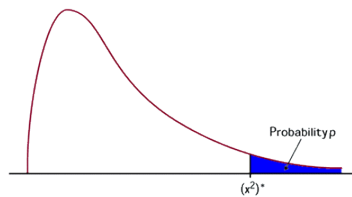


Assessing Fit & Comparing SEMs fit with Likelihood

Jarrett E. K. Byrnes



Outline

1. Assessing model fit: the χ^2
 - Related indices
2. Adjusting for non-normality
3. Model comparison
4. Testing mediation

Evaluating the Discrepancy between Sample Covariances and Model- Implied Covariances

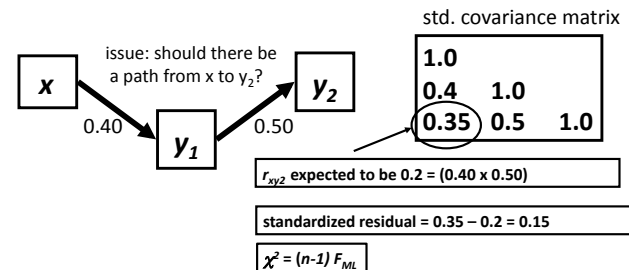
The log likelihood ratio, F_{ML} , and its sample generalization \hat{C} , follow a (chi-square) distribution. It is often used to calculate a model chi-square (χ^2):

$$\chi^2 = \hat{C} = (n-1)F_{ML}$$

Here, n refers to the sample size, thus χ^2 is a direct function of sample size.

The individual discrepancies between observed and implied covariances can also be examined as residuals and through derived "modification indices".

Evaluating Fit of A Model



$\chi^2 = 1.78$ with 50 samples

$\chi^2 = 3.60$ with 100 samples

$\chi^2 = 7.24$ with 200 samples

The Classic Test using P Values

$\chi^2 = 1.78$ with 50 samples
 $p = 0.182$

$\chi^2 = 3.60$ with 100 samples
 $p = 0.058$

$\chi^2 = 7.24$ with 200 samples
 $p = 0.007$

Our ability to detect significant differences from our base model, depends as usual on sample size.

Note that in asking does the modeled covariance matrix match the observed matrix, $p > 0.05$ means no discrepancy

Single-degree-of-freedom chi-square criterion:
 critical χ^2 value for 1 df change = 3.84

Kline (2012) recommends 4 measures of model fit:

- (1) Model Chi-Square with its df and p-value.
 - prefer p-value greater than 0.05
- (2) Root Mean Square Error of Approximation (RMSEA).
 - prefer lower 90%CI to be < 0.05
- (3) Comparative Fit Index (CFI).
 - prefer value greater than 0.90
- (4) Standardized Root Mean Square Residual (SRMR).
 - prefer value less than 0.10

RMSEA for Our Example

Samples	RMSEA	LO90	HI90	PCLOSE
50	.126	.000	.426	.208
100	.162	.000	.356	.089
200	.177	.074	.307	.024

We are still affected by sample size / power.
 (which is reasonable)

As our sample size increases, we can expect our data to support more and more complex models.

Measures of Goodness of Fit that don't involve p-values

CFI: uses Centrality of model χ^2

- 50 samples = 0.96
- 100 samples = 0.94
- 200 samples = 0.94

Diagnosing Causes of Lack of Fit with Residuals (misspecification)

Residuals: These represent the magnitude of deviation between the relationships in the data and those implied by the model. Very important for evaluating model fit.

Sample Covariance Matrix				Implied Covariance Matrix			
	y1	y2	x		y1	y2	x
y1	1.00			y1	1.00		
y2	0.50	1.00		y2	0.50	1.00	
x	0.40	0.35	1.00	x	0.40	0.20	1.00

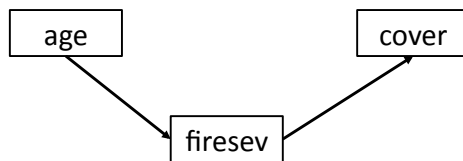
residual = 0.15

Modification Indices: Predicted effects of model modification on model chi-square.

