STA 431s11 Assignment 11

Do this assignment in preparation for the quiz on Friday, April 1st. Answers to the non-SAS questions are practice for the the quiz, and are not to be handed in. For Question 4, bring both your log file and your list file to the quiz; they may (or may not) be handed in.

1. In this factor analysis model, the observed variables are *not* standardized, and the factor loading for D_1 is set equal to one. Let

$$
\begin{aligned}
D_1 &= F + e_1 \\
D_2 &= \lambda_2 F + e_2 \\
D_3 &= \lambda_3 F + e_3,
\end{aligned}
$$

   where $F \sim N(0, \phi)$, e_1, e_2 and e_3 are normal and independent of F and each other with expected value zero, $V(e_1) = \omega_1, V(e_2) = \omega_2, V(e_3) = \omega_3$, and λ_2 and λ_3 are nonzero constants.

   (a) Calculate the variance-covariance matrix of the observed variables.

   (b) Are the model parameters identifiable? Answer Yes or No and prove your answer.

2. We now extend the preceding model by adding another factor. Let

$$
\begin{aligned}
D_1 &= F_1 + e_1 \\
D_2 &= \lambda_2 F_1 + e_2 \\
D_3 &= \lambda_3 F_1 + e_3 \\
D_4 &= F_2 + e_4 \\
D_5 &= \lambda_5 F_2 + e_5 \\
D_6 &= \lambda_6 F_2 + e_6,
\end{aligned}
$$

   where all expected values are zero, $V(e_i) = \omega_i$ for $i = 1, \ldots, 6$,

$$
V \left[\begin{array}{c} F_1 \\ F_2 \end{array} \right] = \left[\begin{array}{cc} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{array} \right],
$$

   and $\lambda_2, \lambda_3, \lambda_5$ and λ_6 are nonzero constants.

   (a) Give the covariance matrix of the observable variables. Show the necessary work. A lot of the work has already been done in Question 1.

   (b) Are the model parameters identifiable? Answer Yes or No and prove your answer.

3. Let's add a third factor to the model of Question 2. That is, we add

$$\begin{array}{rcl} D_7 & = & F_3 + e_7 \\ D_8 & = & \lambda_8 F_3 + e_8 \\ D_9 & = & \lambda_9 F_3 + e_9 \end{array}$$

and

$$V \left[\begin{array}{c} F_1 \\ F_2 \\ F_3 \end{array} \right] = \left[\begin{array}{ccc} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{12} & \phi_{22} & \phi_{23} \\ \phi_{13} & \phi_{23} & \phi_{33} \end{array} \right],$$

with $\lambda_8 \neq 0$, $\lambda_9 \neq 0$ and so on. Are the model parameters identifiable? You don't have to do any calculations if you see the pattern.

4. Return to the little math data set `littlemath.data` used in Assignment 9; see the link on the course web page in case the one from this document does not work. The file contains data from a study of students taking first-year calculus on one campus of a large North American university. The students took a diagnostic test with two parts: pre-calculus and calculus. Marks in high school calculus were available, as were marks in first-year university calculus. You will recall that a single-factor model did not fit the data well. So now please try a 2-factor model confirmatory factor analysis, but it is important to be careful because there are only four observable variables.

   Assume that there are two correlated factors. The first reflects potential performance in a calculus class, and the second reflects potential performance on multiple choice calculus tests. Clearly, variables one and four come from the first (classroom) factor, while variables two and three come from the second (multiple choice test) factor.

   The case of two factors and just four variables is not covered in anything we have done in class, but the parameters may be identifiable. For identifiability, fix one factor loading for each factor to one; this results in a re-parameterization.

   (a) Give a formal statement of the model. Your specification should include
      i. The model equations
      ii. The covariance matrix of the error terms
      iii. The covariance matrix of the factors
      iv. Statements that certain parameters are non-zero.

   (b) Draw a path diagram of the model, with coefficients on the arrows.

   (c) Prove that the parameters are identifiable given that certain parameters are non-zero. Why is it important that $\phi_{12} \neq 0$?

(d) Fit the model using `proc calis`.

i. Does this two-factor model fit the data adequately? Answer Yes or No, and give three numbers to back up your answer.

ii. Suppose you wanted to estimate a *ratio*: the factor loading for high school calculus mark divided by the factor loading for university calculus mark. Because I set the factor loading for university calculus to one, the estimate is a number directly on my printout. If you did it the other way (which is fine), you will have to take the reciprocal. In any case, that is the estimate? The answer is a single number.

iii. Can you figure out how to do a Z-test of the null hypothesis that the ratio of the two factor loadings is equal to one? Carry out this mild calculator work if you can.

(e) You may assume that because the parameters of this model are identifiable, so are the parameters of a similar model for standardized variables. So, please fit a model with standardized factors to the data. This is a second run of `proc calis`. Please put it in the same SAS program.

i. Estimate the correlation between the two factors. Your answer is a single number from the printout.

ii. Using the output from this second model, again please estimate the factor loading for high school calculus mark divided by the factor loading for university calculus mark. It should agree with what you got from the first model. This is quick calculator work. You are using the invariance principle here.

iii. Look at the test for fit, and compare it to the test for the first model you tried. Are they the same?

(f) Well, apparently the two re-parameterized models impose the same equality constraint on the covariances of the observable variables (how do you know there is just one constraint?). Now you will find out what that constraint is.

When you were proving identifiability for the first model, you probably did not use the covariance between the two variables with factor loadings not equal to one. So (still for the first model, the one with two factor loadings set to one), obtain explicit solutions for the three parameters involved in terms of $\sigma_{i,j}$ quantities, and substitute them back into the expression for the unused covariance. My answer has one product of covariances equal to another product.

(g) Show that this equality restriction also holds for the model with standardized factors. To do it, you will need to calculate $\boldsymbol{\Sigma}$ for the second model.