



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΟΡΓΑΝΩΣΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΣΤΗ ΔΙΟΙΚΗΣΗ ΕΠΙΧΕΙΡΗΣΕΩΝ ΓΙΑ ΣΤΕΛΕΧΗ (ΠΜΣ.ΔΕ.Σ)
(Ε-ΜΒΑ)**

Διπλωματική Εργασία

**ΑΝΑΛΥΣΗ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ ΤΩΝ ΣΤΟΙΧΕΙΩΝ
ΑΠΑΣΧΟΛΗΣΗΣ ΣΤΗΝ ΕΥΡΩΠΗ
&
ΟΜΑΔΟΠΟΙΗΣΗ ΧΩΡΩΝ ΜΕ ΤΗ ΣΥΣΣΩΡΕΥΤΙΚΗ ΜΕΘΟΔΟ
ΑΝΑΛΥΣΗΣ ΚΑΤΑ ΣΥΣΤΑΔΕΣ**

Κωνσταντίνος Τσιβερδής

Πειραιάς, Νοέμβριος 2013

Παράρτημα Β: Βεβαίωση Εκπόνησης Διπλωματικής Εργασίας



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΟΡΓΑΝΩΣΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΣΤΗ ΔΙΟΙΚΗΣΗ ΕΠΙΧΕΙΡΗΣΕΩΝ ΓΙΑ ΣΤΕΛΕΧΗ**

ΒΕΒΑΙΩΣΗ ΕΚΠΟΝΗΣΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

(περιλαμβάνεται ως ξεχωριστή (δεύτερη) σελίδα στο σώμα της διπλωματικής εργασίας)

«Δηλώνω υπεύθυνα ότι η διπλωματική εργασία για τη λήψη του μεταπτυχιακού τίτλου σπουδών, του Πανεπιστημίου Πειραιώς, στη Διοίκηση Επιχειρήσεων για Στελέχη : E-MBA» με τίτλο ...**Ανάλυση κυρίων συνιστωσών των στοιχείων απασχόλησης στην Ευρώπη & ομαδοποίηση χωρών με τη συσσωρευτική μέθοδο ανάλυσης κατά συστάδες**... έχει συγγραφεί από εμένα αποκλειστικά και στο σύνολό της. Δεν έχει υποβληθεί ούτε έχει εγκριθεί στο πλαίσιο κάποιου άλλου μεταπτυχιακού προγράμματος ή προπτυχιακού τίτλου σπουδών, στην Ελλάδα ή στο εξωτερικό, ούτε είναι εργασία ή τμήμα εργασίας ακαδημαϊκού ή επαγγελματικού χαρακτήρα.

Δηλώνω επίσης υπεύθυνα ότι οι πηγές στις οποίες ανέτρεξα για την εκπόνηση της συγκεκριμένης εργασίας, αναφέρονται στο σύνολό τους, κάνοντας πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Υπογραφή Μεταπτυχιακού Φοιτητή/τριας.....

Όνοματεπώνυμο.....**Τσιβερδής Κωνσταντίνος**.....

Ημερομηνία.....**27/11/2013**.....

ΑΝΑΛΥΣΗ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ ΤΩΝ ΣΤΟΙΧΕΙΩΝ ΑΠΑΣΧΟΛΗΣΗΣ ΣΤΗΝ ΕΥΡΩΠΗ & ΟΜΑΔΟΠΟΙΗΣΗ ΧΩΡΩΝ ΜΕ ΤΗ ΣΥΣΣΩΡΕΥΤΙΚΗ ΜΕΘΟΔΟ ΑΝΑΛΥΣΗΣ ΚΑΤΑ ΣΥΣΤΑΔΕΣ

Κωνσταντίνος Τσιβερδής

Σημαντικοί όροι: Απασχόληση, Κύριες Συνιστώσες, Συστάδες, Ευρώπη.

ΠΕΡΙΛΗΨΗ

Στο σύγχρονο παγκοσμιοποιημένο περιβάλλον υψηλού ανταγωνισμού, ένας χρηματοπιστωτικός οργανισμός θα πρέπει να είναι σε θέση να αναπτύσσει και να διαμορφώνει κατάλληλες στρατηγικές, πολιτικές και τακτικές, αναλύοντας ταχύτατα μεγάλους όγκους δεδομένων και καταλήγοντας σε αξιόπιστα και χρήσιμα συμπεράσματα, που θα τον βοηθούν να παίρνει σωστές αποφάσεις.

Η κάθε κίνηση θα πρέπει να σχεδιάζεται, να αποφασίζεται και να υλοποιείται άμεσα, καθώς οι εξελίξεις στο παγκόσμιο χρηματο-οικονομικό περιβάλλον εξελίσσονται με ραγδαίους ρυθμούς. Στο σημείο αυτό, η Πολυμεταβλητή Στατιστική Ανάλυση μπορεί να αποτελέσει σημαντικό εργαλείο για την ταχεία επεξεργασία πολλών δεδομένων ταυτόχρονα, συμπιέζοντας τον όγκο της πληροφορίας και οδηγώντας σε χρήσιμα και εύκολα κατανοητά αποτελέσματα για το κέντρο λήψης αποφάσεων ενός χρηματοπιστωτικού οργανισμού.

Σημαντικοί πελάτες ενός χρηματοπιστωτικού οργανισμού σε μία χώρα είναι το κομμάτι του πληθυσμού που εργάζεται και συνεπώς έχει εισόδημα. Τόσο το ποσοστό των εργαζόμενων της χώρας, όσο και το προφίλ τους (ηλικία, μορφωτικό επίπεδο, φύλο), αποτελούν κρίσιμης σημασίας δεδομένα για τις πολιτικές δραστηριοποίησης που θα ακολουθήσει ο συγκεκριμένος χρηματοπιστωτικός οργανισμός. Τα δεδομένα απασχόλησης θα πρέπει πολύ γρήγορα να μετουσιωθούν σε αξιόπιστη πληροφορία για τον οργανισμό, μέσα από την κατάλληλη στατιστική ανάλυση.

Σκοπός της μελέτης είναι η χρήση κατάλληλων τεχνικών Πολυμεταβλητής Στατιστικής Ανάλυσης για τη χαρτογράφηση του εργασιακού τοπίου της Ευρώπης, ούτως ώστε κάποιος χρηματοπιστωτικός οργανισμός που στοχεύει να δραστηριοποιηθεί σε συγκεκριμένες χώρες ή ομάδες χωρών της Ευρώπης, να είναι σε θέση να διαμορφώσει το κατάλληλο μίγμα πολιτικών του.

Η διπλωματική εργασία αποτελείται από έξι κεφάλαια:

Στο 1^ο κεφάλαιο γίνεται μία πρώτη παρουσίαση και σχολιασμός των ανεπεξέργαστων δεδομένων, όπως αυτά ελήφθησαν από τη βάση δεδομένων της Eurostat. Παρατίθενται πίνακες και διαγράμματα για τις 33 χώρες που ανήκουν γεωγραφικά στην Ευρώπη, με τα χαρακτηριστικά απασχόλησής τους (ποσοστά απασχόλησης ανά ηλικιακή ομάδα, ανά φύλο, ανά μορφωτικό επίπεδο, κλπ).

Στο 2^ο κεφάλαιο, επιλέγονται 12 συγκεκριμένα χαρακτηριστικά απασχόλησης που θα αποτελέσουν αντικείμενο περαιτέρω ανάλυσης και γίνεται μία πιο εκτενής περιγραφική στατιστική ανάλυση.

Στο 3^ο κεφάλαιο παρατίθεται το θεωρητικό υπόβαθρο που σχετίζεται με την πολυμεταβλητή ανάλυση και ειδικά με την ανάλυση σε κύριες συνιστώσες.

Στο 4^ο κεφάλαιο, εφαρμόζεται η τεχνική της ανάλυσης κύριων συνιστωσών (Principal Component Analysis) για τη μείωση της διάστασης των 12 αρχικώς επιλεγμένων χαρακτηριστικών απασχόλησης και περιγράφεται η κατασκευή δύο νέων «δεικτών» απασχόλησης.

Στο 5^ο κεφάλαιο, οι χώρες ομαδοποιούνται με την τεχνική της ανάλυσης κατά συστάδες (Cluster Analysis), βάσει των δύο κατασκευασμένων «δεικτών» απασχόλησης. Χρησιμοποιείται η συσσωρευτική μέθοδος ομαδοποίησης και γίνεται σύγκριση των 3 ομάδων που προκύπτουν αφενός από τη χρήση των αρχικώς επιλεγμένων δεδομένων, αφετέρου από τη χρήση των κύριων συνιστωσών που προέκυψαν στο 2^ο κεφάλαιο.

Στο 6^ο κεφάλαιο, διατυπώνονται συμπεράσματα σχετικά με την πιστότητα της ομαδοποίησης βάσει των δύο δεικτών σε σχέση με την ομαδοποίηση αναφοράς, ενώ επαναλαμβάνεται το πείραμα με ομαδοποίηση βάσει των τριών πρώτων κυρίων συνιστωσών. Τέλος, διατυπώνονται κάποιες σκέψεις ως προς την αξία της ομαδοποίησης για τους χρηματοοικονομικούς οργανισμούς.

Σημειώνεται ότι καταβλήθηκε προσπάθεια να μην παρατεθούν (εκτός του 3^{ου} κεφαλαίου) οι πολύπλοκοι μαθηματικοί τύποι που αποτελούν τη θεωρητική βάση της στατιστικής ανάλυσης των κεφαλαίων 4 και 5, καθώς κρίθηκε σκόπιμο το περιεχόμενο της εργασίας να παραμείνει όσο το δυνατόν απλό και κατανοητό. Δόθηκε μεγάλη βαρύτητα στην κατανόηση της όλης φιλοσοφίας της ανάλυσης, η οποία εξηγείται βήμα-βήμα παράλληλα με την παράθεση των απαραίτητων διαγραμμάτων, πινάκων και δένδρογραμμάτων.

Η ανάλυση πραγματοποιήθηκε με το στατιστικό πακέτο SPSS, ενώ χρησιμοποιήθηκε επικουρικά και το MS Excel.

ΕΥΧΑΡΙΣΤΙΕΣ

Ευχαριστώ το Θεό που μου έδωσε όλες τις προϋποθέσεις –με βασικότερη την υγεία– ώστε να μπορέσω να παρακολουθήσω με επιτυχία το μεταπτυχιακό πρόγραμμα.

Ευχαριστώ την οικογένειά μου για τη συνεχή στήριξη σε όλα τα επίπεδα όλα τα χρόνια της ακαδημαϊκής μου καριέρας.

Ευχαριστώ τον επιβλέποντα κ. Σφακιανάκη για την πολύτιμη βοήθειά του σε όλα τα επίπεδα και στάδια της εργασίας αυτής -από την ανάληψή της έως και την τελευταία στιγμή- μα και για το γεγονός ότι η συνεργασία μας ήταν απολύτως αρμονική και αληθινά ευχάριστη.

Ευχαριστώ φίλους, συμφοιτητές και καθηγητές για τις ωραίες στιγμές που μου χάρισαν (ειδική μνεία στα υπέροχα καλοκαίρια στο Πόρτο Χέλι...)!

Ευχαριστώ και τους υπαίτιους των κακών στιγμών που (αναπόφευκτα) προέκυψαν στα χρόνια του μεταπτυχιακού γιατί με τον τρόπο τους με βοήθησαν να ωριμάσω και ίσως να βελτιωθώ.

Τέλος, ευχαριστώ μέσα από την καρδιά μου τον «αιώνιο Πρόεδρο» –κατά κόσμον γνωστό ως Μιχάλη Μποζούδη– για την τεχνική υποστήριξη και όχι μόνο.

ΚΑΤΑΣΤΑΣΗ ΠΙΝΑΚΩΝ

Πίνακας 1- Ετήσιος ρυθμός αύξησης απασχόλησης από το 2000 ως το 2010, για την «Ευρώπη των 27», την Ιαπωνία και τις ΗΠΑ.	3
Πίνακας 2 - Ετήσιος ρυθμός αύξησης απασχόλησης για τις χώρες της Ευρώπης ανά φύλο, για τα έτη 2000, 2005 και 2010.	5
Πίνακας 3 - Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών, Ιαπωνίας και ΗΠΑ, 2000-2010.....	6
Πίνακας 4 - Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών, Ιαπωνίας και ΗΠΑ, για επιλεγμένες κατηγορίες εργαζόμενων, έτη 2000, 2005 και 2010	7
Πίνακας 5 - Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών, ανά κατηγορία μορφωτικού επιπέδου για την ηλικιακή ομάδα 25-64, έτους 2010.....	8
Πίνακας 6 - Ετήσια συνολικά ποσοστά απασχόλησης Ευρωπαϊκών χωρών, έτους 2010.	9
Πίνακας 7 - Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών ανά φύλο, έτους 2010.....	11
Πίνακας 8 - Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών ανά ηλικιακή ομάδα, έτους 2010.....	13
Πίνακας 9 - Ετήσια ποσοστά μερικής απασχόλησης και ατόμων με 2 ^η δουλειά στην Ευρώπη, για τα έτη 2000-2010	15
Πίνακας 10 - Ετήσια ποσοστά μερικής απασχόλησης Ευρωπαϊκών χωρών ανά φύλο, έτους 2010.....	16
Πίνακας 11 - Ετήσια ποσοστά απασχόλησης με συμβόλαιο περιορισμένης διάρκειας στις χώρες της Ευρώπης, έτους 2010.....	18
Πίνακας 12 - Βασικά περιγραφικά μέτρα της μεταβλητής “Male”	20
Πίνακας 13 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Male”	21
Πίνακας 14 - Βασικά περιγραφικά μέτρα της μεταβλητής “Female”	22
Πίνακας 15 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Female”	23
Πίνακας 16 - Βασικά περιγραφικά μέτρα της μεταβλητής “Low_Educated”	25
Πίνακας 17 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Low_Educated”	26
Πίνακας 18 - Βασικά περιγραφικά μέτρα της μεταβλητής “Mid_Educated”	27
Πίνακας 19 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Mid_Educated”	28
Πίνακας 20 - Βασικά περιγραφικά μέτρα της μεταβλητής “High_Educated”	29
Πίνακας 21 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “High_Educated”	30
Πίνακας 22 - Βασικά περιγραφικά μέτρα της μεταβλητής “Young”	31
Πίνακας 23 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Young”	32
Πίνακας 24 - Βασικά περιγραφικά μέτρα της μεταβλητής “Mid_Age”	33
Πίνακας 25 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Mid_Age”	34
Πίνακας 26 - Βασικά περιγραφικά μέτρα της μεταβλητής “Older”	35
Πίνακας 27 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Older”	36
Πίνακας 28 - Βασικά περιγραφικά μέτρα της μεταβλητής “With_Second_Job”	37
Πίνακας 29 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “With_Second_Job”	38
Πίνακας 30 - Βασικά περιγραφικά μέτρα της μεταβλητής “Part_Time_Male”	39
Πίνακας 31- Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Part_Time_Male”	40
Πίνακας 32 - Βασικά περιγραφικά μέτρα της μεταβλητής “Part_Time_Female”	41
Πίνακας 33 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Part_Time_Female”	42
Πίνακας 34 - Βασικά περιγραφικά μέτρα της μεταβλητής “Limited_Duration_Contract”	43
Πίνακας 35 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Limited_Duration_Contract”	44
Πίνακας 36 - Συντελεστές συσχέτισης μεταξύ των 12 μεταβλητών απασχόλησης που εξετάζονται.....	67
Πίνακας 37 - Ιδιοτιμές των κύριων συνιστωσών και ποσοστά κάλυψης της συνολικής αρχικής διακύμανσης	69
Πίνακας 38 - Ποσοστά κάλυψης της αρχικής διακύμανσης ανά μεταβλητή, από τις δύο πρώτες κύριες συνιστώσες.....	70
Πίνακας 39 - Κόστος χρήσης κύριων συνιστωσών.....	72
Πίνακας 40 - Ιδιοδιανύσματα των δύο πρώτων κύριων συνιστωσών	73
Πίνακας 41 - Συντελεστές χαρακτηριστικών απασχόλησης με τις δύο πρώτες κύριες συνιστώσες, μετά την περιστροφή των αξόνων.....	75
Πίνακας 42 - Ερμηνευτική ικανότητα των 2 πρώτων κύριων συνιστωσών πριν και μετά την περιστροφή... ..	77
Πίνακας 43 - Διαδικασία συσσωρευτικής συνένωσης των Ευρωπαϊκών χωρών σε ομάδες, βάσει των 11 αρχικών χαρακτηριστικών απασχόλησης.....	85
Πίνακας 44 - Οι 3 ομάδες των Ευρωπαϊκών χωρών, βάσει των 11 αρχικών χαρακτηριστικών απασχόλησης	87

Πίνακας 45 - Διαδικασία συσσωρευτικής συνένωσης των Ευρωπαϊκών χωρών σε ομάδες, βάσει των 2 «δεικτών» απασχόλησης.....	88
Πίνακας 46 - Οι 3 ομάδες των Ευρωπαϊκών χωρών, βάσει των 2 «δεικτών» απασχόλησης.....	90
Πίνακας 47 – Σύγκριση των ομάδων που προκύπτουν από την ομαδοποίηση των Ευρωπαϊκών χωρών βάσει των 11 αρχικών χαρακτηριστικών απασχόλησης (αριστερά) και των 2 PCA δεικτών απασχόλησης (δεξιά).....	93
Πίνακας 48 - Ποσοστά κάλυψης της αρχικής διακύμανσης ανά μεταβλητή, από τις δύο πρώτες κύριες συνιστώσες.....	95
Πίνακας 49 - Ερμηνευτική ικανότητα των 3 πρώτων κύριων συνιστωσών πριν και μετά την περιστροφή...	96
Πίνακας 50 - Συντελεστές χαρακτηριστικών απασχόλησης με τις τρεις πρώτες κύριες συνιστώσες, μετά την περιστροφή των αξόνων.....	96
Πίνακας 51 - Διαδικασία συσσωρευτικής συνένωσης των Ευρωπαϊκών χωρών σε ομάδες, βάσει των 3 «δεικτών» απασχόλησης.....	98
Πίνακας 52 - Οι 4 ομάδες των Ευρωπαϊκών χωρών, βάσει των 3 «δεικτών» απασχόλησης.....	100

ΚΑΤΑΣΤΑΣΗ ΔΙΑΓΡΑΜΜΑΤΩΝ

Διάγραμμα 1 - Ετήσιος ρυθμός αύξησης απασχόλησης από το 2000 ως το 2010, για την «Ευρώπη των 27», την Ιαπωνία και τις ΗΠΑ.	4
Διάγραμμα 2 - Ετήσια συνολικά ποσοστά απασχόλησης Ευρωπαϊκών χωρών έτους 2010.....	10
Διάγραμμα 3 - Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών ανά φύλο, έτους 2010	12
Διάγραμμα 4 - Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών ανά ηλικιακή ομάδα, έτους 2010	14
Διάγραμμα 5 - Ετήσια ποσοστά μερικής απασχόλησης Ευρωπαϊκών χωρών ανά φύλο, έτους 2010	17
Διάγραμμα 6 - Ετήσια ποσοστά απασχόλησης με συμβόλαιο περιορισμένης διάρκειας στις χώρες της Ευρώπης, έτους 2010.....	19
Διάγραμμα 7 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Male”	22
Διάγραμμα 8 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Female”	24
Διάγραμμα 9 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Low_Educated”	26
Διάγραμμα 10 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Mid_Educated”	28
Διάγραμμα 11 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “High_Educated”	30
Διάγραμμα 12 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Young”.....	32
Διάγραμμα 13 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Mid_Age”.....	34
Διάγραμμα 14 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Older”	36
Διάγραμμα 15 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “With_Second_Job”	38
Διάγραμμα 16 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Part_Time_Male”	40
Διάγραμμα 17 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Part_Time_Female”	42
Διάγραμμα 18 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Limited_Duration_Contract”	44
Διάγραμμα 19 - Οι ιδιοτιμές ταξινομημένες με τη σειρά μεγέθους τους	71
Διάγραμμα 20 - Κόστος χρήσης κύριων συνιστωσών	72
Διάγραμμα 21 - Συντεταγμένες των χαρακτηριστικών απασχόλησης.....	74
Διάγραμμα 22 - Συντεταγμένες των χαρακτηριστικών απασχόλησης μετά την περιστροφή των κύριων αξόνων	76
Διάγραμμα 23 - Γραφική αναπαράσταση των Ευρωπαϊκών χωρών βάσει των δύο «δεικτών» απασχόλησης	79
Διάγραμμα 24 - Γραφική αναπαράσταση των 3 ομάδων Ευρωπαϊκών χωρών βάσει των 2 «δεικτών» απασχόλησης.....	91

ΚΑΤΑΣΤΑΣΗ ΔΕΝΔΡΟΓΡΑΜΜΑΤΩΝ

Δενδρόγραμμα 1 - Απεικόνιση της συσσωρευτικής μεθόδου ομαδοποίησης των Ευρωπαϊκών χωρών, βάσει των 11 αρχικών χαρακτηριστικών απασχόλησης	86
Δενδρόγραμμα 2 - Απεικόνιση της συσσωρευτικής μεθόδου ομαδοποίησης των Ευρωπαϊκών χωρών, βάσει των 2 «δεικτών» απασχόλησης.....	89
Δενδρόγραμμα 3 - Απεικόνιση της συσσωρευτικής μεθόδου ομαδοποίησης των Ευρωπαϊκών χωρών, βάσει των 3 πρώτων κυρίων συνιστωσών	99

ΚΑΤΑΣΤΑΣΗ ΘΗΚΟΓΡΑΜΜΑΤΩΝ

Θηκόγραμμα 1- Μεταβλητή “Male”	21
Θηκόγραμμα 2 - Μεταβλητή “Female”	23
Θηκόγραμμα 3 - Μεταβλητή “Low_Educated”	25
Θηκόγραμμα 4 - Μεταβλητή “Mid_Educated”	27
Θηκόγραμμα 5 - Μεταβλητή “High_Educated”	29
Θηκόγραμμα 6 - Μεταβλητή “Young”	31
Θηκόγραμμα 7 - Μεταβλητή “Mid_Age”	33
Θηκόγραμμα 8 - Μεταβλητή “Older”	35
Θηκόγραμμα 9 - Μεταβλητή “With_Second_Job”	37
Θηκόγραμμα 10 - Μεταβλητή “Part_Time_Male”	39
Θηκόγραμμα 11 - Μεταβλητή “Part_Time_Female”	41
Θηκόγραμμα 12 - Μεταβλητή “Limited_Duration_Contract”	43

ΚΑΤΑΣΤΑΣΗ ΧΑΡΤΩΝ

Χάρτης 1 – Χάρτης των 3 ομάδων Ευρωπαϊκών χωρών βάσει των 2 «δεικτών» απασχόλησης	91
Χάρτης 2 - Μετακίνηση χωρών σε ομάδες ανάμεσα στην ομαδοποίηση με βάσει τα αρχικά χαρακτηριστικά απασχόλησης (αριστερά) και τους PCA δείκτες απασχόλησης (δεξιά).....	95
Χάρτης 3 - Χάρτης των 4 ομάδων Ευρωπαϊκών χωρών βάσει των 3 «δεικτών» απασχόλησης.....	101

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1 - ΔΕΔΟΜΕΝΑ ΑΠΑΣΧΟΛΗΣΗΣ ΕΥΡΩΠΑΪΚΩΝ ΧΩΡΩΝ	1
1.1 Σχετικά με τα δεδομένα απασχόλησης.....	1
1.1.1 Πηγές.....	1
1.1.2 Ορισμός και βασικά χαρακτηριστικά της απασχόλησης	1
1.2 Τα δεδομένα που χρησιμοποιήθηκαν στα πλαίσια της εργασίας	2
1.2.1 Ετήσια μεταβολή απασχόλησης «Ευρώπης των 27», Ιαπωνίας και ΗΠΑ3	
1.2.2 Ετήσια μεταβολή απασχόλησης ανά φύλο στην Ευρώπη	4
1.2.3 Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών, Ιαπωνίας και ΗΠΑ, 2000-2010	5
1.2.4 Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών, Ιαπωνίας και ΗΠΑ, για επιλεγμένες κατηγορίες εργαζόμενων, για τα έτη 2000, 2005 και 2010	6
1.2.5 Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών ανά κατηγορία μορφωτικού επιπέδου, για την ηλικιακή κατηγορία 25-64	7
1.2.6 Ετήσια συνολικά ποσοστά απασχόλησης Ευρωπαϊκών χωρών	8
1.2.7 Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών ανά φύλο.....	11
1.2.8 Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών ανά ηλικιακή ομάδα	13
1.2.9 Ετήσια ποσοστά μερικής απασχόλησης και ατόμων με 2 ^η δουλειά στην Ευρώπη, 2000-2010	15
1.2.10 Ετήσια ποσοστά μερικής απασχόλησης Ευρωπαϊκών χωρών ανά φύλο	16
1.2.11 Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών με συμβόλαιο περιορισμένης διάρκειας.....	18
ΚΕΦΑΛΑΙΟ 2 - ΕΠΙΛΟΓΗ ΜΕΤΑΒΛΗΤΩΝ ΑΠΑΣΧΟΛΗΣΗΣ ΓΙΑ ΠΕΡΑΙΤΕΡΩ ΑΝΑΛΥΣΗ	20
2.1 Μεταβλητή “Male”	20
2.2 Μεταβλητή “Female”	22
2.3 Μεταβλητή “Low_Educated”	24
2.4 Μεταβλητή “Mid_Educated”	26
2.5 Μεταβλητή “High_Educated”	28
2.6 Μεταβλητή “Young”	30
2.7 Μεταβλητή “Mid_Age”.....	32
2.8 Μεταβλητή “Older”	34
2.9 Μεταβλητή “With_Second_Job”	36
2.10 Μεταβλητή “Part_Time_Male”	38
2.11 Μεταβλητή “Part_Time_Female”.....	40
2.12 Μεταβλητή “Limited_Duration_Contract”	42
2.13 ΒΙΒΛΙΟΓΡΑΦΙΑ 2 ^{ου} ΚΕΦΑΛΑΙΟΥ.....	45
ΚΕΦΑΛΑΙΟ 3 – ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ.....	46
3.1 Πολυμεταβλητή Ανάλυση	46

3.1.1	Αλγεβρική παρουσίαση του προβλήματος.....	47
3.1.2	Συνοπτικοί δείκτες θέσης και μεταβλητότητας	47
3.1.3	Γραφική παρουσίαση.....	49
3.1.4	Γεωμετρική ερμηνεία της γενικευμένης δειγματικής διασποράς (GSV) ..	49
3.2	Ανάλυση σε κύριες συνιστώσες.....	50
3.2.1	Εισαγωγή.....	50
3.2.2	Κύριες συνιστώσες από τον πίνακα συνδιασποράς Σ	51
3.2.3	Γεωμετρική ερμηνεία κυρίων συνιστωσών	56
3.2.4	Κύριες συνιστώσες από τον πίνακα συσχέτισης.....	57
3.2.5	Κύριες συνιστώσες για πίνακες συνδιασποράς με ειδική δομή.....	59
3.2.6	Κύριες συνιστώσες από τον πίνακα διεγματικής διασποράς και τον πίνακα δειγματικής συσχέτισης.....	61
3.3	ΒΙΒΛΙΟΓΡΑΦΙΑ 3^{ου} ΚΕΦΑΛΑΙΟΥ.....	64
ΚΕΦΑΛΑΙΟ 4 - ΑΝΑΛΥΣΗ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ (PCA)		65
4.1	Έλεγχος καταλληλότητας δεδομένων για εφαρμογή της PCA.....	66
4.2	Επιλογή δειγματικού πίνακα	68
4.3	Επιλογή βέλτιστου αριθμού κύριων συνιστωσών.	68
4.3.1	Κριτήριο 1ο:	68
4.3.2	Κριτήριο 2 ^ο :.....	69
4.3.3	Κριτήριο 3 ^ο :.....	70
4.3.4	Κριτήριο 4 ^ο :.....	70
4.4	Ιδιοδιανύσματα των δύο πρώτων κύριων συνιστωσών	73
4.5	Περιστροφή των πρώτων κύριων αξόνων.....	75
4.6	Ερμηνεία των πρώτων κύριων συνιστωσών και δημιουργία «δεικτών» απασχόλησης	77
4.6.1	Ερμηνεία 1 ^{ης} κύριας συνιστώσας:.....	77
4.6.2	Ερμηνεία 2 ^{ης} κύριας συνιστώσας:.....	78
4.7	Απεικόνιση των Ευρωπαϊκών χωρών σε σύστημα δύο αξόνων.....	78
4.8	ΒΙΒΛΙΟΓΡΑΦΙΑ 4 ^{ου} ΚΕΦΑΛΑΙΟΥ.....	80
ΚΕΦΑΛΑΙΟ 5 - ΑΝΑΛΥΣΗ ΚΑΤΑ ΣΥΣΤΑΔΕΣ ΜΕ ΤΗ ΣΥΣΣΩΡΕΥΤΙΚΗ ΜΕΘΟΔΟ		81
5.1	Μέτρα απόστασης	81
5.2	Μέτρα ομοιότητας.....	82
5.3	Μέθοδοι ομαδοποίησης	83
5.3.1	Επιλογή του πλήθους των ομάδων	84
5.4	Ομαδοποίηση βάσει των 11 αρχικών χαρακτηριστικών απασχόλησης.84	
5.5	Ομαδοποίηση βάσει των 2 «δεικτών» απασχόλησης	87
5.6	ΒΙΒΛΙΟΓΡΑΦΙΑ 5 ^{ου} ΚΕΦΑΛΑΙΟΥ.....	92
ΚΕΦΑΛΑΙΟ 6 - ΣΥΜΠΕΡΑΣΜΑΤΑ.....		93
6.1	Σύγκριση ομάδων που προέκυψαν από τις δύο ομαδοποιήσεις	93
6.1.1	Ομαδοποίηση βάσει των 11 αρχικών χαρακτηριστικών απασχόλησης .93	
6.1.2	Ομαδοποίηση βάσει των 2 δεικτών απασχόλησης.....	94

6.2	Επανάληψη ομαδοποίησης με διατήρηση τριών PCA κυρίων συνιστωσών	95
6.3	Ερμηνεία των πρώτων κύριων συνιστωσών και δημιουργία «δεικτών» απασχόλησης	96
6.3.1	Ερμηνεία 1 ^{ης} κύριας συνιστώσας:	97
6.3.2	Ερμηνεία 2 ^{ης} κύριας συνιστώσας:	97
6.3.3	Ερμηνεία 3 ^{ης} κύριας συνιστώσας:	97
6.4	Ομαδοποίηση βάσει των 3 «δεικτών» απασχόλησης	97
6.5	Αξία της ομαδοποίησης για ένα χρηματοπιστωτικό οργανισμό	101
ΣΥΝΟΨΗ		103
ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΠΕΡΑΙΤΕΡΩ ΜΕΛΕΤΗ		104
ΒΙΒΛΙΟΓΡΑΦΙΑ		105

ΚΕΦΑΛΑΙΟ 1 - ΔΕΔΟΜΕΝΑ ΑΠΑΣΧΟΛΗΣΗΣ ΕΥΡΩΠΑΪΚΩΝ ΧΩΡΩΝ

1.1 Σχετικά με τα δεδομένα απασχόλησης

1.1.1 Πηγές

Η κύρια πηγή δεδομένων για τις στατιστικές της αγοράς εργασίας είναι η έρευνα εργατικού δυναμικού (Labour Force Survey – LFS) της Ευρωπαϊκής Ένωσης (EU LFS), ενώ μια άλλη συχνά χρησιμοποιούμενη πηγή για τις στατιστικές της απασχόλησης είναι οι εθνικοί λογαριασμοί. Και οι δύο αυτές πηγές χρησιμοποιούν παρόμοιους ορισμούς για την απασχόληση, που βασίζονται σε διεθνή πρότυπα από τη Διεθνή Οργάνωση Εργασίας (International Labor Organization - ILO) και το σύστημα των εθνικών λογαριασμών, αντίστοιχα. Μια τρίτη πιθανή πηγή για πληροφορίες σχετικά με τις στατιστικές για την απασχόληση είναι οι στατιστικές των επιχειρήσεων. Η πηγή δεδομένων για το σύνολο των πληροφοριών που παρουσιάζονται σε αυτή την εργασία είναι η EU LFS.

Η EU LFS είναι μια τριμηνιαία δειγματοληπτική έρευνα η οποία καλύπτει τον πληθυσμό των ιδιωτικών νοικοκυριών στην ΕΕ εκτός από το Λιχτενστάιν και τις υποψήφιες χώρες. Παρέχει ετήσια και τριμηνιαία αποτελέσματα σε σχέση με τη συμμετοχή στην αγορά εργασίας των ατόμων ηλικίας 15 ετών και άνω. Η EU LFS συλλέγει πληροφορίες σχετικά με την κατάσταση του εργατικού δυναμικού (όλα τα πρόσωπα που είτε απασχολούνται, είτε είναι άνεργα ή οικονομικά ανενεργά), τα χαρακτηριστικά της απασχόλησης, το ωράριο εργασίας, την αναζήτηση εργασίας από ανέργους, τα επίπεδα εκπαίδευσης, τις πρόσφατες εκπαιδεύσεις και καταρτίσεις, καθώς και το δημογραφικό υπόβαθρο και την οικογενειακή σύνθεση των ατόμων.

Το μέγεθος του δείγματος της EU LFS ανέρχεται σε περίπου 1,5 εκατομμύρια άτομα κάθε τρίμηνο. Οι τριμηνιαίοι ρυθμοί δειγματοληψίας ποικίλλουν μεταξύ 0,2% και 3,3% σε κάθε χώρα. Η Eurostat άρχισε τη συλλογή μικροδεδομένων της LFS το 1983 με ένα τρίμηνο αναφοράς ανά έτος (συνήθως την άνοιξη). Κατά την περίοδο 1998-2005 η έρευνα μετατράπηκε σε συνεχή τριμηνιαία έρευνα και τα 27 κράτη μέλη παρέχουν πλέον τριμηνιαία αποτελέσματα.

1.1.2 Ορισμός και βασικά χαρακτηριστικά της απασχόλησης

Ο οικονομικά ενεργός πληθυσμός (εργατικό δυναμικό) αποτελείται από εργαζόμενα και μη άτομα. Η EU LFS ορίζει ως απασχολούμενα άτομα αυτά ηλικίας 15 και άνω που κατά την εβδομάδα αναφοράς ασκούσαν κάποια εργασία, έστω και για μία ώρα ανά εβδομάδα με αντίτιμο μισθό, κέρδος ή οικογενειακό όφελος. Το εργατικό δυναμικό επίσης περιλαμβάνει άτομα που δεν εργάζονταν αλλά είχαν μια δουλειά ή επιχείρηση από την οποία απουσίαζαν προσωρινά για λόγους υγείας, διακοπών, εργασιακών διαφορών, εκπαίδευσης ή κατάρτισης.

Η απασχόληση μπορεί να μετρηθεί ως αριθμός ατόμων ή δουλειών, σε ισοδύναμα πλήρους ωραρίου ή σε ώρες εργασίας. Όλες οι εκτιμήσεις που παρουσιάζονται χρησιμοποιούν τον αριθμό των ατόμων, ως επίσης και η πληροφορία που παρουσιάζεται για τα ποσοστά απασχόλησης. Οι στατιστικές απασχολήσεις συχνά εκφράζονται ως ποσοστά απασχόλησης προκειμένου να αντισταθμιστεί το μεταβαλλόμενο μέγεθος του πληθυσμού των χωρών και να καταστούν εφικτές συγκρίσεις μεταξύ χωρών διαφορετικών μεγεθών. Αυτά τα ποσοστά απασχόλησης τυπικά δημοσιεύονται για τον πληθυσμό που είναι σε ηλικία

κατάλληλη για να εργαστεί, η οποία συνήθως θεωρείται αυτή ανάμεσα στα 15 και τα 64 έτη. Αυτή η ηλικιακή ομάδα είναι το standard που χρησιμοποιείται και από άλλους διεθνείς στατιστικούς οργανισμούς.

Ορισμένα κύρια χαρακτηριστικά απασχόλησης όπως ορίζονται από την EU LFS, περιλαμβάνουν:

- Εργαζόμενοι
 - Τα άτομα που δουλεύουν για κάποιο δημόσιο ή ιδιώτη εργοδότη και που λαμβάνουν αποζημίωση σε μορφή μισθών, αμοιβών, αποτελεσμάτων ή σε είδος.
- Αυτοαπασχολούμενοι
 - Άτομα που εργάζονται στη δική τους επιχείρηση, φάρμα ή επάγγελμα.
 - Θεωρείται ότι εργάζονται αν κατά την εβδομάδα αναφοράς, πληρούν ένα από τα παρακάτω κριτήρια:
 - Εργάζονται με σκοπό το κέρδος.
 - Ξοδεύουν χρόνο για τη λειτουργία μιας επιχείρησης.
 - Επί του παρόντος δημιουργούν μια επιχείρηση.
- Η διάκριση πλήρες/μερικό ωράριο στην κύρια εργασία δηλώνεται από τον ερωτώμενο.
- Η ένδειξη «εργαζόμενος με δεύτερη δουλειά» αναφέρεται μόνο σε άτομα με περισσότερες της μία δουλειές ταυτόχρονα και όχι σε άτομα που έχουν αλλάξει δουλειά κατά την εβδομάδα αναφοράς.
- Ένας εργαζόμενος θεωρείται ότι εργάζεται προσωρινά αν ο ίδιος και ο εργοδότης θεωρούν ότι το πέρας της εργασίας ορίζεται από αντικειμενικές συνθήκες, όπως:
 - Συγκεκριμένη ημερομηνία,
 - Ολοκλήρωση ανάθεσης,
 - Η επιστροφή εργαζομένου ο οποίος προσωρινά αναπληρώνεται.
- Τυπικές περιπτώσεις:
 - Άτομα σε εποχιακή απασχόληση,
 - Άτομα που συνδέονται με μια υπηρεσία και προσλαμβάνονται από ένα τρίτο για να φέρουν εις πέρας μια συγκεκριμένη εργασία (εκτός αν υπάρχει γραπτό συμβόλαιο απεριόριστης διάρκειας),
 - Άτομα με συγκεκριμένα εκπαιδευτικά συμβόλαια.

1.2 Τα δεδομένα που χρησιμοποιήθηκαν στα πλαίσια της εργασίας

Τα δεδομένα προέρχονται από το Eurostat. Το Eurostat είναι το γραφείο στατιστικής της Ευρωπαϊκής Ένωσης. Αποστολή του είναι να αποτελεί τον εξέχοντα πάροχο υψηλής ποιότητας στατιστικών στοιχείων στην Ευρώπη. Εργαζόμενο προς αυτή την κατεύθυνση, συλλέγει και αναλύει δεδομένα από τα Εθνικά Ινστιτούτα Στατιστικής των χωρών και παρέχει συγκρίσιμα και εναρμονισμένα δεδομένα που μπορούν να χρησιμοποιηθούν για τη διαμόρφωση, τον ορισμό, την υλοποίηση και την ανάλυση ευρωπαϊκών πολιτικών. Τα προϊόντα και οι υπηρεσίες του Eurostat είναι επίσης υψηλής αξίας για την επιχειρηματική κοινότητα, επαγγελματικούς οργανισμούς, ακαδημαϊκούς, βιβλιοθηκάρχους, μέσα ενημέρωσης και πολίτες.

Στόχος του 1^{ου} κεφαλαίου, πέρα από την παρουσίαση των δεδομένων απασχόλησης που βρίσκονται διαθέσιμα στη Eurostat, είναι αφενός να κάνει αντιληπτό το μέγεθος του όγκου των διαθέσιμων δεδομένων με το οποίο μπορεί να βρεθεί αντιμέτωπος ένας αναλυτής, αφετέρου να καταδείξει την ανάγκη χρήσης κατάλληλων στατιστικών τεχνικών, ούτως ώστε να περιοριστεί ο όγκος της πληροφορίας προκειμένου να είναι εύκολο να προκύψουν κάποια χρήσιμα συμπεράσματα.

Τα δεδομένα που παρουσιάζονται στις επόμενες σελίδες αφορούν στα χαρακτηριστικά απασχόλησης των 33 χωρών της Ευρώπης, όπως ακριβώς ελήφθησαν από τη βάση δεδομένων της Eurostat.

1.2.1 Ετήσια μεταβολή απασχόλησης «Ευρώπης των 27», Ιαπωνίας και ΗΠΑ

Στον πίνακα που ακολουθεί γίνεται μία γενική παρουσίαση του ρυθμού αύξησης απασχόλησης στην «Ευρώπη των 27», στην Ιαπωνία και στην ΗΠΑ από το 2000 ως και το 2010.

