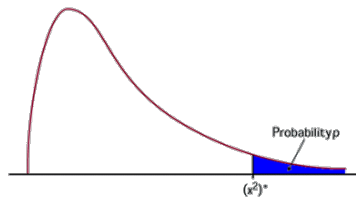


## Assessing Fit & Comparing SEMs with Likelihood

Jarrett E. K. Byrnes



## Outline

1. Assessing model fit: the  $\chi^2$ 
  - Related indices
2. Evaluating Residuals for Normality
3. Adjusting for non-normality
4. Model comparison
5. Testing mediation

## Outline

1. Assessing model fit: the  $\chi^2$ 
  - Related indices
2. Evaluating Residuals for Normality
3. Adjusting for non-normality
4. Model comparison
5. Testing mediation

## Evaluating Fit of Modeled Covariances Matrix

The log likelihood ratio,  $F_{ML}$  follows  $\chi^2$  distribution such that

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

- Note scaling by sample size
- Large  $\chi^2$  implies LACK of fit

### The Classic Test using P Values

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

p>0.05 means no discrepancy between sample and observed covariance matrix

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

As usual, p values decrease with higher n

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

### 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  $\chi^2$**

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

### Evaluating Fit of A Model

issue: should there be a path from x to  $y_2$ ?

std. covariance matrix

1.0		
0.4	1.0	
0.35	0.5	1.0

$r_{xy2}$  expected to be  $0.2 = (0.40 \times 0.50)$

standardized residual =  $0.35 - 0.2 = 0.15$

### Diagnosing Causes of Lack of Fit with Residuals (misspecification)

	y1	y2	x
y1	1.00		
y2	0.50	1.00	
x	0.40	0.35	1.00

	y1	y2	x
y1	1.00		
y2	0.50	1.00	
x	0.40	0.20	1.00

residual = 0.15

**But how much will including a path increase model fit?**

### Modification Indices

- **Lagrange Multipliers:** The amount that  $\chi^2$  would decrease due to including a path.
- **Wald W statistic:** How much  $\chi^2$  would increase if a path is trimmed.
  - Dropping a path can increase parameter variability
- Be very careful here for data dredging.

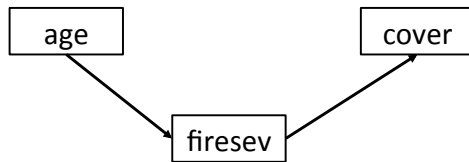
### Fully Mediated Fire

```

fullMedModel<- ' firesev ~ age
                cover ~ firesev '

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

### Fit of the Fully Mediated Model



```

> summary(fullMedSEM)
lavaan (0.5-17) converged normally after 19 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:
SRMR                                  0.062
  
```

### Fit-A-Palooza summary(fullMedSEM, fit.measures=T)

```

> summary(fullMedSEM, fit.measures=T)
...

Full model versus baseline model:

Comparative Fit Index (CFI)           0.943
Tucker-Lewis Index (TLI)              0.828

...

Number of free parameters              4
Akaike (AIC)                           1070.683
Bayesian (BIC)                          1080.682
Sample-size adjusted Bayesian (BIC)    1068.057

