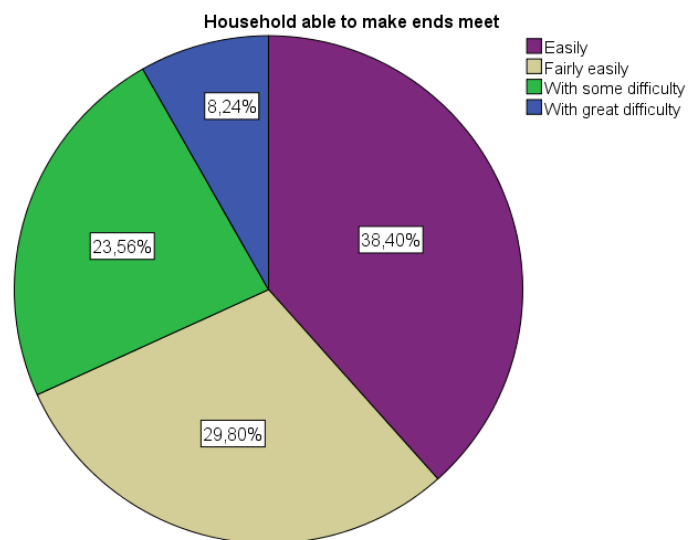
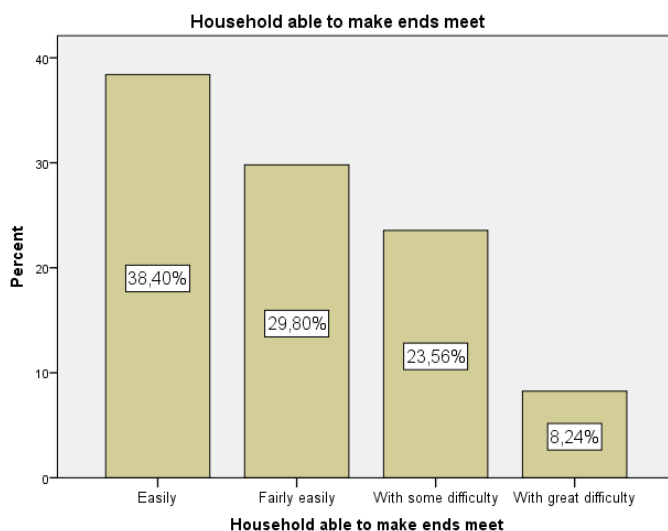
- **Οικονομική κατάσταση νοικοκυριού (FDISTRESS):** Η μεταβλητή αυτή μας πληροφορεί σχετικά με το με πόση ευκολία αντεπεξέρχεται οικονομικά το νοικοκυριό του ατόμου το οποίο ερωτάται.

ΑΓΓΛΙΚΗ ΠΕΡΙΓΡΑΦΗ	ΕΛΛΗΝΙΚΗ ΠΕΡΙΓΡΑΦΗ
With great difficulty	Με μεγάλη δυσκολία
With some difficulty	Με κάποια δυσκολία
Fairly easily	Σχετικά εύκολα
Easily	Εύκολα

Οι συχνότητες εμφάνισης της κάθε κατηγορίας είναι:

Household able to make ends meet	Code	Frequency	Percent	Cumulative Percent
With great difficulty	Too difficult	4958	8,24%	8,24%
With some difficulty	Difficult	14170	23,56%	31,80%
Fairly easily	Easy	17926	29,80%	61,60%
Easily	Too Easy	23099	38,40%	100,00%
Total		60153	100,00%	

Παρατηρούμε ότι η άποψη των ατόμων σχετικά με την οικονομική κατάσταση του νοικοκυριού τους είναι αρκετά καλή καθώς μόνο το 8,24% απάντησε ότι έχει μεγάλη δυσκολία ενώ τα ποσοστά ανεβαίνουν από τη μία κατηγορία στην άλλη και φτάνουμε στο 38,40% να έχει απαντήσει ότι τα βγάζει πέρα με ευκολία. Το ραβδόγραμμα και το κυκλικό διάγραμμα της μεταβλητής εμφανίζονται παρακάτω:



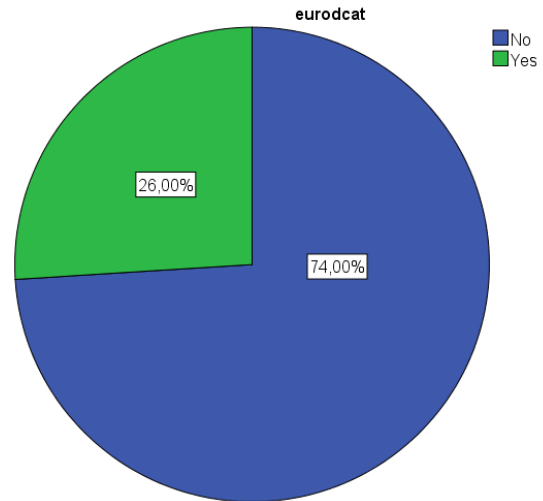
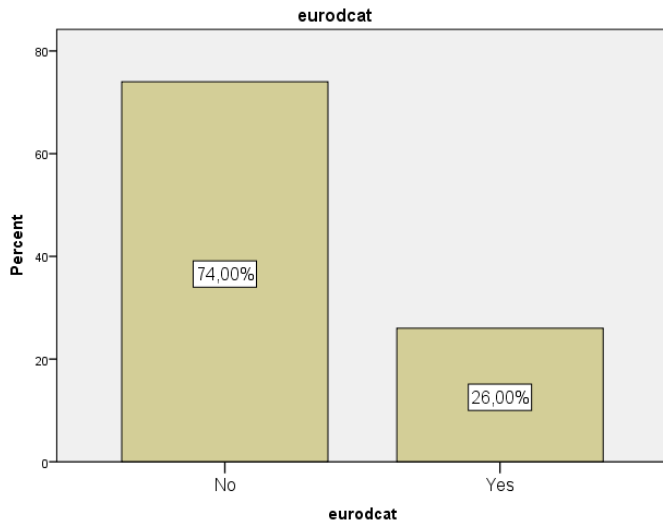
- **Κατάθλιψη (EURODCAT):** είναι μία δίτιμη μεταβλητή στην οποία φαίνεται εάν τα άτομα εμφανίζουν κλινικά σημαντικά συμπτώματα κατάθλιψης. Δημιουργήθηκε από τις παρακάτω 12 μεταβλητές:

ΜΕΤΑΒΛΗΤΕΣ	ΠΕΡΙΓΡΑΦΗ ΜΕΤΑΒΛΗΤΩΝ
euro1	Κατάθλιψη
euro2	Απαισιοδοξία
euro3	Τάσεις αυτοκτονίας
euro4	Ενοχές
euro5	Ύπνος
euro6	Ενδιαφέρον
euro7	Ευερεθιστότητα
euro8	Όρεξη
euro9	Κούραση
euro10	Συγκέντρωση
euro11	Ευχαρίστηση
euro12	Ευσυγκίνητος

Αποτελείται από δύο κατηγορίες, με την τιμή «No» είναι τα άτομα που εμφανίζουν έως τρία από τα παραπάνω συμπτώματα ενώ με την τιμή «Yes» εκείνα που εμφανίζουν πάνω από τρία. Οι συχνότητες εμφάνισης της κάθε κατηγορίας είναι:

EURO-D caseness	Frequency	Percent	Cumulative Percent
Yes	15641	26,00%	26,00%
No	44512	74,00%	100,00%
Total	60153	100,00%	

Παρατηρούμε ότι είναι σχετικά μεγάλο το ποσοστό των ανθρώπων που παρουσιάζουν περισσότερα από 3 συμπτώματα κατάθλιψης, για ακρίβεια 26% ενώ το υπόλοιπο 74% δηλώνει ότι είτε δεν έχει είτε έχει έως 3 συμπτώματα. Τα παραπάνω μπορούν να φανούν και από το παρακάτω ραβδόγραμμα και κυκλικό διάγραμμα:

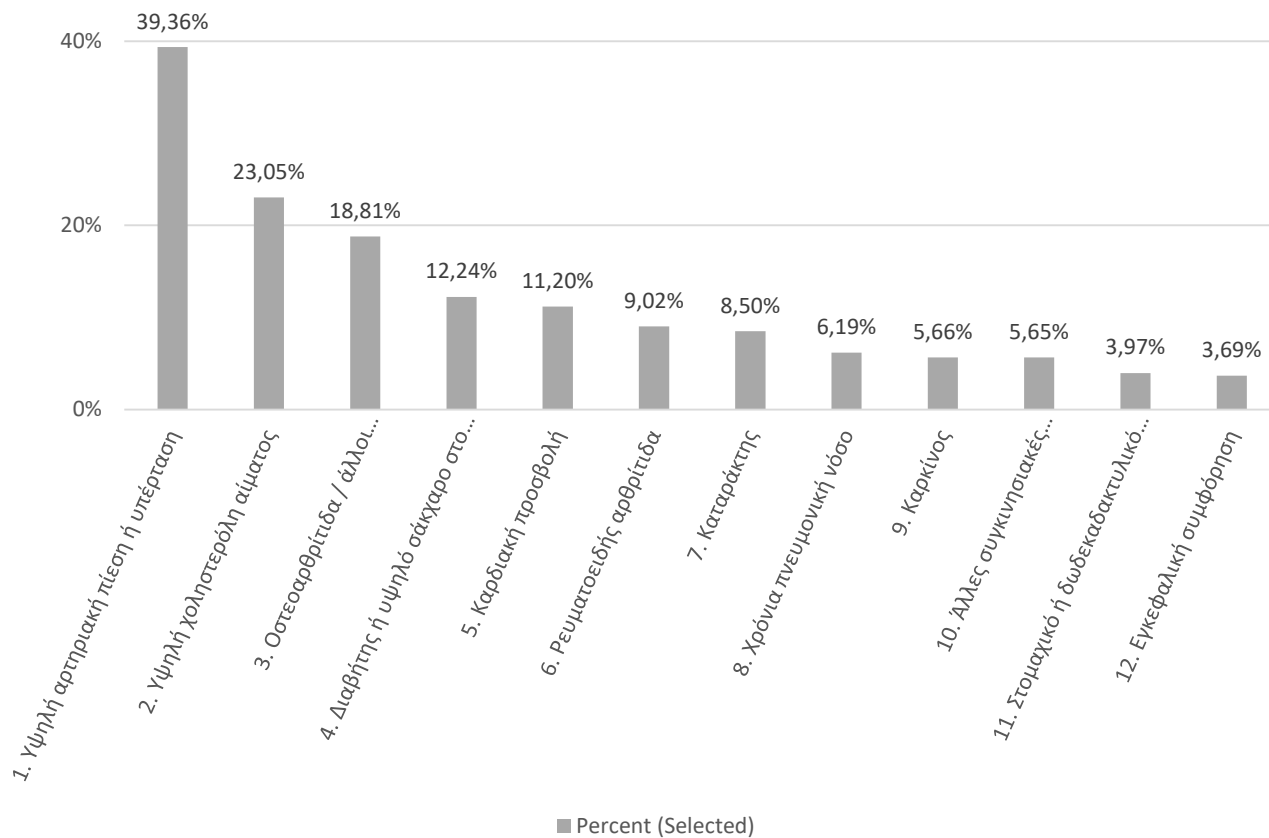


- Μεταβλητές Ασθενειών:** Οι μεταβλητές αυτές αφορούν συγκεκριμένες κατηγορίες ασθενειών και αποτελούνται από δύο κατηγορίες. Η πρώτη αφορά την περίπτωση τα άτομα να μην έχουν διαγνωστεί με την εκάστοτε ασθένεια ούτε στο παρελθόν ούτε στο παρόν και παίρνει την τιμή «Not Selected», αλλιώς παίρνει την τιμή «Selected». Στον παρακάτω συγκεντρωτικό πίνακα μπορούμε να δούμε το πλήθος των ατόμων που δεν έχουν εμφανίσει ούτε έπασχαν τη δεδομένη χρονική στιγμή από την κάθε ασθένεια και το ποσοστό που κατέχει στο σύνολο του δείγματος και έπειτα το πλήθος των ατόμων που είτε έχουν εμφανίσει στο παρελθόν είτε έπασχαν κατά τη διάρκεια της διεξαγωγής της έρευνας την κάθε ασθένεια με το αντίστοιχο ποσοστό. Τα στοιχεία αυτά έχουν διαταχθεί σε φθίνουσα σειρά ανάλογα με το πλήθος των ατόμων που έχουν επιλέξει την τιμή «Selected».

