Psychology 454: Latent Variable Modeling
Types of variables and types of models

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

NORTHWESTERN UNIVERSITY

November, 2012

Outline

Two types of variables, three types of relationships

1. Variables
   1. Observed Variables (X, Y)
   2. Latent Variables (ξ η ϵ ζ)
2. Three kinds of variance/covariances
   1. Observed with Observed C_{xy} or σ_{xy}
   2. Observed with Latent λ
   3. Latent with Latent ϕ
3. Direction
   - Bidirectional (correlation)
   - Directional (regression)

Observed Variables

X	Y
$\boxed{X_1}$	$\boxed{Y_1}$
$\boxed{X_2}$	$\boxed{Y_2}$
$\boxed{X_3}$	$\boxed{Y_3}$
$\boxed{X_4}$	$\boxed{Y_4}$
$\boxed{X_5}$	$\boxed{Y_5}$
$\boxed{X_6}$	$\boxed{Y_6}$

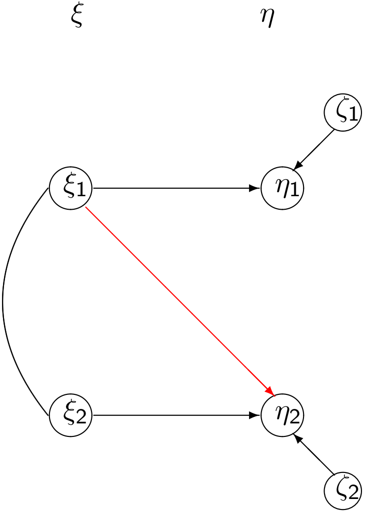
Latent Variables

ξ η

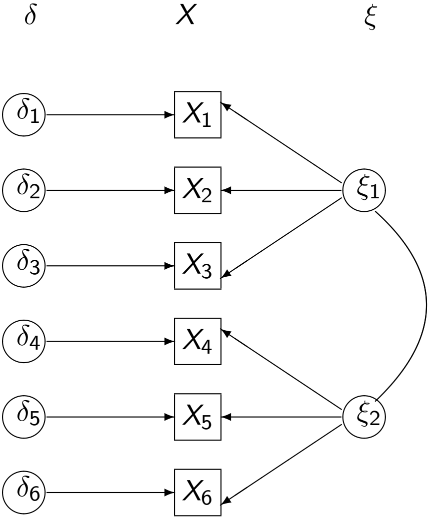
ξ_1 η_1

ξ_2 η_2

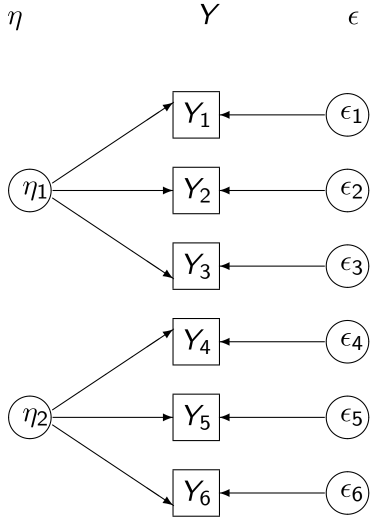
Theory: A regression model of latent variables