Annual employment growth, 2000-2010 (% change in the number of employed persons)

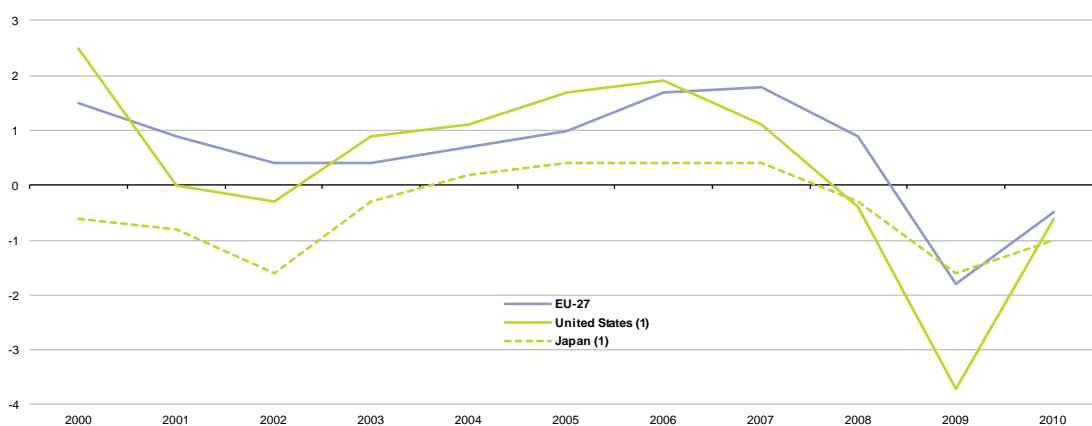
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
EU-27	1,5	0,9	0,4	0,4	0,7	1,0	1,7	1,8	0,9	-1,8	-0,5
Japan (1)	-0,6	-0,8	-1,6	-0,3	0,2	0,4	0,4	0,4	-0,3	-1,6	-1,0
United States (2,5	0,0	-0,3	0,9	1,1	1,7	1,9	1,1	-0,4	-3,7	-0,6

(1) 2009 and 2010, forecasts.

Source: Eurostat (online data code: lfsi_grt_a)

Πίνακας 1- Ετήσιος ρυθμός αύξησης απασχόλησης από το 2000 ως το 2010, για την «Ευρώπη των 27», την Ιαπωνία και τις ΗΠΑ.

Τα παραπάνω στοιχεία απεικονίζονται στο παρακάτω διάγραμμα. Η εικόνα του διαγράμματος επιβεβαιώνει το γεγονός ότι βρισκόμαστε σε μία παγκοσμιοποιημένη οικονομία, καθώς τα μεγέθη της απασχόλησης μεταβάλλονται κατά τον ίδιο τρόπο διαχρονικά, στις σημαντικότερες οικονομίες του πλανήτη. Είναι χαρακτηριστική η ταυτόχρονη μείωση του ρυθμού απασχόλησης στα χαμηλότερα επίπεδα κατά το 2009, όταν η οικονομική κρίση βρισκόταν σε έξαρση.



Διάγραμμα 1 - Ετήσιος ρυθμός αύξησης απασχόλησης από το 2000 ως το 2010, για την «Ευρώπη των 27», την Ιαπωνία και τις ΗΠΑ.

1.2.2 Ετήσια μεταβολή απασχόλησης ανά φύλο στην Ευρώπη

Στον πίνακα που ακολουθεί παρουσιάζεται η ετήσια μεταβολή απασχόλησης ανά φύλο, για τις χώρες της Ευρώπης στα έτη 2000, 2005 και 2010. Εντυπωσιακή είναι η αύξηση του ποσοστού απασχόλησης τόσο των αντρών όσο κυρίως των γυναικών στην Τουρκία, κατά το 2010. Επίσης, αρκετές χώρες του πρώην ανατολικού μπλοκ αντιμετωπίζουν σοβαρή μείωση στο ποσοστό απασχόλησης. Τα εντεινόμενα προβλήματα απασχόλησης που αντιμετωπίζει σήμερα η χώρα μας καθώς και άλλες χώρες του Ευρωπαϊκού νότου, δεν είχαν προλάβει να αποτυπωθούν στα στοιχεία απασχόλησης της Eurostat κατά το 2010.

Annual employment growth by gender, 2000-2010
(% change in the number of employed persons)

	Total			Male			Female		
	2000	2005 (1)	2010 (2)	2000	2005 (3)	2010 (2)	2000	2005 (3)	2010 (2)
EU-27	1,5	1,0	-0,5	1,0	0,7	-0,6	2,2	1,3	-0,3
Euro area	2,3	1,0	-0,5	1,9	0,4	-0,7	3,0	1,8	-0,2
Belgium	2,0	1,4	0,7	1,9	0,5	0,3	2,2	2,6	1,1
Bulgaria	-2,4	2,7	-5,9	·	3,4	-6,9	·	1,9	-4,8
Czech Republic	-0,2	1,0	-0,8	-0,2	1,7	-0,7	-0,1	0,2	-0,9
Denmark	0,5	1,0	-2,1	0,0	0,9	-2,3	-1,1	1,2	-1,8
Germany	1,9	-0,1	0,5	1,4	-1,1	0,9	2,6	1,1	0,2
Estonia	-1,5	2,0	-4,8	-1,1	0,5	-5,1	-1,9	3,6	-4,5
Ireland	4,5	4,9	-4,1	3,9	4,3	-5,2	5,3	5,8	-2,9
Greece	0,5	0,8	-2,1	0,1	0,5	-2,8	1,2	1,3	-1,1
Spain	5,1	4,1	-2,3	3,6	2,7	-3,3	7,7	6,3	-0,9
France	2,6	0,7	0,2	2,5	0,2	0,2	2,7	1,2	0,1
Italy	1,9	0,6	-0,7	1,3	0,7	-1,2	3,1	0,3	0,0
Cyprus	1,7	3,6	0,1	·	3,8	-0,7	·	3,3	1,0
Latvia	-3,2	1,6	-4,8	-4,8	2,3	-5,3	-1,5	0,8	-4,3
Lithuania	-4,0	2,5	-5,1	·	2,2	-5,9	·	2,7	-4,4
Luxembourg	5,6	2,9	1,5	5,3	1,5	0,6	6,1	4,8	2,8
Hungary	1,0	-0,2	0,2	0,9	-0,3	-0,9	-1,1	-0,1	1,5
Malta	·	1,5	2,0	·	0,3	0,9	·	4,2	4,2
Netherlands	2,2	0,5	-0,5	2,0	-0,3	-0,5	2,5	1,4	-0,5
Austria	0,9	1,5	1,0	0,9	1,0	1,1	1,0	2,1	0,9
Poland	-1,6	2,2	0,4	-1,5	3,1	0,1	-1,7	1,1	0,8
Portugal	2,1	-0,3	-1,5	1,9	-1,0	-1,6	2,3	0,5	-1,4
Romania	-0,8	-1,5	-1,8	-1,0	-0,6	-1,5	-0,5	-2,6	-2,1
Slovenia	1,5	-0,5	-2,0	1,4	-0,6	-1,9	1,7	-0,3	-2,2
Slovakia	-2,0	1,6	-1,4	-2,8	2,9	-2,5	-0,9	0,1	0,0
Finland	2,1	1,4	-1,4	2,2	1,0	-0,7	1,9	1,8	-2,2
Sweden	2,5	0,3	1,1	2,3	1,0	1,9	2,6	-0,5	0,3
United Kingdom (4)	1,3	1,3	-0,7	-0,6	1,1	-0,5	3,6	1,6	-0,9
Iceland	2,0	3,3	-0,3	·	3,7	-1,2	·	2,8	0,7
Norway	0,6	1,2	-0,2	·	1,5	0,1	·	0,9	-0,6
Croatia	4,0	0,8	-4,0	·	0,2	-4,5	·	1,5	-3,4
Turkey	-0,4	1,4	6,2	·	·	5,0	·	·	9,5
Japan	-0,6	0,4	-1,0	·	·	·	·	·	·
United States	2,5	1,7	-0,6	·	·	·	·	·	·

(1) Austria, break in series.

(2) The Netherlands, break in series.

(3) Germany, Spain, Italy, Austria, Romania and Sweden, break in series.

(4) Eurostat estimates of persons employed are based on the estimates of jobs transmitted by the United Kingdom.

Source: Eurostat (online data code: lfsi_grt_a)

Πίνακας 2 - Ετήσιος ρυθμός αύξησης απασχόλησης για τις χώρες της Ευρώπης ανά φύλο, για τα έτη 2000, 2005 και 2010.

1.2.3 Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών, Ιαπωνίας και ΗΠΑ, 2000-2010

Στον πίνακα που ακολουθεί παρατίθενται τα ετήσια ποσοστά απασχόλησης των Ευρωπαϊκών χωρών, Ιαπωνίας και ΗΠΑ, για τα έτη 2000-2010. Ο μέσος όρος για όλη την Ευρώπη φαίνεται να παραμένει διαχρονικά σταθερός.

Employment rate, age group 15-64, 2000-2010
(%)

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
EU-27	62,2	62,6	62,4	62,6	63,0	63,5	64,5	65,4	65,9	64,6	64,2
Euro area	61,4	62,1	62,3	62,6	63,1	63,7	64,7	65,6	66,0	64,7	64,2
Belgium	60,5	59,9	59,9	59,6	60,3	61,1	61,0	62,0	62,4	61,6	62,0
Bulgaria	50,4	49,7	50,6	52,5	54,2	55,8	58,6	61,7	64,0	62,6	59,7
Czech Republic	65,0	65,0	65,4	64,7	64,2	64,8	65,3	66,1	66,6	65,4	65,0
Denmark	76,3	76,2	75,9	75,1	75,7	75,9	77,4	77,1	77,9	75,7	73,4
Germany (1)	65,6	65,8	65,4	65,0	65,0	66,0	67,5	69,4	70,7	70,9	71,1
Estonia	60,4	61,0	62,0	62,9	63,0	64,4	68,1	69,4	69,8	63,5	61,0
Ireland	65,2	65,8	65,5	65,5	66,3	67,6	68,7	69,2	67,6	61,8	60,0
Greece	56,5	56,3	57,5	58,7	59,4	60,1	61,0	61,4	61,9	61,2	59,6
Spain (1)	56,3	57,8	58,5	59,8	61,1	63,3	64,8	65,6	64,3	59,8	58,6
France	62,1	62,8	63,0	64,0	63,8	63,7	63,7	64,3	64,9	64,1	64,0
Italy (2)	53,7	54,8	55,5	56,1	57,6	57,6	58,4	58,7	58,7	57,5	56,9
Cyprus	65,7	67,8	68,6	69,2	68,9	68,5	69,6	71,0	70,9	69,9	69,7
Latvia	57,5	58,6	60,4	61,8	62,3	63,3	66,3	68,3	68,6	60,9	59,3
Lithuania	59,1	57,5	59,9	61,1	61,2	62,6	63,6	64,9	64,3	60,1	57,8
Luxembourg	62,7	63,1	63,4	62,2	62,5	63,6	63,6	64,2	63,4	65,2	65,2
Hungary	56,3	56,2	56,2	57,0	56,8	56,9	57,3	57,3	56,7	55,4	55,4
Malta	54,2	54,3	54,4	54,2	54,0	53,9	53,6	54,6	55,3	54,9	56,0
Netherlands (3)	72,9	74,1	74,4	73,6	73,1	73,2	74,3	76,0	77,2	77,0	74,7
Austria (2)	68,5	68,5	68,7	68,9	67,8	68,6	70,2	71,4	72,1	71,6	71,7
Poland	55,0	53,4	51,5	51,2	51,7	52,8	54,5	57,0	59,2	59,3	59,3
Portugal	68,4	69,0	68,8	68,1	67,8	67,5	67,9	67,8	68,2	66,3	65,6
Romania (4)	63,0	62,4	57,6	57,6	57,7	57,6	58,8	58,8	59,0	58,6	58,8
Slovenia	62,8	63,8	63,4	62,6	65,3	66,0	66,6	67,8	68,6	67,5	66,2
Slovakia	56,8	56,8	56,8	57,7	57,0	57,7	59,4	60,7	62,3	60,2	58,8
Finland	67,2	68,1	68,1	67,7	67,6	68,4	69,3	70,3	71,1	68,7	68,1
Sweden (1)	73,0	74,0	73,6	72,9	72,1	72,5	73,1	74,2	74,3	72,2	72,7
United Kingdom	71,2	71,4	71,4	71,5	71,7	71,7	71,6	71,5	71,5	69,9	69,5
Iceland	.	.	.	83,3	82,3	83,8	84,6	85,1	83,6	78,3	78,2
Norway	77,5	77,2	76,8	75,5	75,1	74,8	75,4	76,8	78,0	76,4	75,3
Switzerland	78,3	79,1	78,9	77,9	77,4	77,2	77,9	78,6	79,5	79,0	78,6
Croatia	.	.	53,4	53,4	54,7	55,0	55,6	57,1	57,8	56,6	54,0
FYR of Macedonia	39,6	40,7	41,9	43,3	43,5
Turkey	44,6	44,6	44,9	44,3	46,3
Japan	68,9	68,8	68,2	68,4	68,7	69,3	70,0	70,7	70,7	70,0	70,1
United States	74,1	73,1	71,9	71,2	71,2	71,5	72,0	71,8	70,9	67,6	66,7

(1) Break in series in, 2005.

(2) Break in series in, 2004.

(3) Break in series in, 2010.

(4) Break in series in, 2002.

Source: Eurostat (online data code: lfsi_emp_a)

Πίνακας 3 - Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών, Ιαπωνίας και ΗΠΑ, 2000-2010.

1.2.4 Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών, Ιαπωνίας και ΗΠΑ, για επιλεγμένες κατηγορίες εργαζόμενων, για τα έτη 2000, 2005 και 2010

Στον πίνακα που ακολουθεί παρατίθενται τα ετήσια ποσοστά απασχόλησης των Ευρωπαϊκών χωρών, Ιαπωνίας και ΗΠΑ, για κάθε φύλο και για την ηλικιακή ομάδα 55-64, για τα έτη 2000, 2005 και 2010. Τα ποσοστά απασχόλησης των αντρών είναι μεγαλύτερα από τα αντίστοιχα ποσοστά των γυναικών.

Employment rates for selected population groups, 2000-2010
(%)

	Male			Female			Older workers (55-64)		
	2000	2005	2010	2000	2005	2010	2000	2005	2010
EU-27	70,8	70,8	70,1	53,7	56,3	58,2	36,9	42,3	46,3
Euro area	71,4	71,8	70,4	51,4	55,6	57,9	34,2	40,4	45,8
Belgium	69,5	68,3	67,4	51,5	53,8	56,5	26,3	31,8	37,3
Bulgaria	54,7	60,0	63,0	46,3	51,7	56,4	20,8	34,7	43,5
Czech Republic	73,2	73,3	73,5	56,9	56,3	56,3	36,3	44,5	46,5
Denmark	80,8	79,8	75,8	71,6	71,9	71,1	55,7	59,5	57,6
Germany (1)	72,9	71,3	76,0	58,1	60,6	66,1	37,6	45,4	57,7
Estonia	64,3	67,0	61,5	56,9	62,1	60,6	46,3	56,1	53,8
Ireland	76,3	76,9	63,9	53,9	58,3	56,0	45,3	51,6	50,0
Greece	71,5	74,2	70,9	41,7	46,1	48,1	39,0	41,6	42,3
Spain (1)	71,2	75,2	64,7	41,3	51,2	52,3	37,0	43,1	43,6
France	69,2	69,2	68,3	55,2	58,4	59,9	29,9	38,5	39,7
Italy (1)	68,0	69,9	67,7	39,6	45,3	46,1	27,7	31,4	36,6
Cyprus	78,7	79,2	76,6	53,5	58,4	63,0	49,4	50,6	56,8
Latvia	61,5	67,6	59,2	53,8	59,3	59,4	36,0	49,5	48,2
Lithuania	60,5	66,1	56,8	57,7	59,4	58,7	40,4	49,2	48,6
Luxembourg	75,0	73,3	73,1	50,1	53,7	57,2	26,7	31,7	39,6
Hungary	63,1	63,1	60,4	49,7	51,0	50,6	22,2	33,0	34,4
Malta	75,0	73,8	72,3	33,1	33,7	39,2	28,5	30,8	30,2
Netherlands (2)	82,1	79,9	80,0	63,5	66,4	69,3	38,2	46,1	53,7
Austria (1)	77,3	75,4	77,1	59,6	62,0	66,4	28,8	31,8	42,4
Poland	61,2	58,9	65,6	48,9	46,8	53,0	28,4	27,2	34,0
Portugal	76,5	73,4	70,1	60,5	61,7	61,1	50,7	50,5	49,2
Romania (1)	68,6	63,7	65,7	57,5	51,5	52,0	49,5	39,4	41,1
Slovenia	67,2	70,4	69,6	58,4	61,3	62,6	22,7	30,7	35,0
Slovakia	62,2	64,6	65,2	51,5	50,9	52,3	21,3	30,3	40,5
Finland	70,1	70,3	69,4	64,2	66,5	66,9	41,6	52,7	56,2
Sweden (1)	75,1	74,4	75,1	70,9	70,4	70,3	64,9	69,4	70,5
United Kingdom	77,8	77,7	74,5	64,7	65,8	64,6	50,7	56,8	57,1
Iceland	:	86,9	80,1	:	80,5	76,2	:	84,3	79,8
Norway	81,3	77,8	77,3	73,6	71,7	73,3	65,2	65,5	68,6
Switzerland	87,3	83,9	84,6	69,3	70,4	72,5	63,3	65,1	68,0
Croatia	:	61,7	59,4	:	48,6	48,8	:	32,6	37,6
FYR of Macedonia	:	:	52,8	:	:	34,0	:	:	34,2
Turkey	:	:	66,7	:	:	26,2	:	:	29,6
Japan	80,9	80,4	80,0	56,7	58,1	60,1	62,8	63,9	65,2
United States	80,6	77,6	71,1	67,8	65,6	62,4	57,8	60,8	60,3

(1) Break in series, 2000 to 2005.

(2) Break in series, 2005 to 2010.

Source: Eurostat (online data code: lfsi_emp_a)

Πίνακας 4 - Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών, Ιαπωνίας και ΗΠΑ, για επιλεγμένες κατηγορίες εργαζόμενων, έτη 2000, 2005 και 2010

1.2.5 Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών ανά κατηγορία μορφωτικού επιπέδου, για την ηλικιακή κατηγορία 25-64

Στον πίνακα που ακολουθεί παρατίθενται τα ετήσια ποσοστά απασχόλησης των Ευρωπαϊκών χωρών, ανά κατηγορία μορφωτικού επιπέδου (χαμηλό, ανώτερο, ανώτατο) για την ηλικιακή ομάδα 25-64, για το έτος 2010.

Employment rate by highest level of education, age group 25-64, 2010
(%)

	Pre-primary, primary & lower secondary - ISCED levels 0-2	Upper secondary & post-secondary non-tertiary - ISCED levels 3-4	Tertiary - ISCED levels 5-6
EU-27	53,8	73,1	83,9
Euro area	54,4	74,3	83,4
Belgium	48,9	74,5	84,0
Bulgaria	41,2	71,5	83,8
Czech Republic	43,2	74,5	83,3
Denmark	62,6	79,1	86,3
Germany	55,3	76,3	86,9
Estonia	45,2	68,7	79,9
Ireland	46,8	66,4	81,1
Greece	58,1	66,5	80,0
Spain	52,9	68,9	79,7
France	55,5	74,4	83,7
Italy	50,4	72,6	78,3
Cyprus	66,4	77,6	84,9
Latvia	48,5	66,1	81,1
Lithuania	32,4	63,8	86,9
Luxembourg	61,9	72,1	85,0
Hungary	37,6	66,2	78,6
Malta	49,1	81,0	85,7
Netherlands	61,4	80,3	87,2
Austria	56,1	77,9	85,7
Poland	39,9	65,6	84,8
Portugal	68,2	79,9	85,4
Romania	54,8	68,5	85,3
Slovenia	51,1	73,0	87,3
Slovakia	29,7	69,9	82,2
Finland	55,0	74,1	84,1
Sweden	65,1	83,3	88,1
United Kingdom	56,0	76,8	85,1
Iceland	76,5	82,0	89,1
Norway	63,8	81,4	90,2
Switzerland	69,3	81,2	88,4
Croatia	43,3	62,7	79,0
FYR of Macedonia	33,4	58,4	73,9
Turkey	45,6	60,0	75,7

Source: Eurostat (online data code: lfsa_ergaed)

Πίνακας 5 - Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών, ανά κατηγορία μορφωτικού επιπέδου για την ηλικιακή ομάδα 25-64, έτους 2010.

1.2.6 Ετήσια συνολικά ποσοστά απασχόλησης Ευρωπαϊκών χωρών

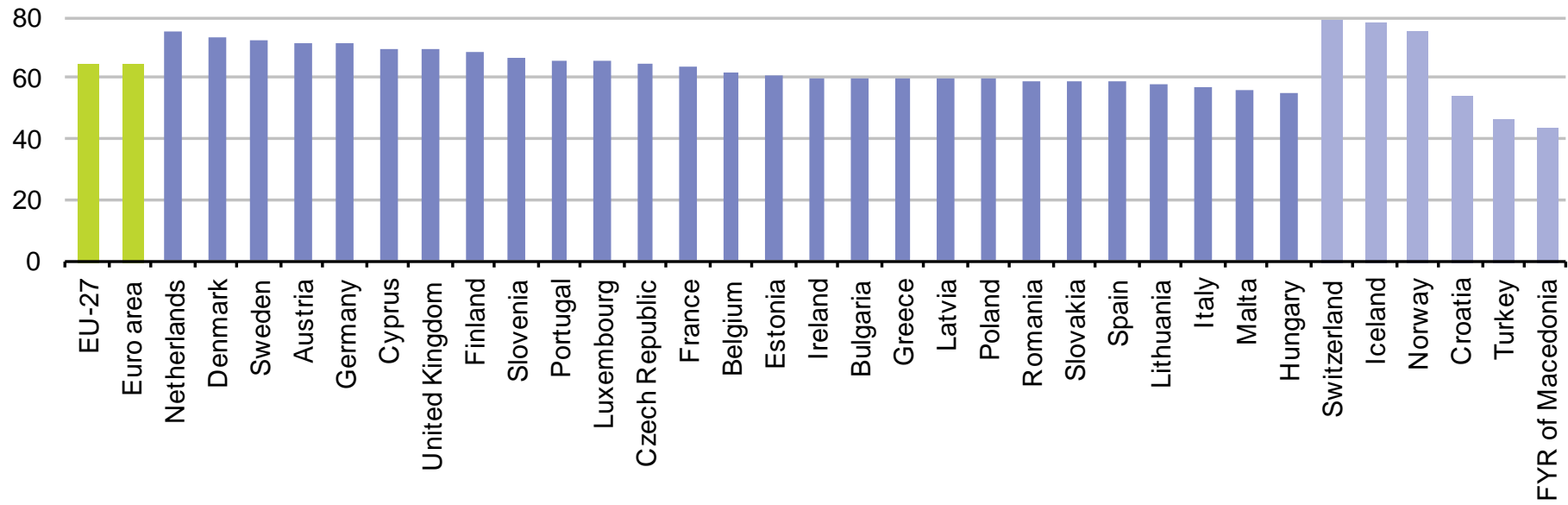
Στον πίνακα και στο διάγραμμα που ακολουθούν παρατίθενται τα συνολικά ετήσια ποσοστά απασχόλησης για τις χώρες της Ευρώπης, για το έτος 2010.

**Employment rate, age group 15-64, 2010
(%)**

	(%)
EU-27	64,2
Euro area	64,2
Netherlands	74,7
Denmark	73,4
Sweden	72,7
Austria	71,7
Germany	71,1
Cyprus	69,7
United Kingdom	69,5
Finland	68,1
Slovenia	66,2
Portugal	65,6
Luxembourg	65,2
Czech Republic	65,0
France	64,0
Belgium	62,0
Estonia	61,0
Ireland	60,0
Bulgaria	59,7
Greece	59,6
Latvia	59,3
Poland	59,3
Romania	58,8
Slovakia	58,8
Spain	58,6
Lithuania	57,8
Italy	56,9
Malta	56,0
Hungary	55,4
Switzerland	78,6
Iceland	78,2
Norway	75,3
Croatia	54,0
Turkey	46,3
FYR of Macedonia	43,5

Source: Eurostat (online data code: lfsi_emp_a)

Πίνακας 6 - Ετήσια συνολικά ποσοστά απασχόλησης Ευρωπαϊκών χωρών, έτους 2010.



Διάγραμμα 2 - Ετήσια συνολικά ποσοστά απασχόλησης Ευρωπαϊκών χωρών έτους 2010

1.2.7 Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών ανά φύλο

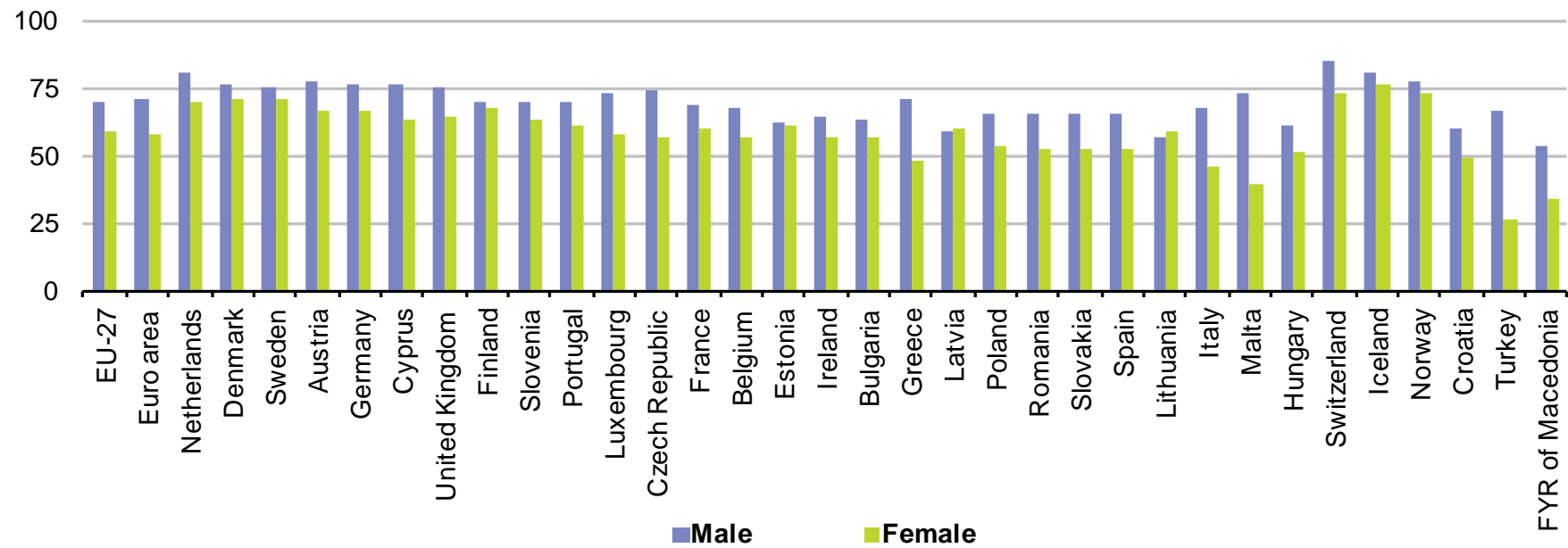
Στον πίνακα και στο διάγραμμα που ακολουθούν παρατίθενται τα ετήσια ποσοστά απασχόλησης ανά φύλο για τις χώρες της Ευρώπης, για το έτος 2010.

Employment rates by gender, 2010 (1) (%)

	Male	Female
EU-27	70,1	58,2
Euro area	70,4	57,9
Netherlands	80,0	69,3
Denmark	75,8	71,1
Sweden	75,1	70,3
Austria	77,1	66,4
Germany	76,0	66,1
Cyprus	76,6	63,0
United Kingdom	74,5	64,6
Finland	69,4	66,9
Slovenia	69,6	62,6
Portugal	70,1	61,1
Luxembourg	73,1	57,2
Czech Republic	73,5	56,3
France	68,3	59,9
Belgium	67,4	56,5
Estonia	61,5	60,6
Ireland	63,9	56,0
Bulgaria	63,0	56,4
Greece	70,9	48,1
Latvia	59,2	59,4
Poland	65,6	53,0
Romania	65,7	52,0
Slovakia	65,2	52,3
Spain	64,7	52,3
Lithuania	56,8	58,7
Italy	67,7	46,1
Malta	72,3	39,2
Hungary	60,4	50,6
Switzerland	84,6	72,5
Iceland	80,1	76,2
Norway	77,3	73,3
Croatia	59,4	48,8
Turkey	66,7	26,2
FYR of Macedonia	52,8	34,0

(1) The figure is ranked on the average of employment rates for males and females.
Source: Eurostat (online data code: lfsi_emp_a)

Πίνακας 7 - Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών ανά φύλο, έτος 2010



Διάγραμμα 3 - Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών ανά φύλο, έτος 2010

1.2.8 Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών ανά ηλικιακή ομάδα

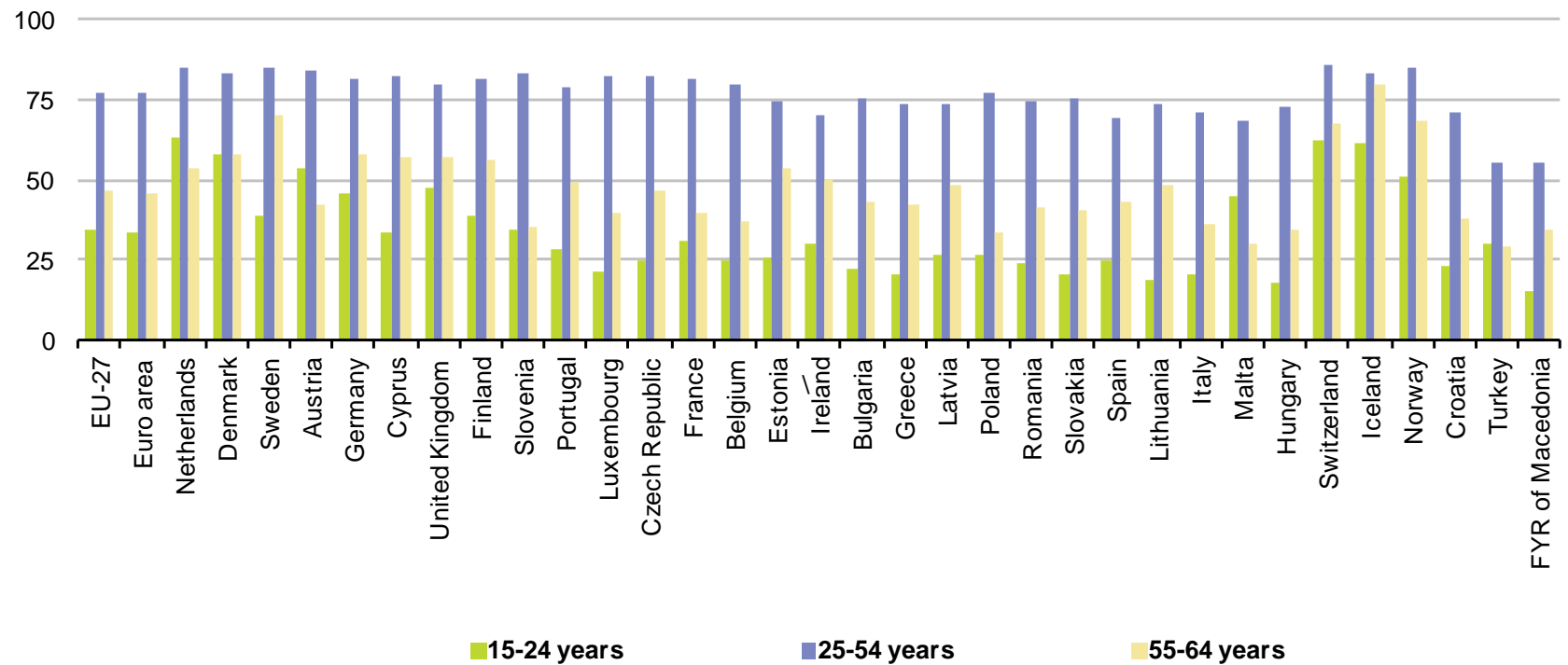
Στον πίνακα και στο διάγραμμα που ακολουθούν παρατίθενται τα ετήσια ποσοστά απασχόλησης ανά ηλικιακή ομάδα για τις χώρες της Ευρώπης, για το έτος 2010.

Employment rates by age group, 2010 (1)
(%)

	15-24 years	25-54 years	55-64 years
EU-27	34,1	77,6	46,3
Euro area	33,9	77,3	45,8
Netherlands	63,0	84,7	53,7
Denmark	58,1	83,2	57,6
Sweden	38,7	85,0	70,5
Austria	53,6	84,2	42,4
Germany	46,2	81,5	57,7
Cyprus	33,8	82,5	56,8
United Kingdom	47,6	79,8	57,1
Finland	38,8	81,6	56,2
Slovenia	34,1	83,7	35,0
Portugal	28,5	79,2	49,2
Luxembourg	21,2	82,3	39,6
Czech Republic	25,2	82,2	46,5
France	30,8	81,8	39,7
Belgium	25,2	80,0	37,3
Estonia	25,7	74,8	53,8
Ireland	30,5	70,3	50,0
Bulgaria	22,2	75,7	43,5
Greece	20,4	73,3	42,3
Latvia	26,4	73,4	48,2
Poland	26,3	77,1	34,0
Romania	24,3	74,4	41,1
Slovakia	20,6	75,8	40,5
Spain	24,9	69,6	43,6
Lithuania	19,2	73,8	48,6
Italy	20,5	71,1	36,6
Malta	44,8	68,7	30,2
Hungary	18,3	72,5	34,4
Switzerland	62,5	85,8	68,0
Iceland	61,7	82,9	79,8
Norway	51,4	84,7	68,6
Croatia	23,2	71,2	37,6
Turkey	30,0	55,4	29,6
FYR of Macedonia	15,4	55,8	34,2

(1) The figure is ranked on the overall employment rate.
Source: Eurostat (online data code: lfsi_emp_a)

Πίνακας 8 - Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών ανά ηλικιακή ομάδα, έτος 2010



Διάγραμμα 4 - Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών ανά ηλικιακή ομάδα, έτος 2010

1.2.9 Ετήσια ποσοστά μερικής απασχόλησης και ατόμων με 2^η δουλειά στην Ευρώπη, 2000-2010

Στον πίνακα που ακολουθεί παρατίθενται τα ετήσια ποσοστά μερικής απασχόλησης και ατόμων με 2^η δουλειά για τις χώρες της Ευρώπης, για τα έτη 2000, 2005 και 2010. Τα ποσοστά είναι επί τοις εκατό του πληθυσμού που εργάζεται.

Persons working part-time or with a second job, 2000-2010
(% of total employment)

	Persons working part-time			Persons with a second job		
	2000	2005	2010	2000	2005	2010
EU-27	16,2	17,8	19,2	3,9	3,7	3,7
Euro area	15,6	18,6	20,4	2,9	3,1	3,3
Belgium	18,9	22,0	24,0	3,8	3,9	4,1
Bulgaria	:	2,1	2,4	2,6	0,6	0,5
Czech Republic	5,3	4,9	5,9	2,7	2,4	2,2
Denmark	21,3	22,1	26,5	10,9	11,0	8,4
Germany (1)	19,4	24,0	26,2	2,4	3,3	3,6
Estonia	8,1	7,8	11,0	6,3	3,3	5,0
Ireland	16,4	:	22,4	1,8	2,1	2,1
Greece	4,5	5,0	6,4	3,8	2,8	3,1
Spain (2)	7,9	12,4	13,3	1,8	2,6	2,2
France	16,7	17,2	17,8	3,5	2,9	3,4
Italy (2)	8,4	12,8	15,0	1,4	1,6	1,5
Cyprus	8,4	8,9	9,3	5,7	6,1	3,3
Latvia	11,3	8,3	9,7	4,7	5,9	4,3
Lithuania	10,2	7,1	8,1	6,9	5,7	5,1
Luxembourg	10,4	17,4	17,9	1,1	1,8	3,0
Hungary	3,5	4,1	5,8	2,0	1,9	1,8
Malta	6,8	9,6	12,4	4,4	4,5	4,9
Netherlands (3)	41,5	46,1	48,9	5,9	6,2	7,2
Austria (2)	16,3	21,1	25,2	5,5	4,0	4,0
Poland	10,5	10,8	8,3	8,5	7,8	7,3
Portugal	10,9	11,2	11,6	6,2	6,4	6,0
Romania (2)	16,5	10,2	11,0	5,3	3,1	2,7
Slovenia	6,5	9,0	11,4	2,7	3,4	3,7
Slovakia	2,1	2,5	3,9	1,0	1,4	1,2
Finland	12,3	13,7	14,6	3,8	4,0	4,5
Sweden (2)	19,5	24,7	26,4	8,8	7,3	8,4
United Kingdom	25,1	25,2	26,9	4,4	3,7	3,8
Iceland	:	22,2	22,9	17,6	10,8	8,7
Norway	25,8	28,2	28,4	8,0	5,9	8,4
Switzerland	30,5	33,1	35,3	5,9	6,4	6,7
Croatia	:	10,1	9,7	:	3,4	2,7
FYR of Macedonia	:	:	5,9	:	:	2,1
Turkey	:	:	11,7	:	:	2,8

(1) Persons working part-time, break in series between 2000 and 2005.

(2) Break in series between 2000 and 2005.

(3) Break in series between 2005 and 2010.

Source: Eurostat (online data codes: tps00159, lfsa_e2gis and lfsa_egan)

Πίνακας 9 - Ετήσια ποσοστά μερικής απασχόλησης και ατόμων με 2^η δουλειά στην Ευρώπη, για τα έτη 2000-2010

1.2.10 Ετήσια ποσοστά μερικής απασχόλησης Ευρωπαϊκών χωρών ανά φύλο

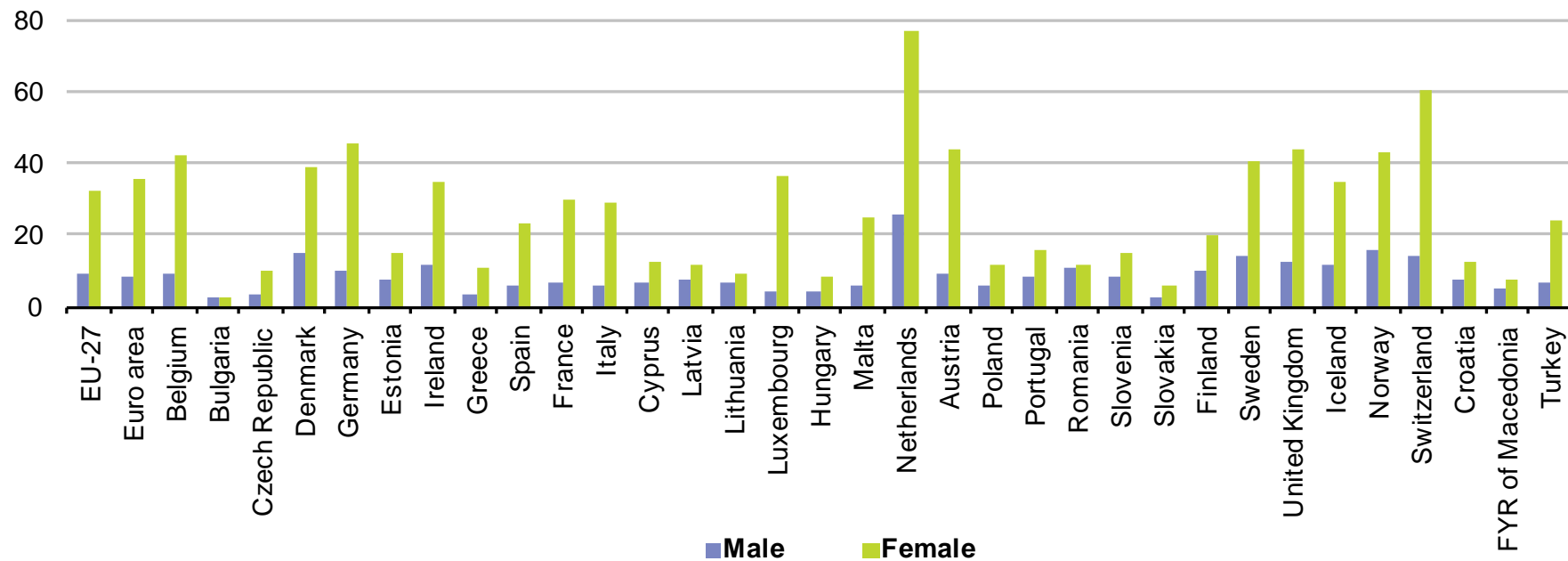
Στον πίνακα και στο διάγραμμα που ακολουθούν παρατίθενται τα ετήσια ποσοστά μερικής απασχόλησης ανά φύλο για τις χώρες της Ευρώπης, για το έτος 2010. Παρατηρείται ότι τα ποσοστά μερικής απασχόλησης των γυναικών είναι υψηλότερα των αντίστοιχων ποσοστών των αντρών.

