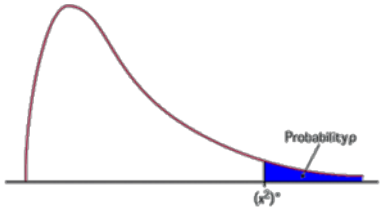
Assessing Fit & Comparing SEMs with Likelihood

Jarrett E. K. Byrnes
University of Massachusetts Boston



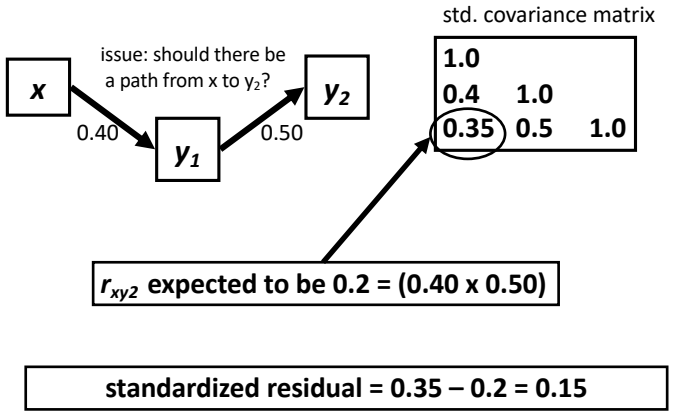
Outline

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

Evaluating Fit of A Model



Diagnosing Causes of Lack of Fit with Residuals (misspecification)

Sample Covariance Matrix

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

Implied Covariance Matrix

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

residual = 0.15

But there will always be residual correlation – is it good enough?

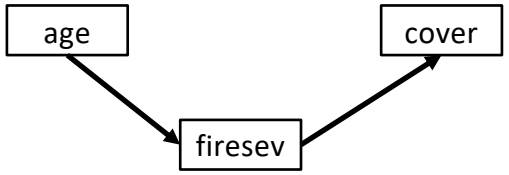
Evaluating Fit of Modeled Covariance Matrices with χ^2

The log likelihood ratio, F_{ML} follows χ^2 distribution such that

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

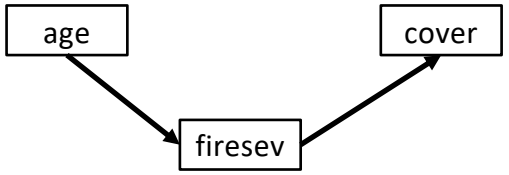
- ***Note scaling by sample size***
- ***Large χ^2 implies LACK of fit***

Fully Mediated Fire




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

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

Fit of the Fully Mediated Model




```
> fullMedSEM
lavaan (0.5-23.1097) converged normally after  22 iterations

  Number of observations                            90
  Estimator                                         ML
  Minimum Function Test Statistic                3.297
  Degrees of freedom                                 1
  P-value (Chi-square)                           0.069
```

But, Sample Size issues?

χ^2 = 1.78 with 50 samples
p = 0.182

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

χ^2 = 3.60 with 100 samples
p = 0.058

As usual, p values decrease with higher n

χ^2 = 7.24 with 200 samples
p = 0.007

Is this a feature or a bug of the technique?

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

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-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
```

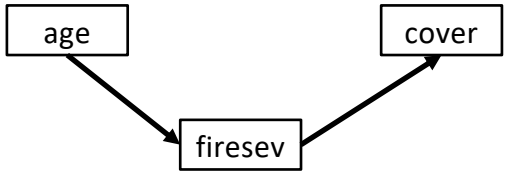
```
Loglikelihood and Information Criteria:

  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-Palooza2
`fitMeasures(fullMedSEM)`

```
> fitMeasures(fullMedSEM)
              npar               fmin              chisq                 df
             6.000              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           1074.683           1089.681
            ntotal               bic2              rmsea      rmsea.ci.lower
            90.000           1070.745              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
             0.051              0.051              0.062              0.051
 srmr_bollen_nomean         srmr_mplus   srmr_mplus_nomean             cn_05
             0.062              0.051              0.062            105.849
            cn_01                gfi               agfi               pgfi
          182.093              0.999              0.987              0.111
               mfi               ecvi
             0.987                 NA
```

OK, my model didn't fit.
What should I have included?

