LUIS FELIPE VILLOTA

LUND UNIVERSITY
QDA USING R - SIMM61

LAB ASSIGNMENT 3 - DIMENSION REDUCTION: ANIMAL RIGHTS SCALE

Introduction

In the present report, we apply exploratory factor analysis (EFA) to the elements in "The Animal Rights Scale (ARS): Measuring Attitudes About Animal Rights and Animal Research" database to detect sets of latent factors (not observable) underneath the answers of the participants.

Model description

Data cleaning. We checked for missing values before recoding, we found a total of 11 cases of missing values in *"ar1", "ar2", "ar10", "ar14", "ar18", "ar19", "ar22", "sex", "party", "liberal"* and were dropped from our original database. We are left with 149 observations (5 observations were finally removed). Then, we changed numeric variables of interest into factors: *"sex"* and *"party"*. Levels of *"sex"* were coded "1" = "*female"* and "2" = "*male"*; and in *"party"* with "1" = "democrat", "2" = "republican", "3" = "other", "4" = "none". Items from *"ar1"* to *"ar28"* are on a 1-5 Likert scale, and were treated as continuous variables here. Same is the case for *"liberal"*.

Checking assumptions. Is there enough correlation between the observations in order to use EFA? Yes. The results of the Bartlett sphericity test show a Chi^2 is 1128.934, $p < 0.001$, df 378 for the correlation matrix of all 28 items. We can see that it is significantly different from the identity matrix (null-correlation matrix). Nevertheless, in our database the ratio of number of observations and number of observed variables: 149/28 = 5.32. So it is slightly higher than the desirable threshold of 5, so the test is not reliable. We also ran a Kaiser-Meyer-Olkin (KMO) test as definitive indicator, to see the contrast between the partial correlation matrix and the regular correlation matrix. It showed large KMO indices for the variables. All observed variables (including the overall MSA) have a KMO above the 0.6 threshold and are close to 1, an indication of reasonable factorability of our data. Regarding multivariate normality, we used the Henze-Zirkler's test, the MVN test based on kurtosis and finally the MVN test based on skewness. All tests show $p < 0.05$, so the assumption of MVN is violated. We proceeded, then, with the Principal Axis Factoring (PAF) extraction method with 6 factors as a starting point.

Factor extraction. The **first EFA model** shows that item *ar5* "It is wrong to wear leather jackets and pants" has the highest proportion of its variance explained (95%) by the

extracted factors in the 6-factor structure. As the least represented in this scheme, we find item ar3 "It is morally wrong to drink milk and eat eggs" with ~23% of its variance explained by the new factors. The average communality was 0.46 (46%). We decide to exclude items with low communalities. According to the literature cited in the course notes, MacCallum et al. suggest that "*the average communality should meet the threshold of 0.6 when the sample size is small (<250 data points)"*. In our case, with the 6-factor structure we have an average of 0.46, so we consider to exclude the (8) worst represented items by the extracted factors: "ar*1", "ar3", "ar8", "ar16", "ar18", "ar22", "ar25", "ar28".* We repeat the exact same process for the **second EFA model** with the remaining 20 items. After excluding the 8 items with the lowest communalities we still have an average of 0.55. We point this as a limitation in our study (the sample size is small). The ideal number of factors (with this last model) according to: *Scree test* is 2 (the last "substantial" break of the slope in the 3th factor; though is not very conclusive); the *Kaiser-Guttman criterion*: keep elements with Eigenvalue > 1 so it suggests to keep 4; *Parallel analysis Scree plot* suggests to keep 2 (see *Figure 1*); the *(VSS) criterion*: 2, the *Velicer MAP* achieves a minimum of 0.02 with 2 factors.

Final EFA Model Table. We have "ar5" = ~70% as the highest communality in the 2-factor structure. The lowest is item ar24 = ~21%, and the average communality is ~41%. In terms of loadings, in Factor 1 *"ar5"* had the highest and *"ar24"* the lowest. In Factor 2, *"ar6"* has the highest and *"ar14"* the lowest. We see that items clustered in Factor 1 associate with "Animal Consumption" and items in Factor 2 with "Animal Research", and thus we name the latent variables (factors) respectively. In *Tables i-ii* we find the report of post-extraction results. **Rotation**. An oblique rotation method (Promax) was chosen because it was possible to assume that the factors were correlated. No factor showed to be significant for predicting *"liberal"*, according to our multiple regression model with our PA1 and PA2 factor scores.

