SurvMeth 687 HW 3

Chia Wen Cheng

2023-11-21

Setup

```r
# set working direction to where the data set was stored locally
setwd("G:/My Drive/0. study abroad/academic/10. 2023 Fall/5. SurvMeth 687 Applications of Statistical Mo
# load required packages
library(ggcorrplot)
library(tidyverse)
library(dplyr)
library(haven) # for read_dta()
library(ggplot2)
library(RColorBrewer)
library(lavaan)
library(semTools)
library(poLCA)
# read the data set from STATA data
ess <- read_dta('ess_belgium.dta') ## 1703 obs. of 26 var.
data <- ess[, -c(1:4, 7, 10, 12:18, 22:24)]
```

For the three TRST variables: −categorical; 00 No trust at all; 01 1; 02 2; 03 3; 04 4; 05 5; 06 6; 07 7; 08 8; 09 9; 10 Complete trust; 77 Refusal; 88 Don't know; 99 No answer For the three IM variables: −categorical; 1 Allow many to come and live here; 2 Allow some; 3 Allow a few; 4 Allow none; 7 Refusal; 8 Don't know; 9 No answer For the STF variable: −categorical; 00 Extremely dissatisfied; 01 1; 02 2; 03 3; 04 4; 05 5; 06 6; 07 7; 08 8; 09 9; 10 Extremely satisfied; 77 Refusal; 88 Don't know; 99 No answer

1. Provide a brief description of the variables of interest using tables or graphs. Are there any missing data? Justify the way that you handle the missing values.

```r
# examine whether any columns contain NAs
apply(is.na(ess), 2, sum) ## there is no NAs
```

```
##    idno  intnum pplfair  pplhlp trstprl trstlgl trstplc trstplt trstprt  stfeco
##       0       0       0       0       0       0       0       0       0       0
##  stfgov  stfdem  stfedu stfhlth gincdif freehms prtyban scnsenv imsmetn imdfetn
##       0       0       0       0       0       0       0       0       0       0
## impcntr imbgeco imueclt imwbcnt  int_rr  weight
##       0       0       0       0       0       0
```

```r
table(data$trstprl) ## invalid: 88
```

```
##
##   0   1   2   3   4   5   6   7   8   9  10  88
## 127  72 149 192 190 383 250 193 102  14   8  23
```

```r
table(data$trstlgl) ## invalid: 88
```

```
##
##   0   1   2   3   4   5   6   7   8   9  10  88
## 101  61 113 171 187 343 239 278 151  30  24   5
```

```r
table(data$trstplt) ## invalid: 88
```

```
##
##   0   1   2   3   4   5   6   7   8   9  10  88
## 174 101 204 227 241 344 219 142  33   9   2   7
```

```r
table(data$trstprt) ## invalid: 88
```

```
##
##   0   1   2   3   4   5   6   7   8   9  10  88
## 170 101 199 246 230 354 214 133  37   5   4  10
```

```r
table(data$imsmetn) ## invalid: 8
```

```
##
##   1   2   3   4   8
## 275 903 369 149   7
```

```r
table(data$imdfetn) ## invalid: 8
```

```
##
##   1   2   3   4   8
## 157 749 502 285  10
```

```r
table(data$impcntr) ## invalid: 8
```

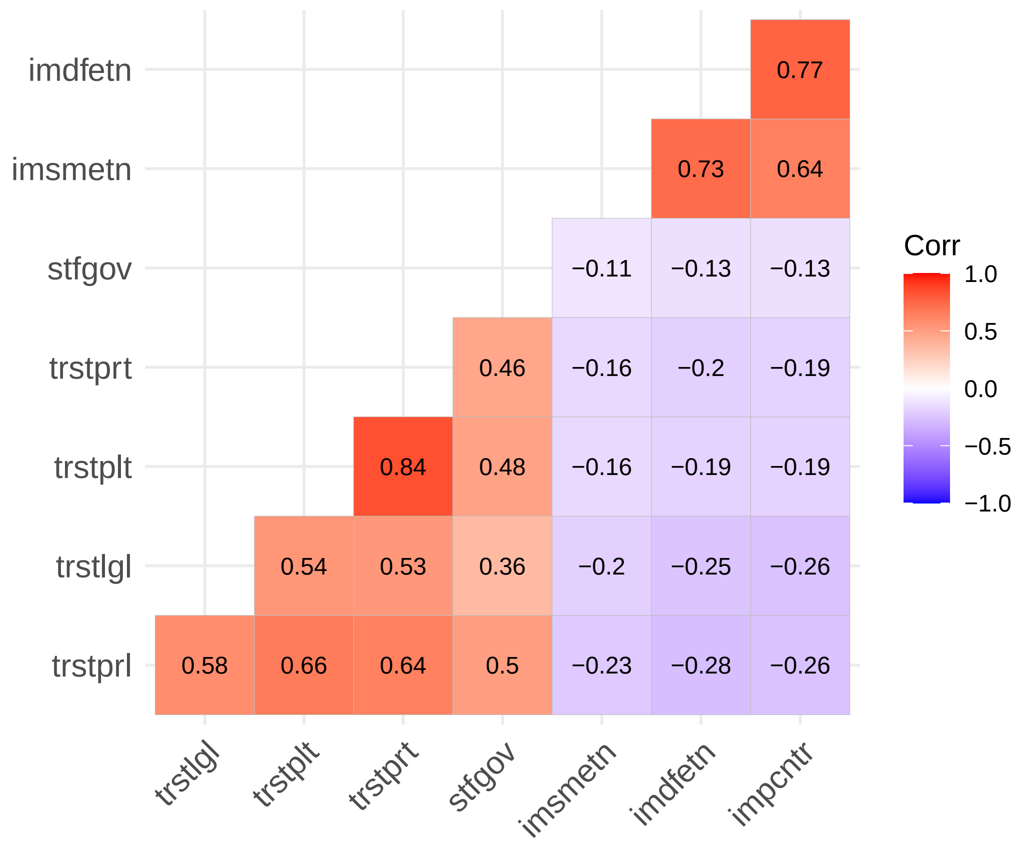
```
##
##   1   2   3   4   8
## 146 755 534 255  13
```

```r
table(data$stfgov) ## invalid: 88
```

```
##
##   0   1   2   3   4   5   6   7   8   9  10  88
## 207 117 247 254 216 265 160 137  60  13   7  20
```

```r
# drop the refusals, DK's, and no answers for all variables
data1 <- data %>%
  subset(trstprl != 88) %>%
  subset(trstlgl != 88) %>%
  subset(trstplt != 88) %>%
  subset(trstprt != 88) %>%
  subset(imsmetn != 8) %>%
  subset(imdfetn != 8) %>%
  subset(impcntr != 8) %>%
  subset(stfgov != 88)

# descriptive statistics
## subsetting a new data frame with only variables of interest
data1_ind <- data1[, -c(9:10)]
## plotting the correlations between variables of interest
model.matrix(~0+., data = data1_ind) %>%
  cor(use = "pairwise.complete.obs") %>%
  ggcorrplot(show.diag = FALSE, type = "lower",
             lab = TRUE, lab_size = 3)
```