Statistical Modification Criteria

- If a path is N.S., dropping it may lead to less variation in other parameter estimates.
- Lagrange Multipliers: The amount that χ^2 would decrease due to including a path.
- Wald W statistic: How much χ^2 would *increase* if a path is trimmed.
- Be very careful here for data dredging.

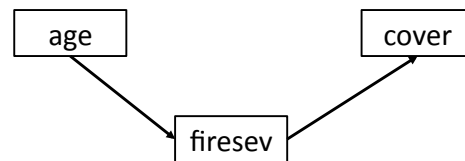
Fully Mediated Fire



```
fullMedModel<- ' firesev ~ age
                cover ~ firesev'
```

```
fullMedSEM<-sem(fullMedModel,
                 data=keeley)
```

Fit of the Fully Mediated Model



```
> summary(fullMedSEM)
lavaan (0.4-12) converged normally after 21 iterations

Number of observations              90

Estimator                          ML
Minimum Function Chi-square         3.297
Degrees of freedom                   1
P-value                              0.069
```

Fit-A-Palooza

summary(fullMedSEM, fit.measures=T)

```

> summary(fullMedSEM, fit.measures=T)
lavaan (0.4-12) converged normally after 21 iterations

Number of observations                    90
Estimator                                ML
Minimum Function Chi-square              3.297
Degrees of freedom                       1
P-value                                  0.069

Chi-square test baseline model:
Minimum Function Chi-square             43.143
Degrees of freedom                      3
P-value                                  0.000

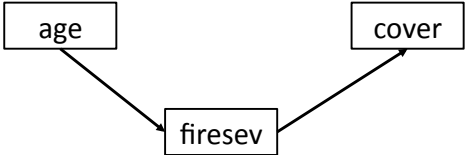
Full model versus baseline model:
Comparative Fit Index (CFI)             0.943
Tucker-Lewis Index (TLI)                0.828

Loglikelihood and Information Criteria:
Loglikelihood user model (H0)            -531.341
Loglikelihood unrestricted model (H1)    -529.693
Number of free parameters                4
Akaike (AIC)                            1070.683
Bayesian (BIC)                           1080.682
Sample-size adjusted Bayesian (BIC)      1068.057

Root Mean Square Error of Approximation:
RMSEA                                    0.160
90 Percent Confidence Interval           0.000 0.365
P-value RMSEA <= 0.05                   0.101

Standardized Root Mean Square Residual:
SMRM                                      0.062
    
```

Observed v. Fitted



```

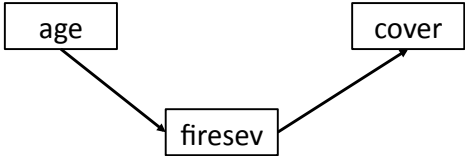
> inspect(fullMedSEM, "sample")
$cov
      firesv cover age
firesev 2.700
cover   -0.227 0.100
age      9.319 -1.381 156.157

$mean
firesev cover age
4.565   0.691 25.567

> fitted(fullMedSEM)
$cov
      firesv cover age
firesev 2.700
cover   -0.227 0.100
age      9.319 -0.782 156.157

$mean
firesev cover age
0         0         0
    
```

Residual Covariance

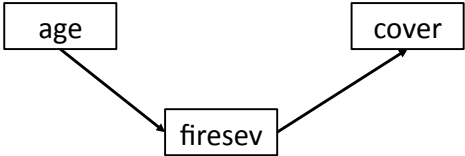


```

> residuals(fullMedSEM)
$cov
      firesv cover age
firesev 0.000
cover   0.000 0.000
age      0.000 -0.599 0.000

$mean
firesev cover age
0         0         0
    
```

