

ΠΑΝΤΕΙΟΝ ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΟΙΝΩΝΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΗΣ & ΠΕΡΙΦΕΡΕΙΑΚΗΣ ΑΝΑΠΤΥΞΗΣ

PANTEION UNIVERSITY OF SOCIAL AND POLITICAL SCIENCES

DEPARTMENT OF ECONOMIC & REGIONAL DEVELOPMENT

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ  
«ΕΦΗΡΜΟΣΜΕΝΩΝ ΟΙΚΟΝΟΜΙΚΩΝ & ΔΙΟΙΚΗΣΗΣ»

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**  
**«Latent Class Analysis»**

Αγγελική Χωρική Α.Μ. 0814Μ025

Επιβλέπων Καθηγητής: Clive Richardson

Εξεταστική Επιτροπή: Σταύρος Ντεγιαννάκης, Β. Αρακελιάν



ΙΑΝΟΥΑΡΙΟΣ 2016

## ***Δήλωση πνευματικών δικαιωμάτων***

*Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω 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. Δηλώνω, συνεπώς, ότι αυτή η ΜΔΕ με τίτλο *Latent Class Analysis* προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά.*

*ΑΓΓΕΛΙΚΗ ΧΩΡΙΚΗ*

*13 ΙΑΝΟΥΑΡΙΟΥ 2016*

## ***Αφιέρωση***

*Σε όσους φανεί λίγο χρήσιμη αυτή η εργασία και  
σε αυτούς που με αγάπη συνέβαλαν στην ολοκλήρωσή της..*

## ΠΕΡΙΕΧΟΜΕΝΑ

<i>Δήλωση πνευματικών δικαιωμάτων</i>	2
<i>Αφιέρωση</i>	3
<b>ΠΕΡΙΛΗΨΗ</b>	6
<b>ABSTRACT</b>	7
<i>Ευχαριστίες</i>	8
<b>ΚΕΦΑΛΑΙΟ 1. Εισαγωγή</b>	9
<b>ΚΕΦΑΛΑΙΟ 2. Η ταξινόμηση των δεδομένων</b>	10
<b>ΚΕΦΑΛΑΙΟ 3. Η θεωρητική προσέγγιση της Latent Class Analysis</b>	11
<b>ΚΕΦΑΛΑΙΟ 4. Υπολογιστικά προγράμματα για τη Latent Class Analysis</b>	13
4.1 Υπολογιστικά προγράμματα γενικά	13
4.2 Αναλυτική περιγραφή χρήσης του R Studio και της εντολής PolCA	13
<b>ΚΕΦΑΛΑΙΟ 5. Εφαρμογή της μεθόδου Latent Class Analysis</b>	21
5.1 Δεδομένα	21
5.2 Προετοιμασία ανάλυσης	23
5.3 Ταξινόμηση	24
5.3.1 ΣΤ' Δημοτικού	24
5.3.2 Β' Γυμνασίου	29
5.3.3 Α' Λυκείου	32
5.3.4 Παρουσίαση συμπερασμάτων	34
5.3.4.1 Περιγραφική στατιστική δεδομένων	34
5.3.4.2 Συμπεράσματα	36
5.5 Ταξινόμηση με ανεξάρτητη μεταβλητή το φύλο	39
5.5.1 Εισαγωγή	39
5.5.2 ΣΤ' Δημοτικού	39
5.5.3 Β' Γυμνασίου	42
5.5.4 Α' Λυκείου	44
<b>ΚΕΦΑΛΑΙΟ 6. Άλλες εφαρμογές της Latent Class Analysis</b>	47
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ</b>	62

## ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ

Εικόνα 1 Το αναγνωριστικό εικονίδιο του προγράμματος R Studio	13
Εικόνα 2 Το πρόγραμμα R Studio	14
Εικόνα 3 Το R Studio κατά την διάρκεια της ανάλυσης	18
Εικόνα 4 Η εκφώνηση της ερώτησης από όπου προέρχονται τα δεδομένα της ανάλυσης (απόσπασμα από την ελληνική έκδοση του ερωτηματολογίου της έρευνας HBSC, 2010)	22
Εικόνα 5 Οι απαντήσεις των παιδιών σε Excel	23

## ΠΙΝΑΚΑΣ ΓΡΑΦΗΜΑΤΩΝ

Γράφημα 1 Στ' Δημοτικού_Κριτήριο BIC	26
Γράφημα 2 Στ' Δημοτικού_Ομαδοποίηση δραστηριοτήτων	28
Γράφημα 3 Στ' Δημοτικού_Δεύτερος τρόπος ομαδοποίησης δραστηριοτήτων	29
Γράφημα 4 Β' Γυμνασίου_Κριτήριο BIC	30
Γράφημα 5 Β' Γυμνασίου_Ομαδοποίηση δραστηριοτήτων	30
Γράφημα 6 Α' Λυκείου_Κριτήριο BIC	32
Γράφημα 7 Α' Λυκείου_Ομαδοποίηση δραστηριοτήτων	32
Γράφημα 8 Πληθυσμός κάθε ηλικίας	34
Γράφημα 9 Ποσοστό παιδιών ανά φύλο των τριών ηλικιών	34
Γράφημα 10 Ποσοστό πληθυσμού που ασχολείται με κάθε δραστηριότητα	35
Γράφημα 11 Οι τρεις ομαδοποιήσεις	36
Γράφημα 12 Πληθυσμός κάθε ταξινομημένης ομάδας των τριών ηλικιών	38
Γράφημα 13 Στ' Δημοτικού_Κριτήριο BIC	40
Γράφημα 14 Στ' Δημοτικού_Ομαδοποίηση δραστηριοτήτων βάσει του φύλου	40
Γράφημα 15 Β' Γυμνασίου_Κριτήριο BIC	42
Γράφημα 16 Β' Γυμνασίου_Ομαδοποίηση δραστηριοτήτων βάσει του φύλου	44
Γράφημα 17 Α' Λυκείου_Κριτήριο BIC	44
Γράφημα 18 Α' Λυκείου_Ομαδοποίηση δραστηριοτήτων βάσει του φύλου	46

## ΠΙΝΑΚΑΣ ΠΙΝΑΚΩΝ

Πίνακας 1 Στ' Δημοτικού_Κριτήριο BIC όλων των ομαδοποιήσεων	25
Πίνακας 2 Στ' Δημοτικού_Πιθανότητες των δραστηριοτήτων	27
Πίνακας 3 Β' Γυμνασίου_Κριτήριο BIC	29
Πίνακας 4 Β' Γυμνασίου_Πιθανότητες των δραστηριοτήτων	31
Πίνακας 5 Α' Λυκείου_Κριτήριο BIC	32
Πίνακας 6 Α' Λυκείου_Πιθανότητες των δραστηριοτήτων	33
Πίνακας 7 Ο πληθυσμός κάθε ηλικίας ανά φύλο	34
Πίνακας 8 Ο πληθυσμός κάθε ομάδας των τριών αναλύσεων	38
Πίνακας 9 Στ' Δημοτικού_κριτήριο BIC	39
Πίνακας 10 Στ' Δημοτικού_Πιθανότητες των δραστηριοτήτων ανά φύλο	41
Πίνακας 11 Β' Γυμνασίου_Κριτήριο BIC	42
Πίνακας 12 Β' Γυμνασίου_Πιθανότητες των δραστηριοτήτων ανά φύλο	43
Πίνακας 13 Α' Λυκείου_Κριτήριο BIC	44
Πίνακας 14 Α' Λυκείου_Πιθανότητες των δραστηριοτήτων ανά φύλο	45
Πίνακας 15 Οι τίτλοι των 20 μεταβλητών	49

## ΠΙΝΑΚΑΣ ΣΧΗΜΑΤΩΝ

Σχήμα 1 Τρόπος ταξινόμησης δεδομένων με τη Latent Class Analysis	11
--	----

## ΠΕΡΙΛΗΨΗ

Η παρούσα εργασία ασχολείται με μια σύγχρονη μέθοδο ανάλυσης δεδομένων, τη Latent Class Analysis. Αυτή η μέθοδος αποτελεί ένα χρήσιμο εργαλείο για την ομαδοποίηση παρατηρήσεων. Αφορά αναλύσεις δεδομένων πολλών μεταβλητών αποτελούμενες από κατηγορίες απαντήσεων. Οι απαντήσεις κάθε ατόμου διανέμονται σε ομάδες ανάλογα με τον συνδυασμό των απαντήσεων σε όλες τις μεταβλητές με τη μέθοδο αυτή. Υπολογίζονται οι πιθανότητες κάθε μεταβλητής και βάσει αυτών κατανέμονται σε κλάσεις, οι πιθανότητες προκύπτουν από τον τρόπο απαντήσεων κάθε ατόμου σε κάθε μεταβλητή ενώ οι κλάσεις προκύπτουν από τον τρόπο απαντήσεων στο σύνολο των μεταβλητών. Οι κλάσεις αυτές αποτελούν τις τάξεις της λανθάνουσας μεταβλητής. Τέλος, υπολογίζεται και το πλήθος του πληθυσμού που ανήκει σε κάθε μια από αυτές τις τάξεις. Η μέθοδος αυτή, γενικά, βοηθά τους αναλυτές να εντοπίζουν τους διάφορους συνδυασμούς μεταβλητών που δημιουργούνται μέσα στον πληθυσμό και τα χαρακτηριστικά κάθε μιας από αυτές τις ομάδες. Αυτή η ταξινόμηση δεδομένων μπορεί να χρησιμοποιηθεί και για ανάλυση παλινδρόμησης με τη βοήθεια μιας επιπλέον μεταβλητής, ανεξάρτητης. Ενώ, μεγάλο μέρος της εργασίας αποτελεί η πρακτική εφαρμογή της μεθόδου με την ταξινόμηση δεδομένων από την έρευνα του HBSC αφού προηγουμένως περιγράφηκε θεωρητικά η μέθοδος. Τα δεδομένα αφορούσαν δραστηριότητες με τις οποίες ασχολούνται τα παιδιά στον ελεύθερό τους χρόνο για τρεις ηλικίες, Στ' Δημοτικού (1339 παιδιά), Β' Γυμνασίου (1370 παιδιά) και Α' Λυκείου (1348 παιδιά), από πολλές χώρες του πλανήτη συμπεριλαμβανομένου και της Ελλάδας, το 2010. Η συγκεκριμένη ανάλυση πραγματοποιήθηκε με τη βοήθεια του προγράμματος R Studio και με τη χρήση του υπολογιστικού πακέτου PoLCA. Μεταβλητές του μοντέλου αποτέλεσαν οι είκοσι δραστηριότητες της ερώτησης 79 ενώ οι τέσσερις κατηγορίες απαντήσεων μετατράπηκαν σε δύο, με τιμές 1 και 2. Οι νέες ομάδες που σχηματίστηκαν με τη βοήθεια της μεθόδου αυτής παρουσίασαν τους διαφορετικούς συνδυασμούς δραστηριοτήτων που ασχολούνται τα παιδιά της ίδιας τάξης στον ελεύθερο χρόνο τους και στις τρεις ηλικίες. Ενώ, παρουσιάσθηκαν οι συνήθειες των παιδιών και βάσει το φύλο. Τελειώνοντας περιγράφηκαν άλλες περιπτώσεις χρήσης της μεθόδου σε διάφορους τομείς.

## Λέξεις-Κλειδιά

Ανάλυση δεδομένων, Ταξινόμηση δεδομένων, Latent Class Analysis, Λανθάνουσα μεταβλητή, Κλάσεις, Ανάλυση παλινδρόμησης, Πρόγραμμα R Studio, Υπολογιστικό πακέτο PoLCA

## **ABSTRACT**

Latent Class Analysis is a new statistical method for clustering multivariate categorical data. Observations are stratified based on an unobserved variable, the latent variable. The latent variable is, usually, categorical.

Each variable has its own probability distribution. Variables are graded into new classes, the latent classes, based on these probabilities. This is the basic model but in an extension to this model covariates which influence the latent class membership can be added. This is the latent class ‘regression’ model.

In this project, we show an application of this method to a practical data analysis. The data come from an HBSC survey recording children’s use of free time. We carried out the analysis using R Studio with the package PoLCA. The analysis includes three age groups from last class of primary school, second class of gymnasium and first class of high school. After fitting the basic model, we use a Latent class regression model to show the classifications of each group based on sex.

## **Keywords**

Classification, Latent Class Analysis, Latent class model, latent variable, latent classes, Latent class ‘regression’ model, R Studio, PoLCA

## *Ευχαριστίες*

Θα ήθελα να ευχαριστήσω τον Καθηγητή Clive-Στέφανο Richardson για την συμβολή του, τις καίριες επισημάνσεις του και για το χρόνο που αφιέρωσε στην προσπάθεια αυτή. Επίσης, θα ήθελα να ευχαριστήσω την Ελευθερία Καναβού για την πρόθυμη προσφορά να μεταδώσει τις γνώσεις της σχετικά με το υπολογιστικό πακέτο PoLCA και το χρόνο που αφιέρωσε γι' αυτό το σκοπό. Η συμβολή της ήταν πολύ χρήσιμη. Τέλος, θέλω να ευχαριστήσω θερμώς όλους εκείνους που αφανώς βοήθησαν να υπάρξουν οι κατάλληλες συνθήκες για να περατωθεί το εκπόνημα αυτό.



## ΚΕΦΑΛΑΙΟ 1. Εισαγωγή

Αρχικά, η απλή συλλογή δεδομένων που χρησιμοποιείται στη Στατιστική σήμερα χρησιμοποιήθηκε από τον Όμηρο κατά τον Τρωικό Πόλεμο για τον κατάλογο των πλοίων (*το νεών κατάλογο*). Να σημειωθεί ότι η πρώτη απλή συλλογή στοιχείων χωρίς ανάλυση έγινε το 2238 π.Χ. στην Κίνα με Αυτοκράτορα τον Γιάο με στόχο την απογραφή του πληθυσμού. Έπειτα, το 1620 ο Τζον Γκράουντ ξεκίνησε δειγματοληπτική έρευνα σε οικογένειες του Λονδίνου για την καταγραφή θανάτων από την ασθένεια πανώλη. Πολλοί θεωρούν το 1663 αφετηρία της Στατιστικής καθώς ο Τζον Γκράουντ εξέδωσε το βιβλίο «Φυσικές και Πολιτικές παρατηρήσεις σχετικά με τους πίνακες Θνησιμότητας». Μέχρι τον 18<sup>ο</sup> αι. η Στατιστική ασχολείται κυρίως με δημογραφικά θέματα και είναι κυρίως Περιγραφική. Στα μέσα του 19<sup>ου</sup> αι στη Ρωσία μετατράπηκε από μια απλή τεχνική σε Επιστήμη λόγω της θεωρίας των πιθανοτήτων (ΠΗΓΗ: <http://www.army.gr/files/File/epitheorisi>).

Επομένως, η Στατιστική αποτελεί μια σύγχρονη επιστήμη προερχόμενη από τον κλάδο των Εφαρμοσμένων Μαθηματικών η οποία ασχολείται με δεδομένα. Η λέξη Στατιστική βγαίνει από τον λατινικό όρο “status” που σημαίνει κράτος αλλά και από το αρχαίο ρήμα “ίστημι” που σημαίνει ταξινομώ (ΠΗΓΗ: <https://el.wikipedia.org>). Σκοπός της είναι να συγκεντρώσει, να ταξινομήσει, να αναλύσει και να εξάγει συμπεράσματα για μια σειρά δεδομένων ώστε να ληφθούν ορθές αποφάσεις.

Όλη αυτή η διαδικασία είναι ιδιαίτερα χρήσιμη στο κλάδο των επιχειρήσεων, των βιομηχανιών, των Επιστημών και στην κοινωνία ολόκληρη. Συγκεκριμένα, στις Ιατρικές επιστήμες/με την βιοστατιστική, στις κοινωνικές επιστήμες για ζητήματα της κοινής γνώμης, σε ερευνητικά κέντρα, σε εταιρίες έρευνας αγοράς, σε επιχειρήσεις με τμήμα R&D, στις ψυχολογικές και στις πολιτικές επιστήμες είναι χρήσιμη κυρίως για την πρόβλεψη αποτελεσμάτων ώστε να είναι εγκυρότερες οι αποφάσεις που λαμβάνονται, είναι όμως χρήσιμη και σε οποιοδήποτε άτομο θέλει να παράγει γνώση και να μειώσει τις συνθήκες αβεβαιότητας σε μια κοινωνία.

Στην εν λόγω εργασία, όμως, θα ασχοληθούμε με ένα τμήμα των λειτουργιών της Στατιστικής, με μια μέθοδο ταξινόμησης δεδομένων. Στη σημερινή εποχή και λόγω της χρήσης των υπολογιστικών συστημάτων παρουσιάζεται συχνά η ανάγκη να ταξινομηθούν και να οργανωθούν δεδομένα που είναι πολύ χρήσιμα για την λήψη αποφάσεων. Όμως, η επεξεργασία αυτών των δεδομένων καθίσταται δύσκολη λόγω της ποσότητάς τους, υπάρχει μεγάλη πιθανότητα να γίνει λάθος, να χαθούν πληροφορίες με αποτέλεσμα να εξαχθούν λάθος συμπεράσματα. Επομένως, υπάρχει ανάγκη για μεθόδους που να διευκολύνουν αυτό το έργο. Αυτές τις μεθόδους τις παίρνουμε από την Στατιστική. Στη παρούσα προσπάθεια θα ασχοληθούμε, ιδιαιτέρως με τη ταξινόμηση δεδομένων αναλύοντας μια νέα αποτελεσματική μεθοδολογία που επιτελεί αυτό το σκοπό, τη Latent Class Analysis.

## ΚΕΦΑΛΑΙΟ 2. Η ταξινόμηση των δεδομένων

Η ταξινόμηση των δεδομένων προηγείται της επεξεργασίας και της ανάλυσης πληροφοριών ενός προβλήματος με πλήθος δεδομένων. Η ταξινόμηση είναι ιδιαίτερα χρήσιμη σε έναν αναλυτή, καθώς ορθή ταξινόμηση σημαίνει και σωστά συμπεράσματα στην ανάλυση. Η αρχική εικόνα των δεδομένων πολλές φορές είναι χαοτική με αποτέλεσμα να είναι ιδιαίτερα πολύπλοκη η διαδικασία εξαγωγής συμπερασμάτων.

Αυτό που συμβαίνει σε αυτό το στάδιο είναι ότι τα δεδομένα αρχικά ομαδοποιούνται βασιζόμενοι σε κάποια κριτήρια και έπειτα ο αναλυτής επεξεργάζεται τις ομάδες αυτές και τα χαρακτηριστικά τους και όχι μια σειρά δεδομένων. Έτσι, γίνεται πιο εύκολη η διαδικασία της ανάλυσης των δεδομένων. Με την ομαδοποίηση γίνονται κατανοητά και εμφανή τα αποτελέσματα, όπως επίσης, εντοπίζονται εύκολα και γρήγορα οι πιο χρήσιμες πληροφορίες. Με την σημερινή ραγδαία εξέλιξη της τεχνολογίας διάφορα πακέτα υπολογιστικά συμβάλουν στο να γίνεται αρκετά γρήγορα και με εύχρηστο τρόπο η ομαδοποίηση. Παίρνοντας έτσι άμεσα αποτελέσματα από την ανάλυση και την ταξινόμηση, απαιτώντας πάντα από τον αναλυτή να τα παρατηρεί με κριτικό μάτι.

Έτσι, κάποιες από τις μεθόδους ταξινόμησης που χρησιμοποιούνται συνήθως πρόκειται να αναφέρουμε παρακάτω με ορισμένα χαρακτηριστικά τους. Κάθε φορά επιλέγεται από τον αναλυτή η μέθοδος που είναι πιο χρήσιμη ανάλογα με το υπό μελέτη θέμα. Ορισμένες από τις μεθόδους είναι η Ανάλυση σε Κύριες Συνιστώσες (Principal Components Analysis), η Factor Analysis, η Ανάλυση Συστάδων (Cluster Analysis) και τέλος η Latent Class Analysis.

Παρακάτω θα αναφερθούν λίγα στοιχεία για κάθε μία από τις παραπάνω μεθόδους (ΠΗΓΗ: [http://books.eudoxus.gr/publishers/CID\\_821/cid\\_00821-0368-ABS.pdf](http://books.eudoxus.gr/publishers/CID_821/cid_00821-0368-ABS.pdf)).

### ➤ Principal Components Analysis

Εντοπίζει γραμμικές σχέσεις μεταξύ των δεδομένων ώστε οι νέες συνιστώσες (δηλαδή οι μεταβλητές) που σχηματίζονται, να είναι ασυσχέτιστες μεταξύ τους.

### ➤ Factor Analysis

Με τη μέθοδο αυτή εντοπίζονται παράγοντες μη μετρήσιμοι που όμως συνδέουν τις παρατηρήσιμες μεταβλητές.

### ➤ Cluster Analysis

Δημιουργούνται ομάδες με παρατηρήσεις που είναι ομοιογενείς σε σχέση με τις μεταβλητές δηλαδή είναι ομάδες που έχουν παρόμοια χαρακτηριστικά. Συνήθως αυτό εξετάζεται με την μέτρηση της απόστασης μεταξύ των παρατηρήσεων.

### ➤ Latent Class Analysis

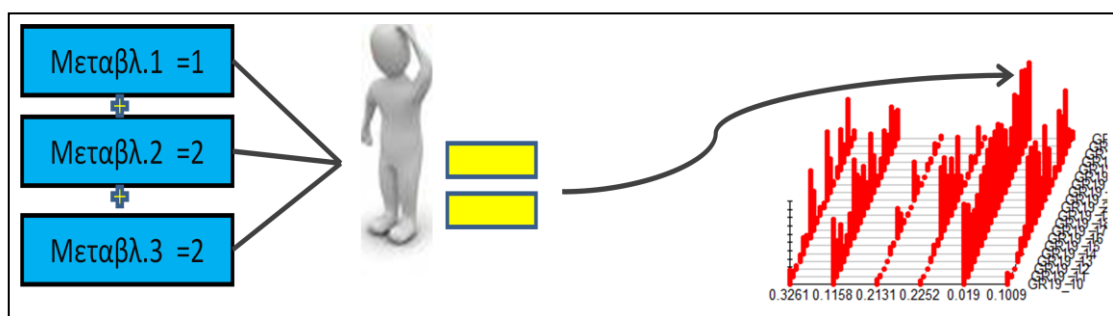
Εντοπίζονται και διαμορφώνονται νέες ομάδες που δεν ήταν εμφανείς, πριν. Κάθε περίπτωση απαντήσεων ταξινομείται σε μια ομάδα ανάλογα με τον συνδυασμό τους. Σε σύγκριση με τις άλλες μεθόδους, το θετικό είναι ότι περιέχει στατιστικό μοντέλο και οι παρατηρήσεις κατανέμονται βάσει πιθανοτήτων κάτι που κάνει την ανάλυση πιο ακριβής, όπως επίσης προσφέρει μια πληθώρα διαθέσιμων εργαλείων για την εκτίμηση του κατάλληλου μοντέλου και τον καταλληλότερο αριθμό ομάδων ώστε να είναι επιτυχημένη η ανάλυση.

Στο επόμενο μέρος της εργασίας θα αναλύσουμε διεξοδικά τη νέα αυτή στατιστική μέθοδο Πολυπαραγοντικής Ανάλυσης, τη Latent Class Analysis. Θα περιγράψουμε την ταξινόμηση με βάση τη μέθοδο αυτή, την ωφέλειά της και τον τρόπο χρήσης της.

### ΚΕΦΑΛΑΙΟ 3. Η θεωρητική προσέγγιση της Latent Class Analysis

Η Latent Class Analysis είναι μία στατιστική μέθοδος που ομαδοποιεί δεδομένα. Χρησιμοποιείται για αναλύσεις πολλών μεταβλητών με κατηγορίες απαντήσεων. Η συνηθέστερη μορφή μεταβλητών είναι οι δίτιμες, εκείνες δηλαδή που έχουν δύο κατηγορίες απαντήσεων όπως είναι ΚΑΤΑΦΑΣΗ και ΑΡΝΗΣΗ.

Σε αυτή την στατιστική μέθοδο περιέχεται στατιστικό μοντέλο με πιθανότητες διανομής σε κλάσεις κάτι που καθιστά ακριβέστερη την ταξινόμηση. Η μέθοδος αυτή λειτουργεί ως εξής: Εντοπίζεται τι απαντήσεις έδωσε κάθε άτομο σε όλες τις μεταβλητές μιας ερώτησης. Ενώ, όλα τα άτομα που απάντησαν με παρόμοιο τρόπο σε όλες τις μεταβλητές δημιουργούν μια ομάδα. Υπολογίζονται, το ποσοστό του πληθυσμού της ομάδας αυτής και η πιθανότητα κάθε μιας μεταβλητής η οποία και προσδιορίζει την τάση του πληθυσμού αυτού να απαντά με έναν τρόπο σε μια μεταβλητή. Έτσι, δημιουργούνται αρκετές νέες ομάδες που η κάθε μια περιλαμβάνει πιθανότητες κατανομής και το σύνολο του πληθυσμού αυτής. Αυτές οι ομάδες πριν ήταν κρυφές ενώ, μετά από αυτή την επεξεργασία παρουσιάστηκαν.



Σχήμα 1 Τρόπος ταξινόμησης δεδομένων με τη Latent Class Analysis

Αυτό που έχει σημασία σε αυτή την ομαδοποίηση είναι η απάντηση και όχι ποια είναι η μεταβλητή της απάντησης. Με αυτό τον τρόπο διαφαίνονται διαφορετικοί τρόποι απαντήσεων των ατόμων σε ένα πληθυσμό. Άρα αυτό που κάνουμε σε αυτή τη μεθοδολογία είναι ότι υποθέτουμε νέες μεταβλητές που δημιουργούν νέες ομάδες στις οποίες μπορούν να αντιστοιχηθούν όλες οι παρατηρήσεις. Ο αριθμός των ομάδων αρχικά βρίσκεται κάνοντας δοκιμές ενώ με τον καιρό γίνεται με την εμπειρία.

Αυτή η ανάλυση είναι χρήσιμη γιατί γνωρίζει κανείς από την αρχή τον διαχωρισμό των ομάδων που σκοπεύει να κάνει, η ομαδοποίηση είναι αρκετά συνοψισμένη και τέλος βγαίνουν συνολικά αποτελέσματα. Βασική και κύρια προϋπόθεση είναι οι μεταβλητές να είναι αμοιβαίως ανεξάρτητες μεταξύ τους μέσα σε κάθε ομάδα.

Γενικά, με αυτή τη μέθοδο υπολογίζεται το ποσοστό του πληθυσμού που ανήκει σε κάθε ομάδα, την πιθανότητα να ανήκει σε μια κατηγορία της μεταβλητής κάθε παρατήρηση και κατά συνέπεια και ένα μέρος του πληθυσμού μιας ομάδας, την πιθανότητα να ανήκει κάθε περίπτωση σε μια ομάδα ανάλογα με τον τρόπο απαντήσεων σε όλες τις παραμέτρους και τις μεταβλητές οι οποίες αντιστοιχούν στον πληθυσμό κάθε ομάδας (Linzer & Lewis, 2011).

Η ιστορική πορεία της μεθόδου αυτής ξεκινά από το **1950** με τον **Lazarsfeld** ο οποίος σχημάτισε μια πιο απλή μορφή με το όνομα 'Latent Structure Analysis' για να φτιάξει τύπους για δίτιμες ευδιάκριτες μεταβλητές. Είκοσι χρόνια μετά (**1970**) ο **Goodman** συνέθεσε έναν αλγόριθμο που εκτιμά τη μέγιστη πιθανοφάνεια των παραμέτρων του

μοντέλου, ενώ εφαρμόστηκε και σε μοντέλα με περισσότερες από μία λανθάνουσες μεταβλητές και σε άλλα με περισσότερες από μία κατηγορίες απαντήσεων. Ο **Haberman** το **1979** φανέρωσε τη σύνδεση των LC μοντέλων με την ανάλυση παλινδρόμησης. Τέλος, το **1990** οι **Hagenaars** και **Vermunt** το **1997** έθεσαν το γενικό πλαίσιο της μεθόδου (Vermunt & Magidson, 2004).