Initially, I examined whether there were NAs and found none. I then scrutinized the presence of invalid numbers in each variable of interest and, surprisingly, identified only one invalid value for each variable. Subsequently, I utilized the `subset()` function to remove rows containing these invalid values in each column, facilitating further analysis. Instead of imputing values, I opted to drop these rows, considering their negligible representation–only 60 out of 1703 observations were affected by this removal.

I proposed that the level of trust in the political system and attitudes toward immigration policies might serve as predictors of government satisfaction. Additionally, I anticipated a correlation between attitudes toward immigration policies and the level of trust in the political system. The correlation matrix plot revealed strong and positive correlations between pairs such as **trstprt** and **trstplr**, **imsmetn** and **imdfetn**, as well as **imdfetn** and **impcntr**. Half of the remaining pairs, including those related to the trust in the political system and satisfaction with the government, showed moderate and positive correlations. The other half, encompassing pairs discussing attitudes toward immigration policies and government satisfaction, as well as attitudes toward immigration policies and trust in the political system, exhibited weak and negative correlations.

2. Create a dichotomous variable to indicate satisfaction with the national government: STF_IND = 0 if STFGOV<=5, and STF_IND = 1 otherwise.

```
# create a dichotomous variable from stfgov
data1["stf_ind"] <- ifelse(data1["stfgov"] <= 5, 0, 1)
# relocate variable locations to streamline the order in the prompts
data1 <- data1 %>%
  relocate(stfgov, .after = impcntr) %>%
  relocate(stf_ind, .after = stfgov)
```

3. Perform a multiple group analysis to examine invariance in the way that a latent construct of trust in leadership is measured between individuals who are satisfied with the government and those who are not. What is your conclusion?

```
# create measurement model (there is no structural model b/c only one latent?)
model_3a <- '
  # Measurement model for trust in leadership
    trstlead =~ trstprl + trstlgl + trstplt + trstprt
'
# fit the model to the overall sample
fit_3 <- cfa(model_3a, data = data1, group = "stf_ind")
summary(fit_3)
```

```
## lavaan 0.6.16 ended normally after 56 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        24
##
##   Number of observations per group:
##     1                                              370
##     0                                             1273
##
## Model Test User Model:
##
##   Test statistic                               123.267
##   Degrees of freedom                                 4
##   P-value (Chi-square)                           0.000
##   Test statistic for each group:
##     1                                           47.911
##     0                                           75.356
##
```

```
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
##
## Group 1 [1]:
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   trstlead =~
##     trstprl           1.000
##     trstlgl           0.876    0.120    7.285    0.000
##     trstplt           1.944    0.177   11.006    0.000
##     trstprt           1.773    0.159   11.153    0.000
##
## Intercepts:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .trstprl           5.962    0.093   64.288    0.000
##    .trstlgl           5.943    0.101   58.817    0.000
##    .trstplt           5.127    0.106   48.199    0.000
##    .trstprt           5.057    0.106   47.619    0.000
##     trstlead          0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .trstprl           2.210    0.171   12.945    0.000
##    .trstlgl           3.031    0.229   13.256    0.000
##    .trstplt           0.514    0.172    2.985    0.003
##    .trstprt           1.117    0.163    6.872    0.000
##     trstlead          0.972    0.176    5.513    0.000
##
##
## Group 2 [0]:
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   trstlead =~
##     trstprl           1.000
##     trstlgl           0.898    0.043   20.713    0.000
##     trstplt           1.213    0.039   30.976    0.000
##     trstprt           1.207    0.039   30.894    0.000
##
## Intercepts:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .trstprl           4.016    0.061   65.899    0.000
##    .trstlgl           4.623    0.065   70.617    0.000
##    .trstplt           3.499    0.058   60.067    0.000
##    .trstprt           3.521    0.058   60.349    0.000
##     trstlead          0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
```

```
##    .trstprl           2.280    0.100   22.745    0.000
##    .trstlgl           3.480    0.145   23.940    0.000
##    .trstplt           0.718    0.062   11.651    0.000
##    .trstprt           0.764    0.062   12.304    0.000
##     trstlead          2.448    0.170   14.417    0.000
```

From the output, there were 370 respondents out of 1643 observations being satisfied with the government with a score higher than 5. There was a difference of 4 between the number of input parameters of observed variables and the number of parameters we estimated. With a p-value of chi-square distribution at 0, we rejected the null hypothesis and conclude that the fits for the two subset of satisfied and dissatisfied respondents were not close; that said, we did not have measurement invariance.

