Latent Variable Modeling using **R**

A. Alexander Beaujean

2015 Summer Quantitative Methods Series

June 2015

Portland State University

R Exercise: Data Entry

Talk Outline

Talk Outline

Data Entry Exercises

Exercise 1

- In **R**, create a dataset called `prac1` with the following data

10 25 63 54 78 41 33

```
prac1 <- c(10,25, 63, 54, 78, 41, 33)
prac1
```

```
## [1] 10 25 63 54 78 41 33
```

Data Entry Exercises

Exercise 1

- In `prac1`, change the 4th observation from 54 to 45

```
prac1[4] <- 45
prac1
```

```
## [1] 10 25 63 45 78 41 33
```

Talk Outline

Data Entry Exercises

- Exercise 1
- Exercise 2
- Exercise 3
- Exercise 4

Data Entry Exercises

Exercise 2

- In **R**, create a data frame called `prac2` with the following data

Last Name	Salary (in $10,000)
Franklin	8.0
Kahn	6.0
Redding	7.8
Cooke	4.1
Pickett	6.3

Data Entry Exercises

Exercise 2

```r
lname <- c("Franklin", "Kahn", "Redding", "Cooke", "Pickett")
salary <- c(8.0, 6.0, 7.8, 4.1, 6.3)
prac2 <- data.frame(lname, salary)
prac2
```

```
##      lname salary
## 1 Franklin    8.0
## 2     Kahn    6.0
## 3  Redding    7.8
## 4    Cooke    4.1
## 5  Pickett    6.3
```

Data Entry Exercises

Exercise 2

- Convert *Salary* to \$1 units (e.g. 3.2 $\rightarrow$ 32000)

```
# remember to save the transformed variable in the dataset
prac2$salary10 <- prac2$salary*10000
prac2
```

```
##      lname salary salary10
## 1 Franklin    8.0    80000
## 2     Kahn    6.0    60000
## 3  Redding    7.8    78000
## 4    Cooke    4.1    41000
## 5  Pickett    6.3    63000
```

Talk Outline

Data Entry Exercises

Exercise 3

- On website, there is a dataset named `SampleData.txt`. (http://blogs.baylor.edu/rlatentvariable/sample-page/courses/)
    - Download the file to your computer
- Import the data into **R** and print out the first 2 observation values on the console

```
x <- read.table(file="SampleData.txt",header=TRUE)
head(x, 2)
```

```
##   Age  IQ Height Sex
## 1  18 100     65   F
## 2  21 110     68   M
```

Talk Outline

Data Entry Exercises

- Exercise 1
- Exercise 2
- Exercise 3
- Exercise 4

Data Entry Exercises

Exercise 4

- The psych package for **R** has a dataset named *peas*, which has the diameter of parent and child for 700 sweet peas

```
##     parent child
## 4       21    15
## 150     20    18
## 300     19    22
```

1. Download the psych package (if it is not already downloaded)
2. Load the psych package
3. Load the peas dataset

```
install.packages("psych", dependencies=TRUE)
library(psych)
data(peas)
```

Data Entry Exercises

Exercise 4

- In the *peas* dataset

1. Find mean parent diameter

```r
mean(peas$parent)
```

```
## [1] 18
```

2. Find the correlation (`cor()`) between parent and child diameter

```r
cor(peas$parent,peas$child)
```

```
## [1] 0.35
```

R Exercise: Path Models

Talk Outline

Path Model Example

Path Model Example

- From Page and Keith (1981)
- n = 18,058 from *High School and Beyond* 1980 data
- Question of Interest: Relationship between school type and school academic achievement

Path Model Example

	Race	SES	CogAbil	PrivSch	AcadAch	Homework
Race	1.00	0.18	0.23	0.11	0.20	-0.00
SES	0.18	1.00	0.33	0.24	0.36	0.23
CogAbil	0.23	0.33	1.00	0.18	0.72	0.28
PrivSch	0.11	0.24	0.18	1.00	0.18	0.21
AcadAch	0.20	0.36	0.72	0.18	1.00	0.30
Homework	-0.00	0.23	0.28	0.21	0.30	1.00

