Kraków, 26.01.2016

Analiza klas i profili ukrytych w segmentacji rynku

- projekt zaliczeniowy

Wykonała:
Agnieszka Kowalik

Spis treści

Zmienne wykorzystane w projekcie

Zmienne binarne (jakościowe):

1) Important child qualities: self-expression (v22) [A],
2) Would not like to have as neighbors: heavy drinkers (v42) [B],
3) Would not like to have as neighbors: people who speak a different language (v44) [H],
4) Sex (v240) [D].

1 – Mentioned/Male; **2** – Not mentioned/Female

Zmienne mierzone na skali Likerta (ilościowe):

1) Satisfaction with your life (v23) [AA],
2) Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair? (v56) [BB],
3) Adventure and taking risks are important to this person; to have an exciting life (v76) [HH],
4) The only acceptable religion is my religion (v154) [DD].

Źródło danych: World Values Survey (2010-2014) - Turkey 2011.

Tabela kontyngencji (tabela chi-kwadrat)

Tabela liczności (daneee_2_binarne)
Liczność oznacz. komórek > 10
(Nie oznaczono sum brzegowych)

V22	V42	V44	V240 1	V240 2	Wiersz Razem
1	1	1	66	60	126
1	1	2	182	223	405
Ogół			248	283	531
1	2	1	7	7	14
1	2	2	49	34	83
Ogół			56	41	97
2	1	1	128	159	287
2	1	2	238	264	502
Ogół			366	423	789
2	2	1	18	17	35
2	2	2	92	61	153
Ogół			110	78	188
Razem w kol.			780	825	1605

Obliczenia zostały wykonane w programie STATISTICA.

W badaniu wzięło udział 780 mężczyzn oraz 825 kobiet, czyli 1605 respondentów.
Spośród wszystkich 1605 przebadanych osób dla 66 mężczyzn oraz 60 kobiet ważne jest wyrażanie opinii przez dzieci, nie chcieliby mieszkać w sąsiedztwie osób nadużywających alkohol oraz w sąsiedztwie osób mówiących różnymi językami. Natomiast 92 mężczyzn oraz 61 kobiet (153 osoby) jest zupełnie odmiennego zdania we wszystkich (trzech) kwestiach.
Dla 182 mężczyzn i 223 przebadanych kobiet w Turcji (405 osób) ważne jest by dzieci wyrażały swoje opinie, a także nie chcieliby mieszkać w sąsiedztwie osób nadużywających alkohol, ale chcieliby mieszkać w sąsiedztwie osób mówiących rożnymi językami.
Dla 49 mężczyzn oraz 34 kobiet (83 osób) ważne jest wyrażanie własnych opinii i jednocześnie chcieliby mieszkać w sąsiedztwie osób nadużywających alkohol oraz w sąsiedztwie osób mówiących różnymi językami.
Dla 128 mężczyzn oraz 159 kobiet (289 osób) nie jest ważne wyrażanie przez dzieci własnych opinii i równocześnie nie chcieliby mieszkać w pobliżu osób nadużywających alkohol oraz mówiących różnymi językami.
Dla 238 mężczyzn oraz 264 kobiet (502 osób) nie jest ważne wyrażanie własnych opinii przez dzieci, co więcej nie chcieliby mieszkać w sąsiedztwie osób nadużywających alkohol, jednakże chcieliby mieszkać w sąsiedztwie osób mówiących różnymi językami.
Dla 110 mężczyzn oraz 78 kobiet (188 osób) nie jest ważne wyrażanie opinii przez dzieci i dodatkowo chcieliby mieszkać w sąsiedztwie osób nadużywających alkohol.
628 respondentów (531+97) twierdzi, iż wyrażanie własnych opinii przez dzieci jest ważne, natomiast 977 respondentów (789+188) jest przeciwnego zdania.

Analiza log –liniowa

```
LEM: log-linear and event history analysis with missing data.
Developed by JeroenVermunt (c), Tilburg University, The Netherlands.
Version 1.0 (September 18, 1997).


*** INPUT ***

   * Analiza logliniowa


   man 4
   dim 2 2 2 2
   lab A B C D
   mod  {AB AC BC BD}
   rec 1605
dat daneee_2_binarne.txt

   * write estimated conditional probabilities to a file
wco hag90_6a.con



*** STATISTICS ***

  Number of iterations = 5
  Converge criterion   = 0.0000000020

  X-squared            = 4.7374 (0.6920)
  L-squared            = 4.6908 (0.6976)
Cressie-Read         = 4.7189 (0.6942)


  Dissimilarity index  = 0.0199 ma być najniższe
  Degrees of freedom   = 7
  Log-likelihood       = -3867.28583
  Number of parameters = 8 (+1)
  Sample size          = 1605.0
  BIC(L-squared)       = -46.9754
AIC(L-squared)       = -9.3092
  BIC(log-likelihood)  = 7793.6187 ma być najniższe
AIC(log-likelihood)  = 7750.5717 ma być najniższe

Eigenvalues information matrix
    2942.6352  2634.4970  2520.5923  1523.7153   866.9301   544.9916
     461.7588   223.3590
```

p> 0,05, nieistotne Chi-kwadrat, odrzucamy H0 o niezależności zmiennych, czyli istnieją jakieś interakcje w tych zmiennych -> zmienne są zależne

Aby mieć model idealnie sklasyfikowany, musimy zmienić 2% przypadków.

*** FREQUENCIES ***

A	B	C	D	observed	estimated	std. res.
1	1	1	1	66.000	59.621	0.826
1	1	1	2	60.000	68.554	-1.033
1	1	2	1	182.000	187.374	-0.393
1	1	2	2	223.000	215.450	0.514
1	2	1	1	7.000	6.887	0.043
1	2	1	2	7.000	4.937	0.928
1	2	2	1	49.000	49.611	-0.087
1	2	2	2	34.000	35.564	-0.262
2	1	1	1	128.000	132.487	-0.390
2	1	1	2	159.000	152.338	0.540
2	1	2	1	238.000	234.518	0.227
2	1	2	2	264.000	269.657	-0.345
2	2	1	1	18.000	21.653	-0.785
2	2	1	2	17.000	15.522	0.375
2	2	2	1	92.000	87.849	0.443
2	2	2	2	61.000	62.976	-0.249

Najmniejsze standaryzowane reszty dla kombinacji odpowiedzi respondentów A1 B2 C1 D1.

*** LOG-LINEAR PARAMETERS ***

* TABLE ABCD [or P(ABCD)] *

effect	beta	std err	z-value	exp(beta)	Wald	dfprob
main	4.0829			59.3180		
A						
1	-0.3425	0.0384	-8.911	0.7100		
2	0.3425			1.4084	79.41	1 0.000
B						
1	0.9032	0.0443	20.399	2.4674		
2	-0.9032			0.4053	416.11	1 0.000
C						
1	-0.6364	0.0438	-14.546	0.5292		
2	0.6364			1.8897	211.57	1 0.000
D						
1	0.0483	0.0330	1.462	1.0495		
2	-0.0483			0.9528	2.14	1 0.144
AB						
1 1	0.0867	0.0347	2.501	1.0906		
1 2	-0.0867			0.9169		
2 1	-0.0867			0.9169		
2 2	0.0867			1.0906	6.25	1 0.012
AC						
1 1	-0.1435	0.0297	-4.840	0.8663		
1 2	0.1435			1.1543		
2 1	0.1435			1.1543		
2 2	-0.1435			0.8663	23.42	1 0.000
BC						
1 1	0.2074	0.0423	4.908	1.2304		
1 2	-0.2074			0.8127		
2 1	-0.2074			0.8127		
2 2	0.2074			1.2304	24.08	1 0.000
BD						
1 1	-0.1181	0.0330	-3.574	0.8886		
1 2	0.1181			1.1254		
2 1	0.1181			1.1254		
2 2	-0.1181			0.8886	12.78	1 0.000

*** (CONDITIONAL) PROBABILITIES ***Warunkowe prawdopodobieństwa przynależności

* P(ABCD) *

1	1	1	1	0.0371	(0.0032)
1	1	1	2	0.0427	(0.0037)
1	1	2	1	0.1167	(0.0061)
1	1	2	2	0.1342	(0.0067)
1	2	1	1	0.0043	(0.0008)
1	2	1	2	0.0031	(0.0006)
1	2	2	1	0.0309	(0.0035)
1	2	2	2	0.0222	(0.0027)
2	1	1	1	0.0825	(0.0050)
2	1	1	2	0.0949	(0.0056)
2	1	2	1	0.1461	(0.0069)
2	1	2	2	0.1680	(0.0075)
2	2	1	1	0.0135	(0.0021)
2	2	1	2	0.0097	(0.0016)
2	2	2	1	0.0547	(0.0050)
2	2	2	2	0.0392	(0.0040)

Najbardziej prawdopodobne jest wystąpienie konfiguracji odpowiedzi A2 B1 C3 D4 (p-stwo równe 0.1680).