The measurement models can be written as the following, denoting the latent variable as **trstlead**. For the group of respondents being satisfied with government: - $trstprl = 5.962 + 1 * trstlead + \epsilon_{trstprl}$, where $\sigma^2_{trstprl} = 2.210$; the estimated error variance of the observed degree of trust in country's parliament is 2.210. - $trstlgl = 5.943 + 0.876 * trstlead + \epsilon_{trstlgl}$, where $\sigma^2_{trstlgl} = 3.031$; the estimated error variance of the observed degree of trust in legal system is 3.031. - $trstplt = 5.127 + 1.944 * trstlead + \epsilon_{trstplt}$, where $\sigma^2_{trstplt} = 0.514$; the estimated error variance of the observed degree of trust in politicians is 0.514. - $trstprt = 5.057 + 1.773 * trstlead + \epsilon_{trstprt}$, where $\sigma^2_{trstprt} = 1.117$; the estimated error variance of the observed degree of trust in political parties is 1.117. - $\sigma^2_{trstlead} = 0.972$; the estimated error variance of the unobserved degree of trust in leadership for satisfied respondents is 0.972.

For the group of respondents being unsatisfied with government: - $trstprl = 4.016 + 1 * trstlead + \epsilon_{trstprl}$, where $\sigma^2_{trstprl} = 2.280$; the estimated error variance of the observed degree of trust in country's parliament is 2.280. - $trstlgl = 4.623 + 0.898 * trstlead + \epsilon_{trstlgl}$, where $\sigma^2_{trstlgl} = 3.480$; the estimated error variance of the observed degree of trust in legal system is 3.480. - $trstplt = 3.499 + 1.213 * trstlead + \epsilon_{trstplt}$, where $\sigma^2_{trstplt} = 0.718$; the estimated error variance of the observed degree of trust in politicians is 0.718. - $trstprt = 3.521 + 1.207 * trstlead + \epsilon_{trstprt}$, where $\sigma^2_{trstprt} = 0.764$; the estimated error variance of the observed degree of trust in political parties is 0.764. - $\sigma^2_{trstlead} = 2.448$; the estimated error variance of the unobserved degree of trust in leadership for unsatisfied respondents is 2.448.

The subset of the sample satisfied with the government exhibited elevated mean trust scores in all four indicators of the political system, demonstrated stronger covariances with the latent variable of degree of trust in leadership (excluding the degree of trust in the legal system), and generally displayed reduced error variance (with the exception of the degree of trust in political parties). Furthermore, with all parameters of estimation having p-values of 0, we can be more confident about our measurement model since the relationships between the latent variable of trust in leadership and the observed indicators are not likely to be zero.

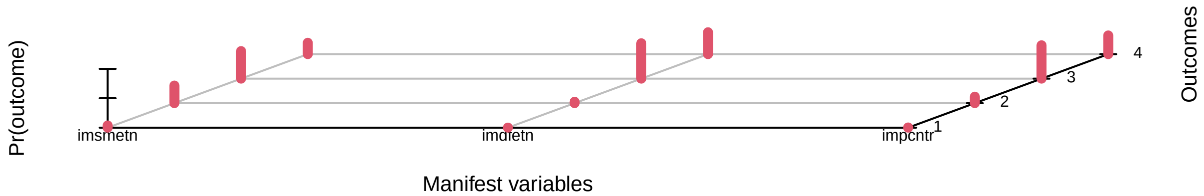
4. Divide the individuals into three groups using a latent class analysis (LCA) based on the three measures of attitudes toward immigrants. What are your qualitative descriptions of each of the three classes?

```r
set.seed(98)
model_lca.3 <- poLCA(cbind(imsmetn, imdfetn, impcntr) ~ 1,
                      ## a. the 1 instructs poLCA to estimate
                      ## the basic latent class model
                      ## b. three measures of attitudes toward
                      ## immigration policies are categorical
                      maxiter = 50000, nclass = 3, nrep = 10,
                      data = data1_ind, graphs = TRUE)
```

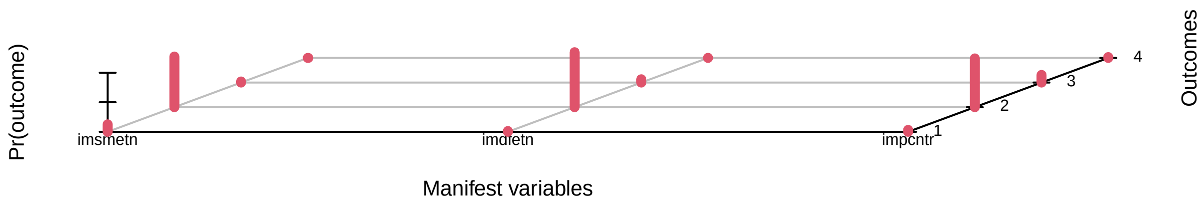
```
## Model 1: llik = -4686.839 ... best llik = -4686.839
## Model 2: llik = -4686.839 ... best llik = -4686.839
## Model 3: llik = -4740.993 ... best llik = -4686.839
```

```
## Model 4: llik = -4820.696 ... best llik = -4686.839
## Model 5: llik = -4861.15 ... best llik = -4686.839
## Model 6: llik = -4709.894 ... best llik = -4686.839
## Model 7: llik = -4709.894 ... best llik = -4686.839
## Model 8: llik = -4740.993 ... best llik = -4686.839
## Model 9: llik = -4709.894 ... best llik = -4686.839
## Model 10: llik = -4686.839 ... best llik = -4686.839
```

