Looking at Familiar Statistical Concepts in a New SEM Light

Jarrett E. K. Byrnes

Data from:
Biodiversity and complex environmental forcing of ecosystem functioning in the marine foundation species, eelgrass:

Matt Whalen, J. Emmett Duffy, Jim Grace

Old Wine in a New Bottle
1. ANOVA and ANCOVA in an SEM context
2. Multiple categorical predictors
3. Nonlinear effects

York River, Virginia:
Major herbivores are invert crustaceans - these grazers control epiphytes and promote the eelgrass
Pesticide effects:
- Crustaceans: reduced 58-96%
- Algal biomass: increased 130-748%
- Nutrients: inconsistent effects

Experimental Design:
- Treatments:
  - pesticide to reduce crustacean grazers
  - nutrient addition
  - combination
  - controls
- 8 reps @ 5 trts = 40 plots

Matt Whalen

Using Summarized Information

\[
\text{anovaModel} \leftarrow \text{`lnchla} - \text{pesticide'}
\]

\[
\text{anovaFit} \leftarrow \text{sem(anovaModel, sample.cov=whalenCov, sample.mean=whalenMeans, sample.nobs=whalenN)}
\]
Seagrass ANCOVA Model

*AMOS v. lavaan*

```
ancovaModel <- 'lnchla ~ pesticide +
               macroalgae + grass
pesticide ~~ 0*macroalgae + 0*grass'
ancovaFit <- sem(ancovaModel, sample.cov=whalenCov,
               sample.mean=whalenMeans, sample.nobs=whalenN,
               fixed.x=F)
```

Mediation Exercise 2!

1. Fit fully mediated and partially mediated model
   with LNGamm
2. Evaluate evidence for Partial or Full Mediation
3. Bonus: fit the full model with LNCaprell

The Models

```
fullModel <- 'lnchla ~ macroalgae + grass + LNGamm
             LNGamm ~ macroalgae + grass + pesticide
             pesticide ~ 0*macroalgae + 0*grass'
partialModel <- 'lnchla ~ macroalgae + grass + LNGamm + pesticide
                 LNGamm ~ macroalgae + grass + pesticide
                 pesticide ~ 0*macroalgae + 0*grass'
```

Likelihood Ratio Comparison

```
> anova(fullFit, partialFit)
Chi Square Difference Test

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>Chisq</th>
<th>Chisq.Df</th>
<th>Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fullFit</td>
<td>2</td>
<td>442.95</td>
<td>463.22</td>
<td>5.1475</td>
<td>4.5422</td>
<td>1</td>
<td>0.03307</td>
</tr>
<tr>
<td>partialFit</td>
<td>3</td>
<td>444.41</td>
<td>462.36</td>
<td>0.6053</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
AICc versus AIC

\[
\begin{array}{c}
\text{pesticide} \\
\text{LNGamm} \\
\text{macroalgae} \\
\text{Inchla} \\
\text{grass}
\end{array}
\]

\[
> \text{aictab.lavaan(list(fullFit, partialFit), c("full", "partial"))}
\]

Model selection based on AICc:

<table>
<thead>
<tr>
<th>K</th>
<th>AICc</th>
<th>Delta AICc</th>
<th>AICcWt</th>
<th>Cum.Wt</th>
<th>LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>partial</td>
<td>12</td>
<td>442.67</td>
<td>0.00</td>
<td>0.69</td>
<td>-207.58</td>
</tr>
<tr>
<td>full</td>
<td>11</td>
<td>444.26</td>
<td>1.59</td>
<td>0.31</td>
<td>1.00 -209.86</td>
</tr>
</tbody>
</table>

AICc versus AIC

\[
\begin{array}{c}
\text{pesticide} \\
\text{LNGamm} \\
\text{macroalgae} \\
\text{Inchla} \\
\text{grass}
\end{array}
\]

\[
> \text{aictab.lavaan(list(fullFit, partialFit), c("full", "partial"), second.ord=F)}
\]

Model selection based on AIC:

<table>
<thead>
<tr>
<th>K</th>
<th>AIC</th>
<th>Delta AIC</th>
<th>AICWt</th>
<th>Cum.Wt</th>
<th>LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>partial</td>
<td>12</td>
<td>439.17</td>
<td>0.00</td>
<td>0.78</td>
<td>-207.58</td>
</tr>
<tr>
<td>full</td>
<td>11</td>
<td>441.71</td>
<td>2.54</td>
<td>0.22</td>
<td>1.00 -209.86</td>
</tr>
</tbody>
</table>

Summary

- Information criteria are suggestive
- Both state that the direct link is a better model, but hard to say...
- LR Test shows that partial mediation model is a better fit to the data
- Given that we may have a 2nd mediator (caprellids), this may lead to weaker performance of AICs

Old Wine in a New Bottle

1. ANOVA and ANCOVA in an SEM context
2. Multiple categorical predictors
3. Nonlinear effects
What about experiment with more than 2 levels of treatment?

1. Can you make the treatment continuous?
   - E.g. nutrient levels

2. Or, treat each level as being present/absent
   \[ y = \gamma_1 x_1 + \gamma_2 x_2 + \zeta \]
   where \( x_i = 0 \) or \( 1 \)

Experiment with 3 Levels

- Exogenous covariance no longer 0.

\[
\begin{align*}
X_1 & \quad X_2 & \quad X_3 \\
X_1 & 1.0 & -0.5 & -0.5 \\
X_2 & -0.5 & 1.0 & -0.5 \\
X_3 & -0.5 & -0.5 & 1.0 \\
\end{align*}
\]

Cannot Include All 3 Variables

- This matrix is singular

\[
\begin{align*}
X_1 & \quad X_2 & \quad X_3 \\
X_1 & 1.0 & -0.5 & -0.5 \\
X_2 & -0.5 & 1.0 & -0.5 \\
X_3 & -0.5 & -0.5 & 1.0 \\
\end{align*}
\]

- If you know \( x_1 \) and \( x_2 \), you know the state of \( x_3 \)

Coefficient judged relative to effect of missing variable
Does Diet Affect Urchin Gonad Development?

- Urchins feeding measured over 6 months
- Gonads and body size assessed at end
- All consumption rates converted to g dry carbon

Urchin Gonad Development Model

- Note that the polyculture is not included.
- Results judged relative to polyculture.

How do we use a categorical variable?