Discussion

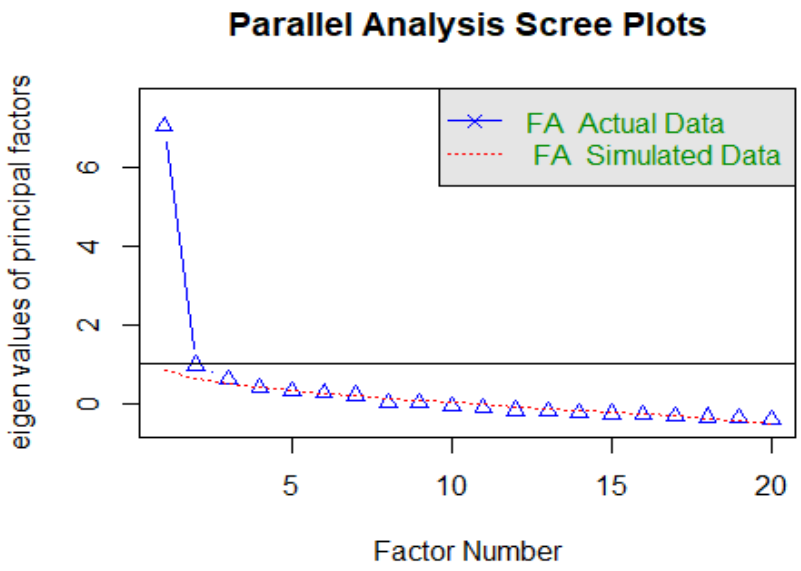
Some items show great skewness. We highlight also the remarkable amount of females in the database ~81% (120) vs 19% (29) of males, which maybe points to a sample bias. Elements *"ar11"* and *"ar14"* had the lowest post-extraction loadings in Animal Consumption Factor and Animal Research Factor, respectively. This could have been

better acknowledged and considered dropping those items. Nevertheless, before reaching the ideal number of factors (final 2-factor structure model), we removed the previously mentioned 8 items with the lowest communalities (leaving us with 20 items in total) and still had an average communality of 0.55. In this way, we had reached the limit for item exclusion for the exercise. We continued the study, pointing finally that the sample size could be bigger and maybe revaluated in terms of representativeness.

R code: https://github.com/FelipeVillota/SIMM61_QDA-with-R/blob/main/animal.R

Appendix

Figure 1. Parallel analysis scree plot.




```
## Parallel analysis suggests that the number of factors = 3 and the numbe
r of components = NA
```

Table i. Post- extraction report[1]. Final EFA Model with Promax as the rotation method

Factor analysis results					
	Factor_1	Factor_2	Communality	Uniqueness	Complexity
ar5	0.898	-0.095	0.70	0.30	1.02
ar10	0.791	-0.134	0.49	0.51	1.06
ar13	0.734	0.130	0.69	0.31	1.06
ar7	0.648	0.008	0.43	0.57	1.00
ar4	0.636	-0.050	0.36	0.64	1.01
ar26	0.562	0.087	0.39	0.61	1.05
ar24	-0.497	0.065	0.21	0.79	1.03
ar23	0.475	0.160	0.36	0.64	1.22
ar11	0.286	0.279	0.27	0.73	2.00
ar6	-0.122	0.875	0.63	0.37	1.04
ar27	-0.148	0.723	0.39	0.61	1.08
ar2	0.084	0.669	0.53	0.47	1.03
ar17	-0.133	0.648	0.32	0.68	1.08
ar9	0.002	0.577	0.33	0.67	1.00
ar20	0.108	0.522	0.36	0.64	1.09
ar15	0.263	0.496	0.50	0.50	1.52
ar12	0.282	0.471	0.49	0.51	1.64
ar19	-0.119	-0.450	0.29	0.71	1.14
ar21	-0.035	-0.444	0.22	0.78	1.01
ar14	0.226	0.265	0.21	0.79	1.95

Table ii. Post- extraction report. Final EFA Model with Promax as the rotation method (continuation)

[1] Tables i-ii were possible thanks to the customized code from Francisco Wilhem (an adaptation from Anthony Smith).

Eigenvalues, Variance Explained, and Factor Correlations for Rotated Factor Solution

Property	Factor_1	Factor_2
SS loadings	4.135	4.042
Proportion Var	0.207	0.202
Cumulative Var	0.207	0.409
Proportion Explained	0.506	0.494
Cumulative Proportion	0.506	1.000
Factor_1	1.000	0.706
Factor_2	0.706	1.000