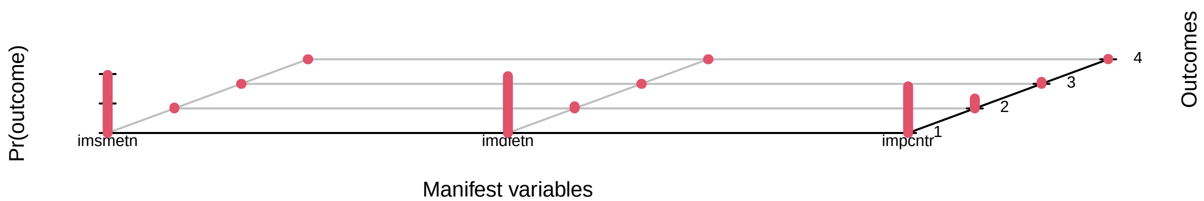
Class 1: population share = 0.452



Class 2: population share = 0.459



Class 3: population share = 0.088




```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $imsmetn
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0376 0.2996 0.4690 0.1937
## class 2:  0.1261 0.8562 0.0177 0.0000
## class 3:  0.9847 0.0153 0.0000 0.0000
##
## $imdfetn
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0013 0.0242 0.6018 0.3727
## class 2:  0.0135 0.9300 0.0552 0.0013
## class 3:  0.9636 0.0364 0.0000 0.0000
##
## $impcntr
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0034 0.1107 0.5676 0.3183
```

```
## class 2:  0.0326 0.8249 0.1288 0.0138
## class 3:  0.7915 0.1675 0.0340 0.0070
##
## Estimated class population shares
##  0.4524 0.4593 0.0883
##
## Predicted class memberships (by modal posterior prob.)
##  0.4474 0.4626 0.0901
##
## =========================================================
## Fit for 3 latent classes:
## =========================================================
## number of observations: 1643
## number of estimated parameters: 29
## residual degrees of freedom: 34
## maximum log-likelihood: -4686.839
##
## AIC(3): 9431.677
## BIC(3): 9588.401
## G^2(3): 715.2258 (Likelihood ratio/deviance statistic)
## X^2(3): 936.5198 (Chi-square goodness of fit)
##
```

```
model_lca.3
```

```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $imsmetn
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0376 0.2996 0.4690 0.1937
## class 2:  0.1261 0.8562 0.0177 0.0000
## class 3:  0.9847 0.0153 0.0000 0.0000
##
## $imdfetn
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0013 0.0242 0.6018 0.3727
## class 2:  0.0135 0.9300 0.0552 0.0013
## class 3:  0.9636 0.0364 0.0000 0.0000
##
## $impcntr
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0034 0.1107 0.5676 0.3183
## class 2:  0.0326 0.8249 0.1288 0.0138
## class 3:  0.7915 0.1675 0.0340 0.0070
##
## Estimated class population shares
##  0.4524 0.4593 0.0883
##
## Predicted class memberships (by modal posterior prob.)
##  0.4474 0.4626 0.0901
##
## =========================================================
## Fit for 3 latent classes:
```

```
## ========================================================
## number of observations: 1643
## number of estimated parameters: 29
## residual degrees of freedom: 34
## maximum log-likelihood: -4686.839
##
## AIC(3): 9431.677
## BIC(3): 9588.401
## G^2(3): 715.2258 (Likelihood ratio/deviance statistic)
## X^2(3): 936.5198 (Chi-square goodness of fit)
##
```

With `nrep=10` it runs every model 10 times and keeps the model with the lowest BIC. Since indicators of immigration policy attitudes were already coded in numbers starting at 1, I didn't have to recode them prior to fitting the `poLCA` model.

In this model, we utilized 1643 observations to estimate 29 parameters, resulting in 34 residual degrees of freedom and a maximum log-likelihood of -4686.839. Class 1 exhibited the highest probabilities of permitting a few immigrants regardless of ethnicity. They expressed a preference for allowing more from the same ethnicity and fewer from poor countries and different ethnicities. Class 2 tended to permit some immigrants for all ethnicities and allowed more from shared ethnicity groups, being less open to immigrants from different ethnicities. Class 3 generally welcomed a substantial number of immigrants and showed openness to both shared and different ethnicities, with a 0 probability of allowing less than some or none. However, for prospective immigrants from poor countries, the probability of allowing many decreased, while the probability of allowing just some increased.

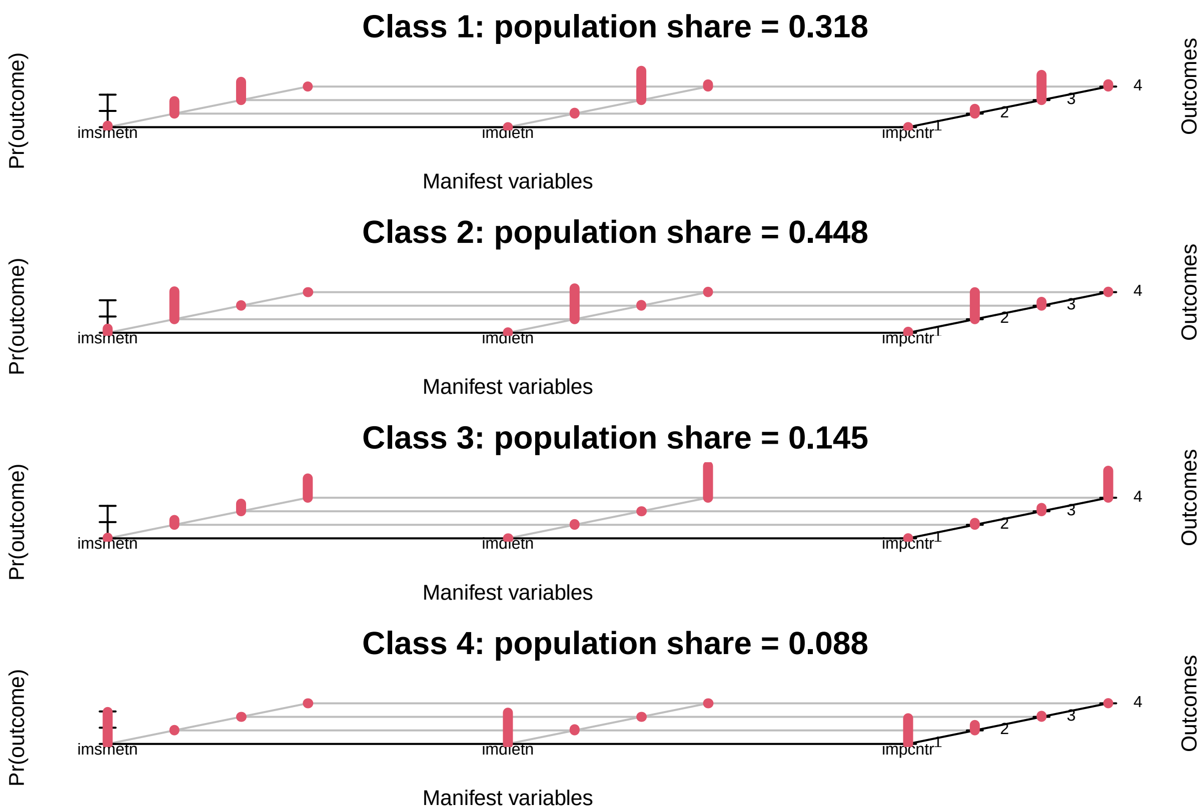
Finally, poLCA outputs a number of goodness of fit statistics in AIC, BIC, G^2, and X^2. With three classes, the AIC was 9431.677 and the BIC was 9588.401. We could use these information criteria scores to compare with LCA with difference numbers of classes.

