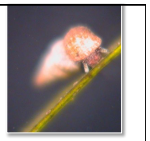
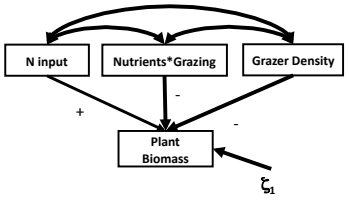


Looking at Familiar Statistical Concepts in a New SEM Light

Jarrett E. K. Byrnes



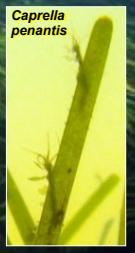
Old Wine in a New Bottle

1. ANOVA and ANCOVA in an SEM context
2. Multiple categorical predictors
3. Nonlinear effects

Data from:
Biodiversity and complex environmental forcing of ecosystem functioning in the marine foundation species, eelgrass:

Matt Whalen, J. Emmett Duffy, Jim Grace

York River, Virginia:
Major herbivores are invert crustaceans - these grazers control epiphytes and promote the eelgrass



Preliminary Study:
Virginia site

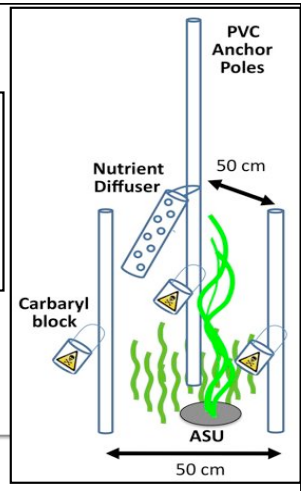
Experimental Design:

Treatments:
- pesticide to reduce crustacean grazers
- nutrient addition
- combination
- controls

8 reps @ 5 trts = 40 plots

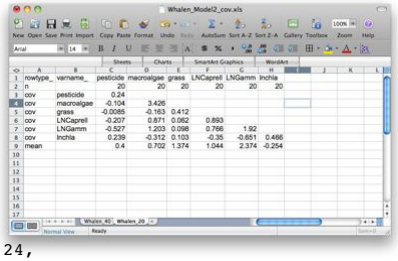
Pesticide effects:

Crustaceans: reduced 58-96%
Algal biomass: increased 130-748%
Nutrients: inconsistent effects



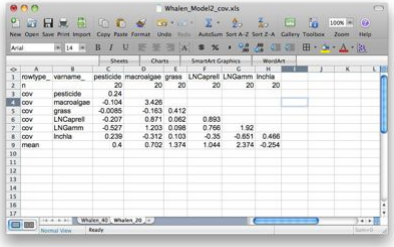
VIMS

Using Summarized Information



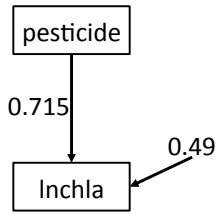
```
lower <- "0.24,
          -0.104,   3.426,
          -0.0085,  -0.163, 0.412,
          -0.207,   0.871,  0.062,   0.893,
          -0.527,   1.203,  0.098,   0.766,   1.92,
          0.239,    -0.312, 0.103,   -0.35,   -0.651,  0.466"

whalenCov <- getCov(lower,
              names=c("pesticide",  "macroalgae",
                      "grass",  "LNCaprell",  "LNGamm",
                      "lnchla"))
```

Using Summarized Information



```
whalenMeans <- c(0.4,  0.702,  1.374, 1.044,
   2.374,  -0.254)

whalenN <- 40
```

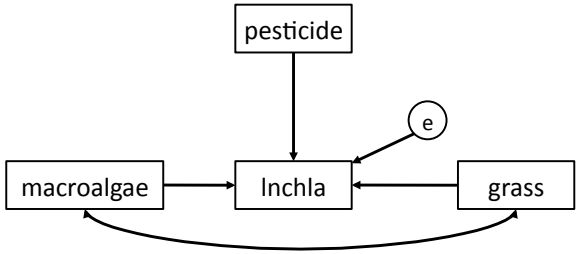
Seagrass ANOVA Model




```
anovaModel<-'lnchla ~ pesticide'

anovaFit<-sem(anovaModel, sample.cov=whalenCov,
  sample.mean=whalenMeans, sample.nobs=whalenN)
```

