Multivariate: Latent class / profile analysis

Table of contents

1 Goals

1.1 Goals

1.1.1 Goals of this lecture

- Latent class / profile analysis (LCA / LPA)
  - **Dimension reduction**: reduce number of variables
- A **large set of (potentially correlated) observed variables**
  - **Discrete** classes or profiles
  - Different **patterns** of responses

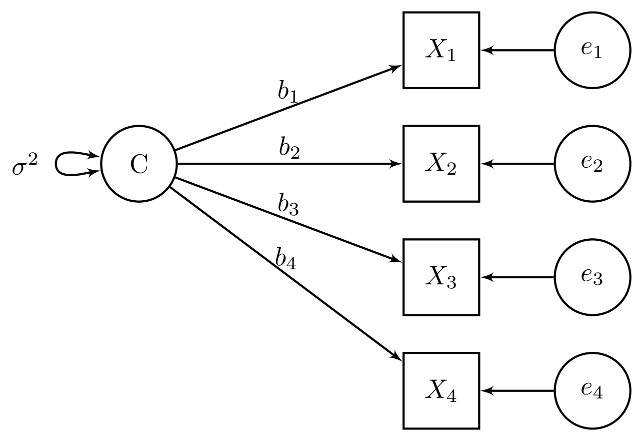
2 Latent class analysis

2.1 Latent class analysis

2.1.1 Latent class analysis

- Latent variable technique that **classifies** people into *previously unknown subgroups* based on their responses to a set of variables
  - Latent class = unknown / unobserved subgroups
  - LCA = only categorical variables; LPA = otherwise
- Groups of people have different *patterns of responses*
  - *Not a single value* on a *single variable*
  - e.g., people in *group 1* are **high on items 1-3**, **low on items 4-6** and people in *group 2* are **low on items 1-2** and **high on items 3-6**

2.1.2 What it looks like



- Latent class (C) is **categorical**: Typically nominal, not ordinal
- X_1 through X_4 are **indicators** of the latent class
- Note that the latent variable **causes** the indicator values

2.1.3 LCA is...

1. Data reduction
   Large number of items → smaller number of classes

2. Mixture model
   Population is a "mixture" of subgroups, LCA "unmixes" them

3. Person-oriented approach
   How do **people** have different *patterns*, rather than how are the **variables** related to one another for all *people*

2.1.4 LCA is similar to...

4. Factor analysis / measurement model
   Very similar, but latent factor is continuous while **latent class is categorical**

2.1.5 Selecting the number of classes

- We have some number of patterns (classes)?
  - How do we decide how many?
- Similar to exploratory factor analysis (EFA) methods
  - Run several models with different numbers of classes
  - Compare these models in terms of **model fit** and **theory**
  - Choose the model that has the **best fit** and **makes the most sense**

2.1.6 Model fit for LCA

- **Chi-square test**
  - Problematic for large samples, do not rely on it
- **Likelihood ratio test** (also bootstrap LR test)
  - Same as previous LR tests, compare two models
- **AIC and BIC**
  - Smaller is better, not a measure of absolute fit
- **Entropy** - certainty of classification
  - Ranges from 0 to 1, closer to 1 is better

- **Predicted vs observed means / probabilities**
  - Mean for each item within each class - do they match?

2.1.7 Model fit for LCA

- BIC and LR test are best for **choosing the number of classes**
- AIC and entropy are bad for choosing the correct number of classes
  - You can *report* AIC and entropy – they tell you about the model – but **don't use them to decide on the number of classes**

2.2 LCA example 1

2.2.1 Mplus example 7.9 (modified)

4 continuous indicators, 2 classes

```
DATA:
FILE IS ex7.9.dat;

VARIABLE:
NAMES ARE y1-y4 x;
USEVARIABLES ARE y1-y4;
CLASSES = c (2);

ANALYSIS:
TYPE = MIXTURE;

PLOT:
TYPE=PLOT3;
SERIES = y1(1) y2(2) y3(3) y4(4);
```

