**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. The obtained metric values are inconclusive regarding the best number of classes.

Table 1. Evaluation metrics for different number of classes

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

nclass AIC BIC LogLike ChiSquare

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

3 8644.280 9125.035 -4191.140 70374479

4 8574.429 9216.658 -4112.215 41807103

5 8549.295 9352.999 -4055.647 22750311

6 8533.964 9499.143 -4003.982 39945034

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

**Results for the solution with 4 classes**

**Interpretation of the classes**

The most probable path for students in **Class 1** (29.2%):

W2: C5(0.35)|C4(0.33) -> W3: C4(0.4)|C1(0.31)|C5(0.24) -> W4: C3(0.45)|C5(0.25)|C2(0.24) -> W5: C2(0.53)|C4(0.35) -> W6: C1(0.45)|C2(0.32) -> W7: C2(0.56)|C4(0.29) -> W8: C3(0.4)|C1(0.39) -> W9: C4(0.56)|C2(0.41) -> W10: C1(0.67)|C4(0.24) -> W11: C2(0.84) -> W12: C4(0.51)|C3(0.44) -> W13: C1(0.54)|C4(0.21)

The most probable path for students in **Class 2** (27.7%):

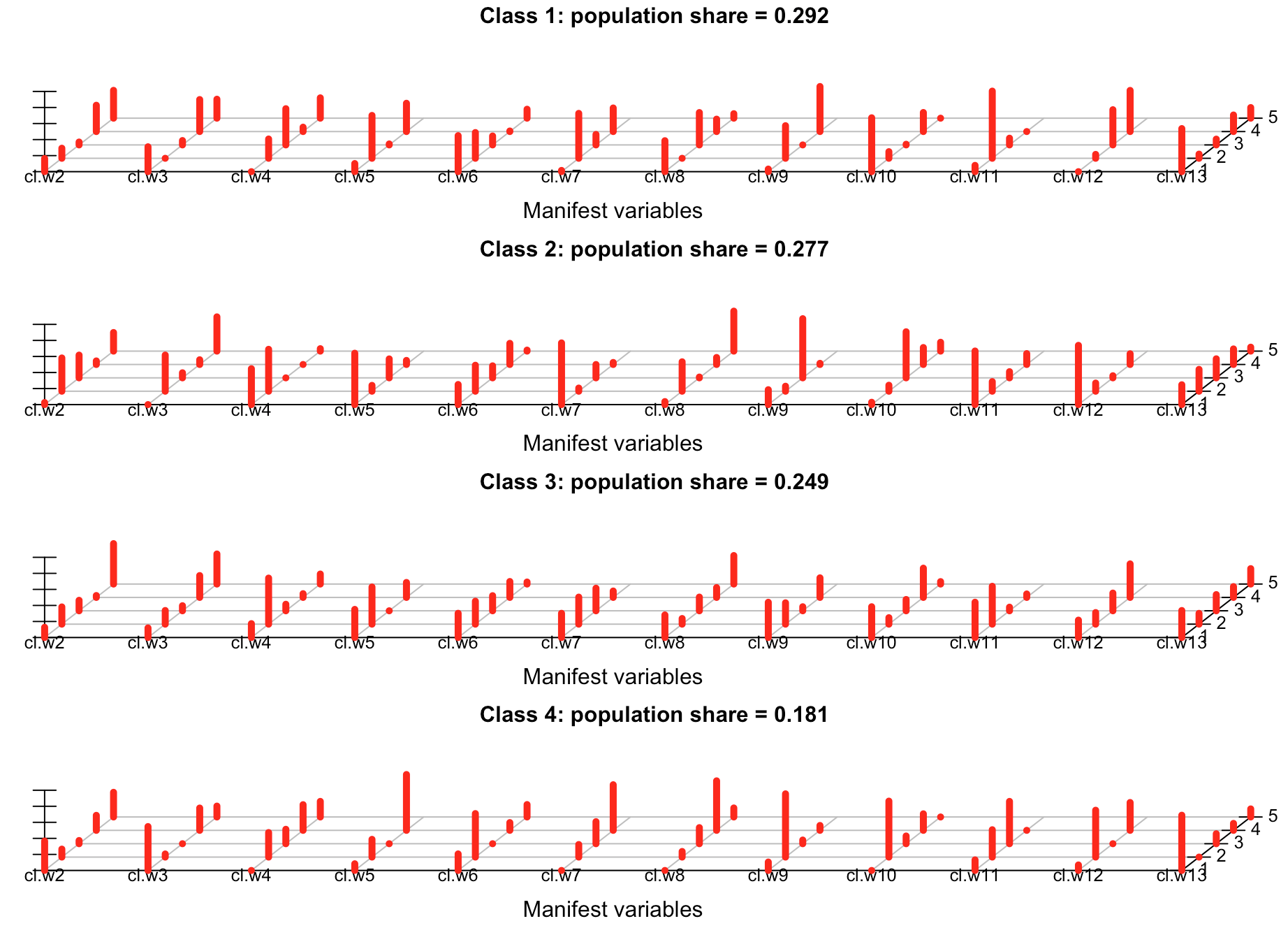
W2: C2(0.41)|C3(0.28)|C5(0.23) -> W3: C2(0.45)|C5(0.43) -> W4: C2(0.52)|C1(0.45) -> W5: C1(0.64)|C3(0.24) -> W6: C2(0.32)|C4(0.26)|C1(0.25) -> W7: C1(0.77) -> W8: C5(0.5)|C2(0.37) -> W9: C3(0.74) -> W10: C3(0.57)|C4(0.21) -> W11: C1(0.67) -> W12: C1(0.74) -> W13: C2(0.27)|C1(0.25)|C3(0.24)