LEM: log-linear and event history analysis with missing data.
Developed by JeroenVermunt (c), Tilburg University, The Netherlands.
Version 1.0 (September 18, 1997).

loading data

iteration	log-likelihood	stop-criterion	L-squared
0	-4450.00489919	-4450.00489919	1170.12889846
1	-3867.48492730	582.51997189	5.08895468
2	-3867.28668967	0.19823764	4.69247940
3	-3867.28583226	0.00085740	4.69076460
4	-3867.28583121	0.00000105	4.69076250
5	-3867.28583121	0.00000000	4.69076249

standard errors
statistics and frequencies
log-linear parameters
cpu-time = 0.00 seconds

CFA – Konfiguracyjna Analiza Częstości (w programie CFA)

CFA (Configural Frequency Analysis) - dopełnienie analizy log-liniowej. Za pomocą CFA identyfikujemy dziwne struktury (konfiguracje klas).

1. Model bazowy – model jednorodny (a CFA model of order zero)
2. Model bazowy – model niezależny (a CFA model of order first)

Ad.1 model bazowy – model jednorodny (a CFA model of order zero)

model jednorodny - wszystkie rozkłady jednorodne w segmentach

```
Configural Frequency Analysis
      author of program: Alexander von Eye, 2000

MarginalFrequencies
VariableFrequencies  (Rozkład częstości dla 4 analizowanych zmiennych)
1    628.  977.
      2   1320.  285.
      3    462. 1143.
      4    780.  825.

  sample size N =  1605

Pearsons chi2 test was used
Bonferroni-adjusted alpha =  .0031250
 a CFA of order   0  was performed
```

Table of results

Configuration	fofe	statistic	p	
1111	66.	100.313	11.737	.00061276 Antitype
1112	60.	100.313	16.200	.00005698 Antitype
1121	182.	100.313	66.521	.00000000 Type
1122	223.	100.313	150.053	.00000000 Type
1211	7.	100.313	86.801	.00000000 Antitype
1212	7.	100.313	86.801	.00000000 Antitype
1221	49.	100.313	26.248	.00000030 Antitype
1222	34.	100.313	43.836	.00000000 Antitype
2111	128.	100.313	7.642	.00570216
2112	159.	100.313	34.335	.00000000 Type
2121	238.	100.313	188.988	.00000000 Type
2122	264.	100.313	267.101	.00000000 Type
2211	18.	100.313	67.542	.00000000 Antitype

2212	17.	100.313	69.193	.00000000	Antitype
2221	92.	100.313	.689	.40656471	
2222	61.	100.313	15.407	.00008669	Antitype

Anty-typy - przypadki odstające, czyli np. konfiguracja 1211 - mężczyzna, dla którego ważne jest wyrażanie swoich opinii oraz który nie chciałby mieć w swoim sąsiedztwie ludzi nadużywających alkohol, a także mógłby mieszkać w sąsiedztwie osób, które mówią innymi językami - segment zbyt mało liczny w stosunku do rozkładu jednorodnego [czyli segment nie jest wart uruchomienia działań marketingowych], (fo = 7).

Typy - przypadki najczęstsze, np. konfiguracja 2122 - kobieta, dla której ważne jest wyrażanie swoich opinii, która nie chciałaby mieszkać w sąsiedztwie osób nadużywających alkohol oraz osób mówiących innymi językami - jest to segment najbardziej liczny w stosunku do rozkładu jednorodnego (fo = 264).

chi2 for CFA model = 1139.0947
df = 15 p = .00000000

Descriptive indicators of types and antitypes

cell	Rel. Risk*	Rank	logP*	Rank
1111	.658	8	3.313	14
1112	.598	10	4.424	12
1121	1.814	4	12.836	8
1122	2.223	3	25.421	5
1211	.070	15	30.066	3
1212	.070	16	30.066	4
1221	.488	11	7.120	11
1222	.339	12	12.464	9
2111	1.276	6	2.525	15
2112	1.585	5	7.465	10
2121	2.373	2	30.871	2
2122	2.632	1	41.270	1
2211	.179	13	21.145	7
2212	.169	14	21.829	6
2221	.917	7	.887	16
2222	.608	9	4.223	13

*Rel.Risk – Ryzyko względne
*logP – prawdopodobieństwo Ni dla rozkładu Poissona dla oczekiwanej wartości Ei

Design Matrix

Ad.2 model bazowy – model niezależny (a CFA model of order first)

model niezależny –brak jakichkolwiek zależności między zmiennymi

```
Configural Frequency Analysis
  author of program: Alexander von Eye, 2000

    Marginal Frequencies
    Variable Frequencies
      1    628.  977.
      2   1320.  285.
      3    462. 1143.
      4    780.  825.

  sample size N =  1605

Pearsons chi2 test was used
Bonferroni-adjusted alpha =  .0031250
 a CFA of order   1  was performed


                      Table of results
                      ----- -- -------
 Configuration   fofe  statistic      p
-------------  ----  -------- ---------   -------
    1111       66.   72.251      .541   .46207844
    1112       60.   76.420     3.528   .06034385
    1121      182.  178.751      .059   .80801676
    1122      223.  189.064     6.091   .01358438
    1211        7.   15.600     4.741   .02945574
    1212        7.   16.500     5.469   .01935195
    1221       49.   38.594     2.806   .09392879
    1222       34.   40.821     1.140   .28572836
    2111      128.  112.404     2.164   .14126958
2112      159.  118.888    13.533   .00023438   Type
    2121      238.  278.089     5.779   .01621641
    2122      264.  294.133     3.087   .07891987
    2211       18.   24.269     1.619   .20318392
    2212       17.   25.669     2.928   .08706832
    2221       92.   60.042    17.010   .00003718   Type
    2222       61.   63.506      .099   .75317078
```

Typy - wartości empiryczne są częstsze niż wartości oczekiwane, np. 2112 - kobieta, brak ważności wyrażanie swoich opinii, może mieszkać w pobliżu osób nadużywających alkohol i mówiących innymi językami - te wartości kategorii są ze sobą najbardziej powiązane, zależne, najsilniej ze sobą skorelowane (są istotnie zależne).

Anty-typy - kategorie zmiennych też są ze sobą silnie skorelowane, ale wartości oczekiwane są wyższe niż wartości empiryczne (w analizowanym zbiorze danych dla modelu niezależnego nie ma żadnych anty-typów).

chi2 for CFA model = 70.5943
df = 11 p = .00000000

Descriptive indicators of types and antitypes

cell	Rel. Risk	Rank	logP	Rank
1111	.913	8	.785	13
1112	.785	12	1.363	7
1121	1.018	6	.957	11
1122	1.179	4	2.326	3
1211	.449	15	1.061	10
1212	.424	16	1.229	9
1221	1.270	3	1.360	8
1222	.833	11	.734	14
2111	1.139	5	1.388	6
2112	1.337	2	3.724	2
2121	.856	10	2.203	4
2122	.898	9	1.633	5
2211	.742	13	.666	16
2212	.662	14	.905	12
2221	1.532	1	4.188	1
2222	.961	7	.703	15

Design Matrix

1.0	1.0	1.0	1.0
1.0	-1.0	1.0	1.0
1.0	1.0	1.0	-1.0
1.0	-1.0	1.0	-1.0
1.0	1.0	-1.0	1.0
1.0	-1.0	-1.0	1.0
1.0	1.0	-1.0	-1.0
1.0	-1.0	-1.0	-1.0
-1.0	1.0	1.0	1.0
-1.0	-1.0	1.0	1.0
-1.0	1.0	1.0	-1.0
-1.0	-1.0	1.0	-1.0
-1.0	1.0	-1.0	1.0
-1.0	-1.0	-1.0	1.0
-1.0	1.0	-1.0	-1.0
-1.0	-1.0	-1.0	-1.0

Analiza porównawcza na podstawie modeli LCA

Wybór modelu ze względu na liczbę klas ukrytych możemy dokonać na podstawie:
- p-value dla statystyki chi-kwadrat –> im wyższe, tym lepiej,
- współczynników AIC, CAIC, BIC, ABIC, i innych –> im mniejsze, tym lepiej,
- testów przyrostowych –> czy model z większą liczba klas jest istotnie lepszy niż model z mniejszą liczbą klas

W poniższej analizie wybór modelu został oparty na podstawie analizy testów przyrostowych, ale zostały także przedstawione wartości współczynników AIC oraz wartości dla statystyki chi-kwadrat.