(Reference: 1. https://www.sscnet.ucla.edu/polisci/faculty/lewis/pdf/poLCA-JSS-final.pdf 2. https://stats.oarc.ucla.edu/sas/dae/latent-class-analysis/#:~:text=Examples%20of%20Latent%20Class%20Analysis&text=For%20e

5. Consider latent class analyses with four or five classes as well. Which of the three LCA models would appear to have the best fit? What do you notice computationally as the hypothesized number of classes increases?

```r
# 4 classes
set.seed(98)
model_lca.4 <- poLCA(cbind(imsmetn, imdfetn, impcntr) ~ 1,
maxiter = 50000, nclass = 4, nrep = 10, data = data1_ind, graphs = TRUE)
```

```
## Model 1: llik = -4338.283 ... best llik = -4338.283
## Model 2: llik = -4338.283 ... best llik = -4338.283
## Model 3: llik = -4338.283 ... best llik = -4338.283
## Model 4: llik = -4338.283 ... best llik = -4338.283
## Model 5: llik = -4338.283 ... best llik = -4338.283
## Model 6: llik = -4338.283 ... best llik = -4338.283
## Model 7: llik = -4338.283 ... best llik = -4338.283
## Model 8: llik = -4338.283 ... best llik = -4338.283
## Model 9: llik = -4338.283 ... best llik = -4338.283
## Model 10: llik = -4338.283 ... best llik = -4338.283
```




```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $imsmetn
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0474 0.3853 0.5631 0.0041
## class 2:  0.1268 0.8574 0.0158 0.0000
## class 3:  0.0208 0.1504 0.2345 0.5943
## class 4:  0.9847 0.0153 0.0000 0.0000
##
## $imdfetn
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0021 0.0284 0.9018 0.0678
## class 2:  0.0137 0.9498 0.0238 0.0127
## class 3:  0.0000 0.0228 0.0000 0.9772
## class 4:  0.9636 0.0364 0.0000 0.0000
##
## $impcntr
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0043 0.1486 0.7765 0.0707
## class 2:  0.0324 0.8322 0.1197 0.0157
## class 3:  0.0042 0.0596 0.1046 0.8316
## class 4:  0.7917 0.1673 0.0340 0.0071
##
## Estimated class population shares
##  0.3182 0.4482 0.1453 0.0883
```

```
##
## Predicted class memberships (by modal posterior prob.)
##  0.3256 0.4425 0.1418 0.0901
##
## =========================================================
## Fit for 4 latent classes:
## =========================================================
## number of observations: 1643
## number of estimated parameters: 39
## residual degrees of freedom: 24
## maximum log-likelihood: -4338.283
##
## AIC(4): 8754.566
## BIC(4): 8965.333
## G^2(4): 18.11507 (Likelihood ratio/deviance statistic)
## X^2(4): 15.73375 (Chi-square goodness of fit)
##
```