ΑΣΘΕΝΕΙΑ	NOT SELECTED	%	SELECTED	%
1. Υψηλή αρτηριακή πίεση ή υπέρταση	36474	60,64%	23679	39,36%
2. Υψηλή χοληστερόλη αίματος	46289	76,95%	13864	23,05%
3. Οστεοαρθρίτιδα / άλλοι ρευματισμοί	48841	81,19%	11312	18,81%
4. Διαβήτης ή υψηλό σάκχαρο στο αίμα	52792	87,76%	7361	12,24%
5. Καρδιακή προσβολή	53418	88,80%	6735	11,20%
6. Ρευματοειδής αρθρίτιδα	54729	90,98%	5424	9,02%
7. Καταράκτης	55040	91,50%	5113	8,50%
8. Χρόνια πνευμονική νόσο	56428	93,81%	3725	6,19%
9. Καρκίνος	56746	94,34%	3407	5,66%
10. Άλλες συγκινησιακές /συναισθηματικές διαταραχές	56756	94,35%	3397	5,65%
11. Στομαχικό ή δωδεκαδακτυλικό έλκος, πεπτικό έλκος	57765	96,03%	2388	3,97%
12. Εγκεφαλική συμφόρηση	57932	96,31%	2221	3,69%

Από τον παραπάνω πίνακα παρατηρούμε ότι εκτός από την υψηλή αρτηριακή πίεση ή υπέρταση που συναντάται στο 39,36% των ανθρώπων, οι υπόλοιπες ασθένειες έχουν ποσοστά μικρότερα του 24%. Πιο συγκεκριμένα, εκτός από την προαναφερθείσα, οι πιο συχνές ασθένειες στο δείγμα μας είναι υψηλή χοληστερόλη αίματος και η οστεοαρθρίτιδα μαζί με άλλους ρευματισμούς με ποσοστά 23,05% και 18,81% αντίστοιχα ενώ στις πιο σπάνιες περιπτώσεις συναντώνται το στομαχικό ή δωδεκαδακτυλικό έλκος, πεπτικό έλκος και η εγκεφαλική συμφόρηση με ποσοστά 3,97% και 3,69% αντίστοιχα. Στο Παράρτημα (Π3) βρίσκεται ο συγκεντρωτικός πίνακας συχνότητας των ασθενειών όπως προκύπτει από το στατιστικό πρόγραμμα SPSS, ενώ παρακάτω μπορούμε να δούμε το συγκεντρωτικό ραβδόγραμμα που περιλαμβάνει τα ποσοστά των ατόμων του δείγματος που είχαν εμφανίσει στο παρελθόν ή έπασχαν κατά τη διάρκεια της συλλογής των δεδομένων από κάποια από τις ερωτηθείσες ασθένειες.

Disease



ΚΕΦΑΛΑΙΟ 4

Εφαρμογή της μεθόδου και παρουσίαση των αποτελεσμάτων

4.1 Εισαγωγή

Σκοπός της ανάλυσης είναι η αναπαράσταση και η ερμηνεία των συσχετίσεων που υπάρχουν στο σύνολο δεδομένων μας και δημιουργούν πιθανές ομαδοποιήσεις μεταξύ των κατηγοριών των δύο ομάδων των μεταβλητών μας: τις δημογραφικές και τις ασθένειες. Η θεώρηση των ομάδων γίνεται με εποπτικό τρόπο, καθώς μία ομάδα αποτελείται από σημεία που είναι κοντινά μεταξύ τους και απομακρυσμένα από τα υπόλοιπα. Η διαγραμματική απεικόνιση των αποτελεσμάτων θα γίνει με τη βοήθεια παραγοντικών διαγραμμάτων σε επίπεδο αξόνων καθώς και συμμετρικών biplots

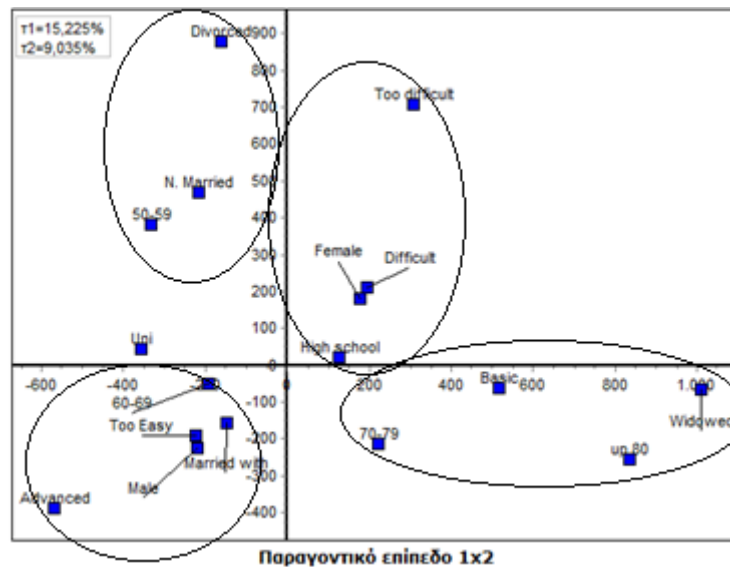
Λόγω της ανομοιογένειας των δύο ομάδων των μεταβλητών μας προτείνεται η ανάλυση να έχει ως βασικές τις μεταβλητές της μίας ομάδας και ως συμπληρωματικές της άλλης. Γενικά συνηθίζεται ως συμπληρωματικές μεταβλητές να θεωρούνται οι δημογραφικές και σαν βασικές να λαμβάνουμε τις υπόλοιπες μεταβλητές (βλ. Greenacre (2007)). Ωστόσο επειδή τα συμπεράσματα βασίζονται στη χρήση των πιο σημαντικών κατηγοριών (με βάση τους δείκτες $COR > 20\%$ και $CTR > 10\%$ όπου οι τιμές έχουν επιλεχθεί μετά από δοκιμές) καθώς και λόγω του μεγάλου πλήθους των μεταβλητών που συμμετέχουν στην ανάλυση είναι πιθανόν αρκετές συσχετίσεις των βασικών μεταβλητών να μην μπορούν να αναπαρασταθούν. Για τους λόγους αυτούς θα παρουσιάσουμε δύο είδη αναλύσεων. Αρχικά θα θεωρήσουμε ως βασικές μεταβλητές τις δημογραφικές με συμπληρωματικές τις ασθένειες και έπειτα θα θεωρήσουμε ως βασικές μεταβλητές τις ασθένειες με συμπληρωματικές τις δημογραφικές.

Το μεγάλο πλήθος των μεταβλητών και των κατηγοριών τους ενδέχεται να απαιτήσει μεγάλο πλήθος κύριων αξόνων για να κρατήσουμε σε υψηλά επίπεδα το

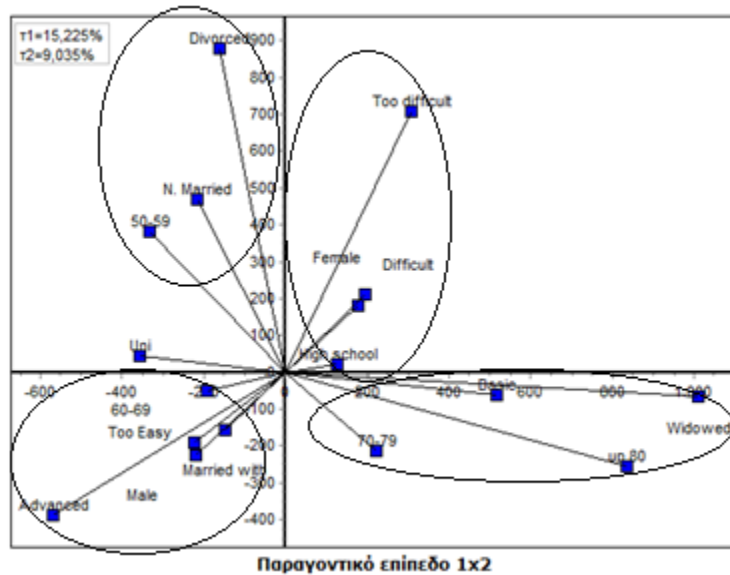
ποσοστό της ερμηνευμένης αδράνειας. Επομένως εκτός από την κλασσική μέθοδο της Πολλαπλής Ανάλυσης Αντιστοιχιών θα εφαρμοσθεί και η βελτίωσή της με την προσαρμοσμένη αδράνεια που έχει προταθεί από τον Greenacre. Τέλος, με βάση τα αποτελέσματα χρειάστηκε να αντικατασταθεί η μεταβλητή της Οικογενειακής Κατάστασης από την νέα μεταβλητή που έχουμε παρουσιάσει στην περιγραφική ανάλυση και συγχωνεύει κάποιες κατηγορίες της και η μεταβλητή Χώρα θα θεωρηθεί και αυτή ως συμπληρωματική μεταβλητή λόγω του μεγάλου πλήθους των κατηγοριών της.

4.2 Πολυμεταβλητή Ανάλυση Αντιστοιχιών στις δημογραφικές μεταβλητές

Θα ξεκινήσουμε με την κλασσική Πολυμεταβλητή Ανάλυση Αντιστοιχιών χρησιμοποιώντας ως βασικές τις δημογραφικές μεταβλητές και εξαιρώντας τις χώρες από την ανάλυση. Το συμμετρικό biplot των κατηγοριών των μεταβλητών είναι:

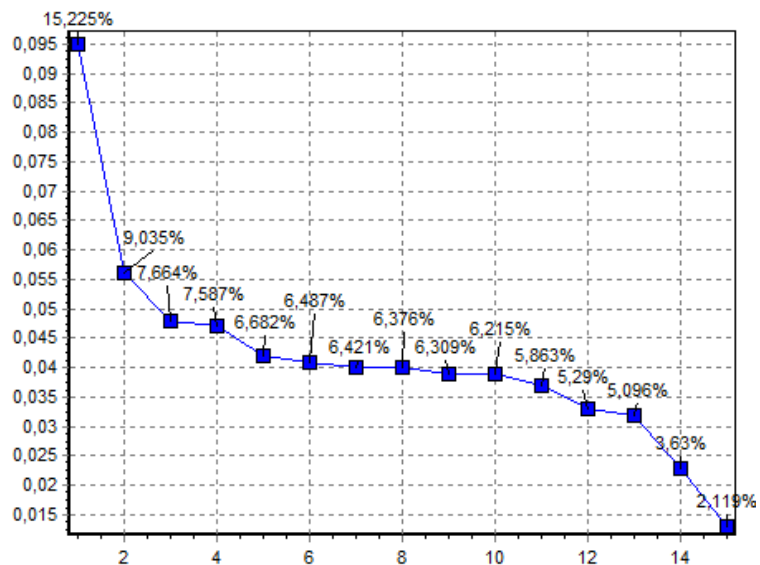


Βοήθεια στην ερμηνεία του παραγοντικού επιπέδου δίνεται από τα διανύσματα θέσης, δηλαδή τα διανύσματα που ξεκινάνε από την αρχή των αξόνων και καταλήγουν στις κατηγορίες. Όσο πιο μικρή είναι η γωνία που σχηματίζεται μεταξύ δύο σημείων τόσο πιο κοντά θεωρούνται ότι βρίσκονται και επομένως ανήκουν στην ίδια ομάδα. Το προηγούμενο συμμετρικό biplot παίρνει τη μορφή:

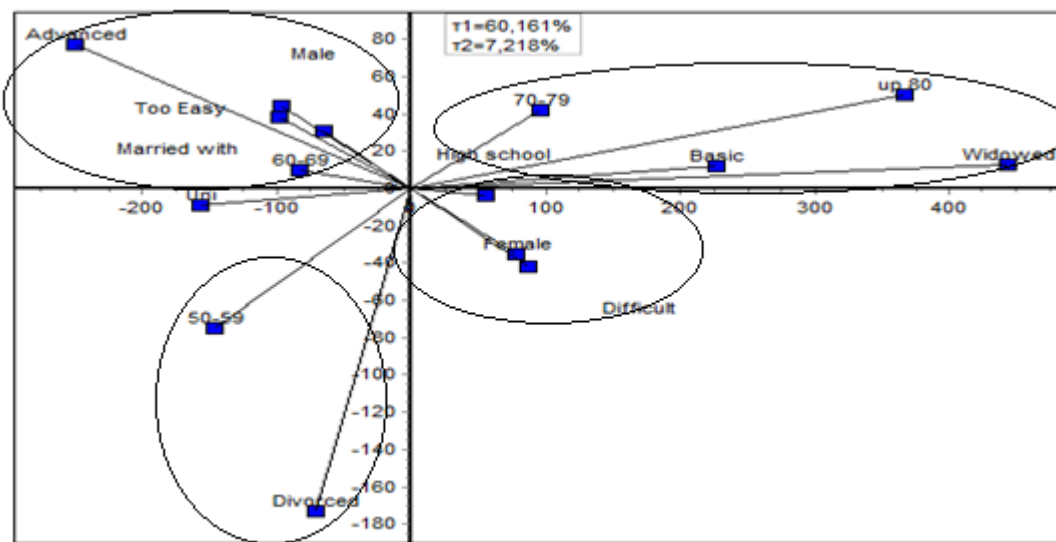


Στο παραπάνω διάγραμμα μπορούμε να διακρίνουμε τέσσερις βασικές ομάδες κατηγοριών. Η πρώτη ομάδα (κάτω αριστερά) περιλαμβάνει άνδρες ηλικίας 60-69 ετών με ανώτατη εκπαίδευση που είναι παντρεμένοι και ζουν με τις συζύγους τους και έχουν πολύ καλή οικονομική κατάσταση. Η δεύτερη ομάδα (πάνω αριστερά) απαρτίζεται από ανθρώπους 50-59 ετών που έχουν πάρει διαζύγιο ή δεν έχουν παντρευτεί ποτέ. Σαν τρίτη ομάδα (κάτω δεξιά) μπορούν να θεωρηθούν οι χήροι/ες ηλικίας άνω των 70 ετών που έχουν πρωτοβάθμια ή δευτεροβάθμια εκπαίδευση. Τέλος, η τέταρτη ομάδα (πάνω δεξιά) απαρτίζεται από γυναίκες με δύσκολη και πολύ δύσκολη οικονομική κατάσταση.