Η μέθοδος αυτή έχει χρησιμοποιηθεί σε διαφορετικά επιστημονικά πεδία όπως είναι η Ιατρική, η Αστρονομία, η Βιολογία, η Οικονομία, η Φαρμακευτική, το Marketing.

Παρακάτω, θα αναλύσουμε τη μέθοδο με τη χρήση μαθηματικών τύπων.

Το πιο σημαντικό στο Latent Class μοντέλο είναι να εκτιμηθεί η πιθανότητα μια περίπτωση ενώ ανήκει σε μια κλάση να έχει ένα συγκεκριμένο συνδυασμό μεταβλητών.

Ο τύπος για να βρίσκεται αυτή η πιθανότητα είναι:

$$P(Y=y | X=x) = \prod_{l=1}^L P(Y_l = y_l | X=x)$$

Όπου  $Y$  συμβολίζει τις γνωστές μεταβλητές με  $Y_l$  μία από αυτές τις γνωστές μεταβλητές πλήθους από 1,2,...,L.  $X$  ορίζεται η λανθάνουσα μεταβλητή η οποία παίρνει τιμές που έχουμε ήδη ορίσει, όσες είναι και οι λανθάνουσες τάξεις. Εκτιμώντας αυτό το τύπο για όλες τις latent τάξεις μπορούμε να συγκρίνουμε τα αποτελέσματα και να προσδιορίσουμε τις διαφορές και τα χαρακτηριστικά της κάθε μιας μεταβλητής αλλά και κλάσης.

Η γενική πιθανότητα ένα άτομο να έχει ένα συγκεκριμένο συνδυασμό αποτελεσμάτων των μεταβλητών είναι η παραπάνω εξίσωση.

$$P(Y=y) = \sum_{x=1}^C P(X=x) \prod_{l=1}^L P(Y_l = y_l | X=x)$$

Ενώ, η πιθανότητα ένα άτομο που έχει ένα συνδυασμό απαντήσεων σε όλες τις μεταβλητές να ανήκει σε μια από τις ομάδες υπολογίζεται από κανόνα του *Bayes* (Vermunt & Magidson, 2004):

$$P(X=x | Y=y) = \frac{P(X=x)P(Y=y|X=x)}{P(Y=y)}$$

Εκτός από το βασικό μοντέλο μπορεί να γίνει ανάλυση παλινδρόμησης του μοντέλου, υπολογίζοντας τις πιθανότητες να ανήκουν οι παρατηρήσεις σε μία τάξη της λανθάνουσας μεταβλητής επηρεαζόμενοι από ανεξάρτητη μεταβλητή. Αυτό είναι το latent class μοντέλο παλινδρόμησης.

Το κριτήριο με το οποίο επιλέγεται το πιο κατάλληλο μοντέλο είναι το Bayesian Information Criterion αλλιώς BIC (από τον Schwartz το 1978). Επιλέγεται συνήθως το μοντέλο με τη χαμηλότερη τιμή του κριτηρίου αυτού.

## ΚΕΦΑΛΑΙΟ 4. Υπολογιστικά προγράμματα για τη Latent Class Analysis

### 4.1 Υπολογιστικά προγράμματα γενικά

Προχωράμε τη μελέτη αυτή περιγράφοντας πως μπορεί να χρησιμοποιηθεί η τεχνολογία για την εφαρμογή αυτής της μεθόδου. Έτσι, αρκετά είναι τα προγράμματα που κυκλοφορούν στο εμπόριο και χρησιμοποιούνται για ανάλυση δεδομένων με τη χρήση της μεθόδου Latent Class Analysis. Η αλήθεια είναι ότι η τεχνολογία πολύ συνέβαλε με αυτό τον τρόπο καθώς γίνεται πιο γρήγορα η ανάλυση εξοικονομώντας έτσι χρόνο και κόστος. Ο αναλυτής δεν χρειάζεται να κάνει όλες αυτές τις μαθηματικές πράξεις που απαιτούνται για την ομαδοποίηση των δεδομένων κινδυνεύοντας να κάνει λάθος. Επίσης, η ανάλυση γίνεται πιο αξιόπιστη καθώς εισάγοντας την αντίστοιχη εντολή πραγματοποιούνται επαναλαμβανόμενοι έλεγχοι στα αποτελέσματα του προγράμματος. Στη συγκεκριμένη μέθοδο επιπλέον να σημειωθεί ότι μπορούμε να είμαστε πιο ασφαλείς καθώς περιέχει στατιστικό μοντέλο με όλους τους ελέγχους που χρειάζεται.

Αρκετά είναι τα προγράμματα που προσφέρονται γι' αυτή τη χρήση. Πρώτο έκανε την εμφάνισή του το MILSA, έπειτα το LEM, Mplus, Latent GOLD και το R Studio μαζί με την εντολή PoLCA (Vermunt & Magidson, 2004). Μερικά από αυτά διανέμονται δωρεάν στο διαδίκτυο, ένα από τα οποία θα χρησιμοποιήσουμε και εμείς στη συγκεκριμένη ανάλυση και είναι το R Studio με την εντολή PoLCA (Linzer & Lewis, 2011). Επιλέχτηκε το συγκεκριμένο πακέτο καθώς είναι αρκετά φιλικό στον χρήστη, προσφέρεται δωρεάν στο διαδίκτυο και επίσης περιέχει αρκετά χρήσιμα εργαλεία για την Latent Class Analysis (Torfs & Brauer, 2014). Στο επόμενο κεφάλαιο πρόκειται να περιγραφεί αναλυτικά ο τρόπος χρήσης του προγράμματος αυτού πριν προχωρήσουμε στην πρακτική εφαρμογή της μεθόδου.

### 4.2 Αναλυτική περιγραφή χρήσης του R Studio και της εντολής PoLCA

Καταρχάς, το πρόγραμμα R Studio έχει για αναγνωριστικό εικονίδιο το παρακάτω σχήμα:



Εικόνα 1 Το αναγνωριστικό εικονίδιο του προγράμματος R Studio

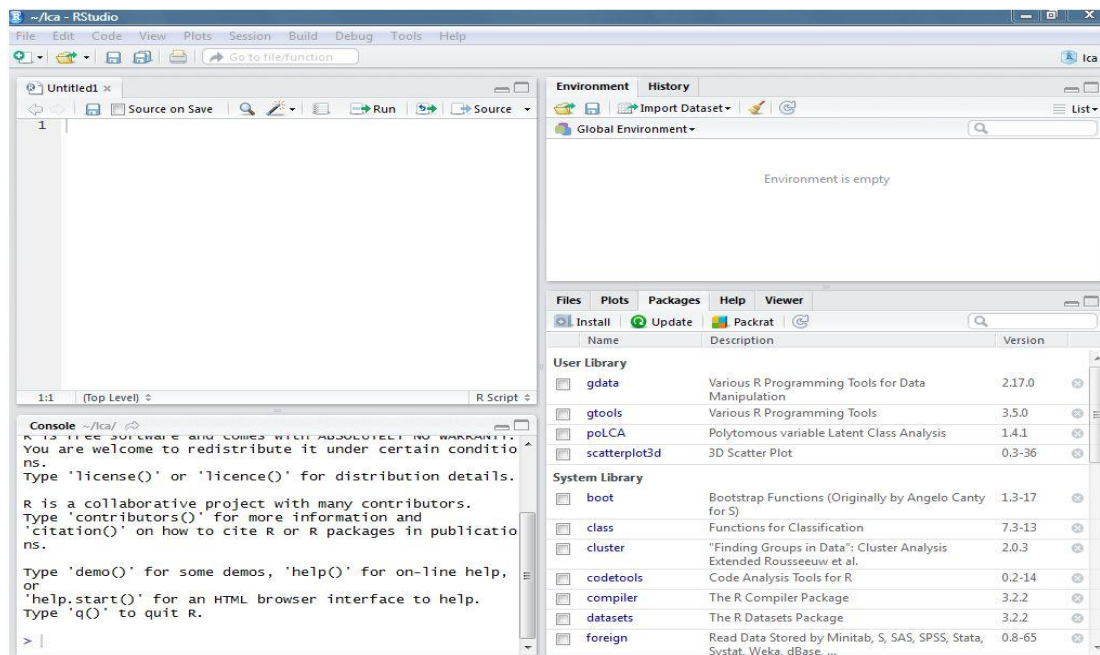
Υπάρχει και το πρόγραμμα R επίσης δωρεάν στο διαδίκτυο αλλά προτιμάται το R Studio καθώς περιέχει περισσότερα εργαλεία. Ακολουθεί η αναλυτική περιγραφή της διαδικασίας για να αποκτήσει κάποιος το πρόγραμμα στον υπολογιστή του.

Αρχικά, πληκτρολογούμε τη διεύθυνση <http://www.rstudio.org/>, έπειτα κάνουμε 'κλικ' στο Download R Studio και επιλέγουμε Desktop. Στο σημείο που μεταφερόμαστε επιλέγουμε το DOWNLOAD RSTUDIO DESKTOP, επιλέγουμε από την λίστα Installers τον κατάλληλο τύπο που ταιριάζει στον υπολογιστή μας, αν για παράδειγμα έχει κάποιος τα Windows 10 θα κάνει κλικ στην πρώτη επιλογή. Πληκτρολογούμε *Συνέχεια* στο παράθυρο που εμφανίζεται, έπειτα *Επόμενο* και



καταχωρούμε τον φάκελο που θέλουμε να αποθηκευτεί το πρόγραμμα και τέλος επιλέγουμε *Εγκατάσταση* (Torfs & Brauer, 2014).

Καθώς επιλέγουμε το πρόγραμμα που έχουμε εγκαταστήσει εισερχόμαστε στο περιβάλλον του προγράμματος R Studio, το οποίο απεικονίζεται στην Εικόνα 2 που ακολουθεί.



Εικόνα 2 Το πρόγραμμα R Studio

Το πρόγραμμα R Studio αποτελεί ένα στατιστικό περιβάλλον για αναλύσεις και επεξεργασία δεδομένων. Είναι μία γλώσσα βασισμένη στη γλώσσα S.

Αναλυτικά, **πάνω αριστερά** είναι το editor window (αν δεν είναι ανοιχτό το παράθυρο αυτό μπορεί να ανοιχτεί από το File -> New File -> R script). Σε αυτό το παράθυρο εισάγεται ο φάκελος και τα δεδομένα που χρειάζονται για την ανάλυση. Επίσης, σε αυτό το παράθυρο καταχωρούνται οι εντολές της ανάλυσης και επεξεργασίας των δεδομένων. Αποθηκεύεται ότι καταχωρείται σε αυτό το σημείο. **Κάτω αριστερά** βρίσκεται το Console Window στο οποίο το πρόγραμμα εξάγει τα αποτελέσματα από της εντολές που έχουμε εισάγει στο πάνω παράθυρο. Εφόσον καταχωρήσουμε στο editor window την εντολή που θέλουμε πατάμε Ctrl + Enter οπότε το αποτέλεσμα τυπώνεται στο Console Window. Επομένως, είναι ένα πολύ βασικό παράθυρο. **Πάνω δεξιά** βρίσκεται το ιστορικό, δηλαδή όλες οι εντολές, τα δεδομένα και οι φάκελοι που έχουν καταχωρηθεί. **Κάτω δεξιά** το πρόγραμμα αποδίδει τις γραφικές παραστάσεις που σχηματίζονται. Επίσης σε αυτό το σημείο παρουσιάζονται τα υπολογιστικά πακέτα που έχουν αποθηκευτεί στο πρόγραμμα ώστε να επιλεγεί το κατάλληλο κάθε φορά ανάλογα με την ανάλυση που θέλει κάποιος να πραγματοποιήσει (Torfs & Brauer, 2014).

Όπως έχουμε προαναφέρει για να πραγματοποιήσουμε την ανάλυση με τη μέθοδο Latent Class Analysis θα χρησιμοποιήσουμε ένα πακέτο στο περιβάλλον του προγράμματος R Studio. Το πακέτο αυτό είναι το PoLCA (Linzer & Lewis, 2011), όπως παρατηρούμε, όμως, δεν περιέχεται στη λίστα με τα υπολογιστικά πακέτα. Έτσι, θα πρέπει να προμηθευτούμε και αυτό από το διαδίκτυο και να το αποθηκεύσουμε στη λίστα με τα πακέτα του προγράμματος. Αυτό θα το πράξουμε ακολουθώντας την παρακάτω διαδικασία. Είτε κάνοντας εγκατάσταση από τη διεύθυνση του διαδικτύου <https://cran.r-project.org/web/packages/poLCA/index.html>, είτε καθώς έχουμε ανοιχτό το R Studio επιλέγουμε από το **κάτω δεξί** παράθυρο packages -> Install και συμπληρώνουμε το όνομα του πακέτου που θέλουμε και τον προορισμό του επιλέγοντας install.

Το PoLCA είναι ένα πακέτο στατιστικής ανάλυσης για πολλές μεταβλητές οι οποίες μπορεί να είναι είτε δίτιμες είτε με περισσότερες από δύο επιλογές απαντήσεων. Η λέξη PoLCA προέρχεται από τα αρχικά της λέξης ‘Polytomous Variable Latent Class Analysis’. Έτσι, δίνεται στο πρόγραμμα η εντολή και πραγματοποιείται η Latent Class ανάλυση χρησιμοποιώντας τη λέξη PoLCA.

Ακολουθεί η αναλυτική διαδικασία για την ανάλυση χρησιμοποιώντας το πρόγραμμα R Studio και την εντολή PoLCA.

Εφόσον καταχωρήσουμε μια εντολή για να τρέξει θα πρέπει είτε να κάνουμε ‘κλικ’ στην επιλογή *run* είτε να πληκτρολογούμε Ctrl + Enter για να παραχθούν τα αποτελέσματα στο κάτω παράθυρο, το Console Window.

Αρχικά, απαιτείται να ορίσουμε τον φάκελο που βρίσκεται στον υπολογιστή και από τον οποίο θα πάρουμε τα δεδομένα που πρόκειται να αναλύσουμε έπειτα. Η εντολή δίνεται πληκτρολογώντας στο πάνω αριστερό παράθυρο (editor) το *setwd* και έπειτα τον φάκελο, για παράδειγμα *setwd('C:/Users/aggeliki/Desktop/ΔΕΔΟΜΕΝΑ')*.

Ύστερα, χρειάζεται να διαβαστεί από το πρόγραμμα το συγκεκριμένο αρχείο του φακέλου. Το αρχείο μπορεί να είναι φύλλο εργασίας του Excel ή SPSS. Να επισημανθεί ότι αν είναι αρχείο του Excel τότε θα πρέπει να γίνει αποθήκευση ως και να αλλάξει ο τύπος του σε ‘.csv’. Για παράδειγμα, θέλουμε να αναλύσουμε τα δεδομένα του αρχείου Grade 6 και να επισημάνουμε ότι ήδη χωρίζονται με ; οι μεταβλητές και ότι περιλαμβάνονται τίτλοι στην πρώτη σειρά του αρχείου, τότε πληκτρολογούμε *grade6<-read.csv("Grade 6.csv", sep=";", header=TRUE)*.

Έπειτα, καλούμε το υπολογιστικό πακέτο της βιβλιοθήκης που πρόκειται να χρησιμοποιήσουμε στην ανάλυση που έπεται. Πληκτρολογούμε *Library(poLCA)*.

Συνεχίζουμε με το να αποθηκεύσουμε το αρχείο των δεδομένων, που είχε διαβαστεί προηγουμένως, στο περιβάλλον του προγράμματος. Για παράδειγμα για το αρχείο Grade 6 εισάγουμε την εντολή στο πάνω αριστερό παράθυρο, όπως και τις προηγούμενες εντολές, *attach(as.data.frame(grade6))*.

Ακολουθώντας, για να είναι εύκολη η ανάλυση και να μην υπάρχει πρόβλημα από το πλήθος των μεταβλητών ονομάζουμε όλες τις μεταβλητές με μία μεταβλητή. Παρακάτω αντικαθιστούμε τις είκοσι μεταβλητές του παραδείγματός μας με την *fr1*:

```
fr1<-cbind(GR19_10, GR19_11, GR19_12, GR19_13,
           GR19_14, GR19_15, GR19_16, GR19_17,
           GR19_18, GR19_19, GR19_20,
           GR19_01, GR19_02, GR19_03, GR19_04,
           GR19_05, GR19_06, GR19_07, GR19_08,
           GR19_09)~1
```

Το ~1 δείχνει ότι το μοντέλο αποτελεί το βασικό μοντέλο της Latent Class Analysis και όχι της παλινδρόμησης. Τέλος, η ανάλυση με τη μέθοδο latent class analysis πραγματοποιείται με την παρακάτω εντολή:

```
lca1<-poLCA(fr1,as.data.frame(όνομα αρχείου),nclass=1,maxiter=2000,nrep=10)
```

Δηλαδή δίνουμε την εντολή να σχηματιστεί μία ομάδα και να υπολογιστούν οι πιθανότητες των παρατηρήσεων να ανήκουν σε αυτή την ομάδα. Το poLCA χρησιμοποιεί τον αλγόριθμο Expectation-Maximization (EM) για να εξάγει το πιο αποδεκτό αποτέλεσμα. Η παραπάνω εντολή περιγράφεται ως εξής: Τα δεδομένα του αρχείου κατανέμονται σε μία τάξη, επαναλαμβάνοντας την διαδικασία 10 φορές για να μειωθούν τα σφάλματα.

Επίσης, μπορεί να κάνει και γράφημα αν συμπληρώσουμε την εντολή όπως παρακάτω:

```
lca2<-poLCA(fr1,as.data.frame(όνομα αρχείου),nclass=2,maxiter=2000,nrep=10,
graphs=TRUE)
```

Να σημειώσουμε εδώ ότι δεν μπορεί να δημιουργηθεί γράφημα για την ταξινόμηση δεδομένων σε μια ομάδα (lca1) γι' αυτό χρησιμοποιήθηκε παράδειγμα με δύο κλάσεις (lca2). Ένα σημαντικό ερώτημα αποτελεί πόσες τάξεις να φτιάξει το σύστημα. Η απάντηση είναι ότι αρχικά κάνουμε υποθέσεις. Δηλαδή δίνουμε εντολή να παρουσιάσει αποτελέσματα ομαδοποιήσεων από την πρώτη περίπτωση της μιας κλάσης μέχρι και τον μέγιστο αριθμό κλάσεων που θεωρούμε πιθανόν να αποτελεί λύση. Για παράδειγμα ομαδοποίηση δεδομένων σε δέκα κλάσεις ίσως δεν έχει νόημα για έναν μικρό αριθμό δεδομένων αλλά προκαλεί και προβλήματα στην ανάλυση. Έτσι αφού πληκτρολογήσουμε την εντολή όσες φορές θέλουμε να δοκιμάσει το σύστημα πατάμε *run*. Για παράδειγμα για να ομαδοποιηθούν τα δεδομένα μέχρι εννιά κλάσεις πληκτρολογούμε:

```
lca1<-poLCA(fr1,as.data.frame(grade6),nclass=1,maxiter=2000,nrep=10)
```

```
lca2<-poLCA(fr1,as.data.frame(grade6),nclass=2,maxiter=2000,nrep=10,  
graphs=TRUE)
```

```
lca3<-poLCA(fr1,as.data.frame(grade6),nclass=3,maxiter=2000,nrep=10,  
graphs=TRUE)
```

```
lca4<-poLCA(fr1,as.data.frame(grade6),nclass=4,maxiter=2000,nrep=10,  
graphs=TRUE)
```

```
lca5<-poLCA(fr1,as.data.frame(grade6),nclass=5,maxiter=2000,nrep=10,  
graphs=TRUE)
```

```
lca6<-poLCA(fr1,as.data.frame(grade6),nclass=6,maxiter=2000,nrep=10,  
graphs=TRUE)
```

```
lca7<-poLCA(fr1,as.data.frame(grade6),nclass=7,maxiter=2000,nrep=10,  
graphs=TRUE)
```

```
lca8<-poLCA(fr1,as.data.frame(grade6),nclass=8,maxiter=2000,nrep=10,  
graphs=TRUE)
```

```
lca9<-poLCA(fr1,as.data.frame(grade6),nclass=9,maxiter=2000,nrep=10,  
graphs=TRUE)
```

Ποια όμως από όλες τις ομαδοποιήσεις θα επιλέξουμε; Αυτό εξαρτάται από ένα κριτήριο που έχει διαπιστωθεί ως το πιο αξιόπιστο, το BIC (Bayesian Information Criterion). Συγκρίνοντας τα BIC όλων των ομαδοποιήσεων επιλέγουμε συνήθως αυτό με τη χαμηλότερη τιμή και με ικανό αριθμό κλάσεων.

Για να εμφανιστούν όλα τα BIC στη σειρά και να μπορέσουμε να τα συγκρίνουμε, καταχωρούμε την παρακάτω εντολή:

```
bics<-cbind(lca1$bic, lca2$bic, lca3$bic, lca4$bic, lca5$bic lca6$bic lca7$bic  
lca8$bic lca9$bic)
```

Έπειτα επιλέγουμε την καταλληλότερη ώστε να ακολουθήσει η επεξεργασία των συμπερασμάτων.

Με την διαδικασία που προηγήθηκε πραγματοποιείται η ομαδοποίηση των δεδομένων με την μέθοδο Latent Class Analysis και η επιλογή της καλύτερης περίπτωσης ομαδοποίησης από όλες. Από κει και έπειτα έγκειται στην ευχέρεια του αναλυτή να επεξεργαστεί και να εξάγει συμπεράσματα για τα αποτελέσματα της ανάλυσης.

Στην περίπτωση, όμως, που θέλουμε να πραγματοποιήσουμε ανάλυση παλινδρόμησης, να προσθέσουμε δηλαδή μια ανεξάρτητη μεταβλητή που θα επηρεάσει την ομαδοποίηση των δεδομένων, θα ακολουθήσουμε τα παρακάτω βήματα:



Στην αντικατάσταση όλων των μεταβλητών με μία θα εισάγουμε:

```
fr2<-cbind(GR19_10, GR19_11, GR19_12, GR19_13,  
          GR19_14, GR19_15, GR19_16, GR19_17,  
          GR19_18, GR19_19, GR19_20,  
          GR19_01, GR19_02, GR19_03, GR19_04,  
          GR19_05, GR19_06, GR19_07, GR19_08,  
          GR19_09)~M1
```

Όπου M1 στο συγκεκριμένο παράδειγμα είναι η ανεξάρτητη μεταβλητή φύλο, ενώ τις γνωστές μεταβλητές τις αντικαθιστούμε με το fr2.

Έπειτα, η εντολή για την ανάλυση είναι ίδια μόνο που δεν δίνουμε το ίδιο όνομα όπως την πρώτη ανάλυση αλλά διαφορετικό. Για παράδειγμα:

```
lca1sex<-poLCA(fr2,as.data.frame(grade6),nclass=1,maxiter=2000,nrep=10)
```

```
lca2sex<-poLCA(fr2,as.data.frame(grade6),nclass=2,maxiter=2000,nrep=10,  
graphs=TRUE)
```

```
lca3sex<-poLCA(fr2,as.data.frame(grade6),nclass=3,maxiter=2000,nrep=10,  
graphs=TRUE)
```

```
lca4sex<-poLCA(fr2,as.data.frame(grade6),nclass=4,maxiter=2000,nrep=10,  
graphs=TRUE)
```

Και η εντολή για το κριτήριο BIC θα πάρει την εξής μορφή:

```
bicssex<-cbind(lca1sex$bic, lca2sex$bic, lca3sex$bic, lca4sex$bic)  
bicssex
```

Τελειώνοντας, να σημειώσουμε ότι αν επιθυμούμε τα διαγράμματα που απεικονίζονται στο δεξί κάτω παράθυρο να είναι ταξινομημένα, δηλαδή οι ομάδες του διαγράμματος να εμφανίζονται από την μεγαλύτερη σε πληθυσμό στην μικρότερη από τα δεξιά στα αριστερά, γράφουμε την εντολή:

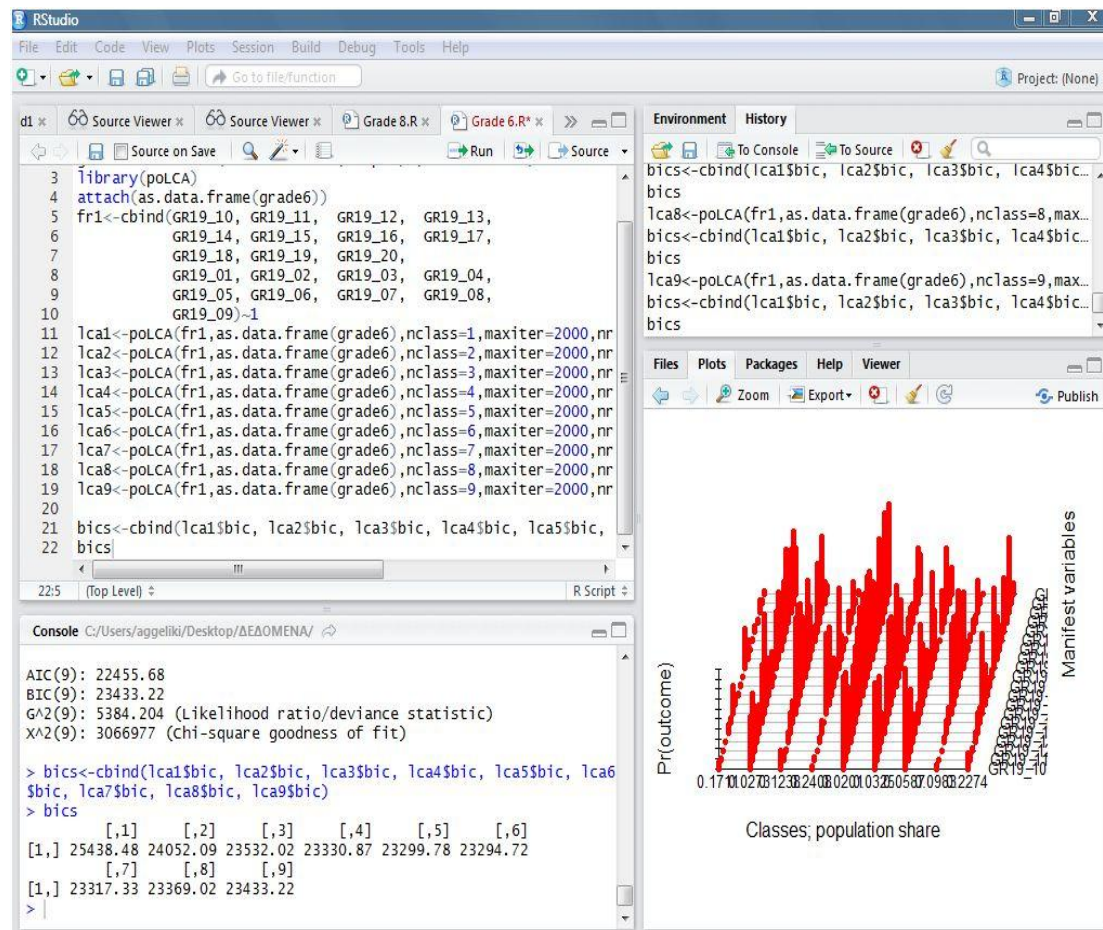
```
lca4sex
```

```
probs.start.new<-poLCA.reorder(lca4sex$probs.start,order(lca4sex$P,decreasing=TRUE))
```

```
lca4sex<-poLCA(fr2,as.data.frame(grade6),nclass=4,probs.start=probs.start.new,  
graphs=TRUE)
```

Αφού λοιπόν πληκτρολογήσουμε τις εντολές κάνουμε ‘κλικ’ στο *run*.

Ενώ, το R Studio κατά την διάρκεια της ανάλυσης θα είναι όπως απεικονίζεται στην Εικόνα 3.