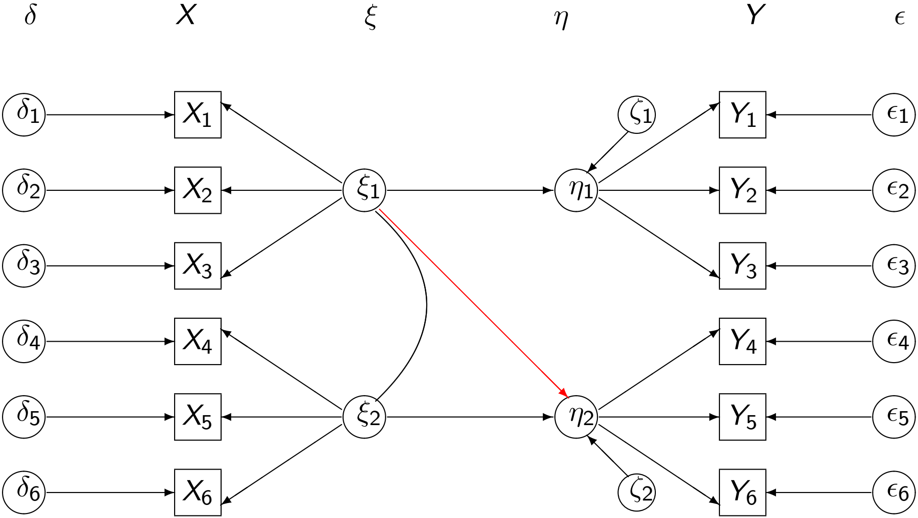
A measurement model for X



A measurement model for Y



A complete structural model



Latent Variable Modeling

1. Requires measuring observed variables
   - Requires defining what is relevant and irrelevant to our theory.
   - Issues in quality of scale information, levels of measurement.
2. Formulating a measurement model of the data: estimating latent constructs
   - Perhaps based upon exploratory and then confirmatory factor analysis, definitely based upon theory.
   - Includes understanding the reliability of the measures.
3. Modeling the structure of the constructs
   - This is a combination of theory and fitting. Do the data fit the theory.
   - Comparison of models. Does one model fit better than alternative models?

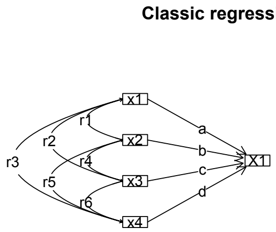
Observed-observed

1. The **?** data set is used in the LISREL manual as an example of regression path models
   - "A sample of boys originally studied as ninth graders in 1969 was recontacted in 1974 to obtain information about high school performance and educational attainment."
   - 767 twelfth grade males
   - A study of background, aspiration and educational attainment
2. Variables
   - Intelligence
   - Number of siblings
   - Fathers Education
   - Father's occupation
   - Grades
   - Educational expectation
   - Occupational aspiration
3. Correlation matrix is available in the sem (**?**) package.

The Kerckhoff correlation matrix (N=767)

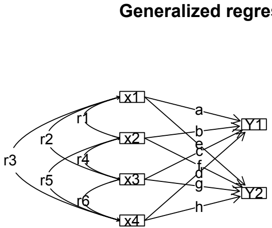
```
> R.kerch
             Intelligence Siblings FatherEd FatherOcc Grades EducExp OccupAsp
Intelligence        1.000   -0.100    0.277     0.250  0.572   0.489    0.335
Siblings           -0.100    1.000   -0.152    -0.108 -0.105  -0.213   -0.153
FatherEd            0.277   -0.152    1.000     0.611  0.294   0.446    0.303
FatherOcc           0.250   -0.108    0.611     1.000  0.248   0.410    0.331
Grades              0.572   -0.105    0.294     0.248  1.000   0.597    0.478
EducExp             0.489   -0.213    0.446     0.410  0.597   1.000    0.651
OccupAsp            0.335   -0.153    0.303     0.331  0.478   0.651    1.000
>
```

The classic regression model



$$\hat{Y} = \beta_x x + \epsilon$$

The generalized regression model

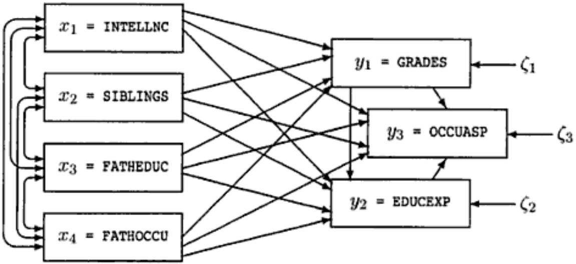


$$\hat{Y} = \beta_x x + \epsilon$$

Conceptual Kerckhoff model

1. Background variables
    - Intelligence
    - Number of siblings
    - Fathers Education
    - Fathers's occupation
2. Intermediate variables
    - Grades
    - Educational expectation
3. Final outcomes
    - Occupational aspiration

The model (From the LISREL manual)



Matrix regression

1. Most regression examples (and functions) use raw data
    - $\hat{Y} = X\beta + \epsilon$
    - $beta = (X'X)^{-1}X'Y$
    - lm(y ~x)
2. Regression is just solving the matrix equation
    - $\beta = R^{-1}r_{xy}$
    - mat.regress(R,x,y) (deprecated)
    - set.cor(y,x,R) (recommended)

