



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**ΜΕΘΟΔΟΙ ΠΟΛΥΜΕΤΑΒΛΗΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ
ΑΝΑΛΥΣΗΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ**

Ναταλία Δ. Μαύρου

ΑΘΗΝΑ, ΑΠΡΙΛΙΟΣ 2012



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**ΜΕΘΟΔΟΙ ΠΟΛΥΜΕΤΑΒΛΗΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ
ΑΝΑΛΥΣΗΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ**

Ναταλία Δ. Μαύρου

Επιβλέπουσα: Χρυσής Καρώνη
Αναπληρώτρια Καθηγήτρια ΕΜΠ

Εξεταστική Επιτροπή: Β. Παπανικολάου, Καθηγητής ΕΜΠ
Ι. Βόντα, Επίκ. Καθηγήτρια ΕΜΠ

ΑΘΗΝΑ, ΑΠΡΙΛΙΟΣ 2012

Στην οικογένεια μου

ΠΡΟΛΟΓΟΣ

Η πολυμεταβλητή στατιστική ανάλυση αποτελεί ένα πολύτιμο εργαλείο σε πολλές επιστήμες. Οι μέθοδοι της πολυμεταβλητής ανάλυσης αναφέρονται σε διαδικασίες και μεθοδολογίες που χρησιμοποιούνται για να καταλήξουμε σε στατιστική συμπερασματολογία με τη χρήση πολλών μεταβλητών. Οι μέθοδοι, αυτοί, είναι ιδιαίτερα χρήσιμες, γιατί βοηθούν στην εξαγωγή συμπερασμάτων από μεγάλο αριθμό δεδομένων και στον περιορισμό της αβεβαιότητας, καθώς και στην εύρεση και στην ερμηνεία συσχετίσεων μεταξύ των μεταβλητών.

Η παρούσα διπλωματική εργασία αφορά την εφαρμογή ποσοτικών μεθόδων της πολυμεταβλητής στατιστικής ανάλυσης σε μεγάλο αριθμό δεδομένων, από έρευνα που διεξάγεται ετησίως και ονομάζεται TGI. Σκοπός της εργασίας, μέσα από αυτά τα δεδομένα, είναι η διερεύνηση του προφίλ των αναγνωστών των κυριακάτικων και των ημερήσιων εφημερίδων, η ομαδοποίηση τους ανάλογα με την αναγνωστική τους συμπεριφορά και στη συνέχεια η ανάλυση του προφίλ αυτών των ομάδων.

Στα πλαίσια της μελέτης εξετάζονται οι περιγραφικοί δείκτες, τα γραφήματα και η παραγοντική δομή του ερωτηματολογίου χρησιμοποιώντας την παραγοντική ανάλυση και την ανάλυση κατά συστάδες.

Ευχαριστώ την Αναπληρώτρια Καθηγήτρια κυρία Χρυσήδα Καρώνη για την ανάθεση της εργασίας, τις λεπτομερείς διορθώσεις της, την καθοδήγηση της και την υπομονή της. Τέλος, ευχαριστώ την οικογένειά μου για την υποστήριξή της.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1	1
<i>ΕΙΣΑΓΩΓΗ</i>	1
ΚΕΦΑΛΑΙΟ 2	3
<i>ΑΝΑΛΥΣΗ ΜΕΤΑΒΛΗΤΩΝ</i>	3
2.1 Περιγραφικά στοιχεία μεταβλητών	3
2.2 Συσχετίσεις μεταβλητών	9
2.3 Αναγνωστική συμπεριφορά ερωτώμενων.....	15
2.4 Δημογραφικό προφίλ αναγνωστών.....	19
ΚΕΦΑΛΑΙΟ 3	21
<i>ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ</i>	21
3.1 Εισαγωγή.....	21
3.2 Το Ορθογώνιο Μοντέλο.....	21
3.3 Υποθέσεις του Ορθογώνιου Μοντέλου	22
3.4 Έλεγχος συσχετίσεων	24
3.5 Αριθμός Παραγόντων και Εκτίμηση των Παραγόντων.....	26
3.5.1 Εκτίμηση με τη μέθοδο Κύριων Συνιστωσών.....	26
3.5.2 Εκτίμηση με τη μέθοδο Μέγιστης Πιθανοφάνειας	27
3.6 Περιστροφή	28
3.7 Υπολογισμός των Σκορ των Παραγόντων	29
3.8 Εφαρμογή των Μεθόδων	30
3.8.1 Καταλληλότητα των δεδομένων	30
3.8.2 Επιλογή Αριθμού Παραγόντων.....	34
ΚΕΦΑΛΑΙΟ 4	53
<i>ΑΝΑΛΥΣΗ ΚΑΤΑ ΣΥΣΤΑΔΕΣ</i>	53
4.1 Εισαγωγή.....	53
4.2 Τι είναι Ανάλυση Κατά Συστάδες	53
4.3 Στόχος της Ανάλυσης Κατά Συστάδες	54

4.4 Η απόσταση.....	54
4.5 Ιεραρχική Ομαδοποίηση	56
4.5.1 Συσσωρευτικές μέθοδοι (<i>Agglomerative methods</i>)	56
4.5.2 Διαιρετικές μέθοδοι (<i>Divisive methods</i>)	59
4.6 Μη Ιεραρχική Ομαδοποίηση.....	59
4.6.1 Η μέθοδος <i>K-Means</i>	60
4.7 Two-Step Ομαδοποίηση.....	61
4.8 Εφαρμογή των μεθόδων <i>K-Means</i> και Two-Step	62
ΚΕΦΑΛΑΙΟ 5	77
ΣΥΜΠΕΡΑΣΜΑΤΑ	77
ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	81

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

Σε αυτό το εισαγωγικό κεφάλαιο θα εξηγηθεί αναλυτικά η προέλευση των δεδομένων, η ταυτότητα της έρευνας που χρησιμοποιήθηκε και οι αναλύσεις που ακολούθησαν σε κάθε κεφάλαιο, με σκοπό την επεξεργασία και την ανάλυση αυτών των δεδομένων.

Τι είναι το TGI

Το TGI (Target Group Index) είναι η παλαιότερη έρευνα «μοναδικής πηγής» (single source), η οποία περιέχει πολύτιμη πληροφόρηση για το σύνολο της αγοράς. Η πρώτη φορά που διεξήχθη ήταν το 1969 στην Αγγλία και από τότε χρησιμοποιείται σε 60 χώρες με πάνω από 700.000 συνεντευξιαζόμενους κάθε χρόνο.

Στη Ελλάδα το TGI εφαρμόζεται για περισσότερα από 23 χρόνια και συλλέγει στοιχεία από ένα αντιπροσωπευτικό δείγμα περίπου 11.000 ατόμων ετησίως, τα οποία δημοσιεύονται ανά τετράμηνο. Η έρευνα περιέχει ερωτήσεις που καλύπτουν τις συμπεριφορές, τα κίνητρα, τις συνήθειες ως προς τα μέσα ενημέρωσης και την καταναλωτική συμπεριφορά των ερωτώμενων. Το γεγονός ότι το TGI είναι έρευνα «μοναδικής πηγής» (single source) σημαίνει ότι οι ερωτώμενοι πρέπει να συμπληρώσουν ολόκληρο το ερωτηματολόγιο, καλύπτοντας όλες τις θεματικές ενότητες. Αυτό, κατ' επέκταση σημαίνει ότι μπορεί να γίνει πολλαπλή ανάλυση μίας μεταβλητής του TGI με μια άλλη μεταβλητή του TGI, καταλήγοντας έτσι σε μια συνολική εικόνα για τον ερωτώμενο. Οι τομείς διερεύνησης είναι η δημογραφική ανάλυση, κατανάλωση MME, τρόπος ζωής, κατανάλωση προϊόντων και υπηρεσιών.

Στην ανάλυση που ακολουθεί χρησιμοποιήθηκε επιλεκτικά μέρος του ερωτηματολογίου του TGI, από το οποίο επιλέχθηκαν ερωτήσεις από τους τρεις πρώτους τομείς που αναφέρθηκαν προηγουμένως και στο οποίο απάντησαν 11.090 ερωτώμενοι. Οι ερωτήσεις που χρησιμοποιήθηκαν στην ανάλυση αφορούν δημογραφικά χαρακτηριστικά, δηλαδή φύλο, ηλικία, εκπαίδευση, οικογενειακή κατάσταση, θέση στην απασχόληση, κοινωνική τάξη και τόπο διαμονής, και ερωτήσεις που αφορούν τα έντυπα και συγκεκριμένα τις κυριακάτικες και τις ημερήσιες εφημερίδες. Οι ερωτήσεις αυτές αφορούν την συχνότητα ανάγνωσης και την ποσότητα αγοράς εφημερίδων, καθώς επίσης και ερωτήσεις που αφορούν τις αναγνωστικές συνήθειες των ερωτώμενων.

Χρησιμοποιώντας, λοιπόν, τα χαρακτηριστικά και το προφίλ των αναγνωστών των εφημερίδων, καθώς και τα κίνητρα, τις συνήθειες αλλά και τους λόγους ανάγνωσης και επιλογής των εφημερίδων θα οδηγηθούμε στην ομαδοποίηση των αναγνωστών.

Ταυτότητα έρευνας

Την έρευνα διεξήγαγε εταιρία ερευνών. Η συλλογή των στοιχείων ήταν συνεχής και μηνιαία, με εξαίρεση τον μήνα Αύγουστο. Η τρέχουσα βάση δεδομένων αφορά τους μήνες Οκτώβριο του 2010 με Σεπτέμβριο του 2011. Είναι συνδρομητική, ποσοτική έρευνα με τη χρήση ερωτηματολογίων στις κατοικίες των ερωτώμενων, που διεξήχθη σε δύο φάσεις (Face to face και αυτοσυμπλήρωση). Το δείγμα της έρευνας, και το οποίο αναλύσαμε, είναι 11.090 άτομα Πανελλαδικά.. Έγινε στρωματοποιημένη πανελλαδική δειγματοληψία, της οποίας η στρωματοποίηση έχει ως αρχικό σημείο αναφοράς εννέα γεωγραφικά διαμερίσματα: Αθήνα (34%), Θεσσαλονίκη (17%), Στερεά (λοιπή) (11%), Πελοπόννησος (9%), Κρήτη (5%), Μακεδονία (λοιπή) (13%), Θράκη (3%), Θεσσαλία (6%), Ήπειρος (2%). Το ποσοστό αντρών – γυναικών του δείγματος είναι 50,1% έναντι 49,9% και η ηλικίες κυμαίνονται από 15 έως 64 ετών. Οι περιοχές που καλύπτει η έρευνα είναι η Αθήνα, η Θεσσαλονίκη, αστικές, ημιαστικές και αγροτικές περιοχές.

Αναλύσεις

Ξεκινώντας την εργασία, στο κεφάλαιο 1, γίνεται μια περιγραφική ανάλυση των μεταβλητών για τη διερεύνηση του δείγματος αλλά και ως μια πρώτη προσέγγιση για την ανάλυση που θα ακολουθήσει. Συγκεκριμένα, αρχικά έγινε δημογραφική ανάλυση του δείγματος και στη συνέχεια έγινε έλεγχος συσχέτισης μεταξύ βασικών δημογραφικών μεταβλητών (φύλο, ηλικία, κοινωνική τάξη) και της αναγνωσιμότητας ημερήσιων και κυριακάτικων εφημερίδων. Στην συνέχεια, διερευνήθηκε η αναγνωστική συμπεριφορά όλων των ερωτώμενων, σε σχέση με το πόσες εφημερίδες διαβάζουν, αλλά και η αναγνωστική συμπεριφορά των αναγνωστών των ημερήσιων και των κυριακάτικων εφημερίδων. Τέλος, εξετάζεται το δημογραφικό προφίλ των αναγνωστών αυτών των εφημερίδων ανά εφημερίδα, αλλά και ανά δημογραφικό χαρακτηριστικό.

Στο επόμενο κεφάλαιο, πραγματοποιείται παραγοντική ανάλυση για τη στάση των ερωτώμενων απέναντι στις εφημερίδες με τη μέθοδο των κύριων συνιστωσών, αλλά και με τη μέθοδο της μέγιστης πιθανοφάνειας και στη συνέχεια γίνεται παρουσίαση των αποτελεσμάτων της.

Στο τελευταίο κεφάλαιο, γίνεται ανάλυση κατά συστάδες, με βάση τα αποτελέσματα της παραγοντικής ανάλυσης με τη μέθοδο εκτίμησης μέγιστης πιθανοφάνειας, με σκοπό την ομαδοποίηση των ερωτώμενων βάσει των αναγνωστικών τους προτιμήσεων και των κινήτρων τους να αγοράσουν εφημερίδα. Στο ίδιο κεφάλαιο υπάρχουν τα αποτελέσματα της εφαρμογής των μεθόδων K-means και Two-Step και στη συνέχεια εξετάζουμε τη σύσταση των ομάδων αυτών, με βάση τα δημογραφικά τους χαρακτηριστικά, αλλά και την αναγνωστική τους συμπεριφορά.

Η ανάλυση των δεδομένων έγινε με την χρήση του στατιστικού πακέτου IBM SPSS Statistics 17.0.

ΚΕΦΑΛΑΙΟ 2

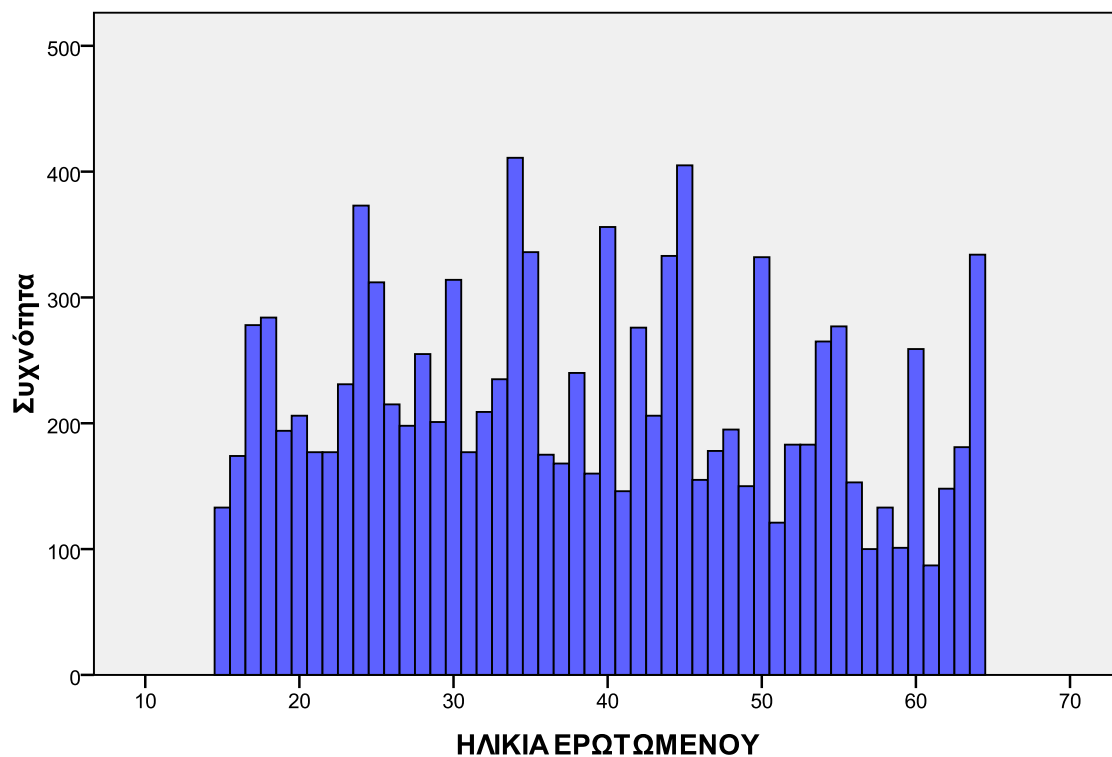
ΑΝΑΛΥΣΗ ΜΕΤΑΒΛΗΤΩΝ

2.1 Περιγραφικά στοιχεία μεταβλητών

Σκοπός της περιγραφικής ανάλυσης είναι να δώσει μια συνοπτική παρουσίαση του δείγματος, καθώς επίσης και να ελέγξει την ορθότητα των τιμών του. Αυτό γίνεται με τη βοήθεια απλών περιγραφικών πινάκων ή γραφημάτων. Η επιλογή των κατάλληλων αριθμητικών και γραφικών μεθόδων γίνεται με βάση τον τύπο της μεταβλητής που θέλουμε να παρουσιάσουμε.

Ο μέσος όρος ηλικίας του δείγματος είναι 38,5, ενώ της διαμέσου είναι 38, συνεπώς δεν έχουν μεγάλη διαφορά μεταξύ τους και άρα μπορούμε να πούμε ότι υπάρχει συμμετρία στην κατανομή. Επίσης, η μικρότερη τιμή που παρατηρείται είναι η ηλικία των 15 ετών και η μεγαλύτερη είναι η ηλικία των 64 ετών.

Ιστόγραμμα



Πίνακας 2.1.2: Πίνακας συχνοτήτων για τη μεταβλητή φύλο

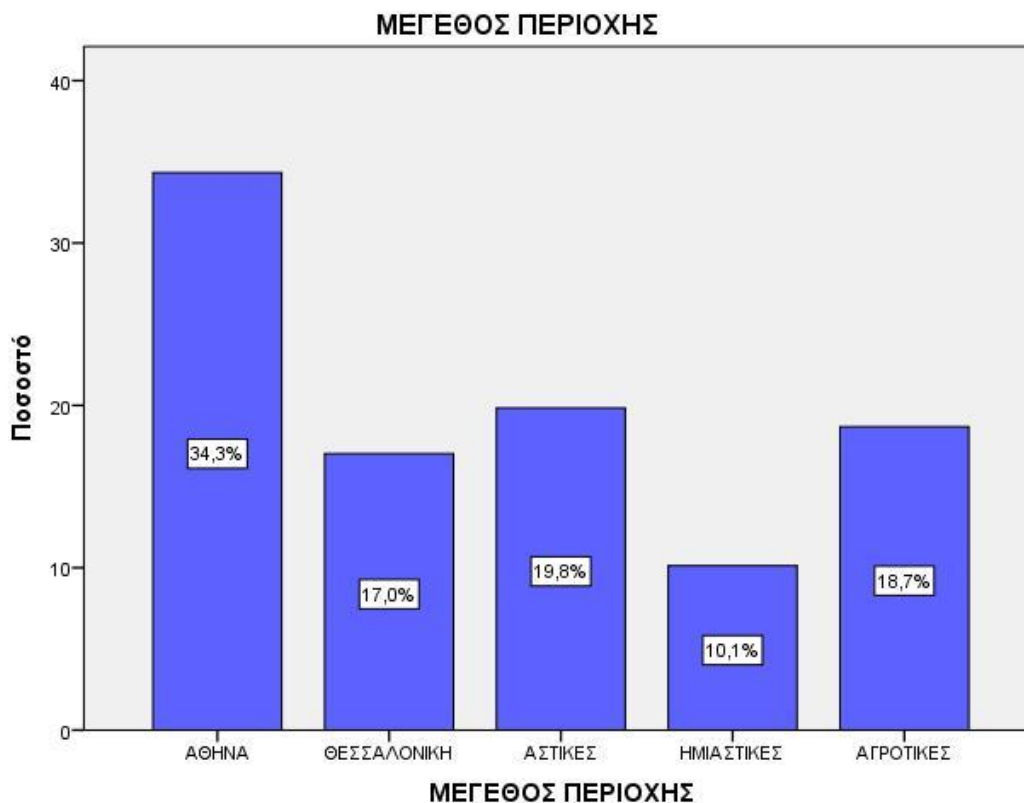
ΦΥΛΟ		
	Συχνότητα	Ποσοστό
ΑΝΔΡΕΣ	5605	50,5
ΓΥΝΑΙΚΕΣ	5485	49,5
ΣΥΝΟΛΟ	11090	100,0

Σύμφωνα με τον πίνακα συχνοτήτων το 50,5% του δείγματος είναι άντρες ενώ το 49,5% είναι γυναίκες.

Πίνακας 2.1.2: Πίνακας συχνοτήτων για τη μεταβλητή περιοχή

ΜΕΓΕΘΟΣ ΠΕΡΙΟΧΗΣ		
	Συχνότητα	Ποσοστό
ΑΘΗΝΑ	3807	34,3
ΘΕΣΣΑΛΟΝΙΚΗ	1888	17,0
ΑΣΤΙΚΕΣ	2200	19,8
ΗΜΙΑΣΤΙΚΕΣ	1123	10,1
ΑΓΡΟΤΙΚΕΣ	2072	18,7
ΣΥΝΟΛΟ	11090	100,0

Από τον πίνακα συχνοτήτων για το μέγεθος της περιοχής βλέπουμε ότι το 34,3% των ερωτώμενων είναι κάτοικοι της Αθήνας και το 17% της Θεσσαλονίκης. Το 19,8% είναι κάτοικοι αστικών κέντρων, το 10% ημιαστικών κέντρων και το 18,7% είναι κάτοικοι αγροτικών περιοχών.

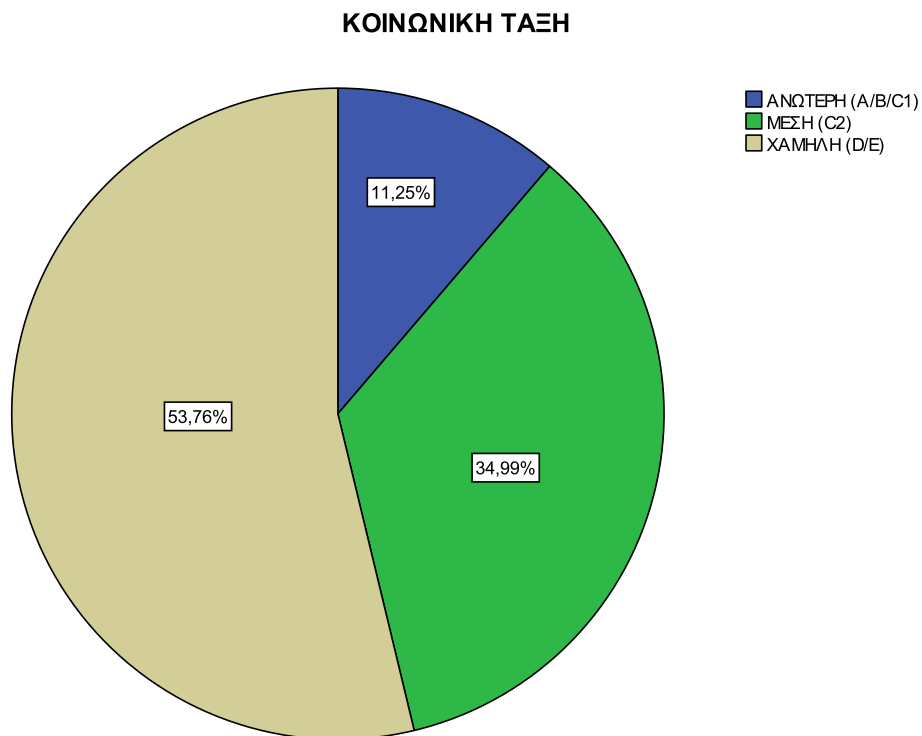


Πίνακας 2.1.3: Πίνακας συχνοτήτων για τη μεταβλητή κοινωνική τάξη

ΚΟΙΝΩΝΙΚΗ ΤΑΞΗ

	Συχνότητα	Ποσοστό
ΑΝΩΤΕΡΗ (Α/Β/С1)	1248	11,3
ΜΕΣΗ (C2)	3880	35,0
ΧΑΜΗΛΗ (D/E)	5962	53,8
ΣΥΝΟΛΟ	11090	100,0

Από τον πίνακα συχνοτήτων για την κοινωνική τάξη βλέπουμε ότι περίπου το 11% των ερωτώμενων ανήκουν στην ανώτερη τάξη, ενώ το 35% ανήκει στη μέση και περίπου το 54% ανήκει στη χαμηλή.

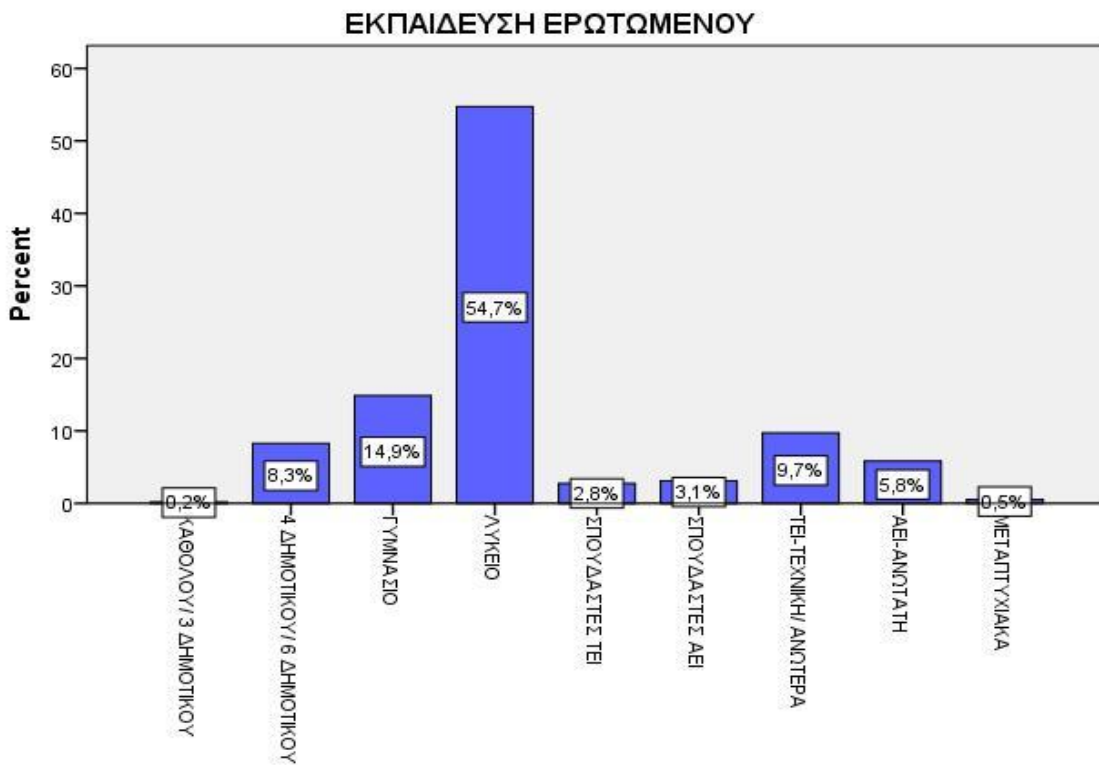


Πίνακας 2.1.4: Πίνακας συχνοτήτων για τη μεταβλητή μόρφωση

ΕΚΠΑΙΔΕΥΣΗ ΕΡΩΤΩΜΕΝΟΥ

	Συχνότητα	Ποσοστό
ΚΑΘΟΛΟΥ/ 3 ΔΗΜΟΤΙΚΟΥ	23	0,2
4 ΔΗΜΟΤΙΚΟΥ/ 6 ΔΗΜΟΤΙΚΟΥ	917	8,3
ΓΥΜΝΑΣΙΟ	1649	14,9
ΛΥΚΕΙΟ	6069	54,7
ΣΠΟΥΔΑΣΤΕΣ ΤΕΙ	305	2,8
ΣΠΟΥΔΑΣΤΕΣ ΑΕΙ	344	3,1
ΤΕΙ-ΤΕΧΝΙΚΗ/ ΑΝΩΤΕΡΑ	1079	9,7
ΑΕΙ-ΑΝΩΤΑΤΗ	646	5,8
ΜΕΤΑΠΤΥΧΙΑΚΑ	58	,5
ΣΥΝΟΛΟ	11090	100,0

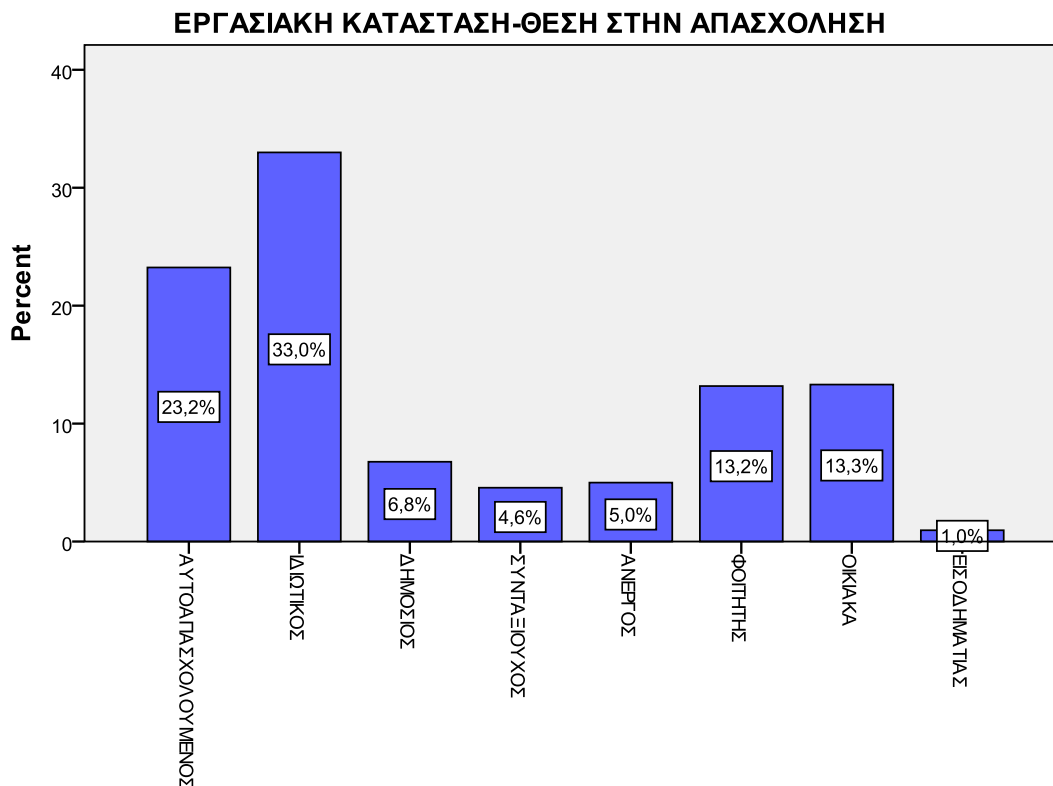
Από τον πίνακα συχνοτήτων για τη μεταβλητή μόρφωση βλέπουμε ότι το μεγαλύτερο ποσοστό των ερωτώμενων, δηλαδή περίπου το 55%, είναι απόφοιτοι λυκείου, ενώ ένα 16% είναι απόφοιτοι ανωτέρας /ανωτάτης σχολής. Μόνο το 0,5% έχουν κάνει μεταπτυχιακά ενώ περίπου το 6% είναι σπουδαστές σε ΤΕΙ και ΑΕΙ.



Πίνακας 2.1.5: Πίνακας συχνοτήτων για τη μεταβλητή εργασιακή κατάσταση

	Συχνότητα	Ποσοστό
ΑΥΤΟΑΠΑΣΧΟΛΟΥΜΕΝΟΣ	2577	23,2
ΙΔΙΩΤΙΚΟΣ	3659	33,0
ΔΗΜΟΣΙΟΣ	750	6,8
ΣΥΝΤΑΞΙΟΥΧΟΣ	506	4,6
ΑΝΕΡΓΟΣ	554	5,0
ΦΟΙΤΗΤΗΣ	1462	13,2
ΟΙΚΙΑΚΑ	1476	13,3
ΕΙΣΟΔΗΜΑΤΙΑΣ	106	1,0
ΣΥΝΟΛΟ	11090	100,0

Από το πίνακα συχνοτήτων για την μεταβλητή εργασιακή κατάσταση βλέπουμε ότι οι ιδιωτικοί υπάλληλοι 33% υπερτερούν των δημοσίων υπαλλήλων σχεδόν 7%. Μπορούμε να πούμε ότι συνολικά οι υπάλληλοι αποτελούν το 40% του δείγματος, ενώ οι αυτοαπασχολούμενοι αποτελούν το 23%. Οι φοιτητές και όσοι ασχολούνται με οικιακά έχουν ένα πολύ κοντινό ποσοστό, περίπου το 13%.



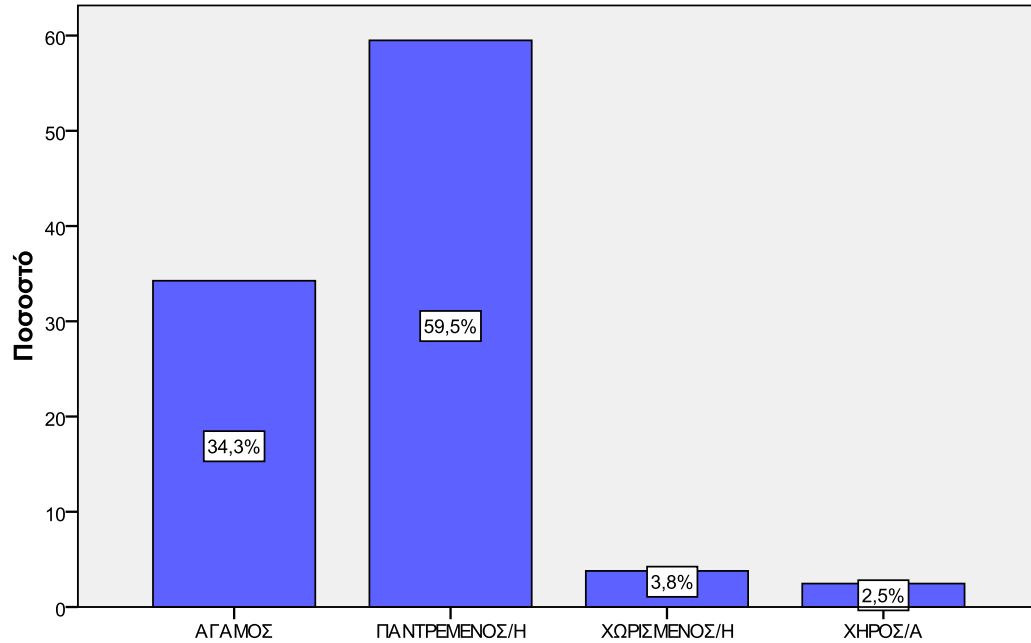
Πίνακας 2.1.6: Πίνακας συχνοτήτων για τη μεταβλητή οικογενειακή κατάσταση

ΟΙΚΟΓΕΝΕΙΑΚΗ ΚΑΤΑΣΤΑΣΗ

	Συχνότητα	Ποσοστό
ΑΓΑΜΟΣ	3799	34,3
ΠΑΝΤΡΕΜΕΝΟΣ/Η	6598	59,5
ΧΩΡΙΣΜΕΝΟΣ/Η	420	3,8
ΧΗΡΟΣ/Α	273	2,5
ΣΥΝΟΛΟ	11090	100,0

Από το πίνακα συχνοτήτων για την μεταβλητή οικογενειακή κατάσταση βλέπουμε ότι άγαμοι είναι το 34% των ερωτώμενων και παντρεμένοι είναι περίπου το 60%. Ενώ παράλληλα το 4% είναι χωρισμένοι και το περίπου 3% είναι χήροι και χήρες.

ΟΙΚΟΓΕΝΕΙΑΚΗ ΚΑΤΑΣΤΑΣΗ



2.2 Συσχετίσεις μεταβλητών

Για να εξετάσουμε αν υπάρχει ή όχι συσχέτιση μεταξύ των διάφορων δημογραφικών χαρακτηριστικών, όπως το φύλο, η ηλικιακή κατηγορία και η κοινωνική τάξη, και της αναγνωσιμότητας των ημερήσιων και κυριακάτικων εφημερίδων έγινε ο έλεγχος ανεξαρτησίας χ^2 . Ο έλεγχος αυτός ελέγχει τη μηδενική υπόθεση, H_0 , ότι οι μεταβλητές είναι ανεξάρτητες, έναντι της εναλλακτικής υπόθεσης, H_1 , ότι είναι εξαρτημένες.

Στον πίνακα διπλής εισόδου 2.2.1, παρατηρείται ότι από το 13% του συνόλου των ερωτώμενων που επέλεξε να διαβάσει ημερήσιες εφημερίδες, το 18,7% είναι άντρες έναντι του 7,3% που είναι γυναίκες. Το μεγαλύτερο ποσοστό, 87%, δεν διαβάζει ημερήσιες, με τις γυναίκες να έχουν μεγαλύτερο ποσοστό έναντι των αντρών.

Πίνακας 2.2.1: Συσχέτιση φύλου – αναγνωσιμότητας ημερήσιων εφημερίδων

			ΗΜΕΡΗΣΙΕΣ		ΣΥΝΟΛΟ
			ΟΧΙ	ΝΑΙ	
ΦΥΛΟ	ΑΝΔΡΕΣ	Count	4559	1046	5605
		% within ΦΥΛΟ	81,3%	18,7%	100,0%
		% within ΗΜΕΡΗΣΙΕΣ	47,3%	72,3%	50,5%
	ΓΥΝΑΙΚΕΣ	Count	5085	400	5485
		% within ΦΥΛΟ	92,7%	7,3%	100,0%
		% within ΗΜΕΡΗΣΙΕΣ	52,7%	27,7%	49,5%
ΣΥΝΟΛΟ	Count	9644	1446	11090	
	% within ΦΥΛΟ	87,0%	13,0%	100,0%	
	% within ΗΜΕΡΗΣΙΕΣ	100,0%	100,0%	100,0%	

Χρησιμοποιώντας τον έλεγχο ανεξαρτησίας χ^2 και βλέποντας ότι η p-τιμή είναι μικρότερη του 0,05, γίνεται αποδεκτή η εναλλακτική υπόθεση (H_1), δηλαδή οι δύο μεταβλητές σχετίζονται. Συνεπώς, το φύλο συσχετίζεται με την αναγνωσιμότητα των ημερήσιων εφημερίδων.