Persons employed part-time, 2010 (% of total employment)

	Male	Female
EU-27	8,7	31,9
Euro area	8,4	35,1
Belgium	9,0	42,3
Bulgaria	2,2	2,6
Czech Republic	2,9	9,9
Denmark	15,2	39,0
Germany	9,7	45,5
Estonia	7,1	14,5
Ireland	11,8	34,7
Greece	3,7	10,4
Spain	5,4	23,2
France	6,7	30,0
Italy	5,5	29,0
Cyprus	6,5	12,7
Latvia	7,8	11,4
Lithuania	6,7	9,3
Luxembourg	4,0	36,0
Hungary	3,9	8,0
Malta	5,9	24,9
Netherlands	25,4	76,5
Austria	9,0	43,8
Poland	5,7	11,5
Portugal	8,2	15,5
Romania	10,6	11,4
Slovenia	8,6	14,7
Slovakia	2,8	5,4
Finland	10,0	19,6
Sweden	14,0	40,4
United Kingdom	12,6	43,3
Iceland	11,9	34,9
Norway	15,4	42,9
Switzerland	14,1	60,6
Croatia	7,3	12,5
FYR of Macedonia	5,0	7,4
Turkey	6,9	23,8

Source: Eurostat (online data code: tps00159)

Πίνακας 10 - Ετήσια ποσοστά μερικής απασχόλησης Ευρωπαϊκών χωρών ανά φύλο, έτος 2010



Διάγραμμα 5 - Ετήσια ποσοστά μερικής απασχόλησης Ευρωπαϊκών χωρών ανά φύλο, έτους 2010

1.2.11 Ετήσια ποσοστά απασχόλησης Ευρωπαϊκών χωρών με συμβόλαιο περιορισμένης διάρκειας

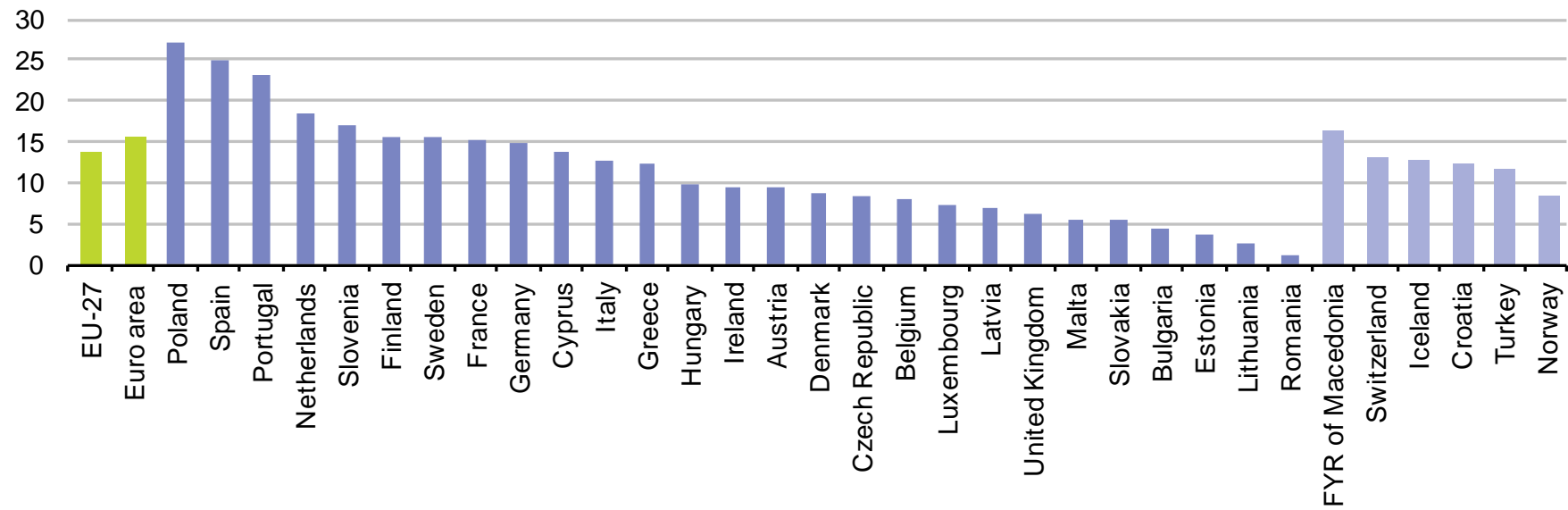
Στον πίνακα και στο διάγραμμα που ακολουθούν παρατίθενται τα ετήσια ποσοστά απασχόλησης με συμβόλαιο περιορισμένης διάρκειας για τις χώρες της Ευρώπης, για το έτος 2010. Τα ποσοστά υπολογίζονται επί τοις εκατό του πληθυσμού που εργάζεται. Παρατηρούνται ιδιαίτερα υψηλά ποσοστά σε Πολωνία, Ισπανία και Πορτογαλία.

Proportion of employees with a contract of limited duration, age group 15-64, 2010
(% of total employees)

	(%)
EU-27	13,9
Euro area	15,6
Poland	27,2
Spain	25,0
Portugal	23,0
Netherlands	18,3
Slovenia	17,1
Finland	15,4
Sweden	15,4
France	15,1
Germany	14,7
Cyprus	13,6
Italy	12,8
Greece	12,4
Hungary	9,6
Ireland	9,3
Austria	9,3
Denmark	8,6
Czech Republic	8,2
Belgium	8,1
Luxembourg	7,1
Latvia	6,8
United Kingdom	6,0
Malta	5,6
Slovakia	5,6
Bulgaria	4,4
Estonia	3,7
Lithuania	2,4
Romania	1,1
FYR of Macedonia	16,4
Switzerland	13,1
Iceland	12,5
Croatia	12,3
Turkey	11,4
Norway	8,4

Source: Eurostat (online data code: lfsa_etpga)

Πίνακας 11 - Ετήσια ποσοστά απασχόλησης με συμβόλαιο περιορισμένης διάρκειας στις χώρες της Ευρώπης, έτος 2010



Διάγραμμα 6 - Ετήσια ποσοστά απασχόλησης με συμβόλαιο περιορισμένης διάρκειας στις χώρες της Ευρώπης, έτους 2010

ΚΕΦΑΛΑΙΟ 2 - ΕΠΙΛΟΓΗ ΜΕΤΑΒΛΗΤΩΝ ΑΠΑΣΧΟΛΗΣΗΣ ΓΙΑ ΠΕΡΑΙΤΕΡΩ ΑΝΑΛΥΣΗ

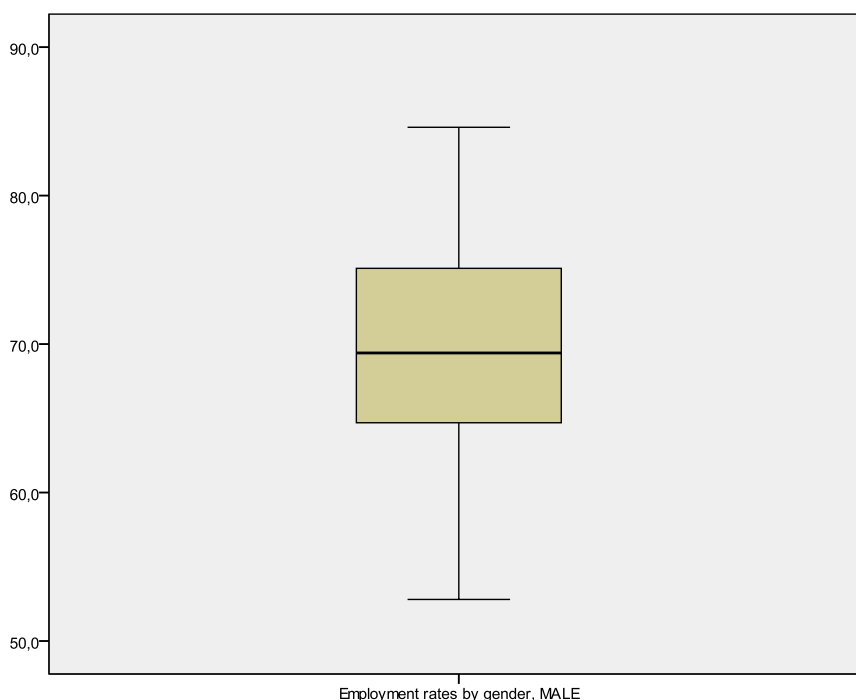
Στόχος του 2^{ου} κεφαλαίου είναι η αναλυτικότερη παρουσίαση 12 συγκεκριμένων χαρακτηριστικών απασχόλησης των Ευρωπαϊκών χωρών, πάνω στα οποία θα βασίσουμε την κατασκευή δύο νέων «δεικτών» απασχόλησης. Ο τρόπος κατασκευής των δύο νέων «δεικτών» θα αναλυθεί σε επόμενο κεφάλαιο, με τη βοήθεια της κατάλληλης μεθόδου πολυμεταβλητής στατιστικής ανάλυσης.

2.1 Μεταβλητή “Male”

“*Employment rates by gender, MALE*”: Εκφράζει το επί τοις εκατό ποσοστό απασχόλησης του αντρικού πληθυσμού στις 33 χώρες της Ευρώπης. Έχει μέση τιμή 69.2 και τυπική απόκλιση 7.4. Η ελάχιστη παρατηρούμενη τιμή είναι 52.8 (FYROM) και η μεγαλύτερη 84.6 (Ελβετία). Τα υπόλοιπα από τα βασικά περιγραφικά μέτρα της μεταβλητής φαίνονται στον πίνακα που ακολουθεί.

Descriptives			Statistic	Std. Error
Employment rates by gender, MALE	Mean		69,221	1,2878
	95% Confidence Interval for Mean	Lower Bound	66,598	
		Upper Bound	71,844	
	5% Trimmed Mean		69,290	
	Median		69,400	
	Variance		54,729	
	Std. Deviation		7,3979	
	Minimum		52,8	
	Maximum		84,6	
	Range		31,8	
	Interquartile Range		11,1	
	Skewness		-,115	,409
	Kurtosis		-,408	,798

Πίνακας 12 - Βασικά περιγραφικά μέτρα της μεταβλητής “Male”



Θηκόγραμμα 1- Μεταβλητή “Male”

Βάσει του p-value του στατιστικού ελέγχου Kolmogorov-Smirnov, σε επίπεδο σημαντικότητας 5% δεν απορρίπτεται η μηδενική υπόθεση ότι η μεταβλητή ακολουθεί την κανονική κατανομή.

Tests of Normality

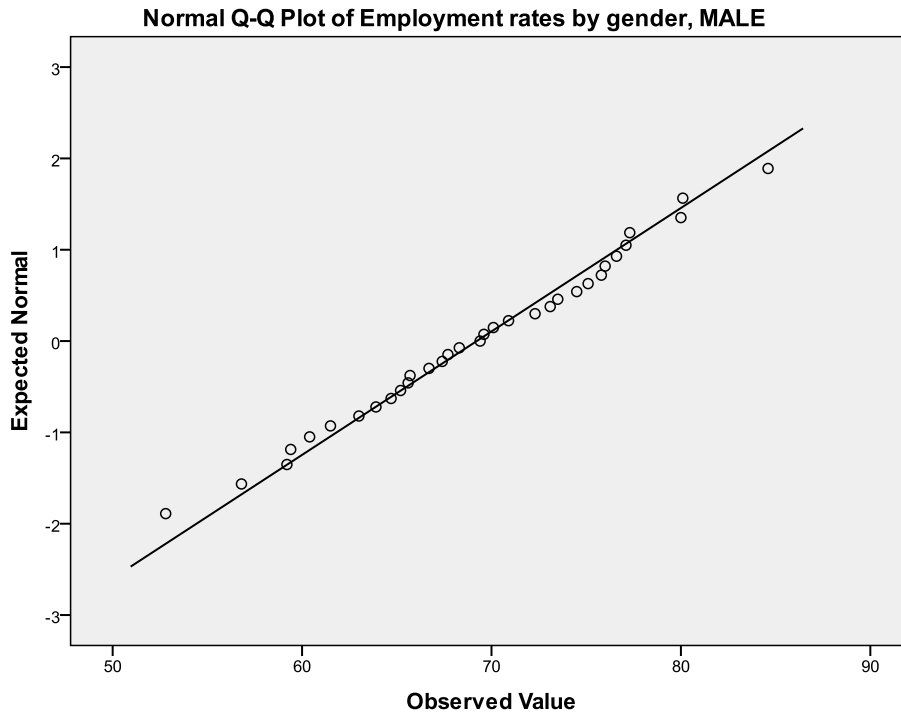
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Employment rates by gender, MALE	,065	33	,200*	,991	33	,993

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Πίνακας 13 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Male”

Το παραπάνω αποτέλεσμα επιβεβαιώνεται και από τον γραφικό έλεγχο κανονικότητας. Στο διάγραμμα “Q-Q Plot” φαίνεται μία πολύ ικανοποιητική προσέγγιση των αναμενόμενων θεωρητικών κανονικών τιμών από τις αντίστοιχες πραγματικές, για το χαρακτηριστικό απασχόλησης που εξετάζεται.



Διάγραμμα 7 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Male”

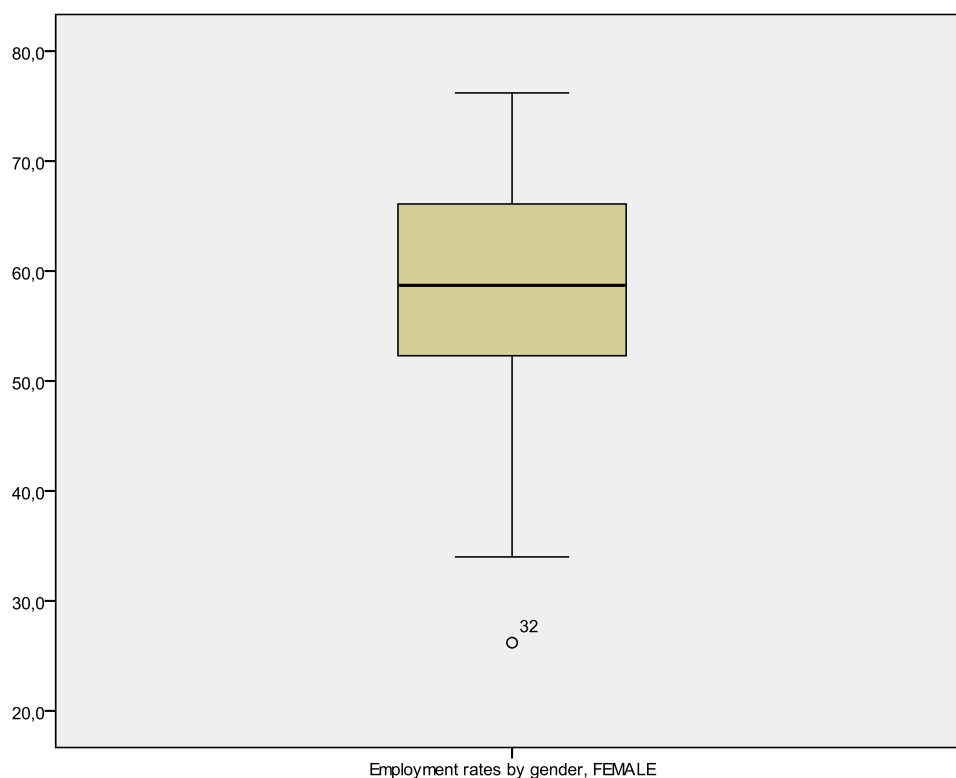
2.2 Μεταβλητή “Female”

“*Employment rates by gender, FEMALE*”: Εκφράζει το επί τοις εκατό ποσοστό απασχόλησης του γυναικείου πληθυσμού στις 33 χώρες της Ευρώπης. Έχει μέση τιμή 57.8 και τυπική απόκλιση 11.2. Η ελάχιστη παρατηρούμενη τιμή είναι 26.2 (Τουρκία) και η μεγαλύτερη 76.2 (Ισλανδία). Τα υπόλοιπα από τα βασικά περιγραφικά μέτρα της μεταβλητής παρουσιάζονται στον πίνακα που ακολουθεί.

Descriptives			Statistic	Std. Error
Employment rates by gender, FEMALE	Mean		57,788	1,9487
	95% Confidence Interval for Mean	Lower Bound	53,818	
		Upper Bound	61,757	
	5% Trimmed Mean		58,413	
	Median		58,700	
	Variance		125,320	
	Std. Deviation		11,1946	
	Minimum		26,2	
	Maximum		76,2	
	Range		50,0	
	Interquartile Range		14,1	
	Skewness		-,807	,409
	Kurtosis		,992	,798

Πίνακας 14 - Βασικά περιγραφικά μέτρα της μεταβλητής “Female”

Στο θηκόγραμμα που ακολουθεί, το ποσοστό απασχόλησης γυναικών της Τουρκίας σημειώνεται ως έκτροπη παρατήρηση (No 32).



Θηκόγραμμα 2 - Μεταβλητή “Female”

Βάσει του p-value του στατιστικού ελέγχου Kolmogorov-Smirnov, σε επίπεδο σημαντικότητας 5% δεν απορρίπτεται η μηδενική υπόθεση ότι η μεταβλητή ακολουθεί την κανονική κατανομή.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Employment rates by gender, FEMALE	,103	33	,200*	,957	33	,206

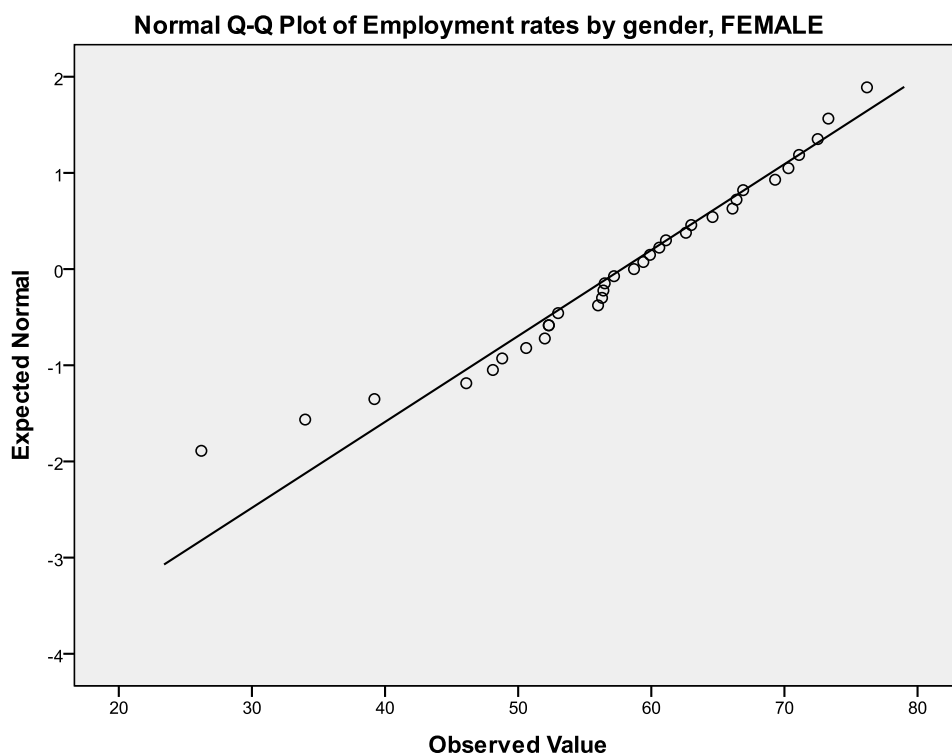
a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Πίνακας 15 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Female”

Το παραπάνω αποτέλεσμα επιβεβαιώνεται και από τον γραφικό έλεγχο κανονικότητας. Στο διάγραμμα “Q-Q Plot” φαίνεται μία ικανοποιητική προσέγγιση των

αναμενόμενων θεωρητικών κανονικών τιμών από τις αντίστοιχες πραγματικές, για το χαρακτηριστικό απασχόλησης που εξετάζεται.



Διάγραμμα 8 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Female”

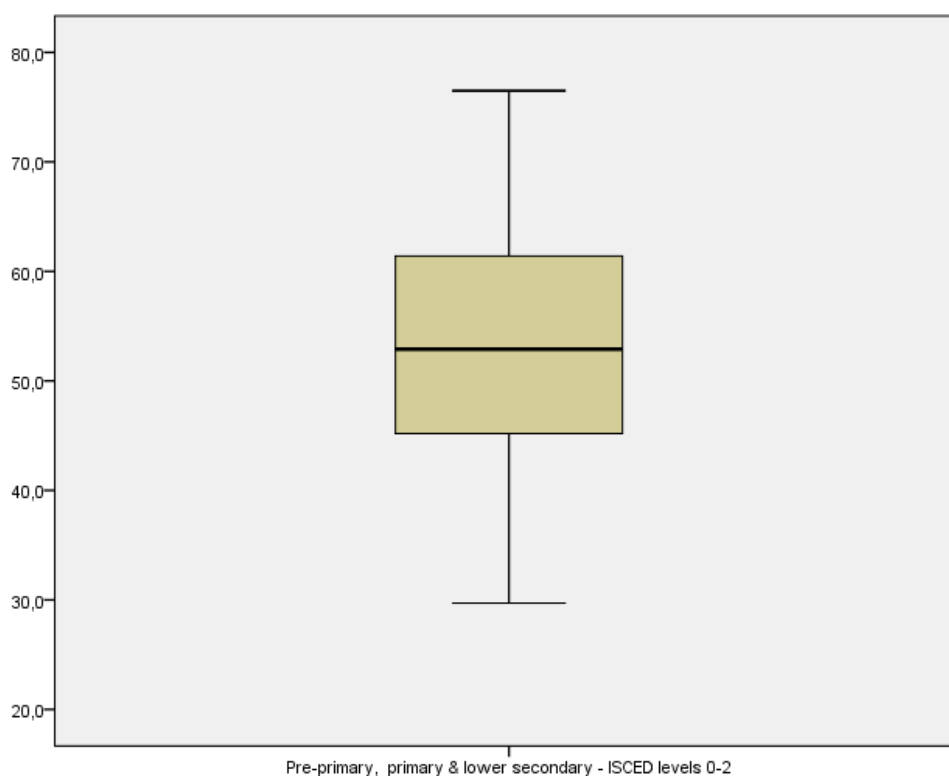
2.3 Μεταβλητή “Low_Educated”

“Pre-primary, primary & lower secondary - ISCED levels 0-2”: Εκφράζει το ποσοστό επί τοις εκατό εργαζόμενων χαμηλού μορφωτικού επιπέδου στην ηλικιακή ομάδα 25-64 ετών, για τις 33 χώρες της Ευρώπης. Έχει μέση τιμή 52.3 και τυπική απόκλιση 11.3. Η χαμηλότερη τιμή είναι 29.7 (Σλοβακία) και η υψηλότερη 76.5 (Ισλανδία). Τα υπόλοιπα από τα βασικά περιγραφικά μέτρα της μεταβλητής παρουσιάζονται στον πίνακα που ακολουθεί.

Descriptives

			Statistic	Std. Error
Pre-primary, primary & lower secondary - ISCED levels 0-2	Mean		52,279	1,9671
	95% Confidence Interval for Mean	Lower Bound	48,272	
		Upper Bound	56,286	
	5% Trimmed Mean		52,286	
	Median		52,900	
	Variance		127,687	
	Std. Deviation		11,2999	
	Minimum		29,7	
	Maximum		76,5	
	Range		46,8	
	Interquartile Range		17,4	
	Skewness		-,050	,409
	Kurtosis		-,407	,798

Πίνακας 16 - Βασικά περιγραφικά μέτρα της μεταβλητής “Low_Educated”



Θηκόγραμμα 3 - Μεταβλητή “Low_Educated”

Βάσει του p-value του στατιστικού ελέγχου Kolmogorov-Smirnov, σε επίπεδο σημαντικότητας 5% δεν απορρίπτεται η μηδενική υπόθεση ότι η μεταβλητή ακολουθεί την κανονική κατανομή.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pre-primary, primary & lower secondary - ISCED levels 0-2	,073	33	,200 [*]	,989	33	,983

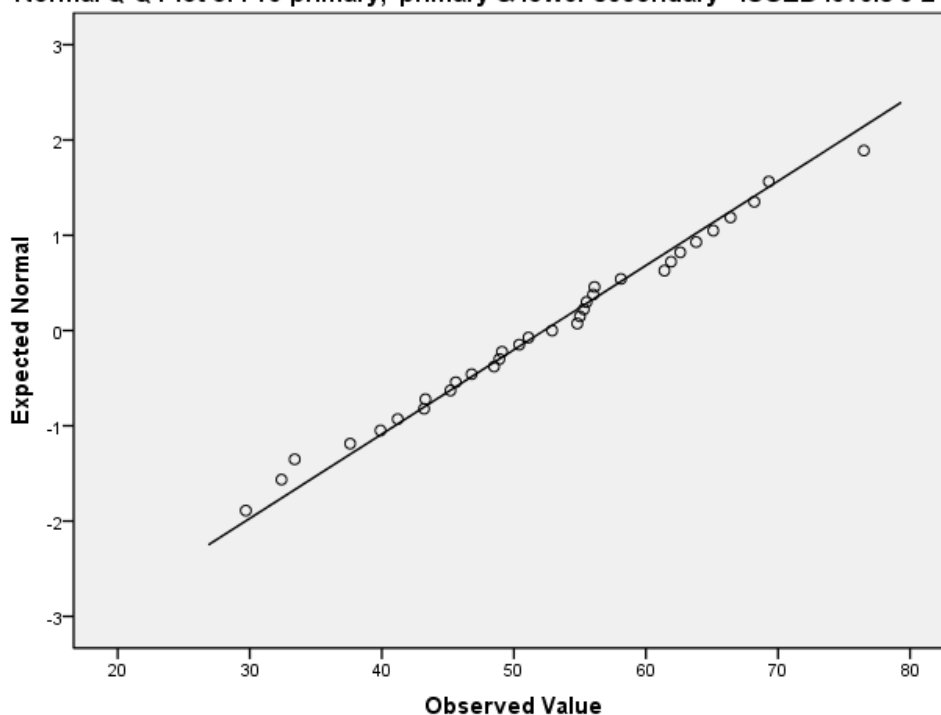
a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Πίνακας 17 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Low_Educated”

Το παραπάνω αποτέλεσμα επιβεβαιώνεται και από τον γραφικό έλεγχο κανονικότητας. Στο διάγραμμα “Q-Q Plot” φαίνεται μία πολύ ικανοποιητική προσέγγιση των αναμενόμενων θεωρητικών κανονικών τιμών από τις αντίστοιχες πραγματικές, για το χαρακτηριστικό απασχόλησης που εξετάζεται.

Normal Q-Q Plot of Pre-primary, primary & lower secondary - ISCED levels 0-2



Διάγραμμα 9 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Low_Educated”

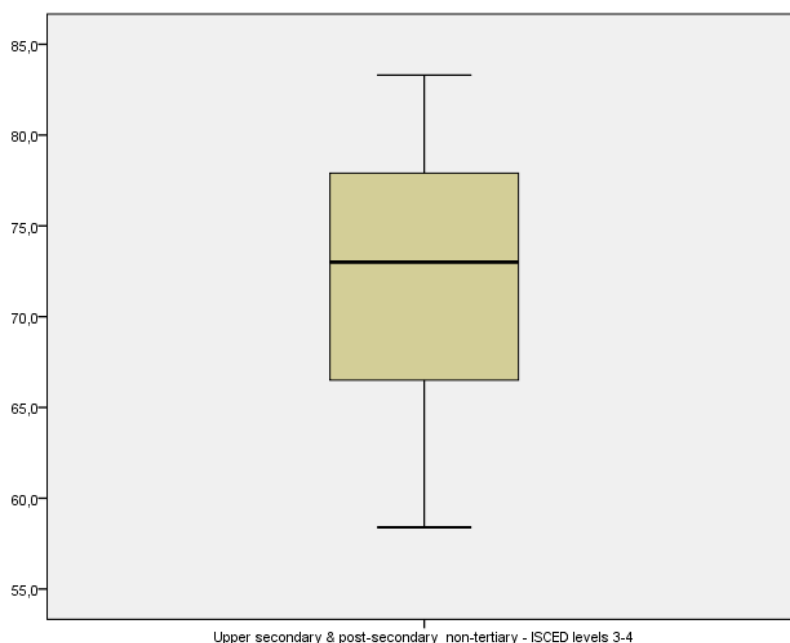
2.4 Μεταβλητή “Mid_Educated”

“Upper secondary & post-secondary non-tertiary - ISCED levels 3-4”: Εκφράζει το επί τοις εκατό ποσοστό εργαζόμενων ανώτερου μορφωτικού επιπέδου στην ηλικιακή ομάδα 25-64 ετών, για τις 33 χώρες της Ευρώπης. Έχει μέση τιμή 72.6

και τυπική απόκλιση 6.8. Η χαμηλότερη τιμή είναι 58.4 (FYROM) και η υψηλότερη 83.3 (Σουηδία). Τα υπόλοιπα από τα βασικά περιγραφικά μέτρα της μεταβλητής παρουσιάζονται στον πίνακα που ακολουθεί.

Descriptives			Statistic	Std. Error
Upper secondary & post-secondary non-tertiary - ISCED levels 3-4	Mean		72,582	1,1819
	95% Confidence Interval for Mean	Lower Bound	70,174	
		Upper Bound	74,989	
	5% Trimmed Mean		72,768	
	Median		73,000	
	Variance		46,097	
	Std. Deviation		6,7894	
	Minimum		58,4	
	Maximum		83,3	
	Range		24,9	
	Interquartile Range		12,0	
	Skewness		-,263	,409
	Kurtosis		-,851	,798

Πίνακας 18 - Βασικά περιγραφικά μέτρα της μεταβλητής “Mid_Educated”



Θηκόγραμμα 4 - Μεταβλητή “Mid_Educated”

Βάσει του p-value του στατιστικού ελέγχου Kolmogorov-Smirnov, σε επίπεδο σημαντικότητας 5% δεν απορρίπτεται η μηδενική υπόθεση ότι η μεταβλητή ακολουθεί την κανονική κατανομή.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Upper secondary & post-secondary non-tertiary - ISCED levels 3-4	,088	33	,200*	,964	33	,334

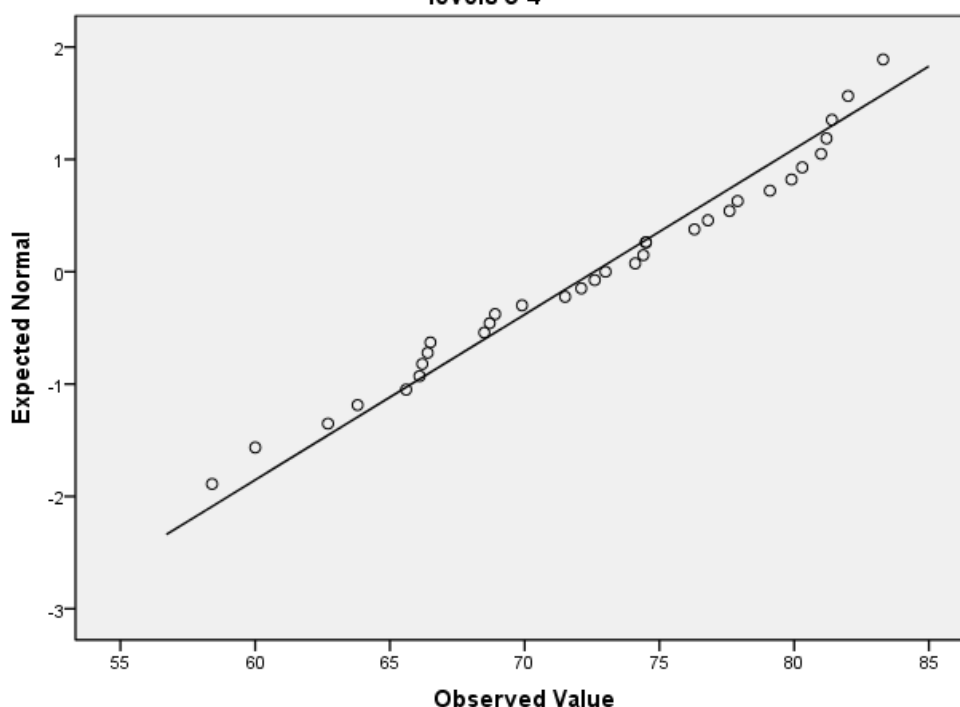
a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Πίνακας 19 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Mid_Educated”

Το παραπάνω αποτέλεσμα επιβεβαιώνεται και από τον γραφικό έλεγχο κανονικότητας. Στο διάγραμμα “Q-Q Plot” φαίνεται μία πολύ ικανοποιητική προσέγγιση των αναμενόμενων θεωρητικών κανονικών τιμών από τις αντίστοιχες πραγματικές, για το χαρακτηριστικό απασχόλησης που εξετάζεται.

Normal Q-Q Plot of Upper secondary & post-secondary non-tertiary - ISCED levels 3-4



Διάγραμμα 10 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Mid_Educated”

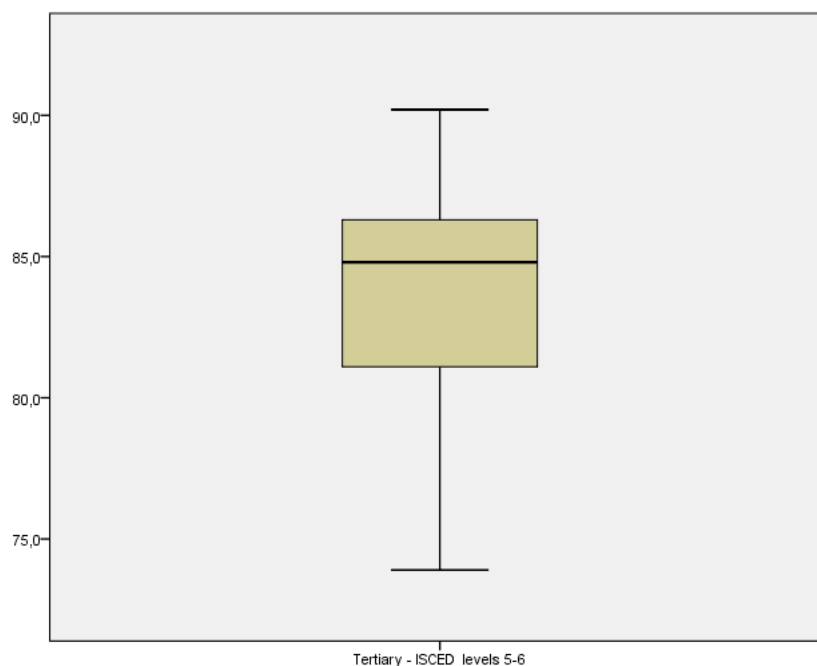
2.5 Μεταβλητή “High_Educated”

“Tertiary - ISCED levels 5-6”: Εκφράζει το επί τοις εκατό ποσοστό εργαζόμενων ανώτατου μορφωτικού επιπέδου στην ηλικιακή ομάδα 25-64 ετών, για τις 33 χώρες της Ευρώπης. Έχει μέση τιμή 83.7 και τυπική απόκλιση 3.9. Η χαμηλότερη τιμή είναι 73.9 (FYROM) και η υψηλότερη 90.2 (Νορβηγία). Τα υπόλοιπα

από τα βασικά περιγραφικά μέτρα της μεταβλητής παρουσιάζονται στον πίνακα που ακολουθεί.

Descriptives			Statistic	Std. Error
Tertiary - ISCED levels 5-6	Mean		83,658	,6760
	95% Confidence Interval for Mean	Lower Bound	82,281	
		Upper Bound	85,035	
	5% Trimmed Mean		83,821	
	Median		84,800	
	Variance		15,082	
	Std. Deviation		3,8835	
	Minimum		73,9	
	Maximum		90,2	
	Range		16,3	
	Interquartile Range		6,0	
	Skewness		-,663	,409
	Kurtosis		,013	,798

Πίνακας 20 - Βασικά περιγραφικά μέτρα της μεταβλητής “High_Educated”



Θηκόγραμμα 5 - Μεταβλητή “High_Educated”

Βάσει του p-value του στατιστικού ελέγχου Kolmogorov-Smirnov, σε επίπεδο σημαντικότητας 5% δεν απορρίπτεται η μηδενική υπόθεση ότι η μεταβλητή ακολουθεί την κανονική κατανομή.

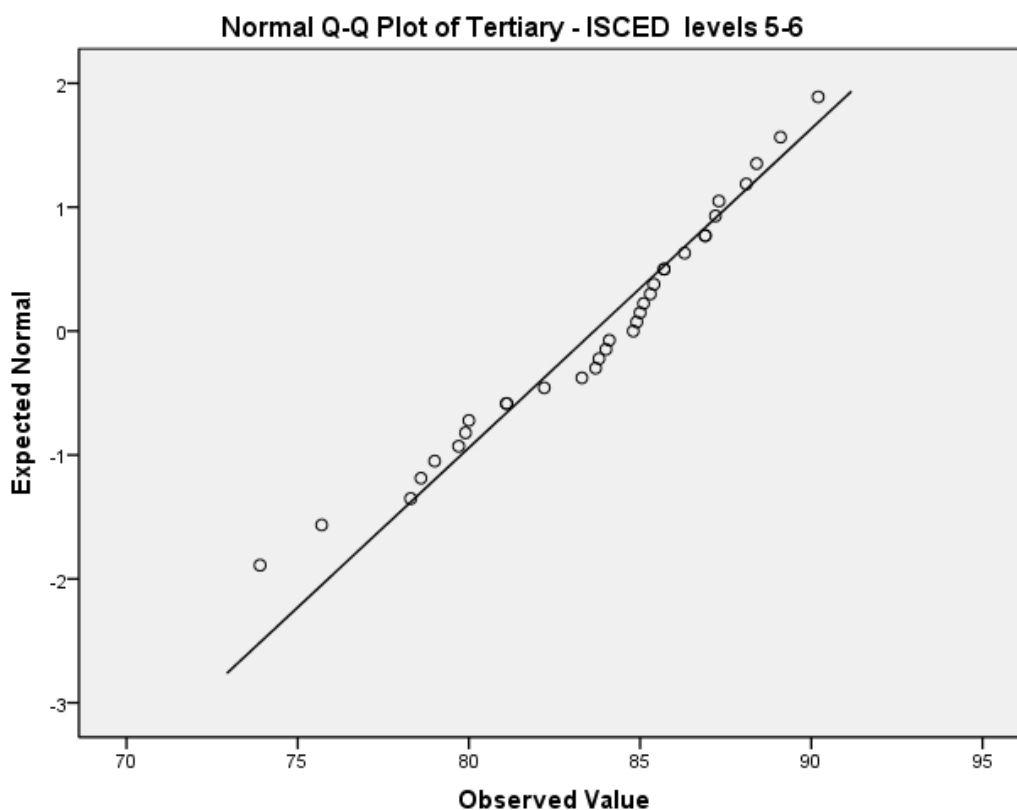
Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Tertiary - ISCED levels 5-6	,141	33	,096	,957	33	,217

a. Lilliefors Significance Correction

Πίνακας 21 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “High_Educated”

Το παραπάνω αποτέλεσμα επιβεβαιώνεται και από τον γραφικό έλεγχο κανονικότητας. Στο διάγραμμα “Q-Q Plot” φαίνεται μία ικανοποιητική προσέγγιση των αναμενόμενων θεωρητικών κανονικών τιμών από τις αντίστοιχες πραγματικές, για το χαρακτηριστικό απασχόλησης που εξετάζεται.



Διάγραμμα 11 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “High_Educated”

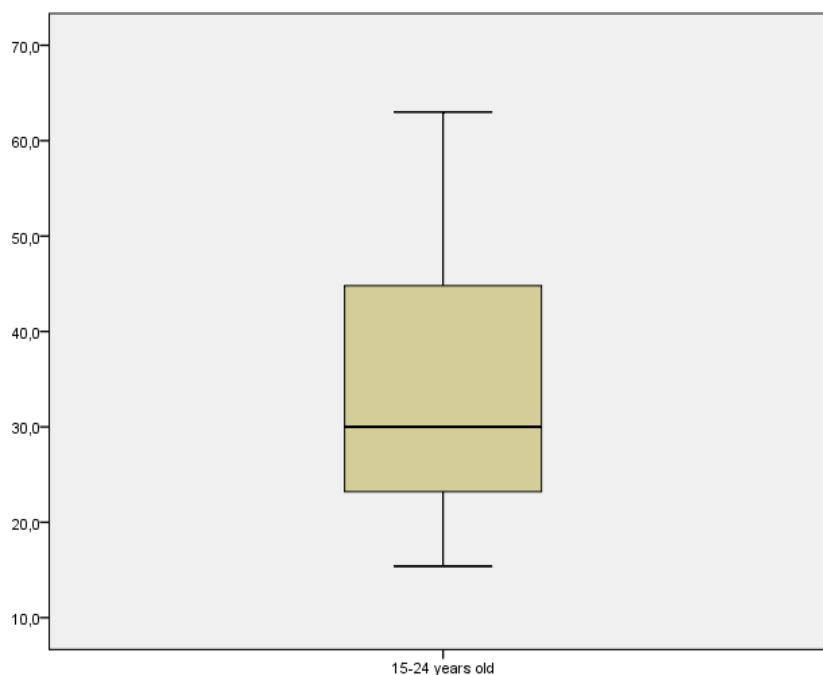
2.6 Μεταβλητή “Young”

“15-24 years old”: Εκφράζει το επί τοις εκατό ποσοστό εργαζόμενων της ηλικιακής ομάδας 16-24 ετών, για τις 33 χώρες της Ευρώπης. Έχει μέση τιμή 34 και τυπική απόκλιση 14.2. Η χαμηλότερη τιμή είναι 15.4 (FYROM) και η υψηλότερη 63

(Ολλανδία). Τα υπόλοιπα από τα βασικά περιγραφικά μέτρα της μεταβλητής παρουσιάζονται στον πίνακα που ακολουθεί.

Descriptives			Statistic	Std. Error
15-24 years old	Mean		33,985	2,4771
	95% Confidence Interval for Mean	Lower Bound	28,939	
		Upper Bound	39,030	
	5% Trimmed Mean		33,353	
	Median		30,000	
	Variance		202,483	
	Std. Deviation		14,2296	
	Minimum		15,4	
	Maximum		63,0	
	Range		47,6	
	Interquartile Range		22,8	
	Skewness		,826	,409
	Kurtosis		-,515	,798

Πίνακας 22 - Βασικά περιγραφικά μέτρα της μεταβλητής “Young”



Θηκόγραμμα 6 - Μεταβλητή “Young”

Βάσει του p-value του στατιστικού ελέγχου Kolmogorov-Smirnov, σε επίπεδο σημαντικότητας 5% απορρίπτεται η μηδενική υπόθεση ότι η μεταβλητή ακολουθεί την κανονική κατανομή.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
15-24 years old	,164	33	,024	,890	33	,003

a. Lilliefors Significance Correction

Πίνακας 23 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Young”

Το παραπάνω αποτέλεσμα επιβεβαιώνεται και από τον γραφικό έλεγχο κανονικότητας. Στο διάγραμμα “Q-Q Plot” φαίνεται ότι δεν υπάρχει ικανοποιητική προσέγγιση των αναμενόμενων θεωρητικών κανονικών τιμών από τις αντίστοιχες πραγματικές, για το χαρακτηριστικό απασχόλησης που εξετάζεται.