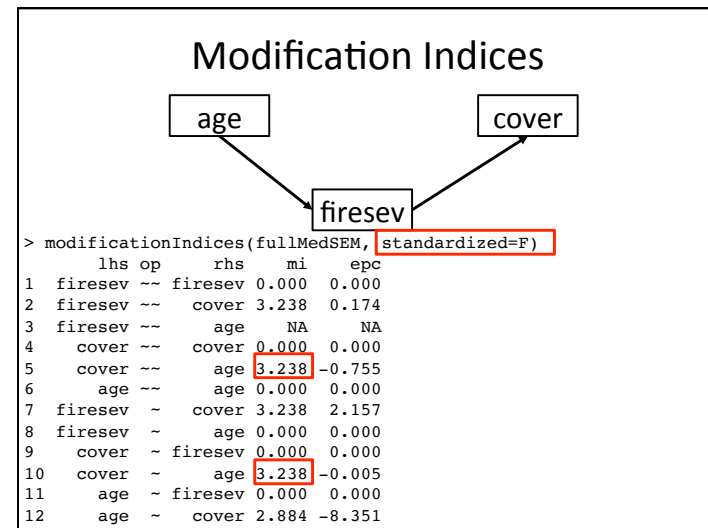
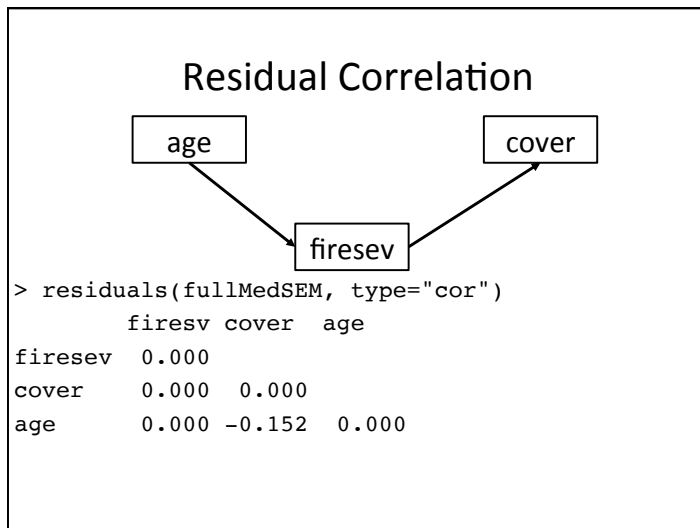
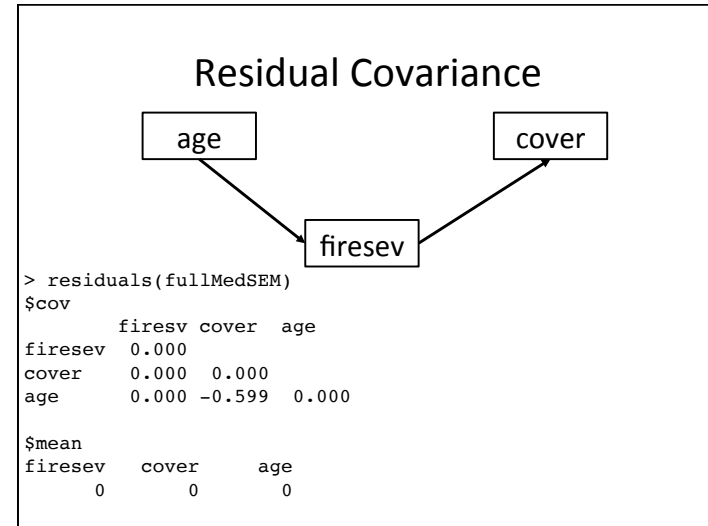
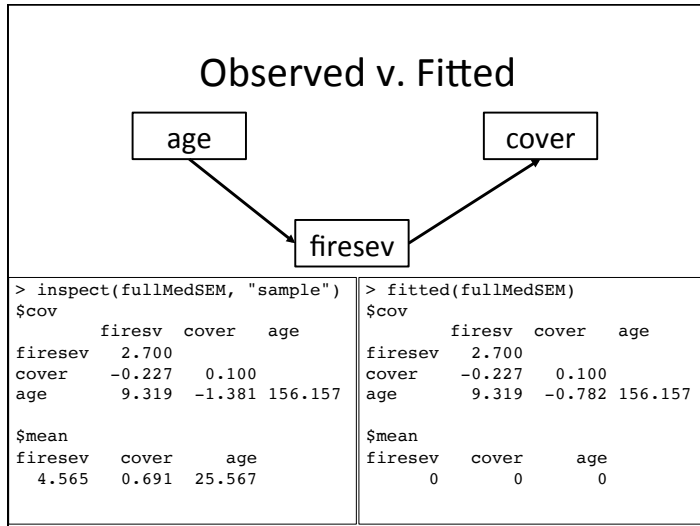
Root Mean Square Error of Approximation:

RMSEA                                 0.160
90 Percent Confidence Interval         0.000 0.365
P-value RMSEA <= 0.05                 0.101
  
```

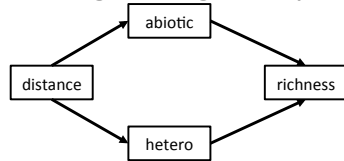
### Fit-A-Palooza2 fitMeasures (fullMedSEM)