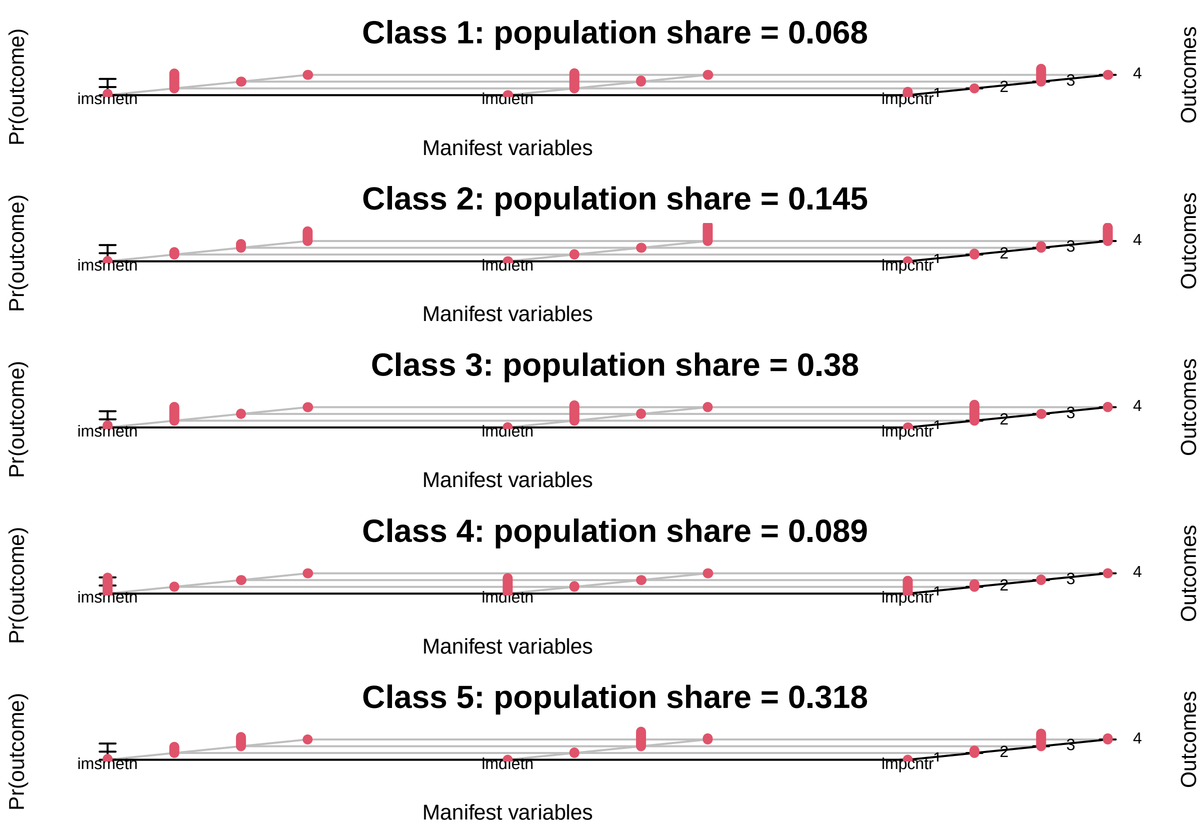
```
model_lca.4
```

```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $imsmetn
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0474 0.3853 0.5631 0.0041
## class 2:  0.1268 0.8574 0.0158 0.0000
## class 3:  0.0208 0.1504 0.2345 0.5943
## class 4:  0.9847 0.0153 0.0000 0.0000
##
## $imdfetn
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0021 0.0284 0.9018 0.0678
## class 2:  0.0137 0.9498 0.0238 0.0127
## class 3:  0.0000 0.0228 0.0000 0.9772
## class 4:  0.9636 0.0364 0.0000 0.0000
##
## $impcntr
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0043 0.1486 0.7765 0.0707
## class 2:  0.0324 0.8322 0.1197 0.0157
## class 3:  0.0042 0.0596 0.1046 0.8316
## class 4:  0.7917 0.1673 0.0340 0.0071
##
## Estimated class population shares
##  0.3182 0.4482 0.1453 0.0883
##
## Predicted class memberships (by modal posterior prob.)
##  0.3256 0.4425 0.1418 0.0901
##
## =========================================================
## Fit for 4 latent classes:
## =========================================================
## number of observations: 1643
```

```
## number of estimated parameters: 39
## residual degrees of freedom: 24
## maximum log-likelihood: -4338.283
##
## AIC(4): 8754.566
## BIC(4): 8965.333
## G^2(4): 18.11507 (Likelihood ratio/deviance statistic)
## X^2(4): 15.73375 (Chi-square goodness of fit)
##
```

```
# 5 classes
set.seed(98)
model_lca.5 <- poLCA(cbind(imsmetn, imdfetn, impcntr) ~ 1,
maxiter = 50000, nclass = 5, nrep = 10, data = data1_ind, graphs = TRUE)
```

```
## Model 1: llik = -4335.552 ... best llik = -4335.552
## Model 2: llik = -4335.209 ... best llik = -4335.209
## Model 3: llik = -4335.209 ... best llik = -4335.209
## Model 4: llik = -4335.649 ... best llik = -4335.209
## Model 5: llik = -4335.552 ... best llik = -4335.209
## Model 6: llik = -4335.649 ... best llik = -4335.209
## Model 7: llik = -4335.542 ... best llik = -4335.209
## Model 8: llik = -4335.552 ... best llik = -4335.209
## Model 9: llik = -4335.552 ... best llik = -4335.209
## Model 10: llik = -4335.542 ... best llik = -4335.209
```



```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $imsmetn
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0831 0.9169 0.0000 0.0000
## class 2:  0.0209 0.1505 0.2324 0.5962
## class 3:  0.1320 0.8516 0.0164 0.0000
## class 4:  0.9828 0.0172 0.0000 0.0000
## class 5:  0.0480 0.3798 0.5680 0.0041
##
## $imdfetn
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.9328 0.0672 0.0000
## class 2:  0.0000 0.0219 0.0000 0.9781
## class 3:  0.0157 0.9473 0.0232 0.0138
## class 4:  0.9542 0.0458 0.0000 0.0000
## class 5:  0.0022 0.0329 0.8947 0.0703
##
## $impcntr
##            Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.2056 0.0025 0.7919 0.0000
## class 2:  0.0042 0.0583 0.1047 0.8328
## class 3:  0.0000 0.9815 0.0000 0.0185
## class 4:  0.7923 0.1652 0.0354 0.0070
## class 5:  0.0030 0.1498 0.7757 0.0715
##
## Estimated class population shares
##  0.0684 0.1448 0.3797 0.0893 0.3177
##
## Predicted class memberships (by modal posterior prob.)
##  0.0676 0.1418 0.3761 0.0901 0.3244
##
## =========================================================
## Fit for 5 latent classes:
## =========================================================
## number of observations: 1643
## number of estimated parameters: 49
## residual degrees of freedom: 14
## maximum log-likelihood: -4335.209
##
## AIC(5): 8768.417
## BIC(5): 9033.227
## G^2(5): 11.96602 (Likelihood ratio/deviance statistic)
## X^2(5): 11.2287 (Chi-square goodness of fit)
##
## ALERT: iterations finished, MAXIMUM LIKELIHOOD NOT FOUND
##
```

```
model_lca.5
```