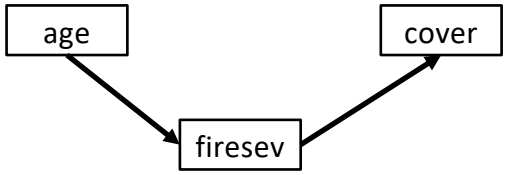
Observed v. Fitted




```
> 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
```

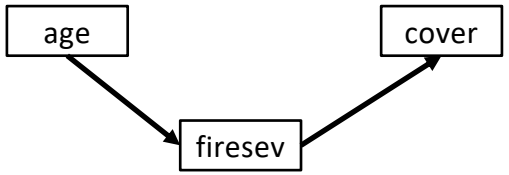
```
> inspect(fullMedSEM, "fitted")
   $cov
        firesv  cover   age
firesev   2.700
cover    -0.227   0.100
age       9.319  -0.782 156.157

$mean
firesev   cover     age
  4.565   0.691  25.567
```

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

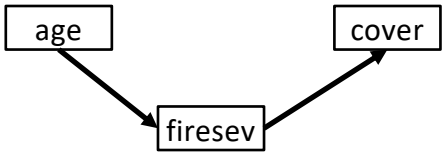



```
> residuals(fullMedSEM, type="cor")
        firesv cover  age
firesev  0.000
cover    0.000  0.000
age      0.000 -0.152  0.000
```

Modification Indices

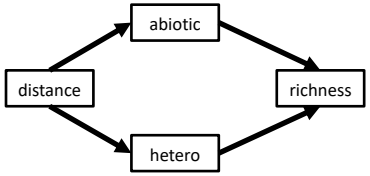
- **Lagrange Multipliers**: The amount that $\chi 2$ would decrease due to including a path.
- **Wald W statistic**: How much $\chi 2$ would *increase* if a path is trimmed.
- Be very careful here for data dredging.

Lagrange Multipliers



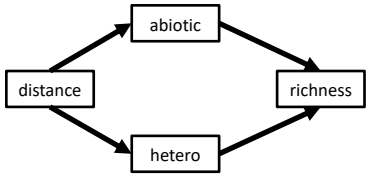

```
> modificationIndices(fullMedSEM,
 standardized=F)
       lhs op   rhs    mi    epc
9  firesev ~~ cover 3.238  0.174
10 firesev  ~ cover 3.238  2.157
11   cover  ~   age 3.238 -0.005
13     age  ~ cover 3.238 -9.375
```

Exercise: Diagnosing Misspecification



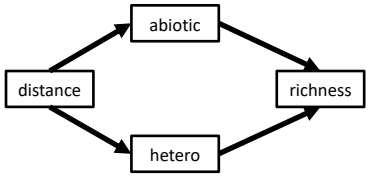
- Fit and assess model
- Look at measures of misspecification

Solution: Diagnosing Misspecification



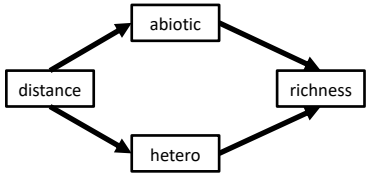

```
#Full Mediation
distMedModel <- '
  rich ~ abiotic + hetero
  hetero ~ distance
  abiotic ~ distance'

distMedFit <- sem(distMedModel, data=keeley)
```

Solution: Model Doesn't Fit Data!