```

> fitMeasures(fullMedSEM)
      npar      fmin      chisq      df
      4.000      0.018      3.297      1.000
      pvalue  baseline.chisq  baseline.df  baseline.pvalue
      0.069      43.143      3.000      0.000
      cfi      tli      nnfi      rfi
      0.943      0.828      0.828      0.771
      nfi      pnfi      ifi      rni
      0.924      0.308      0.945      0.943
      logl  unrestricted.logl      aic      bic
      -531.341      -529.693      1070.683      1080.682
      ntotal      bic2      rmsea      rmsea.ci.lower
      90.000      1068.057      0.160      0.000
      rmsea.ci.upper  rmsea.pvalue      rmr      rmr_nomean
      0.365      0.101      0.245      0.245
      srmr      srmr_bentler  srmr_bentler_nomean  srmr_bollen
      0.062      0.062      0.062      0.062
      srmr_bollen_nomean  srmr_mplus  srmr_mplus_nomean  cn_05
      0.062      0.062      0.062      105.849
      cn_01      gfi      agfi      pgfi
      182.093      0.977      0.859      0.163
      mfi      ecvi
      0.987      0.126
  
```

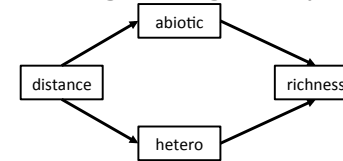


### Exercise: Diagnosing Misspecification



- Fit and assess model
- Look at measures of misspecification

### Solution: Diagnosing Misspecification



#Full Mediation

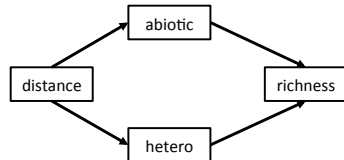
```

distModel2 <- 'rich ~ abiotic + hetero
              hetero ~ distance
              abiotic ~ distance'
  
```

```

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

### Solution: Model Doesn't Fit Data

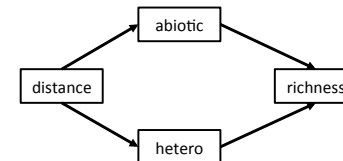


```

> summary(distFit2)
lavaan (0.5-17) converged normally after 36 iterations
  
```

Number of observations	90
Estimator	ML
Minimum Function Test Statistic	17.831
Degrees of freedom	2
P-value (Chi-square)	0.000

### Solution: Large Residual rich->distance correlation

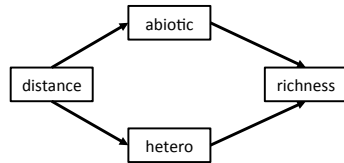


```

> residuals(distFit2, type="cor")
$cor
  
```

	rich	hetero	abiotc	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



#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
  
```

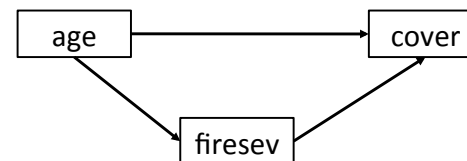
### Additional Points about Overall Model Fit

1. SEM focuses on assessing overall model fit
  - Is your model adequate?
  - Are you missing any paths?
2. When you are missing important paths your parameter estimates may be incorrect
  - your model is ***misspecified***

### Outline

1. Assessing model fit: the  $\chi^2$ 
  - Related indices
2. Evaluating Residuals for Normality
3. Adjusting for non-normality
4. Model comparison
5. Testing mediation

### Partial Mediation Model



```

partialMedModel<- ' firesev ~ age
                  cover ~ firesev + age '
  
```

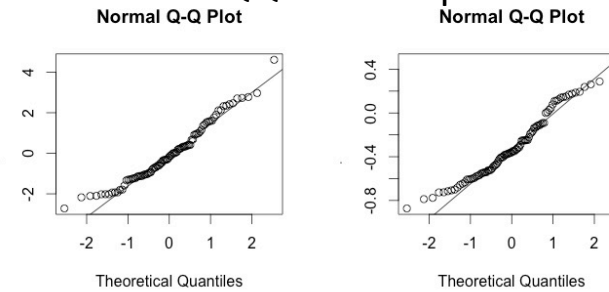
```

partialMedSEM<-sem(partialMedModel,
                  data=keeley,
                  meanstructure=TRUE)
  
```

## What is the distribution of our residuals?

