Assignment 1: Latent Variable Model

Arthur Dent, Ford Prefect, Tricia McMillan

Assignment group number: 42

Utrecht University

Faculty of Social and Behavioural Sciences

Course: Theory Construction and Statistical Modeling

Lecturer: Dr. Kyle M. Lang

Date of submission: October 5, 2022

Data

This assignment used a dataset derived from the German General Social Survey (ALLBUS) which was designed to detect changes in behavior and attitudes in German society (GESIS - Leibniz-Institut für Sozialwissenschaften, 2021). The dataset was reduced to respondents surveyed in the year 2016. The final sample comprises data from n=1377 respondents. The participants were on average 52 years old (SD=16.7) and included 52% men and 48% women.

The ALLBUS survey provides information about several topics. This analysis focused on respondents' attitudes towards naturalization. The item battery used in the latent factor model comprises items mn01-mn09, providing information regarding respondents' attitudes towards naturalization, i.e., how important different requirements for obtaining German citizenship should be. The items include statements such as "Whether the person is of German origin". This was measured in a Likert-scale, ranging from "1- Not at all important" to "7- Very important." (see Table A1).

Research question and theoretical latent variable model

Our research question asks: Which and how many factors underlie individuals' attitudes towards naturalization?

We conducted an exploratory factor analysis (EFA) with the purpose of exploring the underlying factor structure behind individuals' attitudes towards naturalization. All cross-loadings of the items were estimated, i.e., each item loaded on all factors. In the process of extracting factors, we compared the indications of the Guttman-Kaiser criterion, a scree plot and parallel analysis. We ran the EFA using the Bartlett method for estimating factor scores with Weighted Least Squares. This estimation gives more weight to variables with low specific variances. Further, we rotated the factors using the oblique rotation method, allowing for correlations between factors by minimizing the distance between items and factors. This was a more realistic approach compared to the orthogonal rotation method, especially since we did not have a theory in mind about the relationship between the factors.

Analytical model and assumptions

We chose EFA as the latent variable model because we did not have any prior knowledge or theory as to what underlying factors contributed to attitudes regarding naturalization in German society. Confirmatory factor analysis (CFA) would not be a suitable model, as it requires some knowledge of the underlying latent variable structure (Byrne, 2005). Principal components analysis (PCA) was not chosen either, as it reduces correlated observed variables and does not model any underlying factor structure of the observed variables. While PCA directly explains observed variances, EFA allows the observed items to contain its unique factor (Lang, 2022). This makes it possible to quantify the item variance that is left unexplained by the latent variables (Lang, 2022).

Since the items were measured on a 7-point Likert scale, the multivariate normality assumption for EFA by definition was violated. We assessed the extent to which the assumption was violated by individually plotting the univariate distributions of the items in a histogram (Figure B1). The results from the visual test revealed that the normality assumption was severely violated. For the sake of analysis we ignored this violation. The sample size requirements for the model estimation were met with n=1377.

We ran an initial empty EFA model (no rotation, no. of factors = 1) on all nine items and decided on the number of factors to be extracted by comparing different criteria: the Guttman-Kaiser criterion, a scree plot and parallel analysis. The findings of this analysis are reported in the results section. After deciding on the number of factors to be extracted, we examined the underlying factor structure by evaluating the factor loadings of each item. Again, the findings are reported in the next section.

Results

Number of factors. To determine the number of factors to be extracted, we compared the results for the Guttman-Kaiser criterion, a scree plot and parallel analysis. The Guttman-Kaiser criterion suggested extracting factors with an Eigenvalue larger than 1. We found that only one factor met this criterion with an Eigenvalue of 1.894 (Table C1). The visual test from the scree plot suggested extracting two factors (Figure B2), however this largely depends on where the elbow point was set. Since both criteria were not reliable on their own, we compared their suggestions with parallel analysis (Figure B3). Parallel analysis suggested extracting two factors. Due to the contrary findings of the criteria, we compared the factor loadings of two EFA models: model 1 with one factor and model 2 with two factors.