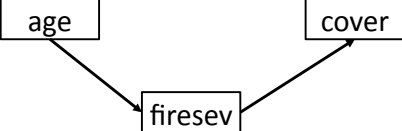
Residual Correlation



```

> residuals(fullMedSEM, type="cor")
      firesv cover age
firesev 0.000
cover   0.000 0.000
age      0.000 -0.152 0.000
    
```

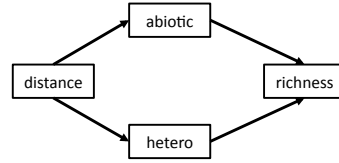
Modification Indices



```

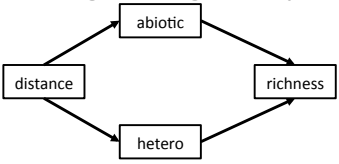
> modificationIndices(fullMedSEM, standardized=F)
  lhs op   rhs   mi   epc
1 firesev ~~ firesev 0.000 0.000
2 firesev ~~ cover 3.238 0.174
3 firesev ~~ age NA NA
4 cover ~~ cover 0.000 0.000
5 cover ~~ age 3.238 -0.755
6 age ~~ age 0.000 0.000
7 firesev ~ cover 3.238 2.157
8 firesev ~ age 0.000 0.000
9 cover ~ firesev 0.000 0.000
10 cover ~ age 3.238 -0.005
11 age ~ firesev 0.000 0.000
12 age ~ cover 2.884 -8.351
    
```

Exercise: Diagnosing Misspecification



- Fit and assess model
- Look at measures of misspecification

Solution: Diagnosing Misspecification

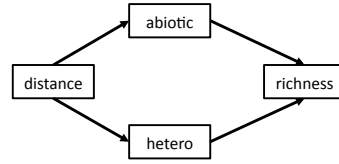


```

#Full Mediation
distModel2 <- 'rich ~ abiotic + hetero
             hetero ~ distance
             abiotic ~ distance'

distFit2 <- sem(distModel2, data=keeley)
    
```

Solution: Model Doesn't Fit Data



```

summary(distFit2)
lavaan (0.4-12) converged normally after 36 iterations

Number of observations              90

Estimator                          ML
Minimum Function Chi-square         17.831
Degrees of freedom                   2
P-value                              0.000
    
```

Solution: Large Residual rich->distance correlation

```

    graph LR
      distance --> abiotic
      distance --> hetero
      abiotic --> richness
      hetero --> richness
  
```

```

> residualsCor(distFit2)
      rich hetero abiotic distnc
rich      0.000
hetero   -0.042  0.000
abiotic  -0.032 -0.118  0.000
distance -0.271  0.000  0.000  0.000
  
```

Solution: Large Residual rich->distance correlation

```

    graph LR
      distance --> abiotic
      distance --> hetero
      abiotic --> richness
      hetero --> richness
  
```

```

#modification indices, with a trick to only see big ones
> modI<-modificationIndices(distFit2, standardized=F)

> modI[which(modI$mi>3),]
      lhs op      rhs      mi      epc
1   rich ~ hetero 15.181 -1.690
2   rich ~ abiotic 15.181 -76.202
3   rich ~ distance 15.181  0.662
4  abiotic ~ rich  3.811 -0.196
5 distance ~ rich 10.672  0.251
  
```

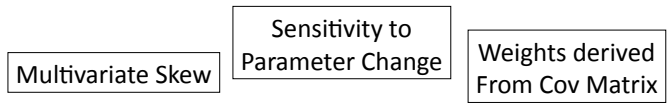
Additional Points about Overall Model Fit

1. Assessing overall model fit is a quest to ensure you have an adequate model (no important missing paths).
2. When you are missing important paths/relationships, your parameter estimates may be incorrect and not reflect the true relationships in the data. Then, your model is said to be **misspecified**.

