

# STA 2101f19 Assignment Nine<sup>1</sup>

1. A farm co-operative (co-op) is an association of farmers. The co-op can buy fertilizer and other supplies in large quantities for a lower price, it often provides a common storage location for harvested crops, and it arranges sale of farm products in large quantities to grocery store chains and other food suppliers. Farm co-ops usually have professional managers, and some do a better job than others.

We have data from a study of farm co-op managers. The variables in the “latent variable” part of the model are the following, but note that one of them is assumed observable.

- Knowledge of business principles and products (economics, fertilizers and chemicals). This is a latent variable measured by `know1` and `know2`.
- Profit-loss orientation (“Tendency to rationally evaluate means to an economic end”). This is a latent variable measured by `ploss1` and `ploss2`.
- Job satisfaction. This is a latent variable measured by `sat1` and `sat2`.
- Formal Education This is an observable variable, assumed to be measured without error.
- Job performance. This is a latent variable measured by `perf1` and `perf2`.

The data file has these observable variables in addition to an identification code for the managers.

`know1`: Knowledge measurement 1

`know2`: Knowledge measurement 2

`ploss1`: Profit-Loss Orientation 1

`ploss2`: Profit-Loss Orientation 2

`sat1`: Job Satisfaction 1

`sat2`: Job Satisfaction 2

`educat`: Number of years of formal schooling divided by 6.

`perf1`: Job Performance 1

`perf2`: Job Performance 2

In this study, the double measurements are obtained by just splitting questionnaires in two, as in split half reliability. Furthermore, all the measurement errors are assumed independent of one another. This is consistent with mainstream psychometric theory,

---

<sup>1</sup>This assignment was prepared by [Jerry Brunner](#), Department of Statistical Sciences, University of Toronto. It is licensed under a [Creative Commons Attribution - ShareAlike 3.0 Unported License](#). Use any part of it as you like and share the result freely. The L<sup>A</sup>T<sub>E</sub>X source code is available from the course website: <http://www.utstat.toronto.edu/~brunner/oldclass/2101f19>

though maybe not with common sense. For this assignment, please assume that the errors are independent of one another, and independent of the exogenous variables. The explanatory variables, of course, should *not* be assumed independent of one another.

In the two main published analyses of these data, the latent exogenous variables were knowledge, profit-loss orientation, education and job satisfaction. The latent response variable was job performance. However, let's make it more interesting. Let's say that the latent exogenous variables are knowledge, education and profit-loss orientation, and that these influence job performance (possibly with a zero regression coefficient; we can test that). Job performance is also influenced by job satisfaction. Job satisfaction, in turn, is influenced by job performance (it feels good to do a good job), but not directly by any of the exogenous variables. So job satisfaction is endogenous too.

- (a) Please make a path diagram. put Greek letters on all the arrows, including curved arrows, unless the coefficient is one.
- (b) List the parameters that appear in the covariance matrix of the observable data.
- (c) Does this model pass the test of the parameter count rule? Answer Yes or no and give the numbers.

The parameters of this model are identifiable in most of the parameter space. Details will be taken up in class.

2. The file [co-opManager.data.txt](#) has raw data for the study described in Question 1. This is a reconstructed data set based on a covariance matrix in Joreskog (1978, p. 465). Joreskog got it from Warren, White and Fuller (1974).

Using `lavaan`, fit the model in your path diagram and look at `summary`. There are 98 co-ops, so please make sure you are reading the correct number of cases. For comparison, my value of the  $G^2$  test statistic for model fit is 29.357. If you got this, we must be fitting the same model.