Εικόνα 3 Το R Studio κατά την διάρκεια της ανάλυσης

Ακολουθούν τα αποτελέσματα που παίρνουμε καθώς επιλέγουμε *run* στο παράθυρο Console. Να σημειωθεί ότι λόγω του μεγάλου όγκου δεδομένων δεν παραθέτονται τα αποτελέσματα για όλες τις ομαδοποιήσεις στο κύριο μέρος της εργασίας αλλά στο Παράρτημα. Ενδεικτικά δίνονται παρακάτω τα αποτελέσματα μιας ομαδοποίησης ώστε να μπορέσει κάποιος να κατανοήσει την λειτουργία του προγράμματος και τον τρόπο παρουσίασης των αποτελεσμάτων.

```

> lca9<-polCA(fr1,as.data.frame(grade6),nclass=9,maxiter=2000,
nrep=10, graphs=TRUE)
Model 1: llik = -11080.46 ... be:
Model 2: llik = -11059.61 ... be:
Model 3: llik = -11075.08 ... be:
Model 4: llik = -11081.91 ... be:
Model 5: llik = -11045.76 ... be:
Model 6: llik = -11039.22 ... be:
Model 7: llik = -11045.76 ... be:
Model 8: llik = -11061.92 ... be:
Model 9: llik = -11051.51 ... be:
Model 10: llik = -11107.97 ... b
Conditional item response (column
by outcome variable, for each c

class 6: 0.2576 0.7424
class 7: 0.8307 0.1693
class 8: 0.4798 0.5202
class 9: 0.0000 1.0000

$GR19_11
      Pr(1) Pr(2)
class 1: 1.0000 0.0000
class 2: 0.1343 0.8657
class 3: 0.9813 0.0187
class 4: 0.9361 0.0639
class 5: 0.9870 0.0130
class 6: 0.4247 0.5753
class 7: 0.9851 0.0149
class 8: 0.3397 0.6603
class 9: 0.8210 0.1790

$GR19_10
      Pr(1) Pr(2)
class 1: 0.9674 0.0326
class 2: 0.0000 1.0000
class 3: 0.9752 0.0248
class 4: 0.8511 0.1489
class 5: 0.9359 0.0641

$GR19_12
      Pr(1) Pr(2)
class 1: 1.0000 0.0000
class 2: 0.3346 0.6654
class 3: 1.0000 0.0000
class 4: 0.9799 0.0201
class 5: 1.0000 0.0000

```

class 6: 0.3513 0.6487  
class 7: 1.0000 0.0000  
class 8: 0.8078 0.1922  
class 9: 0.3957 0.6043

\$GR19\_13

	Pr(1)	Pr(2)
class 1:	0.9719	0.0281
class 2:	0.4027	0.5973
class 3:	0.9645	0.0355
class 4:	0.9162	0.0838
class 5:	1.0000	0.0000
class 6:	0.7244	0.2756
class 7:	0.9249	0.0751
class 8:	0.6512	0.3488
class 9:	0.9503	0.0497

\$GR19\_14

	Pr(1)	Pr(2)
class 1:	0.8816	0.1184
class 2:	0.3690	0.6310
class 3:	0.8879	0.1121
class 4:	0.8068	0.1932
class 5:	0.9328	0.0672
class 6:	0.7231	0.2769
class 7:	0.8181	0.1819
class 8:	0.7378	0.2622
class 9:	0.8397	0.1603

\$GR19\_15

	Pr(1)	Pr(2)
class 1:	0.8423	0.1577
class 2:	0.1976	0.8024
class 3:	1.0000	0.0000
class 4:	0.8497	0.1503
class 5:	0.9562	0.0438
class 6:	0.4203	0.5797
class 7:	0.7989	0.2011
class 8:	0.4185	0.5815
class 9:	0.8170	0.1830

\$GR19\_16

	Pr(1)	Pr(2)
class 1:	0.1063	0.8937
class 2:	0.0839	0.9161
class 3:	0.1975	0.8025
class 4:	0.2022	0.7978
class 5:	0.7214	0.2786
class 6:	0.1091	0.8909
class 7:	0.0434	0.9566
class 8:	0.0000	1.0000
class 9:	0.1045	0.8955

\$GR19\_17

	Pr(1)	Pr(2)
class 1:	0.1276	0.8724
class 2:	0.0458	0.9542
class 3:	0.2234	0.7766
class 4:	0.6039	0.3961
class 5:	0.9207	0.0793
class 6:	0.0946	0.9054
class 7:	0.0000	1.0000
class 8:	0.7692	0.2308
class 9:	0.3824	0.6176

\$GR19\_18

	Pr(1)	Pr(2)
class 1:	0.2018	0.7982
class 2:	0.0000	1.0000
class 3:	0.7795	0.2205
class 4:	0.9637	0.0363
class 5:	0.9792	0.0208
class 6:	0.1891	0.8109
class 7:	0.0924	0.9076
class 8:	0.6376	0.3624

class 9: 0.6773 0.3227

\$GR19\_19

	Pr(1)	Pr(2)
class 1:	0.0456	0.9544
class 2:	0.0423	0.9577
class 3:	0.7270	0.2730
class 4:	0.9163	0.0837
class 5:	0.9861	0.0139
class 6:	0.0781	0.9219
class 7:	0.3009	0.6991
class 8:	0.6622	0.3378
class 9:	0.6098	0.3902

\$GR19\_20

	Pr(1)	Pr(2)
class 1:	0.2981	0.7019
class 2:	0.0000	1.0000
class 3:	0.9165	0.0835
class 4:	0.9277	0.0723
class 5:	1.0000	0.0000
class 6:	0.2605	0.7395
class 7:	0.7173	0.2827
class 8:	1.0000	0.0000
class 9:	0.8911	0.1089

\$GR19\_01

	Pr(1)	Pr(2)
class 1:	0.2489	0.7511
class 2:	0.0000	1.0000
class 3:	0.2130	0.7870
class 4:	0.2001	0.7999
class 5:	0.6045	0.3955
class 6:	0.3062	0.6938
class 7:	0.1232	0.8768
class 8:	0.4379	0.5621
class 9:	0.0605	0.9395

\$GR19\_02

	Pr(1)	Pr(2)
class 1:	0.8359	0.1641
class 2:	0.1635	0.8365
class 3:	0.7617	0.2383
class 4:	0.5901	0.4099
class 5:	0.8172	0.1828
class 6:	0.7185	0.2815
class 7:	0.4821	0.5179
class 8:	0.3350	0.6650
class 9:	0.5344	0.4656

\$GR19\_03

	Pr(1)	Pr(2)
class 1:	0.9805	0.0195
class 2:	0.5861	0.4139
class 3:	0.9611	0.0389
class 4:	0.9536	0.0464
class 5:	0.9976	0.0024
class 6:	0.8624	0.1376
class 7:	0.9509	0.0491
class 8:	0.8644	0.1356
class 9:	0.9613	0.0387

\$GR19\_04

	Pr(1)	Pr(2)
class 1:	0.9524	0.0476
class 2:	0.3211	0.6789
class 3:	0.9876	0.0124
class 4:	0.8553	0.1447
class 5:	0.9958	0.0042
class 6:	0.8091	0.1909
class 7:	0.8569	0.1431
class 8:	0.4211	0.5789
class 9:	1.0000	0.0000

\$GR19\_05

```

      Pr(1)  Pr(2)
class 1: 0.8858 0.1142
class 2: 0.0000 1.0000
class 3: 0.8155 0.1845
class 4: 0.4644 0.5356
class 5: 0.9482 0.0518
class 6: 0.7102 0.2898
class 7: 0.3134 0.6866
class 8: 0.1926 0.8074
class 9: 0.6017 0.3983

```

\$GR19\_06

```

      Pr(1)  Pr(2)
class 1: 0.9397 0.0603
class 2: 0.0936 0.9064
class 3: 0.9159 0.0841
class 4: 0.7689 0.2311
class 5: 0.9790 0.0210
class 6: 0.7382 0.2618
class 7: 0.4907 0.5093
class 8: 0.3657 0.6343
class 9: 0.8383 0.1617

```

\$GR19\_07

```

      Pr(1)  Pr(2)
class 1: 0.6778 0.3222
class 2: 0.0000 1.0000
class 3: 0.7032 0.2968
class 4: 0.2546 0.7454
class 5: 0.8813 0.1187
class 6: 0.5816 0.4184
class 7: 0.2309 0.7691
class 8: 0.3615 0.6385
class 9: 0.3668 0.6332

```

\$GR19\_08

```

      Pr(1)  Pr(2)
class 1: 0.9490 0.0510
class 2: 0.0806 0.9194
class 3: 0.9715 0.0285
class 4: 0.8953 0.1047
class 5: 0.9721 0.0279
class 6: 0.4465 0.5535
class 7: 0.7976 0.2024
class 8: 0.2759 0.7241
class 9: 0.8783 0.1217

```

\$GR19\_09

```

      Pr(1)  Pr(2)
class 1: 0.9900 0.0100
class 2: 0.0639 0.9361
class 3: 0.9714 0.0286
class 4: 0.8936 0.1064
class 5: 0.9699 0.0301
class 6: 0.6169 0.3831
class 7: 0.9125 0.0875
class 8: 0.5485 0.4515
class 9: 0.7322 0.2678

```

Estimated class population shares  
 0.0983 0.0162 0.2399 0.2283 0.1704  
 0.0396 0.1182 0.0272 0.0618

Predicted class memberships (by modal  
 posterior prob.)  
 0.0993 0.0172 0.227 0.236 0.1777  
 0.0358 0.1187 0.0246 0.0635

=====  
 Fit for 9 latent classes:  
 =====

number of observations: 1339  
 number of estimated parameters:  
 188  
 residual degrees of freedom: 1151  
 maximum log-likelihood: -11039.22

AIC(9): 22454.45  
 BIC(9): 23431.99  
 G<sup>2</sup>(9): 5382.966 (Likelihood ratio/deviance  
 statistic)  
 X<sup>2</sup>(9): 1988153 (Chi-square goodness  
 of fit)

```

> bics<-cbind(lca1$bic, lca2$bic, lca3$bic,
  lca4$bic, lca5$bic, lca6$bic, lca7$bic,
  lca8$bic, lca9$bic)
> bics
      [,1]      [,2]      [,3]      [,4]
[1,] 25438.48 24052.09 23532.02 23330.87
      [,5]      [,6]      [,7]
[1,] 23299.78 23294.72 23317.33
      [,8]      [,9]
[1,] 23368.83 23431.99

```

Έτσι, παραθέσαμε μια αναλυτική παρουσίαση του τρόπου με τον οποίο χρησιμοποιούμε το πρόγραμμα και την παραπάνω εντολή για να πραγματοποιηθεί η Latent Class Analysis. Στο επόμενο τμήμα της παρούσας εργασίας προχωράμε στην πρακτική εφαρμογή της μεθόδου εκτελώντας ένα παράδειγμα ανάλυσης δεδομένων.

## ΚΕΦΑΛΑΙΟ 5. Εφαρμογή της μεθόδου Latent Class Analysis

### 5.1 Δεδομένα

Σε αυτό το τμήμα της παρούσας μελέτης πρόκειται να πραγματοποιήσουμε ανάλυση δεδομένων από μια έρευνα με την μέθοδο Latent Class Analysis. Στόχος είναι να ομαδοποιηθούν οι απαντήσεις των ερωτώμενων ώστε να απεικονιστούν ευκρινώς τα αποτελέσματα και τα συμπεράσματα της έρευνας.

Η μέθοδος αυτή μπορεί να χρησιμοποιηθεί στην ανάλυση των συγκεκριμένων δεδομένων καθώς αυτά τα δεδομένα αποτελούν απαντήσεις από πολλές παραμέτρους σε μια ερώτηση της έρευνας και στην συγκεκριμένη ερώτηση περιλαμβάνονται κατηγορίες απαντήσεων για όλες τις παραμέτρους.

Τα δεδομένα πάρθηκαν από έρευνα του HBSC (ΠΗΓΗ: [www.hbsc.org](http://www.hbsc.org)) που διενεργήθηκε σε παιδιά τριών ηλικιών το 2010 σε πολλές χώρες του πλανήτη συμπεριλαμβανομένου και της Ελλάδας. Τις τρεις κατηγορίες ηλικιών αποτελούν τα παιδιά της Στ' Δημοτικού, της Β' Γυμνασίου και της Α' Λυκείου. Το πλήθος των παιδιών αποτελούν 1339 της Στ' Δημοτικού, 1370 της Β' Γυμνασίου και 1348 της Α' Λυκείου. Συνολικά απάντησαν 4057 παιδιά για τις τρεις ηλικίες. Ενώ, έχουν αφαιρεθεί από το σύνολο των απαντημένων ερωτηματολογίων όσα δεν ήταν πλήρως απαντημένα αλλά είχαν κενά. Από το σύνολο της έρευνας πρόκειται να αναλύσουμε τα δεδομένα της ερώτησης 79 της έρευνας η οποία περιέχει μία λίστα με είκοσι δραστηριότητες και η κάθε μία έχει τέσσερις απαντήσεις από τις οποίες τα παιδιά πρέπει να επιλέξουν μόνο μία. Το θέμα της ερώτησης ήταν ο ελεύθερος χρόνος των παιδιών των παραπάνω ηλικιών. Δηλαδή τους δινότανε μία λίστα με είκοσι δραστηριότητες και κλήθηκαν σε κάθε μια από αυτές να απαντήσουν κατά πόσο ασχολούνται μαζί τους στον ελεύθερό τους χρόνο. Όπως κατανοούμε ο όγκος των δεδομένων που συλλέχθηκαν συνολικά ήταν τεράστιος εφόσον κάθε ερωτηματολόγιο από τα 4057 περιείχε 20 απαντήσεις, επομένως  $4057 * 20 = 81140$  πληροφορίες για ταξινόμηση, αυτό σημαίνει ένα αρκετά σύνθετο πρόβλημα όπου απαιτείται αρκετός χρόνος. Εδώ βοηθούν ιδιαιτέρως οι στατιστικές μέθοδοι ανάλυσης και η τεχνολογία στην εξομάλυνση τέτοιων προβλημάτων.

Η ερώτηση αποτελείται από είκοσι δραστηριότητες με τέσσερις πιθανές απαντήσεις. Αυτό που πρόκειται να κάνουμε είναι να ταξινομήσουμε σε κλάσεις ανάλογα με τον τρόπο που απάντησαν σε όλες τις δραστηριότητες, κάθε άτομο.

Στην Εικόνα 4 παρουσιάζεται η εκφώνηση της ερώτησης 79 της έρευνας της οποίας τις απαντήσεις πρόκειται να επεξεργαστούμε.



# Ο ελεύθερος χρόνος μου

79. Ποια από τις παρακάτω δραστηριότητες κάνεις συνήθως στον ελεύθερό σου χρόνο και πόσο συχνά; Εάν δεν κάνεις κάποια δραστηριότητα, μαύρισε τον κύκλο στην πρώτη στήλη και προχώρησε στην επόμενη δραστηριότητα.

[Σημείωσε μία απάντηση σε κάθε σειρά]

	Δεν κάνω αυτή τη δραστηριότητα	2-3 φορές το μήνα ή σπανιότερα	Περίπου μια φορά την εβδομάδα	2 φορές την εβδομάδα ή περισσότερο
Φυσική άσκηση, αθλητισμό (π.χ., γυμναστική, ποδόσφαιρο, κτλ.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Χορό	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Προσκοπισμό	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Γράφω διηγήματα, ποιήματα, επιστολές (όχι για το σχολείο)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ζωγραφική, χειροτεχνία	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Φωτογραφία	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Διαβάζω βιβλία	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Γράφω/συνθέτω μουσική	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Συμμετέχω σε χορωδία	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Παίζω κάποιο μουσικό όργανο	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Τραγουδάω ή παίζω σε κάποιο συγκρότημα, μπάντα ή ορχήστρα	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Κάνω ιδιαίτερα μαθήματα μουσικής	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Παίζω θέατρο	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Συμμετέχω σε δραστηριότητες της εκκλησίας	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Πηγαίνω σε συναυλίες, θέατρο, μουσεία κτλ.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ακούω την αγαπημένη μου μουσική	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Παίζω ηλεκτρονικά παιχνίδια στον υπολογιστή	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Σχεδιάζω ή επεξεργάζομαι εικόνες στον υπολογιστή	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
«Κατεβάζω» μουσική από το διαδίκτυο	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Φτιάχνω δικά μου προγράμματα ή παιχνίδια στον υπολογιστή	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Εικόνα 4 Η εκφώνηση της ερώτησης από όπου προέρχονται τα δεδομένα της ανάλυσης (απόσπασμα από την ελληνική έκδοση του ερωτηματολογίου της έρευνας HBSC, 2010)

Τα αποτελέσματα που παίρνουμε αποτελούν δεδομένα φύλλου εργασίας του Excel. Η λίστα με τις δραστηριότητες αποτελούν την πρώτη οριζόντια γραμμή του Excel. Ενώ, στις οριζόντιες σειρές είναι καταχωρημένες οι απαντήσεις κάθε παιδιού σε κάθε μεταβλητή που είναι κάθε στήλη. Οι είκοσι δραστηριότητες-στήλες είναι οι μεταβλητές του προβλήματος με τίτλο από GR19\_01 μέχρι GR19\_20. Να σημειωθεί ότι στο φύλλο εργασίας η σειρά των μεταβλητών δεν είναι ίδια με αυτή της εκφώνησης αλλά δίνονται πρώτα οι απαντήσεις των έντεκα τελευταίων και έπειτα ακολουθούν οι εννιά πρώτες μεταβλητές. Δηλαδή, πρώτες στήλες από τις είκοσι μεταβλητές του Excel είναι 'Τραγουδάω ή παίζω σε κάποιο συγκρότημα, μπάντα ή ορχήστρα' έως 'Φτιάχνω δικά μου προγράμματα ή παιχνίδια στον υπολογιστή' (GR19\_10 έως GR19\_20). Ενώ, ύστερα ακολουθούν οι επόμενες εννιά από 'Φυσική άσκηση, αθλητισμό (πχ. Γυμναστική, ποδόσφαιρο κτλ.)' έως 'παίζω κάποιο μουσικό όργανο' (GR19\_01 έως GR19\_09). Στις πρώτες στήλες του αρχείου περιέχονται κωδικοί των παιδιών ενώ ακολουθεί η ανεξάρτητη μεταβλητή της ανάλυσης που θα ακολουθήσει 'Φύλο' (M1) όπου παίρνει τιμές για Αγόρι 1 & για Κορίτσι 2 και M2 είναι η τάξη που ανήκουν, πιο συγκεκριμένα, Στ' Τάξη Δημοτικού= 1, Β' Γυμνασίου= 2 και Α' Λυκείου= 3. Πρόκειται να ταξινομήσουμε δεδομένα τριών αρχείων καθώς θέλουμε να εξάγουμε συμπεράσματα για τρεις ηλικίες, τα αρχεία έχουν τα ονόματα Grade 6 παιδιά της Στ' τάξης του Δημοτικού, Grade 8 παιδιά της Β' τάξης Γυμνασίου και Grade 10 της Α' τάξης του Λυκείου.

Στην Εικόνα 5 που ακολουθεί απεικονίζεται η μορφή των δεδομένων ενός από τα αρχεία.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1	ID1GR	ID2	ID3	ID4	M1	M2	GR19_10	GR19_11	GR19_12	GR19_13	GR19_14	GR19_15	GR19_16	GR19_17	GR19_18	GR19_19	GR19_20	GR19_01	GR19_02
2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	2	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
4	1	1	1	4	2	1	2	2	2	1	1	1	2	1	1	2	1	2	2
5	1	1	1	6	2	1	1	2	1	1	1	1	2	1	1	1	1	2	2
6	1	1	1	8	2	1	1	2	2	2	1	2	2	1	1	1	1	2	2
7	1	1	1	9	2	1	1	1	1	1	1	1	2	1	1	1	1	2	2
8	1	1	1	10	2	1	2	1	1	1	1	2	2	1	1	2	1	1	1
9	1	1	1	12	2	1	1	1	1	1	1	1	2	2	2	1	1	1	1
10	1	1	1	13	1	1	1	1	1	1	1	2	2	2	2	2	1	2	2
11	1	1	1	14	1	1	1	1	1	1	1	1	1	2	1	2	1	2	2
12	1	1	1	15	2	1	2	2	1	1	1	2	2	2	1	1	1	2	2
13	1	1	1	16	1	1	1	1	1	1	1	1	2	1	1	1	1	2	2
14	1	2	2	17	1	1	1	1	1	1	1	1	1	2	2	2	1	2	2
15	1	2	2	18	1	1	1	1	1	1	1	1	2	2	1	1	1	2	2

Εικόνα 5 Οι απαντήσεις των παιδιών σε Excel

## 5.2 Προετοιμασία ανάλυσης

Όπως προαναφέραμε η ανάλυση περιέχει πληροφορίες από είκοσι μεταβλητές με απαντήσεις τεσσάρων κατηγοριών. Όμως προς διευκόλυνση της συγκεκριμένης ανάλυσης θα μετατρέψουμε τις τέσσερις κατηγορίες σε δύο. Δηλαδή, οι δύο πρώτες θα γίνουν μία κατηγορία όπου θα περιλαμβάνονται τα παιδιά που σπάνια κάνουν μια δραστηριότητα ή και καθόλου οι άλλες δύο θα αποτελούν την δεύτερη τάξη όπου θα περιέχονται τα παιδιά που κάνουν τακτικά μία δραστηριότητα. Δίνουμε σε κάθε μία κατηγορία έναν αριθμό.

Δεν κάνω αυτή τη δραστηριότητα (1)

2-3 φορές το μήνα ή σπανιότερα (2)

Περίπου μία φορά την εβδομάδα (3)

2 φορές την εβδομάδα ή περισσότερο (4)

Ο χωρισμός των τεσσάρων απαντήσεων θα γίνει ως εξής: Οι δύο νέες κατηγορίες που θα φτιάξουμε θα έχουν τους αριθμούς 1 και 2 στο Excel, όπου 1 θα είναι όσα παιδιά κάνουν μία δραστηριότητα σπάνια ή και καθόλου δηλαδή θα περιέχει τις κατηγορίες 1 & 2 ενώ το 2 θα είναι όσα παιδιά κάνουν μια δραστηριότητα τακτικά, δηλαδή κατηγορίες 3 & 4. Αντικαθιστούμε πρώτα τους αριθμούς 1 & 2 με 1 ενώ έπειτα αντικαθιστούμε τα 3 & 4 με 2 για να μην υπάρξει πρόβλημα με τον αριθμό 2 κατά την δεύτερη αντικατάσταση. Να σημειωθεί σε αυτό το σημείο ότι επιλέχθηκαν αυτοί οι δύο αριθμοί και όχι 0 και 1 διότι το πρόγραμμα δεν μπορεί να αναγνωρίσει το 0 αλλά το απορρίπτει.

Για να χρησιμοποιήσουμε τα αρχεία του Excel στο πρόγραμμα R Studio θα πρέπει να τα μετατρέψουμε προηγουμένως σε μορφή .csv. Την μετατροπή αυτή μπορούμε να την εκτελέσουμε αν επιλέξουμε *αποθήκευση ως* και αλλάξουμε την μορφή του αρχείου επιλέγοντας csv.

Έχοντας ολοκληρώσει με την διαδικασία της προετοιμασίας μπορούμε να προχωρήσουμε στην κυρίως ανάλυση και ομαδοποίηση των δεδομένων με τη μέθοδο Latent Class Analysis. Να υπολογίσουμε δηλαδή την πιθανότητα να κάνει ένα παιδί μία δραστηριότητα. Ενώ, το σύνολο των περιπτώσεων που κάνουν παρόμοιο σύνολο δραστηριοτήτων αποτελούν το πληθυσμό μιας ομάδας. Έτσι, σχηματίζονται οι κλάσεις που η κάθε μία περιέχει το ποσοστό του πληθυσμού με παρόμοιες συνήθειες στον ελεύθερο χρόνο τους. Το πρόγραμμα υπολογίζει το ποσοστό που πληθυσμού που τείνει να παίρνει κάθε μια από τις τιμές. Σε αυτό το σημείο ο αναλυτής καλείται να επιλέξει τον καλύτερο και πιο χρήσιμο συνδυασμό μεταβλητών. Αυτό συμβαίνει συνήθως με την βοήθεια του κριτηρίου BIC. Παρακάτω ακολουθεί η ομαδοποίηση των δεδομένων.

## 5.3 Ταξινόμηση

### 5.3.1 ΣΤ' Δημοτικού

Εισάγουμε τις εντολές στο πρόγραμμα όπως έχουμε αναφέρει στην ενότητα 4.2. Όπως έχουμε αναφέρει σε άλλο σημείο, αρχικά, δοκιμάζουμε διάφορες περιπτώσεις ομαδοποίησης ενώ με τον καιρό αυτό το στάδιο μπορεί να γίνεται γρήγορα επιλέγοντας περισσότερα εμπειρικά παρά δοκιμάζοντας όλες τις πιθανές περιπτώσεις. Η καλύτερη περίπτωση ταξινόμησης όπως έχει ξαναφερθεί επιλέγεται με τη βοήθεια του κριτηρίου BIC. Η περίπτωση με την χαμηλότερη τιμή σε αυτό το κριτήριο μπορεί να κριθεί ως η πιο κατάλληλη παρατηρώντας και την πορεία του. Αν από ένα αριθμό επαναλήψεων και μετά είναι συνεχώς μεγαλύτερο τότε δεν χρειάζεται να εκτελεστούν τόσες επαναλήψεις στην αρχή. Τα αποτελέσματα δίνονται παρακάτω σε μορφή πινάκων, ενώ με τη μορφή που εξάγονται από το πρόγραμμα R Studio δίνονται στο παράρτημα και μόνο για τις επιλεγθείσες περιπτώσεις ταξινόμησης κάθε ανάλυσης.

Το όνομα του αρχείου αυτής της ηλικίας είναι το 'Grade 6'. Οι εντολές εισάγονται στο πάνω αριστερό παράθυρο του R Studio όπως ακολούθως:

```
setwd('C:/Users/aggeliki/Desktop/ΔΕΔΟΜΕΝΑ')
grade6<-read.csv("Grade 6.csv", sep=";", header=TRUE)
library(poLCA)
attach(as.data.frame(grade6))
fr1<-cbind(GR19_10, GR19_11, GR19_12, GR19_13,
           GR19_14, GR19_15, GR19_16, GR19_17,
           GR19_18, GR19_19, GR19_20,
           GR19_01, GR19_02, GR19_03, GR19_04,
           GR19_05, GR19_06, GR19_07, GR19_08,
           GR19_09)~1)

lca1<-poLCA(fr1,as.data.frame(grade6),nclass=1,maxiter=2000,nrep=10)
lca2<-poLCA(fr1,as.data.frame(grade6),nclass=2,maxiter=2000,nrep=10,
graphs=TRUE)
lca3<-poLCA(fr1,as.data.frame(grade6),nclass=3,maxiter=2000,nrep=10,
graphs=TRUE)
lca4<-poLCA(fr1,as.data.frame(grade6),nclass=4,maxiter=2000,nrep=10,
graphs=TRUE)
lca5<-poLCA(fr1,as.data.frame(grade6),nclass=5,maxiter=2000,nrep=10,
graphs=TRUE)
lca6<-poLCA(fr1,as.data.frame(grade6),nclass=6,maxiter=2000,nrep=10,
graphs=TRUE)
lca7<-poLCA(fr1,as.data.frame(grade6),nclass=7,maxiter=2000,nrep=10,
graphs=TRUE)
lca8<-poLCA(fr1,as.data.frame(grade6),nclass=8,maxiter=2000,nrep=10,
graphs=TRUE)
lca9<-poLCA(fr1,as.data.frame(grade6),nclass=9,maxiter=2000,nrep=10,
graphs=TRUE)

bics<-cbind(lca1$bic, lca2$bic, lca3$bic, lca4$bic, lca5$bic, lca6$bic, lca7$bic,
lca8$bic, lca9$bic)

bics
```



Εφόσον επιλέξουμε run και εμφανιστούν τα αποτελέσματα ξεκινά η επεξεργασία από τον αναλυτή ώστε να οδηγηθεί στην εξαγωγή συμπερασμάτων για τις πληροφορίες αυτές. Ακολουθούν οι πίνακες με τα αποτελέσματα.