Simple regression 4 predictors, 3 criteria

set.cor(y=5:7,x=1:4,data=R.kerch)

```
Call: set.cor(y = 5:7, x = 1:4, data = R.kerch,N.obs)

Multiple Regression from matrix input

Beta weights
             Grades EducExp OccupAsp
Intelligence   0.53    0.37     0.25
Siblings      -0.03   -0.12    -0.09
FatherEd       0.12    0.22     0.10
FatherOcc      0.04    0.17     0.20

Multiple R
  Grades  EducExp OccupAsp
    0.59     0.61     0.44

Multiple R2
  Grades  EducExp OccupAsp
    0.35     0.38     0.19
```

More complicated regression

```
> set.cor(y=6:7,x=1:5,data=R.kerch)

Call: set.cor(y = 6:7, x = 1:5, data = R.kerch)

Multiple Regression from matrix input

Beta weights
             EducExp OccupAsp
Intelligence    0.16     0.05
Siblings       -0.11    -0.08
FatherEd        0.17     0.05
FatherOcc       0.15     0.18
Grades          0.41     0.38

Multiple R
 EducExp OccupAsp
    0.70     0.54

Multiple R2
 EducExp OccupAsp
    0.48     0.29
```

Predicting occupational aspirations from the intermediate set

```
> set.cor(y=7,x=5:6,data=R.kerch)

Call: set.cor(y = 7, x = 5:6, data = R.kerch)

Multiple Regression from matrix input

Beta weights
        OccupAsp
Grades      0.14
EducExp     0.57

Multiple R
OccupAsp
    0.66

Multiple R2
OccupAsp
    0.44
```

Predicting occupational aspirations from the entire set

```
> set.cor(y=7,x=1:6,data=R.kerch)

Call: set.cor(y = 7, x = 1:6, data = R.kerch)

Multiple Regression from matrix input

Beta weights
             OccupAsp
Intelligence    -0.04
Siblings        -0.02
FatherEd        -0.04
FatherOcc        0.10
Grades           0.16
EducExp          0.55

Multiple R
OccupAsp
    0.67

Multiple R2
OccupAsp
    0.44
```

Can use sem functions (in either sem or lavaan) to estimate the mediation model

1. Treat all variables as observed (fixed)
   - Specify a limited number of paths rather than the full regression model
2. *sem* commands
   - either in RAM (path) notation or
   - causal notation
3. Rpkglavaan commands similar to causal notation of sem

Kerckhoff-Kenny path analysis (modified from sem help page) to just predict the DVs)

```
model.kerch <- specifyModel()
    Intelligence -> Grades,        gam51
    Siblings -> Grades,            gam52
    FatherEd -> Grades,            gam53
    FatherOcc -> Grades,           gam54
    Intelligence -> EducExp,       gam61
    Siblings -> EducExp,           gam62
    FatherEd -> EducExp,           gam63
    FatherOcc -> EducExp,          gam64
     Intelligence -> OccupAsp,      gam71
    Siblings -> OccupAsp,          gam72
    FatherEd -> OccupAsp,          gam73
    FatherOcc -> OccupAsp,         gam74
  # Grades -> EducExp,             beta65
   # Grades -> OccupAsp,            beta75
 #   EducExp -> OccupAsp,           beta76

sem.kerch <- sem(model.kerch, R.kerch, 737, fixed.x=c(Intelligence,Siblings,
    FatherEd,FatherOcc))
summary(sem.kerch)
```

Fixed path model – not modeling the DV correlations

```
Model Chisquare =  411.72   Df =  3 Pr(>Chisq) = 6.4133e-89
Chisquare (null model) =  1664.3   Df =  21
Goodness-of-fit index =  0.85747
Adjusted goodness-of-fit index =  -0.33031
RMSEA index =  0.43024   90% CI: (0.39572, 0.46581)
Bentler-Bonnett NFI =  0.75262
Tucker-Lewis NNFI =  -0.74103
Bentler CFI =  0.75128
SRMR =  0.099759
AIC =  441.72
AICc =  412.38
BIC =  510.76
CAIC =  388.91

Normalized Residuals
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.000   0.000   0.000   0.956   0.000  10.100

R-square for Endogenous Variables
 Grades  EducExp OccupAsp
 0.3490   0.3765   0.1930
```