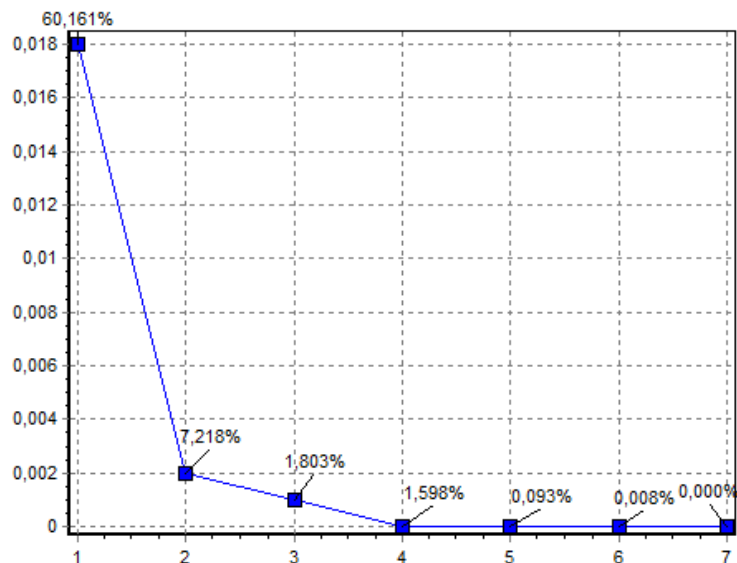
Παρόλα αυτά όπως φαίνεται στο επάνω αριστερά μέρος των διαγραμμάτων, το ποσοστό της ερμηνευμένης αδράνειας είναι μόλις 24,26%. Από αυτό το 15,225% αντιστοιχεί στον 1^ο παραγοντικό άξονα και το υπόλοιπο 9,035% στον 2^ο παραγοντικό άξονα. Σύμφωνα με το Scree test που ακολουθεί θα έπρεπε να κρατήσουμε ίσως και 8 κύριους άξονες για να επιτύχουμε καλύτερη ποιότητα αναπαράστασης (65,478% θα ήταν σε αυτή την περίπτωση το ποσοστό της ερμηνευμένης αδράνειας), γεγονός που θα δυσχέραινε την ανάλυση. Οι τιμές της ερμηνευμένης αδράνειας καθώς και ορισμένοι δείκτες που προκύπτουν από την ανάλυση μπορούν να βρεθούν στο Παράρτημα (Π4).



Σε μία προσπάθεια να βελτιώσουμε την ερμηνευμένη αδράνεια επαναλαμβάνουμε την ανάλυση χρησιμοποιώντας τη διορθωμένη αδράνεια που πρότεινε ο Greenacre. Το συμμετρικό biplot καθώς και το Scree test που προκύπτουν από την ανάλυση είναι:

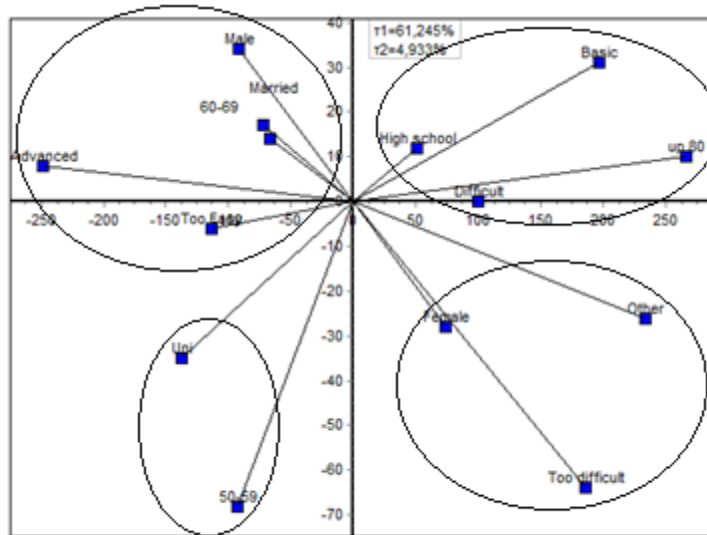


Παραγοντικό επίπεδο 1x2

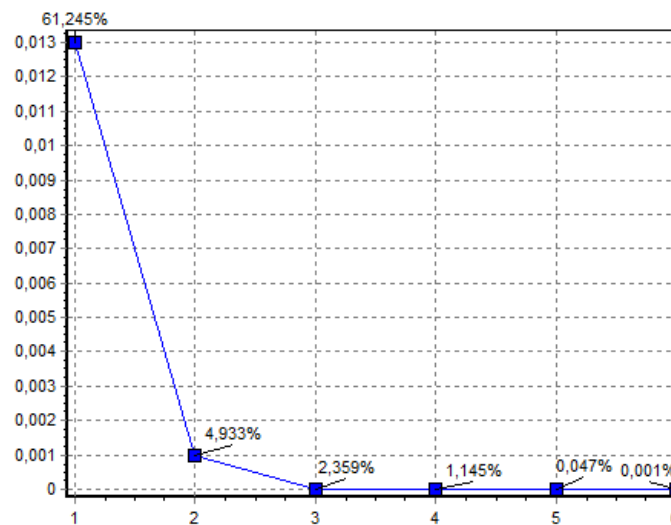


Παρατηρούμε ότι διατηρούνται οι τέσσερις βασικές ομάδες της προηγούμενης ανάλυσης, με διαφορετικό όμως προσανατολισμό καθώς έχουν αντιστραφεί οι ομάδες ως προς τον οριζόντιο άξονα. Όσες προηγουμένως ήταν πάνω τώρα είναι κάτω και το αντίστροφο. Σημαντική αύξηση παρατηρούμε στην ερμηνευμένη αδράνεια που πλέον έχει ποσοστό 67,379% ενώ ο 1^{ος} παραγοντικός άξονας εξηγεί από μόνος του το 60,161% της προσαρμοσμένης αδράνειας. Σύμφωνα με το Scree test δεν χρειαζόμαστε περισσότερους από 2 ή 3 άξονες για να έχουμε πολύ καλή ποιότητα της αναπαράστασης. Οι τιμές της ερμηνευμένης αδράνειας καθώς και ορισμένοι δείκτες που προκύπτουν από την ανάλυση μπορούν να βρεθούν στο Παράρτημα (Π5).

Σε επόμενο στάδιο θα αντικαταστήσουμε την μεταβλητή της Οικογενειακής Κατάστασης με την νέα όπου έχουμε συγχωνεύσει ορισμένες κατηγορίες. Η απόφαση αυτή βασίζεται στο γεγονός ότι οι κατηγορίες με πολύ χαμηλά ποσοστά που περιείχε αρχικά η μεταβλητή αυτή είναι πιθανό να επηρεάζουν την επιλογή των ομάδων και να μην επιτρέπουν την ανάδειξη άλλων κοινών χαρακτηριστικών. Τα αποτελέσματα της ανάλυσης καθώς και το Scree test είναι:



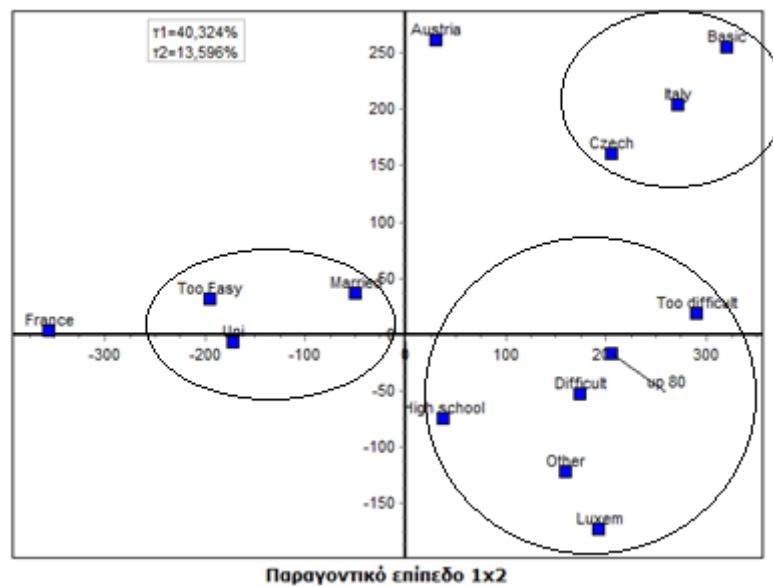
Παραγοντικό επίπεδο 1x2

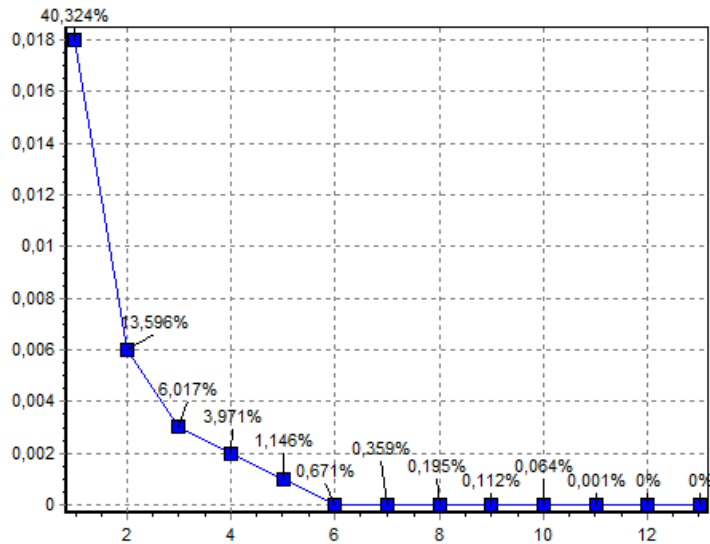


Η αλλαγή στη μεταβλητή της Οικογενειακής Κατάστασης άλλαξε κάπως την περιγραφή των ομάδων. Πιο συγκεκριμένα, πάνω αριστερά παρατηρείται η ομάδα παντρεμένων ανδρών ηλικίας 60-69 ετών με πολύ καλή οικονομική κατάσταση και ανώτατο επίπεδο εκπαίδευσης. Κάτω αριστερά δημιουργείται μία δεύτερη ομάδα που περιλαμβάνει τους ανθρώπους 50-59 ετών με ανώτερη εκπαίδευση. Η τρίτη ομάδα που βρίσκεται πάνω δεξιά έχει αλλάξει ελαφρώς καθώς περιλαμβάνει ηλικίες άνω των 80 ετών με πρωτοβάθμια ή δευτεροβάθμια εκπαίδευση και δυσκολία στα οικονομικά. Τέλος, η τέταρτη ομάδα απαρτίζεται από γυναίκες που ζουν μόνες τους με πολύ κακή οικονομική

κατάσταση. Η ερμηνευμένη αδράνεια είναι 66,178%, ποσοστό μειωμένο σε σχέση με πριν κατά 1,201 μονάδα. Ωστόσο, η ομάδα των γυναικών παριστάνεται καλύτερα και θα επιλέξουμε να κρατήσουμε αυτές τις βασικές ομάδες. Τέλος, από το Scree test μπορούμε να δούμε ότι 2 άξονες είναι αρκετοί για να έχουμε μία ικανοποιητική αναπαράσταση των αρχικών δεδομένων. Οι τιμές της ερμηνευμένης αδράνειας καθώς και ορισμένοι δείκτες που προκύπτουν από την ανάλυση μπορούν να βρεθούν στο Παράρτημα (Π6).

Στο επόμενο στάδιο θα εισάγουμε στην ανάλυση ως βασική και τη μεταβλητή Χώρα για να εξετάσουμε εάν βοηθάει την ανάλυση ή προκαλεί πρόβλημα το μεγάλο πλήθος των κατηγοριών που έχει. Το συμμετρικό biplot των δύο πρώτων κύριων παραγοντικών αξόνων καθώς και το Scree test είναι:

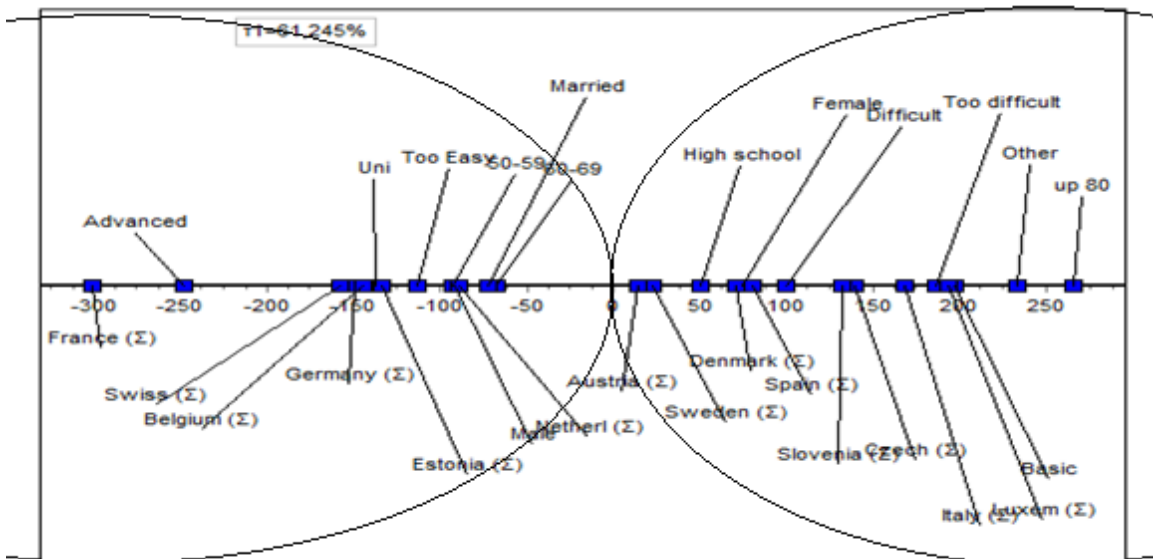




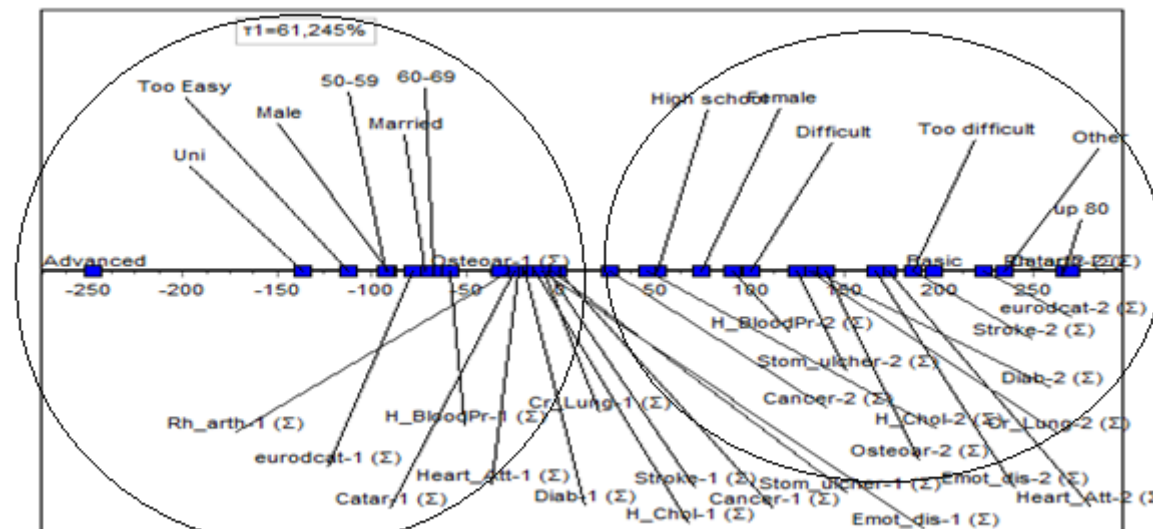
Παρόλο που εμφανίζονται τρεις διακριτές ομάδες μεταβλητών, η εισαγωγή των χωρών στον πίνακα δεδομένων της προηγούμενης ανάλυσης ελαττώνει σημαντικά τον αριθμό των υπόλοιπων μεταβλητών που εμφανίζονται. Πιο συγκεκριμένα, αριστερά εμφανίζεται η ομάδα που περιλαμβάνει τους παντρεμένους με πολύ καλό οικονομικό επίπεδο και ανώτερη εκπαίδευση. Η Γαλλία θα μπορούσε να θεωρηθεί outlier λόγω της υψηλής συσχέτισης με τον 1^ο άξονα (COR=117) ωστόσο έχει χαμηλή μάζα (m=13). Η δεύτερη ομάδα (κάτω δεξιά) περιλαμβάνει κατοίκους Λουξεμβούργου άνω των 80 ετών με δευτεροβάθμια εκπαίδευση που έχουν δύσκολα ή πολύ δύσκολα οικονομικά και είναι είτε χωρισμένοι είτε χήροι είτε δεν ζουν με τους/τις συζύγους τους. Η τρίτη ομάδα (πάνω δεξιά) συγκροτείται από πολίτες της Τσεχίας και της Ιταλίας με βασικό επίπεδο εκπαίδευσης. Η Αυστρία δεν θεωρείται ότι συσχετίζεται με καμία άλλη κατηγορία. Τέλος, το ποσοστό της ερμηνευμένης αδράνειας πέφτει στο 53,92% (40,324% ερμηνεύει ο 1^{ος} κύριος άξονας και 13,596% ο 2^{ος}), δηλαδή μειώνεται κατά 12,258 μονάδες. Οι τιμές της ερμηνευμένης αδράνειας καθώς και ορισμένοι δείκτες που προκύπτουν από την ανάλυση μπορούν να βρεθούν στο Παράρτημα (Π7).

Περνώντας στην τελευταία ανάλυση όπου οι δημογραφικές μεταβλητές χρησιμοποιούνται ως βασικές, επιλέγουμε τελικά να χρησιμοποιήσουμε την Οικογενειακή Κατάσταση αφού έχει δεχθεί συγχώνευση στις κατηγορίες της και να εμφανίσουμε τις χώρες σαν συμπληρωματικές μεταβλητές (τα σημεία που αφορούν τις κατηγορίες των

συμπληρωματικών μεταβλητών φέρουν το διακριτικό (Σ) δίπλα στην ονομασία τους). Λόγω του μεγάλου πλήθους των μεταβλητών θα προβάλλουμε σε ξεχωριστά διαγράμματα τις κατηγορίες της μεταβλητής Χώρα και των ασθενειών. Αρχικά, θα παρουσιάσουμε ξεχωριστά τον 1^ο παραγοντικό άξονα.

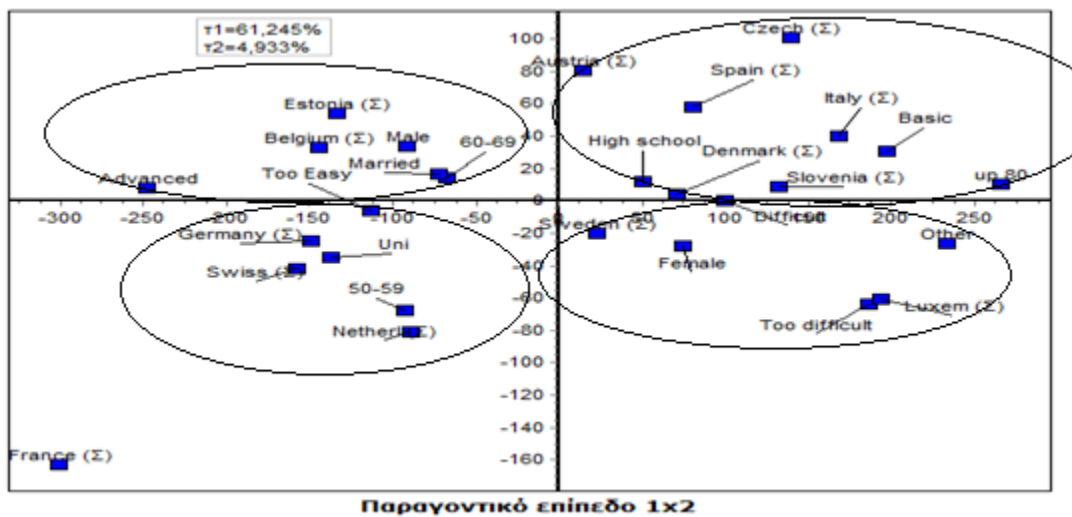


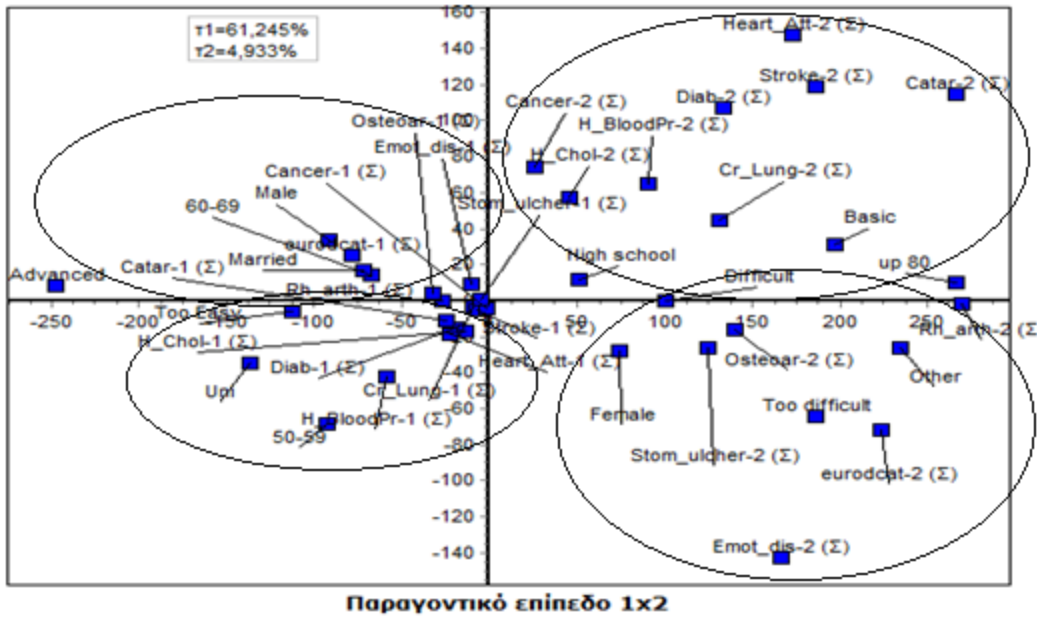
1ος Παραγοντικός άξονας



1ος Παραγοντικός άξονας

Συγκρίνοντας τα δύο παραπάνω διαγράμματα ο 1^{ος} παραγοντικός άξονας χαρακτηρίζεται από δύο κύριες διακριτές ομάδες μεταβλητών που τοποθετούνται στα άκρα του. Στη θετική πλευρά τοποθετούνται οι γυναίκες που ζουν μόνες άνω των 80 ετών με πρωτοβάθμια και δευτεροβάθμια εκπαίδευση και με δύσκολη και πολύ δύσκολη οικονομική κατάσταση. Αυτή η ομάδα σχετίζεται με τις χώρες Αυστρία, Σουηδία, Δανία, Ιταλία, Λουξεμβούργο, Σλοβενία, Τσεχία και Ισπανία ενώ χαρακτηρίζεται από όλες σχεδόν τις ασθένειες και εμφανίζει και κατάθλιψη. Στην αρνητική πλευρά του άξονα τοποθετούνται οι παντρεμένοι άνδρες με πολύ καλή οικονομική κατάσταση και ανώτερες σπουδές ηλικίας 50-69 ετών. Αυτή η ομάδα σχετίζεται με τις χώρες Βέλγιο, Γερμανία, Ολλανδία, Γαλλία, Ελβετία και Εσθονία ενώ ταυτόχρονα σχετίζεται με την απουσία κατάθλιψης. Δεν μπορούμε να εξάγουμε ασφαλή συμπεράσματα για την απουσία των υπόλοιπων ασθενειών καθώς οι εν λόγω κατηγορίες τοποθετούνται πολύ κοντά στην αρχή των αξόνων που αφορά τη μέση κατάσταση. Τέλος, ο 1^{ος} παραγοντικός άξονας θεωρείται ιδιαίτερα σημαντικός καθώς ερμηνεύει ποσοστό αδράνειας ίσο με 61,245%. Συνεχίζουμε την εξαγωγή συμπερασμάτων από την συγκεκριμένη ανάλυση δίνοντας τα συμμετρικά biplot των δύο προηγούμενων διαχωρισμών:

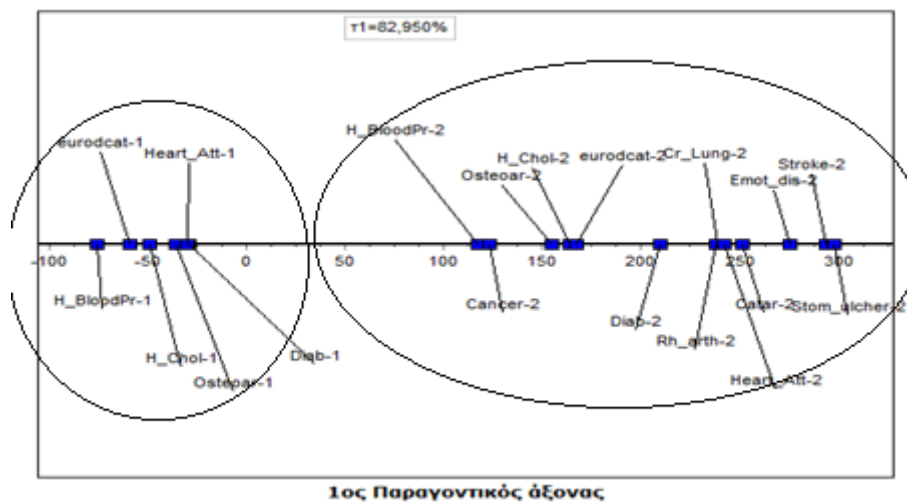




Εισάγοντας και τον 2^ο κύριο άξονα στην ανάλυση μπορούμε να διακρίνουμε και πάλι τέσσερις ομάδες. Η πρώτη ομάδα (πάνω αριστερά) περιλαμβάνει τους εύπορους παντρεμένους άνδρες ηλικίας 60-69 ετών και σε αυτήν προβάλλονται οι Βέλγοι και οι Εσθονοί καθώς και η απουσία κατάθλιψης, οστεοαρθρίτιδας και ρευματοειδούς αρθρίτιδας. Η δεύτερη ομάδα (κάτω αριστερά) αποτελείται από τις ηλικίες 50-59 ετών και το ανώτερο εκπαιδευτικό επίπεδο και εκεί προβάλλονται οι Γερμανοί, οι Ελβετοί και οι Ολλανδοί καθώς και η απουσία υψηλών τιμών σε χοληστερόλη, αρτηριακή πίεση και ζάχαρο. Ακόμη, οι Γάλλοι δεν συμμετέχουν σε καμία ομάδα. Η τρίτη ομάδα που παρατηρούμε συμπεριλαμβάνει υπερήλικες δύσκολου βιοτικού επιπέδου και χαμηλής εκπαίδευσης και εκεί προβάλλονται οι Ισπανοί, οι Ιταλοί, οι Τσέχοι, οι Δανοί και οι Σλοβένοι. Ακόμη, χαρακτηρίζεται από υψηλή χοληστερόλη και αρτηριακή πίεση, καρκίνο, διαβήτη, εγκεφαλικό, χρόνια πνευμονοπάθεια και καταρράκτη. Τέλος, η τέταρτη ομάδα που περιλαμβάνει τις γυναίκες που ζουν μόνες τους με πολύ δύσκολα οικονομικά σχετίζεται με το Λουξεμβούργο και χαρακτηρίζεται από έλκος, συναισθηματικές διαταραχές και κατάθλιψη. Οι ασθενείς με οστεοαρθρίτιδα και ρευματοειδή αρθρίτιδα φαίνεται να τοποθετούνται μεταξύ των δύο παραπάνω ομάδων. Οι τιμές της ερμηνευμένης αδράνειας καθώς και ορισμένοι δείκτες που προκύπτουν από την ανάλυση μπορούν να βρεθούν στο Παράρτημα (Π8).

