



ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΗΣ ΚΑΙ ΠΕΡΙΦΕΡΕΙΑΚΗΣ ΑΝΑΠΤΥΞΗΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΕΦΗΡΜΟΣΜΕΝΑ ΟΙΚΟΝΟΜΙΚΑ ΚΑΙ ΠΕΡΙΦΕΡΕΙΑΚΗ ΑΝΑΠΤΥΞΗ»
ΚΑΤΕΥΘΥΝΣΗ: ΕΦΗΡΜΟΣΜΕΝΑ ΟΙΚΟΝΟΜΙΚΑ ΚΑΙ ΔΙΟΙΚΗΣΗ

Η Ανάλυση Λανθανόντων Προφίλ για δεδομένα από δειγματοληπτικές έρευνες

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ατσάλης Ελευθέριος

Αθήνα, 2021

Τριμελής Επιτροπή

Clive Richardson, Ομότιμος Καθηγητής Παντείου Πανεπιστημίου (Επιβλέπων)

Σταύρος Ντεγιαννάκης, Αναπληρωτής Καθηγητής Παντείου Πανεπιστημίου

Γρηγόρης Σιουρούνης, Επίκουρος Καθηγητής Παντείου Πανεπιστημίου



Copyright © Ελευθέριος Ατσάλης, 2021

All rights reserved. Με επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας διπλωματικής εργασίας εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της διπλωματικής εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Μεταπτυχιακής Διπλωματικής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή.

Η έγκριση της διπλωματικής εργασίας από το Πάντειο Πανεπιστήμιο Κοινωνικών και Πολιτικών Επιστημών δεν δηλώνει αποδοχή των γνώμων του συγγραφέα.

Στη σύζυγο και τα παιδιά μου

Περιεχόμενα

Περίληψη	6
Abstract	7
Εισαγωγή	8
Κεφάλαιο 1. Δειγματοληπτικές έρευνες	9
1.1. Δειγματοληψία	9
1.2. Μέθοδοι δειγματοληψίας	9
1.3. Στοιχεία δειγματοληπτικής έρευνας.....	10
Κεφάλαιο 2. Ομαδοποίηση δεδομένων	12
2.1. Η αξία της ομαδοποίησης των δεδομένων	12
2.2. Μέθοδοι ομαδοποίησης δεδομένων	12
Κεφάλαιο 3. Θεωρητική προσέγγιση της Ανάλυσης Λανθανόντων Προφίλ (LPA) ...	14
3.1. Η Ανάλυση Λανθάνουσας Δομής (Latent Structure Analysis).....	14
3.2. Η Ανάλυση Λανθανόντων προφίλ (LPA).....	14
3.3. Περιγραφή του μοντέλου της Ανάλυσης Λανθανόντων Προφίλ.....	17
Κεφάλαιο 4. Η Ανάλυση Λανθανόντων Προφίλ (LPA) με τη χρήση της γλώσσας προγραμματισμού R.....	19
4.1. Προγράμματα για την ανάλυση δεδομένων με τη χρήση της Ανάλυσης Λανθανόντων Προφίλ (LPA).....	19
4.2. Περιγραφή του προγράμματος RStudio.....	19
4.3. Περιγραφή του υπολογιστικού πακέτου tidyLPA	21
4.4. Διαδικασία εφαρμογής της Ανάλυσης Λανθανόντων Προφίλ (LPA) με τη χρήση του RStudio και του υπολογιστικού πακέτου tidyLPA	22
Κεφάλαιο 5. Εφαρμογή της Ανάλυσης Λανθανόντων Προφίλ (LPA).....	27
5.1. Δεδομένα.....	27
5.2. Παρουσίαση των δεδομένων της έρευνας με Περιγραφική Στατιστική.	29
5.3. Προετοιμασία των δεδομένων	35
5.4. Εφαρμογή της LPA στα δεδομένα	35
5.5. Παρουσίαση συμπερασμάτων.....	39
5.5.1. Συμπεράσματα από την ταξινόμηση με την εφαρμογή της LPA.	39
5.5.2. Συσχέτιση των προφίλ με το φύλο και την τάξη των παιδιών	40
Κεφάλαιο 6. Άλλες εφαρμογές της Ανάλυσης Λανθανόντων Προφίλ (LPA)	48
Συμπεράσματα	50
Βιβλιογραφία	51
Παράρτημα Ι. Ερωτηματολόγιο Δυνατοτήτων και Δυσκολιών (SDQ).....	53
Παράρτημα ΙΙ. Αποτελέσματα εκτέλεσης κώδικα στο RStudio.....	54
1. Αποτελέσματα της εφαρμογής της LPA	54
2. Αποτελέσματα συσχέτισης προφίλ με φύλο και τάξη	56

Πίνακες

Πίνακας 5.1. Πλήθος και ποσοστό παιδιών ανά τάξη και φύλο	27
Πίνακας 5.2. Αντιστοιχία κλιμάκων και μεταβλητών	28
Πίνακας 5.3. Μέση τιμή κλιμάκων του SDQ ανά φύλο και τάξη	30
Πίνακας 5.4. Κριτήριο BIC για κάθε συνδυασμό μοντέλου παραμετροποίησης και αριθμού προφίλ.....	36
Πίνακας 5.5. Εκτιμώμενη μέση τιμή και διακύμανση μεταβλητών σε κάθε προφίλ ..	37
Πίνακας 5.6. Εκτιμώμενο ποσοστό του πληθυσμού σε κάθε προφίλ.....	38
Πίνακας 5.7. Εκτιμώμενο ποσοστό του πληθυσμού σε κάθε προφίλ.....	40
Πίνακας 5.8. Ποσοστό παιδιών ανά φύλο σε κάθε προφίλ	42
Πίνακας 5.9. Ποσοστό παιδιών ανά τάξη σε κάθε προφίλ	42
Πίνακας 5.10. Ποσοστό παιδιών κάθε φύλου σε κάθε προφίλ.....	45
Πίνακας 5.11. Ποσοστό παιδιών κάθε τάξης σε κάθε προφίλ.....	45

Διαγράμματα

Διάγραμμα 1. Γραφική απεικόνιση της Ανάλυσης Λανθανόντων Προφίλ (Latent Profile Analysis - LPA)	17
---	----

Γραφήματα

Γράφημα 1. Πληθυσμός παιδιών ανά τάξη.....	29
Γράφημα 2. Πληθυσμός παιδιών ανά τάξη και φύλο	30
Γράφημα 3. Πλήθος παιδιών ανά σκορ και τάξη για την Κλίμακα 1 του SDQ	31
Γράφημα 4. Πλήθος παιδιών ανά σκορ και τάξη για την Κλίμακα 2 του SDQ	31
Γράφημα 5. Πλήθος παιδιών ανά σκορ και τάξη για την Κλίμακα 3 του SDQ	32
Γράφημα 6. Πλήθος παιδιών ανά σκορ και τάξη για την Κλίμακα 4 του SDQ	33
Γράφημα 7. Πλήθος παιδιών ανά σκορ και τάξη για την Κλίμακα 5 του SDQ	33
Γράφημα 8. Τιμές κριτηρίου BIC για κάθε συνδυασμό μοντέλου παραμετροποίησης και αριθμού προφίλ.....	36
Γράφημα 9. Εκτιμώμενη μέση τιμή και τυπική απόκλιση μεταβλητών σε κάθε προφίλ	38
Γράφημα 10. Εκτιμώμενη μέση τιμή μεταβλητών σε κάθε προφίλ.	39
Γράφημα 11. Ποσοστό παιδιών ανά φύλο σε κάθε προφίλ.....	43
Γράφημα 12. Ποσοστό παιδιών ανά τάξη σε κάθε προφίλ.....	43
Γράφημα 13. Ποσοστό παιδιών κάθε φύλου σε κάθε προφίλ	46
Γράφημα 14. Ποσοστό παιδιών κάθε τάξης σε κάθε προφίλ	47

Εικόνες

Εικόνα 1. Περιβάλλον εργασίας προγράμματος RStudio.....	20
---	----

Περίληψη

Η παρούσα εργασία ασχολείται με μια μέθοδο ανάλυσης δεδομένων, την Ανάλυση Λανθανόντων Προφίλ (Latent Profile Analysis – LPA). Η μέθοδος αυτή αποτελεί ένα ισχυρό εργαλείο για την ομαδοποίηση δεδομένων από δειγματοληπτικές έρευνες και αφορά σε αναλύσεις δεδομένων πολλών μεταβλητών, των οποίων οι απαντήσεις κατηγοριοποιούνται. Τα άτομα που συμμετέχουν στην έρευνα κατανέμονται σε ομάδες ανάλογα με το συνδυασμό των απαντήσεών τους σε όλες τις μεταβλητές. Αυτές οι ομάδες αποτελούν τα προφίλ της λανθάνουσας μεταβλητής. Για κάθε προφίλ υπολογίζονται η μέση τιμή και η διακύμανση όλων των μεταβλητών σε αυτό, ενώ υπολογίζεται και το ποσοστό του πληθυσμού που ανήκει σε κάθε προφίλ. Η Ανάλυση Λανθανόντων Προφίλ βοηθάει αναλυτές από διάφορους επιστημονικούς κλάδους να εντοπίζουν διάφορους συνδυασμούς των μεταβλητών που δημιουργούνται μέσα στον πληθυσμό, αλλά και να εντοπίζουν τα κοινά χαρακτηριστικά των ατόμων που ανήκουν σε κάθε προφίλ. Στο πρώτο μέρος της εργασίας γίνεται μια θεωρητική προσέγγιση της Ανάλυσης Λανθανόντων Προφίλ, ενώ στη συνέχεια παρουσιάζεται μια πρακτική εφαρμογή της με τη χρήση της στατιστικής γλώσσας προγραμματισμού R και συγκεκριμένα με τη χρήση του προγράμματος RStudio και του υπολογιστικού πακέτου tidyLPA. Τα δεδομένα στα οποία θα εφαρμοστεί η Ανάλυση Λανθανόντων Προφίλ προέρχονται από μια από μια έρευνα που έγινε σε 17 σχολεία του Νομού Αττικής το σχολικό έτος 2015-2016, στην οποία 484 παιδιά της Ε΄ Δημοτικού (69 παιδιά), της ΣΤ΄ Δημοτικού (74 παιδιά), της Α΄ Γυμνασίου (307 παιδιά) και της Β΄ Γυμνασίου (34 παιδιά) απάντησαν στο Ερωτηματολόγιο Δυνατοτήτων και Δυσκολιών (SDQ) για τις ηλικίες 4 – 17. Οι μεταβλητές πάνω στις οποίες θα εφαρμοστεί η Ανάλυση Λανθανόντων Προφίλ είναι τα σκορ στις πέντε κλίμακες του SDQ. Τα προφίλ που δημιουργήθηκαν από την εφαρμογή της μεθόδου αντιστοιχούν σε διαφορετικό συνδυασμό σκορ για κάθε κλίμακα του SDQ και ομαδοποιούν τα παιδιά βάσει του επιπέδου των προβλημάτων στα οποία αντιστοιχούν οι κλίμακες 1 έως 4, καθώς και το επίπεδο της κοινωνικής τους συμπεριφοράς (κλίμακα 5). Μετά τη δημιουργία των προφίλ, επιχειρήθηκε να συσχετιστεί ο πληθυσμός κάθε προφίλ με το φύλο και τη σχολική τάξη των παιδιών. Στο τελευταίο κεφάλαιο της εργασίας παρουσιάζονται κάποιες άλλες περιπτώσεις εφαρμογής της Ανάλυσης Λανθανόντων Προφίλ σε διάφορους επιστημονικούς τομείς.

Λέξεις – κλειδιά: Ανάλυση δεδομένων, Ομαδοποίηση δεδομένων, Ανάλυση Λανθανόντων Προφίλ, Προφίλ, Πρόγραμμα RStudio, Υπολογιστικό πακέτο tidyLPA

The Latent Profile Analysis for data from sampling surveys

Eleftherios Atsalis

Abstract

This paper is concerned with a data analysis method, the Latent Profile Analysis (LPA). This method is a powerful tool for grouping data from sample surveys and involves data analysis of many variables, the answers to which are categorized. Persons participating in the survey are divided into groups depending on the combination of their replies to all variables. These groups are the profiles of the latent variable. For each profile the mean and variance of all variables in it, and the percentage of the population belonging to each profile are calculated. Latent Profile Analysis helps analysts from various scientific fields to identify different combinations of variables created within the population, but also to identify the common characteristics of the individuals belonging to each profile. In the first part of this paper, a theoretical approach to Latent Profile Analysis is made and then a practical application is presented using the statistical programming language R and specifically using the RStudio program and the tidyLPA package. The data to which the Analysis of Latent Profiles will be applied are from a survey conducted in 17 schools of the Region of Attica in the school year 2015-2016, in which 484 children of the 5th Elementary School (69 children), of the 6th Elementary School (74 children), of the 1st High School (307 children) and of the 2nd High School (34 children) answered the Strengths and Difficulties Questionnaire (SDQ) for age 4 - 17. The variables on which the Latent Profile Analysis will be applied are the scores on the five SDQ scales. The profiles created by the application of the method correspond to a different combination of scores for each SDQ scale and group the children based on the level of problems to which scales 1 to 4 correspond, as well as their level of social behavior (scale 5). After creating the profiles, an attempt was made to correlate the population of each profile with the gender and school class of the children. The last chapter of the paper presents some other cases of application of Latent Profile Analysis in various scientific fields.

Keywords: Data Analysis, Data Grouping, Latent Profile Analysis, Profiles, RStudio Program, tidyLPA Package

Εισαγωγή

Η λέξη στατιστική έχει στη ρίζα της τη λατινική λέξη «status» που σημαίνει κράτος. Στην αρχή η στατιστική αποτελούσε μέθοδο περιγραφής φαινομένων που σχετίζονταν με το κράτος ή τις ανθρώπινες δραστηριότητες. Στη συνέχεια εξελίχθηκε σε ολόκληρη επιστήμη και χωρίστηκε σε διάφορες υποκατηγορίες όπως η στατιστική του πληθυσμού, του εμπορίου, της εργασίας και άλλα. Η στατιστική αποτελεί ιδιαίτερα χρήσιμο εργαλείο για τη συλλογή, ανάλυση και παρουσίαση αριθμητικών δεδομένων, ενώ η χρήση της έχει επεκταθεί σε πολλούς επιστημονικούς κλάδους, αλλά και περιπτώσεις της καθημερινής ζωής (Ρίτσαρντσον & Βασίλαινας, 1999).

Η στατιστική επιστήμη έχει τις ρίζες της στην αρχαιότητα. Ο Ηρόδοτος, ο Θουκυδίδης, ο Αριστοτέλης και ο Ξενοφώντας συνέλεξαν στοιχεία για το εισόδημα του πληθυσμού, τις στρατιωτικές δυνάμεις και τη φορολογία. Αργότερα, οι Ρωμαίοι συγκέντρωναν δημογραφικά στοιχεία, όπως αριθμό θανάτων, γεννήσεων, πολέμους, γεωργική δραστηριότητα και άλλα.

Η στατιστική σαν αυτοτελής επιστήμη παρουσιάστηκε περί τον 17^ο αιώνα. Η θεμελίωση της στατιστικής επιστήμης ήρθε με τη «θεωρία των πιθανοτήτων» που προήλθε από τα τυχερά παιχνίδια. Οι Bernoulli, Gauss και Laplace ήταν αυτοί που θεμελίωσαν τη θεωρία των πιθανοτήτων ενώ έως τα τέλη του 19^{ου} αιώνα, η στατιστική είχε εξελιχθεί τόσο ώστε να μπορέσει να αναπτυχθεί και κυρίως στο πλαίσιο της επαγωγικής στατιστικής. Σημαντικό είναι επίσης να αναφερθούμε στους σύγχρονους στατιστικούς όπως ο Fischer και ο Pearson οι οποίοι συνέβαλαν στην ανάπτυξη στατιστικών μεθόδων, θεμελιώνοντας κατ' αυτόν τον τρόπο τη σύγχρονη στατιστική.

Όπως κι άλλες επιστήμες έτσι και η στατιστική χρησιμοποιεί τη δική της ορολογία. Όταν ασχολείται με ένα σύνολο, το σύνολο αυτό ονομάζεται στατιστικός πληθυσμός, ανεξάρτητα αν είναι σύνολο ανθρώπων ή όχι. Ταυτόχρονα, τα στοιχεία του συνόλου ονομάζονται στατιστικές μονάδες. Το χαρακτηριστικό του πληθυσμού που θέλουμε να μελετήσουμε ονομάζεται μεταβλητή και ένα μέρος του πληθυσμού δείγμα.

Επίσης, είναι ιδιαίτερα χρήσιμη στην συγκέντρωση, ταξινόμηση και οργάνωση μεγάλου όγκου δεδομένων που χωρίς τα εργαλεία της Στατιστικής δεν θα μπορούσαν να αναλυθούν. Πολλές επιστημονικές κοινότητες χρησιμοποιούν τα εργαλεία της στατιστικής για να μπορέσουν να μετατρέψουν τα δεδομένα που έχουν συγκεντρώσει κατά τις έρευνες τους σε ασφαλή και χρήσιμα συμπεράσματα.

Στην παρούσα εργασία θα ασχοληθούμε με μια μέθοδο ομαδοποίησης δεδομένων. Στη σύγχρονη εποχή που οι ανάγκες αυξάνονται ραγδαία και η χρήση υπολογιστικών συστημάτων είναι ευρεία παρουσιάζεται συχνά η ανάγκη της ομαδοποίησης των δεδομένων. Η ορθή ταξινόμηση και οργάνωση των δεδομένων είναι σημαντικό κομμάτι της λήψης αποφάσεων. Ιδιαίτερα όταν μιλάμε για δεδομένα που έχουν προέλθει από δειγματοληπτικές έρευνες, όπως είναι φυσικό, ο όγκος τους είναι τόσο μεγάλος που η διαχείριση τους χωρίς ομαδοποίηση καθίσταται αδύνατη. Σε αυτό το σημείο η Στατιστική μας βοηθάει με τις μεθόδους ομαδοποίησης δεδομένων οι οποίες ποικίλλουν ανάλογα τον τύπο των δεδομένων που διαθέτουμε. Στην εργασία αυτή, θα ασχοληθούμε με μια μέθοδο ομαδοποίησης δεδομένων, την Ανάλυση Λανθανόντων Προφίλ (Latent Profile Analysis - LPA).

Κεφάλαιο 1. Δειγματοληπτικές έρευνες

1.1. Δειγματοληψία

Με την ανάπτυξη όλων των τεχνολογικών μέσων, η πρόσβαση σε πληροφορίες και δεδομένα έχει γίνει ιδιαίτερος εύκολη συγκριτικά με παλαιότερα. Ο όγκος αυτών των δεδομένων είναι τόσο μεγάλος που η ποσοτικοποίηση τους κρίνεται απαραίτητη για την μετατροπή τους σε δεδομένα που μπορούν να χρησιμοποιηθούν για την εξαγωγή συμπερασμάτων. Η ανάγκη αυτή ποσοτικοποίησης των δεδομένων, οδήγησε στην ανάπτυξη των δειγματοληπτικών ερευνών.

Οι δειγματοληπτικές έρευνες καταγράφουν δεδομένα από ένα υποσύνολο του πληθυσμού, το δείγμα (sample), από τα οποία εξάγονται πληροφορίες για τον πληθυσμό. Οι δειγματοληπτικές έρευνες χρησιμοποιούνται ευρύτατα και έχουν ποικίλη θεματολογία.

Η πιο συχνή δειγματοληπτική έρευνα που είναι ευρέως γνωστή είναι η δημοσκόπηση. Άλλες γνωστές δειγματοληπτικές έρευνες είναι αυτές που διεξάγονται σε επίπεδο χωρών ή άλλων διοικητικών μονάδων και έχουν ως σκοπό την εκτίμηση κάποιων δεικτών. Όμως οι δειγματοληπτικές έρευνες δεν αφορούν απαραίτητα σε πληθυσμούς ανθρώπων. Υπάρχουν και περιπτώσεις που τα ερωτήματα που θέλουμε να απαντηθούν έχουν να κάνουν με το περιβάλλον, τα υλικά αγαθά και άλλα (Παπαγεωργίου, 2015).

1.2. Μέθοδοι δειγματοληψίας

Η έννοια της δειγματοληψίας οφείλεται κυρίως στο Kiaer, ο οποίος ισχυρίστηκε το 1895 ότι η καθολική απογραφή έπρεπε να αντικατασταθεί από την μελέτη επιλεγμένων μονάδων του πληθυσμού. Ο Kiaer έθεσε την ιδέα χωρίς να αποσαφηνίσει τον τρόπο επιλογής των μονάδων του πληθυσμού, αν και πιθανότερο είναι να εννοούσε την καθοδηγούμενη επιλογή και όχι την τυχαία επιλογή που εντέλει αποτελεί τη βασική μέθοδο δειγματοληψίας (Ρίτσαρντσον & Βασίλαινας, 1999).

Η μέθοδος δειγματοληψίας καθορίζεται από τον τρόπο επιλογής του δείγματος. Υπάρχουν δύο κατηγορίες δειγματοληψίας, η δειγματοληψία πιθανότητας και η δειγματοληψία μη πιθανότητας. Τα συμπεράσματα που προκύπτουν από τα δεδομένα της πρώτης κατηγορίας μπορούν να επεκταθούν στον πληθυσμό, ενώ αυτά της δεύτερης κατηγορίας δεν μπορούν να επεκταθούν στον πληθυσμό.

Γνωστές μέθοδοι δειγματοληψίας πιθανότητας είναι επιγραμματικά οι εξής:

- **Τυχαία δειγματοληψία.** Πρόκειται για την κλασική μέθοδο δειγματοληψίας. Στην τυχαία δειγματοληψία οι μονάδες που αποτελούν το δείγμα έχουν μια ορισμένη και γνωστή πιθανότητα, μεγαλύτερη του μηδενός για να επιλεγθούν.
- **Απλή τυχαία δειγματοληψία.** Σε αυτή τη μέθοδο κάθε μονάδα του πληθυσμού έχει την ίδια πιθανότητα να επιλεγεί και οι επιλογές των μονάδων είναι ανεξάρτητες μεταξύ τους.
- **Συστηματική δειγματοληψία.** Αποτελεί εναλλακτικό τρόπο για να πραγματοποιηθεί η τυχαία δειγματοληψία, απλούστερο όμως της απλής τυχαίας

δειγματοληψίας, καθώς απαιτεί μόνο έναν τυχαίο αριθμό που καθορίζει την πρώτη μονάδα που θα επιλέξουμε.

- **Στρωματοποιημένη δειγματοληψία.** Σε αυτή τη μέθοδο ο πληθυσμός χωρίζεται σε στρώματα και η δειγματοληψία γίνεται σε κάθε στρώμα ξεχωριστά.
- **Δειγματοληψία κατά συστάδες ή δειγματοληψία κατά ομάδες.** Και σε αυτή τη μέθοδο ο πληθυσμός χωρίζεται από την αρχή σε συστάδες, ενώ η διαφορά της με την στρωματοποιημένη δειγματοληψία είναι ότι αρχικά γίνεται τυχαία επιλογή συστάδων και έπειτα επιλογή δείγματος μόνο από τις επιλεγμένες συστάδες.
- **Δειγματοληψία με άνισες πιθανότητες ή δειγματοληψία ανάλογα του μεγέθους.** Σε αυτή την περίπτωση ακολουθείται η ίδια διαδικασία με την δειγματοληψία κατά συστάδες, όμως στο δεύτερο στάδιο γίνεται επιλογή μεταξύ δύο σχεδίων. Στο πρώτο επιλέγονται όλες οι μονάδες των επιλεγμένων συστάδων, ενώ στο δεύτερο γίνεται επιλογή απλού τυχαίου δείγματος n μονάδων από κάθε επιλεγμένη συστάδα.
- **Επιφανειακή δειγματοληψία.** Πρόκειται για τη βασική μέθοδο δειγματοληψίας από τον γενικό πληθυσμό σε πολλές χώρες. Στο πρώτο στάδιο της μεθόδου επιλέγονται γεωγραφικές περιοχές και στη συνέχεια επιλέγεται ένα δείγμα ατόμων από τις επιλεγμένες περιοχές με δειγματοληψία κατά συστάδες.