With path coefficients of

```
Parameter Estimates
            Estimate  Std Error z value  Pr(>|z|)
gam51        0.525902 0.031182  16.86530 8.0987e-64 Grades <--- Intelligence
gam52       -0.029942 0.030149  -0.99314 3.2064e-01 Grades <--- Siblings
gam53        0.118966 0.038259   3.10951 1.8740e-03 Grades <--- FatherEd
gam54        0.040603 0.037785   1.07456 2.8257e-01 Grades <--- FatherOcc
gam61        0.373339 0.030517  12.23376 2.0521e-34 EducExp <--- Intelligence
gam62       -0.123910 0.029506  -4.19954 2.6745e-05 EducExp <--- Siblings
gam63        0.220918 0.037442   5.90022 3.6302e-09 EducExp <--- FatherEd
gam64        0.168302 0.036979   4.55125 5.3328e-06 EducExp <--- FatherOcc
gam71        0.248827 0.034718   7.16718 7.6561e-13 OccupAsp <--- Intelligence
gam72       -0.091653 0.033567  -2.73047 6.3245e-03 OccupAsp <--- Siblings
gam73        0.098869 0.042596   2.32109 2.0282e-02 OccupAsp <--- FatherEd
gam74        0.198486 0.042069   4.71809 2.3807e-06 OccupAsp <--- FatherOcc
V[Grades]    0.650995 0.033935  19.18333 5.1010e-82 Grades <--> Grades
V[EducExp]   0.623511 0.032503  19.18333 5.1010e-82 EducExp <--> EducExp
V[OccupAsp]  0.806964 0.042066  19.18333 5.1010e-82 OccupAsp <--> OccupAsp
```

Note, we are not modeling the DV correlations so the residuals will be large

Residuals from this model

```
> lowerMat(resid(sem.kerch))
             Intll Sblng FthrE FthrO Grads EdcEx OccpA
Intelligence 0.00
Siblings     0.00  0.00
FatherEd     0.00  0.00  0.00
FatherOcc    0.00  0.00  0.00  0.00
Grades       0.00  0.00  0.00  0.00  0.00
EducExp      0.00  0.00  0.00  0.00  0.26  0.00
OccupAsp     0.00  0.00  0.00  0.00  0.25  0.38  0.00
```

fixed sem = regression

Compare the coefficients from this sem with the regression β values

```
round(sem.kerch$coeff,3)
     gam51      gam52      gam53      gam54      gam61      gam62      gam63
      gam64      gam71      gam72      gam73    gam74
     0.526     -0.030      0.119      0.041      0.373     -0.124      0.221
     0.168      0.249     -0.092      0.099      0.198
    V[Grades]  V[EducExp] V[OccupAsp]
   0.651      0.624      0.807

 mr.kk <- mat.regress(data=R.kerch,x=c(1:4),y=c(5:7))

 > round(as.vector(mr.kk$beta),3)
[1]  0.526 -0.030  0.119  0.041  0.373 -0.124  0.221
 0.168  0.249 -0.092  0.099  0.198
```

More complicated regression

1. Able to model the intercorrelations of the Y variables
   - This treats the Ys as both predictors and predicted
2. Able to let some of the Ys be part of the regression model

Complete Kerckhoff-Kenny path analysis (taken from sem help page)

```
model.kerch1 <- specifyModel()
    Intelligence -> Grades,       gam51
    Siblings -> Grades,           gam52
    FatherEd -> Grades,           gam53
    FatherOcc -> Grades,          gam54
    Intelligence -> EducExp,      gam61
    Siblings -> EducExp,          gam62
    FatherEd -> EducExp,          gam63
    FatherOcc -> EducExp,         gam64
    Grades -> EducExp,            beta65
    Intelligence -> OccupAsp,     gam71
    Siblings -> OccupAsp,         gam72
    FatherEd -> OccupAsp,         gam73
    FatherOcc -> OccupAsp,        gam74
    Grades -> OccupAsp,           beta75
    EducExp -> OccupAsp,          beta76

sem.kerch1 <- sem(model.kerch1, R.kerch, 737, fixed.x=c(Intelligence,Siblings,
    FatherEd,FatherOcc))
summary(sem.kerch1)
```