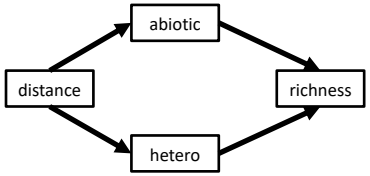

```
> distMedFit
lavaan (0.5-23.1097) converged normally after  36
    iterations

  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(distMedFit, 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[modI$mi>3,]
 lhs op      rhs     mi     epc
9      rich ~~   hetero 15.181  -1.690
10     rich ~~  abiotic 15.181 -76.202
12     rich  ~ distance 15.181   0.662
15  abiotic  ~     rich  3.811  -0.196
17 distance  ~     rich 14.728   0.347
```

Final Points about Assessing Fit

1. In SEM we assess <u>overall model fit</u>
   - Is your model adequate?
   - Are you missing any paths?

1. When you are missing important paths your parameter estimates may be incorrect
   - your model is **misspecified**

2. ***<u>But – what is your modeling goal?</u>***

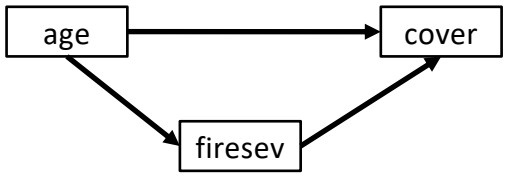
Outline

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

Two Major Assumptions of Covariance-based Estimation

1. Your residuals are normal
   - This is a linear modeling technique
   - Assumption of Gaussian error distribution
   - Violations require… corection
2. Your data is multivariate normal
   - You are fitting based on a covariance matrix
   - Assumption of multivariate normality of **data**
   - Violations can be accommodated

Partial Mediation Model

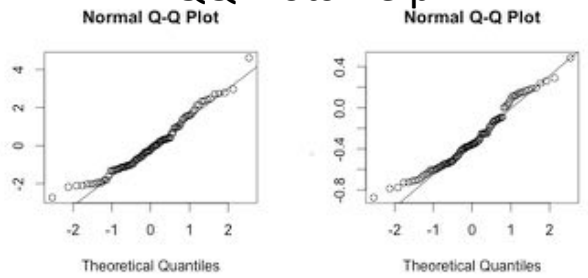



```
partialMedModel<-' firesev ~ age
                   cover ~ firesev + age'

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




```
#qqplot quick function
qqnorm_plot <- function(x){qqnorm(x); qqline(x)}

#2 panels
par(mfrow=c(1,2))
apply(partialResid, 2, qqnorm_plot)
par(mfrow=c(1,1))
```

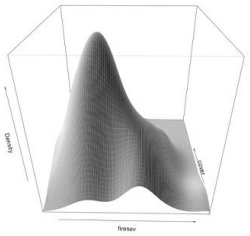
Multivariate Shapiro-Wilks Test

```
library(mvnormtest)

> mshapiro.test(t(partialResid))

   Shapiro-Wilk normality test

data:  Z
W = 0.96889, p-value = 0.02954
```

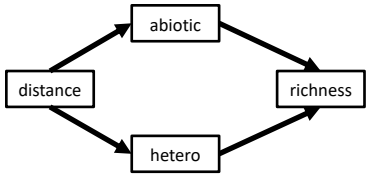
Often too sensitive of a test

Formal Tests from MVN

```
> library(MVN)

> mvn(partialResid, mvnTest="mardia",  multivariatePlot = "persp")
$multivariateNormality
             Test           Statistic            p value Result
1 Mardia Skewness   9.30811608721262 0.0538429114872067    YES
2 Mardia Kurtosis -0.889766844360397  0.373591093104194    YES
3             MVN                <NA>               <NA>
```

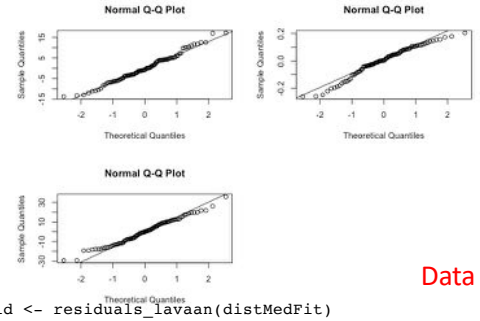


Exercise: Do We Meet Assumptions?