Wybór modelu na podstawie testów przyrostowych
(TECH 11 i TECH 14 w programie Mplus)

Szuka nieoptymalnej liczby punktów startowych

Final stage loglikelihood values at local maxima, seeds, and initial stage start numbers:

-3876.768	195873	6
-3876.768	650371	14
-3876.768	939021	8
-3879.218	93468	3

Model 2-klasowy

```
TITLE:  Analiza klas ukrytych

DATA:
 !FILE IS "C:\Users\Agnieszka\Desktop\dane\Arkusz2bm.txt"; ! PLIK DANYCH
   FILE IS "C:\Users\Agnieszka\Desktop\dane\daneee_2_binarne.txt";

VARIABLE:
 NAMES ARE A B H D;
 USEVARIABLES ARE A B H D;
 CATEGORICAL ARE A B H D; ! WSKAŹNIKI BINARNE/PORZĄDKOWE
 CLASSES = c(2); ! LICZBA KLAS UKRYTYCH

ANALYSIS:
 TYPE IS MIXTURE;
  OPTSEED =  195873;
  PROCESS = 4 (STARTS);
! LRTSTARTS = 6 14 8 3;   ! LICZBA PUNKTÓW STARTOWYCH W POCZĄTKOWYCH ITERACJACH
! ALGORITHM = INTEGRATION;
! INTEGRATION = MONTECARLO (1000);
! STARTS = 0;

Plot:
 type is plot3;
 series is A (1) B (2) H (3) D (4);
 MONITOR = ON;
output: tech11 tech14;
Savedata:
 file is lca1_save_moje.txt ;
 save is cprob;
```

TECHNICAL 11 OUTPUT

Random Starts Specifications for the k-1 Class Analysis Model

Number of initial stage random starts	20
Number of final stage optimizations	4

VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 1 (H0) VERSUS 2 CLASSES

H0 Loglikelihood Value	-3900.109
2 Times the Loglikelihood Difference	46.683
Difference in the Number of Parameters	5
Mean	2.971
Standard Deviation	2.570
P-Value	0.0000

LO-MENDELL-RUBIN ADJUSTED LRT TEST

Value	45.451
P-Value	0.0000

TECHNICAL 14 OUTPUT

Random Starts Specifications for the k-1 Class Analysis Model
- Number of initial stage random starts 20
- Number of final stage optimizations 4

Random Starts Specification for the k-1 Class Model for Generated Data
- Number of initial stage random starts 0
- Number of final stage optimizations for the initial stage random starts 0

Random Starts Specification for the k Class Model for Generated Data
- Number of initial stage random starts 40
- Number of final stage optimizations 8

Number of bootstrap draws requested Varies

PARAMETRIC BOOTSTRAPPED LIKELIHOOD RATIO TEST FOR 1 (H0) VERSUS 2 CLASSES

H0 Loglikelihood Value	-3900.109
2 Times the Loglikelihood Difference	46.683
Difference in the Number of Parameters	5
Approximate P-Value	0.0000
Successful Bootstrap Draws	10

p < 0,05 (wszystkie trzy p-value) -> model 2-klasowy jest istotnie lepszy od modelu 1-klasowego

Model 3-klasowy

```
TITLE:  Analiza klas ukrytych

DATA:
 !FILE IS "C:\Users\Agnieszka\Desktop\dane\Arkusz2bm.txt"; ! PLIK DANYCH
   FILE IS "C:\Users\Agnieszka\Desktop\dane\daneee_2_binarne.txt";

VARIABLE:
 NAMES ARE A B H D;
 USEVARIABLES ARE A B H D;
 CATEGORICAL ARE A B H D; ! WSKAŹNIKI BINARNE/PORZĄDKOWE
 CLASSES = c(3); ! LICZBA KLAS UKRYTYCH

ANALYSIS:
 TYPE IS MIXTURE;
  OPTSEED =  195873;
  PROCESS = 4 (STARTS);
! LRTSTARTS = 6 14 8 3;   ! LICZBA PUNKTÓW STARTOWYCH W POCZĄTKOWYCH ITERACJACH
! ALGORITHM = INTEGRATION;
! INTEGRATION = MONTECARLO (1000);
! STARTS = 0;

Plot:
 type is plot3;
 series is A (1) B (2) H (3) D (4);
 MONITOR = ON;
output: tech11 tech14;
Savedata:
 file is lca1_save_moje.txt ;
 save is cprob;
```

TECHNICAL 11 OUTPUT

Random Starts Specifications for the k-1 Class Analysis Model

Number of initial stage random starts	20
Number of final stage optimizations	4

VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 2 (H0) VERSUS 3 CLASSES

H0 Loglikelihood Value	-3876.768
2 Times the Loglikelihood Difference	21.154
Difference in the Number of Parameters	5
Mean	4.139
Standard Deviation	2.927
P-Value	0.0004

LO-MENDELL-RUBIN ADJUSTED LRT TEST

Value	20.596
P-Value	0.0005

TECHNICAL 14 OUTPUT

Random Starts Specifications for the k-1 Class Analysis Model
- Number of initial stage random starts 20
- Number of final stage optimizations 4

Random Starts Specification for the k-1 Class Model for Generated Data
- Number of initial stage random starts 0
- Number of final stage optimizations for the initial stage random starts 0

Random Starts Specification for the k Class Model for Generated Data
- Number of initial stage random starts 40
- Number of final stage optimizations 8

Number of bootstrap draws requested Varies

PARAMETRIC BOOTSTRAPPED LIKELIHOOD RATIO TEST FOR 2 (H0) VERSUS 3 CLASSES

H0 Loglikelihood Value	-3876.768
2 Times the Loglikelihood Difference	21.154
Difference in the Number of Parameters	5
Approximate P-Value	0.0000
Successful Bootstrap Draws	20

$p < 0,05$ (wszystkie trzy p-value) -> model 3-klasowy jest istotnie lepszy od modelu 2-klasowego

WARNING: OF THE 20 BOOTSTRAP DRAWS, 13 DRAWS HAD BOTH A SMALLER LRT VALUE THAN THE OBSERVED LRT VALUE AND NOT A REPLICATED BEST LOGLIKELIHOOD VALUE FOR THE 3-CLASS MODEL. THIS MEANS THAT THE P-VALUE MAY NOT BE TRUSTWORTHY DUE TO LOCAL MAXIMA. INCREASE THE NUMBER OF RANDOM STARTS USING THE LRTSTARTS OPTION.

Model 4-klasowy

```
TITLE:  Analiza klas ukrytych

DATA:
 !FILE IS "C:\Users\Agnieszka\Desktop\dane\Arkusz2bm.txt"; ! PLIK DANYCH
   FILE IS "C:\Users\Agnieszka\Desktop\dane\daneee_2_binarne.txt";

VARIABLE:
 NAMES ARE A B H D;
 USEVARIABLES ARE A B H D;
 CATEGORICAL ARE A B H D; ! WSKAŹNIKI BINARNE/PORZĄDKOWE
 CLASSES = c(4); ! LICZBA KLAS UKRYTYCH

ANALYSIS:
 TYPE IS MIXTURE;
  OPTSEED =  195873;
  PROCESS = 4 (STARTS);
! LRTSTARTS = 6 14 8 3;   ! LICZBA PUNKTÓW STARTOWYCH W POCZĄTKOWYCH ITERACJACH
! ALGORITHM = INTEGRATION;
! INTEGRATION = MONTECARLO (1000);
! STARTS = 0;

Plot:
 type is plot3;
 series is A (1) B (2) H (3) D (4);
 MONITOR = ON;
output: tech11 tech14;
Savedata:
 file is lca1_save_moje.txt ;
 save is cprob;
```

TECHNICAL 11 OUTPUT

Random Starts Specifications for the k-1 Class Analysis Model

Number of initial stage random starts	20
Number of final stage optimizations	4

VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 3 (H0) VERSUS 4 CLASSES

H0 Loglikelihood Value	-3866.066
2 Times the Loglikelihood Difference	2.201
Difference in the Number of Parameters	5
Mean	2.343
Standard Deviation	2.248
P-Value	0.4137

LO-MENDELL-RUBIN ADJUSTED LRT TEST

Value	2.143
P-Value	0.4247

TECHNICAL 14 OUTPUT

Random Starts Specifications for the k-1 Class Analysis Model

Number of initial stage random starts	20
Number of final stage optimizations	4

Random Starts Specification for the k-1 Class Model for Generated Data

Number of initial stage random starts	0
Number of final stage optimizations for the initial stage random starts	0

Random Starts Specification for the k Class Model for Generated Data

Number of initial stage random starts	40
Number of final stage optimizations	8
Number of bootstrap draws requested	Varies

PARAMETRIC BOOTSTRAPPED LIKELIHOOD RATIO TEST FOR 3 (H0) VERSUS 4 CLASSES

H0 Loglikelihood Value	-3866.066
2 Times the Loglikelihood Difference	2.201
Difference in the Number of Parameters	5
Approximate P-Value	0.6000
Successful Bootstrap Draws	5

p > 0,05 (wszystkie trzy p-value) -> model 4-klasowy nie jest istotnie lepszy od modelu 3-klasowego

Wniosek: Należy przyjąć model 3-klasowy, ponieważ model 4-klasowy nie jest istotnie lepszy od modelu 3-klasowego (p>0,05 dla TECH 11 i dla TECH 14).

