# STA 431s11 Assignment 10

Do this assignment in preparation for the quiz on Friday, March 25th.

1. Consider the general factor analysis model

$$\mathbf{D} = \boldsymbol{\Lambda}\mathbf{F} + \mathbf{e},$$

where $\boldsymbol{\Lambda}$ is a $k \times p$ matrix of factor loadings, the vector of factors $\mathbf{F}$ is a $p \times 1$ multivariate normal with expected value zero and covariance matrix $\boldsymbol{\Phi}$, and $\mathbf{e}$ is multivariate normal with expected value zero and covariance matrix $\boldsymbol{\Omega}$. All covariance matrices are positive definite.

   (a) Calculate the matrix of covariances between the observable variables $\mathbf{D}$ and the underlying factors $\mathbf{F}$.

   (b) Give the covariance matrix of $\mathbf{D}$. Show your work.

   (c) Any positive definite matrix can be written as $\mathbf{SS}'$ (the matrix $\mathbf{S}$ is called the *square root matrix*). Using the square root matrix of $\boldsymbol{\Phi}$, show that the parameters of the general factor analysis model are not identifiable.

   (d) In an attempt to obtain a model whose parameters can be successfully estimated, let $\boldsymbol{\Omega}$ be diagonal (errors are uncorrelated) and set $\boldsymbol{\Phi}$ to the identity matrix (standardizing the factors). Show that the parameters of this revised model are still not identifiable.

2. Here is a factor analysis model in which all the observed variables are *standardized.* That is, they are divided by their standard deviations as well as having the means subtracted off. This gives them mean zero and variance one. Therefore, we work with a correlation matrix rather than a covariance matrix; that's the classical way to do factor analysis.

$$\begin{aligned} Z_1 &= \lambda_1 F_1 + e_1 \\ Z_2 &= \lambda_2 F_2 + e_2 \\ Z_3 &= \lambda_3 F_3 + e_3, \end{aligned}$$

where $F_1$, $F_2$ and $F_3$ are independent $N(0,1)$, $e_1$, $e_2$ and $e_3$ are normal and independent of each other and of $F_1$, $F_2$ and $F_3$, $V(Z_1) = V(Z_2) = V(Z_3) = 1$, and $\lambda_1$, $\lambda_2$ and $\lambda_3$ are nonzero constants. The expected values of all random variables equal zero.

   (a) What is $V(e_1)$? $V(e_2)$? $V(e_3)$?

   (b) What is $Corr(F_1, Z_1)$?

(c) Give the communality of $Z_j$. Recall that the communality is the proportion of variance explained by the common factor(s). That is, it is the proportion of $Var(Z_j)$ that does not come from $e_j$.

(d) Give the variance-covariance matrix (correlation matrix) of the observed variables.

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

(f) Even though the parameters are not identifiable, the model itself is testable. That is, it implies a set of equality restrictions on the correlation matrix $\mathbf{\Sigma}$ that could be tested, and rejecting the null hypothesis would call the model into question. State the null hypothesis. Again, it is a statement about the $\sigma_{i,j}$ values.

3. Here is another factor analysis model. This one has a single underlying factor. Again, all the observed variables are standardized.

$$\begin{aligned} Z_1 &= \lambda_1 F + e_1 \\ Z_2 &= \lambda_2 F + e_2 \\ Z_3 &= \lambda_3 F + e_3, \end{aligned}$$

where $F \sim N(0,1)$, $e_1$, $e_2$ and $e_3$ are normal and independent of $F$ and each other with expected value zero, $V(Z_1) = V(Z_2) = V(Z_3) = 1$, and $\lambda_1$, $\lambda_2$ and $\lambda_3$ are nonzero constants with $\lambda_1 > 0$.

(a) What is $V(e_1)$? $V(e_2)$? $V(e_3)$?

(b) Give the communality of $Z_j$.

(c) Write the reliability of $Z_j$ as a measure of $F$. Recall that the reliability is defined as the squared correlation of the true score with the observed score.

(d) Give the variance-covariance (correlation) matrix of the observed variables.

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

4. Suppose we added another variable to the model of Question 3. That is, we add

$$Z_4 = \lambda_4 F + e_4,$$

with assumptions similar to the ones of Question 3. Now suppose that $\lambda_2 = 0$.

(a) Is $\lambda_2$ identifiable? Justify your answer.

(b) Are the other factor loadings identifiable? Justify your answer.

(c) State the general pattern that is emerging here.

5. Suppose we added a fifth variable to the model of Question 4. That is, we add

$$Z_5 = \lambda_5 F + e_5,$$

with assumptions similar to the ones of Question 3. Now suppose that $\lambda_3 = \lambda_4 = 0$.

   (a) Are $\lambda_3$ and $\lambda_4$ identifiable? Justify your answer.
   (b) Are the other three factor loadings identifiable? Justify your answer.
   (c) State the general pattern that is emerging here.

6. We now extend the model of Question 3 by adding a second factor. Let

$$\begin{aligned} Z_1 &= \lambda_1 F_1 + e_1 \\ Z_2 &= \lambda_2 F_1 + e_2 \\ Z_3 &= \lambda_3 F_1 + e_3 \\ Z_4 &= \lambda_4 F_2 + e_4 \\ Z_5 &= \lambda_5 F_2 + e_5 \\ Z_6 &= \lambda_6 F_2 + e_6, \end{aligned}$$

where all expected values are zero, $V(e_i) = \omega_i$ for $i = 1, \ldots, 6$, $V(F_1) = V(F_2) = 1$, $Cov(F_1, F_2) = \phi_{12}$, the factors are independent of the error terms, and all the error terms are independent of each other. All the factor loadings are non-zero with $\lambda_1 > 0$ and $\lambda_4 > 0$.

   (a) Give the covariance matrix of the observable variables. Show the necessary work. A lot of the work has already been done.
   (b) Are the model parameters identifiable? Answer Yes or No and prove your answer.

7. In Question 6, suppose we added just two variables along with the second factor. That is, we omit the equation for $Z_6$. Are the model parameters identifiable in this case? Answer Yes or No; show your work.

8. Let's add a third factor to the model of Question 6. That is, we add

$$\begin{aligned} Z_7 &= \lambda_7 F_3 + e_7 \\ Z_8 &= \lambda_8 F_3 + e_8 \\ Z_9 &= \lambda_9 F_3 + e_9 \end{aligned}$$

with $\lambda_7 > 0$ and other assumptions similar to the ones we have been using. Are the model parameters identifiable? You don't have to do any calculations if you see the pattern.