2.2.2 Model-implied response pattern

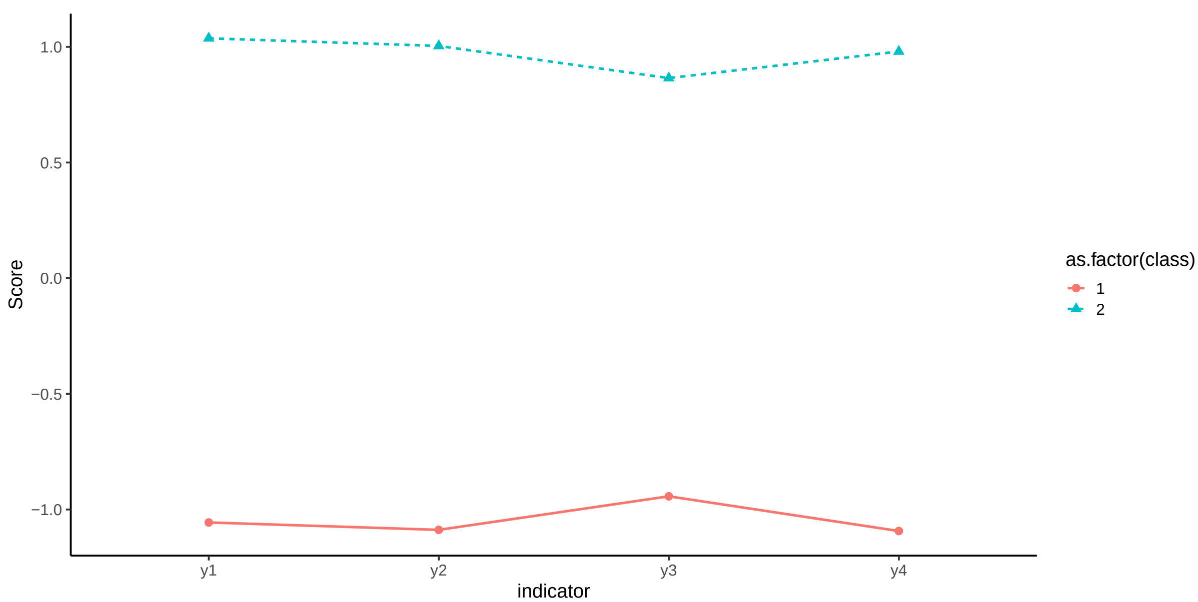
```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

Latent Class 1
```

```
 Means
    Y1                -1.056      0.070    -15.030      0.000
    Y2                -1.088      0.067    -16.255      0.000
    Y3                -0.943      0.063    -15.050      0.000
    Y4                -1.093      0.074    -14.688      0.000

...

Latent Class 2

 Means
    Y1                 1.037      0.070     14.762      0.000
    Y2                 1.004      0.064     15.682      0.000
    Y3                 0.865      0.068     12.763      0.000
    Y4                 0.980      0.060     16.321      0.000

...
```

2.2.3 Model-implied response pattern

GRAPH → VIEW GRAPHS → ESTIMATED MEANS (Windows only)

