

PSQF 6249 HW3: Confirmatory Factor Analysis (CFA) on Your Own Data
15 points; due Monday 3/25/2024 by 11:59 PM under “assignments” in ICON

Please submit this HW3 in an editable format (e.g., .docx or .rtf extension) using this file-naming convention: PSQF6249_Lastname_Firstname_HW3

The goal of HW3 is for you to practice conducting confirmatory factor analyses on data you care about (i.e., ideally using item responses from the instrument that you wrote about for HW1). If you have **six or more items measuring a single factor**, you may use only the items that correspond to a single dimension. If you have **fewer than six items** measuring a single dimension, please use enough items for **two factors** at once so that your model will be testable. Once you know how this analytic process works, you will be able to repeat it as needed for your other dimensions of interest, so the idea is to start with a model of limited size for now (which will also keep this homework of comparable difficulty across students).

Your task is to conduct a series of analyses to answer all questions given below. Use **MLR** (or an analogous robust maximum likelihood estimator) for all analyses and a z-scored factor model identification (factor mean = 0, factor variance = 1). Please note that I am not necessarily expecting you to arrive at a well-fitting model—that may not be possible for your data. Nor do I not want you to eliminate items based on R^2 to create a “short form” as in HW3—that was just an exercise designed to give you practice interacting with the syntax. Likewise, it may not make sense for you to move from one to two factors as I did in Example 4. **What I want to see is evidence that you can understand the information provided by your analysis well enough to make *defendable* revisions to your model as needed (not just following the specific choices I demonstrated, and also not just adding error covariances until it fits).** If you don’t know what to do about a poorly-fitting or non-converging model, please ask us for help! You do not **need** to submit any data, syntax, or output with this assignment. However, it’s fine with me if you **want** to include output in order for me to answer any questions you might have.

Please note the following with respect to your writing and the presentation of your results:

- Although there is a list of points to be included further below, **I DO NOT WANT A NUMBERED LIST BACK FROM YOU**. Instead, your text should read like a **traditional results section** in a published paper; see the Brown CFA (2015) book example results sections or the sample results sections from Example 4 for guidance. Each numbered point below should be answered in a new paragraph. In answering each question, make sure to describe the empirical criteria by which the answer was provided (i.e., what output content or model comparisons support your statements).
- I also want you to get practice with all relevant aspects of technical writing, including transitions between sections, contextual phrasing, and describing the contents of your tables and figures. For example, in a real results section for publication purposes, you would need to introduce each table and tell the reader what information it provides. You would need to provide and interpret the different types of results (i.e., beginning with descriptive analyses, transitioning into factor analyses) and maintain enough context for the reader to follow you.

- You should use not use the same short dash (-) for everything that is dash-like—please find and use the proper punctuation marks. For instance, – is a real minus sign (used for all negative numbers), – is an en-dash (used for compound phrases used as adjectives, such as *item–remainder correlation* or *parent–child conflict*), and — is an em-dash (used to set off phrases that clarify the previous phrase, like above; an en-dash with spaces on both sides could be used in place of an em-dash). I have added keyboard shortcuts on all my machines to make this easier (through the *insert symbol* menu in Word). Paying attention to these small details can help your writing look more professional!
- Use past tense throughout when describing how the data were obtained, how the analyses were conducted, and what you found. Present tense is ok if you are referring to the contents of the current tables and figures (e.g., “Table 1 shows...”).
- In describing the software you used, give the exact version (e.g., Mplus 8.10, not just Mplus 8). Things change quickly enough to where subversion differences can matter! You would also want to provide a reference for your software in a real paper. You can borrow the language I gave you in Example 4 to describe your model identification, which parameters were estimated, and how global and local fit was assessed, but you must change it as needed to fit your data (i.e., number of items, number of factors, and what their latent traits are supposed to be). Be consistent in naming your factors (in technical writing, synonyms make your story harder to follow, not more interesting).
- In tables of correlations, if you have complete data, it can be more efficient to provide the lower boundary for the value of correlation after which all higher correlations would be significant in a note rather than starring every significant value. You could also use bold-face type to denote significant correlations (and explain that in a table note) to reduce visual clutter. In practice you’d only show each correlation value once (either below or above the diagonal) and you do not need a leading 0 (via custom number formats).
- If you are reporting the fit of many models, it can be more efficient to do so in a table, and the same is true for model comparisons (see Model Fit Table 1 and LRTs Table 2 in the Example 4 spreadsheet). Otherwise, for just a few models, reporting their fit and comparisons thereof in the text is fine, but make sure you include all relevant fit stats: χ^2 test, CFI, TLI, RMSEA, and SRMR. For LRT fit comparisons, give the difference in $-2LL$ as $-2\Delta LL$, DF, and p -value.
- In your tables of model parameter estimates, ALL estimated parameters need to be included—this means the loadings, the intercepts, the residual variances, and any factor covariances. Given that you should have fixed factor means to 0 and factor variances to 1, these rows do not need to be included in the table. You do not need to report the z -value column or the p -value column. You should provide unstandardized estimates and their standard errors, and standardized estimates at a minimum (and preferably standard errors for the standardized estimates, if given, too). See Model Estimates Table 3 in the Example 4 spreadsheet for a template. In practice, however, you would not need to give the R^2 values, given they are the squared versions of the standardized loadings (which you would have already included).