Οι βασικότερες μέθοδοι δειγματοληψίας μη-πιθανότητας είναι επιγραμματικά οι εξής:

- **Δειγματοληψία ευκολίας.** Σε αυτή την περίπτωση η επιλογή των μονάδων γίνεται με κριτήριο την ευκολία και όχι την τυχαιότητα ή την αντιπροσωπευτικότητα του πληθυσμού.
- **Καθοδηγούμενη δειγματοληψία.** Πρόκειται για μέθοδο στην οποία το δείγμα που επιλέγεται είναι εξειδικευμένο και υποκειμενικό, με στόχο την εξυπηρέτηση του σκοπού της έρευνας.
- **Δειγματοληψία με επιτρεπόμενους αριθμούς ή δειγματοληψία ποσοστών.** Σε αυτή τη μέθοδο οι επιμέρους μονάδες του δείγματος επιλέγονται σκόπιμα από τον ερευνητή, ώστε το τελικό δείγμα να έχει εκπροσώπους από κάθε κατηγορία του πληθυσμού.

1.3. Στοιχεία δειγματοληπτικής έρευνας

Μια δειγματοληπτική έρευνα αποτελείται από κάποια στοιχεία. Τα στοιχεία αυτά είναι τα εξής (Παπαγεωργίου, 2015):

- **Το δειγματοληπτικό σχέδιο.** Πρόκειται για τον τρόπο επιλογής του δείγματος. Κατά την κατάρτιση του δειγματοληπτικού σχεδίου είναι πολύ σημαντικό να τονιστεί ο σκοπός της έρευνας, καθώς καθορίζει τις μεταβλητές που πρέπει να καταγραφούν. Το δειγματοληπτικό σχέδιο αφορά σε αυτόν που αναθέτει την έρευνα, τους ανθρώπους που θα διεξάγουν τη δειγματοληψία και τον στατιστικό επιστήμονα που θα κληθεί να την αναλύσει.
- **Οι μετρήσεις της έρευνας.** Μετά την κατάρτιση του δειγματοληπτικού σχεδίου πρέπει να καταγραφούν οι μεταβλητές που θα χρησιμοποιηθούν και οι μονάδες μέτρησης της κάθε μιας. Σε αυτή τη φάση πρέπει να δημιουργηθεί και το ερωτηματολόγιο που θα χρησιμοποιηθεί για τη διεξαγωγή της έρευνας. Συχνά, η δημιουργία του ερωτηματολογίου είναι ιδιαίτερα πολύπλοκη. Η σύνταξη του

ερωτηματολογίου πρέπει να υπακούει σε συγκεκριμένους κανόνες όπως να έχει συγκεκριμένο στόχο, να είναι σύντομο, απλό και κατανοητό. Έχοντας προσέξει αυτούς τους κανόνες τότε ελαχιστοποιούνται οι περιπτώσεις λήψης δεδομένων με σφάλματα λόγω παρανόησης των ερωτήσεων και έτσι οδηγούμαστε ευκολότερα σε δεδομένα που ο στατιστικός αναλυτής μπορεί να χρησιμοποιήσει για να εξάγει ασφαλή συμπεράσματα.

- **Η διεξαγωγή της έρευνας.** Σε αυτό το βήμα πρέπει να είμαστε συνεπείς στους αρχικούς σκοπούς της έρευνας και να φροντίσουμε να μην υπάρξει απόκλιση σε κανένα στάδιο διεξαγωγής της. Ταυτόχρονα, θα πρέπει να έχει προβλεφθεί η επίλυση προβλημάτων που μπορεί να προκύψουν κατά τη διάρκεια διεξαγωγής της έρευνας.
- **Η στατιστική ανάλυση.** Μετά την διεξαγωγή της έρευνας έχει συγκεντρωθεί ένα πλήθος δεδομένων που θα ανατεθεί στον στατιστικό αναλυτή. Ο στατιστικός αναλυτής λαμβάνοντας υπόψη τη δειγματοληπτική μέθοδο που εφαρμόστηκε και τους σκοπούς διεξαγωγής της έρευνας, θα χρησιμοποιήσει τα καλύτερα δυνατά εργαλεία από την επιστήμη της Στατιστικής για να οργανώσει και να αναλύσει τα δεδομένα που έχει στην κατοχή του. Κατά την ανάλυση του θα πρέπει να λάβει υπόψη και τα σφάλματα. Τέλος, τα αποτελέσματα της ανάλυσης θα πρέπει να καταγραφούν και να παρουσιαστούν.

Κεφάλαιο 2. **Ομαδοποίηση δεδομένων**

2.1. Η αξία της ομαδοποίησης των δεδομένων

Σε κάθε στατιστική ανάλυση, μετά τη συγκέντρωση των προς επεξεργασία δεδομένων, ένα βασικό «πρόβλημα» είναι η ομαδοποίησή τους. Τα δεδομένα στην αρχική τους μορφή τις περισσότερες φορές παρουσιάζουν μια χαοτική εικόνα, γεγονός που κάνει ιδιαίτερα δύσκολη την εξαγωγή συμπερασμάτων. Έτσι πριν το στάδιο της επεξεργασίας και ανάλυσης των δεδομένων, πρέπει να προηγηθεί ένα ακόμα στάδιο, αυτό της ομαδοποίησής τους. Μια σωστή ομαδοποίηση των δεδομένων κάνει πολύ πιο εύκολη τη δουλειά του στατιστικού αναλυτή.

Κατά το στάδιο της ομαδοποίησης, τα δεδομένα ομαδοποιούνται με κριτήριο κάποιο ή κάποια όμοια χαρακτηριστικά τους. Έτσι ο στατιστικός αναλυτής αντί να επεξεργαστεί τα δεδομένα, επεξεργάζεται τις ομάδες που δημιουργούνται και τα χαρακτηριστικά αυτών. Πλέον ο στατιστικός αναλυτής έχει απέναντί του έναν πολύ μικρότερο αριθμό δεδομένων, τις ομάδες και τα χαρακτηριστικά τους, και όχι τον μεγάλο αριθμό των αρχικών δεδομένων. Η ομαδοποίηση βοηθάει στο να γίνουν ευδιάκριτα, γρήγορα και αποτελεσματικά τα αποτελέσματα της έρευνας, ώστε να εξαχθούν χρήσιμες πληροφορίες.

2.2. Μέθοδοι ομαδοποίησης δεδομένων

Οι μέθοδοι ομαδοποίησης ποικίλλουν, ενώ η πρόοδος της τεχνολογίας έχει συμβάλλει στη γρηγορότερη ομαδοποίηση των δεδομένων, αφού με ειδικά προγράμματα ο στατιστικός αναλυτής μπορεί να ομαδοποιήσει τα δεδομένα πολύ γρήγορα. Η επιλογή της μεθόδου έγκειται στο υπό μελέτη θέμα και είναι στην κρίση του αναλυτή.

Κάποιες από τις βασικότερες μεθόδους ομαδοποίησης, καθώς και τα κύρια χαρακτηριστικά τους, είναι οι εξής:

- **Ανάλυση Κυρίων Συνιστωσών (Principal Components Analysis - PCA)**

Η PCA είναι μία μέθοδος στην οποία δημιουργούνται γραμμικοί συνδυασμοί των αρχικών μεταβλητών, έτσι ώστε αυτοί να είναι ανεξάρτητοι μεταξύ τους, αλλά να περιέχουν όσο το δυνατόν μεγαλύτερο μέρος της διακύμανσης των αρχικών μεταβλητών. Η PCA δεν έχει καμία χρησιμότητα σε ανεξάρτητες μεταβλητές, αφού αυτές θα ταυτιστούν με κάποια συνιστώσα. Οι γραμμικοί συνδυασμοί που προκύπτουν από την PCA αποτελούν τις νέες μεταβλητές, οι οποίες είναι πολύ λιγότερες από τις αρχικές και ανεξάρτητες μεταξύ τους. Επίσης, η PCA είναι χρήσιμη ώστε να ποσοτικοποιήσουμε μη μετρήσιμες ποσότητες και να βρούμε κατά πόσο οι αρχικές μεταβλητές συσχετίζονται ή όχι. Τέλος, ενώ με τη μείωση του αριθμού των μεταβλητών χάνουμε μέρος της πληροφορίας, έχουμε τεράστιο κέρδος στο χώρο και την ταχύτητα επεξεργασίας των δεδομένων (Φαρμάκη, 2012).

- **Ανάλυση Παραγόντων (Factor Analysis - FA)**

Μια άλλη τεχνική που σχετίζεται με την PCA είναι η Ανάλυση Παραγόντων (FA). Σκοπός της είναι επίσης η μείωση του αριθμού των μεταβλητών σε συγκεκριμένο αριθμό γραμμικών συνδυασμών ή παραγόντων. Ο αριθμός των νέων μεταβλητών (παραγόντων) επιλέγεται με στόχο την καλύτερη ερμηνεία της αρχικής κατάστασης. Η

FA στοχεύει στην εύρεση αφανών παραγόντων ικανών να εξηγήσουν τη συσχέτιση μεταξύ των μεταβλητών, δηλαδή δίνει έμφαση στην εξήγηση της συσχέτισης των μεταβλητών, και χρησιμοποιείται για την ανίχνευση και ερμηνεία δομών. Ο αριθμός των παραγόντων που επιλέγεται μπορεί να προκύπτει από τη θεωρία, αλλά και να εξαρτάται από το ποσοστό της διακύμανσης που ερμηνεύεται, ενώ δεν χρησιμοποιείται όλη η διακύμανση μιας μεταβλητής αλλά μόνο αυτή που διαμοιράζεται με τις άλλες μεταβλητές (Πετρίδης, 2015).

- **Ανάλυση κατά Συστάδες (Clusters Analysis - CA)**

Η ανάλυση κατά συστάδες (CA) φανερώνει εσωτερικές διαφορές με στόχο την ομαδοποίηση των δεδομένων βάσει γειτονίας ή ομοιότητας. Στόχος της είναι η οργάνωση των παρατηρήσεων σε συστάδες, ώστε αυτές που ανήκουν στην ίδια συστάδα να είναι μεταξύ τους πιο όμοιες από αυτές που δεν ανήκουν. Το σύνολο των παρατηρήσεων μέσα σε μια συστάδα πρέπει να είναι όσο το δυνατόν πιο ομοιογενές και να διαφέρει όσο το δυνατόν περισσότερο από τις παρατηρήσεις άλλων συστάδων (Φαρμάκη, 2012).

- **Ανάλυση Λανθανουσών Κλάσεων (Latent Class Analysis - LCA)**

Η ανάλυση λανθανουσών κλάσεων (LCA) έχει ως στόχο το διαχωρισμό των ομοιογενών δεδομένων σε δύο ή περισσότερες ομοιογενείς ομάδες ή κλάσεις. Ο συνδυασμός των απαντήσεων κάθε συμμετέχοντα στην έρευνα αποτελούν τη βάση για τον υπολογισμό της πιθανότητας αυτός να τοποθετηθεί σε μία κλάση, ενώ από το μοτίβο των απαντήσεων των συμμετεχόντων σε κάθε κλάση προκύπτουν τα χαρακτηριστικά της κλάσης. Σε αντίθεση με τις προηγούμενες μεθόδους που είναι επικεντρωμένες στις μεταβλητές, η LCA είναι προσωποκεντρική λόγω της εστίασής της στα χαρακτηριστικά των συμμετεχόντων και το μοτίβο των απαντήσεών τους. Στόχος της LCA είναι η αποκάλυψη ετερογένειας από φαινομενικά ομοιογενή δείγματα (Williams & Kibowski, 2016).

- **Ανάλυση Λανθανόντων Προφίλ (Latent Profile Analysis - LPA)**

Η ανάλυση λανθανόντων προφίλ (LPA) είναι παρόμοια με την LCA και η κύρια διαφορά τους έγκειται στον τύπο των μεταβλητών που χρησιμοποιούν. Η LCA πραγματοποιείται σε κατηγορικές μεταβλητές, ενώ η LPA κυρίως σε συνεχείς. Επίσης, στην LCA οι κλάσεις περιγράφονται από την πιθανότητα κάθε μεταβλητής σε αυτές, ενώ στην LPA τα προφίλ περιγράφονται από τη μέση τιμή κάθε μεταβλητής σε αυτά (Williams & Kibowski, 2016).

Κεφάλαιο 3. **Θεωρητική προσέγγιση της Ανάλυσης Λανθανόντων Προφίλ (LPA)**

3.1. Η Ανάλυση Λανθάνουσας Δομής (Latent Structure Analysis)

Η Ανάλυση Λανθάνουσας Δομής αποτελεί ένα αναλυτικό εργαλείο που επιτρέπει στους ερευνητές να διερευνούν ερωτήσεις σχετικά με μοτίβα δεδομένων και να αναλύουν ποια από αυτά τα μοτίβα ταιριάζουν σε σημαντικές μεταβλητές. Πιο συγκεκριμένα η Ανάλυση Λανθάνουσας Δομής είναι ένα προσωποκεντρικό εργαλείο το οποίο εστιάζει στις ομοιότητες και στις διαφορές μεταξύ των μονάδων, αντί των σχέσεων των μεταβλητών.

Η Ανάλυση Λανθάνουσας Δομής αναφέρεται σε μια συλλογή στατιστικών δεδομένων στην οποία οι μονάδες έχουν χωριστεί σε υποπληθυσμούς ή κλάσεις προκειμένου να μελετηθούν. Κύριες μέθοδοι της Ανάλυσης Λανθάνουσας Δομής είναι η Ανάλυση Λανθανουσών Κλάσεων (Latent Classes Analysis - LCA), η Ανάλυση Λανθανόντων Προφίλ (Latent Profile Analysis - LPA) και η Ανάλυση Λανθανόντων Παραγόντων (Latent Factor Analysis - LFA).

Η Ανάλυση Λανθάνουσας Δομής αναπτύχθηκε από τον Paul Lazarsfeld το 1950. Το ενδιαφέρον του Lazarsfeld ήταν η ανάπτυξη της βάσης των μοντέλων λανθάνουσας δομής και στα πρώτα του άρθρα επικεντρώθηκε στην ανάλυση της δομής αυτών των μοντέλων και σε μια περίληψη των ιδιοτήτων τους ως εργαλεία για τη μελέτη κοινωνιολογικών φαινομένων. Αυτό είχε ως αποτέλεσμα οι στατιστικές μέθοδοι που ανέπτυξε να είναι πρωτόγονες και όχι οι πιο αποτελεσματικές. Έτσι, ένας συνάδελφός του, ο T. W. Anderson, το 1954 ανέπτυξε μια μέθοδο εκτίμησης για τις παραμέτρους της LCA. Το 1959 ο Gibson συνέδεσε την LPA με την LCA και έδειξε την ομοιότητά τους, ενώ οι Christoffersson (1975) και Muthén (1978) ανέπτυξαν ένα μοντέλο για την LFA (Andersen, 1982).

Σε αυτή την εργασία ασχολούμαστε με την Ανάλυση Λανθανόντων Προφίλ (LPA) την οποία και θα αναλύσουμε στη συνέχεια.

3.2. Η Ανάλυση Λανθανόντων προφίλ (LPA)

Η Ανάλυση Λανθανόντων Προφίλ (LPA) επικεντρώνεται στον εντοπισμό λανθανόντων υποπληθυσμών εντός ενός πληθυσμού με βάση ένα συγκεκριμένο σύνολο μεταβλητών. Η LPA υποθέτει ότι τα μέλη του πληθυσμού μπορούν να κατηγοριοποιηθούν σε διαφορετικά προφίλ ανάλογα με τα χαρακτηριστικά τους.

Σε σύγκριση με τις παραδοσιακές, μη λανθάνουσες μεθόδους ομαδοποίησης, η LPA αντιμετωπίζει την ιδιότητα μέλους ως μια μη παρατηρούμενη κατηγορική μεταβλητή, η τιμή της οποίας υποδεικνύει σε ποιο προφίλ ανήκει ένα άτομο με συγκεκριμένη πιθανότητα. Σύμφωνα με τους Spurk, Hirschi, Wang, Valero και Kauffeld (2020) τα πλεονεκτήματα της LPA έναντι των παραδοσιακών μεθόδων ομαδοποίησης είναι ότι:

- τα άτομα ταξινομούνται σε συστάδες βάσει των πιθανοτήτων συμμετοχής τους σε αυτές, οι οποίες υπολογίζονται απευθείας από το μοντέλο
- οι μεταβλητές μπορεί να είναι συνεχείς, κατηγορικές, μετρήσεις ή οποιοσδήποτε συνδυασμός αυτών

- για την περιγραφή των προφίλ μπορούν να χρησιμοποιηθούν δημογραφικά στοιχεία ή άλλοι συνδυασμοί στοιχείων.

Η LPA εστιάζει σε μοτίβα μεταβλητών, τους δείκτες LPA. Στην LPA τα προφίλ των ατόμων με παρόμοιο μοτίβο αναγνωρίζονται και συγκρίνονται με άλλα προφίλ, τόσο σε επίπεδο μεταβλητών, όσο και σε επίπεδο συνδυασμού μεταβλητών. Ως διαφορές σχήματος μεταξύ των προφίλ ορίζονται οι ποιοτικές διαφορές στο μοτίβο των μεταβλητών μεταξύ των προφίλ, ενώ ως διαφορές επιπέδου μεταξύ των προφίλ ορίζονται οι ποσοτικές διαφορές στο μοτίβο των μεταβλητών.

Η LPA προϋποθέτει την ετερογένεια του πληθυσμού και χρησιμοποιείται κυρίως όταν υπάρχει μη παρατηρούμενη ετερογένεια, δηλαδή ετερογένεια που δεν μπορεί να παρατηρηθεί εκ των προτέρων. Σε αυτή την περίπτωση οι υποπληθυσμοί είναι λανθάνοντες και πρέπει να εξαχθούν από τα δεδομένα. Η LPA βρίσκει εφαρμογή όταν υπάρχει μη παρατηρούμενη ετερογένεια και ομάδες με συγκεκριμένους συνδυασμούς μεταβλητών. Κάθε συνδυασμός μεταβλητών αναπαριστά ένα λανθάνων προφίλ, τα οποία διαφοροποιούνται σε διάφορες παραμέτρους (π.χ. μέσες τιμές και διακυμάνσεις).

Κάθε άτομο εντάσσεται στο προφίλ με τις υψηλότερες τιμές πιθανότητας για αυτό το άτομο. Η πιθανότητα εκχώρησης προφίλ ονομάζεται και ιδιότητα μέλους (Spurk, Hirschi, Wang, Valero και Kauffeld, 2020).

Όσον αφορά στο σχήμα των λανθανόντων προφίλ η Masyn (2013) πρότεινε τέσσερις διαφορετικές προδιαγραφές παράλληλα με τον καταλληλότερο αριθμό προφίλ:

- Η πρώτη είναι η λιγότερο περιοριστική και περιγράφει ένα μοντέλο που δεν προϋποθέτει ανεξαρτησία εντός προφίλ, ενώ επιτρέπει διαφορετικές διακυμάνσεις και συνδιακυμάνσεις σε κάθε προφίλ. Σε αυτή την περίπτωση ο πίνακας διακύμανσης – συνδιακύμανσης του προφίλ k από τα K προφίλ έχει τη μορφή

$$\Sigma_k = \begin{bmatrix} \theta_{11k} & & & & \\ \theta_{21k} & \theta_{22k} & & & \\ \vdots & \vdots & \ddots & & \\ \theta_{M1k} & \theta_{M2k} & \dots & \theta_{MMk} & \end{bmatrix}$$

όπου M το πλήθος των μεταβλητών, θ_{mmk} είναι η διακύμανση του στοιχείου m στο προφίλ k και θ_{mjk} η συνδιακύμανση μεταξύ των στοιχείων m και j στο προφίλ k .

- Η δεύτερη περιγράφει ένα μοντέλο που δεν προϋποθέτει ανεξαρτησία εντός προφίλ, ενώ ορίζει ίδιες διακυμάνσεις και συνδιακυμάνσεις σε όλα τα προφίλ. Σε αυτή την περίπτωση ο πίνακας διακύμανσης – συνδιακύμανσης του προφίλ k από τα K προφίλ έχει τη μορφή

$$\Sigma_k = \Sigma = \begin{bmatrix} \theta_{11} & & & & \\ \theta_{21} & \theta_{22} & & & \\ \vdots & \vdots & \ddots & & \\ \theta_{M1} & \theta_{M2} & \dots & \theta_{MM} & \end{bmatrix}, \forall k \in (1, \dots, K)$$

όπου M το πλήθος των μεταβλητών, θ_{mm} είναι η διακύμανση του στοιχείου m σε όλα τα προφίλ και θ_{mj} η συνδιακύμανση μεταξύ των στοιχείων m και j σε όλα τα προφίλ.

- Η τρίτη περιγράφει ένα μοντέλο που προϋποθέτει ανεξαρτησία εντός προφίλ, ενώ επιτρέπει διαφορετικές διακυμάνσεις σε κάθε προφίλ. Σε αυτή την περίπτωση ο πίνακας διακύμανσης – συνδιακύμανσης του προφίλ k από τα K προφίλ έχει τη μορφή

$$\Sigma_k = \begin{bmatrix} \theta_{11k} & & & \\ 0 & \theta_{22k} & & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \dots & \theta_{MMk} \end{bmatrix}$$

όπου M το πλήθος των μεταβλητών, θ_{mmk} είναι η διακύμανση του στοιχείου m στο προφίλ k .

- Η τέταρτη είναι η πιο περιοριστική και περιγράφει ένα μοντέλο που προϋποθέτει ανεξαρτησία εντός προφίλ, ενώ ορίζει ίδιες διακυμάνσεις σε όλα τα προφίλ. Σε αυτή την περίπτωση ο πίνακας διακύμανσης – συνδιακύμανσης του προφίλ k από τα K προφίλ έχει τη μορφή

$$\Sigma_k = \Sigma = \begin{bmatrix} \theta_{11} & & & \\ 0 & \theta_{22} & & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \dots & \theta_{MM} \end{bmatrix}, \forall k \in (1, \dots, K)$$

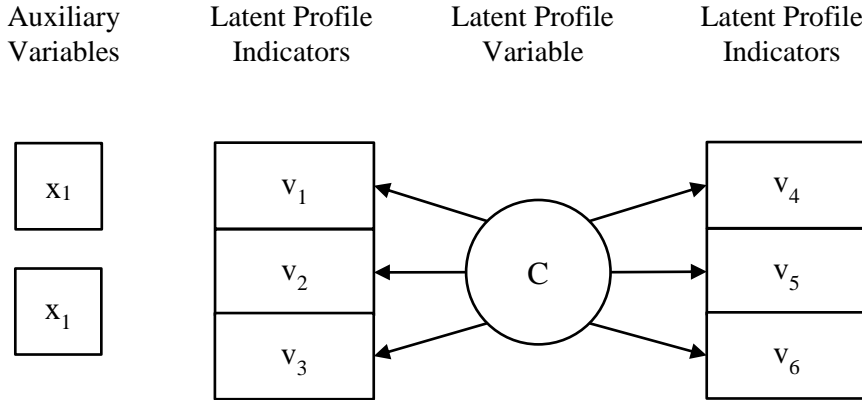
όπου M το πλήθος των μεταβλητών, θ_{mm} είναι η διακύμανση του στοιχείου m σε όλα τα προφίλ.

