**Using LCA to group students pathways throughout the course (weeks 2-13)**

LCA was performed on the dataset (N = 290) consisting of 12 categorical variables, one for each week of the course (week 2 – week 13), representing the cluster each student belonged to in the given week. Variables for weeks 5, 7, 9, 11 and 12 have 4 distinct values, as 4 clusters were detected for those weeks; variables for the other 6 weeks have 5 different values, matching the 5 identified clusters.

To avoid running into local maximum, the model building process was repeated 50 times (using different initial parameter values) for each considered number of classes (3-6). Table 1 shows evaluation metrics. According to the AIC metric, the solution with nclass=5 is the best one

Table 1. Evaluation metrics for different number of classes

====== ======== ======== ========= =========

nclass AIC BIC LogLike ChiSquare

====== ======== ======== ========= =========

3 8481.625 8962.379 -4109.812 55511602

4 8414.179 9056.408 -4032.089 37945574

**5 8397.962 9201.666 -3979.981 23917483**

6 8406.792 9371.971 -3940.396 17088019

====== ======== ======== ========= =========

**Results for the solution with 5 classes**

**Interpretation of the classes**

The most probable path for students in **Class 1** (30.8%):

w2: C5(0.43)|C4(0.29) -> w3: C4(0.32)|C3(0.30)|C1(0.25) -> w4: C3(0.51)|C4(0.31) -> w5: C2(0.61) -> w6: C1(0.51)| C3(0.23) -> w7: C3(0.53)|C2(0.46) -> w8: C5(0.41)|C1(0.37) -> w9: C4(0.61)|C1(0.36) -> w10: C1(0.57)|C5(0.27) -> w11: C3(0.76) -> w12: C1(0.63)|C4(0.28) -> w13: C1(0.36)|C4(0.35)

The most probable path for students in **Class 2** (20.9%):

w2: C5(0.38)|C2(0.32) -> w3: C3(0.48)|C1(0.23) -> w4: C2(0.61)|C1(0.26) -> w5: C1(0.6)|C2(0.29) -> w6: C1(0.54) -> w7: C1(0.38)|C3(0.36) -> w8: C3(0.5) -> w9: C1(0.46)|C2(0.32) -> w10: C1(0.48)|C3(0.2) -> w11: C2(0.34)|C3(0.32) -> w12: C1(0.53)|C3(0.46) -> w13: C4(0.42)|C1(0.32)

The most probable path for students in **Class 3** (20.8%):

w2: C1(0.42)|C5(0.24) -> w3: C5(0.36)|C1(0.28) -> w4: C4(0.33)|C5(0.32) -> w5: C4(0.59) -> w6: C3(0.54)|C1(0.35) -> w7: C3(0.88) -> w8: C4(0.42) -> w9: C1(0.84) -> w10: C2(0.51)|C1(0.32) -> w11: C4(0.47)|C3(0.27) -> w12: C3(0.66) -> w13: C1(0.6)|C3(0.23)

The most probable path for students in **Class 4** (15.1%):

w2: C2(0.45)|C5(0.41) -> w3: C3(0.72) -> w4: C2(0.58) -> w5: C1(0.62) -> w6: C1(0.37)|C4(0.34) -> w7: C1(0.71) -> w8: C2(0.75) -> w9: C2(0.78) -> w10: C3(0.65) -> w11: C1(0.69)|C2(0.26) -> w12: C2(0.83) -> w13: C4(0.49)|C3(0.23)

The most probable path for students in **Class 5** (12.4%):