Στον πίνακα διπλής εισόδου 2.2.2, παρατηρείται ότι από το 25,2% του συνόλου των ερωτώμενων που επιλέγει να διαβάσει κυριακάτικες εφημερίδες, το 31,7% είναι άντρες έναντι του 18,5% που είναι γυναίκες. Το μεγαλύτερο ποσοστό, 74,8%, δεν διαβάζει κυριακάτικες εφημερίδες, με τις γυναίκες να έχουν μεγαλύτερο ποσοστό έναντι των αντρών.

Πίνακας 2.2.2: Συσχέτιση φύλου – αναγνωσιμότητας κυριακάτικων εφημερίδων

			ΚΥΡΙΑΚΑΤΙΚΕΣ		ΣΥΝΟΛΟ
			ΟΧΙ	ΝΑΙ	
ΦΥΛΟ	ΑΝΔΡΕΣ	Count	3829	1776	5605
		% within ΦΥΛΟ	68,3%	31,7%	100,0%
		% within ΚΥΡΙΑΚΑΤΙΚΕΣ	46,1%	63,6%	50,5%
	ΓΥΝΑΙΚΕΣ	Count	4469	1016	5485
		% within ΦΥΛΟ	81,5%	18,5%	100,0%
		% within ΚΥΡΙΑΚΑΤΙΚΕΣ	53,9%	36,4%	49,5%
ΣΥΝΟΛΟ	Count	8298	2792	11090	
	% within ΦΥΛΟ	74,8%	25,2%	100,0%	
	% within ΚΥΡΙΑΚΑΤΙΚΕΣ	100,0%	100,0%	100,0%	

Χρησιμοποιώντας τον έλεγχο ανεξαρτησίας χ^2 και βλέποντας ότι η p-τιμή είναι μικρότερη του 0,05, γίνεται αποδεκτή η εναλλακτική υπόθεση (H_1), δηλαδή οι δύο μεταβλητές σχετίζονται.. Συνεπώς, το φύλο συσχετίζεται με την αναγνωσιμότητα των κυριακάτικων εφημερίδων.

Ως προς την ηλικία παρατηρείται, στον πίνακα 2.2.3, ότι το μεγαλύτερο ποσοστό των αναγνωστών, 22,8%, των ημερήσιων εφημερίδων ανήκει στην ηλικιακή κατηγορία των 25-34, ενώ αρκετά υψηλό είναι και το ποσοστό της ηλικιακής κατηγορίας 35-44, 22,8%. Συνεπώς, οι αναγνώστες των ημερήσιων εφημερίδων είναι κυρίως νεανικά κοινά.

Πίνακας 2.2.3: Συσχέτιση ηλικιακής κατηγορίας – αναγνωσιμότητας ημερήσιων εφημερίδων

			ΗΜΕΡΗΣΙΕΣ		ΣΥΝΟΛΟ
			ΟΧΙ	ΝΑΙ	
ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	15-24	Count	1993	234	2227
		% within ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	89,5%	10,5%	100,0%
		% within ΗΜΕΡΗΣΙΕΣ	20,7%	16,2%	20,1%
	25-34	Count	2192	335	2527
		% within ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	86,7%	13,3%	100,0%
		% within ΗΜΕΡΗΣΙΕΣ	22,7%	23,2%	22,8%
	35-44	Count	2067	329	2396
		% within ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	86,3%	13,7%	100,0%
		% within ΗΜΕΡΗΣΙΕΣ	21,4%	22,8%	21,6%
	45-54	Count	1862	305	2167
		% within ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	85,9%	14,1%	100,0%
		% within ΗΜΕΡΗΣΙΕΣ	19,3%	21,1%	19,5%
	55-64	Count	1530	243	1773
		% within ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	86,3%	13,7%	100,0%
		% within ΗΜΕΡΗΣΙΕΣ	15,9%	16,8%	16,0%
ΣΥΝΟΛΟ	Count	9644	1446	11090	
	% within ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	87,0%	13,0%	100,0%	
	% within ΗΜΕΡΗΣΙΕΣ	100,0%	100,0%	100,0%	

Χρησιμοποιώντας τον έλεγχο ανεξαρτησίας χ^2 και βλέποντας ότι η p-τιμή = 0,002 < 0,05, γίνεται αποδεκτή η εναλλακτική υπόθεση (H_1), δηλαδή οι δύο μεταβλητές σχετίζονται. Συνεπώς, η ηλικιακή κατηγορία συσχετίζεται με την αναγνωσιμότητα των ημερήσιων εφημερίδων.

Στον παρακάτω πίνακα διπλής εισόδου, παρατηρείται ότι το μεγαλύτερο ποσοστό των

αναγνωστών, 29%, των κυριακάτικων εφημερίδων ανήκει στην ηλικιακή κατηγορία των 35-44, ενώ αρκετά υψηλό είναι και το ποσοστό της ηλικιακής κατηγορίας 25-34, 24,6%. Συνεπώς, οι αναγνώστες την κυριακάτικων εφημερίδων είναι κυρίως νεανικά κοινά.

Πίνακας 2.2.4: Συσχέτιση ηλικιακής κατηγορίας – αναγνωσιμότητας κυριακάτικων εφημερίδων

			ΚΥΡΙΑΚΑΤΙΚΕΣ		ΣΥΝΟΛΟ
			ΟΧΙ	ΝΑΙ	
ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	15-24	Count	1840	387	2227
		% within ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	82,6%	17,4%	100,0%
		% within ΚΥΡΙΑΚΑΤΙΚΕΣ	22,2%	13,9%	20,1%
	25-34	Count	1839	688	2527
		% within ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	72,8%	27,2%	100,0%
% within ΚΥΡΙΑΚΑΤΙΚΕΣ		22,2%	24,6%	22,8%	
35-44	Count	1701	695	2396	
	% within ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	71,0%	29,0%	100,0%	
	% within ΚΥΡΙΑΚΑΤΙΚΕΣ	20,5%	24,9%	21,6%	
45-54	Count	1556	611	2167	
	% within ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	71,8%	28,2%	100,0%	
	% within ΚΥΡΙΑΚΑΤΙΚΕΣ	18,8%	21,9%	19,5%	
55-64	Count	1362	411	1773	
	% within ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	76,8%	23,2%	100,0%	
	% within ΚΥΡΙΑΚΑΤΙΚΕΣ	16,4%	14,7%	16,0%	
ΣΥΝΟΛΟ	Count	8298	2792	11090	
	% within ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	74,8%	25,2%	100,0%	
	% within ΚΥΡΙΑΚΑΤΙΚΕΣ	100,0%	100,0%	100,0%	

Χρησιμοποιώντας τον έλεγχο ανεξαρτησίας χ^2 και βλέποντας ότι η p-τιμή είναι μικρότερη του

0,05, γίνεται αποδεκτή η εναλλακτική υπόθεση (H_1), δηλαδή οι δύο μεταβλητές σχετίζονται. Συνεπώς, η ηλικιακή κατηγορία συσχετίζεται με την αναγνωσιμότητα των κυριακάτικων εφημερίδων.

Ως προς την κοινωνική τάξη παρατηρείται, στον παρακάτω πίνακα διπλής εισόδου, το 19,2% της ανώτερης τάξης διαβάζει ημερήσιες εφημερίδες, το οποίο είναι το μεγαλύτερο ποσοστό αναγνωσιμότητας σε σχέση με τα ποσοστά των υπολοίπων κοινωνικών τάξεων. Βλέπουμε, όμως, ότι από τους αναγνώστες των ημερήσιων εφημερίδων, τα άτομα που ανήκουν στην ανώτερη κοινωνική τάξη αποτελούν το μικρότερο ποσοστό, 16,6%. Ενώ, το 45,8% των αναγνωστών των ημερήσιων εφημερίδων ανήκουν στην χαμηλή κοινωνική τάξη.

Πίνακας 2.2.5: Συσχέτιση κοινωνικής τάξης – αναγνωσιμότητας ημερήσιων εφημερίδων

			ΗΜΕΡΗΣΙΕΣ		ΣΥΝΟΛΟ
			ΟΧΙ	ΝΑΙ	
ΚΟΙΝΩΝΙΚΗ ΤΑΞΗ	ΑΝΩΤΕΡΗ (A/B/C1)	Count	1008	240	1248
		% within ΚΟΙΝΩΝΙΚΗ ΤΑΞΗ	80,8%	19,2%	100,0%
		% within ΗΜΕΡΗΣΙΕΣ	10,5%	16,6%	11,3%
	ΜΕΣΗ (C2)	Count	3336	544	3880
		% within ΚΟΙΝΩΝΙΚΗ ΤΑΞΗ	86,0%	14,0%	100,0%
		% within ΗΜΕΡΗΣΙΕΣ	34,6%	37,6%	35,0%
	ΧΑΜΗΛΗ (D/E)	Count	5300	662	5962
		% within ΚΟΙΝΩΝΙΚΗ ΤΑΞΗ	88,9%	11,1%	100,0%
		% within ΗΜΕΡΗΣΙΕΣ	55,0%	45,8%	53,8%
ΣΥΝΟΛΟ	Count	9644	1446	11090	
	% within ΚΟΙΝΩΝΙΚΗ ΤΑΞΗ	87,0%	13,0%	100,0%	
	% within ΗΜΕΡΗΣΙΕΣ	100,0%	100,0%	100,0%	

Χρησιμοποιώντας τον έλεγχο ανεξαρτησίας χ^2 και βλέποντας ότι η p-τιμή είναι μικρότερη του 0,05, γίνεται αποδεκτή η εναλλακτική υπόθεση (H_1), δηλαδή οι δύο μεταβλητές σχετίζονται. Συνεπώς, η κοινωνική τάξη συσχετίζεται με την αναγνωσιμότητα των ημερήσιων εφημερίδων.

Παρόμοια, στον πίνακα 2.2.6, το 40,7% της ανώτερης τάξης διαβάζει κυριακάτικες εφημερίδες, το οποίο είναι το μεγαλύτερο ποσοστό αναγνωσιμότητας σε σχέση με τα ποσοστά των υπολοίπων κοινωνικών τάξεων. Βλέπουμε, όμως, ότι από τους αναγνώστες των κυριακάτικων εφημερίδων, τα άτομα που ανήκουν στην ανώτερη κοινωνική τάξη αποτελούν το μικρότερο ποσοστό, 18,2%. Ενώ, το 42,8% των αναγνωστών των κυριακάτικων εφημερίδων ανήκουν στην

χαμηλή κοινωνική τάξη.

Πίνακας 2.2.6: Συσχέτιση κοινωνικής τάξης – αναγνωσιμότητας κυριακάτικων εφημερίδων

			ΚΥΡΙΑΚΑΤΙΚΕΣ		ΣΥΝΟΛΟ
			ΟΧΙ	ΝΑΙ	
ΚΟΙΝΩΝΙΚΗ ΤΑΞΗ	ΑΝΩΤΕΡΗ (Α/Β/Σ1)	Count	740	508	1248
		% within ΚΟΙΝΩΝΙΚΗ ΤΑΞΗ	59,3%	40,7%	100,0%
		% within ΚΥΡΙΑΚΑΤΙΚΕΣ	8,9%	18,2%	11,3%
	ΜΕΣΗ (Σ2)	Count	2790	1090	3880
		% within ΚΟΙΝΩΝΙΚΗ ΤΑΞΗ	71,9%	28,1%	100,0%
		% within ΚΥΡΙΑΚΑΤΙΚΕΣ	33,6%	39,0%	35,0%
	ΧΑΜΗΛΗ (Δ/Ε)	Count	4768	1194	5962
		% within ΚΟΙΝΩΝΙΚΗ ΤΑΞΗ	80,0%	20,0%	100,0%
		% within ΚΥΡΙΑΚΑΤΙΚΕΣ	57,5%	42,8%	53,8%
ΣΥΝΟΛΟ	Count	8298	2792	11090	
	% within ΚΟΙΝΩΝΙΚΗ ΤΑΞΗ	74,8%	25,2%	100,0%	
	% within ΚΥΡΙΑΚΑΤΙΚΕΣ	100,0%	100,0%	100,0%	

Χρησιμοποιώντας τον έλεγχο ανεξαρτησίας χ^2 και βλέποντας ότι η p-τιμή είναι μικρότερη του 0,05, γίνεται αποδεκτή η εναλλακτική υπόθεση (H_1), δηλαδή οι δύο μεταβλητές σχετίζονται. Συνεπώς, η κοινωνική τάξη συσχετίζεται με την αναγνωσιμότητα των κυριακάτικων εφημερίδων.

2.3 Αναγνωστική συμπεριφορά ερωτώμενων

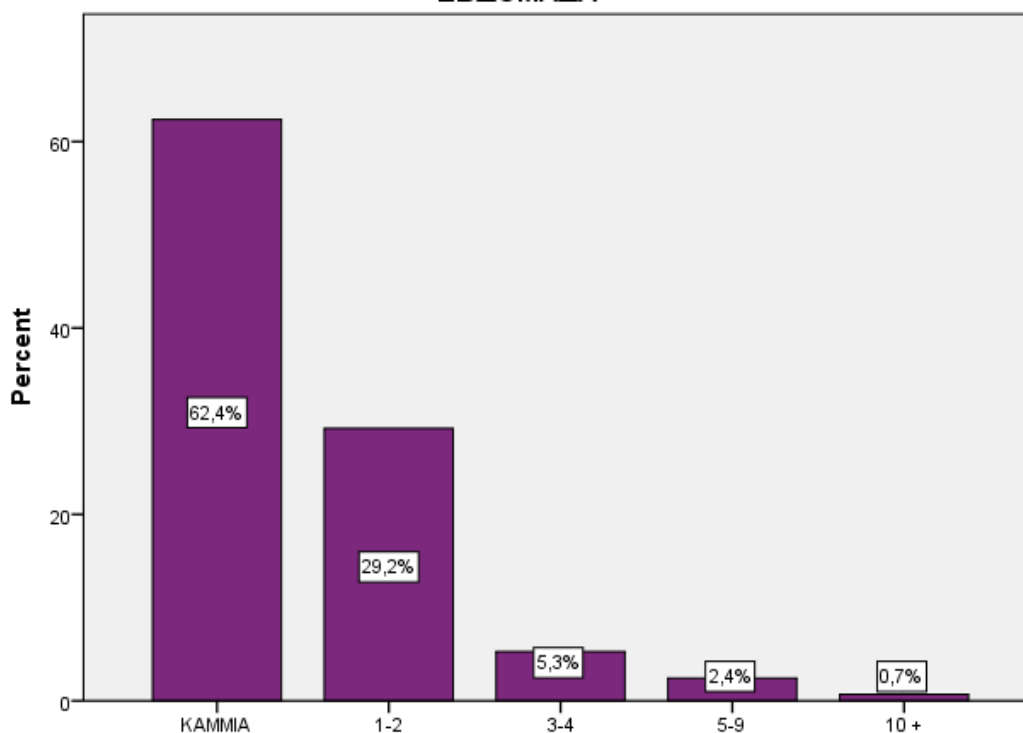
Σε αυτή την παράγραφο θα αναλύσουμε την αναγνωστική συμπεριφορά των ερωτώμενων. Θα εντοπίσουμε τον αριθμό των εφημερίδων που διαβάζουν στο σύνολο οι αναγνώστες εφημερίδων, θα βρούμε το ποσοστό των αναγνωστών που διαβάζουν μόνο ημερήσιες, μόνο κυριακάτικες ή και τα δύο, αλλά και τον αριθμό των εφημερίδων που διαβάζουν αυτά τα ποσοστά των αναγνωστών.

Στο πίνακα 2.3.1 βλέπουμε το πλήθος και τα ποσοστά των ερωτώμενων που απάντησαν στο πόσες εφημερίδες έχουν διαβάσει ή ξεφυλλίσει την τελευταία βδομάδα.

Πίνακας 2.3.1: ΑΝΑΓΝΩΣΤΙΚΗ ΣΥΜΠΕΡΙΦΟΡΑ-ΑΡΙΘΜΟΣ ΕΦΗΜΕΡΙΔΩΝ

		Συχνότητα	Ποσοστό
	ΚΑΜΜΙΑ	6912	62,3
	1-2	3240	29,2
	3-4	585	5,3
	5-9	269	2,4
	10 +	77	,7
	ΣΥΝΟΛΟ	11083	99,9
	ΔΞ/ ΔΑ	7	,1
ΣΥΝΟΛΟ		11090	100,0

Από τον παραπάνω πίνακα βλέπουμε ότι το μεγαλύτερο ποσοστό των ερωτώμενων, 62,3%, έχουν απαντήσει ότι δεν έχουν διαβάσει καμία εφημερίδα την τελευταία εβδομάδα. Το αμέσως μεγαλύτερο ποσοστό, 29,2%, απάντησε ότι έχουν διαβάσει μια με δυο εφημερίδες, ενώ παρατηρούμε ότι όσο αυξάνει ο αριθμός των εφημερίδων, τόσο φθίνει το ποσοστό των ερωτώμενων που έχουν απαντήσει ότι διαβάζουν αυτόν τον αριθμό εφημερίδων, με μικρότερο ποσοστό, 0,7%, αυτών που έχουν πει ότι έχουν διαβάσει πάνω από 10 εφημερίδες την τελευταία εβδομάδα.

ΠΟΣΕΣ ΕΦΗΜΕΡΙΔΕΣ ΕΧΕΤΕ ΔΙΑΒΑΣΕΙ/ ΞΕΦΥΛΙΣΕΙ ΤΗΝ ΤΕΛΕΥΤΑΙΑ ΕΒΔΟΜΑΔΑ

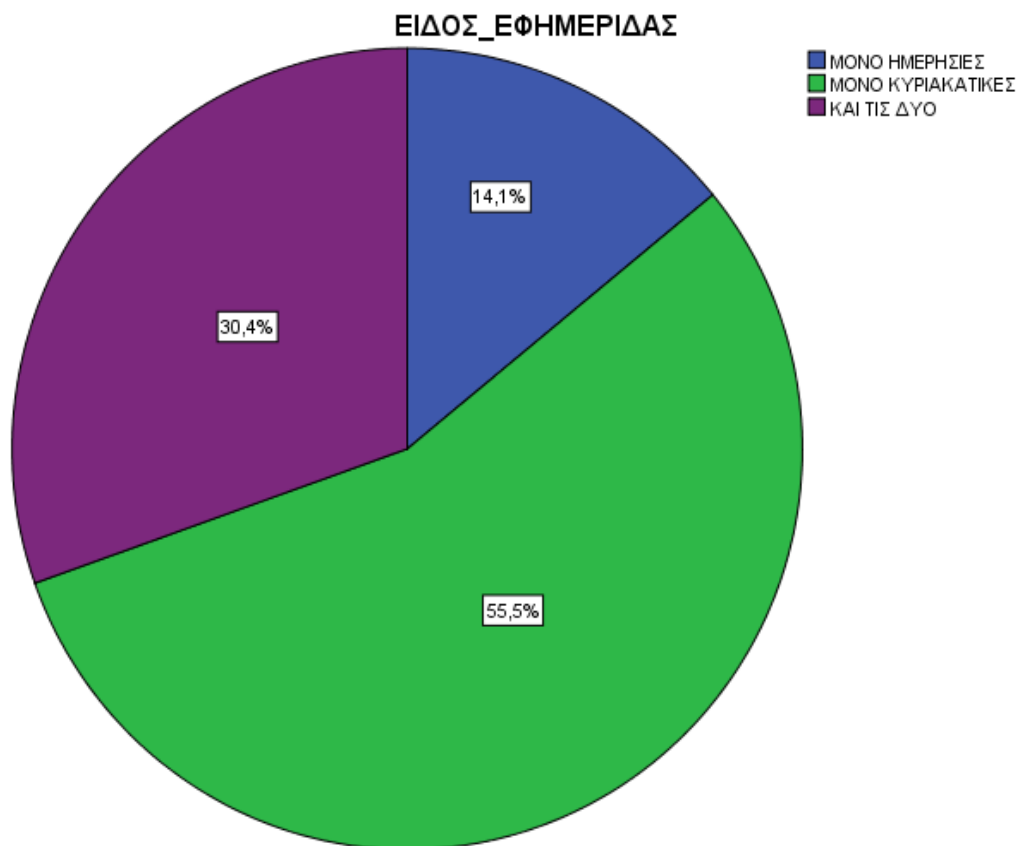
Στο πίνακα 2.3.2 βλέπουμε το πλήθος και τα ποσοστά των ερωτώμενων που διαβάζουν μόνο ημερήσιες ή μόνο κυριακάτικες ή και τους δύο τύπους εφημερίδων.

Πίνακας 2.3.2: ΑΝΑΓΝΩΣΤΙΚΗ ΣΥΜΠΕΡΙΦΟΡΑ-ΕΦΗΜΕΡΙΔΕΣ ΠΟΥ ΔΙΑΒΑΖΟΥΝ

		Συχνότητα	Ποσοστό
	ΜΟΝΟ ΗΜΕΡΗΣΙΕΣ	457	4,1
	ΜΟΝΟ ΚΥΡΙΑΚΑΤΙΚΕΣ	1803	16,3
	ΚΑΙ ΤΙΣ ΔΥΟ	989	8,9
	ΣΥΝΟΛΟ	3249	29,3
	ΔΞ/ΔΑ	7841	70,7
ΣΥΝΟΛΟ		11090	100,0

Από τον παραπάνω πίνακα παρατηρούμε ότι το μεγαλύτερο ποσοστό των ερωτώμενων, 16,3%, διαβάζουν μόνο κυριακάτικες, ενώ το 8,9% διαβάζουν και τους δύο τύπους εφημερίδων. Συνεπώς, παρατηρούμε ότι το μικρότερο ποσοστό των ερωτώμενων, 4,1%, διαβάζουν μόνο ημερήσιες εφημερίδες.

Ακολουθεί το γράφημα που περιγράφει αυτόν τον πίνακα και διαγραμματικά.

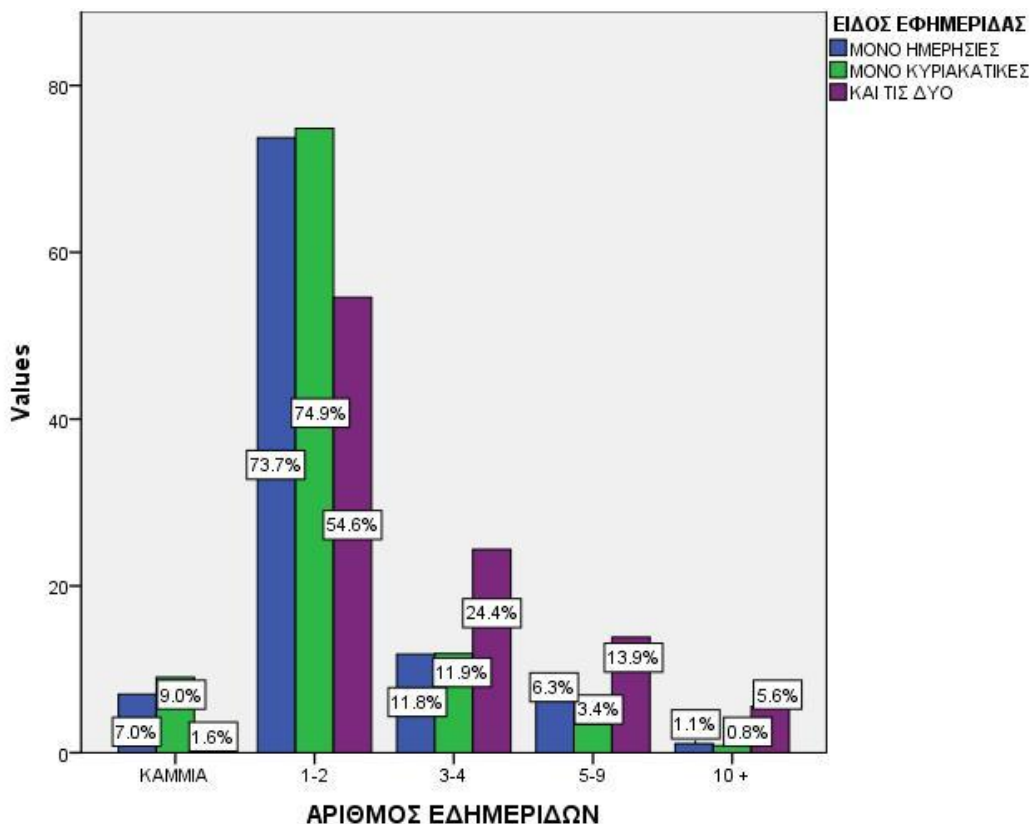


Στο πίνακα 2.3.3 βλέπουμε τα ποσοστά των ερωτώμενων που απάντησαν στο πόσες εφημερίδες έχουν διαβάσει ή ξεφυλλίσει την τελευταία βδομάδα ανάλογα με το αν διαβάζουν μόνο ημερήσιες ή μόνο κυριακάτικες ή και τους δύο τύπους εφημερίδων.

Πίνακας 2.3.3: ΑΝΑΓΝΩΣΤΙΚΗ ΣΥΜΠΕΡΙΦΟΡΑ-ΑΡΙΘΜΟΣ ΕΦΗΜΕΡΙΔΩΝ

		ΠΟΣΕΣ ΕΦΗΜΕΡΙΔΕΣ ΕΧΕΤΕ ΔΙΑΒΑΣΕΙ/ ΞΕΦΥΛΛΙΣΕΙ ΤΗΝ ΤΕΛΕΥΤΑΙΑ ΕΒΔΟΜΑΔΑ				
		ΚΑΜΜΙΑ	1-2	3-4	5-9	10 +
ΕΙΔΟΣ ΕΦΗΜΕΡΙΔΑΣ	ΜΟΝΟ ΗΜΕΡΗΣΙΕΣ	7,0%	73,7%	11,8%	6,3%	1,1%
	ΜΟΝΟ ΚΥΡΙΑΚΑΤΙΚΕΣ	9,0%	74,9%	11,9%	3,4%	,8%
	ΚΑΙ ΤΙΣ ΔΥΟ	1,6%	54,6%	24,4%	13,9%	5,6%

Στον παραπάνω πίνακα βλέπουμε ότι το μεγαλύτερο ποσοστό αναγνωστών διαβάζει 1-2 εφημερίδες ανεξάρτητα από το αν οι αναγνώστες διαβάζουν μόνο ημερήσιες, μόνο κυριακάτικες ή και τα δύο είδη. Επίσης, παρατηρούμε ότι είναι μεγαλύτερο το ποσοστό των αναγνωστών που διαβάζουν πάνω από 3 εφημερίδες σε αυτούς που διαβάζουν και ημερήσιες και κυριακάτικες εφημερίδες.



2.4 Δημογραφικό προφίλ αναγνωστών

Μετά την ανάλυση που έγινε παραπάνω για κάθε είδος εφημερίδας ξεχωριστά με τα δημογραφικά χαρακτηριστικά (φύλο, ηλικιακή κατηγορία, κοινωνική τάξη) των ερωτώμενων, τώρα θα τα δούμε σε συγκεντρωτικούς πίνακες προκειμένου να δημιουργήσουμε το δημογραφικό προφίλ των αναγνωστών κάθε εφημερίδας.

Στο πίνακα 2.4.1 βλέπουμε τα ποσοστά ανά στήλη. Δηλαδή βλέπουμε μέσα σε κάθε είδος εφημερίδα πως είναι κατανομημένα το φύλο, οι ηλικιακές κατηγορίες και οι κοινωνικές τάξεις. Παρατηρούμε ότι το ποσοστό των αντρών που διαβάζουν μόνο ημερήσιες, μόνο κυριακάτικες ή και τις δύο είναι μεγαλύτερο από το ποσοστό των γυναικών. Η διαφορά των ποσοστών των δύο φύλων είναι αρκετά μεγάλη ως προς την ανάγνωση των ημερήσιων εφημερίδων αλλά και ως προς την ανάγνωση και των δύο. Αύξηση του ποσοστού των γυναικών παρατηρείται στην ανάγνωση κυριακάτικων εφημερίδων όπου το ποσοστό τους γίνεται 42% έναντι των αντρών 58,4%. Ως προς την ηλικιακή κατηγορία παρατηρούμε ότι το ποσοστό των νέων από 15 έως 44 στις κυριακάτικες εφημερίδες, 63%, είναι μεγαλύτερο από το ποσοστό των νέων που διαβάζουν ημερήσιες, 56%, ενώ το μεγαλύτερο ποσοστό στους νέους εμφανίζεται στους αναγνώστες και δύο ειδών εφημερίδων, 65%. Ως προς την κοινωνική τάξη βλέπουμε ότι το μεγαλύτερο ποσοστό αναγνωστών που ανήκουν στην ανώτερη κοινωνική τάξη εμφανίζεται στους αναγνώστες που διαβάζουν και ημερήσιες και κυριακάτικες εφημερίδες, 19,4%.

Πίνακας 2.4.1: ΔΗΜΟΓΡΑΦΙΚΟ ΠΡΟΦΙΛ ΑΝΑ ΕΙΔΟΣ ΕΦΗΜΕΡΙΔΑΣ

		ΕΙΔΟΣ ΕΦΗΜΕΡΙΔΑΣ		
		ΜΟΝΟ ΗΜΕΡΗΣΙΕΣ	ΜΟΝΟ ΚΥΡΙΑΚΑΤΙΚΕΣ	ΚΑΙ ΤΙΣ ΔΥΟ
ΦΥΛΟ	ΑΝΔΡΕΣ	70,7%	58,4%	73,1%
	ΓΥΝΑΙΚΕΣ	29,3%	41,6%	26,9%
ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	15-24	16,4%	12,6%	16,1%
	25-34	18,6%	24,3%	25,3%
	35-44	20,8%	25,6%	23,7%
	45-54	23,2%	22,9%	20,1%
	55-64	21,0%	14,6%	14,9%
ΚΟΙΝΩΝΙΚΗ ΤΑΞΗ	ΑΝΩΤΕΡΗ (Α/Β/Σ1)	10,5%	17,5%	19,4%
	ΜΕΣΗ (Σ2)	38,7%	40,1%	37,1%
	ΧΑΜΗΛΗ (Σ/Ε)	50,8%	42,4%	43,5%

Στον πίνακα 2.4.2 βλέπουμε τα ποσοστά ανά γραμμή. Βλέπουμε, δηλαδή, ποιο είδος εφημερίδας επιλέγεται από τους αναγνώστες ανάλογα από το φύλο τους, την ηλικιακή κατηγορία και την κοινωνική τάξη στην οποία ανήκουν. Αυτό που παρατηρείται είναι ότι σε όλα τα δημογραφικά χαρακτηριστικά το μεγαλύτερο ποσοστό προτίμησης εμφανίζεται στις κυριακάτικες εφημερίδες.

Συνολικά, συνεπώς, η μεγαλύτερη μερίδα των αναγνωστών επιλέγουν να διαβάσουν μόνο κυριακάτικες εφημερίδες, ενώ ένα αρκετά μεγάλο ποσοστό επιλέγει να διαβάσει και τα δύο είδη εφημερίδων. Αλλά είναι αισθητά μικρότερο το ποσοστό των αναγνωστών, που ανεξαρτήτως δημογραφικών χαρακτηριστικών επιλέγει να διαβάσει μόνο ημερήσιες.

Πίνακας 2.4.2: ΔΗΜΟΓΡΑΦΙΚΟ ΠΡΟΦΙΛ ΑΝΑ ΔΗΜΟΓΡΑΦΙΚΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ

		ΕΙΔΟΣ ΕΦΗΜΕΡΙΔΑΣ		
		ΜΟΝΟ ΗΜΕΡΗΣΙΕΣ	ΜΟΝΟ ΚΥΡΙΑΚΑΤΙΚΕΣ	ΚΑΙ ΤΙΣ ΔΥΟ
ΦΥΛΟ	ΑΝΔΡΕΣ	15,4%	50,2%	34,4%
	ΓΥΝΑΙΚΕΣ	11,7%	65,2%	23,1%
ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	15-24	16,2%	49,4%	34,4%
	25-34	11,0%	56,7%	32,3%
	35-44	12,0%	58,4%	29,6%
	45-54	14,8%	57,5%	27,8%
	55-64	18,9%	52,1%	29,0%
ΚΟΙΝΩΝΙΚΗ ΤΑΞΗ	ΑΝΩΤΕΡΗ (Α/Β/С1)	8,6%	56,8%	34,5%
	ΜΕΣΗ (С2)	14,0%	57,1%	29,0%
	ΧΑΜΗΛΗ (D/E)	16,3%	53,6%	30,2%

ΚΕΦΑΛΑΙΟ 3

ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ

3.1 Εισαγωγή

Η ανάλυση παραγόντων (factor analysis) είναι μια στατιστική μέθοδος που έχει ως στόχο τη δημιουργία κρυφών, μη παρατηρήσιμων ποσοτήτων που ονομάζονται *παράγοντες*. Το παραγοντικό μοντέλο στηρίζεται στην υπόθεση ότι οι μεταβλητές μπορούν να ομαδοποιηθούν με βάση τις συσχετίσεις που υπάρχουν μεταξύ τους. Έτσι, όλες οι μεταβλητές που βρίσκονται μέσα στην ίδια ομάδα είναι υψηλά συσχετισμένες μεταξύ τους, αλλά έχουν σχετικά μικρές συσχετίσεις με τις μεταβλητές διαφορετικών ομάδων. Συνεπώς, γίνεται αντιληπτό, ότι κάθε ομάδα μεταβλητών αναπαριστά έναν κρυφό παράγοντα, που ευθύνεται για τις παρατηρούμενες συσχετίσεις.

Η παραγοντική ανάλυση είναι μια τεχνική που έχει πολλές ομοιότητες με την τεχνική της ανάλυσης σε κύριες συνιστώσες. Η διαφορά της μεθόδου της παραγοντικής ανάλυσης από αυτήν των κύριων συνιστωσών είναι ότι, ενώ η τελευταία κατασκευάζει έναν ορθογώνιο μετασχηματισμό των μεταβλητών ο οποίος δεν εξαρτάται από το μοντέλο, η παραγοντική ανάλυση βασίζεται σε ένα κατάλληλο στατιστικό μοντέλο.

Επίσης, η παραγοντική ανάλυση εστιάζεται περισσότερο στην εξήγηση της δομής της συνδιακύμανσης των μεταβλητών παρά με την εξήγηση της διακύμανσης, κάτι που συμβαίνει στην ανάλυση σε κύριες συνιστώσες. Οποιαδήποτε διακύμανση δεν εξηγείται από τους κοινούς παράγοντες, θεωρείται ότι προέρχεται από τους όρους των καταλοίπων.

3.2 Το Ορθογώνιο Μοντέλο

Στο ορθογώνιο μοντέλο της παραγοντικής ανάλυσης, το οποίο είναι και το πιο διαδεδομένο, υποθέτουμε ότι οι όποιες συσχετίσεις μεταξύ των μεταβλητών οφείλονται αποκλειστικά στην ύπαρξη κάποιων κοινών παραγόντων τους οποίους δεν ξέρουμε και θέλουμε να εκτιμήσουμε.

Σύμφωνα με το παραγοντικό μοντέλο, θεωρούμε ένα τυχαίο διάνυσμα \mathbf{X} το οποίο μπορεί να παρατηρηθεί με p συνιστώσες, με μέσο μ και πίνακα συνδιακύμανσης Σ . Το \mathbf{X} εξαρτάται γραμμικά από μια σειρά τυχαίων μεταβλητών F_1, F_2, \dots, F_k οι οποίες δεν είναι δυνατόν να παρατηρηθούν. Οι μεταβλητές αυτές ονομάζονται *κοινοί παράγοντες* (common factors). Το \mathbf{X} εξαρτάται επίσης από p πρόσθετες πηγές μεταβλητότητας $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ που ονομάζονται σφάλματα (errors).

Ειδικότερα το μοντέλο της παραγοντικής ανάλυσης γράφεται ως εξής:

$$\mathbf{X}-\boldsymbol{\mu}=\mathbf{LF}+\boldsymbol{\varepsilon} \quad (3.1)$$

όπου

\mathbf{X} είναι το διάνυσμα των αρχικών μεταβλητών μεγέθους $p \times 1$,

$\boldsymbol{\mu}$ είναι το διάνυσμα των μέσων μεγέθους $p \times 1$,

\mathbf{L} είναι ένας πίνακας $p \times k$ όπου το L_{ij} είναι η επιβάρυνση (loading) του παράγοντα L_j στη μεταβλητή X_i . Ονομάζεται *πίνακας των επιβαρύνσεων* (matrix of factor loadings),

\mathbf{F} είναι ένα $k \times 1$ διάνυσμα με τους (κοινούς) παράγοντες (common factors) και

$\boldsymbol{\varepsilon}$ είναι το *σφάλμα* (error) ή *ειδικός παράγοντας* (specific factor). Το σφάλμα ε_i είναι ο μοναδικός παράγοντας της i μεταβλητής και είναι το μέρος της μεταβλητής το οποίο δεν μπορεί να εξηγηθεί από τους παράγοντες (Johnson & Wichern, 1998).

Μπορούμε να υποθέσουμε ότι όλες οι μεταβλητές έχουν μέσο 0, οπότε διάνυσμα $\boldsymbol{\mu}$ δεν χρειάζεται στο παραπάνω μοντέλο. Επίσης, είναι προφανές ότι $k < p$, δηλαδή ο αριθμός των παραγόντων πρέπει να είναι μικρότερος του αριθμού των μεταβλητών, γιατί αλλιώς θα ήταν χωρίς νόημα να γίνει παραγοντική ανάλυση. Σύμφωνα με τα παραπάνω υποθέτουμε ότι κάθε μεταβλητή μπορούμε να τη γράψουμε με τη μορφή

$$X_1 = L_{11} F_1 + L_{12} F_2 + \dots + L_{1k} F_k + \varepsilon_1$$

$$X_2 = L_{21} F_1 + L_{22} F_2 + \dots + L_{2k} F_k + \varepsilon_2$$

...

$$X_p = L_{p1} F_1 + L_{p2} F_2 + \dots + L_{pk} F_k + \varepsilon_p$$

Είναι σημαντικό να σημειωθεί ότι:

- Το παραπάνω μοντέλο, αν και μοιάζει, με ένα γραμμικό μοντέλο παλινδρόμησης, έχει μερικές διαφορές. Πρώτον, τα X_i δεν είναι παρατηρήσεις αλλά μεταβλητές. Δεύτερον, το δεξί μέλος της εξίσωσης δεν είναι παρατηρήσιμο και έτσι πρέπει να εκτιμηθεί.
- Οι παράγοντες F_i μπορούν να γραφούν και αυτοί σαν γραμμικός συνδυασμός των μεταβλητών. Αυτό είναι χρήσιμο να γίνεται, όταν θέλουμε να δημιουργήσουμε νέες μεταβλητές. Αυτοί οι συντελεστές όμως διαφέρουν από τις επιβαρύνσεις. Οι συντελεστές κάθε μεταβλητής όταν εκφράζουμε κάθε παράγοντα σαν γραμμικό συνδυασμό των μεταβλητών καλούνται *συντελεστές των σκορ* (factor scores coefficients).
- Οι παράγοντες έχουν την ίδια διακύμανση. Συνεπώς οι παράγοντες που προκύπτουν δεν είναι απαραίτητως σε μια σειρά.
- Το μοντέλο αυτό προσπαθεί να εκφράσει τις μεταβλητές ως γραμμικό συνδυασμό των παραγόντων.