Factor loadings. Both model 1 and 2 were estimated using Weighted Least Squares and rotated with the promax oblique rotation method. The factor loadings for both models are shown in Table C2 and Table C3. Comparing the models in light of the theoretical meaning of the items and the suggestions by parallel analysis, we decided to keep the two-factor solution. In the final two-factor model the factors had the Eigenvalues 1.65 and 1.54, and combined they explained 35% of the variance in the variables (Table C4). The correlation between the factors was .40 (Table C5). The items mn01, mn02 and mn06 loaded highest on factor 1, whereas items mn03, mn05, mn07, mn08 and mn09 loaded highest on factor 2. Item mn04 didn't load high on either of the factors.

Model fit. According to the Chi-squared criterion, the model didn't fit the data well, $X^2(9, N=1377) = 50.38$, $p = <.001$ (Table C6). However, the RMSEA fit measure suggested a close model fit, RMSEA = .04, 90% CI [.02, .05] (Table C7).

Discussion

The aim of this analysis was to understand the underlying constructs behind individuals' attitudes towards naturalization in Germany. We therefore conducted an EFA revealing the presence and nature of latent factors behind respondents' attitudes. The results of the EFA hint at the presence of two latent factors that could explain respondents' attitudes towards naturalization. We labeled factor one "ascriptive citizenship" because the high-loading items captured respondents' attitudes towards naturalization criteria that are ascribed by birth and cannot be controlled by individuals. Naturalization based on ascriptive citizenship is therefore harder to achieve and more exclusive. Factor two was labeled "achieved citizenship" as its high-loading items capture respondents' attitudes towards naturalization criteria that can be achieved through some form of investment (e.g., time or education) and are thus less exclusive. In summary, we found that respondents' attitudes towards naturalization were influenced by two underlying latent factors: ascriptive and achieved citizenship. Further research should be conducted to investigate whether the results of this EFA can be reproduced with other data.

There are some limitations to our analyses that potentially restrict the explanatory power of our model. With regards to determining the nature of the underlying factor structure, there are no clear rules about the threshold for a "high" factor loading in the literature. The selection of items with high loadings therefore might be considered arbitrary. Moreover, the items analyzed heavily violated the multivariate normality assumption. Despite treating the items as continuous, they were in fact measured in a categorical way and therefore by definition violate the multivariate normality assumption. As a consequence, the results of the EFA are likely biased. Lastly, we conducted a list wise deletion to deal with missing data. This too could contribute to biased results, providing less representative data for the German population.

References

Byrne, B. (2005). Factor analytic models: Viewing the structure of an assessment instrument from three perspectives, Journal of Personality Assessment, 85(1), 17–32.

GESIS - Leibniz-Institut für Sozialwissenschaften (2021). German General Social Survey (ALLBUScompact) - Cumulation 1980-2018. *GESIS Data Archive, Cologne. ZA5277 Data file Version 1.1.0, https://doi.org/10.4232/1.13775*.

Lang, K. M. (2022, 26th of September). *EFA*. Theory Construction and Statistical Modeling. Retrieved 29th of September, 2022 from: https://www.kylemlang.com/tcsm/reading-1.html

Appendix A - Materials

Table A1. Item description.

Variable name	Item
mn01	Whether the person was born in Germany
mn02	Whether the person is of German origin
mn03	Whether the person is fluent in German.
mn04	Whether the person has lived in Germany a long time.
mn05	Whether the person is prepared to adapt to the German way of life.
mn06	Whether the person belongs to a Christian denomination
mn07	Whether the person has committed any crimes.
mn08	Whether the person can support himself or herself.
mn09	Whether the person accepts and affirms the liberal democratic constitutional order.

Note. Respondents were asked the following question: "I will tell you a few things which may play a role in the decision whether or not to grant German citizenship. Using the scale, please tell me how important these things should be IN YOUR OPINION." Respondents could answer the question on a scale from "1 - Not at all important" to "7 - Very important".

Appendix B - Figures

Figure B1. Histogram with univariate distribution of items.