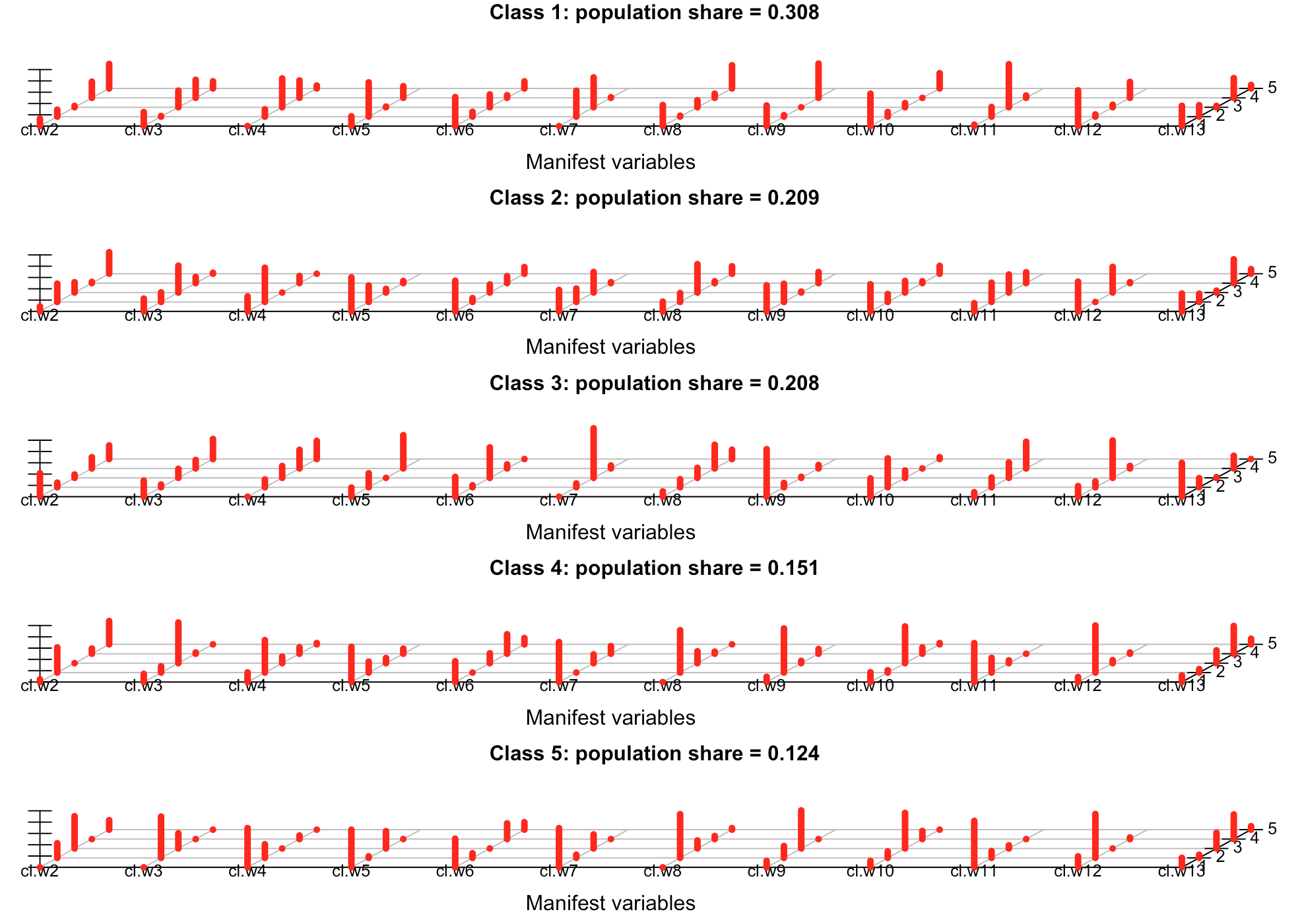
w2: C3(0.57)|C2(0.26) -> w3: C2(0.73)|C3(0.27) -> w4: C1(0.69)|C2(0.24) -> w5: C1(0.66)|C3(0.31) ->w6: C1(0.5)|C4(0.28) -> w7: C1(0.69)|C3(0.25) -> w8: C2(0.77) -> w9: C3(0.67)|C2(0.21) -> w10: C3(0.63) -> w11: C1(0.82) -> w12: C2(0.78) -> w13: C4(0.44)|C3(0.28)

