



The Test of Mediation: Practice Exercise

Jim Grace

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

1

This module contains a practice exercise on the test of mediation. This module builds on the teaching modules for Model Evaluation and The Test of Mediation.

An appropriate general citation for this material is

Grace, J.B. (2006) *Structural Equation Modeling and Natural Systems*. 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. Questions about this material can be sent to sem@usgs.gov.

Last revised 15.04.08.

**The Example:
Field-based Experimental Study of the Importance of
Small Herbivores in a Seagrass Ecosystem:**

Matthew A Whalen and J Emmett Duffy

Whalen, Duffy, and Grace, 2013. *Ecology* 94:510-520.



This exercise relates to a study conducted in seagrass beds in Virginia coastal waters. The full citation for the source work is

Whalen, M.A., Duffy, J.E. and Grace, J.B. 2013. Temporal shifts in top-down versus bottom-up control of epiphytic algae in a seagrass ecosystem. *Ecology* 94:510-520.

(<http://www.esajournals.org/doi/abs/10.1890/12-0156.1>)

More information about the umbrella program, the Zostera Experimental Network, ZEN, can be found at <http://zenscience.org/>.

Experimental Design:

Treatments:

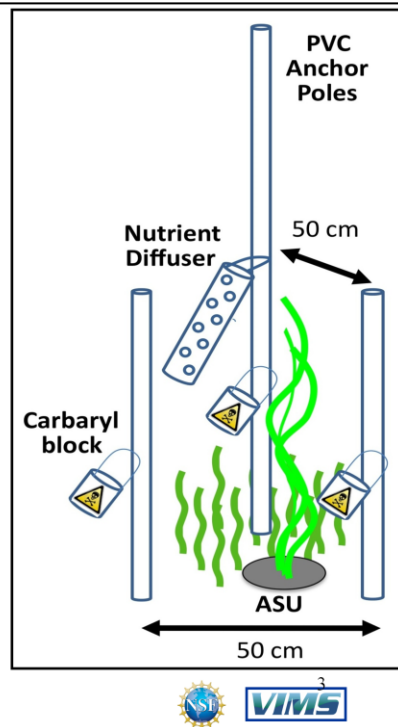
- pesticide (Carbaryl)
- nutrient addition
- combination
- controls

8 reps @ 5 trts = 40 plots

Pesticide/Deterrent effects:

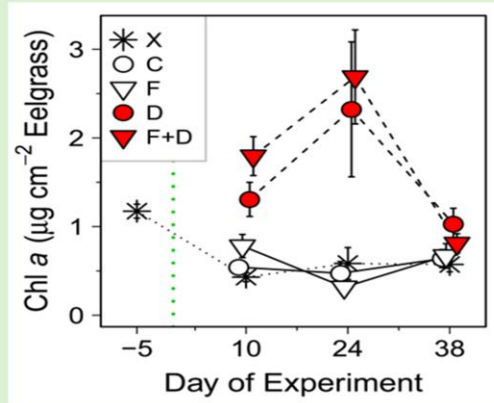
Crustaceans: reduced 58-96%
Algal biomass: incr. 130-748%

Nutrients: no significant effects



Here is the part of the experimental study I use for this example. The study used slow-release methods to expose plots of seagrass beds to the pesticide Carbaryl, to kill microcrustaceans, and nutrients, to stimulate the production of epiphytes. Since the fertilizer treatment had no discernable effect for the time period used in this analysis, I will only focus on the pesticide (aka deterrent) effect in this example.

ANOVA results:
Means for epiphytes increased by deterrent.



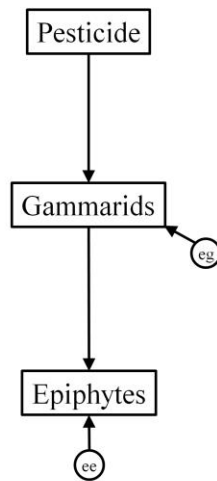
USGS

4

Anova results showed a net effect of deterrent/pesticide (in red) on epiphyte biomass (Chl a), but no effect of fertilization.

