Assignment 1 Multivariate Statistics 2025-2026

Task 1

Description data

The data set for this assignment (**life.Rdata**) is extracted from the database Wave 6 of the World Values Survey (www.worldvaluessurvey.org). The data set contains, for each of 52 countries, the average score of respondents on 6 items. The items measure to what extent the aspects "family", "friends", "leisure time", "politics", "work" and "religion" are important in their life. Items were measured using a 4-point likert scale (1=not at all important, .., 4=very important).

Questions

a. Conduct principal components analysis on the standardized variables. Use the procedure of Horn to evaluate how many principal components are needed to explain the data. Briefly discuss the results of the PCA and the output of the procedure of Horn.

b. Make a biplot that represents countries using standardized component scores and that represents items using component loadings and discuss what you can conclude from the biplot.

Task 2

Description of the data

The data set for this task (**fsdata.Rdata**) is extracted from the Cupesse data on Cultural Pathways to Economic Self-Sufficiency and Entrepreneurship (Tosun et al.,2016). Besides the variable country (country), the data set includes the responses of 4829 persons of two countries (Germany and UK) on 19 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 "current financial situation" (FS), "financial situation of family when person was 14 years old" (FSF), "lack of support in finding a job" (SFJ), "importance of job for self development" (SDJ), and "bad health" (HEALTH).

Questions

a. Use the fa() function to conduct exploratory factor analysis (EFA) on the standardized variables. Use the following specifications in the EFA: maximum likelihood estimation, 5 factors, oblique rotation. Discuss the output of the analysis. Investigate to what extent the EFA can reproduce the observed correlation matrix.

b. Conduct a confirmatory factor analysis on the covariance matrix of the items (i.e., on centered items) in Table 1 assuming a model with 5 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.

c. Use modification indices to select a few correlated error terms that can be added in order to obtain an acceptable descriptive model fit in terms of TLI and CFI (i.e. both CFI and TLI above .90). Try to interpret the correlated error terms from a substantive point of view (i.e., which specific aspect do items with correlated error terms have in common?)

d. Starting from the measurement model selected in step 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 variable "financial situation" can be explained by the other 4 latent variables.

   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.).

References

Tosun, J., Hörisch, F., Schuck, B. (2016). (ZA7475; Version 1.0.0) [Data set]. GESIS, Cologne. https://doi.org/10.4232/1.13042

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 14 pages (including one title page)**.
- Script File with the R-code (.R file)
- Completed code of conduct on the use of genAI for each team member

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
- If certain lines of the code were generated using GenAI, add a comment about this.

Table 1: Description of variables

latent variable	item	label	coding
financial situation	FS_pay_bills	Financial situation: I was able to pay bills by myself (over last 6 months)	1=never, ..,4=always
financial situation	FS_afford_extras	Financial situation: I could afford extras for myself (over last 6 months)	1=never, ..,4=always
financial situation	FS_afford_housing	Financial situation: I could afford decent housing (over last 6 months)	1=never, ..,4=always
financial situation	FS_save_money	Financial situation: I was able to save money (over last 6 months)	1=never, ..,4=always
financial situation family	FSF_pay_bills	Financial situation: My family was able to pay its bills (RECALL WHEN 14)	1=never, ..,4=always
financial situation family	FSF_afford_extras	Financial situation: We could afford extras for ourselves (RECALL WHEN 14)	1=never, ..,4=always
financial situation family	FSF_afford_housing	Financial situation: We could afford to live in decent housing (RECALL WHEN 14)	1=never, ..,4=always
financial situation family	FSF_save_money	Financial situation: My family was able to put money in a savings account (RECALL WHEN 14)	1=never, ..,4=always
lack of support finding job	SFJ_no_info	Not enough detailed information about job opportunities.	1=strongly disagree,..,4=strongly agree
lack of support finding job	SFJ_no_chance_show	Job applicants are not given a chance to demonstrate they could do a good job.	1=strongly disagree,..,4=strongly agree
lack of support finding job	SFJ_no_training	No opportunities to get the training needed to get a good job.	1=strongly disagree,..,4=strongly agree
lack of support finding job	SFJ_no_support_findjob	People who are unemployed don't get the right support in finding a job.	1=strongly disagree,..,4=strongly agree
importance job for self development	SDJ_help_people	How important: job allowing me to help other people	1=very unimportant, ..,4=very important
importance job for self development	SDJ_learn_new_things	How important: job allowing me to learn new things	1=very unimportant, ..,4=very important
importance job for self development	SDJ_develop_creativity	How important: job allowing me to develop my creativity	1=very unimportant, ..,4=very important
importance job for self development	SDJ_meet_people	How important: job allowing me to meet and interact with people	1=very unimportant, ..,4=very important
importance job for self development	SDJ_feeling_self_worth	How important: job giving me a feeling of self-worth	1=very unimportant, ..,4=very important
bad health	HEALTH_felt_down	HEALTH: How often you have felt down and that nothing could cheer you up (over last 3 months)	1=never, ..,4=always
bad health	HEALTH_limitation	HEALTH: How often your physical health has limited you in your daily life (over last 3 months)	1=never, ..,4=always