M1:(usually we do LCA starting with 2 classes, the result of M1 can be ignored)

Conditional item response (column) probabilities,

by outcome variable, for each class (row)

$income

Pr(1) Pr(2) Pr(3) Pr(4) Pr(5) Pr(6)

class 1: 0 0.0971 0.1849 0.2909 0.2473 0.1797

$health\_cover

Pr(1) Pr(2)

class 1: 0.1784 0.8216

$house

Pr(1) Pr(2)

class 1: 0.5591 0.4409

$wc

Pr(1) Pr(2)

class 1: 0.4948 0.5052

$phone

Pr(1) Pr(2)

class 1: 0.111 0.889

$car

Pr(1) Pr(2)

class 1: 0.4747 0.5253

$consumption

Pr(1) Pr(2)

class 1: 0.9027 0.0973

Estimated class population shares

1

Predicted class memberships (by modal posterior prob.)

1

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

Fit for 1 latent classes:

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

number of observations: 98443

number of estimated parameters: 11

residual degrees of freedom: 372

maximum log-likelihood: -468521.8

AIC(1): 937065.7

BIC(1): 937170.1

G^2(1): 103464.3 (Likelihood ratio/deviance statistic)

X^2(1): 242432.4 (Chi-square goodness of fit)

M2:

> M2 <- poLCA(f, data = latent\_class\_data, nclass = 2, graphs = TRUE, na.rm = TRUE)

Conditional item response (column) probabilities,

by outcome variable, for each class (row)

$income

Pr(1) Pr(2) Pr(3) Pr(4) Pr(5) Pr(6)

class 1: 0 0.0033 0.0262 0.2314 0.4120 0.3271

class 2: 0 0.2109 0.3776 0.3632 0.0475 0.0009

$health\_cover

Pr(1) Pr(2)

class 1: 0.0829 0.9171

class 2: 0.2943 0.7057

For example, a person classified in class 1 will have 0.0829 chance(8.29%) of having high level of health insurance coverage.

$house

Pr(1) Pr(2)

class 1: 0.3682 0.6318

class 2: 0.7908 0.2092

$wc

Pr(1) Pr(2)

class 1: 0.2098 0.7902

class 2: 0.8407 0.1593

$phone

Pr(1) Pr(2)

class 1: 0.0020 0.9980

class 2: 0.2432 0.7568

$car

Pr(1) Pr(2)

class 1: 0.3618 0.6382

class 2: 0.6117 0.3883

$consumption

Pr(1) Pr(2)

class 1: 0.8338 0.1662

class 2: 0.9864 0.0136

Estimated class population shares

0.5483 0.4517

Predicted class memberships (by modal posterior prob.)

0.5703 0.4297

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

Fit for 2 latent classes:

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

number of observations: 98443

number of estimated parameters: 23

residual degrees of freedom: 360

maximum log-likelihood: -432540.1

AIC(2): 865126.2

BIC(2): 865344.7

G^2(2): 31500.84 (Likelihood ratio/deviance statistic)

X^2(2): 35732.49 (Chi-square goodness of fit)

M3:

Conditional item response (column) probabilities,

by outcome variable, for each class (row)

$income

Pr(1) Pr(2) Pr(3) Pr(4) Pr(5) Pr(6)

class 1: 0 0.0430 0.2754 0.5137 0.1627 0.0053

class 2: 0 0.0059 0.0002 0.1231 0.4320 0.4389

class 3: 0 0.5113 0.4150 0.0714 0.0000 0.0023

$health\_cover

Pr(1) Pr(2)

class 1: 0.2190 0.7810

class 2: 0.0577 0.9423

class 3: 0.3846 0.6154

$house

Pr(1) Pr(2)

class 1: 0.6838 0.3162

class 2: 0.3117 0.6883

class 3: 0.8559 0.1441

$wc

Pr(1) Pr(2)

class 1: 0.6855 0.3145

class 2: 0.1231 0.8769

class 3: 0.9308 0.0692

$phone

Pr(1) Pr(2)

class 1: 0.0480 0.9520

class 2: 0.0014 0.9986

class 3: 0.6026 0.3974

$car

Pr(1) Pr(2)

class 1: 0.4703 0.5297

class 2: 0.3355 0.6645

class 3: 0.8691 0.1309

$consumption

Pr(1) Pr(2)

class 1: 0.9594 0.0406

class 2: 0.8055 0.1945

class 3: 0.9960 0.0040

Estimated class population shares

0.4491 0.4034 0.1475

Predicted class memberships (by modal posterior prob.)

0.4493 0.4107 0.14

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

Fit for 3 latent classes:

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

number of observations: 98443

number of estimated parameters: 35

residual degrees of freedom: 348

maximum log-likelihood: -425938.3

AIC(3): 851946.5

BIC(3): 852278.9

G^2(3): 18297.13 (Likelihood ratio/deviance statistic)

X^2(3): 18451.99 (Chi-square goodness of fit)

ALERT: iterations finished, MAXIMUM LIKELIHOOD NOT FOUND

M4:

Conditional item response (column) probabilities,

by outcome variable, for each class (row)