```
> urchinData <- read.csv("./urchin_ex_sem.csv")
> summary(urchinData)
```

```
Box     treatment
Min.   : 1   MAPY: 7
1st Qu.:10   POLY: 7
Median :18   R    : 7
Mean   :18   3rd Qu.:26
Max.   :35   Min. :1
```
# Make treatment into a series of binary variables
source("./makeBinaryTreatments.R")
binTrt <- makeBinaryTreatments(urchinData, "treatment")
head(binTrt)

> cor(binTrt)
MAPY POLY R
MAPY 1.0 -0.5 -0.5
POLY -0.5 1.0 -0.5
R -0.5 -0.5 1.0
>solve(cor(binTrt))
Error in solve.default(cor(binTrt)) : 
  Lapack routine dgesv: system is exactly singular

# add new columns to data frame
urchinData <- cbind(urchinData, binTrt)
urchinModel <-'
   Feeding.rate.dry ~ MAPY + R + GONAD_INDEX ~ MAPY + R + Feeding.rate.dry',
urchinSEM <- sem(urchinModel, data=urchinData)
#add new columns to data frame
urchinData<-cbind(urchinData, binTrt)

urchinModel<-
  Feeding_rate.dry ~ MAPY + R
  GONAD_INDEX ~ MAPY + R + Feeding_rate.dry

urchinSEM<-sem(urchinModel, data=urchinData)

## Fit

### Rhodymenia is not good food.
- Urchins eat more, but produce less gonad

### Performance is similar with *Macrocystis* or Mixture diet
1. But what if coefficient changes when the feeding rate -> gonad link was dropped?
2. Extra: try test growth

```r
urchinModel2 <-
  Feeding.rate.dry ~ MAPY + R
  GONAD_INDEX ~ MAPY + R

urchinSEM2 <- sem(urchinModel2, data=urchinData)

> urchinSEM2
lavaan (0.5-12) converged normally after 73 iterations

<table>
<thead>
<tr>
<th></th>
<th>Used</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>Estimator</td>
<td>ML</td>
<td></td>
</tr>
<tr>
<td>Minimum Function Test Statistic</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>P-value (Chi-square)</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>
```