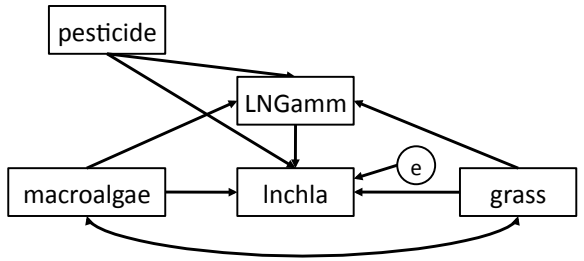
Seagrass ANCOVA Model

AMOS v. lavaan



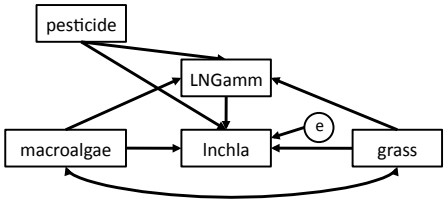

```
ancovaModel <- 'lnchla ~ pesticide +
                         macroalgae + grass

               pesticide ~~ 0*macroalgae + 0*grass'

ancovaFit <- sem(ancovaModel, sample.cov=whalenCov,
  sample.mean=whalenMeans, sample.nobs=whalenN,
  fixed.x=F)
```

Mediation Exercise 2!



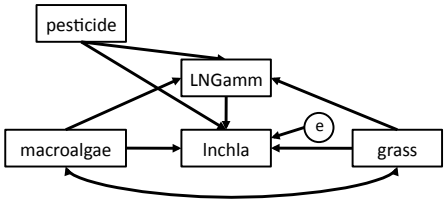
1. Fit fully mediated and partially mediated model with LNGamm
2. Evaluate evidence for Partial or Full Mediation
3. Bonus: fit the full model with LNCaprell

The Models




```
fullModel <- 'lnchla ~ macroalgae + grass + LNGamm
              LNGamm ~ macroalgae + grass + pesticide
                pesticide ~~ 0*macroalgae + 0*grass'

partialModel <- 'lnchla ~ macroalgae + grass + LNGamm + pesticide
                  LNGamm ~ macroalgae + grass + pesticide
                  pesticide ~~ 0*macroalgae + 0*grass'
```

Likelihood Ratio Comparison

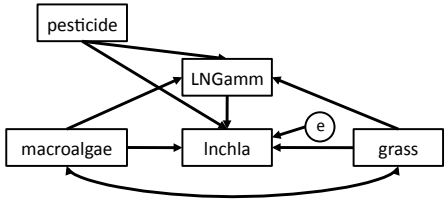



```
> anova(fullFit, partialFit)

Chi Square Difference Test
```

They are different
Partial Mediation Favored

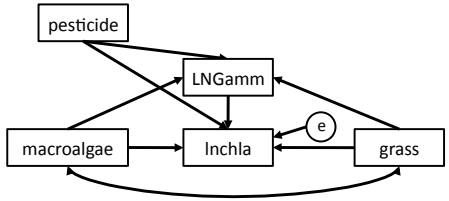
```
           Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
partialFit  2 440.41 462.36 0.6053
fullFit     3 442.95 463.22 5.1475     4.5422       1    0.03307
```

AICc versus AIC



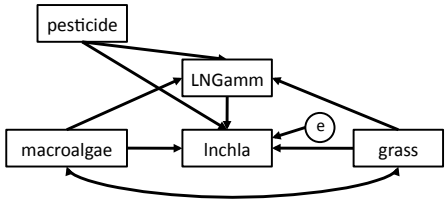

```
> aictab.lavaan(list(fullFit, partialFit), c("full",
   "partial"))

Model selection based on AICc :

         K   AICc Delta_AICc AICcWt Cum.Wt      LL
partial 12 442.67       0.00   0.69   0.69 -207.58
full    11 444.26       1.59   0.31   1.00 -209.86
```

AICc versus AIC




```
> aictab.lavaan(list(fullFit, partialFit), c("full",
   "partial"), second.ord=F)

Model selection based on AIC :

         K    AIC Delta AIC AICWt Cum.Wt      LL
partial 12 439.17      0.00  0.78   0.78 -207.58
full    11 441.71      2.54  0.22   1.00 -209.86
```

Summary

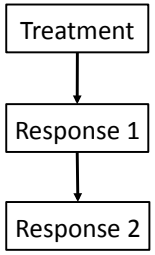


- Information criteria are suggestive
- Both state that the direct link is a better model, but hard to say…
- LR Test shows that partial mediation model is a better fit to the data
- Given that we may have a 2nd mediator (caprellids), this may lead to weaker performance of AICs

Old Wine in a New Bottle

1. ANOVA and ANCOVA in an SEM context
2. Multiple categorical predictors
3. Nonlinear effects

What about experiment with more than 2 levels of treatment?