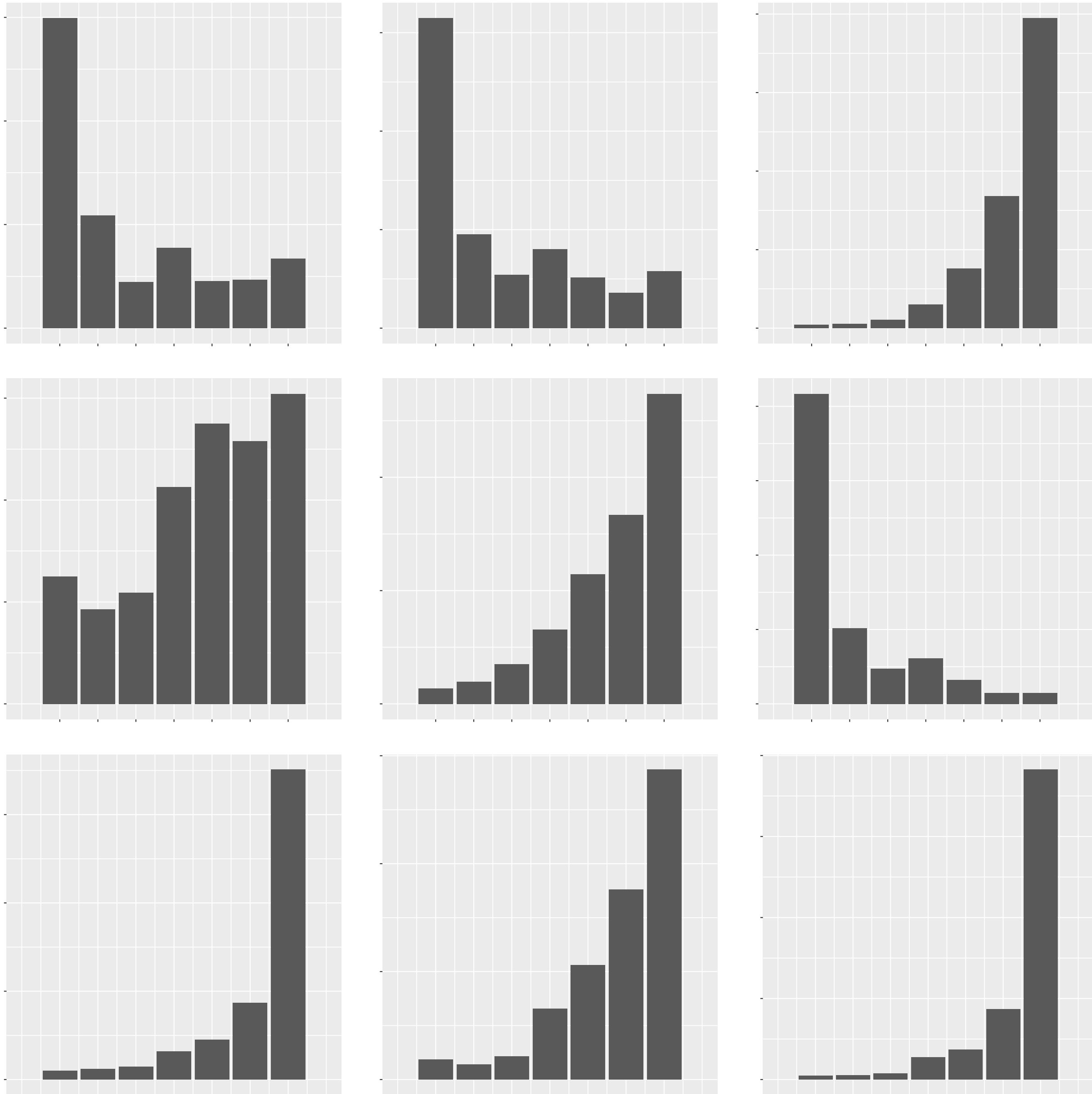


Figure B2. Scree plot.