`Race`: White/Hispanic=1, Black=0

`CogAbil`: composite of verbal and nonverbal subtests

`PrivSch`: Private School=1, Public=0

`AcadAch`: composite of Reading Math subtests

`Homework`: Time spent on homework (not used)

Path Model Example

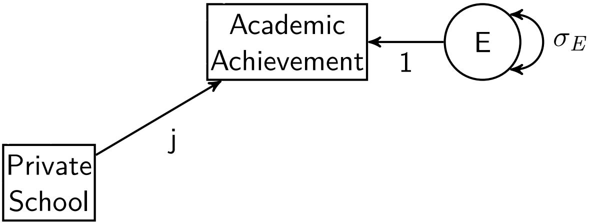
	Race	SES	CogAbil	PrivSch	AcadAch	Homework
Race	1.00	0.18	0.23	0.11	0.20	-0.00
SES	0.18	1.00	0.33	0.24	0.36	0.23
CogAbil	0.23	0.33	1.00	0.18	0.72	0.28
PrivSch	0.11	0.24	0.18	1.00	0.18	0.21
AcadAch	0.20	0.36	0.72	0.18	1.00	0.30
Homework	-0.00	0.23	0.28	0.21	0.30	1.00

- Enter data into **R**
- Name variables

Path Model Example

```r
# load lavaan
library(lavaan)
# enter upper matrix of correlation values
privSchool.cor <- c(1.000, .178, .230, .106, .195, -.001, 1.000, .327, .245, .356,
        .229, 1.000, .183, .721, .278, 1.000, .178, .211, 1.000, .304, 1.000)
# make full matrix
privSchool.cor <- lav_matrix_upper2full(privSchool.cor)
# name variables
dimnames(privSchool.cor) <- list(c("Race", "SES", "CogAbil", "PrivSch", "AcadAch",
"Homework"),c("Race", "SES", "CogAbil", "PrivSch", "AcadAch", "Homework"))
```

Path Model Example



Initial Path Model

- Specify simple regression in **R**
- Fit model to data

Path Model Example

```
initial.model <- '
AcadAch ~ j*PrivSch
'
```

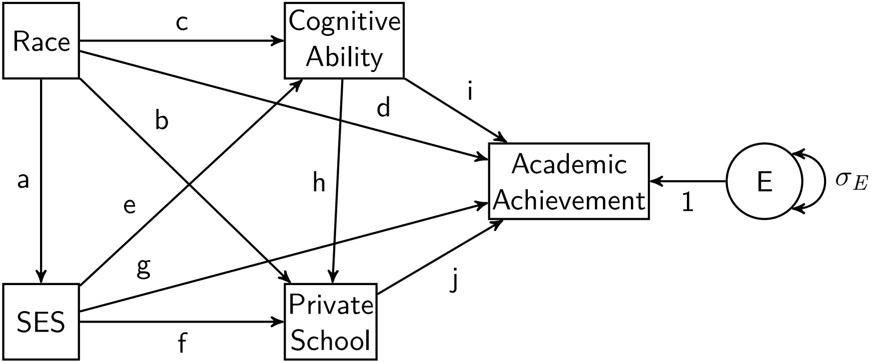
Path Model Example

```r
initial.fit <- sem(initial.model, sample.cov=privSchool.cor, sample.nobs=18058)
parameterEstimates(initial.fit)
```

```
##       lhs op     rhs label  est    se  z pvalue ci.lower ci.upper
## 1 AcadAch  ~ PrivSch     j 0.18 0.007 24      0     0.16     0.19
## 2 AcadAch ~~ AcadAch       0.97 0.010 95      0     0.95     0.99
## 3 PrivSch ~~ PrivSch       1.00 0.000 NA     NA     1.00     1.00
```

- `parameterEstimates()` is an alternative to using `summary()`
  - Returns path estimates, standard errors, and 95% confidence intervals