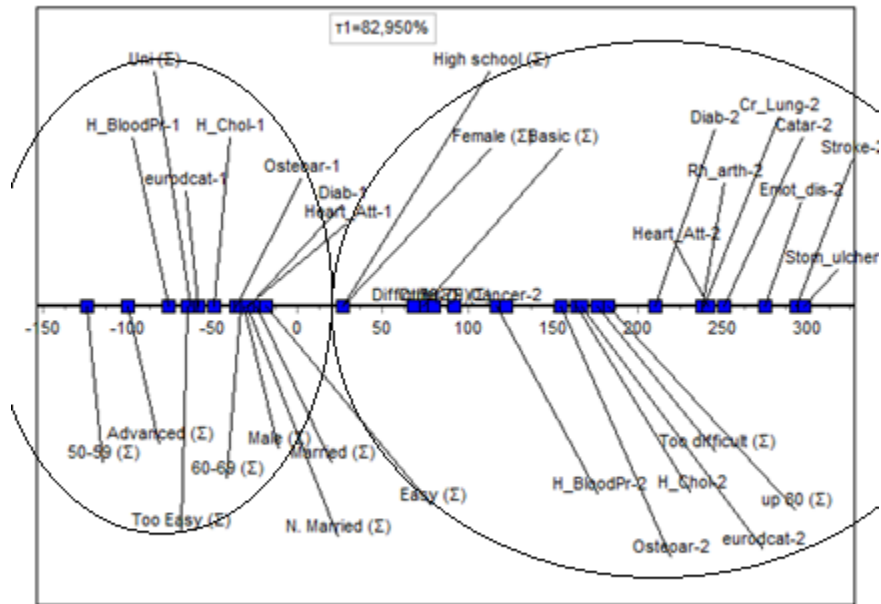
4.3 Πολυμεταβλητή Ανάλυση Αντιστοιχιών στις μεταβλητές των ασθενειών

Για το δεύτερο μέρος της ανάλυσης θα χρησιμοποιήσουμε ως βασικές μεταβλητές τις ασθένειες και θα προβάλλουμε στο παραγοντικό επίπεδο ως συμπληρωματικές τις δημογραφικές. Η μέθοδος που θα χρησιμοποιηθεί είναι η Πολυμεταβλητή Ανάλυση Αντιστοιχιών με τη διορθωμένη αδράνεια του Greenacre. Όπως και πριν θα προβάλλουμε σε δύο διαφορετικά διαγράμματα τις χώρες από τις υπόλοιπες δημογραφικές μεταβλητές λόγω του μεγάλου πλήθους των κατηγοριών που έχουν. Τα αποτελέσματα της ανάλυσης των βασικών μεταβλητών στον 1^ο παραγοντικό άξονα είναι:

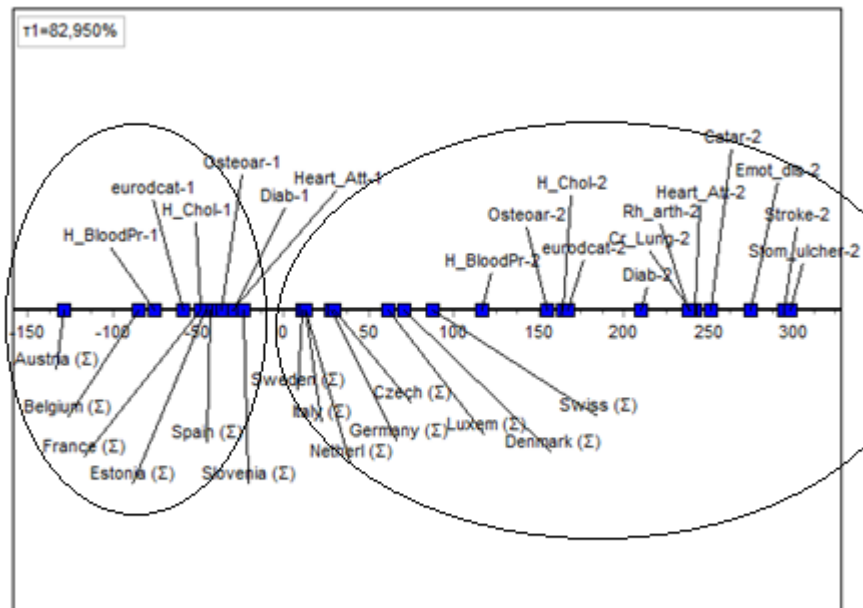


Ο 1^{ος} παραγοντικός άξονας είναι ιδιαίτερα σημαντικός καθώς ερμηνεύει αδράνεια 82,95%. Στο αρνητικό μέρος του έχουμε την απουσία υψηλής πίεσης, υψηλής χοληστερόλης, οστεοαρθρίτιδας, διαβήτη, καρδιακής προσβολής και κατάθλιψης. Αντίθετα, στο δεξιό μέρος έχουμε υψηλή πίεση, υψηλή χοληστερόλη, οστεοαρθρίτιδα, ρευματοειδή αρθρίτιδα, καρδιακή προσβολή, καρκίνο, διαβήτη, καταρράκτη, έλκος, εγκεφαλικό, συναισθηματικές διαταραχές και κατάθλιψη. Επομένως, μπορούμε να πούμε ότι ο 1^{ος} παραγοντικός άξονας αντικατοπτρίζει το επίπεδο της υγείας, το οποίο χειροτερεύει όσο τείνουμε από τα αριστερά προς τα δεξιά. Προβάλλοντας τα δημογραφικά χαρακτηριστικά βλέπουμε από το παρακάτω διάγραμμα ότι η καλή υγεία σχετίζεται με τους παντρεμένους άνδρες ηλικίας 50-69 ετών με ανώτερη και ανώτατη εκπαίδευση που

έχουν πολύ καλό βιοτικό επίπεδο. Αντιθέτως, η κακή υγεία σχετίζεται με τους υπερήλικες άνω των 80 ετών πρωτοβάθμιας ή δευτεροβάθμιας εκπαίδευσης με κακή και πολύ κακή οικονομική κατάσταση.



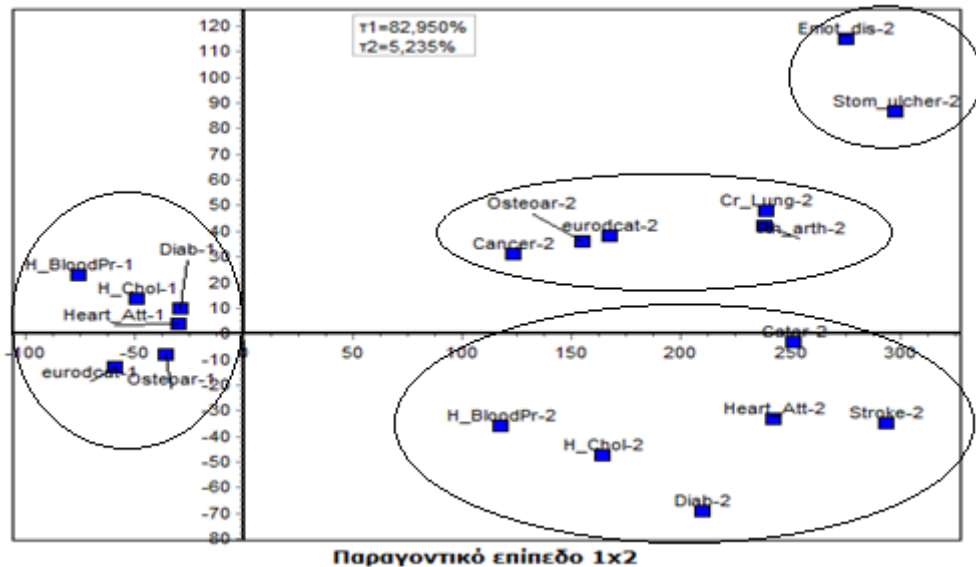
1ος Παραγοντικός άξονας

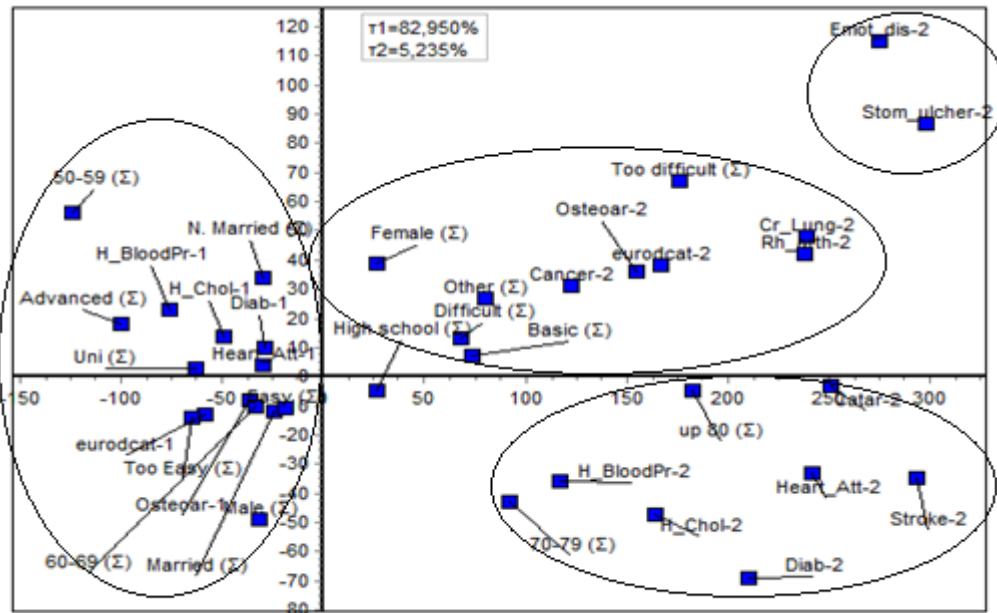


1ος Παραγοντικός άξονας

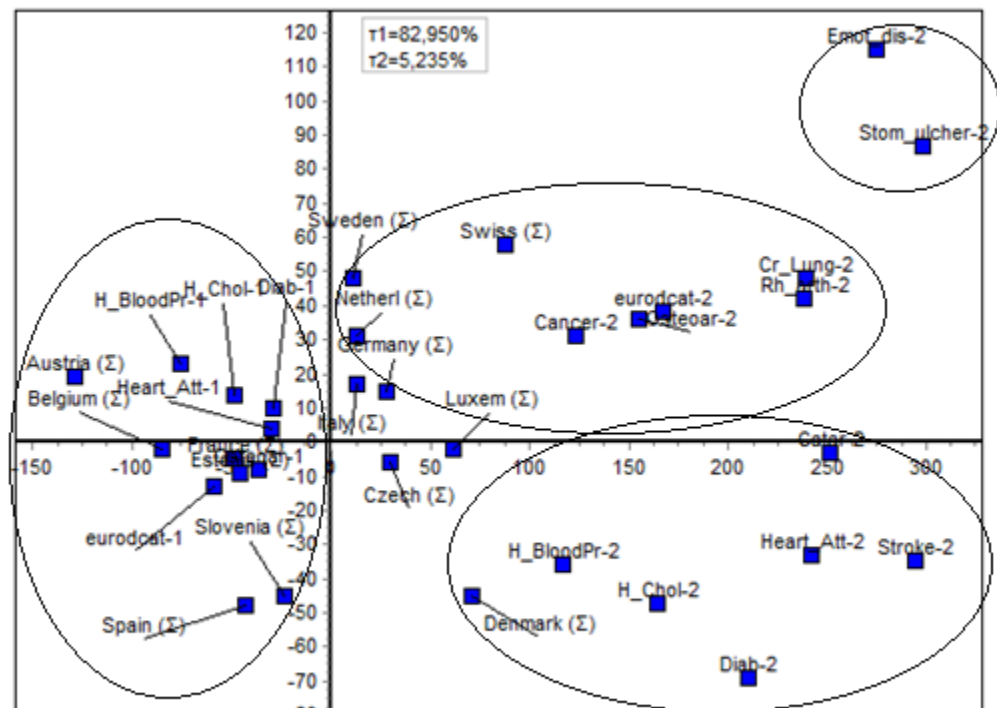
Στο παραπάνω διάγραμμα μπορούμε να δούμε την προβολή των χωρών στον 1^ο παραγοντικό άξονα. Στο αρνητικό μέρος του διαγράμματος που έχουμε τις κατηγορίες που αφορούν την καλή υγεία βλέπουμε τώρα ότι αυτές σχετίζονται με τις χώρες: Αυστρία, Βέλγιο, Γαλλία, Εσθονία, Ισπανία και Σλοβενία. Αντιθέτως, η κακή υγεία σχετίζεται κυρίως με την Ελβετία, τη Δανία και το Λουξεμβούργο και ίσως και με την Τσεχία, τη Γερμανία, την Ολλανδία, την Ιταλία και τη Σουηδία.

Στα επόμενα διαγράμματα εισάγουμε και τον 2^ο κύριο άξονα και δημιουργούμε το παραγοντικό επίπεδο. Σε αυτό προβάλλουμε έπειτα τις δημογραφικές μεταβλητές και τέλος τις χώρες.