Διάγραμμα 12 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Young”

2.7 Μεταβλητή “Mid_Age”

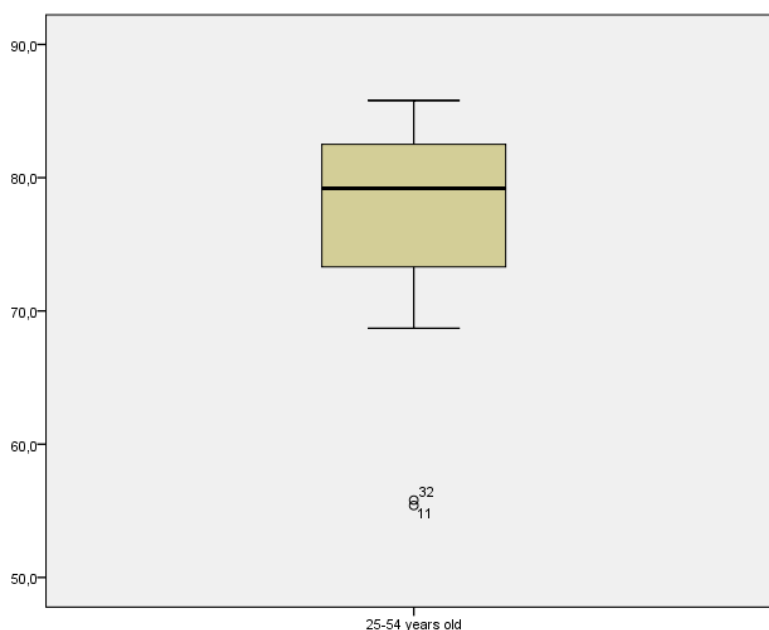
“25-54 years old”: Εκφράζει το επί τοις εκατό ποσοστό εργαζόμενων της ηλικιακής ομάδας 25-54 ετών, για τις 33 χώρες της Ευρώπης. Έχει μέση τιμή 77 και τυπική απόκλιση 7.5. Η χαμηλότερη τιμή είναι 55.4 (Τουρκία) και η υψηλότερη 85.8

(Ελβετία). Τα υπόλοιπα από τα βασικά περιγραφικά μέτρα της μεταβλητής παρουσιάζονται στον πίνακα που ακολουθεί.

Descriptives			Statistic	Std. Error
25-54 years old	Mean		76,994	1,3135
	95% Confidence Interval for Mean	Lower Bound	74,318	
		Upper Bound	79,669	
	5% Trimmed Mean		77,713	
	Median		79,200	
	Variance		56,934	
	Std. Deviation		7,5455	
	Minimum		55,4	
	Maximum		85,8	
	Range		30,4	
	Interquartile Range		9,8	
	Skewness		-1,352	,409
	Kurtosis		2,056	,798

Πίνακας 24 - Βασικά περιγραφικά μέτρα της μεταβλητής “Mid_Age”

Στο θηκόγραμμα που ακολουθεί σημειώνονται ως έκτροπες παρατηρήσεις τα ποσοστά που αντιστοιχούν στην Τουρκία (No 32) και στη FYROM (No 11).



Θηκόγραμμα 7 - Μεταβλητή “Mid_Age”

Βάσει του p-value του στατιστικού ελέγχου Kolmogorov-Smirnov, σε επίπεδο σημαντικότητας 5% δεν απορρίπτεται (οριακά) η μηδενική υπόθεση ότι η μεταβλητή ακολουθεί την κανονική κατανομή.

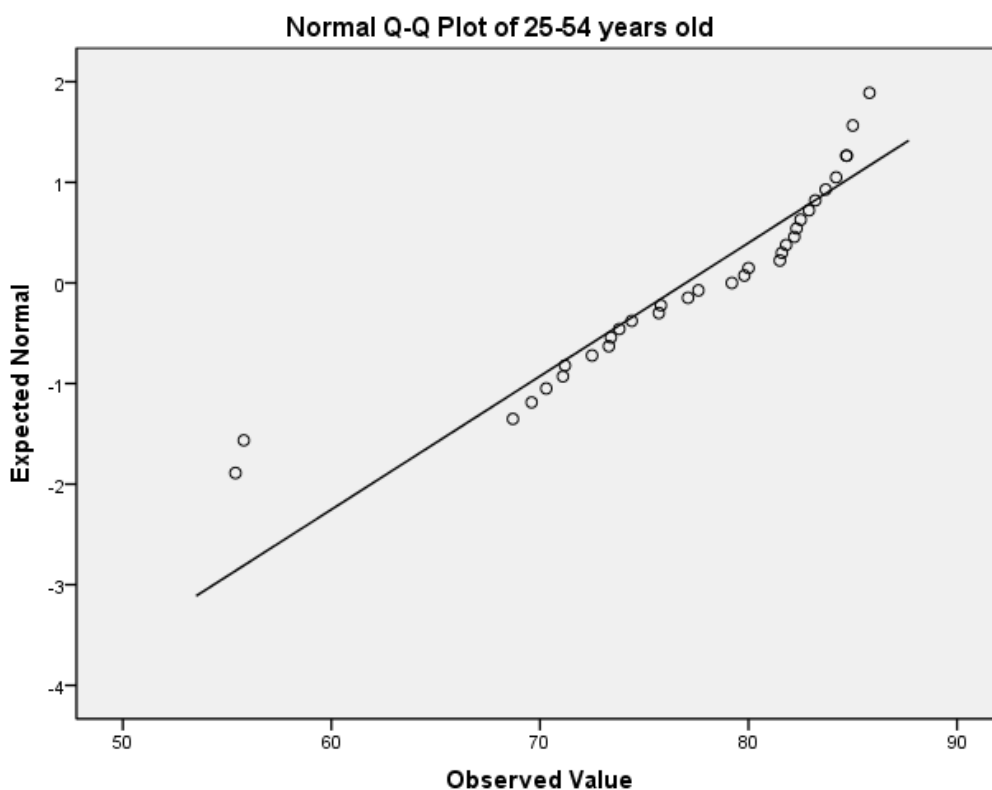
Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
25-54 years old	,149	33	,061	,867	33	,001

a. Lilliefors Significance Correction

Πίνακας 25 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Mid_Age”

Το παραπάνω αποτέλεσμα επιβεβαιώνεται και από τον γραφικό έλεγχο κανονικότητας. Στο διάγραμμα “Q-Q Plot”, με εξαίρεση τις έκτροπες παρατηρήσεις που επισημάνθηκαν, φαίνεται κατά τα άλλα μία ικανοποιητική προσέγγιση των αναμενόμενων θεωρητικών κανονικών τιμών από τις αντίστοιχες πραγματικές, για το χαρακτηριστικό απασχόλησης που εξετάζεται.



Διάγραμμα 13 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Mid_Age”

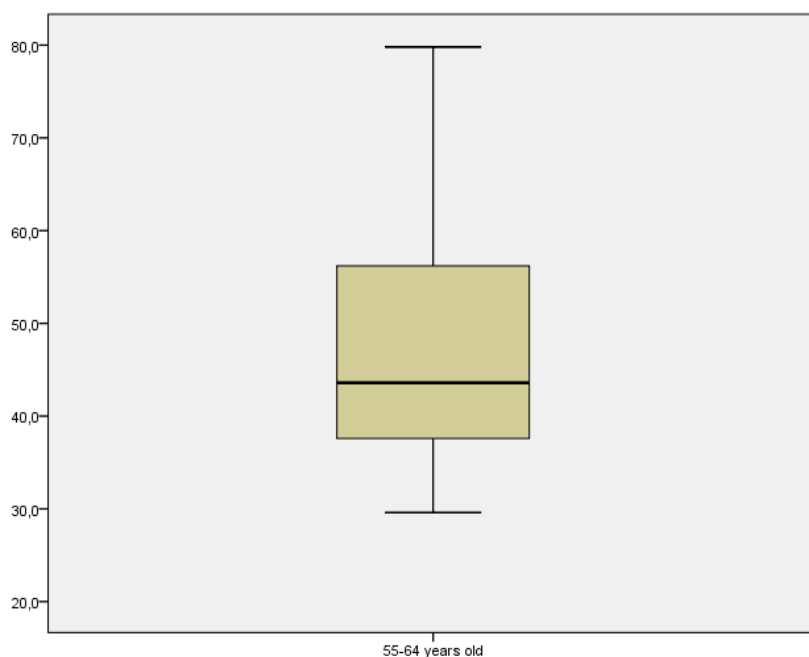
2.8 Μεταβλητή “Older”

“55-64 years old”: Εκφράζει το επί τοις εκατό ποσοστό εργαζόμενων της ηλικιακής ομάδας 55-64 ετών, για τις 33 χώρες της Ευρώπης. Έχει μέση τιμή 47.2 και τυπική απόκλιση 12.3. Η χαμηλότερη τιμή είναι 29.6 (Τουρκία) και η υψηλότερη

79.8 (Ισλανδία). Τα υπόλοιπα από τα βασικά περιγραφικά μέτρα της μεταβλητής παρουσιάζονται στον πίνακα που ακολουθεί.

Descriptives			Statistic	Std. Error
55-64 years old	Mean		47,164	2,1478
	95% Confidence Interval for Mean	Lower Bound	42,789	
		Upper Bound	51,538	
	5% Trimmed Mean		46,517	
	Median		43,600	
	Variance		152,224	
	Std. Deviation		12,3379	
	Minimum		29,6	
	Maximum		79,8	
	Range		50,2	
	Interquartile Range		19,0	
	Skewness		,855	,409
	Kurtosis		,289	,798

Πίνακας 26 - Βασικά περιγραφικά μέτρα της μεταβλητής “Older”



Θηκόγραμμα 8 - Μεταβλητή “Older”

Βάσει του p-value του στατιστικού ελέγχου Kolmogorov-Smirnov, σε επίπεδο σημαντικότητας 5% δεν απορρίπτεται η μηδενική υπόθεση ότι η μεταβλητή ακολουθεί την κανονική κατανομή.

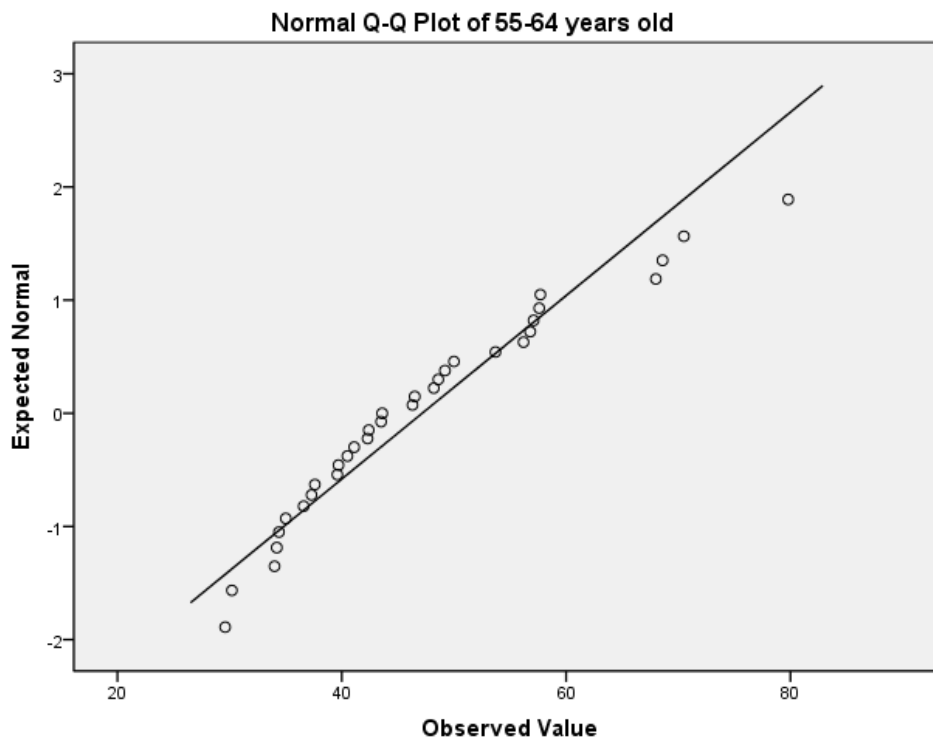
Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
55-64 years old	,129	33	,180	,936	33	,050

a. Lilliefors Significance Correction

Πίνακας 27 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Older”

Το παραπάνω αποτέλεσμα επιβεβαιώνεται και από τον γραφικό έλεγχο κανονικότητας. Στο διάγραμμα “Q-Q Plot” φαίνεται γενικά μία ικανοποιητική προσέγγιση των αναμενόμενων θεωρητικών κανονικών τιμών από τις αντίστοιχες πραγματικές, για το χαρακτηριστικό απασχόλησης που εξετάζεται.



Διάγραμμα 14 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Older”

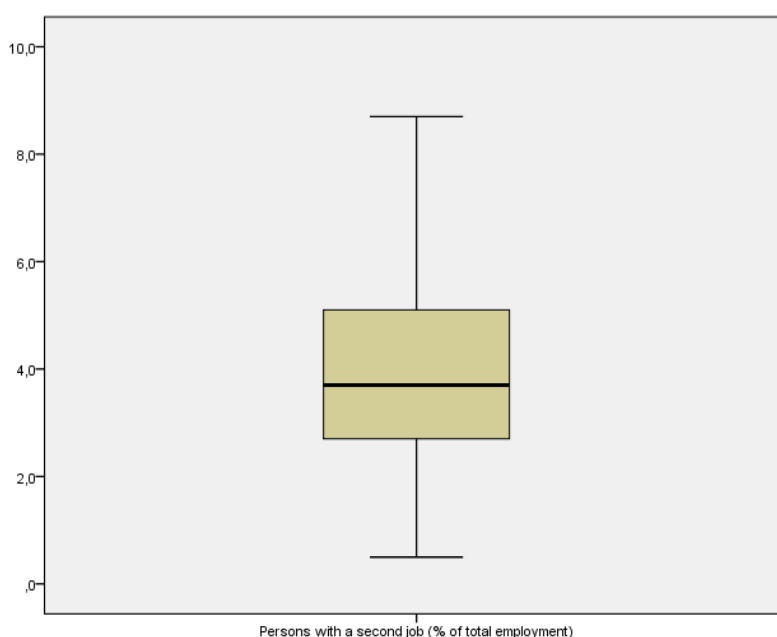
2.9 Μεταβλητή “With_Second_Job”

“Persons with a second job (% of total employment)”: Εκφράζει το επί τοις εκατό ποσοστό εργαζόμενων με δεύτερη δουλειά, ως προς το σύνολο των εργαζόμενων, για τις 33 χώρες της Ευρώπης. Έχει μέση τιμή 4.2 και τυπική απόκλιση 2.3. Η χαμηλότερη τιμή είναι 0.5 (Βουλγαρία) και η υψηλότερη 8.7

(Ισλανδία). Τα υπόλοιπα από τα βασικά περιγραφικά μέτρα της μεταβλητής παρουσιάζονται στον πίνακα που ακολουθεί.

Descriptives			Statistic	Std. Error
Persons with a second job (% of total employment)	Mean		4,203	,3980
	95% Confidence Interval for Mean	Lower Bound	3,392	
		Upper Bound	5,014	
	5% Trimmed Mean		4,150	
	Median		3,700	
	Variance		5,227	
	Std. Deviation		2,2862	
	Minimum		,5	
	Maximum		8,7	
	Range		8,2	
	Interquartile Range		3,1	
	Skewness		,645	,409
	Kurtosis		-,530	,798

Πίνακας 28 - Βασικά περιγραφικά μέτρα της μεταβλητής “With_Second_Job”



Θηκόγραμμα 9 - Μεταβλητή “With_Second_Job”

Βάσει του p-value του στατιστικού ελέγχου Kolmogorov-Smirnov, σε επίπεδο σημαντικότητας 5% δεν απορρίπτεται η μηδενική υπόθεση ότι η μεταβλητή ακολουθεί την κανονική κατανομή.

Tests of Normality

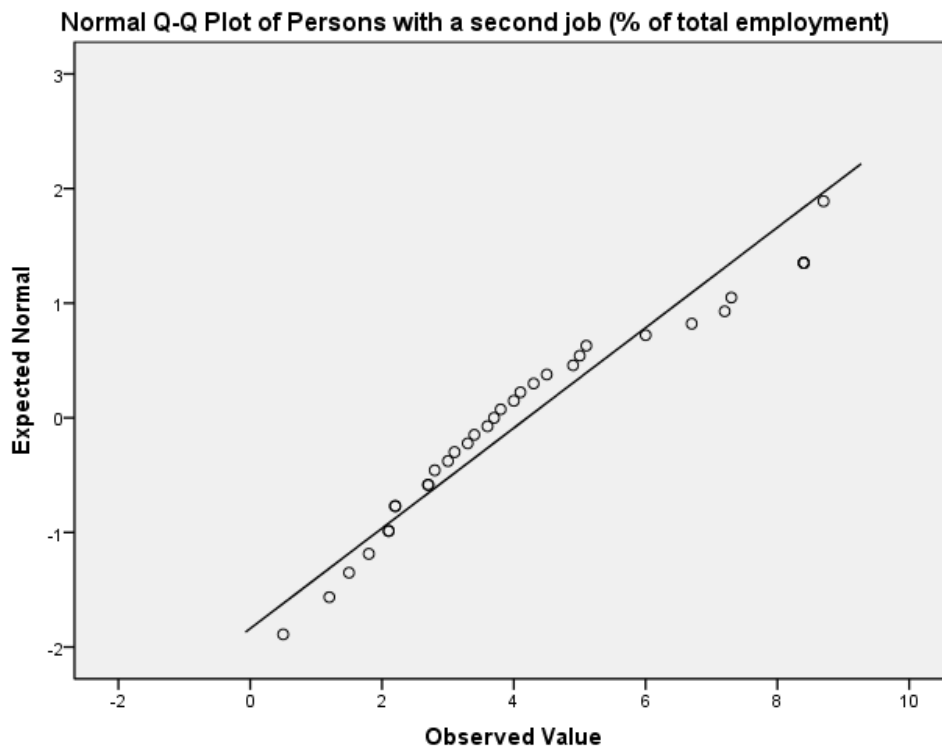
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Persons with a second job (% of total employment)	,124	33	,200 [*]	,928	33	,030

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Πίνακας 29 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “With_Second_Job”

Το παραπάνω αποτέλεσμα επιβεβαιώνεται και από τον γραφικό έλεγχο κανονικότητας. Στο διάγραμμα “Q-Q Plot” φαίνεται γενικά μία ικανοποιητική προσέγγιση των αναμενόμενων θεωρητικών κανονικών τιμών από τις αντίστοιχες πραγματικές, για το χαρακτηριστικό απασχόλησης που εξετάζεται.



Διάγραμμα 15 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “With_Second_Job”

2.10 Μεταβλητή “Part_Time_Male”

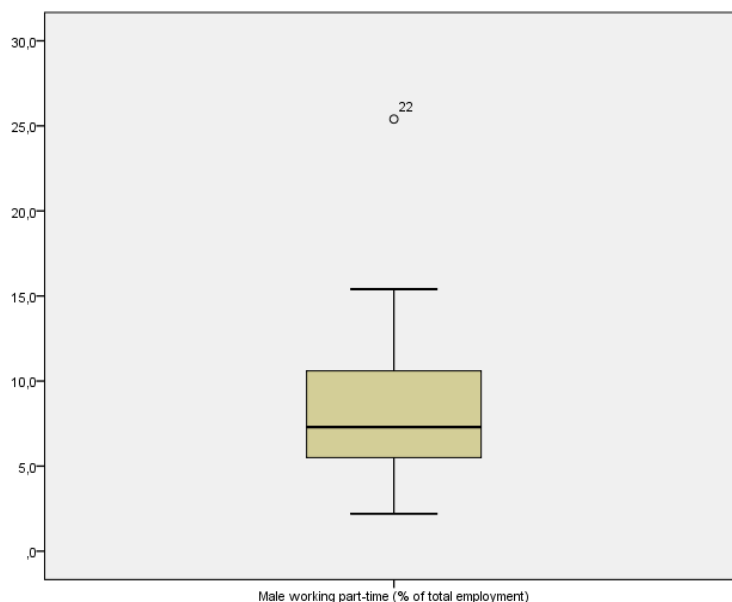
“Male working part-time (% of total employment)”: Εκφράζει το επί τοις εκατό ποσοστό μερικώς απασχολούμενων αντρών, ως προς το σύνολο των εργαζόμενων, για τις 33 χώρες της Ευρώπης. Έχει μέση τιμή 8.5 και τυπική απόκλιση 4.7. Η

χαμηλότερη τιμή είναι 2.2 (Βουλγαρία) και η υψηλότερη 25.4 (Ολλανδία). Τα υπόλοιπα από τα βασικά περιγραφικά μέτρα της μεταβλητής παρουσιάζονται στον πίνακα που ακολουθεί.

Descriptives			Statistic	Std. Error
Male working part-time (% of total employment)	Mean		8,530	,8246
	95% Confidence Interval for Mean	Lower Bound	6,851	
		Upper Bound	10,210	
	5% Trimmed Mean		8,151	
	Median		7,300	
	Variance		22,438	
	Std. Deviation		4,7369	
	Minimum		2,2	
	Maximum		25,4	
	Range		23,2	
	Interquartile Range		5,7	
	Skewness		1,507	,409
	Kurtosis		3,677	,798

Πίνακας 30 - Βασικά περιγραφικά μέτρα της μεταβλητής “Part_Time_Male”

Στο θηκόγραμμα που ακολουθεί σημειώνεται ως έκτροπη παρατήρηση το ποσοστό που αντιστοιχεί στην Ολλανδία (No 22).



Θηκόγραμμα 10 - Μεταβλητή “Part_Time_Male”

Βάσει του p-value του στατιστικού ελέγχου Kolmogorov-Smirnov, σε επίπεδο σημαντικότητας 5% δεν απορρίπτεται η μηδενική υπόθεση ότι η μεταβλητή ακολουθεί την κανονική κατανομή.

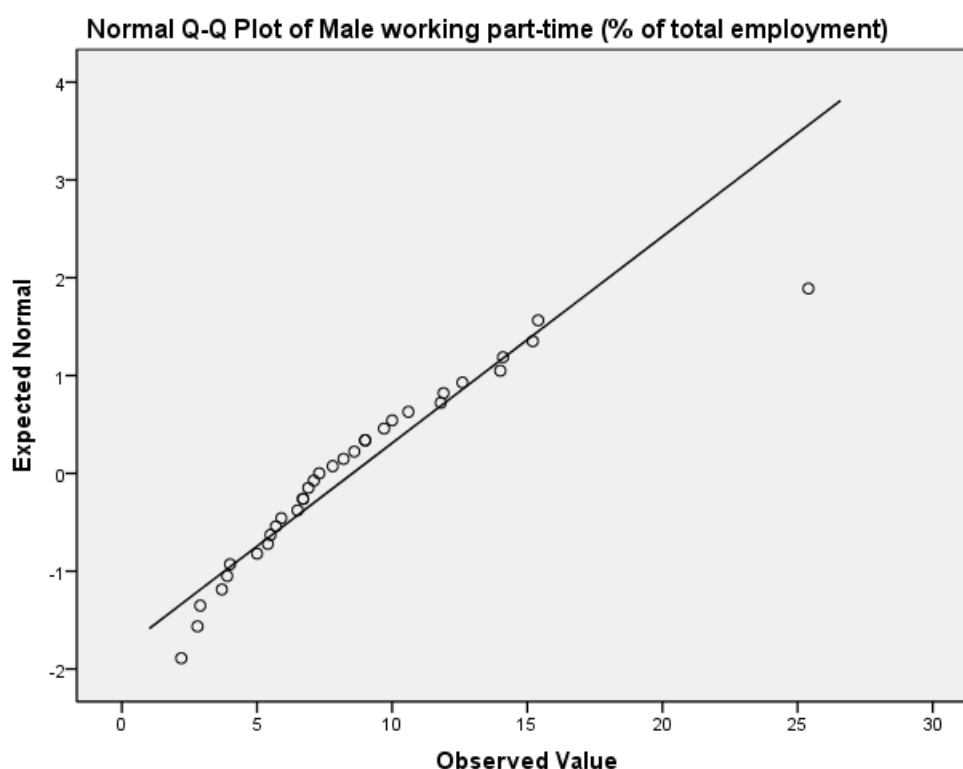
Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Male working part-time (% of total employment)	,127	33	,192	,892	33	,003

a. Lilliefors Significance Correction

Πίνακας 31- Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Part_Time_Male”

Το παραπάνω αποτέλεσμα επιβεβαιώνεται και από τον γραφικό έλεγχο κανονικότητας. Στο διάγραμμα “Q-Q Plot” με εξαίρεση την έκτροπη παρατήρηση για την Ολλανδία, φαίνεται γενικά μία ικανοποιητική προσέγγιση των αναμενόμενων θεωρητικών κανονικών τιμών από τις αντίστοιχες πραγματικές, για το χαρακτηριστικό απασχόλησης που εξετάζεται.



Διάγραμμα 16 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Part_Time_Male”

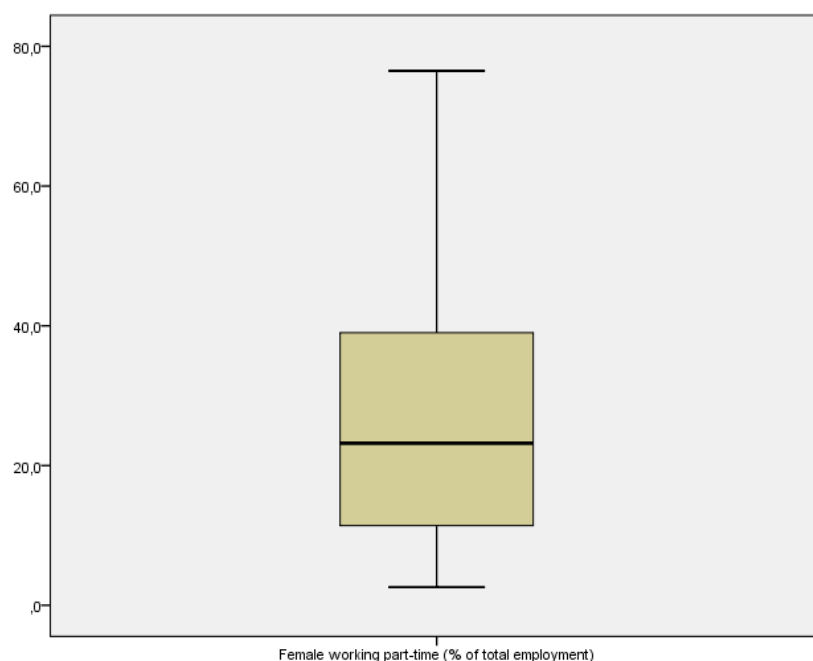
2.11 Μεταβλητή “Part_Time_Female”

“Female working part-time (% of total employment)”: Εκφράζει το επί τοις εκατό ποσοστό μερικώς απασχολούμενων γυναικών, ως προς το σύνολο των εργαζόμενων, για τις 33 χώρες της Ευρώπης. Έχει μέση τιμή 25.7 και τυπική απόκλιση 17.5. Η χαμηλότερη τιμή είναι 2.6 (Βουλγαρία) και η υψηλότερη 76.5

(Ολλανδία). Τα υπόλοιπα από τα βασικά περιγραφικά μέτρα της μεταβλητής παρουσιάζονται στον πίνακα που ακολουθεί.

Descriptives			Statistic	Std. Error
Female working part-time (% of total employment)	Mean		25,685	3,0451
	95% Confidence Interval for Mean	Lower Bound	19,482	
		Upper Bound	31,888	
	5% Trimmed Mean		24,431	
	Median		23,200	
	Variance		306,005	
	Std. Deviation		17,4930	
	Minimum		2,6	
	Maximum		76,5	
	Range		73,9	
	Interquartile Range		28,3	
	Skewness		,929	,409
	Kurtosis		,665	,798

Πίνακας 32 - Βασικά περιγραφικά μέτρα της μεταβλητής “Part_Time_Female”



Θηκόγραμμα 11 - Μεταβλητή “Part_Time_Female”

Βάσει του p-value του στατιστικού ελέγχου Kolmogorov-Smirnov, σε επίπεδο σημαντικότητας 5% απορρίπτεται η μηδενική υπόθεση ότι η μεταβλητή ακολουθεί την κανονική κατανομή.

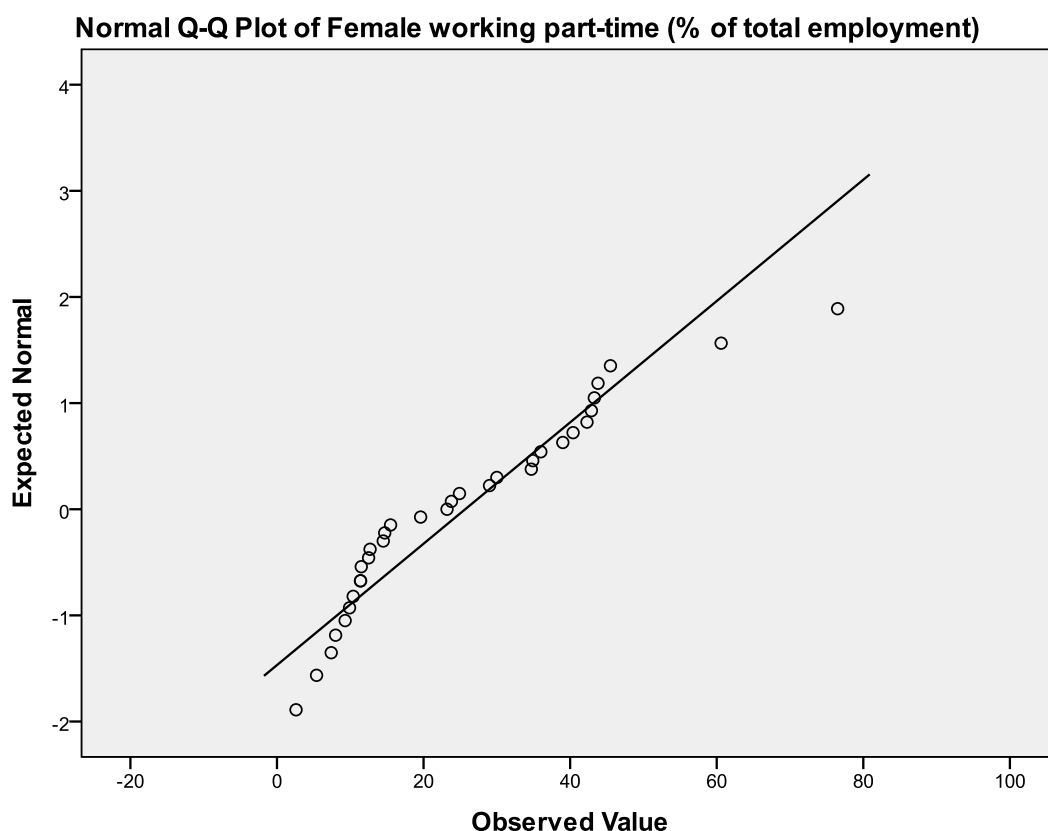
Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Female working part-time (% of total employment)	,174	33	,012	,909	33	,009

a. Lilliefors Significance Correction

Πίνακας 33 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή “Part_Time_Female”

Το παραπάνω αποτέλεσμα επιβεβαιώνεται και από τον γραφικό έλεγχο κανονικότητας. Στο διάγραμμα “Q-Q Plot” φαίνεται η μη ικανοποιητική προσέγγιση των αναμενόμενων θεωρητικών κανονικών τιμών από τις αντίστοιχες πραγματικές, για το χαρακτηριστικό απασχόλησης που εξετάζεται.



Διάγραμμα 17 - Γραφικός έλεγχος κανονικότητας της μεταβλητής “Part_Time_Female”

2.12 Μεταβλητή “Limited_Duration_Contract”

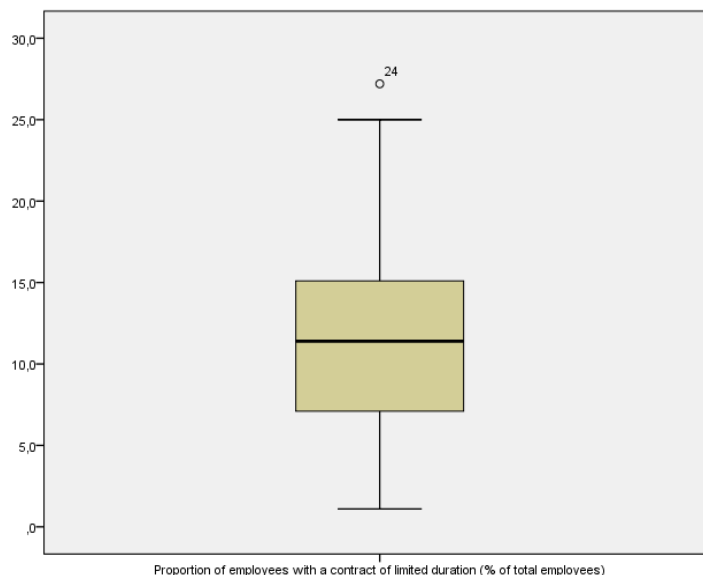
“Proportion of employees with a contract of limited duration (% of total employees)”: Εκφράζει το επί τοις εκατό ποσοστό εργαζόμενων με σύμβαση περιορισμένης διάρκειας, ως προς το σύνολο των εργαζόμενων, για τις 33 χώρες της

Ευρώπης. Έχει μέση τιμή 11.5 και τυπική απόκλιση 6.2. Η χαμηλότερη τιμή είναι 1.1 (Ρουμανία) και η υψηλότερη 27.2 (Πολωνία). Τα υπόλοιπα από τα βασικά περιγραφικά μέτρα της μεταβλητής παρουσιάζονται στον πίνακα που ακολουθεί.

Descriptives			Statistic	Std. Error
Proportion of employees with a contract of limited duration (% of total employees)	Mean		11,512	1,0769
	95% Confidence Interval for Mean	Lower Bound	9,319	
		Upper Bound	13,706	
	5% Trimmed Mean		11,239	
	Median		11,400	
	Variance		38,270	
	Std. Deviation		6,1863	
	Minimum		1,1	
	Maximum		27,2	
	Range		26,1	
	Interquartile Range		8,3	
	Skewness		,702	,409
	Kurtosis		,433	,798

Πίνακας 34 - Βασικά περιγραφικά μέτρα της μεταβλητής “Limited_Duration_Contract”

Στο θηκόγραμμα που ακολουθεί σημειώνεται ως έκτροπη παρατήρηση το ποσοστό που αντιστοιχεί στην Πολωνία (No 24).



Θηκόγραμμα 12 - Μεταβλητή “Limited_Duration_Contract”

Βάσει του p-value του στατιστικού ελέγχου Kolmogorov-Smirnov, σε επίπεδο σημαντικότητας 5% δεν απορρίπτεται η μηδενική υπόθεση ότι η μεταβλητή ακολουθεί την κανονική κατανομή.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Proportion of employees with a contract of limited duration (% of total employees)	,106	33	,200 [*]	,959	33	,242

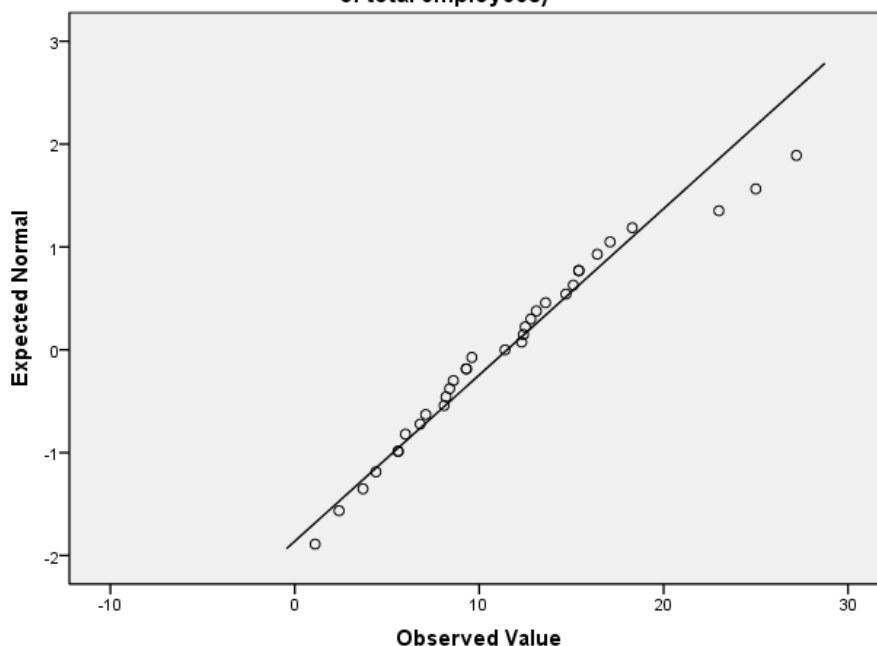
a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Πίνακας 35 - Στατιστικός έλεγχος κανονικότητας για τη μεταβλητή "Limited_Duration_Contract"

Το παραπάνω αποτέλεσμα επιβεβαιώνεται και από τον γραφικό έλεγχο κανονικότητας. Στο διάγραμμα "Q-Q Plot", με εξαίρεση τις 3 παρατηρήσεις με τα υψηλότερα ποσοστά, φαίνεται γενικά μία ικανοποιητική προσέγγιση των αναμενόμενων θεωρητικών κανονικών τιμών από τις αντίστοιχες πραγματικές, για το χαρακτηριστικό απασχόλησης που εξετάζεται.

Normal Q-Q Plot of Proportion of employees with a contract of limited duration (% of total employees)



Διάγραμμα 18 - Γραφικός έλεγχος κανονικότητας της μεταβλητής "Limited_Duration_Contract"

2.13 ΒΙΒΛΙΟΓΡΑΦΙΑ 2^{ου} ΚΕΦΑΛΑΙΟΥ

- [1] Barks, Robert A. (1972) *Descriptive Statistics and Probability Theory*, Hutchinson.
- [2] Holcomb, Zealure C. (1998) *Fundamentals of Descriptive Statistics*, Amazon.com.
- [3] Mann, Prem S. (1995) *Introductory Statistics, 2nd Edition*, Wiley.
- [4] Rice, John A. (2007) *Mathematical Statistics and Data Analysis*, Cengage Learning.

ΚΕΦΑΛΑΙΟ 3 – ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

3.1 Πολυμεταβλητή Ανάλυση

Η Πολυμεταβλητή Ανάλυση ασχολείται με στατιστικές μεθόδους συλλογής, περιγραφής και ανάλυσης δεδομένων που αποτελούνται από μετρήσεις πολλών μεταβλητών σε ένα πλήθος ατόμων ή γενικότερα πειραματικών μονάδων. Για παράδειγμα, τα δεδομένα μπορεί να είναι μετρήσεις του ύψους και του βάρους κάποιων ατόμων. Στην περίπτωση αυτή, οι μεταβλητές είναι δύο: το ύψος και το βάρος των ατόμων.


Γενικά εάν υπάρχουν p υπό μελέτη μεταβλητές και αυτές μετρούνται σε n πειραματικές μονάδες, τα προς ανάλυση δεδομένα μπορούν να παρασταθούν υπό την μορφή ενός $n \times p$ πίνακα $X = (x_{i,j})$, όπου $x_{i,j}$ η τιμή της j μεταβλητής στην i πειραματική μονάδα. Όλες μαζί οι μετρήσεις των p μεταβλητών σε κάθε συγκεκριμένη πειραματική μονάδα (δηλαδή κάθε γραμμή του πίνακα X) είναι μία πολυμεταβλητή παρατήρηση. Έτσι, εδώ, τα προς ανάλυση δεδομένα αποτελούνται από n πολυμεταβλητές παρατηρήσεις.

Περαιτέρω, για να εξαχθούν γενικά συμπεράσματα από τα διαθέσιμα δεδομένα, είναι απαραίτητο κάθε μεταβλητή να θεωρηθεί ως τυχαία μεταβλητή (με την πιθανοθεωρητική έννοια του όρου), έστω, X_i , $i = 1, 2, \dots, p$. Το διάνυσμα με συνιστώσες τις τυχαίες μεταβλητές $X_i, Y = \{X_1, X_2, \dots, X_p\}$ λέγεται τυχαίο διάνυσμα. Συνεπώς οι πολυμεταβλητές παρατηρήσεις είναι παρατηρηθείσες τιμές τυχαίων διανυσμάτων και προέρχονται από πολυμεταβλητούς στατιστικούς πληθυσμούς.

Ένα μέρος του γνωστικού αντικείμενου της πολυμεταβλητής στατιστικής είναι ανάλογο αυτού της μονομεταβλητής στατιστικής. Μία, όμως, πολύ σημαντική πλευρά της πολυμεταβλητής στατιστικής (που δεν υφίσταται στην μονομεταβλητή στατιστική) είναι η αναζήτηση και ανάλυση της εξάρτησης μεταξύ των μεταβλητών ή μεταξύ ομάδων μεταβλητών, καθώς και η αναζήτηση «κύριων» μεταβλητών. Για αυτά και άλλα προβλήματα νέα μεθοδολογία έχει αναπτυχθεί.