Path Model Example



Full Path Model

- Specify and fit full model in **R**

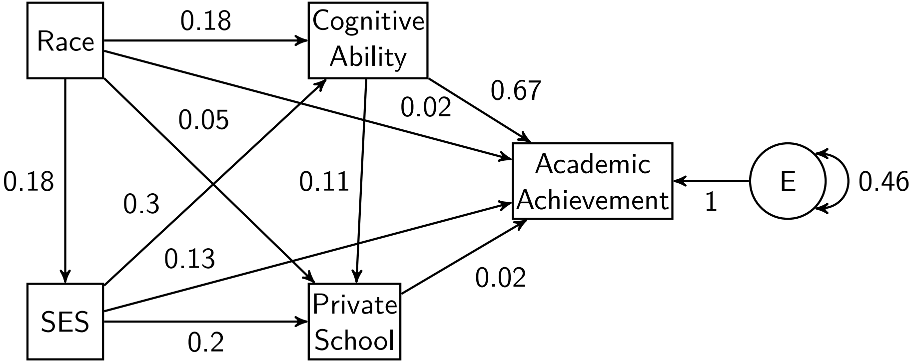
Path Model Example

```
full.model <- '
AcadAch ~ j*PrivSch + g*SES + d*Race + i*CogAbil
PrivSch ~ f*SES + b*Race + h*CogAbil
CogAbil ~ e*SES + c*Race
SES ~ a*Race

# error variance
AcadAch~~ev*AcadAch
'
full.fit <- sem(full.model, sample.cov=privSchool.cor, sample.nobs=18058)
```

Path Model Example

```
parameterEstimates(full.fit)

##        lhs op     rhs label   est    se     z pvalue ci.lower ci.upper
## 1  AcadAch  ~ PrivSch     j 0.022 0.005   4.2  0.000    0.012    0.032
## 2  AcadAch  ~     SES     g 0.128 0.005  23.3  0.000    0.118    0.139
## 3  AcadAch  ~    Race     d 0.015 0.005   2.9  0.003    0.005    0.026
## 4  AcadAch  ~ CogAbil     i 0.671 0.005 122.5  0.000    0.661    0.682
## 5  PrivSch  ~     SES     f 0.202 0.008  26.5  0.000    0.187    0.217
## 6  PrivSch  ~    Race     b 0.046 0.007   6.2  0.000    0.031    0.060
## 7  PrivSch  ~ CogAbil     h 0.106 0.008  13.8  0.000    0.091    0.122
## 8  CogAbil  ~     SES     e 0.295 0.007  42.1  0.000    0.282    0.309
## 9  CogAbil  ~    Race     c 0.177 0.007  25.3  0.000    0.164    0.191
## 10     SES  ~    Race     a 0.178 0.007  24.3  0.000    0.164    0.192
## 11 AcadAch ~~ AcadAch    ev 0.463 0.005  95.0  0.000    0.454    0.473
## 12 PrivSch ~~ PrivSch       0.926 0.010  95.0  0.000    0.907    0.945
## 13 CogAbil ~~ CogAbil       0.863 0.009  95.0  0.000    0.845    0.880
## 14     SES ~~     SES       0.968 0.010  95.0  0.000    0.948    0.988
## 15    Race ~~    Race       1.000 0.000    NA     NA    1.000    1.000
```

Path Model Example



Full Path Model

R Exercise: Simple Latent Variable Models

Talk Outline

Simple Latent Variable Model Example

Simple Latent Variable Model Example

- Example from Little, Slegers, and Card (2006)
- Enter the correlations and standard deviations ($n = 380$)
- Fit the path model using the standardized LV and marker variable methods to scale the the LV
  - For the marker variable method, use indicators 1 and 5 as the marker variables

