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

**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.
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. *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/.
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.

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

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.
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.

[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


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).

The slide says it all that needs saying here.
Preliminary steps in R.

```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")
```

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

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

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

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

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

<table>
<thead>
<tr>
<th>lhs</th>
<th>op</th>
<th>rhs</th>
<th>mi</th>
<th>epc</th>
<th>sepc.lv</th>
<th>sepc.all</th>
<th>sepc.nox</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gammarids</td>
<td>~</td>
<td>Epiphytes</td>
<td>6.307</td>
<td>0.221</td>
<td>0.221</td>
<td>0.234</td>
<td>0.234</td>
</tr>
<tr>
<td>Epiphytes</td>
<td>~</td>
<td>Pesticide</td>
<td>6.307</td>
<td>0.153</td>
<td>0.153</td>
<td>0.457</td>
<td>0.153</td>
</tr>
<tr>
<td>Gammarids</td>
<td>~</td>
<td>Epiphytes</td>
<td>6.307</td>
<td>0.901</td>
<td>0.901</td>
<td>0.444</td>
<td>0.444</td>
</tr>
<tr>
<td>Epiphytes</td>
<td>~</td>
<td>Pesticide</td>
<td>6.307</td>
<td>0.637</td>
<td>0.637</td>
<td>0.457</td>
<td>0.933</td>
</tr>
</tbody>
</table>

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

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")
```

<table>
<thead>
<tr>
<th>Gmrmrs</th>
<th>Epphyt</th>
<th>Pestcd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gammarids</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Epiphytes</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Pesticide</td>
<td>0.000</td>
<td>1.739</td>
</tr>
</tbody>
</table>

$\text{mean}$

<table>
<thead>
<tr>
<th>Gmrmrs</th>
<th>Epiphytes</th>
<th>Pesticide</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

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.

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

```r
### 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.
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.
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.

```r
### 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.

And here is the lavaan code for the no-mediation model.
No-mediation also sufficient. This leads us to the question of which model is best.
We can use AICc to compare the models.

```r
## 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"))
```

<table>
<thead>
<tr>
<th></th>
<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>5</td>
<td>220.25</td>
<td>0.00</td>
<td>0.63</td>
<td>0.63 -104.79</td>
<td>-</td>
</tr>
<tr>
<td>None</td>
<td>4</td>
<td>221.69</td>
<td>1.44</td>
<td>0.31</td>
<td>0.93 -106.68</td>
<td>-</td>
</tr>
<tr>
<td>Full</td>
<td>4</td>
<td>224.77</td>
<td>4.52</td>
<td>0.07</td>
<td>1.00 -108.22</td>
<td>-</td>
</tr>
</tbody>
</table>

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

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:
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.

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