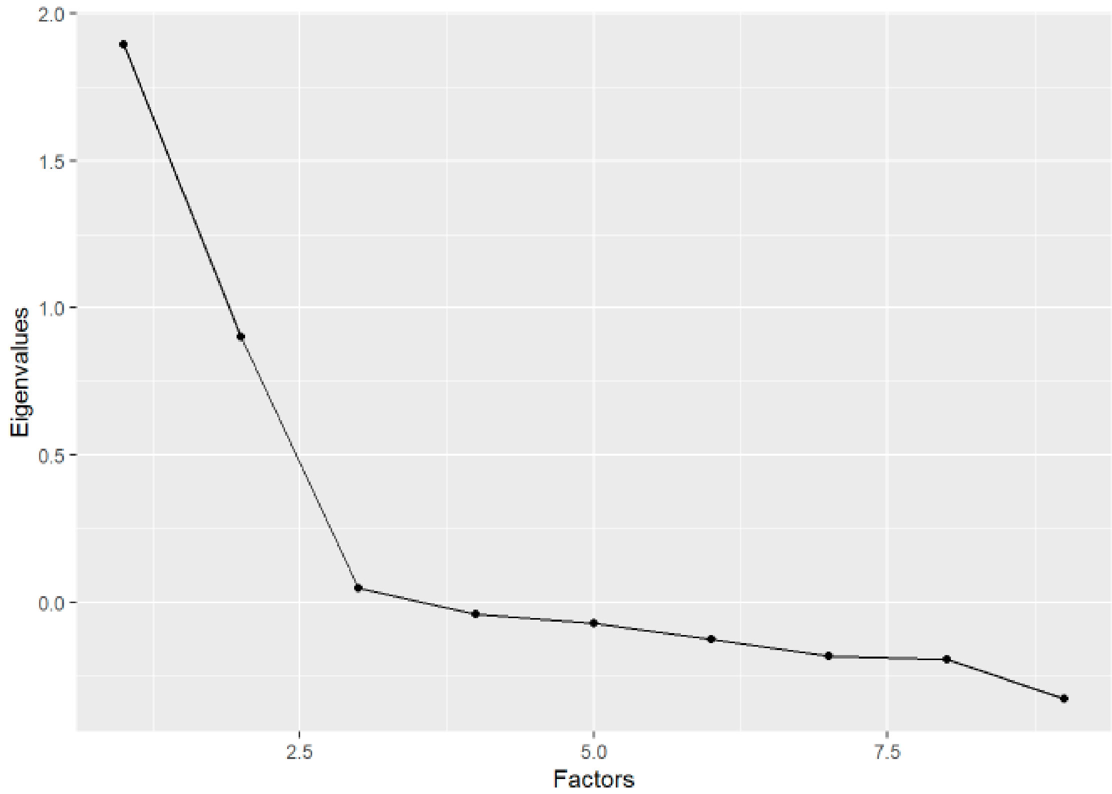
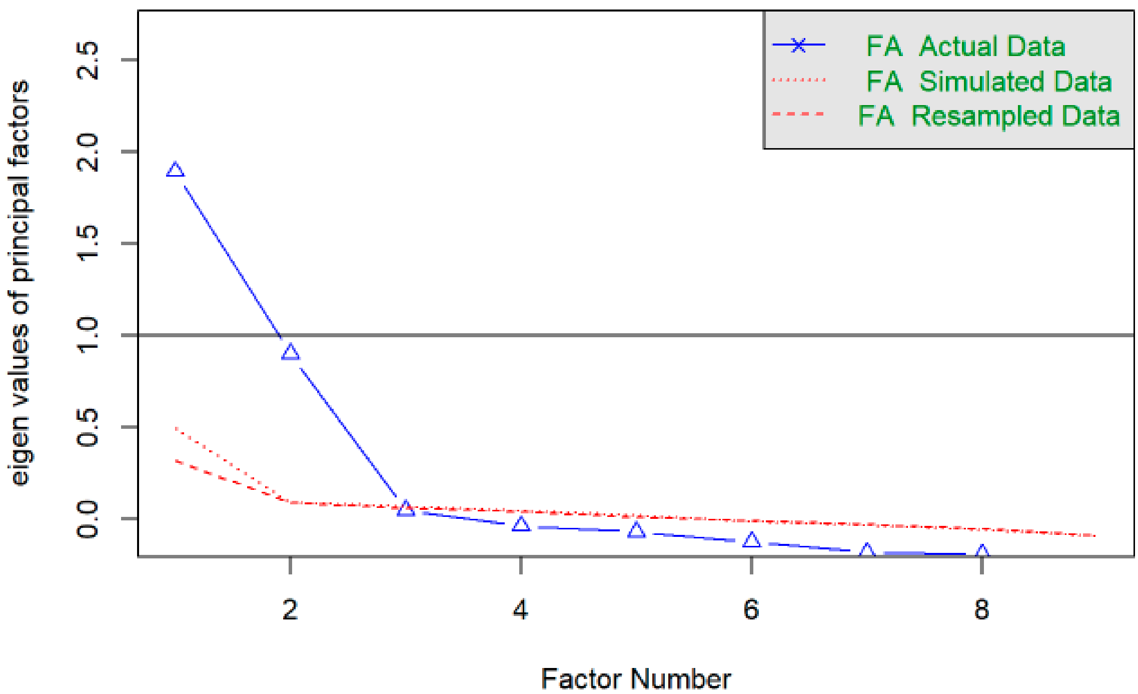