2.2.4 Model-implied response pattern



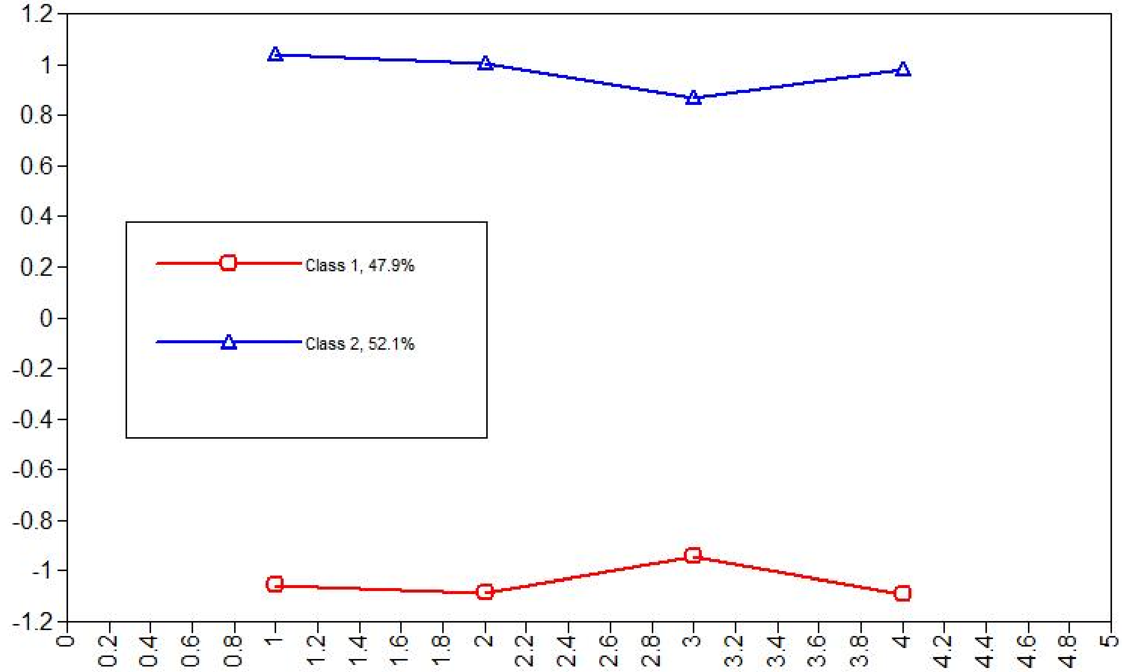


Figure 1: Response pattern

2.3 LCA example 2

2.3.1 Another example: Video game preferences

- Quaiser-Pohl, C., Geiser, C., & Lehmann, W. (2006). The relationship between computer-game preference, gender, and mental-rotation ability. Personality and Individual differences, 40(3), 609-619.
  - Classify people into types of video game players
  - 8 binary indicators (0 = never or rarely, 1 = often or very often)
  - 3 class solution

2.3.2 Indicators

- How often do you play the following types of computer games?
  1. Adventure
  2. Action
  3. Sport
  4. Fantasy role playing

5. Logic
6. Skill training
7. Simulation
8. Driving simulation

2.3.3 Mplus syntax

```
DATA: FILE = computer_games.dat;

VARIABLE: NAMES = gender c1-c8;
          USEVARIABLES = c1-c8;
          CATEGORICAL = c1-c8;
          CLASSES = L(3);

ANALYSIS: TYPE = MIXTURE;

PLOT:   TYPE = PLOT3;
        SERIES = c1(1) c2(2) c3(3) c4(4)
                 c5(5) c6(6) c7(7) c8(8);

SAVEDATA: FILE = computer_games_3_classes.dat;
          SAVE = CPROBABILITIES;
```

2.3.4 Model-implied response pattern

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

Latent Class 1

 Thresholds
    C1$1               0.555      0.144      3.852      0.000
    C2$1              -0.591      0.191     -3.101      0.002
    C3$1              -0.097      0.163     -0.593      0.553
    C4$1               0.463      0.143      3.244      0.001
    C5$1               1.154      0.247      4.680      0.000
    C6$1               1.542      0.291      5.290      0.000
    C7$1              -0.628      0.152     -4.134      0.000
    C8$1              -0.576      0.159     -3.633      0.000
```

```
Latent Class 2

 Thresholds
    C1$1               1.398      0.506      2.761      0.006
    C2$1               1.920      0.706      2.717      0.007
    C3$1               1.675      0.423      3.961      0.000
    C4$1               1.489      0.523      2.850      0.004
    C5$1              -0.981      0.406     -2.413      0.016
    C6$1              -1.480      0.654     -2.263      0.024
    C7$1               0.187      0.357      0.524      0.600
    C8$1               0.685      0.334      2.052      0.040

Latent Class 3

 Thresholds
    C1$1               4.582      0.780      5.872      0.000
    C2$1               4.334      0.804      5.388      0.000
    C3$1               2.723      0.322      8.455      0.000
    C4$1               4.583      0.927      4.947      0.000
    C5$1               2.216      0.300      7.391      0.000
    C6$1               2.166      0.366      5.924      0.000
    C7$1               2.873      0.326      8.817      0.000
    C8$1               2.845      0.410      6.937      0.000

...

RESULTS IN PROBABILITY SCALE

Latent Class 1

 C1
    Category 1         0.635      0.033     19.021      0.000
    Category 2         0.365      0.033     10.915      0.000
 C2
    Category 1         0.356      0.044      8.152      0.000
    Category 2         0.644      0.044     14.721      0.000
 C3
    Category 1         0.476      0.041     11.727      0.000
    Category 2         0.524      0.041     12.916      0.000
 C4
    Category 1         0.614      0.034     18.144      0.000
    Category 2         0.386      0.034     11.423      0.000
```

```
 C5
    Category 1         0.760      0.045     16.915      0.000
    Category 2         0.240      0.045      5.332      0.000
 C6
    Category 1         0.824      0.042     19.465      0.000
    Category 2         0.176      0.042      4.165      0.000
 C7
    Category 1         0.348      0.034     10.094      0.000
    Category 2         0.652      0.034     18.917      0.000
 C8
    Category 1         0.360      0.037      9.847      0.000
    Category 2         0.640      0.037     17.521      0.000