Το πρόγραμμα εμφανίζει τους αριθμούς του κριτηρίου BIC και για τις εννιά περιπτώσεις ομαδοποίησης όπως παρακάτω:

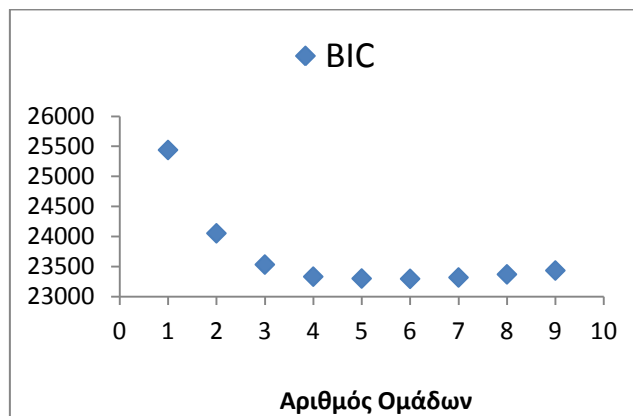
[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]  
 [1,] 25438.48 24052.09 23532.02 23330.87 23299.78 23294.72 23317.33 23369.02  
 [,9]  
 [1,] 23433.22

Στον Πίνακα 1 δίνονται οι αριθμοί του κριτηρίου BIC μαζί με το μέγεθος κάθε ομάδας κάθε ταξινόμησης.

Αριθμός Ομάδων		Μέγεθος ομάδων								
	BIC	1	2	3	4	5	6	7	8	9
<b>1</b>	25438									
<b>2</b>	24052	0.6114	0.3886							
<b>3</b>	23532	0.1606	0.3345	0.5049						
<b>4</b>	23331	0.1063	0.2975	0.2972	0.299					
<b>5</b>	23300	0.0146	0.2905	0.3018	0.1271	0.266				
<b>6</b>	<b>23295</b>	<b>0.3261</b>	<b>0.1158</b>	<b>0.2131</b>	<b>0.2252</b>	<b>0.019</b>	<b>0.1009</b>			
<b>7</b>	23317	0.342	0.0317	0.078	0.1071	0.0278	0.2007	0.2127		
<b>8</b>	23369	0.0548	0.0199	0.3421	0.2095	0.0335	0.0296	0.1999	0.1107	
<b>9</b>	23433	0.0317	0.2442	0.2098	0.0671	0.0347	0.0982	0.1194	0.0201	0.1748

Πίνακας 1 Στ' Δημοτικού\_Κριτήριο BIC όλων των ομαδοποιήσεων

Στο Γράφημα 1 παρατηρείται η πορεία του κριτηρίου BIC. Στον άξονα των Χ τοποθετείται ο αριθμός των κλάσεων από την περίπτωση μίας ομάδας μέχρι την περίπτωση εννιά ομάδων και στον άξονα των Υ είναι οι τιμές του κριτηρίου που παίρνει σε κάθε περίπτωση ομαδοποίησης.



Γράφημα 1 Στ' Δημοτικού\_Κριτήριο BIC

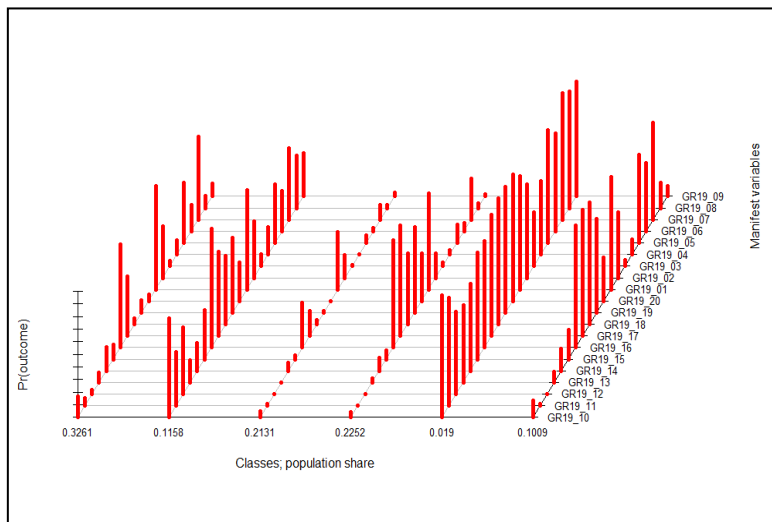
Όπως παρατηρούμε, από ένα σημείο και μετά η τιμή του κριτηρίου παραμένει στα ίδια επίπεδα με οριακές βελτιώσεις. Σύμφωνα με το κριτήριο BIC τη χαμηλότερη τιμή παίρνει στην περίπτωση των έξι κλάσεων, με τιμή 23295, πολύ κοντά βρίσκεται βέβαια και η περίπτωση των πέντε κλάσεων. Θα μπορούσε κάποιος να επιλέξει και αυτή των πέντε, στη συγκεκριμένη περίπτωση επιλέγουμε έξι κλάσεις για να εντοπίσουμε το πλήθος των υποομάδων μεταξύ των παιδιών της Στ' Δημοτικού σχετικά με τις συνήθειες.

Ακολουθεί ο Πίνακας 2 με τις πιθανότητες να απαντά με συγκεκριμένο τρόπο σε όλες τις δραστηριότητες ο πληθυσμός κάθε ομάδας.

	Πιθανότητες ανά ομάδα					
Δραστηριότητες	1	2	3	4	5	6
1 – Αθλητικά	0.2121	0.5433	0.2396	0.1051	0.1111	0.1813
2 – Χορός	0.5535	0.8220	0.8106	0.2607	0.4830	0.5940
3 – Προσκοπισμός	0.9031	0.9898	0.9719	0.5719	0.9520	0.9582
4 – Συγγραφή	0.7834	0.9973	0.9664	0.4216	0.8841	0.8874
5 – Καλλιτεχνικά	0.5364	0.9375	0.8535	0.1063	0.3036	0.5252
6 – Φωτογραφία	0.6832	0.9723	0.9353	0.2247	0.4565	0.7908
7 – Διάβασμα βιβλίων	0.4318	0.8814	0.6751	0.0000	0.2351	0.3465
8 – Σύνθεση Μουσικής	0.5858	0.9772	0.9661	0.0772	0.7976	0.9078
9 – Χορωδία	0.6574	0.9720	0.9820	0.0893	0.9211	0.9024
10 – Μουσικά Όργανα	0.2171	0.9500	0.9597	0.0362	0.8687	0.8352
11 – Συγκρότημα	0.5759	0.9859	1.0000	0.1456	0.9885	0.9406
12 – Ιδιαίτερα μαθήματα μουσικής	0.4728	1.0000	1.0000	0.3506	1.0000	0.9686
13 – Θέατρο	0.8226	1.0000	0.9686	0.3891	0.9188	0.9232
14 – Δραστηριότητες της Εκκλησίας	0.7819	0.9333	0.8975	0.3091	0.8251	0.8143
15 – Συναντήσεις, θέατρο, μουσεία	0.6144	0.9701	0.9321	0.1579	0.7686	0.8819
16 – Ακούω μουσική	0.0618	0.6445	0.1529	0.1535	0.0294	0.1823
17 – Ηλεκτρονικά παιχνίδια Η/Υ	0.3320	0.8010	0.1227	0.0384	0.0000	0.5291
18 – Επεξεργασία εικόνων Η/Υ	0.4591	0.9677	0.4386	0.0000	0.0343	0.9501
19 – “Κατέβασμα μουσικής online”	0.4090	0.9842	0.3229	0.0000	0.2599	0.9054
20 – Δημιουργία προγραμμάτων/παιχνιδιών στον Η/Υ	0.6948	1.0000	0.6239	0.0000	0.6530	0.9473
<b>ΣΥΝΟΛΟ ΠΛΗΘΥΣΜΟΥ</b>	<b>0.1158</b>	<b>0.2131</b>	<b>0.2252</b>	<b>0.019</b>	<b>0.1009</b>	<b>0.3261</b>

**Πίνακας 2 Στ' Δημοτικού\_ Πιθανότητες των δραστηριοτήτων**

Ενώ, ακολούθως στο Γράφημα 2 που εξάγει το R Studio στο δεξί κάτω παράθυρο απεικονίζεται ο τρόπος απαντήσεων των παιδιών σε κάθε ομάδα. Σε αυτό το σημείο να αναφέρουμε ένα τεχνικό ζήτημα ότι η σειρά των έξι ομάδων με τους πληθυσμούς όπως εμφανίζονται στο γράφημα δεν είναι ίδια με αυτή στο Console window όπου παρουσιάζονται αναλυτικά οι πιθανότητες, όπως και στον παραπάνω πίνακα.



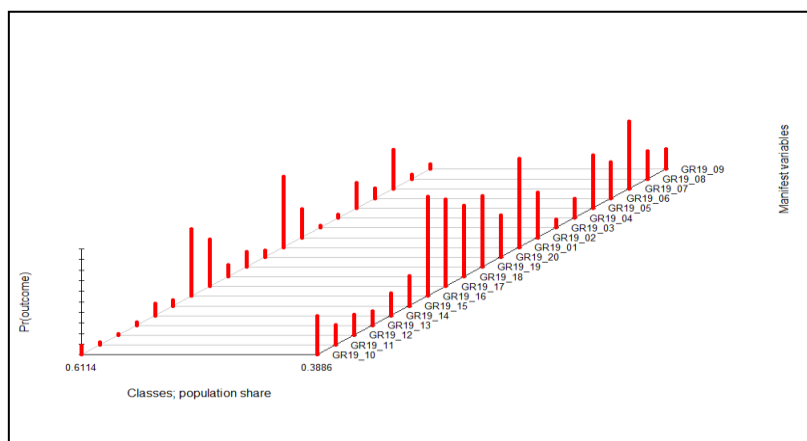
**Γράφημα 2 Στ' Δημοτικού\_Ομαδοποίηση δραστηριοτήτων**

Οι κόκκινοι ράβδοι συμβολίζουν το πλήθος των μαθητών που τείνουν να απαντούν θετικά στις δραστηριότητες ενώ, ο αριθμός που φαίνεται στη βάση του γραφήματος είναι το ποσοστό του πληθυσμού κάθε ομάδας. Δεξιά φαίνεται ο τίτλος της δραστηριότητας υπενθυμίζοντας ότι στην συγκεκριμένη ανάλυση το πρόγραμμα παρουσιάζει πρώτα τις έντεκα τελευταίες δραστηριότητες της ερώτησης και μετά τις επόμενες εννιά όπως παρατηρούμε το γράφημα από κάτω προς τα πάνω.

Αυτό που συναιεί κατά την διάρκεια της ανάλυσης με την Latent Class Analysis είναι ότι πάρθηκαν οι απαντήσεις κάθε ενός παιδιού από τα 1339 και μελετήθηκαν όπως παρουσιάζεται παρακάτω. Εξετάστηκε για κάθε περίπτωση παιδιού πως απάντησε σε κάθε μια δραστηριότητα. Κάθε περίπτωση παιδιού καταχωρήθηκε σε μια ομάδα ανάλογα με τον τρόπο που απάντησε. Η κάθε ομάδα περιλάμβανε τα παιδιά που απάντησαν με παρόμοιο τρόπο σε όλες τις δραστηριότητες. Επαναλαμβάνοντας την διαδικασία αυτή και για τα 1339 παιδιά, δημιουργήθηκαν έξι ομάδες, όπου η κάθε μία είχε ένα ξεχωριστό μοτίβο απαντήσεων (με 1 και 2). Δηλαδή στην πρώτη ομάδα καταχωρήθηκαν τα παιδιά που απαντούν θετικά (τιμή 2) περισσότερο στις δραστηριότητες 7,1 και 16 και κάποια από αυτά και στις 5,2 και 17. Στην προτελευταία ομάδα καταχωρήθηκαν τα παιδιά που απαντούν θετικά (επιλέγουν την τιμή 2) σχεδόν σε όλες τις δραστηριότητες. Έπειτα, υπολογίζεται το ποσοστό του πληθυσμού που ανήκει σε κάθε μια ομάδα βάσει τους αριθμούς των παιδιών που ανήκουν σε κάθε μία από τις ομάδες. Τέλος, υπολογίζεται η πιθανότητα να απαντά θετικά (2) σε κάθε δραστηριότητα ο πληθυσμός μιας ομάδας αλλά και η αντίστοιχη να απαντά αρνητικά (1). Η λανθάνουσα μεταβλητή που υποθέσουμε σε αυτή την ανάλυση είναι ότι τα παιδιά χωρίζονται σε έξι ομάδες, η οποία δεν υπήρχε πριν ή τουλάχιστον δεν φαινόταν. Οι κλάσεις της λανθάνουσας μεταβλητής που με δική μας επεξεργασία εμφανίστηκαν είναι οι έξι ενώ η τιμή της λανθάνουσας μεταβλητής για κάθε τάξη είναι ο συνδυασμός των 1 και 2. Στο παραπάνω γράφημα φαίνονται με κόκκινους ράβδους οι θετικές απαντήσεις (2) των παιδιών ενώ το ύψος τους δείχνει κατά πόσο ο πληθυσμός της ομάδας απαντά θετικά σε κάθε δραστηριότητα. Δηλαδή στην πρώτη κλάση από τα αριστερά στα δεξιά όπως παρατηρούμε το γράφημα βλέπουμε ότι ο πληθυσμός της ομάδας είναι 33% του συνόλου. Όπως παρατηρούμε η δραστηριότητα 16 έχει υψηλή κόκκινη ράβδο κάτι που μας φανερώνει ότι το μεγάλο ποσοστό παιδιών από το 33% ασχολείται με την δραστηριότητα 16 'Ακούω την αγαπημένη μου μουσική'.

Ακολουθώντας, θα θέλαμε να παραθέσουμε άλλη μία περίπτωση ομαδοποίησης ανάλυση της Στ' Τάξης του Δημοτικού. Ο λόγος είναι να διευκρινιστούν οι διαφορές μεταξύ των ομαδοποιήσεων καθώς επιλέγεται μία από αυτές τις περιπτώσεις. Η δεύτερη ομαδοποίηση που πρόκειται να αναλυθεί είναι η ταξινόμηση σε δύο ομάδες. Το

κριτήριο BIC αυτής της περίπτωσης είναι 24052, αρκετά μεγαλύτερο από το 23295 των έξι κλάσεων. Ακολουθεί το Γράφημα 3:



Γράφημα 3 Στ' Δημοτικού\_Δεύτερος τρόπος ομαδοποίησης δραστηριοτήτων

Όπως παρατηρούμε συγχωνεύονται κατά πολύ οι ομάδες και δεν διαφαίνονται οι υποομάδες των παιδιών με τις συνήθειές τους. Επίσης, δεν φαίνεται ποιο ποσοστό των παιδιών κάνει τις περισσότερες δραστηριότητες και ποιο δεν ασχολείται με δραστηριότητες στον ελεύθερο χρόνο. Δεν θα λέγαμε, λοιπόν, ότι αυτή η ταξινόμηση δίνει επαρκείς πληροφορίες για τα δεδομένα.

### 5.3.2 Β' Γυμνασίου

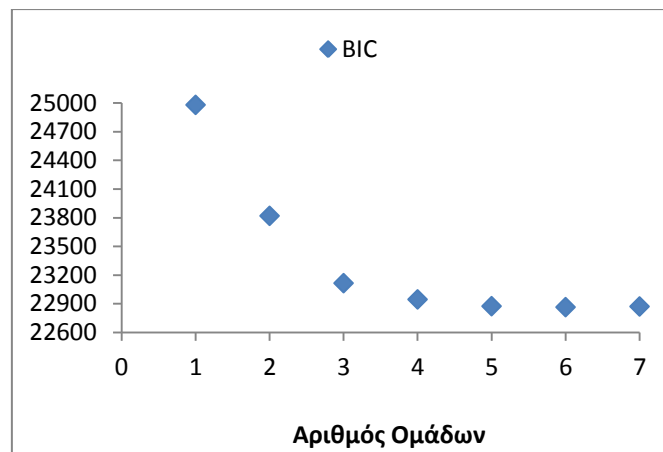
Η διαδικασία της ανάλυσης είναι η ίδια με τη προηγούμενη, αλλά αντί για Grade 6 πληκτρολογούμε το Grade 8. Ο πληθυσμός των παιδιών της Β' Γυμνασίου είναι 1370. Στον Πίνακα 3 καταγράφονται οι αριθμοί του κριτηρίου BIC για επτά περιπτώσεις ομαδοποίησης και δεξιά είναι το ποσοστό του πληθυσμού που βρίσκεται σε κάθε κλάση, όπως και στις προηγούμενες αναλύσεις.

Αριθμός Ομάδων	BIC	Μέγεθος ομάδων						
		1	2	3	4	5	6	7
1	24980							
2	23819	0.2749	0.7251					
3	23115	0.1989	0.3782	0.4229				
4	22945	0.0894	0.4333	0.3702	0.1071			
5	<b>22874</b>	<b>0.0456</b>	<b>0.2966</b>	<b>0.1184</b>	<b>0.3556</b>	<b>0.1838</b>		
6	<b>22864</b>	<b>0.3415</b>	<b>0.097</b>	<b>0.1462</b>	<b>0.2579</b>	<b>0.1107</b>	<b>0.0467</b>	
7	22871	0.0871	0.0377	0.1991	0.1515	0.1147	0.1106	0.2992

Πίνακας 3 Β' Γυμνασίου\_Κριτήριο BIC

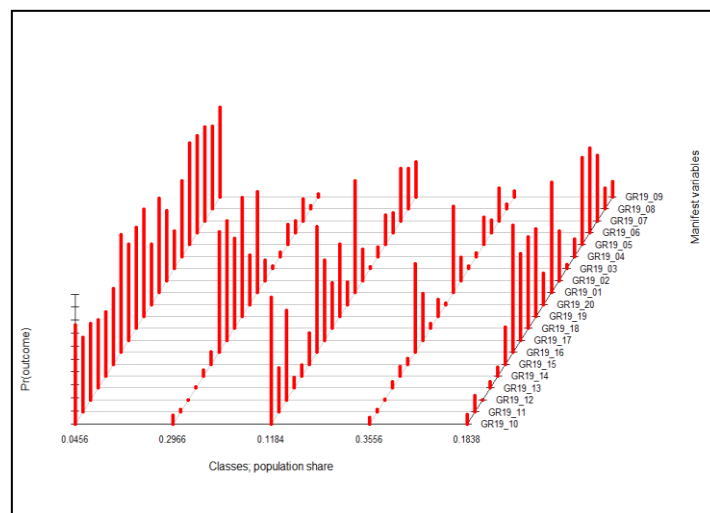
Όπως παρατηρούμε στο Γράφημα 4, από τέσσερις ομάδες και πάνω η μεταβολή του κριτηρίου είναι ασήμαντη με μικρότερη τιμή ομαδοποίησης αυτής των έξι ομάδων.

Παρατηρούμε ότι οι διαφορές μεταξύ των πέντε και των έξι αριθμό ομάδων είναι ασήμαντες, επομένως επιλέγουμε την πέμπτη επιλογή εφόσον οι διαφορές δεν είναι σημαντικές.



Γράφημα 4 Β' Γυμνασίου\_Κριτήριο BIC

Στο Γράφημα 5 παρουσιάζεται εικονικά η ταξινόμηση των ομάδων ανάλογα με τις πιθανότητες των μεταβλητών.



Γράφημα 5 Β' Γυμνασίου\_Ομαδοποίηση δραστηριοτήτων

Στον Πίνακα 4 οι πιθανότητες για κάθε μια από τις πέντε ομάδες τα παιδιά να ασχολούνται με τις δραστηριότητες.

	Πιθανότητες ανά ομάδα				
Ερώτηση	1	2	3	4	5
1 – Αθλητικά	0.2758	0.2243	0.1428	0.3390	0.1522
2 – Χορός	0.4633	0.8455	0.7592	0.8208	0.6173
3 – Προσκοπισμός	0.7146	0.9807	0.9866	0.9807	0.9692
4 – Συγγραφή	0.4178	0.9667	0.9275	0.9706	0.8687
5 – Καλλιτεχνικά	0.2154	0.8435	0.7744	0.7904	0.3302
6 – Φωτογραφία	0.2503	0.9104	0.8489	0.9021	0.3465
7 – Διάβασμα βιβλίων	0.2805	0.8358	0.5998	0.7508	0.4956
8 – Σύνθεση Μουσικής	0.3637	0.9747	0.6904	0.9679	0.8446
9 – Χορωδία	0.3133	0.9808	0.7322	0.9569	0.8853
10 – Μουσικά Όργανα	0.2384	0.9359	0.0252	0.9489	0.9272
11 – Συγκρότημα	0.4259	0.9822	0.6593	0.9793	0.8746
12 – Ιδιαίτερα μαθήματα μουσικής	0.4101	1.0000	0.3122	0.9926	1.0000
13 – Θέατρο	0.4730	0.9908	0.9207	0.9510	0.9498
14 – Δραστηριότητες της Εκκλησίας	0.5083	0.9554	0.9138	0.9270	0.9338
15 – Συναυλίες, θέατρο, μουσεία	0.4194	0.9055	0.7627	0.9588	0.7148
16 – Ακούω μουσική	0.0920	0.0733	0.0337	0.3187	0.0229
17 – Ηλεκτρονικά παιχνίδια Η/Υ	0.2623	0.0790	0.3843	0.6423	0.3306
18 – Επεξεργασία εικόνων Η/Υ	0.2212	0.3060	0.6483	0.9634	0.2952
19 – “Κατέβασμα μουσικής online”	0.1742	0.0839	0.4453	0.8689	0.3276
20 – Δημιουργία προγραμμάτων/παιχνιδιών στον Η/Υ	0.5364	0.6234	0.8299	0.9894	0.7607
<b>ΣΥΝΟΛΟ ΠΛΗΘΥΣΜΟΥ</b>	<b>0.0456</b>	<b>0.2966</b>	<b>0.1184</b>	<b>0.3556</b>	<b>0.1838</b>

Πίνακας 4 Β' Γυμνασίου\_Πιθανότητες των δραστηριοτήτων

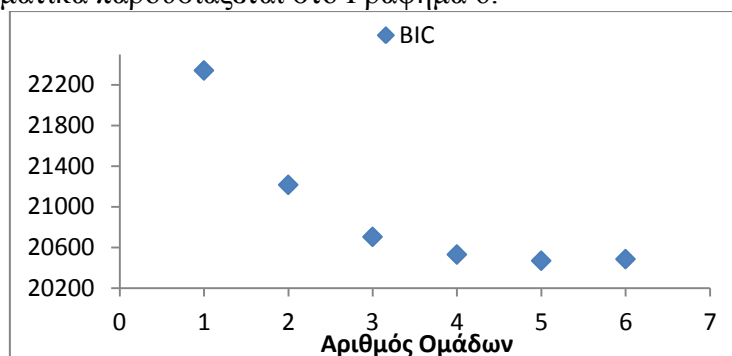
### 5.3.3 Α' Λυκείου

Περνώντας τώρα στα παιδιά της Α' Λυκείου το σύνολο των παιδιών που απάντησαν σωστά στην ερώτηση αυτή της έρευνας είναι 1348. Αφού υπολογίστηκαν οι πιθανότητες κάθε μεταβλητής να παίρνει την τιμή 2 έπειτα έγινε η ομαδοποίηση των πιθανοτήτων σε έξι παραλλαγές, για να επιλέξουμε την καταλληλότερη παρατηρούμε στον Πίνακα την τιμή του κριτηρίου BIC.

Αριθμός Ομάδων	BIC	Μέγεθος ομάδων					
		1	2	3	4	5	6
1	22341						
2	21215	0.8158	0.1842				
3	20703	0.4171	0.1709	0.412			
4	20529	0.4158	0.063	0.1088	0.4124		
5	<b>20468</b>	<b>0.0437</b>	<b>0.3597</b>	<b>0.2925</b>	<b>0.1088</b>	<b>0.1953</b>	
6	20484	0.0425	0.056	0.0643	0.2835	0.1956	0.358

Πίνακας 5 Α' Λυκείου\_Κριτήριο BIC

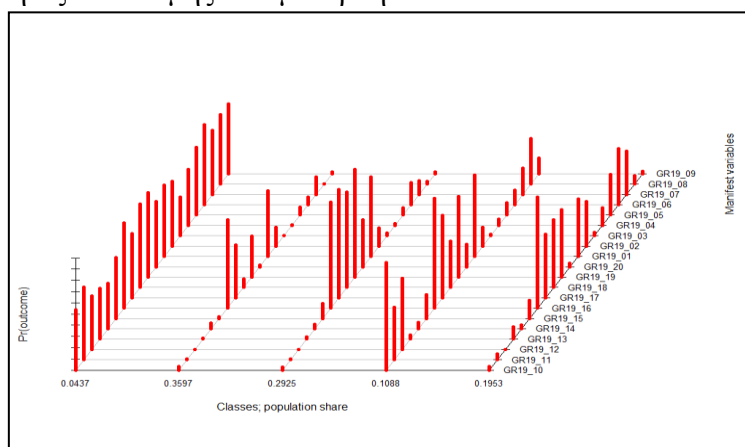
Ενώ, διαγραμματικά παρουσιάζεται στο Γράφημα 6.



Γράφημα 6 Α' Λυκείου\_Κριτήριο BIC

Παρατηρούμε ότι από τις τέσσερις τάξεις και μετά το BIC δεν παρουσιάζει σημαντικές μεταβολές αλλά κρατά μια πιο σταθερή πορεία με χαμηλότερη τιμή στον αριθμό πέντε. Επομένως, είναι ένας ικανός αριθμός ομάδων, οι πέντε τάξεις.

Στο Γράφημα 7 παρατίθεται η ταξινόμηση του πληθυσμού σε πέντε κλάσεις ανάλογα με τις πιθανότητες κατανομής των μεταβλητών.



Γράφημα 7 Α' Λυκείου\_Ομαδοποίηση δραστηριοτήτων



Στον Πίνακα 6, παρουσιάζονται οι πιθανότητες κατανομής κάθε δραστηριότητας.