Παραγοντικό επίπεδο 1x2



Παραγοντικό επίπεδο 1x2

Το ποσοστό αδράνειας που ερμηνεύεται με τους 2 πρώτους κύριους άξονες έχει ανέβει πλέον στο 88,185% και μπορούμε να διακρίνουμε τέσσερις βασικές ομάδες κατηγοριών. Η πρώτη ομάδα βρίσκεται στο αρνητικό μέρος του 1^{ου} κύριου άξονα και χαρακτηρίζεται από χαμηλή πίεση, χαμηλή χοληστερόλη και απουσία καρδιακής προσβολής, οστεοαρθρίτιδας και κατάθλιψης. Στην ομάδα αυτή προβάλλονται οι παντρεμένοι και μη άνδρες 50-69 ετών υψηλού εκπαιδευτικού και βιοτικού επιπέδου. Οι χώρες της ομάδας είναι η Αυστρία, το Βέλγιο, η Γαλλία και η Εσθονία. Οριακά θα μπορούσαμε να συμπεριλάβουμε τη Σλοβενία και την Ισπανία. Η δεύτερη ομάδα βρίσκεται στο 4^ο τεταρτημόριο και χαρακτηρίζεται από υψηλή πίεση και χοληστερόλη και παρουσία καρδιακής προσβολής, διαβήτη και καταρράκτη. Αυτή η ομάδα περιλαμβάνει κυρίως ανθρώπους άνω των 70 ετών με βασική χώρα τη Δανία. Η τρίτη ομάδα που δημιουργείται και βρίσκεται κοντά στην αρχή των αξόνων στο 1^ο τεταρτημόριο χαρακτηρίζεται από παρουσία καρκίνου, οστεοαρθρίτιδας, ρευματοπάθειας, χρόνιας πνευμονοπάθειας και κατάθλιψης ενώ προβάλλονται σε αυτήν γυναίκες που ζουν μόνες πρωτοβάθμιας και δευτεροβάθμιας εκπαίδευσης με δύσκολη και πολύ δύσκολη οικονομική κατάσταση. Η κύρια χώρα στην ομάδα αυτή είναι η Ελβετία. Τέλος η τέταρτη ομάδα χαρακτηρίζεται από έλκος και συναισθηματική διαταραχή. Δεν υπάρχει κάποιο δημογραφικό χαρακτηριστικό που να συσχετίζεται με αυτή την ομάδα. Με την εισαγωγή του 2^{ου} παραγοντικού άξονα μπορούμε να πούμε ότι διαχωρίζονται στο θετικό του κομμάτι οι ηλικίες 50-59 που ζουν με τους συζύγους τους και στο αρνητικό κομμάτι οι μεγαλύτερες ηλικίες που ζουν μόνοι τους. Ακόμη, ενώ το φύλο και το επίπεδο εκπαίδευσης διακρίνονται ξεκάθαρα στον 1^ο παραγοντικό άξονα, δεν συμβαίνει το ίδιο και με τον 2^ο. Οι τιμές της ερμηνευμένης αδράνειας καθώς και ορισμένοι δείκτες που προκύπτουν από την ανάλυση μπορούν να βρεθούν στο Παράρτημα (Π9).

ΚΕΦΑΛΑΙΟ 5

Συμπεράσματα

Συνοψίζοντας όλα τα αποτελέσματα των αναλύσεων του Κεφαλαίου 4 εξάγονται τα εξής γενικά συμπεράσματα:

- Οι εύποροι άνδρες, μέσης ηλικίας 50-69 ετών με υψηλό μορφωτικό επίπεδο κυρίως παντρεμένοι εμφανίζουν χαμηλά ποσοστά πίεσης, χοληστερόλης, καρδιακής προσβολής, οστεοαρθρίτιδας και κατάθλιψης. Η απουσία των συμπτωμάτων αυτών μάλλον οφείλεται στον υγιεινό τρόπο ζωής που διασφαλίζεται από την καλή οικονομική τους κατάσταση. Οι περισσότεροι εκπρόσωποι της ομάδας αυτής τοποθετούνται στις λεγόμενες πλούσιες χώρες (Γερμανία, Γαλλία, Βέλγιο, Ελβετία, Ολλανδία, Ισπανία, Αυστρία, Σλοβενία) καθώς και στην Εσθονία.
- Τα άτομα μεγάλης ηλικίας (άνω των 70 ετών), κυρίως γυναίκες, μόνες με χαμηλά εισοδήματα και χαμηλό μορφωτικό επίπεδο εμφανίζουν πληθώρα ασθενειών καθώς και κατάθλιψη. Η ομάδα αυτή περιλαμβάνει κυρίως τις χώρες Δανία, Σουηδία, Λουξεμβούργο, Τσεχία, Αυστρία, Ιταλία, Σλοβενία και Ισπανία.
- Τα άτομα μεγάλης ηλικίας (άνω των 70 ετών) εμφανίζουν ασθένειες της τρίτης ηλικίας: υψηλή χοληστερόλη, υψηλή αρτηριακή πίεση, καταρράκτη, διαβήτη, οστεοαρθρίτιδα και ρευματοειδή αρθρίτιδα. Σε περίπτωση δύσκολου βιοτικού επιπέδου και χαμηλής εκπαίδευσης μπορούν ακόμη να εμφανίσουν εγκεφαλικό, καρκίνο, χρόνια πνευμονοπάθεια και κατάθλιψη. Η ομάδα αυτή περιλαμβάνει κυρίως Ισπανούς, Ιταλούς, Τσέχους, Δανούς και Σλοβένους.
- Οι γυναίκες του Λουξεμβούργου που ζουν μόνες με δύσκολο βιοτικό επίπεδο εμφανίζουν έλκος, συναισθηματικές διαταραχές και κατάθλιψη. Σε αυτή την κατηγορία ανήκουν και οι αντίστοιχες γυναίκες της Ελβετίας οι οποίες μπορούν να εμφανίσουν επιπλέον καρκίνο και οστεοαρθρίτιδα.

ΑΝΑΦΟΡΕΣ

Αναφορές ως προς τη χρήση των δεδομένων

This paper uses data from SHARE Wave 5 (DOIs: [10.6103/SHARE.w5.600](https://doi.org/10.6103/SHARE.w5.600))

The SHARE data collection has been primarily funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812) and FP7 (SHARE-PREP: N°211909, SHARE-LEAP: N°227822, SHARE M4: N°261982). Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

- Börsch-Supan, A. (2017). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5*. Release version: 6.0.0. SHARE-ERIC. Data set. DOI: [10.6103/SHARE.w5.600](https://doi.org/10.6103/SHARE.w5.600).
- Forschungsdatenzentrum der Rentenversicherung, Max-Planck-Institut für Sozialrecht und Sozialpolitik (2017). *SHARE-RV*. Release version: 6.0.0. SHARE-ERIC. Data set. DOI: [10.6103/SHARE.SHARE-RV.600](https://doi.org/10.6103/SHARE.SHARE-RV.600).
- Börsch-Supan, A., M. Brandt, C. Hunkler, T. Kneip, J. Korbmacher, F. Malter, B. Schaan, S. Stuck, S. Zuber (2013). *Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE)*. International Journal of Epidemiology. DOI: [10.1093/ije/dyt088](https://doi.org/10.1093/ije/dyt088).
- Malter, F. and A. Börsch-Supan (Eds.) (2015). *SHARE Wave 5: Innovations & Methodology*. Munich: MEA, Max Planck Institute for Social Law and Social Policy.
- Hausstein, B., Schleinstein, N., Koch, U., Meichsner, J., Becker, K. & Stahn, L.-L. (2014). *da/ra Metadata Schema*. Version: 3.0. GESIS Leibniz Institute for the Social Sciences. DOI: [10.4232/10.mdsdoc.3.0](https://doi.org/10.4232/10.mdsdoc.3.0).
- DataCite Metadata Working Group (2013). *DataCite Metadata Schema for the Publication and Citation of Research Data*. Version 3.1. DataCite. Text. DOI: [10.5438/00010](https://doi.org/10.5438/00010).

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνικές Βιβλιογραφικές Αναφορές

- A1. Ηλίας Π. Αθανασιάδης (1995). *Παραγοντική Ανάλυση Αντιστοιχιών και Ιεραρχική Ταξινόμηση*. Αθήνα: Εκδόσεις Νέων Τεχνολογιών.
- A2. Θεόδωρος Ε. Μπεχράκης (1999). *Πολυδιάστατη Ανάλυση Δεδομένων: Μέθοδοι και Εφαρμογές*. Αθήνα: Εκδοτικός Οργανισμός Λιβάνη.
- A3. Δημήτρης Καρλής (2005). *Πολυμεταβλητή Στατιστική Ανάλυση*. Αθήνα: Εκδόσεις Σταμούλη Α.Ε..
- A4. Άγγελος Ι. Μάρκος (2006). *Βοήθεια στην Ερμηνεία των Αποτελεσμάτων της Παραγοντικής Ανάλυσης των Αντιστοιχιών & Αλγόριθμοι Κατασκευής και Ανάλυσης Ειδικών Πινάκων Εισόδου: Η Περίπτωση του Λογισμικού CHIC Analysis*. Διδακτορική Διατριβή, Τμήμα Εφαρμοσμένης Πληροφορικής, Πανεπιστήμιο Μακεδονίας.
- A5. Βασίλειος Ι. Καδδίτης (2008). *Ανάλυση Ερωτηματολογίου μέσω Ανάλυσης Αντιστοιχιών και Μοντέλων Συνάφειας*. Διπλωματική Εργασία, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Μεταπτυχιακό πρόγραμμα σπουδών στην Εφαρμοσμένη Στατιστική, Πανεπιστήμιο Πειραιώς.
- A6. Άγγελος Γ. Γεωργιτσόπουλος (2013). *Απλή & Πολλαπλή Ανάλυση Αντιστοιχιών (Simple & Multiple Correspondence Analysis) και πρακτική εφαρμογή τους σε Έρευνα Αγοράς (Market Research) με τη χρήση στατιστικών πακέτων SPSS & MINITAB*. Διπλωματική Εργασία, Τμήμα Μαθηματικών, διατμηματικό ΠΜΣ Μαθηματικά των Υπολογιστών & των Αποφάσεων, Πανεπιστήμιο Πατρών.

Ξένες Βιβλιογραφικές Αναφορές

- B1. Michael J. Greenacre (1984). *Theory and Applications of Correspondence Analysis*. London: Academic Press.
- B2. Lebart L., Morineau A. & Warwick K. (1984). *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices*. New York: John Wiley & Sons.
- B3. Sten-Erik Clausen (1998). *Applied Correspondence Analysis: An introduction*. Sage Publications Inc..
- B4. Michael J. Greenacre (2007). *Correspondence Analysis in Practice: Second Edition*. Taylor & Francis Group, LLC
- B5. Phillip M Yelland (2010). *An Introduction to Correspondence Analysis*. The Mathematica Journal 12 © 2010 Wolfram Media, Inc.

- B6. Benzécri, J. P (1979). Sur les calcul des taux d' inertie dans l' analyse d' un questionnaire. *Les Cahiers de l' Analyse des Données* **4** 377-378.
- B7. Greenacre, M. J. (1988). Correspondence analysis of multivariate categorical data by weighted least squares. *Biometrika* **75** 457-467.
- B8. A. Gifi (1990). *Nonlinear multivariate analysis*. John Wiley & Sons Ltd., New York.
- B9. C. Hayashi (1950). On the Quantification of Qualitative Data from the Mathematico-statistical Point of View, *Annals of the Institute of Statist. Math.* **2** 35-47.
- B10. P. Horst (1935). Measuring complex attitudes, *Journal of Social Psychology* **6** 369-374.
- B11. R. A. Fisher (1940). The precision of discriminant functions. *Annals of Eugenics* **10** 422-29.
- B12. L. Guttman (1941). The quantification of a class of attributes: A theory and method of scale construction. In the Committee on Social Adjustment, *The prediction of personal adjustment* 319-348. New York: Social Science Research Council.
- B13. C. Burt (1950). The factorial analysis of qualitative data. *British Journal of Psychology* **3** 166-185.
- B14. Camiz S., Gomes G.C. (2013) *Joint Correspondence Analysis Versus Multiple Correspondence Analysis: A Solution to an Undetected Problem*. In: Giusti A., Ritter G., Vichi M. (eds) *Classification and Data Mining. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, Berlin, Heidelberg.