Η Masyn (2013) πρότεινε ότι αυτές οι τέσσερις προδιαγραφές πρέπει να αξιολογηθούν παράλληλα με τον αριθμό των προφίλ, ώστε η λύση που θα προκύψει από την LPA να λαμβάνει υπόψιν τόσο το καταλληλότερο σχήμα, όσο και τον καταλληλότερο αριθμό προφίλ.

Επιγραμματικά, η LPA συγκρίνει τις απαντήσεις που έδωσε κάθε άτομο και εντοπίζει ομοιότητες μεταξύ των απαντήσεων. Το άτομα με παρόμοιες απαντήσεις εντάσσονται σε μια ομάδα. Στη συνέχεια υπολογίζονται η μέση τιμή και η διακύμανση κάθε μεταβλητής σε κάθε προφίλ, η πιθανότητα κάθε ατόμου να βρίσκεται σε κάθε προφίλ και το ποσοστό του πληθυσμού που εντάσσεται σε κάθε προφίλ. Η μέση τιμή των μεταβλητών είναι αυτή που προσδιορίζει και τα χαρακτηριστικά του προφίλ.

Αυτό που υπολογίζεται με τη βοήθεια της LPA είναι μια νέα λανθάνουσα μεταβλητή η οποία παίρνει για κάθε προφίλ τιμή ίση με το συνδυασμό της μέσης τιμής των μεταβλητών σε αυτό το προφίλ. Στο Διάγραμμα 1 βλέπουμε τη γραφική απεικόνιση της LPA.

Διάγραμμα 1. Γραφική απεικόνιση της Ανάλυσης Λανθανόντων Προφίλ (Latent Profile Analysis - LPA)



3.3. Περιγραφή του μοντέλου της Ανάλυσης Λανθανόντων Προφίλ

Υποθέτουμε ότι έχουμε M συνεχείς μεταβλητές, y_1, y_2, \dots, y_M , στις οποίες αντιστοιχούν οι απαντήσεις από n άτομα, όπου y_{mi} είναι η απάντηση στο αντικείμενο m από το i άτομο. Υποθέτουμε ότι υπάρχει μία κατηγορική λανθάνουσα μεταβλητή c , με K προφίλ, όπου $c_i = k$ αν το άτομο i ανήκει στο προφίλ k . Το ποσοστό των ατόμων στο προφίλ k συμβολίζεται με π_k . Τα K προφίλ είναι τέτοια ώστε κάθε άτομο να ανήκει ακριβώς σε ένα προφίλ και $\sum \pi_k = 1$. Η σχέση μεταξύ των M μεταβλητών και της λανθάνουσας μεταβλητής c δίνεται από τη σχέση

$$f(y_i) = \sum_{k=1}^K [\pi_k \cdot f_k(y_i)]$$

όπου $y_i = (y_{1i}, y_{2i}, \dots, y_{Mi})$, $f(y_i)$ είναι η πολυμεταβλητή συνάρτηση πυκνότητας πιθανότητας για όλο τον πληθυσμό και $f_k(y_i) = f(y_i | c_i = k)$ είναι η συνάρτηση πυκνότητας για το προφίλ k . Συνεπώς, το μοντέλο της LPA προσδιορίζει ότι η συνολική κοινή κατανομή των M μεταβλητών είναι το αποτέλεσμα της μίξης των K κατανομών των M μεταβλητών, με την $f_k(y_i)$ να αντιπροσωπεύει την κατανομή για το y_i .

Οι δομικές παράμετροι του μοντέλου της LPA είναι οι αναλογίες προφίλ π_k . Οι παράμετροι μέτρησης είναι όλες αυτές που σχετίζονται με τις κατανομές πιθανότητας για κάθε προφίλ. Συνήθως, η εντός προφίλ κατανομή των συνεχών μεταβλητών θεωρείται ότι είναι πολυμεταβλητή κανονική κατανομή, δηλαδή

$$[y_i | c_i = k] \sim MVN(\alpha_k, \Sigma_k),$$

όπου α_k είναι το διάνυσμα των μέσων τιμών για το $y = (y_1, y_2, \dots, y_M)$ στο προφίλ k , δηλαδή $E(y_{i|k}) = \alpha_k$, και Σ_k είναι ο πίνακας διακύμανσης – συνδιακύμανσης για το $y = (y_1, y_2, \dots, y_M)$ στο προφίλ k , δηλαδή $Var(y_{i|k}) = \Sigma_k$. Διαφορετικά η τελευταία εξίσωση μπορεί να γραφεί ως

$$y_{i|k} = \alpha_k + \varepsilon_{ik},$$

$$\varepsilon_{ik} = MVN(0, \Sigma_k).$$

Οι παράμετροι μέτρησης στη συνέχεια είναι οι μέσες τιμές, οι διακυμάνσεις και οι συνδιακυμάνσεις των μεταβλητών για κάθε προφίλ. Παρατηρούμε ότι ενώ υπάρχει συγκεκριμένη κατανομή των μεταβλητών σε κάθε προφίλ, δεν εξάγεται κάποιο συμπέρασμα για την κατανομή των μεταβλητών στον πληθυσμό.

Οι μέσες τιμές του $y = (y_1, y_2, \dots, y_M)$ είναι πάντα διαφορετικές σε κάθε προφίλ, ενώ οι διακυμάνσεις και οι συνδιακυμάνσεις εντός προφίλ μπορεί να είναι διαφορετικές σε κάθε προφίλ ή ίδιες. Επίσης, η ανεξαρτησία εντός προφίλ δεν είναι απαραίτητη (Masyn, 2013).

Οι παράμετροι του μοντέλου εκτιμώνται με τη μέθοδο μέγιστης πιθανοφάνειας. Έστω θ_k το σύνολο των παραμέτρων του προφίλ k . Το πλήρες σύνολο των παραμέτρων του μοντέλου μπορεί να γραφεί ως

$$\Theta = (\theta_1, \theta_2, \dots, \theta_k, \pi_1, \pi_2, \dots, \pi_k)$$

Δεδομένου του συνόλου δεδομένων Y με N ανεξάρτητες παρατηρήσεις η λογαριθμική συνάρτηση πιθανοφάνειας του μοντέλου με K προφίλ είναι

$$\log f(Y|\Theta) = \log \prod_{i=1}^N f(y_i|\Theta) = \sum_{i=1}^N \log \sum_{k=1}^K [\pi_k \cdot f_k(y_i|\theta_i)]$$

όπου το Θ μπορεί να εκτιμηθεί χρησιμοποιώντας τη λύση του εκτιμητή μέγιστης πιθανοφάνειας, που δίνεται από τον τύπο

$$\hat{\Theta} = \arg \max_{\Theta} \log f(Y|\Theta).$$

Συχνά, για τη μεγιστοποίηση των εκτιμήσεων της πιθανότητας των παραμέτρων του μοντέλου χρησιμοποιείται ο αλγόριθμος μεγιστοποίησης προσδοκίας. Ο αλγόριθμος μεγιστοποίησης προσδοκίας αντιμετωπίζει τη μη παρατηρούμενη ετερογένεια των δεδομένων (δηλαδή την ομαδοποίησή τους στα προφίλ) ως δεδομένα που λείπουν και, στη συνέχεια, δημιουργεί ένα σύνολο δεδομένων για το μοντέλο, βάσει του οποίου εκτιμώνται οι παράμετροι της LPA (He & Fan, 2019).

Κεφάλαιο 4. Η Ανάλυση Λανθανόντων Προφίλ (LPA) με τη χρήση της γλώσσας προγραμματισμού R

4.1. Προγράμματα για την ανάλυση δεδομένων με τη χρήση της Ανάλυσης Λανθανόντων Προφίλ (LPA)

Στη συνέχεια θα περιγράψουμε πώς μπορεί να εφαρμοστεί η Ανάλυση Λανθανόντων Προφίλ (LPA) με τη βοήθεια της τεχνολογίας και συγκεκριμένα με τη βοήθεια υπολογιστικών προγραμμάτων. Τα προγράμματα αυτά συμβάλουν στην αξιοπιστία και τη γρήγορη ανάλυση των δεδομένων, ενώ εξασφαλίζουν ότι η ανάλυση θα γίνει χωρίς λάθη, αφού ο αναλυτής δεν χρειάζεται να κάνει τις μαθηματικές πράξεις που απαιτούνται.

Υπάρχουν αρκετά προγράμματα με τη βοήθεια των οποίων εφαρμόζεται η Ανάλυση Λανθανόντων Προφίλ (LPA), όπως το Mplus, το Latent GOLD, το πρόσθετο Proc LCA & LTA του στατιστικού λογισμικού SAS (Hickendorff, Edelsbrunner, McMullen, Schneider & Trezise, 2018) και το RStudio με τα πακέτα mclust και tidyLPA (Wardenaar, 2021). Κάποια από αυτά διατίθενται δωρεάν στο διαδίκτυο. Ένα από αυτά θα χρησιμοποιήσουμε κι εμείς και συγκεκριμένα το Rstudio με το υπολογιστικό πακέτο tidyLPA.

Η R είναι μια στατιστική γλώσσα προγραμματισμού με ραγδαία ανάπτυξη, διατίθεται δωρεάν και συνοδεύεται από πολλά πακέτα (packages) τα οποία επίσης διατίθενται δωρεάν από χρήστες σε όλο τον κόσμο (Φωκιανός & Χαραλάμπους, 2010).

Στη συνέχεια του κεφαλαίου θα γίνει αναλυτική περιγραφή του προγράμματος RStudio και του υπολογιστικού πακέτου tidyLPA, καθώς και του τρόπου χρήσης τους, ενώ στο επόμενο κεφάλαιο θα προχωρήσουμε στην πρακτική εφαρμογή της Ανάλυσης Λανθανόντων Προφίλ (LPA).

4.2. Περιγραφή του προγράμματος RStudio

Για την γλώσσα προγραμματισμού R υπάρχει το περιβάλλον R, αλλά θα προτιμήσουμε το πρόγραμμα RStudio που έχει πιο οργανωμένη διάταξη και περισσότερες επιλογές.

Αρχικά εγκαθιστούμε στον υπολογιστή μας την R από την ιστοσελίδα <http://www.r-project.org> ακολουθώντας την εξής διαδικασία:

- Κάνουμε κλικ στο «download CRAN» στο αριστερό μενού
- Επιλέγουμε μια τοποθεσία (CRAN Mirror) για να κατεβάσουμε την R και συγκεκριμένα την Ελλάδα (<https://ftp.cc.uoc.gr/mirrors/CRAN/>)
- Επιλέγουμε το λειτουργικό μας σύστημα (π.χ. Windows)
- Κάνουμε κλικ στη βασική έκδοση (base distribution)
- Επιλέγουμε «Download R 4.1.0 for Windows»¹ και εγκαθιστούμε την R επιλέγοντας τις προεπιλεγμένες απαντήσεις.

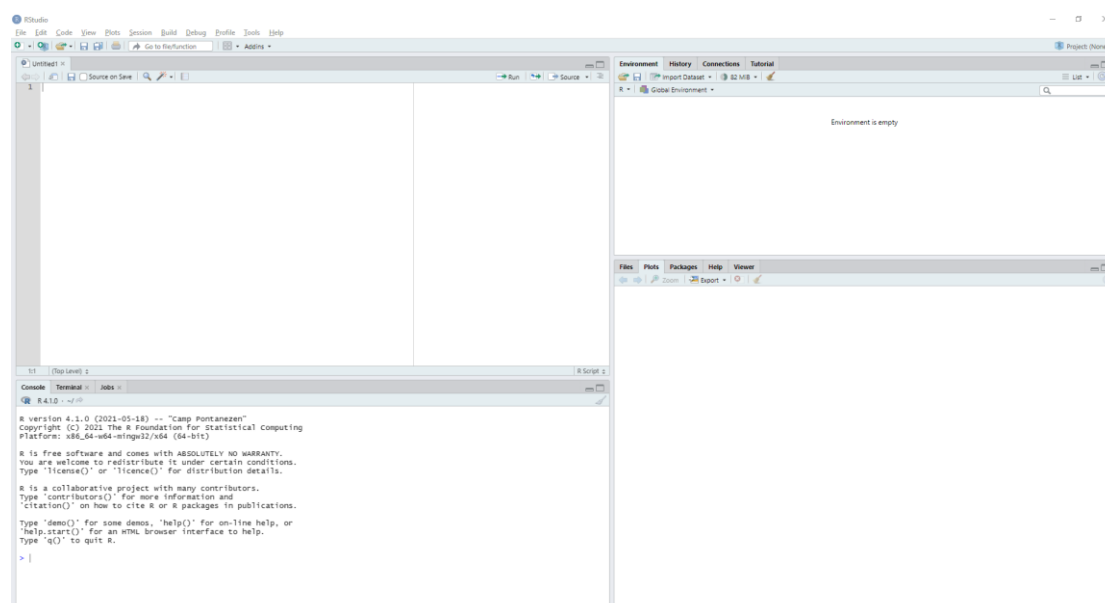
¹ Τη στιγμή της συγγραφής της εργασίας η 4.1.0 ήταν η τελευταία έκδοση. Επιλέγουμε την πιο πρόσφατη έκδοση.

Στη συνέχεια, εγκαθιστούμε το RStudio στον υπολογιστή μας, το οποίο και κατεβάζουμε από την ιστοσελίδα <http://www.rstudio.org> ακολουθώντας την εξής διαδικασία:

- Κάνουμε κλικ στο «DOWNLOAD» από το μενού
- Κάνουμε κλικ στο «Download RStudio Desktop»
- Κάνουμε κλικ στο «Download RStudio Desktop for Windows 1.4.1717»²
- Κατεβάζουμε το .exe αρχείο και εγκαθιστούμε το RStudio επιλέγοντας τις προεπιλεγμένες απαντήσεις (Torfs & Brauer, 2014).

Μετά την εγκατάσταση τρέχουμε το πρόγραμμα και βλέπουμε το περιβάλλον εργασίας, το οποίο φαίνεται στην Εικόνα 1.

Εικόνα 1. Περιβάλλον εργασίας προγράμματος RStudio



Το περιβάλλον εργασίας αποτελείται από 4 παράθυρα:

- Κάτω αριστερά είναι το console window ή command window. Πρόκειται για το πιο σημαντικό παράθυρο αφού σε αυτό η R κάνει όλη τη δουλειά. Σε αυτό μπορούμε να γράψουμε και να εκτελέσουμε μια εντολή μετά το «>».
- Πάνω αριστερά είναι το editor window ή script window. Αν δεν είναι ανοιχτό, μπορούμε να το ανοίξουμε επιλέγοντας File → New File → R script. Σε αυτό το παράθυρο καταχωρείται και αποθηκεύεται ο κώδικας, ενώ το «τρέξιμο» μίας εντολής ή όλου του κώδικα γίνεται με κλικ στο κουμπί «Run» ή με ctrl+Enter. Κατά το «τρέξιμο» των εντολών, αυτές περνάνε πρώτα στο console window και μετά εκτελούνται.

² Τη στιγμή της συγγραφής της εργασίας η 1.4.1717 ήταν η τελευταία έκδοση. Επιλέγουμε την πιο πρόσφατη έκδοση.

- Πάνω δεξιά είναι το workspace/history window. Σε αυτό το παράθυρο μπορούμε να δούμε τα δεδομένα στη μνήμη, καθώς και το ιστορικό των εντολών.
- Κάτω δεξιά είναι το files/plots/packages/help window. Σε αυτό το παράθυρο μπορούμε να ανοίξουμε αρχεία, να δούμε διαγράμματα, να εγκαταστήσουμε και να φορτώσουμε πακέτα ή να προσπελάσουμε τα αρχεία βοήθειας της R (Torfs & Brauer, 2014).

4.3. Περιγραφή του υπολογιστικού πακέτου tidyLPA

Για την Ανάλυση Λανθανόντων προφίλ (LPA) θα χρησιμοποιήσουμε το υπολογιστικό πακέτο tidyLPA μέσα από το RStudio. Ανοίγοντας τη λίστα των πακέτων στο packages window βλέπουμε ότι το tidyLPA δεν υπάρχει. Επομένως, πρέπει να το κατεβάσουμε από το διαδίκτυο και να το εγκαταστήσουμε. Η εγκατάσταση γίνεται είτε από τη διεύθυνση <https://CRAN.R-project.org/package=tidyLPA>, είτε μέσα από το RStudio και το packages window. Θα χρησιμοποιήσουμε τη δεύτερη μέθοδο επιλέγοντας Packages → Install, γράφουμε το όνομα του πακέτου που θέλουμε (tidyLPA) και προχωράμε σε εγκατάσταση.

Το tidyLPA είναι ένα πακέτο που επιτρέπει τον καθορισμό διαφορετικών μοντέλων με τα οποία εκτιμώνται κάποιες παράμετροι, καθώς και τον προσδιορισμό και τη σύγκριση διαφορετικών λύσεων βάσει του αριθμού των προφίλ που εκτιμώνται. Έχει σχεδιαστεί ώστε να είναι εύκολο στη χρήση ειδικά για αρχάριους στην LPA, αλλά με διαθέσιμες επιλογές για την εκτίμηση και αξιολόγηση πιο σύνθετων αναλύσεων. Το πακέτο tidyLPA λειτουργεί ως διεπαφή προς δύο διαφορετικά εργαλεία για την εκτίμηση μοντέλων, το πακέτο ανοιχτού κώδικα mclust και το εμπορικό λογισμικό MPlus.

Το πακέτο tidyLPA επιτρέπει τον καθορισμό των τεσσάρων μοντέλων παραμετροποίησης που παρουσιάστηκαν νωρίτερα στην Ενότητα 3.2:

- Ίσες διακυμάνσεις σε όλα τα προφίλ και συνδιακυμάνσεις σταθερές στο 0 (Μοντέλο 1 - EEI)
- Διαφορετικές διακυμάνσεις σε κάθε προφίλ και συνδιακυμάνσεις σταθερές στο 0 (Μοντέλο 2 - VVI)
- Ίσες διακυμάνσεις και ίσες συνδιακυμάνσεις σε όλα τα προφίλ (Μοντέλο 3 - EEE)
- Διαφορετικές διακυμάνσεις και διαφορετικές συνδιακυμάνσεις σε κάθε προφίλ (Μοντέλο 6 - VVV)

Τα άλλα δύο μοντέλα παραμετροποίησης (Μοντέλα 4 και 5) είναι διαθέσιμα μόνο μέσω του εμπορικού λογισμικού MPlus.

Οι δύο κύριες συναρτήσεις του πακέτου είναι οι `estimate_profiles()` και η `compare_solutions()`. Η πρώτη χρησιμοποιείται για την εκτίμηση ενός μοντέλου και η δεύτερη για τη σύγκριση διαφορετικών μοντέλων και πλήθους προφίλ. Η `estimate_profiles()` επιστρέφει την προβλεπόμενη πιθανότητα συμμετοχής κάθε μονάδας σε κάθε προφίλ και επιτρέπει την ερμηνεία της εξόδου του μοντέλου, ιδιαίτερα με τον συνδυασμό της συνάρτησης `plot_profiles()`, η οποία εμφανίζει τη μέση τιμή των μεταβλητών σε κάθε προφίλ. Η συνάρτηση `compare_solutions()` συγκρίνει

διάφορους δείκτες προσαρμογής διάφορων μοντέλων για κάθε παραμετροποίηση και αριθμό προφίλ (Rosenberg, Beymer, Anderson, van Lissa & Schmidt, 2018).

4.4. Διαδικασία εφαρμογής της Ανάλυσης Λανθανόντων Προφίλ (LPA) με τη χρήση του RStudio και του υπολογιστικού πακέτου tidyLPA

Οι εντολές πληκτρολογούνται στο editor window του RStudio και τρέχουν κάνοντας κλικ στο κουμπί «Run» ή πληκτρολογώντας Ctrl+Enter. Τα αποτελέσματα παράγονται στο console window.

Ξεκινώντας τη διαδικασία ορίζουμε το φάκελο στον οποίο βρίσκονται τα προς ανάλυση δεδομένα με την εντολή `setrw("διαδρομή_φακέλου")`. Για παράδειγμα:

```
setwd("/Users/Λευτέρης/Desktop/Διπλωματική/Data")
```

Στη συνέχεια, επειδή τα δεδομένα είναι αποθηκευμένα σε ένα αρχείο .sav (αρχείο SPSS), πρέπει να εισαχθούν με τις κατάλληλες εντολές στο RStudio. Για τη δουλειά αυτή χρησιμοποιούμε το πακέτο haven, το οποίο είναι ήδη εγκατεστημένο στο RStudio και απλώς το φορτώνουμε με τη χρήση της εντολής `library(haven)`, η οποία φορτώνει την ανάλογη βιβλιοθήκη. Αφότου φορτωθεί η βιβλιοθήκη, με τη χρήση της εντολής

```
data <- read_sav("Data.sav")
```

τα δεδομένα αποθηκεύονται στη μεταβλητή *data*, η οποία είναι τύπου data frame, δηλαδή ένας πίνακας με ονόματα στις στήλες.

Πλέον έφτασε η ώρα της ανάλυσης των δεδομένων, οπότε και καλούμε τη βιβλιοθήκη του πακέτου tidyLPA με την εντολή `library(tidyLPA)`. Επίσης, με την εντολή `library(dplyr)` φορτώνουμε το πακέτο dplyr το οποίο θα μας επιτρέψει τη χρήση του τελεστή `%>%` (pipe operator), ο οποίος επιτρέπει το άμεσο πέρασμα του αποτελέσματος μιας συνάρτησης ως όρισμα σε μια άλλη συνάρτηση.