3.3 Υποθέσεις του Ορθογώνιου Μοντέλου

Προκειμένου να καταστεί δυνατός ο έλεγχος του μοντέλου που εκφράζεται με τη σχέση (3.1), είναι απαραίτητο να γίνουν κάποιες υποθέσεις για τα τυχαία διανύσματα \mathbf{F} και $\boldsymbol{\varepsilon}$. Με τον τρόπο αυτό εισάγουμε σχέσης διακύμανσης που είναι εύκολο να μελετηθούν.

Οι υποθέσεις τις οποίες κάνουμε είναι οι εξής (Johnson, 1998; Πανάρετου & Ξεκαλάκη, 1995):

$$1. E(\mathbf{F}) = 0, \text{Cov}(\mathbf{F}) = E(\mathbf{F}, \mathbf{F}') = \mathbf{I}$$

Υποθέτουμε ότι οι παράγοντες είναι ασυσχέτιστοι μεταξύ τους και έχουν μηδενικές μέσες τιμές. Επίσης, οι παράγοντες είναι ορθογώνιοι μεταξύ τους και εφόσον οι διακυμάνσεις τους είναι ίσες με τη μονάδα, συμπεραίνουμε ότι όλοι οι παράγοντες έχουν την ίδια διακύμανση.

$$2. E(\boldsymbol{\varepsilon}) = 0, \text{Cov}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}') = \boldsymbol{\Psi} \text{ όπου } \boldsymbol{\Psi} \text{ είναι ένας διαγώνιος πίνακας της μορφής}$$

$$\boldsymbol{\Psi} = \text{diag}(\psi_1, \dots, \psi_p)$$

Υποθέτουμε ότι τα σφάλματα είναι ασυσχέτιστα μεταξύ τους και έχουν μηδενικές μέσες τιμές.

$$3. \text{Cov}(\varepsilon_i, F_j) = E(\varepsilon_i, F_j') = 0, \text{ για κάθε } i \neq j.$$

Υποθέτουμε ότι τα σφάλματα και οι παράγοντες είναι ασυσχέτιστοι μεταξύ τους.

Επίσης, υποθέτουμε ότι τα δεδομένα προέρχονται από πολυμεταβλητούς κανονικούς πληθυσμούς. Αυτή η υπόθεση χρησιμοποιείται ως βάση για ελέγχους καλής προσαρμογής και για την εκτίμηση με τη μέθοδο της μέγιστης πιθανοφάνειας. Συνεπώς, μπορεί να αγνοηθεί στην περίπτωση που δουλεύουμε με άλλες μεθόδους εκτίμησης.

Οι υποθέσεις (1) και (2) και η σχέση (3.1) αποτελούν αυτό που ονομάζουμε *ορθογώνιο παραγοντικό μοντέλο* (orthogonal factor model).

Από τις παραπάνω υποθέσεις μπορεί ναδειχθεί ότι:

$$\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}) = \text{Cov}(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon}) = \mathbf{L} \text{Cov}(\mathbf{F})\mathbf{L}' + \text{Cov}(\boldsymbol{\varepsilon}) = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$$

$$\text{Cov}(\mathbf{X}, \mathbf{F}) = E(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon})\mathbf{F}' = \mathbf{L}E(\mathbf{F}, \mathbf{F}') + E(\boldsymbol{\varepsilon}, \mathbf{F}') = \mathbf{L}$$

Συνεπώς βλέπουμε ότι ο πίνακας διακύμανσης μπορεί να διασπαστεί σε δύο μέρη διακύμανσης, το πρώτο είναι το κομμάτι που ερμηνεύουν οι κοινοί παράγοντες και ονομάζεται *εταιρικήτητα* (communality) και το δεύτερο μέρος που οφείλεται στα σφάλματα και άρα το μοντέλο δεν μπορεί να ερμηνεύσει και ονομάζεται *ιδιαιτερότητα* (specificity). Επομένως, έχουμε

$$\text{Var}(X_i) = \text{εταιρικήτητα} + \text{ιδιαιτερότητα}.$$

Δηλαδή,

$$\sigma_{ii} = L_{i1}^2 + L_{i2}^2 + \dots + L_{ik}^2 + \psi_i.$$

όπου αν συμβολίσουμε με h_i^2 την *i*-εταιρικήτητα θα έχουμε τις σχέσεις

$$h_i^2 = L_{i1}^2 + L_{i2}^2 + \dots + L_{ik}^2 \quad (3.2)$$

$$\sigma_{ii} = h_i^2 + \psi_i, \quad i=1, 2, \dots, p \quad (3.3)$$

Επομένως, η *i*-εταιρικήτητα είναι το άθροισμα των τετραγώνων των επιβαρύνσεων των *k* παραγόντων στην *i*-μεταβλητή.

Στην παραγοντική ανάλυση σκοπός μας είναι να εκτιμήσουμε τους πίνακες \mathbf{L} και $\boldsymbol{\Psi}$, να αναπαραστήσουμε δηλαδή τον πίνακα διακύμανσης του πληθυσμού.

Τα βήματα, για να κάνουμε παραγοντική ανάλυση, πρέπει να είναι τα εξής:

- Έλεγχος για το αν υπάρχουν συσχετίσεις ικανοποιητικές για να κάνουμε παραγοντική ανάλυση.
- Εύρεση του αριθμού των παραγόντων και εκτίμηση των παραμέτρων του μοντέλου.
- Περιστροφή του μοντέλου με σκοπό να αυξήσουμε την ερμηνευτική του ικανότητα.

3.4 Έλεγχος συσχετίσεων

Για την παραγοντική ανάλυση είναι σημαντικό να υπάρχουν συσχετίσεις ανάμεσα στις μεταβλητές, καθώς αυτές τις συσχετίσεις θα προσπαθήσουμε να εξηγήσουμε. Βάσει της σημαντικότητας της συσχέτισης, ο υπολογισμός του πίνακα συσχετίσεων των αρχικών μεταβλητών είναι συχνά το πρώτο βήμα της παραγοντικής ανάλυσης. Βάσει της δύναμης της συσχέτισης των αρχικών μεταβλητών θα προσπαθήσουμε:

- Να προσδιορίσουμε το μικρότερο δυνατό αριθμό κοινών παραγόντων, που προσδιορίζουν ή μετρούν τη συσχέτιση μεταξύ των αρχικών μεταβλητών.
- Να εκτιμήσουμε τη σχέση σύνδεσης και τα βάρη των κοινών και των μοναδικών παραγόντων.
- Να επεξηγήσουμε τους κοινούς παράγοντες.
- Να υπολογίσουμε, αν κριθεί απαραίτητο, τις τιμές των παραγόντων.

Συνεπώς, αυτό που μας ενδιαφέρει είναι να υπάρχουν μεγάλες συσχετίσεις τουλάχιστον σε μεγάλο ποσοστό του πίνακα συσχετίσεων. Σε αντίθετη περίπτωση δεν έχει νόημα να συνεχίσουμε. Αν υπάρχουν κάποια ή κάποιες μεταβλητές που είναι ασυσχέτιστες με τις υπόλοιπες, καλό είναι να τις αγνοήσουμε καθώς, επειδή δε σχετίζονται με άλλες, θα προκύψουν από μόνες τους ως ένας ξεχωριστός παράγοντας.

Για να ελέγξουμε τη στατιστική σημαντικότητα ενός δειγματικού συντελεστή συσχέτισης, χρειαζόμαστε κάποιον έλεγχο. Έτσι, για να ελέγξουμε την

$H_0: \rho = 0$ έναντι της εναλλακτικής

$H_1: \rho \neq 0$

Υπολογίζουμε την ελεγχοσυνάρτηση
$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

η οποία ακολουθεί την t κατανομή με $n-2$ βαθμούς ελευθερίας. Συνεπώς, απορρίπτουμε τη μηδενική υπόθεση, όταν $|t| \geq t_{1-\alpha/2, n-2}$ όπου $t_{\alpha, n-2}$ είναι το $\alpha\%$ ποσοστιαίο σημείο από την κατανομή t με $n-2$ βαθμούς ελευθερίας.

Απόρριψη της μηδενικής υπόθεσης σημαίνει ότι υπάρχει μη μηδενική συσχέτιση.

Για να ελέγξουμε αν υπάρχουν συσχετίσεις στα δεδομένα μας, μπορούμε να χρησιμοποιήσουμε τον έλεγχο σφαιρικότητας του **Bartlett** (Bartlett's test of sphericity) (Rencher, 1992). Αν τα

δεδομένα ήταν ασυσχέτιστα, τότε το νέφος των σημείων θα ήταν μια υπερσφαίρα στο πολυεπίπεδο. Ο έλεγχος για έναν πίνακα συσχέτισης ελέγχει την υπόθεση

$H_0: R = I_p$ έναντι της εναλλακτικής

$H_1: R \neq I_p$,

δηλαδή ελέγχουμε την υπόθεση ότι ο πίνακας συσχετίσεων του πληθυσμού είναι ο μοναδιαίος.

Η ελεγχοσυνάρτηση που χρησιμοποιούμε είναι η

$$L = - \left[n - \frac{1}{6(2p+5)} \right] \log |R|$$

η οποία ακολουθεί την χ^2 κατανομή ασυμπτωτικά, με $p(p-1)/2$ βαθμούς ελευθερίας. Θα πρέπει να τονιστεί ότι ο έλεγχος σφαιρικότητας αφορά όλον τον πίνακα και όχι μόνο μεμονωμένα στοιχεία του.

Μερικός συντελεστής συσχέτισης

Ένα μέτρο για να συγκρίνουμε το σχετικό μέγεθος των συντελεστών συσχέτισης σχετικά με τους μερικούς συντελεστές συσχέτισης είναι το Kaiser-Meyer-Olkin (Ηλιόπουλος, 2008) στατιστικό το οποίο υπολογίζεται ως

$$KMO = \frac{\sum_{i=1}^{p-1} \sum_{j=i+1}^p r_{ij}^2}{\sum_{i=1}^{p-1} \sum_{j=i+1}^p r_{ij}^2 + \sum_{i=1}^{p-1} \sum_{j=i+1}^p \alpha_{ij}^2}$$

όπου r_{ij} και α_{ij} είναι οι δειγματικοί συντελεστές συσχέτισης και μερικής συσχέτισης αντίστοιχα. Αν η τιμή KMO είναι μεγάλη, τότε τα δεδομένα μας είναι κατάλληλα για παραγοντική ανάλυση. Τιμές κάτω από 0,5 είναι πολύ κακές τιμές γιατί αποτελούν ένδειξη ότι η παραγοντική ανάλυση δεν θα μας δώσει ικανοποιητικά αποτελέσματα. Στην πράξη τιμές γύρω στο 0,8 θεωρούνται αρκετά καλές για να προχωρήσουμε.

Τέλος, ένα άλλο μέτρο που μας επιτρέπει να εξτάσουμε μια μια τις μεταβλητές και κατά πόσο είναι κατάλληλες, για να χρησιμοποιηθούν στην ανάλυση, είναι το μέτρο της δειγματικής καταλληλότητας (measure of sampling adequacy) (Καρλής, 2005) το οποίο υπολογίζεται για την i μεταβλητή ως

$$MSA = \frac{\sum_j r_{ij}^2}{\sum_j r_{ij}^2 + \sum_j a_{ij}^2}$$

Τιμές κοντά στο 1 είναι ενδείξεις ότι η μεταβλητή είναι πολύ καλή για να χρησιμοποιηθεί στην ανάλυση, σε περίπτωση μικρής τιμής η μεταβλητή δεν έχει λόγο να συμμετάσχει στην παραγοντική ανάλυση. Παρατηρούμε ότι, ενώ το KMO αφορά όλα τα δεδομένα, το MSA υπολογίζεται για κάθε μεταβλητή ξεχωριστά.

3.5 Αριθμός Παραγόντων και Εκτίμηση των Παραγόντων

Πολύ βασικό ερώτημα στην παραγοντική ανάλυση είναι ο καθορισμός του αριθμού των παραγόντων που θα χρησιμοποιηθούν. Ο αριθμός, όμως, δεν είναι γνωστός και υπάρχουν διάφοροι μέθοδοι για να εκτιμηθεί. Για να βρεθεί, λοιπόν, ο αριθμός των παραγόντων, πρέπει να χρησιμοποιηθούν διάφορες τεχνικές. Δηλαδή, οι τιμές των ιδιοτιμών του πίνακα διακύμανσης, τιμές που εξηγούν κάποιο ποσοστό διακύμανσης ή το scree plot (το γράφημα των ιδιοτιμών) ως προς τον αύξοντα αριθμό τους.

Επομένως, κάποιος θα μπορούσε να δουλέψει με διαδοχικά αύξοντα αριθμό παραγόντων και να κρατήσει το μοντέλο με βάση κάποιο κριτήριο προσαρμογής. Τέτοια κριτήρια είναι:

- Από τον πίνακα των επιβαρύνσεων μπορεί κάποιος να εκτιμήσει τον πίνακα Σ . Οι αποκλίσεις του πραγματικού πίνακα με τον εκτιμημένο (reproduced matrix) θα πρέπει να είναι μικρές.
- Έλεγχος λόγου πιθανοφανειών, αν οι εκτιμήσεις έχουν γίνει με τη μέθοδο μέγιστης πιθανοφάνειας. Τέτοιοι έλεγχοι στηρίζονται σε υποθέσεις για την κατανομή του πληθυσμού.

Οι δύο βασικές μέθοδοι εκτίμησης των παραγόντων που χρησιμοποιούνται στην πράξη είναι η μέθοδος των κύριων συνιστωσών και η μέθοδος της μέγιστης πιθανοφάνειας.

3.5.1 Εκτίμηση με τη μέθοδο Κύριων Συνιστωσών

Μια πολύ βασική μέθοδος εκτίμησης που χρησιμοποιείται στην ανάλυση είναι η μέθοδος των κύριων συνιστωσών. Η εκτίμηση με τη μέθοδο των κύριων συνιστωσών βασίζεται στην φασματική ανάλυση του πίνακα διακύμανσης (συσχέτισης) Σ . Ο πίνακας Σ αποτελείται από ζεύγη ιδιοτιμών-ιδιοδιανυσμάτων $(\lambda_i, \mathbf{e}_i)$ με $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Οπότε

$$\Sigma = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \dots + \lambda_p \mathbf{e}_p \mathbf{e}_p' = \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1 & & & \\ & \sqrt{\lambda_2} \mathbf{e}_2 & & \\ & & \ddots & \\ & & & \sqrt{\lambda_p} \mathbf{e}_p \end{bmatrix} \quad (3.4)$$

Η παραπάνω αναπαράσταση αφορά το μοντέλο της παραγοντικής ανάλυσης, το οποίο έχει τόσους παράγοντες όσες και μεταβλητές ($m=p$) και ειδική διακύμανση $\psi_i=0$ για όλα τα i . Τα φορτία του j κοινού παράγοντα ισούνται με τους συντελεστές της j κύριας συνιστώσας πολλαπλασιασμένους με $\sqrt{\lambda_j} \mathbf{e}_j$. Συνεπώς, μπορούμε να γράψουμε ότι

$$\underset{(p \times p)}{\Sigma} = \underset{(p \times p)}{\mathbf{L}} \underset{(p \times p)}{\mathbf{L}'} + \underset{(p \times p)}{\mathbf{0}} = \underset{(p \times p)}{\mathbf{L}\mathbf{L}'} \quad (3.5)$$

Όμως, η αναπαράσταση του πίνακα Σ στη σχέση (3.5) δεν είναι ιδιαίτερα χρήσιμη γιατί δημιουργεί τόσους κοινούς παράγοντες όσες είναι οι μεταβλητές και δεν επιτρέπει καμία μεταβολή των ειδικών παραγόντων ϵ της σχέσης (3.1). Προτιμούνται μοντέλα τα οποία εξηγούν

τη διακύμανση χρησιμοποιώντας λίγους παράγοντες. Μία προσέγγιση είναι, όταν οι $p-m$ τελευταίες ιδιοτιμές είναι μικρές, να αγνοούμε τη συνεισφορά των $\lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \dots + \lambda_p \mathbf{e}_p \mathbf{e}_p'$ στον Σ στη σχέση (3.4). Αγνοώντας αυτή τη συνεισφορά, καταλήγουμε στον πίνακα

$$\Sigma = \underset{(p \times m)}{\mathbf{L}} \underset{(p \times m)}{\mathbf{L}'} \quad (3.6)$$

Η αναπαράσταση (3.6) υποθέτει ότι οι ειδικοί παράγοντες ε στη σχέση (3.1) είναι αμελητέας σημαντικότητας και μπορούν, επίσης, να αγνοηθούν στην παραγοντοποίηση του πίνακα Σ . Αν συμπεριληφθούν οι ειδικοί παράγοντες στο μοντέλο, οι διακυμάνσεις τους θα θεωρηθούν ότι είναι τα διαγώνια στοιχεία του $\Sigma - \mathbf{L}\mathbf{L}'$. Θεωρούμε τότε την προσέγγιση (Johnson, 1998; Πανάρετου & Ξεκαλάκη, 1995):

$$\Sigma \approx \mathbf{L}\mathbf{L}' + \Psi$$

$$= \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1 & \sqrt{\lambda_2} \mathbf{e}_2 & \dots & \sqrt{\lambda_m} \mathbf{e}_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1' \\ \sqrt{\lambda_2} \mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_m} \mathbf{e}_m' \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{bmatrix} \quad (3.7)$$

όπου $\psi_i = \sigma_{ii} - \sum_{j=1}^m l_{ij}^2$ για i από $1, 2, \dots, p$.

Η αναπαράσταση της σχέσης (3.7), όταν εφαρμόζεται στον δειγματικό πίνακα διακύμανσης ή στον δειγματικό πίνακα συσχέτισης, είναι γνωστή ως η λύση των κύριων συνιστωσών.

Συνεπώς, η παραγοντική ανάλυση με τη μέθοδο των κύριων συνιστωσών για τον δειγματικό πίνακα διακύμανσης είναι η εξής:

Έστω $(\hat{\lambda}_1, \hat{\mathbf{e}}_1), (\hat{\lambda}_2, \hat{\mathbf{e}}_2), \dots, (\hat{\lambda}_p, \hat{\mathbf{e}}_p)$, όπου $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$. Τότε στο μοντέλο με $m < p$ αριθμό των κοινών παραγόντων, εκτιμούμε τον πίνακα των εκτιμώμενων φορτίων των παραγόντων $\{\tilde{l}_{ij}\}$, ο οποίος δίνεται από

$$\hat{\mathbf{L}} = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1 & \sqrt{\hat{\lambda}_2} \hat{\mathbf{e}}_2 & \dots & \sqrt{\hat{\lambda}_m} \hat{\mathbf{e}}_m \end{bmatrix} \quad (3.8)$$

και τις ειδικές διασπορές $\tilde{\psi}_{ii} = s_{ii} - \sum_{j=1}^m \tilde{l}_{ij}^2$ (3.9)

Οι εταιρικότητες υπολογίζονται ως εξής

$$\tilde{h}_i^2 = \tilde{l}_{i1}^2 + \tilde{l}_{i2}^2 + \dots + \tilde{l}_{im}^2$$

3.5.2 Εκτίμηση με τη μέθοδο Μέγιστης Πιθανοφάνειας

Για να χρησιμοποιήσουμε τη μέθοδο μέγιστης πιθανοφάνειας υποθέτουμε ότι οι κατανομές των παραγόντων και των σφαλμάτων ακολουθούν πολυμεταβλητή κανονική κατανομή με σκοπό να

βρούμε τους εκτιμητές μέγιστης πιθανοφάνειας των \mathbf{L} και Ψ (Johnson & Wichern, 1998). Συγκεκριμένα αν στο μοντέλο $\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon}$ υποθέσουμε ότι $\boldsymbol{\varepsilon} \sim N_p(\mathbf{0}, \Psi)$ και $\mathbf{F} \sim N_k(\mathbf{0}, \mathbf{I})$, τότε προκύπτει πως, $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \mathbf{L}\mathbf{L}' + \Psi)$.

Θέτοντας στη συνάρτηση πιθανοφάνειας $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \Psi$, προσπαθούμε να τη μεγιστοποιήσουμε ως προς \mathbf{L} και Ψ . Επειδή όμως ο \mathbf{L} δεν είναι μοναδικός, πρέπει να εισάγουμε κάποιον επιπλέον περιορισμό έτσι ώστε να πάρουμε μοναδική λύση. Ο συνήθης περιορισμός που τίθεται είναι $\mathbf{L}'\Psi^{-1}\mathbf{L} = \text{διαγώνιος}$.

Η πολυμεταβλητή κανονική κατανομή έχει $p + p(p+1)/2$ παραμέτρους. Με την αντικατάσταση $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \Psi$ και λόγω του περιορισμού (3.4) έχουμε $p + pm + p - m(m-1)/2$ παραμέτρους. Θα πρέπει λοιπόν $p + p(p+1)/2 \geq p + pm + p - m(m-1)/2 \Leftrightarrow \frac{(p-m)^2 - (p+m)}{2} \geq 0$.

Υπό κανονικότητα μπορούμε να ελέγξουμε και την υπόθεση των m παραμέτρων,

$$H_0: \boldsymbol{\Sigma} = \underset{p \times p}{\mathbf{L}} \underset{p \times m}{\mathbf{L}'} + \underset{p \times p}{\Psi}$$

χρησιμοποιώντας τον έλεγχο λόγου πιθανοφανειών. Ασυμπτωτικά, και υπό την υπόθεση H_0

$$\text{έχουμε} \quad 2 \log \frac{f(x/\boldsymbol{\Sigma} = \mathbf{S}_n)}{f(x/\boldsymbol{\Sigma} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi})} \sim X^2_{[(p-m)^2 - (p+m)]/2}$$

όπου $\hat{\mathbf{L}}, \hat{\Psi} := \mathbf{S}_n - \hat{\mathbf{L}}\hat{\mathbf{L}}'$ είναι οι εκτιμήτριες μέγιστης πιθανοφάνειας των \mathbf{L} και Ψ .

3.6 Περιστροφή

Με την περιστροφή (rotation) των παραγόντων απλοποιείται η ερμηνεία των παραγόντων (I. Πανάρετος & E. Ξεκαλάκη, 1995). Με την περιστροφή δεν αλλάζουν κάποια από τα χαρακτηριστικά του μοντέλου, όπως η καλή του προσαρμοστικότητα και το ποσοστό της διακύμανσης που ερμηνεύει το μοντέλο, παρά μόνο οι τιμές των επιβαρύνσεων.

Αν \mathbf{L} είναι ένας πίνακας που περιέχει τις επιβαρύνσεις και \mathbf{G} ένας ορθογώνιος πίνακας, τότε ισχύει πως $\mathbf{L}\mathbf{G}(\mathbf{L}\mathbf{G})' = \mathbf{L}\mathbf{G}\mathbf{G}'\mathbf{L}' = \mathbf{L}\mathbf{L}'$ και επομένως και ο πίνακας $\mathbf{L}\mathbf{G}$ μπορεί να θεωρηθεί ως ένας πίνακας επιβαρύνσεων. Μαθηματικά ο πίνακας \mathbf{G} ορίζει έναν ορθογώνιο μετασχηματισμό (Δ.Καρλής, 2005)

Κάνοντας, λοιπόν, την περιστροφή, ελπίζουμε ότι οι επιβαρύνσεις κάποιων παραγόντων θα είναι μεγάλες σε απόλυτη κλίμακα μόνο για κάποιες από τις μεταβλητές και έτσι, βλέποντας ποιες μεταβλητές εξαρτώνται με ποιους παράγοντες, να μπορέσουμε να δώσουμε ερμηνεία σε αυτούς. Οι βασικές μέθοδοι περιστροφής είναι:

- Varimax: Ελαχιστοποιεί τον αριθμό των μεταβλητών που έχουν μεγάλες επιβαρύνσεις για κάθε παράγοντα.
- Quartimax: Ελαχιστοποιεί τον αριθμό των παραγόντων που εξηγούν μια μεταβλητή.
- Equimax: Συνδυασμός των varimax και quartimax.

- Oblique: Μη ορθογώνια περιστροφή, οι άξονες που προκύπτουν δεν είναι πια ορθογώνιοι και άρα οι παράγοντες δεν είναι ανεξάρτητοι.

3.7 Υπολογισμός των Σκορ των Παραγόντων

Σκοπός της παραγοντικής ανάλυσης είναι να μείωση των αριθμό των αρχικών μεταβλητών και για να επιτευχθεί αυτό δημιουργούμε καινούριες μεταβλητές, τους παράγοντες, ως γραμμικό συνδυασμό των αρχικών μεταβλητών. Κάθε παράγοντας τότε θα έχει τη μορφή:

$$\begin{aligned} F_1 &= \alpha_{11}X_1 + \alpha_{12}X_2 + \dots + \alpha_{1p}X_p \\ F_2 &= \alpha_{21}X_1 + \alpha_{22}X_2 + \dots + \alpha_{2p}X_p \\ &\dots \\ F_k &= \alpha_{k1}X_1 + \alpha_{k2}X_2 + \dots + \alpha_{kp}X_p \end{aligned}$$

Οι συντελεστές α_{ij} είναι το σκορ της μεταβλητής X_j στον παράγοντα F_i και δεν πρέπει να συγχέονται με τις επιβαρύνσεις. Τα σκορ αυτά είναι εκτιμητές του μη παρατηρήσιμου, τυχαίου διανύσματος των παραγόντων \mathbf{F}_i , όπου $i=1,2,\dots,k$.

Τυπικά, περισσότερο οι εκτιμητές των περιστραμμένων (rotated) επιβαρύνσεων, παρά οι αρχικά εκτιμώμενες επιβαρύνσεις, χρησιμοποιούνται για να υπολογιστούν τα σκορ των παραγόντων (Johnson & Wichern, 1998).

Υπάρχουν διάφορες μέθοδοι για την εκτίμηση των τιμών των παραγόντων. Οι μέθοδοι που προσφέρονται είναι οι εξής:

- Μέθοδος Παλινδρόμησης (regression). Το διάνυσμα F των καινούριων μεταβλητών υπολογίζεται ως εξής

$$\mathbf{F} = (\hat{\mathbf{L}}' \hat{\mathbf{L}})^{-1} \hat{\mathbf{L}}' \mathbf{X} .$$

Η μέθοδος αυτή στηρίζεται στη μέθοδο των ελαχίστων τετραγώνων ανάμεσα στις πραγματικές τιμές και αυτές που το παραγοντικό μοντέλο προβλέπει (Johnson & Wichern, 1998; Καρλής, 2005).

- Μέθοδος Bartlett. Η μέθοδος αυτή χρησιμοποιεί γενικευμένα ελάχιστα τετράγωνα, καθώς η διακύμανση δεν είναι ίδια για όλες τις παρατηρήσεις. Επομένως εκτιμά τους παράγοντες ως εξής

$$\mathbf{F} = (\hat{\mathbf{L}}' \hat{\Psi} \hat{\mathbf{L}})^{-1} \hat{\mathbf{L}}' \hat{\Psi}^{-1} \mathbf{X} .$$

- Μέθοδος Anderson. Η μέθοδος αυτή χρησιμοποιεί τον τύπο

$$\mathbf{F} = (\hat{\mathbf{L}}' \hat{\Psi} \hat{\mathbf{L}})^{-1} (\mathbf{I} + \hat{\mathbf{L}}' \hat{\Psi}^{-1} \mathbf{X})^{-1/2} \hat{\mathbf{L}}' \hat{\Psi}^{-1} \mathbf{X} .$$

3.8 Εφαρμογή των Μεθόδων

Παραγοντική ανάλυση για τη στάση των ερωτώμενων απέναντι στις εφημερίδες με τη μέθοδο των κυρίων συνιστωσών και της μέγιστης πιθανοφάνειας.

Για την εφαρμογή της παραγοντικής ανάλυσης για τη στάση των ερωτώμενων απέναντι στις εφημερίδες χρησιμοποιήθηκαν 17 φράσεις/ προτάσεις σχετικά με τα έντυπα και την ενημέρωση που βρίσκονται στο κομμάτι του ερωτηματολογίου που αναφέρονται ο τρόπος ζωής και οι συνήθειες. Οι ερωτώμενοι απάντησαν στο κατά πόσο συμφωνούν με αυτές τις φράσεις/ προτάσεις σε μια κλίμακα από το 1 έως το 5, όπου 1 είναι το συμφωνώ απόλυτα και 5 είναι το διαφωνώ απόλυτα. Οι μεταβλητές που θα μας απασχολήσουν εμφανίζονται στον πίνακα 3.1.

Πίνακας 3.1

ΜΕΤΑΒΛΗΤΕΣ			
1	ΔΕΝ ΕΝΔΙΑΦΕΡΟΜΑΙ ΓΙΑ ΤΙΣ ΠΟΛΙΤΙΚΕΣ ΕΙΔΗΣΕΙΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	10	ΔΙΑΒΑΖΩ ΤΙΣ ΟΙΚΟΝΟΜΙΚΕΣ ΣΕΛΙΔΕΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ
2	ΤΑ ΕΝΘΕΤΑ ΚΑΝΟΥΝ ΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΠΙΟ ΕΝΔΙΑΦΕΡΟΥΣΕΣ	11	Η ΕΦΗΜΕΡΙΔΑ ΑΠΑΙΤΕΙ ΧΡΟΝΟ ΚΑΙ ΑΥΤΟΣΥΓΚΕΝΤΡΩΣΗ, ΕΝΩ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΕΝΗΜΕΡΩΝΕΣΑΙ ΓΡΗΓΟΡΑ ΚΑΙ ΧΩΡΙΣ ΚΟΠΟ
3	ΔΕΝ ΘΑ ΑΛΛΑΖΑ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΠΟΥ ΔΙΑΒΑΖΩ	12	ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΑΝΑΛΟΓΑ ΜΕ ΤΑ ΕΝΘΕΤΑ ΠΕΡΙΟΔΙΚΑ ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΟΥΝ
4	ΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΗΜΕΡΕΣ ΔΙΑΒΑΖΩ ΕΦΗΜΕΡΙΔΕΣ	13	ΔΕΝ ΑΛΛΑΖΩ ΕΥΚΟΛΑ ΕΦΗΜΕΡΙΔΑ ΓΙΑΤΙ ΜΕ ΕΝΔΙΑΦΕΡΕΙ Η ΑΠΟΨΗ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΗΜΟΣΙΟΓΡΑΦΩΝ
5	ΤΑ ΕΝΘΕΤΑ ΤΩΝ ΕΦΗΜΕΡΙΔΩΝ ΠΡΑΓΜΑΤΙΚΑ ΑΞΙΖΟΥΝ	14	ΣΥΧΝΑ ΣΥΜΜΕΤΕΧΩ ΣΕ ΔΙΑΓΩΝΙΣΜΟΥΣ ΕΦΗΜΕΡΙΔΩΝ ΚΑΙ ΠΕΡΙΟΔΙΚΩΝ
6	ΠΙΣΤΕΥΩ ΠΕΡΙΣΣΟΤΕΡΟ ΤΙΣ ΤΟΠΙΚΕΣ ΕΦΗΜΕΡΙΔΕΣ ΠΑΡΑ ΤΙΣ ΠΑΝΕΛΛΑΔΙΚΗΣ ΚΑΛΥΨΗΣ	15	ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ-ΠΕΡΙΟΔΙΚΟ ΑΝΑΛΟΓΑ ΜΕ ΤΟ ΔΩΡΟ-DVD-CD-ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΕΙ
7	ΔΙΑΒΑΖΩ ΚΑΘΗΜΕΡΙΝΕΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΔΙΑΣΚΕΔΑΣΗ ΚΑΙ ΟΧΙ ΓΙΑ ΕΝΗΜΕΡΩΣΗ	16	ΟΤΑΝ ΧΡΕΙΑΖΟΜΑΙ ΜΙΑ ΠΛΗΡΟΦΟΡΙΑ, ΤΟ ΠΡΩΤΟ ΜΕΡΟΣ ΠΟΥ ΚΟΙΤΑΖΩ ΕΙΝΑΙ ΤΟ INTERNET
8	ΒΑΣΙΖΟΜΑΙ ΣΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΤΗΝ ΕΝΗΜΕΡΩΣΗ ΜΟΥ	17	ΒΑΣΙΖΟΜΑΙ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΓΙΑ ΝΑ ΕΝΗΜΕΡΩΝΟΜΑΙ
9	ΔΕΝ ΜΠΟΡΩ ΝΑ ΑΝΤΙΣΤΑΘΩ ΣΤΟ ΝΑ ΑΓΟΡΑΖΩ ΠΕΡΙΟΔΙΚΑ		

3.8.1 Καταλληλότητα των δεδομένων

Αρχικά εξετάζουμε περιγραφικά τα δεδομένα μας, όπως φαίνεται στον παρακάτω πίνακα 3.2. Στον πίνακα αυτόν μπορεί να δει κανείς τη μέση τιμή, την τυπική απόκλιση και τον αριθμό των παρατηρήσεων για κάθε μεταβλητή που συμμετέχει στην ανάλυση.

Πίνακας 3.2: Περιγραφικά στατιστικά για τις μεταβλητές

	Mean	Std. Deviation	Analysis N
ΔΕΝ ΕΝΔΙΑΦΕΡΟΜΑΙ ΓΙΑ ΤΙΣ ΠΟΛΙΤΙΚΕΣ ΕΙΔΗΣΕΙΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	2,92	1,218	6549
ΤΑ ΕΝΘΕΤΑ ΚΑΝΟΥΝ ΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΠΙΟ ΕΝΔΙΑΦΕΡΟΥΣΕΣ	2,66	1,104	6549
ΔΕΝ ΘΑ ΑΛΛΑΞΑ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΠΟΥ ΔΙΑΒΑΖΩ	2,78	1,131	6549
ΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΗΜΕΡΕΣ ΔΙΑΒΑΖΩ ΕΦΗΜΕΡΙΔΕΣ	2,98	1,141	6549
ΤΑ ΕΝΘΕΤΑ ΤΩΝ ΕΦΗΜΕΡΙΔΩΝ ΠΡΑΓΜΑΤΙΚΑ ΞΕΙΖΟΥΝ	2,66	1,077	6549
ΠΙΣΤΕΥΩ ΠΕΡΙΣΣΟΤΕΡΟ ΤΙΣ ΤΟΠΙΚΕΣ ΕΦΗΜΕΡΙΔΕΣ ΠΑΡΑ ΤΙΣ ΠΑΝΕΛΛΑΔΙΚΗΣ ΚΑΛΥΨΗΣ	2,91	1,122	6549
ΔΙΑΒΑΖΩ ΚΑΘΗΜΕΡΙΝΕΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΔΙΑΣΚΕΔΑΣΗ ΚΑΙ ΟΧΙ ΓΙΑ ΕΝΗΜΕΡΩΣΗ	3,11	1,165	6549
ΒΑΣΙΖΟΜΑΙ ΣΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΤΗΝ ΕΝΗΜΕΡΩΣΗ ΜΟΥ	2,93	1,151	6549
ΔΕΝ ΜΠΟΡΩ ΝΑ ΑΝΤΙΣΤΑΘΩ ΣΤΟ ΝΑ ΑΓΟΡΑΣΩ ΠΕΡΙΟΔΙΚΑ	3,11	1,183	6549
ΔΙΑΒΑΖΩ ΤΙΣ ΟΙΚΟΝΟΜΙΚΕΣ ΣΒΛΙΔΕΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	2,81	1,172	6549
Η ΕΦΗΜΕΡΙΔΑ ΑΠΑΙΤΕΙ ΧΡΟΝΟ ΚΑΙ ΑΥΤΟΣΥΓΚΕΝΤΡΩΣΗ, ΕΝΩ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΕΝΗΜΕΡΩΝΕΣΑΙ ΓΡΗΓΟΡΑ ΚΑΙ ΧΩΡΙΣ ΚΟΠΟ	2,59	1,098	6549
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΑΝΑΛΟΓΑ ΜΕ ΤΑ ΕΝΘΕΤΑ ΠΕΡΙΟΔΙΚΑ ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΟΥΝ	3,01	1,166	6549
ΔΕΝ ΑΛΛΑΖΩ ΕΥΚΟΛΑ ΕΦΗΜΕΡΙΔΑ ΓΙΑΤΙ ΜΕ ΕΝΔΙΑΦΕΡΕΙ Η ΑΠΟΨΗ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΗΜΟΣΙΟΓΡΑΦΩΝ	2,82	1,156	6549
ΣΥΧΝΑ ΣΥΜΜΕΤΕΧΩ ΣΕ ΔΙΑΓΩΝΙΣΜΟΥΣ ΕΦΗΜΕΡΙΔΩΝ ΚΑΙ ΠΕΡΙΟΔΙΚΩΝ	3,25	1,261	6549
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΠΕΡΙΟΔΙΚΟ ΑΝΑΛΟΓΑ ΜΕ ΤΟ ΔΩΡΟ DVD-CD-ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΕΙ	3,04	1,198	6549
ΟΤΑΝ ΧΡΕΙΑΖΟΜΑΙ ΜΙΑ ΠΛΗΡΟΦΟΡΙΑ, ΤΟ ΠΡΩΤΟ ΜΕΡΟΣ ΠΟΥ ΚΟΙΤΑΣΩ ΕΙΝΑΙ ΤΟ INTERNET	2,89	1,271	6549
ΒΑΣΙΖΟΜΑΙ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΓΙΑ ΝΑ ΕΝΗΜΕΡΩΝΟΜΑΙ	2,58	1,072	6549

Βασική προϋπόθεση για την εφαρμογή της παραγοντικής ανάλυσης είναι η ύπαρξη συσχετίσεων μεταξύ των μεταβλητών. Επειδή οι μεταβλητές που θα χρησιμοποιηθούν είναι διατάξιμες (ordinal), μπορούμε να τις διαχειριστούμε ως συνεχή και να χρησιμοποιήσουμε τον συντελεστή συσχέτισης r του Pearson. Ο έλεγχος του Pearson παίρνει τιμές στο διάστημα $[-1,1]$. Θετικές τιμές του ελέγχου αφορούν τη θετική συσχέτιση, και αντίστοιχα, αρνητικές τιμές την αρνητική συσχέτιση μεταξύ των μεταβλητών.