Outline

1. Assessing model fit: the χ^2
 - Related indices
2. Adjusting for non-normality
3. Model comparison
4. Testing mediation

Correcting for Violation of Normality: The Satorra-Bentler Chi Square

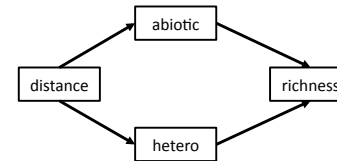


Correction coefficient for χ^2 and Standard Errors

```
distFitSB<-sem(distModel2, data=keeley,
               estimator="mlm")
```

- GLS, WLS are other fitting estimators
- MLF, MLR use ML but implement other corrections

Satorra-Bentler Output

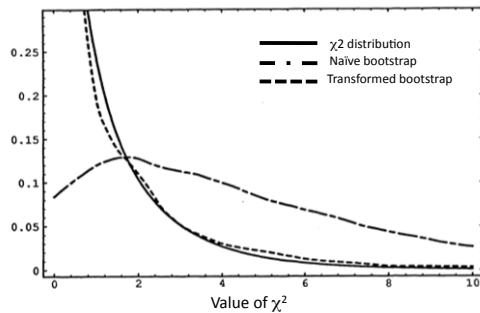


```
> summary(distFit2SB)
lavaan (0.5-12) converged normally after 51 iterations

Number of observations              90

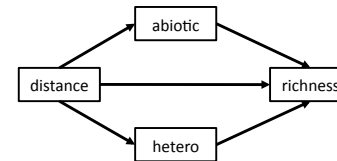
Estimator                          ML      Robust
Minimum Function Test Statistic    17.831  17.854
Degrees of freedom                  2       2
P-value (Chi-square)                0.000   0.000
Scaling correction factor           0.999
```

Violation of Multivariate Normality: The Bollen-Stine Bootstrap



To get accurate bootstrap, you can calculate a bootstrapped χ^2 , but you must transform your data first.

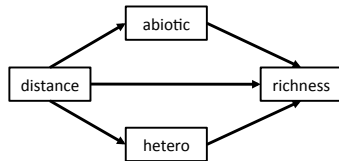
Bollen-Stine Bootstrap Output



```
> distFitBoot<-sem(distModel, data=keeley,
                  test="bollen.stine", se="boot", bootstrap=100)
```

Typically want ~ 1000 bootstrap replicates

Bollen-Stine Bootstrap Output



```

> summary(distFitBoot)
lavaan (0.5-12) converged normally after 40 iterations

Number of observations              90

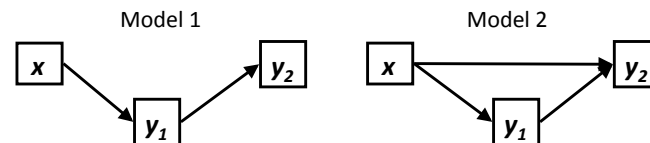
Estimator                          ML
Minimum Function Test Statistic     1.810
Degrees of freedom                   1
P-value (Chi-square)                 0.178
P-value (Bollen-Stine Bootstrap)    0.194
  
```

Questions?

Outline

1. Assessing model fit: the χ^2
 - Related indices
2. Adjusting for non-normality
3. Model comparison
4. Testing mediation

The Likelihood Ratio Test Revisited for Mediation



•Previously, we used the LRT to compare a saturated model to a non-saturated model.

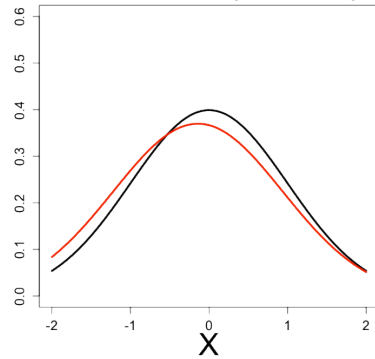
for n = 50 samples,

	χ^2	DF	p
Model 1	1.78	1	
Model 2	0.00	0	
diff	1.78	1	0.18

which would suggest Model 1 fits as well as model 2 with fewer paths – parsimony wins!