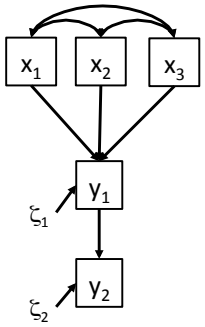


1. Can you make the treatment continuous?
   - E.g. nutrient levels
2. Or, treat each level as being present/absent

$y = \gamma_1 x_1 + \gamma_2 x_2 + \zeta$

where x_i=0 or 1

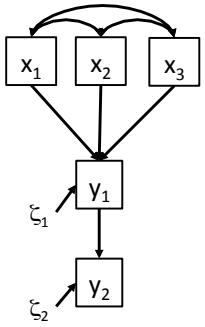
Experiment with 3 Levels



- Exogenous covariance no longer 0.

	X1	X2	X3
X1	1.0	-0.5	-0.5
X2	-0.5	1.0	-0.5
X3	-0.5	-0.5	1.0

Cannot Include All 3 Variables



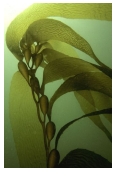
- This matrix is singular

	X1	X2	X3
X1	1.0	-0.5	-0.5
X2	-0.5	1.0	-0.5
X3	-0.5	-0.5	1.0

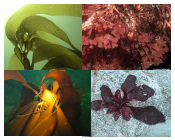
- If you know x1 and x2, you know the state of x3

Coefficient judged relative to effect of missing variable

Does Diet Affect Urchin Gonad Development?



Macrocystis

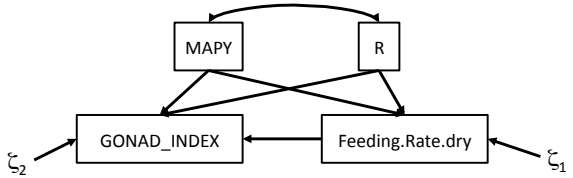


Mixture

Rhodymenia

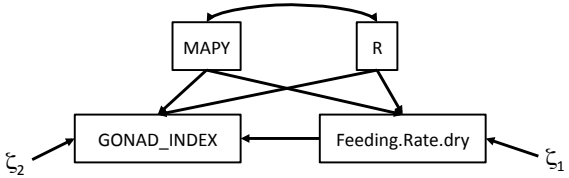
- Urchins feeding measured over 6 months
- Gonads and body size assessed at end
- All consumption rates converted to g dry carbon

Urchin Gonad Development Model



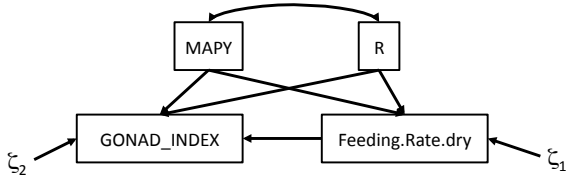
- Note that the polyculture is not included.
- Results judged relative to polyculture.

How do we use a categorical variable?



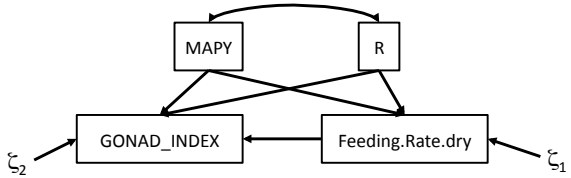

```
> urchinData<-read.csv("./urchin_ex_sem.csv")

> summary(urchinData)
```

How do we use a categorical variable?