The test of mediation

Does reduction of Gammarids explain promotion of epiphytes by pesticide?



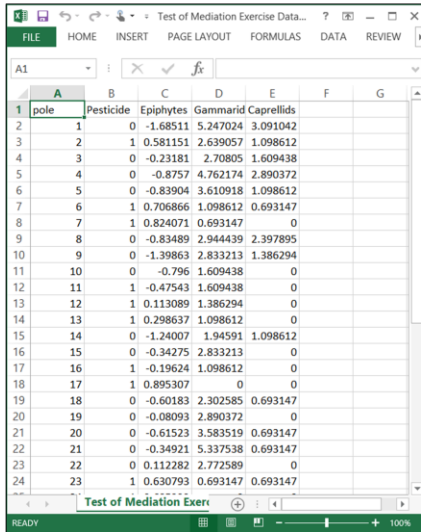
note that Gammarids are class of crustations whose abundance is most strongly reduced by the pesticide.



5

In this exercise, your task will be to test the mediation model shown here. This model hypothesizes that negative effects of pesticide on the most abundance class of microcrustaceans, the Gammarids, is sufficient to explain the observed increase in Epiphytes.

The Data



	A	B	C	D	E	F	G
1	pole	Pesticide	Epiphytes	Gammarid	Caprellids		
2	1	0	-1.68511	5.247024	3.091042		
3	2	1	0.581151	2.639057	1.098612		
4	3	0	-0.23181	2.70805	1.609438		
5	4	0	-0.8757	4.762174	2.890372		
6	5	0	-0.83904	3.610918	1.098612		
7	6	1	0.706866	1.098612	0.693147		
8	7	1	0.824071	0.693147	0		
9	8	0	-0.83489	2.944439	2.397895		
10	9	0	-1.39863	2.833213	1.386294		
11	10	0	-0.796	1.609438	0		
12	11	1	-0.47543	1.609438	0		
13	12	1	0.113089	1.386294	0		
14	13	1	0.298637	1.098612	0		
15	14	0	-1.24007	1.94591	1.098612		
16	15	0	-0.34275	2.833213	0		
17	16	1	-0.19624	1.098612	0		
18	17	1	0.895307	0	0		
19	18	0	-0.60183	2.302585	0.693147		
20	19	0	-0.08093	2.890372	0		
21	20	0	-0.61523	3.583519	0.693147		
22	21	0	-0.34921	5.337538	0.693147		
23	22	0	0.112282	2.772589	0		
24	23	1	0.630793	0.693147	0.693147		

The data to be used in this exercise can be found in the file “Test of Mediation Exercise.csv”.

There are a total of 40 rows of data in the file.

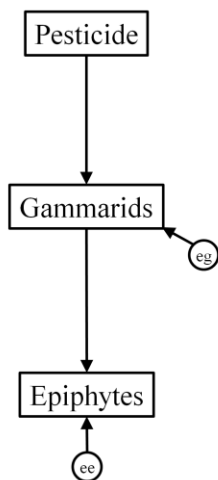
The survey design is a simple random sample.

We are only using the week #2 data.



This image shows the data associated with this exercise.

The Model



This model asks whether pesticide effects on just the Gammarid microcrustaceae are sufficient to explain treatment effects on epiphytes.

There are several obvious alternative models. This one is given as a starting point for your exercise.



Again, here is the initial hypothesis, which is an example of “full mediation” (i.e., Gammarids fully mediate the effects of Pesticides on Epiphytes).

The Exercise

To complete this exercise:

- (1) Use this data to estimate the model on the next page using lavaan.
- (2) Check model fit and, if necessary, compare to alternative models.
- (3) Once you have selected a model, examine the output to draw interpretations.