- (a) Based on the number of covariance structure equations and the number of unknown parameters, how many equality restrictions should the model impose on the covariance matrix? The answer is a single number; fortunately, you need not say exactly what the equality restrictions are.
- (b) Does your model fit the data adequately? Answer Yes or No and give three numbers: a chisquared statistic, the degrees of freedom, and a  $p$ -value. The degrees of freedom should agree with your answer to Question 2(a)i.
- (c) In plain, non-statistical language, what are the main conclusions of this study? Be able to back up your conclusions with hypothesis tests that reject  $H_0$  at  $\alpha = 0.05$ . Of course you keep quiet about it in your plain-language conclusions.
- (d) It's remarkable that one can assess the effect of satisfaction on performance *and* the effect of performance on satisfaction. Be able to give the value of the test

statistics and the  $p$ -values. It's a little disappointing, but these data are a recreation of a real data set. Measurable job satisfaction is notoriously unrelated to any actual behavior — unless that behaviour consists of more talk.

- (e) Carry out a Wald test of all the regression coefficients in the latent variable model at once; I count three  $\gamma_j$  and two  $\beta_j$ . Be able to give the value of the chi-squared test statistic, the degrees of freedom, and the  $p$ -value – all numbers from your printout. Using the usual  $\alpha = 0.05$  significance level, is there evidence that at least one regression coefficient must be non-zero? You can tell which ones from the output of `summary`.
- (f) Estimate the reliability of knowledge measure one and knowledge measure two; give 95% confidence intervals as well. There is an easy way to do this. I almost asked about the reliability of job satisfaction, which is a nightmare for this model.
- (g) Test whether the reliabilities of the two knowledge measures are equal; as you know, this is equivalent to testing equality of the measurement error variances. Be able to give the value of the test statistic, the  $p$ -value, and draw a directional conclusion if one is warranted.
- (h) There is another way to estimate reliability. Suppose that  $D_1 = F + e_1$  and  $D_2 = F + e_2$ . If  $Var(e_1) = Var(e_2) = \omega$ , we call the measurements “equivalent,” and their common reliability is  $\phi/(\omega + \phi)$ . Calculate  $Corr(D_1, D_2)$ . This suggests a sample correlation as an estimate of reliability.
- (i) Use the `cor` function to get a sample correlation matrix of all the observable variables. Assuming the measurements of knowledge are equivalent, can you find another estimate of the common reliability? How does it compare to your earlier estimates?

### 3. Consider the general factor analysis model

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

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

- (a) How do you know that  $\mathbf{D}_i$  is multivariate normal?
- (b) Calculate the matrix of covariances between the observable variables  $\mathbf{D}_i$  and the underlying factors  $\mathbf{F}_i$ .
- (c) Give the covariance matrix of  $\mathbf{D}_i$ . Show your work.
- (d) Because  $\mathbf{\Phi}$  symmetric and positive definite, it has a square root matrix that is also symmetric. Using this, show that the parameters of the general factor analysis model are not identifiable.

- (e) In an attempt to obtain a model whose parameters can be successfully estimated, let  $\mathbf{\Omega}$  be diagonal (errors are uncorrelated) and set  $\mathbf{\Phi}$  to the identity matrix (standardizing the factors). Show that the parameters of this revised model are still not identifiable. Hint: An orthogonal matrix  $\mathbf{R}$  (corresponding to a rotation) is one satisfying  $\mathbf{R}\mathbf{R}^\top = \mathbf{I}$ .
4. 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,  $Var(e_1) = \omega_1$ ,  $Var(e_2) = \omega_2$ ,  $Var(e_3) = \omega_3$ , and  $\lambda_2$  and  $\lambda_3$  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.
5. 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,  $Var(e_i) = \omega_i$  for  $i = 1, \dots, 6$ ,

$$cov \begin{pmatrix} F_1 \\ F_2 \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{pmatrix},$$

and  $\lambda_2, \lambda_3, \lambda_5$  and  $\lambda_6$  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 4.
- (b) Are the model parameters identifiable? Answer Yes or No and prove your answer.

6. Let's add a third factor to the model of Question 5. That is, we add

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

and

$$\text{cov} \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{12} & \phi_{22} & \phi_{23} \\ \phi_{13} & \phi_{23} & \phi_{33} \end{pmatrix},$$

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.

**Bring a printout with your R input and output to the quiz. Please remember that while the questions may appear in comment statements, answers and interpretation may not, except for numerical answers generated by R.**