```
>source("../fitted_lavaan.R")
> partialResid <- residuals_lavaan(partialMedSEM)
> head(partialResid)
  firesev      cover
1 -1.9263673  0.4752431
2 -0.4811819 -0.2186521
3 -1.3343917  0.1642312
4 -1.0343917  0.4101956
5 -0.1118239  0.5842525
6 -0.4715029  0.4683961
```

## QQ Plots Help



```
par(mfrow=c(1,2))
  apply(partialResid, 2, function(x){
    qqnorm(x)
    qqline(x)})
par(mfrow=c(1,1))
```

## Multivariate Shapiro-Wilks Test

```
library(mvnormtest)
> mshapiro.test(t(partialResid))
Shapiro-Wilk normality test
data: Z
W = 0.96889, p-value = 0.02954
```

Often too sensitive of a test

## Formal Tests from MVN

```
library(MVN)
mt <- mardiaTest(partialResid, qqplot=F)
Mardia's Multivariate Normality Test
-----
data : partialResid

g1p      : 0.6205411
chi.skew : 9.308116
p.value.skew : 0.05384291

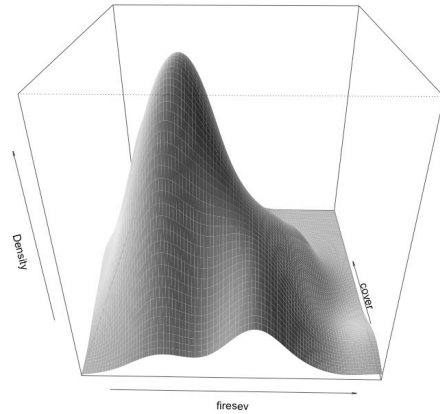
g2p      : 7.249683
z.kurtosis : -0.8897668
p.value.kurt : 0.3735911

chi.small.skew : 9.83453
p.value.small : 0.04330918

Result   : Data are multivariate normal.
```



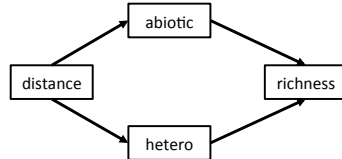
```
mvnPlot(mt, type = "persp", default = T)
```



## Outline

1. Assessing model fit: the  $\chi^2$ 
  - Related indices
2. Evaluating Residuals for Normality
3. Adjusting for non-normality
4. Model comparison
5. Testing mediation

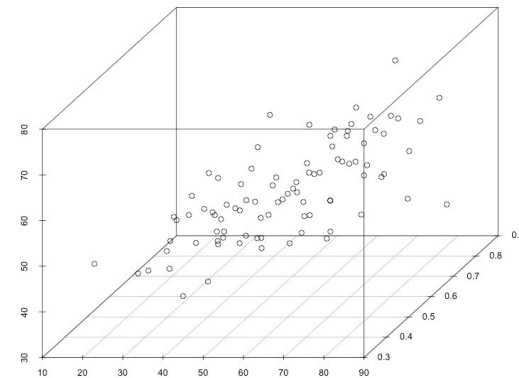
## Our Model for Correction



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

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

## Are these Residuals Multivariate Normal?



### Multivariate Shapiro-Wilks Test

```
> library(mvnormtest)
> mshapiro.test(t(res))
```

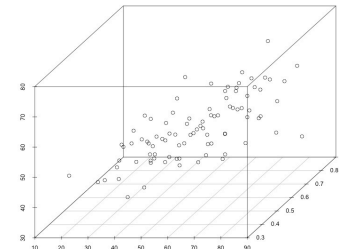
Shapiro-Wilk normality test

```
data: Z
W = 0.98579, p-value = 0.4367
```

These residuals are fine

- Can be overly sensitive
- Skew most important

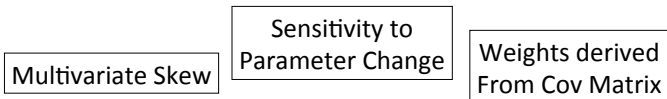
### Mardia's Multivariate Skew