Wybór modelu na podstawie współczynników AIC oraz statystyki Chi-kwadrat (w programie Mplus)

Model 2-klasowy

THE MODEL ESTIMATION TERMINATED NORMALLY

MODEL FIT INFORMATION
Number of Free Parameters 9

Loglikelihood
H0 Value -3876.768
H0 Scaling Correction Factor 1.0000
for MLR

Information Criteria
Akaike (AIC) 7771.536
Bayesian (BIC) 7819.964
Sample-Size Adjusted BIC 7791.372
(n* = (n + 2) / 24)

Chi-Square Test of Model Fit for the Binary and Ordered Categorical (Ordinal) Outcomes
Pearson Chi-Square
Value 24.011
Degrees of Freedom 6
P-Value 0.0005

Likelihood Ratio Chi-Square
Value 23.655
Degrees of Freedom 6
P-Value 0.0006

Model 3-klasowy

THE MODEL ESTIMATION TERMINATED NORMALLY

MODEL FIT INFORMATION
Number of Free Parameters 14

Loglikelihood
H0 Value -3866.191
H0 Scaling Correction Factor 0.9345
for MLR

Information Criteria

Akaike (AIC)	7760.382
Bayesian (BIC)	7835.714
Sample-Size Adjusted BIC	7791.239

(n* = (n + 2) / 24)

Chi-Square Test of Model Fit for the Binary and Ordered Categorical (Ordinal) Outcomes

Pearson Chi-Square

Value	2.510
Degrees of Freedom	1
P-Value	0.1132

Likelihood Ratio Chi-Square

Value	2.501
Degrees of Freedom	1
P-Value	0.1138

Model 4-klasowy

THE MODEL ESTIMATION TERMINATED NORMALLY

THE DEGREES OF FREEDOM FOR THIS MODEL ARE NEGATIVE. THE MODEL IS NOT IDENTIFIED OR TOO MANY CELLS WERE DELETED. A CHI-SQUARE TEST IS NOT AVAILABLE.

MODEL FIT INFORMATION

Number of Free Parameters 19

Loglikelihood

H0 Value	-3864.966
H0 Scaling Correction Factor for MLR	0.9047

Information Criteria

Akaike (AIC)	7767.932
Bayesian (BIC)	7870.168
Sample-Size Adjusted BIC	7809.809

(n* = (n + 2) / 24)

Parametry modelu LCA (Mplus) – Model 3-klasowy

```
TITLE:  Analiza klas ukrytych

DATA:
 !FILE IS "C:\Users\Agnieszka\Desktop\dane\Arkusz2bm.txt"; ! PLIK DANYCH
   FILE IS "C:\Users\Agnieszka\Desktop\dane\daneee_2_binarne.txt";

VARIABLE:
 NAMES ARE A B H D;
 USEVARIABLES ARE A B H D;
 CATEGORICAL ARE A B H D; ! WSKAŹNIKI BINARNE/PORZĄDKOWE
 CLASSES = c(3); ! LICZBA KLAS UKRYTYCH

ANALYSIS:
 TYPE IS MIXTURE;
  OPTSEED =  195873;
  PROCESS = 4 (STARTS);
! LRTSTARTS = 6 14 8 3;   ! LICZBA PUNKTÓW STARTOWYCH W POCZĄTKOWYCH ITERACJACH
! ALGORITHM = INTEGRATION;
! INTEGRATION = MONTECARLO (1000);
! STARTS = 0;

Plot:
 type is plot3;
 series is A (1) B (2) H (3) D (4);
 MONITOR = ON;
output: tech11 tech14;
Savedata:
 file is lca1_save_moje.txt ;
 save is cprob;
```

```
 * Analiza klas ukrytych
 * Moje dane

 lat 1
 man 4
 dim 3 2 2 2 2
 lab X A B H D
 mod X
    A|X
    B|X
    H|X
    D|X

 rec 1605
 dat daneee_2_binarne.txt

* write estimated conditional probabilities to a file
wco hag90_6777a.con
```

Prawdopodobieństwa przynależności do klas ukrytych

- **Mplus**

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THE ESTIMATED MODEL

Latent Classes		
1	200.60236	0.12499
2	144.42207	0.08998
3	1259.97557	0.78503

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON ESTIMATED POSTERIOR PROBABILITIES

Latent Classes		
1	200.60177	0.12499
2	144.42233	0.08998
3	1259.97590	0.78503

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP

Class Counts and Proportions

Latent Classes		
1	0	0.00000
2	175	0.10903
3	1430	0.89097

- **LEM**

*** LATENT CLASS OUTPUT ***

X 1	X 2	X 3
0.2199	0.4484	0.3317

Prawdopodobieństwo, że dany respondent należy do pierwszej klasy ukrytej (X1) wynosi 0.2199 (–> najmniej ważny segment).

Prawdopodobieństwo, że dany respondent należy do drugiej klasy ukrytej (X2) wynosi 0.4484 (–> najbardziej ważny segment).

Prawdopodobieństwa warunkowe

- **Mplus**

RESULTS IN PROBABILITY SCALE

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Latent Class 1				
A				
Category 1	0.981	1.064	0.922	0.357
Category 2	0.019	1.064	0.018	0.986
B				
Category 1	0.956	0.104	9.151	0.000
Category 2	0.044	0.104	0.419	0.675
H				
Category 1	0.000	0.000	0.000	1.000
Category 2	1.000	0.000	0.000	1.000
D				
Category 1	0.420	0.062	6.789	0.000
Category 2	0.580	0.062	9.363	0.000
Latent Class 2				
A				
Category 1	0.342	0.104	3.301	0.001
Category 2	0.658	0.104	6.356	0.000
B				
Category 1	0.013	0.339	0.038	0.970
Category 2	0.987	0.339	2.913	0.004
H				
Category 1	0.000	0.000	0.000	1.000
Category 2	1.000	0.000	0.000	1.000
D				
Category 1	0.682	0.071	9.568	0.000
Category 2	0.318	0.071	4.469	0.000
Latent Class 3				
A				
Category 1	0.303	0.021	14.171	0.000
Category 2	0.697	0.021	32.593	0.000
B				
Category 1	0.894	0.014	62.343	0.000
Category 2	0.106	0.014	7.397	0.000
H				
Category 1	0.367	0.085	4.302	0.000
Category 2	0.633	0.085	7.431	0.000
D				
Category 1	0.474	0.024	19.590	0.000
Category 2	0.526	0.024	21.738	0.000

- **LEM**

*** LATENT CLASS OUTPUT ***

		X 1	X 2	X 3
		0.2199	0.4484	0.3317
A	1	0.3320	0.3021	0.5511
A	2	0.6680	0.6979	0.4489
B	1	0.4316	0.9155	0.9557
B	2	0.5684	0.0845	0.0443
H	1	0.0585	0.6132	0.0002
H	2	0.9415	0.3868	0.9998
D	1	0.6370	0.4662	0.4127
D	2	0.3630	0.5338	0.5873

Jeżeli dany respondent znajduje się w pierwszej klasie ukrytej (X1), to p-stwo, że na pierwsze pytanie (pytanie A - Czy jest ważne wyrażanie swoich opinii przez dzieci?) odpowiedział twierdząco wynosi 0.332; natomiast p-stwo, że ten respondent odpowie na owe pytanie przecząco, wynosi 0.668.

Jeżeli dany respondent znajduje się w drugiej klasie ukrytej (X2), to p-stwo, że dany respondent jest mężczyzną wynosi 0.4662, podczas gdy p-stwo, że dany respondent jest kobietą wynosi 0.5338.

Dopasowanie wzorców odpowiedzi

- **Mplus**

Szacowanie wzorców odpowiedzi. Gdy obserwowane rozkłady są podobne do oszacowanych, czyli są małe reszty, wówczas model jest dobrze dopasowany.