The most probable path for students in **Class 3** (24.9%):

W2: C5(0.5)|C2(0.21) -> W3: C5(0.37)|C4(0.27) -> W4: C2(0.57) -> W5: C2(0.46)|C1(0.35) -> W6: C1(0.3)|C2(0.28)|C4(0.2) -> W7: C2(0.33)|C1(0.3)|C3(0.28) -> W8: C5(0.36)|C1(0.28) -> W9: C1(0.44)|C2(0.27)|C4(0.24) -> W10: C1(0.38)|C4(0.36) -> W11: C2(0.47)|C1(0.44) -> W12: C4(0.42)|C3(0.22)|C1(0.21) -> W13: C1(0.34)|C3(0.2)

The most probable path for students in **Class 4** (18.1%):

W2: C1(0.37)|C5(0.31) -> W3: C1(0.55)|C4(0.28) -> W4: C4(0.32)|C2(0.31) -> W5: C4(0.7)|C2(0.22) -> W6: C2(0.54)|C1(0.21) -> W7: C4(0.57)|C3(0.27) -> W8: C4(0.62)|C3(0.2) -> W9: C2(0.79) -> W10: C2(0.7)|C4(0.21) -> W11: C3(0.53)|C2(0.34) -> W12: C2(0.58)|C4(0.35) -> W13: C1(0.69)



**INTERPRETATION BASED ON CLUSTERS IDENTIFIED WITH ABELARDO**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CLASS 1 (29.2%)** | **CLASS 2 (27.7%)** | **CLASS 3 (24.9%)** | **CLASS 4 (18.1%)** |
| **WEEK 2** | C1|B1 | D|E|C1 | C1|D | A|C1 |
| **WEEK 3** | B1|C1|A | E|A | A|B1 | C1|B1 |
| **WEEK 4** | B1|A|C1 | C1|E | C1 | B2|C1 |
| **WEEK 5** | B1|A | D|E | B1|D | A|B1 |
| **WEEK 6** | C2|C1 | C1|D|C2 | C2|C1|D | C1|C2 |
| **WEEK 7** | B1|D | E | B1|E|C1 | D|C1 |
| **WEEK 8** | C1|B1 | B2|E | B2|B1 | A|C1 |
| **WEEK 9** | B1|D | E | B2|D|B1 | D |
| **WEEK 10** | B2|E | D|E | B2|E | A|E |
| **WEEK 11** | C1 | E | C1|E | B1|C1 |
| **WEEK 12** | B1|F | B2 | B1|F|B2 | E|B1 |
| **WEEK 13** | A|C2 | C1|A|D | A|F | A |

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

* A – disengaged
* B1, B2 – gaming/guessing assessment
* C1, C2 – low engaged, assessment driven
* D (Eff) – engaged and effective/efficient
* E (Top) - engage frequently in all kinds of activities
* F - cheaters