```
#Reminder – our model
distMedModel <- '
  rich ~ abiotic + hetero
  hetero ~ distance
  abiotic ~ distance'

distMedFit <- sem(distMedModel,
                  data=keeley,
                  meanstructure=TRUE)
```

Exercise: Do We Meet Assumptions?



Data is fine!

```
dist_resid <- residuals_lavaan(distMedFit)

#plot it
mvn(dist_resid, mvnTest="mardia",  univariatePlot = "qqplot")

$multivariateNormality
             Test          Statistic            p value Result
1 Mardia Skewness   8.83179215872101 0.548138184908851    YES
2 Mardia Kurtosis -0.880217174390066 0.378741671360096    YES
```

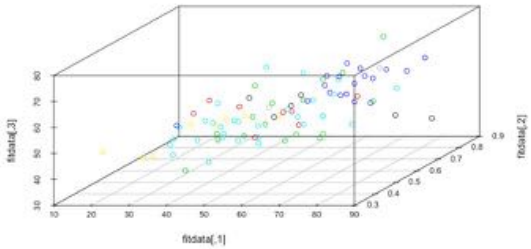
Multivariate Mardia's Test

```
> mshapiro.test(t(fitdata))

   Shapiro-Wilk normality test

data:  Z
W = 0.97472, p-value = 0.07636
```

Data is fine

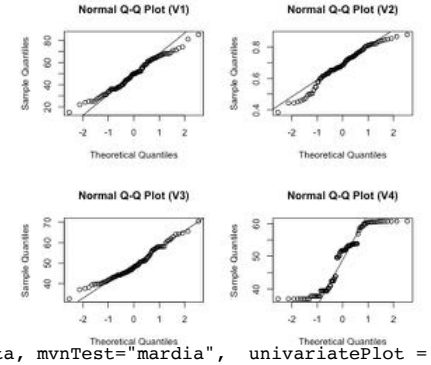
- Can be overly sensitive
- Skew most important

Are these Data Multivariate Normal?



```
#Get just the data used for fitting
fitdata <- inspect(distMedFit, "data")

#fun plots!
library(scatterplot3d)
scatterplot3d(fitdata)
```

Are these Data Multivariate Normal?




```
> mvn(fitdata, mvnTest="mardia",  univariatePlot = "qqplot")
$multivariateNormality
             Test         Statistic            p value Result
1 Mardia Skewness  17.4699469907421  0.622280935570818    YES
2 Mardia Kurtosis -1.72048778060888 0.0853438129583823    YES
```

Help! I Violated Assumptions!

1. My residuals are not normal
   - If this is simple nonlinearity, build it into model or transform data
   - If error generating process is non-gaussian, *piecewiseSEM*
2. My data is not normal
   - This can just be a feature of the data, and residuals may still be normal
   - If so, many techniques to get unbiased fit and error statistics!

Outline

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

Alternatives to FML

F_{ML} is unbiased, scale invariant, best estimator

$F_{GLS}=0.5*tr[(S-\Sigma(\Theta))^2]$

- A.K.A. the ULS criterion
- Least squares!
- Sensitive to scale of variables

$F_{WLS}=0.5*tr[\{(S-\Sigma(\Theta))W^{-1}\}^2)]$

- A.K.A. the ADF criterion – ***no normality assumption***
- Weighted: flexible
- Scale free
- Asymptotically χ^2 distributed
- Sensitive to fat or thin tailed data
- Sensitive to sample size (n>1000)

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

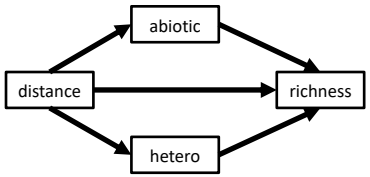
Multivariate Skew | Sensitivity to Parameter Change | Weights derived From Cov Matrix

Correction coefficient for χ^2 and Standard Errors

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

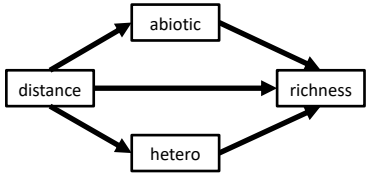
- GLS, WLS are other fitting estimators
- MLF, MLR use ML but implement other corrections

Satorra-Bentler Output