TECHNICAL 10 OUTPUT

MODEL FIT INFORMATION FOR THE LATENT CLASS INDICATOR MODEL PART

RESPONSE PATTERNS – WZORCE ODPOWIEDZI

No.	Pattern	No.	Pattern	No.	Pattern	No.	Pattern
1	1011	2	1010	3	0110	4	0011
5	0010	6	0001	7	1001	8	0111
9	1000	10	0000	11	1111	12	0100
13	1110	14	0101	15	1100	16	1101

RESPONSE PATTERN FREQUENCIES AND CHI-SQUARE CONTRIBUTIONS

Response Pattern	Frequency Observed	Estimated	Standardized Residual	Chi-square Contribution Pearson	Loglikelihood	(z-score)
1	264.00	263.99		0.00	0.00	0.02
2	238.00	238.01		0.00	0.00	-0.02
3	49.00	49.00		0.00	0.00	0.00
4	223.00	223.00		0.00	0.00	0.00
5	182.00	182.00		0.00	0.00	0.00
6	60.00	65.83	-0.73	0.52	-11.12	
7	159.00	151.41		0.65	0.38	15.56
8	34.00	34.00		0.00	0.00	0.00

9	128.00	136.44		-0.76			0.52	-16.35	
10	66.00	59.32			0.88		0.75	14.08	
11	61.00	61.00			0.00		0.00		0.00
12	7.00	7.04	-0.01			0.00	-0.08		
13	92.00	92.00			0.00		0.00		0.00
14	7.00	7.81	-0.29			0.08	-1.53		
15	18.00	16.19			0.45		0.20		3.82
16	17.00	17.96		-0.23			0.05	-1.87	

Analizując reszty standaryzowane można zauważyć, iż są małe. W związku z tym, możemy przyjąć, że model jest dobrze dopasowany.
(Największe reszty standaryzowane odpowiadają wzorcom odpowiedzi tj.: 0000 (0.88); 1000 (-0.76); 0001 (-0.73)).

- **LEM**

*** FREQUENCIES ***

A B H D	observed	estimated	std.res.
1 1 1 1	66.000	58.793	0.940
1 1 1 2	60.000	66.243	-0.767
1 1 2 1	182.000	181.922	0.006
1 1 2 2	223.000	223.059	-0.004
1 2 1 1	7.000	7.732	-0.263
1 2 1 2	7.000	7.428	-0.157
1 2 2 1	49.000	48.624	0.054
1 2 2 2	34.000	34.199	-0.034
2 1 1 1	128.000	135.251	-0.623
2 1 1 2	159.000	152.693	0.510
2 1 2 1	238.000	238.180	-0.012
2 1 2 2	264.000	263.860	0.009
2 2 1 1	18.000	17.123	0.212
2 2 1 2	17.000	16.737	0.064
2 2 2 1	92.000	92.375	-0.039
2 2 2 2	61.000	60.781	0.028

Jeżeli reszty standaryzowane są duże, to model nie jest najlepszym modelem pod względem jakości odwzorowania.

Tutaj reszty standaryzowane są dość niskie. (Najwyższe wartości bezwzględne tych reszt to: 0.94 (1111); -0.767 (1112); -0.623 (2111) – najsłabsze dopasowania częstości obserwowanych do częstości rzeczywistych).
Stąd można stwierdzić, iż model jest dobrze dopasowany.

(Najniższe reszty to m.in.: -0.004 (1122); 0.006 (1121); 0.009 (2122) – najlepsze dopasowania częstości obserwowanych do częstości rzeczywistych)

Dopasowanie modelu

- **Mplus**

Interpretacje –> patrz wyżej –> analiza porównawcza modeli

THE MODEL ESTIMATION TERMINATED NORMALLY

MODEL FIT INFORMATION
Number of Free Parameters 14

Loglikelihood
H0 Value -3866.191
H0 Scaling Correction Factor 0.9345
for MLR

Information Criteria
Akaike (AIC) 7760.382
Bayesian (BIC) 7835.714
Sample-Size Adjusted BIC 7791.239
(n* = (n + 2) / 24)

Chi-Square Test of Model Fit for the Binary and Ordered Categorical
(Ordinal) Outcomes
Pearson Chi-Square
Value 2.510
Degrees of Freedom 1
P-Value 0.1132

Likelihood Ratio Chi-Square
Value 2.501
Degrees of Freedom 1
P-Value 0.1138

QUALITY OF NUMERICAL RESULTS

Condition Number for the Information Matrix 0.988E-06
(ratio of smallest to largest eigenvalue)

0,000000988 < 1/5000

Jest mniejszy, czyli jest **źle** – mogły być problemy z macierzą informacyjną (macierz informacyjna mogła nie być dodatnio określona -> mógł być problem z odwracaniem macierzy podczas estymacji.

- **LEM**

*** STATISTICS ***

Number of iterations	= 1987
Converge criterion	= 0.0000009987
Seed random values	= 531
X-squared	= 2.2706 (0.1318)
L-squared	= 2.2616 (0.1326)
Cressie-Read	= 2.2672 (0.1321)
Dissimilarity index	= 0.0096
Degrees of freedom	= 1
Log-likelihood	= -3866.07125
Number of parameters	= 14 (+1)
Sample size	= 1605.0
BIC(L-squared)	= -5.1193
AIC(L-squared)	= 0.2616
BIC(log-likelihood)	= 7835.4748
AIC(log-likelihood)	= 7760.1425

Eigenvalues information matrix

1609.4803 1496.9215 1345.9066 705.2307 598.9663 535.7704
169.4782 160.6226 132.3879 3.8462 1.3612 0.8620
0.0031 -0.0992

P-stwo dla statystyki chi-kwadrat jest istotne, p=0.1318 > 0.05; a także indeks niepodobieństwa jest mniejszy od 0.05 (D=0.0096).
Stąd wskazuje to na akceptowalne dopasowanie modelu.

Gdyby p-value było zbyt wysokie (np. 0.7), wówczas należało by zastosować walidację krzyżową.
Uwaga: wysokie p-value niekoniecznie świadczy o tym, że model jest dobrze dopasowany.

WARNING: 2 (nearly) boundary or non-identified (log-linear) parameters

1609.4803 / (-0.0992) ≈ 16224.6
–> stosunek jest duży, więc model jest słabo identyfikowalny

*** PSEUDO R-SQUARED MEASURES ***

* P(A|X) *

	baseline	fitted	R-squared
entropy	0.6693	0.6427	0.0398
qualitative variance	0.2382	0.2254	0.0538
classification error	0.3913	0.3574	0.0867
-2/N*log-likelihood	1.3386	1.2853	0.0398/0.0506
likelihood^(-2/N)	3.8138	3.6158	0.0519/0.0704

* P(B|X) *

	baseline	fitted	R-squared
entropy	0.4677	0.3404	0.2721
qualitative variance	0.1460	0.1027	0.2969
classification error	0.1776	0.1475	0.1694
-2/N*log-likelihood	0.9354	0.6809	0.2721/0.2029
likelihood^(-2/N)	2.5482	1.9756	0.2247/0.3699

* P(H|X) *

	baseline	fitted	R-squared
entropy	0.6002	0.3487	0.4190
qualitative variance	0.2050	0.1185	0.4219
classification error	0.2879	0.1864	0.3526
-2/N*log-likelihood	1.2004	0.6974	0.4190/0.3347
likelihood^(-2/N)	3.3216	2.0085	0.3953/0.5656

* P(D|X) *

	baseline	fitted	R-squared
entropy	0.6928	0.6787	0.0203
qualitative variance	0.2498	0.2428	0.0279
classification error	0.4860	0.4257	0.1239
-2/N*log-likelihood	1.3855	1.3574	0.0203/0.0274
likelihood^(-2/N)	3.9969	3.8859	0.0278/0.0370

classification error – błądklasyfikacji

Entropia – ilość informacji zawarta w danym zdarzeniu jest odwrotnie proporcjonalna do p-stwa wystąpienia tego zdarzenia.

Pseudo R^2 –> które wskaźniki przynależności do klas ukrytych dobrze wyjaśnia model.

Niskie współczynniki R^2, więc model słabo wyjaśnia zmienność klas ukrytych (zmienność klas ukrytych jest słabo wyjaśniona przez wskaźniki klas ukrytych**).**

Gdy zmienna jakościowa jest zmienną zależną, wtedy nie można policzyć wariancji w sposób klasyczny (tak jak dla zmiennych ilościowych). W związku z tym, zakres wyjaśnionej wariancji zmiennych jakościowych najczęściej liczymy za pomocą entropii –> pseudo R^2.