Ερώτηση	Πιθανότητες ανά ομάδα				
	1	2	3	4	5
1 – Αθλητικά	0.3654	0.4127	0.2938	0.2724	0.4883
2 – Χορός	0.4211	0.8312	0.8374	0.8313	0.6041
3 – Προσκοπισμός	0.6579	1.0000	0.9741	0.9863	0.9710
4 – Συγγραφή	0.5011	0.9961	0.9833	0.9448	0.8441
5 – Καλλιτεχνικά	0.3954	0.9308	0.9342	0.8903	0.6399
6 – Φωτογραφία	0.2891	0.9350	0.8036	0.8720	0.5009
7 – Διάβασμα βιβλίων	0.4291	0.8459	0.8812	0.7677	0.6134
8 – Σύνθεση Μουσικής	0.3825	0.9971	0.9747	0.5926	0.9261
9 – Χορωδία	0.3773	0.9861	0.9869	0.8593	0.9782
10 – Μουσικά Όργανα	0.4620	0.9653	0.9720	0.0417	0.9701
11 – Συγκρότημα	0.3499	0.9909	0.9915	0.5311	0.9462
12 – Ιδιαίτερα μαθήματα μουσικής	0.5247	0.9981	1.0000	0.3656	1.0000
13 – Θέατρο	0.5485	0.9908	0.9821	0.9634	0.8917
14 – Δραστηριότητες της Εκκλησίας	0.5956	0.9493	0.9587	0.9430	0.9660
15 – Συναυλίες, θέατρο, μουσεία	0.4582	0.9865	0.8667	0.7887	0.8367
16 – Ακούω μουσική	0.2394	0.2152	0.0532	0.0203	0.0099
17 – Ηλεκτρονικά παιχνίδια Η/Υ	0.4294	0.5319	0.0330	0.2663	0.4359
18 – Επεξεργασία εικόνων Η/Υ	0.2566	0.9229	0.1514	0.5899	0.3995
19 – “Κατέβασμα μουσικής online”	0.2482	0.6376	0.0427	0.2838	0.3981
20 – Δημιουργία προγραμμάτων/παιχνιδιών στον Η/Υ	0.4212	0.9871	0.5839	0.8021	0.9676
<b>ΣΥΝΟΛΟ ΠΛΗΘΥΣΜΟΥ</b>	<b>0.0437</b>	<b>0.3597</b>	<b>0.2925</b>	<b>0.1088</b>	<b>0.1953</b>

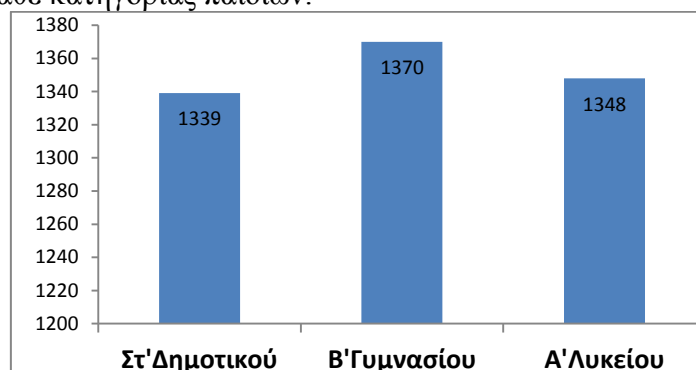
Πίνακας 6 Α' Λυκείου\_Πιθανότητες των δραστηριοτήτων

Τα συμπεράσματα από τα αποτελέσματα και των τριών αναλύσεων θα παρουσιαστούν αναλυτικά και συγκριτικά στο αμέσως επόμενο κεφάλαιο.

### 5.3.4 Παρουσίαση συμπερασμάτων

#### 5.3.4.1 Περιγραφική στατιστική δεδομένων

Σε αυτό το σημείο πρόκειται να περιγράψουμε τα αποτελέσματα της ερώτησης 79 της έρευνας με την βοήθεια της Περιγραφικής Στατιστικής. Στο Γράφημα 8 απεικονίζεται ο πληθυσμός κάθε κατηγορίας παιδιών.



Γράφημα 8 Πληθυσμός κάθε ηλικίας

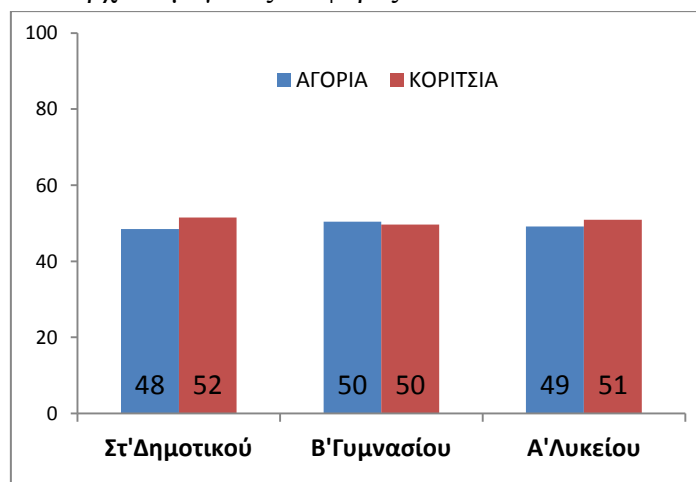
Όπως παρατηρούμε ο πληθυσμός των παιδιών της Β' τάξης Γυμνασίου είναι λίγο μεγαλύτερος χωρίς να έχει και μεγάλη απόκλιση από τις άλλες δύο τάξεις. Σε γενικές γραμμές ο πληθυσμός και των τριών ηλικιών είναι στα ίδια επίπεδα. Τέλος, ο αριθμός των απαντήσεων των τριών ηλικιών είναι ικανοποιητικός ώστε τα συμπεράσματα της ανάλυσης να είναι έγκυρα.

Ενώ, στον Πίνακα 7 μπορούμε να παρατηρήσουμε τον αριθμό των αγοριών και των κοριτσιών κάθε ηλικίας:

	ΑΓΟΡΙΑ	ΚΟΡΙΤΣΙΑ	ΣΥΝΟΛΟ
Στ' Δημοτικού	649	690	1339
Β' Γυμνασίου	690	680	1370
Α' Λυκείου	662	686	1348

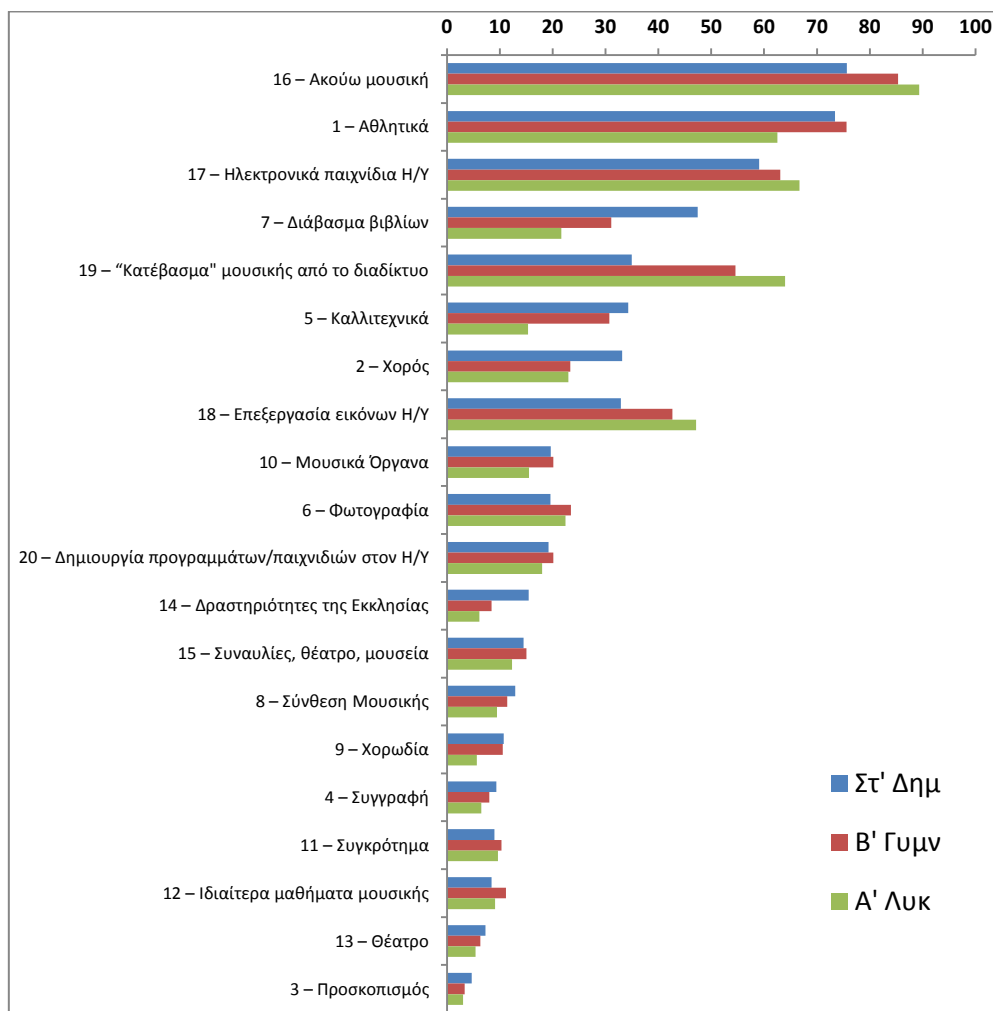
Πίνακας 7 Ο πληθυσμός κάθε ηλικίας ανά φύλο

Όπως παρατηρούμε τα κορίτσια δύο ηλικιών υπερисχύουν σε σχέση με τα αγόρια ενώ είναι λιγότερα τα κορίτσια της Β' Γυμνασίου από τα αγόρια. Στο Γράφημα 9 απεικονίζεται το ποσοστό των παιδιών ανά φύλο και των τριών ηλικιών. Οι μπάρες με μπλε χρώμα είναι το πλήθος των αγοριών ενώ με κόκκινο είναι τα κορίτσια. Όπως παρατηρούμε δεν υπάρχουν μεγάλες διαφορές.



Γράφημα 9 Ποσοστό παιδιών ανά φύλο των τριών ηλικιών

Παρακάτω στο Γράφημα 10 συνεχίζουμε περιγράφοντας τις συνήθειες των παιδιών κάθε τάξης. Στο διάγραμμα αυτό απεικονίζονται οι συχνότητες των παιδιών όπου στον κάθετο άξονα (Υ) είναι οι είκοσι δραστηριότητες και τα τρία χρώματα δείχνουν τις τρεις σχολικές τάξεις. Με μπλε απεικονίζονται τα παιδιά που κάνουν τις δραστηριότητες και φοιτούν στην Έκτη Δημοτικού, ενώ με κόκκινο τα παιδιά της Δευτέρας Γυμνασίου και τέλος πράσινο είναι το χρώμα των παιδιών της Πρώτης Λυκείου.

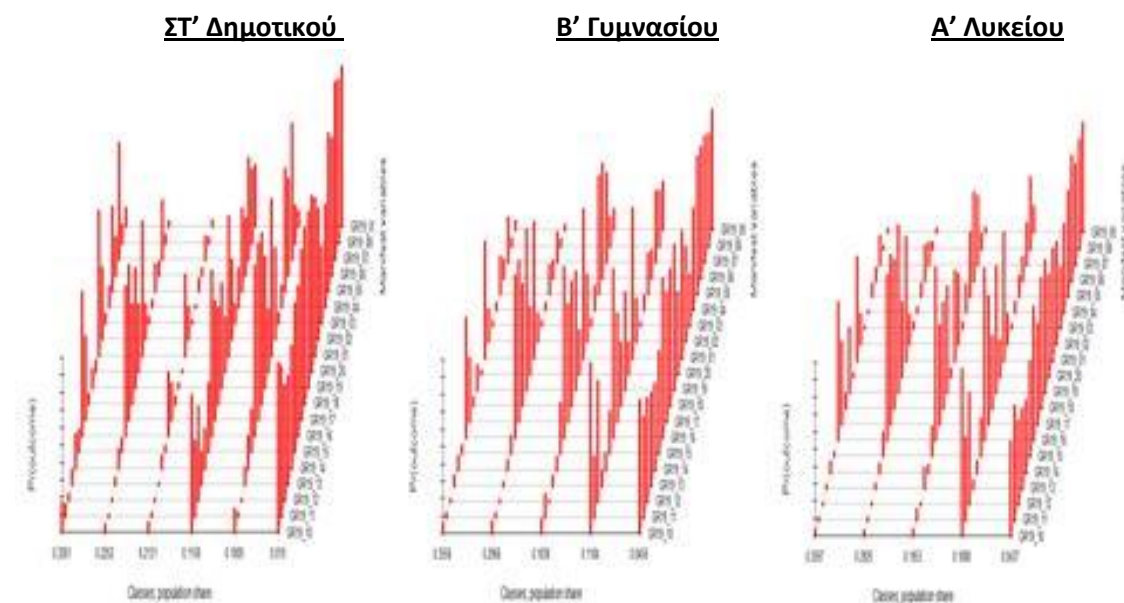


Γράφημα 10 Ποσοστό πληθυσμού που ασχολείται με κάθε δραστηριότητα

Όπως προσέχουμε οι συνήθειες των παιδιών είναι παρόμοιες και στις τρεις ηλικίες με ορισμένες διαφορές. Η διαφορά έγκειται στην χρήση του ηλεκτρονικού υπολογιστή. Τα παιδιά του Δημοτικού τον χρησιμοποιούν κυρίως για ηλεκτρονικά παιχνίδια και λιγότερο για ‘κατέβασμα’ μουσικής, και σχεδιασμό εικόνων όπως συνηθίζουν τα παιδιά των άλλων δύο ηλικιών. Γενικά, τα παιδιά συνήθως ακούνε μουσική σύμφωνα με τις απαντήσεις τους, τα περισσότερα ασχολούνται με τον αθλητισμό και με τον υπολογιστή, με τα παιδιά της Α’ Λυκείου να τείνουν πολύ περισσότερο προς αυτές τις δραστηριότητες. Τα παιδιά της Στ’ Δημοτικού φαίνεται να διαβάζουν βιβλία στον ελεύθερο χρόνο τους σε αντίθεση με τα παιδιά της Α’ Λυκείου που σχεδιάζουν εικόνες. Όπως προσέχουμε ο προσκοπισμός δεν συνηθίζεται στα παιδιά αυτών των ηλικιών αλλά και οι καλλιτεχνικές δραστηριότητες που θέλουν περισσότερο ενεργή συμμετοχή από τα παιδιά βρίσκονται σε χαμηλά επίπεδα. Ύστερα από την ανάλυση και την ταξινόμηση που πραγματοποιήσαμε στα προηγούμενα κεφάλαια, συνεχίζουμε με την μελέτη των συμπερασμάτων της ανάλυσης αυτής.

### 5.3.4.2 Συμπεράσματα

Στο Γράφημα 11 απεικονίζονται τα διαγράμματα των ομαδοποιήσεων για τις τρεις ηλικίες αλλά ταξινομημένα, δηλαδή από την μεγαλύτερη σε πληθυσμό στην μικρότερη όπως κοιτάμε κάθε γράφημα από τα αριστερά στα δεξιά. Ενώ, παρακάτω θα ακολουθήσουν τα συμπεράσματα για τις ομάδες παιδιών που σχηματίζονται ανάλογα με το σύνολο των δραστηριοτήτων που ασχολούνται.



Γράφημα 11 Οι τρεις ομαδοποιήσεις

### Συμπεράσματα από την ταξινόμηση των παιδιών της Στ' Δημοτικού σε ομάδες

- \* Υπάρχει μια μικρή ομάδα μαθητών ( $P=0.019$ ) που τείνουν να απαντούν θετικά στις **περισσότερες δραστηριότητες**, δηλαδή ασχολούνται με αυτές στον ελεύθερο χρόνο τους.
- \* Το μεγαλύτερο ποσοστό των παιδιών της Στ' Δημοτικού ( $P=0.3261$  ή 33%) ασχολείται κυρίως με τον **αθλητισμό, ακούει μουσική και διαβάζει βιβλία** (1), (16), (7) αντίστοιχα.
- \* Ένα λίγο μικρότερο ποσοστό παιδιών αυτής της ηλικίας ( $P=0.2252$  ή 23%) ασχολείται επίσης με **τα αθλητικά, ακούει μουσική** όπως η προηγούμενη ομάδα αλλά παίζει **ηλεκτρονικά παιχνίδια** στον υπολογιστή.
- \* Σε μια ομάδα παρόμοιου μεγέθους ( $P=0.2131$ ) φαίνεται ότι ασχολούνται με τις δύο κύριες δραστηριότητες που συνηθίζουν και οι δύο προηγούμενες ομάδες, **τον αθλητισμό και την ακρόαση μουσικής**.
- \* Μία μικρότερη ομάδα ( $P=0.1158$  ή 11%) φαίνεται ότι ασχολείται με αρκετές δραστηριότητες αλλά σε μικρότερο ποσοστό από τη πρώτη που αναφέραμε παραπάνω. Ιδιαίτερα ασχολείται με **τον αθλητισμό, την ακρόαση μουσικής και την εκμάθηση μουσικού οργάνου**.
- \* Τέλος, ένα μικρό σχετικά ποσοστό παιδιών ( $P=0.1009$ ) ασχολείται εκτός από τις δραστηριότητες **αθλητισμό, μελέτη βιβλίων, ακρόαση μουσικής** και με τον **ηλεκτρονικό υπολογιστή** δηλαδή παίζει ηλεκτρονικά παιχνίδια και σχεδιάζει

εικόνες. Συνολικά, αξιoσημείωτο είναι ότι οι δραστηριότητες που συνήθως ασχολούνται στην ηλικία της Στ' τάξης του Δημοτικού τα παιδιά είναι **ο αθλητισμός και η ακρόαση μουσικής**.

Καθώς για λόγους διδακτικούς αναλύσαμε και την επιλογή της ταξινόμησης του πληθυσμού σε δύο κλάσεις θα αναφέρουμε παρακάτω τα συμπεράσματα της ανάλυσης αυτής.

\* Η μεγαλύτερη ομάδα με ποσοστό πάνω από το μισό των παιδιών ( $P=0.6114$  ή 61%) αναφέρει ότι ασχολείται με τον αθλητισμό, την ανάγνωση βιβλίων και την ακρόαση μουσικής.

\* Ενώ, η δεύτερη ομάδα με μικρότερο αριθμό παιδιών αναφέρει ότι ασχολείται με τις παραπάνω δραστηριότητες και με τον ηλεκτρονικό υπολογιστή (17), (18), (19). Αξιoσημείωτο είναι ότι με την ομαδοποίηση σε λιγότερες ομάδες δεν διακρίνονται τα δύο μεγάλα ποσοστά πληθυσμού που ενώ συνηθίζουν την ακρόαση μουσικής και τον αθλητισμό, η πρώτη ασχολείται επιπλέον με την ανάγνωση βιβλίων ενώ η δεύτερη με τα παιχνίδια του ηλεκτρονικού υπολογιστή. Επομένως με την ταξινόμηση σε περισσότερες κλάσεις τα συμπεράσματα είναι πιο λεπτομερείς και αναλυτικά εγγυώντας έτσι την εγκυρότητα και την χρησιμότητα της ανάλυσης. Με την ομαδοποίηση σε δύο κλάσεις εξάγονται τα βασικά συμπεράσματα που παρατηρούνται πιο εύκολα αλλά συμπυκνώνονται οι ειδικές περιπτώσεις συνδυασμών των μεταβλητών. Έτσι, η ανάλυση είναι ελλιπής μη αναλύοντας αρκετά όλους τους συνδυασμούς δραστηριοτήτων στο πλήθος των παιδιών.

### **Συμπεράσματα από την ταξινόμηση των παιδιών της Β' Γυμνασίου σε ομάδες**

\* Το μεγαλύτερο ποσοστό των παιδιών αυτής της ηλικίας ( $P=0.3556$  αλλιώς 35%) ασχολείται με τον **αθλητισμό** (1) και την **ακρόαση μουσικής** (16).

\* Η αμέσως επόμενη ομάδα με λίγο μικρότερο ποσοστό πληθυσμού από την μεγαλύτερη ομάδα ( $P=0.2966$  ή 30%) εκτός από τις **παραπάνω δραστηριότητες** ασχολείται και με τον **υπολογιστή** (17), (18), (19) δηλαδή παίζει παιχνίδια στον υπολογιστή, σχεδιάζει εικόνες και κατεβάζει μουσική από το διαδίκτυο.

\* Ένα μικρό ποσοστό παιδιών ( $P=0.1838$ ) τείνει να ασχολείται με όλα τα **παραπάνω** που αναφέραμε αλλά και με **καλλιτεχνικές δραστηριότητες** όπως τη ζωγραφική και την φωτογραφία (5), (6).

\* Η μικρότερη ομάδα παιδιών της Δευτέρας Γυμνασίου (ένα 5% αλλιώς  $P=0.0456$ ) ασχολείται με τις **περισσότερες δραστηριότητες**.

\* Τέλος, υπάρχει ένα μικρό ποσοστό πληθυσμού ( $P=0.1184$ ) που ασχολείται με τις **περισσότερες δραστηριότητες σε μικρότερο βαθμό** όμως από την μικρότερη ομάδα παραπάνω.

Συμπερασματικά, τα περισσότερα παιδιά της Β' Γυμνασίου ασχολούνται όπως και της Στ' Δημοτικού με τον αθλητισμό και την μουσική χωρίζόμενα έπειτα σε ομάδες ανάλογα με τις επιπλέον δραστηριότητες.

## Συμπεράσματα από την ταξινόμηση των παιδιών της Α΄ Λυκείου σε ομάδες

\* Όπως και στις προηγούμενες ηλικίες, η μικρότερη ομάδα ( $P=0.0437$  το 4%) ασχολείται **σχεδόν με όλες** τις δραστηριότητες.

\* Το μεγαλύτερο ποσοστό του πληθυσμού ( $P=0.3597$  ή 36%) ασχολείται με τον **αθλητισμό** (1) και την **ακρόαση μουσικής** (16).

\* Ένα λίγο μικρότερο σύνολο παιδιών της Α΄ Λυκείου ( $P=0.2925$  ή 29%) από την προηγούμενη ομάδα ασχολείται με τις **παραπάνω δραστηριότητες** και με τον **υπολογιστή** και συγκεκριμένα παίζουν παιχνίδια, σχεδιάζουν εικόνες και ‘κατεβάζουν’ μουσική από το διαδίκτυο.

\* Ένα μικρό ποσοστό παιδιών ( $P=0.1953$ ) τείνει να ασχολείται σε μικρότερο ποσοστό, με όσες δραστηριότητες έχουμε ήδη αναφέρει **παραπάνω**, με την **ανάγνωση βιβλίων** και τα **καλλιτεχνικά** δηλαδή το χορό, τη ζωγραφική, τη φωτογραφία.

\* Τέλος, η τάση της ομάδας με μικρό ποσοστό πληθυσμού (11%) είναι να ασχολείται με **τις περισσότερες** από τις δραστηριότητες αλλά σε μικρότερο βαθμό από ότι της μικρότερης ομάδας που αναφέραμε παραπάνω, και κυρίως ασχολείται με τον αθλητισμό, τον υπολογιστή και περισσότερο με μουσικά όργανα και τη μουσική.

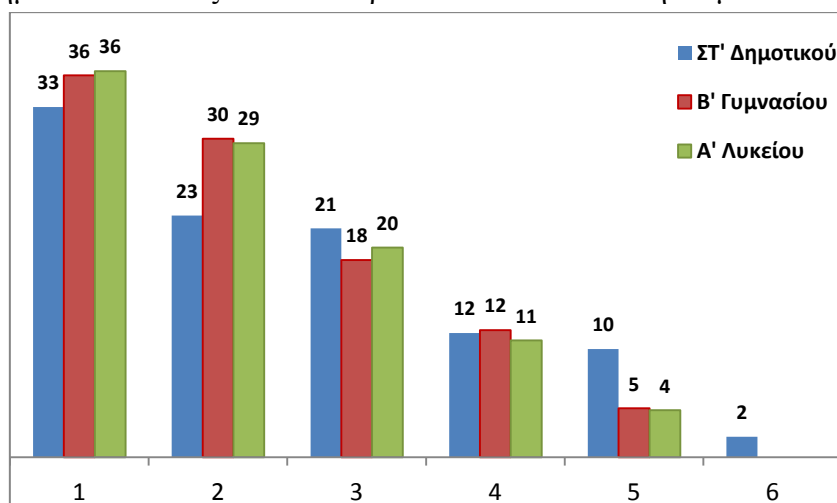
Συγκρίνοντας τα συμπεράσματα των τριών αναλύσεων μπορούμε να συμπεράνουμε ότι δεν υπάρχουν μεγάλες αποκλίσεις στις συνήθειες των παιδιών των τριών ηλικιών. Ενώ μεταξύ των παιδιών της ίδιας ηλικίας υπάρχουν διαφορές στις συνήθειες.

Στον Πίνακα 8 παρουσιάζεται ο πληθυσμός κάθε ομάδας ταξινομημένες από την μεγαλύτερη στη μικρότερη. Ενώ, ακολουθούν τα διαγράμματα των ομάδων ταξινομημένων και των τριών ηλικιών ώστε να απεικονιστούν οι προτιμήσεις των παιδιών.

Ομάδα	Πληθυσμός ανά ομάδα		
	ΣΤ΄ Δημοτικό	Β΄ Γυμνασίου	Α΄ Λυκείου
1	0.3261	0.3556	0.3597
2	0.2252	0.2966	0.2925
3	0.2131	0.1838	0.1953
4	0.1158	0.1184	0.1088
5	0.1009	0.0456	0.0437
6	0.019		

Πίνακας 8 Ο πληθυσμός κάθε ομάδας των τριών αναλύσεων

Στο Γράφημα 12 απεικονίζεται καλύτερα το ποσοστό του πληθυσμού κάθε ομάδας.



Γράφημα 12 Πληθυσμός κάθε ταξινομημένης ομάδας των τριών ηλικιών

## 5.5 Ταξινόμηση με ανεξάρτητη μεταβλητή το φύλο

### 5.5.1 Εισαγωγή

Συνεχίζοντας την εφαρμογή της Latent Class Analysis στην ανάλυση δεδομένων θα προχωρήσουμε ξανά στην ταξινόμηση δεδομένων των τριών ηλικιών χρησιμοποιώντας, πλέον, άλλη μία μεταβλητή. Πρόκειται να μελετήσουμε κατά πόσο επηρεάζει το φύλο την ταξινόμηση των δεδομένων δηλαδή πως κατανέμονται τα παιδιά ανά φύλο ανάλογα με τις προτιμήσεις τους. Επομένως, προχωράμε σε ανάλυση παλινδρόμησης του μοντέλου με ανεξάρτητη μεταβλητή το φύλο. Η διαδικασία της ανάλυσης αυτής έχει περιγραφεί αναλυτικά στο κεφάλαιο 4.2 παραπάνω. Ενώ, όπως ίσχυε στις προηγούμενες αναλύσεις τα εξαγόμενα αποτελέσματα κάθε επιλεγείσας ομαδοποίησης που παίρνουμε από το πρόγραμμα R Studio με την εντολή PoLCA με την μορφή που τα λαμβάνουμε δίνονται στο Παράρτημα. Παραθέτουμε παρακάτω τις εντολές της ανάλυσης παλινδρόμησης όπως θα εισαχθούν στο πρόγραμμα, ενώ οι πρώτες εντολές παραμένουν ίδιες με την ανάλυση του βασικού μοντέλου.

```
fr2<-cbind(GR19_10,      GR19_11,      GR19_12,      GR19_13,
           GR19_14,      GR19_15,      GR19_16,      GR19_17,
           GR19_18,      GR19_19,      GR19_20,
           GR19_01,      GR19_02,      GR19_03,      GR19_04,
           GR19_05,      GR19_06,      GR19_07,      GR19_08,
           GR19_09)~M1
lca1sex<-poLCA(fr2,as.data.frame(grade6),nclass=1,maxiter=2000,nrep=10)
lca2sex<-poLCA(fr2,as.data.frame(grade6),nclass=2,maxiter=2000,nrep=10,
graphs=TRUE)
lca3sex<-poLCA(fr2,as.data.frame(grade6),nclass=3,maxiter=2000,nrep=10,
graphs=TRUE)
lca4sex<-poLCA(fr2,as.data.frame(grade6),nclass=4,maxiter=2000,nrep=10,
graphs=TRUE)
lca5sex<-poLCA(fr2,as.data.frame(grade6),nclass=5,maxiter=2000,nrep=10,
graphs=TRUE)
lca6sex<-poLCA(fr2,as.data.frame(grade6),nclass=6,maxiter=2000,nrep=10,
graphs=TRUE)
bicssex<-cbind(lca1sex$bic, lca2sex$bic, lca3sex$bic, lca4sex$bic, lca5sex$bic,
lca6sex$bic)
bicssex
```

### 5.5.2 ΣΤ' Δημοτικού

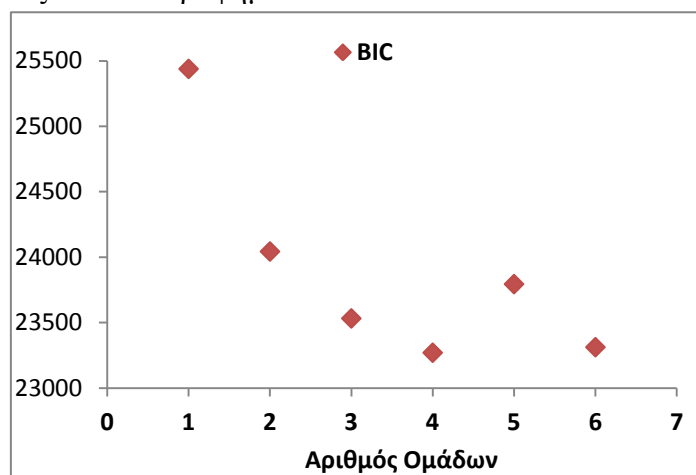
Στον Πίνακα 9 εμφανίζονται οι αριθμοί του κριτηρίου BIC για τις έξι περιπτώσεις ομαδοποιήσεων ώστε να επιλεγεί ο κατάλληλος αριθμός ομάδων.