ΠΑΡΑΡΤΗΜΑ

Π1. Αναλυτικός πίνακας συχνοτήτων ανά ηλικία

Age	Frequency	Percent	Cumulative Percent	Age	Frequency	Percent	Cumulative Percent
50	313	0,52%	0,52%	78	1266	2,10%	85,80%
51	1193	1,98%	2,50%	79	1166	1,94%	87,74%
52	1415	2,35%	4,86%	80	1044	1,74%	89,47%
53	1794	2,98%	7,84%	81	1011	1,68%	91,15%
54	1757	2,92%	10,76%	82	896	1,49%	92,64%
55	1744	2,90%	13,66%	83	828	1,38%	94,02%
56	1880	3,13%	16,78%	84	679	1,13%	95,15%
57	1964	3,27%	20,05%	85	640	1,06%	96,21%
58	2065	3,43%	23,48%	86	510	0,85%	97,06%
59	2146	3,57%	27,05%	87	409	0,68%	97,74%
60	2143	3,56%	30,61%	88	379	0,63%	98,37%
61	2134	3,55%	34,16%	89	275	0,46%	98,83%
62	2129	3,54%	37,70%	90	205	0,34%	99,17%
63	2236	3,72%	41,42%	91	167	0,28%	99,45%
64	2170	3,61%	45,02%	92	123	0,20%	99,65%
65	2300	3,82%	48,85%	93	80	0,13%	99,78%
66	2221	3,69%	52,54%	94	48	0,08%	99,86%
67	2199	3,66%	56,20%	95	27	0,04%	99,91%
68	1868	3,11%	59,30%	96	14	0,02%	99,93%
69	1951	3,24%	62,54%	97	16	0,03%	99,96%
70	1860	3,09%	65,64%	98	6	0,01%	99,97%
71	1721	2,86%	68,50%	99	5	0,01%	99,98%
72	1642	2,73%	71,23%	100	3	0,00%	99,98%
73	1788	2,97%	74,20%	101	6	0,01%	99,99%
74	1561	2,60%	76,79%	102	3	0,00%	100,00%
75	1473	2,45%	79,24%	103	1	0,00%	100,00%
76	1363	2,27%	81,51%	104	1	0,00%	100,00%
77	1315	2,19%	83,69%	Total	60153	100,00%	

Π2. Αναλυτικός πίνακας συχνοτήτων ανά έτη εκπαίδευσης

Years of education	Frequency	Percent	Cumulative Percent
0	494	0,82%	0,82%
1	325	0,54%	1,36%
2	663	1,10%	2,46%
3	1624	2,70%	5,16%
4	1228	2,04%	7,20%
5	2204	3,66%	10,87%
6	1760	2,93%	13,79%
7	2432	4,04%	17,84%
8	6367	10,58%	28,42%
9	3543	5,89%	34,31%
10	4451	7,40%	41,71%
10,5	44	0,07%	41,79%
11	6081	10,11%	51,89%
12	7697	12,80%	64,69%
13	5009	8,33%	73,02%
14	3862	6,42%	79,44%
15	3155	5,24%	84,68%
16	3064	5,09%	89,78%
17	1876	3,12%	92,89%
17,5	185	0,31%	93,20%
18	1627	2,70%	95,91%
19	797	1,32%	97,23%
20	747	1,24%	98,47%
20,5	3	0,00%	98,48%
21	294	0,49%	98,97%
22	225	0,37%	99,34%
23	147	0,24%	99,59%
24	83	0,14%	99,72%
25	166	0,28%	100,00%
Total	60153	100,00%	

Π3. Συγκεντρωτικός πίνακας συχνότητας ασθενειών

Disease	Category	Frequency	Percent	Cumulative Percent
Heart Attack	Not selected	53418	88,80%	88,80%
	Selected	6735	11,20%	100,00%
	Total	60153	100,00%	
High blood pressure or hypertension	Not selected	36474	60,64%	60,64%
	Selected	23679	39,36%	100,00%
	Total	60153	100,00%	
High blood cholesterol	Not selected	46289	76,95%	76,95%
	Selected	13864	23,05%	100,00%
	Total	60153	100,00%	
Stroke	Not selected	57932	96,31%	96,31%
	Selected	2221	3,69%	100,00%
	Total	60153	100,00%	
Diabetes or high blood sugar	Not selected	52792	87,76%	87,76%
	Selected	7361	12,24%	100,00%
	Total	60153	100,00%	
Chronic lung disease	Not selected	56428	93,81%	93,81%
	Selected	3725	6,19%	100,00%
	Total	60153	100,00%	
Cancer	Not selected	56746	94,34%	94,34%
	Selected	3407	5,66%	100,00%
	Total	60153	100,00%	
Stomach or duodenal ulcer, peptic ulcer	Not selected	57765	96,03%	96,03%
	Selected	2388	3,97%	100,00%
	Total	60153	100,00%	
Cataracts	Not selected	55040	91,50%	91,50%
	Selected	5113	8,50%	100,00%
	Total	60153	100,00%	
Other affective/emotional disorders	Not selected	56756	94,35%	94,35%
	Selected	3397	5,65%	100,00%
	Total	60153	100,00%	
Rheumatoid arthritis	Not selected	54729	90,98%	90,98%
	Selected	5424	9,02%	100,00%
	Total	60153	100,00%	
Osteoarthritis/other rheumatism	Not selected	48841	81,19%	81,19%
	Selected	11312	18,81%	100,00%
	Total	60153	100,00%	

Π4. Αδράνεια και δείκτες που προκύπτουν από την 1^η ανάλυση των δημογραφικών μεταβλητών

Αξονας	Αδράνεια	%Ερμηνείας	Αθρ. %Ερμην.	Ραβδόγραμμα Χαρ. Ριζών
1	0,095	15,225	15,225	*****
2	0,056	9,035	24,261	*****
3	0,048	7,664	31,925	****
4	0,047	7,587	39,512	****
5	0,042	6,682	46,194	***
6	0,041	6,487	52,682	***
7	0,04	6,421	59,103	***
8	0,04	6,376	65,478	***
9	0,039	6,309	71,787	***
10	0,039	6,215	78,002	***
11	0,037	5,863	83,866	***
12	0,033	5,29	89,156	***
13	0,032	5,096	94,251	***
14	0,023	3,63	97,881	**
15	0,013	2,119	100	*

Κατηγορία	Ποιότητα	Βάρος	Αδράνεια	F1	COR1	CTR1	F2	COR2	CTR2
50-59	0,445	0,054	0,031	-0,332	0,192	0,063	0,381	0,253	0,139
60-69	0,106	0,071	0,026	-0,192	0,099	0,027	-0,05	0,007	0,003
70-79	0,156	0,05	0,031	0,22	0,079	0,026	-0,215	0,076	0,041
up 80	0,48	0,025	0,039	0,836	0,439	0,18	-0,254	0,04	0,028
Basic	0,206	0,028	0,036	0,516	0,203	0,077	-0,062	0,003	0,002
High school	0,082	0,102	0,02	0,125	0,08	0,017	0,02	0,002	0,001
Uni	0,278	0,062	0,029	-0,357	0,273	0,084	0,045	0,004	0,002
Advanced	0,1	0,008	0,039	-0,571	0,068	0,028	-0,388	0,032	0,022
Too difficult	0,262	0,016	0,038	0,308	0,042	0,016	0,709	0,22	0,147
Difficult	0,126	0,047	0,031	0,195	0,058	0,019	0,212	0,068	0,037
Easy	0,033	0,06	0,028	0,051	0,005	0,002	-0,115	0,028	0,014
Too Easy	0,263	0,077	0,026	-0,225	0,151	0,041	-0,193	0,111	0,051
Married with Partner	0,488 0,021	0,139 0,003	0,013 0,04	-0,148 -0,39	0,232 0,012	0,032 0,005	-0,156 0,352	0,256 0,01	0,059 0,007
Marr. without N. Married	0,025 0,078	0,002 0,011	0,04 0,038	-0,138 -0,217	0,001 0,014	0 0,006	0,617 0,468	0,024 0,064	0,017 0,043
Divorced Widowed	0,379 0,682	0,018 0,027	0,037 0,041	-0,162 1,011	0,012 0,679	0,005 0,291	0,878 -0,066	0,367 0,003	0,242 0,002
Male Female	0,375 0,375	0,089 0,111	0,023 0,019	-0,22 0,177	0,184 0,184	0,045 0,036	-0,224 0,18	0,191 0,191	0,079 0,063

Π5. Αδράνεια και δείκτες που προκύπτουν από την 2^η ανάλυση των δημογραφικών μεταβλητών

Αξονας	Διορθ/νη Αδράνεια	%Ερμηνείας	Αθρ. %Ερμην.	Ραβδόγραμμα Χαρ. Ριζών
1	0,018	60,161	60,161	*****
2	0,002	7,218	67,378	****
3	0,001	1,803	69,181	*
4	0	1,598	70,779	*
5	0	0,093	70,872	
6	0	0,008	70,881	
7	0	0	70,881	

Κατηγορία	Ποιότητα	Βάρος	Αδράνεια	F1	COR1	CTR1	F2	COR2	CTR2
Male	0,046	0,089	0,023	-0,097	0,036	0,045	0,044	0,007	0,079
Female	0,046	0,111	0,019	0,078	0,036	0,036	-0,035	0,007	0,063
50-59	0,048	0,054	0,031	-0,146	0,037	0,063	-0,075	0,01	0,139
60-69	0,02	0,071	0,026	-0,084	0,019	0,027	0,01	0	0,003
70-79	0,02	0,05	0,031	0,096	0,015	0,026	0,042	0,003	0,041
up 80	0,088	0,025	0,039	0,367	0,085	0,18	0,05	0,002	0,028
Basic	0,043	0,028	0,036	0,226	0,039	0,077	0,012	0	0,002
High school	0,018	0,102	0,02	0,055	0,015	0,017	-0,004	0	0,001
Uni	0,055	0,062	0,029	-0,157	0,053	0,084	-0,009	0	0,002
Advanced	0,016	0,008	0,039	-0,251	0,013	0,028	0,077	0,001	0,022
Too difficult	0,019	0,016	0,038	0,135	0,008	0,016	-0,14	0,009	0,147
Difficult	0,016	0,047	0,031	0,086	0,011	0,019	-0,042	0,003	0,037
Easy	0,003	0,06	0,028	0,022	0,001	0,002	0,023	0,001	0,014
Too Easy	0,037	0,077	0,026	-0,099	0,029	0,041	0,038	0,004	0,051
Married with Partner	0,055	0,139	0,013	-0,065	0,045	0,032	0,031	0,01	0,059
Marr. without N. Married	0,003	0,003	0,04	-0,171	0,002	0,005	-0,07	0	0,007
Divorced	0,002	0,002	0,04	-0,061	0	0	-0,122	0,001	0,017
Widowed	0,007	0,011	0,038	-0,095	0,003	0,006	-0,092	0,003	0,043
	0,017	0,018	0,037	-0,071	0,002	0,005	-0,173	0,014	0,242
	0,132	0,027	0,041	0,444	0,131	0,291	0,013	0	0,002

Π6. Αδράνεια και δείκτες που προκύπτουν από την 3^η ανάλυση των δημογραφικών μεταβλητών

Αξονας	Διορθ/νη Αδράνεια	%Ερμηνείας	Αθρ. %Ερμην.	Ραβδόγραμμα Χαρ. Ριζών
1	0,013	61,245	61,245	*****
2	0,001	4,933	66,178	**
3	0	2,359	68,537	*
4	0	1,145	69,682	*
5	0	0,047	69,729	
6	0	0,001	69,73	

Κατηγορία	Ποιότητα	Βάρος	Αδράνεια	F1	COR1	CTR1	F2	COR2	CTR2
Male	0,039	0,089	0,023	-0,092	0,033	0,061	0,034	0,005	0,105
Female	0,039	0,111	0,019	0,074	0,033	0,049	-0,028	0,005	0,084
50-59	0,024	0,054	0,03	-0,093	0,016	0,038	-0,068	0,008	0,246
60-69	0,013	0,071	0,026	-0,067	0,012	0,025	0,014	0,001	0,014
70-79	0,011	0,05	0,03	0,065	0,007	0,017	0,048	0,004	0,117
up 80	0,048	0,025	0,037	0,266	0,046	0,138	0,01	0	0,002
Basic	0,034	0,028	0,036	0,197	0,03	0,085	0,031	0,001	0,027
High school	0,016	0,102	0,02	0,051	0,013	0,021	0,012	0,001	0,015
Uni	0,044	0,062	0,029	-0,137	0,041	0,094	-0,035	0,003	0,074
Advanced	0,015	0,008	0,039	-0,248	0,013	0,04	0,008	0	0,001
Too difficult	0,019	0,016	0,038	0,186	0,015	0,046	-0,064	0,002	0,068
Difficult	0,016	0,047	0,031	0,1	0,015	0,038	0	0	0
Easy	0,002	0,06	0,028	0,014	0	0,001	0,026	0,001	0,04
Too Easy	0,04	0,077	0,026	-0,113	0,038	0,078	-0,006	0	0,003
Married	0,062	0,142	0,013	-0,072	0,059	0,058	0,017	0,003	0,042
N. Married	0,008	0,011	0,038	-0,077	0,002	0,005	-0,109	0,003	0,131
Other	0,079	0,047	0,034	0,234	0,077	0,207	-0,026	0,001	0,033

Π7. Αδράνεια και δείκτες που προκύπτουν από την 4^η ανάλυση των δημογραφικών μεταβλητών

Αξονας	Διορθ/νη Αδράνεια	%Ερμηνείας	Αθρ. %Ερμην.	Ραβδόγραμμα Χαρ. Ριζών
1	0,018	40,324	40,324	*****
2	0,006	13,596	53,919	*****
3	0,003	6,017	59,937	***
4	0,002	3,971	63,908	**
5	0,001	1,146	65,054	*
6	0	0,671	65,725	
7	0	0,359	66,084	
8	0	0,195	66,279	
9	0	0,112	66,391	
10	0	0,064	66,455	
11	0	0,001	66,456	
12	0	0	66,456	
13	0	0	66,456	