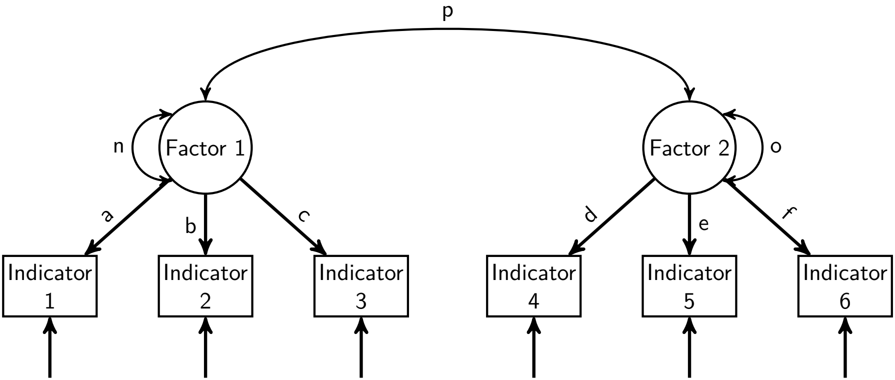
Simple Latent Variable Model Example

	I1	I2	I3	I4	I5	I6
Indicator 1	1.000					
Indicator 2	0.759	1.000				
Indicator 3	0.762	0.787	1.000			
Indicator 4	0.028	0.010	−0.058	1.000		
Indicator 5	−0.061	−0.061	−0.141	0.785	1.000	
Indicator 6	−0.022	−0.052	−0.102	0.816	0.816	1.000
SD	0.668	0.685	0.707	0.714	0.663	0.653

Simple Latent Variable Model Example

```r
library(lavaan)
# Little et al correlations
little.cor <- c(1.00000, 0.759, 1.000, 0.762, 0.787, 1.000, 0.0278, 0.010, -0.058,
        1.000, -0.061, -0.061,  -0.141, 0.785, 1.000, -0.022, -0.052, -0.102, 0.816,
        0.816, 1.000)
# make full matrix
little.cor <- lav_matrix_lower2full(little.cor)
# name columns and rows in correlation matrix
rownames(little.cor) <- colnames(little.cor) <- paste("I",seq(1,6),sep="")
# SDs
little.sd <- c(0.6678, 0.685, 0.707, 0.714, 0.663, 0.653)
# create covariance matrix
little.cov <- cor2cov(little.cor, little.sd)
```

Simple Latent Variable Model Example



Simple Latent Variable Model Example

```
# marker variable
little.marker.model <- '
# factor model
F1 =~ a*I1 + b*I2 + c*I3
F2 =~ e*I5 + d*I4 + f*I6

# label latent (co)variances
F1~~n*F1
F2~~o*F2
F1~~p*F2
'
## fit model
little.marker.fit <- cfa(little.marker.model, sample.cov=little.cov, sample.nobs=380)
```

```r
summary(little.marker.fit)
```

```
## lavaan (0.5-18) converged normally after  31 iterations
##
##   Number of observations                           380
##
##   Estimator                                         ML
##   Minimum Function Test Statistic               17.330
##   Degrees of freedom                                 8
##   P-value (Chi-square)                           0.027
##
## Parameter estimates:
##
##   Information                                 Expected
##   Standard Errors                             Standard
##
##                    Estimate  Std.err  Z-value  P(>|z|)
## Latent variables:
##   F1 =~
##     I1        (a)     1.000
##     I2        (b)     1.059    0.049   21.688    0.000
##     I3        (c)     1.100    0.050   21.817    0.000
##   F2 =~
##     I5        (e)     1.000
##     I4        (d)     1.076    0.044   24.252    0.000
##     I6        (f)     1.023    0.040   25.698    0.000
##
```

```
## Covariances:
##   F1 ~~
##     F2        (p)    -0.024    0.019   -1.252    0.211
##
## Variances:
##     F1        (n)     0.326    0.032
##     F2        (o)     0.344    0.032
##     I1                0.118    0.012
##     I2                0.102    0.012
##     I3                0.104    0.013
##     I5                0.094    0.010
##     I4                0.110    0.012
##     I6                0.064    0.009
```

Simple Latent Variable Model Example

```r
# standardized LV
little.standardized.model <- '
# factor model
F1 =~ NA*I1 + a*I1 + b*I2 + c*I3
F2 =~ NA*I4 + d*I4 + e*I5 + f*I6

# fix latent variances
F1~~1*F1
F2~~1*F2

# label latent covariance
F1~~p*F2
'
## fit model
little.standardized.fit <- cfa(little.standardized.model, sample.cov=little.cov,
        sample.nobs=380)
```

```r
summary(little.standardized.fit)
```

```
## lavaan (0.5-18) converged normally after  34 iterations
##
##   Number of observations                           380
##
##   Estimator                                         ML
##   Minimum Function Test Statistic               17.330
##   Degrees of freedom                                 8
##   P-value (Chi-square)                           0.027
##
## Parameter estimates:
##
##   Information                                 Expected
##   Standard Errors                             Standard
##
##                    Estimate  Std.err  Z-value  P(>|z|)
## Latent variables:
##   F1 =~
##     I1        (a)     0.571    0.028   20.216    0.000
##     I2        (b)     0.605    0.029   21.213    0.000
##     I3        (c)     0.628    0.029   21.403    0.000
##   F2 =~
##     I4        (d)     0.631    0.029   21.542    0.000
##     I5        (e)     0.587    0.027   21.582    0.000
##     I6        (f)     0.601    0.026   22.945    0.000
##
```

```
## Covariances:
##   F1 ~~
##     F2        (p)    -0.070    0.056   -1.262    0.207
##
## Variances:
##     F1                 1.000
##     F2                 1.000
##     I1                 0.118    0.012
##     I2                 0.102    0.012
##     I3                 0.104    0.013
##     I4                 0.110    0.012
##     I5                 0.094    0.010
##     I6                 0.064    0.009
```