```
lavaan 0.6-3 ended normally after 37 iterations

  Optimization method                           NLMINB
  Number of free parameters                          8

  Number of observations                            90

  Estimator                                         ML      Robust
  Model Fit Test Statistic                       1.810       1.712
  Degrees of freedom                                 1           1
  P-value (Chi-square)                           0.178       0.191
  Scaling correction factor                                  1.058
    for the Satorra-Bentler correction
```

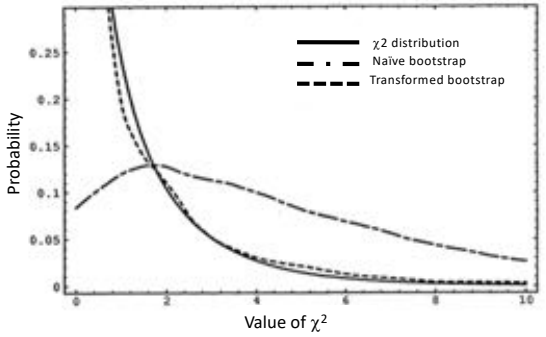
Bollen-Stine Bootstrap Output




```
distFitBoot<-sem(distModel2, data=keeley,
 test="bollen.stine", se="boot", bootstrap=100)
```

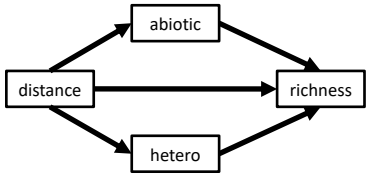
Typically want ~ 1000 bootstrap replicates

What is the BS Boot?

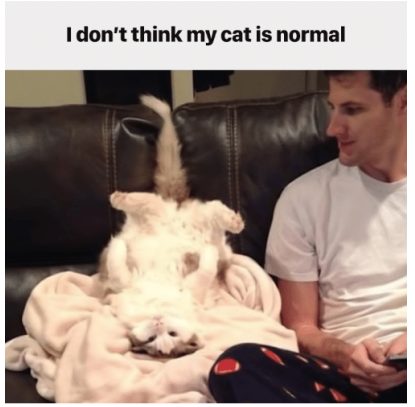


- Bootstrapped SEMs don't produce valid χ^2 statistics
- To get correct χ^2, you need to transform the data

Bollen-Stine Bootstrap Output




```
lavaan 0.6-3 ended normally after 37 iterations

  Optimization method                           NLMINB
  Number of free parameters                          8