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

```
## $imsmetn
##           Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0831 0.9169 0.0000 0.0000
## class 2:  0.0209 0.1505 0.2324 0.5962
## class 3:  0.1320 0.8516 0.0164 0.0000
## class 4:  0.9828 0.0172 0.0000 0.0000
## class 5:  0.0480 0.3798 0.5680 0.0041
##
## $imdfetn
##           Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.0000 0.9328 0.0672 0.0000
## class 2:  0.0000 0.0219 0.0000 0.9781
## class 3:  0.0157 0.9473 0.0232 0.0138
## class 4:  0.9542 0.0458 0.0000 0.0000
## class 5:  0.0022 0.0329 0.8947 0.0703
##
## $impcntr
##           Pr(1)  Pr(2)  Pr(3)  Pr(4)
## class 1:  0.2056 0.0025 0.7919 0.0000
## class 2:  0.0042 0.0583 0.1047 0.8328
## class 3:  0.0000 0.9815 0.0000 0.0185
## class 4:  0.7923 0.1652 0.0354 0.0070
## class 5:  0.0030 0.1498 0.7757 0.0715
##
## Estimated class population shares
##  0.0684 0.1448 0.3797 0.0893 0.3177
##
## Predicted class memberships (by modal posterior prob.)
##  0.0676 0.1418 0.3761 0.0901 0.3244
##
## =========================================================
## Fit for 5 latent classes:
## =========================================================
## number of observations: 1643
## number of estimated parameters: 49
## residual degrees of freedom: 14
## maximum log-likelihood: -4335.209
##
## AIC(5): 8768.417
## BIC(5): 9033.227
## G^2(5): 11.96602 (Likelihood ratio/deviance statistic)
## X^2(5): 11.2287 (Chi-square goodness of fit)
##
## ALERT: iterations finished, MAXIMUM LIKELIHOOD NOT FOUND
##
```

For the four-class LCA, we utilized 1643 observations to estimate 39 parameters, resulting in 24 residual degrees of freedom and a maximum log-likelihood greater than the three-class model at -4338.283. Class 1 demonstrated a more conservative stance, favoring only a few immigrants, but particularly showing a greater acceptance for those from the same ethnicity. Class 2 tended to permit some immigrants for all ethnicities and allowed more from different ethnicity groups, being the least open to immigrants from poor countries among themselves. Class 3 exhibited the highest probabilities of permitting none immigrants regardless of ethnicity. However, among themselves, they were the most open to same ethnicity potential immigrants, decreasing openness to immigrants from poor countries and shared different ethnicity in that order. Class 4 generally welcomed a substantial number of immigrants and showed openness to both shared and different ethnicities, with a 0 probability of allowing less than some or none. However, they hesitated more for prospective immigrants from poor countries with the increased probability in allowing just some or lesser.

With four classes, the AIC decreased to 8754.566 and the BIC decreased to 8965.333. A four-class latent class analysis approach is proved to be better than a three-class LCA.

For the five-class model, it took substantially longer to generate the results. In addition, while the number of estimated parameters increased, with the same number of input parameters, the residual degrees of freedom gradually decreased. We utilized 1643 observations to estimate 49 parameters, resulting in 14 residual degrees of freedom and a maximum log-likelihood greater than the three-class model but less than the four-class model at -4334.934.

Class 2 tended to refuse immigrants from different ethnicities and those from poor countries. However, among themselves, they moderately accepted immigrants from the same ethnicity, with a noticeable increase in both “allowing for some” and “allowing for a few,” shifted from “allowing for none.” Class 4 was the most open among all classes, with the highest probabilities of allowing many immigrants from both the same and different ethnicities, but not from poor countries among themselves. Class 5 welcomed only a few immigrants but expressed a desire to allow more from the same ethnicity, with a significant shift from “allowing a few” and “allowing none” to “allowing some.” Class 3 predominantly had high probabilities of allowing for some immigrants among themselves, but the acceptance descended across different poor countries, different ethnicities, and the same ethnicities. Class 1 exhibited a more conservative attitude toward immigrants, with no clear pattern in shifts of attitudes toward different ethnicities.

The AIC rose to 8768.417 and the BIC increased to 9033.227 compared to the four-class LCA. The four-class latent class analysis approach is demonstrated to be superior to both the three-class and the five-class LCA with the smallest AIC and BIC scores.

6. Fit a structural equation model, using the latent trust in leadership measure indicated by the four measures to predict the latent immigration attitude measure indicated by the three measures. Interpret the estimated coefficient for the structural component of this model.

```r
model_6 <- '
  # seasurement model for trust in leadership
    trstlead =~ trstprl + trstlgl + trstplt + trstprt

  # seasurement model for immigration attitude
    imlat =~ imsmetn + imdfetn + impcntr

  # structural model
    trstlead ~ b*imlat
'
fit_6 <- sem(model_6, data = data1)
summary(fit_6)
```

```
## lavaan 0.6.16 ended normally after 33 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        15
##
##   Number of observations                          1643
##
## Model Test User Model:
##
```

```
##   Test statistic                               211.890
##   Degrees of freedom                                13
##   P-value (Chi-square)                           0.000
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   trstlead =~
##     trstprl           1.000
##     trstlgl           0.863    0.036   24.160    0.000
##     trstplt           1.226    0.034   36.493    0.000
##     trstprt           1.200    0.033   36.177    0.000
##   imlat =~
##     imsmetn           1.000
##     imdfetn           1.288    0.034   37.698    0.000
##     impcntr           1.091    0.031   35.648    0.000
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   trstlead ~
##     imlat      (b)   -0.663    0.070   -9.505    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .trstprl           2.355    0.091   25.831    0.000
##    .trstlgl           3.383    0.124   27.225    0.000
##    .trstplt           0.715    0.056   12.715    0.000
##    .trstprt           0.842    0.057   14.860    0.000
##    .imsmetn           0.269    0.012   23.180    0.000
##    .imdfetn           0.091    0.012    7.824    0.000
##    .impcntr           0.236    0.012   20.445    0.000
##    .trstlead          2.505    0.150   16.663    0.000
##     imlat             0.409    0.023   17.975    0.000
```

From the output, there were 1643 observations. There was a difference of 13 between the number of input parameters of observed variables and the number of parameters we estimated. With a p-value of chi-square distribution at 0, we rejected the null hypothesis and concluded that the relationship between the latent variables and their indicators were not likely to be zero.