Πριν προχωρήσουμε στην εκτίμηση των μοντέλων, με τη χρήση της εντολής `select()` επιλέγουμε τις μεταβλητές που θα χρησιμοποιήσουμε κατά την εκτίμηση. Αν εφαρμόσουμε την LPA στις μεταβλητές *var1*, *var2*, *var3*, *var4*, *var5* και αποθηκεύσουμε την επιλογή στη μεταβλητή *selected_data*. Δίνουμε την εντολή:

```
selected_data <- select(data, var1, var2, var3, var4, var5)
```

ή με τη χρήση του pipe operator

```
selected_data <- data %>% select(var1, var2, var3, var4, var5).
```

Η εκτίμηση των μοντέλων, όπως αναφέραμε, γίνεται με τη χρήση της συνάρτησης `estimate_profiles()`. Η συνάρτηση παίρνει ως ορίσματα το data frame με τα δεδομένα, δηλαδή τη μεταβλητή *selected_data*, τον αριθμό των προφίλ για τα οποία θέλουμε να γίνει εκτίμηση και το είδος του μοντέλου παραμετροποίησης ή των μοντέλων παραμετροποίησης που θέλουμε να εκτιμηθούν (1, 2, 3 ή 6). Αν θέλουμε να εκτιμήσουμε τα LPA μοντέλα με 1 έως 9 προφίλ και παραμετροποίηση EEI (ίσες διακυμάνσεις σε όλα τα προφίλ και συνδιακύμανσεις σταθερές στο 0) και να

αποθηκεύσουμε το αποτέλεσμα στη μεταβλητή *results* θα δώσουμε την εξής εντολή:

```
results <- estimate_profiles(selected_data, 1:9, models = 1)
```

ή χρησιμοποιώντας τον pipe operator:

```
results <- selected_data %>% estimate_profiles(1:9, models = 1).
```

Για να εμφανίσουμε το αποτέλεσμα της εκτίμησης δίνουμε την εντολή:

```
results.
```

Με τον ίδιο τρόπο μπορούμε να εκτιμήσουμε ταυτόχρονα μοντέλα LPA με διαφορετικές παραμετροποιήσεις. Για παράδειγμα, αν θέλουμε να εκτιμήσουμε τα LPA μοντέλα με 1 έως 9 προφίλ και παραμετροποιήσεις EEI (ίσες διακυμάνσεις σε όλα τα προφίλ και συνδιακυμάνσεις σταθερές στο 0) και VVI (διαφορετικές διακυμάνσεις σε κάθε προφίλ και συνδιακυμάνσεις σταθερές στο 0) και να αποθηκεύσουμε το αποτέλεσμα στη μεταβλητή *results* θα δώσουμε την εξής εντολή:

```
results <- estimate_profiles(selected_data, 1:9, models = c(1, 2))
```

ή χρησιμοποιώντας τον pipe operator:

```
results <- selected_data %>% estimate_profiles(1:9, models = c(1, 2)).
```

Για να εμφανίσουμε το αποτέλεσμα της εκτίμησης δίνουμε την εντολή:

```
results.
```

Για τη επιλογή του βέλτιστου μοντέλου παραμετροποίησης και αριθμού προφίλ συγκρίνουμε τις τιμές των στατιστικών κριτηρίων ή χρησιμοποιούμε τη συνάρτηση *compare_solutions()* η οποία κάνει αυτόματα τη σύγκριση. Έτσι δίνουμε την εξής εντολή:

```
comp <- compare_solutions(results)
```

ή χρησιμοποιώντας τον pipe operator:

```
comp <- results %>% compare_solutions().
```

Με την εντολή

```
comp$fits
```

εμφανίζονται οι τιμές των στατιστικών κριτηρίων για κάθε συνδυασμό μοντέλου παραμετροποίησης και αριθμού προφίλ, ενώ με την εντολή

```
comp$best
```

εμφανίζεται ποιος συνδυασμός είναι ο βέλτιστος βάσει κάθε κριτηρίου.

Κατά την εκτέλεση της συνάρτησης *compare_solutions()* μπορούμε να ορίσουμε συγκεκριμένο ή συγκεκριμένα στατιστικά κριτήρια βάσει των οποίων θα γίνει η σύγκριση. Αυτό γίνεται δίνοντας ένα ακόμα όρισμα στη συνάρτηση, οπότε στην

περίπτωση που θέλουμε να γίνει σύγκριση μόνο με τα κριτήρια BIC και AIC, η εντολή γίνεται:

```
comp <- compare_solutions(results, statistics = c("BIC",  
  "AIC"))
```

ή χρησιμοποιώντας τον pipe operator:

```
comp <- results %>% compare_solutions(statistics = c("BIC",  
  "AIC"))
```

Με την εντολή

```
comp
```

εμφανίζονται οι τιμές των κριτηρίων BIC και AIC για κάθε συνδυασμό μοντέλου παραμετροποίησης και αριθμού προφίλ, καθώς και ο βέλτιστος συνδυασμός βάσει κάθε επιλεγμένου στατιστικού κριτηρίου.

Τα κριτήρια BIC (Bayesian Information Criterion) και AIC (Akaike's Information Criterion) είναι οι δύο θεωρητικές μέθοδοι πληροφόρησης που χρησιμοποιούνται πιο συχνά για την επιλογή μοντέλων. Και τα δύο βασίζονται στην εκτίμηση μέγιστης πιθανότητας των παραμέτρων των προς επιλογή μοντέλων για την επιλογή του πιο σωστού (Tein, Coxe & Cham, 2013). Ο υπολογισμός τους γίνεται από τους εξής τύπους:

$$AIC = -2\log L(\hat{\theta}) + 2k$$

$$BIC = -2\log L(\hat{\theta}) + 2k \cdot \ln(n)$$

Όπου $L(\hat{\theta})$ η μέγιστη τιμή της συνάρτησης πιθανοφάνειας, k ο αριθμός των εκτιμώμενων παραμέτρων και n ο αριθμός των παρατηρήσεων.

Στην ανάλυση των δεδομένων μας, για την επιλογή του καλύτερου μοντέλου παραμετροποίησης και αριθμού προφίλ, θα χρησιμοποιήσουμε το κριτήριο BIC (Bayesian Information Criterion), αφού έχει διαπιστωθεί ότι είναι το πιο αξιόπιστο. Συγκρίνοντας, λοιπόν, τις τιμές του κριτηρίου BIC για κάθε ομαδοποίηση, επιλέγουμε την ομαδοποίηση με τη χαμηλότερη τιμή BIC, αλλά και με ικανό αριθμό προφίλ.

Αφού έχουμε επιλέξει το καλύτερο μοντέλο παραμετροποίησης και αριθμό προφίλ ξανατρέχουμε τη συνάρτηση `estimate_profiles()`, αυτή τη φορά με ορίσματα τον αριθμό προφίλ που επιλέξαμε και το μοντέλο παραμετροποίησης που επιλέξαμε. Αν για παράδειγμα τη μικρότερη τιμή BIC την είχε το μοντέλο παραμετροποίησης 1 (EEI - ίσες διακυμάνσεις σε όλα τα προφίλ και συνδιακύμανσεις σταθερές στο 0) με αριθμό προφίλ ίσο με 4, η εντολή διαμορφώνεται ως εξής:

```
best_model <- estimate_profiles(selected_data, 4, models =  
  1)
```

ή χρησιμοποιώντας τον pipe operator:

```
best_model <- selected_data %>% estimate_profiles(4,  
  models = 1)
```

Η παράμετρος `models = 1` θα μπορούσε να παραληφθεί αφού το μοντέλο 1 είναι το default μοντέλο.

Εναλλακτικά, δεδομένου ότι με την πρώτη εφαρμογή της συνάρτησης *estimate_profiles()* έχουν εκτιμηθεί όλα τα μοντέλα, αντί να ξανακάνουμε εκτίμηση του καλύτερου μοντέλου, μπορούμε να ανακτήσουμε τα δεδομένα για αυτό μέσω της εντολής

```
best_model <- results[[names(results)[comp$best[4]]]]
```

αφού το κριτήριο BIC εμφανίζεται τέταρτο στο αποτέλεσμα της εντολής *comp\$best*.

Στην περίπτωση που κατά την εκτέλεση της *compare_solutions()* έχουμε επιλέξει να μη γίνει σύγκριση των μοντέλων με όλα τα κριτήρια η παραπάνω εντολή πρέπει να τροποποιηθεί κατάλληλα, αφού το κριτήριο BIC μπορεί να μην είναι το τέταρτο στη σειρά των κριτηρίων.

Για την ανάλυση των αποτελεσμάτων της εκτίμησης χρησιμοποιούμε κάποιες συναρτήσεις ακόμα, τις *get_data()*, *get_estimates()* και *get_fit()*.

Για να εμφανίσουμε την πιθανότητα κάθε μονάδας να ανήκει σε κάθε προφίλ χρησιμοποιούμε την εντολή

```
get_data(best_model)
```

ενώ για να εμφανίσουμε την εκτίμηση για την μέση τιμή και τη διακύμανση κάθε μεταβλητής σε κάθε προφίλ χρησιμοποιούμε την εντολή

```
get_estimates(best_model).
```

Ακόμη, για να εμφανίσουμε τις τιμές των στατιστικών κριτηρίων του καλύτερου μοντέλου χρησιμοποιούμε την εντολή

```
get_fit(best_model)
```

ενώ με τη συνάρτηση *plot_profiles()* μπορούμε να εμφανίσουμε το διάγραμμα με τις εκτιμήσεις για τη μέση τιμή και την τυπική απόκλιση κάθε μεταβλητής σε κάθε προφίλ. Για τη δημιουργία του διαγράμματος, αν θέλουμε να εμφανιστούν μόνο οι εκτιμήσεις για τη μέση τιμή, χρησιμοποιούμε την εντολή

```
plot_profiles(best_model, add_line = TRUE, sd = FALSE, ci = NULL)
```

ενώ αν θέλουμε να εμφανιστούν και οι εκτιμήσεις για τις τυπικές αποκλίσεις, χρησιμοποιούμε την εντολή

```
plot_profiles(best_model, add_line = TRUE, sd = TRUE, ci = NULL).
```

Αν θέλουμε να εμφανίσουμε το ποσοστό του πληθυσμού που εκτιμάται ότι θα ανήκει σε κάθε προφίλ χρησιμοποιούμε την εντολή

```
best_model$model$parameters$pro
```

ενώ αν θέλουμε να εμφανίσουμε μόνο τις εκτιμώμενες μέσες τιμές κάθε μεταβλητής σε κάθε προφίλ χρησιμοποιούμε την εντολή

```
best_model$model$parameters$mean.
```

Τέλος, αν θέλουμε να εμφανίσουμε τους πίνακες διακύμανσης – συνδιακύμανσης για όλα τα προφίλ χρησιμοποιούμε την εντολή

```
best_model$model$parameters$variance$sigma.
```

Σε αυτό το σημείο έχουμε ολοκληρώσει την παρουσίαση της διαδικασίας με την οποία εφαρμόζεται η LPA στο RStudio με τη βοήθεια του υπολογιστικού πακέτου tidyLPA³. Στο επόμενο κεφάλαιο θα παρουσιάσουμε μια πρακτική εφαρμογή της διαδικασίας, εκτελώντας ένα παράδειγμα ανάλυσης δεδομένων με τη χρήση της LPA.

³ Για πιο αναλυτική περιγραφή του πακέτου tidyLPA μπορούμε να ανατρέξουμε στο εγχειρίδιο αναφοράς (Reference manual) του πακέτου που βρίσκεται στη διεύθυνση <https://CRAN.R-project.org/package=tidyLPA>.

Κεφάλαιο 5. Εφαρμογή της Ανάλυσης Λανθανόντων Προφίλ (LPA)

5.1. Δεδομένα

Στη συνέχεια της παρούσης εργασίας θα πραγματοποιήσουμε ανάλυση δεδομένων από μία δειγματοληπτική έρευνα με τη βοήθεια της Ανάλυσης Λανθανόντων Προφίλ (LPA). Στόχος της ανάλυσης είναι η ομαδοποίηση των αποτελεσμάτων της έρευνας, ώστε να γίνει πιο εύκολη η απεικόνισή τους και η εξαγωγή των συμπερασμάτων.

Τα δεδομένα αντλήθηκαν από μια έρευνα που έγινε σε 17 σχολεία του Νομού Αττικής το σχολικό έτος 2015-2016. Το πλήθος και το ποσοστό των παιδιών ανά τάξη και φύλο φαίνεται στον Πίνακα 5.1.

Πίνακας 5.1. Πλήθος και ποσοστό παιδιών ανά τάξη και φύλο

Τάξη	Φύλο		Σύνολο
	Αγόρια	Κορίτσια	
Ε΄ Δημοτικού	43	26	69 (14,3%)
ΣΤ΄ Δημοτικού	39	35	74 (15,3%)
Α΄ Γυμνασίου	132	175	307 (63,4%)
Β΄ Γυμνασίου	9	25	34 (7%)
Σύνολο	223 (46%)	261 (54%)	484

Το δείγμα αποτελούταν από 484 παιδιά, με τα κορίτσια να αποτελούν το 54% του δείγματος. Όσον αφορά στη σχολική τάξη η πλειοψηφία των παιδιών πηγαίνει στην Α΄ Γυμνασίου (63,4%), ένα μικρό ποσοστό (7%) πηγαίνει στην Β΄ Γυμνασίου και το υπόλοιπο ποσοστό σχεδόν μοιράζεται στην Ε΄ και την ΣΤ΄ Δημοτικού (14,3% και 15,3% αντίστοιχα).

Στην έρευνα χρησιμοποιήθηκαν δύο «εργαλεία» μέτρησης: το ερωτηματολόγιο Μαθητών του Ευρωπαϊκού Προγράμματος «ENABLE», ενδυναμώνοντας τα παιδιά και το Ερωτηματολόγιο Δυνατοτήτων και Δυσκολιών (SDQ) για τις ηλικίες 4 – 17⁴.

Σε αυτή την εργασία θα ασχοληθούμε με την ανάλυση των αποτελεσμάτων που προέκυψαν από το Ερωτηματολόγιο Δυνατοτήτων και Δυσκολιών (SDQ). Το Ερωτηματολόγιο Δυνατοτήτων και Δυσκολιών (SDQ) κατασκευάστηκε από τον Goodman το 1997 και είναι ένα σύντομο εργαλείο ανίχνευσης προβλημάτων στη συμπεριφορά και τα συναισθήματα σε παιδιά και εφήβους 4 έως 17 ετών.

Το SDQ είναι ερωτηματολόγιο αυτοαναφοράς και περιέχει 25 ερωτήσεις, οι οποίες χωρίζονται σε 5 κλίμακες των 5 ερωτήσεων η καθεμία. Για κάθε ερώτηση υπάρχουν οι απαντήσεις «Δεν ισχύει», «Ισχύει κάπως» και «Ισχύει σίγουρα». Η απάντηση «Ισχύει κάπως» βαθμολογείται πάντα με 1. Οι απαντήσεις «Δεν ισχύει» και «Ισχύει σίγουρα» στις περισσότερες ερωτήσεις (ερωτήσεις με «αρνητική» διατύπωση) βαθμολογούνται με 0 και 2 αντίστοιχα, ενώ σε ερωτήσεις με «θετική» διατύπωση βαθμολογούνται με 2

⁴ Το Ερωτηματολόγιο Δυνατοτήτων και Δυσκολιών (SDQ) παρατίθεται στο Παράρτημα Ι

και 0 αντίστοιχα. Έτσι κάθε κλίμακα βαθμολογείται από 0 έως 10, αν έχουν απαντηθεί όλες οι ερωτήσεις της κλίμακας.

Οι πέντε κλίμακες και οι ερωτήσεις που περιλαμβάνονται στην κάθε μία είναι οι εξής:

- Κλίμακα 1 - Συναισθηματικά προβλήματα (Emotional problems scale): ερωτήσεις 3, 8, 13, 16, 24
- Κλίμακα 2 - Προβλήματα συμπεριφοράς (Conduct problems scale): ερωτήσεις 5, 7, 12, 18, 22
- Κλίμακα 3 – Υπερκινητικότητα, διάσπαση προσοχής (Hyperactivity scale): ερωτήσεις 2, 10, 15, 21, 25
- Κλίμακα 4 - Προβλήματα με τους συνομηλίκους (Peer problems scale): ερωτήσεις 6, 11, 14, 19, 23
- Κλίμακα 5 - Κοινωνική συμπεριφορά (Prosocial scale): ερωτήσεις 1, 4, 9, 17, 20

Τα δεδομένα που θα αναλύσουμε προκύπτουν από το σκορ των μαθητών στις 5 αυτές κλίμακες. Ο όγκος των δεδομένων ($5 * 484 = 2.420$ σκορ) απαιτεί πολύ χρόνο για να ταξινομηθεί και σε αυτό βοηθάνε οι στατιστικές μέθοδοι ανάλυσης με την ανάλογη τεχνολογία. Αυτό που θα κάνουμε είναι να ομαδοποιήσουμε τους μαθητές, ανάλογα με το σκορ που συγκέντρωσαν, με τη βοήθεια της LPA. Το μέγεθος του δείγματος που έχουμε (484) είναι κοντά στο 500 και επομένως ικανό να οδηγήσει με αρκετή ακρίβεια στον εντοπισμό ενός σωστού αριθμού λανθανόντων προφίλ (Spruk κ.ά., 2020)

Τα δεδομένα απαρτίζουν έναν πίνακα δεδομένων του SPSS (αρχείο Data.sav), όπου κάθε γραμμή αποτελεί απάντηση ενός παιδιού και κάθε στήλη αντιστοιχεί σε μια μεταβλητή. Στην πρώτη στήλη περιέχεται η μεταβλητή Gender (Φύλο), η οποία παίρνει τις τιμές 1 για αγόρι ή 2 για κορίτσι, ενώ στη δεύτερη στήλη περιέχεται η μεταβλητή Grade (Τάξη), η οποία παίρνει τις τιμές 1 για την Ε΄ Δημοτικού, 2 για την ΣΤ΄ Δημοτικού, 3 για την Α΄ Γυμνασίου και 4 για την Β΄ Γυμνασίου. Ακολουθούν κάποιες άλλες μεταβλητές που αντιστοιχούν σε άλλες ερωτήσεις της έρευνας, και στις τελευταίες 5 στήλες περιέχονται οι μεταβλητές που αντιστοιχούν στις 5 κλίμακες του SDQ. Αυτές είναι οι *sdq_em*, *sdq_cp*, *sdq_hy*, *sdq_pp* και *sdq_ps* και η αντιστοιχία τους με τις κλίμακες του SDQ φαίνεται στον Πίνακα 5.2.

Πίνακας 5.2. Αντιστοιχία κλιμάκων και μεταβλητών

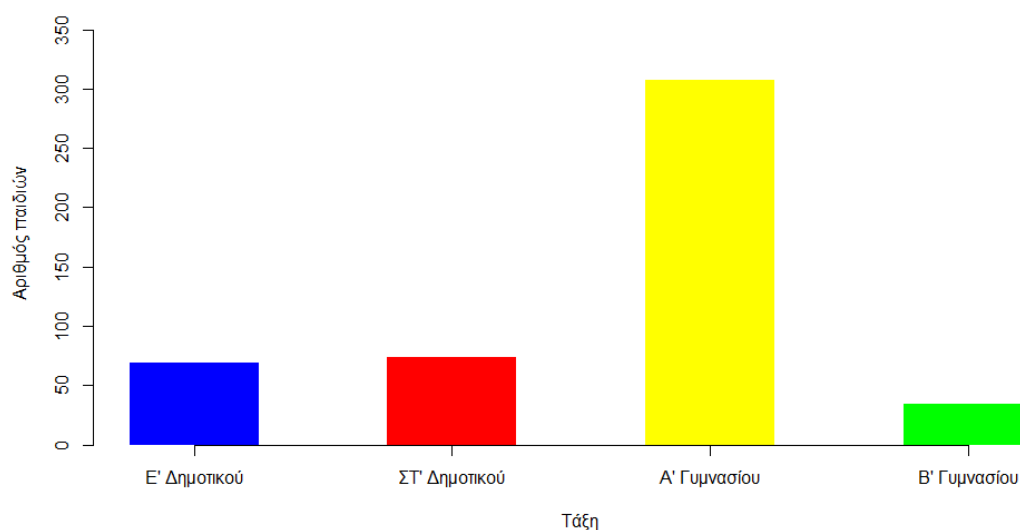
Κλίμακα	Μεταβλητή
Συναισθηματικά προβλήματα (Emotional problems scale)	<i>sdq_em</i>
Προβλήματα συμπεριφοράς (Conduct problems scale)	<i>sdq_cp</i>
Υπερκινητικότητα (Hyperactivity scale)	<i>sdq_hy</i>
Προβλήματα με τους συνομηλίκους (Peer problems scale)	<i>sdq_pp</i>
Κοινωνική συμπεριφορά (Prosocial scale)	<i>sdq_ps</i>

5.2. Παρουσίαση των δεδομένων της έρευνας με Περιγραφική Στατιστική.

Πριν προχωρήσουμε στην ανάλυση των δεδομένων με τη βοήθεια της LPA, θα περιγράψουμε τα αποτελέσματα του SDQ με τη βοήθεια της Περιγραφικής Στατιστικής.

Στο Γράφημα 1 απεικονίζεται ο πληθυσμός των παιδιών σε κάθε τάξη. Παρατηρούμε ότι ο πληθυσμός των παιδιών της Α΄ Γυμνασίου είναι σημαντικά μεγαλύτερος από τον πληθυσμό των παιδιών των άλλων τάξεων. Επίσης, ο πληθυσμός των παιδιών των τάξεων του Δημοτικού είναι στο ίδιο επίπεδο, ενώ ο πληθυσμός των παιδιών της Β΄ Γυμνασίου είναι πολύ μικρός.

Γράφημα 1. Πληθυσμός παιδιών ανά τάξη

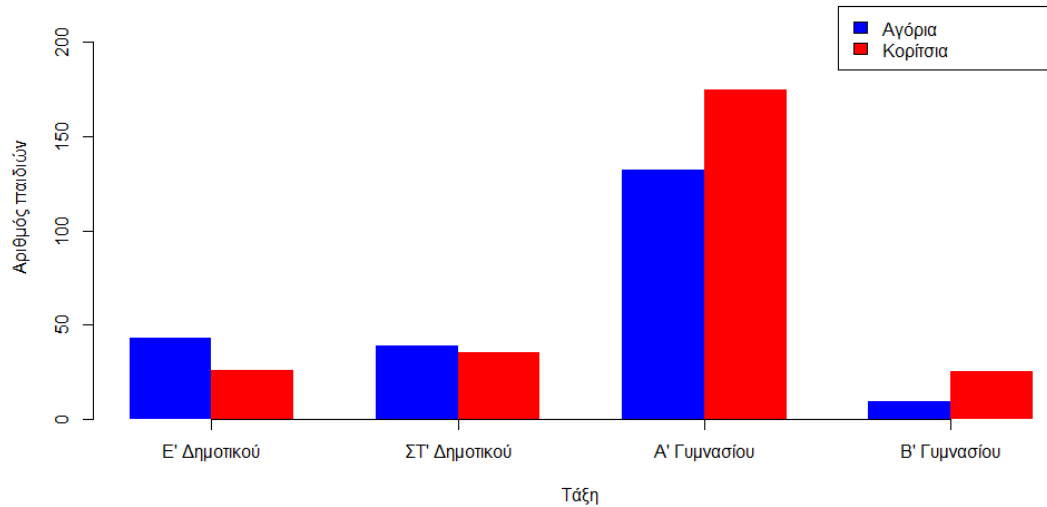


Στο Γράφημα 2 απεικονίζεται ο πληθυσμός των παιδιών ανά τάξη και φύλο. Παρατηρούμε ότι στις τάξεις του Δημοτικού ο πληθυσμός των αγοριών είναι μεγαλύτερος από αυτόν των κοριτσιών, ενώ στις τάξεις του Γυμνασίου συμβαίνει το αντίθετο. Μάλιστα στην Ε΄ Δημοτικού και στις τάξεις του Γυμνασίου η διαφορά είναι μεγάλη.

Στον Πίνακα 5.3 βλέπουμε τη μέση τιμή κάθε κλίμακας του SDQ ανά φύλο και τάξη. Παρατηρούμε ότι:

- στην κλίμακα 1 τα αγόρια έχουν σημειώσει μικρότερο μέσο σκορ από τα κορίτσια σε όλες τις τάξεις πλην της Ε΄ Δημοτικού
- στην κλίμακα 2 και την κλίμακα 3 τα αγόρια έχουν σημειώσει μεγαλύτερο μέσο σκορ από τα κορίτσια σε όλες τις τάξεις
- στην κλίμακα 4 τα αγόρια έχουν σημειώσει μεγαλύτερο μέσο σκορ από τα κορίτσια σε όλες τις τάξεις πλην της Β΄ Γυμνασίου
- στην κλίμακα 5 τα κορίτσια έχουν σημειώσει μεγαλύτερο μέσο σκορ από τα αγόρια σε όλες τις τάξεις.

