



Multi-Group Modeling

Jim Grace

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

1

In this module I illustrate a long-standing major option in SEM, what is called "multi-group" modeling.

An appropriate citation for this material is

Grace, J.B. 2003. Comparing groups using structural equations. Chapter 11, pp 281-296. In: Pugesek, B.H., Tomer, A., and von Eye. A. (eds.). *Structural Equation Modeling*, Cambridge University Press

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 17.02.15.

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

1. Multigroup modeling involves situations where there are discrete groups in the data that you want to compare.

- males versus females in a population
- treated versus control plots
- areas with different disturbance histories

In multigroup modeling, we develop a common model for different groups and then ask what parameters are the same or different between groups.

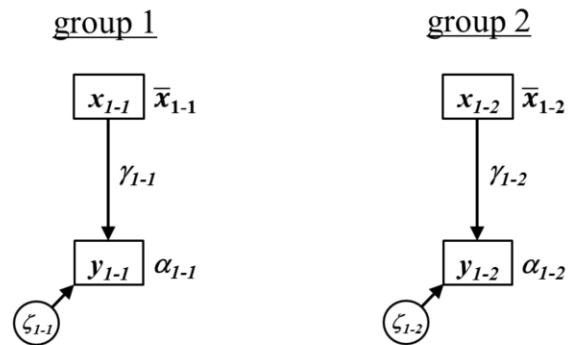
We can use the classical likelihood-based measures (e.g., model chi-square) to test for constraints across groups.



2

The ability to formally compare groups is very important in certain situations.

2. There are several parameters that can be compared.



A number of hypotheses we can test:

- (1) equal slopes: $\gamma_{I-1} = \gamma_{I-2}$
- (2) equal intercepts: $\alpha_{I-1} = \alpha_{I-2}$
- (3) equal means: $\bar{x}_{1-1} = \bar{x}_{1-2}$
- (4) equal errors: $\zeta_{I-1} = \zeta_{I-2}$

increasingly strong
constraints

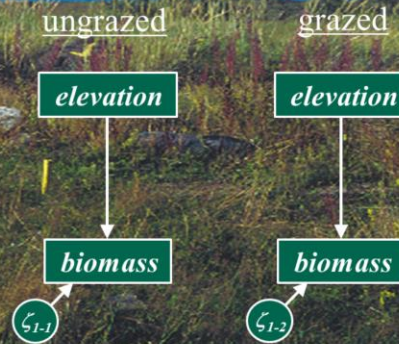


We are proposing an overall model that applies to both groups and then testing to see if the raw parameter estimates are essentially the same across groups (meaning a process is common to both groups).

Effects of Grazing on Finnish Coastal Meadows*

How does biomass respond to the interaction between grazing & elevation?

Data from 1-m² plots arrayed along an elevation gradient in each of several paired grazed and ungrazed meadows in SW Finland.



Grace, J.B. and Jutila, H. (1999) The relationship between species density and community biomass in grazed and ungrazed coastal meadows. *Oikos*, 85:398-408.

Here is an example.

1. Llavaan uses a ‘group=’ command to invoke a multigroup analysis.

```
#lavaan code for basic model
mod1 <- 'biomass ~ elev'

#fit the model, specifying groups
mod1.fit <- sem(mod1, data=sem.dat, group="grazed")

#request output
summary(mod1.fit)
```



5

The lavaan software provides some special commands to ease the process of testing constraints.

We use the “group=“ command to invoke a multi-group modeling setup. Note that here the grouping variable “grazed” is a 0/1 dummy variable with 1==grazed.

2. Default allows all parameters to differ between groups.

```
> summary(mod1.fit)
lavaan (0.5-12) converged normally after 24 iterations

Number of observations per group
1                                     165
0                                     189

Estimator                           ML
Minimum Function Test Statistic      0.000
Degrees of freedom                    0
P-value (Chi-square)                 0.000

Chi-square for each group:

1                                     0.000
0                                     0.000
```

There are no equality constraints and therefore no chi-square tests. 6

Default is to permit all parameters to be unique across groups. This is very handy. Models contain many parameters and if we set all parameters equal across groups initially, diagnosing misspecifications could be very difficult.

lavaan reports an overall Test Statistic for the two models (i.e., the multi-group model). Since all parameters are freely estimated for each group, there are no constraints and therefore no degrees of freedom.

Notice the number of observations for the groups is given.

3. Default allows all parameters to differ between groups.

Group 1 [1]: (notice Group=1 = grazed)				
	Estimate	Std.err	Z-value	P(> z)
Regressions:				
biomass ~				
elev	-0.474	0.205	-2.311	0.021
Intercepts:				
biomass	5.263	0.117	45.072	0.000
Variances:				
biomass	0.534	0.059		
Group 2 [0]:				
	Estimate	Std.err	Z-value	P(> z)
Regressions:				
biomass ~				
elev	-0.798	0.145	-5.523	0.000
Intercepts:				
biomass	5.926	0.065	91.069	0.000
Variances:				
biomass	0.296	0.030	0.000	7

Group results are presented separately.