- APA style (or whatever style you should be publishing in) should be used for all tables and figures—please do not paste in unformatted program output and call it a table. Btw, tables can start in excel so that you can use the number formatting options to control how many digits show after the decimal consistently within columns, but then you'd bring the table into your document and provide a proper title (with font large enough to be legible).
- In plotting estimated factor scores, Mplus gives a bar chart for factor scores by default (i.e., a plot of the frequency of each unique score, where the x-axis is categorical, not numeric). To make a histogram instead, under the *Plot* menu in our output, select *view plots*, then *histograms*, under *plot properties* (left tab) scroll down until you see the name of your factor. Then under the center tab *display properties*, pick the second option of *histogram/density plot* and hit *ok*. Under the *Plot* menu, select *axis properties* and *edit settings*, and then you can customize your axes and titles.
- In creating a factor–response plot using my Example 4 excel spreadsheet, you should have entered the unstandardized intercepts and factor loadings (as the slopes), and then predicted item responses for ± 3 SDs should have been calculated and plotted for you, as well as the lowest and highest item responses. You can check to see if the plot is correct by noting the predicted item response when Factor=0, which should match the intercept, or else something is wrong. To change the x-axis labels, right-click on the plot, choose *select data*, and then select *edit under horizontal (category) axis labels*. You should be able to highlight the cells holding the values to be plotted on the x-axis.
- The factor–response plots were designed to illustrate whether a linear relationship between the factor and the predicted items responses were plausible. I also wanted you to link these results to the distribution of the predicted factor scores. Btw, on your reading list, Fernando (2009) provides the derivations for the trait level at which predicted item responses will go out of bounds, as well as an analogous version of item difficulty for CFA models (like b_i in IRT models), which are implemented in the Example 4 spreadsheet.

Items (and their point values) to be included (you can earn up to 2 points for writing quality and proper use of APA style):

1. Begin by summarizing the construct(s) being measured and the items themselves, including how many items there are and what their response options are. Also provide your sample size and briefly describe the sample. Provide all relevant modeling info: program, estimator, how each model was identified, how models will be compared, and what criteria you are using to indicate “good fit” (i.e., cut-off values) both globally and locally. The idea is that a reader should be able to replicate your analyses given the information provided. **(1 point)**
2. Provide a table of item descriptive statistics, including columns for each item’s content (if possible), sample size, mean, standard deviation, minimum, maximum; also provide each item’s item–remainder correlation *per hypothesized latent factor*. Also provide a Pearson correlation matrix for your items (also organized by factor if your indicators belong to more than a single factor). Note that you should be able to paste output directly into excel in order to make these tables easily (i.e., no manual typing of rows of numbers should be required).

Comment on your items' difficulty and discrimination, as well as the size and heterogeneity of your inter-item correlations. **(3 points)**

3. Estimate a CFA model that corresponds to your hypothesized dimensionality. Report the relevant fit statistics and describe by which indices good fit has been achieved globally. Provide the range of effect sizes across indicators (i.e., standardized loadings). Examine and describe any local misfit. If your model fit is not adequate, considering its sources of local misfit *and* the item content, re-specify your model to try to improve fit. Note that any model modifications should be theoretically defensible, so provide a rationale for these modifications. Describe the model modification process you followed, and conduct any relevant model comparisons to support your modifications. **(4 points)**
4. Once your fit is as good as it is going to get and your model is still theoretically defensible, you can call it a final model. Compute omega for each retained factor. Provide and reference a table of ALL estimate model parameters, including columns for unstandardized estimates, their standard errors, and standardized estimates. Use the "text to columns" feature in the Data menu of Excel to make this easier, but make sure each parameter is clearly labeled (i.e., do not leave the impoverished labels used by Mplus). **(2 points)**
5. Provide and reference a histogram of your sample's FSCORE distribution. Note that Mplus in Windows will do this for you as a PLOT option; otherwise, import the saved FSCORE file into another program to do so. Also provide and reference a factor-response plot that shows the predicted indicator responses at ± 2 SD of the factor for the two items with the lowest and highest CFA difficulty (see the Example 4 spreadsheet for help). Comment on how plausible a linear model predicting the indicator responses from your factor is for your data. **(3 points)**