```r
urchinSEM2 <- lavaan(urchinModel2, data=urchinData, auto.cov.y = FALSE, auto.var=TRUE)
```
Old Wine in a New Bottle

1. ANOVA and ANCOVA in an SEM context
2. Multiple categorical predictors
3. Nonlinear effects
Careful in Interpreting Results of Centering

- In uncentered model, additive paths estimate the effect of one variable in the absence of the other.
- In centered model, additive paths estimate the effect of one variable at the average level of the other.
- E.g., nutrients have an effect, but only when grazers are absent.

Does Diversity = Productivity of vice-versa?

Cardinale et al 2009

A Multi-Stream Experiment

Cardinale et al 2009

Nonlinear Relationship Between Nutrient Addition and Richness

Cardinale et al 2009
Nonlinear Nutrient Effect on Richness

Create a Nonlinear Variable

Note that Treatment's Don't Covary with Regional Richness

Model Fits Quite Well

```r
#read in the data
cards<-read.table("./cardainel_et_al_2009.csv")

#make a new nonlinear column
cards$logN2 <- cards$logN^2

cardModel<-
  SA ~ logN + logNcen2 + SR
  logChl ~ SA + logN
  SR ~ 0*logN + 0*logNcen2
  logN ~ logNcen2

  cardFit <- sem(cardModel, data=cards, fixed.x=T)

lavaan (0.4-12) converged normally after 64 iterations

Number of observations       127
Estimator                     ML
Minimum Function Chi-square   0.545
Degrees of freedom            4
P-value                       0.969
```

Cardinale et al. 2009
But...no Nutrient Effect?

Regressions:  
\[ \text{SA} \sim \log N \]
-0.97

Regressions:
|          | Estimate | Std.err | Z-value | P(>|z|) | Std.lv | Std.all |
|----------|----------|---------|---------|---------|--------|---------|
| SA       | 0.368    | 0.368   | 0.368   | 0.368   | 0.062  |
| logN     | 0.368    | 0.368   | 0.368   | 0.368   | 0.062  |
| logN2Cen | -0.475   | 0.240   | -1.974  | 0.048   | -0.475 | -0.147  |
| SR       | 0.384    | 0.384   | 10.859  | 0.000   | 0.384  | 0.688   |
| logChl   | 0.020    | 0.004   | 4.667   | 0.000   | 0.020  | 0.393   |

Refit with Centered Nutrients

Regressions:
|          | Estimate | Std.err | Z-value | P(>|z|) | Std.lv | Std.all |
|----------|----------|---------|---------|---------|--------|---------|
| SA       | 0.368    | 0.368   | 0.368   | 0.368   | 0.062  |
| logN     | 0.368    | 0.368   | 0.368   | 0.368   | 0.062  |
| logN2Cen | -0.475   | 0.240   | -1.974  | 0.048   | -0.475 | -0.147  |
| SR       | 0.384    | 0.384   | 10.859  | 0.000   | 0.384  | 0.688   |
| logChl   | 0.020    | 0.004   | 4.667   | 0.000   | 0.020  | 0.393   |

Refit with Centered Nutrients

Regressions:
|          | Estimate | Std.err | Z-value | P(>|z|) | Std.lv | Std.all |
|----------|----------|---------|---------|---------|--------|---------|
| SA       | 0.368    | 0.452   | 0.815   | 0.415   | 0.368  | 0.062  |
| logN     | 0.368    | 0.452   | 0.815   | 0.415   | 0.368  | 0.062  |
| logN2Cen | -0.475   | 0.240   | -1.974  | 0.048   | -0.475 | -0.147  |
| SR       | 0.384    | 0.035   | 10.859  | 0.000   | 0.384  | 0.688   |
| logChl   | 0.020    | 0.004   | 4.667   | 0.000   | 0.020  | 0.393   |

Refit with Centered Nutrients

Regressions:
|          | Estimate | Std.err | Z-value | P(>|z|) | Std.lv | Std.all |
|----------|----------|---------|---------|---------|--------|---------|
| SA       | 0.368    | 0.452   | 0.815   | 0.415   | 0.368  | 0.062  |
| logN     | 0.368    | 0.452   | 0.815   | 0.415   | 0.368  | 0.062  |
| logN2Cen | -0.475   | 0.240   | -1.974  | 0.048   | -0.475 | -0.147  |
| SR       | 0.384    | 0.035   | 10.859  | 0.000   | 0.384  | 0.688   |
| logChl   | 0.020    | 0.004   | 4.667   | 0.000   | 0.020  | 0.393   |

Questions?

Cardinale et al. 2009

#make a new nonlinear column

cards$logN2Cen <- (cards$logN-mean(cards$logN))^2