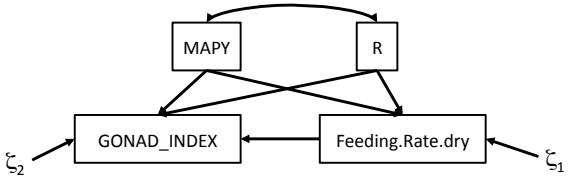

```
   Box      treatment
Min.   : 1   MAPY:7
1st Qu.:10   POLY:7
Median :18   R   :7
Mean   :18
3rd Qu.:26
Max.   :35
```

Script to Turn Make Variables Binary



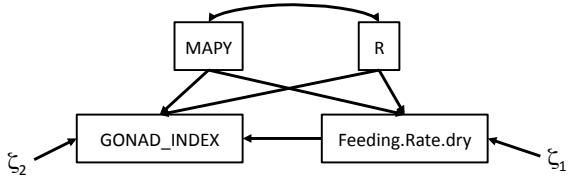

```
#Make treatment into a series of binary variables
source("./makeBinaryTreatments.R")

binTrt<-makeBinaryTreatments(urchinData,
  "treatment")

head(binTrt)
```

Script to Turn Make Variables Binary



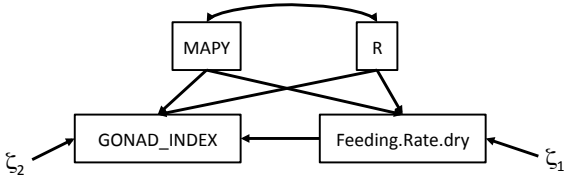

```
  MAPY POLY R
1    1    0 0
2    0    0 1
3    0    1 0
4    1    0 0
5    0    0 1
6    0    1 0
```

Cannot Use All 3 Variables



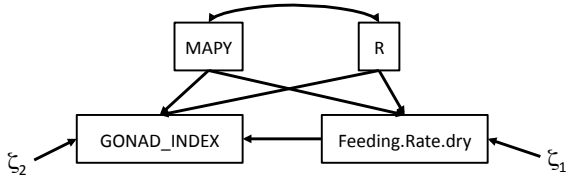

```
> cor(binTrt)
     MAPY POLY    R
MAPY  1.0 -0.5 -0.5
POLY -0.5  1.0 -0.5
R    -0.5 -0.5  1.0

> solve(cor(binTrt))
Error in solve.default(cor(binTrt)) :
  Lapack routine dgesv: system is exactly singular
```

Cannot Use All 3 Variables



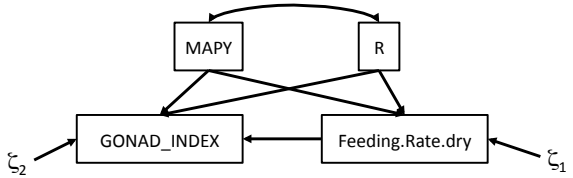

```
#add new columns to data frame
urchinData<-cbind(urchinData, binTrt)

urchinModel<-'
  Feeding.rate.dry ~ MAPY + R
  GONAD_INDEX ~ MAPY + R + Feeding.rate.dry
'

urchinSEM<-sem(urchinModel, data=urchinData)
```

Fitting the Model



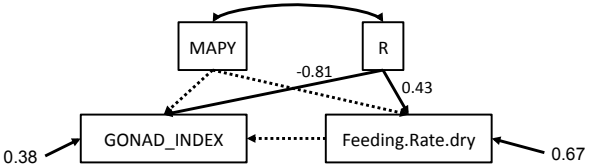

```
#add new columns to data frame
urchinData<-cbind(urchinData, binTrt)

urchinModel<-'
   Feeding.rate.dry ~ MAPY + R
   GONAD_INDEX ~ MAPY + R + Feeding.rate.dry
'

urchinSEM<-sem(urchinModel, data=urchinData)
```

Fitting the Model



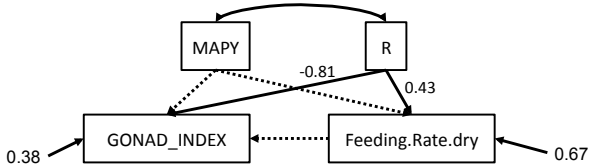

```
lavaan (0.4-12) converged normally after 68 iterations

                                                  Used       Total
  Number of observations                            20          21

  Estimator                                         ML
  Minimum Function Chi-square                    0.000
  Degrees of freedom                                 0
  P-value                                        1.000
```