3.1.1 Αλγεβρική παρουσίαση του προβλήματος

$X_{(n \times p)} = (x_{ij})$, όπου x_{ij} η τιμή της j μεταβλητής στην i πειραματική μονάδα.

	1	2	...	j	...	p
1	x_{11}	x_{12}		x_{1j}		x_{1p}
2	x_{21}	x_{22}		x_{2j}		x_{2p}
⋮						
i	x_{i1}	x_{i2}		x_{ij}		x_{ip}  μία πολυμεταβλητή παρατήρηση
⋮						
n	x_{n1}	x_{n2}		x_{nj}		x_{np}
	\bar{x}_1	\bar{x}_2	...	\bar{x}_j	...	\bar{x}_p

Όπου: $1, 2, \dots, i, \dots, n$: πειραματικές μονάδες και $1, 2, \dots, j, \dots, p$: μεταβλητές.

Ή σε μορφή πινάκων:

$$X = \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix}$$

3.1.2 Συνοπτικοί δείκτες θέσης και μεταβλητότητας

1. Δειγματικός μέσος:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{pmatrix}$$

2. Πίνακας δειγματικής συνδιασποράς και πίνακας δειγματικής συσχέτισης

$$S = (s_{jk})$$

$$s_{jj} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$$

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

$$S = \frac{1}{n-1} \sum_{i=1}^n (\underline{X}_i - \bar{\underline{X}})(\underline{X}_i - \bar{\underline{X}})'$$

$$R = (r_{jk})$$

$$r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}s_{kk}}}$$

Οι πίνακες S και R περιέχουν όλες τις πληροφορίες αναφορικά με την μεταβλητότητα σε κάθε μεταβλητή ξεχωριστά και τις συσχετίσεις μεταξύ των μεταβλητών.

3. Ολική δειγματική διασπορά (Total Sample Variance)

$$TSV = s_{11} + s_{22} + \dots + s_{pp} (= \text{tr}S)$$

α. Εκφράζει τη μεταβλητότητα σε όλες μαζί τις μεταβλητές.

β. Δεν περιέχει πληροφορίες σχετικά με τις συσχετίσεις μεταξύ μεταβλητών.

γ. Αποτελεί μία γενίκευση της έννοιας της δειγματικής διασποράς (για $p=1$ συμπίπτει με αυτήν) και εκτιμά την ολική πληθυσμιακή διασπορά $\text{tr}\Sigma = s_{11} + s_{22} + \dots + s_{pp}$

4. Γενικευμένη δειγματική διασπορά (Sample Variance)

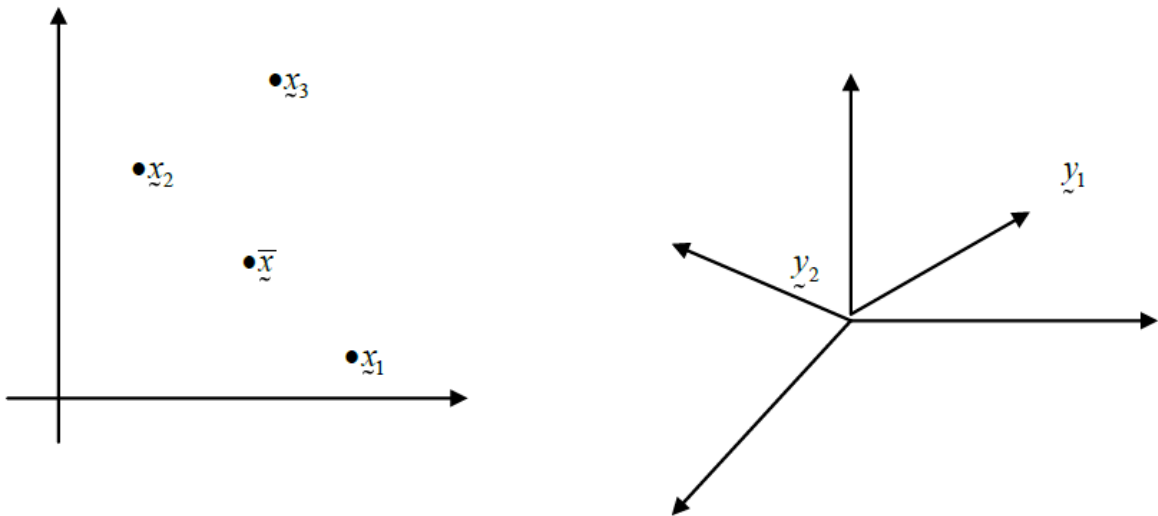
$$GSV = |S|$$

Αποτελεί και αυτή γενίκευση της έννοιας της δειγματικής διασποράς (για $p=1$ συμπίπτει με αυτήν). Όπως θα δούμε αποτελεί ένα δείκτη του πόσο διασκορπισμένες είναι οι παρατηρήσεις x_1, x_2, \dots, x_n γύρω από το διάνυσμα \bar{x} , εάν θεωρηθούν ως n σημεία στον ευκλείδειο χώρο διάστασης $p(\mathbb{R}^p)$.

3.1.3 Γραφική παρουσίαση

1. Α τρόπος: τα δεδομένα παριστάνονται ως n σημεία στο χώρο \mathbb{R}^p . Άξονες είναι οι p μεταβλητές.
2. Β τρόπος: τα δεδομένα παριστάνονται ως p διανύσματα στο χώρο \mathbb{R}^n . Άξονες είναι οι n πειραματικές μονάδες.

π.χ $p=2, n=3$.



Εδώ $X_{3 \times 2} = (x_{ij})$, οι γραμμές του X είναι x'_1, x'_2, x'_3 και οι στήλες του y_1, y_2 . Φυσικά οι γραφικές αυτές παραστάσεις είναι δύσκολο έως αδύνατο να γίνουν για $p>3$ (Α τρόπος) και $n>3$ (Β τρόπος). Εναλλακτική λύση είναι να γίνει το διάγραμμα σκέδασης (scatter plot) για όλα τα ζεύγη των μεταβλητών.

3.1.4 Γεωμετρική ερμηνεία της γενικευμένης δειγματικής διασποράς (GSV)

Θεωρούμε το ελλειψοειδές στον \mathbb{R}^p με κέντρο \bar{x} που ορίζεται από τη σχέση:

$$(\underline{x} - \bar{x})' S^{-1} (\underline{x} - \bar{x}) = c^2$$

$$E = \left\{ \underline{x} \in \mathbb{R}^p : (\underline{x} - \bar{x})' S^{-1} (\underline{x} - \bar{x}) \leq c^2 \right\}$$

Μικρή τιμή της GSV, $|S|$, αντιστοιχεί σε μικρή τιμή του όγκου του E.

Εάν τώρα πάρουμε, $c^2 = \max(x_i - \bar{x})S^{-1}(x_i - \bar{x})$, τότε προφανώς $x_i \in E, \forall i = 1, 2, \dots, n$ δηλαδή όλες οι παρατηρήσεις θεωρούμενες ως σημεία είναι (μέσα) στο ελλειψοειδές E. Συνεπώς, όσο πιο μικρή τιμή έχει η $|S|$, τόσο πιο μικρό όγκο έχει το E, άρα τόσο «πιο κοντά» οι παρατηρήσεις x_i είναι στο δειγματικό μέσο \bar{x} .

Συμπέρασμα:

Η GSV, $|S|$, είναι ένας δείκτης συγκεντρωτικότητας των παρατηρήσεων γύρω από τη μέση τιμή τους. Η ερμηνεία αυτή είναι ανάλογη της ερμηνείας που έχει η δειγματική διασπορά στην μονομεταβλητή στατιστική.

3.2 Ανάλυση σε κύριες συνιστώσες

3.2.1 Εισαγωγή

Η ανάλυση πολυμεταβλητών δεδομένων καθίσταται ιδιαίτερα δύσκολη όταν το πλήθος των μεταβλητών p είναι μεγάλο. Το p αναφέρεται ως η διάσταση των δεδομένων. Ένας τρόπος γραφικής παρουσίασης πολυμεταβλητών δεδομένων είναι ως n σημεία στον \mathbb{R}^n χώρο ($n =$ μέγεθος του δείγματος). Έτσι για $p \geq 4$ είναι αδύνατο να έχουμε έστω και μία οπτική ιδέα, μια εικόνα των δεδομένων. Επίσης δυσκολία υπάρχει στην ανάλυση όταν οι μεταβλητές είναι υψηλά συσχετισμένες μεταξύ τους.

Η ανάλυση κύριων συνιστωσών είναι μια τεχνική που έχει ως στόχο την μείωση της διάστασης των δεδομένων διατηρώντας συγχρόνως ένα ποσοστό της ολικής μεταβλητότητας $\sum_{i=1}^p Var(X_i)$ των αρχικών μεταβλητών X_1, X_2, \dots, X_p . Συγκεκριμένα, επιδιώκεται με την τεχνική αυτή η ανεύρεση λιγότερων μεταβλητών $Y_1, Y_2, \dots, Y_k, k < p$, που είναι γραμμικοί συνδυασμοί των αρχικών X_1, X_2, \dots, X_p , ασυσχέτιστες μεταξύ τους και έχουν ολική μεταβλητότητα $\sum_{i=1}^k Var(Y_i)$ μικρότερη από αυτή των αρχικών. Δηλαδή: $\sum_{i=1}^p Var(X_i) < \sum_{i=1}^k Var(Y_i)$. Οι μεταβλητές $Y_1, Y_2, \dots, Y_k, k < p$ λέγονται κύριες συνιστώσες. Οι κύριες συνιστώσες μπορούν στην συνέχεια να αντικαταστήσουν τις αρχικές μεταβλητές και το αρχικό σύνολο δεδομένων - p μετρήσεις σε καθένα από τα n άτομα - να μειωθεί σε σύνολο δεδομένων με k μετρήσεις σε καθένα από τα n άτομα, όπου οι μετρήσεις τώρα γίνονται επί των κύριων συνιστωσών. Με αυτό τον τρόπο η αρχική διάσταση p μειώνεται σε k και αντί να αναλύσουμε δεδομένα στον \mathbb{R}^p , αναλύουμε δεδομένα στον \mathbb{R}^k . Σε ορισμένες περιπτώσεις το k , η νέα διάσταση, είναι 2 ή 3 και αυτό αποτελεί σημαντικό όφελος της τεχνικής. Η ανάλυση σε κύριες συνιστώσες προσπαθεί να ερμηνεύσει τη δομή διασποράς - συνδιασποράς, χρησιμοποιώντας κάποιους γραμμικούς συνδυασμούς των αρχικών μεταβλητών με γενικό στόχο τη μείωση των δεδομένων και την ερμηνεία τους. Μια ανάλυση σε κύριες συνιστώσες συχνά αποκαλύπτει σχέσεις οι οποίες δεν φαίνονταν προηγουμένως.

Κλείνοντας την εισαγωγή, σημειώνεται ότι η τεχνική κύριων συνιστωσών δεν επιτυγχάνει πάντοτε την μείωση της διάστασης (π.χ αυτό συμβαίνει όταν οι αρχικές μεταβλητές είναι ασυσχέτιστες μεταξύ τους). Τότε άλλες τεχνικές πρέπει να αναζητηθούν.

3.2.2 Κύριες συνιστώσες από τον πίνακα συνδιασποράς Σ

Αλγεβρικά, οι κύριες συνιστώσες είναι συγκεκριμένοι γραμμικοί συνδυασμοί των p τυχαίων μεταβλητών X_1, X_2, \dots, X_p . Γεωμετρικά, αυτοί οι γραμμικοί συνδυασμοί παριστάνουν την επιλογή ενός νέου συστήματος συντεταγμένων που δίνεται με την περιστροφή του αρχικού συστήματος που έχει τις μεταβλητές X_1, X_2, \dots, X_p ως άξονες συντεταγμένων. Οι νέοι άξονες παριστάνουν τις κατευθύνσεις με τη μέγιστη διασπορά και δίνουν μια πιο απλή περιγραφή της δομής διασποράς-συνδιασποράς των δεδομένων.

Όπως θα δούμε οι κύριες συνιστώσες εξαρτώνται από τον πίνακα συνδιασπορών Σ (ή των πίνακα συσχέτισεων ρ) των μεταβλητών X_1, X_2, \dots, X_p . Η ανάπτυξή τους δεν απαιτεί την πολυδιάστατη κανονική υπόθεση. Όμως οι κύριες συνιστώσες που εξάγονται από πληθυσμούς με πολυδιάστατη κανονική κατανομή έχουν χρήσιμες γεωμετρικές ερμηνείες.

Έστω ότι το τυχαίο διάνυσμα $X'=[X_1, X_2, \dots, X_p]$ έχει πίνακα συνδιασπορών Σ με ιδιοτιμές $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Θεωρούμε τους γραμμικούς συνδυασμούς:

$$\begin{aligned} Y_1 &= l_1'X = l_{11}X_1 + l_{21}X_2 + \dots + l_{p1}X_p \\ Y_2 &= l_2'X = l_{12}X_1 + l_{22}X_2 + \dots + l_{p2}X_p \\ &\vdots \\ Y_p &= l_p'X = l_{1p}X_1 + l_{2p}X_2 + \dots + l_{pp}X_p \end{aligned} \tag{1}$$

Τότε:

$$Var(Y_i) = l_i'\Sigma l_i, \quad i = 1, 2, \dots, p \tag{2}$$

$$Cov(Y_i, Y_k) = l_i'\Sigma l_k, \quad i, k = 1, 2, \dots, p \tag{3}$$

Οι κύριες συνιστώσες είναι εκείνοι οι γραμμικοί συνδυασμοί των ασυσχέιστων τυχαίων μεταβλητών Y_1, Y_2, \dots, Y_p των οποίων η διασπορά στη σχέση (2) είναι όσο το δυνατόν μεγαλύτερη. Η πρώτη κύρια συνιστώσα είναι εκείνος ο γραμμικός συνδυασμός

με τη μεγαλύτερη διασπορά. Έτσι έχουμε μέγιστη διασπορά $Var(Y_1) = l_1' \Sigma l_1$. Είναι προφανές ότι η τιμή αυτή μπορεί να μεγαλώσει αν πολλαπλασιαστεί με κάποια σταθερά.

Προκειμένου να να αποφευχθεί αυτό, ορίζουμε:

- Πρώτη Κύρια Συνιστώσα = γραμμικός συνδυασμός $l_1'X$ που μεγιστοποιεί την $Var(l_1'X)$, υπό τον περιορισμό $l_1'l_1 = 1$.
- Δεύτερη Κύρια Συνιστώσα = γραμμικός συνδυασμός $l_2'X$ που μεγιστοποιεί την $Var(l_2'X)$, υπό τον περιορισμό $l_2'l_2 = 1$ και $Cov(l_1'X, l_2'X) = 0$.
- i -οστή Κύρια Συνιστώσα = γραμμικός συνδυασμός $l_i'X$ που μεγιστοποιεί την $Var(l_i'X)$, υπό τον περιορισμό $l_i'l_i = 1$ και $Cov(l_i'X, l_k'X) = 0$, για $k < i$.

Αποτέλεσμα 1:

Έστω Σ ο πίνακας συνδιασπορών του τυχαίου διανύσματος $X' = [X_1, X_2, \dots, X_p]$.

Έστω επίσης ότι ο πίνακας Σ έχει ζεύγη ιδιοτιμών-ιδιοδιανυσμάτων $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ με $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Τότε υπάρχει $p \times p$ ορθογώνιος πίνακας $Q = (e_1, e_2, \dots, e_p)$ τέτοιος ώστε:

$$Q' \Sigma Q = \Lambda = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p \end{pmatrix}$$

Από την σχέση διαγωνοποίησης συμπεραίνουμε ότι τα διανύσματα $e_i \in \mathcal{R}^p$ είναι τα ορθογώνια ιδιοδιανύσματα του πίνακα Σ με μοναδιαίο μήκος που αντιστοιχούν στις ιδιοτιμές λ_i .

Θέτουμε $Y = (Y_1, Y_2, \dots, Y_p)' = Q'X$. Τότε:

$Cov(Y) = Cov(Q'X) = Q' Cov(X) Q = Q' \Sigma Q = \Lambda$ και οι τυχαίες μεταβλητές $Y_1 = e_1'X, Y_2 = e_2'X, \dots, Y_p = e_p'X$ είναι ασυσχέτιστες γραμμικές συναρτήσεις των αρχικών μεταβλητών X_1, X_2, \dots, X_p .

Επομένως έχουμε:

$$Y_i = e_i'X = e_{1i}X_1 + e_{2i}X_2 + \dots + e_{pi}X_p, \quad i = 1, 2, \dots, p$$

$$Var(Y_i) = e_i' \Sigma e_i = \lambda_i, \quad i = 1, 2, \dots, p$$

$$Cov(Y_i, Y_k) = e_i' \Sigma e_k = 0, \quad i \neq k$$

Φαίνεται ότι οι κύριες συνιστώσες είναι ασυσχέτιστες μεταξύ τους και έχουν διασπορές ίσες με τις ιδιοτιμές του πίνακα Σ .

Άρα, όπως ορίσαμε τις κύριες συνιστώσες και χρησιμοποιώντας τα αποτελέσματα της πρότασης 1 παρατηρούμε ότι οι κύριες συνιστώσες έχουν την εξής βέλτιστη ιδιότητα: Η πρώτη κύρια συνιστώσα $Y_1 = e_1'X$ έχει την μεγαλύτερη διασπορά λ_1 , μεταξύ όλων των γραμμικών συνδυασμών της μορφής $l'X$ με $l'l = 1$. Στη συνέχεια, μεταξύ όλων των γραμμικών συνδυασμών $l'X$ με $l'l = 1$ που είναι ασυσχέτιστοι με την Y_1 , η κύρια συνιστώσα $Y_2 = e_2'X$ έχει την μεγαλύτερη διασπορά,

λ_2 . Γενικά μεταξύ όλων των γραμμικών συνδυασμών $l'X$ με $l'l = 1$ που είναι ασυσχέτιστοι με τις Y_1, Y_2, \dots, Y_{m-1} , η κύρια συνιστώσα $Y_m = e_m'X$ έχει την μεγαλύτερη διασπορά λ_m .

Επίσης, σημειώνουμε ότι η γραμμική συνάρτηση $l'X$ έχει διασπορά $Var(l'X) = l'\Sigma l$ και η συνθήκη να είναι η $l'X$ ασυσχέτιστη με την i -κύρια συνιστώσα $Y_i = e_i'X$ σημαίνει: $0 = Cov(l'X, e_i'X) = l'Cov(X)e_i = l'\Sigma e_i = \lambda_i l'e_i$, επειδή $\Sigma e_i = \lambda_i e_i$. Επομένως πρέπει $l'e_i = 0$.

Η προαναφερθείσα ιδιότητα των κύριων συνιστωσών είναι άμεση συνέπεια της επόμενης πρότασης. Η ιδιότητα αυτή μπορεί να χρησιμοποιηθεί και σαν άλλος τρόπος ορισμού των κύριων συνιστωσών και έχει ευνότητα, από στατιστικής πλευράς, ερμηνεία αναφορικά με την μέγιστη δυνατή μείωση της διάστασης των δεδομένων.

Πρόταση:

Έστω $Q = (e_1, \dots, e_p)$, $p \times p$ ορθογώνιος πίνακας τέτοιος ώστε:

$$Q'\Sigma Q = \Lambda = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p \end{pmatrix}, \text{όπου } \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0. \text{ Τότε έχουμε:}$$

$$\max\{l'\Sigma l : l'l = 1 \text{ και } l'e_i = 0, i = 1, \dots, m-1\} = \lambda_m, \forall m = 1, 2, \dots, p.$$

Το maximum επιτυγχάνεται για $l = e_m$.

Απόδειξη

Έστω $l \in \mathbb{R}^p, l'l = 1$. Θέτουμε $k = Q'l = (k_1 \dots k_p)'$. Οπότε έχουμε:

$$k'k = l'QQ'l = l'l = 1$$

$$\text{και } l'\Sigma l = (Qk)' \Sigma (Qk) = k'Q'\Sigma Qk = k'\Lambda k = \sum_{i=1}^p \lambda_i k_i^2$$

$$\text{Συνεπώς: } l'\Sigma l = \sum_{i=1}^p \lambda_i k_i^2 \leq \lambda_1 (k_1^2 + k_2^2 + \dots + k_p^2) = \lambda_1 k'k = \lambda_1$$

Και η ισότητα επιτυγχάνεται για $\underline{k} = (1, 0, \dots, 0)'$ δηλαδή για $\underline{l} = Q\underline{k} = Q \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \underline{e}_1$. Άρα

$$\lambda_1 = \max\{\underline{l}'\Sigma\underline{l} : \underline{l}'\underline{l} = 1\} = \underline{e}_1'\Sigma\underline{e}_1.$$

Στη συνέχεια η συνθήκη $\underline{l}'\underline{e}_1 = 0$ είναι ισοδύναμη με $\underline{k}'Q\underline{e}_1 = 0$ ή με $k_1 = 0$ γιατί $\underline{k}'Q\underline{e}_1 = k_1$. Επομένως, όταν $\underline{l}'\underline{e}_1 = 0$ και $\underline{l}'\underline{l} = \underline{k}'\underline{k} = 1$ έχουμε

$$\underline{l}'\Sigma\underline{l} = \underline{k}'\Lambda\underline{k} = \sum_{i=1}^p \lambda_i k_i^2 \leq \lambda_2 (k_2^2 + \dots + k_p^2) = \lambda_2 \quad \text{και η ισότητα επιτυγχάνεται όταν}$$

$\underline{k} = (0, 1, \dots, 0)'$ δηλαδή όταν $\underline{l} = Q\underline{k} = Q \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \underline{e}_2$. Άρα

$\lambda_2 = \max\{\underline{l}'\Sigma\underline{l} : \underline{l}'\underline{l} = 1, \underline{l}'\underline{e}_1 = 0\} = \underline{e}_2'\Sigma\underline{e}_2$. Η υπόλοιπη απόδειξη γίνεται ακολουθώντας την ίδια ακριβώς μέθοδο.

Αποτέλεσμα 2:

Έστω ότι το τυχαίο διάνυσμα $X' = [X_1, X_2, \dots, X_p]$ έχει πίνακα συνδιασπορών Σ με ζεύγη ιδιοτιμών-ιδιοδιανυσμάτων $(\lambda_1, \underline{e}_1), (\lambda_2, \underline{e}_2), \dots, (\lambda_p, \underline{e}_p)$ με $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Έστω επίσης $Y_1 = \underline{e}_1'X, Y_2 = \underline{e}_2'X, \dots, Y_p = \underline{e}_p'X$ οι κύριες συνιστώσες.

Τότε ισχύει:

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^p \text{Var}(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p \text{Var}(Y_i)$$

Απόδειξη:

Γενικά ισχύει ότι: $\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \text{tr}(\Sigma)$.

Όπως έχουμε δει όμως, ο πίνακας Σ μπορεί να γραφεί και ως: $\Sigma = Q\Lambda Q'$ όπου Λ ο διαγώνιος πίνακας ιδιοτιμών ενώ Q είναι ο πίνακας με στήλες τα ιδιοδιανύσματα του Σ δηλαδή $Q = [\underline{e}_1, \underline{e}_2, \dots, \underline{e}_p]$ ώστε: $QQ' = Q'Q = I$.

Τότε:

$$\text{tr}(\Sigma) = \text{tr}(Q\Lambda Q') = \text{tr}(\Lambda Q Q') = \text{tr}(\Lambda) = \lambda_1 + \lambda_2 + \dots + \lambda_p$$

Δηλαδή:

$$\sum_{i=1}^p \text{Var}(X_i) = \text{tr}(\Sigma) = \text{tr}(\Lambda) = \sum_{i=1}^p \text{Var}(Y_i)$$

Άρα παρατηρούμε ότι ισχύει το εξής:

$$\text{Συνολική Διασπορά Πληθυσμού} = \sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \lambda_1 + \lambda_2 + \dots + \lambda_p$$

Κατά συνέπεια το ποσοστό της συνολικής διασποράς που ερμηνεύει η k κύρια συνιστώσα θα είναι:

ποσοστό της συνολικής

$$\text{διασποράς που ερμηνεύει} = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p}, k = 1, 2, \dots, p$$

η k κύρια συνιστώσα

Εάν το μεγαλύτερο ποσοστό (π.χ 80-90%) της συνολικής διασποράς του πληθυσμού, για μεγάλο p , μπορεί να ερμηνευθεί από τις πρώτες μία, δύο ή και τρεις κύριες συνιστώσες τότε αυτές οι συνιστώσες μπορούν να «αντικαταστήσουν» τις αρχικές p μεταβλητές χωρίς να χαθεί αρκετό μέρος πληροφορίας.

Αποτέλεσμα 3:

Εάν $Y_1 = e_1'X, Y_2 = e_2'X, \dots, Y_p = e_p'X$ είναι οι κύριες συνιστώσες που προέρχονται από τον πίνακα συνδιασπορών Σ , τότε:

$$\rho_{Y_i, X_k} = \frac{e_{ik} \sqrt{\lambda_i}}{\sqrt{\sigma_{ik}}} \quad i, k = 1, 2, \dots, p$$

όπου $Q = (e_{ij})$ και $\sigma_{ii} = \text{Var}(X_i)$.

Η παραπάνω σχέση παρέχει τη συσχέτιση ανάμεσα στην κύρια συνιστώσα Y_i και στην μεταβλητή X_k .

Απόδειξη:

Έστω $l'_k = [0, \dots, 0, 1, 0, \dots, 0]$ έτσι ώστε να ισχύει $X_k = l'_k X$ και

$Cov(X_k, Y_i) = Cov(l'_k X, e'_i X) = l'_k \Sigma e_i$. Επειδή $\Sigma e_i = \lambda_i e_i$, $Cov(X_k, Y_i) = l'_k \lambda_i e_i = \lambda_i e_{ki}$ και

$Var(Y_i) = \lambda_i$, $Var(X_k) = \sigma_{kk}$ έχουμε:

$$\rho_{Y_i, X_k} = \frac{Cov(Y_i, X_k)}{\sqrt{Var(Y_i)} \sqrt{Var(X_k)}} = \frac{\lambda_i e_{ki}}{\sqrt{\lambda_i} \sqrt{\sigma_{kk}}} = \frac{e_{ki} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \quad i, k = 1, 2, \dots, p$$

Παρατήρηση:

Είναι $Y_j = \sum_{i=1}^p e_{ij} X_i$ και ο συντελεστής συσχέτισης του X_i στη γραμμικό συνδυασμό που εκφράζει την Y_j είναι e_{ij} . Άρα η συσχέτιση των Y_j και X_i έχει το ίδιο πρόσημο όπως ο συντελεστής e_{ij} και επιπλέον είναι ανάλογη αυτού (αφού είναι $\frac{e_{ij} \sqrt{\lambda_j}}{\sqrt{\sigma_{ii}}}$). Αυτή η παρατήρηση παρέχει στατιστική ερμηνεία στους τελεστές e_{ij} των γραμμικών συνδυασμών $Y_j = \sum_{i=1}^p e_{ij} X_i$.

3.2.3 Γεωμετρική ερμηνεία κυρίων συνιστωσών

Η κατασκευή των κύριων συνιστωσών έχει να κάνει μόνο με τον πίνακα συνδιασποράς Σ του διανύσματος X , ($Cov(X) = \Sigma$) και όχι με την κατανομή του. Θα υποθέσουμε όμως τώρα ότι οι κύριες συνιστώσες προέρχονται από μεταβλητές που ακολουθούν πολυδιάστατη κανονική κατανομή. Έτσι θεωρούμε ότι το διάνυσμα X ακολουθεί $N_p(\mu, \Sigma)$. Τότε τα ελλειψοειδή σταθερής πυκνότητας είναι σύνολα σημείων x στο χώρο \mathbb{R}^p που ικανοποιούν:

$$(x-\mu)' \Sigma^{-1} (x-\mu) = c^2$$

Αυτά έχουν κέντρο μ και άξονες $\pm c \sqrt{\lambda_i} e_i$, $i=1, 2, \dots, p$ όπου (λ_i, e_i) είναι οι ιδιοτιμές και τα ιδιοδιανύσματα του πίνακα Σ . Ένα σημείο που βρίσκεται στον i άξονα του ελλειψοειδούς θα έχει συντεταγμένες ανάλογες προς το διάνυσμα $e'_i = [e_{1i}, e_{2i}, \dots, e_{pi}]$ του

συστήματος συντεταγμένων του οποίου οι άξονες είναι παράλληλοι προς τους αρχικούς άξονες x_1, x_2, \dots, x_p . Είναι βολικό να θέσουμε $\mu=0$ χωρίς βλάβη της γενικότητας αφού το

τυχαίο διάνυσμα X μπορεί να γραφεί ως το κανονικό τυχαίο διάνυσμα $W=X-\mu$, όπου $E(W)=0$ και ταυτόχρονα $Cov(X)=Cov(W)$.

Έτσι μπορούμε να γράψουμε:

$$c^2 = x' \Sigma^{-1} x = \frac{1}{\lambda_1} (e_1' x)^2 + \frac{1}{\lambda_2} (e_2' x)^2 + \dots + \frac{1}{\lambda_p} (e_p' x)^2$$

όπου $e_1' x, e_2' x, \dots, e_p' x$ είναι οι κύριες συνιστώσες του διανύσματος x . Θέτοντας

$$y_1 = e_1' x, y_2 = e_2' x, \dots, y_p = e_p' x \text{ έχουμε}$$

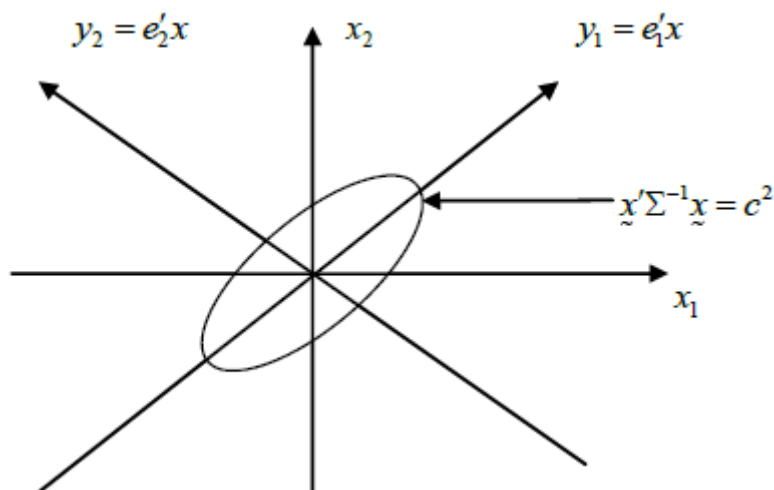
$$c^2 = \frac{1}{\lambda_1} y_1^2 + \frac{1}{\lambda_2} y_2^2 + \dots + \frac{1}{\lambda_p} y_p^2$$

και αυτή εξίσωση παριστάνει την επιφάνεια ελλειψοειδούς (αφού $\lambda_1, \lambda_2, \dots, \lambda_p$ θετικοί αριθμοί) με κέντρο 0 σε ένα σύστημα συντεταγμένων με τους άξονες y_1, y_2, \dots, y_p να βρίσκονται αντίστοιχα στις κατευθύνσεις των διανυσμάτων e_1, e_2, \dots, e_p . Εάν λ_1 είναι η μεγαλύτερη ιδιοτιμή, τότε ο κύριος άξονας βρίσκεται στην κατεύθυνση του e_1 . Οι υπόλοιποι άξονες βρίσκονται στις κατευθύνσεις που καθορίζονται από τα διανύσματα e_2, \dots, e_p .

Επιπλέον εάν $x \in \mathbb{R}^p$, τότε η προβολή του στο e_i είναι $\frac{e_i' x}{e_i' e_i} = y_i e_i$, δηλαδή το x έχει συντεταγμένες (y_1, y_2, \dots, y_p) στο σύστημα των αξόνων που καθορίζουν τα e_1, e_2, \dots, e_p και οι συντεταγμένες αυτές, $y_i = e_i' x$, συμπίπτουν με τις τιμές των κύριων συνιστωσών για το x

Επομένως οι κύριες συνιστώσες του x , $y_1 = e_1' x, y_2 = e_2' x, \dots, y_p = e_p' x$, είναι οι συντεταγμένες του x σε ένα σύστημα ορθογώνιων αξόνων που προκύπτει από το αρχικό με περιστροφή μέχρι οι αρχικοί άξονες να συμπίψουν με τις διευθύνσεις των ιδιοδιανυσμάτων e_1, e_2, \dots, e_p .

Έτσι οποιοδήποτε σημείο στον i άξονα του ελλειψοειδούς έχει x συντεταγμένες ανάλογες στο διάνυσμα $e_i = [e_{1i}, e_{2i}, \dots, e_{pi}]$ και συντεταγμένες όσον αφορά τις κύριες συνιστώσες της μορφής $[0, \dots, 0, y_i, 0, \dots, 0]$. Όταν το μ δεν είναι ίσο με μηδέν τότε οι κεντροποιημένες κύριες συνιστώσες της μορφής $y_i = e_i' (x - m)$ έχουν μέσο 0 και βρίσκονται στις κατευθύνσεις των e_i



Στο σχήμα απεικονίζεται η έλλειψη $x' \Sigma^{-1} x = c^2$ και οι κύριες συνιστώσες y_1, y_2 για το διδιάστατο κανονικό διάνυσμα X με μέσο μηδέν.

3.2.4 Κύριες συνιστώσες από τον πίνακα συσχέτισης

Οι κύριες συνιστώσες είναι δύσκολο να ερμηνευθούν όταν οι μεταβλητές μετρούνται σε διαφορετικές μονάδες μέτρησης. Ακόμη και όταν όλες οι μεταβλητές έχουν την ίδια μονάδα μέτρησης, μπορεί τα πεδία τιμών τους να είναι πολύ διαφορετικά και το εύρος τους πολύ μεγάλο. Τότε οι κύριες συνιστώσες επηρεάζονται πολύ από τις μεταβλητές που παρουσιάζουν μεγάλη διασπορά. Σε αυτές τις περιπτώσεις είναι προτιμότερο να τυποποιούνται οι μεταβλητές και στη συνέχεια να εφαρμόζεται η τεχνική των κύριων συνιστωσών στον πίνακα συνδιασποράς των τυποποιημένων μεταβλητών που δεν είναι άλλος από τον πίνακα συσχέτισης των αρχικών μεταβλητών. Αναλυτικότερα έχουμε τα εξής:

Οι τυποποιημένες μεταβλητές είναι:

$$\begin{aligned} Z_1 &= \frac{(X_1 - \mu_1)}{\sqrt{\sigma_{11}}} \\ Z_2 &= \frac{(X_2 - \mu_2)}{\sqrt{\sigma_{22}}} \\ &\vdots \\ Z_p &= \frac{(X_p - \mu_p)}{\sqrt{\sigma_{pp}}} \end{aligned}$$

ή σε μορφή πινάκων $Z = (V^{1/2})^{-1} (X - \mu)$

Όπου $V^{1/2}$ είναι ο διαγώνιος πίνακας των τυπικών αποκλίσεων. Ισχύουν τα εξής:

$$E(Z) = 0$$

$$Cov(Z) = (V^{1/2})^{-1} \Sigma (V^{1/2})^{-1} = \rho$$

Όπου ρ ο πίνακας συσχέτισης των X_1, \dots, X_p .

Σημειώνεται ότι οι διασπορές των Z_i , δηλαδή τα διαγώνια στοιχεία του ρ είναι 1 και συνεπώς η ολική μεταβλητότητα των Z_i είναι ίση με p (= πλήθος μεταβλητών).

Οι κύριες συνιστώσες του Z μπορούν να βρεθούν από τα ιδιοδιανύσματα του πίνακα συσχέτισης ρ του διανύσματος X . Όλα τα προηγούμενα αποτελέσματα μπορούν να εφαρμοσθούν και σε αυτή την περίπτωση με κάποιες απλοποιήσεις. Θα συνεχίσουμε να χρησιμοποιούμε τον συμβολισμό Y_i όταν αναφερόμαστε στην i κύρια συνιστώσα και τον συμβολισμό (λ_i, e_i) για το ζεύγος ιδιοτιμών-ιδιοδιανυσμάτων. Όμως αυτές οι ποσότητες που προέρχονται από τον πίνακα συνδιασπορών Σ είναι γενικά διαφορετικές από όσες προέρχονται από τον πίνακα συσχέτισης ρ .

Αποτέλεσμα 4:

Η i -οστή κύρια συνιστώσα των κανονικοποιημένων μεταβλητών $Z' = [Z_1, Z_2, \dots, Z_p]$ με $Cov(Z) = \rho$, δίνεται από τη σχέση:

$$Y_i = e_i' Z = e_i' (V^{1/2})^{-1} (X - \mu), \quad i = 1, 2, \dots, p.$$

Επιπλέον,

$$\sum_{i=1}^p \text{Var}(Y_i) = \sum_{i=1}^p \text{Var}(Z_i) = p,$$

$$\rho_{Y_i, Z_k} = e_{ki} \sqrt{\lambda_i}, \quad i, k = 1, 2, \dots, p.$$

Απόδειξη:

Η απόδειξη προκύπτει από τις προηγούμενες αποδείξεις στα αποτελέσματα 1,2 και 3 αν αντικαταστήσουμε στη θέση των μεταβλητών τις Z_1, Z_2, \dots, Z_p και στην θέση του

πίνακα συνδιασπορών Σ τον πίνακα συσχετίσεων ρ .

Βλέπουμε από την παραπάνω σχέση ότι η συνολική πληθυσμιακή διασπορά (για τις τυποποιημένες μεταβλητές) είναι ίση με p , το άθροισμα δηλαδή των διαγώνιων στοιχείων του πίνακα συσχετίσεων ρ . Χρησιμοποιώντας την παρακάτω σχέση:

ποσοστό της συνολικής

$$\text{διασποράς που ερμηνεύει} = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p}, \quad k = 1, 2, \dots, p$$

η k κύρια συνιστώσα

Και αντικαθιστώντας με Z στη θέση του X , παρατηρούμε ότι το ποσοστό της συνολικής διασποράς που ερμηνεύεται από την k κύρια συνιστώσα του Z είναι:

$$\left(\begin{array}{l} \text{ποσοστό συνολικής πληθυσμιακής} \\ \text{διασποράς που ερμηνεύεται από την} \\ \text{\textit{k} - κύρια συνιστώσα} \end{array} \right) = \frac{\lambda_k}{p}, \quad k = 1, 2, \dots, p$$

Όπου $\lambda_k, k=1, 2, \dots, p$ οι ιδιοτιμές του πίνακα συσχετίσεων ρ .

3.2.5 Κύριες συνιστώσες για πίνακες συνδιασποράς με ειδική δομή

Υπάρχουν κάποιοι πίνακες συνδιασπορών και συσχετίσεων των οποίων οι κύριες συνιστώσες μπορούν να εκφραστούν σε απλές μορφές. Υποθέτουμε ότι Σ είναι ο διαγώνιος πίνακας:

$$\Sigma = \begin{pmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \dots & \sigma_{pp} \end{pmatrix}$$

Θέτοντας $e'_i = (0, \dots, 0, i, 0, \dots, 0)$ με τη μονάδα στην i -οστή θέση παρατηρούμε ότι:

$$\begin{pmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \dots & \sigma_{pp} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \sigma_{ii} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{ή} \quad \Sigma e_i = \sigma_{ii} e_i$$

Συμπεραίνουμε έτσι ότι (σ_{ii}, e_i) είναι το i -οστό ζεύγος ιδιοτιμών-ιδιοδιανυσμάτων. Αφού ο γραμμικός συνδυασμός είναι $e'_i X = X_i$, το σύνολο των κύριων συνιστωσών είναι απλά το αρχικό σύνολο των ασυσχέτιστων τυχαίων μεταβλητών. Άρα για ένα πίνακα της μορφής αυτής (13) η ανάλυση σε κύριες συνιστώσες δεν αλλάζει κάτι.

Ένας άλλος πίνακας συνδιασπορών ο οποίος συχνά περιγράφει την ανταπόκριση ανάμεσα σε συγκεκριμένες βιολογικές μεταβλητές, όπως το μέγεθος των ζωντανών υπάρξεων, έχει την εξής γενική μορφή:

$$\Sigma = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \dots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \dots & \rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \dots & \sigma^2 \end{pmatrix}$$

Ο πίνακας συσχετίσεων τότε θα είναι:

$$\rho = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix}$$

ο οποίος είναι επίσης ο πίνακας συνδιασπορών των τυποποιημένων μεταβλητών. Ο παραπάνω πίνακας δηλώνει ότι οι μεταβλητές X_1, X_2, \dots, X_p είναι ίσα συσχετισμένες.