Για να υπολογίσουμε το r υπολογίζουμε την ελεγχοσυνάρτηση

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Ο πίνακας 3.3 είναι ένας συμμετρικός πίνακας στον οποίο θα απεικονίζονται οι τιμές των συσχετίσεων για κάθε ζευγάρι μεταβλητών. Στα διαγώνια κελιά του πίνακα απεικονίζεται η συσχέτιση κάθε μεταβλητής με τον εαυτό της, για αυτό τον λόγο η τιμή των διαγώνιων κελιών του πίνακα είναι 1. Κάτω από κάθε συσχέτιση υπάρχει το p-value (Sig.) για τον έλεγχο της μηδενικής υπόθεσης, ότι η τιμή του συντελεστή συσχέτισης στον πληθυσμό είναι 0, έναντι της εναλλακτικής ότι είναι διάφορη του 0. Θα μας ορίσει, δηλαδή, αν η συσχέτιση μεταξύ των μεταβλητών είναι στατιστικά σημαντική ή όχι.

Ο ένας αστερίσκος στην τιμή της συσχέτισης είναι ένδειξη ότι η συγκεκριμένη συσχέτιση είναι σημαντική σε επίπεδο σημαντικότητας 0,05 ($\alpha=5\%$), ενώ οι δύο αστερίσκοι είναι ένδειξη ότι η τιμή της συσχέτισης είναι στατιστικά σημαντική σε επίπεδο σημαντικότητας 0,01 ($\alpha=1\%$). Κοιτάζοντας, λοιπόν, τον παρακάτω πίνακα συσχετίσεων, παρατηρούμε μια έντονη συσχέτιση μεταξύ των μεταβλητών. Βλέπουμε ότι στις περισσότερες περιπτώσεις η τιμή του Sig. είναι μικρότερη του 0,01 και άρα είναι στατιστικά σημαντικές οι συσχετίσεις.

Πίνακας 3.3: Πίνακας συσχετίσεων για τα δεδομένα

Correlations																					
Spearman's rho	ΣΠΑΝΙΟΣ ΠΡΟΣΧΩ ΤΙΣ ΔΙΑΦΗΜΙΣΕΙΣ ΣΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΚΑΙ ΤΑ ΠΕΡΙΟΔΙΚΑ	Κοrelation Coefficient	1,000	,161**	,078**	,112**	,048**	,101**	,161**	,140**	,130**	,039**	,080**	,150**	,119**	,087**	,108**	,023*	,135**		
		Sig. (2-tailed)		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,028	,000		
		N	10272	9449	9694	8942	9125	9690	9367	8953	9171	9259	8876	9953	9099	9135	9102	9131	9255	9991	
	ΔΕΝ ΕΝΔΙΑΦΕΡΟΜΑΙ ΓΙΑ ΤΙΣ ΠΟΛΥΤΙΚΕΣ ΕΠΙΧΕΙΡΗΣΕΙΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	Correlation Coefficient	,161**	1,000	,120**	,015	,004	,108**	,149**	,212**	,058**	,196**	,208**	,126**	,208**	,258**	,238**	,158**	,135**		
		Sig. (2-tailed)	,000		,000	,171	,680	,000	,000	,000	,000	,000	,000	,000	,000	,039	,000	,000	,000		
		N	9449	9781	9366	8759	8911	9318	8774	8930	8758	8934	8968	8724	9545	8834	8963	8798	8913	8991	9539
	ΤΑ ΕΝΘΕΤΑ ΚΑΝΟΥΝ ΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΠΙΟ ΕΝΔΙΑΦΕΡΟΥΣΕΣ	Correlation Coefficient	,078**	,120**	1,000	,196**	,204**	,414**	,125**	,107**	,163**	,213**	,227**	,151**	,253**	,179**	,177**	,241**	,218**	,082**	
		Sig. (2-tailed)	,000	,000		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000		
		N	9694	9366	10205	9366	8879	9786	8924	9182	9192	8818	8948	9133	9138	9156	9134	9220	9947		
	ΔΕΝ ΘΑ ΑΛΛΑΞΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΠΟΥ ΔΙΑΒΑΖΩ	Correlation Coefficient	,112**	,015	,186**	1,000	,345**	,224**	,230**	,174**	,318**	,153**	,346**	,064**	,066**	,387**	,079**	,049**	,169**	,098**	
		Sig. (2-tailed)	,000	,171	,000		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000		
		N	8942	8759	8879	9179	8679	8882	8726	8584	8686	8470	8537	8979	8569	8693	8390	8529	8471	8978	
	ΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΗΜΕΡΕΣ ΔΙΑΒΑΖΩ ΕΦΗΜΕΡΙΔΕΣ	Correlation Coefficient	,048**	,004	,204**	,345**	1,000	,161**	,223**	,226**	,376**	,217**	,358**	,042**	,162**	,346**	,196**	,277**	,037**	,037**	
		Sig. (2-tailed)	,000	,680	,000	,000		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000		
		N	9125	8811	9152	8679	9453	9093	8861	8722	8887	8735	8608	9230	8808	8880	8716	8780	8813	9230	
	ΤΑ ΕΝΘΕΤΑ ΤΩΝ ΕΦΗΜΕΡΙΔΩΝ ΠΡΑΓΜΑΤΙΚΑ ΑΞΙΖΟΥΝ	Correlation Coefficient	,101**	,108**	,414**	,224**	,181**	1,000	,136**	,129**	,202**	,245**	,223**	,108**	,136**	,219**	,166**	,219**	,187**	,022**	
		Sig. (2-tailed)	,000	,000	,000	,000	,000		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000		
		N	9690	9318	9786	8882	9093	10185	9353	8915	9185	8834	9922	9103	9089	9115	9108	9195	9950		
	ΠΙΣΤΕΥΩ ΠΕΡΙΣΣΟΤΕΡΟ ΤΙΣ ΤΟΠΙΚΕΣ ΕΦΗΜΕΡΙΔΕΣ ΠΑΡΑ ΤΙΣ ΠΑΝΕΛΛΑΔΙΚΗΣ ΚΑΛΥΨΗΣ	Correlation Coefficient	,161**	,149**	,125**	,230**	,223**	,136**	1,000	,263**	,268**	,167**	,196**	,097**	,192**	,231**	,207**	,188**	,118**	,120**	
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000		
		N	9367	9080	9350	8726	8861	9353	9729	8784	8931	8657	8663	9503	8841	8875	8832	8834	9523		
	ΔΙΑΒΑΖΩ ΚΑΘΗΜΕΡΙΝΕΣ ΕΦΗΜΕΡΙΔΕΣ ΠΑΡΑ ΤΙΣ ΔΙΑΣΚΕΔΑΣΗ ΚΑΙ ΟΧΙ ΓΙΑ ΕΝΗΜΕΡΩΣΗ	Correlation Coefficient	,140**	,212**	,107**	,174**	,226**	,129**	,263**	1,000	,329**	,323**	,172**	,017	,249**	,135**	,274**	,154**	,086**		
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000		,000	,000	,000	,116	,000	,000	,000	,000			
		N	8953	8774	8924	8584	8722	8915	8784	9227	8781	8630	8575	9025	8652	8695	8545	8616	9033		
	ΒΑΣΙΖΟΜΑΙ ΣΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΤΗΝ ΕΝΗΜΕΡΩΣΗ ΜΟΥ	Correlation Coefficient	,130**	,059**	,163**	,318**	,376**	,202**	,268**	,329**	1,000	,262**	,359**	,030**	,153**	,282**	,188**	,155**	,163**	,030**	
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000		,000	,000	,004	,000	,000	,000	,000	,004		
		N	9171	8930	9182	8686	8887	9165	8761	9501	8809	8669	8274	8785	8808	8708	8780	8738	9295		
	ΔΕΝ ΜΠΟΡΩ ΝΑ ΑΝΤΙΣΤΑΘΩ ΣΤΟ ΝΑ ΑΓΟΡΑΣΩ ΠΕΡΙΟΔΙΚΑ	Correlation Coefficient	,039**	,196**	,213**	,153**	,217**	,245**	,167**	,323**	,262**	1,000	,187**	,-0,040**	,257**	,104**	,319**	,285**	,227**	,005	
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000		,000	,000	,000	,000	,000	,000	,640		
		N	9259	8868	9192	8470	8735	9185	8957	8630	8909	9616	8498	9363	8779	8700	8864	8783	8914	9418	
	ΔΙΑΒΑΖΩ ΤΙΣ ΟΙΚΟΝΟΜΙΚΕΣ ΣΕΛΙΔΕΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	Correlation Coefficient	,080**	,126**	,227**	,346**	,358**	,252**	,196**	,172**	,359**	,187**	1,000	,001	,091**	,280**	,075**	,087**	,161**	,060**	
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000		,931	,000	,000	,000	,000	,000		
		N	8876	8724	8818	8537	8608	8834	8683	8575	8669	8498	9114	8915	8540	8603	8396	8520	8423	8915	
	Η ΕΦΗΜΕΡΙΔΑ ΑΠΑΙΤΕΙ ΧΡΟΝΟ ΚΑΙ ΑΥΤΟΣΥΓΚΕΝΤΡΩΣΗ. ΕΝΟ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΕΝΗΜΕΡΩΝΕΣΑΙ ΓΡΗΓΟΡΑ ΚΑΙ ΧΩΡΙΣ ΚΟΠΟ	Correlation Coefficient	,150**	,208**	,151**	,064**	,-0,042**	,108**	,097**	,017	,-0,030**	,-0,040**	,001	1,000	,170**	,123**	,-0,009	,109**	,067**	,333**	
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,116	,004	,000	,931		,000	,000	,410	,000	,000		
		N	9953	9545	9948	8979	9230	9922	9503	9525	9274	9363	8915	10685	9263	9259	9366	9271	9484	10411	
	ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΑΝΑΛΟΓΑ ΜΕ ΤΑ ΕΝΘΕΤΑ ΠΕΡΙΟΔΙΚΑ ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΟΥΝ	Correlation Coefficient	,119**	,248**	,253**	,066**	,162**	,223**	,192**	,249**	,153**	,257**	,091**	,170**	1,000	,212**	,437**	,496**	,191**	,022**	
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000		,000	,000	,000	,000		
		N	9099	8934	9133	8568	8808	9103	8841	8652	8785	8779	8540	9263	9446	8924	8800	8967	8713	9242	
	ΔΕΝ ΑΛΛΑΞΩ ΕΥΚΟΛΑ ΕΦΗΜΕΡΙΔΑ ΠΑΤΙ ΜΕ ΕΝΔΙΑΦΕΡΕΙ Η ΑΠΟΨΗ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΗΜΟΣΙΟΓΡΑΦΩΝ	Correlation Coefficient	,056**	,022*	,179**	,387**	,346**	,166**	,231**	,135**	,282**	,104**	,280**	,123**	,212**	1,000	,178**	,128**	,151**	,059**	
		Sig. (2-tailed)	,000	,039	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000		,000	,000	,000		
		N	9135	8963	9138	8693	8880	9089	8875	8695	8827	8700	8603	9259	8924	9456	8719	8823	8692	9232	
	ΣΥΧΝΑ ΣΥΜΜΕΤΕΧΩ ΣΕ ΔΙΑΓΩΝΙΣΜΟΥΣ ΕΦΗΜΕΡΙΔΩΝ ΚΑΙ ΠΕΡΙΟΔΙΚΩΝ	Correlation Coefficient	,087**	,258**	,177**	,079**	,196**	,180**	,207**	,274**	,188**	,075**	,319**	,009	,437**	,178**	1,000	,479**	,151**	,-0,074**	
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,410	,000	,000		,000	,000		
		N	9102	8798	9156	8390	8716	9115	8791	8545	8708	8864	8396	9366	8800	8719	9564	8890	8837	9364	
	ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΠΕΡΙΟΔΙΚΟ ΑΝΑΛΟΓΑ ΜΕ ΤΟ ΔΩΡΟ DVD-CD-ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΕΙ	Correlation Coefficient	,108**	,239**	,241**	,049**	,188**	,219**	,188**	,235**	,155**	,087**	,109**	,285**	,498**	,128**	1,000	,209**	,032**		
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000		,000	,002		
		N	9131	8913	9134	8529	8780	9108	8832	8616	8750	8783	8520	9271	8967	8823	8890	9463	8754	9257	
	ΟΤΑΝ ΧΡΕΙΑΖΟΜΑΙ ΜΙΑ ΠΛΗΡΟΦΟΡΙΑ ΤΟ ΠΡΩΤΟ ΜΕΡΟΣ ΠΟΥ ΚΟΤΑΖΩ ΕΙΝΑΙ ΤΟ INTERNET	Correlation Coefficient	,023*	,159**	,216**	,169**	,275**	,187**	,118**	,154**	,163**	,187**	,067**	,191**	,151**	,151**	1,000	,209**	,705**		
		Sig. (2-tailed)	,028	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000		,000	,000		
		N	9255	8891	9220	8471	8813	9195	8834	8561	8738	8914	8423	9484	8713	8692	8837	8754	9780	9501	
	ΒΑΣΙΖΟΜΑΙ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΓΙΑ ΝΑ ΕΝΗΜΕΡΩΝΟΜΑΙ	Correlation Coefficient	,135**	,135**	,082**	,098**	,-0,037**	,082**	,120**	,086**	,030**	,005	,060**	,333**	,022*	,059**	,474**	,032**	,004	1,000	
		Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,004	,640	,000	,000	,033	,000	,000	,002	,705		
		N	9991	9539	9947	8978	9230	9950	9523	9033	9295	9418	8915	10411	9242	9232	9364	9257	9501	10739	

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

* Correlation is significant at the 0.05 level (2-tailed).

Η τιμή του μέτρου Kaiser-Meyer-Olkin που υπολογίστηκε είναι αρκετά υψηλή (0,834), το οποίο υποδεικνύει ότι οι συσχετίσεις ανάμεσα στα δεδομένα μας είναι αρκετά υψηλές. Επίσης, ο έλεγχος σφαιρικότητας του Bartlett απορρίπτει τη μηδενική υπόθεση ότι ο πίνακας συσχέτισης είναι ο μοναδιαίος (τιμή της ελεγχουσυνάρτησης 20770,6, βαθμοί ελευθερίας 136, p-value<0,001).

Όλα, λοιπόν, τα στοιχεία δείχνουν ότι τα δεδομένα μας είναι κατάλληλα για παραγοντική ανάλυση και ότι μπορούμε να συνεχίσουμε. Πρέπει, όμως, να εξετάσουμε και αν όλες οι μεταβλητές είναι κατάλληλες να χρησιμοποιηθούν στο μοντέλο. Αυτό θα το εξετάσουμε χρησιμοποιώντας την τιμή MSA, η οποία υπάρχει για κάθε μεταβλητή στον πίνακα 3.4. Ο πίνακας αυτός περιέχει την τιμή του συντελεστή συσχέτισης για κάθε μία μεταβλητή. Παρατηρούμε ότι η μεταβλητή « Βασίζομαι στην τηλεόραση για να ενημερώνομαι» έχει την μικρότερη τιμή (0,617) και δηλώνει πως αυτή η μεταβλητή είναι λιγότερο «σχετική» με τις υπόλοιπες. Παρ'όλα αυτά, όλες οι τιμές κρίνονται ικανοποιητικές και δεν υπάρχει λόγος να διώξουμε κάποια μεταβλητή.

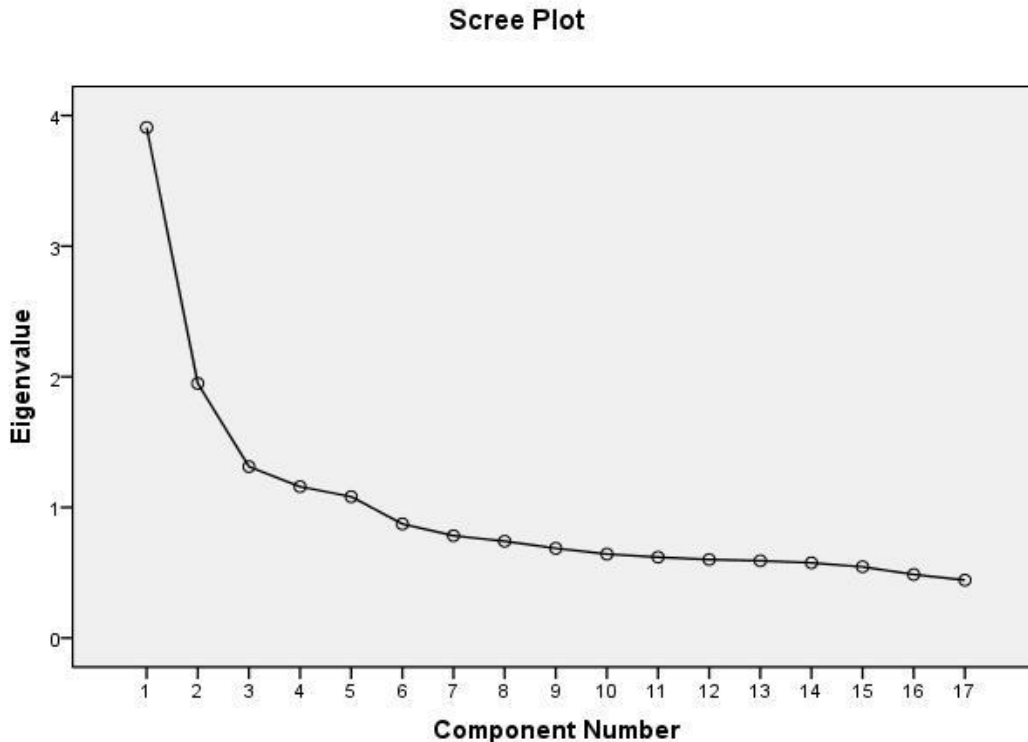
Πίνακας 3.4: Οι τιμές MSA των μεταβλητών

MSA			
ΔΕΝ ΕΝΔΙΑΦΕΡΟΜΑΙ ΓΙΑ ΤΙΣ ΠΟΛΙΤΙΚΕΣ ΕΙΔΗΣΕΙΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	0,784	ΔΙΑΒΑΖΩ ΤΙΣ ΟΙΚΟΝΟΜΙΚΕΣ ΣΕΛΙΔΕΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	0,812
ΤΑ ΕΝΘΕΤΑ ΚΑΝΟΥΝ ΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΠΙΟ ΕΝΔΙΑΦΕΡΟΥΣΕΣ	0,844	Η ΕΦΗΜΕΡΙΔΑ ΑΠΑΙΤΕΙ ΧΡΟΝΟ ΚΑΙ ΑΥΤΟΣΥΓΚΕΝΤΡΩΣΗ, ΕΝΩ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΕΝΗΜΕΡΩΝΕΣΑΙ ΓΡΗΓΟΡΑ ΚΑΙ ΧΩΡΙΣ ΚΟΠΟ	0,684
ΔΕΝ ΘΑ ΑΛΛΑΖΑ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΠΟΥ ΔΙΑΒΑΖΩ	0,835	ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΑΝΑΛΟΓΑ ΜΕ ΤΑ ΕΝΘΕΤΑ ΠΕΡΙΟΔΙΚΑ ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΟΥΝ	0,837
ΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΗΜΕΡΕΣ ΔΙΑΒΑΖΩ ΕΦΗΜΕΡΙΔΕΣ	0,851	ΔΕΝ ΑΛΛΑΖΩ ΕΥΚΟΛΑ ΕΦΗΜΕΡΙΔΑ ΓΙΑΤΙ ΜΕ ΕΝΔΙΑΦΕΡΕΙ Η ΑΠΟΨΗ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΗΜΟΣΙΟΓΡΑΦΩΝ	0,815
ΤΑ ΕΝΘΕΤΑ ΤΩΝ ΕΦΗΜΕΡΙΔΩΝ ΠΡΑΓΜΑΤΙΚΑ ΑΞΙΖΟΥΝ	0,835	ΣΥΧΝΑ ΣΥΜΜΕΤΕΧΩ ΣΕ ΔΙΑΓΩΝΙΣΜΟΥΣ ΕΦΗΜΕΡΙΔΩΝ ΚΑΙ ΠΕΡΙΟΔΙΚΩΝ	0,823
ΠΙΣΤΕΥΩ ΠΕΡΙΣΣΟΤΕΡΟ ΤΙΣ ΤΟΠΙΚΕΣ ΕΦΗΜΕΡΙΔΕΣ ΠΑΡΑ ΤΙΣ ΠΑΝΕΛΛΑΔΙΚΗΣ ΚΑΛΥΨΗΣ	0,903	ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ-ΠΕΡΙΟΔΙΚΟ ΑΝΑΛΟΓΑ ΜΕ ΤΟ ΔΩΡΟ-DVD-CD-ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΕΙ	0,825
ΔΙΑΒΑΖΩ ΚΑΘΗΜΕΡΙΝΕΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΔΙΑΣΚΕΔΑΣΗ ΚΑΙ ΟΧΙ ΓΙΑ ΕΝΗΜΕΡΩΣΗ	0,859	ΟΤΑΝ ΧΡΕΙΑΖΟΜΑΙ ΜΙΑ ΠΛΗΡΟΦΟΡΙΑ, ΤΟ ΠΡΩΤΟ ΜΕΡΟΣ ΠΟΥ ΚΟΙΤΑΖΩ ΕΙΝΑΙ ΤΟ INTERNET	0,881
ΒΑΣΙΖΟΜΑΙ ΣΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΤΗΝ ΕΝΗΜΕΡΩΣΗ ΜΟΥ	0,864	ΒΑΣΙΖΟΜΑΙ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΓΙΑ ΝΑ ΕΝΗΜΕΡΩΝΟΜΑΙ	0,617
ΔΕΝ ΜΠΟΡΩ ΝΑ ΑΝΤΙΣΤΑΘΩ ΣΤΟ ΝΑ ΑΓΟΡΑΖΩ ΠΕΡΙΟΔΙΚΑ	0,878		

3.8.2 Επιλογή Αριθμού Παραγόντων

Η επιλογή του αριθμού των παραγόντων είναι μια διαδικασία που προϋποθέτει επαναληπτικά την εκτίμηση και αξιολόγηση του μοντέλου. Επομένως, μπορεί κάποιος να χρησιμοποιήσει τον κανόνα του Kaiser, το ποσοστό της διακύμανσης που εξηγείται ή το scree plot.

Με τη βοήθεια του Scree Plot μπορούμε να δούμε πολύ παραστατικά το ποσοστό της διακύμανσης του δείγματος που επεξηγείται από κάθε παράγοντα.



Συγκεκριμένα στον οριζόντιο άξονα του γραφήματος απεικονίζονται οι παραγόμενοι παράγοντες και στον κάθετο το ποσοστό της διακύμανσης που επεξηγεί ο καθένας από αυτούς. Η ερμηνεία (ποσοστό διακύμανσης που ερμηνεύει) είναι σωστή μόνο αν χρησιμοποιηθεί η μέθοδος των κύριων συνιστωσών. Συνεπώς, το πρόβλημα επιλογής αριθμού παραγόντων δεν είναι άσχετο με την επιλογή μεθόδου εκτίμησης. Παρ'όλα αυτά, κριτήρια βασισμένα στις ιδιοτιμές χρησιμοποιούνται συχνά στην πράξη άσχετα με τη μέθοδο εκτίμησης που διαλέγει κανείς. Μόνο από το Scree Plot δεν είναι ξεκάθαρο πόσους παράγοντες θα κρατήσουμε.

Ο πίνακας 3.5 περιέχει τις ιδιοτιμές και το ποσοστό της διακύμανσης που κάθε ιδιοτιμή ερμηνεύει. Χρησιμοποιώντας τους 6 πρώτους θεωρητικούς παράγοντες που προκύπτουν από την εφαρμογή της παραγοντικής ανάλυσης, με τη μέθοδο των κύριων συνιστωσών, στο στατιστικό πακέτο SPSS, επιτυγχάνουμε να ερμηνεύσουμε σχεδόν το 60,5% της συνολικής μεταβλητότητας ενώ αν προχωρήσουμε με περισσότερους θεωρητικούς παράγοντες θα ερμηνεύουμε όλο και περισσότερη μεταβλητότητα. Ωστόσο δεν έχει νόημα να προχωρήσουμε σε λύση που θα υποδεικνύει περισσότερους από 6 παράγοντες καθότι στόχος είναι η απλοποίηση των δεδομένων. Για τη συμμετοχή κάθε θεωρητικού παράγοντα στην ερμηνεία της συνολικής μεταβλητότητας αλλά και τη φυσική ερμηνεία τους μέσω των factor loadings χρησιμοποιούμε ως τελική λύση αυτή που προκύπτει από την ορθογώνια περιστροφή των παραγόντων (factor rotation).

Πίνακας 3.5: Το ποσοστό της διακύμανσης που εξηγούν οι 6 παράγοντες**Total Variance Explained**

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,908	22,991	22,991	3,908	22,991	22,991	2,273	13,368	13,368
2	1,949	11,465	34,455	1,949	11,465	34,455	2,256	13,271	26,639
3	1,312	7,718	42,174	1,312	7,718	42,174	1,694	9,963	36,601
4	1,159	6,816	48,990	1,159	6,816	48,990	1,537	9,039	45,640
5	1,082	6,366	55,356	1,082	6,366	55,356	1,372	8,070	53,710
6	,874	5,139	60,494	,874	5,139	60,494	1,153	6,784	60,494
7	,783	4,608	65,102						
8	,741	4,358	69,461						
9	,687	4,039	73,499						
10	,643	3,782	77,281						
11	,619	3,643	80,924						
12	,601	3,533	84,457						
13	,591	3,478	87,935						
14	,576	3,387	91,323						
15	,545	3,208	94,530						
16	,487	2,862	97,392						
17	,443	2,608	100,000						

Extraction Method: Principal Component Analysis.

3.8.3 Εκτίμηση των παραμέτρων

Μέθοδος Κυρίων Συνιστωσών

Με τη μέθοδο των κύριων συνιστωσών η εκτιμήτρια του πίνακα L είναι ο πίνακας

$$\hat{L} = \left[\sqrt{\lambda_1} e_1 \quad \sqrt{\lambda_2} e_2 \right],$$

όπου λ_i , e_i είναι οι ιδιοτιμές και τα ιδιοδιανύσματα στήλη που αντιστοιχούν σε κάθε μια ιδιοτιμή. Στην περίπτωσή μας παίρνουμε τα πρώτα 6 ζεύγη ιδιοτιμών και ιδιοδιανυσμάτων, επειδή διαλέξαμε έξι παράγοντες. Στον πίνακα 3.6 βλέπουμε τα στοιχεία του πίνακα \hat{L} , δηλαδή τις επιβαρύνσεις των παραγόντων, που προκύπτουν για το μοντέλο για έξι παράγοντες.

Από τον πίνακα αυτό έχουν αφαιρεθεί οι τιμές των βαρών των οποίων η απόλυτη τιμή ήταν μικρότερη του 0,3. Μπορεί κανείς να δει πώς εκφράζεται η κάθε μεταβλητή με τη χρήση των 6 παραγόντων που χρησιμοποιήσαμε. Έτσι, για παράδειγμα έχουμε:

Επιλέγω με βάση τα ένθετα περιοδικά = 0,587F₁-0,422F₂-0,312F₃

Επιλέγω ανάλογα με το δώρο = 0,575 F₁-0,452 F₂

.....

Κοιτάζοντας, λοιπόν, τις σχέσεις των μεταβλητών και τον παραγόντων μπορούμε να δούμε ότι κάποιες μεταβλητές έχουν αρνητικά πρόσημα, ενώ οι υπόλοιπες θετικά πρόσημα, επομένως θα μπορούσε κάποιος να προσπαθήσει να ερμηνεύσει τους παράγοντες.

Πίνακας 3.6: Πίνακας με τις επιβαρύνσεις των παραγόντων

Component Matrix^a

	Component					
	1	2	3	4	5	6
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΑΝΑΛΟΓΑ ΜΕ ΤΑ ΕΝΘΕΤΑ ΠΕΡΙΟΔΙΚΑ ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΟΥΝ	,587	-,422			-,312	
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ-ΠΕΡΙΟΔΙΚΟ ΑΝΑΛΟΓΑ ΜΕ ΤΟ ΔΩΡΟ-DVD-CD-ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΕΙ	,575	-,452				
ΣΥΧΝΑ ΣΥΜΜΕΤΕΧΩ ΣΕ ΔΙΑΓΩΝΙΣΜΟΥΣ ΕΦΗΜΕΡΙΔΩΝ ΚΑΙ ΠΕΡΙΟΔΙΚΩΝ	,575	-,424				
ΔΕΝ ΜΠΟΡΩ ΝΑ ΑΝΤΙΣΤΑΘΩ ΣΤΟ ΝΑ ΑΓΟΡΑΣΩ ΠΕΡΙΟΔΙΚΑ	,550				,388	
ΔΙΑΒΑΣΩ ΚΑΘΗΜΕΡΙΝΕΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΔΙΑΣΚΕΔΑΣΗ ΚΑΙ ΟΧΙ ΓΙΑ ΕΝΗΜΕΡΩΣΗ	,544			,421	,303	
ΒΑΣΙΖΟΜΑΙ ΣΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΤΗΝ ΕΝΗΜΕΡΩΣΗ ΜΟΥ	,535	,405				
ΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΗΜΕΡΕΣ ΔΙΑΒΑΣΩ ΕΦΗΜΕΡΙΔΕΣ	,522	,432				
ΠΙΣΤΕΥΩ ΠΕΡΙΣΣΟΤΕΡΟ ΤΙΣ ΤΟΠΙΚΕΣ ΕΦΗΜΕΡΙΔΕΣ ΠΑΡΑ ΤΙΣ ΠΛΑΝΗΛΛΑΔΙΚΗΣ ΚΑΛΥΨΗΣ	,479			,417		
ΔΙΑΒΑΣΩ ΤΙΣ ΟΙΚΟΝΟΜΙΚΕΣ ΣΕΛΙΔΕΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	,437	,545				-,335
ΔΕΝ ΕΝΔΙΑΦΕΡΟΜΑΙ ΓΙΑ ΤΙΣ ΠΟΛΙΤΙΚΕΣ ΕΙΔΗΣΕΙΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	,374	-,529				,330
ΔΕΝ ΘΑ ΑΛΛΑΞΑ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΠΟΥ ΔΙΑΒΑΣΩ	,451	,501				
Η ΕΦΗΜΕΡΙΔΑ ΑΠΑΙΤΕΙ ΧΡΟΝΟ ΚΑΙ ΑΥΤΟ ΣΥΓΚΕΝΤΡΩΣΗ ΕΝΩ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΕΝΗΜΕΡΩΝΕΣΑΙ ΓΡΗΓΟΡΑ ΚΑΙ ΧΩΡΙΣ ΚΟΠΟ			,695			
ΒΑΣΙΖΟΜΑΙ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΓΙΑ ΝΑ ΕΝΗΜΕΡΩΝΟΜΑΙ			,627	,361	,315	
ΤΑ ΕΝΘΕΤΑ ΚΑΝΟΥΝ ΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΠΙΟ ΕΝΔΙΑΦΕΡΟΥΣΕΣ	,482			-,525		
ΤΑ ΕΝΘΕΤΑ ΤΩΝ ΕΦΗΜΕΡΙΔΩΝ ΠΡΑΓΜΑΤΙΚΑ ΑΞΙΖΟΥΝ	,467			-,469		
ΔΕΝ ΑΛΛΑΖΩ ΕΥΚΟΛΑ ΕΦΗΜΕΡΙΔΑ ΓΙΑΤΙ ΜΕ ΕΝΔΙΑΦΕΡΕΙ Η ΑΠΟΨΗ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΗΜΟΣΙΟΓΡΑΦΩΝ	,477	,332			-,514	
ΟΤΑΝ ΧΡΕΙΑΖΟΜΑΙ ΜΙΑ ΠΛΗΡΟΦΟΡΙΑ, ΤΟ ΠΡΩΤΟ ΜΕΡΟΣ ΠΟΥ ΚΟΙΤΑΖΩ ΕΙΝΑΙ ΤΟ INTERNET	,471					,615

Extraction Method: Principal Component Analysis.

a. 6 components extracted.

Στον πίνακα 3.7 βλέπουμε τις εταιρικότητες, δηλαδή τις διακυμάνσεις που εξηγούν οι παράγοντες που προσαρμόσαμε. Αυτό είναι ένας αριθμός από 0 έως 1 και είναι το ποσοστό της διακύμανσης κάθε μεταβλητής που εξηγείται από τον αριθμό των παραγόντων που προσαρμόσαμε. Η πρώτη στήλη (initial) είναι 1, αν έχουμε χρησιμοποιήσει τη μέθοδο των κύριων συνιστωσών. Παρατηρούμε επομένως ότι με το μοντέλο που προσαρμόσαμε

ερμηνεύουμε το 60,5% της μεταβλητής-φράσης «Δεν ενδιαφέρομαι για τις πολιτικές ειδήσεις της εφημερίδας μου». Το ποσοστό αυτό προκύπτει από το άθροισμα των τετραγώνων των επιβαρύνσεων των παραγόντων σε αυτή τη μεταβλητή. Βλέπουμε, επίσης, ότι το μοντέλο δεν καταφέρνει να εξηγήσει παρά μόνο το 44% της μεταβλητής-φράσης «Πιστεύω περισσότερο τις τοπικές εφημερίδες παρά τις πανελλαδικής κάλυψης» και αυτό είναι ίσως μια ένδειξη πως πρέπει να προσθέσουμε και άλλον παράγοντα, αν θέλουμε να αυξήσουμε την ερμηνεία για αυτή τη μεταβλητή.

Πίνακας 3.7: Οι εταιρικότητες των μεταβλητών για το παραγοντικό μοντέλο που προσαρμόσαμε

Communalities		
	Initial	Extraction
ΔΕΝ ΕΝΔΙΑΦΕΡΟΜΑΙ ΓΙΑ ΤΙΣ ΠΟΛΙΤΙΚΕΣ ΕΙΔΗΣΕΙΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	1,000	,605
ΤΑ ΕΝΘΕΤΑ ΚΑΝΟΥΝ ΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΠΙΟ ΕΝΔΙΑΦΕΡΟΥΣΕΣ	1,000	,595
ΔΕΝ ΘΑ ΑΛΛΑΞΑ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΠΟΥ ΔΙΑΒΑΖΩ	1,000	,512
ΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΗΜΕΡΕΣ ΔΙΑΒΑΖΩ ΕΦΗΜΕΡΙΔΕΣ	1,000	,571
ΤΑ ΕΝΘΕΤΑ ΤΩΝ ΕΦΗΜΕΡΙΔΩΝ ΠΡΑΓΜΑΤΙΚΑ ΑΞΙΖΟΥΝ	1,000	,629
ΠΙΣΤΕΥΩ ΠΕΡΙΣΣΟΤΕΡΟ ΤΙΣ ΤΟΠΙΚΕΣ ΕΦΗΜΕΡΙΔΕΣ ΠΑΡΑ ΤΙΣ ΠΑΝΕΛΛΑΔΙΚΗΣ ΚΑΛΥΨΗΣ	1,000	,438
ΔΙΑΒΑΖΩ ΚΑΘΗΜΕΡΙΝΕΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΔΙΑΣΚΕΔΑΣΗ ΚΑΙ ΟΧΙ ΓΙΑ ΕΝΗΜΕΡΩΣΗ	1,000	,597
ΒΑΣΙΖΟΜΑΙ ΣΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΤΗΝ ΕΝΗΜΕΡΩΣΗ ΜΟΥ	1,000	,531
ΔΕΝ ΜΠΟΡΩ ΝΑ ΑΝΤΙΣΤΑΘΩ ΣΤΟ ΝΑ ΑΓΟΡΑΖΩ ΠΕΡΙΟΔΙΚΑ	1,000	,559
ΔΙΑΒΑΖΩ ΤΙΣ ΟΙΚΟΝΟΜΙΚΕΣ ΣΕΛΙΔΕΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	1,000	,618
Η ΕΦΗΜΕΡΙΔΑ ΑΠΑΙΤΕΙ ΧΡΟΝΟ ΚΑΙ ΑΥΤΟΣΥΓΚΕΝΤΡΩΣΗ, ΕΝΩ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΕΝΗΜΕΡΩΝΕΣΑΙ ΓΡΗΓΟΡΑ ΚΑΙ ΧΩΡΙΣ ΚΟΠΟ	1,000	,656
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΑΝΑΛΟΓΑ ΜΕ ΤΑ ΕΝΘΕΤΑ ΠΕΡΙΟΔΙΚΑ ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΟΥΝ	1,000	,651
ΔΕΝ ΑΛΛΑΖΩ ΕΥΚΟΛΑ ΕΦΗΜΕΡΙΔΑ ΓΙΑΤΙ ΜΕ ΕΝΔΙΑΦΕΡΕΙ Η ΑΠΟΨΗ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΗΜΟΣΙΟΓΡΑΦΩΝ	1,000	,642
ΣΥΧΝΑ ΣΥΜΜΕΤΕΧΩ ΣΕ ΔΙΑΓΩΝΙΣΜΟΥΣ ΕΦΗΜΕΡΙΔΩΝ ΚΑΙ ΠΕΡΙΟΔΙΚΩΝ	1,000	,655
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ-ΠΕΡΙΟΔΙΚΟ ΑΝΑΛΟΓΑ ΜΕ ΤΟ ΔΩΡΟ-DVD-CD-ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΕΙ	1,000	,647
ΟΤΑΝ ΧΡΕΙΑΖΟΜΑΙ ΜΙΑ ΠΛΗΡΟΦΟΡΙΑ, ΤΟ ΠΡΩΤΟ ΜΕΡΟΣ ΠΟΥ ΚΟΙΤΑΖΩ ΕΙΝΑΙ ΤΟ INTERNET	1,000	,725
ΒΑΣΙΖΟΜΑΙ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΓΙΑ ΝΑ ΕΝΗΜΕΡΩΝΟΜΑΙ	1,000	,652

Extraction Method: Principal Component Analysis.