Latent Class 2

 C1
    Category 1         0.802      0.080      9.967      0.000
    Category 2         0.198      0.080      2.463      0.014
 C2
    Category 1         0.872      0.079     11.068      0.000
    Category 2         0.128      0.079      1.623      0.105
 C3
    Category 1         0.842      0.056     14.991      0.000
    Category 2         0.158      0.056      2.808      0.005
 C4
    Category 1         0.816      0.078     10.399      0.000
    Category 2         0.184      0.078      2.345      0.019
 C5
    Category 1         0.273      0.081      3.383      0.001
    Category 2         0.727      0.081      9.020      0.000
 C6
    Category 1         0.185      0.099      1.877      0.060
    Category 2         0.815      0.099      8.247      0.000
 C7
    Category 1         0.547      0.088      6.179      0.000
    Category 2         0.453      0.088      5.125      0.000
 C8
    Category 1         0.665      0.074      8.941      0.000
    Category 2         0.335      0.074      4.508      0.000

Latent Class 3

 C1
```

```
   Category 1          0.990      0.008    126.531      0.000
   Category 2          0.010      0.008      1.295      0.195
C2
   Category 1          0.987      0.010     96.062      0.000
   Category 2          0.013      0.010      1.259      0.208
C3
   Category 1          0.938      0.019     50.383      0.000
   Category 2          0.062      0.019      3.309      0.001
C4
   Category 1          0.990      0.009    106.680      0.000
   Category 2          0.010      0.009      1.090      0.276
C5
   Category 1          0.902      0.027     33.919      0.000
   Category 2          0.098      0.027      3.699      0.000
C6