Ocena założenia lokalnej niezależności (Mplus - TECH10)

UNIVARIATE MODEL FIT INFORMATION

Variable	Estimated Probabilities H1	H0	Standardized Residual (z-score)
A			
Category 1	0.391	0.391	0.000
Category 2	0.609	0.609	0.000
Univariate Pearson Chi-Square			0.000
Univariate Log-Likelihood Chi-Square			0.000
B			
Category 1	0.822	0.822	0.000
Category 2	0.178	0.178	0.000
Univariate Pearson Chi-Square			0.000
Univariate Log-Likelihood Chi-Square			0.000
H			
Category 1	0.288	0.288	0.000
Category 2	0.712	0.712	0.000
Univariate Pearson Chi-Square			0.000
Univariate Log-Likelihood Chi-Square			0.000
D			
Category 1	0.486	0.486	0.000
Category 2	0.514	0.514	0.000
Univariate Pearson Chi-Square			0.000
Univariate Log-Likelihood Chi-Square			0.000

Overall Univariate Pearson Chi-Square	0.000
Overall Univariate Log-Likelihood Chi-Square	0.000

Wszystkie reszty były nieistotne statystycznie, bo mieściły się pomiędzy (-2) i 2.
Stąd nie zostały złamane założenia.

Ocena reszt dwuzmiennowych BVR (Mplus – TECH10)

BIVARIATE MODEL FIT INFORMATION

Variable	Variable	Estimated Probabilities H1	H0	Standardized Residual (z-score)
A	B			
Category 1	Category 1	0.331	0.330	0.045
Category 1	Category 2	0.060	0.061	-0.089
Category 2	Category 1	0.492	0.492	-0.042
Category 2	Category 2	0.117	0.117	0.066
Bivariate Pearson Chi-Square				0.013
Bivariate Log-Likelihood Chi-Square				0.013
A	H			
Category 1	Category 1	0.087	0.087	0.000
Category 1	Category 2	0.304	0.304	0.000
Category 2	Category 1	0.201	0.201	0.000
Category 2	Category 2	0.408	0.408	0.000
Bivariate Pearson Chi-Square				0.000
Bivariate Log-Likelihood Chi-Square				0.000
A	D			
Category 1	Category 1	0.189	0.185	0.427
Category 1	Category 2	0.202	0.206	-0.410
Category 2	Category 1	0.297	0.301	-0.361
Category 2	Category 2	0.312	0.308	0.359
Bivariate Pearson Chi-Square				0.462
Bivariate Log-Likelihood Chi-Square				0.462
B	H			
Category 1	Category 1	0.257	0.257	0.000
Category 1	Category 2	0.565	0.565	0.000
Category 2	Category 1	0.031	0.031	0.000
Category 2	Category 2	0.147	0.147	0.000
Bivariate Pearson Chi-Square				0.000
Bivariate Log-Likelihood Chi-Square				0.000
B	D			
Category 1	Category 1	0.383	0.384	-0.091
Category 1	Category 2	0.440	0.439	0.089
Category 2	Category 1	0.103	0.102	0.146
Category 2	Category 2	0.074	0.075	-0.168
Bivariate Pearson Chi-Square				0.055
Bivariate Log-Likelihood Chi-Square				0.055
H	D			
Category 1	Category 1	0.136	0.136	0.001
Category 1	Category 2	0.151	0.151	-0.001
Category 2	Category 1	0.350	0.350	0.000
Category 2	Category 2	0.363	0.363	0.000
Bivariate Pearson Chi-Square				0.000
Bivariate Log-Likelihood Chi-Square				0.000

Overall Bivariate Pearson Chi-Square	0.530
Overall Bivariate Log-Likelihood Chi-Square	0.530

Wszystkie reszty dwuzmiennowe są nieistotne statystycznie (są mniejsze od 1), dlatego w tym modelu założenie lokalnej niezależności nie jest złamane. Stąd nie ma potrzeby aby wprowadzić nowe zmienne ukryte (dodatkowe czynniki).

Gdyby standaryzowane reszty były większe od 1 dla jednej pary zmiennych, wówczas należałoby wprowadzić jedną dodatkową zmienną ukrytą.

Gdyby standaryzowane reszty były większe od 1 dla dwóch par zmiennych, wówczas należałoby wprowadzić dwie dodatkowe zmienne ukryte.

Ocena jakości modelu LCA

CLASSIFICATION QUALITY

Entropy	0.724

Tutaj miara entropii dla oszacowanego modelu wynosi 0.724, czyli możemy stwierdzić, iż ten model cechuje się dobrą jakością klasyfikacji.

Miara entropii informuje czy model poprawnie klasyfikuje respondentów na podstawie wybranych klas. Wartości E>0,8 wskazują na wysoką jakość klasyfikacji, a E< 0.4 na niską jakość.

(Miara entropii nie mówi, czy model jest dobrze dopasowany. Miara ta nie może być podstawą do wyboru tego ilu klasowy ma być model, ponieważ modele o niskiej entropii mogą mieć dobre dopasowanie.)

Prawdopodobieństwo przynależności do klas ukrytych (trzeci sposób klasyfikacji)
FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP
Class Counts and Proportions

Latent Classes		
1	0	0.00000
2	175	0.10903
3	1430	0.89097

Zwracając uwagę na prawdopodobieństwo przynależności do klas ukrytych według jednego z trzech sposobów klasyfikacji należy stwierdzić, iż tak naprawdę istnieją dwie klasy. Stąd zerowe prawdopodobieństwa dla pierwszej klasy ukrytej.

Następnym krokiem w analizie byłaby zmiana modelu 3-klasowego na 2-klasowy. Jednakże tutaj pozostanę przy modelu 3-klasowym.

Average Latent Class Probabilities for Most Likely Latent Class Membership (Row) by Latent Class (Column) – Średnie prawdopodobieństwo klas ukrytych

	1	2	3
1	0.000	0.000	0.000
2	0.050	0.644	0.306
3	0.134	0.022	0.844

Prawdopodobieństwo, że respondent należy do klasy 2 po warunkiem, że został zaklasyfikowany do klasy 2 wynosi 0.644.
Prawdopodobieństwo, że respondent jest w klasie 3, ale ma etykietę z klasy 2 wynosi 0.022.
Prawdopodobieństwo, że respondent należy do klasy 3 pod warunkiem, że został zaklasyfikowany do klasy 3 wynosi 0.844.
Prawdopodobieństwo, że respondent należy do klasy 2 pod warunkiem, że został zaklasyfikowany do klasy 3 wynosi 0.306.

Na głównej przekątnej wartości powinny być wysokie (i tutaj są 0.644 oraz 0.306, oprócz wartości zerowej dla pierwszej klasy) natomiast poza główną przekątną niskie (tak jak tutaj). W takim przypadku model został poprawnie sklasyfikowany.
Tutaj problem jest dla pierwszej klasy, o czym mówią nam wartości zerowe w pierwszym wierszu. **Świadczy to o tym, iż ten model jest tak naprawdę 2-klasowy i należałoby taki oszacować**. (Jednakże tutaj pozostałam przy 3-klasowy, ponieważ wcześniejsze interpretacje dotyczyły tego modelu.)

Classification Probabilities for the Most Likely Latent Class Membership (Column) by Latent Class (Row) – Prawdopodobieństwo klasyfikacji dla najbardziej prawdopodobnych klas

	1	2	3
1	0.000	0.043	0.957
2	0.000	0.780	0.220
3	0.000	0.043	0.957

Prawdopodobieństwo, że respondent jest w klasie 1 przy założeniu, że jest w klasie 2 wynosi 0.043.

Prawdopodobieństwo, że respondent jest w klasie 2 przy założeniu, że jest w klasie 2 wynosi 0.78.

Prawdopodobieństwo, że respondent należy do klasy 2 przy założeniu, że jest w klasie 3 wynosi 0.22.

Ocena homogeniczności i separacji klas ukrytych

[TECHNICAL 7 OUTPUT] UNIVARIATE SAMPLE DISTRIBUTIONS

Variable	CLASS 1	CLASS 2	CLASS 3
A			
Category 1	0.981	0.342	0.303
Category 2	0.019	0.658	0.697
B			
Category 1	0.956	0.013	0.894
Category 2	0.044	0.987	0.106
H			
Category 1	0.000	0.000	0.367
Category 2	1.000	1.000	0.633
D			
Category 1	0.420	0.682	0.474
Category 2	0.580	0.318	0.526

HOMOGENICZNOŚĆ

Dość wysoki stopień homogeniczności i niski stopień separacji klas.

