



Latent Variable Confirmatory Factor Analysis Practice Exercise

Jim Grace

U.S. Department of the Interior
U.S. Geological Survey

1

This module provides some practice with latent variable modeling. This module builds on the teaching module for confirmatory factor analysis.

Notes: IP-056512; Support provided by the USGS Climate & Land Use R&D and Ecosystems Programs. I would like to acknowledge formal review of this material by Jesse Miller and Phil Hahn, University of Wisconsin. Many helpful informal comments have contributed to the final version of this presentation. The use of trade names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Last revised 17.02.08.

Source: <https://www.usgs.gov/centers/wetland-and-aquatic-research-center/science/quantitative-analysis-using-structural-equation>

Example is drawn from
Jimenez-Alfaro (2016). Ecol. and Evol. 6:1515-1526.

Ecology and Evolution

Open Access

**Disentangling vegetation diversity from climate–energy
and habitat heterogeneity for explaining animal
geographic patterns**

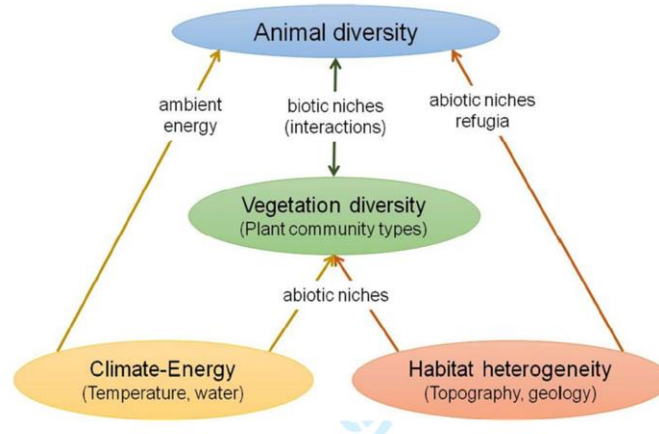
Borja Jiménez-Alfaro¹, Milan Chytrý¹, Ladislav Mucina^{2,3}, James B. Grace⁴ & Marcel Rejmánek⁵



2

Here is an example dealing with diversity patterns in animal groups.

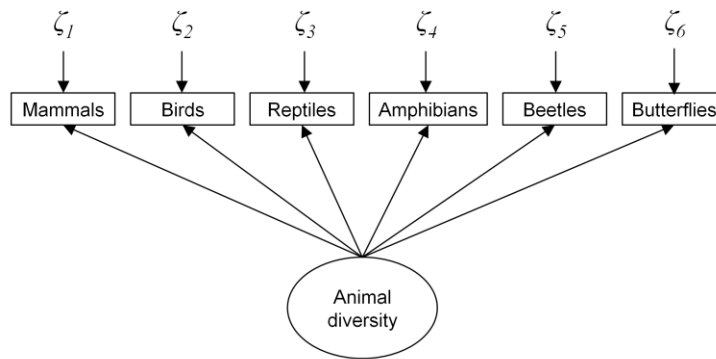
The Overall Hypothesis



3

The general hypothesis, as represented by this meta-model, is that climate effects are partly mediated through climate effects on vegetation. Natural variations in habitat heterogeneity are likely also important and must be controlled for in the analysis.

Our research question: Is there support for the idea of a general diversity response?



4

In this model, the measures of diversity for individual animal groups are hypothesized to be parallel reflections of a common response in Animal Diversity overall. If this hypothesis is true, it represents a striking demonstration of parallel evolution in the diversity of different groups.

The Exercise

To complete this exercise:

- (1) First, use the code on the next two pages to examine the pattern of correlations amongst the indicators. Do you think there is a chance that a general diversity effect will be at least partly supported?
- (2) Use lavaan and the data provided to fit the model shown.
- (3) Check model fit to first see if any important links appear to be missing.
- (4) Add additional links and reevaluate model if and as needed.
- (5) Request summary of model results and consider your recommendation for whether the data support a general response by diversity measures.



5

Here is a description of the exercise illustrated in this module.

Preliminary steps in R.

```
### CFA Exercise (adapted from Jimenez-Alfaro et al. 2016)

setwd("")
dat <- read.csv("AnimalDiversity_CFA_exercise_data.csv")

# load libraries and functions
library(lavaan)
library(AICcmodavg)
source("lavaan_modavg.R")

# Recode data and use dat2 for the CFA
dat2 <- with(dat, data.frame(Country))
dat2$Mammals <- dat$Mammals/100
dat2$Birds <- dat$Birds/100
dat2$Amphibians <- dat$Amphibians/100
dat2$Reptiles <- dat$Reptiles/100
dat2$Beetles <- dat$Beetles/1000
dat2$Butterflies <- dat$Butterflies/1000
summary(dat2)
```



6

lavaan.mod.avg.R can be obtained from
"http://jarrettburnes.info/ubc_sem/lavaan_materials/lavaan.modavg.R"
if need be.

