EPSY 8265: Take-home final exam
Due 5PM, December 20 (Wed), 2022

I will post a datafile named “personality.csv” to Canvas. This dataset has survey ratings for 452 individuals to 44 items on a personality questionnaire applied to employees from a company. The items are all scored on a 1-5 scale (1=not true of employee to 5=very true of employee). The variable labels for the survey items are provided in a separate work document named “personalityitems.docx”.

1. [20pt] We will use a subset of items from the dataset. You will first subset data by selecting ‘pers3’, ‘pers7’, ‘pers11’, ‘pers13’, ‘pers16’, ‘pers26’, and ‘pers36’ which, respectively, correspond to ‘does a thorough job’, ‘helpful’, ‘full of energy’, ‘reliable’, ‘generates enthusiasm’, ‘assertive’, and ‘outgoing’. You may, for example, use this code to subset the data: `subdat <- data[, c(3, 7, 11, 13, 16, 26, 36)]`

   (a) Is a single-factor model with seven indicators statistically identified? Attempt to verify its identification status using one or more of the identification rules discussed in class. [4pt]
   (b) Fit a single-factor model in ‘lavaan’ package and assess the model fit. Does this model fit the data? If so, interpret the resulting parameter estimates. [4pt]
   (c) Based on the inspection of local fit measures (e.g., standardized residual matrix, modification index), identify the largest (even if negligible) sources of misfit for this model. What might be done to respecify (change) the model to address this source of misfit? [4pt]
   (d) Fit the respecified model based on your evaluation in (c). Does this model fit better than the original one-factor model? Are the two models nested? Compare the fit of the two models. [4pt]
   (e) Interpret the estimated parameters derived from one of the two models that fits better to the data. [4pt]

2. Using the entire personality dataset, apply an EFA to all 44 items and answer the following. [20pt]

   (a) Are the data suitable for factor analysis? Explain why or why not. [2pt]
   (b) Determine an appropriate number of factors. Using the correlation matrix, perform a maximum likelihood factor analysis requesting a solution with the determined number of factors. Does the goodness-of-fit test suggest that the model fits with your determined number of factors? Explain the result based on what you know about the goodness-of-fit-test. [4pt]

(c) Use a rotation method to interpret the factor solution in (b). Do you prefer an orthogonal or oblique method? Why? Also, why are some of the factor loadings of largest magnitude negative? [4pt]

(d) Inspect the communalities of the items. Based on your interpretation of the factors, give an explanation as to why the communalities for some of the variables are so low (what it means to have low communalities). [3pt]

(e) For this dataset, why might factor analysis be a more appropriate form of analysis than principal component analysis? Explain how a principal component analysis would provide a different way of studying these data. [4pt]

(f) With these data, would it be reasonable to perform an exploratory factor analysis using the covariance matrix? Explain why or why not. How would you expect the results to change if the covariance matrix were used? [3pt]

3. A two-factor model with orthogonal factors was fit to a correlation matrix, producing following results: [10pt]

$$\widehat{\boldsymbol{\Lambda}} = \begin{bmatrix} .4 & .8 \\ .7 & -.4 \\ .1 & .7 \\ .5 & -.7 \\ .5 & -.3 \\ .4 & .2 \end{bmatrix}, \widehat{\boldsymbol{\Psi}} = \begin{bmatrix} .20 & & & & & \\ 0 & .35 & & & & \\ 0 & 0 & .50 & & & \\ 0 & 0 & 0 & .26 & & \\ 0 & 0 & 0 & 0 & .66 & \\ 0 & 0 & 0 & 0 & 0 & .80 \end{bmatrix}$$

(a) Compute by hand the reproduced (or fitted) correlation matrix for these variables [5pt]

(b) Compute by hand the estimated communalities and uniquenesses for the variables. [5pt]