	Αριθμός Ομάδων					
	1	2	3	4	5	6
BIC	25438	24043	23532	<b>23270</b>	23794	23312

Πίνακας 9 Στ' Δημοτικού\_κριτήριο BIC



Γραφικά απεικονίζεται στο Γράφημα 13 που ακολουθεί.



Γράφημα 13 Στ' Δημοτικού\_Κριτήριο BIC

Όπως παρατηρούμε το κριτήριο παρουσιάζει μεταβολές με χαμηλότερη τιμή το σημείο των τεσσάρων ομάδων. Επομένως, κρίνεται η καλύτερη επιλογή στη παρούσα φάση.

Η ανεξάρτητη μεταβλητή ΦΥΛΟ είναι στατιστικά σημαντική στην ομαδοποίηση των δεδομένων τεσσάρων κλάσεων όπως μπορούμε να παρατηρήσουμε από τις παρακάτω πληροφορίες που παίρνουμε από το πρόγραμμα.

Fit for 4 latent classes:

2 / 1

	Coefficient	Std. error	t value	Pr(> t )
(Intercept)	3.03943	0.40937	7.425	0
M1	-1.83289	0.23310	-7.863	0

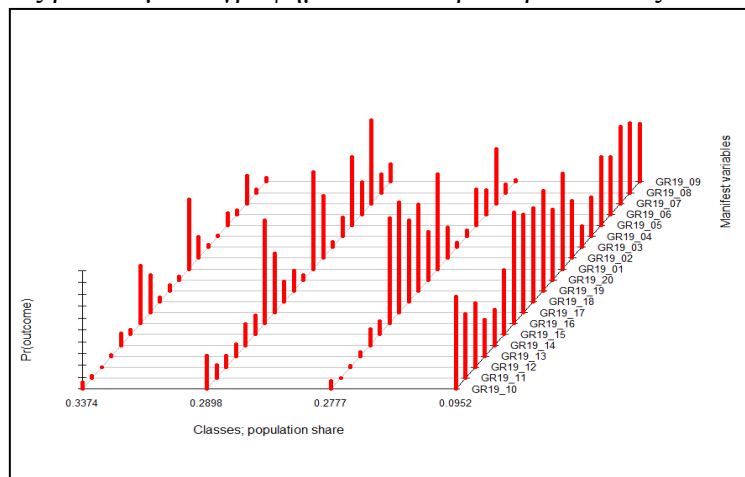
3 / 1

	Coefficient	Std. error	t value	Pr(> t )
(Intercept)	2.44418	0.42994	5.685	0
M1	-1.54483	0.24400	-6.331	0

4 / 1

	Coefficient	Std. error	t value	Pr(> t )
(Intercept)	0.61431	0.51331	1.197	0.232
M1	-1.03474	0.29410	-3.518	0.000

Στο Γράφημα 14 απεικονίζονται οι ομάδες των μεταβλητών από την μεγαλύτερη στην μικρότερη, όπως βλέπουμε το γράφημα από τα αριστερά στα δεξιά.



Γράφημα 14 Στ' Δημοτικού\_Ομαδοποίηση δραστηριοτήτων βάσει του φύλου



Οι πιθανότητες ανά φύλο να συμμετέχουν σε κάθε δραστηριότητα δίνονται από τον Πίνακα 10, παρακάτω.

	Πιθανότητες ανά ομάδα							
	1		2		3		4	
Ερώτηση	ΚΟΡΙΤΣΙΑ	ΑΓΟΡΙΑ	ΚΟΡΙΤΣΙΑ	ΑΓΟΡΙΑ	ΚΟΡΙΤΣΙΑ	ΑΓΟΡΙΑ	ΚΟΡΙΤΣΙΑ	ΑΓΟΡΙΑ
1 – Αθλητικά	0.4139	0.8095	0.1827	0.7987	0.2013	0.5861	0.1905	0.8173
2 – Χορός	0.8207	0.4864	0.4744	0.2584	0.7416	0.1793	0.5136	0.5256
3 – Προσκοπισμός	0.9784	0.1773	0.9585	0.0364	0.9636	0.0216	0.8227	0.0415
4 – Συγγραφή	0.9952	0.3317	0.8400	0.0470	0.9530	0.0048	0.6683	0.1600
5 – Καλλιτεχνικά	0.8978	0.5803	0.4198	0.3033	0.6967	0.1022	0.4197	0.5802
6 – Φωτογραφία	0.9616	0.4877	0.7278	0.2041	0.7959	0.0384	0.5123	0.2722
7 – Διάβασμα βιβλίων	0.7664	0.6526	0.2944	0.4613	0.5387	0.2336	0.3474	0.7056
8 – Σύνθεση Μουσικής	0.9748	0.5865	0.8454	0.0697	0.9303	0.0252	0.4135	0.1546
9 – Χορωδία	0.9642	0.4904	0.8486	0.0151	0.9849	0.0358	0.5096	0.1514
10 – Μουσικά Όργανα	0.9410	0.7776	0.7163	0.0704	0.9296	0.0590	0.2224	0.2837
11 – Συγκρότημα	0.9831	0.5432	0.8913	0.0000	1.0000	0.0169	0.4568	0.1087
12 – Ιδιαίτερα μαθήματα μουσικής	0.9978	0.5396	0.9009	0.0102	0.9898	0.0022	0.4604	0.0991
13 – Θέατρο	0.9891	0.3121	0.9008	0.0371	0.9629	0.0109	0.6879	0.0992
14 – Δραστηριότητες της Εκκλησίας	0.8967	0.3014	0.8209	0.1384	0.8616	0.1033	0.6986	0.1791
15 – Συναυλίες, θέατρο, μουσεία	0.9588	0.5427	0.8389	0.1150	0.8850	0.0412	0.4573	0.1611
16 – Ακούω μουσική	0.5102	0.9405	0.1260	0.8923	0.1077	0.4898	0.0595	0.8740
17 – Ηλεκτρονικά παιχνίδια Η/Υ	0.6805	0.8265	0.4996	0.9312	0.0688	0.3195	0.1735	0.5004
18 – Επεξεργασία εικόνων Η/Υ	0.9621	0.7911	0.8286	0.6874	0.3126	0.0379	0.2089	0.1714
19 – “Κατέβασμα μουσικής online”	0.9525	0.8475	0.8297	0.7302	0.2698	0.0475	0.1525	0.1703
20 – Δημιουργία προγραμμάτων/παιχνιδιών στον Η/Υ	0.9711	0.5981	0.9576	0.4043	0.5957	0.0289	0.4019	0.0424
<b>ΠΛΗΘΥΣΜΟΣ ΑΝΑ ΦΥΛΟ</b>	<b>0.233</b>	<b>0.448</b>	<b>0.436</b>	<b>0.134</b>	<b>0.229</b>	<b>0.330</b>	<b>0.102</b>	<b>0.088</b>
<b>ΣΥΝΟΛΟ ΠΛΗΘΥΣΜΟΥ</b>	<b>0.3374</b>		<b>0.2898</b>		<b>0.2777</b>		<b>0.0952</b>	

Πίνακας 10 Στ' Δημοτικού\_Πιθανότητες των δραστηριοτήτων ανά φύλο

Για να υπολογίσουμε το ποσοστό των αγοριών και των κοριτσιών που ανήκουν σε κάθε κλάση παίρνουμε τον τύπο:  $\ln(p_2/p_1) = Z\beta_2$  για παράδειγμα για την δεύτερη κλάση όπου  $p_2$  είναι η πιθανότητα να είναι στην τάξη 2,  $Z$  το διάνυσμα  $(1, z)$  με  $z = 1$  για αγόρι και  $z = 2$  για κορίτσι, και  $\beta_2$  είναι το αντίστοιχο διάνυσμα των συντελεστών /Coefficients, στην εξίσωση (12) των Linzer & Lewis. Για  $\beta_1 = 0$  ο τύπος σχηματίζεται για τις 4 επί μέρους πιθανότητες:  $p_i = e^{Z\beta_i} / \sum e^{Z\beta_j}$

## Συμπεράσματα ομαδοποίησης

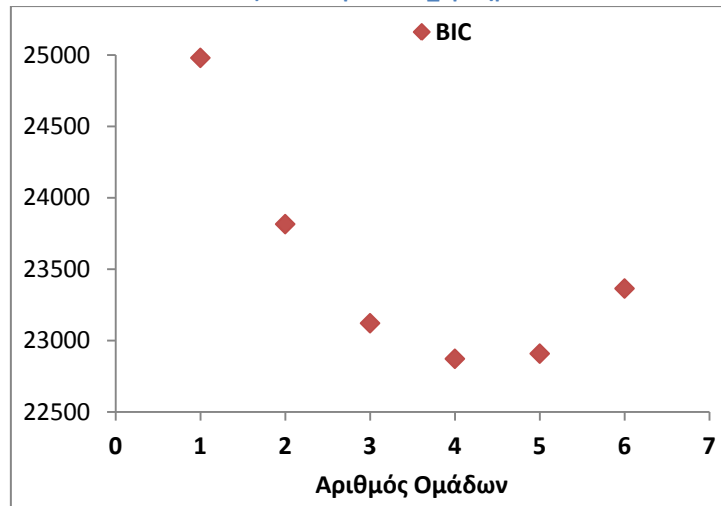
- \* Το μικρότερο ποσοστό του πληθυσμού των παιδιών ( $P=0.0952$ ) αναφέρει ότι ασχολείται με **όλες σχεδόν τις δραστηριότητες** όπου υπερέχουν λίγο τα κορίτσια.
- \* Το μεγαλύτερο ποσοστό (34% με  $P=0.3374$ ) ασχολείται με τον **αθλητισμό, τη μουσική** και την **ανάγνωση βιβλίων**, με τα αγόρια να υπερισχύουν.
- \* Ένα λίγο μικρότερο από το προηγούμενο ( $P=0.2898$ ) ασχολείται και με **καλλιτεχνικές δραστηριότητες** όπως ζωγραφική και χορό εκτός των **υπόλοιπων δραστηριοτήτων** και όπως θα περιμέναμε τα κορίτσια είναι περισσότερα από τα αγόρια.
- \* Τέλος, το 28% των μαθητών ( $P=0.2777$ ) **εκτός των άλλων ασχολείται και με τον υπολογιστή**, με μεγάλη διαφορά να ασχολούνται κυρίως τα αγόρια με αυτές τις δραστηριότητες.

### 5.5.3 Β' Γυμνασίου

Όπως παρατηρούμε από τις έξι περιπτώσεις ταξινόμησης ικανός αριθμός αποτελούν οι τέσσερις κλάσεις όπως φαίνεται και στον Πίνακα 11 παρακάτω.

	Αριθμός Ομάδων					
	1	2	3	4	5	6
BIC	24980	23816	23122	<b>22873</b>	22909	23365

Πίνακας 11 Β' Γυμνασίου\_Κριτήριο BIC



Γράφημα 15 Β' Γυμνασίου\_Κριτήριο BIC

Το κριτήριο παρουσιάζει μεταβολές ανά περίπτωση ομαδοποίησης, όπως επισημαίνεται όμως στο Γράφημα 15 η περίπτωση των τεσσάρων με τις πέντε ομάδες δεν διαφέρουν πολύ, με χαμηλότερη τιμή την περίπτωση τεσσάρων ομάδων. Επιλέγουμε τις τέσσερις ομάδες ως ικανοποιητικός αριθμός ομάδων. Μελετώντας αν η μεταβλητή M1 είναι στατιστικά σημαντική σε αυτή την ταξινόμηση, παρατηρούμε ότι σχεδόν σε όλες τις ομάδες είναι στατιστικά σημαντική σύμφωνα με το Pr.

Fit for 4 latent classes:

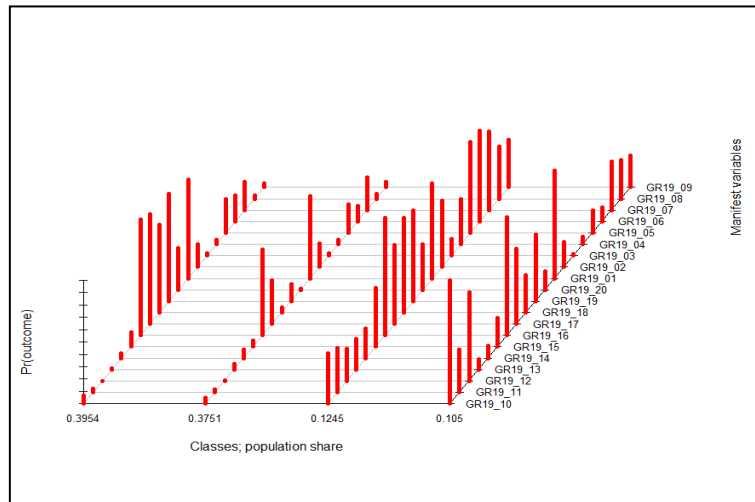
=====				
2 / 1				
	Coefficient	Std. error	t value	Pr(> t )
(Intercept)	1.53962	0.29356	5.245	0
M1	-1.28986	0.23167	-5.568	0
=====				
3 / 1				
	Coefficient	Std. error	t value	Pr(> t )
(Intercept)	-4.51971	0.71756	-6.299	0
M1	2.41760	0.36563	6.612	0
=====				
4 / 1				
	Coefficient	Std. error	t value	Pr(> t )
(Intercept)	-0.67420	0.33205	-2.03	0.043
M1	-0.04557	0.21746	-0.21	0.834
=====				

Στον Πίνακα 12 καταχωρήθηκαν οι πιθανότητες για κάθε φύλο να εκτελεί κάθε δραστηριότητα.

	Πιθανότητες ανά ομάδα							
	1		2		3		4	
Ερώτηση	ΚΟΡΙΤΣΙΑ	ΑΓΟΡΙΑ	ΚΟΡΙΤΣΙΑ	ΑΓΟΡΙΑ	ΚΟΡΙΤΣΙΑ	ΑΓΟΡΙΑ	ΚΟΡΙΤΣΙΑ	ΑΓΟΡΙΑ
1 – Αθλητικά	0.2003	0.7997	0.3304	0.6696	0.2262	0.7738	0.1260	0.8740
2 – Χορός	0.8152	0.1848	0.8089	0.1911	0.4628	0.5372	0.7981	0.2019
3 – Προσκοπισμός	0.9826	0.0174	0.9808	0.0192	0.8624	0.1376	0.9846	0.0154
4 – Συγγραφή	0.9628	0.0372	0.9644	0.0356	0.6362	0.3638	0.9408	0.0592
5 – Καλλιτεχνικά	0.7281	0.2719	0.7657	0.2343	0.2621	0.7379	0.8160	0.1840
6 – Φωτογραφία	0.7906	0.2094	0.8736	0.1264	0.2671	0.7329	0.8848	0.1152
7 – Διάβασμα βιβλίων	0.7729	0.2271	0.7355	0.2645	0.3603	0.6397	0.6040	0.3960
8 – Σύνθεση Μουσικής	0.9695	0.0305	0.9604	0.0396	0.5738	0.4262	0.6841	0.3159
9 – Χορωδία	0.9690	0.0302	0.9524	0.0476	0.6122	0.3878	0.7425	0.2575
10 – Μουσικά Όργανα	0.9330	0.0670	0.9521	0.0479	0.5886	0.4114	0.0000	1.0000
11 – Συγκρότημα	0.9705	0.0295	0.0759	0.0241	0.6386	0.3614	0.6525	0.3475
12 – Ιδιαίτερα μαθήματα μουσικής	1.0000	0.0000	0.9923	0.0077	0.7391	0.2609	0.2806	0.7194
13 – Θέατρο	0.9888	0.0112	0.9505	0.0495	0.7489	0.2511	0.9186	0.0814
14 – Δραστηριότητες της Εκκλησίας	0.9588	0.0412	0.9274	0.0726	0.7603	0.2397	0.8994	0.1006
15 – Συναυλίες, θέατρο, μουσεία	0.8824	0.1176	0.9499	0.0501	0.5209	0.4791	0.7650	0.2350
16 – Ακούω μουσική	0.0598	0.9402	0.3036	0.6964	0.0458	0.9542	0.0404	0.9596
17 – Ηλεκτρονικά παιχνίδια Η/Υ	0.1071	0.8929	0.6453	0.3547	0.3573	0.6427	0.3922	0.6078
18 – Επεξεργασία εικόνων Η/Υ	0.2888	0.7112	0.9548	0.0452	0.2352	0.7648	0.6932	0.3068
19 – “Κατέβασμα μουσικής online”	0.1302	0.8698	0.8586	0.1414	0.2628	0.7372	0.4614	0.5386
20 – Δημιουργία προγραμμάτων/παιχνιδιών στον Η/Υ	0.6583	0.3417	0.9901	0.0099	0.6323	0.3677	0.8464	0.1536
<b>ΠΛΗΘΥΣΜΟΣ ΑΝΑ ΦΥΛΟ</b>	<b>0.314</b>	<b>0.346</b>	<b>0.111</b>	<b>0.444</b>	<b>0.430</b>	<b>0.042</b>	<b>0.146</b>	<b>0.168</b>
<b>ΣΥΝΟΛΟ ΠΛΗΘΥΣΜΟΥ</b>	<b>0.3954</b>		<b>0.3751</b>		<b>0.1245</b>		<b>0.105</b>	

Πίνακας 12 Β' Γυμνασίου\_Πιθανότητες των δραστηριοτήτων ανά φύλο

Ακολούθως, στο γράφημα 16 απεικονίζεται η ομαδοποίηση των δεδομένων.



Γράφημα 16 Β' Γυμνασίου\_Ομαδοποίηση δραστηριοτήτων βάσει του φύλου

### Συμπεράσματα ομαδοποίησης

- \* Στην μεγαλύτερη ομάδα ( $P=0.3954$ ) με 40% ποσοστό πληθυσμού, τα παιδιά ασχολούνται με τις δραστηριότητες **αθλητισμό, μουσική, υπολογιστή** (1,16, 17,19) με τα αγόρια να είναι λίγο περισσότερα.
- \* Το αμέσως μικρότερο ποσοστό ( $P=0.3751$ ) ασχολείται λίγο με τον **αθλητισμό και την ακρόαση μουσικής** με τα αγόρια και πάλι να προπορεύονται αρκετά.
- \* Ένα σχετικά μικρό ποσοστό ( $P=0.1245$ ) ασχολείται με την **πλειονότητα των δραστηριοτήτων** με την πλειονότητα του πληθυσμού να είναι κορίτσια.
- \* Τέλος, το μικρότερο πλήθος παιδιών αυτής της ηλικίας ( $P=0.105$ ) εκτός από τις συνηθισμένες δραστηριότητες, **αθλητισμό και την ακρόαση μουσικής** (1 και 16), δείχνει ενδιαφέρον σε δραστηριότητες σχετικές με τη μουσική όπως είναι **μουσικά όργανα και ιδιαίτερα για την εκμάθησή τους** (10 και 12) με τα αγόρια να είναι λίγο περισσότερα.

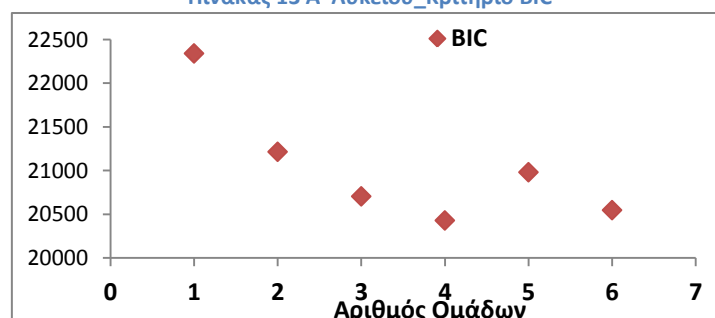
Όπως κατανοούμε τα συμπεράσματα είναι παρόμοια με αυτά του βασικού μοντέλου αλλά με το μοντέλο παλινδρόμησης παίρνουμε ακόμα περισσότερες πληροφορίες, για την κατανομή των φύλων.

### 5.5.4 Α' Λυκείου

Ο Πίνακας 13 με τις τιμές του κριτηρίου BIC.

	Αριθμός Ομάδων					
	1	2	3	4	5	6
BIC	22341	21215	20705	<b>20429</b>	20981	20547

Πίνακας 13 Α' Λυκείου\_Κριτήριο BIC



Γράφημα 17 Α' Λυκείου\_Κριτήριο BIC

Όπως παρατηρούμε οι τιμές του κριτηρίου έχουν αποκλίσεις μεταξύ τους. Επιλέγουμε την περίπτωση τεσσάρων ομάδων όπως τις προηγούμενες αναλύσεις γιατί είναι χαμηλότερη τιμή. Επιβεβαιώνεται ότι το φύλο είναι σημαντικό στατιστικά σύμφωνα με τις πιθανότητες των παρακάτω πληροφοριών.

Fit for 4 latent classes:

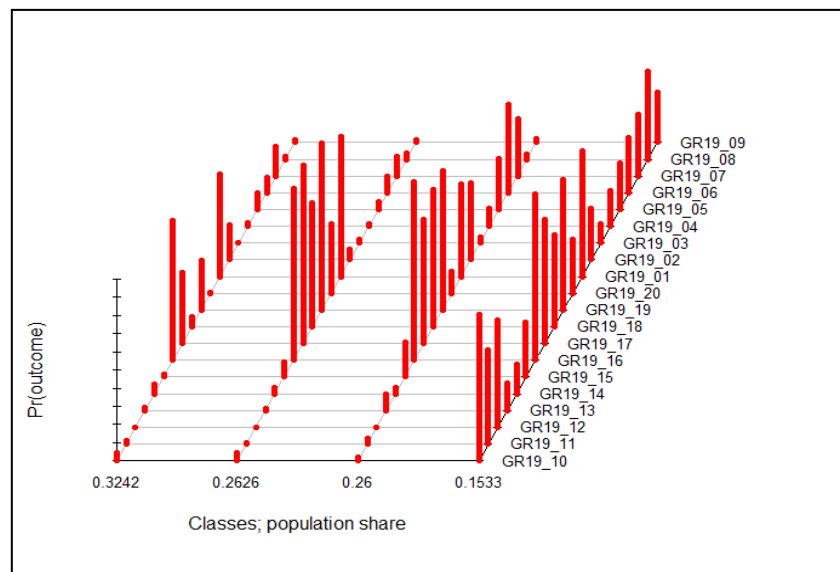
=====				
2 / 1				
	Coefficient	Std. error	t value	Pr(> t )
(Intercept)	2.42450	0.35273	6.873	0
M1	-1.99317	0.28973	-6.880	0
=====				
3 / 1				
	Coefficient	Std. error	t value	Pr(> t )
(Intercept)	-3.87812	0.65051	-5.962	0
M1	2.08003	0.33348	6.237	0
=====				
4 / 1				
	Coefficient	Std. error	t value	Pr(> t )
(Intercept)	0.26063	0.32921	0.792	0.429
M1	-0.68992	0.21889	-3.152	0.002
=====				

Οι πιθανότητες της ταξινόμησης ανά φύλο δίνονται στον Πίνακα 14 που ακολουθεί.

	Πιθανότητες ανά ομάδα							
	1		2		3		4	
Ερώτηση	ΚΟΡΙΤΣΙΑ	ΑΓΟΡΙΑ	ΚΟΡΙΤΣΙΑ	ΑΓΟΡΙΑ	ΚΟΡΙΤΣΙΑ	ΑΓΟΡΙΑ	ΚΟΡΙΤΣΙΑ	ΑΓΟΡΙΑ
1 – Αθλητικά	0.4363	0.5637	0.2262	0.7738	0.4909	0.5091	0.3060	0.6940
2 – Χορός	0.8099	0.1901	0.9430	0.0570	0.5796	0.4204	0.7172	0.2828
3 – Προσκοπισμός	1.0000	0.0000	0.9822	0.0178	0.9677	0.0323	0.8918	0.1082
4 – Συγγραφή	0.9828	0.0172	0.9868	0.0132	0.9024	0.0976	0.8035	0.1965
5 – Καλλιτεχνικά	0.9073	0.0927	0.9579	0.0421	0.7199	0.2801	0.7462	0.2538
6 – Φωτογραφία	0.9149	0.0851	0.9119	0.0881	0.5162	0.4838	0.6948	0.3052
7 – Διάβασμα βιβλίων	0.8347	0.1653	0.8935	0.1065	0.6840	0.3160	0.6596	0.3404
8 – Σύνθεση Μουσικής	0.9873	0.0127	0.9680	0.0320	0.9710	0.0290	0.5163	0.4837
9 – Χορωδία	0.9862	0.0138	0.9877	0.0123	0.9784	0.0216	0.7240	0.2760
10 – Μουσικά Όργανα	0.9793	0.0207	0.9589	0.0411	0.1937	0.8063	0.9552	0.0448
11 – Συγκρότημα	0.9668	0.0332	0.9933	0.0067	0.4807	0.5193	0.9825	0.0175
12 – Ιδιαίτερα μαθήματα μουσικής	1.0000	0.0000	1.0000	0.0000	0.4097	0.5903	1.0000	0.0000
13 – Θέατρο	0.9125	0.0875	0.9916	0.0084	0.8481	0.1519	0.9840	0.0160
14 – Δραστηριότητες της Εκκλησίας	0.9629	0.0371	0.9638	0.0362	0.8377	0.1623	0.9482	0.0518
15 – Συναυλίες, θέατρο, μουσεία	0.8098	0.1902	0.9163	0.0837	0.6983	0.3017	0.9855	0.0145
16 – Ακούω μουσική	0.0180	0.9820	0.0570	0.9430	0.0871	0.9129	0.2300	0.7700
17 – Ηλεκτρονικά παιχνίδια Η/Υ	0.3193	0.6807	0.0168	0.9832	0.3145	0.6855	0.6114	0.3886
18 – Επεξεργασία εικόνων Η/Υ	0.2423	0.7577	0.3205	0.6795	0.4958	0.5142	0.9456	0.0544
19 – “Κατέβασμα μουσικής online”	0.2347	0.7653	0.0838	0.9162	0.2820	0.7180	0.7250	0.2750
20 – Δημιουργία προγραμμάτων/παιχνιδιών στον Η/Υ	0.9929	0.0071	0.6178	0.3822	0.8790	0.1210	0.7036	0.2964
<b>ΠΛΗΘΥΣΜΟΣ ΑΝΑ ΦΥΛΟ</b>	<b>0.349</b>	<b>0.298</b>	<b>0.073</b>	<b>0.459</b>	<b>0.463</b>	<b>0.049</b>	<b>0.114</b>	<b>0.194</b>
<b>ΣΥΝΟΛΟ ΠΛΗΘΥΣΜΟΥ</b>	<b>0.3242</b>		<b>0.2626</b>		<b>0.26</b>		<b>0.1533</b>	

Πίνακας 14 Α' Λυκείου\_Πιθανότητες των δραστηριοτήτων ανά φύλο

Εικονικά παριστάνεται η ομαδοποίηση στο Γράφημα 18 που ακολουθεί.