The Fit



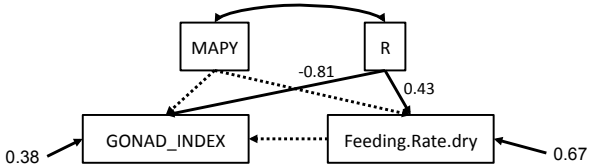

```
                   Estimate  Std.err  Z-value  P(>|z|)   Std.lv  Std.all
Regressions:
  Feeding.rate.dry ~
    MAPY             -0.001    0.001   -1.083    0.279   -0.001   -0.229
    R                 0.002    0.001    2.013    0.044    0.002    0.425
  GONAD_INDEX ~
    MAPY             -0.009    0.008   -1.038    0.299   -0.009   -0.171
    R                -0.041    0.009   -4.644    0.000   -0.041   -0.814
    Feedng.rt.dry    -1.027    2.218   -0.463    0.643   -1.027   -0.078
```

Interpretation



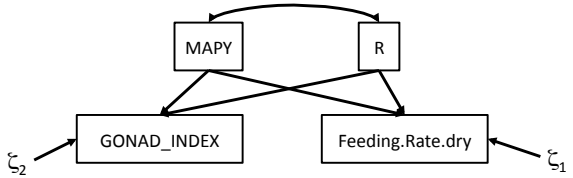
1. *Rhodymenia* is not good food.
   - Urchins eat more, but produce less gonad
2. Performance is similar with *Macrocystis* or Mixture diet

Exercise



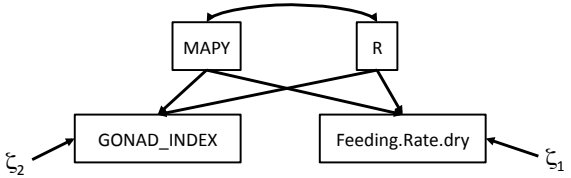
1. But what if coefficient changes when the feeding rate -> gonad link was dropped?
2. Extra: try test growth

Exercise



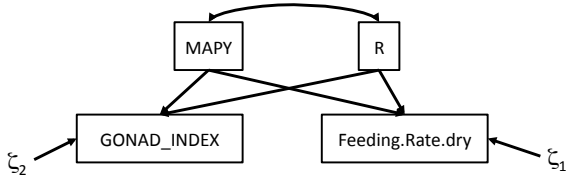

```
urchinModel2<-'
   Feeding.rate.dry ~ MAPY + R
   GONAD_INDEX ~ MAPY + R
'
urchinSEM2<-sem(urchinModel2, data=urchinData)
```

Exercise



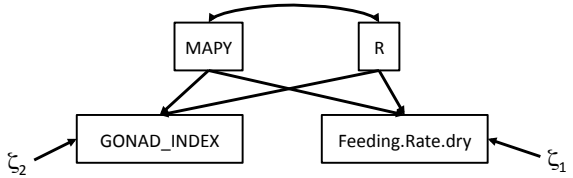

```
> urchinSEM2
lavaan (0.5-12) converged normally after  73 iterations

                                                  Used       Total
  Number of observations                            20          21

  Estimator                                         ML
  Minimum Function Test Statistic                0.000
  Degrees of freedom                                 0
  P-value (Chi-square)                           1.000
```

Exercise



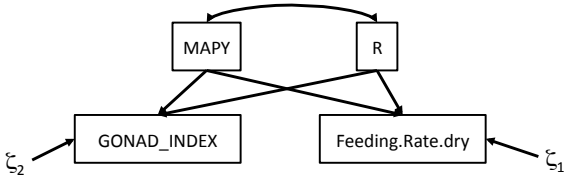

```
urchinSEM2<-lavaan(urchinModel2,
  data=urchinData, auto.cov.y =
  FALSE, auto.var=TRUE)
```

Exercise




```
urchinModel2a<-'
  Feeding.rate.dry ~ MAPY + R
  GONAD_INDEX ~ MAPY + R
  GONAD_INDEX ~~ 0*Feeding.rate.dry
'
```

Exercise

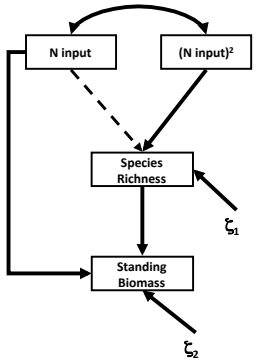



```
> anova(urchinSEM, urchinSEM2a)
Chi Square Difference Test

           Df     AIC     BIC  Chisq Chisq diff Df diff Pr(>Chisq)
urchinSEM   0 -254.32 -247.35 0.0000
urchinSEM2a 1 -256.10 -250.12 0.2245    0.22447       1     0.6357
```