```
> library(semTools)
> mardiaSkew(res)
> mardiaSkew(res)
      bld      chi      df      p
0.5693772 8.5406580 10.0000000 0.5761788
```

This is fine

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

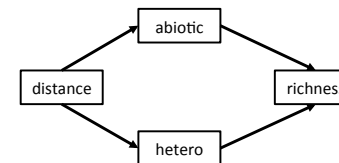


Correction coefficient for  $\chi^2$  and Standard Errors

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

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

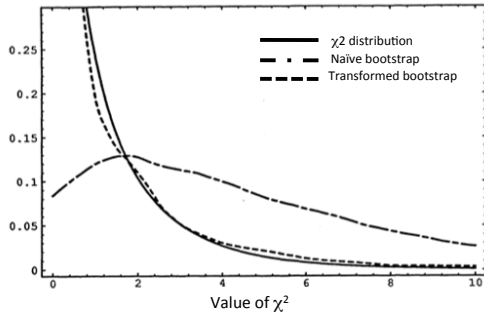
### Satorra-Bentler Output



```
> summary(distFit2SB)
lavaan (0.5-17) converged normally after 44 iterations
```

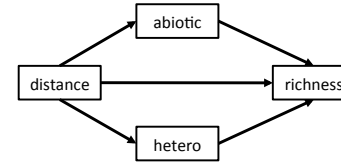
Number of observations	90	
Estimator	ML	Robust
Minimum Function Test Statistic	17.831	17.854
Degrees of freedom	2	2
P-value (Chi-square)	0.000	0.000
Scaling correction factor		0.999

### Violation of Multivariate Normality: The Bollen-Stine Bootstrap



To get accurate bootstrap, you can calculate a bootstrapped  $\chi^2$  on transformed data

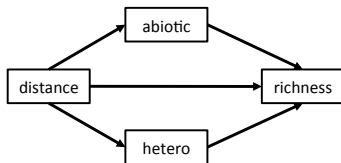
### Bollen-Stine Bootstrap Output



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

*Typically want ~ 1000 bootstrap replicates*

### Bollen-Stine Bootstrap Output



```
> summary(distFitBoot)
lavaan (0.5-17) converged normally after 37 iterations

