NES8010 Data Analaysis and Modelling

Structural Equation Modelling

February 2021

Structural Equation Modelling is a method for investigating the impacts of variables in systems where there are multiple pathways to the final effect. The approach is used to challenge a hypothetical representation of the system pathways using data. A good text book for this is Grace 2006 ‘Structural Equation Modelling and Natural Systems’, also see the website <http://www.structuralequations.com/>. The underlying assumption of SEM is that the data are approximately multivariate normal. This assumption may not always hold. In which case the model may appear to be good but in reality flawed by failure of this assumption. You will recall that we always do normality tests with GLM, the same assumptions apply here. Please consult Grace for the underlying theory.

Here we are going to use the approach to investigate the factors impacting on the trends in fox and lynx in Sweden, as explored using linear mixed effects models.

There are several packages for fitting SEM, the most popular are

Lavaan [Yves Rosseel (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. URL <http://www.jstatsoft.org/v48/i02/>]

and piecewiseSEM [Lefcheck, Jonathan S. (2016) piecewiseSEM: Piecewise structural equation modeling in R for ecology, evolution, and systematics. Methods in Ecology and Evolution. 7(5): 573-579. DOI: 10.1111/2041-210X.12512]

This is also a good, well laid out tutorial on using the package piecewiseSEM in R. <https://cran.r-project.org/web/packages/piecewiseSEM/vignettes/piecewiseSEM.html#worked-example>

NB this topic is developing - for latest notes and code please see:

<https://jonlefcheck.net/teaching/>

We will use `piecewiseSEM as the syntax is very similar to lme and it allows us to build hierarchical models.

Load the libraries, dependencies and data

install.packages("piecewiseSEM")  
install.packages("lmerTest")  
install.packages("nlme")  
install.packages("here")

library(piecewiseSEM)  
library(lmerTest)  
library(nlme)  
library(here)  
  
rawd<-read.csv(here("Data", "sweden\_simple.csv"))

After carrying out the usual data exploration and linear modeling the next step is to design a conceptual model (you may wish to do this on paper!)

We then code the conceptual model as a series of lme models. here is some code to get started on the modelling the red fox and lynx data.

sem.m1 = psem(lme(red~seed+pop+I(lynx+1) ,   
 random = ~1|name, na.action = na.omit, data = rawd),   
 lme(seed~pop, random = ~1|name, na.action = na.omit,data = rawd),   
 lme(I(lynx+1)~pop+north, random = ~1|name, na.action = na.omit, data = rawd),   
 lme(pop~north,random = ~1|name,data = rawd)   
 )

We can then extract the model outputs such as the model fit, and coefficients.

summary(sem.m1)

## | | | 0% | |======================= | 33% | |=============================================== | 67% | |======================================================================| 100%

##   
## Structural Equation Model of sem.m1   
##   
## Call:  
## red ~ seed + pop + I(lynx + 1)  
## seed ~ pop  
## I(lynx + 1) ~ pop + north  
## pop ~ north  
##   
## AIC BIC  
## 154.812 257.307  
##   
## ---  
## Tests of directed separation:  
##   
## Independ.Claim Test.Type DF Crit.Value P.Value   
## seed ~ north + ... coef 21 -2.2073 0.0386 \*  
## red ~ north + ... coef 21 0.2203 0.8278   
## I(lynx + 1) ~ seed + ... coef 1602 -10.4086 0.0000 \*\*\*  
##   
## Global goodness-of-fit:  
##   
## Fisher's C = 116.812 with P-value = 0 and on 6 degrees of freedom  
##   
## ---  
## Coefficients:  
##   
## Response Predictor Estimate Std.Error DF Crit.Value P.Value  
## red seed 0.0029 0.0002 1601 15.2454 0.0000  
## red pop 0.3290 0.1390 1601 2.3668 0.0181  
## red I(lynx + 1) -7.8294 1.3057 1601 -5.9965 0.0000  
## seed pop 490.9280 14.3749 1603 34.1516 0.0000  
## I(lynx + 1) pop -0.0292 0.0020 1603 -14.2814 0.0000  
## I(lynx + 1) north -0.0009 0.0003 21 -3.1229 0.0051  
## pop north -0.0324 0.0080 21 -4.0551 0.0006  
## Std.Estimate   
## 0.7769 \*\*\*  
## 0.1214 \*  
## -0.1052 \*\*\*  
## 0.6676 \*\*\*  
## -0.8021 \*\*\*  
## -0.4523 \*\*  
## -0.6219 \*\*\*  
##   
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05  
##   
## ---  
## Individual R-squared:  
##   
## Response method Marginal Conditional  
## red none 0.53 0.77  
## seed none 0.51 0.90  
## lynx none 0.25 0.52  
## pop none 0.37 0.87