Before running multi-indicator LV models, inspect patterns of correlations.

```
# Inspect indicator correlation patterns

indicators <- with(dat2, data.frame(Mammals, Birds,
Amphibians, Reptiles, Beetles, Butterflies))

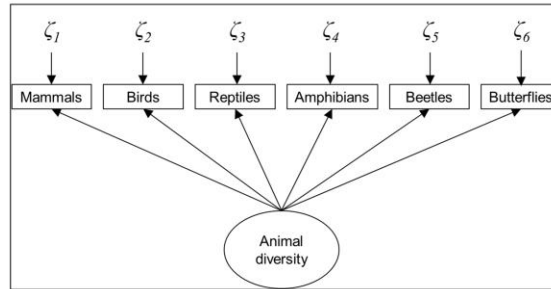
print(cor(indicators), digits=2)
```



7

There are logical expectations for patterns of correlations. The concept of convergent validity refers to the expectation that indicators for an LV should be well correlated with each other.

Go ahead and evaluate this hypothesis with the available data.



[when you have finished with your work, go to the next slides to compare with those anticipated for this exercise]



8

Perfectly legal to go ahead and look ahead to the answers I came up with if you don't have time or inclination to work out your own answer.

Initial CFA

```
### CFA Evaluation of "Interrelated Diversity" Hypothesis
# Specify model
mod1 <- 'Diversity =~ Mammals +Birds +Amphibians +Reptiles
+Beetles +Butterflies'

# Fit model
mod1.fit <- cfa(mod1, data=dat2)

# Examine overall fit
print(mod1.fit)
```

```
> print(mod1.fit)
lavaan (0.5-20) converged normally after 60 iterations

    Number of observations              20

    Estimator                          ML
    Minimum Function Test Statistic    20.817
    Degrees of freedom                  9
    P-value (Chi-square)                0.013
```



Large chi-square for such a small sample size.

9

Fit of this simplest model looks like it may be missing something.

Check Modification Indices for Missing Linkages

```
## Modification Indices
# error correlations only
subset(modindices(mod1.fit), mi > 3.8 & op == "~")

# between-indicator effects
subset(modindices(mod1.fit), mi > 3.8 & op == "~")

> # error correlations only
> subset(modindices(mod1.fit), mi > 3.8 & op == "~")
      lhs op      rhs      mi      epc
19 Birds ~~ Amphibians 3.990 -0.003
21 Birds ~~      Beetles 4.442  0.012

> # between-indicator effects
> subset(modindices(mod1.fit), mi > 3.8 & op == "~")
[1] lhs      op      rhs      mi      epc
<0 rows> (or 0-length row.names)
```



Do either error correlations make ecological sense?

10

Consulting modification indices is usually the way we look for hints at what is missing when model discrepancy is noticeable.

New Model - Add Error Correlation: Birds \sim Beetles

```
# Specify new model
mod2 <- 'Diversity =~ Mammals +Birds +Amphibians
        +Reptiles +Beetles +Butterflies
        Birds ~~ Beetles'
```

```
> print(mod2.fit)
Number of observations                20

Estimator                           ML
Minimum Function Test Statistic      16.013
Degrees of freedom                    8
P-value (Chi-square)                 0.042
```

Still room for improvement in fit.



11

We dropped the test statistic from 20.8 to 16.0, a drop of 4.8, which is greater than the single-degree-of-freedom criterion of 3.84. Still, indications are fit might be improved.

Compare models

```
# Compare models  
anova(mod1.fit, mod2.fit)
```

```
> anova(mod1.fit, mod2.fit)  
Chi Square Difference Test
```

	Df	AIC	BIC	Chisq	Chisqdiff	Dfdiff	
Pr(>Chisq)							
mod2.fit	8	-138.76	-125.82	16.013			
mod1.fit	9	-135.96	-124.01	20.817	4.804	1	0.02839 *

Chi-square test indicates mod2 is significant improvement over mod1.
both AIC and BIC favor agree.



12

Here is an easy way to do the formal test for adding a link by using the anova command.

Compare models (cont.)

```
### Compare all mods using AICc criterion
aictab.lavaan(list(mod1.fit, mod2.fit),
              c("Model1", "Model2"))
```

Model selection based on AICc:

	K	AICc	Delta_AICc	AICcWt	Cum.Wt	LL
Model1	12	-129.50	0.00	0.51	0.51	79.98
Model2	13	-129.43	0.07	0.49	1.00	82.38*

AICc says we lack the power to tell the models apart. So, we may reject Model2 unless we have a compelling reason to include the extra relationship in our model.



13

Generally, we get a more nuanced comparison using AICc comparison.