Επίσης, αν αφαιρέσουμε τη δεύτερη στήλη από τη μονάδα (και όχι από τη στήλη initial που τυγχάνει να είναι 1, επειδή χρησιμοποιήσαμε τη μέθοδο των κυρίων συνιστωσών) έχουμε τις εκτιμήσεις των ιδιοτεροτήτων ψ_i για κάθε μεταβλητή, δηλαδή του κομματιού εκείνου της διακύμανσης κάθε μεταβλητής που δεν μπορεί να εξηγήσει το παραγοντικό μοντέλο. Έτσι από

τον πίνακα 3.7 βλέπουμε ότι η ιδιαιτερότητα της μεταβλητής-φράσης φράσης «Δεν ενδιαφέρομαι για τις πολιτικές ειδήσεις της εφημερίδας μου» είναι 0,395 (=1-0,605).

Περιστροφή

Αν κοιτάξουμε τον πίνακα 3.8 βλέπουμε το ακριβές βάρος της κάθε μεταβλητής στους έξι παράγοντες. Εδώ χρησιμοποιήθηκε η μέθοδος varimax, η οποία προσπαθεί να κάνει τα στοιχεία της κάθε στήλης να φαίνονται μικρά. Η varimax περιστροφή είναι προτιμότερη σε σχέση με τις άλλες που έχουμε αναφέρει γιατί σκοπός μας είναι να δούμε ποιες μεταβλητές επιδρούν σε κάθε παράγοντα, για να αναγνωρίσουμε τον παράγοντα (Καρλής,2005). Από τον πίνακα Rotated Component Matrix έχουν αφαιρεθεί οι τιμές των βαρών των οποίων η απόλυτη τιμή είναι μικρότερη του 0,3 για να διευκολυνθεί η ερμηνεία των παραγόντων. Διαβάζοντας τον πίνακα αυτόν μπορούμε να εντοπίσουμε την ερμηνεία του κάθε παράγοντα.

Πίνακας 3.8: Ο πίνακας των επιβαρύνσεων μετά την περιστροφή

	Rotated Component Matrix ^a					
	Component					
	1	2	3	4	5	6
ΔΕΝ ΑΝΤΑΖΩ ΕΥΚΟΛΑ ΕΦΗΜΕΡΙΔΑ ΓΙΑΤΙ ΜΕ ΕΝΔΙΑΦΕΡΕΙ Η ΑΠΟΨΗ ΣΥΓΚΕΚΡΙΜΕΝΟΝ ΔΗΜΟΣΙΟΓΡΑΦΟΝ	,744					
ΔΕΝ ΘΑ ΑΝΤΑΖΑ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΠΟΥ ΔΙΑΒΑΖΩ	,674					
ΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΗΜΕΡΕΣ ΔΙΑΒΑΖΩ ΕΦΗΜΕΡΙΔΕΣ	,656					
ΔΙΑΒΑΖΩ ΤΙΣ ΟΙΚΟΝΟΜΙΚΕΣ ΣΕΛΙΔΕΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	,520			,409		
ΠΙΣΤΕΥΩ ΠΕΡΙΣΣΟΤΕΡΟ ΤΙΣ ΤΟΠΙΚΕΣ ΕΦΗΜΕΡΙΔΕΣ ΠΑΡΑ ΤΙΣ ΠΑΝΕΛΛΑΔΙΚΗΣ ΚΑΛΥΨΗΣ	,408		,375			
ΣΥΝΑ ΣΥΜΜΕΤΕΧΩ ΣΕ ΔΙΑΓΩΝΙΣΜΟΥΣ ΕΦΗΜΕΡΙΔΟΝ ΚΑΙ ΠΕΡΙΟΔΙΚΟΝ		,771				
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΑΝΑΛΟΓΑ ΜΕ ΤΑ ΕΝΘΕΤΑ ΠΕΡΙΟΔΙΚΑ ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΟΥΝ		,770				
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΠΕΡΙΟΔΙΚΟ ΑΝΑΛΟΓΑ ΜΕ ΤΟ ΔΩΡΟ-DVD-CD-ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΕΙ		,765				
ΔΙΑΒΑΖΩ ΚΑΘΗΜΕΡΙΝΕΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΔΙΑΣΚΕΔΑΣΗ ΚΑΙ ΟΧΙ ΓΙΑ ΕΝΗΜΕΡΩΣΗ			,713			
ΔΕΝ ΜΠΩΡΩ ΝΑ ΑΝΤΙΣΤΑΘΩ ΣΤΟ ΝΑ ΑΓΟΡΑΖΩ ΠΕΡΙΟΔΙΚΑ			,591			
ΒΑΣΙΖΟΜΑΙ ΣΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΤΗΝ ΕΝΗΜΕΡΩΣΗ ΜΟΥ	,480		,525			
ΤΑ ΕΝΘΕΤΑ ΤΩΝ ΕΦΗΜΕΡΙΔΟΝ ΠΡΑΓΜΑΤΙΚΑ ΑΞΙΖΟΥΝ				,764		
ΤΑ ΕΝΘΕΤΑ ΚΑΝΟΥΝ ΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΠΙΟ ΕΝΔΙΑΦΕΡΟΥΣΕΣ				,710		
ΒΑΣΙΖΟΜΑΙ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΓΙΑ ΝΑ ΕΝΗΜΕΡΩΘΩΜΑΙ					,743	
Η ΕΦΗΜΕΡΙΔΑ ΑΠΑΙΤΕΙ ΧΡΟΝΟ ΚΑΙ ΑΥΤΟΣΥΓΚΕΝΤΡΩΣΗ, ΕΝΩ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΕΝΗΜΕΡΩΝΕΣΑΙ ΓΡΗΓΟΡΑ ΚΑΙ ΧΩΡΙΣ ΚΟΠΟ					,713	
ΟΤΑΝ ΧΡΕΙΑΖΟΜΑΙ ΜΙΑ ΠΛΗΡΟΦΟΡΙΑ, ΤΟ ΠΡΩΤΟ ΜΕΡΟΣ ΠΟΥ ΚΟΙΤΑΖΩ ΕΙΝΑΙ ΤΟ INTERNET						,774
ΔΕΝ ΕΝΔΙΑΦΕΡΟΜΑΙ ΓΙΑ ΤΙΣ ΠΟΛΙΤΙΚΕΣ ΕΙΔΗΣΕΙΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ		,350			,350	,508

Extraction Method: Principal Component Analysis.
 Rotation Method: Varimax with Kaiser Normalization.
 a. Rotation converged in 7 iterations.

Ο πίνακας Component Transformation Matrix (πίνακας 3.9) είναι ο πίνακας με τον οποίο πολλαπλασιάσαμε τον αρχικό πίνακα επιβαρύνσεων για να οδηγηθούμε στον τελικό πίνακα επιβαρύνσεων. Δηλαδή προκύπτει:

Rotated Component Matrix = Component Matrix × Component Transformation Matrix
(Σιώμκος & Βασιλακοπούλου, 2005; Καρλής, 2005)

Όπως είπαμε και στην αρχή, η περιστροφή έχει σκοπό να μας βοηθήσει να «δούμε» καλύτερα τι σημαίνει κάθε παράγοντας, αφού ελπίζουμε πως θα μας ξεχωρίσει καλύτερα τις μεταβλητές.

Πίνακας 3.9: Ο πίνακας που χρησιμοποιήθηκε για την varimax περιστροφή

Component Transformation Matrix						
Component	1	2	3	4	5	6
1	,521	,548	,451	,377	,143	,249
2	,703	-,616	,080	,076	-,244	-,233
3	,088	-,175	-,301	,289	,867	-,187
4	,088	-,015	,512	-,748	,380	-,164
5	-,432	-,481	,621	,378	,083	,216
6	,180	-,240	-,228	-,257	,129	,880

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Δημιουργία των factor scores

Ένας από τους σκοπούς, όπως είπαμε και στην αρχή, είναι η δημιουργία καινούριων μεταβλητών για κάθε παρατήρηση, τόσες όσοι και οι παράγοντες στο μοντέλο που χρησιμοποιήσαμε, ώστε να μειώσουμε τις διαστάσεις του προβλήματος και να δημιουργήσουμε καινούρια δεδομένα για κάθε μια παρατήρηση με σκοπό περαιτέρω στατιστική επεξεργασία.

Σκοπός είναι δηλαδή να εκφράσουμε κάθε παράγοντα ως γραμμικό συνδυασμό των αρχικών μας μεταβλητών, συγκεκριμένα

$$F_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

$$F_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p$$

.....

$$F_k = a_{k1}X_1 + a_{k2}X_2 + \dots + a_{kp}X_p$$

ή αλλιώς $\mathbf{F} = \mathbf{AX}$, όπου ο πίνακας A ονομάζεται factor score coefficient matrix και διαφέρει από τον πίνακα των επιβαρύνσεων.

Για το μοντέλο που χρησιμοποιήσαμε, δηλαδή τους έξι παράγοντες, με τη μέθοδο των κύριων συνιστωσών, ο πίνακας 3.10 περιέχει τον πίνακα με τους συντελεστές των σκορ.

Πίνακας 3.10: Ο πίνακας των συντελεστών των factor scores

	Component					
	1	2	3	4	5	6
ΔΕΝ ΕΝΔΙΑΦΕΡΟΜΑΙ ΓΙΑ ΤΙΣ ΠΟΛΙΤΙΚΕΣ ΕΙΔΗΣΕΙΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	-,132	,036	,122	-,132	,234	,424
ΤΑ ΕΝΘΕΤΑ ΚΑΝΟΥΝ ΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΠΙΟ ΕΝΔΙΑΦΕΡΟΥΣΕΣ	-,040	-,006	-,138	,502	,021	,071
ΔΕΝ ΘΑ ΑΛΛΑΞΑ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΠΟΥ ΔΙΑΒΑΖΩ	,325	-,115	-,025	-,028	,086	,075
ΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΗΜΕΡΕΣ ΔΙΑΒΑΖΩ ΕΦΗΜΕΡΙΔΕΣ	,313	-,054	-,056	-,042	-,137	,245
ΤΑ ΕΝΘΕΤΑ ΤΩΝ ΕΦΗΜΕΡΙΔΩΝ ΠΡΑΓΜΑΤΙΚΑ ΑΞΙΖΟΥΝ	-,109	-,042	,006	,565	,027	-,082
ΠΙΣΤΕΥΩ ΠΕΡΙΣΣΟΤΕΡΟ ΤΙΣ ΤΟΠΙΚΕΣ ΕΦΗΜΕΡΙΔΕΣ ΠΑΡΑ ΤΙΣ ΠΑΝΕΛΛΑΔΙΚΗΣ ΚΑΛΥΨΗΣ	,170	,089	,191	-,224	,148	-,152
ΔΙΑΒΑΖΩ ΚΑΘΗΜΕΡΙΝΕΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΔΙΑΣΚΕΔΑΣΗ ΚΑΙ ΟΧΙ ΓΙΑ ΕΝΗΜΕΡΩΣΗ	-,047	-,019	,478	-,130	,063	-,006
ΒΑΣΙΖΟΜΑΙ ΣΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΤΗΝ ΕΝΗΜΕΡΩΣΗ ΜΟΥ	,137	-,093	,305	,014	-,032	-,089
ΔΕΝ ΜΠΟΡΩ ΝΑ ΑΝΤΙΣΤΑΘΩ ΣΤΟ ΝΑ ΑΓΟΡΑΖΩ ΠΕΡΙΟΔΙΚΑ	-,163	,005	,373	,131	-,128	,093
ΔΙΑΒΑΖΩ ΤΙΣ ΟΙΚΟΝΟΜΙΚΕΣ ΣΕΛΙΔΕΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	,158	-,044	,131	,263	-,119	-,351
Η ΕΦΗΜΕΡΙΔΑ ΑΠΑΙΤΕΙ ΧΡΟΝΟ ΚΑΙ ΑΥΤΟΣΥΓΚΕΝΤΡΩΣΗ, ΕΝΩ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΕΝΗΜΕΡΩΝΕΣΑΙ ΓΡΗΓΟΡΑ ΚΑΙ ΧΩΡΙΣ ΚΟΠΟ	,096	,059	-,268	,038	,518	,050
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΑΝΑΛΟΓΑ ΜΕ ΤΑ ΕΝΘΕΤΑ ΠΕΡΙΟΔΙΚΑ ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΟΥΝ	,013	,397	-,128	,032	-,002	-,112
ΔΕΝ ΑΛΛΑΖΩ ΕΥΚΟΛΑ ΕΦΗΜΕΡΙΔΑ ΓΙΑΤΙ ΜΕ ΕΝΔΙΑΦΕΡΕΙ Η ΑΠΟΨΗ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΗΜΟΣΙΟΓΡΑΦΩΝ	,428	,137	-,267	-,160	,072	-,020
ΣΥΧΝΑ ΣΥΜΜΕΤΕΧΩ ΣΕ ΔΙΑΓΩΝΙΣΜΟΥΣ ΕΦΗΜΕΡΙΔΩΝ ΚΑΙ ΠΕΡΙΟΔΙΚΩΝ	-,007	,393	,017	-,119	-,100	-,055
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ-ΠΕΡΙΟΔΙΚΟ ΑΝΑΛΟΓΑ ΜΕ ΤΟ ΔΩΡΟ-DVD-CD-ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΕΙ	-,070	,390	-,011	,060	-,083	-,144
ΟΤΑΝ ΧΡΕΙΑΖΟΜΑΙ ΜΙΑ ΠΛΗΡΟΦΟΡΙΑ, ΤΟ ΠΡΩΤΟ ΜΕΡΟΣ ΠΟΥ ΚΟΙΤΑΖΩ ΕΙΝΑΙ ΤΟ INTERNET	,087	-,138	-,079	,034	-,088	,746
ΒΑΣΙΖΟΜΑΙ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΓΙΑ ΝΑ ΕΝΗΜΕΡΩΝΟΜΑΙ	-,067	-,177	,218	,036	,567	-,088

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Αν αποφασίσουμε να προσθέσουμε έναν ακόμη παράγοντα, οι επιβαρύνσεις των έξι παραγόντων που είχαμε ήδη δεν θα αλλάξουν, ούτε αρχικά ούτε μετά την περιστροφή. Αυτό ισχύει μόνο στην περίπτωση της εκτίμησης με τη μέθοδο των κύριων συνιστωσών. Αυτό είναι εξαιρετικά χρήσιμο, καθώς η προσθήκη νέων παραγόντων είναι σχετικά εύκολη και δεν αλλάζει την όποια ερμηνεία έχουμε ήδη δώσει σε αυτούς τους παράγοντες.

Μέθοδος Μέγιστης Πιθανοφάνειας

Θα χρησιμοποιήσουμε τη μέθοδο μέγιστης πιθανοφάνειας, για να εκτιμήσουμε το παραγοντικό μοντέλο. Διαλέξαμε να πάρουμε 6 παράγοντες, κυρίως για να μπορέσουμε να κάνουμε σύγκριση ανάμεσα στις δύο μεθόδους.

Ο αλγόριθμος, για να ξεκινήσει χρειάζεται αρχικές τιμές για τις ιδιαιτερότητες και τις εταιρικότητες. Οι αρχικές τιμές των ιδιαιτεροτήτων για κάθε μεταβλητή είναι οι συντελεστές προσδιορισμού της παλινδρόμησης, κρατώντας ως εξαρτημένη μια μεταβλητή και ανεξάρτητες όλες τις υπόλοιπες. Αυτές τις αρχικές τιμές τις βλέπουμε στη στήλη Initial του πίνακα communalities (πίνακας 3.12). Στην περίπτωση της εκτίμησης με τη μέθοδο των κύριων συνιστωσών η αρχική τιμή ήταν 1.

Ο αλγόριθμος εργάζεται επαναληπτικά, χρησιμοποιώντας κάθε φορά τις ιδιαιτερότητες ως γνωστές, εκτιμά τις επιβαρύνσεις και στη συνέχεια, χρησιμοποιώντας τις επιβαρύνσεις ως γνωστές, επανεκτιμά τις ιδιαιτερότητες. Αυτή η διαδικασία επαναλαμβάνεται μέχρι οι εκτιμήσεις των παραγόντων να πάψουν να αλλάζουν.

Ο πίνακας 3.11 είναι ο πίνακας με τις επιβαρύνσεις, όταν τις εκτιμήσουμε με τη μέθοδο μέγιστης πιθανοφάνειας. Συγκρίνοντας τα αποτελέσματα με αυτά που πήραμε από την προηγούμενη μέθοδο εκτίμησης, μπορεί κανείς να δει ότι υπάρχουν διαφορές. Ως προς τον δεύτερο παράγοντα, για παράδειγμα, οι επιβαρύνσεις έχουν αλλάξει σε μερικές μεταβλητές. Για την μεταβλητή «Διαβάζω τις οικονομικές σελίδες της εφημερίδας μου» έχουμε τώρα 0,496 έναντι 0,545 που είχαμε πριν.

Πίνακας 3.11: Ο πίνακας των συντελεστών των παραγόντων

Factor Matrix^a

	Factor					
	1	2	3	4	5	6
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ-ΠΕΡΙΟΔΙΚΟ ΑΝΑΛΟΓΑ ΜΕ ΤΟ ΔΩΡΟ-DVD-CD-ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΕΙ	,600	-,369				
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΑΝΑΛΟΓΑ ΜΕ ΤΑ ΕΝΘΕΤΑ ΠΕΡΙΟΔΙΚΑ ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΟΥΝ	,597	-,314				
ΣΥΧΝΑ ΣΥΜΜΕΤΕΧΩ ΣΕ ΔΙΑΓΩΝΙΣΜΟΥΣ ΕΦΗΜΕΡΙΔΩΝ ΚΑΙ ΠΕΡΙΟΔΙΚΩΝ	,589	-,339				
ΔΕΝ ΜΠΟΡΩ ΝΑ ΑΝΤΙΣΤΑΘΩ ΣΤΟ ΝΑ ΑΓΟΡΑΣΩ ΠΕΡΙΟΔΙΚΑ	,498					
ΔΙΑΒΑΣΩ ΚΑΘΗΜΕΡΙΝΕΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΔΙΑΣΚΕΔΑΣΗ ΚΑΙ ΟΧΙ ΓΙΑ ΕΝΗΜΕΡΩΣΗ	,490					
ΒΑΣΙΖΟΜΑΙ ΣΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΤΗΝ ΕΝΗΜΕΡΩΣΗ ΜΟΥ	,458	,378				
ΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΗΜΕΡΕΣ ΔΙΑΒΑΣΩ ΕΦΗΜΕΡΙΔΕΣ	,452	,391				
ΔΕΝ ΑΛΛΑΖΩ ΕΥΚΟΛΑ ΕΦΗΜΕΡΙΔΑ ΓΙΑΤΙ ΜΕ ΕΝΔΙΑΦΕΡΕΙ Η ΑΠΟΨΗ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΗΜΟΣΙΟΓΡΑΦΩΝ	,434	,324				
ΤΑ ΕΝΘΕΤΑ ΚΑΝΟΥΝ ΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΠΙΟ ΕΝΔΙΑΦΕΡΟΥΣΕΣ	,432				,327	
ΠΙΣΤΕΥΩ ΠΕΡΙΣΣΟΤΕΡΟ ΤΙΣ ΤΟΠΙΚΕΣ ΕΦΗΜΕΡΙΔΕΣ ΠΑΡΑ ΤΙΣ ΠΑΝΕΛΛΑΔΙΚΗΣ ΚΑΛΥΨΗΣ	,410					
ΤΑ ΕΝΘΕΤΑ ΤΩΝ ΕΦΗΜΕΡΙΔΩΝ ΠΡΑΓΜΑΤΙΚΑ ΑΞΙΖΟΥΝ	,408				,301	
ΟΤΑΝ ΧΡΕΙΑΖΟΜΑΙ ΜΙΑ ΠΛΗΡΟΦΟΡΙΑ, ΤΟ ΠΡΩΤΟ ΜΕΡΟΣ ΠΟΥ ΚΟΙΤΑΣΩ ΕΙΝΑΙ ΤΟ INTERNET	,407					
ΔΙΑΒΑΣΩ ΤΙΣ ΟΙΚΟΝΟΜΙΚΕΣ ΣΕΛΙΔΕΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	,372	,496				
ΔΕΝ ΘΑ ΑΛΛΑΞΑ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΠΟΥ ΔΙΑΒΑΣΩ	,372	,444				
ΕΝΔΙΑΦΕΡΟΜΑΙ ΓΙΑ ΤΙΣ ΠΟΛΙΤΙΚΕΣ ΕΙΔΗΣΕΙΣ	-,352	,396				
Η ΕΦΗΜΕΡΙΔΑ ΑΠΑΙΤΕΙ ΧΡΟΝΟ ΚΑΙ ΑΥΤΟΣΥΓΚΕΝΤΡΩΣΗ, ΕΝΩ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΕΝΗΜΕΡΩΝΕΣΑΙ ΓΡΗΓΟΡΑ ΚΑΙ ΧΩΡΙΣ ΚΟΠΟ			,502			
ΒΑΣΙΖΟΜΑΙ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΓΙΑ ΝΑ ΕΝΗΜΕΡΩΝΟΜΑΙ						

Extraction Method: Maximum Likelihood.
a. 6 factors extracted. 5 iterations required.

Από τον πίνακα αυτό έχουν αφαιρεθεί οι τιμές των βαρών των οποίων η απόλυτη τιμή ήταν μικρότερη του 0,3. Μπορεί κανείς να δει πώς εκφράζεται η κάθε μεταβλητή με την χρήση των 6 παραγόντων που χρησιμοποιήσαμε. Έτσι, για παράδειγμα έχουμε:

Επιλέγω με βάση τα ένθετα περιοδικά = $0,597F_1 - 0,314F_2$
 Επιλέγω ανάλογα με το δώρο = $0,600 F_1 - 0,369 F_2$

.....

Πίνακας 3.12: Οι εταιρικότητες των μεταβλητών για το παραγοντικό μοντέλο που προσαρμόσαμε με τη μέθοδο της μέγιστης πιθανοφάνειας

Communalities		
	Initial	Extraction
ΕΝΔΙΑΦΕΡΟΜΑΙ ΓΙΑ ΤΙΣ ΠΟΛΙΤΙΚΕΣ ΕΙΔΗΣΕΙΣ	,227	,434
ΤΑ ΕΝΘΕΤΑ ΚΑΝΟΥΝ ΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΠΙΟ ΕΝΔΙΑΦΕΡΟΥΣΕΣ	,215	,383
ΔΕΝ ΘΑ ΑΛΛΑΞΑ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΠΟΥ ΔΙΑΒΑΖΩ	,260	,364
ΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΗΜΕΡΕΣ ΔΙΑΒΑΖΩ ΕΦΗΜΕΡΙΔΕΣ	,292	,395
ΤΑ ΕΝΘΕΤΑ ΤΩΝ ΕΦΗΜΕΡΙΔΩΝ ΠΡΑΓΜΑΤΙΚΑ ΑΞΙΖΟΥΝ	,205	,338
ΠΙΣΤΕΥΩ ΠΕΡΙΣΣΟΤΕΡΟ ΤΙΣ ΤΟΠΙΚΕΣ ΕΦΗΜΕΡΙΔΕΣ ΠΑΡΑ ΤΙΣ ΠΑΝΕΛΛΑΔΙΚΗΣ ΚΑΛΥΨΗΣ	,174	,236
ΔΙΑΒΑΖΩ ΚΑΘΗΜΕΡΙΝΕΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΔΙΑΣΚΕΔΑΣΗ ΚΑΙ ΟΧΙ ΓΙΑ ΕΝΗΜΕΡΩΣΗ	,254	,408
ΒΑΣΙΖΟΜΑΙ ΣΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΤΗΝ ΕΝΗΜΕΡΩΣΗ ΜΟΥ	,286	,397
ΔΕΝ ΜΠΟΡΩ ΝΑ ΑΝΤΙΣΤΑΘΩ ΣΤΟ ΝΑ ΑΓΟΡΑΖΩ ΠΕΡΙΟΔΙΚΑ	,252	,358
ΔΙΑΒΑΖΩ ΤΙΣ ΟΙΚΟΝΟΜΙΚΕΣ ΣΕΛΙΔΕΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	,278	,447
Η ΕΦΗΜΕΡΙΔΑ ΑΠΑΙΤΕΙ ΧΡΟΝΟ ΚΑΙ ΑΥΤΟ ΣΥΓΚΕΝΤΡΩΣΗ, ΕΝΩ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΕΝΗΜΕΡΩΝΕΣΑΙ ΓΡΗΓΟΡΑ ΚΑΙ ΧΩΡΙΣ ΚΟΠΟ	,154	,359
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΑΝΑΛΟΓΑ ΜΕ ΤΑ ΕΝΘΕΤΑ ΠΕΡΙΟΔΙΚΑ ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΟΥΝ	,374	,498
ΔΕΝ ΑΛΛΑΖΩ ΕΥΚΟΛΑ ΕΦΗΜΕΡΙΔΑ ΓΙΑΤΙ ΜΕ ΕΝΔΙΑΦΕΡΕΙ Η ΑΠΟΨΗ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΗΜΟΣΙΟΓΡΑΦΩΝ	,255	,470
ΣΥΧΝΑ ΣΥΜΜΕΤΕΧΩ ΣΕ ΔΙΑΓΩΝΙΣΜΟΥΣ ΕΦΗΜΕΡΙΔΩΝ ΚΑΙ ΠΕΡΙΟΔΙΚΩΝ	,377	,516
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ-ΠΕΡΙΟΔΙΚΟ ΑΝΑΛΟΓΑ ΜΕ ΤΟ ΔΩΡΟ-DVD-CD-ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΕΙ	,383	,554
ΟΤΑΝ ΧΡΕΙΑΖΟΜΑΙ ΜΙΑ ΠΛΗΡΟΦΟΡΙΑ, ΤΟ ΠΡΩΤΟ ΜΕΡΟΣ ΠΟΥ ΚΟΙΤΑΖΩ ΕΙΝΑΙ ΤΟ INTERNET	,173	,230
ΒΑΣΙΖΟΜΑΙ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΓΙΑ ΝΑ ΕΝΗΜΕΡΩΝΟΜΑΙ	,092	,234

Extraction Method: Maximum Likelihood.

Στον πίνακα 3.12 μπορούμε να δούμε τη διακύμανση κάθε μεταβλητής που καταφέραμε να εξηγήσουμε με 6 παράγοντες. Ο αριθμός προκύπτει, όπως και πριν, ως το άθροισμα τετραγώνων των επιβαρύνσεων όλων των παραγόντων. Αυτό που παρατηρεί κανείς είναι ότι με διαφορετικές μεθόδους παίρνουμε διαφορετικά αποτελέσματα ως προς το ποσοστό διακύμανσης κάθε μεταβλητής. Με τη μέθοδο μέγιστης πιθανοφάνειας εξηγούμε το 23,4% της μεταβλητής

«Βασίζομαι στην τηλεόραση για να ενημερώνομαι», ενώ με την προηγούμενη μέθοδο είχαμε εξηγήσει το 65,2%. Γενικότερα, παρατηρούμε ότι με αυτή τη μέθοδο εξηγούμε μικρότερα ποσοστά σε όλες τις μεταβλητές σε σχέση με τη μέθοδο των κύριων συνιστωσών.

Πίνακας 3.13: Το ποσοστό της διακύμανσης που εξηγούν οι 6 παράγοντες

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,913	23,017	23,017	3,305	19,444	19,444	1,728	10,166	10,166
2	1,943	11,429	34,446	1,400	8,233	27,676	1,681	9,887	20,053
3	1,313	7,722	42,168	,654	3,847	31,524	1,051	6,180	26,233
4	1,159	6,818	48,985	,507	2,980	34,504	1,024	6,025	32,258
5	1,077	6,337	55,322	,493	2,901	37,405	,668	3,930	36,188
6	,869	5,111	60,434	,262	1,540	38,945	,469	2,758	38,945
7	,784	4,610	65,043						
8	,741	4,360	69,403						
9	,686	4,035	73,438						
10	,641	3,771	77,209						
11	,618	3,635	80,843						
12	,602	3,540	84,383						
13	,590	3,473	87,856						
14	,577	3,397	91,252						
15	,551	3,242	94,494						
16	,489	2,874	97,369						
17	,447	2,631	100,000						

Extraction Method: Maximum Likelihood.

Ο πίνακας 3.13 περιέχει τις ιδιοτιμές και το ποσοστό της διακύμανσης που κάθε ιδιοτιμή ερμηνεύει. Στην αριστερή στήλη μπορεί να δει κανείς τις ιδιοτιμές και το ποσοστό διακύμανσης που θα εξηγούσε κάθε παράγοντας. Για τον j παράγοντα, η διακύμανση που εξηγεί αυτός ο παράγοντας είναι το άθροισμα τετραγώνων των επιβαρύνσεων του για όλες τις μεταβλητές, δηλαδή $\sum_{i=1}^p \hat{L}_{ij}^2$. Η τρίτη στήλη έχει να κάνει με το ποσοστό της διακύμανσης για κάθε μεταβλητή.

Παρατηρούμε ότι το ποσοστό κάθε παράγοντα έχει αλλάξει μετά την περιστροφή, αν και το συνολικό ποσοστό παραμένει το ίδιο.

Χρησιμοποιώντας τους 6 πρώτους θεωρητικούς παράγοντες που προκύπτουν από την εφαρμογή της παραγοντικής ανάλυσης, ερμηνεύσουμε το 39% της συνολικής μεταβλητότητας ενώ αν προχωρήσουμε με περισσότερους θεωρητικούς παράγοντες θα ερμηνεύουμε όλο και

περισσότερη μεταβλητότητα. Αυτός ο πίνακας παρουσιάζει ενδιαφέρον γιατί παρατηρείται διαφορά ανάμεσα στις δυο μεθόδους εκτίμησης, καθώς με αυτή τη μέθοδο ερμηνεύουμε το 39% της συνολικής μεταβλητότητας, ενώ με την προηγούμενη μέθοδο ερμηνεύσαμε το 60,5%. Επίσης, μια σημαντική διαφορά με τη μέθοδο των κύριων συνιστωσών είναι πως κάθε φορά που προσθέτουμε έναν παράγοντα, οι επιβαρύνσεις των προηγούμενων αλλάζουν.

Αξιολόγηση του μοντέλου

Η χρήση της μεθόδου της μέγιστης πιθανοφάνειας μας επιτρέπει τη στατιστική αξιολόγηση του μοντέλου. Συγκεκριμένα, μπορούμε να ελέγξουμε τη μηδενική υπόθεση

$$H_0 : \Sigma = LL' + \Psi \text{ έναντι της εναλλακτικής}$$

H_1 : δεν υπάρχει περιορισμός στον πίνακα Σ .

Η ελεγχοσυνάρτηση είναι η $LR = n(\text{tr}D - \log|D| - p)$, όπου $D = (\hat{L}\hat{L}' + \Psi)^{-1}S$. Η μηδενική υπόθεση είναι ισοδύναμη με την υπόθεση ότι το μοντέλο προσαρμόζει καλά τα δεδομένα. Η τιμή του LR συγκρίνεται με την τιμή χ^2 κατανομής με $s = [(p-k)^2 - (p+k)]/2$ βαθμούς ελευθερίας και, αν είναι μεγαλύτερη, απορρίπτουμε την καλή προσαρμοστικότητα του μοντέλου. Για τα δεδομένα μας απορρίπτουμε τη μηδενική υπόθεση, καθώς ο έλεγχος έχει p -τιμή $< 0,001$, συνεπώς το μοντέλο δεν είναι καλά προσαρμοσμένο.

Η απόρριψη της μηδενικής υπόθεσης ίσως να οφείλεται στο γεγονός ότι η υπόθεση της κανονικότητας μπορεί και να μην ισχύει.

Πίνακας 3.14: Έλεγχοι καλής προσαρμογής για διάφορα μοντέλα

Αριθμός παραγόντων k	Τιμή ελεγχοσυνάρτησης	Βαθμοί ελευθερίας	p-τιμή
1	8190,556	119	<0,001
2	3416,060	103	<0,001
3	2407,881	88	<0,001
4	1512,510	74	<0,001
5	623,554	61	<0,001
6	299,514	49	<0,001

Στην πράξη μια σωστή πρακτική είναι να προσθέτουμε παράγοντες μέχρι το μοντέλο να γίνει στατιστικά σημαντικό και άρα χρειαζόμαστε κάποιο κριτήριο επιλογής του καλύτερου μοντέλου, όπως αυτά που ακολουθούν (Hu & Bentler, 1999).

1) Tucker Lewis Index ή Non-normed Fit Index (NNFI)

Το Tucker Lewis Index υπολογίζεται από τον τύπο

$$\frac{x^2(0)/df(0) - x^2(k)/df(k)}{(x^2(0)/df(0)) - 1},$$

όπου $df(k)$ με $k=0,1,2,\dots,6$ είναι οι βαθμοί ελευθερίας και $\chi^2(k)$ με $k=0,1,2,\dots,6$ είναι οι τιμές της ελεγχοσυνάρτησης και οι βαθμοί ελευθερίας που υπάρχουν στον πίνακα 3.14 για κάθε μοντέλο με k διαφορετικούς παράγοντες.

2) Comparative Fit Index (CFI)

Το Comparative Fit Index υπολογίζεται από τον τύπο

$$1 - \frac{x^2(k) - df(k)}{x^2(0) - df(0)},$$

όπου $df(k)$ με $k=0,1,2,\dots,6$ είναι οι βαθμοί ελευθερίας και $\chi^2(k)$ με $k=0,1,2,\dots,6$ είναι οι τιμές της ελεγχοσυνάρτησης και οι βαθμοί ελευθερίας που υπάρχουν στον πίνακα 3.14 για κάθε μοντέλο με k διαφορετικούς παράγοντες.

Το Tucker Lewis Index περιλαμβάνει μία προσαρμογή για την πολυπλοκότητα του μοντέλου και όταν η τιμή του δείκτη είναι μεγαλύτερη του 0,9 έχουμε καλή προσαρμογή του μοντέλου. Το Comparative Fit Index υποδεικνύει το ποσοστό βελτίωσης του υπό διερεύνηση μοντέλου συγκριτικά με το μηδενικό μοντέλο, δηλαδή το «ανεξάρτητο» μοντέλο. Το ανεξάρτητο μοντέλο είναι εκείνο στο οποίο όλες οι μεταβλητές είναι ασυσχέτιστες. Ένας CFI δείκτης με τιμή 0,80 σημαίνει ότι η προσαρμογή του υπό διερεύνηση μοντέλου είναι 80% καλύτερη από αυτή του ανεξάρτητου μοντέλου. Επίσης, ο δείκτης CFI δεν επηρεάζεται από το μέγεθος του δείγματος. Όσο μεγαλύτερη είναι η τιμή του από 0,9 τόσο καλύτερη είναι η προσαρμογή του μοντέλου.