Κατηγορία	Ποιότητα	Βάρος	Αδράνεια	F1	COR1	CTR1	F2	COR2	CTR2
Male	0,041	0,074	0,016	-0,062	0,018	0,016	0,055	0,014	0,037
Female	0,041	0,092	0,013	0,05	0,018	0,013	-0,044	0,014	0,03
50-59	0,035	0,045	0,021	-0,087	0,016	0,019	0,01	0	0,001
60-69	0,015	0,059	0,018	-0,056	0,01	0,01	0,002	0	0
70-79	0,015	0,042	0,021	0,071	0,01	0,012	-0,005	0	0
up 80	0,051	0,02	0,026	0,206	0,033	0,048	-0,017	0	0,001
Basic	0,136	0,023	0,029	0,321	0,082	0,132	0,255	0,052	0,247
High school	0,05	0,085	0,014	0,038	0,008	0,007	-0,074	0,032	0,078
Uni	0,084	0,052	0,021	-0,172	0,073	0,086	-0,006	0	0
Advanced	0,021	0,007	0,027	-0,241	0,015	0,022	0,115	0,003	0,015
Too difficult	0,052	0,014	0,027	0,29	0,042	0,064	0,019	0	0,001
Difficult	0,066	0,039	0,023	0,174	0,051	0,066	-0,053	0,005	0,018
Easy	0,014	0,05	0,02	0,035	0,003	0,003	-0,004	0	0
Too Easy	0,134	0,064	0,02	-0,196	0,12	0,137	0,032	0,003	0,011
Married	0,069	0,118	0,009	-0,05	0,034	0,017	0,037	0,019	0,027
N. Married	0,009	0,009	0,027	-0,033	0	0,001	0,038	0,001	0,002
Other	0,084	0,039	0,024	0,159	0,042	0,055	-0,121	0,024	0,096
Spain	0,012	0,011	0,026	0,059	0,001	0,002	0,048	0,001	0,004
Germany	0,017	0,015	0,026	-0,151	0,014	0,019	-0,022	0	0,001
Estonia	0,023	0,012	0,027	-0,178	0,015	0,022	-0,015	0	0
Belgium	0,015	0,011	0,026	-0,167	0,012	0,017	-0,016	0	0
Czech	0,047	0,017	0,026	0,206	0,028	0,041	0,161	0,017	0,073
Italy	0,061	0,013	0,028	0,272	0,033	0,052	0,204	0,019	0,087
Sweden	0,007	0,012	0,026	-0,013	0	0	-0,101	0,005	0,02
France	0,06	0,011	0,028	-0,357	0,05	0,078	0,004	0	0
Austria	0,028	0,008	0,028	0,03	0	0	0,261	0,02	0,092
Netherl	0,011	0,015	0,026	-0,116	0,008	0,011	-0,038	0,001	0,004
Denmark	0,023	0,015	0,026	0,057	0,002	0,003	-0,17	0,016	0,07
Swiss	0,005	0,004	0,027	-0,165	0,004	0,006	0,042	0	0,001
Slovenia	0,016	0,008	0,027	0,157	0,007	0,011	-0,087	0,002	0,01
Luxem	0,042	0,015	0,027	0,193	0,021	0,031	-0,172	0,017	0,073

Π8. Αδράνεια και δείκτες που προκύπτουν από την τελική ανάλυση των δημογραφικών μεταβλητών

Αξονας	Διορθ/νη Αδράνεια	%Ερμηνείας	Αθρ. %Ερμην.	Ραβδόγραμμα Χαρ. Ριζών
1	0,013	61,245	61,245	*****
2	0,001	4,933	66,178	**
3	0	2,359	68,537	*
4	0	1,145	69,682	*
5	0	0,047	69,729	
6	0	0,001	69,73	

Κατηγορία	Ποιότητα	Βάρος	Αδράνεια	F1	COR1	CTR1	F2	COR2	CTR2
Male	0,039	0,089	0,023	-0,092	0,033	0,061	0,034	0,005	0,105
Female	0,039	0,111	0,019	0,074	0,033	0,049	-0,028	0,005	0,084
50-59	0,024	0,054	0,03	-0,093	0,016	0,038	-0,068	0,008	0,246
60-69	0,013	0,071	0,026	-0,067	0,012	0,025	0,014	0,001	0,014
70-79	0,011	0,05	0,03	0,065	0,007	0,017	0,048	0,004	0,117
up 80	0,048	0,025	0,037	0,266	0,046	0,138	0,01	0	0,002
Basic	0,034	0,028	0,036	0,197	0,03	0,085	0,031	0,001	0,027
High school	0,016	0,102	0,02	0,051	0,013	0,021	0,012	0,001	0,015
Uni	0,044	0,062	0,029	-0,137	0,041	0,094	-0,035	0,003	0,074
Advanced	0,015	0,008	0,039	-0,248	0,013	0,04	0,008	0	0,001
Married	0,062	0,142	0,013	-0,072	0,059	0,058	0,017	0,003	0,042
N. Married	0,008	0,011	0,038	-0,077	0,002	0,005	-0,109	0,003	0,131
Other	0,079	0,047	0,034	0,234	0,077	0,207	-0,026	0,001	0,033
Too difficult	0,019	0,016	0,038	0,186	0,015	0,046	-0,064	0,002	0,068
Difficult	0,016	0,047	0,031	0,1	0,015	0,038	0	0	0
Easy	0,002	0,06	0,028	0,014	0	0,001	0,026	0,001	0,04
Too Easy	0,04	0,077	0,026	-0,113	0,038	0,078	-0,006	0	0,003

Συμπληρωματικές Μεταβλητές	Ποιότητα	Βάρος	Αδράνεια	F1	COR1	F2	COR2
Cr_Lung-1	0,822	0,188	0	-0,009	0,609	-0,003	0,071
Cr_Lung-2	0,822	0,012	0	0,131	0,609	0,045	0,071
Diab-1	0,847	0,176	0	-0,019	0,433	-0,015	0,28
Diab-2	0,847	0,024	0,001	0,133	0,433	0,107	0,28
Stroke-1	0,678	0,193	0	-0,007	0,443	-0,005	0,182
Stroke-2	0,678	0,007	0,001	0,186	0,443	0,119	0,182
H_Chol-1	0,736	0,154	0	-0,014	0,236	-0,017	0,374
H_Chol-2	0,736	0,046	0	0,046	0,236	0,057	0,374
H_BloodPr-1	0,799	0,121	0,001	-0,059	0,522	-0,042	0,265
H_BloodPr-2	0,799	0,079	0,001	0,091	0,522	0,065	0,265
Heart_Att-1	0,735	0,178	0	-0,022	0,366	-0,019	0,27
Heart_Att-2	0,735	0,022	0,002	0,173	0,366	0,148	0,27
Cancer-1	0,422	0,189	0	-0,002	0,036	-0,004	0,294
Cancer-2	0,422	0,011	0	0,026	0,036	0,074	0,294
Stom_ulcher-1	0,674	0,192	0	-0,005	0,523	0,001	0,022
Stom_ulcher-2	0,674	0,008	0	0,125	0,523	-0,026	0,022
Catar-1	0,757	0,183	0	-0,025	0,505	-0,011	0,094
Catar-2	0,757	0,017	0,002	0,266	0,505	0,115	0,094
Emot_dis-1	0,877	0,189	0	-0,01	0,462	0,009	0,345
Emot_dis-2	0,877	0,011	0,001	0,166	0,462	-0,143	0,345
Rh_arth-1	0,888	0,182	0	-0,027	0,858	0	0
Rh_arth-2	0,888	0,018	0,002	0,27	0,858	-0,002	0
Osteoar-1	0,813	0,162	0	-0,032	0,683	0,004	0,009
Osteoar-2	0,813	0,038	0,001	0,14	0,683	-0,016	0,009
eurodcatt-1	0,916	0,148	0,001	-0,079	0,789	0,025	0,082
eurodcatt-2	0,916	0,052	0,003	0,223	0,789	-0,072	0,082

Συμπληρωματικές Μεταβλητές	Ποιότητα	Βάρος	Αδράνεια	F1	COR1	F2	COR2
Spain	0,234	0,013	0,001	0,081	0,108	0,058	0,055
Germany	0,885	0,018	0,001	-0,149	0,613	-0,025	0,017
Estonia	0,589	0,015	0,001	-0,134	0,234	0,054	0,038
Belgium	0,723	0,013	0,001	-0,145	0,402	0,033	0,021
Czech	0,735	0,021	0,002	0,14	0,179	0,101	0,093
Italy	0,863	0,015	0,003	0,169	0,136	0,04	0,008
Sweden	0,751	0,014	0	0,023	0,02	-0,02	0,015
France	0,888	0,013	0,003	-0,301	0,43	-0,163	0,126
Austria	0,401	0,01	0,002	0,015	0,001	0,081	0,036
Netherl	0,873	0,018	0	-0,089	0,291	-0,081	0,243
Denmark	0,601	0,017	0,001	0,071	0,074	0,004	0
Swiss	0,693	0,005	0	-0,158	0,493	-0,042	0,035
Slovenia	0,661	0,009	0,001	0,132	0,199	0,009	0,001
Luxem	0,573	0,018	0,002	0,194	0,311	-0,061	0,031

Π9. Αδράνεια και δείκτες που προκύπτουν από την ανάλυση των μεταβλητών των ασθενειών

Άξονας	Διορθ/νη Αδράνεια	%Ερμηνείας	Αθρ. %Ερμην.	Ραβδόγραμμα Χαρ. Ριζών
1	0,006	82,95	82,95	*****
2	0	5,235	88,186	***
3	0	0,118	88,303	

Κατηγορία	Ποιότητα	Βάρος	Αδράνεια	F1	COR1	CTR1	F2	COR2	CTR2
H_BloodPr-1	0,111	0,047	0,003	-0,076	0,101	0,048	0,023	0,009	0,071
H_BloodPr-2	0,111	0,03	0,004	0,117	0,101	0,074	-0,036	0,009	0,109
H_Chol-1	0,101	0,059	0,002	-0,049	0,093	0,025	0,014	0,008	0,033
H_Chol-2	0,101	0,018	0,005	0,164	0,093	0,084	-0,047	0,008	0,11
Stroke-1	0,042	0,074	0	-0,011	0,041	0,002	0,001	0,001	0
Stroke-2	0,042	0,003	0,006	0,294	0,041	0,043	-0,035	0,001	0,01
Diab-1	0,081	0,068	0,001	-0,029	0,073	0,01	0,01	0,008	0,017
Diab-2	0,081	0,009	0,006	0,21	0,073	0,073	-0,069	0,008	0,124
Cr_Lung-1	0,049	0,072	0	-0,016	0,047	0,003	-0,003	0,002	0,002
Cr_Lung-2	0,049	0,005	0,006	0,239	0,047	0,048	0,048	0,002	0,031
Cancer-1	0,013	0,073	0	-0,007	0,012	0,001	-0,002	0,001	0,001
Cancer-2	0,013	0,004	0,006	0,123	0,012	0,012	0,031	0,001	0,012
Stom_ulcher-1	0,049	0,074	0	-0,012	0,045	0,002	-0,004	0,004	0,003
Stom_ulcher-2	0,049	0,003	0,006	0,298	0,045	0,048	0,087	0,004	0,065
Catar-1	0,071	0,07	0,001	-0,023	0,071	0,007	0	0	0
Catar-2	0,071	0,007	0,006	0,251	0,071	0,073	-0,003	0	0
Emot_dis-1	0,064	0,073	0	-0,016	0,055	0,003	-0,007	0,009	0,01
Emot_dis-2	0,064	0,004	0,006	0,275	0,055	0,058	0,115	0,009	0,16
Rh_arth-1	0,07	0,07	0,001	-0,024	0,068	0,007	-0,004	0,002	0,003
Rh_arth-2	0,07	0,007	0,006	0,238	0,068	0,07	0,042	0,002	0,034
Osteoar-1	0,071	0,062	0,001	-0,036	0,067	0,014	-0,008	0,004	0,012
Osteoar-2	0,071	0,014	0,005	0,155	0,067	0,062	0,036	0,004	0,052
eurodcacat-1	0,117	0,057	0,002	-0,059	0,111	0,035	-0,013	0,006	0,029
eurodcacat-2	0,117	0,02	0,005	0,167	0,111	0,099	0,038	0,006	0,082
Heart_Att-1	0,089	0,068	0,001	-0,03	0,087	0,011	0,004	0,002	0,003
Heart_Att-2	0,089	0,009	0,006	0,242	0,087	0,089	-0,033	0,002	0,026

Συμπληρωματικές Μεταβλητές	Ποιότητα	Βάρος	Αδράνεια	F1	COR1	F2	COR2
Male	0,738	0,034	0	-0,032	0,169	-0,049	0,394
Female	0,738	0,043	0	0,026	0,169	0,039	0,394
50-59	0,87	0,021	0	-0,124	0,714	0,056	0,143
60-69	0,593	0,027	0	-0,034	0,474	-0,01	0,043
70-79	0,888	0,019	0	0,092	0,709	-0,043	0,153
up 80	0,71	0,009	0	0,183	0,677	-0,005	0
Basic	0,739	0,011	0	0,074	0,666	0,007	0,007
High school	0,924	0,039	0	0,026	0,89	-0,005	0,032
Uni	0,89	0,024	0	-0,063	0,882	0,003	0,002
Advanced	0,893	0,003	0	-0,1	0,824	0,018	0,026
Too difficult	0,788	0,006	0	0,176	0,679	0,067	0,099
Difficult	0,781	0,018	0	0,068	0,734	0,013	0,027
Easy	0,84	0,023	0	-0,019	0,617	-0,011	0,223
Too Easy	0,745	0,03	0	-0,065	0,693	-0,014	0,031
Married	0,856	0,054	0	-0,024	0,69	-0,012	0,162
N. Married	0,644	0,004	0	-0,03	0,282	0,034	0,354
Other	0,864	0,018	0	0,08	0,771	0,027	0,09
Spain	0,377	0,005	0	-0,043	0,161	-0,048	0,203
Germany	0,439	0,007	0	0,028	0,166	0,015	0,046
Estonia	0,448	0,006	0	-0,046	0,156	-0,009	0,006
Belgium	0,694	0,005	0	-0,085	0,585	-0,002	0
Czech	0,195	0,008	0	0,03	0,064	-0,006	0,003
Italy	0,297	0,006	0	0,013	0,037	0,017	0,062
Sweden	0,116	0,006	0	0,011	0,006	0,048	0,104
France	0,233	0,005	0	-0,049	0,213	-0,005	0,002
Austria	0,88	0,004	0	-0,129	0,86	0,019	0,019
Netherl	0,32	0,007	0	0,013	0,025	0,031	0,148
Denmark	0,584	0,007	0	0,071	0,394	-0,045	0,163
Swiss	0,336	0,002	0	0,088	0,226	0,058	0,099
Slovenia	0,173	0,004	0	-0,023	0,034	-0,045	0,132
Luxem	0,229	0,007	0	0,062	0,227	-0,002	0