Variable	CLASS 1	CLASS 2	CLASS 3
A			
Category 1	0.981	0.342	0.303
Category 2	0.019	0.658	0.697
B			
Category 1	0.956	0.013	0.894
Category 2	0.044	0.987	0.106
H			
Category 1	0.000	0.000	0.367
Category 2	1.000	1.000	0.633
D			
Category 1	0.420	0.682	0.474
Category 2	0.580	0.318	0.526

SEPARACJA

Wielogrupowe modele LCA

Jednoczesna estymacja modeli z kowariantami

```
TITLE:  Analiza klas ukrytych z kowariantą

DATA:
 FILE IS "C:\Users\Agnieszka\Desktop\dane\daneee_2_binarne.txt";

VARIABLE:
 NAMES ARE A B H m1;
 USEVARIABLES ARE A B H m1;
 CATEGORICAL ARE A B H;
 CLASSES = c(3);
 !NOMINAL ARE A B H;
ANALYSIS:
 TYPE IS MIXTURE;
   !PROCESS = 4 (STARTS);
   optseed = 285380 ;
MODEL:
         %OVERALL%
         c ON m1;
Plot:
 type is plot3;
 series is A (1) B (2) H (3);
   !MONITOR = ON;
output: tech11 tech14;
Savedata:
 file is lca2_save.txt ;
 save is cprob;
```

Jednoczesna estymacja modeli z kowariantami może spowodować zmianę struktury klas i ich znaczeń → zmienne kowariancyjne jako zmienne dodatkowe (auxiliary variables)

LOGISTIC REGRESSION ODDS RATIO RESULTS

Categorical Latent Variables

C#1	ON	
M1		0.798
C#2	ON	
M1		0.415

Fakt bycia kobietą powoduje wzrost prawdopodobieństwa przynależności do pierwszej (1) klasy ukrytej o 0.798 w porównaniu do klasy referencyjnej, czyli klasy 3.

Fakt bycia kobietą powoduje wzrost prawdopodobieństwa przynależności do drugiej (2) klasy ukrytej o 0.415 w porównaniu do klasy referencyjnej, czyli klasy 3.

ALTERNATIVE PARAMETERIZATIONS FOR THE CATEGORICAL LATENT VARIABLE REGRESSION

Parameterization using Reference Class 1

C#2	ON				
M1		-0.654	0.503	-1.302	0.193
C#3	ON				
M1		0.225	0.383	0.589	0.556

Intercepts

Fakt bycia kobietą powoduje spadek prawdopodobieństwa przynależności do drugiej klasy ukrytej o exp(-0.654) w porównaniu do klasy referencyjnej, czyli klasy 1.

Fakt bycia kobietą powoduje wzrost prawdopodobieństwa przynależności do trzeciej klasy ukrytej o exp(0.225) w porównaniu do klasy referencyjnej, czyli klasy 1.

C#2	1.013	0.831	1.220	0.222
C#3	0.138	1.185	0.117	0.907

Parameterization using Reference Class 2

C#1 ON				
M1	0.654	0.503	1.302	0.193
C#3 ON				
M1	0.880	0.428	2.054	0.040

Intercepts

C#1	-1.013	0.831	-1.220	0.222
C#3	-0.875	1.835	-0.477	0.633

Fakt bycia kobietą powoduje wzrost prawdopodobieństwa przynależności do pierwszej klasy ukrytej o exp(0.654) w porównaniu do klasy referencyjnej, czyli klasy 2.

Fakt bycia kobietą powoduje wzrost prawdopodobieństwa przynależności do trzeciej klasy ukrytej o exp(0.88) w porównaniu do klasy referencyjnej, czyli klasy 2.

Podejście trójetapowe

```
DATA:
 FILE IS "C:\Users\Agnieszka\Desktop\dane\daneee_2_binarne.txt";

VARIABLE:
 NAMES ARE A B H m1;
 USEVARIABLES ARE A B H;
 CATEGORICAL ARE A B H;
 CLASSES = c(3);
 AUXILIARY = m1 (R3STEP);  ! traktuje zmieną jako predyktor, a nie jako zmienną zależną
 ! NOMINAL = A B H;

ANALYSIS:
 TYPE IS MIXTURE;
 ! ALGORITHM = INTEGRATION;
 ! STARTS 100 20;
 ! STITERATIONS = 10;
  OPTSEED = 253358;
  PROCESS = 4 (STARTS);
!MODEL:              ! Test inwariancji IRP
!%OVERALL%
! [P51$1](1);
! [P52$1](2);
! [P53$1](3);
! [P54$1](4);
! [P55$1](5);
! [P56$1](6);
 Plot:
  type is plot3;
  series is A (1) B (2) H (3);
  !MONITOR = ON;
 output: tech7 tech10 tech11 tech14;
 Savedata:
  file is lca1_save.txt ;
  save is cprob;
```

Podejście trójetapowe (1) oszacowanie modelu klas ukrytych (2) określenie przynależności dla klas na podstawie prawdopodobieństwa a posteriori (3) model regresji wielomianowej dodatkowych predyktorów:
- AUXILIARY = m1 (R) – metoda pseudo klas (PC) z wieloraką imputacją m1;
- AUXILIARY = m1 (E) – metoda pseudo klas (PC) z m1 jako zmienną zależną (distaloutcome);
- **AUXILIARY = m1 (R3STEP) – metoda trzyetapowa z m1 jako predyktorem**;
- AUXILIARY = m1 (DU3STEP/DE3STEP) – metody trzyetapowe z m1 jako zmienną zależną;
- AUXILIARY = m1 (DCON/DCAT) – metody Lanzy z m1 jako zmienną zależną (ciągłą lub kategorialną).

THE 3-STEP PROCEDURE

		Estimate	S.E.	Two-Tailed Est./S.E.	P-Value
C#1	ON				
M1		0.828	0.292	2.832	0.005
C#2	ON				
M1		0.771	0.308	2.503	0.012
Intercepts					
C#1		-0.092	0.431	-0.214	0.830
C#2		-1.320	0.462	-2.855	0.004

Klasa referencyjna: klasa 3.

Fakt bycia kobietą powoduje wzrost prawdopodobieństwa przynależności do klasy ukrytej 1 o exp(0.828) w porównaniu do klasy 3.

Fakt bycia kobietą powoduje wzrost prawdopodobieństwa przynależności do klasy ukrytej 2 o exp(0.771) w porównaniu do klasy 3.

(Wnioskując: W klasie 1 i klasie 2 przeważają kobiety w stosunku do klasy 3.)

Parameterization using Reference Class 1

C#2	ON				
M1		-0.057	0.222	-0.257	0.797
C#3	ON				
M1		-0.828	0.292	-2.832	0.005
Intercepts					
C#2		-1.228	0.366	-3.354	0.001
C#3		0.092	0.431	0.214	0.830

Klasa referencyjna: klasa 1.

Fakt bycia kobietą powoduje spadek prawdopodobieństwa przynależności do drugiej (2) klasy ukrytej o exp(-0.057) w porównaniu do klasy referencyjnej, czyli klasy 1.

Fakt bycia kobietą powoduje spadek prawdopodobieństwa przynależności do trzeciej (3) klasy ukrytej o exp(-0.828) w porównaniu do klasy referencyjnej, czyli klasy 1.

(Wnioskując: W klasie 3 przeważają mężczyźni w porównaniu z klasą 1.)

Parameterization using Reference Class 2

C#1	ON				
M1		0.057	0.222	0.257	0.797
C#3	ON				
M1		-0.771	0.308	-2.503	0.012
Intercepts					
C#1		1.228	0.366	3.354	0.001
C#3		1.320	0.462	2.855	0.004

Klasa referencyjna: klasa 2.

Fakt bycia kobietą powoduje wzrost prawdopodobieństwa przynależności do klasy ukrytej 1 o exp(0.057) w porównaniu do klasy referencyjnej, czyli klasy 2.

Fakt bycia kobietą powoduje spadek prawdopodobieństwa przynależności do klasy ukrytej 3 o exp(-0.771) w porównaniu do klasy referencyjnej, czyli klasy 2.

(Wnioskując: W klasie 3 w porównaniu do klasy pierwszej większość stanowią mężczyźni.)