Figure B3. Parallel analysis scree plots.



Appendix C - Tables

Table C1. Guttman-Kaiser criterion.

	Eigenvalue
Factor 1	1.894
Factor 2	0.900
Factor 3	0.046
Factor 4	-0.041
Factor 5	-0.071
Factor 6	-0.126
Factor 7	-0.184
Factor 8	-0.195
Factor 9	-0.329

Table C2. Model 1 factor loadings.

	Factor 1	Uniqueness
mn08	0.535	0.714
mn05	0.531	0.718
mn02	0.515	0.735
mn01	0.510	0.740
mn04	0.488	0.761
mn03	0.448	0.800
mn06		0.850
mn07		0.862
mn09		0.927

Note. Applied rotation method is promax.

Table C3. Model 2 factor loadings.

	Factor 1	Factor 2	Uniqueness
mn02	0.804		0.392
mn01	0.771		0.438
mn06	0.508		0.750
mn08		0.643	0.570
mn03		0.574	0.675
mn05		0.562	0.643
mn07		0.482	0.774
mn09		0.430	0.833
mn04	0.359	0.213	0.774

Note. Applied rotation method is promax.

Table C4. Factor characteristics.

	Rotated solution		
	SumSq. Loadings	Proportion var.	Cumulative
Factor 1	1.648	0.183	0.183
Factor 2	1.536	0.171	0.354

Table C5. Factor correlations.

	Factor 1	Factor 2
Factor 1	1.000	0.400
Factor 2	0.400	1.000

Table C6. Chi-squared test.

	Value	df	p
Model	50.380	19	< .001

Table C7. Additional fit indices.

RMSEA	RMSEA 90% confidence	TLI	BIC
0.035	0.023 - 0.046	0.971	-86.946

Appendix D – R script

```
# Latent variable model

## Guttman-Kaiser criterion: Extract eigenvalues from EFA model
efa_nat0 <- fa(allbus_final[18:26], nfactors = 1, rotate = "none")
round(efa_nat0$values, digits = 3) # extract eigenvalues

## Scree plot
qplot(y = efa_nat0$values) +
  geom_path() +
  ylab("Eigenvalues") +
  xlab("Factors")

## Parallel analysis
# Set the random number seed:
set.seed(235717)
# Run the parallel analysis:
pa_nat <- fa.parallel(allbus_final[18:26], fa = "fa")

## One-factor and two-factor model
## Define an empty list to hold all of our fitted EFA objects:
efa_nat <- list()

## Loop through the interesting numbers of factors and estimate an
EFA for each:
for(i in 1:2)
  efa_nat[[i]] <- fa(allbus_final[18:26],
                          nfactors = i,          # Number of
factors = Loop index
                          rotate   = "promax",   # Oblique rotation
                          scores   = "Bartlett") # Estimate factor
scores with WLS

# Compare the factor loading matrices from both models
for(x in efa_nat) print(x$loadings)

## Final two-factor model
efa_nat_final <- fa(allbus_final[18:26],
                          nfactors = 2,          # Number of
factors = Loop index
                          rotate   = "promax",   # Oblique rotation
                          scores   = "Bartlett") # Estimate factor
scores with WLS

summary(efa_nat_final,
        standardized = TRUE,
        fit.measures = TRUE,
        rsquare = TRUE)
```