R Exercise: Latent Variable Models with Multiple Groups in R

Talk Outline

Multiple Groups Example

Multiple Groups Example

- Example from Little et al. (2006), now with data from two groups
  - Group 1 $n = 380$; Group 2 $n = 379$
- Enter the correlations and standard deviations
- Examine configural, metric, and scalar invariance using the marker variable method to scale the the LV
  - For the marker variable method, use indicators 1 and 5 as the marker variables

Multiple Groups Example

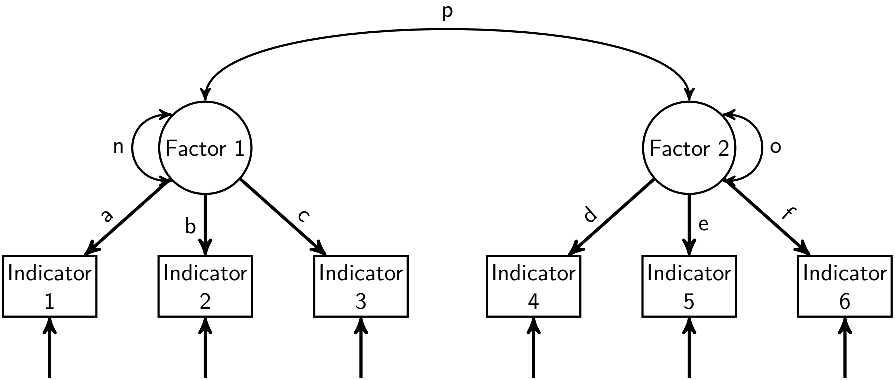
	I1	I2	I3	I4	I5	I6
Indicator 1	1.000	0.813	0.850	−0.188	−0.289	−0.293
Indicator 2	0.759	1.000	0.835	−0.155	−0.250	−0.210
Indicator 3	0.762	0.787	1.000	−0.215	−0.338	−0.306
Indicator 4	0.028	0.010	−0.058	1.000	0.784	0.800
Indicator 5	−0.061	−0.061	−0.141	0.785	1.000	0.832
Indicator 6	−0.022	−0.052	−0.102	0.816	0.816	1.000
Group One SD	0.668	0.685	0.707	0.714	0.663	0.653
Group One Mean	3.135	2.991	3.069	1.701	1.527	1.545
Group Two SD	0.703	0.718	0.762	0.650	0.602	0.614
Group Two Mean	3.073	2.847	2.979	1.717	1.580	1.550

Correlations for group one are given in the lower diagonal, and those for group two are given in the upper diagonal.

Multiple Groups Example

```r
library(lavaan)
# group 1
group1.cor <- c(1.00000, 0.759, 1.000, 0.762, 0.787, 1.000, 0.0278, 0.010, -0.058,
        1.000, -0.061, -0.061,  -0.141, 0.785, 1.000, -0.022, -0.052, -0.102, 0.816,
        0.816, 1.000)
group1.cor <- lav_matrix_lower2full(group1.cor)
rownames(group1.cor) <- colnames(group1.cor) <- paste("I",seq(1,6),sep="")
group1.sd <- c(0.6678, 0.685, 0.707, 0.714, 0.663, 0.653)
group1.cov <- cor2cov(group1.cor, group1.sd)
group1.mean <- c(3.135, 2.991, 3.069, 1.701, 1.527, 1.545)
# group 2
group2.cor <- c(1.000, 0.813, 1.000, 0.850, 0.835, 1.000, -0.188, -0.155, -0.215,
        1.000, -0.289, -0.250, -0.338, 0.784, 1.000, -0.293, -0.210, -0.306, 0.800,
        0.832, 1.000)
group2.cor <- lav_matrix_lower2full(group2.cor)
rownames(group2.cor) <- colnames(group2.cor) <- paste("I",seq(1,6),sep="")
group2.sd <- c(0.703, 0.718, 0.762, 0.650, 0.602, 0.614)
group2.cov <- cor2cov(group2.cor, group2.sd)
group2.mean <- c(3.073, 2.847, 2.979, 1.717, 1.580, 1.550)
# combine the data
little.cov <- list(group1=group1.cov, group2=group2.cov)
little.mean <- list(group1=group1.mean, group2=group2.mean)
little.n <- list(group1=380, group2=379)
```