## Assessing a SEM

SEM outputs appear complex. There are many measures that can be used to assess model fit. In piecewise SEM we can use the AIC to compare nested models:

To assess the model first look at the significance of the coefficients. Any Non-Significant variables should be removed to create a simpler better fitting model. RUn the revised model then compare the two using the AIC.

sem.m2 = psem(lme(red~seed+pop+I(lynx+1) ,   
 random = ~1|name, na.action = na.omit, data = rawd),   
 lme(seed~pop, random = ~1|name, na.action = na.omit,data = rawd),   
 lme(I(lynx+1)~pop+north, random = ~1|name, na.action = na.omit, data = rawd),   
 lme(pop~north,random = ~1|name,data = rawd)   
 )

The two models can be compared using the AIC

summary(sem.m2)

## | | | 0% | |======================= | 33% | |=============================================== | 67% | |======================================================================| 100%

##   
## Structural Equation Model of sem.m2   
##   
## Call:  
## red ~ seed + pop + I(lynx + 1)  
## seed ~ pop  
## I(lynx + 1) ~ pop + north  
## pop ~ north  
##   
## AIC BIC  
## 154.812 257.307  
##   
## ---  
## Tests of directed separation:  
##   
## Independ.Claim Test.Type DF Crit.Value P.Value   
## seed ~ north + ... coef 21 -2.2073 0.0386 \*  
## red ~ north + ... coef 21 0.2203 0.8278   
## I(lynx + 1) ~ seed + ... coef 1602 -10.4086 0.0000 \*\*\*  
##   
## Global goodness-of-fit:  
##   
## Fisher's C = 116.812 with P-value = 0 and on 6 degrees of freedom  
##   
## ---  
## Coefficients:  
##   
## Response Predictor Estimate Std.Error DF Crit.Value P.Value  
## red seed 0.0029 0.0002 1601 15.2454 0.0000  
## red pop 0.3290 0.1390 1601 2.3668 0.0181  
## red I(lynx + 1) -7.8294 1.3057 1601 -5.9965 0.0000  
## seed pop 490.9280 14.3749 1603 34.1516 0.0000  
## I(lynx + 1) pop -0.0292 0.0020 1603 -14.2814 0.0000  
## I(lynx + 1) north -0.0009 0.0003 21 -3.1229 0.0051  
## pop north -0.0324 0.0080 21 -4.0551 0.0006  
## Std.Estimate   
## 0.7769 \*\*\*  
## 0.1214 \*  
## -0.1052 \*\*\*  
## 0.6676 \*\*\*  
## -0.8021 \*\*\*  
## -0.4523 \*\*  
## -0.6219 \*\*\*  
##   
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05  
##   
## ---  
## Individual R-squared:  
##   
## Response method Marginal Conditional  
## red none 0.53 0.77  
## seed none 0.51 0.90  
## lynx none 0.25 0.52  
## pop none 0.37 0.87

AIC(sem.m1, sem.m2)

## df AIC  
## x 19 154.812  
## y 19 154.812

## Interpreting a SEM

The standardised coefficients give an estimate of the relative contribution of key variables to the variation in others. However, if you have pathways between three or more variables then the net contribution is the product of the coefficients. So if the coefficients of variables A->B is 0.6; B->C is 0.3, then the impact of A on C is the product of 0.6\*0.3 (0.18). If there are multiple pathways between A and C (e.g. involving D), then the effect of A is the sum of all of the potential pathway products.