Assignment 1 Multivariate Statistics 2023-2024

Description of the data

The data set for this task (**ess.Rdata**) is extracted from Round 6 of the European Social Survey (ESS Round 6, 2012). Besides the variable "country", the data set includes the responses of 4046 persons of two countries (France and United Kingdom) on 13 items. Table 1 lists the items that are included in the data set and the latent variables they aim to measure. The items included in the data set measure the concepts "social trust", "trust in public institutions" and "well-being". As "social trust" and "trust in public institutions" can be considered as components of social capital (Gómez-Balcácer et al., 2023), we will use the data to investigate how social capital affects well-being.

Questions

a. Conduct a confirmatory factor analysis on the covariance matrix of the items (i.e., on centered items) in Table 1 assuming a model with 3 correlated latent variables (one for each concept) and assuming that each item only has a loading on the concept it aims to measure.
Discuss convergent validity, divergent validity, composite reliability of the factor scores and fit measures of the fitted CFA model.

b. Use modification indices to see how you can obtain a model that meets the criteria of good fit in (in terms of TLI, CFI, RMSEA, SRMR) by including a few well-chosen correlated error terms for pairs of items. Try to justify the correlated error terms from a substantive point of view.

c. Fit a multi-group structural equation model (with country as the group variable) on the matrix of centered variables to investigate how the latent variables "social trust" and "trust in public institutions" affect the latent variable "well-being".

   Estimate four versions of the multi-group structural equation model:
   1) a configural measurement invariance model with country-specific regression coefficients in the regression equation of the structural model
   2) a configural measurement invariance model with regression coefficients that are constrained to be equal across countries
   3) a metric measurement invariance model with country-specific regression coefficients in the regression equation of the structural model
   4) a metric measurement invariance model with regression coefficients that are constrained to be equal across countries

   Compare the fit measures of the four estimated models and/or use model comparison tests to select the best model. Next discuss the results of this final model (e.g., model fit, estimated intercepts, (standardized) regression coefficients, etc.).

d. Conduct a canonical correlation analysis on standardized variables to investigate the relations between the following two sets of variables:

Set of X variables: items that measure "social trust" or "trust in public institutions", i.e., ppltrst, pplfair, pplhlp, trstprl, trstlgl, trstplc, trstplt

Set of Y variables: items that measure "well-being", i.e., fltdpr, fltsd,   fltanx, wrhpp, enjlf,   fltpcfl

Discuss the results of the canonical correlation analysis. How many canonical correlations are significant? How much of the variance in the Y variables can be explained by the X variables?

e. Use the split-half approach to assess the validity of the solution. Assign even-numbered observations to the calibration set and assign odd-numbered observations to the validation set when conducting this analysis. Discuss what you can conclude about the validity of the solution.

f. Which pairs of canonical variates are both important and reliable? How can you interpret these pairs of canonical variates?

References

ESS Round 6: European Social Survey Round 6 Data (2012). Data file edition 2.4. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS6-2012

Gómez-Balcácer, L., Somarriba Arechavala, N., & Gómez-Costilla, P. (2023). The importance of different forms of social capital for happiness in Europe: A multilevel structural equation model (GSEM). *Applied Research in Quality of Life, 18*, 601-624.

Submission of the assignment

For this assignment, one member of each team should upload the following files on Toledo:
- Report with answers to questions of both tasks (word document or .pdf file). The length of the report is limited to **maximum 12 pages (including one title page)**.
- Script File with the R-code (.R file)

Report with answers to the questions
For each question show the R-code followed by the relevant analysis output or graphs generated by R, and discuss in sufficient detail the results of the analysis to answer the question. Remark: include the R output using an appropriate font (e.g., courier) and layout.

Script file with R-code
- Include for each question all the R-code of the fitted models
- Add comments to the R-code
- Write the code so that it can be used to replicate all the reported analyses

Table 1: Description of variables

name	item	scale	concept
ppltrst	Most people can be trusted or you can't be too careful	0=you can't be too careful,..,10=most people can be trusted	social trust
pplfair	Most people try to take advantage of you, or try to be fair	0=most people try to take advantage of me, …, 10=most people try to be fair	social trust
pplhlp	Most of the time people are helpful or mostly looking out for themselves	0=people mostly look out for themselves, …, 10=people mostly try to be helpful	social trust
trstprl	Trust in country's parliament	0=no trust at all, …, 10=complete trust	trust institutions
trstlgl	Trust in the legal system	0=no trust at all, …, 10=complete trust	trust institutions
trstplc	Trust in the police	0=no trust at all, …, 10=complete trust	trust institutions
trstplt	Trust in politicians	0=no trust at all, …, 10=complete trust	trust institutions
fltdpr	Felt depressed, how often past week	4=none or almost none of the time, …, 1= all or almost all of the time	well-being
fltsd	Felt sad, how often past week	4=none or almost none of the time, …, 1= all or almost all of the time	well-being
fltanx	Felt anxious, how often past week	4=none or almost none of the time, …, 1= all or almost all of the time	well-being
wrhpp	Were happy, how often past week	1=none or almost none of the time, …, 4= all or almost all of the time	well-being
enjlf	Enjoyed life, how often past week	1=none or almost none of the time, …, 4= all or almost all of the time	well-being
fltpcfl	Felt calm and peaceful, how often past week	1=none or almost none of the time, …, 4= all or almost all of the time	well-being