Γράφημα 2. Πληθυσμός παιδιών ανά τάξη και φύλο

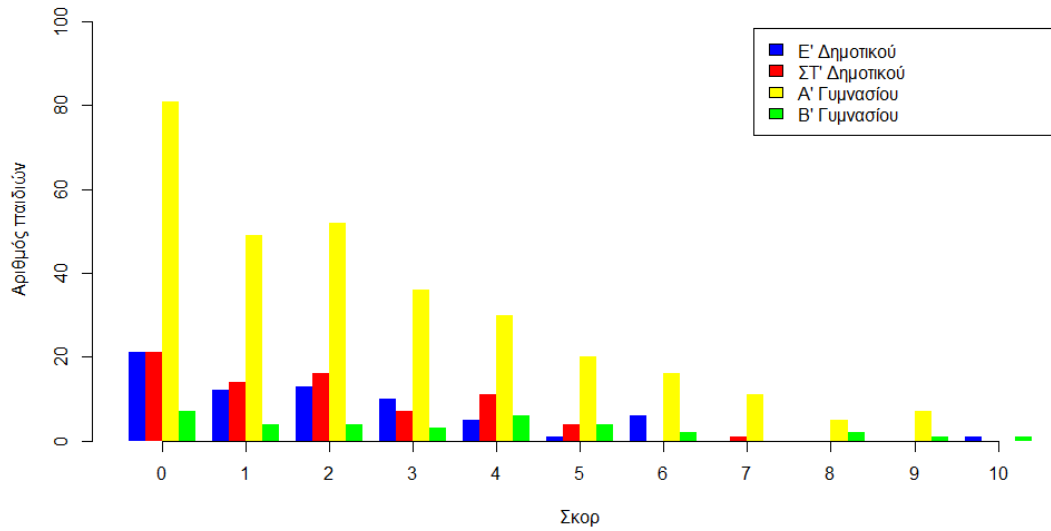


Πίνακας 5.3. Μέση τιμή κλιμάκων του SDQ ανά φύλο και τάξη

Φύλο/Τάξη	Κλίμακα				
	1	2	3	4	5
Αγόρια					
Ε' Δημοτικού	2,140	2,721	3,000	1,651	8,581
ΣΤ' Δημοτικού	1,667	2,359	2,465	1,641	7,308
Α' Γυμνασίου	1,939	2,667	2,871	1,568	7,947
Β' Γυμνασίου	1,444	2,333	3,556	1,444	7,667
Κορίτσια					
Ε' Δημοτικού	1,808	1,962	2,231	0,8846	9,000
ΣΤ' Δημοτικού	2,086	1,857	2,257	1,171	8,800
Α' Γυμνασίου	2,863	2,434	2,737	1,469	8,286
Β' Γυμνασίου	3,960	2,120	2,800	1,520	8,960

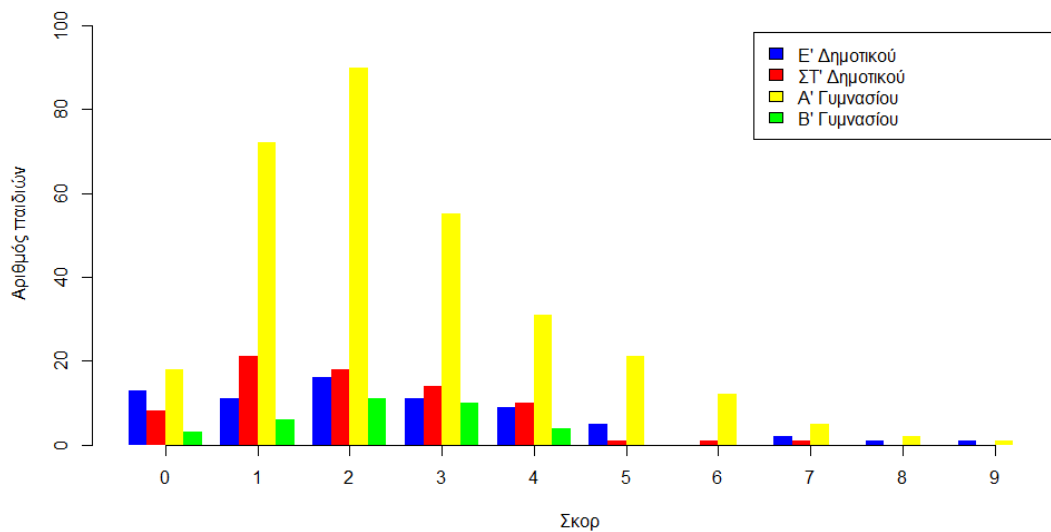
Στο Γράφημα 3 βλέπουμε το πλήθος των παιδιών ανά σκορ και τάξη για τη μεταβλητή *sdq_em* που αντιστοιχεί στην Κλίμακα 1 - Συναισθηματικά προβλήματα (Emotional problems scale) του SDQ. Παρατηρούμε ότι τα παιδιά όλων των τάξεων στην Κλίμακα 1 του SDQ συγκέντρωσαν χαμηλό σκορ και μάλιστα όσο αυξάνεται το σκορ μειώνεται το πλήθος των παιδιών που το συγκέντρωσε. Τα αποτελέσματα για την Κλίμακα 1 του SDQ δείχνουν ότι τα περισσότερα παιδιά αντιμετωπίζουν ελάχιστα έως καθόλου συναισθηματικά προβλήματα, ενώ λίγα είναι αυτά που αντιμετωπίζουν σοβαρά συναισθηματικά προβλήματα.

Γράφημα 3. Πλήθος παιδιών ανά σκορ και τάξη για την Κλίμακα 1 του SDQ



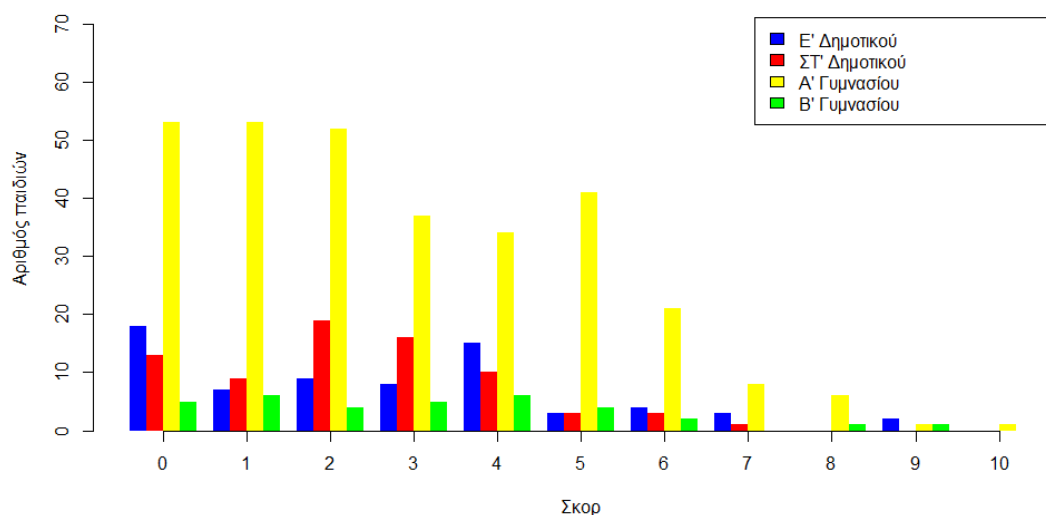
Στο Γράφημα 4 βλέπουμε το πλήθος των παιδιών ανά σκορ και τάξη για τη μεταβλητή *sdq_cr* που αντιστοιχεί στην Κλίμακα 2 - Προβλήματα συμπεριφοράς (Conduct problems scale) του SDQ. Παρατηρούμε ότι τα περισσότερα παιδιά όλων των τάξεων στην Κλίμακα 2 του SDQ συγκέντρωσαν σκορ κοντά στο 2. Και σε αυτή την Κλίμακα, στα μεγάλα σκορ το πλήθος των παιδιών είναι μικρό, ενώ κανένα παιδί δεν συγκέντρωσε σκορ ίσο με 10. Τα αποτελέσματα για την Κλίμακα 2 του SDQ δείχνουν ότι τα περισσότερα παιδιά αντιμετωπίζουν ελάχιστα προβλήματα συμπεριφοράς, ενώ ο αριθμός αυτών που αντιμετωπίζουν σοβαρά προβλήματα συμπεριφοράς είναι πολύ μικρός.

Γράφημα 4. Πλήθος παιδιών ανά σκορ και τάξη για την Κλίμακα 2 του SDQ



Στο Γράφημα 5 βλέπουμε το πλήθος των παιδιών ανά σκορ και τάξη για τη μεταβλητή *sdq_hy* που αντιστοιχεί στην Κλίμακα 3 – Υπερκινητικότητα, διάσπαση προσοχής (Hyperactivity scale) του SDQ. Παρατηρούμε ότι και στην Κλίμακα 3 του SDQ, όπως και στις προηγούμενες κλίμακες, τα περισσότερα παιδιά όλων των τάξεων συγκέντρωσαν σκορ μικρότερο ή του 5. Επίσης, και σε αυτή την κλίμακα ελάχιστα παιδιά συγκέντρωσαν υψηλό σκορ. Τα αποτελέσματα για την Κλίμακα 3 του SDQ δείχνουν ότι τα περισσότερα παιδιά αντιμετωπίζουν μέτρια έως καθόλου προβλήματα υπερκινητικότητας και διάσπασης προσοχής. Ένας πολύ μικρός αριθμός παιδιών και σε αυτή τη κλίμακα αντιμετωπίζει σοβαρά προβλήματα υπερκινητικότητας και διάσπασης προσοχής.

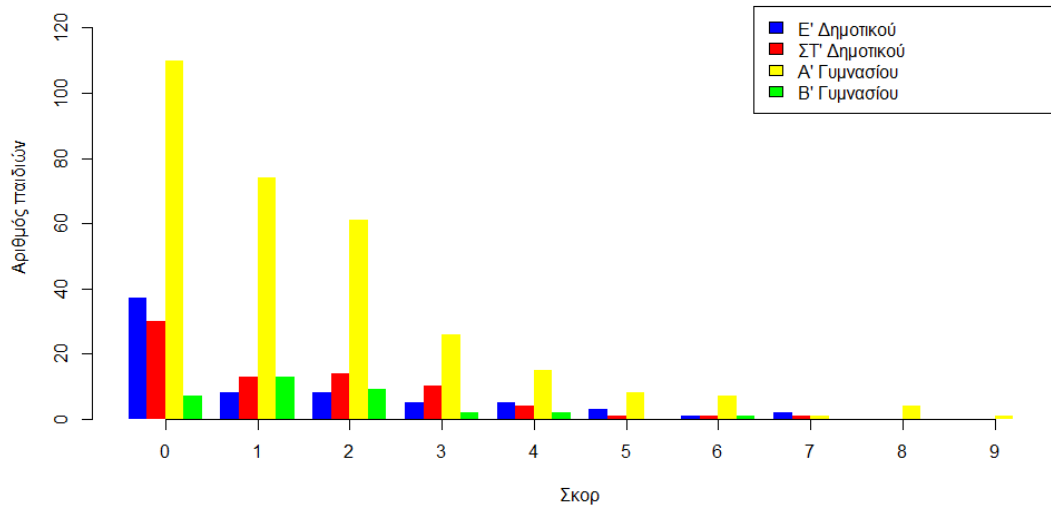
Γράφημα 5. Πλήθος παιδιών ανά σκορ και τάξη για την Κλίμακα 3 του SDQ



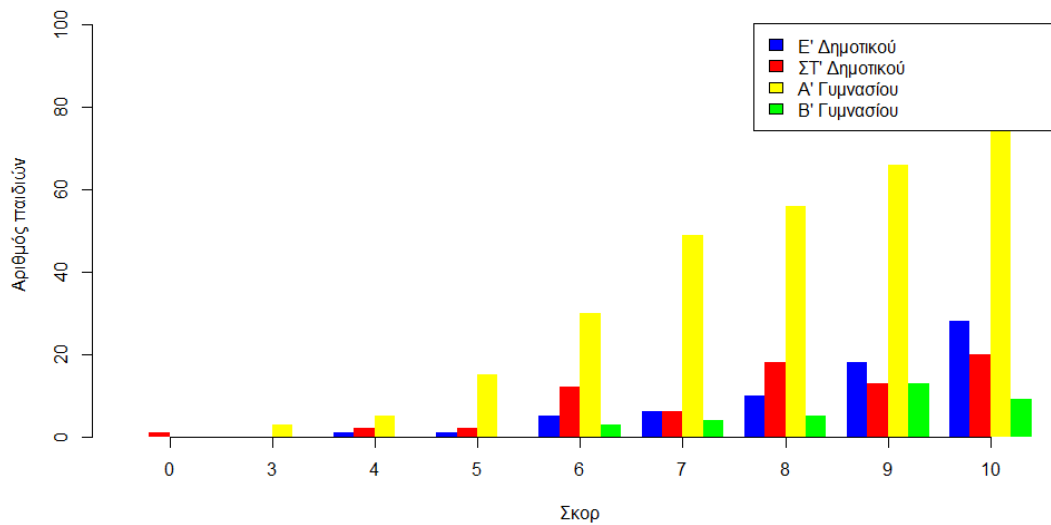
Στο Γράφημα 6 βλέπουμε το πλήθος των παιδιών ανά σκορ και τάξη για τη μεταβλητή *sdq_pp* που αντιστοιχεί στην Κλίμακα 4 - Προβλήματα με τους συνομηλίκους (Peer problems scale) του SDQ. Παρατηρούμε ότι τα περισσότερα παιδιά όλων των τάξεων συγκέντρωσαν πολύ χαμηλό σκορ στην κλίμακα 4 του SDQ. Επίσης, ελάχιστα παιδιά συγκέντρωσαν υψηλό σκορ, ενώ κανένα παιδί δεν συγκέντρωσε σκορ ίσο με 10. Τα αποτελέσματα για την Κλίμακα 4 του SDQ δείχνουν ότι τα περισσότερα παιδιά δεν αντιμετωπίζουν ιδιαίτερα προβλήματα με τους συνομηλίκους τους. Και σε αυτή την κλίμακα ελάχιστα είναι αυτά που αποτελούν την εξαίρεση.

Στο Γράφημα 7 βλέπουμε το πλήθος των παιδιών ανά σκορ και τάξη για τη μεταβλητή *sdq_ps* που αντιστοιχεί στην Κλίμακα 5 - Κοινωνική συμπεριφορά (Prosocial scale) του SDQ. Παρατηρούμε ότι τα περισσότερα παιδιά όλων των τάξεων στην Κλίμακα 5 του SDQ συγκέντρωσαν πολύ υψηλό σκορ. Επίσης, ελάχιστα παιδιά συγκέντρωσαν χαμηλό σκορ, ενώ κανένα δεν συγκέντρωσε σκορ ίσο με 1 ή 2. Τα αποτελέσματα για την Κλίμακα 5 του SDQ δείχνουν ότι τα περισσότερα παιδιά έχουν καλή κοινωνική συμπεριφορά.

Γράφημα 6. Πλήθος παιδιών ανά σκορ και τάξη για την Κλίμακα 4 του SDQ



Γράφημα 7. Πλήθος παιδιών ανά σκορ και τάξη για την Κλίμακα 5 του SDQ



Τα παραπάνω γραφήματα παράχθηκαν από το RStudio με τη χρήση του παρακάτω κώδικα:

```
barplot(table(data$Grade), names.arg = c("Ε' Δημοτικού",  
"ΣΤ' Δημοτικού", "Α' Γυμνασίου", "Β' Γυμνασίου"), width  
= 1, ylim = c(0, max(table(data$Grade))+50), col =  
c("blue", "red", "yellow", "green"), axis.lty = 1,  
border = NA, xlab = "Τάξη", ylab = "Αριθμός παιδιών",  
space = 1, legend = FALSE).
```

```

barplot(table(data$Gender, data$Grade), names.arg = c("Ε'
  Δημοτικού", "ΣΤ' Δημοτικού", "Α' Γυμνασίου", "Β'
  Γυμνασίου"), width = 1, beside = TRUE, ylim = c(0,
  max(table(data$Gender, data$Grade))+50), legend.text =
  c("Αγόρια", "Κορίτσια"), col = c("blue", "red"),
  axis.lty = 1, border = NA, xlab = "Τάξη", ylab =
  "Αριθμός παιδιών").

arplot(table(data$Grade, data$sdq_em), names.arg =
  colnames(y), width = 1, beside = TRUE, ylim = c(0,
  max(table(data$Grade, data$sdq_em))+20), legend.text =
  c("Ε' Δημοτικού", "ΣΤ' Δημοτικού", "Α' Γυμνασίου", "Β'
  Γυμνασίου"), col = c("blue", "red", "yellow", "green"),
  axis.lty = 1, border = NA, xlab = "Σκορ", ylab =
  "Αριθμός παιδιών").

barplot(table(data$Grade, data$sdq_cp), names.arg =
  colnames(y), width = 1, beside = TRUE, ylim = c(0,
  max(table(data$Grade, data$sdq_em))+20), legend.text =
  c("Ε' Δημοτικού", "ΣΤ' Δημοτικού", "Α' Γυμνασίου", "Β'
  Γυμνασίου"), col = c("blue", "red", "yellow", "green"),
  axis.lty = 1, border = NA, xlab = "Σκορ", ylab =
  "Αριθμός παιδιών").

barplot(table(data$Grade, data$sdq_hy), names.arg =
  colnames(table(data$Grade, data$sdq_hy)), width = 1,
  beside = TRUE, ylim = c(0, max(table(data$Grade,
  data$sdq_hy))+20), legend.text = c("Ε' Δημοτικού",
  "ΣΤ' Δημοτικού", "Α' Γυμνασίου", "Β' Γυμνασίου"), col
  = c("blue", "red", "yellow", "green"), axis.lty = 1,
  border = NA, xlab = "Σκορ", ylab = "Αριθμός παιδιών").

barplot(table(data$Grade, data$sdq_pp), names.arg =
  colnames(table(data$Grade, data$sdq_pp)), width = 1,
  beside = TRUE, ylim = c(0, max(table(data$Grade,
  data$sdq_pp))+20), legend.text = c("Ε' Δημοτικού",
  "ΣΤ' Δημοτικού", "Α' Γυμνασίου", "Β' Γυμνασίου"), col
  = c("blue", "red", "yellow", "green"), axis.lty = 1,
  border = NA, xlab = "Σκορ", ylab = "Αριθμός παιδιών").

barplot(table(data$Grade, data$sdq_ps), names.arg =
  colnames(table(data$Grade, data$sdq_ps)), width = 1,
  beside = TRUE, ylim = c(0, max(table(data$Grade,
  data$sdq_ps))+20), legend.text = c("Ε' Δημοτικού",
  "ΣΤ' Δημοτικού", "Α' Γυμνασίου", "Β' Γυμνασίου"), col
  = c("blue", "red", "yellow", "green"), axis.lty = 1,
  border = NA, xlab = "Σκορ", ylab = "Αριθμός παιδιών").

```

5.3. Προετοιμασία των δεδομένων

Όπως αναφέραμε τα δεδομένα περιέχονται σε ένα αρχείο .sav, οπότε και τα εισάγουμε στο RStudio με τις κατάλληλες εντολές:

```
library(haven)
setwd("/Users/Λευτέρης/Desktop/Διπλωματική/Data")
data <- read_sav("Data.sav").
```

Μετά την εκτέλεση των παραπάνω εντολών τα δεδομένα έχουν αποθηκευτεί στη μεταβλητή *data* η οποία είναι τύπου data frame.

Πλέον μπορούμε να προχωρήσουμε στην εφαρμογή της LPA για να αναλύσουμε και να ομαδοποιήσουμε τα δεδομένα. Με την LPA θα σχηματίσουμε κάποια προφίλ που το καθένα θα περιέχει ένα μέρος του πληθυσμού με παρόμοια σκορ στις κλίμακες του SDQ. Το RStudio με τη χρήση του πακέτου tidyLPA θα υπολογίσει το εκτιμώμενο ποσοστό του πληθυσμού σε κάθε προφίλ, αλλά και την εκτιμώμενη μέση τιμή και διακύμανση των κλιμάκων του SDQ σε κάθε προφίλ.

5.4. Εφαρμογή της LPA στα δεδομένα

Για την εφαρμογή της LPA γράφουμε τις εντολές που έχουμε αναφέρει στην Ενότητα 4.4 στο editor window του RStudio. Εκτελούμε την ανάλυση για όλα τα μοντέλα παραμετροποίησης (1, 2, 3 και 6) και για 1 έως 9 προφίλ. Μετά τη σύγκριση των αποτελεσμάτων, επιλέγουμε εκείνη την επιλογή μοντέλου παραμετροποίησης και αριθμού προφίλ με την μικρότερη τιμή στο κριτήριο BIC. Τέλος, χρησιμοποιώντας τις εντολές που είδαμε στην Ενότητα 4.4, εμφανίζουμε το εκτιμώμενο ποσοστό του πληθυσμού σε κάθε προφίλ, καθώς και την εκτιμώμενη μέση τιμή και διακύμανση των μεταβλητών σε κάθε προφίλ.

Ο κώδικας που χρησιμοποιούμε είναι ο παρακάτω:

```
library(tidyLPA)
library(dplyr)
results <- data %>%
  select(sdq_em, sdq_cp, sdq_hy, sdq_pp, sdq_ps) %>%
  single_imputation() %>%
  estimate_profiles(1:9, models = c(1, 2, 3, 6))
comp <- compare_solutions(results)
comp$fits
comp$best
best_model <- results[[names(results)[comp$best[4]]]]
get_data(best_model)
get_estimates(best_model)
get_fit(best_model)
```

```

plot_profiles(best_model, add_line = TRUE, sd = TRUE)
best_model$model$parameters$pro
best_model$model$parameters$mean
best_model$model$parameters$variance$sigma

```

Τα αποτελέσματα με τη μορφή που παράγονται στο RStudio δίνονται στο Παράρτημα II, ενώ εδώ παρατίθεται ένα μέρος από αυτά υπό μορφή πίνακα.

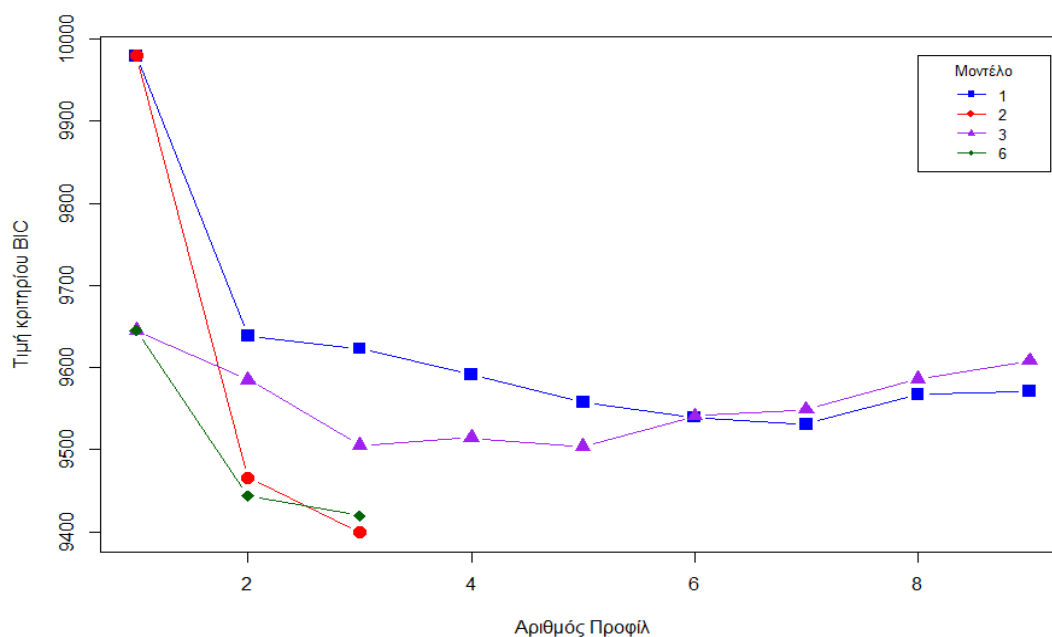
Από το αποτέλεσμα της σύγκρισης και με τη βοήθεια της εντολής `comp$fits` παίρνουμε τις τιμές του κριτηρίου BIC για όλες τις περιπτώσεις. Οι τιμές φαίνονται στον Πίνακα 5.4, ενώ στο Γράφημα 8 μπορούμε να δούμε την πορεία της τιμής του κριτηρίου BIC.