Number of observations              90

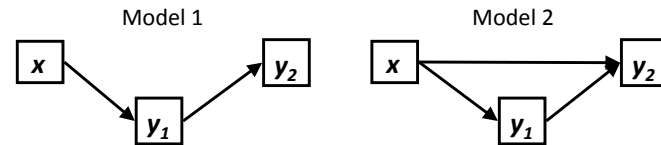
Estimator                          ML
Minimum Function Test Statistic     1.810
Degrees of freedom                  1
P-value (Chi-square)                0.178
P-value (Bollen-Stine Bootstrap)    0.140
```

Questions?

### Outline

1. Assessing model fit: the  $\chi^2$ 
  - Related indices
2. Evaluating Residuals for Normality
3. Adjusting for non-normality
4. Model comparison
5. Testing mediation

### The Likelihood Ratio Test Revisited for Mediation



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

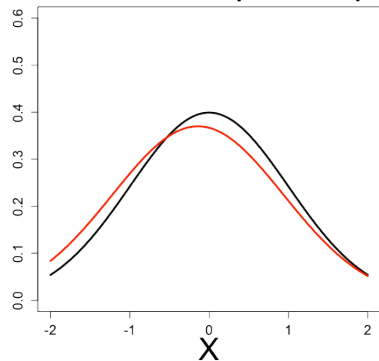
• We can use LRTs to compare any set of nested models that differ in DF

for n = 50 samples,

	$\chi^2$	DF	p
Model 1	1.78	1	
Model 2	0.00	0	
diff	1.78	1	0.18

Suggests Model 1 fits as well as model 2 with fewer paths – parsimony wins!

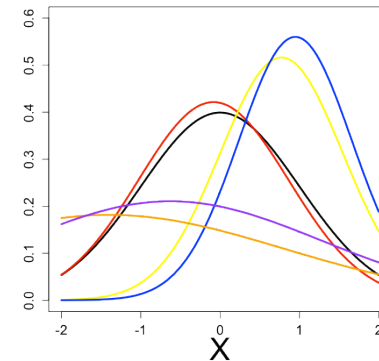
### AIC Comparisons: Because You Will Only Ever Know Your Sampled Population



$f(x)$  = "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  $\theta$

### 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 – most predictive model  
 $AIC = \chi^2 + 2K$
- Small Sample-Size Adjusted AIC  
 $AICc = \chi^2 + 2K * (K-1) / (N-K-1)$
- Bayesian Information Criterion – most 'correct' model  
 $BIC = \chi^2 - DF * \log(N)$

### 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

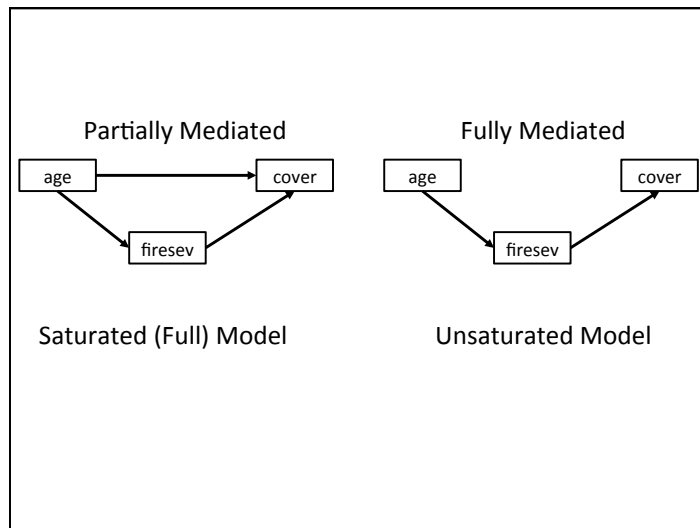
Burnham, K.P. and Anderson, D.R. 2002. Model Selection and Multimodel Inference. Springer Verlag. (second edition), p 70.

## LR Testing v. AIC

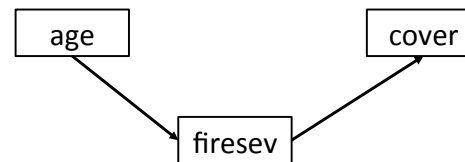
1. SEM provides a framework that aids the application of scientific judgment to selecting an appropriate model of the world
2. Growing interest in an information-based approach that focuses on model selection and effect sizes.
3. Many viewpoints on utility of Neyman-Pearson hypothesis testing
4. The two can be used complementarily, however!

## Outline

1. Assessing model fit: the  $\chi^2$ 
  - Related indices
2. Evaluating Residuals for Normality
3. Adjusting for non-normality
4. Model comparison
5. Testing mediation



## Fully Mediated Model



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

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

### Partially Mediated Model

```

partialMedModel<-' firesev ~ age
                    cover ~ firesev + age'

partialMedSEM<-sem(partialMedModel,
                  data=keeley)
    
```

### Comparing Models with a Likelihood Ratio Test

```

> anova(partialMedSEM, fullMedSEM)
Chi Square Difference Test

      Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
partialMedSEM  0 1069.4 1081.9  0.0000
fullMedSEM     1 1070.7 1080.7  3.2974    3.2974    1  0.06939 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    
```

### 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

Bonus: Calculate summed direct and indirect effects



### 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          16.02      1 6.267e-05
distFit2  2 1815.8 1833.2 17.8307
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    
```

### Solution 3: Model Comparison with AICc

```

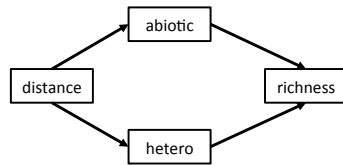
> aictab.lavaan(list(distFit2, distFit),
                c("Partial", "Full"))
Model selection based on AICc :

      K   AICc Delta_AICc AICcWt Cum.Wt   LL
Partial 8 1802.44      0.00      1     1 -892.87
Full    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.

## We Should Not have Used the Fully Mediated Model for AIC Analyses



lavaan (0.5-17) converged normally after 36 iterations

Number of observations	90
Estimator	ML
Minimum Function Test Statistic	17.831
Degrees of freedom	2
P-value (Chi-square)	0.000

Questions?