Multiple Groups Example



Multiple Groups Example

```
# marker variable
little.marker.model <- '
# factor model
F1 =~ a*I1 + b*I2 + c*I3
F2 =~ e*I5 + d*I4 + f*I6
'

## configural
little.configural.fit <- cfa(little.marker.model, sample.cov=little.cov,
        sample.mean=little.mean, sample.nobs=little.n)
## metric
little.metric.fit <- cfa(little.marker.model, sample.cov=little.cov,
        sample.mean=little.mean, sample.nobs=little.n,group.equal=c("loadings"))
## scalar
little.scalar.fit <- cfa(little.marker.model, sample.cov=little.cov,
        sample.mean=little.mean, sample.nobs=little.n,
        group.equal=c("loadings", "intercepts"))
```

Multiple Groups Example

```
# examine model fit
# selected fit measures
fit.indices <- c("chisq", "df", "cfi", "rmsea", "srmr", "mfi")

fitMeasures(little.configural.fit, fit.indices)

##  chisq     df    cfi  rmsea   srmr    mfi
## 46.224 16.000  0.992  0.071  0.033  0.980

fitMeasures(little.metric.fit, fit.indices)

##  chisq     df    cfi  rmsea   srmr    mfi
## 49.014 20.000  0.992  0.062  0.036  0.981

fitMeasures(little.scalar.fit, fit.indices)

##  chisq     df    cfi  rmsea   srmr    mfi
## 58.786 24.000  0.990  0.062  0.033  0.977
```

R Exercise: Sample Size Determination

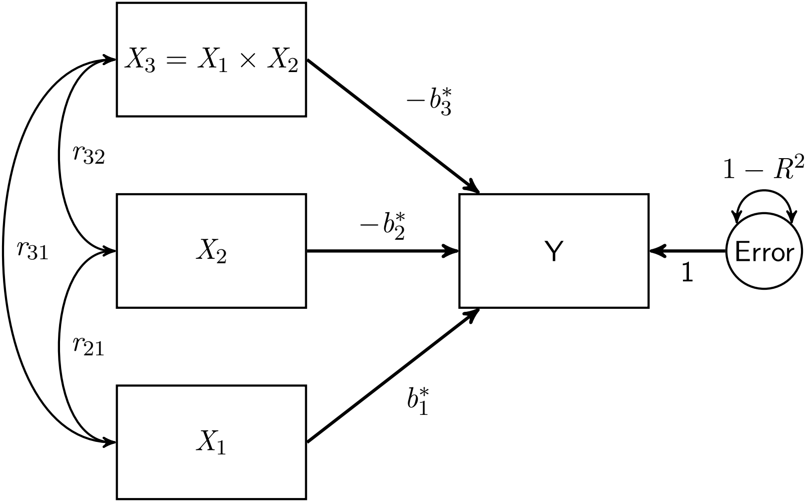
Talk Outline

Sample Size Determination Example

Sample Size Determination Example

- Is there an interaction between the antisocial behavior and hyperactivity in predicting academic achievement?