4. We can test equality constraints by labeling parameters.

```
#lavaan code naming the path coefficient "b1"  
mod2 <- 'biomass ~ c("b1","b1")*elev'  
  
#fit the model, specifying groups  
mod2.fit <- sem(mod2, data=sem.dat, group="grazed")
```

When you label a parameter across groups, you have to pass to lavaan a vector of labels, one for each group. Here, `c("b1","b1")` is a vector of labels.



8

So, our initial analysis allows all parameters to be different between groups. We then might like to add constraints sequentially to determine what is the same across groups. There are some general commands in lavaan for this task, but let's start with a simple, general approach – setting a single parameter equal across groups.

5. With constraints imposed, we can test if models sig. different.

```
> summary(mod2.fit)
lavaan (0.5-12) converged normally after 19 iterations

Estimator              ML
Minimum Function Test Statistic    1.668
Degrees of freedom                1
P-value (Chi-square)              0.197

Chi-square for each group:

1              1.116
0              0.552
```

The overall model chi-square of 1.668 with 1 df would traditionally be interpreted as a non-significant difference between models.



9

Equality constraints reduce the number of parameters being estimated and provide model degrees of freedom for hypothesis tests.

6. With constraints imposed, we can test if parameters different.

Group 1 [1]:

	Estimate	Std.err	Z-value	P(> z)
Regressions:				
biomass ~				
elev (b1)	-0.691	0.118	-5.836	0.000
Intercepts:				
biomass	5.371	0.082	65.493	0.000
Variances:				
biomass	0.538	0.059		

We get one best estimate for both groups.

Group 2 [0]:

	Estimate	Std.err	Z-value	P(> z)
Regressions:				
biomass ~				
elev (b1)	-0.691	0.118	-5.836	0.000
Intercepts:				
biomass	5.888	0.058	101.552	0.000
Variances:				
biomass	0.297	0.031		

10

Here we see what is going on.

7. We can also use an 'equal' or a 'group.equal' command.

```
mod2a <- 'biom.log ~ equal("b1")*elev.m'  
mod2a.fit <- sem(mod2a, data=sem.dat, group="grazed")
```

Produces exactly the same results as the previous command.

```
mod3 <- 'biom.log ~ elev.m'  
mod3.fit <- sem(mod3, data=sem.dat, group="grazed",  
group.equal="regressions")
```



11

Lavann has other options.

8. There are a number of helpful options for the 'group.equal' command.

```
group.equal=c(  
  "intercepts",  
  "means",  
  "regressions",  
  "residuals",  
  "residual.covariances")
```

This gives you a taste of the possibilities for testing equality constraints across groups.

For more, consult the tutorial

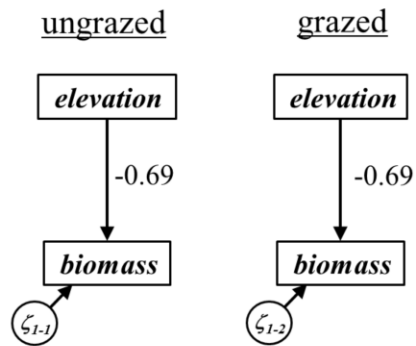
<http://lavaan.ugent.be/tutorial/tutorial.pdf>



12

Lavaan makes this as automated as possible.

9. In this case, we can conclude that the slope of the relationship between biomass and elevation is not significantly different between grazed and ungrazed meadows.



13

Here we are looking at the raw parameter estimates, which are the ones being tested.

A more complete view of a multi-group analysis.

This figure, from the example below, shows the parameter estimates that were significantly different among groups (those with asterisks in the lower panel). Note that these are standardized parameters, which is why the "non-different" parameters between models are not identical.

Grace, J.B. and Jutila, H. (1999) The relationship between species density and community biomass in grazed and ungrazed coastal meadows. *Oikos*, 85:398-408.

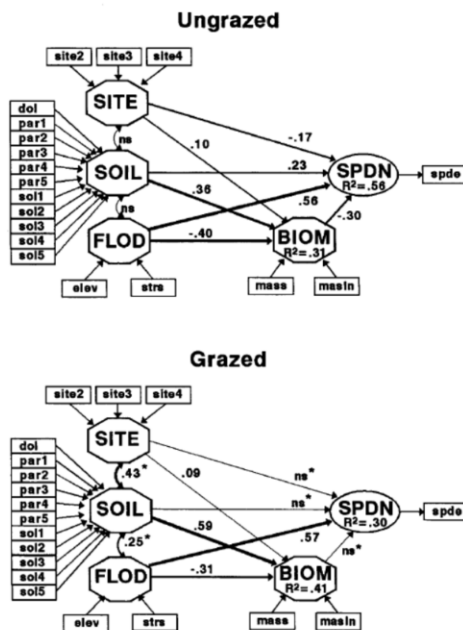


Fig. 4. Summary path models for grazed and ungrazed sites. Notation is as shown in the caption for Fig. 3. Paths indicated by an asterisk for the grazed group are statistically different between groups based on the multigroup analysis.

Refer to the original reference for more details on what the mult-group analysis revealed.