Οι p ιδιοτιμές του πίνακα συσχετίσεων αυτού μπορούν να χωριστούν σε δύο ομάδες. Όταν η συσχέτιση ρ είναι θετικός αριθμός, η μεγαλύτερη ιδιοτιμή είναι: $\lambda_1 = 1 + (p - 1)\rho$, με αντίστοιχο ιδιοδιάνυσμα:

$$e'_1 = \left(\frac{1}{\sqrt{p}}, \frac{1}{\sqrt{p}}, \dots, \frac{1}{\sqrt{p}} \right)$$

Οι υπόλοιπες $p-1$ ιδιοτιμές είναι: $\lambda_2 = \lambda_3 = \dots = \lambda_p = 1 - \rho$

Μία επιλογή για τα ιδιοδιανύσματα θα μπορούσε να είναι η εξής:

$$e'_2 = \left(\frac{1}{\sqrt{1 \times 2}}, \frac{-1}{\sqrt{1 \times 2}}, 0, \dots, 0 \right)$$

$$e'_3 = \left(\frac{1}{\sqrt{2 \times 3}}, \frac{1}{\sqrt{2 \times 3}}, \frac{-2}{\sqrt{2 \times 3}}, 0, \dots, 0 \right)$$

$$e'_i = \left(\frac{1}{\sqrt{(i-1)i}}, \dots, \frac{1}{\sqrt{(i-1)i}}, \frac{-(i-1)}{\sqrt{(i-1)i}}, 0, \dots, 0 \right)$$

$$e'_p = \left(\frac{1}{\sqrt{(p-1)p}}, \dots, \frac{1}{\sqrt{(p-1)p}}, \frac{-(p-1)}{\sqrt{(p-1)p}} \right).$$

Η πρώτη κύρια συνιστώσα

$$Y_1 = e'_1 X = \frac{1}{\sqrt{p}} \sum_{i=1}^p X_i$$

είναι ανάλογη στο άθροισμα των p αρχικών μεταβλητών. Αυτή η κύρια συνιστώσα ερμηνεύει το εξής ποσοστό της συνολικής μεταβλητότητας:

$$\frac{\lambda_1}{p} = \frac{1 + (p-1)\rho}{p} = \rho + \frac{1-\rho}{p}$$

Παρατηρούμε ότι $\frac{\lambda_1}{p} = \rho$, όταν ρ τείνει στο 1 ή όταν το p είναι πολύ μεγάλο. Για παράδειγμα εάν $\rho=0.80$ και $p = 5$, η πρώτη κύρια συνιστώσα ερμηνεύει 84% της συνολικής διασποράς. Όταν το ρ τείνει στο 1 οι τελευταίες $p-1$ συνιστώσες συνεισφέρουν πάρα πολύ λίγο στην συνολική διασπορά και άρα μπορούν να παραληφθούν.

3.2.6 Κύριες συνιστώσες από τον πίνακα διεγματικής διασποράς και τον πίνακα δειγματικής συσχέτισης

Στις περισσότερες πρακτικές εφαρμογές ο πίνακας Σ (ή ο πίνακας ρ) είναι άγνωστος και συνεπώς το ίδιο είναι και οι ιδιοτιμές του λ_i , τα αντίστοιχα

ιδιοδιανύσματα του e_i και φυσικά οι κύριες συνιστώσες, $y_i = e_i'x$. Στην πράξη αυτό που συνήθως έχουμε στη διάθεσή μας είναι ένα τυχαίο δείγμα $x_1, x_2, \dots, x_n \in \mathbb{R}^p$ ή ισοδύναμα ένα $n \times p$ πίνακα δεδομένων με στοιχεία τις μετρήσεις p μεταβλητών σε κάθε μία από τις n πειραματικές μονάδες. Αυτά τα δεδομένα έχουν δειγματικό μέσο \bar{x} , πίνακα δειγματικής συνδιασποράς S και πίνακα δειγματικής συσχέτισης R .

Επειδή τα λ_i και e_i που αντιστοιχούν στον Σ είναι άγνωστα, εκτιμούνται από τις αντίστοιχες ποσότητες του πίνακα S (που είναι εκτιμητής του Σ) και στη συνέχεια εξάγονται οι κύριες συνιστώσες. Δηλαδή, η όλη διαδικασία (εύρεση ιδιοτιμών και ιδιοδιανυσμάτων) γίνεται επί του πίνακα:

$$S = \frac{1}{n-1} \sum (X_i - \bar{X})(X_i - \bar{X})'$$

Έστω λοιπόν $\hat{Q} = (\hat{e}_1, \dots, \hat{e}_p)$ ο ορθογώνιος πίνακας $p \times p$, ώστε:

$$\hat{Q}'S\hat{Q} = \hat{\Lambda} = \begin{pmatrix} \hat{\lambda}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \hat{\lambda}_p \end{pmatrix},$$

όπου $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p \geq 0$ είναι οι ιδιοτιμές του πίνακα S . Τότε οι (στατιστικές συναρτήσεις) $\hat{Y}_j = \hat{e}_j'X$, $j=1,2,\dots,p$ λέγονται δειγματικές κύριες συνιστώσες (πρώτη, δεύτερη, κ.λ.π.). Οι \hat{Y}_j είναι εκτιμητές των Y_j (που προκύπτουν από τον Σ) και έχουν ανάλογες ιδιότητες, δηλαδή:

A) Δειγματική διασπορά της $\hat{Y}_j = \hat{\lambda}_j$, $j=1, 2, \dots, p$, αφού:

$$\frac{1}{n-1} \sum_{i=1}^n (\hat{e}_j'X_i - \hat{e}_j'\bar{X})^2 = \frac{1}{n-1} \hat{e}_j' \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' \hat{e}_j = \hat{e}_j'S\hat{e}_j = \hat{\lambda}_j$$

B) Δειγματική συνδιασπορά \hat{Y}_j και $\hat{Y}_m = 0$.

Γ) Ολική δειγματική διασπορά των $\hat{Y}_j = trS = \hat{\lambda}_1 + \dots + \hat{\lambda}_p$

Δ) Δειγματικός συντελεστής συσχέτισης \hat{Y}_j και X_i :

$$r_{\hat{Y}_j, X_i} = \frac{\hat{e}_j \sqrt{\hat{\lambda}_j}}{\sqrt{s_{ii}}}$$

Οι παρατηρήσεις x_1, x_2, \dots, x_n συχνά «κεντράρονται» αφαιρώντας από κάθε μία τον δειγματικό μέσο \bar{x} . Αυτό δεν έχει καμία επίδραση στον πίνακα S (που παραμένει ο ίδιος) και οι κύριες συνιστώσες παίρνουν την μορφή: $\hat{Y}_j = \hat{e}_j'(x - \bar{x})$, $j=1,2,\dots,p$. Εάν τώρα υπολογιστούν οι τιμές της \hat{Y}_j για $x = x_i$, $i=1,2,\dots,n$ (δηλαδή για το δείγμα), τότε ο δειγματικός μέσος

$$\bar{\hat{Y}} = \frac{1}{n} \sum_{i=1}^n \hat{e}_j'(x_i - \bar{x}) = 0. \text{ (Η δειγματική διασπορά παραμένει } \hat{\lambda}_j \text{).}$$

Όταν οι μεταβλητές μετρούνται σε διαφορετικές μονάδες μέτρησης είναι προτιμότερο τα δεδομένα να τυποποιούνται και να εφαρμόζεται ανάλυση κύριων συνιστωσών επί του πίνακα δειγματικής συσχέτισης R. Εάν συμβολίσουμε πάλι με $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \dots \geq \widehat{\lambda}_p \geq 0$ και $\widehat{e}_1, \widehat{e}_2, \dots, \widehat{e}_p$ τις ιδιοτιμές και τα ιδιοδιανύσματα του πίνακα R, οι κύριες συνιστώσες τότε θα είναι:

$$\widehat{Y}_j = \widehat{e}_j' \underline{z}, \text{ όπου } \underline{z} = (z_i)' \text{ και } z_i = \frac{x_i - \bar{x}_i}{\sqrt{s_{ii}}}, \quad i = 1, 2, \dots, p$$

Σχόλιο: X_i είναι ο γενικός συμβολισμός για την i αρχική μεταβλητή, \bar{x}_i είναι η i συνιστώσα του \bar{x} δηλαδή ο δειγματικός μέσος που αντιστοιχεί στην i μεταβλητή και s_{ii} είναι το i διαγώνιο στοιχείο του S, δηλαδή η δειγματική διασπορά που αντιστοιχεί στην i μεταβλητή.

Οι δειγματικές κύριες συνιστώσες που υπολογίζονται από τον πίνακα S (ή τον R) έχουν αντίστοιχη γεωμετρική ερμηνεία όπως αυτές που υπολογίζονται από τον Σ. Έτσι εάν $\widehat{Y}_j = \widehat{e}_j'(x - \bar{x})$, $j=1, 2, \dots, p$ είναι οι δειγματικές κύριες συνιστώσες και θεωρήσουμε το ελλειψοειδές: $(x - \bar{x})S^{-1}(x - \bar{x}) = c^2$

Αυτό έχει κέντρο τον δειγματικό μέσο των δεδομένων \bar{x} και άξονες $\pm \sqrt{\widehat{\lambda}_j} c \widehat{e}_j$.

Επί πλέον, οι συντεταγμένες του $x_i - \bar{x}$ στο σύστημα ορθογώνιων αξόνων που ορίζουν τα ιδιοδιανύσματα $\widehat{e}_1, \widehat{e}_2, \dots, \widehat{e}_p$ και έχει κέντρο \bar{x} είναι οι δειγματικές κύριες συνιστώσες \widehat{Y}_j , $j=1, 2, \dots, p$. Ακόμη το \widehat{e}_1 που αντιστοιχεί στην μεγαλύτερη ιδιοτιμή $\widehat{\lambda}_1$, δείχνει την διεύθυνση κατά την οποία τα δεδομένα παρουσιάζουν την μεγαλύτερη μεταβλητότητα.

Ολοκληρώνοντας τα περί των κύριων συνιστωσών αναφέρουμε ότι η επιλογή του k (του πλήθους των κύριων συνιστωσών που θα «κρατήσουμε») δεν είναι γενικά απλή υπόθεση. Πέραν του καθαρά εμπειρικού τρόπου υπάρχουν και άλλοι στηριζόμενοι σε στατιστικούς ελέγχους. Όταν οι κύριες συνιστώσες υπολογίζονται από τους πίνακες ρ ή R ένας εμπειρικός κανόνας είναι να κρατήσουμε εκείνες τις κύριες συνιστώσες για τις οποίες οι αντίστοιχες ιδιοτιμές είναι μεγαλύτερες της μονάδας. Το σκεπτικό είναι ότι αυτές οι κύριες συνιστώσες εξηγούν τουλάχιστον το $1/p$ της ολικής μεταβλητότητας.

3.3 ΒΙΒΛΙΟΓΡΑΦΙΑ 3^{ου} ΚΕΦΑΛΑΙΟΥ

- [1] Aczel A., Sounderpandian J. - Επιμέλεια ελληνικής έκδοσης Μιχάλης Σφακιανάκης (2013) - Στατιστική σκέψη στον κόσμο των επιχειρήσεων, Εκδόσεις: Π.Χ. Πασχαλίδης
- [2] Παπαγεωργίου Ανδρέας (2009-2010) Διπλωματική Εργασία: «Παραγοντική ανάλυση σε κύριες συνιστώσες», Πανεπιστήμιο Πατρών, Διατμηματικό Π.Μ.Σ. «Μαθηματικά των Υπολογιστών και των Αποφάσεων»

ΚΕΦΑΛΑΙΟ 4 - ΑΝΑΛΥΣΗ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ (PCA)

Η ανάλυση κύριων συνιστωσών (Principal Component Analysis, PCA) είναι μία από τις τεχνικές της πολυμεταβλητής στατιστικής ανάλυσης (Multivariate Statistical Analysis), που χρησιμοποιείται από τους αναλυτές προκειμένου να εξετάσουν περισσότερα από ένα χαρακτηριστικά συγχρόνως.

Στόχος της PCA είναι να δημιουργήσει ένα μικρό και διαχειρίσιμο αριθμό από ασυσχέτιστους γραμμικούς συνδυασμούς των αρχικών μεταβλητών, οι οποίοι θα περιέχουν κατά το δυνατόν μεγαλύτερο μέρος της πληροφορίας που υπάρχει στις αρχικές μεταβλητές. Έτσι, τα αρχικά πολυπληθή δεδομένα, μπορούν να αναπαρασταθούν γραφικά μέσω των πρώτων κύριων συνιστωσών, παρέχοντας μία απλή και γλαφυρή απεικόνιση της αρχικής πληροφορίας.

Για την κατασκευή των κύριων συνιστωσών απαιτείται η εύρεση των ιδιοτιμών και των ιδιοδιανυσμάτων του δειγματικού πίνακα διακυμάνσεων – συνδιακυμάνσεων ή του πίνακα συσχέτισεων των αρχικών δεδομένων. Η μεγαλύτερη ιδιοτιμή και το αντίστοιχο ιδιοδιάνυσμα αντιστοιχούν στην πρώτη κύρια συνιστώσα, η δεύτερη μεγαλύτερη ιδιοτιμή στη δεύτερη κύρια συνιστώσα, κ.ο.κ. Η χρήση όλων των κύριων συνιστωσών συνεπάγεται την πλήρη κάλυψη της διακύμανσης των αρχικών δεδομένων, ενώ αν παραληφθούν κάποιες κύριες συνιστώσες θα χαθεί ένα μέρος της αρχικής πληροφορίας.

Η εφαρμογή της PCA επιτρέπει την ταχύτερη και οικονομικότερη διαχείριση και αποθήκευση των αρχικών δεδομένων, λόγω της σημαντικής ελάττωσης του όγκου τους. Το «κόστος» της PCA είναι η απώλεια ενός μέρους της αρχικής πληροφορίας.

Το προϊόν της PCA είναι ένα σύνολο ασυσχέτιστων μεταβλητών, γεγονός που μπορεί να βοηθήσει στην περαιτέρω ανάλυση των δεδομένων με τη χρήση στατιστικών τεχνικών, οι οποίες δεν μπορούσαν να εφαρμοστούν αρχικά.

Συνήθως, απώτερος στόχος για έναν αναλυτή που εφαρμόζει την PCA, είναι να κατορθώσει να δώσει κάποια ερμηνεία στις κύριες συνιστώσες και να τις χρησιμοποιήσει ως «δείκτες» για τα χαρακτηριστικά που εξετάζει.

4.1 Έλεγχος καταλληλότητας δεδομένων για εφαρμογή της PCA

Αρχικά θα πρέπει να ελεγχθεί αν υπάρχουν στατιστικά σημαντικές συσχετίσεις μεταξύ των 12 μεταβλητών που θα αναλυθούν, ούτως ώστε να διαπιστωθεί αν έχει νόημα η προσπάθεια μείωσης της διάστασης των δεδομένων. Σε περίπτωση που εντοπιστεί κάποιο χαρακτηριστικό απασχόλησης που δεν συσχετίζεται με τα υπόλοιπα, αυτό θα πρέπει να εξαιρεθεί από την περαιτέρω ανάλυση. Ο λόγος της εξαίρεσής του είναι ότι κάποια από τις κύριες συνιστώσες θα ταυτιστεί με το συγκεκριμένο χαρακτηριστικό, και συνεπώς είναι προτιμότερο αυτό να μελετηθεί ξεχωριστά από τα υπόλοιπα.

Στον πίνακα συσχετίσεων (correlation matrix) που ακολουθεί στην επόμενη σελίδα, φαίνεται ότι όλα τα χαρακτηριστικά απασχόλησης παρουσιάζουν στατιστικά σημαντικές συσχετίσεις ανά δύο. Εξαίρεση αποτελεί το χαρακτηριστικό "*Limited_Duration_Contract*", το οποίο δεν συσχετίζεται σημαντικά με κανένα από τα υπόλοιπα χαρακτηριστικά, και συνεπώς δεν θα ληφθεί υπόψη κατά την εφαρμογή της PCA.

Correlations

		Employment rates by gender, MALE	Employment rates by gender, FEMALE	Pre-primary, primary & lower secondary - ISCED levels 0-2	Upper secondary & post-secondary non-tertiary - ISCED levels 3-4	Tertiary - ISCED levels 5-6	15-24 years old	25-54 years old	55-64 years old	Persons with a second job (% of total employment)	Male working part-time (% of total employment)	Female working part-time (% of total employment)	Proportion of employees with a contract of limited duration (% of total employees)
Employment rates by gender, MALE	Pearson Correlation	1	,624**	,807**	,855**	,704**	,610**	,554**	,526**	,536**	,553**	,719**	,131
	Sig. (2-tailed)		,000	,000	,000	,000	,000	,001	,002	,001	,001	,000	,468
	N	33	33	33	33	33	33	33	33	33	33	33	33
Employment rates by gender, FEMALE	Pearson Correlation	,624**	1	,628**	,725**	,805**	,543**	,846**	,764**	,622**	,595**	,514**	,037
	Sig. (2-tailed)	,000		,000	,000	,000	,001	,000	,000	,000	,000	,002	,839
	N	33	33	33	33	33	33	33	33	33	33	33	33
Pre-primary, primary & lower secondary - ISCED levels 0-2	Pearson Correlation	,807**	,628**	1	,752**	,588**	,489**	,467**	,558**	,592**	,555**	,594**	,233
	Sig. (2-tailed)	,000	,000		,000	,000	,004	,006	,001	,000	,001	,000	,191
	N	33	33	33	33	33	33	33	33	33	33	33	33
Upper secondary & post-secondary non-tertiary - ISCED levels 3-4	Pearson Correlation	,855**	,725**	,752**	1	,794**	,619**	,677**	,565**	,585**	,530**	,619**	,066
	Sig. (2-tailed)	,000	,000	,000		,000	,000	,000	,001	,000	,001	,000	,715
	N	33	33	33	33	33	33	33	33	33	33	33	33
Tertiary - ISCED levels 5-6	Pearson Correlation	,704**	,805**	,588**	,794**	1	,531**	,754**	,564**	,666**	,533**	,489**	-,036
	Sig. (2-tailed)	,000	,000	,000	,000		,001	,000	,001	,000	,001	,004	,843
	N	33	33	33	33	33	33	33	33	33	33	33	33
15-24 years old	Pearson Correlation	,610**	,543**	,489**	,619**	,531**	1	,560**	,669**	,573**	,644**	,695**	-,014
	Sig. (2-tailed)	,000	,001	,004	,000	,001		,001	,000	,000	,000	,000	,940
	N	33	33	33	33	33	33	33	33	33	33	33	33
25-54 years old	Pearson Correlation	,554**	,846**	,467**	,677**	,754**	,560**	1	,606**	,489**	,418*	,434*	-,032
	Sig. (2-tailed)	,001	,000	,006	,000	,000	,001		,000	,004	,015	,012	,862
	N	33	33	33	33	33	33	33	33	33	33	33	33
55-64 years old	Pearson Correlation	,526**	,764**	,558**	,565**	,564**	,669**	,606**	1	,617**	,560**	,490**	-,028
	Sig. (2-tailed)	,002	,000	,001	,001	,001	,000	,000		,000	,001	,004	,879
	N	33	33	33	33	33	33	33	33	33	33	33	33
Persons with a second job (% of total employment)	Pearson Correlation	,536**	,622**	,592**	,585**	,666**	,573**	,489**	,617**	1	,682**	,512**	,227
	Sig. (2-tailed)	,001	,000	,000	,000	,000	,000	,004	,000		,000	,002	,203
	N	33	33	33	33	33	33	33	33	33	33	33	33
Male working part-time (% of total employment)	Pearson Correlation	,553**	,595**	,555**	,530**	,533**	,644**	,418*	,560**	,682**	1	,805**	,116
	Sig. (2-tailed)	,001	,000	,001	,001	,001	,000	,015	,001	,000		,000	,520
	N	33	33	33	33	33	33	33	33	33	33	33	33
Female working part-time (% of total employment)	Pearson Correlation	,719**	,514**	,594**	,619**	,489**	,695**	,434*	,490**	,512**	,805**	1	,123
	Sig. (2-tailed)	,000	,002	,000	,000	,004	,000	,012	,004	,002	,000		,494
	N	33	33	33	33	33	33	33	33	33	33	33	33
Proportion of employees with a contract of limited duration (% of total employees)	Pearson Correlation	,131	,037	,233	,066	-,036	-,014	-,032	-,028	,227	,116	,123	1
	Sig. (2-tailed)	,468	,839	,191	,715	,843	,940	,862	,879	,203	,520	,494	
	N	33	33	33	33	33	33	33	33	33	33	33	33

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

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

Αφού εξαιρεθεί το χαρακτηριστικό “*Limited_Duration_Contract*”, η PCA μπορεί να εφαρμοστεί στα υπόλοιπα 11 χαρακτηριστικά απασχόλησης.

4.2 Επιλογή δειγματικού πίνακα

Κατά την έναρξη εφαρμογής της PCA υπάρχουν δύο επιλογές: Η πρώτη επιλογή είναι να χρησιμοποιηθούν «κεντροποιημένα» δεδομένα, γεγονός που αντιστοιχεί στην ανάλυση του πίνακα διακυμάνσεων-συνδιακυμάνσεων. Η δεύτερη επιλογή είναι να προηγηθεί τυποποίηση των δεδομένων, ώστε να αναλυθεί ο πίνακας συσχετίσεων. Τα αποτελέσματα της PCA θα διαφέρουν ανάλογα με το ποιος πίνακας τελικά θα επιλεγεί για ανάλυση.

Η επιλογή του δειγματικού πίνακα είναι υποκειμενική και εξαρτάται από το κατά πόσο εκτιμάται ότι τα κεντροποιημένα δεδομένα θα οδηγήσουν σε «μεροληπτικά» αποτελέσματα. Κάτι τέτοιο μπορεί να συμβεί αν υπάρχουν μεταβλητές με μεγάλη διαφορά διακύμανσης ως προς τις υπόλοιπες. Αυτό είναι σύνθητες φαινόμενο, όταν οι μεταβλητές είναι εκφρασμένες σε διαφορετικές κλίμακες ή/και με διαφορετικές μονάδες μέτρησης. Από την άλλη μεριά, δεν θα πρέπει να επιλέγεται άκριτα ο δειγματικός πίνακας συσχετίσεων, γιατί σε ορισμένες περιπτώσεις ενδέχεται να χαθεί πολύτιμη πληροφορία που «κρύβεται» στις διαφορετικές διακυμάνσεις μεταξύ των μεταβλητών.

Στο πρόβλημα που μελετάμε, κάποιες μεταβλητές είναι εκφρασμένες ως ποσοστά επί του συνόλου του πληθυσμού των χωρών, ενώ κάποιες άλλες είναι εκφρασμένες ως ποσοστά επί των εργαζόμενων. Λαμβάνοντας υπόψη το στόχο της εργασίας, θεωρήσαμε ότι κάποιες μεταβλητές με μικρή διακύμανση μπορεί να είναι αρκετά σημαντικές για τη διαμόρφωση των τελικών συμπερασμάτων. Για το λόγο αυτό, προχωρήσαμε σε τυποποίηση των δεδομένων και επιλέξαμε για την ανάλυση PCA τον δειγματικό πίνακα συσχετίσεων.

4.3 Επιλογή βέλτιστου αριθμού κύριων συνιστωσών.

Τα κριτήρια επιλογής του βέλτιστου αριθμού κύριων συνιστωσών είναι τα ακόλουθα:

4.3.1 Κριτήριο 1ο:

Ποσοστό της αρχικής διακύμανσης που αθροιστικά αποτυπώνουν οι κύριες 68 συνιστώσες. Σύμφωνα με αυτό το κριτήριο, τίθεται κάποιο ελάχιστο αποδεκτό όριο

και στη συνέχεια επιλέγεται ο ελάχιστος αριθμός κύριων συνιστωσών, οι οποίες αθροιστικά καλύπτουν τουλάχιστον αυτό το αποδεκτό όριο διακύμανσης. Είναι υποκειμενικό κριτήριο και εξαρτάται από τη φύση του προβλήματος και τις απαιτήσεις των αναλυτών. Αν επιλεγεί υψηλό όριο κάλυψης της αρχικής διακύμανσης, ενδεχομένως να απαιτηθεί η χρήση πολλών κύριων συνιστωσών, οπότε τα οφέλη από την εφαρμογή της μεθόδου περιορίζονται σημαντικά.

Στο συγκεκριμένο πρόβλημα που εξετάζουμε, τα ποσοστά κάλυψης της αρχικής διακύμανσης που καλύπτουν οι κύριες συνιστώσες (μεμονωμένα και αθροιστικά) φαίνονται στον πίνακα που ακολουθεί. Έτσι, αν επιλέξουμε να χρησιμοποιήσουμε τις δύο πρώτες κύριες συνιστώσες, αυτό σημαίνει ότι θεωρούμε ικανοποιητική την κάλυψη του 74,5% της αρχικής διακύμανσης των δεδομένων.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7,164	65,124	65,124	7,164	65,124	65,124
2	1,032	9,380	74,504	1,032	9,380	74,504
3	,811	7,371	81,875			
4	,556	5,058	86,933			
5	,464	4,220	91,153			
6	,368	3,343	94,496			
7	,194	1,767	96,263			
8	,141	1,284	97,547			
9	,123	1,117	98,664			
10	,082	,747	99,411			
11	,065	,589	100,000			

Extraction Method: Principal Component Analysis.

Πίνακας 37 - Ιδιοτιμές των κύριων συνιστωσών και ποσοστά κάλυψης της συνολικής αρχικής διακύμανσης

4.3.2 Κριτήριο 2^ο:

Το δεύτερο κριτήριο, γνωστό και ως «κριτήριο του Kaiser», προτείνει να επιλέγουμε εκείνες τις κύριες συνιστώσες, των οποίων οι ιδιοτιμές είναι μεγαλύτερες από τη μέση τιμή όλων των ιδιοτιμών. Στην περίπτωση που εξετάζουμε, επειδή για την ανάλυση επιλέξαμε τη χρήση του πίνακα συσχετίσεων, η μέση τιμή των ιδιοτιμών είναι μονάδα. Συνεπώς, μπορούμε να επιλέξουμε εκείνες τις κύριες συνιστώσες που έχουν ιδιοτιμή μεγαλύτερη της μονάδας. Στο πρόβλημα που εξετάζουμε και όπως φαίνεται από τις ιδιοτιμές στο αριστερό μέρος του παραπάνω πίνακα, το κριτήριο

Kaiser μας οδηγεί στην επιλογή των δύο πρώτων κύριων συνιστωσών, με ιδιοτιμές $\lambda_1 = 7.164$ και $\lambda_2 = 1.032$ αντίστοιχα.

4.3.3 Κριτήριο 3^ο:

Ποσοστό διακύμανσης που ερμηνεύεται για την κάθε μεταβλητή. Σύμφωνα με αυτό το κριτήριο τίθεται κάποιο ελάχιστο αποδεκτό όριο κάλυψης για τη διακύμανση της κάθε αρχικής μεταβλητής. Στη συνέχεια επιλέγονται τόσες κύριες συνιστώσες, ούτως ώστε να καλύπτεται για την κάθε μεταβλητή ποσοστό τουλάχιστον ίσο με το ελάχιστο όριο που τέθηκε. Το κριτήριο αυτό είναι υποκειμενικό. Αν η διακύμανση κάποιας μεμονωμένης μεταβλητής δεν καλύπτεται ικανοποιητικά, το κριτήριο θα οδηγήσει στην επιλογή μεγάλου αριθμού κύριων συνιστωσών και πιθανόν να ακυρώσει τα οφέλη της μεθόδου. Στον πίνακα που ακολουθεί, φαίνεται η κάλυψη της διακύμανσης ανά μεταβλητή έπειτα από την επιλογή των δύο πρώτων κύριων συνιστωσών. Στην περίπτωση αυτή, το χειρότερο ποσοστό κάλυψης αρχικής διακύμανσης είναι 60,2% και παρατηρείται στην μεταβλητή “Persons with a second job (% of total employment)”.

Communalities		
	Initial	Extraction
Employment rates by gender, MALE	1,000	,728
Employment rates by gender, FEMALE	1,000	,878
Pre-primary, primary & lower secondary - ISCED levels 0-2	1,000	,640
Upper secondary & post-secondary non-tertiary - ISCED levels 3-4	1,000	,782
Tertiary - ISCED levels 5-6	1,000	,819
15-24 years old	1,000	,666
25-54 years old	1,000	,821
55-64 years old	1,000	,615
Persons with a second job (% of total employment)	1,000	,602
Male working part-time (% of total employment)	1,000	,793
Female working part-time (% of total employment)	1,000	,852

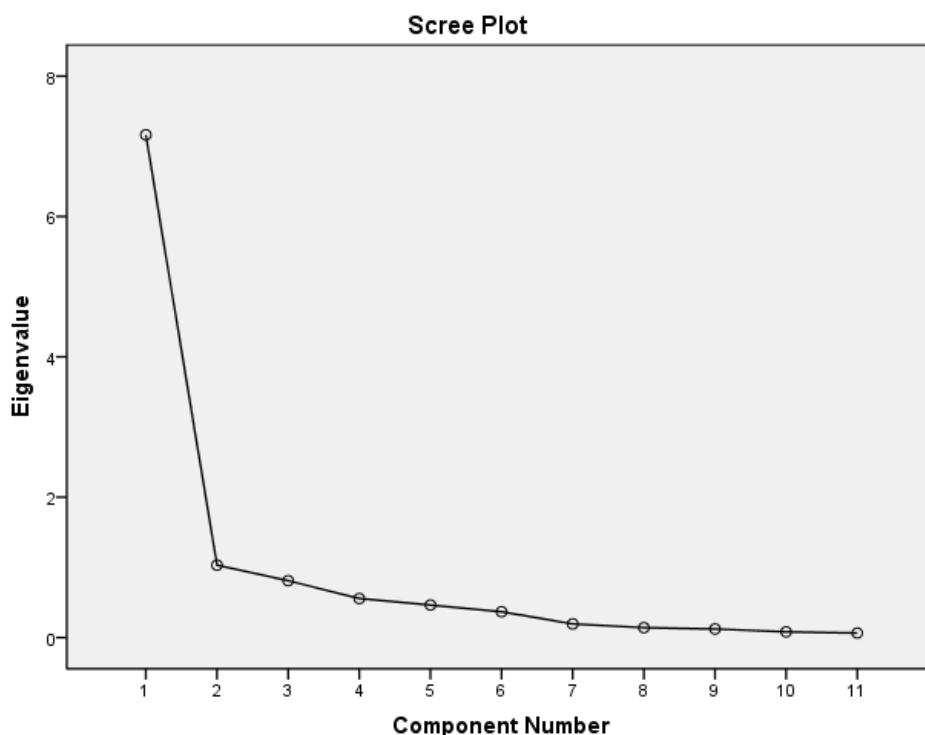
Extraction Method: Principal Component Analysis.

Πίνακας 38 - Ποσοστά κάλυψης της αρχικής διακύμανσης ανά μεταβλητή, από τις δύο πρώτες κύριες συνιστώσες

4.3.4 Κριτήριο 4^ο:

Διάγραμμα «αγκώνα» (scree plot). Το διάγραμμα αυτό απεικονίζει τις ιδιοτιμές πάνω σε μία τεθλασμένη γραμμή, σύμφωνα με τη σειρά μεγέθους τους. Εξετάζοντας 70 οπτικά την εικόνα που δίνει το διάγραμμα, εντοπίζεται το σημείο στο οποίο η

τεθλασμένη γραμμή γίνεται σχεδόν οριζόντια και επιλέγεται ο αριθμός των κύριων συνιστωσών που αντιστοιχεί στο συγκεκριμένο σημείο. Με άλλα λόγια, δεξιότερα από το σημείο αυτό το «κόστος» χρήσης επιπρόσθετων κύριων συνιστωσών δεν αντισταθμίζεται από αντίστοιχο «όφελος», όσον αφορά την κάλυψη περαιτέρω διακύμανσης. Στο πρόβλημα που εξετάζουμε, βάσει του κριτηρίου αυτού, φαίνεται ως βέλτιστη επιλογή η χρήση των δύο πρώτων κύριων συνιστωσών.

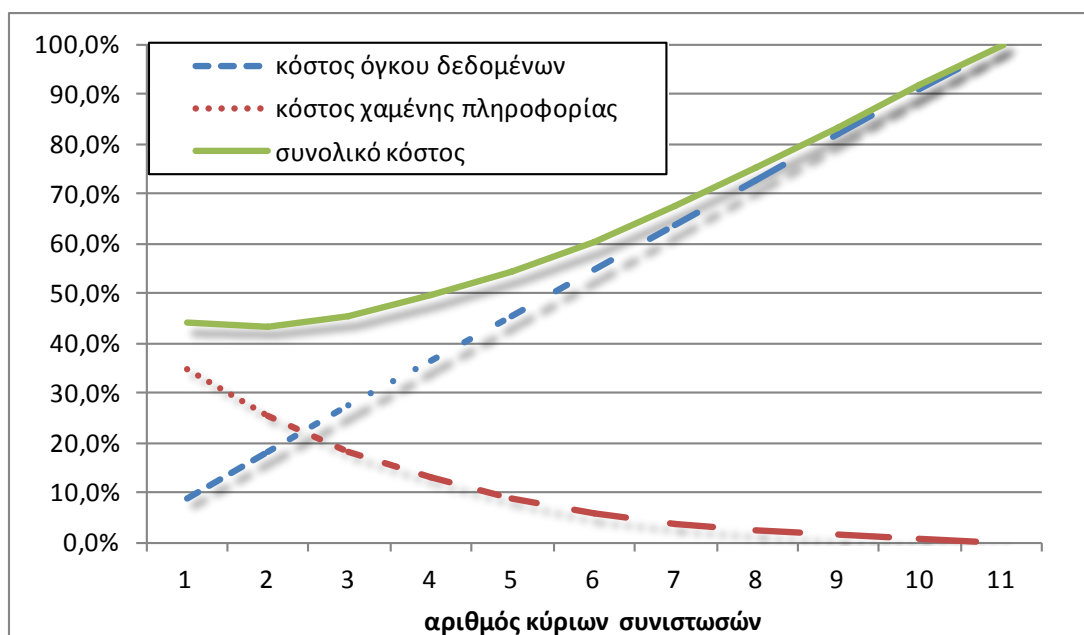


Διάγραμμα 19 - Οι ιδιοτιμές ταξινομημένες με τη σειρά μεγέθους τους

Συγκεκριμένα όλα τα παραπάνω κριτήρια, θεωρούμε ως βέλτιστη επιλογή τη χρήση των δύο πρώτων κύριων συνιστωσών, οι οποίες κρίνεται ότι ικανοποιούν τα κριτήρια 2 και 4, ενώ ελαχιστοποιούν το συνολικό «κόστος» της ανάλυσης, όπως φαίνεται στον επόμενο πίνακα.

αριθμός κύριων συνιστωσών	"κόστος" όγκου δεδομένων	"κόστος" χαμένης πληροφορίας	συνολικό "κόστος"
1	9,1%	34,9%	44,0%
2	18,2%	25,5%	43,7%
3	27,3%	18,1%	45,4%
4	36,4%	13,1%	49,4%
5	45,5%	8,8%	54,3%
6	54,5%	5,5%	60,0%
7	63,6%	3,7%	67,4%
8	72,7%	2,5%	75,2%
9	81,8%	1,3%	83,2%
10	90,9%	0,6%	91,5%
11	100,0%	0,0%	100,0%

Πίνακας 39 - Κόστος χρήσης κύριων συνιστωσών



Διάγραμμα 20 - Κόστος χρήσης κύριων συνιστωσών

Αξίζει να σημειωθεί ότι η χρήση των δύο πρώτων κύριων συνιστωσών θα επιτρέψει μία παραστατική απεικόνιση της αρχικής πληροφορίας των δεδομένων σε ένα σύστημα δύο κάθετων αξόνων (ονομάζονται πρώτοι κύριοι άξονες), μετατρέποντας έτσι σε μία εύληπτη «εικόνα» το 74,5% της πληροφορίας που κρύβεται στα αρχικά δεδομένα.

4.4 Ιδιοδιανύσματα των δύο πρώτων κύριων συνιστωσών

Τα ιδιοδιανύσματα u_1 και u_2 των δύο πρώτων κύριων συνιστωσών φαίνονται αντίστοιχα στον πίνακα που ακολουθεί.

Component Matrix^a

	Component	
	1	2
Employment rates by gender, MALE	,848	,093
Employment rates by gender, FEMALE	,867	-,355
Pre-primary, primary & lower secondary - ISCED levels 0-2	,795	,092
Upper secondary & post-secondary non-tertiary - ISCED levels 3-4	,876	-,120
Tertiary - ISCED levels 5-6	,842	-,331
15-24 years old	,778	,247
25-54 years old	,769	-,479
55-64 years old	,778	-,102
Persons with a second job (% of total employment)	,772	,077
Male working part-time (% of total employment)	,770	,447
Female working part-time (% of total employment)	,771	,508

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

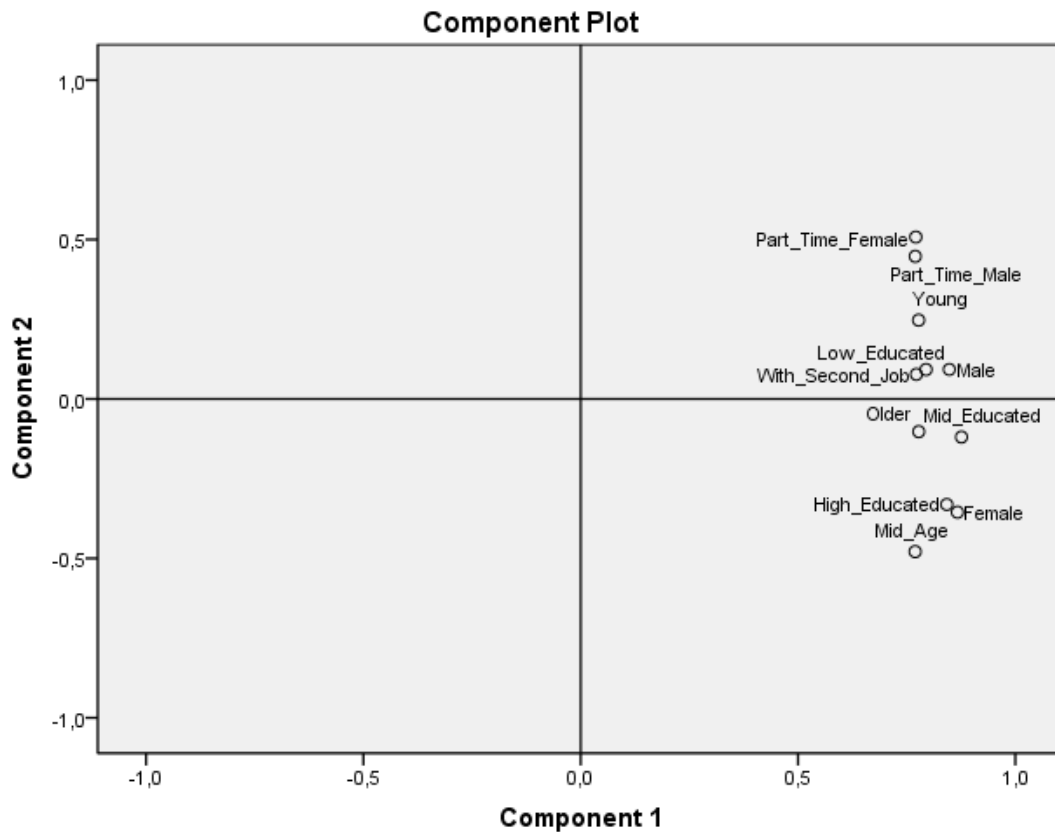
Πίνακας 40 - Ιδιοδιανύσματα των δύο πρώτων κύριων συνιστωσών

Τα ιδιοδιανύσματα είναι:

$$u_1 = (.848, .867, .795, .876, .842, .778, .769, .778, .772, .770, .771)'$$

$$u_2 = (.093, -.355, .092, -.120, -.331, .247, -.479, -.102, .077, .447, .508)'$$

Αν τα παραπάνω ιδιοδιανύσματα απεικονιστούν σε ένα σύστημα δύο αξόνων για κάθε χαρακτηριστικό απασχόλησης, προκύπτει το ακόλουθο διάγραμμα.



Διάγραμμα 21 - Συντεταγμένες των χαρακτηριστικών απασχόλησης

Έστω ότι για μία συγκεκριμένη χώρα, κατασκευάζουμε το διάνυσμα-γραμμή \mathbf{X}' που περιλαμβάνει τις z-τιμές των 11 τυποποιημένων χαρακτηριστικών απασχόλησης, δηλαδή $\mathbf{X}' = (X_1, X_2, \dots, X_{11})$. Τότε:

- Το score f_1 που δίνει η 1^η κύρια συνιστώσα στη συγκεκριμένη χώρα προκύπτει από το γινόμενο $\mathbf{X}' \cdot \mathbf{u}_1$
- Το score f_2 που δίνει η 2^η κύρια συνιστώσα στη συγκεκριμένη χώρα προκύπτει από το γινόμενο $\mathbf{X}' \cdot \mathbf{u}_2$

Αφού κατασκευάσουμε δύο νέες μεταβλητές με τα scores f_1 και f_2 για την κάθε Ευρωπαϊκή χώρα, στη συνέχεια ελέγχουμε πόσο ισχυρά συσχετίζεται το κάθε χαρακτηριστικό απασχόλησης με τις δύο νέες μεταβλητές. Έτσι, θα δούμε ποια είναι εκείνα τα χαρακτηριστικά απασχόλησης που συμβάλλουν περισσότερο στη διαμόρφωση της κάθε κύριας συνιστώσας. Στην προκειμένη περίπτωση, επειδή χρησιμοποιήσαμε το δειγματικό πίνακα συσχετίσεων, οι συντελεστές συσχέτισης συμπίπτουν με τις αντίστοιχες τιμές των ιδιοδιανυσμάτων, όπως ακριβώς φαίνονται στον Πίνακα 40.

Παρατηρούμε ένα ανεπιθύμητο φαινόμενο: Όλα τα χαρακτηριστικά απασχόλησης παρουσιάζουν ισχυρή θετική συσχέτιση με την 1^η κύρια συνιστώσα. Σε περιπτώσεις όπως αυτή, δεν μπορούμε να δώσουμε κάποια «ερμηνεία» στις δύο κύριες συνιστώσες.