Στις περιπτώσεις που δεν υπάρχει τιμή τα μοντέλα δεν συγκλίνουν, λόγω του μικρού συνόλου των δεδομένων (Commeford, Brewe & Traxler, 2021).

Πίνακας 5.4. Κριτήριο BIC για κάθε συνδυασμό μοντέλου παραμετροποίησης και αριθμού προφίλ

Μοντέλο παραμετροποίησης	Αριθμός Προφίλ								
	1	2	3	4	5	6	7	8	9
1	9979	9639	9623	9592	9557	9539	9531	9568	9571
2	9979	9466	9399						
3	9645	9585	9505	9514	9504	9541	9549	9586	9608
6	9645	9443	9419						

Γράφημα 8. Τιμές κριτηρίου BIC για κάθε συνδυασμό μοντέλου παραμετροποίησης και αριθμού προφίλ



Παρατηρούμε ότι τη μικρότερη τιμή στο κριτήριο BIC την έχει ο συνδυασμός του μοντέλου παραμετροποίησης 2 (διαφορετικές διακυμάνσεις σε κάθε προφίλ και συνδιακυμάνσεις σταθερές στο 0) με αριθμό προφίλ ίσο με 3. Αμέσως καλύτερος συνδυασμός με μικρή διαφορά στην τιμή του κριτηρίου BIC είναι αυτός του μοντέλου παραμετροποίησης 6 (διαφορετικές διακυμάνσεις και διαφορετικές συνδιακυμάνσεις σε κάθε προφίλ) με αριθμό προφίλ ίσο με 3. Θα μπορούσαμε να διαλέξουμε οποιονδήποτε από τους δύο συνδυασμούς, αλλά θα συνεχίσουμε την ανάλυση μας με αυτόν που δίνει το αποτέλεσμα της εντολής `comp$best` ως καλύτερο βάσει του κριτηρίου BIC, δηλαδή το συνδυασμό του μοντέλου παραμετροποίησης 2 (διαφορετικές διακυμάνσεις σε κάθε προφίλ και συνδιακυμάνσεις σταθερές στο 0) με 3 προφίλ.

Αφού επιλέξαμε το καλύτερο συνδυασμό μοντέλου παραμετροποίησης και αριθμού προφίλ συνεχίζουμε πλέον την ανάλυσή μας χρησιμοποιώντας αυτό το συνδυασμό. Με την εντολή `get_data(best_model)` παίρνουμε τις πιθανότητες κάθε μέλους του πληθυσμού να ανήκει σε κάθε προφίλ, αλλά και το προφίλ στο οποίο ταξινομείται. Επίσης, με την εντολή `get_estimates(best_model)` παίρνουμε την εκτίμηση για τη μέση τιμή και τη διακύμανση κάθε μεταβλητής σε κάθε προφίλ, ενώ με την εντολή `get_fit(best_model)` παίρνουμε τις τιμές των στατιστικών κριτηρίων για το μοντέλο που επιλέξαμε ως καλύτερο.

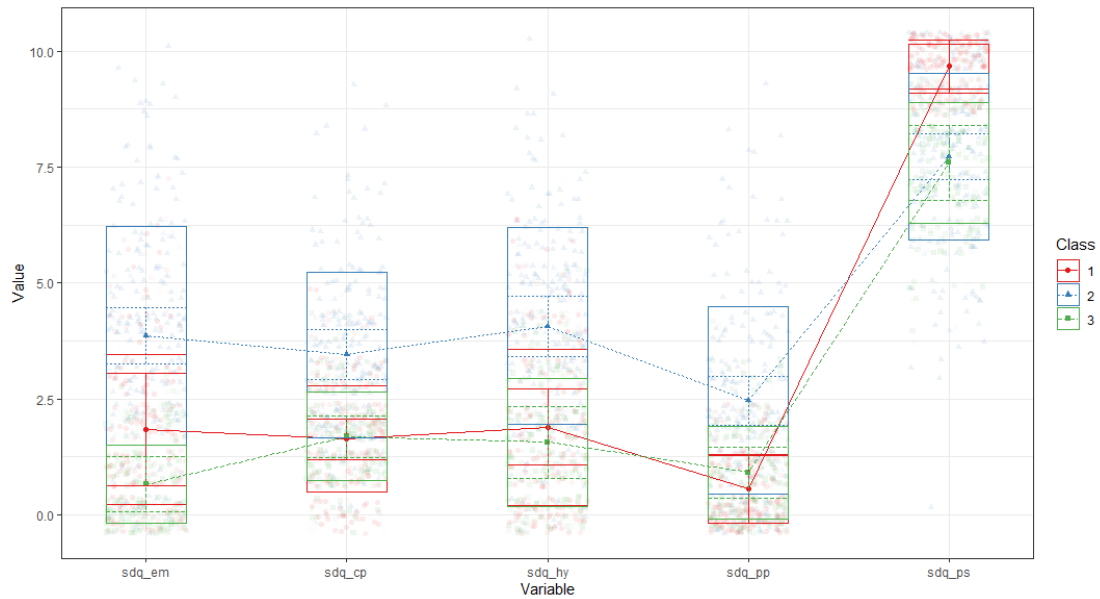
Οι εκτιμήσεις για τη μέση τιμή και τη διακύμανση των μεταβλητών σε κάθε προφίλ φαίνονται στον Πίνακα 5.5.

Πίνακας 5.5. Εκτιμώμενη μέση τιμή και διακύμανση μεταβλητών σε κάθε προφίλ

Μεταβλητή	Μέση τιμή			Διακύμανση		
	Προφίλ 1	Προφίλ 2	Προφίλ 3	Προφίλ 1	Προφίλ 2	Προφίλ 3
sdq_em	1.840	3.867	0.663	2.6297	5.5281	0.7099
sdq_cp	1.633	3.458	1.695	1.3159	3.1872	0.9193
sdq_hy	1.896	4.072	1.564	2.8306	4.5211	1.9065
sdq_pp	0.554	2.464	0.915	0.5243	4.0674	1.0132
sdq_ps	9.671	7.727	7.593	0.2283	3.2435	1.7117

Στη συνέχεια, με την εντολή `plot_profiles(best_model, add_line = TRUE, sd = TRUE)` εμφανίζεται το διάγραμμα με την εκτιμώμενη μέση τιμή και την εκτιμώμενη τυπική απόκλιση κάθε μεταβλητής σε κάθε προφίλ, το οποίο φαίνεται στο Γράφημα 9. Στον άξονα x βρίσκονται οι μεταβλητές στις οποίες εφαρμόζεται η LPA, ενώ στον άξονα y εμφανίζονται οι τιμές που παίρνουν αυτές οι μεταβλητές. Οι αχνές χρωματισμένες κουκίδες αναπαριστούν σε ποιο προφίλ ανήκει κάθε μέλος του πληθυσμού ανάλογα με το χρώμα, ενώ οι γραμμές ενώνουν τις εκτιμώμενες μέσες τιμές των μεταβλητών σε κάθε προφίλ.

Γράφημα 9. Εκτιμώμενη μέση τιμή και τυπική απόκλιση μεταβλητών σε κάθε προφίλ



Τέλος, με την εντολή `best_model$model$parameters$pro` βλέπουμε το ποσοστό του πληθυσμού που εκτιμάται ότι θα ανήκει σε κάθε προφίλ. Το αποτέλεσμα της εντολής φαίνεται στον Πίνακα 5.6.

Πίνακας 5.6. Εκτιμώμενο ποσοστό του πληθυσμού σε κάθε προφίλ

Προφίλ	Ποσοστό πληθυσμού
1	0,2836
2	0,4279
3	0,2885

Κατά την εφαρμογή της LPA στις μεταβλητές του SDQ εξετάστηκαν τα σκορ κάθε παιδιού και κάθε παιδί καταχωρήθηκε σε ένα προφίλ ανάλογα με αυτά τα σκορ. Έτσι κάθε προφίλ περιλαμβάνει παιδιά που συγκέντρωσαν παρόμοιο σκορ στις 5 κλίμακες του SDQ. Πιο συγκεκριμένα:

- Στην κλίμακα 1 του SDQ τα παιδιά που ανήκουν στο προφίλ 1 συγκέντρωσαν μέσο σκορ 1,840, τα παιδιά που ανήκουν στο προφίλ 2 συγκέντρωσαν μέσο σκορ 3,867 και τα παιδιά που ανήκουν στο προφίλ 3 συγκέντρωσαν μέσο σκορ 0,663.
- Στην κλίμακα 2 του SDQ τα παιδιά που ανήκουν στο προφίλ 1 συγκέντρωσαν μέσο σκορ 1,633, τα παιδιά που ανήκουν στο προφίλ 2 συγκέντρωσαν μέσο σκορ 3,458 και τα παιδιά που ανήκουν στο προφίλ 3 συγκέντρωσαν μέσο σκορ 1,695.
- Στην κλίμακα 3 του SDQ τα παιδιά που ανήκουν στο προφίλ 1 συγκέντρωσαν μέσο σκορ 1,896, τα παιδιά που ανήκουν στο προφίλ 2 συγκέντρωσαν μέσο σκορ 4,072 και τα παιδιά που ανήκουν στο προφίλ 3 συγκέντρωσαν μέσο σκορ 1,564.

- Στην κλίμακα 4 του SDQ τα παιδιά που ανήκουν στο προφίλ 1 συγκέντρωσαν μέσο σκορ 0,554 , τα παιδιά που ανήκουν στο προφίλ 2 συγκέντρωσαν μέσο σκορ 2,464 και τα παιδιά που ανήκουν στο προφίλ 3 συγκέντρωσαν μέσο σκορ 0,915.
- Στην κλίμακα 5 του SDQ τα παιδιά που ανήκουν στο προφίλ 1 συγκέντρωσαν μέσο σκορ 9,671, τα παιδιά που ανήκουν στο προφίλ 2 συγκέντρωσαν μέσο σκορ 7,727 και τα παιδιά που ανήκουν στο προφίλ 3 συγκέντρωσαν μέσο σκορ 7,593.

Με τη βοήθεια της LPA υπολογίσαμε μια λανθάνουσα μεταβλητή, της οποίας η τιμή είναι ο συνδυασμός των τιμών των 5 μεταβλητών για κάθε προφίλ. Αυτή η λανθάνουσα μεταβλητή είναι ότι τα παιδιά χωρίζονται σε τρεις ομάδες και δεν υπήρχε εξ αρχής.

5.5. Παρουσίαση συμπερασμάτων

5.5.1. Συμπεράσματα από την ταξινόμηση με την εφαρμογή της LPA.

Στο Γράφημα 10 βλέπουμε το αποτέλεσμα της εφαρμογής της LPA στις μεταβλητές του SDQ. Προκύπτει από την εκτέλεση της εντολής

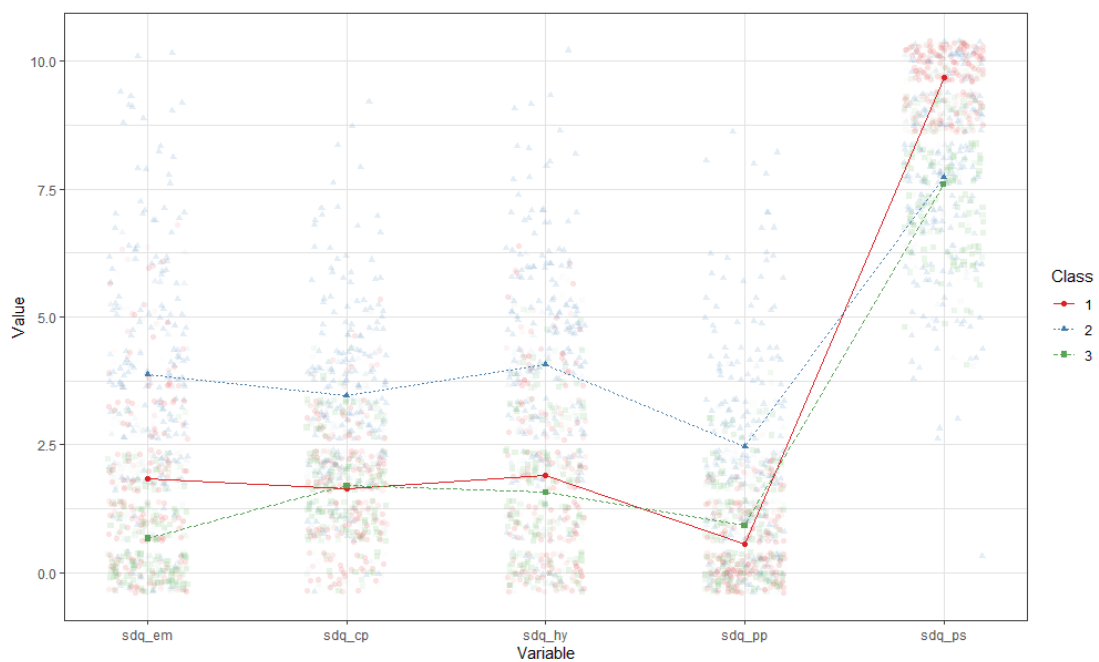
```
plot_profiles(best_model, add_line = TRUE, sd = FALSE, ci = NULL)
```

στο RStudio και παρουσιάζει γραφικά την εκτιμώμενη μέση τιμή κάθε μεταβλητής σε κάθε προφίλ.

Στον Πίνακα 5.7 βλέπουμε το εκτιμώμενο ποσοστό του πληθυσμού που ανήκει σε κάθε προφίλ. Το αποτέλεσμα αυτό προκύπτει από την εκτέλεση της εντολής

```
best_model$model$parameters$pro
```

Γράφημα 10. Εκτιμώμενη μέση τιμή μεταβλητών σε κάθε προφίλ.



Πίνακας 5.7. Εκτιμώμενο ποσοστό του πληθυσμού σε κάθε προφίλ

Προφίλ	Ποσοστό πληθυσμού
1	0,2836
2	0,4279
3	0,2885

Από το παραπάνω γράφημα και τον παραπάνω πίνακα συμπεραίνουμε ότι:

- Μια ομάδα παιδιών (28,36%) ανήκει στο προφίλ 1 και σημειώνει χαμηλό μέσο σκορ στις Κλίμακες 1 έως 3 του SDQ από 1,3 έως 2,5, πολύ χαμηλό μέσο σκορ στην Κλίμακα 4 (κάτω από 1) και πολύ υψηλό μέσο σκορ στην Κλίμακα 5 (πάνω από 9). Από αυτό συμπεραίνουμε ότι στο προφίλ 1 ανήκουν παιδιά με λίγα συναισθηματικά προβλήματα, λίγα προβλήματα συμπεριφοράς, λίγη υπερκινητικότητα και διάσπαση προσοχής και ελάχιστα προβλήματα με τους συνομηλίκους τους. Τα παιδιά που ανήκουν σε αυτό το προφίλ έχουν πολύ καλή κοινωνική συμπεριφορά
- Το μεγαλύτερο ποσοστό των παιδιών (42,79%) ανήκει στο προφίλ 2 και σημειώνει μέσο σκορ στις Κλίμακες 1 έως 4 του SDQ από 2,5 έως 5, ενώ στην Κλίμακα 5 σημειώνει μέσο σκορ περίπου 7,5. Αυτό σημαίνει ότι στο προφίλ αυτό ανήκουν παιδιά με μέτρια συναισθηματικά προβλήματα, μέτρια προβλήματα συμπεριφοράς, μέτρια υπερκινητικότητα και διάσπαση προσοχής, αλλά και κάποια προβλήματα με τους συνομηλίκους τους. Επίσης, τα παιδιά που ανήκουν σε αυτό το προφίλ έχουν αρκετά καλή κοινωνική συμπεριφορά.
- Μια ομάδα παιδιών παρόμοιου μεγέθους (28,85%) με αυτή του προφίλ 1 αποτελεί τον πληθυσμό του προφίλ 3 και σημειώνει πολύ χαμηλό μέσο σκορ (κάτω από 1) στις Κλίμακες 1 και 4 του SDQ, χαμηλό μέσο σκορ (λίγο πάνω από 1) στις Κλίμακες 2 και 3 και υψηλό μέσο σκορ στην Κλίμακα 5 (περίπου 7,5). Συμπερασματικά, στο προφίλ 3 ανήκουν παιδιά με ελάχιστα συναισθηματικά προβλήματα, λίγα προβλήματα συμπεριφοράς, λίγη υπερκινητικότητα και διάσπαση προσοχής και ελάχιστα προβλήματα με τους συνομηλίκους τους, ενώ έχουν αρκετά καλή κοινωνική συμπεριφορά

Συνοψίζοντας, παρατηρούμε ότι το μεγαλύτερο ποσοστό των παιδιών (προφίλ 2, 42,79%) παρουσιάζει μέτρια προβλήματα στις Κλίμακες 1 έως 4 του SDQ και ταυτοχρόνως έχει μειωμένη κοινωνική συμπεριφορά. Τα υπόλοιπα παιδιά χωρίζονται σε δύο ίσα σε μέγεθος προφίλ (προφίλ 1 – 28,36%, προφίλ 3 – 28,85%) με λιγότερα προβλήματα στις Κλίμακες 1 έως 4 από τα παιδιά του προφίλ 2, αλλά διαφορετικό επίπεδο κοινωνικής συμπεριφοράς. Αυτά του προφίλ 1 έχουν πολύ υψηλό επίπεδο κοινωνικής συμπεριφοράς, ενώ αυτά του προφίλ 3 ίδιο επίπεδο με τα παιδιά του προφίλ 2.

5.5.2. Συσχέτιση των προφίλ με το φύλο και την τάξη των παιδιών

Στη συνέχεια, θα ερευνήσουμε πιθανή συσχέτιση των δημιουργημένων προφίλ με το φύλο (μεταβλητή Gender) και την ηλικία των παιδιών (μεταβλητή Grade).

Για τον υπολογισμό των ποσοστών των αγοριών και των κοριτσιών σε κάθε προφίλ, αλλά και το ποσοστό των παιδιών κάθε τάξης σε κάθε προφίλ εκτελούμε τον εξής κώδικα:


```

datawithclass <- cbind(select(data, Gender, Grade),
  select(get_data(best_model), Class))
class1 <- filter(datawithclass, Class == 1)
class2 <- filter(datawithclass, Class == 2)
class3 <- filter(datawithclass, Class == 3)
boys <- c(round(table(class1$Gender)[1]/nrow(class1), 4),
  round(table(class2$Gender)[1]/nrow(class2), 4),
  round(table(class3$Gender)[1]/nrow(class3), 4))
girls <- c(round(table(class1$Gender)[2]/nrow(class1), 4),
  round(table(class2$Gender)[2]/nrow(class2), 4),
  round(table(class3$Gender)[2]/nrow(class3), 4))
pgender <- data.frame(Class = c(1, 2, 3), Αγόρια = boys,
  Κορίτσια = girls)
pgender
edim <- c(round(table(class1$Grade)[1]/nrow(class1), 4),
  round(table(class2$Grade)[1]/nrow(class2), 4),
  round(table(class3$Grade)[1]/nrow(class3), 4))
stdim <- c(round(table(class1$Grade)[2]/nrow(class1), 4),
  round(table(class2$Grade)[2]/nrow(class2), 4),
  round(table(class3$Grade)[2]/nrow(class3), 4))
agimn <- c(round(table(class1$Grade)[3]/nrow(class1), 4),
  round(table(class2$Grade)[3]/nrow(class2), 4),
  round(table(class3$Grade)[3]/nrow(class3), 4))
bgimn <- c(round(table(class1$Grade)[4]/nrow(class1), 4),
  round(table(class2$Grade)[4]/nrow(class2), 4),
  round(table(class3$Grade)[4]/nrow(class3), 4))
pgrade <- data.frame(Class = c(1, 2, 3), "Ε Δημοτικού" =
  edim, "ΣΤ Δημοτικού" = stdim, "Α Γυμνασίου" = agimn,
  "Β Γυμνασίου" = bgimn)
pgrade.

```

Αρχικά δημιουργείται ένας νέος πίνακας με στήλες που αντιστοιχούν στο φύλο, την τάξη και το προφίλ στο οποίο κατατάχθηκε κάθε μαθητής. Στη συνέχεια δημιουργούνται τρεις νέοι πίνακες, ένας για κάθε προφίλ. Τέλος, υπολογίζονται και εμφανίζονται τα ποσοστά των παιδιών ανά φύλο σε κάθε προφίλ και εν συνεχεία τα ποσοστά των παιδιών ανά τάξη σε κάθε προφίλ.

Το αποτέλεσμα της εκτέλεσης του κώδικα φαίνεται στους Πίνακες 5.8 και 5.9, ενώ τα αποτελέσματα με τη μορφή που παράγονται στο RStudio δίνονται στο Παράρτημα II.

Παρατηρούμε ότι το ποσοστό των αγοριών και των κοριτσιών στα δύο πρώτα προφίλ είναι σχεδόν το ίδιο, ενώ αλλάζει στο τρίτο προφίλ, όπου έχουμε μεγαλύτερο ποσοστό αγοριών. Αυτό σημαίνει ότι στο προφίλ 3 έχουμε μεγαλύτερη συμμετοχή αγοριών απ' ότι στα άλλα δύο προφίλ.