Γράφημα 18 Α' Λυκείου\_Ομαδοποίηση δραστηριοτήτων βάσει του φύλου

### Συμπεράσματα ομαδοποίησης

- \* Το 32% στην μεγαλύτερη ομάδα παιδιών ( $P=0.3242$ ) ασχολείται περισσότερο με τον **αθλητισμό** και την **ακρόαση μουσικής**, με τα κορίτσια να είναι περισσότερα.
- \* Στην αμέσως επόμενη ομάδα ( $P=0.2626$ ) όπου είναι περισσότερα τα αγόρια ο πληθυσμός ασχολείται αρκετά με τον **υπολογιστή**, πέραν των άλλων δραστηριοτήτων που αναφέραμε παραπάνω.
- \* Ενώ η επόμενη ομάδα πληθυσμού ( $P=0.26$ ) συνηθίζει τις προηγούμενες δραστηριότητες και τις **καλλιτεχνικές δραστηριότητες** όπως ζωγραφική και χορός με τα κορίτσια να υπερισχύουν σημαντικά όπως είναι φυσικό.
- \* Τέλος, με τις **περισσότερες δραστηριότητες** φαίνεται να ασχολείται η μικρότερη ομάδα παιδιών ( $P=0.1533$ ) με λίγο περισσότερα αγόρια.

## ΚΕΦΑΛΑΙΟ 6. Άλλες εφαρμογές της Latent Class Analysis

Η μέθοδος αυτή είναι ευρέως γνωστή στους επιστημονικούς κλάδους. Κατά καιρούς έχει χρησιμοποιηθεί για την επίλυση διαφόρων ειδών προβλημάτων. Ένα από αυτά είναι και η προσπάθεια να δοθεί απάντηση στο ερώτημα αν οι καταναλωτές αγοράζουν υγιεινές τροφές για λόγους υγείας ή λόγω κοινωνικής αναγνώρισης καθώς είναι περισσότερο ποιοτικές τροφές ενώ έχει αποδειχτεί ότι συνδέεται αρκετά στη συνείδηση των καταναλωτών η ποιότητα με την τιμή. Συγκεκριμένα, πραγματοποιήθηκε πείραμα για να εξεταστεί πόσο πρόθυμοι είναι οι καταναλωτές να πληρώσουν αρκετά χρήματα μία τροφή που προσφέρει ποιότητα (Palma, Ness, & Anderson, 2015). Διενεργήθηκε έρευνα σε 201 άτομα μιας πόλης μεσαίου μεγέθους το 2014. Στόχος ήταν να προσδιοριστεί ο τρόπος με τον οποίο καταναλώνουν οι σημερινοί αγοραστές που εκτίθενται τόσο πολύ στα ΜΜΕ και στις διαφημίσεις προσμετρώντας και το εισόδημά τους αλλιώς μελετάται αν αγοράζουν ποιοτικές τροφές που είναι και ακριβές για λόγους υγείας ή λόγω κοινωνικής αναγνώρισης και οι διαφορές στα κίνητρα μεταξύ των καταναλωτών μιας κοινωνίας. Χρησιμοποιήθηκε η μέθοδος Latent Class Analysis για να χωριστεί το κοινό σε ομάδες ανάλογα με τον τρόπο απάντησης στις δώδεκα ερωτήσεις που τους δόθηκαν. Μεταβλητές αποτέλεσαν οι παρακάτω: πόσες φορές την εβδομάδα ασκείται κάποιος, πόσα ξοδεύει σε λαχανικά κάθε εβδομάδα, αν πιστεύει ότι τον παρατηρούν οι άλλοι όταν αγοράζει την πιο ακριβή μάρκα ενός προϊόντος, αν απολαμβάνει κοινωνική αναγνώριση όταν αγοράζει ακριβά τρόφιμα, αν πιστεύει ότι κάτι σημαίνει για τους άλλους ανθρώπους να καταναλώνει κάποιος ακριβές τροφές, αν πιστεύει ότι οι φίλοι θα τον θεωρούν φτωχό διαφορετικά, οι άλλοι κρίνουν το είδος των προϊόντων που αγοράζει κάποιος, αν πιστεύει ότι ακόμα και για φθηνό προϊόν προκαλεί εντύπωση όταν αγοράζει το ακριβότερο, αν κάνει συχνά εξετάσεις και τέλος αν έχει γνώση και προσέχει την κατάσταση της υγείας του. Επίσης, σημαντικός παράγοντας αποτέλεσαν τα δημογραφικά στοιχεία των ατόμων. Τα άτομα σύμφωνα με τις απαντήσεις τους διανεμήθηκαν σε τέσσερις κλάσεις κατά την ανάλυση. Η πρώτη ομάδα αποτελεί το 38% του πληθυσμού χαρακτηριστικό αυτής της ομάδας είναι ότι βασικό κίνητρο είναι η **υγεία και λιγότερο η κοινωνική αναγνώριση**. Είναι αρκετά δραστήριοι και ξοδεύουν αρκετά χρήματα σε λαχανικά την εβδομάδα. Το μέσο εισόδημα αυτής της ομάδας είναι 55,000\$, το μεγαλύτερο όλων των κλάσεων. Στην δεύτερη κλάση καταχωρήθηκε το 24% των συμμετεχόντων με **υψηλό** σκορ σε **υγεία και κοινωνική αναγνώριση**, εφόσον υπάρχει 72% πιθανότητα οι άνθρωποι να πιστεύουν ότι οι άλλοι τους παρατηρούν όταν αγοράζουν τα πιο ακριβά προϊόντα και 90% να γνωρίζουν και να βελτιώνουν την υγεία τους. Η τρίτη κλάση αποτελείται από το 33% όπου ο πληθυσμός δίνει **χαμηλό** σκορ στην **υγεία** και την **κοινωνική αναγνώριση**. Τέλος, η τέταρτη κλάση είναι το 6% του πληθυσμού που δίνει **λιγότερο** προσοχή στην **υγεία** και **περισσότερο στο κοινωνικό στάτους**. Ενώ, σύμφωνα με τον δείκτη WTP, που δείχνει κατά πόσο είναι διαθέσιμοι να πληρώσουν περισσότερα χρήματα για πιο ποιοτικές τροφές, είναι χαμηλότερος στην τρίτη κλάση. Την υψηλότερη τιμή παρουσιάζει η ομάδα 4 κάτι που δείχνει ότι πιστεύουν τα άτομα αυτής της ομάδας ότι ανεβαίνουν κοινωνική τάξη με αυτό τον τρόπο κατανάλωσης. Παρουσιάστηκε λοιπόν ο συσχετισμός του εισοδήματος με τον λόγο αγοράς υγιεινών τροφών.



Επίσης, η Latent Class Analysis αποτελεί μια πολύ χρήσιμη μέθοδος ανάλυσης για του παιδοψυχιάτρους. Καθώς είναι ιδιαίτερα σημαντικό να ομαδοποιούν τις περιπτώσεις παιδιών με παρόμοιες ενδείξεις ώστε να διευκολύνονται στην διάγνωση και στη θεραπεία του ασθενούς. Τα χαμηλά επίπεδα δραστηριότητας, οι καθιστική ζωή και η κατανάλωση περισσότερων λιπαρών τροφών οδηγεί τα παιδιά να έχουν περιττό λίπος. Θα αναφέρουμε ένα παράδειγμα ομαδοποίησης παιδιών σε ομάδες ανάλογα με τα επίπεδα δραστηριότητάς τους, κατά πόσο συνηθίζουν να κάνουν καθιστική ζωή και αν καταναλώνουν επαρκείς θρεπτικές τροφές και λιγότερο λιπαρές. Το σύνολο των παιδιών ταξινομήθηκαν σε τρεις κλάσεις ανάλογα με τα επίπεδα στις παραπάνω παραμέτρους. Έτσι, η μεγαλύτερη ομάδα με το 89% των παιδιών χαρακτηρίζεται από **μέσα επίπεδα δραστηριότητας και καθιστικής ζωής αλλά με καλή διατροφή**. Η δεύτερη ομάδα με το 8% του πληθυσμού αποτελεί ομάδα υψηλού κινδύνου καθώς το σκορ είναι **πολύ χαμηλό σε όλες τις παραμέτρους**. Στην τελευταία ομάδα που είναι και η μικρότερη (2%) οι συνδυασμοί των τιμών των παραμέτρων είναι πιο σύνθετοι, δηλαδή τα **επίπεδα δραστηριότητας είναι υψηλά, αλλά και της καθιστικής ζωής** ενώ καταναλώνονται **σαλάτες και αναψυκτικά**. Από την παραπάνω ανάλυση αντιλαμβανόμαστε ότι η μέθοδος αυτή αποτελεί ένα χρήσιμο εργαλείο στους γιατρούς με τις τόσες περιπτώσεις που φαίνονται αρχικά διαφορετικές αλλά αν ερευνηθούν μπορούν να ομαδοποιηθούν σε περιπτώσεις παρόμοιων κρουσμάτων. Όπως επίσης είναι αρκετά χρήσιμο για την ασφάλεια και την αξιοπιστία των διαγνώσεων να έχουν ομαδοποιηθεί οι συνδυασμοί αιτιών που προκαλούν μια ασθένεια (Berlin, Williams, & Parra, 2014).

Εκτός των παραπάνω εφαρμογών, η Latent Class Analysis έχει χρησιμοποιηθεί και για άλλους σκοπούς όπως είναι η καταγραφή και ομαδοποίηση των απόψεων των πολιτών της Πολωνίας σχετικά με την είσοδό τους στη ζώνη του Ευρώ (Genge, 2013), την ταξινόμηση των τάσεων σύγχρονων καταναλωτών σχετικά με την μουσική. Δηλαδή κατά πόσο προτιμούν ακόμα τον παραδοσιακό τρόπο αγοράζοντας CD ή κατεβάζουν από το διαδίκτυο τραγούδια και βίντεο ή αγοράζουν μουσική από το διαδίκτυο ως πιο νόμιμα (Weijters & Goedertier, 2015).

## **Συμπεράσματα**

Όπως παρατηρούμε η μέθοδος αυτή μπορεί να χρησιμοποιηθεί για εμπορικούς σκοπούς από τμήματα μάρκετινγκ, για επιστημονικούς από γιατρούς αλλά, τέλος, και για πολιτικούς σκοπούς. Οποιοσδήποτε γνωρίζει αυτό το εργαλείο αρκετά μπορεί να το εφαρμόσει για την επίλυση προβλημάτων ανάλυσης δεδομένων διαφορετικών κλάδων.

Στην παρούσα εργασία έγινε προσπάθεια να προσεγγιστεί, αρχικά, το θεωρητικό υπόβαθρο της Latent Class Analysis, της παραπάνω μεθόδου ανάλυσης δεδομένων, ενώ, ακολούθησε η πρακτική εφαρμογή της για να μπορέσει να γίνει περισσότερο κατανοητός ο σκοπός χρήσης αυτής της μεθόδου και ο τρόπος λειτουργίας της.



# ΠΑΡΑΡΤΗΜΑ

## 1. Μεταβλητές και δραστηριότητες

A/A	ΣΕΙΡΑ ΣΤΟ excel	ΜΕΤΑΒΛΗΤΗ-ΔΡΑΣΤΗΡΙΟΤΗΤΑ
1	GR19_10	Παίζω κάποιο μουσικό όργανο
2	GR19_11	Τραγουδάω ή παίζω σε κάποιο συγκρότημα, μπάντα ή ορχήστρα
3	GR19_12	Κάνω ιδιαίτερα μαθήματα μουσικής
4	GR19_13	Παίζω θέατρο
5	GR19_14	Συμμετέχω σε δραστηριότητες της εκκλησίας
6	GR19_15	Πηγαίνω σε συναυλίες, θέατρο, μουσεία κτλ.
7	GR19_16	Ακούω την αγαπημένη μου μουσική
8	GR19_17	Παίζω ηλεκτρονικά παιχνίδια στον υπολογιστή
9	GR19_18	Σχεδιάζω ή επεξεργάζομαι εικόνες στον υπολογιστή
10	GR19_19	"Κατεβάζω" μουσική από το διαδίκτυο
11	GR19_20	Φτιάχνω δικά μου προγράμματα ή παιχνίδια στον υπολογιστή
12	GR19_01	Φυσική άσκηση, αθλητισμό (π.χ. γυμναστική, ποδόσφαιρο, κτλ.)
13	GR19_02	Χορό
14	GR19_03	Προσκοπισμό
15	GR19_04	Γράφω διηγήματα, ποιήματα, επιστολές (όχι για το σχολείο)
16	GR19_05	Ζωγραφική, χειροτεχνία
17	GR19_06	Φωτογραφία
18	GR19_07	Διαβάζω βιβλία
19	GR19_08	Γράφω/Συνθέτω μουσική
20	GR19_09	Συμμετέχω σε χορωδία

Πίνακας 15 Οι τίτλοι των 20 μεταβλητών

## 2. ΣΤ' Δημοτικού\_Latent Class Analysis

Τα αποτελέσματα δύο περιπτώσεων ομαδοποίησης δεδομένων όπως εξάγονται από το πρόγραμμα R Studio για τα παιδιά της Στ' τάξης του Δημοτικού σχολείου. Πρώτη περίπτωση ομαδοποίησης είναι δύο ομάδες ταξινόμησης και δεύτερη περίπτωση έξι ομάδες, με επιλεγθείσα ομαδοποίηση αυτή των έξι κλάσεων.

### 2.1 Ταξινόμηση δύο ομάδων

```
lca2<-poLCA(fr1,as.data.frame(grade6),nclass=2,maxiter=2000,nrep=10, graphs=TRUE)
```

Model 1: llik = -11878.45 ... best llik = -11878.45

Model 2: llik = -11878.45 ... best llik = -11878.45

Model 3: llik = -11878.45 ... best llik = -11878.45

Model 4: llik = -11878.45 ... best llik = -11878.45

Model 5: llik = -11878.45 ... best llik = -11878.45

Model 6: llik = -11878.45 ... best llik = -11878.45

Model 7: llik = -11878.45 ... best llik = -11878.45

Model 8: llik = -11878.45 ... best llik = -11878.45

Model 9: llik = -11878.45 ... best llik = -11878.45

Model 10: llik = -11878.45 ... best llik = -11878.45

Conditional item response (column) probabilities,

by outcome variable, for each class (row)

\$GR19\_10

Pr(1) Pr(2)

class 1: 0.6316 0.3684

class 2: 0.9141 0.0859

Pr(1) Pr(2)

class 1: 0.8110 0.1890

class 2: 0.9748 0.0252

\$GR19\_12

\$GR19\_11

Pr(1) Pr(2)

class 1: 0.8053 0.1947	Pr(1) Pr(2)
class 2: 0.9869 0.0131	class 1: 0.1613 0.8387
	class 2: 0.3335 0.6665
\$GR19_13	
Pr(1) Pr(2)	\$GR19_02
class 1: 0.8627 0.1373	Pr(1) Pr(2)
class 2: 0.9688 0.0312	class 1: 0.5705 0.4295
	class 2: 0.7319 0.2681
\$GR19_14	
Pr(1) Pr(2)	\$GR19_03
class 1: 0.7891 0.2109	Pr(1) Pr(2)
class 2: 0.8824 0.1176	class 1: 0.9196 0.0804
	class 2: 0.9754 0.0246
\$GR19_15	
Pr(1) Pr(2)	\$GR19_04
class 1: 0.7153 0.2847	Pr(1) Pr(2)
class 2: 0.9452 0.0548	class 1: 0.8181 0.1819
	class 2: 0.9641 0.0359
\$GR19_16	
Pr(1) Pr(2)	\$GR19_05
class 1: 0.0576 0.9424	Pr(1) Pr(2)
class 2: 0.3628 0.6372	class 1: 0.4959 0.5041
	class 2: 0.7610 0.2390
\$GR19_17	
Pr(1) Pr(2)	\$GR19_06
class 1: 0.1783 0.8217	Pr(1) Pr(2)
class 2: 0.5573 0.4427	class 1: 0.6527 0.3473
	class 2: 0.9020 0.0980
\$GR19_18	
Pr(1) Pr(2)	\$GR19_07
class 1: 0.3307 0.6693	Pr(1) Pr(2)
class 2: 0.8880 0.1120	class 1: 0.3660 0.6340
	class 2: 0.6286 0.3714
\$GR19_19	
Pr(1) Pr(2)	\$GR19_08
class 1: 0.3264 0.6736	Pr(1) Pr(2)
class 2: 0.8577 0.1423	class 1: 0.7363 0.2637
	class 2: 0.9575 0.0425
\$GR19_20	
Pr(1) Pr(2)	\$GR19_09
class 1: 0.6048 0.3952	Pr(1) Pr(2)
class 2: 0.9385 0.0615	class 1: 0.8063 0.1937
	class 2: 0.9484 0.0516
\$GR19_01	
Estimated class population shares	
0.3886 0.6114	
Predicted class memberships (by modal posterior prob.)	
0.3801 0.6199	
=====	
Fit for 2 latent classes:	
=====	
number of observations: 1339	
number of estimated parameters: 41	
residual degrees of freedom: 1298	
maximum log-likelihood: -11878.45	
AIC(2): 23838.91	
BIC(2): 24052.09	
G^2(2): 7061.427 (Likelihood ratio/deviance statistic)	
X^2(2): 286818303 (Chi-square goodness of fit)	

## 2.2 Ταξινόμηση έξι ομάδων

```
> lca6<-polCA(fr1,as.data.frame(grade6),nclass=6,maxiter=2000,nrep=10, graphs=TRUE)
```

Model 1: llik = -11198.07 ... best llik = -11198.07

Model 2: llik = -11197.38 ... best llik = -11197.38

Model 3: llik = -11200.69 ... best llik = -11197.38

Model 4: llik = -11200.69 ... best llik = -11197.38

Model 5: llik = -11197.38 ... best llik = -11197.38

Model 6: llik = -11197.38 ... best llik = -11197.38

Model 7: llik = -11197.38 ... best llik = -11197.38

Model 8: llik = -11218.86 ... best llik = -11197.38

Model 9: llik = -11218.86 ... best llik = -11197.38

Model 10: llik = -11217.34 ... best llik = -11197.38

Conditional item response (column) probabilities,  
by outcome variable, for each class (row)

\$GR19\_10

Pr(1) Pr(2)

class 1: 0.2171 0.7829

class 2: 0.9500 0.0500

class 3: 0.9597 0.0403

class 4: 0.0362 0.9638

class 5: 0.8687 0.1313

class 6: 0.8352 0.1648

\$GR19\_11

Pr(1) Pr(2)

class 1: 0.5759 0.4241

class 2: 0.9859 0.0141

class 3: 1.0000 0.0000

class 4: 0.1456 0.8544

class 5: 0.9885 0.0115

class 6: 0.9406 0.0594

\$GR19\_12

Pr(1) Pr(2)

class 1: 0.4728 0.5272

class 2: 1.0000 0.0000

class 3: 1.0000 0.0000

class 4: 0.3506 0.6494

class 5: 1.0000 0.0000

class 6: 0.9686 0.0314

\$GR19\_13

Pr(1) Pr(2)

class 1: 0.8226 0.1774

class 2: 1.0000 0.0000

class 3: 0.9686 0.0314

class 4: 0.3891 0.6109

class 5: 0.9188 0.0812

class 6: 0.9232 0.0768

\$GR19\_14

Pr(1) Pr(2)

class 1: 0.7819 0.2181

class 2: 0.9333 0.0667

class 3: 0.8975 0.1025

class 4: 0.3091 0.6909

class 5: 0.8251 0.1749

class 6: 0.8143 0.1857

\$GR19\_15

Pr(1) Pr(2)

class 1: 0.6144 0.3856

class 2: 0.9701 0.0299

class 3: 0.9321 0.0679

class 4: 0.1579 0.8421

class 5: 0.7686 0.2314

class 6: 0.8819 0.1181

\$GR19\_16

Pr(1) Pr(2)

class 1: 0.0618 0.9382

class 2: 0.6445 0.3555

class 3: 0.1529 0.8471

class 4: 0.1535 0.8465

class 5: 0.0294 0.9706

class 6: 0.1823 0.8177

\$GR19\_17

Pr(1) Pr(2)

class 1: 0.3320 0.6680

class 2: 0.8010 0.1990

class 3: 0.1227 0.8773

class 4: 0.0384 0.9616

class 5: 0.0000 1.0000

class 6: 0.5291 0.4709

\$GR19\_18

Pr(1) Pr(2)

class 1: 0.4591 0.5409

class 2: 0.9677 0.0323

class 3: 0.4386 0.5614

class 4: 0.0000 1.0000

class 5: 0.0343 0.9657

class 6: 0.9501 0.0499

\$GR19\_19

Pr(1) Pr(2)

class 1: 0.4090 0.5910

class 2: 0.9842 0.0158

class 3: 0.3229 0.6771

class 4: 0.0000 1.0000

class 5: 0.2599 0.7401

class 6: 0.9054 0.0946

\$GR19\_20

Pr(1) Pr(2)  
class 1: 0.6948 0.3052  
class 2: 1.0000 0.0000  
class 3: 0.6239 0.3761  
class 4: 0.0000 1.0000  
class 5: 0.6530 0.3470  
class 6: 0.9473 0.0527

\$GR19\_01

Pr(1) Pr(2)  
class 1: 0.2121 0.7879  
class 2: 0.5433 0.4567  
class 3: 0.2396 0.7604  
class 4: 0.1051 0.8949  
class 5: 0.1111 0.8889  
class 6: 0.1813 0.8187

\$GR19\_02

Pr(1) Pr(2)  
class 1: 0.5535 0.4465  
class 2: 0.8220 0.1780  
class 3: 0.8106 0.1894  
class 4: 0.2607 0.7393  
class 5: 0.4830 0.5170  
class 6: 0.5940 0.4060

\$GR19\_03

Pr(1) Pr(2)  
class 1: 0.9031 0.0969  
class 2: 0.9898 0.0102  
class 3: 0.9719 0.0281  
class 4: 0.5719 0.4281  
class 5: 0.9520 0.0480  
class 6: 0.9582 0.0418

\$GR19\_04

Pr(1) Pr(2)  
class 1: 0.7834 0.2166  
class 2: 0.9973 0.0027  
class 3: 0.9664 0.0336  
class 4: 0.4216 0.5784  
class 5: 0.8841 0.1159  
class 6: 0.8874 0.1126

Estimated class population shares

0.1158 0.2131 0.2252 0.019 0.1009 0.3261

Predicted class memberships (by modal posterior prob.)

0.1075 0.227 0.2166 0.0194 0.0993 0.3301

=====

Fit for 6 latent classes:

=====

number of observations: 1339

number of estimated parameters: 125

residual degrees of freedom: 1214

maximum log-likelihood: -11197.38

AIC(6): 22644.76

BIC(6): 23294.72

G^2(6): 5699.277 (Likelihood ratio/deviance statistic)

X^2(6): 2540368 (Chi-square goodness of fit)

\$GR19\_05

Pr(1) Pr(2)  
class 1: 0.5364 0.4636  
class 2: 0.9375 0.0625  
class 3: 0.8535 0.1465  
class 4: 0.1063 0.8937  
class 5: 0.3036 0.6964  
class 6: 0.5252 0.4748

\$GR19\_06

Pr(1) Pr(2)  
class 1: 0.6832 0.3168  
class 2: 0.9723 0.0277  
class 3: 0.9353 0.0647  
class 4: 0.2247 0.7753  
class 5: 0.4565 0.5435  
class 6: 0.7908 0.2092

\$GR19\_07

Pr(1) Pr(2)  
class 1: 0.4318 0.5682  
class 2: 0.8814 0.1186  
class 3: 0.6751 0.3249  
class 4: 0.0000 1.0000  
class 5: 0.2351 0.7649  
class 6: 0.3465 0.6535

\$GR19\_08

Pr(1) Pr(2)  
class 1: 0.5858 0.4142  
class 2: 0.9772 0.0228  
class 3: 0.9661 0.0339  
class 4: 0.0772 0.9228  
class 5: 0.7976 0.2024  
class 6: 0.9078 0.0922

\$GR19\_09

Pr(1) Pr(2)  
class 1: 0.6574 0.3426  
class 2: 0.9720 0.0280  
class 3: 0.9820 0.0180  
class 4: 0.0893 0.9107  
class 5: 0.9211 0.0789  
class 6: 0.9024 0.0976

### 3. Β' Γυμνασίου\_ Latent Class Analysis πέντε ομάδων

lca5<-poLCA(fr1,as.data.frame(grade8),nclass=5,maxiter=2000,nrep=10, graphs=TRUE)

Model 1: llik = -11071.08 ... best llik = -11071.08

Model 2: llik = -11126.02 ... best llik = -11071.08

Model 3: llik = -11061.42 ... best llik = -11061.42

Model 4: llik = -11061.42 ... best llik = -11061.42

Model 5: llik = -11061.42 ... best llik = -11061.42

Model 6: llik = -11061.42 ... best llik = -11061.42

Model 7: llik = -11071.08 ... best llik = -11061.42

Model 8: llik = -11071.08 ... best llik = -11061.42

Model 9: llik = -11061.42 ... best llik = -11061.42

Model 10: llik = -11126.02 ... best llik = -11061.42

Conditional item response (column) probabilities,

by outcome variable, for each class (row)

\$GR19\_10

Pr(1) Pr(2)

class 1: 0.2384 0.7616

class 2: 0.9359 0.0641

class 3: 0.0252 0.9748

class 4: 0.9489 0.0511

class 5: 0.9272 0.0728

\$GR19\_11

Pr(1) Pr(2)

class 1: 0.4259 0.5741

class 2: 0.9822 0.0178

class 3: 0.6593 0.3407

class 4: 0.9793 0.0207

class 5: 0.8746 0.1254

\$GR19\_12

Pr(1) Pr(2)

class 1: 0.4101 0.5899

class 2: 1.0000 0.0000

class 3: 0.3122 0.6878

class 4: 0.9926 0.0074

class 5: 1.0000 0.0000

\$GR19\_13

Pr(1) Pr(2)

class 1: 0.4730 0.5270

class 2: 0.9908 0.0092

class 3: 0.9207 0.0793

class 4: 0.9510 0.0490

class 5: 0.9498 0.0502

\$GR19\_14

Pr(1) Pr(2)

class 1: 0.5083 0.4917

class 2: 0.9554 0.0446

class 3: 0.9138 0.0862

class 4: 0.9270 0.0730

class 5: 0.9338 0.0662

\$GR19\_15

Pr(1) Pr(2)

class 1: 0.4194 0.5806

class 2: 0.9055 0.0945

class 3: 0.7627 0.2373

class 4: 0.9588 0.0412

class 5: 0.7148 0.2852

\$GR19\_16

Pr(1) Pr(2)

class 1: 0.0920 0.9080

class 2: 0.0733 0.9267

class 3: 0.0337 0.9663

class 4: 0.3187 0.6813

class 5: 0.0229 0.9771

\$GR19\_17

Pr(1) Pr(2)

class 1: 0.2623 0.7377

class 2: 0.0790 0.9210

class 3: 0.3843 0.6157

class 4: 0.6423 0.3577

class 5: 0.3306 0.6694

\$GR19\_18

Pr(1) Pr(2)

class 1: 0.2212 0.7788

class 2: 0.3060 0.6940

class 3: 0.6483 0.3517

class 4: 0.9634 0.0366

class 5: 0.2952 0.7048

\$GR19\_19

Pr(1) Pr(2)

class 1: 0.1742 0.8258

class 2: 0.0839 0.9161

class 3: 0.4453 0.5547

class 4: 0.8689 0.1311

class 5: 0.3276 0.6724

\$GR19\_20

Pr(1) Pr(2)

class 1: 0.5364 0.4636

class 2: 0.6234 0.3766

class 3: 0.8299 0.1701

class 4: 0.9894 0.0106

class 5: 0.7607 0.2393

\$GR19\_01

Pr(1) Pr(2)

```

class 1: 0.2758 0.7242
class 2: 0.2243 0.7757
class 3: 0.1428 0.8572
class 4: 0.3390 0.6610
class 5: 0.1522 0.8478

$GR19_02
  Pr(1) Pr(2)
class 1: 0.4633 0.5367
class 2: 0.8455 0.1545
class 3: 0.7592 0.2408
class 4: 0.8208 0.1792
class 5: 0.6173 0.3827

$GR19_03
  Pr(1) Pr(2)
class 1: 0.7146 0.2854
class 2: 0.9807 0.0193
class 3: 0.9866 0.0134
class 4: 0.9807 0.0193
class 5: 0.9692 0.0308

$GR19_04
  Pr(1) Pr(2)
class 1: 0.4178 0.5822
class 2: 0.9667 0.0333
class 3: 0.9275 0.0725
class 4: 0.9706 0.0294
class 5: 0.8687 0.1313

$GR19_05
  Pr(1) Pr(2)
class 1: 0.2154 0.7846
class 2: 0.8435 0.1565
class 3: 0.7744 0.2256
Estimated class population shares
0.0456 0.2966 0.1184 0.3556 0.1838
Predicted class memberships (by modal posterior prob.)
0.0431 0.3109 0.1153 0.3613 0.1693
=====
Fit for 5 latent classes:
=====
number of observations: 1370
number of estimated parameters: 104
residual degrees of freedom: 1266
maximum log-likelihood: -11061.42
AIC(5): 22330.83
BIC(5): 22873.98
G^2(5): 5438.17 (Likelihood ratio/deviance statistic)
X^2(5): 18813415 (Chi-square goodness of fit)