Πίνακας 3.15: Κριτήρια επιλογής αριθμού παραγόντων

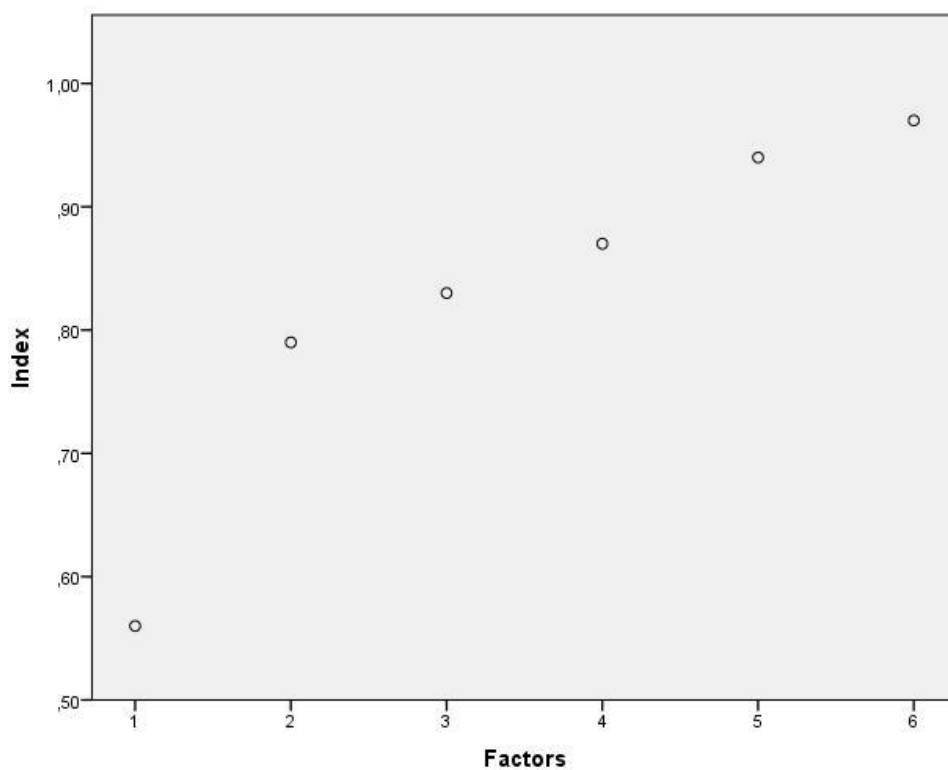
Παράγοντες	Tucker-Lewis Index (NNFI)	Comparative Fit Index (CFI)
1	0,56	0,61
2	0,79	0,84
3	0,83	0,89
4	0,87	0,93
5	0,94	0,97
6	0,97	0,99

Από τον πίνακα 3.15 παρατηρούμε ότι το χαμηλότερο index και στα δύο κριτήρια παρουσιάζει το παραγοντικό μοντέλο με έναν παράγοντα. Με βάση το κριτήριο Tucker-Lewis Index βλέπουμε ότι τα μοντέλα με πέντε και έξι παράγοντες έχουν τιμές δεικτών πάνω από 0,9. Το μοντέλο με τους 6

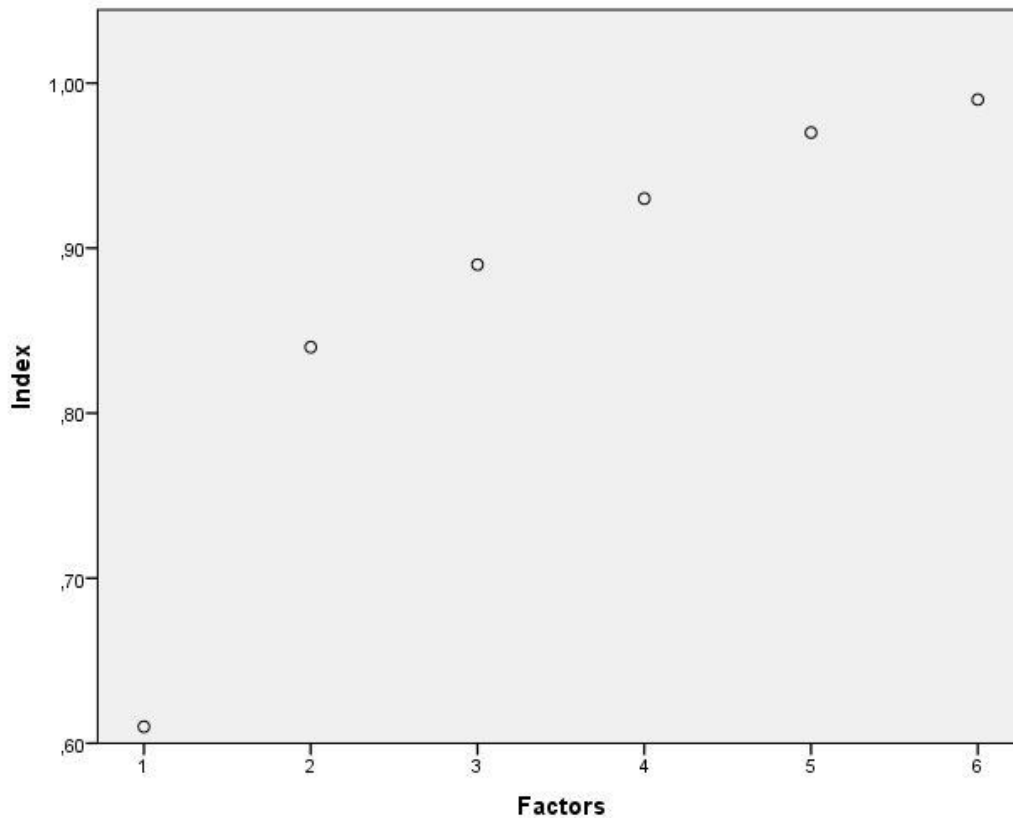
παράγοντες, όμως, έχει το μεγαλύτερο index , 0,97, συνεπώς είναι το καλύτερο μοντέλο. Με βάση το κριτήριο Comparative Fit Index βλέπουμε ότι και το μοντέλο με τους 4 παράγοντες θεωρείται καλό γιατί η τιμή του δείκτη είναι μεγαλύτερη του 0,9, αλλά και πάλι το μοντέλο με τους έξι παράγοντες είναι το καλύτερο.

Ακολουθούν γραφικές απεικονίσεις των δύο κριτηρίων, όπου στον οριζόντιο άξονα απεικονίζονται τα μοντέλα για τον διαφορετικό, κάθε φορά, αριθμό παραγόντων και στον κάθετο άξονα απεικονίζονται οι τιμές των δεικτών για κάθε ένα μοντέλο .

Γράφημα για το Tucker-Lewis Index / Non-normed Fit Index (NNFI)



Γράφημα για το Comparative Fit Index (CFI)



Περιστροφή

Αν κοιτάξουμε τον πίνακα 3.16 βλέπουμε το ακριβές βάρος της κάθε μεταβλητής στους έξι παράγοντες. Επίσης, και εδώ χρησιμοποιήθηκε η μέθοδος varimax, η οποία προσπαθεί να κάνει τα στοιχεία της κάθε στήλης να φαίνονται μικρά. Από τον πίνακα Rotated Component Matrix έχουν αφαιρεθεί οι τιμές των βαρών των οποίων η απόλυτη τιμή είναι μικρότερη του 0,3 για να διευκολυνθεί η ερμηνεία των παραγόντων. Διαβάζοντας τον πίνακα αυτόν μπορούμε να εντοπίσουμε την ερμηνεία του κάθε παράγοντα.

Πίνακας 3.16: Ο πίνακας των επιβαρύνσεων μετά την περιστροφή

Rotated Factor Matrix^a

	Factor					
	1	2	3	4	5	6
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ-ΠΕΡΙΟΔΙΚΟ ΑΝΑΛΟΓΑ ΜΕ ΤΟ ΔΩΡΟ-DVD-CD-ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΕΙ	,701					
ΣΥΧΝΑ ΣΥΜΜΕΤΕΧΩ ΣΕ ΔΙΑΓΩΝΙΣΜΟΥΣ ΕΦΗΜΕΡΙΔΩΝ ΚΑΙ ΠΕΡΙΟΔΙΚΩΝ	,661					
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΑΝΑΛΟΓΑ ΜΕ ΤΑ ΕΝΘΕΤΑ ΠΕΡΙΟΔΙΚΑ ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΟΥΝ	,647					
ΔΕΝ ΑΛΛΑΖΩ ΕΥΚΟΛΑ ΕΦΗΜΕΡΙΔΑ ΓΙΑΤΙ ΜΕ ΕΝΔΙΑΦΕΡΕΙ Η ΑΠΟΨΗ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΗΜΟΣΙΟΓΡΑΦΩΝ		,648				
ΔΕΝ ΘΑ ΑΛΛΑΞΑ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΠΟΥ ΔΙΑΒΑΖΩ		,559				
ΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΗΜΕΡΕΣ ΔΙΑΒΑΖΩ ΕΦΗΜΕΡΙΔΕΣ		,542				
ΒΑΣΙΖΟΜΑΙ ΣΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΤΗΝ ΕΝΗΜΕΡΩΣΗ ΜΟΥ		,441	,410			
ΔΙΑΒΑΖΩ ΤΙΣ ΟΙΚΟΝΟΜΙΚΕΣ ΣΕΛΙΔΕΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ		,437				,377
ΠΙΣΤΕΥΩ ΠΕΡΙΣΣΟΤΕΡΟ ΤΙΣ ΤΟΠΙΚΕΣ ΕΦΗΜΕΡΙΔΕΣ ΠΑΡΑ ΤΙΣ ΠΑΝΕΛΛΑΔΙΚΗΣ ΚΑΛΥΨΗΣ		,326				
ΔΙΑΒΑΖΩ ΚΑΘΗΜΕΡΙΝΕΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΔΙΑΣΚΕΔΑΣΗ ΚΑΙ ΟΧΙ ΓΙΑ ΕΝΗΜΕΡΩΣΗ			,556			
ΔΕΝ ΜΠΟΡΩ ΝΑ ΑΝΤΙΣΤΑΘΩ ΣΤΟ ΝΑ ΑΓΟΡΑΖΩ ΠΕΡΙΟΔΙΚΑ			,441			
ΤΑ ΕΝΘΕΤΑ ΚΑΝΟΥΝ ΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΠΙΟ ΕΝΔΙΑΦΕΡΟΥΣΕΣ				,567		
ΤΑ ΕΝΘΕΤΑ ΤΩΝ ΕΦΗΜΕΡΙΔΩΝ ΠΡΑΓΜΑΤΙΚΑ ΑΞΙΖΟΥΝ				,528		
ΟΤΑΝ ΧΡΕΙΑΖΟΜΑΙ ΜΙΑ ΠΛΗΡΟΦΟΡΙΑ, ΤΟ ΠΡΩΤΟ ΜΕΡΟΣ ΠΟΥ ΚΟΙΤΑΖΩ ΕΙΝΑΙ ΤΟ INTERNET				,305		
Η ΕΦΗΜΕΡΙΔΑ ΑΠΑΙΤΕΙ ΧΡΟΝΟ ΚΑΙ ΑΥΤΟΣΥΓΚΕΝΤΡΩΣΗ, ΕΝΩ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΕΝΗΜΕΡΩΝΕΣΑΙ ΓΡΗΓΟΡΑ ΚΑΙ ΧΩΡΙΣ ΚΟΠΟ					,532	
ΒΑΣΙΖΟΜΑΙ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΓΙΑ ΝΑ ΕΝΗΜΕΡΩΝΟΜΑΙ					,461	
ΕΝΔΙΑΦΕΡΟΜΑΙ ΓΙΑ ΤΙΣ ΠΟΛΙΤΙΚΕΣ ΕΙΔΗΣΕΙΣ						,487

Extraction Method: Maximum Likelihood.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Σύμφωνα με τις τιμές των φορτίων προχωρήσαμε στην εξής ομαδοποίηση και ονομασία των θεωρητικών παραγόντων:

Πίνακας 3.17

ΠΡΟΣΦΟΡΑ&ΔΩΡΟ	ΑΠΟΨΕΙΣ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΗΜΟΣΙΟΓΡΑΦΩΝ	ΨΥΧΑΓΩΓΙΚΑ ΘΕΜΑΤΑ	ΕΝΘΕΤΑ	ΓΡΗΓΟΡΗ & ΕΥΚΟΛΗ ΕΝΗΜΕΡΩΣΗ	ΠΟΛΙΤΙΚΗ & ΟΙΚΟΝΟΜΙΚΗ ΕΝΗΜΕΡΩΣΗ
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ-ΠΕΡΙΟΔΙΚΟ ΑΝΑΛΟΓΑ ΜΕ ΤΟ ΔΩΡΟ-DVD-CD ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΕΙ	ΔΕΝ ΑΛΛΑΖΩ ΕΥΚΟΛΑ ΕΦΗΜΕΡΙΔΑ ΓΙΑΤΙ ΜΕ ΕΝΔΙΑΦΕΡΕΙ Η ΑΠΟΨΗ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΗΜΟΣΙΟΓΡΑΦΩΝ	ΔΙΑΒΑΖΩ ΚΑΘΗΜΕΡΙΝΕΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΔΙΑΣΚΕΔΑΣΗ ΚΑΙ ΟΧΙ ΓΙΑ ΕΝΗΜΕΡΩΣΗ	ΤΑ ΕΝΘΕΤΑ ΚΑΝΟΥΝ ΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΠΙΟ ΕΝΔΙΑΦΕΡΟΥΣΕΣ	Η ΕΦΗΜΕΡΙΔΑ ΑΠΑΙΤΕΙ ΧΡΟΝΟ ΚΑΙ ΑΥΤΟΣΥΓΚΕΝΤΡΩΣΗ, ΕΝΩ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΕΝΗΜΕΡΩΝΕΣΑΙ ΓΡΗΓΟΡΑ ΚΑΙ ΧΩΡΙΣ ΚΟΠΟ	ΕΝΔΙΑΦΕΡΟΜΑΙ ΓΙΑ ΤΙΣ ΠΟΛΙΤΙΚΕΣ ΕΙΔΗΣΕΙΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ
ΣΥΝΗΘΩΣ ΣΥΜΜΕΤΕΧΩ ΣΕ ΔΙΑΓΩΝΙΣΜΟΥΣ ΕΦΗΜΕΡΙΔΩΝ ΚΑΙ ΠΕΡΙΟΔΙΚΩΝ	ΔΕΝ ΘΑ ΑΛΛΑΞΑ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΠΟΥ ΔΙΑΒΑΖΩ	ΔΕΝ ΜΠΟΡΩ ΝΑ ΑΝΤΙΣΤΑΘΩ ΣΤΟ ΝΑ ΑΓΟΡΑΖΩ ΠΕΡΙΟΔΙΚΑ	ΤΑ ΕΝΘΕΤΑ ΤΩΝ ΕΦΗΜΕΡΙΔΩΝ ΠΡΑΓΜΑΤΙΚΑ ΑΞΙΖΟΥΝ	ΒΑΣΙΖΟΜΑΙ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΓΙΑ ΝΑ ΕΝΗΜΕΡΩΝΟΜΑΙ	ΔΙΑΒΑΖΩ ΤΙΣ ΟΙΚΟΝΟΜΙΚΕΣ ΣΕΛΙΔΕΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΑΝΑΛΟΓΑ ΜΕ ΤΑ ΕΝΘΕΤΑ ΠΕΡΙΟΔΙΚΑ ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΟΥΝ	ΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΗΜΕΡΕΣ ΔΙΑΒΑΖΩ ΕΦΗΜΕΡΙΔΕΣ		ΟΤΑΝ ΧΡΕΙΑΖΟΜΑΙ ΜΙΑ ΠΛΗΡΟΦΟΡΙΑ, ΤΟ ΠΡΩΤΟ ΜΕΡΟΣ ΠΟΥ ΚΟΙΤΑΖΩ ΕΙΝΑΙ ΤΟ INTERNET		
	ΒΑΣΙΖΟΜΑΙ ΣΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΤΗΝ ΕΝΗΜΕΡΩΣΗ ΜΟΥ				
	ΔΙΑΒΑΖΩ ΤΙΣ ΟΙΚΟΝΟΜΙΚΕΣ ΣΕΛΙΔΕΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ				
	ΠΙΣΤΕΥΩ ΠΕΡΙΣΣΟΤΕΡΟ ΤΙΣ ΤΟΠΙΚΕΣ ΕΦΗΜΕΡΙΔΕΣ ΠΑΡΑ ΤΙΣ ΠΑΝΕΛΛΑΔΙΚΗΣ ΚΑΛΥΨΗΣ				

Ο πίνακας Component Transformation Matrix (πίνακας 3.18), όπως αναφέραμε και στην μέθοδο εκτίμησης των κύριων συνιστωσών, είναι ο πίνακας με τον οποίο πολλαπλασιάσαμε τον αρχικό πίνακα επιβαρύνσεων για να οδηγηθούμε στον τελικό πίνακα επιβαρύνσεων.

Πίνακας 3.18: Ο πίνακας μετασχηματισμού που χρησιμοποιήθηκε για τη varimax περιστροφή

Factor Transformation Matrix

Factor	1	2	3	4	5	6
1	,637	,491	,406	,404	,139	-,076
2	-,589	,697	,100	,089	-,109	,370
3	-,098	,074	-,556	,366	,736	-,017
4	-,454	-,302	,665	,242	,346	-,288
5	-,027	-,348	-,096	,766	-,414	,333
6	,175	-,235	,255	-,223	,370	,814

Extraction Method: Maximum Likelihood.
Rotation Method: Varimax with Kaiser Normalization.

Για το μοντέλο που χρησιμοποιήσαμε, δηλαδή τους έξι παράγοντες, με τη μέθοδο της μέγιστης πιθανοφάνειας, ο πίνακας 3.19 περιέχει τον πίνακα με τους συντελεστές των σκορ.

Πίνακας 3.19: Ο πίνακας των συντελεστών των factor scores

Factor Score Coefficient Matrix

	Factor					
	1	2	3	4	5	6
ΕΝΔΙΑΦΕΡΟΜΑΙ ΓΙΑ ΤΙΣ ΠΟΛΙΤΙΚΕΣ ΕΙΔΗΣΕΙΣ	-,007	,057	-,131	-,038	-,100	,431
ΤΑ ΕΝΘΕΤΑ ΚΑΝΟΥΝ ΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΠΙΟ ΕΝΔΙΑΦΕΡΟΥΣΕΣ	-,022	-,019	-,090	,388	,032	-,011
ΔΕΝ ΘΑ ΑΛΛΑΞΑ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΠΟΥ ΔΙΑΒΑΖΩ	-,085	,243	,006	,023	,035	-,056
ΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΗΜΕΡΕΣ ΔΙΑΒΑΖΩ ΕΦΗΜΕΡΙΔΕΣ	-,017	,232	,010	,055	-,156	-,031
ΤΑ ΕΝΘΕΤΑ ΤΩΝ ΕΦΗΜΕΡΙΔΩΝ ΠΡΑΓΜΑΤΙΚΑ ΔΕΙΣΟΥΝ	-,042	-,038	-,012	,332	,061	,029
ΠΙΣΤΕΥΩ ΠΕΡΙΣΣΟΤΕΡΟ ΤΙΣ ΤΟΠΙΚΕΣ ΕΦΗΜΕΡΙΔΕΣ ΠΑΡΑ ΤΙΣ ΠΑΝΕΛΛΑΔΙΚΗΣ ΚΑΛΥΨΗΣ	,023	,095	,104	-,093	,093	,008
ΔΙΑΒΑΖΩ ΚΑΘΗΜΕΡΙΝΕΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΔΙΑΣΚΕΔΑΣΗ ΚΑΙ ΟΧΙ ΓΙΑ ΕΝΗΜΕΡΩΣΗ	-,007	,007	,369	-,099	,074	-,030
ΒΑΣΙΖΟΜΑΙ ΣΤΙΣ ΕΦΗΜΕΡΙΔΕΣ ΓΙΑ ΤΗΝ ΕΝΗΜΕΡΩΣΗ ΜΟΥ	-,062	,133	,227	-,006	,004	,072
ΔΕΝ ΜΠΟΡΩ ΝΑ ΑΝΤΙΣΤΑΘΩ ΣΤΟ ΝΑ ΑΓΟΡΑΖΩ ΠΕΡΙΟΔΙΚΑ	,014	-,057	,237	,111	-,091	-,056
ΔΙΑΒΑΖΩ ΤΙΣ ΟΙΚΟΝΟΜΙΚΕΣ ΣΕΛΙΔΕΣ ΤΗΣ ΕΦΗΜΕΡΙΔΑΣ ΜΟΥ	-,034	,121	,103	,110	,006	,352
Η ΕΦΗΜΕΡΙΔΑ ΑΠΑΙΤΕΙ ΧΡΟΝΟ ΚΑΙ ΑΥΤΟΣΥΓΚΕΝΤΡΩΣΗ, ΕΝΩ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΕΝΗΜΕΡΩΝΕΣΑΙ ΓΡΗΓΟΡΑ ΚΑΙ ΧΩΡΙΣ ΚΟΠΟ	,023	,028	-,132	,041	,402	-,041
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ ΑΝΑΛΟΓΑ ΜΕ ΤΑ ΕΝΘΕΤΑ ΠΕΡΙΟΔΙΚΑ ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΟΥΝ	,306	,008	-,104	,026	,066	,035
ΔΕΝ ΑΛΛΑΖΩ ΕΥΚΟΛΑ ΕΦΗΜΕΡΙΔΑ ΓΙΑΤΙ ΜΕ ΕΝΔΙΑΦΕΡΕΙ Η ΑΠΟΨΗ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΗΜΟΣΙΟΓΡΑΦΩΝ	,066	,403	-,218	-,113	,055	-,084
ΣΥΧΝΑ ΣΥΜΜΕΤΕΧΩ ΣΕ ΔΙΑΓΩΝΙΣΜΟΥΣ ΕΦΗΜΕΡΙΔΩΝ ΚΑΙ ΠΕΡΙΟΔΙΚΩΝ	,323	,023	,028	-,104	-,103	-,079
ΣΥΝΗΘΩΣ ΕΠΙΛΕΓΩ ΤΗΝ ΕΦΗΜΕΡΙΔΑ-ΠΕΡΙΟΔΙΚΟ ΑΝΑΛΟΓΑ ΜΕ ΤΟ ΔΩΡΟ-DVD-CD-ΠΟΥ ΜΕ ΕΝΔΙΑΦΕΡΕΙ	,394	-,105	,003	,030	-,024	,175
ΟΤΑΝ ΧΡΕΙΑΖΟΜΑΙ ΜΙΑ ΠΛΗΡΟΦΟΡΙΑ, ΤΟ ΠΡΩΤΟ ΜΕΡΟΣ ΠΟΥ ΚΟΙΤΑΖΩ ΕΙΝΑΙ ΤΟ INTERNET	-,002	,036	,027	,142	-,117	-,118
ΒΑΣΙΖΟΜΑΙ ΣΤΗΝ ΤΗΛΕΟΡΑΣΗ ΓΙΑ ΝΑ ΕΝΗΜΕΡΩΝΟΜΑΙ	-,053	-,030	,091	-,017	,319	,038

Extraction Method: Maximum Likelihood.
Rotation Method: Varimax with Kaiser Normalization.

ΚΕΦΑΛΑΙΟ 4

ΑΝΑΛΥΣΗ ΚΑΤΑ ΣΥΣΤΑΔΕΣ

4.1 Εισαγωγή

Η ανάλυση κατά συστάδες (cluster analysis) είναι ένα βασικό εργαλείο για τμηματοποίηση δεδομένων και δημιουργία ομάδων. Είναι μια διερευνητική στατιστική τεχνική σχεδιασμένη για τον εντοπισμό εσωτερικών σχέσεων στα δεδομένα. Για παράδειγμα, θα μπορούσαμε να χρησιμοποιήσουμε την ανάλυση κατά συστάδες για τον εντοπισμό ομάδων μεταξύ των καταναλωτών ενός προϊόντος. Η ιδέα της ανάλυσης κατά συστάδες βασίζεται στην απλή λογική ότι: «τα δεδομένα μπορούν να διαχωριστούν σε ομάδες βάση της απόστασης μεταξύ τους».

Εκείνο που θα εξετάσουμε στις επόμενες παραγράφους είναι συστηματικές μεθόδους ομαδοποίησης οι οποίες έχουν κάποια μαθηματική/ στατιστική βάση. Οι μέθοδοι ομαδοποίησης μπορούν να χωριστούν σε δύο διαφορετικές κατηγορίες ανάλογα με τον τρόπο που προχωρούν στη διαμόρφωση των ομάδων: στις *ιεραρχικές* και στις *μη ιεραρχικές* μεθόδους.

4.2 Τι είναι Ανάλυση Κατά Συστάδες

Η ανάλυση κατά συστάδες χρησιμοποιείται για να ομαδοποιήσει ή να τμηματοποιήσει τις παρατηρήσεις έτσι ώστε:

1. Κάθε ομάδα να είναι ομοιογενής με βάση κάποια χαρακτηριστικά έτσι ώστε οι παρατηρήσεις να είναι όμοιες μεταξύ τους.
2. Κάθε ομάδα να είναι διαφορετική από την άλλη με βάση τα χαρακτηριστικά τους. (Sharma, 1995)

Το βασικό κριτήριο της ανάλυσης είναι η απόσταση. Βάσει της λογικής της ανάλυσης κατά συστάδες, οι κοντινές παρατηρήσεις θα ανήκουν στην ίδια ομάδα, ενώ οι μακρινές παρατηρήσεις θα ανήκουν σε διαφορετικές ομάδες.

Οι βασικότερες και πιο διαδεδομένες προσεγγίσεις είναι:

- **Ιεραρχικές μέθοδοι:** Οι ομάδες σχηματίζονται σταδιακά είτε με συνένωση μικρότερων ομάδων σχηματίζοντας συνεχώς μεγαλύτερες ομάδες μέχρι να φτάσουμε να έχουμε όλα τα δεδομένα σε μια ομάδα (*συσσωρευτικές μέθοδοι*), είτε με διαίρεση ομάδων σε μικρότερες μέχρι να φτάσουμε σε μια κατάσταση όπου κάθε παρατήρηση να είναι από μόνη της μια ομάδα (*διαιρετικές μέθοδοι*).
- **Μη ιεραρχικές μέθοδοι:** Τα δεδομένα διαιρούνται σε k τμήματα. Κάθε ένα από τα τμήματα αυτά αντιστοιχεί σε μία ομάδα. Σε αντίθεση, λοιπόν, με τις ιεραρχικές μεθόδους ο αριθμός των ομάδων που θα δημιουργηθούν θα πρέπει να είναι γνωστός εκ των προτέρων.

4.3 Στόχος της Ανάλυσης Κατά Συστάδες

Στόχος της ανάλυσης κατά συστάδες είναι οι ομαδοποιήσεις των παρατηρήσεων, έτσι ώστε οι ομάδες να είναι όσο ομοιογενείς είναι δυνατόν, βάσει των μεταβλητών που συμμετέχουν στην ανάλυση. Το πρώτο βήμα της ανάλυσης είναι η επιλογή του *μέτρου ομοιότητας* που θα χρησιμοποιηθεί. Έπειτα, επιλέγεται το *είδος της τεχνικής ομαδοποίησης* που θα χρησιμοποιηθεί (ιεραχική ή μη ιεραχική). Το τρίτο βήμα είναι η *επιλογή της μεθόδου* του είδους ομαδοποίησης που έχει επιλεγεί. Το τέταρτο βήμα είναι η *επιλογή του αριθμού των ομάδων* που θα δημιουργηθούν και τέλος ακολουθεί η *ερμηνεία των ομάδων* που δημιουργήθηκαν.

4.4 Η απόσταση

Κρίνεται σκόπιμο να αναφερθεί η έννοια της *απόστασης* και κατ' επέκταση της *ομοιότητας*. Η απόσταση είναι μια θεμελιώδης έννοια στην πολυμεταβλητή ανάλυση και όχι μόνο για την ανάλυση δεδομένων. Σκοπός της απόστασης είναι να μετρήσει κανείς πόσο απέχουν δύο παρατηρήσεις, να ποσοτικοποιήσει δηλαδή αν μοιάζουν ή όχι οι παρατηρήσεις. Γίνεται κατανοητό πως ένα μέτρο απόστασης μπορεί να μετασχηματιστεί σε μέτρο ομοιότητας και το αντίστροφο. Από στατιστική άποψη, η απόσταση επιτρέπει τις συγκρίσεις ανάμεσα στις μεταβλητές, αλλά το βασικό της μειονέκτημα είναι ότι αγνοεί τις συνδιακυμάνσεις μεταξύ των μεταβλητών, γεγονός που θα γίνει κατανοητό στη συνέχεια.

Για παράδειγμα, αν υποθέσουμε ότι έχουμε κάποιες υποθετικές παρατηρήσεις που αφορούν δύο μεταβλητές, το εισόδημα και την εκπαίδευση έξι ατόμων, η γεωμετρική απεικόνιση αυτών θα αποτελείται από σημεία, στον χώρο των δύο διαστάσεων. Αν συμβολίσουμε τις δύο μεταβλητές $y = (y_1, y_2)$ και $x = (x_1, x_2)$, τότε μια προσέγγιση για την επιλογή μιας απόστασης ανάμεσα στις δύο παρατηρήσεις θα ήταν η *ευκλείδεια απόσταση*

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

την οποία μπορούμε να γενικεύσουμε για την περίπτωση που έχουμε p παρατηρήσεις, δηλαδή $y = (y_1, y_2, \dots, y_p)$ και $x = (x_1, x_2, \dots, x_p)$. Τότε η αντίστοιχη απόσταση μπορεί να οριστεί ως

$$d(x, y) = \sqrt{\sum_{i=1}^p (x_i - y_i)^2} \quad (4.1).$$

Η ευκλείδεια απόσταση δεν είναι, ωστόσο, επαρκής και αυτό συμβαίνει εξαιτίας των μεταβλητών που δεν είναι σε συγκρίσιμη κλίμακα. Ένας τρόπος για να αποφύγουμε αυτή την ανεπάρκεια είναι να διαιρέσουμε όλες τις μεταβλητές με τις τυπικές τους αποκλίσεις. Θεωρούμε s_i^2 τη διακύμανση της i μεταβλητής και έχουμε τον τύπο της μεταβλητής με την εξής μορφή:

$$d(x, y) = \sqrt{\sum_{i=1}^p \left(\frac{x_i - y_i}{s_i} \right)^2} \quad (4.2).$$

Ένας τύπος που υπολογίζει τις συνδιακυμάνσεις των μεταβλητών είναι η *απόσταση Mahalanobis*, ο οποίος είναι ο εξής:

$$d^2(x, y) = (x - y)^T \cdot S^{-1} \cdot (x - y) \quad (4.3),$$

Όπου S είναι ο δειγματικός πίνακας διακυμάνσεων και x, y είναι διανύσματα. Αν οι τιμές των

διανυσμάτων είναι ασυσχέτιστες τότε ο S είναι διαγώνιος με διαγώνια στοιχεία s_k^2 η απόσταση του Mahalanobis γίνεται ίδια με τη σταθμισμένη ευκλείδεια απόσταση με $w_k = 1/s_k$.

Έπειτα, έχουμε την απόσταση *City-block* ή *Manhattan*

$$d(x, y) = \sum_{i=1}^p |x_i - y_i| \quad (4.4),$$

η οποία μοιάζει με την ευκλείδεια απόσταση με τη διαφορά ότι αντί για τετραγωνικές αποκλίσεις χρησιμοποιεί απόλυτες αποκλίσεις. Δίνει περίπου ίδια αποτελέσματα με την ευκλείδεια απόσταση, αλλά απαλείφει την αδυναμία της ευκλείδειας απόστασης να επηρεάζεται από τις ακραίες τιμές των παρατηρήσεων (outliers) επειδή τους δίνει μικρότερο βάρος.

Σε αντιστοιχία με τη σταθμισμένη ευκλείδεια απόσταση, μπορεί να οριστεί και η σταθμισμένη απόσταση Manhattan μέσω του τύπου

$$d(x, y) = \sum_{i=1}^p w_r |x_i - y_j|,$$

όπου $w_k, k = 1, 2, \dots, p$ είναι μη αρνητικά βάρη.

Ένας εναλλακτικός τύπος είναι η απόσταση *Minkowski* ή *Lq norm* (Johnson & Wichern, 1998) ο οποίος ορίζεται ως

$$d(x, y) = \left\{ \sum_{i=1}^p (|x_i - y_i|)^q \right\}^{1/q} \quad (4.5),$$

όπου $q \geq 1$ δεδομένη παράμετρος.

Αν ισχύει $q=2$ τότε έχουμε πάλι την ευκλείδεια απόσταση και αν $q=1$ προκύπτει η απόσταση city-block (Manhattan).

Επίσης, υπάρχει η απόσταση *Chebyshev*

$$d(x, y) = \max \{|x_i - y_i|, i = 1, \dots, p\} \quad (4.6),$$

η οποία χρησιμοποιεί μόνο τη μεγαλύτερη απόκλιση. Σύμφωνα με την απόσταση αυτή, δύο παρατηρήσεις είναι διαφορετικές αν έχουν μεγάλες διαφορές σε μια τουλάχιστον μεταβλητή. Μπορεί να θεωρηθεί ως ειδική περίπτωση της Minkowski, αφού ισχύει

$$\lim_{q \rightarrow \infty} \left\{ \sum_{i=1}^p (|x_i - y_i|)^q \right\}^{1/q} = \max |x_i - y_i|.$$

Η απόσταση *Gower*

$$d(x, y) = \sum_{r=1}^n \frac{|x_r - y_r|}{R_r} \quad (4.7),$$

όπου $R_r = \max x_r - \min y_r$. Η απόσταση αυτή, που αποτελεί ειδική περίπτωση της σταθμισμένης city block, χρησιμοποιείται συνήθως όταν στα δεδομένα μας περιλαμβάνονται και δίτιμες τυχαίες μεταβλητές.

Τέλος, η απόσταση *Bhattacharyya*

$$d(x, y) = \left\{ \sum_{i=1}^p (\sqrt{x_i} - \sqrt{y_i})^2 \right\}^{1/2} \quad (4.8),$$

η οποία χρησιμοποιείται συνήθως όταν τα δεδομένα μας αποτελούνται από ποσοστά.

Όλες οι παραπάνω αποστάσεις αφορούν συνεχή δεδομένα που είναι και τα πιο διαδεδομένα.

Κλείνοντας, την παράγραφο αυτή θα αναφέρουμε ακόμη ότι:

A. όταν οι προς μελέτη μεταβλητές είναι κατηγορικές μη διατάξιμες (π.χ. χρώμα ματιών, μάρκα αυτοκινήτου, τύπος μηχανήματος) τότε θα μπορούσαμε να προχωρήσουμε αρχικά στην κατασκευή δίτιμων μεταβλητών, μια για την κάθε κατηγορία κάθε κατηγορικής μεταβλητής, και στη συνέχεια να υπολογίσουμε την αντίστοιχη απόσταση για τα δίτιμα δεδομένα που θα προκύψουν. Η μεθοδολογία αυτή θα δώσει την επόμενη απόσταση (simple matching distance) μεταξύ των ατόμων/ αντικειμένων i και j :

$$d_{ij} = \frac{p-u}{p} = 1 - \frac{u}{p} \quad (4.9),$$

όπου u είναι ο αριθμός των συμφωνιών, δηλαδή ο αριθμός των μεταβλητών για τις οποίες τα αντικείμενα i και j εμφανίζουν την ίδια κατάσταση, και p είναι ο συνολικός αριθμός των μεταβλητών.

B. στην περίπτωση διατάξιμων κατηγορικών μεταβλητών, για παράδειγμα, όταν ζητείται από κάποιον να απαντήσει δίδοντας βαθμό σε κάποια κλίμακα (καθόλου=0, ελάχιστα=1, μέτρια=2, αρκετά=3, πολύ=4), συνήθως θεωρούμε τις μεταβλητές ως συνεχείς και χρησιμοποιούμε μια κατάλληλη απόσταση (Κούτρας, 2008).

4.5 Ιεραρχική Ομαδοποίηση

Σε αυτήν την παράγραφο θα εξετάσουμε μια σειρά από αλγορίθμους που παράγουν μια ιεραρχία «δενδροειδούς μορφής» όπου στα διάφορα στάδια το πλήθος k των ομάδων παίρνει όλες τις δυνατές τιμές από 1 έως n . Στο ένα άκρο αυτής της ιεραρχίας υπάρχει μια μόνο ομάδα που περιέχει n άτομα και στο άλλο άκρο υπάρχουν n ομάδες όπου η κάθε μία περιέχει ένα μόνο άτομο.

Όπως έχει ήδη αναφερθεί, υπάρχουν δύο βασικές κατηγορίες ιεραρχικών μεθόδων: οι συσσωρευτικές μέθοδοι (agglomerative methods) και οι διαιρετικές μέθοδοι (divisive methods). Οι ιεραρχικές μέθοδοι, επειδή σε κάθε βήμα χρησιμοποιούν έναν πίνακα αποστάσεων, δηλαδή τις αποστάσεις όλων των παρατηρήσεων από τις υπόλοιπες, χρειάζονται πολύ χρόνο και χώρο στον υπολογιστή και για αυτό είναι ασύμφορες για μεγάλα σετ δεδομένων.

4.5.1 Συσσωρευτικές μέθοδοι (Agglomerative methods)

Υποθέτουμε ότι διαθέτουμε n παρατηρήσεις και έχουν υπολογιστεί οι αποστάσεις τους $d(x_i, y_i)$, $i=1,2,\dots,n$ οι οποίες έχουν τοποθετηθεί σε έναν πίνακα **D**.

Θα μπορούσαμε να περιγράψουμε τον αλγόριθμο για τις συσσωρευτικές μεθόδους ως εξής (Κούτρας, 2008):

Βήμα 1^ο. Ξεκίνα με n ομάδες (clusters) του ενός ατόμου η κάθε μία και με τον πίνακα των αποστάσεων **D**.

Βήμα 2^ο. Εντόπισε στον πίνακα **D** το ζεύγος των πλησιέστερων (πιο όμοιων) ομάδων έστω Q και R.

Βήμα 3^ο. Συγχώνευσε τις ομάδες Q και R σε μια ομάδα, την $P=(QR)$ μειώνοντας έτσι τον αριθμό των ομάδων κατά ένα. Ανανέωσε τον πίνακα αποστάσεων **D** διαγράφοντας τις γραμμές και τις στήλες που αντιστοιχούσαν στις ομάδες Q και R, και προσθέτοντας μια γραμμή και μια στήλη που περιέχει τις αποστάσεις της ομάδας $P=(QR)$ από τις υπόλοιπες.

Βήμα 4^ο. Επανάλαβε τα βήματα 2 και 3 συνολικά $n-1$ φορές έτσι ώστε με τη λήξη του αλγορίθμου όλες οι παρατηρήσεις να αποτελούν μία μόνο ομάδα.

Αν και έχουμε ορίσει μέτρα αποστάσεων μεταξύ στοιχείων, δεν έχουμε, μέχρι στιγμής, ορίσει μέτρα αποστάσεων μεταξύ ομάδων. Για τον καθορισμό τέτοιων αποστάσεων υπάρχει ένας αριθμός από διαφορετικές τεχνικές για τον υπολογισμό των αποστάσεων αυτών (S. Sharma, 1995). Οι πιο γνωστές μέθοδοι είναι:

1. Μέθοδος της απλής συνένωσης (Nearest neighbour ή Single linkage method).
2. Μέθοδος της πλήρους συνένωσης (Farthest neighbour ή Complete linkage method).
3. Μέθοδος των σταθμισμένων μέσων (Average linkage method).
4. Μέθοδος των κέντρων βάρους (Centroid method).
5. Μέθοδος του Ward (Ward's method).

1. Μέθοδος της απλής συνένωσης (Nearest neighbour ή Single linkage method)

Εδώ ο ορίζουμε την απόσταση μεταξύ δύο ομάδων, τη μικρότερη απόσταση από μια παρατήρηση μέσα στη μια ομάδα με κάποια παρατήρηση στην άλλη ομάδα. Στον ορισμό αυτόν οφείλεται η δεύτερη ονομασία της μεθόδου που είναι **μέθοδος του πλησιέστερου (κοντινότερου) γείτονα (nearest neighbour method)**. Η μέθοδος έχει κάποιες χρήσιμες μαθηματικές ιδιότητες, αλλά παράγει ομάδες που δεν είναι συμπαγείς και συνήθως δημιουργεί μερικές πολύ μεγάλες ομάδες και κάποιες πολύ μικρές.