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

8

[Initiate theme music from the TV series Mission Impossible]

Your task is to evaluate this initial hypothesis, essentially performing “The Test of Mediation”.

If you are adverse to statistical hypothesis testing, which is a popular aversion in some quarters these days, you can perhaps be comforted by the fact that this is the testing of a scientific hypothesis, not a statistical null hypothesis.

If you would like a refresher on this material
and don't have my workshop notes handy,
just go to

<http://www.nwrc.usgs.gov/SEM/index.html>

and look at

SEM.3- Model Evaluation
and
SEM.5- The Test of Mediation

On the next few slides I show one approach to the task
(and certainly it is not the only way to approach the
problem).



9

The slide says it all that needs saying here.

Preliminary steps in R.

```
### Test of Mediation Exercise
### 3-variable submodel from Whalen et al.

setwd("./SEM.5-The Test of Mediation")

dat <- read.csv("Test of Mediation Exercise
Data.csv")

attach(dat)

### Examine Group Difference

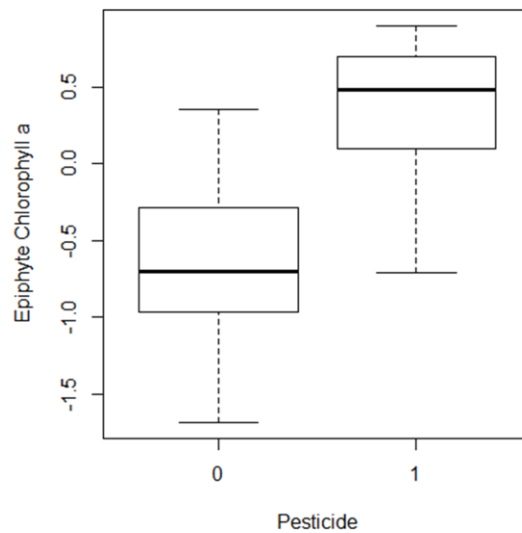
boxplot(Epiphytes ~ Pesticide, xlab="Pesticide",
ylab="Epiphyte Chlorophyll a")
```



10

OK, here is the code I developed to read in the data and make a cursory examination of the mean treatment differences (shown on next slide).

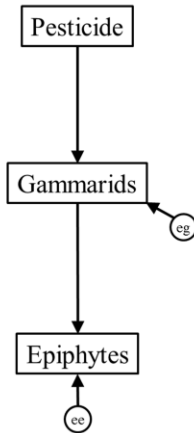
Boxplots summarize net difference between groups.



11

Box plots are my preferred method over bar plots, as they reveal more of the features of the data. Big increase in epiphytes for plots exposed to pesticide.

Lavaan code for Full-Mediation model.



```
### lavaan modeling
library(lavaan)

### Full mediation model

# Specify model
full.med <- 'Gammarids ~ Pesticide
            Epiphytes ~ Gammarids'

# Fit model
full.med.fit <- sem(full.med, dat)
```



12

Here is the code for this first “full-mediation” model.

Lavaan results.

```
# Examine model fit
# Note: executing the object name yields basic
# fit statistics below.
show(full.med.fit)
```

```
lavaan (0.5-17) converged normally after 18
iterations

    Number of observations              40

    Estimator                          ML
    Minimum Function Test Statistic    6.863
    Degrees of freedom                  1
    P-value (Chi-square)               0.009
```



13

Often we use

```
summary(fit.object)
```

which gives us a lot of output.

Here I illustrate the `show()` extractor function, which gives us just the top part of the full summary, which is all we want to look at at this point.

I would interpret these results as an indication that the model is not complete.

Next, examine goodness of fit statistics and diagnostics.

```
# Examine diagnostics - modification indices  
subset(modindices(full.med.fit), mi > 3.8)
```

	lhs	op	rhs	mi	epc	sepc.lv	sepc.all	sepc.nox
1	Gammarids	~~	Epiphytes	6.307	0.221	0.221	0.234	0.234
2	Epiphytes	~~	Pesticide	6.307	0.153	0.153	0.457	0.153
3	Gammarids	~	Epiphytes	6.307	0.901	0.901	0.444	0.444
4	Epiphytes	~	Pesticide	6.307	0.637	0.637	0.457	0.933

Remember, you focus on the “mi” column for large values.