**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 84 14 11.0 18.00

2 81 20 16.0 32.00

3 70 21 14.0 28.75

4 55 16 12.5 22.00

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

Pairwise comparisons with the FDR correction

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

\ c1 c2 Z p effect.size significant

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

1 1 2 -5.6803 0.000000 0.4422 YES

2 1 3 -4.4338 0.000006 0.3573 YES

5 2 4 3.4648 0.000462 0.2971 YES

6 3 4 2.1858 0.028564 0.1955 YES

3 1 4 -1.9672 0.049057 0.1669 NO

4 2 3 1.2587 0.209143 0.1024 NO

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

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

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: 0.1638 0.1254 0.0360 0.3279 0.3469

class 2: 0.0254 0.4146 0.2837 0.0444 0.2319

class 3: 0.1273 0.2148 0.1305 0.0235 0.5039

class 4: 0.3665 0.0993 0.0399 0.1869 0.3074

$cl.w3

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

class 1: 0.3101 0.0000 0.0545 0.3990 0.2364

class 2: 0.0000 0.4507 0.0634 0.0589 0.4270

class 3: 0.1195 0.1702 0.0631 0.2724 0.3747

class 4: 0.5473 0.0388 0.0000 0.2781 0.1358

$cl.w4

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

class 1: 0.0000 0.2403 0.4517 0.0545 0.2534

class 2: 0.4474 0.5223 0.0000 0.0000 0.0303

class 3: 0.1730 0.5744 0.0813 0.0465 0.1249

class 4: 0.0000 0.3060 0.1798 0.3201 0.1942

$cl.w5

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

class 1: 0.1011 0.5342 0.0121 0.3526

class 2: 0.6405 0.0728 0.2361 0.0505

class 3: 0.3506 0.4629 0.0000 0.1865

class 4: 0.0831 0.2209 0.0000 0.6960

$cl.w6

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

class 1: 0.4490 0.3209 0.1115 0.0069 0.1117

class 2: 0.2504 0.3239 0.1499 0.2634 0.0124

class 3: 0.3024 0.2859 0.1882 0.1982 0.0253

class 4: 0.2060 0.5423 0.0000 0.0967 0.1549

$cl.w7

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

class 1: 0.0154 0.5599 0.1308 0.2940

class 2: 0.7690 0.0394 0.1663 0.0253

class 3: 0.3020 0.3371 0.2812 0.0797

class 4: 0.0000 0.1565 0.2749 0.5686

$cl.w8

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

class 1: 0.3860 0.0000 0.4034 0.1546 0.0559

class 2: 0.0377 0.3668 0.0099 0.0872 0.4984

class 3: 0.2816 0.0702 0.1707 0.1222 0.3553

class 4: 0.0000 0.0699 0.2005 0.6172 0.1123

$cl.w9

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

class 1: 0.0321 0.4063 0.0000 0.5616

class 2: 0.1855 0.0622 0.7377 0.0145

class 3: 0.4395 0.2662 0.0503 0.2440

class 4: 0.1050 0.7884 0.0459 0.0608

$cl.w10

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

class 1: 0.6717 0.0824 0.0084 0.2375 0.0000

class 2: 0.0291 0.0747 0.5745 0.2114 0.1103

class 3: 0.3838 0.0805 0.1407 0.3646 0.0303

class 4: 0.0000 0.6978 0.0963 0.2060 0.0000

$cl.w11

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

class 1: 0.0784 0.8377 0.0839 0.0000

class 2: 0.6679 0.1196 0.0769 0.1356

class 3: 0.4402 0.4718 0.0450 0.0430

class 4: 0.1316 0.3411 0.5274 0.0000

$cl.w12

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

class 1: 0.0000 0.0484 0.4389 0.5128

class 2: 0.7384 0.1017 0.0263 0.1336

class 3: 0.2168 0.1445 0.2208 0.4179

class 4: 0.0686 0.5842 0.0000 0.3472

$cl.w13

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

class 1: 0.5370 0.0521 0.0721 0.2065 0.1322

class 2: 0.2473 0.2723 0.2374 0.1939 0.0492

class 3: 0.3368 0.1346 0.2018 0.1347 0.1921

class 4: 0.6887 0.0000 0.1241 0.0866 0.1006

Estimated class population shares

0.2921 0.2775 0.2494 0.181

Predicted class memberships (by modal posterior prob.)

0.2897 0.2793 0.2414 0.1897

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

Fit for 4 latent classes:

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

number of observations: 290

number of fully observed cases: 239

number of estimated parameters: 175

residual degrees of freedom: 115

maximum log-likelihood: -4112.292

AIC(4): 8574.584

BIC(4): 9216.814

G^2(4): 4385.674 (Likelihood ratio/deviance statistic)

X^2(4): 39148808 (Chi-square goodness of fit)