**INTERPRETATION BASED ON CLUSTERS IDENTIFIED WITH ABELARDO**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **CLASS 1 (30.8%)** | **CLASS 2 (20.9)** | **CLASS 3 (20.8%)** | **CLASS 4 (15.1%)** | **CLASS 5 (12.4%)** |
| **WEEK 2** | C|B | C|Eff | A|C | Eff|C | Top|Eff |
| **WEEK 3** | B|Eff|C | Eff|C | A|C | Eff | Top|Eff |
| **WEEK 4** | B|B’ | C|Top | B’|A | C | Top|C |
| **WEEK 5** | B | Eff|B | A | Eff | Eff|Top |
| **WEEK 6** | C’|A | C’ | A|C’ | C’|Eff | C’|Eff |
| **WEEK 7** | C|B | Top|C | C | Top | Top|C |
| **WEEK 8** | B’|B | C | A | Top | Top |
| **WEEK 9** | B|B’ | B’|Eff | B’ | Eff | Top|Eff |
| **WEEK 10** | B’|B | B’|Eff | A|B’ | Eff | Eff |
| **WEEK 11** | B | C|B | F|B | Top|C | Top |
| **WEEK 12** | B’|B | B’|F | F | Top | Top |
| **WEEK 13** | A|C’ | C’|A | A|Eff | C’|Eff | C’|Eff |

**The abbreviations used in the class-comparisons table**:

* A – disengaged
* B, B’ – gaming/guessing assessment
* C, C’ – low engaged, assessment driven
* D (Top) - engage frequently in all kinds of activities
* E (Eff) – engaged and effective/efficient
* F - cheaters



**INITIAL INTERPRETATION**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **CLASS 1 (30.8%)** | **CLASS 2 (20.9)** | **CLASS 3 (20.8%)** | **CLASS 4 (15.1%)** | **CLASS 5 (12.4%)** |
| **w2** | MA\_ALL-ACT\_BAL |  SAS-ONLY\_BAL | MA\_ALL-ACT\_BAL|  A\_POS-DIFF | DENG|  MA\_ALL-ACT\_BAL | A\_POS-DIFF|  MA\_ALL-ACT\_BAL | A\_POS-DIFF\_MC|  A\_POS-DIFF |
| **w3** | SAS-ONLY\_NEG-DIFF| A\_NEG-DIFF |  SAS-MOSTLY\_NEG-DIFF | A\_NEG-DIFF|  SAS-MOSTLY\_NEG-DIFF | DENG|  SAS-MOSTLY\_NEG-DIFF | A\_NEG-DIFF | HA\_NEG-DIFF\_MC|  A\_NEG-DIFF |
| **w4** | SAS-ONLY\_NEG-DIFF|  SAS-MOSTLY\_NEG-DIFF | A\_NEG-DIFF|  HA\_BAL\_MC | SAS-MOSTLY\_NEG-DIFF|  DENG | A\_NEG-DIFF | HA\_BAL\_MC|  A\_NEG-DIFF |
| **w5** | SAS-ONLY\_NEG-DIFF | MA\_ALL-ACT\_BAL|  SAS-ONLY\_NEG-DIFF | SAS-ONLY\_BAL | MA\_ALL-ACT\_BAL | MA\_ALL-ACT\_BAL|  HA\_NEG-DIFF\_MC |
| **w6** | A\_POS-DIFF|  DENG | A\_POS-DIFF | DENG|  A\_POS-DIFF | A\_POS-DIFF|  HA\_BAL\_MC | A\_POS-DIFF|  HA\_BAL\_MC |
| **w7** | SAS-ONLY\_BAL|  SAS-ONLY\_NEG-DIFF | A\_NEG-DIFF|  SAS-ONLY\_BAL | SAS-ONLY\_BAL | A\_NEG-DIFF | A\_NEG-DIFF|  SAS-ONLY\_BAL |
| **w8** | SAS-ONLY\_NEG-DIFF | SAS-MOSTLY\_NEG-DIFF | SAS-ONLY\_BAL | MA\_ALL-ACT\_BAL | MA\_ALL-ACT\_BAL |
| **w9** | SAS-ONLY\_NEG-DIFF | SAS-ONLY\_NEG-DIFF|  A\_NEG-DIFF | SAS-ONLY\_NEG-DIFF | A\_NEG-DIFF | HA\_NEG-DIFF\_MC|  A\_NEG-DIFF |
| **w10** | SAS-ONLY\_NEG-DIFF | SAS-ONLY\_NEG-DIFF|  MA\_ALL-ACT\_BAL | SAS-ONLY\_NEG-DIFF|  DENG | MA\_ALL-ACT\_BAL | MA\_ALL-ACT\_BAL |
| **w11** | SAS-ONLY\_NEG-DIFF | SAS-ONLY\_NEG-DIFF | SAS-ONLY\_BAL|  SAS-ONLY\_NEG-DIFF | A\_NEG-DIFF|  SAS-ONLY\_NEG-DIFF | A\_NEG-DIFF |
| **w12** | SAS-ONLY\_NEG-DIFF | SAS-ONLY\_NEG-DIFF|  SAS-ONLY\_BAL | SAS-ONLY\_BAL | MA\_ALL-ACT\_BAL | MA\_ALL-ACT\_BAL |
| **w13** | SAS-ONLY\_BAL|  A\_POS-DIFF | A\_POS-DIFF|  SAS-ONLY\_BAL | SAS-ONLY\_BAL|  HA\_BAL | A\_POS-DIFF|  HA\_BAL | A\_POS-DIFF|  HA\_BAL |
| **FE\_TOT** | 14.5 (11.0 19.0) | 19.0 (15.0 27.5) | 16.0 (11.5 22.5) | 24.0 (15.75 34.0) | 18.0 (14.0 29.5) |