14

Since our model does not fit the data well, I request to see the modification indices.

By using the “subset” function, we can request only those modification indices with “mi” values greater than 3.8, the single-degree-of-freedom chi-square criterion.

These results show residual correlation “~~” between Gammarids and Epiphytes (which makes no sense as a suggestion for model modification since these are directly linked in the model). There is also a residual relation between Pesticide and Epiphytes, suggesting you examine a model in which those are directly linked, as shown on the next page.

We can also look directly at the covariance residuals.

```
# Examine diagnostics - residuals  
residuals(full.med.fit, type="standardized")
```

```
      Gmmrds Epiphyt Pestcd  
Gammarids 0.000  
Epiphytes 0.000 0.000  
Pesticide 0.000 1.739 0.000  
  
$mean  
Gammarids Epiphytes Pesticide  
      0      0      0
```

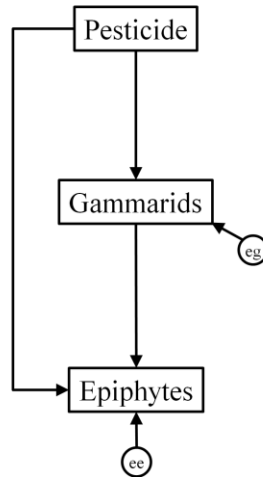
There is no simple guideline for an absolute value of residual that would be deemed important. What we see here is that there is a residual relationship between Pesticide and Epiphytes.



15

Residual covariances are often useful and I often consult them when looking for information. I generally request standardized residuals so I can compare paths on an equal basis.

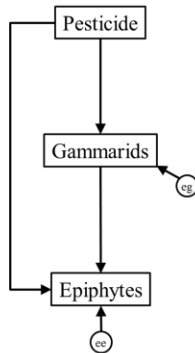
Alternative model suggested by the diagnostics. This is a partial-mediation model, as the effect of pesticide is only partly mediated by Gammarids.



16

So, this shows the alternative model suggested by the modindices, presuming we feel the residual relationship between Pesticide and Epiphytes is best expressed as a directed one.

Lavaan code for Partial-Mediation model.



```
### Partial mediation model

# Specify model
part.med <- 'Gammarids ~ Pesticide
            Epiphytes ~ Gammarids
            + Pesticide'

# Fit model
part.med.fit <- sem(part.med, dat)

# Examine model fit and parameters
summary(part.med.fit)
```

Note here we recognize that the model is saturated; thus, there will be no overall model discrepancy. For this reason, we ask for the full summary so we can get some idea of how strong the support is for the individual links.



17

Now we have a saturated model, so I want to use the `summary()` function to see the parameter statistics.

Lavaan results for partial-mediation model.

```
lavaan (0.5-17) converged normally after 22 iterations

Number of observations                    40

Estimator                               ML
Minimum Function Test Statistic          0.000
Degrees of freedom                       0

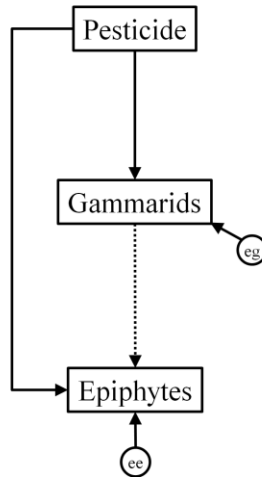
                                Estimate  Std.err  Z-value  P(>|z|)
Regressions:
  Gammarids ~
    Pesticide          -2.196    0.282   -7.792    0.000
  Epiphytes ~
    Gammarids          -0.164    0.082   -1.992    0.046
    Pesticide           0.637    0.233    2.736    0.006
Variances:
  Gammarids            0.762    0.170
  Epiphytes            0.207    0.046
```



18

What we see here is that the p-value for the link from Gammarids to Epiphytes is in that grey zone near 0.05. While we don't use parameter p-values as more than a guideline, the result suggest we should consider yet another alternative model, as shown on the next slide.