Old Wine in a New Bottle

1. ANOVA and ANCOVA in an SEM context
2. Multiple categorical predictors
3. Nonlinear effects

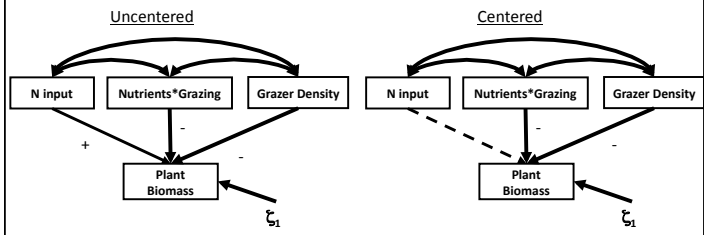
Nonlinearities in Observed Variable SEM



- Nonlinearities are just another variable
- But, nonlinearities may be collinear with their predictor
- Incorporate collinearities into path structure (simple for exogenous variables)
- If necessary (r>0.9), consider centering variables before transforming
- However, best solution for nonlinearities is a larger sample size!

Cardinale et al. 2009 Ecology

Careful in Interpreting Results of Centering

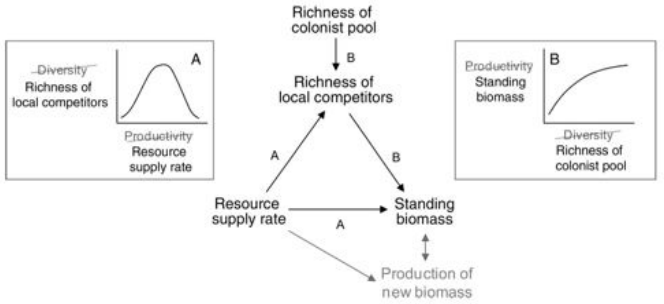


• In uncentered model, additive paths estimate the effect of one variable in the absence of the other.

• In centered model, additive paths estimate the effect of one variable at the average level of the other.

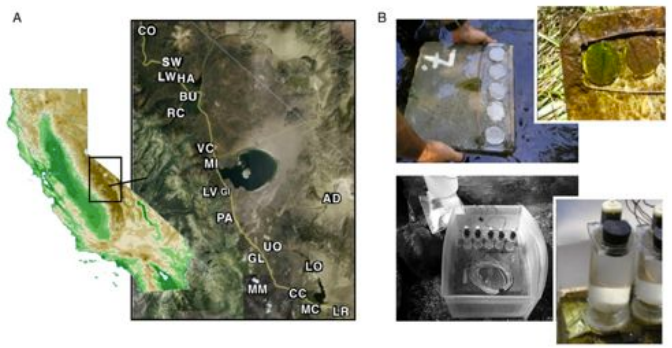
• E.g., nutrients have an effect, but only when grazers are absent.

Does Diversity = Productivity of vice-versa?