2. Μέθοδος της πλήρους συνένωσης (Complete linkage method)

Η μέθοδος αυτή, σε σχέση με τη μέθοδο του πλησιέστερου γείτονα βασίζεται στην ακριβώς αντίθετη έκφραση της απόστασης μεταξύ των ομάδων. Εδώ, η απόσταση ανάμεσα σε δύο ομάδες ορίζεται ως η μεγαλύτερη απόσταση από μια παρατήρηση μέσα στη μια ομάδα με κάποια παρατήρηση στην άλλη ομάδα. Αυτό δικαιολογεί και τη χρήση της εναλλακτικής ορολογίας **μέθοδος του μακρινότερου γείτονα (furthest neighbour method)**. Οι ομάδες που δημιουργούνται είναι συνήθως μεγάλες συμπαγείς, όμως, η μέθοδος αρκετά συχνά αποτυγχάνει να ξεχωρίσει κάποιες πολύ συμπαγείς μικρές ομάδες.

3. Μέθοδος σταθμισμένων μέσων (Weighted Average Linkage method)

Σε αυτή την περίπτωση η απόσταση είναι ο μέσος της απόστασης ανάμεσα σε όλες τις αποστάσεις της μιας ομάδας με τα στοιχεία της άλλης. Αν, για παράδειγμα, η μια ομάδα περιλαμβάνει τις παρατηρήσεις 1,2,4 και η άλλη τις παρατηρήσεις 3,5, τότε η απόστασή τους είναι ο μέσος των αποστάσεων d_{13} , d_{15} , d_{23} , d_{25} , d_{43} , d_{45} , δηλαδή

$$d((124), (3,5)) = \frac{d_{13} + d_{15} + d_{23} + d_{25} + d_{43} + d_{45}}{6}$$

Μια παραλλαγή της μεθόδου είναι η Group Average Method, όπου η απόσταση είναι ο μέσος όλων των αποστάσεων που προκύπτουν αν ενώσουμε τις δύο ομάδες. Στην περίπτωση των ομάδων που είχαμε πριν (124) και (35) η νέα απόσταση θα είναι ο μέσος των αποστάσεων d_{12} , d_{13} , d_{14} , d_{15} , d_{23} , d_{24} , d_{25} , d_{34} , d_{35} , d_{45} , δηλαδή

$$d((124), (3,5)) = \frac{d_{12} + d_{13} + d_{14} + d_{15} + d_{23} + d_{24} + d_{25} + d_{34} + d_{35} + d_{45}}{10}.$$

4. Μέθοδος των κέντρων βάρους (Centroid method)

Η απόσταση υπολογίζεται ως η απόσταση των κέντρων των ομάδων. Ένα μειονέκτημα της μεθόδου είναι ότι μπορεί να εφαρμοστεί μόνο σε ποσοτικά δεδομένα, λόγω της χρήσης της ευκλείδειας απόστασης. Η μέθοδος αυτή έχει μερικές καλές ιδιότητες και παράγει συνήθως ομάδες συμπαγείς και ελλειπτικές.

5. Μέθοδος του Ward (Ward's method)

Η μέθοδος του Ward διαφέρει από τις υπόλοιπες και είναι σχεδιασμένη να ελαχιστοποιεί τη διακύμανση μέσα στις ομάδες. Για κάθε παρατήρηση μπορούμε να υπολογίσουμε την απόστασή της από το κέντρο της ομάδας. Αν αθροίσουμε για όλες τις ομάδες, έχουμε μια τιμή που είναι το συνολικό άθροισμα. Αρχικά, αυτό το άθροισμα είναι 0, αφού κάθε παρατήρηση είναι και μια ομάδα, άρα η απόστασή της από το κέντρο είναι 0. Σε κάθε βήμα ενώνουμε τις ομάδες τις οποίες, αν ενωθούν, οδηγούν στη μικρότερη αύξηση του συνολικού αθροίσματος αποστάσεων. Η μέθοδος έχει μερικές καλές ιδιότητες και συνήθως δημιουργεί ομάδες με παρόμοιο αριθμό παρατηρήσεων (Mardia et al., 1979).

Συγκρίνοντας τις μεθόδους μεταξύ τους θα πρέπει να γνωρίζουμε ότι από πειράματα προσομοίωσης οι μέθοδοι με την καλύτερη επίδοση είναι η μέθοδος του Ward και η group average method. Η μέθοδος του πλησιέστερου γείτονα είναι αυτή με τη χειρότερη επίδοση. Είναι σημαντικό να γνωρίζουμε ότι αν οι ομάδες είναι αρκετά διαφορετικές μεταξύ τους, κάθε μέθοδος θα βρει τη σωστή ομαδοποίηση και, επίσης, ότι κάθε μέθοδος δουλεύει καλύτερα με συγκεκριμένη μορφή δεδομένων (Καρλής, 2005).

Τα βασικά μειονεκτήματα των αλγορίθμων ιεραρχικής ομαδοποίησης είναι τα εξής:

- Συνήθως είναι ασύμφορες από άποψη υπολογιστικού φόρτου όταν θέλουμε να αναλύσουμε μεγάλα σύνολα δεδομένων. Αυτό οφείλεται στο γεγονός ότι πρέπει κανείς να σχηματίσει και να αποθήκευση στη μνήμη του υπολογιστή ολόκληρο τον πίνακα αποστάσεων των παρατηρήσεων.
- Οι ομάδες που φτιάχνονται σε αρχικά βήματα δεν μπορούν να χωρίσουν και επομένως οι παρατηρήσεις που ενώνονται σε αρχικά βήματα μένουν μαζί για πάντα. Πολύ συχνά καταλήγει στη δημιουργία ενός μικρού πλήθους ομάδων με πολλές παρατηρήσεις και αφήνει αρκετές παρατηρήσεις να είναι από μόνες τους ανεξάρτητες ομάδες.

4.5.2 Διαιρετικές μέθοδοι (*Divisive methods*)

Παρότι οι ιεραρχικές μέθοδοι είναι σχεδόν συνυφασμένες με τις συσσωρευτικές μεθόδους, υπάρχει και μια άλλη μεγάλη κατηγορία, αυτή των διαιρετικών μεθόδων, οι οποίες εκτελούν την αντίστροφη ακριβώς διαδικασία από τις συσσωρευτικές (Κούτρας, 2008).

Οι διαιρετικές μέθοδοι ξεκινούν με μια μόνο ομάδα που περιέχει τις n παρατηρήσεις, για τις οποίες έχουν καταμετρηθεί τα χαρακτηριστικά (μεταβλητές) και τη διαιρούν σε όλο και μικρότερες ομάδες. Η λογική στην οποία βασίζονται οι διαιρετικοί αλγόριθμοι είναι, να βρίσκουν υποομάδες των ήδη διαμορφωμένων ομάδων που είναι περισσότερο απομακρυσμένες και να τις διαχωρίζουν. Έτσι, σε κάθε βήμα διαμερίζουν μια ομάδα σε δύο άλλες έως ότου φτάσουν στο σημείο όπου όλες οι ομάδες περιέχουν ένα μόνο στοιχείο.

Ο κύριος λόγος που οι διαιρετικές μέθοδοι δεν είναι αρκετά διαδεδομένες στην πράξη, είναι ότι απαιτούν πολύ περισσότερους υπολογισμούς από ότι οι συσσωρευτικές μέθοδοι.

4.6 Μη Ιεραρχική Ομαδοποίηση

Ο στόχος των μη ιεραρχικών μεθόδων είναι να ομαδοποιήσουν τις n μονάδες των δεδομένων σε k ομάδες, όπου το k είναι καθορισμένο από την αρχή. Αυτό αποτελεί έναν περιορισμό της μεθόδου, καθώς είτε πρέπει να τρέξουμε τον αλγόριθμο με διαφορετικές επιλογές ως προς το πλήθος των ομάδων, είτε πρέπει με κάποιον άλλο τρόπο να έχουμε καταλήξει στον αριθμό των ομάδων.

Ο μηχανισμός λειτουργίας των περισσότερων μη ιεραρχικών μεθόδων είναι:

- να θεωρούν k συγκεκριμένα άτομα (μητρικά σημεία- *seed points*) και γύρω από αυτά να ταξινομηθούν τα υπόλοιπα στοιχεία έως ότου διαμορφωθούν οι επιθυμητές ομάδες ή

- να ξεκινούν με ένα αρχικό διαμερισμό (initial partition) των ατόμων σε k ομάδες και στη συνέχεια να μετακινούν τις παρατηρήσεις μεταξύ των ομάδων έως ότου πετύχουν τον καλύτερο διαμερισμό.

Θα δούμε, στη συνέχεια, μία από τις πιο δημοφιλής μη ιεραρχικές μεθόδους, την K-Means μέθοδο (Johnson & Wichern, 1998).

4.6.1 Η μέθοδος K-Means

Ο αλγόριθμος k-means ανήκει σε μια μεγάλη οικογένεια αλγορίθμων ομαδοποίησης που είναι γνωστοί ως αλγόριθμοι διαμέρισης (partitioning algorithms). Ουσιαστικά οι αλγόριθμοι είναι έτσι φτιαγμένοι ώστε να διαμερίζουν το πολυεπίπεδο που δημιουργούν τα δεδομένα σε περιοχές και να αντιστοιχούν μια περιοχή σε κάθε ομάδα.

Η μέθοδος θεωρεί ότι ο αριθμός των ομάδων που θα προκύψουν είναι γνωστός εκ των προτέρων. Αυτό αποτελεί έναν περιορισμό της μεθόδου καθώς είτε πρέπει να τρέξουμε τον αλγόριθμο με διαφορετικές επιλογές ως προς το πλήθος των ομάδων είτε πρέπει με κάποιον άλλο τρόπο να έχουμε καταλήξει στον αριθμό των ομάδων.

Η μέθοδος δουλεύει επαναληπτικά. Χρησιμοποιεί την έννοια του κέντρου της ομάδας και στη συνέχεια κατατάσσει τις παρατηρήσεις ανάλογα με την απόσταση τους από τα κέντρα όλων των ομάδων. Το κέντρο της ομάδας είναι η μέση τιμή για κάθε μεταβλητή όλων των παρατηρήσεων των ομάδων, δηλαδή αντιστοιχεί στο διάνυσμα των μέσων.

Ο αλγόριθμος k-means για την ταξινόμηση n παρατηρήσεων σε k ομάδες αποτελείται από τα ακόλουθα βήματα (Johnson & Wichern, 1998; Κούτρας, 2008).

Βήμα 1^ο. Καθόρισε αρχικά ένα αρχικό σύνολο από k μητρικά σημεία χρησιμοποιώντας k από τις n παρατηρήσεις που είναι διαθέσιμες.

Βήμα 2^ο. Κατάταξε κάθε μία από τις εναπομένουσες $n-k$ παρατηρήσεις στην ομάδα της οποίας το κέντρο έχει τη μικρότερη απόσταση από την παρατήρηση. Μετά από κάθε τοποθέτηση υπολόγισε ξανά το κέντρο βάρους της αλλαγμένης πλέον ομάδας.

Βήμα 3^ο. Όταν όλες οι παρατηρήσεις έχουν τοποθετηθεί σε ομάδες μέσω του βήματος 2, θεώρησε τα δημιουργηθέντα κέντρα βάρους ως μητρικά σημεία και εκτέλεσε μια ακόμη σάρωση στα δεδομένα τοποθετώντας κάθε παρατήρηση στο πλησιέστερο μητρικό σημείο.

Αυτός ο αλγόριθμος είναι ιδιαίτερα γρήγορος αφού στην πράξη τερματίζει συνήθως μετά από σχετικά λίγες επαναλήψεις. Αυτό τον κάνει ιδιαίτερα χρήσιμο για τις περιπτώσεις που απαιτείται ομαδοποίηση σε μεγάλα σύνολα δεδομένων. Επίσης, δεν χρειάζεται να κρατά στη μνήμη πολλά στοιχεία και επομένως δεν απαιτεί τεράστιες χωρητικότητες και μεγάλη υπολογιστική ισχύ.

Συνήθως, η τελική ομαδοποίηση που δημιουργεί ο αλγόριθμος περιέχει ομάδες με περίπου ίσο αριθμό παρατηρήσεων.

Το μεγάλο μειονέκτημα του αλγορίθμου είναι ότι εξαρτάται από τα αρχικά μητρικά σημεία ή τις αρχικές διαμερίσεις, τα οποία αν δεν είναι σωστά επιλεγμένα μπορεί να οδηγήσουν σε εντελώς διαφορετική ομαδοποίηση από τη φυσική ομαδοποίηση που υπάρχει στα δεδομένα. Για να το ξεπεράσουμε αυτό, τρέχουμε τη μέθοδο με διαφορετικές επιλογές, ώστε να είμαστε σίγουροι για την βέλτιστη λύση (Κούτρας, 2008).

4.7 Two-Step Ομαδοποίηση

Σε σύγκριση με τις προηγούμενες μεθόδους ανάλυσης συστάδων, η Two-Step ανάλυση συστάδων παρέχει περισσότερα πλεονεκτήματα. Με τη μέθοδο αυτή, μπορεί να γίνει αυτόματα η επιλογή του αριθμού των συστάδων, βασισμένη σε στατιστικά κριτήρια. Επίσης, δεν απαιτείται μεγάλη χωρητικότητα και, τέλος, μπορεί να δημιουργήσει ομάδες με κατηγορικές και συνεχείς μεταβλητές λαμβάνοντας υπόψη τα διαφορετικά τους χαρακτηριστικά. Συνοπτικά μπορούμε να πούμε ότι έχει την ικανότητα να αναλύει μεγάλο όγκο δεδομένων αποτελεσματικά. Αυτή η μέθοδος ομαδοποίησης βασίζεται σε ένα μέτρο απόστασης το οποίο δίνει τα καλύτερα αποτελέσματα αν όλες οι μεταβλητές είναι ανεξάρτητες, συνεχείς, ακολουθούν κανονική κατανομή και αν οι κατηγορικές μεταβλητές είναι ανεξάρτητες και ακολουθούν την πολυωνυμική κατανομή. Παρόλο που αυτές οι υποθέσεις σπάνια ικανοποιούνται στην πράξη, ο αλγόριθμος της Two-Step ανταποκρίνεται αρκετά καλά ακόμα και όταν αυτές οι υποθέσεις δεν ικανοποιούνται. Επειδή η μέθοδος ομαδοποίησης δεν περιλαμβάνει κάποιον έλεγχο υπόθεσης και τον υπολογισμό κάποιων επιπέδων σημαντικότητας, πέρα από κάποια περιγραφικά, είναι απολύτως αποδεκτό να ομαδοποιούνται δεδομένα, τα οποία μπορεί να μην ικανοποιούν τις υποθέσεις τις μεθόδου.

Ο αλγόριθμος αυτής της ανάλυσης αναφέρεται σε δύο βήματα. Στο πρωταρχικό βήμα συστάδας (pre-clustering) και στο τελικό βήμα συστάδας (clustering step). Τα βήματα αυτά αναλύονται στη συνέχεια.

Βήμα 1^ο: Pre-clustering: Δημιουργία μικρών ομάδων

Το πρώτο βήμα της two-step ανάλυσης είναι οι δημιουργία πρωταρχικών συστάδων, που ακολουθεί μια διαδοχική διαδικασία ομαδοποίησης (Theodoridis & Koutroumbas, 1999). Ο στόχος του βήματος αυτού είναι να μειώσει το μέγεθος του πίνακα που περιέχει τις αποστάσεις όλων των πιθανών ζευγαριών των εγγραφών. Καθώς διαβάζεται μία εγγραφή ο αλγόριθμος αποφασίζει, βασιζόμενος σε ένα μέτρο απόστασης, αν η συγκεκριμένη εγγραφή θα πρέπει να ομαδοποιηθεί σε κάποια ήδη υπάρχουσα πρωταρχική συστάδα ή θα δημιουργήσει μία καινούρια. Όταν η διαδικασία δημιουργίας πρωταρχικών συστάδων ολοκληρωθεί, όλες οι εγγραφές που βρίσκονται στην ίδια συστάδα θα αντιμετωπίζονται σαν μια οντότητα. Το μέγεθος του πίνακα αποστάσεων δεν εξαρτάται πλέον από τον αριθμό των εγγραφών αλλά από τον αριθμό των πρωταρχικών συστάδων.

Βήμα 2^ο: Ιεραρχική ανάλυση των πρωταρχικών συστάδων

Στο δεύτερο βήμα, πραγματοποιείται ο αλγόριθμος της ιεραρχικής ανάλυσης συστάδων στις πρωταρχικές συστάδες. Ο σχηματίζοντας των συστάδων ιεραρχικά επιτρέπει τη δημιουργία ενός εύρους λύσεων με διαφορετικό αριθμό συστάδων.

Μέτρα αποστάσεις

Αν τα δεδομένα είναι ένας συνδυασμός συνεχών και κατηγορικών μεταβλητών, μπορεί κανείς να χρησιμοποιήσει μόνο το log-likelihood κριτήριο. Όταν τα δεδομένα είναι μόνο συνεχή μπορούμε να χρησιμοποιήσουμε και την ευκλείδεια απόσταση μεταξύ δύο ομάδων. Αναλόγως με το μέτρο της απόστασης που θα επιλεγθεί, οι εγγραφές κατατάσσονται στη συστάδα που οδηγεί στο μεγαλύτερο log-likelihood ή στη συστάδα με την μικρότερη ευκλείδεια απόσταση.

4.8 Εφαρμογή των μεθόδων K-Means και Two-Step

K-means Ομαδοποίηση

Η μέθοδος K-Means δουλεύει επαναληπτικά. Κάνει χρήση της έννοιας του κέντρου της ομάδας και στη συνέχεια προχωρά σε κατάταξη των παρατηρήσεων με βάση την απόστασή τους από τα κέντρα όλων των ομάδων. Όσον αφορά το κέντρο της ομάδας, δεν είναι τίποτε άλλο από τη μέση τιμή για κάθε μεταβλητή όλων των παρατηρήσεων της ομάδας. Για κάθε παρατήρηση, υπολογίζουμε την απόστασή της από τα κέντρα των ομάδων που έχουμε και την κατατάσσουμε στην ομάδα που είναι πιο κοντά.

Τελικά Κέντρα Συστάδων

Αφού κατατάξουμε όλες τις παρατηρήσεις, υπολογίζουμε εκ νέου τα κέντρα και η διαδικασία επαναλαμβάνεται μέχρις ότου δεν υπάρχουν διαφορές ανάμεσα σε δυο διαδοχικές επαναλήψεις. Χρησιμοποιώντας τα τελικά κέντρα των συστάδων, μπορούμε να περιγράψουμε τις ομαδοποιήσεις.

Όταν σε μια στατιστική μέθοδο υπολογίζονται αποστάσεις, θα πρέπει να είμαστε πολύ προσεκτικοί με το γεγονός ότι μπορεί οι μεταβλητές να έχουν διαφορετική μονάδα μέτρησης μεταξύ τους (π.χ. βάρος, ύψος, ηλικία). Μεταβλητές με μεγάλες τιμές θα επηρεάζουν περισσότερο την απόσταση, συγκριτικά με μεταβλητές μικρότερων τιμών. Στην συγκεκριμένη εφαρμογή, από τη στιγμή που τα σκορ των παραγόντων είναι κανονικοποιημένες τιμές, δεν έχουμε τέτοιο πρόβλημα. Συνεπώς, το συνολικό άθροισμα των μέσων είναι 0 και, έτσι, οι θετικές τιμές υποδηλώνουν σκορ πάνω από τον μέσο όρο, ενώ αρνητικές τιμές υποδηλώνουν τιμές κάτω από τον μέσο όρο. Επίσης, μετά από δοκιμές εντοπίστηκε ότι η καλύτερη ερμηνευτικά λύση είναι αυτή των τεσσάρων ομάδων, συνεπώς, αυτή θα παρουσιαστεί στη συνέχεια.

Στον πίνακα 4.1 βλέπουμε ότι η ομάδα 1 έχει σε όλους τους παράγοντες τιμές πάνω από τον μέσο όρο εκτός από τον παράγοντα «Εύκολη και γρήγορη ενημέρωση». Στη συγκεκριμένη συστάδα βλέπουμε ότι οι αναγνώστες παρουσιάζουν το μεγαλύτερο ενδιαφέρον για τις απόψεις συγκεκριμένων δημοσιογράφων σε σχέση με τους αναγνώστες των υπόλοιπων ομάδων.

Πίνακας 4.1

Τελικά Κέντρα Ομάδων

	Ομάδα			
	1	2	3	4
ΠΡΟΣΦΟΡΑ & ΔΩΡΟ	0,22	0,64	0,35	-0,6
ΑΠΟΨΕΙΣ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΗΜΟΣΙΟΓΡΑΦΩΝ	0,85	-0,6	0,27	-0,35
ΨΥΧΑΓΩΓΙΚΑ ΘΕΜΑΤΑ	0,47	0,17	0,04	-0,33
ΕΝΘΕΤΑ	0,43	0,54	-0,21	-0,29
ΕΥΚΟΛΗ & ΓΡΗΓΟΡΗ ΕΝΗΜΕΡΩΣΗ	-0,39	0,03	0,62	-0,24
ΠΟΛΙΤΙΚΗ & ΟΙΚΟΝΟΜΙΚΗ ΕΝΗΜΕΡΩΣΗ	0,35	-0,52	-0,07	0,09

Η ομάδα 2 περιέχει αναγνώστες με μέτριο ενδιαφέρον για τα ψυχαγωγικά θέματα και την εύκολη και γρήγορη ενημέρωση, κυρίως μέσω τηλεόρασης, αλλά ενδιαφέρονται κυρίως για την προσφορά ή το δώρο που περιέχει η εφημερίδα. Το κύριο χαρακτηριστικό της ομάδας 3 είναι η προτίμηση για γρήγορη και χωρίς κόπο ενημέρωσης, κυρίως μέσω της τηλεόρασης, αλλά θα μπορούσαμε να πούμε ότι ενδιαφέρονται και για την προσφορά ή το δώρο που περιέχει η εφημερίδα. Τέλος, στην ομάδα 4 βλέπουμε ότι οι πλειοψηφία των τιμών είναι κάτω από τον μέσο όρο. Θα μπορούσαμε να πούμε ότι παρατηρείται ένα ενδιαφέρον πάνω από τον μέσο όρο για πολιτική και οικονομική ενημέρωση, αλλά γενικότερα είναι αδιάφοροι ως προς τους υπόλοιπους παράγοντες επιλογής εφημερίδας.

Αποστάσεις μεταξύ των Ομάδων

Στον πίνακα 4.2 βλέπουμε τις αποστάσεις μεταξύ των κέντρων των ομάδων και στον πίνακα 4.3 βλέπουμε πόσες εγγραφές έχουν καταταχθεί σε κάθε ομάδα. Συνήθως δε θέλουμε ομάδες με πολύ λίγες εγγραφές εκτός αν αυτές είναι πραγματικά πολύ διαφοροποιημένες από τις υπόλοιπες. Στο δικό μας παράδειγμα δε συμβαίνει κάτι τέτοιο. Σε όλες τις συστάδες έχουμε ικανοποιητικά μεγάλο αριθμό εγγραφών.

Πίνακας 4.2

Αποστάσεις μεταξύ των Τελικών Κέντρων των Ομάδων

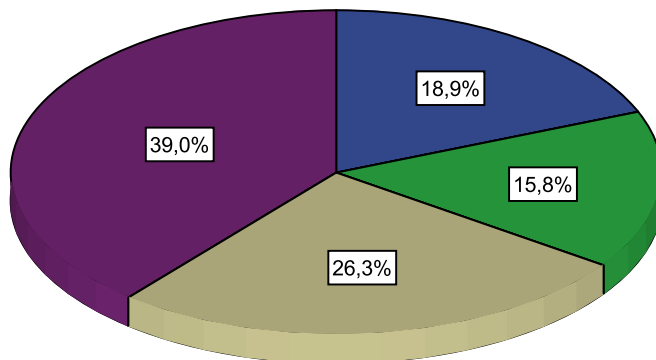
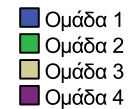
Ομάδα	1	2	3	4
1		1,817	1,464	1,828
2	1,817		1,401	1,725
3	1,464	1,401		1,483
4	1,828	1,725	1,483	

Πίνακας 4.3

Αριθμός εγγραφών σε κάθε Ομάδα

Ομάδα	1	1245,000
	2	1041,000
	2	1737,000
	4	2570,000
	Έγκυρες	6593,000
	Ελλείπουσες	4497,000

Ποσοστό Εγγραφών σε κάθε Ομάδα



Two-Step Ομαδοποίηση

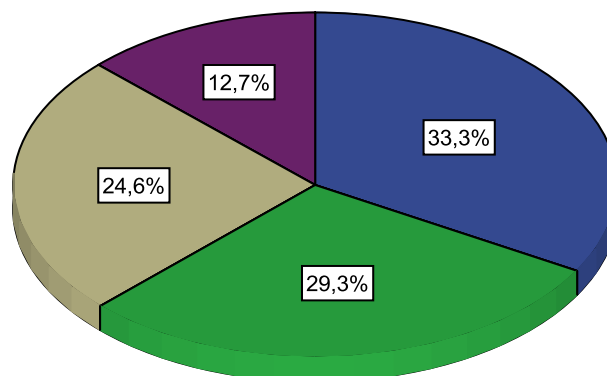
Τέλος, θα εφαρμόσουμε τη μέθοδο Two-Step, η οποία θα προσδιορίσει τον βέλτιστο αριθμό ομάδων αυτόματα. Στον πίνακα 4.4 βλέπουμε τον αριθμό των εγγραφών στις τελικές ομάδες. Η μεγαλύτερη ομάδα, ομάδα 1, περιέχει το 33,3% των ομαδοποιημένων εγγραφών και η μικρότερη, ομάδα 4, το 12,7%.

Πίνακας 4.4

Αριθμός εγγραφών σε κάθε Ομάδα

		N	% του Αθροιστικού	% του Συνόλου
Ομάδα	1	2195	33,3%	19,8%
	2	1933	29,3%	17,4%
	3	1625	24,6%	14,7%
	4	840	12,7%	7,6%
	Αθροιστικά	6593	100,0%	59,4%
	Εξαιρούμενες Εγγραφές	4497		40,6%
	Σύνολο	11090		100,0%

Ποσοστό εγγραφών σε κάθε Ομάδα



Διερευνώντας τη σημαντικότητα κάθε μεταβλητής ξεχωριστά

Όταν ομαδοποιούμε τις εγγραφές, θέλουμε να ξέρουμε πόσο σημαντική είναι η κάθε μεταβλητή για την ομάδα. Στους παρακάτω πίνακες βλέπουμε τη μέση τιμή και την τυπική απόκλιση για κάθε μεταβλητή στις τέσσερις ομάδες. Και εδώ οι τιμές είναι κανονικοποιημένες, οπότε θετικές τιμές των μέσων υποδεικνύουν σκορ πάνω από το μέσο όρο, ενώ αρνητικές τιμές υποδεικνύουν τιμές κάτω από το μέσο όρο.

Πίνακας 4.5**Κέντρα**

		ΠΡΟΣΦΟΡΑ & ΔΩΡΟ		ΑΠΟΨΕΙΣ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΗΜΟΣΙΟΓΡΑΦΩΝ	
		Μέση τιμή	Τυπική απόκλιση	Μέση τιμή	Τυπική απόκλιση
Ομάδα	1	-,5586978	,62783306	-,4660260	,56918201
	2	,0645596	,59913166	,2984994	,53297449
	3	,1744835	,85607414	,6761699	,70334329
	4	,9738240	,65834843	-,7772005	,73283163
	Αθροιστικά	,0000000	,84000955	,0000000	,81818396

Πίνακας 4.6**Κέντρα**

		ΨΥΧΑΓΩΓΙΚΑ ΘΕΜΑΤΑ		ΕΝΘΕΤΑ	
		Μέση τιμή	Τυπική απόκλιση	Μέση τιμή	Τυπική απόκλιση
Ομάδα	1	-,4570698	,69722737	-,3084100	,53013947
	2	-,1074881	,58617818	,4234792	,54131492
	3	,4917199	,54606191	,0680239	,74807533
	4	,4904737	,52579096	-,3001955	,88679203
	Αθροιστικά	,0000000	,73099705	,0000000	,71711875

Πίνακας 4.7

Κέντρα

		ΕΥΚΟΛΗ & ΓΡΗΓΟΡΗ ΕΝΗΜΕΡΩΣΗ		ΠΟΛΙΤΙΚΗ & ΟΙΚΟΝΟΜΙΚΗ ΕΝΗΜΕΡΩΣΗ	
		Μέση τιμή	Τυπική απόκλιση	Μέση τιμή	Τυπική απόκλιση
Ομάδα	1	-,2180065	,48373540	,1675880	,45154523
	2	,5862532	,44402445	-,1024402	,47996322
	3	-,4962636	,49694711	,3250370	,61833561
	4	,1806251	,86338419	-,8309808	,52430580
	Αθροιστικά	,0000000	,68764512	,0000000	,62531785

Παρατηρούμε ότι όλες οι μεταβλητές έχουν τιμές πάνω από το μέσο σε κάποιες ομάδες και κάτω από το μέσο όρο στις υπόλοιπες. Αυτό αποδεικνύει ότι σε κάθε ομάδα είναι διαφορετικές μεταβλητές σημαντικές.

Για παράδειγμα, ο παράγοντας «Προσφορά & Δώρο» έχει σε όλες τις ομάδες τιμές πάνω από τον μέσο όρο, εκτός από τη ομάδα 1. Επίσης, τη μεγαλύτερη τιμή την εμφανίζει στη ομάδα 4, επομένως σε αυτή τη ομάδα έχει τη μεγαλύτερη βαρύτητα ο παράγοντας αυτός.

Διερευνώντας τη σύσταση κάθε ομάδας

Από τη στιγμή που διαμορφωθούν οι ομάδες, σκοπός είναι να δούμε πώς διαφέρουν μεταξύ τους. Στη συνέχεια, θα δούμε διάφορους πίνακες και γραφήματα που θα μας βοηθήσουν να προσδιορίσουμε τη σύνθεση των ομάδων, της Two-Step ομαδοποίησης, και τη σημαντικότητα κάθε μεταβλητής στις ομάδες, με σκοπό να δημιουργήσουμε το προφίλ των αναγνωστών που ανήκουν στην κάθε ομάδα.

Στον πίνακα 4.8 βλέπουμε τα ποσοστά διαφόρων δημογραφικών χαρακτηριστικών σε κάθε ομάδα, όπως το φύλο, η ηλικιακή κατηγορία και η κοινωνική τάξη. Ως προς το φύλο παρατηρούμε διαφοροποιήσεις στις ομάδες 1 και 2 και σημαντική διαφοροποίηση στη ομάδα 4. Ως προς την ηλικία, επίσης, παρατηρούνται διαφοροποιήσεις. Για παράδειγμα, η ομάδα 1 έχει το μεγαλύτερο ποσοστό, 69%, στις ηλικίες από 15 έως 44 χρονών, συγκριτικά με τις υπόλοιπες ομάδες, συνεπώς θα λέγαμε ότι έχει το «νεανικότερο» αναγνωστικό κοινό. Αντίθετα, η ομάδα 4 έχει το μεγαλύτερο ποσοστό, 41%, στις ηλικίες άνω των 45 ετών, σε σχέση με τις υπόλοιπες ομάδες. Συνεπώς, θα λέγαμε ότι έχει τους περισσότερους αναγνώστες μεγαλύτερης ηλικίας. Τέλος, ως προς την κοινωνική τάξη βλέπουμε ότι οι ομάδες διαφοροποιούνται, επίσης. Το μεγαλύτερο ποσοστό αναγνωστών, 56%, που ανήκουν στη χαμηλή κοινωνική τάξη εμφανίζεται στη ομάδα 3, ενώ το μεγαλύτερο ποσοστό αναγνωστών που ανήκουν στην ανώτερη τάξη εμφανίζεται στην ομάδα 4.

Πίνακας 4.8

		Ποσοστό εγγραφών σε κάθε Συστάδα			
		Ομάδα 1	Ομάδα 2	Ομάδα 3	Ομάδα 4
ΦΥΛΟ	ΑΝΔΡΕΣ	48,5%	52,9%	50,0%	67,0%
	ΓΥΝΑΙΚΕΣ	51,5%	47,1%	50,0%	33,0%
ΗΛΙΚΙΑΚΗ ΚΑΤΗΓΟΡΙΑ	15-24	21,3%	20,5%	22,2%	10,1%
	25-34	24,4%	23,1%	23,3%	23,6%
	35-44	23,4%	21,0%	21,4%	25,1%
	45-54	18,4%	18,3%	19,1%	22,3%
	55-64	12,6%	17,2%	14,0%	18,9%
ΚΟΙΝΩΝΙΚΗ ΤΑΞΗ	ΑΝΩΤΕΡΗ (Α/Β/С1)	11,5%	10,6%	9,7%	15,1%
	ΜΕΣΗ (C2)	37,4%	35,8%	34,1%	38,1%
	ΧΑΜΗΛΗ (D/E)	51,1%	53,6%	56,2%	46,8%

Στον πίνακα 4.9 διερευνούμε τις ομάδες με βάση τον αριθμό των εφημερίδων που διαβάζουν οι αναγνώστες τους αλλά και το αν επιλέγουν να διαβάσουν μόνο ημερήσιες ή μόνο κυριακάτικες ή και τα δύο είδη εφημερίδων. Βλέπουμε ότι οι αναγνώστες της ομάδας 2 διαβάζουν τις λιγότερες εφημερίδες σε σχέση με τους αναγνώστες των υπόλοιπων ομάδων, καθώς το 65% έχει απαντήσει ότι δεν διαβάζει καμία εφημερίδα, όμως, ως προς το είδος της εφημερίδας που επιλέγουν να διαβάζουν, το 55%, επιλέγει τις κυριακάτικες εφημερίδες.

Η ομάδα 1 έχει και το μεγαλύτερο ποσοστό αναγνωστών κυριακάτικων εφημερίδων. Παρατηρούμε, επίσης, ότι τις περισσότερες εφημερίδες, δηλαδή πάνω από μία, τις διαβάζουν οι αναγνώστες της ομάδας 4, με ποσοστό 83% (53,6% + 16,7% + 12% + 0,7%), ενώ οι συστάδες 1 και 3 έχουν 42% και 35%, αντίστοιχα.

Πίνακας 4.9

		Ποσοστό εγγραφών σε κάθε Συστάδα			
		Ομάδα 1	Ομάδα 2	Ομάδα 3	Ομάδα 4
ΠΟΣΕΣ ΕΦΗΜΕΡΙΔΕΣ ΕΧΕΤΕ ΔΙΑΒΑΣΕΙ/ΞΕΦΥΛΛΙΣΕΙ ΤΗΝ ΤΕΛΕΥΤΑΙΑ ΕΒΔΟΜΑΔΑ	ΚΑΜΜΙΑ	57,6%	65,3%	59,7%	17,0%
	1-2	34,9%	27,3%	33,6%	53,6%
	3-4	4,9%	5,6%	4,3%	16,7%
	5-9	1,4%	1,1%	2,0%	12,0%
	10 +	1,1%	,7%	,4%	,7%
ΕΙΔΟΣ ΕΦΗΜΕΡΙΔΑΣ	ΜΟΝΟ ΗΜΕΡΗΣΙΕΣ	11,6%	16,1%	15,9%	7,3%
	ΜΟΝΟ ΚΥΡΙΑΚΑΤΙΚΕΣ	57,9%	54,9%	57,0%	57,2%
	ΚΑΙ ΤΙΣ ΔΥΟ	30,5%	29,0%	27,1%	35,4%

Η διαδικασία προσδιορισμού των προφίλ ολοκληρώνεται με την εφαρμογή στατιστικών ελέγχων υποθέσεων στις μεταβλητές που χρησιμοποιήσαμε για ομαδοποίηση. Για τις κατηγορικές μεταβλητές υπολογίζεται η τιμή του χ^2 ελέγχου, ο οποίος συγκρίνει την παρατηρούμενη κατανομή των τιμών μιας μεταβλητής μέσα σε μία ομάδα με τη συνολική κατανομή των τιμών. Μεγάλες τιμές του ελέγχου σε μια ομάδα, υποδηλώνουν ότι η κατανομή των τιμών της μεταβλητής στη συγκεκριμένη ομάδα διαφέρει από τη συνολική κατανομή. Για συνεχείς μεταβλητές, αντί για χ^2 έλεγχο, υπολογίζεται η τιμή του t ελέγχου, ο οποίος συγκρίνει το μέσο της μεταβλητής μέσα στην ομάδα με το συνολικό μέσο.

Στη συγκεκριμένη εφαρμογή, συμμετέχουν μόνο συνεχείς μεταβλητές και επομένως θα χρησιμοποιήσουμε τον t έλεγχο. Τα αποτελέσματα του ελέγχου θα απεικονιστούν γραφικά, με τη χρήση ραβδογράμματος. Η κατεύθυνση των ράβδων υποδηλώνει την σχέση μεταξύ των ομάδων και του συνολικού πληθυσμού. Οι διακεκομμένες γραμμές που εμφανίζονται στα ραβδογράμματα, απεικονίζουν την κρίσιμη τιμή του ελέγχου και με βάση αυτές κρίνουμε αν οι παρατηρηθείσες διαφορές μπορούν να θεωρηθούν στατιστικά σημαντικές. Επιπλέον, οι ράβδοι στην δεξιά ή στην αριστερή πλευρά του Y- άξονα, υποδηλώνουν μέσους ομάδων μεγαλύτερους ή μικρότερους του συνολικού μέσου. Επίσης, όταν μια ράβδος περνάει τις διακεκομμένες γραμμές σημαίνει ότι υπάρχει στατιστικά σημαντική διαφορά μεταξύ του μέσου της ομάδας και του συνολικού μέσου (K. Tsiptsis & A. Chorianoopoulos, 2010). Αυτές οι γραφικές απεικονίσεις αναδεικνύουν τα διαφοροποιά χαρακτηριστικά των ομάδων και συμπληρώνουν το συνολικό προφίλ της κάθε ομάδας, όπως ακολουθεί.