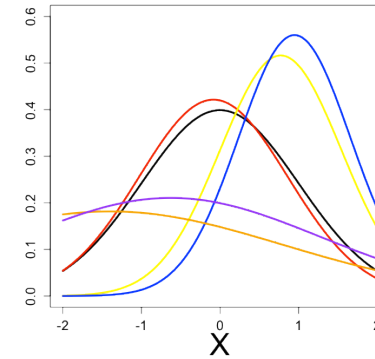
•We can use it to compare any set of nested models that differ in DF

AIC Comparisons: Because You Will Only Ever Know Your Sampled Population



f(x) = "True" value at point x
Discrepancy between fit model and f(x) conveys information loss

Models Provide Varying Degrees of Information about Reality



$G_i(x|\theta)$ = estimate of model i at point x given parameters θ

Kulback-Leibler Information

$$I(f, g) = \int f(x) \log \frac{f(x)}{g(x|\theta)} dx$$

I(f,g) = information loss when g is used to approximate f – integrated over all values of x

Note: f(x) can be pulled out as a constant when comparing multiple models! No need to know the true value of f(x)

Likelihood and Information

For likelihood, information loss is conveyed by the following with K = # of parameters:

$$\log(L(\hat{\theta} | data)) - K = constant - I(f, \hat{g})$$

This gives rise to Akaike's Information Criterion – lower AIC means less information is lost by a model

$$\text{AIC} = -2\log(L(\hat{\theta} | data)) + 2K$$

Principal of Parsimony:
How many parameters does it take to fit an
elephant?



Correcting for Sample Size: the AICc

$$AICc = AIC + \left(\frac{2t(t+1)}{n-t-1} \right)$$

where t = number of estimated parameters in the model and
 n = the number of samples

Note, this is not the "consistent AIC" reported as CAIC by many pieces of software

Model Weights to Compare Models

- In a set of models, the difference between model i and the model with the best fit is $\Delta_i = AIC_i - AIC_{\min}$
- We can then define the relative support for a model as a model weight

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta_i\right)}{\sum_{r=1}^R \exp\left(-\frac{1}{2}\Delta_r\right)}$$

- N.B. model weights summed together = 1

AIC and SEM

- $AIC = \chi^2 + 2K$
- Small Sample-Size Adjusted AIC
 $AIC_c = \chi^2 + 2K * (K-1) / (N-K-1)$
- Complexity Adjusted CAIC
 $CAIC = \chi^2 - DF * (1 + \log(N))$
- Bayesian Information Criterion
 $BIC = \chi^2 - DF * \log(N)$