The structural regression model showed $trstlead = -0.663 * imlat$, signifying a moderate negative correlation between the latent variables of the degree of trust in leadership and attitude towards immigrants. For every one-unit increase in the unobserved attitude toward immigrants, reflecting a decrease in willingness to open up for immigrants, the unobserved respondent's degree of trust in political leadership decreased by 0.663 unit. With a p-value at 0, we failed to reject the null hypothesis. We concluded that the relationship between the latent variables of trust in leadership and attitude towards immigrants was not likely to be zero. Therefore, changes in the attitude towards immigrants have a discernible impact on the degree of trust in political leadership, and this relationship is not solely the result of random fluctuations in the data.

- $\sigma^2_{trstprl} = 2.355$; the estimated error variance of the observed degree of trust in country's parliament is 2.355.

- $\sigma^2_{trstlgl} = 3.383$; the estimated error variance of the observed degree of trust in legal system is 3.383.
- $\sigma^2_{trstplt} = 0.715$; the estimated error variance of the observed degree of trust in politicians is 0.715.
- $\sigma^2_{trstprt} = 0.842$; the estimated error variance of the observed degree of trust in political parties is 0.842.
- $\sigma^2_{imsmetn} = 0.269$; the estimated error variance of the observed attitude towards same ethnicity immigrants is 0.269.
- $\sigma^2_{imdfetn} = 0.091$; the estimated error variance of the observed attitude towards different ethnicity immigrants is 0.091.
- $\sigma^2_{impcntr} = 0.236$; the estimated error variance of the observed attitude towards immigrants from poor countries is 0.236.
- $\sigma^2_{trstlead} = 2.505$; the estimated error variance of the latent variable of the degree of trust in political leadership is 2.505.
- $\sigma^2_{imlat} = 0.409$; the estimated error variance of the latent variable of attitudes toward immigrants is 0.409.

7. How might you improve the fitted model in part 6?

The output didn't print intercepts, which are the mean values, for each measurement model. Since I had ordered variables, both with more than two categories, I had to free the thresholds, k-1 where k was the number of categories, so that the models could be built completely.

8. Write a one-page summary of the results and conclusions of the above analyses.

To address the three study objectives listed in the following, I applied several structural equation models.

a. Is there measurement invariance in the way that a latent construct of trust in leadership is measured by the four TRST variables across groups of respondents defined by different satisfaction with the national government?
b. Allowing for possible measurement error, cluster individuals based on the three IM measures of attitudes toward immigrants.
c. Use the latent trust in leadership measure indicated by the four TRST measures to predict the latent immigration attitude measure indicated by the three IM measures.

Using the four observed indicators of the degree of trust in different perspectives of the political system, the model had a p-value of 0. We concluded that there is no measurement invariance across the two respondent groups having opposite satisfaction with the national government. In addition, the subset of the sample satisfied with the government exhibited higher mean scores of trust in all four indicators of the political system, demonstrated stronger covariances with the latent variable of degree of trust in leadership (excluding the degree of trust in the legal system), and generally displayed reduced error variance (with the exception of the degree of trust in political parties). We also concluded that the relationships between the latent variable of trust in leadership and the observed indicators were not likely to be zero, with p-values of 0.

As the three latent class analysis models indicated, the four-class model performed the best compared to the three- and the five-class model in both shorter processing time and smaller AIC and BIC scores. It revealed that class 1 demonstrated a more conservative stance, favoring only a few immigrants, but particularly showing a greater acceptance for those from the same ethnicity. Class 2 tended to permit some immigrants for all ethnicities and allowed more from different ethnicity groups, being the least open to immigrants from poor countries among themselves. Class 3 exhibited the highest probabilities of permitting none immigrants regardless of ethnicity. However, among themselves, they were the most open to same ethnicity potential immigrants, decreasing openness to immigrants from poor countries and shared different ethnicity in that order. Class 4 generally welcomed a substantial number of immigrants and showed openness to both shared and different ethnicities, with a 0 probability of allowing less than some or none. However, they hesitated more for prospective immigrants from poor countries with the increased probability in allowing just some or lesser.

With four classes, the AIC decreased to 8754.566 and the BIC decreased to 8965.333. A four-class latent class analysis approach is proved to be better than a three-class LCA.

Across the three latent class analysis models, when the number of estimated parameters increased, with the same number of input parameters, the residual degrees of freedom gradually decreased.

To address the last research question, we applied a structural equation model with both measurement and structural models in it. With a p-value of chi-square distribution at 0, we rejected the null hypothesis and concluded that the relationship between the latent variables and their indicators were not likely to be zero.

The structural regression model showed $trstlead = -0.663 * imlat$, signifying a moderate negative correlation between the latent variables of the degree of trust in leadership and attitude towards immigrants. For every one-unit increase in the unobserved attitude toward immigrants, reflecting a decrease in willingness to open up for immigrants, the unobserved respondent's degree of trust in political leadership decreased by 0.663 unit. With a p-value at 0, we failed to reject the null hypothesis. We thus concluded that the relationship between the latent variables of trust in leadership and attitude towards immigrants was not likely to be zero. Therefore, changes in the attitude towards immigrants have a discernible impact on the degree of trust in political leadership, and this relationship is not solely the result of random fluctuations in the data. P-values of zero in each estimated error variance also implied that the relationships between each indicator and their corresponding latent variables were not likely to be zero.