

Farm Co-op Managers

The only R questions on the final exam will be based on the Co-op manager data. You will answer based on my R input and output.

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

The data file has the following 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. For this assignment, please assume that all the measurement errors are independent of one another. This is consistent with mainstream psychometric theory, though maybe not with common sense. Also, assume that the measurement errors are independent of the exogenous variables. The exogenous 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 zero regression coefficients; 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.

1. Please make a path diagram. Put Greek letters on all the arrows, including curved arrows, unless the coefficient is one.
2. The file <https://www.utstat.toronto.edu/brunner/openSEM/data/co-opManager.data.txt> has raw data. 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`. Do some more tests. Try to understand the output. The analyses I will do for the exam will be straightforward.

My model has 21 parameters, and so the hypothesis matrices for Wald tests are big. To get around this, one can set up the matrices without typing them in. Suppose you want to test $H_0 : \theta_3 = \theta_7 = 0$. The following will work.

```
L = matrix(0,2,21); L[1,3] = L[2,7] = 1
```

That's how I'm going to do it for the final exam. You might want to try a few examples so you get used to this approach.

Also, to save space I'm going to use `show` and `parameterEstimates` in place of `summary`.