   Category 1          0.897      0.034     26.596      0.000
   Category 2          0.103      0.034      3.049      0.002
C7
   Category 1          0.947      0.017     57.362      0.000
   Category 2          0.053      0.017      3.242      0.001
C8
   Category 1          0.945      0.021     44.386      0.000
   Category 2          0.055      0.021      2.580      0.010
```

2.3.5 Model-implied response pattern

GRAPH → VIEW GRAPHS → ESTIMATED PROBABILITIES (Windows only)

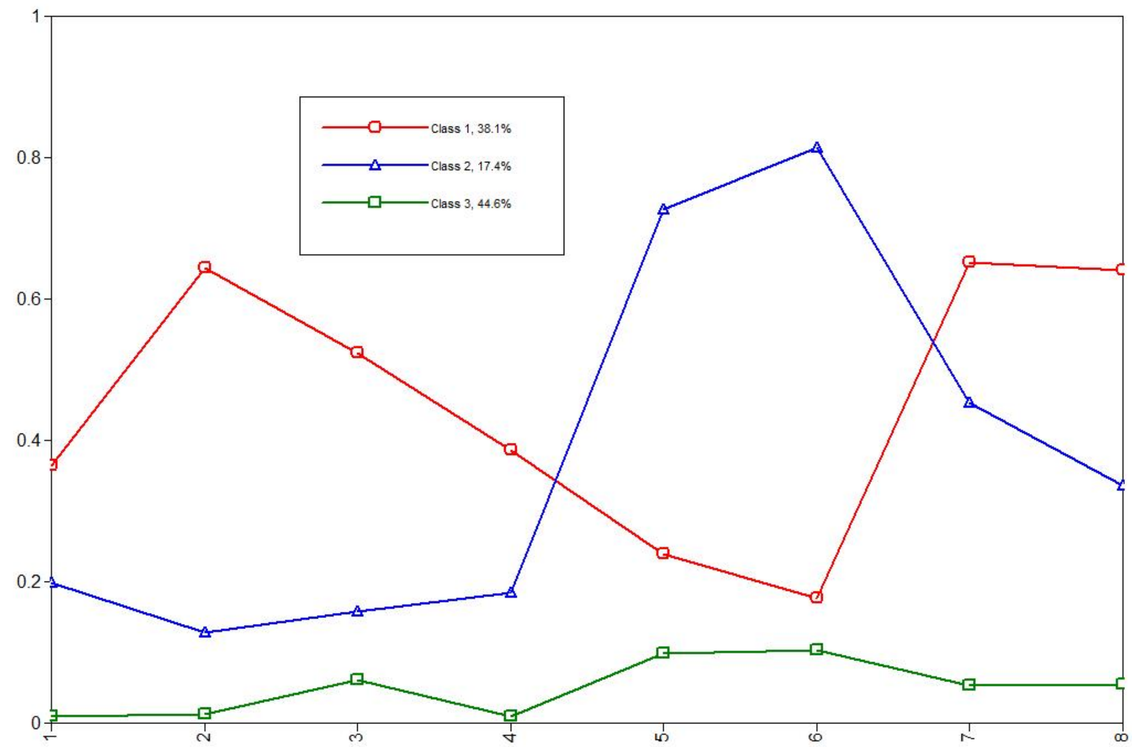
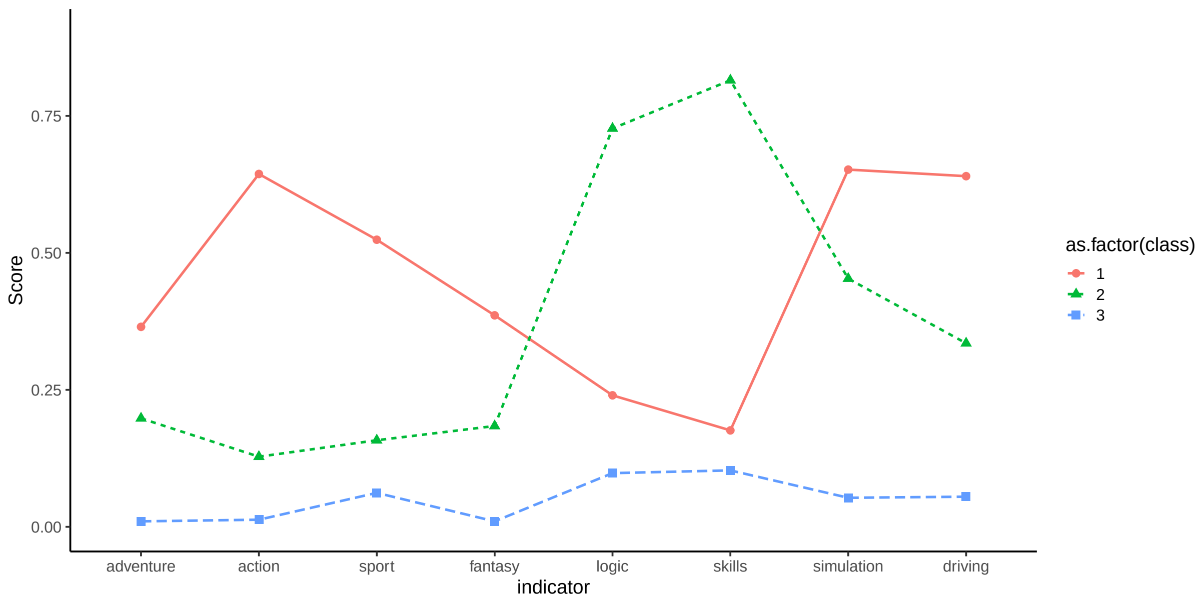
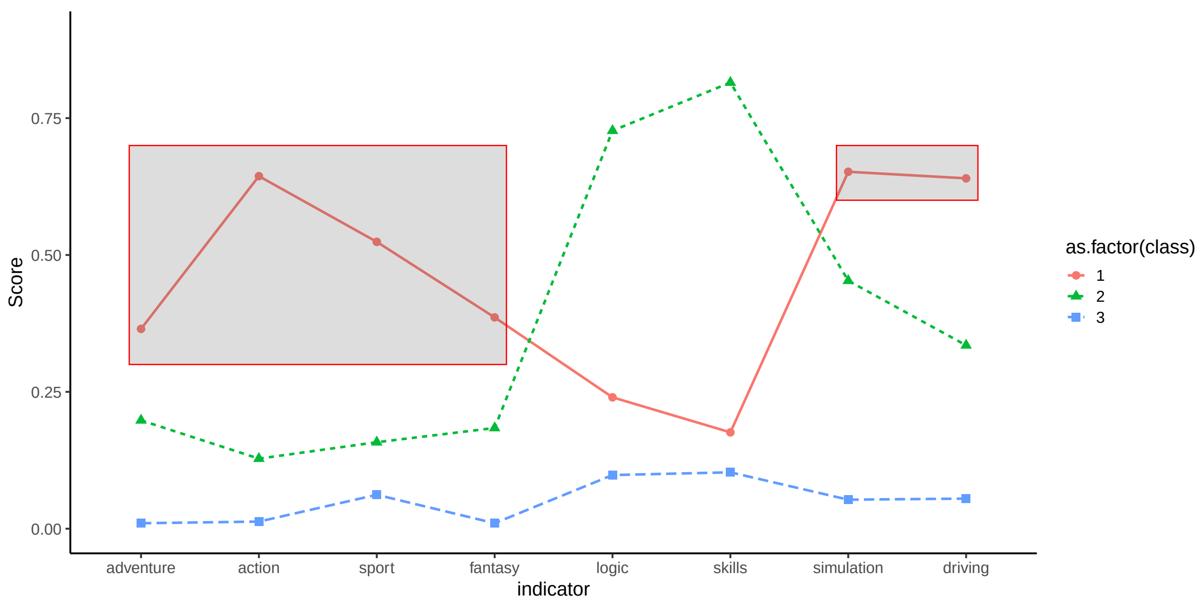


Figure 2: Three class binary

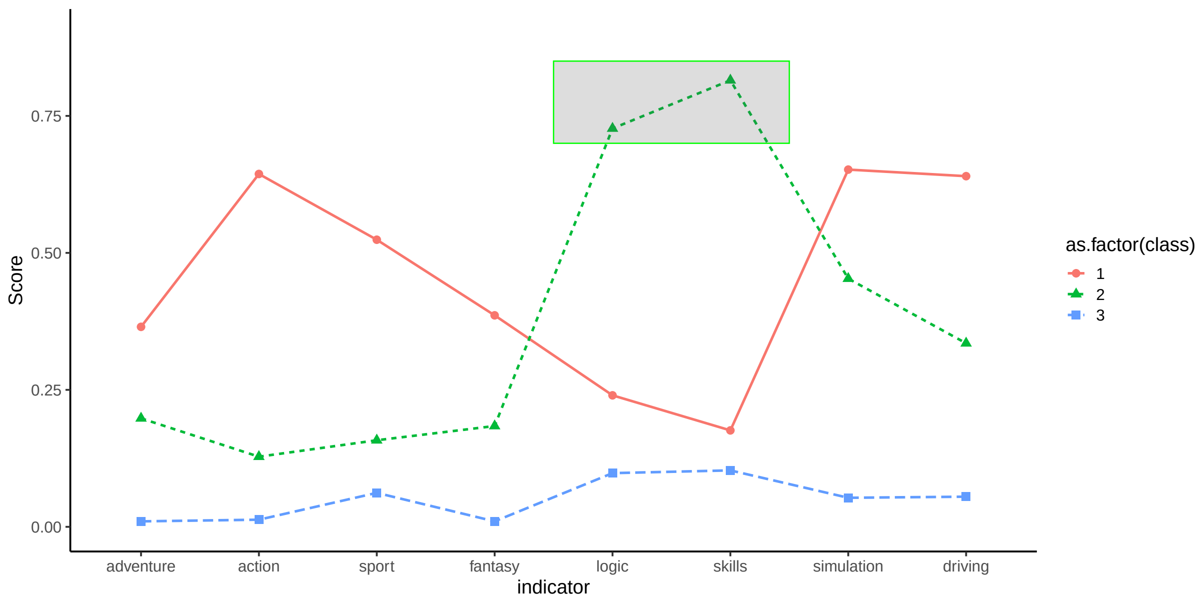
2.3.6 Model-implied response pattern



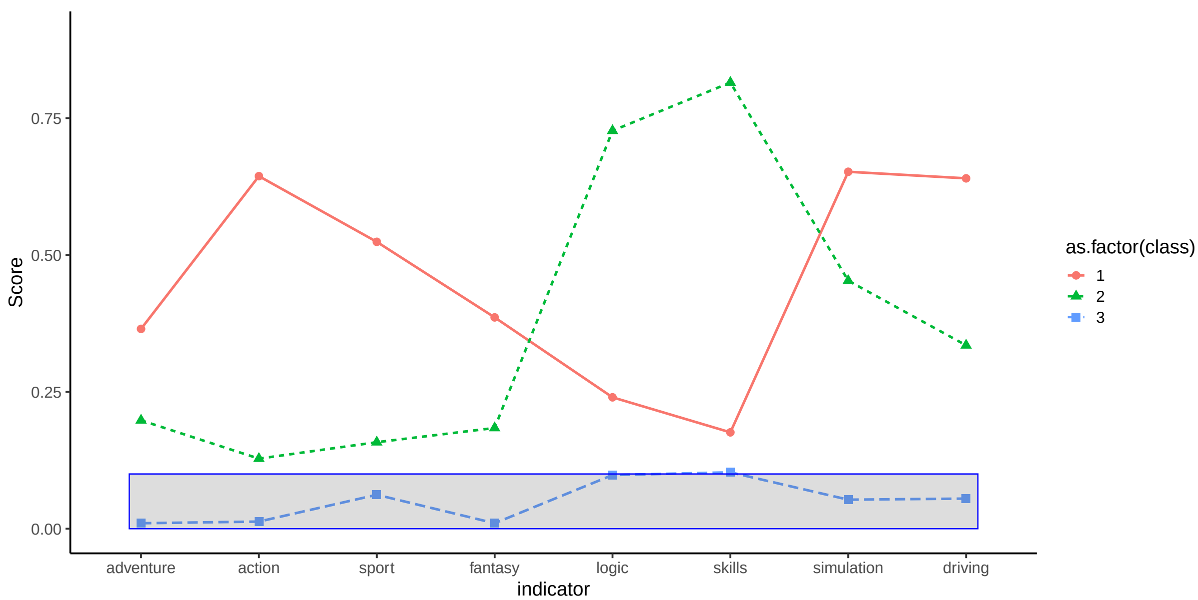
2.3.7 Model-implied response pattern



2.3.8 Model-implied response pattern



2.3.9 Model-implied response pattern