**The abbreviations used in the class-comparisons table**

|  |  |
| --- | --- |
| HA\_BAL | Highly active; good balance is kept between correct and incorrect solutions on all forms of assessment |
| HA\_BAL\_MC | Highly active; good balance is kept between correct and incorrect solutions on all forms of assessment; (meta-cog) evaluation activities are also present |
| HA\_NEG-DIFF\_MC | Highly active; incorrect solutions outnumber correct ones *on summative assessment*; (meta-cog) evaluation activities are also present |
| A\_POS-DIFF | Active, engaged in all kinds of course activities; correct solutions outnumber incorrect ones on all forms of assessment |
| A\_NEG-DIFF | Active, engaged in all kinds of course activities; incorrect solutions outnumber correct ones *on summative assessment* |
| MA\_ALL-ACT\_BAL | Moderately active; engaged in all kinds of course activities; good balance is kept between correct and incorrect solutions on all forms of assessment |
| MA\_MANY-ACT\_BAL\_MC | Moderately active; engaged in many (but not all) kinds of course activities; good balance is kept between correct and incorrect solutions on all forms of assessment; meta-cog activities are also present |
| SAS-ONLY\_BAL | Restricted to summative assessment; balance is kept between correct and incorrect solutions |
| SAS-ONLY\_NEG-DIFF | Restricted to summative assessment; incorrect solutions outnumber correct ones |
| SAS-MOSTLY\_NEG-DIFF | Mostly focused on summative assessment, though low level of other activities is also present incorrect solutions outnumber correct ones *on summative assessment* |
| DENG | Disengaged |

**Comparison of LCA classes based on the students’ final exam score**

Since data about students’ final exam score are not normally distributed, non-parametric tests were performed: Kruskal-Wallis test followed by Mann-Whitney U test for pair-wise comparison.

Descriptive statistics

===== === ====== ===== ====

class N median Q1 Q3

===== === ====== ===== ====

1 92 14.5 11.00 19.0

2 59 19.0 15.00 27.5

3 59 16.0 11.50 22.5

4 44 24.0 15.75 34.0

5 36 18.0 14.00 29.5

===== === ====== ===== ====

Pairwise comparisons with the FDR correction

=== === === ======= ======== =========== ===========

\ c1 c2 Z p effect.size significant

=== === === ======= ======== =========== ===========

2 1 4 -4.9556 0.000000 0.2910 YES

1 1 2 -4.5426 0.000004 0.2668 YES

7 3 4 -3.2462 0.001025 0.1906 YES

3 1 5 -3.2086 0.001171 0.1884 YES

4 2 3 2.3734 0.017300 0.1394 YES

8 3 5 -1.8052 0.071220 0.1060 NO

5 2 4 -1.2683 0.206247 0.0745 NO

9 4 5 1.2591 0.210137 0.0739 NO

=== === === ======= ======== =========== ===========

To conclude, significant difference with respect to the final exam score are detected between the following pairs of classes: 1 – 2, 1 – 4, 1 – 5, 2 – 3, and 3 – 4.