New Model - Add Error Correlation: Birds \sim Amphibians

```
# Specify new model
mod3 <- 'Diversity =~ Mammals +Birds +Amphibians
          +Reptiles +Beetles +Butterflies
          Birds ~~ Amphibians'
```

lavaan (0.5-20) converged normally after 59 iterations

Number of observations	20
Estimator	ML
Minimum Function Test Statistic	12.923
Degrees of freedom	8
P-value (Chi-square)	0.115

Now model fit is tolerable.



14

Compare models

```
# Compare models  
anova(mod1.fit, mod3.fit)
```

```
> anova(mod1.fit, mod3.fit)  
Chi Square Difference Test
```

	Df	AIC	BIC	Chisq	Chisqdiff	Dfdiff	Pr
mod3.fit	8	-141.85	-128.91	12.924			
mod1.fit	9	-135.96	-124.01	20.817	7.8934	1	0.0049**

Chi-square says adding error correlation an improvement, and AIC and BIC also pick mod3 over mod1 (lower values, i.e., more negative).



Compare models (cont.)

```
### Compare all mods using AICc criterion  
aictab.lavaan(list(mod1.fit, mod3.fit),  
              c("Model1", "Model3"))
```

Model selection based on AICc:

	K	AICc	Delta_AICc	AICcWt	Cum.Wt	LL
Model3	13	-132.52	0.00	0.82	0.82	83.93
Model1	12	-129.50	3.02	0.18	1.00	79.98

AICc also indicates model 3 is an improvement over model 1.



Result details

Latent Variables:

	Estimate	Std.Err	Z-value	P(> z)	Std.lv	Std.all
Diversity =~						
Mammals	1.000				0.088	0.770
Birds	1.756	0.590	2.978	0.003	0.154	0.635
Amphibians	0.694	0.162	4.281	0.000	0.061	0.859
Reptiles	1.034	0.370	2.793	0.005	0.091	0.600
Beetles	2.469	0.564	4.380	0.000	0.216	0.874
Butterflies	4.672	0.942	4.960	0.000	0.409	0.971

Covariances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.Birds ~~						
.Amphibians	-0.004	0.002	-2.615	0.009	-0.004	-0.789

Pretty good fit for such a simple cfa and small sample.

Note that Jimenez-Alfaro et al. used Bayesian estimates to confirm model selection and parameter estimates.



17

The rationale for Jimenez-Alfaro et al. confirming their estimates using Bayesian MCMC methods is simply because likelihood estimates are based on large-sample theory while Bayesian estimates are not.

Result details

Variances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.Mammals	0.006	0.002	3.048	0.002	0.006	0.475
.Birds	0.027	0.009	2.877	0.004	0.027	0.461
.Amphibians	0.001	0.000	2.451	0.014	0.001	0.232
.Reptiles	0.013	0.004	3.087	0.002	0.013	0.575
.Beetles	0.010	0.004	2.506	0.012	0.010	0.161
.Butterflies	0.029	0.012	2.526	0.012	0.029	0.165
Diversity	0.007	0.004	1.866	0.062	1.000	1.000

R-Square:

	Estimate
Mammals	0.593
Birds	0.404
Amphibians	0.737
Reptiles	0.360
Beetles	0.764
Butterflies	0.943

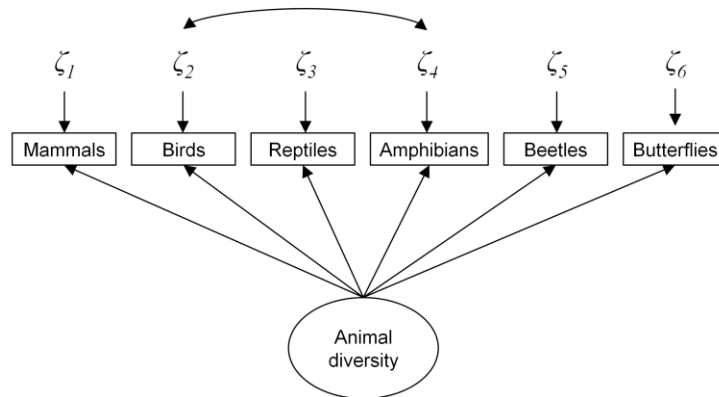
- None of the variance estimates are negative, which is the way we want it. Reptiles are least well predicted by the model.



18

We often don't pay much attention to the variances, but for latent variable models, we do wish to avoid models with highly negative variances since that indicates some sort of mis-fit.

Final Model Selected



We conclude that there is a general diversity response (diversity of all groups are rising and falling in concert. However, birds and amphibians tend to also show some separation in responses (our interpretation since the error correlation is negative).



19

All things considered, this is a fairly remarkable result. Note, however, that once this submodel is included in the total model, which includes a number of other variables, we could reach a different conclusion.

Want to take the example further?

For an additional exercise:

- (1) Use your own data to construct a cfa for evaluation.