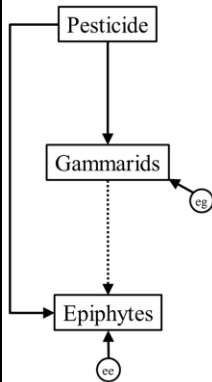
Yet another alternative model suggested by the diagnostics. This is a no-mediation model, as the effect of pesticide does not propagate through Gammarids to epiphytes.



19

Results suggest an alternative “no mediation” model in which the path from Gammarids to Epiphytes is too weak to include. Having established this alternative model, we can use model comparison to evaluate the possibilities.

Lavaan code for No-Mediation model.



```
### No mediation model

# Specify model
no.med <- 'Gammarids ~ Pesticide
          Epiphytes ~ 0*Gammarids
          + Pesticide'

# Fit model
no.med.fit <- sem(no.med, dat)
```

By pre-multiplying Gammarids by 0, we hypothesize no effect.



20

And here is the lavaan code for the no-mediation model.

Lavaan results for no-mediation model.

```
# Examine model fit  
show(no.med.fit)
```

```
lavaan (0.5-17) converged normally after 17  
iterations  
  
Number of observations              40  
  
Estimator                          ML  
Minimum Function Test Statistic    3.782  
Degrees of freedom                  1  
P-value (Chi-square)               0.052
```

So, in terms of GOF based on a chi-square test, our model might be considered to be sufficient. This alone does not provide us with a comprehensive approach to model selection.



21

No-mediation also sufficient. This leads us to the question of which model is best.

We can use AICc to compare the the models.

```
### Model Comparison - AICc
library(AICcmodavg)
source("lavaan.modavg.R")
aictab.lavaan(list(full.med.fit, part.med.fit,
no.med.fit), c("Full", "Partial", "None"))
```

Model selection based on AICc :

	K	AICc	Delta_AICc	AICcWt	Cum.Wt	LL
Partial	5	220.25	0.00	0.63	0.63	-104.79
None	4	221.69	1.44	0.31	0.93	-106.68
Full	4	224.77	4.52	0.07	1.00	-108.22

AICc weights suggest partial mediation is best model, though the Delta_AICc is less than 2 from the second-best model.



22

The tutorial module “Model Evaluation” shows you the source for the “lavaan.modavg.R” function developed by Jarrett Byrnes, which is also given in the notes section of this slide*.

Again, we focus on the Delta_AICc and then AICcWt columns. This comparison favors the partial mediation model over the no-mediation or full-mediation models.

*Jarrett Byrnes from Univ. Mass at Boston has developed a function for computing a AICc table for lavaan models. It can be obtained from his website at:

http://jarrettbyrnes.info/ubc_sem/lavaan_materials/lavaan.modavg.R

Want to take the example further?

For an additional exercise:

- (1) Use the other variable in the data set, “Caprellids” and test whether it mediates the remaining effect of Pesticide on Epiphytes.
- (2) Check model fit and, if necessary, compare to alternative models.
- (3) Once you have selected a model, examine the output to draw interpretations.
- (4) Compute indirect and total effects for the final model.



23

If you would like to go further, you can bring in the other variable in the dataset, Caprellids” to see if that is the second mediator.

A more complete treatment of this example can be found in the tutorial module “SEM versus ANOVA and ANCOVA (SEM.8)” located at

<http://www.nwrc.usgs.gov/SEM/index.html>

Additional information about structural equation
modeling can be found at
www.nwrc.usgs.gov/SEM/index.html