**Comparison of LCA classes based on the students’ scores on the MSLQ and SPQ questionnaires**

Abelardo’s dataset contains the following 5 variables computed based on the students’ responses to the Motivated Strategies for Learning Questionnaire (MSLQ):

* MSLQ\_IVAL: Intrinsic Value
* MSLQ\_SEFF: Self-efficacy
* MSLQ\_TANX: Test anxiety
* MSLQ\_CSUS: Cognitive Strategy Use
* MSLQ\_SREL: Self regulation

The five LCA classes are compared based on these 5 variables. For all the variables except one (Test anxiety, MSLQ\_TANX), conditions were satisfied for the use of one-way ANOVA, so, this test was used to compare the group means; only for the MSLQ\_TANX, Kruskal-Wallis test was used. However, no statistically significant difference was found for any of these 5 variables.

Note that out of 290 students for whom we have log data and who were subjects of the analyses reported in this document, 144 students filled in the MSLQ questionnaire, so, the reported finding that there is no stat. significant difference between LCA classes is not fully trustworthy.

The dataset also contains the following 6 variables computed based on the students’ responses to the Study Process Questionnaire (SPQ):

* SPQ\_DA: Deep approach
* SPQ\_SA: Surface approach
* SPQ\_DM: Deep motive
* SPQ\_SM: Surface motive
* SPQ\_DS: Deep strategy
* SPQ\_SS: Surface strategy

None of the conducted one-way ANOVA tests (conditions for its use were satisfied for all the variables) revealed statistically significant difference among the LCA classes.

This test was filled out by the same group of students as the MSLQ test, so, here, again, we are missing responses from 146 students (out of 290 for whom log data is available).

Raw output obtained from poLCA

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

$cl.w2

Pr(1) Pr(2) Pr(3) Pr(4) Pr(5)

class 1: 0.1321 0.1238 0.0246 0.2864 0.4331

class 2: 0.0883 0.3243 0.1832 0.0226 0.3817

class 3: 0.4159 0.0826 0.0594 0.2007 0.2413

class 4: 0.0472 0.4464 0.0000 0.0933 0.4131

class 5: 0.0000 0.2600 0.5718 0.0000 0.1683

$cl.w3

Pr(1) Pr(2) Pr(3) Pr(4) Pr(5)

class 1: 0.2482 0.0114 0.2953 0.3190 0.1262

class 2: 0.2266 0.1699 0.4768 0.1090 0.0178

class 3: 0.2844 0.0491 0.1604 0.1506 0.3555

class 4: 0.1447 0.1112 0.7207 0.0234 0.0000

class 5: 0.0000 0.7314 0.2686 0.0000 0.0000

$cl.w4

Pr(1) Pr(2) Pr(3) Pr(4) Pr(5)

class 1: 0.0000 0.1309 0.5099 0.3085 0.0507

class 2: 0.2658 0.6065 0.0000 0.1277 0.0000

class 3: 0.0000 0.1387 0.2101 0.3297 0.3214

class 4: 0.1695 0.5751 0.1166 0.1155 0.0233

class 5: 0.6946 0.2421 0.0000 0.0633 0.0000

$cl.w5

Pr(1) Pr(2) Pr(3) Pr(4)

class 1: 0.1737 0.6079 0.0116 0.2068

class 2: 0.6027 0.2922 0.0658 0.0394

class 3: 0.1614 0.2483 0.0000 0.5902

class 4: 0.6232 0.1940 0.0919 0.0909

class 5: 0.6658 0.0278 0.3064 0.0000

$cl.w6

Pr(1) Pr(2) Pr(3) Pr(4) Pr(5)