Στην επόμενη ενότητα θα πραγματοποιήσουμε μία περιστροφή των κύριων αξόνων, προκαλώντας διαφοροποιήσεις στους συντελεστές συσχέτισης ανά χαρακτηριστικό απασχόλησης και ανά κύρια συνιστώσα. Εξετάζοντας τους νέους συντελεστές συσχέτισης που θα προκύψουν, θα επιχειρήσουμε εκ νέου να ερμηνεύσουμε τις πρώτες κύριες συνιστώσες.

4.5 Περιστροφή των πρώτων κύριων αξόνων

Μετά την περιστροφή των αξόνων, οι συντελεστές συσχέτισης των χαρακτηριστικών απασχόλησης με την κάθε κύρια συνιστώσα φαίνονται στον επόμενο πίνακα.

Rotated Component Matrix^a

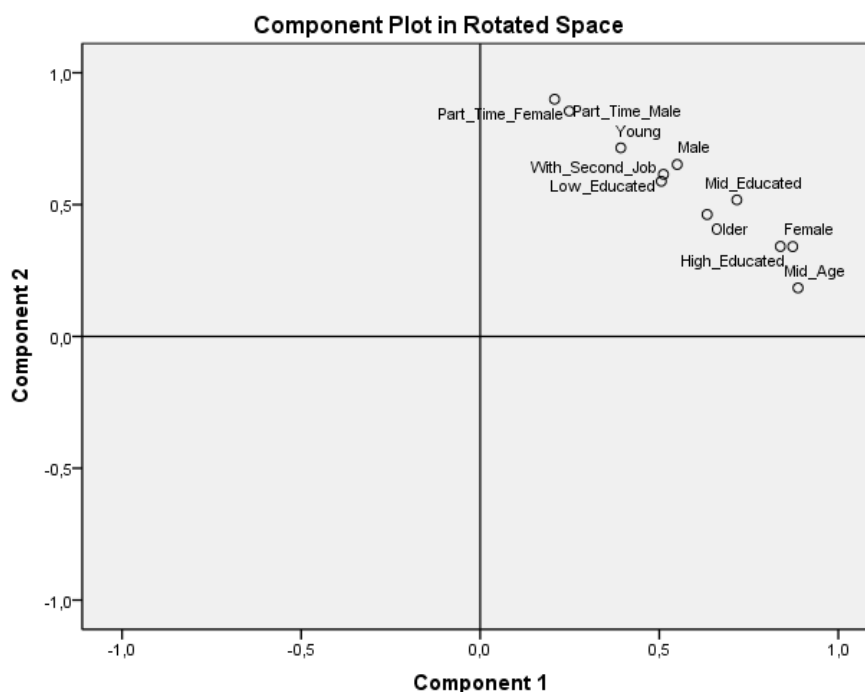
	Component	
	1	2
Employment rates by gender, MALE	,550	,852
Employment rates by gender, FEMALE	,873	,341
Pre-primary, primary & lower secondary - ISCED levels 0-2	,512	,615
Upper secondary & post-secondary non-tertiary - ISCED levels 3-4	,717	,518
Tertiary - ISCED levels 5-6	,838	,342
15-24 years old	,392	,715
25-54 years old	,887	,184
55-64 years old	,634	,462
Persons with a second job (% of total employment)	,506	,589
Male working part-time (% of total employment)	,249	,855
Female working part-time (% of total employment)	,208	,900

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

a. Rotation converged in 3 iterations.

Πίνακας 41 - Συντελεστές χαρακτηριστικών απασχόλησης με τις δύο πρώτες κύριες συνιστώσες, μετά την περιστροφή των αξόνων 75

Όπως φαίνεται από τον παραπάνω πίνακα, κάποια χαρακτηριστικά απασχόλησης τώρα έχουν μεγάλο συντελεστή συσχέτισης με την 1^η κύρια συνιστώσα, ενώ κάποια άλλα με τη 2^η κύρια συνιστώσα.



Διάγραμμα 22 - Συντεταγμένες των χαρακτηριστικών απασχόλησης μετά την περιστροφή των κύριων αξόνων

Επιπλέον, όπως φαίνεται και στον πίνακα της επόμενης σελίδας, όσον αφορά στην ερμηνευτική «βαρύτητα» που έχουν πλέον οι δύο πρώτες κύριες συνιστώσες μετά την περιστροφή, επισημαίνονται τα εξής:

- το συνολικό ποσοστό από την αρχική διακύμανση που ερμηνεύεται από τις δύο πρώτες κύριες συνιστώσες παραμένει το ίδιο (74.5%)
- η συνεισφορά των δύο πρώτων κύριων συνιστωσών έχει διαφοροποιηθεί από 65.1% vs 9.4% και έχει «ισορροπήσει» στο 38.6% vs 35.9%. Αυτό σημαίνει ότι μετά την περιστροφή των κύριων αξόνων και οι δύο πρώτες κύριες συνιστώσες έχουν αποκτήσει περίπου την ίδια βαρύτητα ως προς την ερμηνευτική τους ικανότητα.

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	7,164	65,124	65,124	7,164	65,124	65,124	4,245	38,594	38,594
2	1,032	9,380	74,504	1,032	9,380	74,504	3,950	35,910	74,504
3	,811	7,371	81,875						
4	,556	5,058	86,933						
5	,464	4,220	91,153						
6	,368	3,343	94,496						
7	,194	1,767	96,263						
8	,141	1,284	97,547						
9	,123	1,117	98,664						
10	,082	,747	99,411						
11	,065	,589	100,000						

Extraction Method: Principal Component Analysis.

Πίνακας 42 - Ερμηνευτική ικανότητα των 2 πρώτων κύριων συνιστωσών πριν και μετά την περιστροφή

4.6 Ερμηνεία των πρώτων κύριων συνιστωσών και δημιουργία «δεικτών» απασχόλησης

Η ερμηνεία των πρώτων κύριων συνιστωσών είναι ίσως το πιο ενδιαφέρον αλλά και πιο δύσκολο κομμάτι της μεθόδου PCA. Με τη βοήθεια της περιστροφής αξόνων που πραγματοποιήσαμε στην προηγούμενη ενότητα, προέκυψαν αρκετά διαφοροποιημένες συσχετίσεις για τα χαρακτηριστικά απασχόλησης και τις πρώτες κύριες συνιστώσες, γεγονός που επιτρέπει μία προσπάθεια ερμηνείας. Στη συνέχεια θα προσπαθήσουμε να ερμηνεύσουμε τις πρώτες κύριες συνιστώσες, λαμβάνοντας υπόψη και θεωρώντας σημαντικές μόνο εκείνες τις περιπτώσεις που οι συντελεστής συσχέτισης είναι κατ' απόλυτη τιμή μεγαλύτερος του 0.8, στον Πίνακα 41.

4.6.1 Ερμηνεία 1^{ης} κύριας συνιστώσας:

Η 1^η κύρια συνιστώσα φαίνεται ότι έχει ισχυρή θετική συσχέτιση με υψηλά ποσοστά μόνιμης απασχόλησης γυναικών, υψηλά ποσοστά απασχόλησης εργατικού δυναμικού με ανώτατο μορφωτικό επίπεδο και υψηλά ποσοστά απασχόλησης ατόμων στις πλέον παραγωγικές ηλικίες, 25-54 ετών.

Συνεπώς, η 1^η κύρια συνιστώσα θα μπορούσε να χρησιμοποιηθεί ως ένας «δείκτης» για να απεικονίσει ένα εξελιγμένο κοινωνικό μοντέλο απασχόλησης. Πράγματι, παρατηρούμε ότι οι χώρες με μεγάλη τιμή στο συγκεκριμένο «δείκτη» έχουν υψηλά ποσοστά μόνιμης απασχόλησης γυναικών, υψηλά ποσοστά απασχόλησης εργατικού δυναμικού με ανώτατο μορφωτικό επίπεδο και υψηλά ποσοστά απασχόλησης ατόμων στις πλέον παραγωγικές ηλικίες, 25-54 ετών. Αντίθετα, για χώρες με αρνητικές τιμές σε αυτόν τον «δείκτη», παρατηρούνται τα χαμηλότερα ποσοστά αναφορικά με τα συγκεκριμένα χαρακτηριστικά απασχόλησης.

4.6.2 Ερμηνεία 2^{ης} κύριας συνιστώσας:

Η 2^η κύρια συνιστώσα φαίνεται ότι έχει ισχυρή θετική συσχέτιση με υψηλά ποσοστά μερικής απασχόλησης τόσο σε άντρες, όσο και σε γυναίκες.

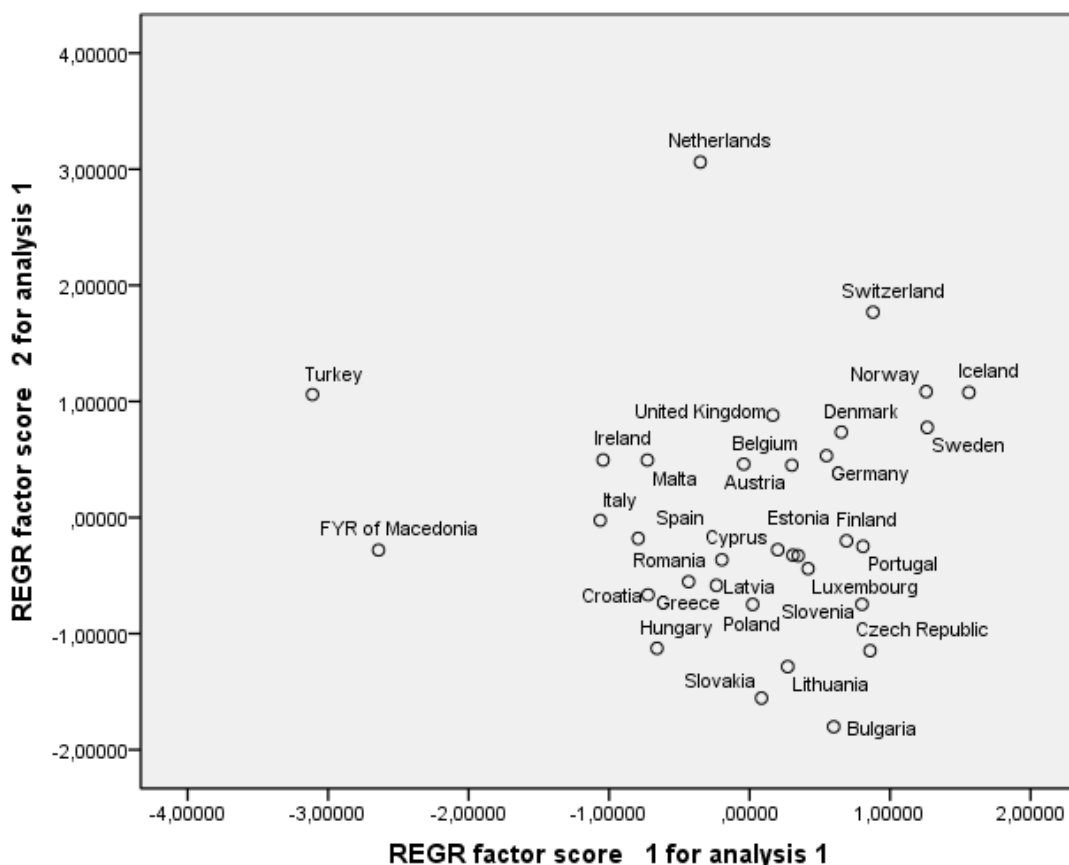
Συνεπώς, η 2^η κύρια συνιστώσα θα μπορούσε να χρησιμοποιηθεί ως ένας «δείκτης» για να απεικονίσει χώρες με υψηλά ποσοστά μερικής απασχόλησης και στα δύο φύλα. Πράγματι, παρατηρούμε ότι χώρες με μεγάλη τιμή στο συγκεκριμένο «δείκτη» έχουν υψηλά ποσοστά μερικής απασχόλησης και στα δύο φύλα. Αντίθετα, για χώρες με αρνητικές τιμές σε αυτόν τον «δείκτη», παρατηρούνται τα χαμηλότερα ποσοστά αναφορικά με τα συγκεκριμένα χαρακτηριστικά απασχόλησης.

4.7 Απεικόνιση των Ευρωπαϊκών χωρών σε σύστημα δύο αξόνων

Σε κάθε χώρα αντιστοιχούν πλέον δύο νέες μεταβλητές με τα scores f_1 και f_2 , οι οποίες μπορούν να πάρουν τη θέση των αρχικών 11 μεταβλητών απασχόλησης, διατηρώντας το 74.5% της συνολικής διακύμανσής τους. Έχουμε δηλαδή μία απώλεια πληροφορίας 25.5%, ως αντιστάθμισμα για τη συρρίκνωση των διαστάσεων του προβλήματος από 11 σε μόλις 2.

Με τον τρόπο αυτό κατορθώνουμε μία 2-διάστατη «αντιπροσώπευση» του 74.5% μίας 11-διάστατης πληροφορίας, γεγονός που επιτρέπει πλέον την εύκολη γραφική αναπαράστασή της.

Στο διάγραμμα που ακολουθεί στην επόμενη σελίδα, φαίνεται η θέση των Ευρωπαϊκών χωρών σε σύστημα δύο αξόνων, οι οποίοι αντιστοιχούν στους δύο «δείκτες» (πρώτες κύριες συνιστώσες) που ερμηνεύσαμε στην προηγούμενη ενότητα.



Διάγραμμα 23 - Γραφική αναπαράσταση των Ευρωπαϊκών χωρών βάσει των δύο «δεικτών» απασχόλησης

Όσο πιο δεξιά βρίσκεται μία χώρα, τόσο μεγαλύτερη τιμή έχει ο 1^{ος} «δείκτης» της, ο οποίος υποδηλώνει υψηλά ποσοστά μόνιμης απασχόλησης γυναικών, υψηλά ποσοστά απασχόλησης εργατικού δυναμικού με ανώτατο μορφωτικό επίπεδο και υψηλά ποσοστά απασχόλησης ατόμων στις πλέον παραγωγικές ηλικίες, 25-54 ετών.

Όσο πιο πάνω βρίσκεται μία χώρα, τόσο μεγαλύτερη τιμή έχει ο 2^{ος} «δείκτης» της, ο οποίος υποδηλώνει υψηλά ποσοστά μερικής απασχόλησης και στα δύο φύλα.

Η απεικόνιση των χωρών στις δύο διαστάσεις και η δημιουργία αποστάσεων μεταξύ τους είναι κρίσιμης σημασίας για την υλοποίηση της ανάλυσης κατά συστάδες, που στοχεύει στην ομαδοποίηση των χωρών και περιγράφεται στο επόμενο Κεφάλαιο.

4.8 ΒΙΒΛΙΟΓΡΑΦΙΑ 4^{ου} ΚΕΦΑΛΑΙΟΥ

- [1] Everitt, B. S. and Dunn, G. (1991) *Applied Multivariate Data Analysis*, Arnold, New York.
- [2] Flury, B. and Riedwyl, H. (1988) *Multivariate Statistics: A practical approach*, Prentice Hall, New York.
- [3] Johnson, R. A. and Wichern, D. W. (1998) *Applied Multivariate Statistical Analysis*, Prentice Hall, New Jersey.
- [4] Krzanowski, W. J. (1988) *Principles of Multivariate Analysis: A user's perspective*, Oxford University Press, U.K.
- [5] Manly, B. F. J. (1986) *Multivariate Statistical Methods: A primer*, Chapman and Hall, London.
- [6] Toit, S. H. C., Steyn, A. G. W. and Stumpf, R. H. (1986) *Graphical Exploratory Data Analysis*, Springer-Verlag, New York.
- [7] Κούτρας, Μάρκος (2012) *Εφαρμοσμένη Πολυμεταβλητή Ανάλυση, Σημειώσεις Τμήματος Εφαρμοσμένης Στατιστικής Πανεπιστημίου Πειραιώς*

ΚΕΦΑΛΑΙΟ 5 - ΑΝΑΛΥΣΗ ΚΑΤΑ ΣΥΣΤΑΔΕΣ ΜΕ ΤΗ ΣΥΣΣΩΡΕΥΤΙΚΗ ΜΕΘΟΔΟ

Η διαδικασία της ανάλυσης κατά συστάδες (Cluster Analysis) προϋποθέτει την ύπαρξη κάποιων μέτρων που εκφράζουν πόσο όμοιες ή ανόμοιες είναι οι παρατηρήσεις μεταξύ τους. Τα μέτρα αυτά ονομάζονται μέτρα απόστασης (distance measures) και μέτρα ομοιότητας (similarity measures). Όσο περισσότερο θεωρούμε ότι μοιάζουν οι παρατηρήσεις, τόσο μικρότερη τιμή θα πρέπει να λαμβάνει το μέτρο της απόστασής τους και μεγαλύτερη τιμή το μέτρο ομοιότητάς τους. Αντίθετα, όσο λιγότερο θεωρούμε ότι μοιάζουν οι παρατηρήσεις, τόσο μεγαλύτερη τιμή θα πρέπει να λαμβάνει το μέτρο της απόστασής τους και μικρότερη τιμή το μέτρο ομοιότητάς τους.

Η ανάλυση κατά συστάδες αποσκοπεί στη δημιουργία κατάλληλων ομάδων, ούτως ώστε οι παρατηρήσεις που ανήκουν στην ίδια ομάδα να έχουν μικρή τιμή στο μέτρο απόστασης (ή μεγάλη τιμή στο μέτρο ομοιότητας), ενώ παρατηρήσεις που ανήκουν σε διαφορετικές ομάδες να έχουν μεγάλη τιμή στο μέτρο απόστασης (ή μικρή τιμή στο μέτρο ομοιότητας).

5.1 Μέτρα απόστασης

Μία συνάρτηση $d_{ij} = d(\mathbf{x}_i, \mathbf{x}_j)$ μπορεί να δώσει ένα μέτρο απόστασης, αρκεί να ικανοποιεί τις ακόλουθες ιδιότητες:

- $d_{ij} \geq 0$ για κάθε i, j , και $d_{ij} = 0$ αν και μόνον αν $i = j$
- $d_{ij} \leq d_{is} + d_{sj}$ (τριγωνική ανισότητα)
- $d_{ij} = d_{ji}$ (συμμετρική ιδιότητα)

Το συνηθέστερο μέτρο απόστασης είναι η ευκλείδεια απόσταση, η οποία αφενός πληροί τις παραπάνω ιδιότητες και αφετέρου μπορεί εύκολα να αναπαρασταθεί γραφικά. Ωστόσο, όταν η κλίμακα των παρατηρήσεων είναι διαφορετική, συνηθίζουμε πρώτα να τυποποιούμε τις μεταβλητές, ούτως ώστε το μέτρο απόστασης να είναι περισσότερο αμερόληπτο. Το μέτρο απόστασης που παίρνουμε στην περίπτωση αυτή λέγεται «απόσταση Pearson». Άλλα μέτρα απόστασης είναι:

- Απόσταση Manhattan ή City-block metric

- Απόσταση Minkowski
- Απόσταση Max ή Chebyshev
- Απόσταση Gower
- Απόσταση Bhattacharyya
- Canberra metric

Στις περιπτώσεις δίτιμων μεταβλητών, τα μέτρα απόστασης που χρησιμοποιούνται συνήθως είναι:

- Απόσταση simple matching
- Απόσταση Rogers & Tanimoto
- Απόσταση Sokal & Sneath
- Απόσταση Jaccard
- Απόσταση Dice & Sorrensen

5.2 Μέτρα ομοιότητας

Μία συνάρτηση $s_{ij} = s(\mathbf{x}_i, \mathbf{x}_j)$ μπορεί να δώσει ένα μέτρο ομοιότητας, αρκεί να ικανοποιεί τις ακόλουθες ιδιότητες:

- $s_{ij} \geq 0$ για κάθε i, j , και αν $i = j$ συνεπάγεται $s_{ij} = 1$
- $s_{ij} \leq 1$
- $s_{ij} = s_{ji}$ (συμμετρική ιδιότητα)

Το συνηθέστερο μέτρο ομοιότητας είναι ο συντελεστής συσχέτισης. Άλλα μέτρα ομοιότητας που χρησιμοποιούνται συνήθως είναι:

- Simple matching
- Rogers & Tanimoto
- Sokal & Sneath (3 παραλλαγές)
- Jaccard coefficient
- Dice & Sorensen

- Russel & Rao
- Kulczynski

5.3 Μέθοδοι ομαδοποίησης

Οι μέθοδοι ομαδοποίησης διακρίνονται στις μη ιεραρχικές και στις ιεραρχικές:

α. Οι μη ιεραρχικές μέθοδοι ομαδοποιούν τα δεδομένα σε συγκεκριμένο (προκαθορισμένο) k αριθμό ομάδων, είτε θεωρώντας συγκεκριμένα k δεδομένα ως μητρικά σημεία και ταξινομώντας γύρω από αυτά τα υπόλοιπα δεδομένα, είτε θεωρώντας έναν αρχικό διαμερισμό των δεδομένων σε k ομάδες και αναδιατάσσοντας τα δεδομένα μεταξύ των ομάδων, μέχρι να πετύχουν τη βέλτιστη ομαδοποίηση.

β. Οι ιεραρχικές μέθοδοι δημιουργούν μία ιεραρχία με τη μορφή δέντρου, κατά την οποία είτε τα κοντινότερα δεδομένα ενώνονται αρχικά σε μικρές ομάδες και στη συνέχεια σχηματίζονται σταδιακά όλο και μεγαλύτερες ομάδες μέχρι να δημιουργηθεί μία ομάδα με όλα τα δεδομένα (συσσωρευτικές μέθοδοι), είτε οι μεγαλύτερες ομάδες διαιρούνται σε μικρότερες, μέχρις ότου όλα τα δεδομένα να αποτελούν από μόνα τους ομάδες (διαιρητικές μέθοδοι).

Η επιλογή μη ιεραρχικής μεθόδου ομαδοποίησης δεν ενδείκνυται για το πρόβλημα που μελετάμε, για τους ακόλουθους λόγους:

α. Δεν γνωρίζουμε εκ των προτέρων το βέλτιστο πλήθος k των ομάδων των Ευρωπαϊκών χωρών, βάσει των στοιχείων απασχόλησης.

β. Το γεγονός ότι έχουμε αρκετές έκτροπες παρατηρήσεις στα δεδομένα μας, όπως διαπιστώθηκε κατά την παρουσίαση των μεταβλητών απασχόλησης στο 1^ο Κεφάλαιο, ενδεχομένως να οδηγήσει στη δημιουργία ομάδων με πολύ διεσπαρμένες χώρες.

γ. Ακόμα και αν προκαθορίσουμε έναν αριθμό k ομάδων, η ομαδοποίηση των χωρών ενδεχομένως να επηρεαστεί από την επιλογή των αρχικών (μητρικών) χωρών γύρω από τις οποίες θα ταξινομηθούν οι υπόλοιπες χώρες.

Για την ομαδοποίηση των Ευρωπαϊκών χωρών θα χρησιμοποιήσουμε κάποια συσσωρευτική ιεραρχική μέθοδο. Οι δυνατές επιλογές μας είναι οι εξής:

- Μέθοδος απλής συνένωσης
- Μέθοδος πλήρους συνένωσης

- Μέθοδος σταθμισμένων μέσων
- Μέθοδος κέντρων βάρους
- Μέθοδος Ward
- Μέθοδος Grower
- Μέθοδος ευέλικτης στρατηγικής

Από τις παραπάνω μεθόδους θα επιλέξουμε τη μέθοδο Ward, η οποία στοχεύει στην ελαχιστοποίηση της διακύμανσης μέσα στις ομάδες που δημιουργούνται. Η συγκεκριμένη μέθοδος έχει μερικές πολύ καλές ιδιότητες και συνήθως χρησιμοποιείται στην πράξη συχνότερα από τις άλλες.

5.3.1 Επιλογή του πλήθους των ομάδων

Ως κριτήριο για την επιλογή του βέλτιστου πλήθους των ομάδων, στην πράξη χρησιμοποιούνται οι αποστάσεις δημιουργίας των ομάδων στο δένδρογραμμα, το οποίο που απεικονίζει τη διαδικασία της συσσωρευτικής μεθόδου ομαδοποίησης. Συγκεκριμένα, εξετάζοντας τις αποστάσεις δημιουργίας των ομάδων, εντοπίζεται το σημείο στο οποίο οι αποστάσεις αυτές γίνονται ξαφνικά πολύ μεγάλες μέχρι το ανώτερο επίπεδο ομαδοποίησης. Σε εκείνο ακριβώς το σημείο, το μέτρο απόστασης μεταξύ των ομάδων που σχηματίστηκαν θεωρείται πλέον σημαντικό. Από την άλλη, εκείνο το σημείο μπορεί να θεωρηθεί το όριο μέχρι το οποίο η κάθε ομάδα μπορεί να χαρακτηριστεί ομοιογενής.

5.4 Ομαδοποίηση βάσει των 11 αρχικών χαρακτηριστικών απασχόλησης

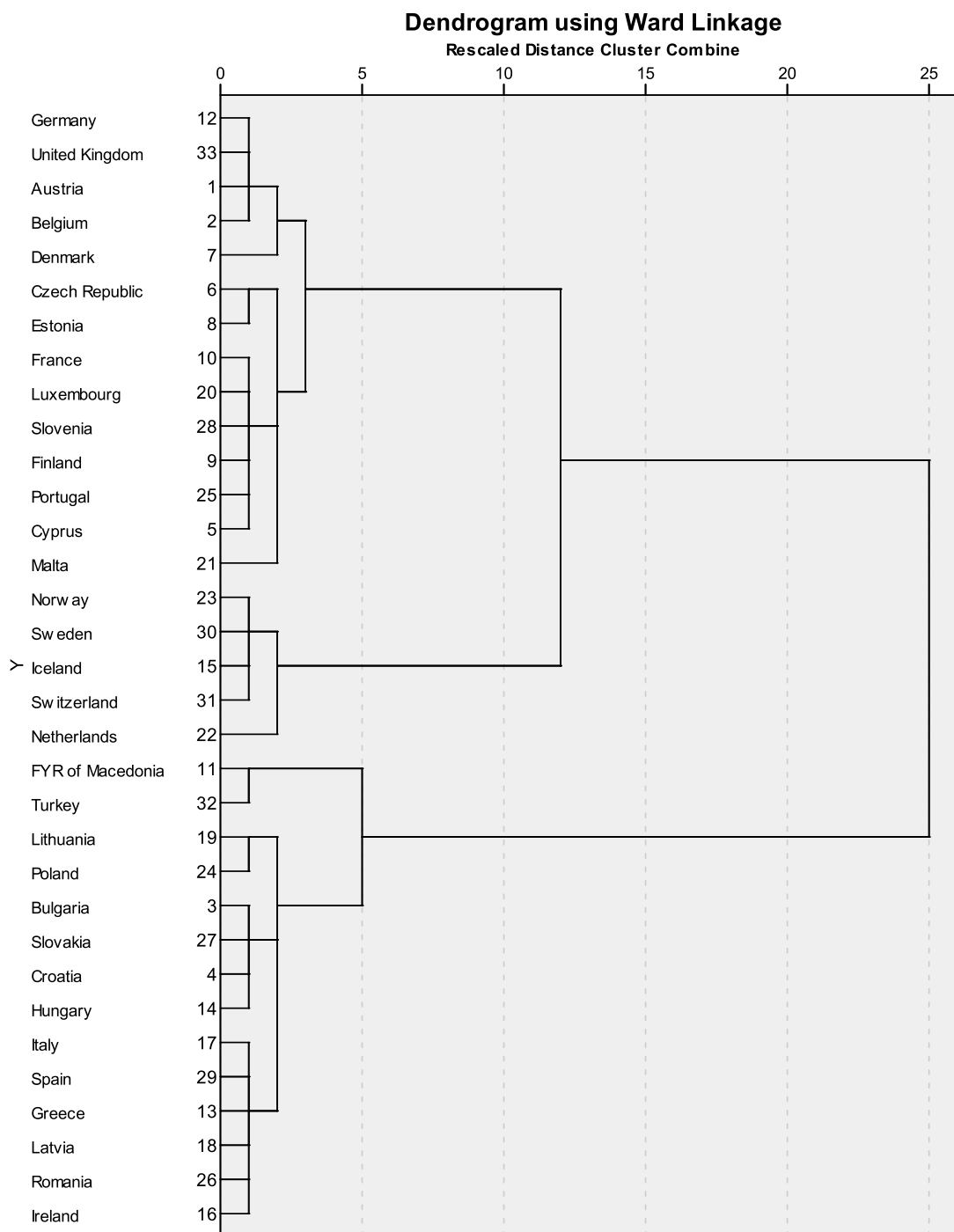
Αν για την ομαδοποίηση των Ευρωπαϊκών χωρών λάβουμε υπόψη και τα 11 χαρακτηριστικά απασχόλησης, ουσιαστικά λαμβάνουμε υπόψη το 100% της αρχικής πληροφορίας. Η συσσωρευτική διαδικασία ομαδοποίησης στην περίπτωση αυτή, χρησιμοποιώντας ως μέτρο απόστασης την απόσταση Pearson, περιγράφεται βήμα – βήμα στον πίνακα που ακολουθεί.

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	12	33	,372	0	0	7
2	23	30	1,109	0	0	15
3	17	29	1,913	0	0	8
4	10	20	2,893	0	0	11
5	3	27	3,943	0	0	19
6	4	14	4,998	0	0	19
7	1	12	6,415	0	1	16
8	13	17	8,181	0	3	22
9	18	26	10,193	0	0	12
10	9	25	12,223	0	0	17
11	10	28	14,430	4	0	20
12	16	18	16,784	0	9	22
13	15	31	19,352	0	0	15
14	19	24	22,004	0	0	26
15	15	23	25,238	13	2	27
16	1	2	28,727	7	0	23
17	5	9	32,517	0	10	20
18	11	32	36,407	0	0	30
19	3	4	40,334	5	6	24
20	5	10	44,613	17	11	25
21	6	8	49,408	0	0	28
22	13	16	54,515	8	12	24
23	1	7	61,446	16	0	29
24	3	13	69,852	19	22	26
25	5	21	78,830	20	0	28
26	3	19	88,964	24	14	30
27	15	22	99,175	15	0	31
28	5	6	109,593	25	21	29
29	1	5	122,239	23	28	31
30	3	11	149,310	26	18	32
31	1	15	212,120	29	27	32
32	1	3	352,000	31	30	0

Πίνακας 43 - Διαδικασία συσσωρευτικής σύνεωσης των Ευρωπαϊκών χωρών σε ομάδες, βάσει των 11 αρχικών χαρακτηριστικών απασχόλησης

Η παραπάνω διαδικασία ομαδοποίησης αποτυπώνεται γραφικά στο δένδρογραμμα που ακολουθεί.



Δενδρόγραμμα 1 - Απεικόνιση της συσσωρευτικής μεθόδου ομαδοποίησης των Ευρωπαϊκών χωρών, βάσει των 11 αρχικών χαρακτηριστικών απασχόλησης

Ως βέλτιστη επιλογή πλήθους ομάδων φαίνεται να είναι οι 3 ομάδες. Πράγματι, από την εικόνα που δίνει το παραπάνω δενδρόγραμμα, μέχρι να καλυφθεί η απόσταση 5 περίπου της κλίμακας συμβαίνουν συχνά συνενώσεις χωρών, ενώ η επόμενη συνένωση συμβαίνει πολύ αργότερα, σε απόσταση περίπου 12.

Οι χώρες που εντάχθηκαν στις 3 ομάδες φαίνονται στον πίνακα που ακολουθεί.

1η ΟΜΑΔΑ		2η ΟΜΑΔΑ		3η ΟΜΑΔΑ	
	Χώρα		Χώρα		Χώρα
1	1:Austria	1	3:Bulgaria	1	15:Iceland
2	2:Belgium	2	4:Croatia	2	22:Netherlands
3	5:Cyprus	3	11:FYR of Macedonia	3	23:Norway
4	6:Czech Republic	4	13:Greece	4	30:Sweden
5	7:Denmark	5	14:Hungary	5	31:Switzerland
6	8:Estonia	6	16:Ireland		
7	9:Finland	7	17:Italy		
8	10:France	8	18:Latvia		
9	12:Germany	9	19:Lithuania		
10	20:Luxembourg	10	24:Poland		
11	21:Malta	11	26:Romania		
12	25:Portugal	12	27:Slovakia		
13	28:Slovenia	13	29:Spain		
14	33:United Kingdom	14	32:Turkey		

Πίνακας 44 - Οι 3 ομάδες των Ευρωπαϊκών χωρών, βάσει των 11 αρχικών χαρακτηριστικών απασχόλησης

Τα παραπάνω αποτελέσματα προέκυψαν με τη χρήση του 100% της αρχικής πληροφορίας και θα χρησιμοποιηθούν ως σημείο αναφοράς, προκειμένου να ελεγχθεί η αξιοπιστία της ομαδοποίησης βάσει των 2 «δεικτών» απασχόλησης που κατασκευάσαμε στο 2^ο Κεφάλαιο.

5.5 Ομαδοποίηση βάσει των 2 «δεικτών» απασχόλησης

Αν για την ομαδοποίηση των Ευρωπαϊκών χωρών λάβουμε υπόψη τους 2 «δείκτες» απασχόλησης, ουσιαστικά λαμβάνουμε υπόψη το 74,5% της αρχικής πληροφορίας. Ωστόσο, θα ελέγξουμε κατά πόσο τα αποτελέσματα που θα προκύψουν τώρα, θα συμφωνήσουν με τα αποτελέσματα της προηγούμενης ενότητας, όταν για την ομαδοποίηση των χωρών χρησιμοποιήθηκε το 100% της αρχικής πληροφορίας.

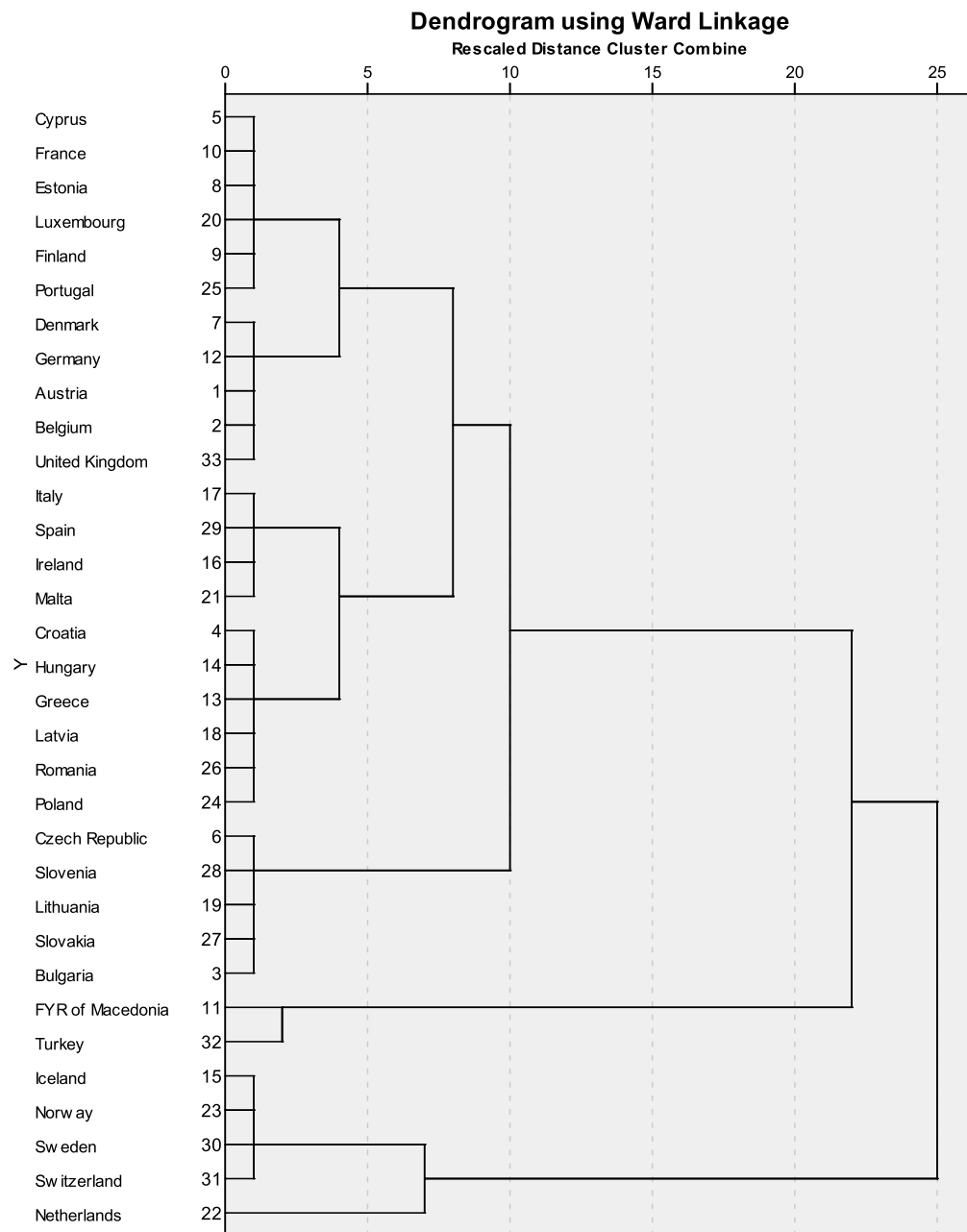
Η συσσωρευτική διαδικασία ομαδοποίησης περιγράφεται βήμα – βήμα στον πίνακα που ακολουθεί.

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	5	10	,001	0	0	3
2	9	25	,009	0	0	20
3	5	8	,021	1	0	5
4	13	18	,041	0	0	7
5	5	20	,067	3	0	20
6	7	12	,093	0	0	19
7	13	26	,134	4	0	16
8	15	23	,180	0	0	13
9	17	29	,229	0	0	21
10	16	21	,279	0	0	21
11	19	27	,333	0	0	18
12	1	2	,391	0	0	17
13	15	30	,467	8	0	23
14	6	28	,549	0	0	24
15	4	14	,657	0	0	22
16	13	24	,776	7	0	22
17	1	33	,898	12	0	19
18	3	19	1,113	0	11	24
19	1	7	1,367	17	6	26
20	5	9	1,633	5	2	26
21	16	17	1,991	10	9	27
22	4	13	2,445	15	16	27
23	15	31	3,087	13	0	28
24	3	6	3,833	18	14	30
25	11	32	4,841	0	0	31
26	1	5	7,176	19	20	29
27	4	16	9,680	22	21	29
28	15	22	14,546	23	0	32
29	1	4	20,631	26	27	30
30	1	3	27,970	29	24	31
31	1	11	44,770	30	25	32
32	1	15	64,000	31	28	0

Πίνακας 45 - Διαδικασία συσσωρευτικής συνένωσης των Ευρωπαϊκών χωρών σε ομάδες, βάσει των 2 «δεικτών» απασχόλησης

Η παραπάνω διαδικασία ομαδοποίησης αποτυπώνεται γραφικά στο δενδρόγραμμα που ακολουθεί.



Δενδρόγραμμα 2 - Απεικόνιση της συσσωρευτικής μεθόδου ομαδοποίησης των Ευρωπαϊκών χωρών, βάσει των 2 «δεικτών» απασχόλησης

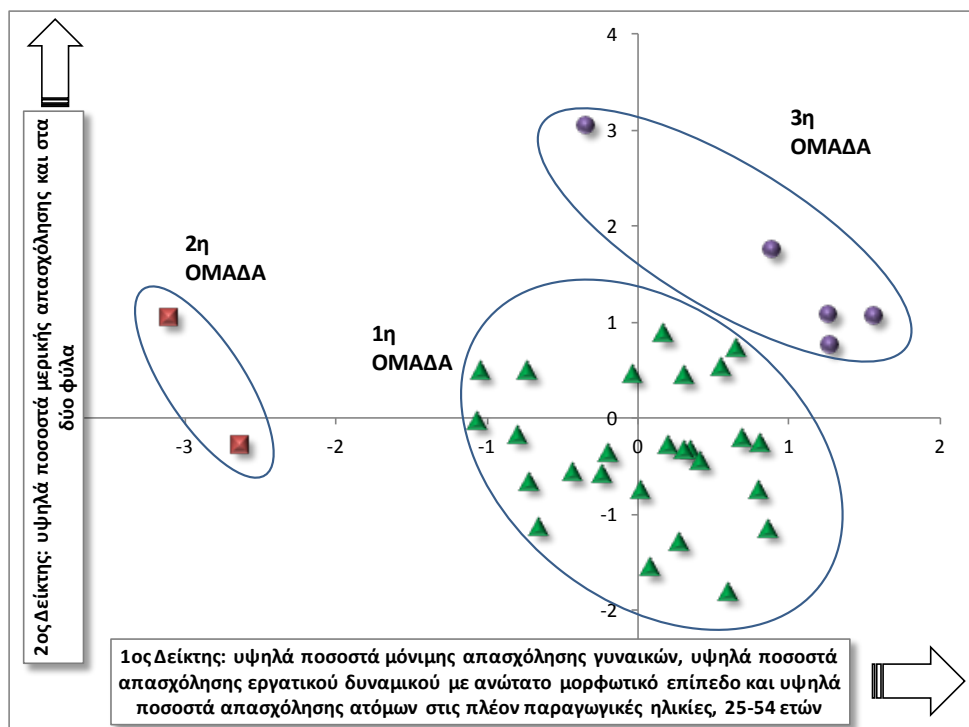
Ως βέλτιστη επιλογή πλήθους ομάδων φαίνεται να είναι οι 3 ομάδες. Πράγματι, από την εικόνα που δίνει το παραπάνω δενδρόγραμμα, μέχρι να καλυφθεί η απόσταση 10 περίπου της κλίμακας συμβαίνουν συχνά συνενώσεις χωρών μέχρι να σχηματιστούν 3 ομάδες, ενώ η επόμενη συνένωση συμβαίνει πολύ αργότερα, σε απόσταση περίπου 22.