Sample Size Determination Example



Sample Size Determination Example

- Review the literature and determine that:
  - The main effects should be approximately twice the size of the interaction
  - The correlation between antisocial behavior and hyperactivity is ≈ 0.3
  - After controlling for hyperactivity
    - Antisocial behavior is positively related to academic achievement
    - Hyperactivity is negatively related to academic achievement
  - The interaction has a negative relationship to academic achievement

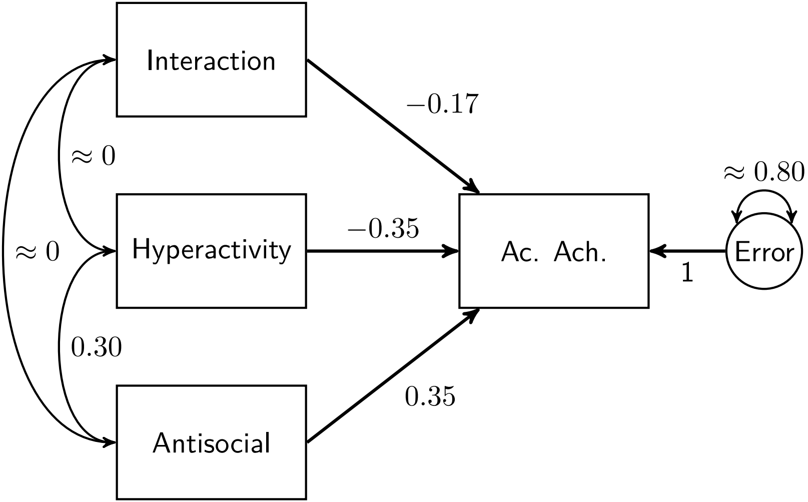
Sample Size Determination Example

- Say we have the following data

$$b_{YX} = \begin{array}{l} \text{anti} = \\ \text{hyper} = \\ \text{int} = \end{array} \begin{bmatrix} 0.35 \\ -0.35 \\ -0.17 \end{bmatrix} \text{ and } R_{YX} = V_{YX} = \begin{bmatrix} 1.00 & 0.30 & 0.00 \\ 0.30 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$$

$$R^2 = 0.2$$

Sample Size Determination Example



- Specify population and analysis models using *lavaan* syntax

Sample Size Determination Example

```
# population model
antiPop.model<-'
academ ~  0.35*anti + -0.35*hyp + -0.17*int
anti~~0.3*hyp
int ~~ 0*anti + 0*hyp
academ~~0.80*academ
'
# analysis model
antiReg.model <- '
academ ~  anti + hyp + int
int ~~ 0*anti + 0*hyp
'
```

Sample Size Determination Example

- Double check your model
  - "Fit" population model in *lavaan*
    - Hint: Use the `fixed.x=FALSE` argument in the `sem()` function
  - Calculate implied correlations
  - Compare implied correlations to hypothesized values

```r
library(lavaan)
# fit model with population moments
antiPop.fit <- sem(antiPop.model,fixed.x=FALSE)
summary(antiPop.fit, std=TRUE, rsquare=TRUE)
```

```
## ** WARNING ** lavaan (0.5-18) model has NOT been fitted
## ** WARNING ** Estimates below are simply the starting values
##
##   Number of observations                             0
##
##   Estimator                                         ML
##
## Parameter estimates:
##
##   Information                                 Expected
##   Standard Errors                                 None
##
##                    Estimate  Std.err  Z-value  P(>|z|)   Std.lv  Std.nox
## Regressions:
##   academ ~
##     anti              0.350                                0.350    0.350
##     hyp              -0.350                               -0.350   -0.350
##     int              -0.170                               -0.170   -0.170
##
## Covariances:
##   anti ~~
##     hyp               0.300                                0.300    0.300
```

```
##     int                0.000                                    0.000    0.000
##   hyp ~~
##     int                0.000                                    0.000    0.000
##
## Variances:
##     academem           0.800                                    0.800    0.800
##     anti               1.000                                    1.000    1.000
##     hyp                1.000                                    1.000    1.000
##     int                1.000                                    1.000    1.000
##
## R-Square:
##
##     academem           0.200
```

```r
# implied correlations
fitted(antiPop.fit)
```

```
## $cov
##        academ anti  hyp   int
## academ  1.00
## anti    0.24   1.00
## hyp    -0.24   0.30  1.00
## int    -0.17   0.00  0.00  1.00
##
## $mean
## academ   anti    hyp    int
##      0      0      0      0
```

Sample Size Determination Example

- Determine if a sample size of 200 is sufficient to find an interaction effect.