Modele profili ukrytych

```
TITLE:  Analiza profili ukrytych (model mieszany);

DATA:
 FILE IS "C:\Users\Agnieszka\Desktop\dane_metryczne_4c.txt";

VARIABLE:
 NAMES ARE  AA BB HH DD m1 m2;
 USEVARIABLES ARE AA BB HH DD;
 !CATEGORICAL ARE AA BB HH DD;
 CLASSES = c(2);
 MISSING ARE ALL (-9);      !Braki danych zostały zastąpione wartością "-9".

ANALYSIS:
 TYPE IS MIXTURE;
 OPTSEED = 903420;

Plot:
   type is plot3;
   series is AA (1) BB (2) HH (3) DD(4);
  !MONITOR = ON;

Output: tech11 tech14;
Savedata:
   file is lca1_save.txt ;
   save is cprob;
```

!CATEGORICAL → dane nie są kategorialne, ale metryczne (dane ilościowe, tutaj dane na skali Likerta, które zostały poddane standaryzacji)

TECHNICAL 11 OUTPUT

Random Starts Specifications for the k-1 Class Analysis Model

Number of initial stage random starts	20
Number of final stage optimizations	4

VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 1 (H0) VERSUS 2 CLASSES

H0 Loglikelihood Value	-18384.752
2 Times the Loglikelihood Difference	626.174
Difference in the Number of Parameters	5
Mean	0.617
Standard Deviation	15.490
P-Value	0.0000

LO-MENDELL-RUBIN ADJUSTED LRT TEST

Value	609.654
P-Value	0.0000

p-value< 0.05

Model 2-klasowy jest nieistotnie lepszy od modelu 1-klasowego na poziomie istotności 5%.

TECHNICAL 14 OUTPUT

Random Starts Specifications for the k-1 Class Analysis Model
Number of initial stage random starts 20
Number of final stage optimizations 4

Random Starts Specification for the k-1 Class Model for Generated Data
Number of initial stage random starts 0
Number of final stage optimizations for the
initial stage random starts 0
Random Starts Specification for the k Class Model for Generated Data
Number of initial stage random starts 40
Number of final stage optimizations 8
Number of bootstrap draws requested Varies

PARAMETRIC BOOTSTRAPPED LIKELIHOOD RATIO TEST FOR 1 (H0) VERSUS 2 CLASSES
H0 Loglikelihood Value -18384.752
2 Times the Loglikelihood Difference 626.174
Difference in the Number of Parameters 5
Approximate P-Value 0.0000
Successful Bootstrap Draws 5

p-value<0.05

Model 2-klasowy jest nieistotnie lepszy od modelu 1-klasowego na poziomie istotności 5%.

WARNING: OF THE 5 BOOTSTRAP DRAWS, 4 DRAWS HAD BOTH A SMALLER LRT VALUE THAN THE OBSERVED LRT VALUE AND NOT A REPLICATED BEST LOGLIKELIHOOD VALUE FOR THE 2-CLASS MODEL.
THIS MEANS THAT THE P-VALUE MAY NOT BE TRUSTWORTHY DUE TO LOCAL MAXIMA.
INCREASE THE NUMBER OF RANDOM STARTS USING THE LRTSTARTS OPTION.

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Latent Class 1				
Means				
AA	-2.202	0.110	-20.007	0.000
BB	17.151	1.980	8.662	0.000
HH	-0.380	0.064	-5.907	0.000
DD	54.639	3.245	16.836	0.000
Variances				
AA	0.274	0.012	23.079	0.000
BB	625.763	16.708	37.452	0.000
HH	0.307	0.011	26.861	0.000
DD	886.326	18.535	47.819	0.000
Latent Class 2				

Ładunki czynnikowe dla zmiennych AA i HH są ujemne lub bliskie zero.

Means

AA	0.023	0.014	1.621	0.105
BB	39.948	0.659	60.577	0.000
HH	-0.064	0.015	-4.388	0.000
DD	50.567	0.771	65.628	0.000

Ładunki czynnikowe dla zmiennych AA i HH są ujemne lub bliskie zero. Podobnie jak dla klasy 1. Stąd być może należałoby usunąć z modelu te dwie zmienne i wówczas otrzymalibyśmy lepszy model.

Variances

AA	0.274	0.012	23.079	0.000
BB	625.763	16.708	37.452	0.000
HH	0.307	0.011	26.861	0.000
DD	886.326	18.535	47.819	0.000

Categorical Latent Variables

Means

C#1	-2.626	0.122	-21.445	0.000

Średnie dla klasy pierwszej.
Średnie dla klasy drugiej są zerowe.

Mieszane modele czynnikowe

```
TITLE:  Czynnikowy model mieszany (factor mixture) z scisla inwariancja pomiaru
DATA:
 FILE IS "C:\Users\Agnieszka\Desktop\dane_metryczne_4c.txt";

VARIABLE:
 NAMES ARE  AA BB HH DD m1 m2;
 USEVARIABLES ARE AA BB HH DD;
 CLASSES = c(2);

ANALYSIS:
 TYPE IS MIXTURE;
 OPTSEED = 27071;

MODEL:
%overall%

F1 by AA* BB HH DD;
F1@1;

Plot:
 type is plot3;
 !  MONITOR = ON;
Output: tech11 tech14;
Savedata:
 file is lca1_save.txt ;
 save is cprob;
```

TECHNICAL 11 OUTPUT

Random Starts Specifications for the k-1 Class Analysis Model

Number of initial stage random starts	20
Number of final stage optimizations	4

VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 1 (H0) VERSUS 2 CLASSES

H0 Loglikelihood Value	-19585.122
2 Times the Loglikelihood Difference	1801.905
Difference in the Number of Parameters	2
Mean	-666.262
Standard Deviation	959.510
P-Value	0.0000

LO-MENDELL-RUBIN ADJUSTED LRT TEST

Value	1687.584
P-Value	0.0000

p-value< 0.05

Model 2-klasowy jest nieistotnie lepszy od modelu 1-klasowego na poziomie istotności 5%.

TECHNICAL 14 OUTPUT

Random Starts Specifications for the k-1 Class Analysis Model

Number of initial stage random starts	20
Number of final stage optimizations	4

Random Starts Specification for the k-1 Class Model for Generated Data

Number of initial stage random starts	0
Number of final stage optimizations for the initial stage random starts	0

Random Starts Specification for the k Class Model for Generated Data

Number of initial stage random starts	40
Number of final stage optimizations	8
Number of bootstrap draws requested	Varies

PARAMETRIC BOOTSTRAPPED LIKELIHOOD RATIO TEST FOR 1 (H0) VERSUS 2 CLASSES

H0 Loglikelihood Value	-19585.122
2 Times the Loglikelihood Difference	1801.905
Difference in the Number of Parameters	2
Approximate P-Value	0.0000
Successful Bootstrap Draws	5

p-value<0.05

Model 2-klasowy jest nieistotnie lepszy od modelu 1-klasowego na poziomie istotności 5%.

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Latent Class 1				
F1 BY				
AA	-0.062	0.024	-2.622	0.009
BB	-1.179	0.526	-2.241	0.025
HH	-0.743	0.292	-2.548	0.011
DD	-2.001	1.073	-1.864	0.062
Means				
F1	11.997	4.703	2.551	0.011
Intercepts				
AA	-0.141	0.021	-6.602	0.000
BB	38.550	0.644	59.874	0.000
HH	-0.086	0.014	-6.090	0.000
DD	51.082	0.758	67.414	0.000

Ładunki czynnikowe dla wszystkich zmiennych są ujemne lub bliskie zero.
Wniosek: W modelu nie ma żadnej zmiennej ukrytej/czynnika wyjaśniającego wszystkie zmienne.

Variances

F1	1.000	0.000	999.000	999.000

Residual Variances

AA	0.769	0.098	7.816	0.000
BB	655.115	17.288	37.895	0.000
HH	-0.242	0.433	-0.559	0.576
DD	877.675	21.963	39.961	0.000

Latent Class 2

F1 BY

AA	-0.062	0.024	-2.622	0.009
BB	-1.179	0.526	-2.241	0.025
HH	-0.743	0.292	-2.548	0.011
DD	-2.001	1.073	-1.864	0.062

Means

F1	0.000	0.000	999.000	999.000

Intercepts

AA	-0.141	0.021	-6.602	0.000
BB	38.550	0.644	59.874	0.000
HH	-0.086	0.014	-6.090	0.000
DD	51.082	0.758	67.414	0.000

Variances

F1	1.000	0.000	999.000	999.000

Residual Variances

AA	0.769	0.098	7.816	0.000
BB	655.115	17.288	37.895	0.000
HH	-0.242	0.433	-0.559	0.576
DD	877.675	21.963	39.961	0.000

Categorical Latent Variables

Means

C#1	-4.598	0.251	-18.301	0.000

Gdyby tylko dla jednej zmiennej ładunek byłby bliski zero lub ujemny, wówczas należałoby usunąć tą zmienną i ponownie oszacować model.
Jednakże gdyby w tym modelu występowały dwa lub więcej ujemnych, bądź bliskich zera ładunków (tutaj jest ich 4) to nie można było by zrekonstruować tego modelu. Analiza czynnikowa ma sens wówczas gdy w modelu występują przynajmniej 3 zmienne. (Ewentualnie należałoby dobrać do modelu inne zmienne.)