  Number of observations                            90

  Estimator                                         ML
  Model Fit Test Statistic                       1.810
  Degrees of freedom                                 1
  P-value (Chi-square)                           0.178
  P-value (Bollen-Stine Bootstrap)               0.210
```



Questions?

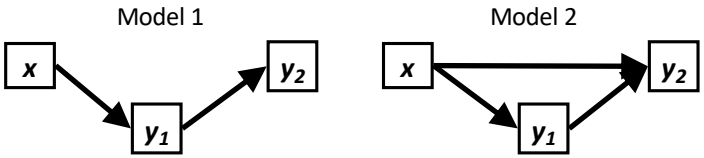
Outline

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

Model Comparison Paradigms

1. Does a simpler model still reproduce the more complex model's covariance matrix?
   - Likelihood Ratio Testing
2. Compare the weight of evidence across multiple models
   - Information Theoretic Approaches

The Likelihood Ratio Test Revisited for Mediation



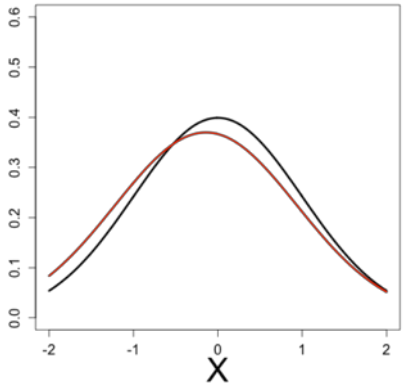
- Previously, we used a LRT to compare a saturated model to a non-saturated model.
- We can use LRTs to compare any set of nested models that differ in DF

for n = 50 samples,

	χ^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!

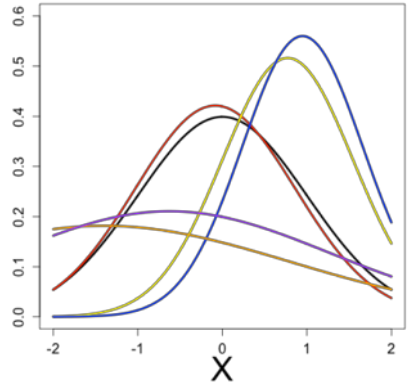
AIC Comparisons: Because You Will Only Ever Know Your Sampled Population



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 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

AND… 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 related via the following with K = # of parameters:

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

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

AIC = -2log(L(θhat|data))+2K

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

AIC and SEM

- AIC – most predictive model

  AIC = χ2+2K

- Small Sample-Size Adjusted AIC

  AICc=χ2+2K*(K-1)/(N-K-1)

- Bayesian Information Criterion – most 'correct' model

  BIC=χ2-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.

Model Weights Provide Intuitive Comparison

- 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

LR Testing v. AIC

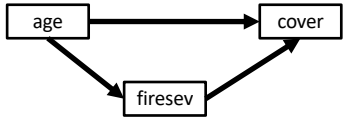
1. SEM provides a framework that aids the application of scientific judgment to selecting an appropriate model of the world
2. Growing interest in an information-based approach that focuses on model selection and effect sizes.
3. Many viewpoints on utility of Neyman-Pearson hypothesis testing
4. The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

Outline

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

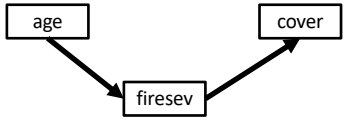
How is this relationship Mediated?

Partially Mediated



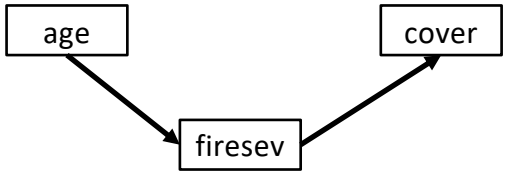
Saturated (Full) Model

Fully Mediated



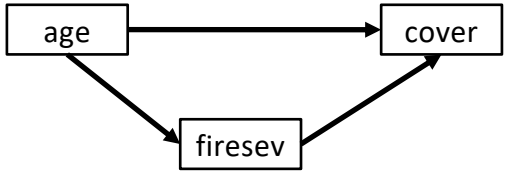
Unsaturated Model

Fully Mediated Model



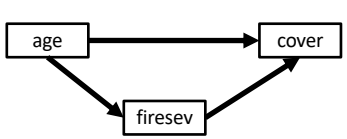

```
fullMedModel<-' firesev ~ age
                cover ~ firesev'

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

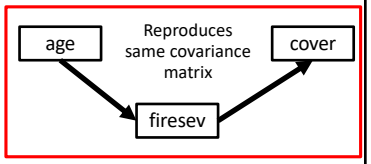
Partially Mediated Model




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

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

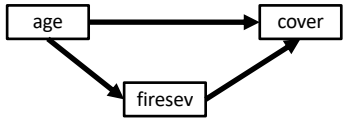
Comparing Models with a Likelihood Ratio Test

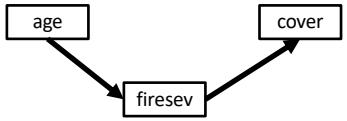





```
> anova(partialMedSEM, fullMedSEM)
Chi Square Difference Test

             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
```