Πίνακας 5.8. Ποσοστό παιδιών ανά φύλο σε κάθε προφίλ

Προφίλ	Αγόρια (%)	Κορίτσια (%)
1	40,00	60,00
2	43,43	56,57
3	56,62	43,38

Πίνακας 5.9. Ποσοστό παιδιών ανά τάξη σε κάθε προφίλ

Προφίλ	Ε' Δημοτικού (%)	ΣΤ' Δημοτικού (%)	Α' Γυμνασίου (%)	Β' Γυμνασίου (%)
1	22,67	13,33	57,33	6,67
2	12,12	13,64	66,67	7,58
3	8,09	19,85	65,44	6,62

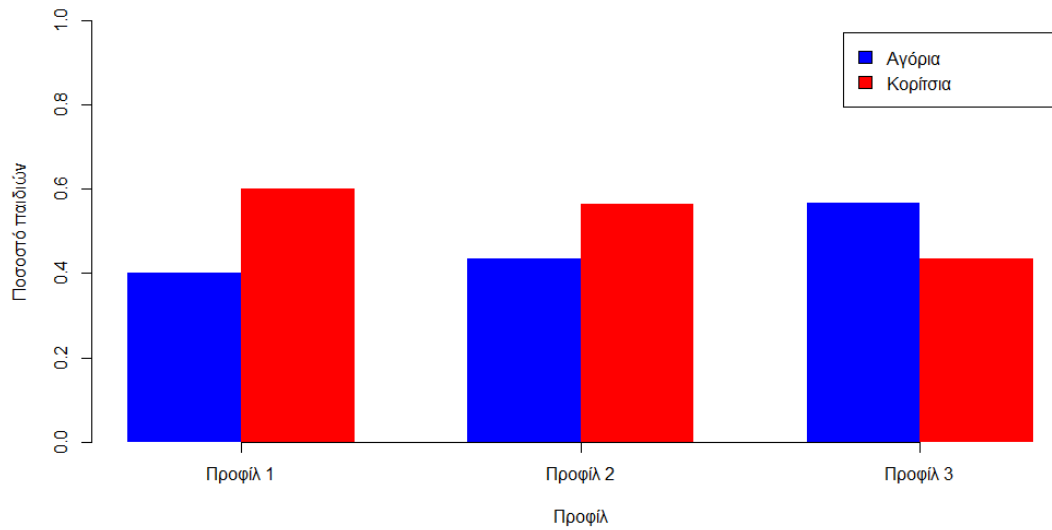
Όσον αφορά στις τάξεις, παρατηρούμε ότι το ποσοστό των παιδιών ΣΤ' Δημοτικού και Β' Γυμνασίου είναι σχεδόν ίδιο σε όλα τα προφίλ και παρουσιάζονται διαφοροποιήσεις στα παιδιά της Ε' Δημοτικού και της Α' Γυμνασίου. Στο πρώτο προφίλ το ποσοστό των παιδιών Ε' Δημοτικού είναι μεγαλύτερο από αυτό στα άλλα δύο προφίλ, ενώ για το ποσοστό των παιδιών Α' Γυμνασίου συμβαίνει το αντίθετο, δηλαδή το ποσοστό των παιδιών της Α' Γυμνασίου στο πρώτο προφίλ είναι μικρότερο από αυτό στα άλλα δύο προφίλ. Αυτό σημαίνει ότι στο προφίλ 1 έχουμε μεγαλύτερη συμμετοχή των παιδιών της Ε' Δημοτικού και μικρότερη των παιδιών της Α' Γυμνασίου απ' ότι στα άλλα δύο προφίλ. Τα προφίλ 2 και 3 δεν παρουσιάζουν ιδιαίτερες διαφορές όσον αφορά στη συμμετοχή των παιδιών ανάλογα με την τάξη.

Ακολουθούν τα γραφήματα με τις παραπάνω διαπιστώσεις. Τα γραφήματα παράγονται με την εκτέλεση του εξής κώδικα:

```
y <- rbind(boys, girls)
barplot(y, names.arg = c("Προφίλ 1", "Προφίλ 2", "Προφίλ 3"), width = 1, beside = TRUE, ylim = c(0, 1), legend.text = c("Αγόρια", "Κορίτσια"), col = c("blue", "red"), axis.lty = 1, border = NA, xlab = "Προφίλ", ylab = "Ποσοστό παιδιών")
x <- rbind(edim, stdim, agimn, bgimn)
barplot(x, names.arg = c("Προφίλ 1", "Προφίλ 2", "Προφίλ 3"), width = 1, beside = TRUE, ylim = c(0, 1), legend.text = c("Ε' Δημοτικού", "ΣΤ' Δημοτικού", "Α' Γυμνασίου", "Β' Γυμνασίου"), col = c("blue", "red", "yellow", "green"), axis.lty = 1, border = NA, xlab = "Προφίλ", ylab = "Ποσοστό παιδιών").
```

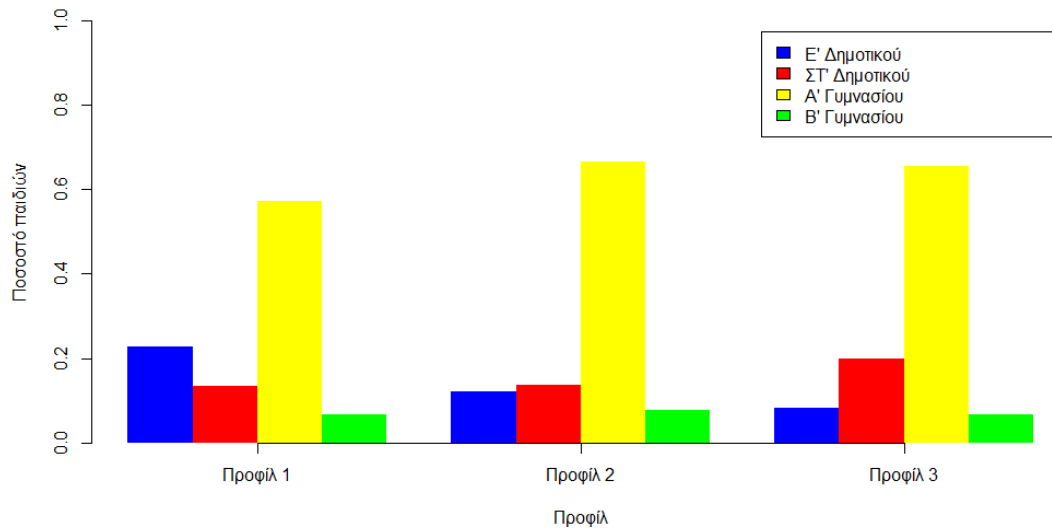
Στο Γράφημα 11 φαίνεται το ποσοστό των παιδιών ανά φύλο σε κάθε προφίλ και στο Γράφημα 12 φαίνεται το ποσοστό των παιδιών ανά τάξη σε κάθε προφίλ.

Γράφημα 11. Ποσοστό παιδιών ανά φύλο σε κάθε προφίλ



Από το Γράφημα 11 παρατηρούμε ότι τα προφίλ 1 και 2 έχουν το ίδιο μοτίβο όσον αφορά στη συμμετοχή αγοριών και κοριτσιών σε αυτά, κάτι που αλλάζει στο προφίλ 3.

Γράφημα 12. Ποσοστό παιδιών ανά τάξη σε κάθε προφίλ



Τέλος, από το γράφημα 12 παρατηρούμε ότι τα προφίλ 2 και 3 έχουν το ίδιο μοτίβο όσον αφορά στη συμμετοχή των τάξεων σε αυτά, γεγονός που αλλάζει στο προφίλ 1.

Ένα άλλο στοιχείο που μπορούμε να ερευνήσουμε είναι το ποσοστό των παιδιών κάθε φύλου που ανήκει σε κάθε προφίλ, καθώς και το ποσοστό των παιδιών κάθε τάξης που ανήκει σε κάθε προφίλ.

Για τον υπολογισμό αυτών των ποσοστών εκτελούμε τον εξής κώδικα:

```
class1gender <- c(round(nrow(filter(class1, Gender ==
  1))/nrow(filter(datawithclass, Gender == 1)), 4),
  round(nrow(filter(class1, Gender ==
  2))/nrow(filter(datawithclass, Gender == 2)), 4))

class2gender <- c(round(nrow(filter(class2, Gender ==
  1))/nrow(filter(datawithclass, Gender == 1)), 4),
  round(nrow(filter(class2, Gender ==
  2))/nrow(filter(datawithclass, Gender == 2)), 4))

class3gender <- c(round(nrow(filter(class3, Gender ==
  1))/nrow(filter(datawithclass, Gender == 1)), 4),
  round(nrow(filter(class3, Gender ==
  2))/nrow(filter(datawithclass, Gender == 2)), 4))

pclassgender <- data.frame(Φύλο = c("Αγόρια", "Κορίτσια"),
  "Class 1" = class1gender, "Class 2" = class2gender,
  "Class 3" = class3gender)

pclassgender

class1grade <- c(round(nrow(filter(class1, Grade ==
  1))/nrow(filter(datawithclass, Grade == 1)), 4),
  round(nrow(filter(class1, Grade ==
  2))/nrow(filter(datawithclass, Grade == 2)), 4),
  round(nrow(filter(class1, Grade ==
  3))/nrow(filter(datawithclass, Grade == 3)), 4),
  round(nrow(filter(class1, Grade ==
  4))/nrow(filter(datawithclass, Grade == 4)), 4))

class2grade <- c(round(nrow(filter(class2, Grade ==
  1))/nrow(filter(datawithclass, Grade == 1)), 4),
  round(nrow(filter(class2, Grade ==
  2))/nrow(filter(datawithclass, Grade == 2)), 4),
  round(nrow(filter(class2, Grade ==
  3))/nrow(filter(datawithclass, Grade == 3)), 4),
  round(nrow(filter(class2, Grade ==
  4))/nrow(filter(datawithclass, Grade == 4)), 4))

class3grade <- c(round(nrow(filter(class3, Grade ==
  1))/nrow(filter(datawithclass, Grade == 1)), 4),
  round(nrow(filter(class3, Grade ==
  2))/nrow(filter(datawithclass, Grade == 2)), 4),
  round(nrow(filter(class3, Grade ==
  3))/nrow(filter(datawithclass, Grade == 3)), 4),
  round(nrow(filter(class3, Grade ==
  4))/nrow(filter(datawithclass, Grade == 4)), 4))
```

```
pclassgrade <- data.frame(Τάξη = c("Ε Δημοτικού", "ΣΤ
  Δημοτικού", "Α Γυμνασίου", "Β Γυμνασίου"), "Class 1" =
  class1grade, "Class 2" = class2grade, "Class 3" =
  class3grade)
```

```
pclassgrade.
```

Το αποτέλεσμα της εκτέλεσης του κώδικα φαίνεται στους Πίνακες 5.10 και 5.11.

Πίνακας 5.10. Ποσοστό παιδιών κάθε φύλου σε κάθε προφίλ

Φύλο	Προφίλ		
	1 (%)	2 (%)	3 (%)
Αγόρια	26,91	38,57	34,53
Κορίτσια	34,48	42,91	22,61

Πίνακας 5.11. Ποσοστό παιδιών κάθε τάξης σε κάθε προφίλ

Τάξη	Προφίλ		
	1 (%)	2 (%)	3 (%)
Ε' Δημοτικού	49,28	34,78	15,94
ΣΤ' Δημοτικού	27,03	36,49	36,49
Α' Γυμνασίου	28,01	43,00	28,99
Β' Γυμνασίου	29,41	44,12	26,47

Παρατηρούμε ότι τα περισσότερα αγόρια ανήκουν στο προφίλ 2, όπως και τα περισσότερα κορίτσια. Από τα υπόλοιπα αγόρια τα περισσότερα ανήκουν στο προφίλ 3, ενώ από τα υπόλοιπα κορίτσια τα περισσότερα ανήκουν στο προφίλ 1.

Όσον αφορά στις τάξεις, παρατηρούμε ότι τα περισσότερα παιδιά της Ε' Δημοτικού ανήκουν στο προφίλ 1, τα περισσότερα παιδιά της ΣΤ' Δημοτικού ανήκουν στα προφίλ 2 και 3 με ίδιο ποσοστό και τα περισσότερα παιδιά της Α' Γυμνασίου και της Β' Γυμνασίου ανήκουν στο προφίλ 2. Για την Ε' Δημοτικού βλέπουμε ότι ακολουθεί το προφίλ 2, ενώ για τις Α' και Β' Γυμνασίου βλέπουμε ότι τα υπόλοιπα παιδιά μοιράζονται στα προφίλ 1 και 3.

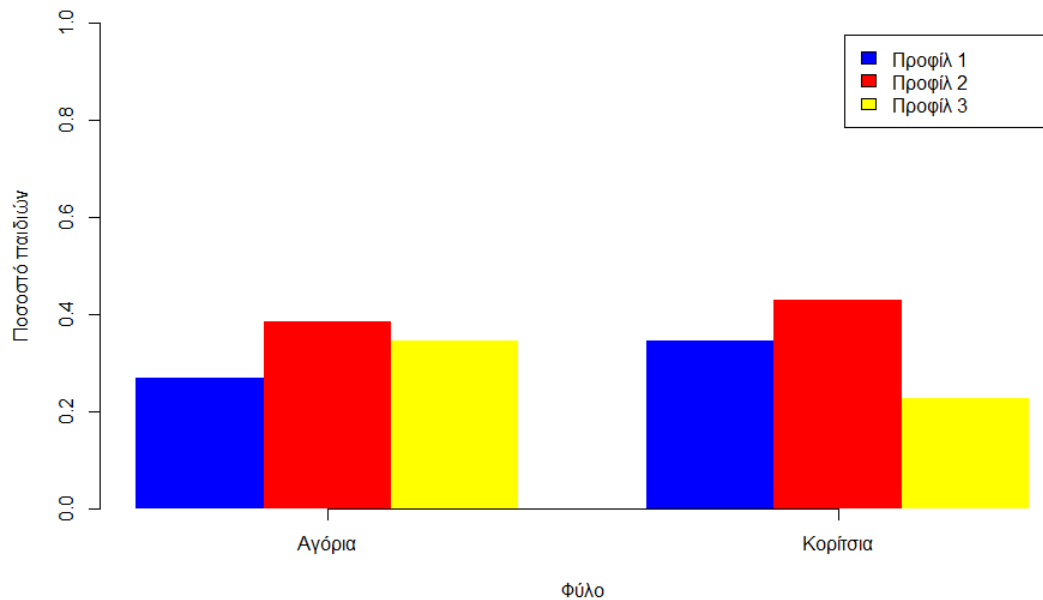
Ακολουθούν τα γραφήματα με τις παραπάνω διαπιστώσεις. Τα γραφήματα παράγονται με την εκτέλεση του εξής κώδικα:

```
y <- rbind(class1gender, class2gender, class3gender)
barplot(y, names.arg = c("Αγόρια", "Κορίτσια"), width = 1,
  beside = TRUE, ylim = c(0, 1), legend.text = c("Προφίλ
  1", "Προφίλ 2", "Προφίλ 3"), col = c("blue", "red",
  "yellow"), axis.lty = 1, border = NA, xlab = "Φύλο",
  ylab = "Ποσοστό παιδιών")
```

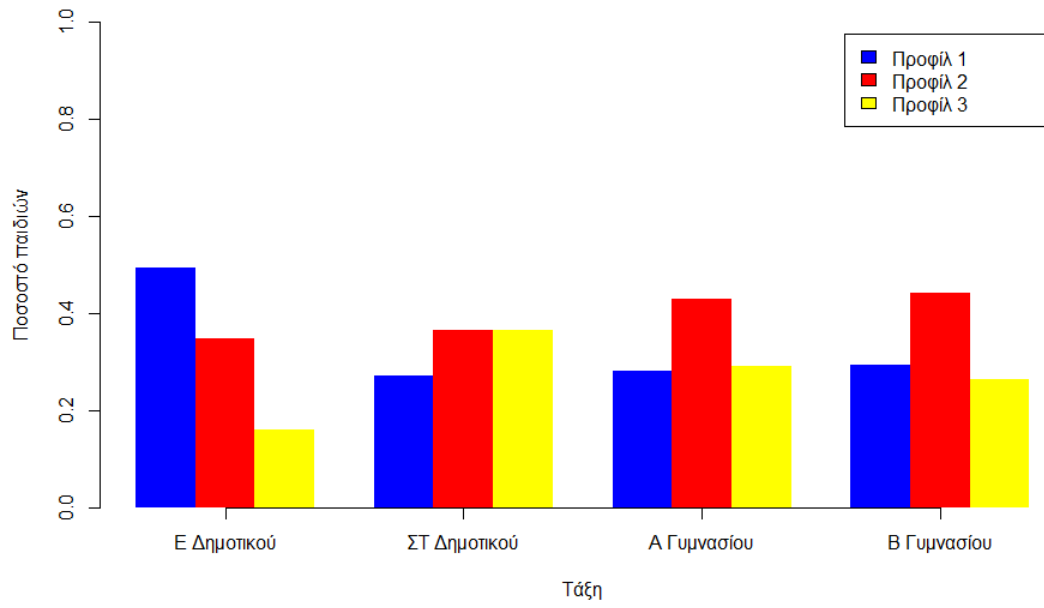
```
x <- rbind(class1grade, class2grade, class3grade)
barplot(x, names.arg = c("Ε Δημοτικού", "ΣΤ Δημοτικού", "Α
  Γυμνασίου", "Β Γυμνασίου"), width = 1, beside = TRUE,
  ylim = c(0, 1), legend.text = c("Προφίλ 1", "Προφίλ 2",
  "Προφίλ 3"), col = c("blue", "red", "yellow"), axis.lty
  = 1, border = NA, xlab = "Τάξη", ylab = "Ποσοστό
  παιδιών").
```

Στο Γράφημα 13 φαίνεται το ποσοστό των παιδιών κάθε φύλου σε κάθε προφίλ και στο Γράφημα 14 φαίνεται το ποσοστό των παιδιών κάθε τάξης σε κάθε προφίλ.

Γράφημα 13. Ποσοστό παιδιών κάθε φύλου σε κάθε προφίλ



Γράφημα 14. Ποσοστό παιδιών κάθε τάξης σε κάθε προφίλ



Κεφάλαιο 6. **Άλλες εφαρμογές της Ανάλυσης Λανθανόντων Προφίλ (LPA)**

Η Ανάλυση Λανθανόντων Προφίλ (LPA) χρησιμοποιείται σε πολλούς επιστημονικούς τομείς, όπως οι κοινωνικές και πολιτικές επιστήμες, η ιατρική και το μάρκετινγκ, για την ανάλυση, μελέτη και επίλυση προβλημάτων.

Η μελέτη της επαγγελματικής συμπεριφοράς είναι ένας τομέας στον οποίο υπάρχει πλήθος αναλύσεων με τη βοήθεια της LPA. Ένα παράδειγμα είναι η μελέτη του πώς επιδρούν στους εργαζόμενους τα κίνητρα στον εργασιακό χώρο. Έτσι πραγματοποιήθηκε μια μελέτη για να εξεταστεί η ταυτόχρονη εμφάνιση διαφορετικών τύπων κινήτρων σε μεμονωμένους εργαζόμενους (Howard, Gagné, Morin & Van den Broeck, 2016). Η μελέτη βασίστηκε σε δύο δείγματα που συλλέχθηκαν από το 2008 έως το 2013. Το πρώτο δείγμα περιλάμβανε 723 Καναδούς υπαλλήλους από τρεις διαφορετικούς κλάδους: 105 από τον τεχνολογικό τομέα, 319 από τον κυβερνητικό τομέα και 299 από τον μεταποιητικό τομέα. Το δεύτερο δείγμα περιλάμβανε 286 Βέλγους υπαλλήλους. Χρησιμοποιήθηκε η LPA για να χωριστούν οι υπάλληλοι σε ομάδες, ανάλογα με τις απαντήσεις που έδωσαν στις ερωτήσεις που τους δόθηκαν, με στόχο την εξέταση της εμφάνισης πολλαπλών κανονιστικών συμπεριφορών μεμονωμένων υπαλλήλων από την αλληλεπίδραση πολλών τύπων κινήτρων. Οι μεταβλητές της μελέτης ήταν οι εξής: έλλειψη κινήτρου, εξωτερικό κίνητρο, εξωτερικοί κανόνες, εσωτερικοί κανόνες, καθορισμένοι κανόνες, εσωτερικό κίνητρο, απόδοση εντός καθηκόντων, απόδοση εκτός καθηκόντων, εργασιακή δέσμευση, εξάντληση και εργασιακή ικανοποίηση. Για το πρώτο δείγμα χρησιμοποιήθηκαν οι πρώτες 6 μεταβλητές, ενώ για το δεύτερο και οι 11 μεταβλητές. Οι υπάλληλοι σύμφωνα με τις απαντήσεις τους κατανεμήθηκαν σε τέσσερα προφίλ. Τα αποτελέσματα ήταν παρόμοια και για τα δύο δείγματα. Το πρώτο προφίλ, που αποτελεί το 27,6% του πρώτου δείγματος και το 13,1% του δεύτερου, περιλαμβάνει υπαλλήλους χωρίς κίνητρα και παρουσιάζει πολύ υψηλή τιμή στη μεταβλητή έλλειψη κινήτρου και μέσες προς χαμηλές τιμές στις άλλες μεταβλητές. Το δεύτερο προφίλ (11,5% του πρώτου δείγματος και 27,8% του δεύτερου) περιλαμβάνει υπαλλήλους με μέτρια κίνητρα και παρουσιάζει πολύ χαμηλά επίπεδα εξωτερικών κινήτρων και κανόνων, χαμηλά επίπεδα έλλειψης κινήτρων και εσωτερικών κανόνων και μέσο επίπεδο καθορισμένων κανόνων και εσωτερικών κινήτρων. Το τρίτο προφίλ (25,6% του πρώτου δείγματος και 22% του δεύτερου) περιλαμβάνει υπαλλήλους με υψηλά κίνητρα και παρουσιάζει χαμηλό επίπεδο έλλειψης κινήτρων, και μέσο προς υψηλό επίπεδο στις άλλες μεταβλητές. Τέλος, το τέταρτο προφίλ (35,3% του πρώτου δείγματος και 37,1% του δεύτερου) περιλαμβάνει υπαλλήλους με ισορροπημένα κίνητρα και παρουσιάζει μέσο επίπεδο σε όλες τις μεταβλητές. Επίσης, για τις πέντε επιπλέον μεταβλητές που χρησιμοποιούνται μόνο στο δείγμα 2 εξάγονται κάποια συμπεράσματα. Η απόδοση εντός και εκτός καθηκόντων έχουν υψηλή τιμή στο τρίτο και στο δεύτερο προφίλ και χαμηλή στο πρώτο και το τέταρτο προφίλ. Η εργασιακή ικανοποίηση και η εργασιακή δέσμευση έχουν υψηλή τιμή στο τρίτο και το δεύτερο προφίλ, ισορροπημένη στο τέταρτο και χαμηλή στο πρώτο. Τέλος, η εξάντληση έχει υψηλή τιμή στο πρώτο και το τέταρτο προφίλ και χαμηλή στο δεύτερο και το τρίτο προφίλ.

Ένας ακόμα τομέας στο οποίο είναι χρήσιμη η LPA είναι η παιδοψυχιατρική, ώστε η ομαδοποίηση των παιδιών με παρόμοια χαρακτηριστικά να συμβάλλει στη διάγνωση και θεραπεία προβλημάτων. Το χαμηλό επίπεδο σωματικής άσκησης, η καθιστική ζωή και η μη βέλτιστη διατροφή συμβάλουν στην αύξηση του λίπους στους νέους. Με τη βοήθεια της LPA τα παιδιά ομαδοποιούνται βάσει αυτών των τριών χαρακτηριστικών, ώστε να φανεί κατά πόσο αυτά συμβάλουν στην ύπαρξη υψηλού Δείκτη Μάζας

Σώματος (ΔΜΣ) (Berlin, Williams & Parra 2014). Η εφαρμογή της LPA οδήγησε στη δημιουργία τριών προφίλ παιδιών. Το πρώτο προφίλ αποτελεί το μεγαλύτερο προφίλ, συγκεντρώνοντας το 89,4% των παιδιών, και χαρακτηρίζεται από μέσο επίπεδο σωματικής άσκησης και καθιστικής ζωής, αλλά και άνω του μέσου όρου επίπεδο διατροφής. Το δεύτερο προφίλ συγκεντρώνει το 8,3% των παιδιών και χαρακτηρίζεται από μέσο κίνδυνο, αφού όλα τα χαρακτηριστικά έχουν μικρή τιμή. Το τρίτο προφίλ αποτελεί το μικρότερο προφίλ (2,3% των παιδιών) και χαρακτηρίζεται από υψηλό επίπεδο σωματικής άσκησης, αλλά και καθιστικής ζωής με υψηλή κατανάλωση σαλάτας και σόδας. Από τη συσχέτιση των χαρακτηριστικών των προφίλ με την τιμή του ΔΜΣ των μελών κάθε προφίλ, μπορούν να προκύψουν χρήσιμα συμπεράσματα για τα αίτια που προκαλούν την αύξηση του λίπους στους νέους. Συνεπώς, η LPA είναι ένα σημαντικό εργαλείο για τους παιδοψυχολόγους, καθώς μπορεί να ομαδοποιήσει περιπτώσεις ασθενών και να αποκαλύψει τα κοινά αίτια της ασθένειας.