class 1: 0.5134 0.0865 0.2276 0.0468 0.1256

class 2: 0.5429 0.0693 0.1441 0.1243 0.1194

class 3: 0.3510 0.0369 0.5397 0.0723 0.0000

class 4: 0.3677 0.0000 0.1749 0.3454 0.1120

class 5: 0.5059 0.0839 0.0000 0.2776 0.1326

$cl.w7

Pr(1) Pr(2) Pr(3) Pr(4)

class 1: 0.0000 0.4599 0.5289 0.0112

class 2: 0.3805 0.2375 0.3646 0.0174

class 3: 0.0000 0.0692 0.8782 0.0526

class 4: 0.7072 0.0000 0.1561 0.1367

class 5: 0.6942 0.0566 0.2492 0.0000

$cl.w8

Pr(1) Pr(2) Pr(3) Pr(4) Pr(5)

class 1: 0.3735 0.0175 0.1245 0.0738 0.4107

class 2: 0.1723 0.1583 0.5033 0.0332 0.1329

class 3: 0.0936 0.1439 0.1771 0.4199 0.1656

class 4: 0.0000 0.7501 0.2112 0.0387 0.0000

class 5: 0.0000 0.7749 0.1424 0.0548 0.0279

$cl.w9

Pr(1) Pr(2) Pr(3) Pr(4)

class 1: 0.3649 0.0281 0.0000 0.6070

class 2: 0.4627 0.3232 0.0167 0.1974

class 3: 0.8413 0.0778 0.0195 0.0614

class 4: 0.0932 0.7816 0.0411 0.0841

class 5: 0.1135 0.2118 0.6747 0.0000

$cl.w10

Pr(1) Pr(2) Pr(3) Pr(4) Pr(5)

class 1: 0.5738 0.0827 0.0737 0.0000 0.2698

class 2: 0.4812 0.1404 0.2057 0.0337 0.1390

class 3: 0.3248 0.5123 0.1297 0.0000 0.0331

class 4: 0.1731 0.0410 0.6506 0.1121 0.0232

class 5: 0.1123 0.1144 0.6297 0.1436 0.0000

$cl.w11

Pr(1) Pr(2) Pr(3) Pr(4)

class 1: 0.0247 0.1714 0.7620 0.0419

class 2: 0.1456 0.3418 0.3194 0.1932

class 3: 0.0794 0.1748 0.2733 0.4726

class 4: 0.6884 0.2612 0.0504 0.0000

class 5: 0.8228 0.1214 0.0557 0.0000

$cl.w12

Pr(1) Pr(2) Pr(3) Pr(4)

class 1: 0.6319 0.0399 0.0458 0.2824

class 2: 0.5266 0.0000 0.4555 0.0179

class 3: 0.1886 0.1036 0.6621 0.0457

class 4: 0.0886 0.8336 0.0543 0.0234

class 5: 0.1944 0.7769 0.0000 0.0287

$cl.w13

Pr(1) Pr(2) Pr(3) Pr(4) Pr(5)

class 1: 0.3592 0.2040 0.0247 0.3504 0.0618

class 2: 0.3160 0.1464 0.0341 0.4220 0.0816

class 3: 0.5987 0.1615 0.0135 0.2263 0.0000

class 4: 0.1179 0.0612 0.2280 0.4927 0.1001

class 5: 0.1682 0.0558 0.2778 0.4367 0.0615

Estimated class population shares

0.3079 0.2091 0.208 0.1514 0.1236

Predicted class memberships (by modal posterior prob.)

0.3172 0.2034 0.2034 0.1517 0.1241

=========================================================

Fit for 5 latent classes:

=========================================================

number of observations: 290

number of fully observed cases: 239

number of estimated parameters: 219

residual degrees of freedom: 71

maximum log-likelihood: -3981.609

AIC(5): 8401.219

BIC(5): 9204.922

G^2(5): 4115.308 (Likelihood ratio/deviance statistic)

X^2(5): 25077250 (Chi-square goodness of fit)