2.4 LPA example 3

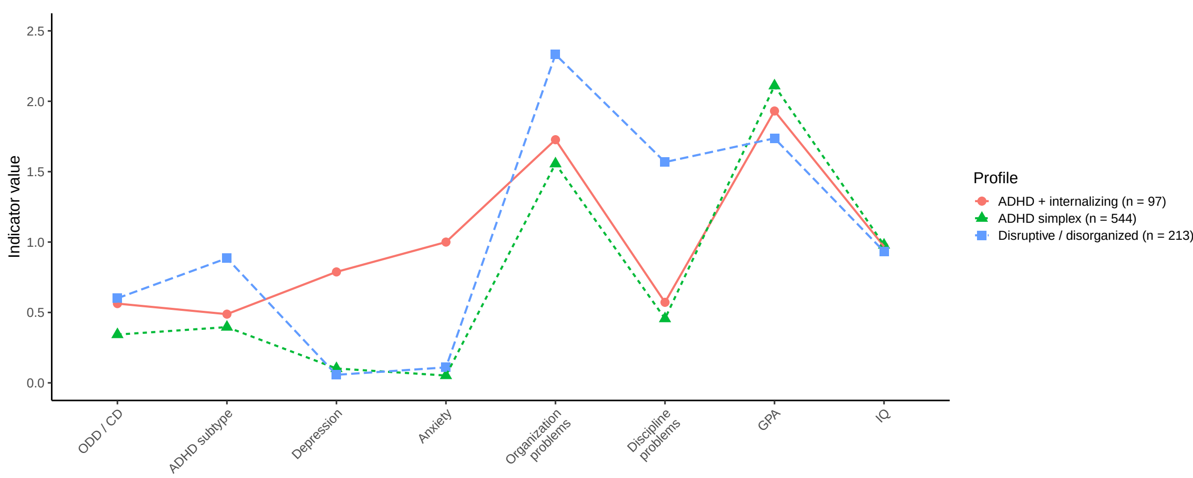
2.4.1 Another example: ADHD subtypes

- Coxe, S., Sibley, M. H., & Becker, S. P. (2021). Presenting problem profiles for adolescents with ADHD: differences by sex, age, race, and family adversity. Child and Adolescent Mental Health, 26(3), 228-237.
  - Characterize adolescents seeking treatment for ADHD
  - 8 indicators (next slide)
  - 3 class solution

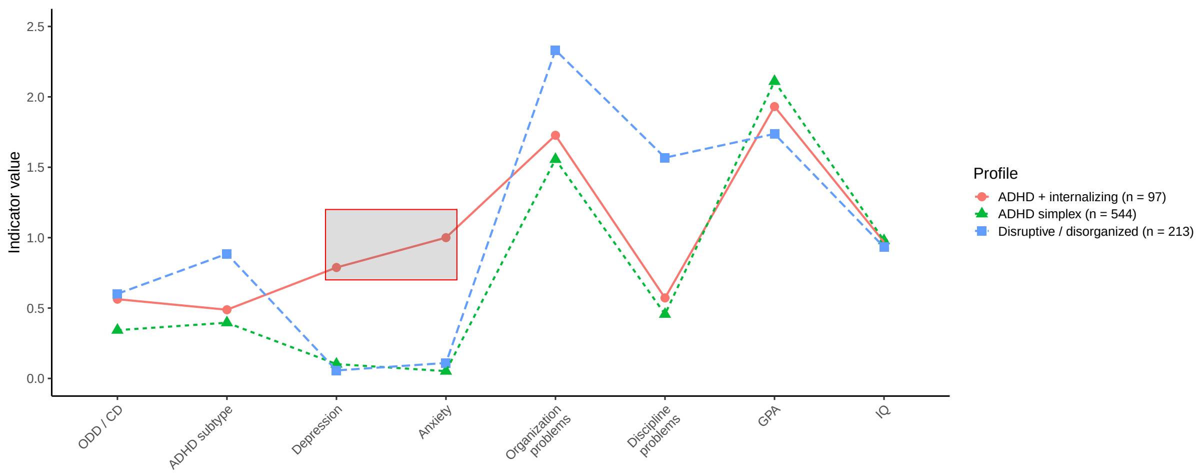
2.4.2 Indicators

- **Opposition-defiant disorder / conduct disorder**: Binary
- **ADHD combined subtype** (inattention and hyperactivity): Binary
- **Depression diagnosis**: Binary
- **Anxiety diagnosis**: Binary
- **Organization problems**: 0 to 3 (higher is worse, continuous)
- **Discipline problems**: 0 to 3 (higher is worse, continuous)
- **GPA**: 0 to 4 (higher is better)