AIC difference criteria

AIC diff	support for equivalency of models
0-2	substantial
4-7	weak
> 10	none

Note: Models are not required to be nested, as in using LRT tests

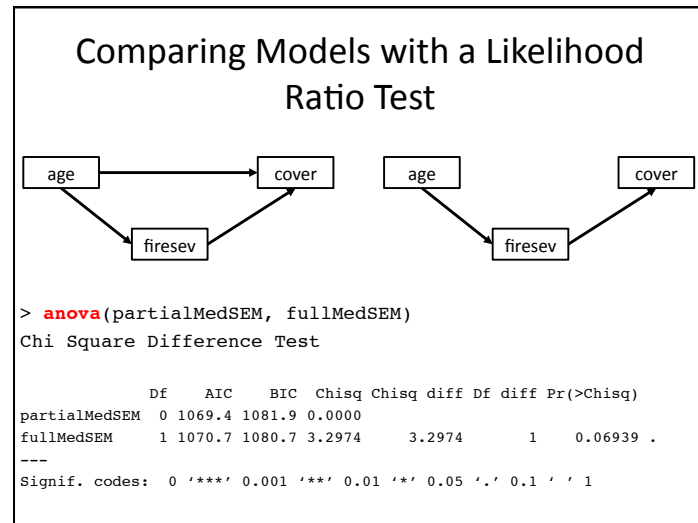
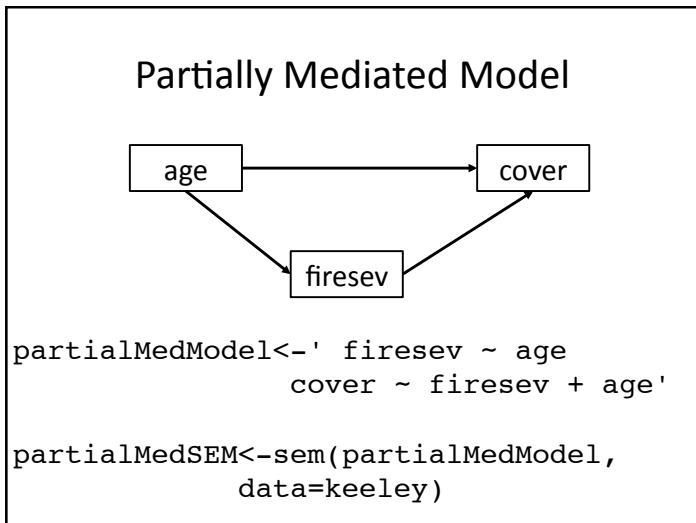
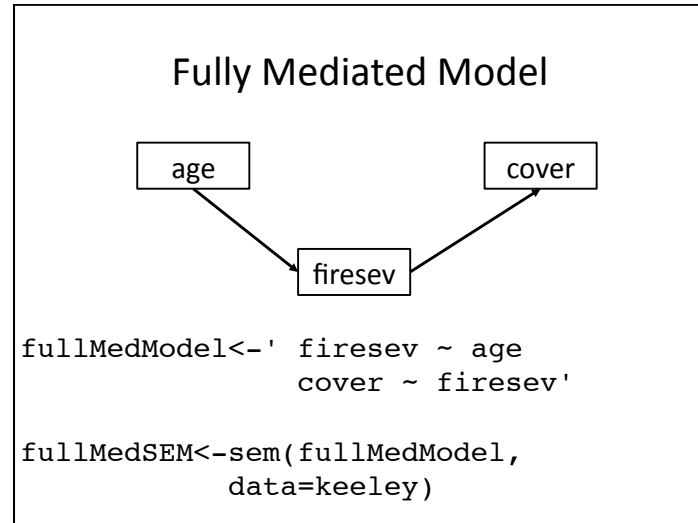
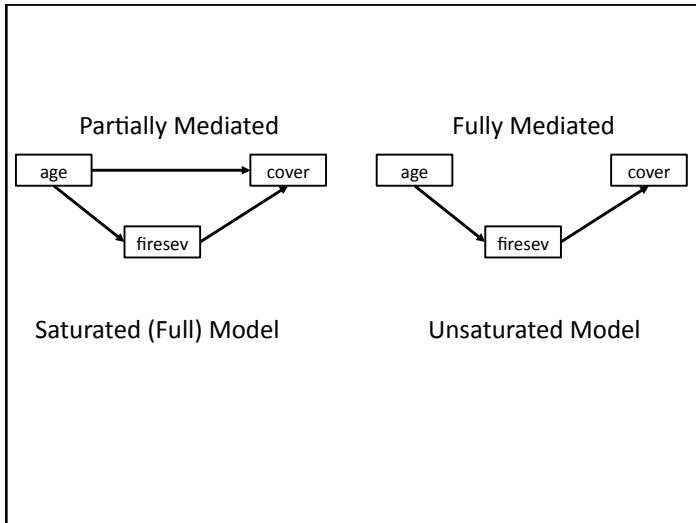
Burnham, K.P. and Anderson, D.R. 2002. Model Selection and Multimodel Inference. Springer Verlag. (second edition), p 70.