$income

Pr(1) Pr(2) Pr(3) Pr(4) Pr(5) Pr(6)

class 1: 0 0.0177 0.0732 0.3280 0.3623 0.2188

class 2: 0 0.4730 0.4318 0.0934 0.0006 0.0013

class 3: 0 0.0419 0.2834 0.5044 0.1604 0.0100

class 4: 0 0.0059 0.0132 0.1294 0.4034 0.4480

$health\_cover

Pr(1) Pr(2)

class 1: 0.1107 0.8893

class 2: 0.3912 0.6088

class 3: 0.2095 0.7905

class 4: 0.0714 0.9286

$house

Pr(1) Pr(2)

class 1: 0.6097 0.3903

class 2: 0.8457 0.1543

class 3: 0.7004 0.2996

class 4: 0.1969 0.8031

$wc

Pr(1) Pr(2)

class 1: 0.0765 0.9235

class 2: 0.9060 0.0940

class 3: 0.8655 0.1345

class 4: 0.1396 0.8604

$phone

Pr(1) Pr(2)

class 1: 0.0078 0.9922

class 2: 0.5523 0.4477

class 3: 0.0549 0.9451

class 4: 0.0012 0.9988

$car

Pr(1) Pr(2)

class 1: 0.9992 0.0008

class 2: 0.8935 0.1065

class 3: 0.3373 0.6627

class 4: 0.0138 0.9862

$consumption

Pr(1) Pr(2)

class 1: 0.7858 0.2142

class 2: 0.9950 0.0050

class 3: 0.9727 0.0273

class 4: 0.8551 0.1449

Estimated class population shares

0.2115 0.1641 0.3345 0.29

Predicted class memberships (by modal posterior prob.)

0.2206 0.1524 0.333 0.2941

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

Fit for 4 latent classes:

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

number of observations: 98443

number of estimated parameters: 47

residual degrees of freedom: 336

maximum log-likelihood: -420857.1

AIC(4): 841808.2

BIC(4): 842254.6

G^2(4): 8134.834 (Likelihood ratio/deviance statistic)

X^2(4): 8664.601 (Chi-square goodness of fit)

ALERT: iterations finished, MAXIMUM LIKELIHOOD NOT FOUND

M5:

Conditional item response (column) probabilities,

by outcome variable, for each class (row)

$income

Pr(1) Pr(2) Pr(3) Pr(4) Pr(5) Pr(6)

class 1: 0 0.0000 0.0000 0.1351 0.4274 0.4375

class 2: 0 0.0554 0.2077 0.5308 0.2061 0.0000

class 3: 0 0.0433 0.2834 0.4910 0.1648 0.0175

class 4: 0 0.0061 0.0179 0.1227 0.4144 0.4389

class 5: 0 0.4858 0.4240 0.0888 0.0005 0.0009

$health\_cover

Pr(1) Pr(2)

class 1: 0.0415 0.9585

class 2: 0.2634 0.7366

class 3: 0.1889 0.8111

class 4: 0.0715 0.9285

class 5: 0.3893 0.6107

$house

Pr(1) Pr(2)

class 1: 0.5817 0.4183

class 2: 0.6190 0.3810

class 3: 0.7156 0.2844

class 4: 0.0649 0.9351

class 5: 0.8524 0.1476

$wc

Pr(1) Pr(2)

class 1: 0.0834 0.9166

class 2: 0.0156 0.9844

class 3: 1.0000 0.0000

class 4: 0.1421 0.8579

class 5: 0.9463 0.0537

$phone

Pr(1) Pr(2)

class 1: 0.0018 0.9982

class 2: 0.0376 0.9624

class 3: 0.0532 0.9468

class 4: 0.0014 0.9986

class 5: 0.5784 0.4216

$car

Pr(1) Pr(2)

class 1: 0.6473 0.3527

class 2: 0.7769 0.2231

class 3: 0.3446 0.6554

class 4: 0.0000 1.0000

class 5: 0.8917 0.1083

$consumption

Pr(1) Pr(2)

class 1: 0.6822 0.3178

class 2: 0.9335 0.0665

class 3: 0.9703 0.0297

class 4: 0.9107 0.0893

class 5: 0.9948 0.0052

Estimated class population shares

0.1848 0.1472 0.3016 0.213 0.1535

Predicted class memberships (by modal posterior prob.)

0.1723 0.1334 0.3196 0.2365 0.1381

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

Fit for 5 latent classes:

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

number of observations: 98443

number of estimated parameters: 59

residual degrees of freedom: 324

maximum log-likelihood: -419520.9

AIC(5): 839159.9

BIC(5): 839720.2

G^2(5): 5462.509 (Likelihood ratio/deviance statistic)

X^2(5): 5815.432 (Chi-square goodness of fit)

ALERT: iterations finished, MAXIMUM LIKELIHOOD NOT FOUND

In LCA, we do not know which model or how many classes are appropriate for our data yet. So we have to try out different models and then compare them.

Akaike Information Criterion(AIC): a measure of relative fit or quality of the model to the data, which considered model parsimony(the number of parameters). AIC is estimated based on log-likelihood squared and log likelihood. In either case. Lower AIC indices indicate a better fit.

Bayesian Information Criterion(BIC): based on LL: BIC is used to compare the fit of different models with lower indices indicating better fir. BIC imposes a stronger penalty on the number of parameters than AIC and is therefore in case where there is a discrepancy between AIC and other fit statistics, the former index is often prioritized to assess model fit. Lower BIC indices indicate a better fit.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | BIC Value | AIC Value | Log-Likelihood |  |  |
| M1 | 937170.1 | 937065.7 | -468521.8 |  |  |
| M2 | 865344.7 | 865126.2 | -432540.1 |  |  |
| M3 | 852278.9 | 851946.5 | -425938.3 |  |  |
| M4 | 842254.6 | 841808.2 | -420857.1 |  |  |
| M5 | 839720.2 | 839159.9 | -419520.9 |  |  |