Sample Size Determination Example

```r
library(simsem)
intReg.sim.n200 <- sim(nRep=1000, model=antiReg.model, n=200,
        generate=antiPop.model, lavaanfun = "sem", multicore=TRUE)
antiPop.pwr.n200 <- summaryParam(intReg.sim.n200)
antiPop.pwr.n200["academ~int",]
```

```
##            Estimate.Average Estimate.SD Average.SE Power..Not.equal.0. Std.Est
## academ~int            -0.17       0.065      0.063                0.78   -0.17
##            Std.Est.SD Average.Param Average.Bias Coverage Rel.Bias Std.Bias
## academ~int      0.062         -0.17      -0.0035     0.94    0.021   -0.055
##            Rel.SE.Bias Not.Cover.Below Not.Cover.Above Average.CI.Width
## academ~int      -0.031           0.023           0.039             0.25
##            SD.CI.Width
## academ~int       0.018
```

- Will a sample size of 200 be sufficient to find an interaction effect?
    - From a power perspective, it will not
    - From an AIPE perspective, it might

Sample Size Determination Example

- Determine the sample size for a power of .80
  - Simulate $m = 10$ data sets from $n = 200$ to $n = 300$ increasing by units of 20

Sample Size Determination Example

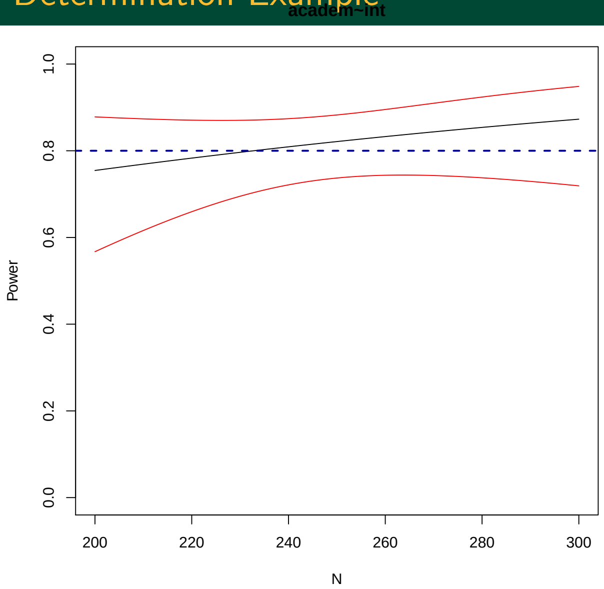
```r
# simulate data
antiPop.n <- sim(nRep = NULL, model=antiReg.model,  n = rep(seq(200,300,20), 10),
generate=antiPop.model, lavaanfun = "sem", multicore=TRUE)
```

```r
# find sample size for desire power
antiPop.pwr.n <- getPower(antiPop.n)
findPower(antiPop.pwr.n, "N", 0.8)
```

```
##    academ~anti     academ~hyp     academ~int academ~~academ     anti~~anti
##            Inf            Inf            233            Inf            Inf
##       hyp~~hyp       int~~int
##            Inf            Inf
```

```r
# plot power curve
plotPower(antiPop.n, powerParam="academ~int")
# line at power = 0.80
abline(h=.8, lwd=2, lty=2, col="blue4")
```

Sample Size Determination Example



Sample Size Determination Example

Advanced

- Equation 1 can estimate R^2

$$R^2 = \rho'_{YX} R_{XX}^{-1} \rho_{YX} = \frac{b'_{YX} V_{XX} b_{YX}}{\sigma_Y^2} \tag{1}$$

where

- ρ_{YX} is the population $k \times 1$ column vector of correlations of each of the *k* predictors with the outcome
- b_{YX} is the population $k \times 1$ column vector of regression coefficients of each of the *k* predictors with the outcome
- R_{XX} and V_{XX} are the $k \times k$ population correlation and covariance matrixes, respectively, of the predictor variables,
- σ_Y^2 is the variance of the outcome, and
- $'$ is the transpose

References

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*, 59-72. doi: 10.1207/s15328007sem1301_3

Page, E. B., & Keith, T. Z. (1981). Effects of U.S. private schools: A technical analysis of two recent claims. *Educational Researcher*, *10*, 7-17. doi: 10.2307/1174256