```

```

class 4: 0.7904 0.2096
class 5: 0.3302 0.6698

$GR19_06
  Pr(1) Pr(2)
class 1: 0.2503 0.7497
class 2: 0.9104 0.0896
class 3: 0.8489 0.1511
class 4: 0.9021 0.0979
class 5: 0.3465 0.6535

$GR19_07
  Pr(1) Pr(2)
class 1: 0.2805 0.7195
class 2: 0.8358 0.1642
class 3: 0.5998 0.4002
class 4: 0.7508 0.2492
class 5: 0.4956 0.5044

$GR19_08
  Pr(1) Pr(2)
class 1: 0.3637 0.6363
class 2: 0.9747 0.0253
class 3: 0.6904 0.3096
class 4: 0.9679 0.0321
class 5: 0.8446 0.1554

$GR19_09
  Pr(1) Pr(2)
class 1: 0.3133 0.6867
class 2: 0.9808 0.0192
class 3: 0.7322 0.2678
class 4: 0.9569 0.0431
class 5: 0.8853 0.1147

```

#### 4. Α' Λυκείου\_ Latent Class Analysis πέντε ομάδων

```

> lca5<-poLCA(fr1,as.data.frame(grade10),nclass=5,maxiter=2000,nrep=10, graphs=TRUE)
Model 1: llik = -9859.441 ... best llik = -9859.441
Model 2: llik = -9859.441 ... best llik = -9859.441
Model 3: llik = -9859.441 ... best llik = -9859.441
Model 4: llik = -9896.743 ... best llik = -9859.441
Model 5: llik = -9859.441 ... best llik = -9859.441
Model 6: llik = -9859.441 ... best llik = -9859.441

```

Model 7: llik = -9968.889 ... best llik = -9859.441  
 Model 8: llik = -9859.441 ... best llik = -9859.441  
 Model 9: llik = -9859.441 ... best llik = -9859.441  
 Model 10: llik = -9859.441 ... best llik = -9859.441  
 Conditional item response (column) probabilities,  
 by outcome variable, for each class (row)

\$GR19\_10

	Pr(1)	Pr(2)
class 1:	0.4620	0.5380
class 2:	0.9653	0.0347
class 3:	0.9720	0.0280
class 4:	0.0417	0.9583
class 5:	0.9701	0.0299

\$GR19\_11

	Pr(1)	Pr(2)
class 1:	0.3499	0.6501
class 2:	0.9909	0.0091
class 3:	0.9915	0.0085
class 4:	0.5311	0.4689
class 5:	0.9462	0.0538

\$GR19\_12

	Pr(1)	Pr(2)
class 1:	0.5247	0.4753
class 2:	0.9981	0.0019
class 3:	1.0000	0.0000
class 4:	0.3656	0.6344
class 5:	1.0000	0.0000

\$GR19\_13

	Pr(1)	Pr(2)
class 1:	0.5485	0.4515
class 2:	0.9908	0.0092
class 3:	0.9821	0.0179
class 4:	0.9634	0.0366
class 5:	0.8917	0.1083

\$GR19\_14

	Pr(1)	Pr(2)
class 1:	0.5956	0.4044
class 2:	0.9493	0.0507
class 3:	0.9587	0.0413
class 4:	0.9430	0.0570
class 5:	0.9660	0.0340

\$GR19\_15

	Pr(1)	Pr(2)
class 1:	0.4582	0.5418
class 2:	0.9865	0.0135
class 3:	0.8667	0.1333
class 4:	0.7887	0.2113
class 5:	0.8367	0.1633

\$GR19\_16

	Pr(1)	Pr(2)
class 1:	0.2394	0.7606
class 2:	0.2152	0.7848
class 3:	0.0532	0.9468
class 4:	0.0203	0.9797
class 5:	0.0099	0.9901

\$GR19\_17

	Pr(1)	Pr(2)
class 1:	0.4294	0.5706
class 2:	0.5319	0.4681
class 3:	0.0330	0.9670
class 4:	0.2663	0.7337
class 5:	0.4359	0.5641

\$GR19\_18

	Pr(1)	Pr(2)
class 1:	0.2566	0.7434
class 2:	0.9229	0.0771
class 3:	0.1514	0.8486
class 4:	0.5899	0.4101
class 5:	0.3995	0.6005

\$GR19\_19

	Pr(1)	Pr(2)
class 1:	0.2482	0.7518
class 2:	0.6376	0.3624
class 3:	0.0427	0.9573
class 4:	0.2838	0.7162
class 5:	0.3981	0.6019

\$GR19\_20

	Pr(1)	Pr(2)
class 1:	0.4212	0.5788
class 2:	0.9871	0.0129
class 3:	0.5839	0.4161
class 4:	0.8021	0.1979
class 5:	0.9676	0.0324

\$GR19\_01

	Pr(1)	Pr(2)
class 1:	0.3654	0.6346
class 2:	0.4127	0.5873
class 3:	0.2938	0.7062
class 4:	0.2724	0.7276
class 5:	0.4883	0.5117

\$GR19\_02

	Pr(1)	Pr(2)
class 1:	0.4211	0.5789
class 2:	0.8312	0.1688
class 3:	0.8374	0.1626
class 4:	0.8313	0.1687
class 5:	0.6041	0.3959

\$GR19\_03

	Pr(1)	Pr(2)
class 1:	0.6579	0.3421
class 2:	1.0000	0.0000
class 3:	0.9741	0.0259
class 4:	0.9863	0.0137



class 5: 0.9710 0.0290

\$GR19\_04

Pr(1) Pr(2)

class 1: 0.5011 0.4989

class 2: 0.9961 0.0039

class 3: 0.9833 0.0167

class 4: 0.9448 0.0552

class 5: 0.8441 0.1559

\$GR19\_05

Pr(1) Pr(2)

class 1: 0.3954 0.6046

class 2: 0.9308 0.0692

class 3: 0.9342 0.0658

class 4: 0.8903 0.1097

class 5: 0.6399 0.3601

\$GR19\_06

Pr(1) Pr(2)

class 1: 0.2891 0.7109

class 2: 0.9350 0.0650

class 3: 0.8036 0.1964

class 4: 0.8720 0.1280

class 5: 0.5009 0.4991

Estimated class population shares

0.0437 0.3597 0.2925 0.1088 0.1953

Predicted class memberships (by modal posterior prob.)

0.0408 0.4043 0.2752 0.1076 0.1721

=====

Fit for 5 latent classes:

=====

number of observations: 1348

number of estimated parameters: 104

residual degrees of freedom: 1244

maximum log-likelihood: -9859.441

AIC(5): 19926.88

BIC(5): 20468.35

G<sup>2</sup>(5): 4299.208 (Likelihood ratio/deviance statistic)

X<sup>2</sup>(5): 1952786 (Chi-square goodness of fit)

\$GR19\_07

Pr(1) Pr(2)

class 1: 0.4291 0.5709

class 2: 0.8459 0.1541

class 3: 0.8812 0.1188

class 4: 0.7677 0.2323

class 5: 0.6134 0.3866

\$GR19\_08

Pr(1) Pr(2)

class 1: 0.3825 0.6175

class 2: 0.9971 0.0029

class 3: 0.9747 0.0253

class 4: 0.5926 0.4074

class 5: 0.9261 0.0739

\$GR19\_09

Pr(1) Pr(2)

class 1: 0.3773 0.6227

class 2: 0.9861 0.0139

class 3: 0.9869 0.0131

class 4: 0.8593 0.1407

class 5: 0.9782 0.0218

## 5. ΣΤ' Δημοτικού\_Ταξινόμηση τεσσάρων κλάσεων βάσει του φύλου

```
> lca4sex<-poLCA(fr2,as.data.frame(grade6),nclass=4,maxiter=2000,nrep=10,
graphs=TRUE)
```

```
Model 1: llik = -11325.59 ... best llik = -11325.59
```

```
Model 2: llik = -11325.59 ... best llik = -11325.59
```

```
Model 3: llik = -11898.03 ... best llik = -11325.59
```

```
Model 4: llik = -11489.14 ... best llik = -11325.59
```

```
Model 5: llik = -11325.59 ... best llik = -11325.59
```

```
Model 6: llik = -11527.82 ... best llik = -11325.59
```

```
Model 7: llik = -11325.59 ... best llik = -11325.59
```

```
Model 8: llik = -11387.31 ... best llik = -11325.59
```

```
Model 9: llik = -11527.82 ... best llik = -11325.59
```

```
Model 10: llik = -11325.59 ... best llik = -11325.59
```

```
Conditional item response (column) probabilities,
by outcome variable, for each class (row)
```

\$GR19\_10

Pr(1) Pr(2)

class 1: 0.9410 0.0590

class 2: 0.2224 0.7776

class 3: 0.9296 0.0704

class 4: 0.7163 0.2837

\$GR19\_11

	Pr(1)	Pr(2)
class 1:	0.9831	0.0169
class 2:	0.4568	0.5432
class 3:	1.0000	0.0000
class 4:	0.8913	0.1087

\$GR19\_12

	Pr(1)	Pr(2)
class 1:	0.9978	0.0022
class 2:	0.4604	0.5396
class 3:	0.9898	0.0102
class 4:	0.9009	0.0991

\$GR19\_13

	Pr(1)	Pr(2)
class 1:	0.9891	0.0109
class 2:	0.6879	0.3121
class 3:	0.9629	0.0371
class 4:	0.9008	0.0992

\$GR19\_14

	Pr(1)	Pr(2)
class 1:	0.8967	0.1033
class 2:	0.6986	0.3014
class 3:	0.8616	0.1384
class 4:	0.8209	0.1791

\$GR19\_15

	Pr(1)	Pr(2)
class 1:	0.9588	0.0412
class 2:	0.4573	0.5427
class 3:	0.8850	0.1150
class 4:	0.8389	0.1611

\$GR19\_16

	Pr(1)	Pr(2)
class 1:	0.5102	0.4898
class 2:	0.0595	0.9405
class 3:	0.1077	0.8923
class 4:	0.1260	0.8740

\$GR19\_17

	Pr(1)	Pr(2)
class 1:	0.6805	0.3195
class 2:	0.1735	0.8265
class 3:	0.0688	0.9312
class 4:	0.4996	0.5004

\$GR19\_18

	Pr(1)	Pr(2)
class 1:	0.9621	0.0379
class 2:	0.2089	0.7911
class 3:	0.3126	0.6874
class 4:	0.8286	0.1714

\$GR19\_19

	Pr(1)	Pr(2)
class 1:	0.9525	0.0475
class 2:	0.1525	0.8475
class 3:	0.2698	0.7302
class 4:	0.8297	0.1703

\$GR19\_20

	Pr(1)	Pr(2)
class 1:	0.9711	0.0289
class 2:	0.4019	0.5981

Estimated class population shares  
0.3374 0.0952 0.2777 0.2898

class 3:	0.5957	0.4043
class 4:	0.9576	0.0424

\$GR19\_01

	Pr(1)	Pr(2)
class 1:	0.4139	0.5861
class 2:	0.1905	0.8095
class 3:	0.2013	0.7987
class 4:	0.1827	0.8173

\$GR19\_02

	Pr(1)	Pr(2)
class 1:	0.8207	0.1793
class 2:	0.5136	0.4864
class 3:	0.7416	0.2584
class 4:	0.4744	0.5256

\$GR19\_03

	Pr(1)	Pr(2)
class 1:	0.9784	0.0216
class 2:	0.8227	0.1773
class 3:	0.9636	0.0364
class 4:	0.9585	0.0415

\$GR19\_04

	Pr(1)	Pr(2)
class 1:	0.9952	0.0048
class 2:	0.6683	0.3317
class 3:	0.9530	0.0470
class 4:	0.8400	0.1600

\$GR19\_05

	Pr(1)	Pr(2)
class 1:	0.8978	0.1022
class 2:	0.4197	0.5803
class 3:	0.6967	0.3033
class 4:	0.4198	0.5802

\$GR19\_06

	Pr(1)	Pr(2)
class 1:	0.9616	0.0384
class 2:	0.5123	0.4877
class 3:	0.7959	0.2041
class 4:	0.7278	0.2722

\$GR19\_07

	Pr(1)	Pr(2)
class 1:	0.7664	0.2336
class 2:	0.3474	0.6526
class 3:	0.5387	0.4613
class 4:	0.2944	0.7056

\$GR19\_08

	Pr(1)	Pr(2)
class 1:	0.9748	0.0252
class 2:	0.4135	0.5865
class 3:	0.9303	0.0697
class 4:	0.8454	0.1546

\$GR19\_09

	Pr(1)	Pr(2)
class 1:	0.9642	0.0358
class 2:	0.5096	0.4904
class 3:	0.9849	0.0151
class 4:	0.8486	0.1514

Predicted class memberships (by modal posterior prob.)  
 0.3495 0.0941 0.2778 0.2786  
 Fit for 4 latent classes:

=====					
2 / 1					
	Coefficient	Std. error	t value	Pr(> t )	
(Intercept)	3.03943	0.40937	7.425	0	
M1	-1.83289	0.23310	-7.863	0	
=====					
3 / 1					
	Coefficient	Std. error	t value	Pr(> t )	
(Intercept)	2.44418	0.42994	5.685	0	
M1	-1.54483	0.24400	-6.331	0	
=====					
4 / 1					
	Coefficient	Std. error	t value	Pr(> t )	
(Intercept)	0.61431	0.51331	1.197	0.232	
M1	-1.03474	0.29410	-3.518	0.000	
=====					

number of observations: 1339  
 number of estimated parameters: 86  
 residual degrees of freedom: 1253  
 maximum log-likelihood: -11325.59  
 AIC(4): 22823.18  
 BIC(4): 23270.36  
 $\chi^2(4)$ : 2060726 (Chi-square goodness of fit)

## 6. Β' Γυμνασίου\_Ταξινόμηση τεσσάρων κλάσεων βάσει του φύλου

```
> probs.start.new <-
poLCA.reorder(lca4sex$probs.start,order(lca4sex$P,decreasing=TRUE))
> lca4sex <-
poLCA(fr2,as.data.frame(grade8),nclass=4,probs.start=probs.start.new,
graphs=TRUE)
Conditional item response (column) probabilities,
by outcome variable, for each class (row)
```

\$GR19_10				Pr(1)	Pr(2)
class 1:	0.9330	0.0670	class 1:	0.9588	0.0412
class 2:	0.9521	0.0479	class 2:	0.9274	0.0726
class 3:	0.5886	0.4114	class 3:	0.7603	0.2397
class 4:	0.0000	1.0000	class 4:	0.8994	0.1006
\$GR19_15					
class 1:	0.9705	0.0295	class 1:	0.8824	0.1176
class 2:	0.9759	0.0241	class 2:	0.9499	0.0501
class 3:	0.6386	0.3614	class 3:	0.5209	0.4791
class 4:	0.6525	0.3475	class 4:	0.7650	0.2350
\$GR19_16					
class 1:	1.0000	0.0000	class 1:	0.0598	0.9402
class 2:	0.9923	0.0077	class 2:	0.3036	0.6964
class 3:	0.7391	0.2609	class 3:	0.0458	0.9542
class 4:	0.2806	0.7194	class 4:	0.0404	0.9596
\$GR19_17					
class 1:	0.9888	0.0112	class 1:	0.1071	0.8929
class 2:	0.9505	0.0495	class 2:	0.6453	0.3547
class 3:	0.7489	0.2511	class 3:	0.3573	0.6427
class 4:	0.9186	0.0814	class 4:	0.3922	0.6078
\$GR19_18					
			Pr(1)	Pr(2)	

class 1:	0.2888	0.7112			Pr(1)	Pr(2)
class 2:	0.9548	0.0452			class 1:	0.9628 0.0372
class 3:	0.2352	0.7648			class 2:	0.9644 0.0356
class 4:	0.6932	0.3068			class 3:	0.6362 0.3638
					class 4:	0.9408 0.0592

\$GR19\_19

		Pr(1)	Pr(2)
class 1:	0.1302	0.8698	
class 2:	0.8586	0.1414	
class 3:	0.2628	0.7372	
class 4:	0.4614	0.5386	

\$GR19\_20

		Pr(1)	Pr(2)
class 1:	0.6583	0.3417	
class 2:	0.9901	0.0099	
class 3:	0.6323	0.3677	
class 4:	0.8464	0.1536	

\$GR19\_01

		Pr(1)	Pr(2)
class 1:	0.2003	0.7997	
class 2:	0.3304	0.6696	
class 3:	0.2262	0.7738	
class 4:	0.1260	0.8740	

\$GR19\_02

		Pr(1)	Pr(2)
class 1:	0.8152	0.1848	
class 2:	0.8089	0.1911	
class 3:	0.4628	0.5372	
class 4:	0.7981	0.2019	

\$GR19\_03

		Pr(1)	Pr(2)
class 1:	0.9826	0.0174	
class 2:	0.9808	0.0192	
class 3:	0.8624	0.1376	
class 4:	0.9846	0.0154	

\$GR19\_04

Estimated class population shares  
0.3954 0.3751 0.1245 0.105  
Predicted class memberships (by modal posterior prob.)  
0.4 0.3803 0.1182 0.1015

		Pr(1)	Pr(2)
class 1:	0.7281	0.2719	
class 2:	0.7657	0.2343	
class 3:	0.2621	0.7379	
class 4:	0.8160	0.1840	

\$GR19\_05

		Pr(1)	Pr(2)
class 1:	0.7906	0.2094	
class 2:	0.8736	0.1264	
class 3:	0.2671	0.7329	
class 4:	0.8848	0.1152	

\$GR19\_06

		Pr(1)	Pr(2)
class 1:	0.7729	0.2271	
class 2:	0.7355	0.2645	
class 3:	0.3603	0.6397	
class 4:	0.6040	0.3960	

\$GR19\_07

		Pr(1)	Pr(2)
class 1:	0.9695	0.0305	
class 2:	0.9604	0.0396	
class 3:	0.5738	0.4262	
class 4:	0.6841	0.3159	

\$GR19\_08

		Pr(1)	Pr(2)
class 1:	0.9698	0.0302	
class 2:	0.9524	0.0476	
class 3:	0.6122	0.3878	
class 4:	0.7425	0.2575	

\$GR19\_09

Fit for 4 latent classes:

```
=====
2 / 1
      Coefficient  Std. error  t value  Pr(>|t|)
(Intercept)    1.53962    0.29356    5.245    0
M1             -1.28986    0.23167   -5.568    0
=====
3 / 1
      Coefficient  Std. error  t value  Pr(>|t|)
(Intercept)   -4.51971    0.71756   -6.299    0
M1              2.41760    0.36563    6.612    0
=====
4 / 1
      Coefficient  Std. error  t value  Pr(>|t|)
(Intercept)   -0.67420    0.33205   -2.03    0.043
M1             -0.04557    0.21746   -0.21    0.834
=====
```

number of observations: 1370  
number of estimated parameters: 86  
residual degrees of freedom: 1284  
maximum log-likelihood: -11159.33  
AIC(4): 22490.66

BIC(4): 22939.8  
 $\chi^2(4)$ : 14097564 (Chi-square goodness of fit)

## 7. Α' Λυκείου\_ Ταξινόμηση τεσσάρων κλάσεων βάσει του φύλου

```
probs.start.new <-
poLCA.reorder(lca4sex$probs.start,order(lca4sex$P,decreasing=TRUE))
> lca4sex <-
poLCA(fr2,as.data.frame(grade10),nclass=4,probs.start=probs.start.new,
graphs=TRUE)
Conditional item response (column) probabilities,
by outcome variable, for each class (row)
```

\$GR19_10			\$GR19_18		
	Pr(1)	Pr(2)		Pr(1)	Pr(2)
class 1:	0.9552	0.0448	class 1:	0.9456	0.0544
class 2:	0.9589	0.0411	class 2:	0.3205	0.6795
class 3:	0.9793	0.0207	class 3:	0.2423	0.7577
class 4:	0.1937	0.8063	class 4:	0.4958	0.5042
\$GR19_11			\$GR19_19		
	Pr(1)	Pr(2)		Pr(1)	Pr(2)
class 1:	0.9825	0.0175	class 1:	0.7250	0.2750
class 2:	0.9933	0.0067	class 2:	0.0838	0.9162
class 3:	0.9668	0.0332	class 3:	0.2347	0.7653
class 4:	0.4807	0.5193	class 4:	0.2820	0.7180
\$GR19_12			\$GR19_20		
	Pr(1)	Pr(2)		Pr(1)	Pr(2)
class 1:	1.0000	0.0000	class 1:	0.9929	0.0071
class 2:	1.0000	0.0000	class 2:	0.6178	0.3822
class 3:	1.0000	0.0000	class 3:	0.8790	0.1210
class 4:	0.4097	0.5903	class 4:	0.7036	0.2964
\$GR19_13			\$GR19_01		
	Pr(1)	Pr(2)		Pr(1)	Pr(2)
class 1:	0.9840	0.0160	class 1:	0.4363	0.5637
class 2:	0.9916	0.0084	class 2:	0.2262	0.7738
class 3:	0.9125	0.0875	class 3:	0.4909	0.5091
class 4:	0.8481	0.1519	class 4:	0.3060	0.6940
\$GR19_14			\$GR19_02		
	Pr(1)	Pr(2)		Pr(1)	Pr(2)
class 1:	0.9482	0.0518	class 1:	0.8099	0.1901
class 2:	0.9638	0.0362	class 2:	0.9430	0.0570
class 3:	0.9629	0.0371	class 3:	0.5796	0.4204
class 4:	0.8377	0.1623	class 4:	0.7172	0.2828
\$GR19_15			\$GR19_03		
	Pr(1)	Pr(2)		Pr(1)	Pr(2)
class 1:	0.9855	0.0145	class 1:	1.0000	0.0000
class 2:	0.9163	0.0837	class 2:	0.9822	0.0178
class 3:	0.8098	0.1902	class 3:	0.9677	0.0323
class 4:	0.6983	0.3017	class 4:	0.8918	0.1082
\$GR19_16			\$GR19_04		
	Pr(1)	Pr(2)		Pr(1)	Pr(2)
class 1:	0.2300	0.7700	class 1:	0.9828	0.0172
class 2:	0.0570	0.9430	class 2:	0.9868	0.0132
class 3:	0.0180	0.9820	class 3:	0.9024	0.0976
class 4:	0.0871	0.9129	class 4:	0.8035	0.1965
\$GR19_17			\$GR19_05		
	Pr(1)	Pr(2)		Pr(1)	Pr(2)
class 1:	0.6114	0.3886	class 1:	0.9073	0.0927
class 2:	0.0168	0.9832	class 2:	0.9579	0.0421
class 3:	0.3193	0.6807	class 3:	0.7199	0.2801
class 4:	0.3145	0.6855	class 4:	0.7462	0.2538
			\$GR19_06		

	Pr(1)	Pr(2)
class 1:	0.9149	0.0851
class 2:	0.9119	0.0881
class 3:	0.5162	0.4838
class 4:	0.6948	0.3052

\$GR19\_07

	Pr(1)	Pr(2)
class 1:	0.8347	0.1653
class 2:	0.8935	0.1065
class 3:	0.6840	0.3160
class 4:	0.6596	0.3404

\$GR19\_08

	Pr(1)	Pr(2)
class 1:	0.9873	0.0127
class 2:	0.9680	0.0320
class 3:	0.9710	0.0290
class 4:	0.5163	0.4837

\$GR19\_09

	Pr(1)	Pr(2)
class 1:	0.9862	0.0138
class 2:	0.9877	0.0123
class 3:	0.9784	0.0216
class 4:	0.7240	0.2760

Estimated class population shares  
0.3242 0.2626 0.26 0.1533

Predicted class memberships (by modal posterior prob.)  
0.3264 0.2648 0.2582 0.1506

Fit for 4 latent classes:

```
=====
2 / 1
      Coefficient  Std. error  t value  Pr(>|t|)
(Intercept)      2.42450    0.35273    6.873    0
M1               -1.99317    0.28973   -6.880    0
=====
3 / 1
      Coefficient  Std. error  t value  Pr(>|t|)
(Intercept)     -3.87812    0.65051   -5.962    0
M1               2.08003    0.33348    6.237    0
=====
4 / 1
      Coefficient  Std. error  t value  Pr(>|t|)
(Intercept)      0.26063    0.32921    0.792    0.429
M1              -0.68992    0.21889   -3.152    0.002
=====
```

number of observations: 1348  
number of estimated parameters: 86  
residual degrees of freedom: 1262  
maximum log-likelihood: -9904.603

AIC(4): 19981.21  
BIC(4): 20428.95  
 $\chi^2(4)$ : 62667061 (Chi-square goodness of fit)

## **ΒΙΒΛΙΟΓΡΑΦΙΑ**

- Berlin, K. S., Williams, N. A., & Parra, G. R. (2014). An Introduction to Latent Variable Mixture Modeling (Part 1): Overview and Cross-Sectional Latent Class and Latent Profile Analysis. *Journal of Pediatric Psychology* 39, 174-187.
- Genge, E. (2014). A latent class analysis of the public attitude towards the euro adoption in Poland. *Advances in Data Analysis and Classification* 8, 427-442.
- Goodman, L. A. (2002). Latent class analysis: The empirical study of latent types, latent variables, and latent structures. In: J.A. Hagenaars & A.L. McCutcheon (Eds.). *Applied Latent Class Analysis*. Cambridge: Cambridge University.
- Goodman, L. A. (1978). *Analyzing Qualitative / Categorical Data: Log-linear Models and Latent Structure Analysis*. Cambridge, MA: Abt Books.
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.
- Hagenaars, J., & McCutcheon, A. (Eds) (2002). *Applied latent class analysis*. Cambridge, UK: Cambridge University Press.
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent Structure Analysis*. Boston: Houghton Mifflin.
- Lazarsfeld, P. F. (1950b). "Some Latent Structures", Chapter 11 in Stouffer (1950).
- Lazarsfeld, P. F. (1950a). "The Logical and Mathematical Foundations of Latent Structure Analysis", Chapter 10 in Stouffer (1950).
- Linzer, D. A., & Lewis, J. B. (2011). polCA: An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software*, 42 (10).
- Palma, M. A., Ness, M. L., & Anderson, D. P. (2015). *Buying More than Taste? A Latent Class Analysis of Health and Prestige Determinants of Healthy Food*. Rotorua, New Zealand: Australian Agricultural & Resource Economics Society.
- Torfs, P., & Brauer, C. (2014). *A (very) short introduction to R*. The Netherlands: Hydrology and Quantitative Water Management Group, Wageningen University.
- Vermunt, J. K., & Magidson, J. (2004). Latent Class Analysis. *The Sage Handbook of Quantitative Methodology for the Social Sciences*, pp. 175-198. Sage Publications: Thousand Oaks, CA.
- Weijters, B., & Goedertier, F. (2015). Understanding today's music acquisition mix: A latent class analysis of consumers' combined use of music platforms. *Marketing Letters*, 1-8.