- **IQ**: divided by 100 (population mean = 1.0, population SD = 0.15)

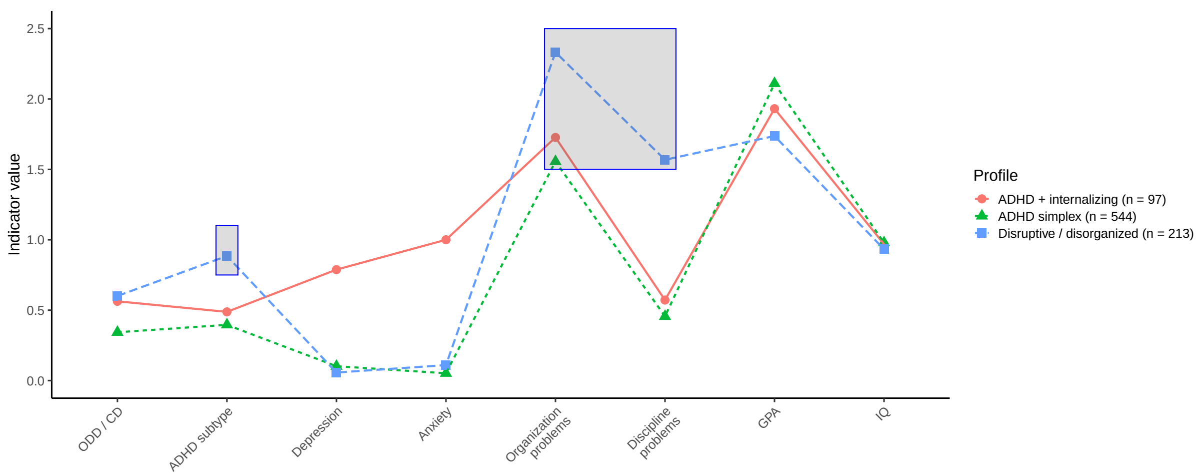
2.4.3 Model-implied response pattern



2.4.4 Model-implied response pattern



2.4.5 Model-implied response pattern



3 Conclusion

3.1 Summary of this week

3.1.1 Summary of this week

- Latent class / profile analysis
  - Latent or unknown groups

  - Discrete groups
    * Generally not just high, medium, low
    * Patterns of responses

- Conceptually similar to factor analysis

  - Only with a discrete factor

3.2 Summary of this section

3.2.1 Summary of this section

- There are a lot of ways to **reduce the dimension** of your data
- Which one you use depends on (among other things)

  - What you think about measurement error
  - Which way the causal arrow is pointing
  - Whether the reduced dimension is continuous or categorical