Επίσης, σημαντική είναι η χρήση της LPA στην μελέτη κοινωνικών φαινομένων, όπως τι είναι αυτό που παρακινεί τους ανθρώπους στον εθελοντισμό και την παροχή απλήρωτων υπηρεσιών στην κοινότητα γενικά (Geiser, Okun & Grano 2014). Σε έρευνα που πραγματοποιήθηκε σε δύο δείγματα, ένα με 589 Αμερικανούς φοιτητές και ένα με 993 Ιταλούς φοιτητές, χρησιμοποιήθηκε η LPA για να εξεταστεί η σχέση των κινήτρων και της συχνότητας εθελοντισμού. Οι μεταβλητές που χρησιμοποιήθηκαν στην εφαρμογή της ανάλυσης ήταν έξι: εσωτερικά κίνητρα, εξωγενή κίνητρα σύμφωνα με τις αξίες του ατόμου, πλήρως ενσωματωμένα κίνητρα, κίνητρα από τη συνείδηση, εξωτερικά κίνητρα και έλλειψη κινήτρων. Οι φοιτητές μετά την εφαρμογή της LPA κατανεμήθηκαν σε έξι προφίλ. Το πρώτο προφίλ είχε χαμηλές τιμές σε όλες τις μεταβλητές και αποτελεί ένα προφίλ με χαμηλό κίνητρο για εθελοντισμό. Το δεύτερο προφίλ είχε ίδιο μοτίβο τιμών με το πρώτο, αλλά με υψηλότερες τιμές στα ενδογενή κίνητρα και αποτελεί ένα προφίλ με μεσαία ενδογενή και χαμηλά εξωγενή κίνητρα για εθελοντισμό. Και το τρίτο προφίλ είχε ίδιο μοτίβο τιμών, αλλά με ακόμα υψηλότερες τιμές στα ενδογενή κίνητρα και αποτελεί ένα προφίλ με υψηλά ενδογενή και χαμηλά εξωγενή κίνητρα. Το τέταρτο προφίλ διαφέρει από τα προηγούμενα αφού έχει υψηλές τιμές σε όλες τις μεταβλητές εκτός αυτής της έλλειψης κινήτρων και αποτελεί ένα προφίλ με υψηλά ενδογενή και εξωγενή κίνητρα. Το πέμπτο προφίλ είναι εντελώς διαφορετικό με τα προηγούμενα, αφού παρουσιάζει χαμηλές τιμές στα ενδογενή και εξωγενή κίνητρα, αλλά υψηλή τιμή στην έλλειψη κινήτρων και αποτελεί ένα προφίλ στο οποίο υπάρχει έλλειψη κινήτρων. Τέλος, το έκτο προφίλ παρουσιάζει ένα μοτίβο στο οποίο όλες οι μεταβλητές έχουν μεσαίες τιμές και αποτελεί ένα προφίλ στο οποίο πιθανόν ανήκουν άτομα που επιλέγουν πάντα το μεσαίο σημείο της κλίμακας των μεταβλητών. Δεν εντοπίστηκαν διαφορές μεταξύ ανδρών και γυναικών, ενώ οι Ιταλοί ανήκαν περισσότερο στο πέμπτο προφίλ. Ερευνώντας τη συσχέτιση της συμμετοχής σε κάποιο προφίλ με τη συχνότητα εθελοντισμού, η μελέτη κατέληξε στο συμπέρασμα ότι οι συμμετέχοντες στο τρίτο προφίλ παρουσίαζαν μεγαλύτερη συχνότητα εθελοντισμού, κάτι που υποδεικνύει ότι τα εξωγενή κίνητρα μπορεί να υπονομεύσουν τα ενδογενή κίνητρα ενός ατόμου για εθελοντισμό.

Συμπεράσματα

Όπως έγινε εμφανές η Ανάλυση Λανθανόντων Προφίλ (LPA) πρόκειται για ένα πολύ χρήσιμο και δυνατό εργαλείο που προσφέρει σημαντική βοήθεια σε όποιον θέλει να αναλύσει και να ομαδοποιήσει δεδομένα σε οποιονδήποτε κλάδο. Η LPA μπορεί να χρησιμοποιηθεί για πολλούς σκοπούς και από πολλές επιστήμες. Στην κοινωνιολογία μπορεί να χρησιμοποιηθεί για την ανάλυση κοινωνικών φαινομένων, στην ιατρική για την πρόβλεψη και αντιμετώπιση ασθενειών, στην ψυχολογία για την ανάλυση συμπεριφορών, αλλά και στο μάνατζμεντ για λήψη αποφάσεων.

Στην εργασία αυτή αρχικά η Ανάλυση Λανθανόντων Προφίλ (LPA) προσεγγίστηκε θεωρητικά, ενώ στη συνέχεια πραγματοποιήθηκε μια πρακτική εφαρμογή της με τη βοήθεια της γλώσσας προγραμματισμού R. Μέσω της πρακτικής εφαρμογής έγινε προσπάθεια ώστε να γίνει περισσότερο κατανοητός ο σκοπός χρήσης της LPA, καθώς και ο τρόπος εφαρμογής και λειτουργίας της. Τέλος, στόχος αποτέλεσε και η ανάδειξη του τρόπου ερμηνείας των αποτελεσμάτων της εφαρμογής της.

Βιβλιογραφία

Ελληνόγλωσση

- Παπαγεωργίου, Ι. (2015). *Θεωρία Δειγματοληψίας*. [ηλεκτρονικό βιβλίο] Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: <http://hdl.handle.net/11419/1296>
- Πετρίδης, Δ. (2015). *Ανάλυση πολυμεταβλητών τεχνικών*. [ηλεκτρονικό βιβλίο] Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: <http://hdl.handle.net/11419/2126>
- Ρίτσαρντσον, Κ., & Βασίλαινας, Α. (1999). *Εισαγωγή στη στατιστική επιστήμη*. Αθήνα: Κάκτος
- Φαρμάκη, Ε. (2012). *Εφαρμογές πολυπαραμετρικών στατιστικών τεχνικών στη χημική ανάλυση* (Διδακτορική διατριβή). Αθήνα: Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών. Ανακτήθηκε από: <https://www.didaktorika.gr/eadd/handle/10442/27807>
- Φωκιανός, Κ., & Χαραλάμπους, Χ. (2010). *Εισαγωγή στηνR - Πρόχειρες Σημειώσεις*. Λευκωσία: Τμήμα Μαθηματικών & Στατιστικής Πανεπιστήμιο Κύπρου. Ανακτήθηκε από: <https://cran.r-project.org/doc/contrib/mainfokianoscharalambous.pdf>

Ξενόγλωσση

- Andersen, E. (1982). Latent Structure Analysis: A Survey. *Scandinavian Journal of Statistics*, 9 (1), 1-12. Ανακτήθηκε από: <http://www.jstor.org/stable/4615848>
- Berlin, K., Williams, N., & Parra, G. (2014). An Introduction to Latent Variable Mixture Modeling (Part 1): Overview and Cross-Sectional Latent Class and Latent Profile Analyses. *Journal of Pediatric Psychology*, 39 (2), 174-187. doi: 10.1093/jpepsy/jst084
- Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, 40 (1), 5-32. doi: 10.1007/BF02291477
- Commeford, K., Brewe, E., & Traxler, A. (2021). Characterizing active learning environments in physics using latent profile analysis. Ανακτήθηκε από: <https://arxiv.org/abs/2105.02897v1>
- Geiser, C., Okun, M., & Grano, C. (2014). Who is motivated to volunteer? A latent profile analysis linking volunteer motivation to frequency of volunteering. *Psychological Test and Assessment Modeling*, 56 (1), 3-24. Ανακτήθηκε από: <https://www.researchgate.net/publication/268688031>
- Gibson, W. (1959). Three multivariate models: Factor analysis, latent structure analysis and latent profile analysis. *Psychometrika*, 24 (3), 229-252. doi: 10.1007/BF02289845
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, 38 (5), 581-586. doi: 10.1111/j.1469-7610.1997.tb01545.x

- He, J., & Fan, X. (2019). Latent Profile Analysis. Στο V. Zeigler-Hill & T. K. Shackelford (επιμ.), *Encyclopedia of Personality and Individual Differences*. Cham: Springer. doi: 10.1007/978-3-319-28099-8_2316-1
- Hickendorff, M., Edelsbrunner, P., McMullen, J., Schneider, M., & Trezise, K. (2018). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. *Learning and Individual Differences*, 66, 4-15. doi: 10.1016/j.lindif.2017.11.001
- Howard, J., Gagné, M., Morin, A., & Van den Broeck, A. (2016). Motivation profiles at work: A self-determination theory approach. *Journal of Vocational Behavior*, 95–96, 74-89. doi: 10.1016/j.jvb.2016.07.004
- Masyn, K. (2013). Latent class analysis and finite mixture modeling. Στο T. D. Little (επιμ.), *The Oxford handbook of quantitative methods: Statistical analysis* (σ. 551-611). New York, NY: Oxford University Press.
- Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43 (4), 551-560. doi: 10.1007/BF02293813
- Rosenberg, J., Beymer, P., Anderson, D., van Lissa, C.J., & Schmidt, J. (2018). tidyLPA: An R Package to Easily Carry Out Latent Profile Analysis (LPA) Using Open-Source or Commercial Software. *Journal of Open Source Software*, 3(30), 978. doi: 10.21105/joss.00978
- Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and “how to” guide of its application within vocational behavior research. *Journal of Vocational Behavior*, 120, 103445. doi: 10.1016/j.jvb.2020.103445
- Tein, J.Y., Coxe, S., & Cham, H. (2013). Statistical Power to Detect the Correct Number of Classes in Latent Profile Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 20 (4), 640-657. doi: 10.1080/10705511.2013.824781
- Torfs, P., & Brauer, C. (2014). A (very) short introduction to R. The Netherlands: Hydrology and Quantitative Water Management Group, Wageningen University. Ανακτήθηκε από: <https://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf>
- Wardenaar, K. J. (2021, April 9). Latent Profile Analysis in R: A tutorial and comparison to Mplus. doi: 10.31234/osf.io/wzfttr
- Williams, G., & Kibowski, F. (2016) Latent Class Analysis and Latent Profile Analysis. Στο L.A. Jason & D.S. Glenwick (επιμ.), *Handbook of Methodological Approaches to Community-Based Research: Qualitative, Quantitative, and Mixed Methods* (σ. 143-151). New York, NY: Oxford University Press

Παράρτημα Ι.
Ερωτηματολόγιο Δυνατοτήτων και Δυσκολιών (SDQ)

	Δεν ισχύει	Ισχύει κάπως	Ισχύει σίγουρα
1. Προσπαθώ να είμαι αρεστός/ή στους άλλους. Νοιάζομαι για τα αισθήματά τους	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Είμαι ανήσυχος/η. Δεν μπορώ να καθίσω ήσυχα για πολλή ώρα	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Έχω συχνά πονοκεφάλους, πόνο στο στομάχι ή αδιαθεσία	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Συνήθως μοιράζομαι με τους άλλους (φαγητό, παιχνίδια, στυλό κτλ)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Θυμώνω πολύ και συχνά χάνω την ψυχραιμία μου	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Συνήθως είμαι μόνος/η. Γενικά παίζω μόνος/η ή μένω με τον εαυτό μου	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Συνήθως κάνω αυτό που μου ζητάνε	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Ανησυχώ πολύ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Βοηθώ κάποιον/α που είναι πληγωμένος/η, αναστατωμένος/η ή αισθάνεται άρρωστος/η	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Συνεχώς στριφογυρίζω και κουνιέμαι	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Έχω ένα/μία ή περισσότερους/ες καλούς/ες φίλους/ες	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Μαλώνω πολύ. Μπορώ να αναγκάζω τους άλλους να κάνουν το δικό μου	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Αισθάνομαι συχνά δυστυχισμένος/η, αποκαρδιωμένος/η ή κλαίω	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Τα άτομα της ηλικίας μου γενικά με συμπαθούν	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. Είμαι συχνά αφηρημένος/η, δυσκολεύομαι να συγκεντρωθώ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. Είμαι νευρικός/η σε νέες καταστάσεις. Χάνω εύκολα την αυτοπεποίθησή μου	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17. Είμαι καλός/η με τα μικρότερα παιδιά	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18. Συχνά με κατηγορούν ότι λέω ψέματα ή εξαπατώ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19. Τα άλλα παιδιά ή οι νέοι/ες με κοροϊδεύουν ή με αγριεύουν	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20. Συχνά προσφέρομαι να βοηθήσω τους άλλους (Γονείς, καθηγητές, παιδιά)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21. Σκέφτομαι πριν κάνω κάτι	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22. Παίρνω πράγματα που δεν μου ανήκουν από το σπίτι, το σχολείο ή αλλού	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23. Τα πηγαίνω καλύτερα με τους ενήλικες παρά με άτομα της ηλικίας μου	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24. Έχω πολλούς φόβους, τρομάζω εύκολα	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25. Τελειώνω τη δουλειά που κάνω. Η προσοχή μου είναι καλή	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Παράρτημα II. Αποτελέσματα εκτέλεσης κώδικα στο RStudio

1. Αποτελέσματα της εφαρμογής της LPA

Αποτέλεσμα εντολής `comp$fits`:

```
> comp$fits
# A tibble: 36 x 19
  Model Classes LogLik AIC AWE BIC CAIC CLC KIC SABIC ICL Entropy prob_min prob_max
  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 1 1 -4959. 9937. 10069. 9979. 9989. 9919. 9950. 9948. -9979. 1 1 1
2 2 1 2 -4770. 9572. 9784. 9639. 9655. 9541. 9591. 9588. -9685. 0.856 0.860 0.983
3 3 1 3 -4744. 9531. 9824. 9623. 9645. 9489. 9556. 9553. -9739. 0.780 0.683 0.951
4 4 1 4 -4709. 9475. 9847. 9592. 9620. 9420. 9506. 9503. -9766. 0.735 0.731 0.918
5 5 1 5 -4674. 9415. 9868. 9557. 9591. 9349. 9452. 9449. -9693. 0.822 0.702 0.956
6 6 1 6 -4646. 9372. 9905. 9539. 9579. 9294. 9415. 9412. -9715. 0.787 0.743 0.922
7 7 1 7 -4623. 9339. 9952. 9531. 9577. 9248. 9388. 9385. -9712. 0.797 0.726 0.923
8 8 1 8 -4623. 9350. 10044. 9568. 9620. 9247. 9405. 9403. -9972. 0.669 0.00311 0.921
9 9 1 9 -4606. 9329. 10102. 9571. 9629. 9214. 9390. 9387. -9989. 0.681 0 0.937
10 2 1 -4959. 9937. 10069. 9979. 9989. 9919. 9950. 9948. -9979. 1 1 1
# ... with 26 more rows, and 5 more variables: n_min <dbl>, n_max <dbl>, BLRT_val <dbl>, BLRT_p <dbl>,
# Warnings <chr>
```

Αποτέλεσμα εντολής `comp$best`:

```
> comp$best
LogLik AIC AWE BIC CAIC CLC KIC SABIC ICL
30 30 11 12 12 30 30 30 30
```

Αποτέλεσμα εντολής `get_data(best_model)`:

```
> get_data(best_model)
# A tibble: 484 x 11
  model_number classes_number sdq_em sdq_cp sdq_hy sdq_pp sdq_ps CPROB1 CPROB2 CPROB3 class
  <dbl> <dbl> <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 2 2 3 0 3 5 1 8 4.85e-3 0.476 5.19e-1 3
2 2 2 3 2 1 5 4 10 3.20e-4 0.997 2.82e-3 2
3 2 2 3 4 5 7 2 10 9.57e-4 0.999 5.81e-9 2
4 2 2 3 3 1 2 1 9 9.40e-1 0.0310 2.90e-2 1
5 2 2 3 9 6 6 3 7 1.10e-15 1.00 9.88e-27 2
6 2 2 3 0 1 0 1 8 2.18e-3 0.00292 9.95e-1 3
7 2 2 3 1 1 4 2 5 8.59e-21 0.243 7.57e-1 3
8 2 2 3 0 3 0 0 9 4.77e-1 0.00916 5.14e-1 3
9 2 2 3 0 3 1 2 5 3.86e-21 0.0797 9.20e-1 3
10 2 2 3 0 1 2 0 10 8.47e-1 0.00213 1.51e-1 1
# ... with 474 more rows
```

Αποτέλεσμα εντολής `get_estimates(best_model)`:

```
> get_estimates(best_model)
Category Parameter Estimate se p class Model classes
1 Means sdq_em 1.8398905 0.6425326 4.189922e-03 1 2 3
2 Means sdq_cp 1.6327188 0.2702688 1.530738e-09 1 2 3
3 Means sdq_hy 1.8962111 0.4555948 3.153947e-05 1 2 3
4 Means sdq_pp 0.5536891 0.4058158 1.724463e-01 1 2 3
5 Means sdq_ps 9.6710996 0.3481643 8.134410e-170 1 2 3
6 Variances sdq_em 2.6296660 0.7368462 3.586030e-04 1 2 3
7 Variances sdq_cp 1.3158830 0.2428915 6.041261e-08 1 2 3
8 Variances sdq_hy 2.8305968 0.4772539 3.010808e-09 1 2 3
9 Variances sdq_pp 0.5242571 0.5106312 3.045691e-01 1 2 3
10 Variances sdq_ps 0.2283100 0.4485074 6.107215e-01 1 2 3
11 Means sdq_em 3.8667673 0.3361370 1.266053e-30 2 2 3
12 Means sdq_cp 3.4581337 0.3135300 2.748607e-28 2 2 3
13 Means sdq_hy 4.0715929 0.3653361 7.592643e-29 2 2 3
14 Means sdq_pp 2.4637311 0.2530713 2.131389e-22 2 2 3
15 Means sdq_ps 7.7268445 0.2845892 2.472569e-162 2 2 3
16 Variances sdq_em 5.5281105 0.7947665 3.509616e-12 2 2 3
17 Variances sdq_cp 3.1871504 0.4193526 2.957455e-14 2 2 3
18 Variances sdq_hy 4.5210642 0.5728282 2.961186e-15 2 2 3
19 Variances sdq_pp 4.0673587 0.6142332 3.547360e-11 2 2 3
20 Variances sdq_ps 3.2434598 0.5446633 2.601006e-09 2 2 3
21 Means sdq_em 0.6628861 0.3047150 2.959790e-02 3 2 3
22 Means sdq_cp 1.6948540 0.1942158 2.623164e-18 3 2 3
23 Means sdq_hy 1.5642743 0.4008456 9.522928e-05 3 2 3
```

24	Means	sdq_pp	0.9146661	0.2930220	1.799369e-03	3	2	3
25	Means	sdq_ps	7.5931421	0.4106861	2.535796e-76	3	2	3
26	Variances	sdq_em	0.7099161	0.4390322	1.058778e-01	3	2	3
27	Variances	sdq_cp	0.9192863	0.1872244	9.104004e-07	3	2	3
28	Variances	sdq_hy	1.9065442	0.6492894	3.320942e-03	3	2	3
29	Variances	sdq_pp	1.0132426	0.3528695	4.086097e-03	3	2	3
30	Variances	sdq_ps	1.7117302	0.3053532	2.073683e-08	3	2	3

Αποτέλεσμα εντολής get_fit (best_model):

```
> get_fit(best_model)
  Model      Classes      LogLik      AIC      AWE      BIC      CAIC
2.000000e+00 3.000000e+00 -4.600583e+03 9.265167e+03 9.691349e+03 9.398993e+03 9.430993e+03
          CLC      KIC      SABIC      ICL      Entropy      prob_min      prob_max
9.202637e+03 9.300167e+03 9.297428e+03 -9.544810e+03 7.353155e-01 8.270149e-01 9.036804e-01
          n_min      n_max      BLRT_val      BLRT_p
2.809917e-01 4.090909e-01 1.345179e+02 9.900990e-03
```

Αποτέλεσμα εντολής best_model\$model\$parameters\$pro:

```
> best_model$model$parameters$pro
[1] 0.2836007 0.4279445 0.2884549
```

Αποτέλεσμα εντολής best_model\$model\$parameters\$mean:

```
> best_model$model$parameters$mean
      [,1]      [,2]      [,3]
sdq_em 1.8398905 3.866767 0.6628861
sdq_cp 1.6327188 3.458134 1.6948540
sdq_hy 1.8962111 4.071593 1.5642743
sdq_pp 0.5536891 2.463731 0.9146661
sdq_ps 9.6710996 7.726844 7.5931421
```

Αποτέλεσμα εντολής best_model\$model\$parameters\$variance\$sigma:

```
> best_model$model$parameters$variance$sigma
, , 1
      sdq_em sdq_cp sdq_hy sdq_pp sdq_ps
sdq_em 2.629666 0.000000 0.000000 0.000000 0.000000
sdq_cp 0.000000 1.315883 0.000000 0.000000 0.000000
sdq_hy 0.000000 0.000000 2.830597 0.000000 0.000000
sdq_pp 0.000000 0.000000 0.000000 0.5242571 0.000000
sdq_ps 0.000000 0.000000 0.000000 0.000000 0.22831

, , 2
      sdq_em sdq_cp sdq_hy sdq_pp sdq_ps
sdq_em 5.52811 0.000000 0.000000 0.000000 0.000000
sdq_cp 0.000000 3.18715 0.000000 0.000000 0.000000
sdq_hy 0.000000 0.000000 4.521064 0.000000 0.000000
sdq_pp 0.000000 0.000000 0.000000 4.067359 0.000000
sdq_ps 0.000000 0.000000 0.000000 0.000000 3.24346

, , 3
      sdq_em sdq_cp sdq_hy sdq_pp sdq_ps
sdq_em 0.7099161 0.0000000 0.000000 0.000000 0.000000
sdq_cp 0.0000000 0.9192863 0.000000 0.000000 0.000000
sdq_hy 0.0000000 0.0000000 1.906544 0.000000 0.000000
sdq_pp 0.0000000 0.0000000 0.000000 1.013243 0.000000
sdq_ps 0.0000000 0.0000000 0.000000 0.000000 1.71173
```

2. Αποτελέσματα συσχέτισης προφίλ με φύλο και τάξη

Αποτέλεσμα εντολής pgender:

```
> pgender
class Αγόρια Κορίτσια
1 1 0.4000 0.6000
2 2 0.4343 0.5657
3 3 0.5662 0.4338
```

Αποτέλεσμα εντολής pgrade:

```
> pgrade
class Ε. Δημοτικού ΣΤ. Δημοτικού Α. Γυμνασίου Β. Γυμνασίου
1 1 0.2267 0.1333 0.5733 0.0667
2 2 0.1212 0.1364 0.6667 0.0758
3 3 0.0809 0.1985 0.6544 0.0662
```

Αποτέλεσμα εντολής pclassgender:

```
> pclassgender
φύλο class.1 class.2 class.3
1 Αγόρια 0.2691 0.3857 0.3453
2 Κορίτσια 0.3448 0.4291 0.2261
```

Αποτέλεσμα εντολής pclassgrade:

```
> pclassgrade
τάξη class.1 class.2 class.3
1 Ε Δημοτικού 0.4928 0.3478 0.1594
2 ΣΤ Δημοτικού 0.2703 0.3649 0.3649
3 Α Γυμνασίου 0.2801 0.4300 0.2899
4 Β Γυμνασίου 0.2941 0.4412 0.2647
```