Ομάδα 1 – Περιστασιακοί της ενημέρωσης

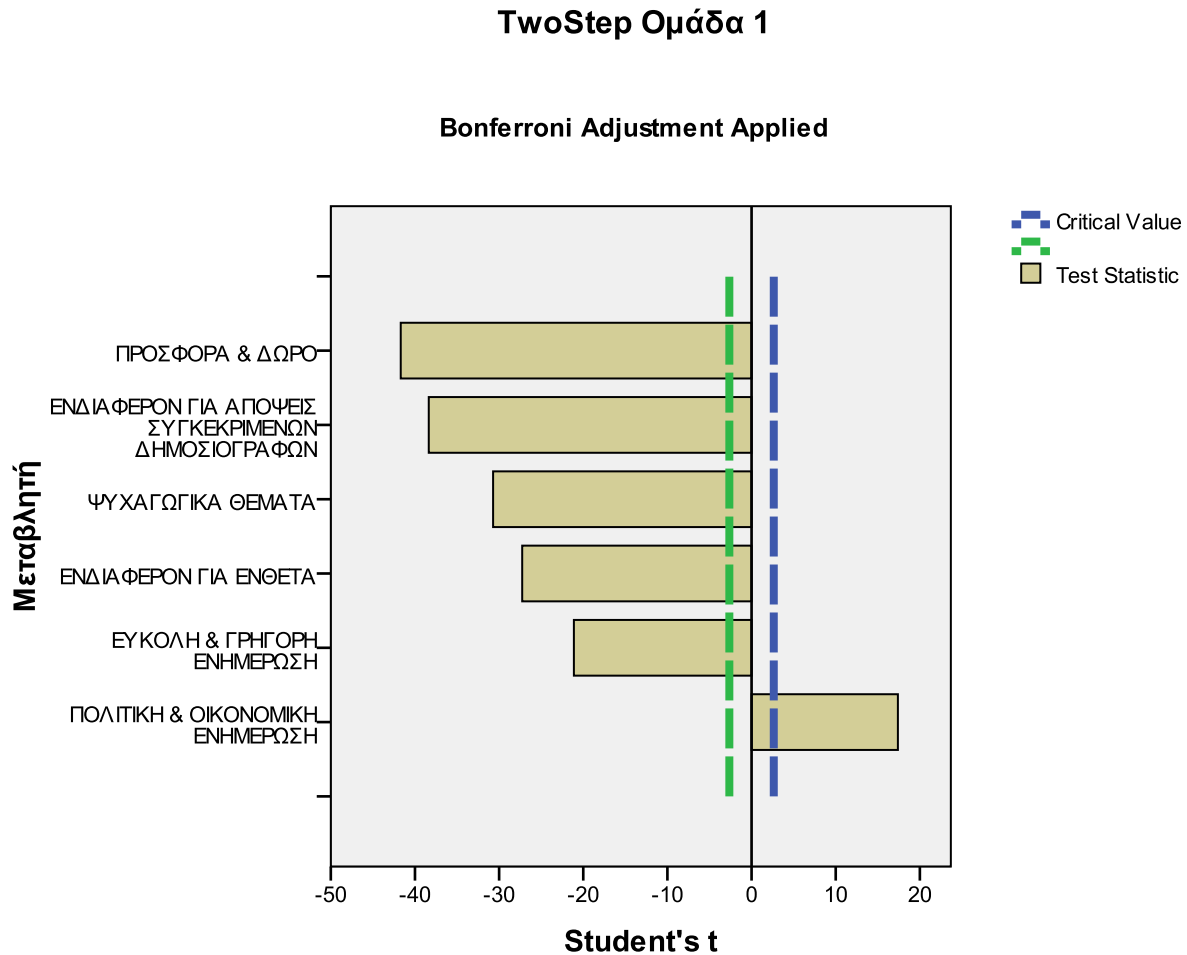
Μέγεθος ομάδας

33,3 %

Προφίλ αναγνωστών

Στην ομάδα αυτή παρατηρούμε ότι οι μέσοι όλων των μεταβλητών, εκτός της μεταβλητής «Πολιτική & Οικονομική ενημέρωση», είναι μικρότεροι του συνολικού μέσου. Συνεπώς, σε αυτή την ομάδα οι αναγνώστες ενδιαφέρονται μόνο για πολιτική και οικονομική ενημέρωση και δεν έχουν κάποιο μέσο ενημέρωσης που βασίζονται για να ενημερωθούν. Η μεταβλητή αυτή χαρακτηρίζει αυτή την ομάδα, αλλά, όπως βλέπουμε και στον πίνακα 4.7, συγκριτικά με τις υπόλοιπες ομάδες η μεταβλητή αυτή δεν έχει σε αυτή την ομάδα την μεγαλύτερη τιμή της. Συνεπώς θα λέγαμε ότι είναι άτομα που ενημερώνονται περιστασιακά, ενδεχομένως για θέματα επικαιρότητας και μόνο, αλλά γενικότερα δεν επιζητούν την ενημέρωση και δεν ενδιαφέρονται για τις προσφορές και τα δώρα ή τα ένθετα που περιέχουν οι εφημερίδες.

Το 35% αυτών των αναγνωστών διαβάζει 1-2 εφημερίδες, κυρίως μόνο κυριακάτικες, 58%, αλλά το 30,5%, διαβάζει και ημερήσιες και κυριακάτικες εφημερίδες. Επίσης, οι αναγνώστες αυτής της ομάδας είναι κυρίως νέοι, ηλικίας 15 έως 44 ετών και ως προς την κοινωνική τάξη, δεν υπάρχει κάποια τάξη που να συγκεντρώνει μεγαλύτερο αριθμό ατόμων από το αναμενόμενο.

Εικόνα 4.1: Σημαντικότητα κάθε μεταβλητής μέσα στην ομάδα 1

Ομάδα 2 – Αναζητητές της είδησης

Μέγεθος ομάδας

29,3 %

Προφίλ αναγνωστών

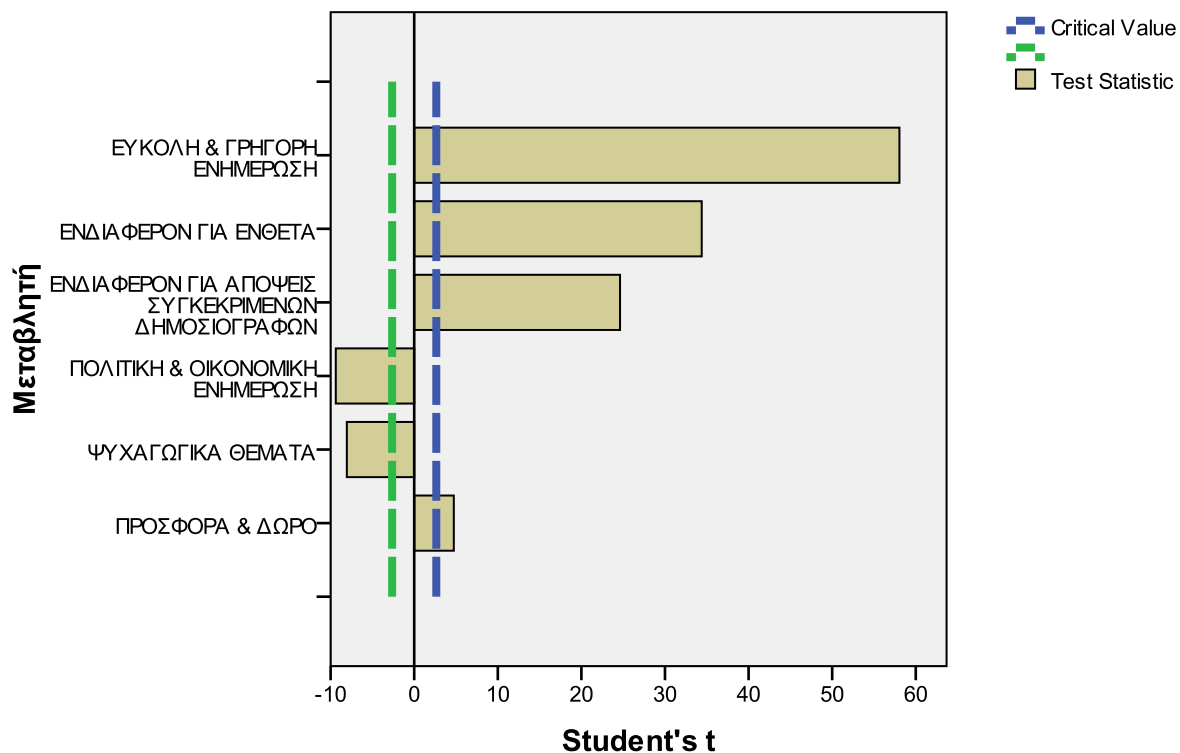
Στην ομάδα 2 παρατηρούμε ότι οι αναγνώστες αναζητούν κυρίως εύκολη και γρήγορη ενημέρωση, αλλά εμφανίζουν και το μεγαλύτερο ενδιαφέρον για τα ένθετα-περιοδικά που υπάρχουν στις εφημερίδες, σε σχέση με τις υπόλοιπες ομάδες, όπως φαίνεται και στον πίνακα 4.6. Με βάση τις αρχικές μεταβλητές που αποτελούν τους παράγοντες «Εύκολη και γρήγορη ενημέρωση» και «Ένθετα» (κεφάλαιο 3, πίνακας 3.17), συμπεραίνουμε ότι τα άτομα αυτής της ομάδας δεν είναι, κατά κύριο λόγο, αναγνώστες εφημερίδων. Ενημερώνονται, κυρίως, από την τηλεόραση αλλά και το ίντερνετ. Θεωρούν ότι οι εφημερίδες απαιτούν πολύ χρόνο, αλλά τα

ένθετά τους αξίζουν και τις κάνουν πιο ενδιαφέρουσες. Παρατηρούμε ότι και οι μεταβλητές «Ενδιαφέρον για απόψεις συγκεκριμένων δημοσιογράφων» και « Προσφορά και Δώρο» έχουν τιμές πάνω από τον μέσο όρο όμως δεν είναι αυτές που χαρακτηρίζουν την ομάδα. Ως προς την καταναλωτική τους συνήθεια, αυτά τα άτομα δε διαβάζουν καμία εφημερίδα, σε ποσοστό 65%, όπως ήταν αναμενόμενο, ενώ το 27% διαβάζει 1-2 εφημερίδες. Ως προς το είδος των εφημερίδων προτιμούν κυρίως τις κυριακάτικες, αλλά έχουν και αρκετά υψηλό ποσοστό στις ημερήσιες, σε σχέση με τις υπόλοιπες ομάδες. Ως προς το δημογραφικό τους προφίλ, οι άντρες (53%) είναι σχετικά περισσότεροι έναντι των γυναικών (47%). Η ομάδα αυτή περιέχει κυρίως άτομα ηλικίας 15 έως 44 και ως προς την κοινωνική τάξη, δεν υπάρχει κάποια τάξη που να συγκεντρώνει μεγαλύτερο αριθμό ατόμων από το αναμενόμενο.

Εικόνα 4.2: Σημαντικότητα κάθε μεταβλητής μέσα στην ομάδα 2

TwoStep Ομάδα 2

Bonferroni Adjustment Applied



Ομάδα 3 – Γεννημένοι αναγνώστες

Μέγεθος ομάδας

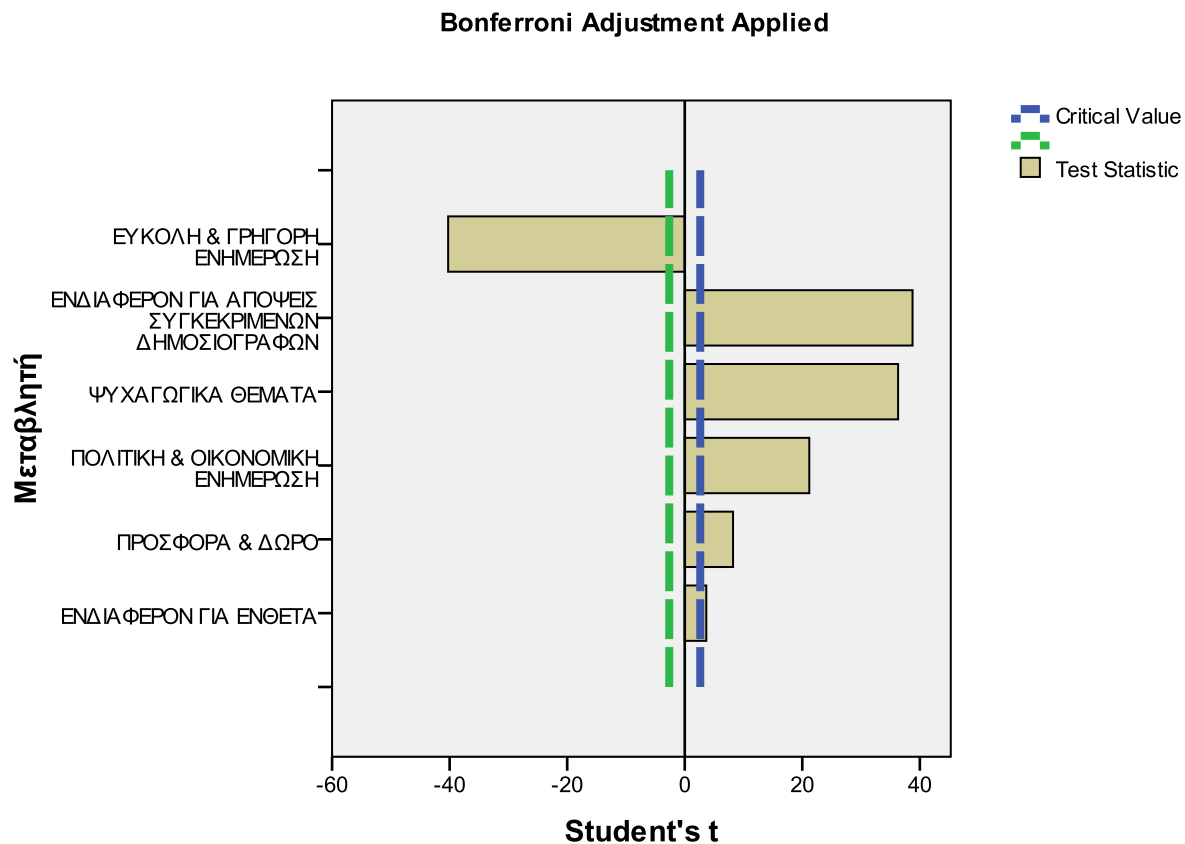
24,6 %

Προφίλ αναγνωστών

Τα άτομα αυτής της ομάδας ενδιαφέρονται κυρίως για τις απόψεις συγκεκριμένων δημοσιογράφων και την πολιτική και οικονομική ενημέρωση, αλλά και για ψυχαγωγικά θέματα. Ενώ, λίγο πάνω από το μέσο όρο βρίσκεται και το ενδιαφέρον τους για τα ένθετα. Από τις αρχικές μεταβλητές που δημιούργησαν αυτούς τους παράγοντες, προκύπτει ότι είναι αναγνώστες που δεν αλλάζουν εύκολα την εφημερίδα τους, διαβάζουν καθημερινά εφημερίδες και για ενημέρωση και για ψυχαγωγία και βασίζονται σε αυτές για την ενημέρωσή τους. Επίσης, διαβάζουν και αρκετά περιοδικά. Ως προς την καταναλωτική τους συμπεριφορά, σε ποσοστό περίπου 40% διαβάζουν περισσότερες από μία εφημερίδες, κυρίως κυριακάτικες αλλά και ημερήσιες, Ως προς το φύλο δεν υπάρχει κάποια διαφοροποίηση. Είναι, κυρίως, άτομα μεσοκατώτερης κοινωνικής τάξης και νεαρά σε ηλικία.

Εικόνα 4.3: Σημαντικότητα κάθε μεταβλητής μέσα στην ομάδα 3

TwoStep Ομάδα 3



Ομάδα 4 - Συλλέκτες

Μέγεθος ομάδας

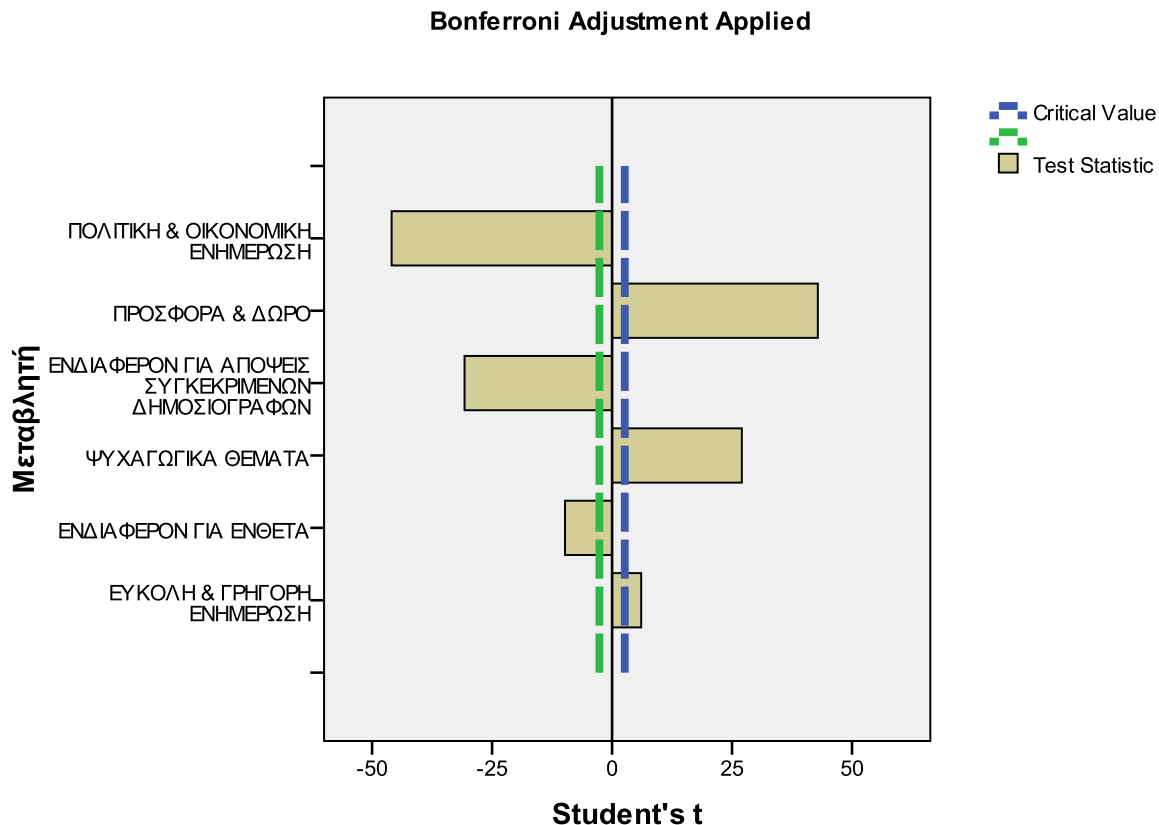
12,7 %

Προφίλ αναγνωστών

Τα άτομα της ομάδας 4 επιλέγουν εφημερίδες και περιοδικά ανάλογα με το δώρο (DVD, CD) και τα ένθετα περιοδικά που τους ενδιαφέρουν, ενώ συμμετέχουν και σε διαγωνισμούς. Είναι αναγνώστες που διαβάζουν εφημερίδες και περιοδικά κυρίως για ψυχαγωγία και όχι για ενημέρωση. Είναι στην πλειοψηφία τους άντρες, 67%, άνω των 35 ετών, σε ποσοστό 66%, και μεσοανώτερης κοινωνικής τάξης. Αγοράζει τις περισσότερες εφημερίδες σε σχέση με τους αναγνώστες των υπόλοιπων ομάδων, καθώς το 83% αγοράζει πάνω από μία εφημερίδες τη βδομάδα, ενώ το 50% διαβάζει περισσότερες από τρεις. Είναι αναγνώστες κυρίως κυριακάτικες εφημερίδων, σε ποσοστό 57%, αλλά έχει, επίσης, και το μεγαλύτερο ποσοστό αναγνωστών, 35%, που διαβάζουν και τα δύο είδη εφημερίδων.

Εικόνα 4.4: Σημαντικότητα κάθε μεταβλητής μέσα στην ομάδα 4

TwoStep Ομάδα 4



Θέλουμε, σαν συνέχεια της ανάλυσης, να εξετάσουμε το προφίλ των αναγνωστών των ημερήσιων και των κυριακάτικων εφημερίδων με βάση της ομάδες των αναγνωστών που δημιουργήσαμε, παρατηρώντας το πώς κατανέμονται οι αναγνώστες της κάθε ομάδας ανάλογα με το είδος της εφημερίδας.

Στον πίνακα 4.10 βλέπουμε ποια εφημερίδα επιλέγουν οι αναγνώστες κάθε ομάδας. Παρατηρούμε ότι η πλειοψηφία των αναγνωστών σε κάθε ομάδα, αν κοιτάξουμε τα ποσοστά ανά γραμμή, επιλέγει να διαβάσει κυριακάτικες εφημερίδες, το οποίο είναι εμφανές και στην ανάλυση που έγινε παραπάνω. Αν εξετάσουμε, όμως, πώς κατανέμονται οι αναγνώστες της κάθε ομάδας μέσα στα δύο είδη των εφημερίδων, κοιτώντας τα ποσοστά ανά στήλη, θα παρατηρήσουμε ότι υπάρχουν διαφοροποιήσεις.

Παρατηρούμε ότι οι αναγνώστες των ημερήσιων εφημερίδων προέρχονται κυρίως από τις ομάδες των «Περιστασιακών της ενημέρωσης» (29,6%), των «Αναζητητών της είδησης» (29,9%) και των «Γεννημένων αναγνωστών» (25,9%). Συνεπώς, θα λέγαμε ότι οι αναγνώστες των ημερήσιων εφημερίδων, ως επί το πλείστον, δεν έχουν ως πρώτη επιλογή την εφημερίδα ως μέσο ενημέρωσης και προτιμούν την τηλεόραση ή το διαδίκτυο. Αγοράζουν, όμως, εφημερίδες κυρίως για τα ένθετα, αλλά στην περίπτωση των «Αναζητητών της είδησης» και για να ενημερώνονται για τα όσα συμβαίνουν καθημερινά. Υπάρχουν, όμως, και αναγνώστες από την ομάδα των «Γεννημένων αναγνωστών», οι οποίοι διαβάζουν καθημερινά εφημερίδες και για ενημέρωση και για ψυχαγωγία και βασίζονται στις εφημερίδες ως κύριο μέσο ενημέρωσης.

Αντίθετα, οι αναγνώστες των κυριακάτικων εφημερίδων προέρχονται κυρίως από τις ομάδες των «Περιστασιακών της ενημέρωσης» και των «Συλλεκτών». Συνεπώς, θα μπορούσαμε να πούμε ότι οι αναγνώστες των κυριακάτικων εφημερίδων είναι άτομα που τις επιλέγουν κυρίως για τα δώρα ή τις προσφορές τους, αλλά και για ψυχαγωγία. Ένα αρκετά μεγάλο ποσοστό, όμως, αγοράζουν κυριακάτικες εφημερίδες και για να ενημερωθούν, όταν θελήσουν, για πολιτικά και οικονομικά θέματα, όχι όμως σε σταθερή βάση.

Και τα δύο είδη εφημερίδων επιλέγουν να διαβάσουν η ομάδα των «Περιστασιακών της ενημέρωσης», αλλά κυρίως η ομάδα των «Συλλεκτών», οι οποίοι διαβάζουν τις κυριακάτικες για ψυχαγωγία και αγοράζουν κυριακάτικες για τα δώρα και τις προσφορές.

Πίνακας 4.10

			ΕΙΔΟΣ ΕΦΗΜΕΡΙΔΑΣ		
			ΜΟΝΟ ΗΜΕΡΗΣΙΕΣ	ΜΟΝΟ ΚΥΡΙΑΚΑΤΙΚΕΣ	ΚΑΙ ΤΙΣ ΔΥΟ
TwoStep Αριθμός Ομάδων	Περιστασιακοί της ενημέρωσης	Στήλη N %	29,6%	32,2%	31,4%
		Γραμμή N %	11,6%	57,9%	30,5%
	Αναζητητές της είδησης	Στήλη N %	29,9%	22,4%	21,9%
		Γραμμή N %	16,1%	54,9%	29,0%
	Γεννημένοι αναγνώστες	Στήλη N %	25,9%	20,3%	17,9%
		Γραμμή N %	15,9%	57,0%	27,1%
	Συλλέκτες	Στήλη N %	14,6%	25,0%	28,7%
		Γραμμή N %	7,3%	57,2%	35,4%

ΚΕΦΑΛΑΙΟ 5

ΣΥΜΠΕΡΑΣΜΑΤΑ

Από το δείγμα των 11.090 ερωτώμενων που είχαμε στην διάθεσή μας και με βάση το μέρος του ερωτηματολογίου που αναλύθηκε, τα κυριότερα ευρήματα είναι τα εξής:

➤ Το 13% του συνόλου των ερωτώμενων διαβάζουν ημερήσιες εφημερίδες, ενώ 25,2% διαβάζουν κυριακάτικες εφημερίδες.

➤ Η αναγνωσιμότητα ημερήσιων και κυριακάτικων εφημερίδων σχετίζεται με το φύλο, την ηλικιακή κατηγορία και την κοινωνική τάξη.

➤ Ως προς την αναγνωστική του συμπεριφορά, η πλειοψηφία του συνόλου των ερωτώμενων απάντησε ότι δεν διαβάζει καμία εφημερίδα, σε ποσοστό 62,3%.

➤ Το μεγαλύτερο ποσοστό των ερωτώμενων, 16,3%, διαβάζουν μόνο κυριακάτικες, ενώ το 8,9% διαβάζουν και τους δύο τύπους εφημερίδων. Το μικρότερο ποσοστό των ερωτώμενων, 4,1%, διαβάζουν μόνο καθημερινές εφημερίδες.

➤ Το μεγαλύτερο ποσοστό αναγνωστών διαβάζει 1-2 εφημερίδες ανεξάρτητα από το αν είναι αναγνώστες μόνο ημερήσιων, μόνο κυριακάτικων ή και των δύο ειδών. Επίσης, οι αναγνώστες που διαβάζουν και ημερήσιες και κυριακάτικες εφημερίδες διαβάζουν, σε μεγαλύτερο ποσοστό, πάνω από 3 εφημερίδες την εβδομάδα.

➤ Οι άντρες είναι οι κύριοι αναγνώστες των ημερήσιων και των κυριακάτικων εφημερίδων. Το μεγαλύτερο ποσοστό αναγνωστριών, όμως, 41,6%, έχουν οι αναγνώστες μόνο κυριακάτικων εφημερίδων. Πλειοψηφικά, όμως, και οι άντρες και οι γυναίκες επιλέγουν να διαβάζουν μόνο κυριακάτικες εφημερίδες, με ποσοστά 50,2% και 65,2% αντίστοιχα.

➤ Το μεγαλύτερο ποσοστό, 65%, στις νεαρότερες ηλικίες, από 15 έως 44, εμφανίζεται στους αναγνώστες που διαβάζουν και ημερήσιες και κυριακάτικες εφημερίδες και δεν επιλέγουν μόνο ένα είδος εξ αυτών. Το μικρότερο ποσοστό στους νέους εμφανίζουν οι αναγνώστες των ημερήσιων εφημερίδων.

➤ Το μεγαλύτερο ποσοστό αναγνωστών που ανήκουν στην ανώτερη τάξη, εμφανίζεται στους αναγνώστες που διαβάζουν και ημερήσιες και κυριακάτικες εφημερίδες, 19,4%. Αντίθετα, το μεγαλύτερο ποσοστό αναγνωστών που ανήκουν στην χαμηλή κοινωνική τάξη, εμφανίζεται στους αναγνώστες μόνο των ημερήσιων εφημερίδων, 50,8%.

➤ Εφαρμόζοντας την παραγοντική ανάλυση με τις μεθόδους εκτίμησης των κύριων συνιστωσών και της μέγιστης πιθανοφάνειας παρατηρήσαμε τις διαφορές και τις ομοιότητές τους. Συγκεκριμένα, με τη μέθοδο της μέγιστης πιθανοφάνειας ερμηνεύσαμε το 39% της συνολικής μεταβλητότητας, ενώ με τη μέθοδο των κύριων συνιστωσών το 60,5%. Χρησιμοποιώντας την τεχνική της ορθογώνιας περιστροφής και επιλέγοντας, στην συσχετισμένη περίπτωση, ως κατώτατο όριο για τις επιβαρύνσεις την τιμή 0,30, γίνεται κατανοητό ποιες μεταβλητές επιβαρύνουν τον κάθε παράγοντα. Συγκρίνοντας τις

δυο αυτές μεθόδους εκτίμησης, παρατηρούμε ότι έχουν διαφορετικές επιβαρύνσεις σε όλες τις μεταβλητές, αλλά έχουν κάνει σχεδόν τις ίδιες ομαδοποιήσεις μεταβλητών. Μόνο στον παράγοντα 6 παρατηρείται διαφοροποίηση, η οποία κάνει τον παράγοντα καλύτερα επεξηγήσιμο με τη μέθοδο της μέγιστης πιθανοφάνειας.

➤ Μετά την εφαρμογή της παραγοντικής ανάλυσης για τη στάση των ερωτώμενων απέναντι στις εφημερίδες, στην οποία χρησιμοποιήθηκαν 17 φράσεις/ προτάσεις σχετικά με τα έντυπα και την ενημέρωση, προέκυψαν, με τη μέθοδο εκτίμησης της μέγιστης πιθανοφάνειας, 6 παράγοντες που ονομάσαμε:

- ❖ ΠΡΟΣΦΟΡΑ ΚΑΙ ΔΩΡΟ
- ❖ ΑΠΟΨΕΙΣ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΔΗΜΟΣΙΟΓΡΑΦΩΝ
- ❖ ΨΥΧΑΓΩΓΙΚΑ ΘΕΜΑΤΑ
- ❖ ΕΝΘΕΤΑ
- ❖ ΕΥΚΟΛΗ ΚΑΙ ΓΡΗΓΟΡΗ ΕΝΗΜΕΡΩΣΗ
- ❖ ΠΟΛΙΤΙΚΗ ΚΑΙ ΟΙΚΟΝΟΜΙΚΗ ΕΝΗΜΕΡΩΣΗ

➤ Βασιζόμενοι στους παράγοντες που προέκυψαν από την παραγοντική ανάλυση προχωρήσαμε σε ανάλυση κατά συστάδες, εφαρμόζοντας την K-means ομαδοποίηση και την Two-Step. Οι δύο αυτές μέθοδοι καταλήγουν σε παρόμοιες ομαδοποιήσεις εγγραφών.

➤ Από την ανάλυση κατά συστάδες καταλήξαμε σε 4 ομάδες. Διερευνώντας τις ομαδοποιήσεις, στις οποίες αυτόματα κατέληξε η μέθοδος Two-Step, δημιουργήθηκαν οι εξής ομάδες: οι «Περιστασιακοί της ενημέρωσης» με ποσοστό 33,3%, οι «Αναζητητές της είδησης» με ποσοστό 29,3%, οι «Γεννημένοι αναγνώστες» με ποσοστό 24,6% και οι «Συλλέκτες» με ποσοστό 12,7%.

➤ Οι αναγνώστες της ομάδας «Περιστασιακοί της ενημέρωσης» ενδιαφέρονται μόνο για πολιτική και οικονομική ενημέρωση και δεν έχουν κάποιο μέσο ενημέρωσης που βασίζονται για να ενημερωθούν. Θα λέγαμε ότι είναι άτομα που ενημερώνονται περιστασιακά, ενδεχομένως για θέματα τα επικαιρότητας και μόνο, αλλά γενικότερα δεν επιζητούν την ενημέρωση και δεν ενδιαφέρονται για τις προσφορές και τα δώρα ή τα ένθετα που περιέχουν οι εφημερίδες.

➤ Οι αναγνώστες τις ομάδας «Αναζητητές της είδησης» δεν είναι, κατά κύριο λόγο, αναγνώστες εφημερίδων. Ενημερώνονται, κυρίως, από την τηλεόραση αλλά και το διαδίκτυο. Ως προς το είδος των εφημερίδων προτιμούν κυρίως τις κυριακάτικες, αλλά έχουν και αρκετά υψηλό ποσοστό στις ημερήσιες, σε σχέση με τις υπόλοιπες ομάδες. Ως προς το δημογραφικό τους προφίλ, οι άντρες (53%) είναι σχετικά περισσότεροι έναντι των γυναικών (47%). Η ομάδα αυτή περιέχει κυρίως άτομα ηλικίας 15 έως 44.

➤ Οι αναγνώστες τις ομάδας «Γεννημένοι αναγνώστες» ενδιαφέρονται κυρίως για τις απόψεις συγκεκριμένων δημοσιογράφων και την πολιτική και οικονομική ενημέρωση, αλλά και για ψυχαγωγικά θέματα. Διαβάζουν κυρίως κυριακάτικες αλλά και ημερήσιες και σε ποσοστό περίπου 40% διαβάζουν περισσότερες από μία εφημερίδες

➤ Οι αναγνώστες της ομάδας «Συλλέκτες» επιλέγουν εφημερίδες και περιοδικά ανάλογα με το δώρο (DVD, CD) και τα ένθετα περιοδικά που τους ενδιαφέρουν, ενώ συμμετέχουν και σε διαγωνισμούς. Είναι αναγνώστες που διαβάζουν εφημερίδες και περιοδικά κυρίως για ψυχαγωγία και όχι για ενημέρωση. Είναι στην πλειοψηφία τους

άντρες, 67%, άνω των 35 ετών σε ποσοστό 66% και μεσοανώτερης κοινωνικής τάξης. Είναι αναγνώστες, κυρίως, κυριακάτικων εφημερίδων, αλλά έχει, επίσης, και το μεγαλύτερο ποσοστό αναγνωστών που διαβάζουν και τα δύο είδη εφημερίδων. Το 83% αγοράζει πάνω από μία εφημερίδες τη βδομάδα, ενώ το 50% διαβάζει περισσότερες από τρεις.

➤ Η πλειοψηφία των αναγνωστών σε κάθε ομάδα επιλέγει να διαβάσει κυριακάτικες εφημερίδες.

➤ Οι αναγνώστες των ημερήσιων εφημερίδων, ως επί το πλείστον, δεν έχουν ως πρώτη επιλογή την εφημερίδα ως μέσο ενημέρωσης και προτιμούν την τηλεόραση ή το διαδίκτυο. Αγοράζουν, όμως, εφημερίδες κυρίως για τα ένθετα, αλλά στην περίπτωση των «Αναζητητών της είδησης» και για να ενημερώνονται για τα όσα συμβαίνουν καθημερινά. Υπάρχουν, όμως, και αναγνώστες ημερήσιων εφημερίδων που διαβάζουν καθημερινά εφημερίδες για ενημέρωση και για ψυχαγωγικά θέματα και έχουν τις εφημερίδες ως κύριο μέσο ενημέρωσης.

➤ Οι αναγνώστες των κυριακάτικων εφημερίδων είναι άτομα που τις επιλέγουν κυρίως για τα δώρα ή τις προσφορές τους, αλλά και για ψυχαγωγία. Ένα αρκετά μεγάλο ποσοστό, όμως, αγοράζουν κυριακάτικες εφημερίδες και για να ενημερωθούν για πολιτικά και οικονομικά θέματα, όχι όμως σε σταθερή βάση.

➤ Οι αναγνώστες που διαβάζουν και τα δύο είδη εφημερίδων, διαβάζουν τις ημερήσιες για ενημέρωση και αγοράζουν κυριακάτικες για τα δώρα και τις προσφορές.

ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ

ΕΛΛΗΝΙΚΕΣ

Ηλιόπουλος, Γ. (2008). *Πολυμεταβλητή Ανάλυση. Πανεπιστημιακές Σημειώσεις*. Πανεπιστήμιο Πειραιώς, Πειραιάς.

Καρλής, Δ. (2005). *Πολυμεταβλητή Στατιστική Ανάλυση*. Εκδόσεις: Σταμούλης Α., Αθήνα.

Κούτρας, Μ. (2008). *Εφαρμοσμένη Πολυμεταβλητή Ανάλυση-Ανάλυση κατά συστάδες*. Πανεπιστήμιο Πειραιώς, Πειραιάς.

Πανάρετος, Ι. και Ξεκαλάκη, Ε. (1995). *Εισαγωγή στην Πολυμεταβλητή Στατιστική Ανάλυση*. Εκδόσεις: Πανάρετος Ι., Αθήνα.

Σιώμκος, Γ. Ι. και Βασιλακοπούλου, Α. Ι. (2005). *Εφαρμογή Μεθόδων Ανάλυσης στην Έρευνα Αγοράς*. Εκδόσεις: Σταμούλης Α., Αθήνα.

ΞΕΝΟΓΛΩΣΣΕΣ

Hu, L. and Bentler, M. P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternative. *Structural Equation Modelling*, **6**, 1-55.

Johnson, D. E. (1998). *Applied Multivariate Methods for Data Analysis*. Pacific Grove: Duxbury Press.

Johnson, R. A. and Wichern, W. D. (1998). *Applied Multivariate Statistical Analysis*. 4th ed. New Jersey: Prentice Hall.

Mardia, K.V. , Kent, J.T. & Bibby, J.M. (1979). *Multivariate Analysis*. London: Academic Press.

Rencher, C. A. (2002). *Methods of Multivariate Analysis*. 2nd ed. New Jersey: John Wiley & Sons, Inc.

Sharma, S. (1995). *Applied Multivariate Techniques*. New Jersey: John Wiley & Sons, Inc.

Theodoridis, S. and Koutroumpas, K. (1999). *Pattern recognition*. New York: American Press.

Tsiptsis, P. and Chorianopoulos, A. (2010). *Data Mining Techniques in CRM*. New Jersey: John Wiley & Sons.

ΙΣΤΟΣΕΛΙΔΕΣ

http://www.norusis.com/pdf/SPC_v13.pdf