1η ΟΜΑΔΑ		2η ΟΜΑΔΑ		3η ΟΜΑΔΑ	
		Χώρα		Χώρα	
1	1:Austria	1	11:FYR of Macedonia	1	15:Iceland
2	2:Belgium	2	32:Turkey	2	22:Netherlands
3	3:Bulgaria			3	23:Norway
4	4:Croatia			4	30:Sweden
5	5:Cyprus			5	31:Switzerland
6	6:Czech Republic				
7	7:Denmark				
8	8:Estonia				
9	9:Finland				
10	10:France				
11	12:Germany				
12	13:Greece				
13	14:Hungary				
14	16:Ireland				
15	17:Italy				
16	18:Latvia				
17	19:Lithuania				
18	20:Luxembourg				
19	21:Malta				
20	24:Poland				
21	25:Portugal				
22	26:Romania				
23	27:Slovakia				
24	28:Slovenia				
25	29:Spain				
26	33:United Kingdom				

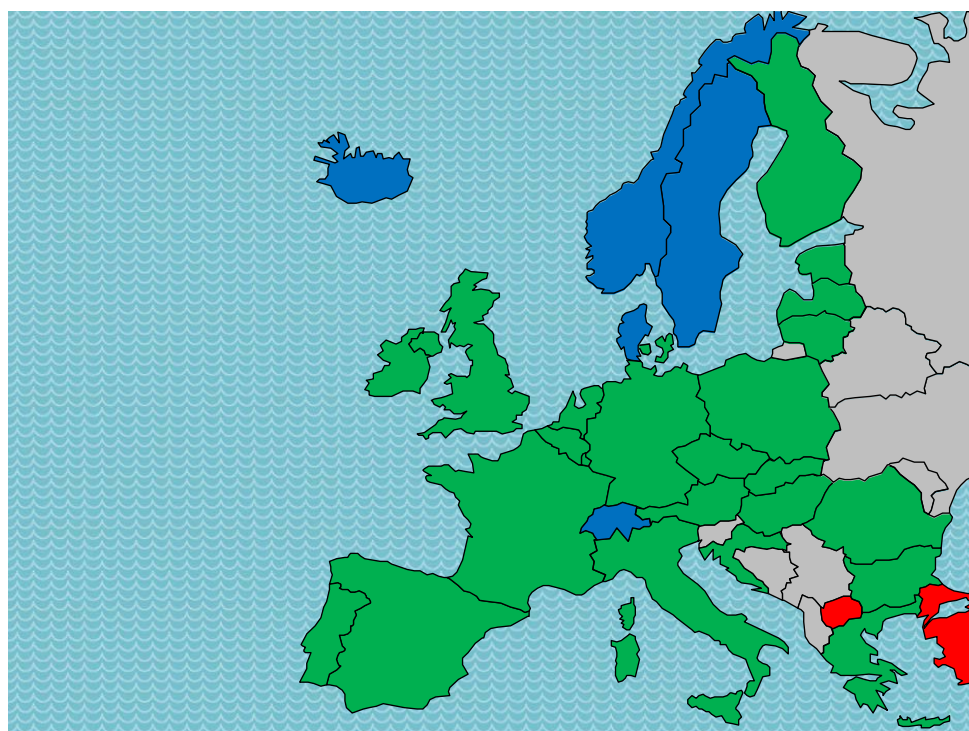
Πίνακας 46 - Οι 3 ομάδες των Ευρωπαϊκών χωρών, βάσει των 2 «δεικτών» απασχόλησης

Παρατηρούμε ότι σχεδόν όλες οι χώρες της 2^{ης} ομάδας, η οποία είχε προκύψει βάσει του 100% της αρχικής πληροφορίας, έχουν ενσωματωθεί τώρα στην 1^η ομάδα. Στη 2^η ομάδα παραμένουν μόνο η FYROM και η Τουρκία, οι οποίες φαίνεται να διατηρούν σημαντικές διαφοροποιήσεις από τις υπόλοιπες χώρες που μετακινήθηκαν στην 1^η ομάδα. Τέλος, η σύνθεση της 3^{ης} ομάδας παραμένει αμετάβλητη.

Συνεπώς, αν και το βέλτιστο πλήθος ομάδων παραμένει 3, η εφαρμογή της PCA οδήγησε στην αλλαγή του πληθάριμου των ομάδων από 14-14-5 σε 26-2-5. Ουσιαστικά, χρησιμοποιώντας το 74,5% της αρχικής πληροφορίας, καταλήξαμε στο συμπέρασμα ότι η μεγάλη πλειοψηφία των Ευρωπαϊκών χωρών (26 από τις 33) έχουν μεταξύ τους παρόμοια χαρακτηριστικά απασχόλησης. Η ομαδοποίηση που προέκυψε, φαίνεται στο διάγραμμα που ακολουθεί.



Διάγραμμα 24 - Γραφική αναπαράσταση των 3 ομάδων Ευρωπαϊκών χωρών βάσει των 2 «δεικτών» απασχόλησης



Χάρτης 1 – Χάρτης των 3 ομάδων Ευρωπαϊκών χωρών βάσει των 2 «δεικτών» απασχόλησης

5.6 ΒΙΒΛΙΟΓΡΑΦΙΑ 5^{ου} ΚΕΦΑΛΑΙΟΥ

- [1] Aldenderfer, M. S., Blashfield, R. K. (1984). *Cluster Analysis*, Sage Publ., Beverly Hills and London.
- [2] Anderberg, M. R. (1973). *Cluster analysis for applications*, Academic press, New York.
- [3] Everitt, B.(1981). *Cluster Analysis*, Heinemann Educational Books, London.
- [4] Gordon, A. D. (1999). *Classification (2nd ed.)*, Chapman and Hall, London.
- [5] Kaufman, L. and Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York.
- [6] Κούτρας, Μάρκος (2012) *Εφαρμοσμένη Πολυμεταβλητή Ανάλυση*, Σημειώσεις Τμήματος Εφαρμοσμένης Στατιστικής Πανεπιστημίου Πειραιώς

ΚΕΦΑΛΑΙΟ 6 - ΣΥΜΠΕΡΑΣΜΑΤΑ

6.1 Σύγκριση ομάδων που προέκυψαν από τις δύο ομαδοποιήσεις

Η αρχική ομαδοποίηση με αξιοποίηση του 100% της διαθέσιμης πληροφορίας οδήγησε στο παρακάτω σετ ομάδων:



Πίνακας 47 – Σύγκριση των ομάδων που προκύπτουν από την ομαδοποίηση των Ευρωπαϊκών χωρών βάσει των 11 αρχικών χαρακτηριστικών απασχόλησης (αριστερά) και των 2 PCA δεικτών απασχόλησης (δεξιά)

Η ίδια ιεραρχική μέθοδος ομαδοποίησης χρησιμοποιήθηκε και στις δύο περιπτώσεις και οδήγησε στο ίδιο πλήθος ομάδων (στις ιεραρχικές μεθόδους δεν προδιαγράφεται το πλήθος των ομάδων), με διαφορετικό όμως πληθάρημο.

6.1.1 Ομαδοποίηση βάσει των 11 αρχικών χαρακτηριστικών απασχόλησης

- Ομάδα 1
 - Κατά κύριο λόγο χώρες της κεντρικής Ευρώπης
- Ομάδα 2
 - Μεσογειακές χώρες
 - Βαλκάνια
 - Ανατολική Ευρώπη
- Ομάδα 3
 - Σκανδιναβία μείον τη Φινλανδία
 - Ελβετία
 - Ισλανδία

Η ομαδοποίηση φαίνεται διαισθητικά σωστή, καθώς η κεντρική Ευρώπη είναι αναμενόμενο να εμφανίζει κοινά χαρακτηριστικά απασχόλησης, ως επίσης και η Μεσόγειος, τα Βαλκάνια και η Ανατολική Ευρώπη είναι λογικό να μπαίνουν στην ίδια ομάδα. Οι χώρες Δανία, Σουηδία, Νορβηγία, Ισλανδία, Ελβετία τέλος αποτελούν μια ιδιαίτερη ομάδα με συγκεκριμένα χαρακτηριστικά απασχόλησης που τις διαφοροποιούν από τις άλλες δύο ομάδες.

Μπορεί επίσης να ανιχνευτεί η υψηλή συσχέτιση της συγκεκριμένης ομαδοποίησης με την κατάταξη των Ευρωπαϊκών χωρών σε οικονομικοκοινωνικά:

- Ευημερούσες (Ομάδα 3)
- Σταθερές (Ομάδα 1)
- Αναπτυσσόμενες (Ομάδα 2)

6.1.2 Ομαδοποίηση βάσει των 2 δεικτών απασχόλησης

- Ομάδα 1
 - Όλες οι χώρες της Ευρώπης, πλην των παρακάτω:
- Ομάδα 2
 - FYROM
 - Τουρκία
- Ομάδα 3
 - Σκανδιναβία μείον τη Φινλανδία
 - Ελβετία
 - Ισλανδία

Η ομαδοποίηση διακρίνει τις χώρες (Δανία, Σουηδία, Νορβηγία, Ισλανδία, Ελβετία) που αποτελούν μια ιδιαίτερη ομάδα με συγκεκριμένα χαρακτηριστικά απασχόλησης, καθώς και τις χώρες (FYROM, Τουρκία) που επίσης διαφοροποιούνται -προς την αντίθετη κατεύθυνση- ως προς τα χαρακτηριστικά απασχόλησης. Όλες οι υπόλοιπες Ευρωπαϊκές χώρες εντάσσονται στην ίδια υπερ-ομάδα.

Η παραπάνω ομαδοποίηση έγινε με χρήση μόνο των δύο δεικτών απασχόλησης που κατασκευάστηκαν μετά την PCA στην παράγραφο 3.6 της παρούσης εργασίας. Οι δύο δείκτες αποτυπώνουν το 74,5% της αρχικής πληροφορίας, μειώνοντας ταυτόχρονα το πλήθος των μεταβλητών από έντεκα σε μόλις δύο.

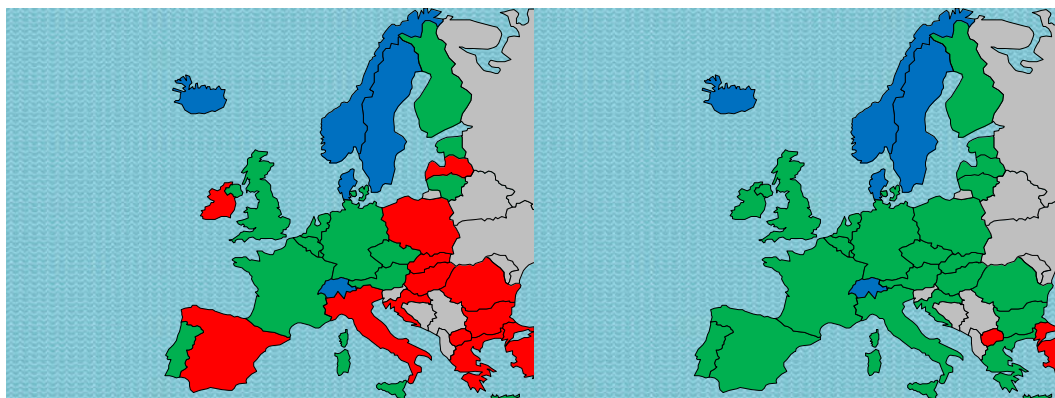
Αυτό είναι το όφελος της PCA που πραγματοποιήθηκε, ότι συμπίεσε τον αρχικό αριθμό μεταβλητών σε δύο, με αποτέλεσμα την εύκολη και εύληπτη γραφική αναπαράσταση (Διάγραμμα 24) σε ένα σύστημα δύο καθέτων αξόνων.

Το προφανές αρνητικό επακόλουθο της PCA είναι πως απωλέστηκε το 25,5% της αρχικής πληροφορίας, το οποίο ήταν αρκετό για να μεταβάλλει την αρχική διασθητικά συμβατή με τις προσδοκίες μας ομαδοποίηση. Η νέα ομαδοποίηση δεν κατορθώνει να διαχωρίσει παρά μόνο τα ακραία απκλίνοντα ως προς τα χαρακτηριστικά απασχόλησης υποσύνολα Ευρωπαϊκών χωρών, αφήνοντας 26 από τις 33 Ευρωπαϊκές χώρες που μετέχουν στη μελέτη στην ίδια ομάδα.

Το παραπάνω trade-off είναι απολύτως αναμενόμενο και εναπόκειται στον εκάστοτε μελετητή να αποφασίσει αν προτιμά να θυσιάσει (σε ελεγχόμενο φυσικά βαθμό) την ακρίβεια της μελέτης του προκειμένου να μπορέσει να μεταβεί από ένα δύσκολο διαχείρισιμο και μη απεικονίσιμο αρχικό αριθμό χαρακτηριστικών απασχόλησης σε ένα μικρό πλήθος δεικτών απασχόλησης που διευκολύνουν και επιταχύνουν την απεικόνιση, κατανόηση και περεταίρω μελέτη του προβλήματος.

Στο χάρτη που ακολουθεί, μπορεί κανείς εποπτικά να αντιληφθεί τη «μετακίνηση» των χωρών ανάμεσα στις δύο ομαδοποιήσεις:

- Ομάδα 1: Πράσινο χρώμα
- Ομάδα 2: Κόκκινο χρώμα
- Ομάδα 3: Μπλε χρώμα
- Χώρες που δε μετέχουν στη μελέτη: Γκρι χρώμα



Χάρτης 2 - Μετακίνηση χωρών σε ομάδες ανάμεσα στην ομαδοποίηση με βάση τα αρχικά χαρακτηριστικά απασχόλησης (αριστερά) και τους PCA δείκτες απασχόλησης (δεξιά)

6.2 Επανάληψη ομαδοποίησης με διατήρηση τριών PCA κυρίων συνιστωσών

Η χρήση 3 κύριων συνιστωσών αποτυπώνει τα ακόλουθα ποσοστά από την αρχική πληροφορία της κάθε μεταβλητής:

Communalities		
	Initial	Extraction
Employment rates by gender, MALE	1,000	,922
Employment rates by gender, FEMALE	1,000	,899
Pre-primary, primary & lower secondary - ISCED levels 0-2	1,000	,788
Upper secondary & post-secondary non-tertiary - ISCED levels 3-4	1,000	,892
Tertiary - ISCED levels 5-6	1,000	,825
15-24 years old	1,000	,726
25-54 years old	1,000	,837
55-64 years old	1,000	,758
Persons with a second job (% of total employment)	1,000	,656
Male working part-time (% of total employment)	1,000	,850
Female working part-time (% of total employment)	1,000	,855

Extraction Method: Principal Component Analysis.

Πίνακας 48 - Ποσοστά κάλυψης της αρχικής διακύμανσης ανά μεταβλητή, από τις δύο πρώτες κύριες συνιστώσες

Αν συγκρίνει κανείς τον παραπάνω πίνακα με τον Πίνακα 38 της παρούσης εργασίας, θα διαπιστώσει τη βελτίωση που επιτυγχάνεται ανά μεταβλητή, με την αξιοποίηση 3^{ης} συνιστώσας. Συνολικά, το ποσοστό της αρχικής πληροφορίας που διατηρείται, ανάλογα με τη χρήση συγκεκριμένου πλήθους κύριων συνιστωσών, φαίνεται παρακάτω. Η χρήση 3 κύριων συνιστωσών αποτυπώνει τώρα το 81,9% της αρχικής πληροφορίας.

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	7,164	65,124	65,124	7,164	65,124	65,124	3,247	29,516	29,516
2	1,032	9,380	74,504	1,032	9,380	74,504	2,958	26,894	56,410
3	,811	7,371	81,875	,811	7,371	81,875	2,801	25,465	81,875
4	,556	5,058	86,933						
5	,464	4,220	91,153						
6	,368	3,343	94,496						
7	,194	1,767	96,263						
8	,141	1,284	97,547						
9	,123	1,117	98,664						
10	,082	,747	99,411						
11	,065	,589	100,000						

Extraction Method: Principal Component Analysis.

Πίνακας 49 - Ερμηνευτική ικανότητα των 3 πρώτων κύριων συνιστωσών πριν και μετά την περιστροφή

Οι συντελεστές των χαρακτηριστικών απασχόλησης τώρα είναι (βλ. πίνακα 41 της διπλωματικής):

Rotated Component Matrix^a

	Component		
	1	2	3
Employment rates by gender, MALE	,296	,346	,845
Employment rates by gender, FEMALE	,829	,312	,337
Pre-primary, primary & lower secondary - ISCED levels 0-2	,283	,341	,769
Upper secondary & post-secondary non-tertiary - ISCED levels 3-4	,503	,265	,754
Tertiary - ISCED levels 5-6	,722	,215	,508
15-24 years old	,375	,722	,253
25-54 years old	,853	,163	,288
55-64 years old	,672	,541	,115
Persons with a second job (% of total employment)	,490	,596	,247
Male working part-time (% of total employment)	,225	,853	,268
Female working part-time (% of total employment)	,084	,765	,512

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

a. Rotation converged in 9 iterations.

Πίνακας 50 - Συντελεστές χαρακτηριστικών απασχόλησης με τις τρεις πρώτες κύριες συνιστώσες, μετά την περιστροφή των αξόνων

6.3 Ερμηνεία των πρώτων κύριων συνιστωσών και δημιουργία «δεικτών» απασχόλησης

Στη συνέχεια θα προσπαθήσουμε να ερμηνεύσουμε τις πρώτες κύριες 96 συνιστώσες, λαμβάνοντας υπόψη και θεωρώντας σημαντικές μόνο εκείνες τις

περιπτώσεις που οι συντελεστής συσχέτισης είναι κατ' απόλυτη τιμή μεγαλύτερος του 0.67, στον Πίνακα 50.

6.3.1 Ερμηνεία 1^{ης} κύριας συνιστώσας:

Η 1^η κύρια συνιστώσα φαίνεται ότι έχει ισχυρή θετική συσχέτιση με υψηλά ποσοστά μόνιμης απασχόλησης γυναικών, υψηλά ποσοστά απασχόλησης εργατικού δυναμικού με ανώτατο μορφωτικό επίπεδο και υψηλά ποσοστά απασχόλησης ατόμων στις ηλικίες 25-54 και 55-64 ετών.

Συνεπώς, η 1^η κύρια συνιστώσα θα μπορούσε και πάλι να χρησιμοποιηθεί ως ένας «δείκτης» για να απεικονίσει ένα εξελιγμένο κοινωνικό μοντέλο απασχόλησης.

6.3.2 Ερμηνεία 2^{ης} κύριας συνιστώσας:

Η 2^η κύρια συνιστώσα φαίνεται ότι έχει ισχυρή θετική συσχέτιση με υψηλά ποσοστά μερικής απασχόλησης τόσο σε άντρες, όσο και σε γυναίκες.

Συνεπώς, η 2^η κύρια συνιστώσα θα μπορούσε να χρησιμοποιηθεί ως ένας «δείκτης» για να απεικονίσει ένα ευμετάβλητο, περιστασιακό, καταχρηστικό κοινωνικό σύστημα απασχόλησης.

6.3.3 Ερμηνεία 3^{ης} κύριας συνιστώσας:

Η 3^η κύρια συνιστώσα φαίνεται ότι έχει ισχυρή θετική συσχέτιση με υψηλά ποσοστά μόνιμης απασχόλησης ανδρών και με υψηλά ποσοστά απασχόλησης εργατικού δυναμικού με κατώτερο και μέσο μορφωτικό επίπεδο.

Συνεπώς, η 3^η κύρια συνιστώσα θα μπορούσε να χρησιμοποιηθεί ως ένας «δείκτης» για να απεικονίσει ένα σταθερό αλλά μάλλον αναχρονιστικό σύστημα απασχόλησης, όπου εργάζονται κυρίως οι άνδρες σε εργασίες που κατά κύριο λόγο δεν προϋποθέτουν ανώτερο μορφωτικό επίπεδο.

6.4 Ομαδοποίηση βάσει των 3 «δεικτών» απασχόλησης

Αν για την ομαδοποίηση των Ευρωπαϊκών χωρών λάβουμε υπόψη τους 3 «δείκτες» απασχόλησης, ουσιαστικά λαμβάνουμε υπόψη το 81,9% της αρχικής πληροφορίας. Και πάλι θα ελέγξουμε κατά πόσο τα αποτελέσματα που θα προκύπτουν τώρα, συμφωνούν με τα αποτελέσματα της ομαδοποίησης των χωρών με αξιοποίηση του 100% της αρχικής πληροφορίας.

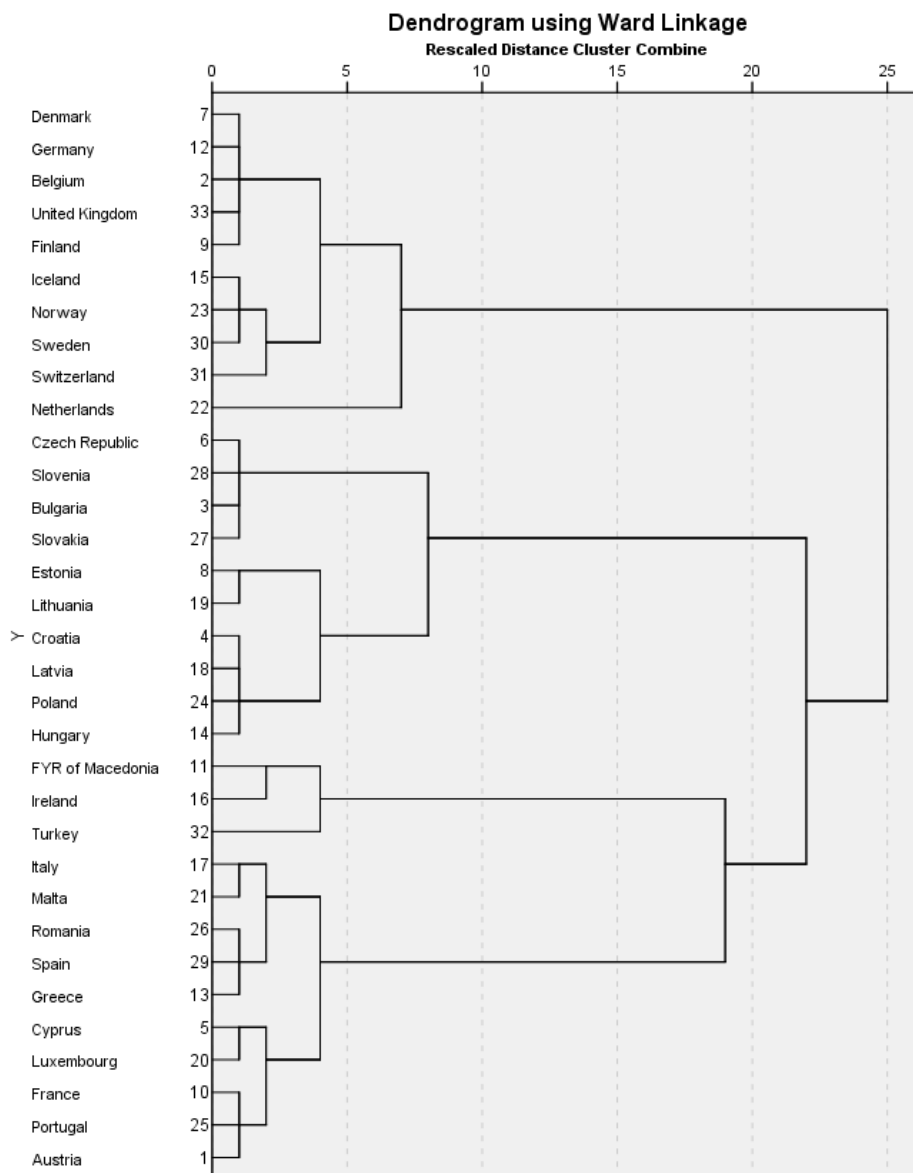
Η συσσωρευτική διαδικασία ομαδοποίησης στην περίπτωση αυτή, χρησιμοποιώντας ως μέτρο απόστασης την απόσταση Pearson, περιγράφεται βήμα – βήμα στον πίνακα που ακολουθεί:

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	7	12	,059	0	0	16
2	15	23	,137	0	0	3
3	15	30	,228	2	0	20
4	6	28	,333	0	0	18
5	4	18	,468	0	0	12
6	2	33	,634	0	0	16
7	10	25	,804	0	0	14
8	26	29	1,010	0	0	13
9	5	20	1,244	0	0	22
10	3	27	1,504	0	0	18
11	17	21	1,771	0	0	21
12	4	24	2,060	5	0	17
13	13	26	2,439	0	8	21
14	1	10	2,895	0	7	22
15	8	19	3,476	0	0	25
16	2	7	4,071	6	1	19
17	4	14	4,748	12	0	25
18	3	6	5,544	10	4	29
19	2	9	6,489	16	0	24
20	15	31	7,464	3	0	24
21	13	17	9,014	13	11	27
22	1	5	10,593	14	9	27
23	11	16	12,179	0	0	26
24	2	15	14,925	19	20	28
25	4	8	17,867	17	15	29
26	11	32	20,949	23	0	30
27	1	13	24,292	22	21	30
28	2	22	30,288	24	0	32
29	3	4	37,126	18	25	31
30	1	11	53,933	27	26	31
31	1	3	73,582	30	29	32
32	1	2	96,000	31	28	0

Πίνακας 51 - Διαδικασία συσσωρευτικής συνένωσης των Ευρωπαϊκών χωρών σε ομάδες, βάσει των 3 «δεικτών» απασχόλησης

Η παραπάνω διαδικασία ομαδοποίησης αποτυπώνεται γραφικά στο δενδρόγραμμα που ακολουθεί:



Δενδρόγραμμα 3 - Απεικόνιση της συσσωρευτικής μεθόδου ομαδοποίησης των Ευρωπαϊκών χωρών, βάσει των 3 πρώτων κυρίων συνιστωσών

Ως βέλτιστη επιλογή πλήθους ομάδων φαίνεται να είναι οι 4 ομάδες. Πράγματι, από την εικόνα που δίνει το παραπάνω δενδρόγραμμα, μέχρι να καλυφθεί η απόσταση 8 περίπου της κλίμακας συμβαίνουν συχνά συνενώσεις χωρών μέχρι να σχηματιστούν 4 ομάδες, ενώ η επόμενη συνένωση συμβαίνει πολύ αργότερα, σε απόσταση περίπου 19.

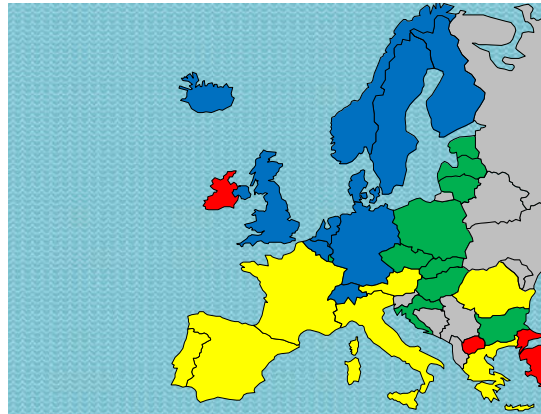
1η ΟΜΑΔΑ		2η ΟΜΑΔΑ		3η ΟΜΑΔΑ		4η ΟΜΑΔΑ	
	Χώρα		Χώρα		Χώρα		Χώρα
1	Denmark	1	Czech Republic	1	FYROM	1	Italy
2	Germany	2	Slovenia	2	Ireland	2	Malta
3	Belgium	3	Bulgaria	3	Turkey	3	Romania
4	United Kingdom	4	Slovakia			4	Spain
5	Finland	5	Estonia			5	Greece
6	Iceland	6	Lithuania			6	Cyprus
7	Norway	7	Coratia			7	Luxembourg
8	Sweden	8	Latvia			8	France
9	Switzerland	9	Poland			9	Portugal
10	Netherlands	10	Hungary			10	Austria

Πίνακας 52 - Οι 4 ομάδες των Ευρωπαϊκών χωρών, βάσει των 3 «δεικτών» απασχόλησης

Παρατηρούμε ότι πλέον η μέθοδος οδηγεί σε 4 ομάδες χωρών:

- Ομάδα 1
 - Η κεντρική και βόρεια Ευρώπη
- Ομάδα 2
 - Η ανατολική Ευρώπη
- Ομάδα 3
 - FYROM
 - Τουρκία
 - Ιρλανδία
- Ομάδα 4
 - Μεσόγειος
 - Πορτογαλία
 - Αυστρία

Στο χάρτη που ακολουθεί απεικονίζεται η νέα ομαδοποίηση με αξιοποίηση των 3 πρώτων συνιστωσών της PCA:



Χάρτης 3 - Χάρτης των 4 ομάδων Ευρωπαϊκών χωρών βάσει των 3 «δεικτών» απασχόλησης

Η ομαδοποίηση που προέκυψε δε φαίνεται απόλυτα σωστή διαισθητικά, καθώς:

- Εντάσσει στο μπλοκ των αναπτυσσομένων χωρών την Ιρλανδία.
- Δεν εντάσσει τη Γαλία και την Αυστρία στο μπλοκ της κεντρικής Ευρώπης.
- Εντάσσει τη Ρουμανία στο μπλοκ των Μεσογειακών χωρών.

Μολαταύτα αποτελεί βελτίωση σε σχέση με την ομαδοποίηση με αξιοποίηση των δύο κυρίων συνιστωσών της PCA, στην οποία όπως είδαμε οι 26 από τις 33 ευρωπαϊκές χώρες εντάσσονται στην ίδια ομάδα.

Κατά συνέπεια, επιβεβαιώνεται η αίσθησή μας πως όταν η αξιοποιούμενη για την ομαδοποίηση πληροφορία ανεβαίνει από το 74,5% στο 81,9% το αποτέλεσμα της διαδικασίας πλησιάζει περισσότερο σε αυτό της αρχικής ομαδοποίησης, όπου αξιοποιήθηκε το 100% της διαθέσιμης αρχικής πληροφορίας.

6.5 Αξία της ομαδοποίησης για ένα χρηματοπιστωτικό οργανισμό

Δεδομένου του πολύ μεγάλου μεγέθους ορισμένων χρηματοπιστωτικών ιδρυμάτων που δραστηριοποιούνται στον ευρωπαϊκό χώρο, μα και των διαδικασιών συγχώνευσης στις οποίες η οικονομική κρίση έχει δράσει μάλλον καταλυτικά υπέρ, καθίσταται προφανές πως τα όρια των τραπεζικών ομίλων μπορούν να υπερβούν αυτά των χωρών.

Τέτοιοι οργανισμοί είναι πιθανό να αναπτύσσουν προϊόντα και υπηρεσίες που απευθύνονται σε ομάδες χωρών με βάση τα ιδιαίτερα χαρακτηριστικά τους και την αγοραστική δύναμη των πολιτών τους, που σχετίζονται άμεσα με τις συνθήκες απασχόλησης.

Κατά συνέπεια, είναι ιδιαίτερα χρήσιμη η a priori γνώση των ομάδων χωρών με συναφή χαρακτηριστικά απασχόλησης, ώστε να μπορούν οι χρηματοπιστωτικοί οργανισμοί να αναπτύξουν στοχευμένες λύσεις για κάθε ομάδα χωρών.

Δεχόμενοι την ομαδοποίηση με αξιοποίηση των 2 δεικτών απασχόλησης για παράδειγμα, θα μπορούσε ένας οργανισμός να προσφέρει προϊόντα/υπηρεσίες:

- Στην ομάδα των Σκανδιναβικών χωρών - Ελβετίας – Ισλανδίας:
 - Ιδιώτες:
 - Αποταμιευτικά
 - Επενδυτικά
 - Τραπεζοασφαλιστικά
 - Πιστωτικές κάρτες
 - Επιχειρήσεις:
 - Επενδυτικά
 - Δάνεια (για επεκτάσεις, βελτιώσεις)
 - Συμβουλευτικές υπηρεσίες
- Στην ομάδα FYROM – Τουρκίας:
 - Ιδιώτες:
 - Δάνεια καταναλωτικά/στεγαστικά
 - Χρηματοδότηση επιχειρηματικότητας
 - Επιχειρήσεις:
 - Δάνεια (για εξοπλισμούς, πάγιες ανάγκες)
 - Χρηματοδότηση επιχειρηματικότητας
- Στην ομάδα των υπολοίπων Ευρωπαϊκών χωρών:
 - Ιδιώτες
 - Μίγμα προϊόντων
 - Επιχειρήσεις
 - Μίγμα προϊόντων

ΣΥΝΟΨΗ

Με την παρούσα διπλωματική εργασία επιχειρήσαμε την κατηγοριοποίηση των Ευρωπαϊκών χωρών σε ομοιογενείς ομάδες, έχοντας ως κριτήρια ομαδοποίησης 2 «δείκτες» τους οποίους κατασκευάσαμε μετά από κατάλληλη ανάλυση 11 μεταβλητών απασχόλησης της Eurostat. Η ομαδοποίηση έγινε με στόχο να προτείνουμε σε έναν χρηματοπιστωτικό οργανισμό εκείνες τις πολιτικές δραστηριοποίησης που ταιριάζουν καλύτερα στα χαρακτηριστικά κάθε ομάδας και συνεπώς μεγιστοποιούν τις πιθανότητες κερδοφορίας του οργανισμού.

Η ομαδοποίηση των χωρών υλοποιήθηκε με τη χρήση κατάλληλων μεθόδων πολυμεταβλητής στατιστικής ανάλυσης, μέσα από τα ακόλουθα βήματα:

α. Επιλέξαμε 11 μεταβλητές απασχόλησης της Eurostat, ως μεγέθη μέτρησης του εργασιακού τοπίου σε κάθε χώρα.

β. Με την εφαρμογή της ανάλυσης κύριων συνιστωσών, μειώσαμε το αρχικό πλήθος των μεγεθών μέτρησης από έντεκα σε μόλις δύο μεταβλητές, δεχόμενοι ένα «κόστος» απώλειας 25.5% από την αρχική πληροφορία.

γ. Δώσαμε στις δύο νέες μεταβλητές το ρόλο «δεικτών» απασχόλησης.

δ. Με την εφαρμογή της συσσωρευτικής μεθόδου ανάλυσης κατά συστάδες, κατηγοριοποιήσαμε τις 33 Ευρωπαϊκές χώρες σε 3 ομάδες, χρησιμοποιώντας ως μέτρα τη θέση των χωρών ως προς τους δύο νέους «δείκτες» που κατασκευάσαμε.

ε. Απεικονίσαμε σε ένα σύστημα δύο αξόνων τις Ευρωπαϊκές χώρες και τις 3 ομάδες που σχηματίστηκαν.

Η 1^η ομάδα περιλαμβάνει την πλειοψηφία των χωρών (26 χώρες), η 2^η μόλις δύο χώρες (Τουρκία και FYROM) και η 3^η πέντε χώρες (Νορβηγία, Σουηδία, Ισλανδία, Ελβετία, Ολλανδία).

Το αποτέλεσμα της ομαδοποίησης αξιολογήθηκε σε σχέση με το αποτέλεσμα αναφοράς και εξήχθηκαν συμπεράσματα.

Επαναλήφθηκε η μελέτη με αξιοποίηση τριών κυρίων συνιστωσών με «κόστος» απώλειας 18,1% από την αρχική πληροφορία και πραγματοποιήθηκε σύγκριση με τα προηγούμενα αποτελέσματα ομαδοποίησης.

Τέλος, επισημάνθηκε η αξία της ομαδοποίησης που υλοποιήθηκε για
103
χρηματοπιστωτικούς οργανισμούς μεγάλου μεγέθους.

ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΠΕΡΑΙΤΕΡΩ ΜΕΛΕΤΗ

Επανάληψη της μελέτης με νεότερα δεδομένα και εντοπισμός των μετακινήσεων της κάθε Ευρωπαϊκής χώρας στη νέα της σχετική θέση ως προς τις υπόλοιπες. Εξέταση αν οι δύο πρώτες κύριες συνιστώσες αντανακλούν μεγαλύτερο ή μικρότερο ποσοστό της αρχικής πληροφορίας, καθώς επίσης και αν η ομοιογένεια και η σύνθεση των ομάδων παραμένει η ίδια, αν δημιουργούνται νέες ομάδες, κλπ.

Πρόταση κατάλληλων μιγμάτων πολιτικών δραστηριοποίησης για έναν χρηματοπιστωτικό οργανισμό, προσαρμοσμένων στα χαρακτηριστικά απασχόλησης της κάθε ομάδας χωρών.

Εξέταση των εφαρμοζόμενων πολιτικών από διάφορους χρηματοπιστωτικούς οργανισμούς που δραστηριοποιούνται στην Ευρώπη και σύγκριση με τις πολιτικές που προτείνονται στην παρούσα μελέτη. Διερεύνηση αν και κατά πόσο η κερδοφορία των οργανισμών σχετίζεται με το βαθμό ταύτισης της πολιτικής που ακολουθούν οι οργανισμοί και της πολιτικής που προτείνεται στη διπλωματική εργασία.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Aczel A., Sounderpandian J. - Επιμέλεια ελληνικής έκδοσης Μιχάλης Σφακιανάκης (2013) - Στατιστική σκέψη στον κόσμο των επιχειρήσεων, Εκδόσεις: Π.Χ. Πασχαλίδης
- [2] Aldenderfer, M. S., Blashfield, R. K. (1984). *Cluster Analysis*, Sage Publ., Beverly Hills and London.
- [3] Anderberg, M. R. (1973). *Cluster analysis for applications*, Academic press, New York.
- [4] Barks, Robert A. (1972) *Descriptive Statistics and Probability Theory*, Hutchinson.
- [5] Bourgeois L. et. al. (1999) *Strategic Management*, The Dryden Press. Everitt, B. (1981). *Cluster Analysis*, Heinemann Educational Books, London.
- [6] Certo S. and Peter P. (1999) *Strategic Management*, McGraw-Hill.
- [7] David F. (2003) *Strategic Management Concepts*, Prentice Hall.
- [8] Everitt, B. S. and Dunn, G. (1991). *Applied Multivariate Data Analysis*, Arnold, New York.
- [9] Flury, B. and Riedwyl, H. (1988). *Multivariate Statistics : A practical approach*, Prentice Hall, New York.
- [10] Gordon, A. D. (1999). *Classification (2nd ed.)*, Chapman and Hall, London.
- [11] Hax A. and Majluf N. (1996) *The Strategy Concept and Process: A Pragmatic Approach*, Prentice Hall.
- [12] Hill C. and Jones G. (1998) *Strategic Management*, Houghton Mifflin Co.
- [13] Holcomb, Zealure C. (1998) *Fundamentals of Descriptive Statistics*, Amazon.com.
- [14] Hunger D and Wheelen T. (2008) *Concepts in Strategic Management and Business Policy*, Prentice Hall.
- [15] Hunger D and Wheelen T. (2003) *Essentials of Strategic Management*, Prentice Hall.
- [16] Johnson, R. A. and Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*, Prentice Hall, New Jersey.
- [17] Johnson M and Scholes K. (2002) *Exploring Corporate Strategy*, Prentice Hall.
- [18] Kaufman, L. and Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York.

- [19] Kotler, Philip (2001) *Marketing Management*, Prentice Hall.
- [20] Krzanowski, W. J. (1988). *Principles of Multivariate Analysis: A user's perspective*, Oxford University Press, U.K.
- [21] Manly, B. F. J. (1986). *Multivariate Statistical Methods: A primer*, Chapman and Hall, London.
- [22] Mann, Prem S. (1995) *Introductory Statistics, 2nd Edition*, Wiley.
- [23] Miller A. and Dess G. (1996) *Strategic Management*, McGraw-Hill.
- [24] Mintzberg H, Lambel J., Quinn J. and Ghoshal S. (2003) *The Strategy Process*.
- [25] Porter M. (1980) *Competitive Strategy*, Free Press.
- [26] Porter M. (1985) *Competitive Advantage*, Free Press.
- [27] Toit, S. H. C., Steyn, A. G. W. and Stumpf, R. H. (1986). *Graphical Exploratory Data Analysis*, Springer-Verlag, New York.
- [28] Rice, John A. (2007) *Mathematical Statistics and Data Analysis*, Cengage Learning.
- [29] Κούτρας, Μάρκος (2012) *Εφαρμοσμένη Πολυμεταβλητή Ανάλυση*, Σημειώσεις Τμήματος Εφαρμοσμένης Στατιστικής Πανεπιστημίου Πειραιώς.
- [30] Μάλλιαρης, Πέτρος (2001) *Εισαγωγή στο Μάρκετινγκ*, Εκδ. Σταμούλη.
- [31] Παπαγεωργίου Ανδρέας (2009-2010) Διπλωματική Εργασία: «Παραγοντική ανάλυση σε κύριες συνιστώσες», Πανεπιστήμιο Πατρών, Διατμηματικό Π.Μ.Σ. «Μαθηματικά των Υπολογιστών και των Αποφάσεων»
- [32] Σιώμκος, Γεώργιος (1999) *Στρατηγικό Μάρκετινγκ*, Εκδ. Σταμούλη.
- [33] Τσακλαγκάνος, Άγγελος (2005) *Βασικές Αρχές του Μάρκετινγκ*, Εκδ. Κυριακίδη Αφοι.