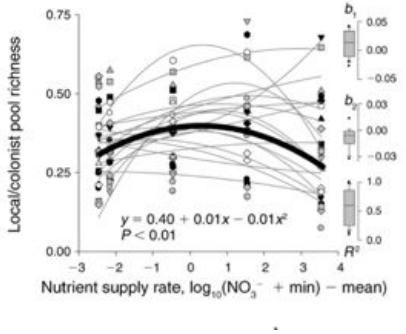
Cardinale et al 2009

A Multi-Stream Experiment



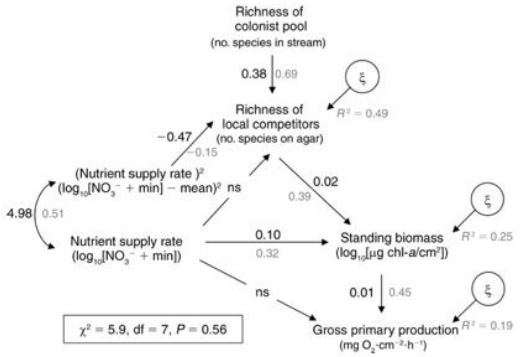
Cardinale et al 2009

Nonlinear Relationship Between Nutrient Addition and Richness



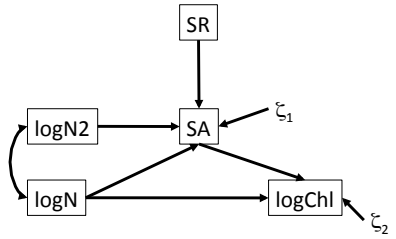
Cardinale et al 2009

Nonlinear Nutrient Effect on Richness



Cardinale et al 2009

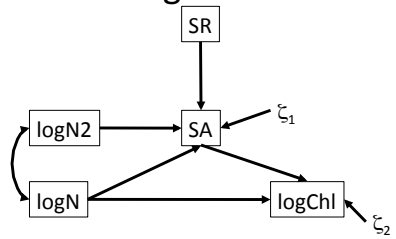
Create a Nonlinear Variable




```
#read in the data
cards<-read.table("./cardainel_et_al_2009.csv")

#make a new nonlinear column
cards$logN2 <- cards$logN^2
```

Cardinale et al 2009

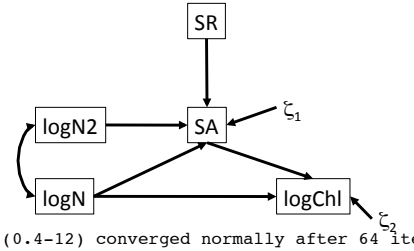
Note that Treatment's Don't Covary with Regional Richness




```
cardModel<-'
   SA ~ logN + logNcen2 + SR
   logChl ~ SA + logN
   SR ~~ 0*logN + 0*logNcen2
   logN ~~ logNcen2
'
cardFit <- sem(cardModel, data=cards, fixed.x=F)
```

Cardinale et al 2009

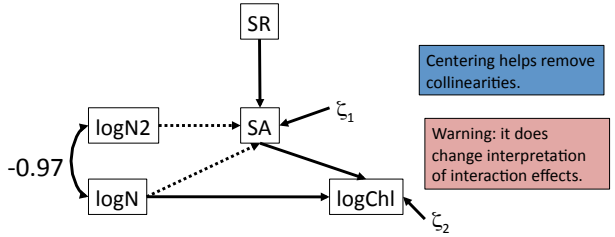
Model Fits Quite Well




```
lavaan (0.4-12) converged normally after 64 iterations

  Number of observations                           127

  Estimator                                         ML
  Minimum Function Chi-square                    0.545
  Degrees of freedom                                 4
  P-value                                        0.969
```

Cardinale et al 2009

But…no Nutrient Effect?



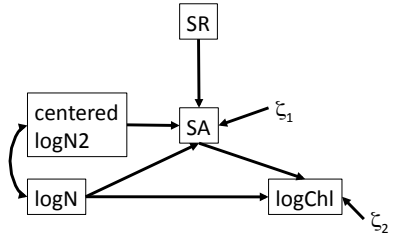
Centering helps remove collinearities.

Warning: it does change interpretation of interaction effects.

	Estimate	Std.err	Z-value	P(>|z|)	Std.lv	Std.all
Regressions:						
SA ~						
logN	-2.996	2.960	-1.012	0.311	-2.996	-0.504
logN2	-0.475	0.456	-1.041	0.298	-0.475	-0.506
SR	0.384	0.035	10.859	0.000	0.384	0.688
logChl ~						
SA	0.020	0.005	3.768	0.000	0.020	0.393
logN	0.097	0.023	4.194	0.000	0.097	0.321

Cardinale et al 2009

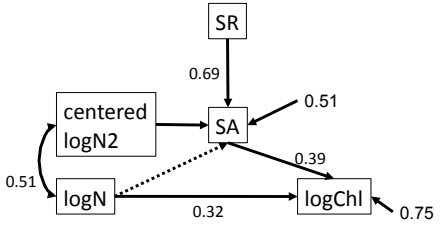
Refit with Centered Nutrients




```
#make a new nonlinear column
cards$logN2Cen <- (cards$logN-mean(cards$logN))^2
```

Cardinale et al 2009

Refit with Centered Nutrients



	Estimate	Std.err	Z-value	P(>|z|)	Std.lv	Std.all
Regressions:						
SA ~						
logN	0.368	0.452	0.815	0.415	0.368	0.062
logN2Cen	-0.475	0.240	-1.974	0.048	-0.475	-0.147
SR	0.384	0.035	10.859	0.000	0.384	0.688
logChl ~						
SA	0.020	0.004	4.667	0.000	0.020	0.393
logN	0.097	0.023	4.194	0.000	0.097	0.321

Cardinale et al 2009

Questions?