LR Testing v. AIC

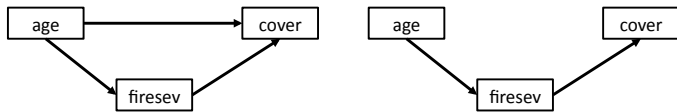
1. SEM provides a framework that aids the application of scientific judgement to the decision problem.
2. There is a growing interest in an information-based approach that avoids p-values and hypothesis tests while focusing on model selection and effect sizes.
3. There remain multiple viewpoints on whether hypothesis tests can or should be avoided.
4. The two can be used complementarily, as we shall see

Outline

1. Assessing model fit: the χ^2
 - Related indices
2. Adjusting for non-normality
3. Model comparison
4. Testing mediation



Comparing Models with AICc



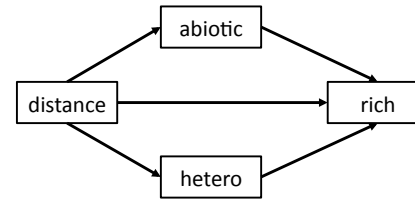
```
> source("./lavaan.modavg.R")
> aictab.lavaan(list(fullMedSEM, partialMedSEM),
               c("Full", "Partial"))
```

Model selection based on AICc :

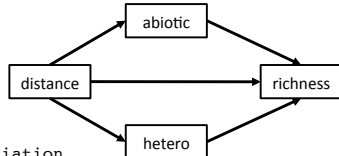
	K	AICc	Delta_AICc	AICcWt	Cum.Wt	LL
Partial	5	1069.66	0.00	0.64	0.64	-529.69
Full	4	1070.82	1.16	0.36	1.00	-531.34

Exercises

Perform a test of mediation for the following model



Solution: The Models



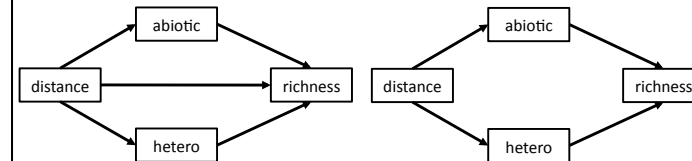
```
#Partial Mediation
distModel <- 'rich ~ distance + abiotic + hetero
             hetero ~ distance
             abiotic ~ distance'
```

```
distFit <- sem(distModel, data=keeley)
```

```
#Full Mediation
distModel2 <- 'rich ~ abiotic + hetero
              hetero ~ distance
              abiotic ~ distance'
```

```
distFit2 <- sem(distModel2, data=keeley)
```

Solution 3: Model Comparison with LRT

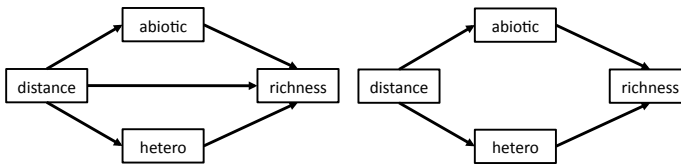


```
> anova(distFit, distFit2)
Chi Square Difference Test
```

	Df	AIC	BIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)				
distFit	1	1801.7	1821.7	1.8104							
distFit2	2	1815.8	1833.2	17.8307	16.02	1	6.267e-05				

Signif. codes:	0	'****'	0.001	'***'	0.01	'**'	0.05	'.'	0.1	' '	1

Solution 3: Model Comparison with AICc



```
> aictab.lavaan(list(distFit2, distFit),
  c("Full", "Partial"))
```

Model selection based on AICc :

	K	AICc	Delta_AICc	AICcWt	Cum.Wt	LL
Full	8	1802.44	0.00	1	1	-892.87
Partial	7	1816.22	13.78	0	1	-900.88

Mediation & SEM

- A central goal of SEM analyses is the evaluation of mediation
- We can use complementary sources of information to determine mediation
- Models that we evaluate for AIC analyses, etc, must fit the data before using in calculating AIC differences, etc.

Questions?