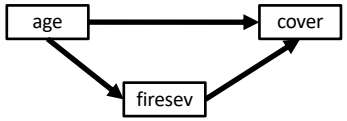
Comparing Models with AICc

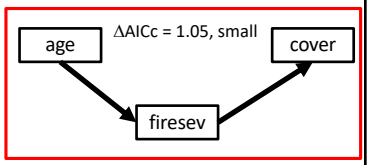





```
library(AICcmodavg)

aictab(cand.set = list(fullMedSEM, partialMedSEM),
       modnames = c("Full", "Partial"))
```

Comparing Models with AICc

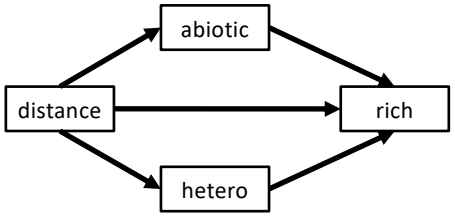




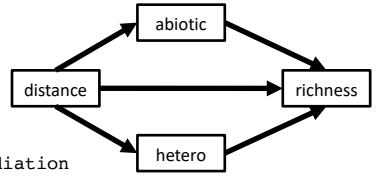

```
Model selection based on AICc:

        K   AICc Delta_AICc AICcWt Cum.Wt      LL
Partial 5 360.11       0.00   0.63   0.63 -174.70
Full    4 361.17       1.05   0.37   1.00 -176.35
```

Exercise

Perform a test of mediation for the following model to evaluate if the distance effect is partially or fully mediated by abiotic conditions and soil heterogeneity

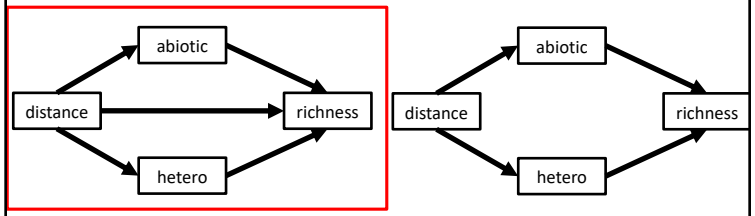


Solution: The Models



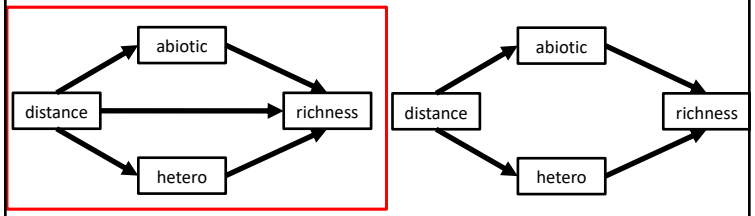

```
#Partial Mediation
distModel <- 'rich ~ distance + abiotic + hetero
              hetero ~ distance
              abiotic ~ distance'

distFit <- sem(distModel, data=keeley)

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

distMedFit <- sem(distMedModel, data=keeley)
```

Solution 3: Model Comparison with LRT




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

Chi Square Difference Test

           Df    AIC    BIC   Chisq Chisq diff Df diff Pr(>Chisq)
distFit     1 1155.3 1175.3  1.8104
distMedFit  2 1169.3 1186.8 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

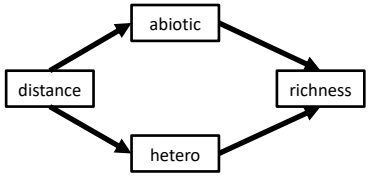



```
> aictab(cand.set = list(distMedFit, distFit),
       modnames = c("Full", "Partial"))

Model selection based on AICc:

        K    AICc Delta_AICc AICcWt Cum.Wt      LL
Partial 8 1157.05       0.00      1      1 -569.63
Full    7 1170.66      13.61      0      1 -577.64
```

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 0.6-3 ended normally after 36 iterations

  Number of observations                            90

  Estimator                                         ML
  Model Fit Test Statistic                      17.831
  Degrees of freedom                                 2
  P-value (Chi-square)                           0.000
```

Questions?
Then assess your fits from yesterday!