sem output

```
 Model Chisquare =  3.2685e-13   Df =  0 Pr(>Chisq) = NA
 Chisquare (null model) =  1664.3   Df =  21
 Goodness-of-fit index =  1
 AIC =  36
 AICc =  0.95265
 BIC =  118.85
 CAIC =  3.2685e-13
 Normalized Residuals
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-4.26e-15 -1.35e-15  0.00e+00 -4.17e-16  0.00e+00  1.49e-15
 Parameter Estimates
           Estimate  Std Error z value  Pr(>|z|)
gam51       0.525902 0.031182  16.86530 8.0987e-64 Grades <--- Intelligence
gam52      -0.029942 0.030149  -0.99314 3.2064e-01 Grades <--- Siblings
gam53       0.118966 0.038259   3.10951 1.8740e-03 Grades <--- FatherEd
gam54       0.040603 0.037785   1.07456 2.8257e-01 Grades <--- FatherOcc
gam61       0.160270 0.032710   4.89979 9.5940e-07 EducExp <--- Intelligence
gam62      -0.111779 0.026876  -4.15899 3.1966e-05 EducExp <--- Siblings
gam63       0.172719 0.034306   5.03461 4.7882e-07 EducExp <--- FatherEd
gam64       0.151852 0.033688   4.50758 6.5571e-06 EducExp <--- FatherOcc
beta65      0.405150 0.032838  12.33799 5.6552e-35 EducExp <--- Grades
gam71      -0.039405 0.034500  -1.14215 2.5339e-01 OccupAsp <--- Intelligence
gam72      -0.018825 0.028222  -0.66700 5.0477e-01 OccupAsp <--- Siblings
gam73      -0.041333 0.036216  -1.14126 2.5376e-01 OccupAsp <--- FatherEd
gam74       0.099577 0.035446   2.80924 4.9658e-03 OccupAsp <--- FatherOcc
beta75      0.157912 0.037443   4.21738 2.4716e-05 OccupAsp <--- Grades
beta76      0.549593 0.038260  14.36486 8.5976e-47 OccupAsp <--- EducExp
V[Grades]   0.650995 0.033935  19.18333 5.1010e-82 Grades <--> Grades
V[EducExp]  0.516652 0.026932  19.18333 5.1010e-82 EducExp <--> EducExp
V[OccupAsp] 0.556617 0.029016  19.18333 5.1010e-82 OccupAsp <--> OccupAsp
```

Note that these models give different path coefficients

```
> round(sem.kerch$coeff,3)
      gam51       gam52       gam53       gam54       gam61       gam62       gam63       gam64
       gam71       gam72       gam73
      0.526      -0.030       0.119       0.041       0.373      -0.124       0.221       0.168
      0.249      -0.092       0.099
      gam74   V[Grades]  V[EducExp] V[OccupAsp]
      0.198       0.651       0.624       0.807
> round(sem.kerch1$coeff,3)
      gam51       gam52       gam53       gam54       gam61       gam62       gam63       gam64
      beta65       gam71       gam72
      0.526      -0.030       0.119       0.041       0.160      -0.112       0.173       0.152
      0.405      -0.039      -0.019
      gam73       gam74      beta75      beta76   V[Grades]  V[EducExp] V[OccupAsp]
     -0.041       0.100       0.158       0.550       0.651       0.517       0.557

> round(mr.kk$beta,2)
             Grades EducExp OccupAsp
Intelligence   0.53    0.37     0.25
Siblings      -0.03   -0.12    -0.09
FatherEd       0.12    0.22     0.10
FatherOcc      0.04    0.17     0.20
```

A latent variable structural model

1. Taken from the LISREL User's reference guide
    -
2. Data from, Caslyn and Kenny (1977)
    - Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement
3. Variables
    - self concept
    - parental evaluation
    - teacher evaluation
    - friend evaluation
    - educational aspiration
    - college plans

Caslyn and Kenny data

	self	parent	teacher	friend	edu_asp	college
self_concept	1.00	0.73	0.70	0.58	0.46	0.56
parental_eval	0.73	1.00	0.68	0.61	0.43	0.52
teacher_eval	0.70	0.68	1.00	0.57	0.40	0.48
friend_eval	0.58	0.61	0.57	1.00	0.37	0.41
edu_aspir	0.46	0.43	0.40	0.37	1.00	0.72
college_plans	0.56	0.52	0.48	0.41	0.72	1.00

Creating the model using structure.diagam

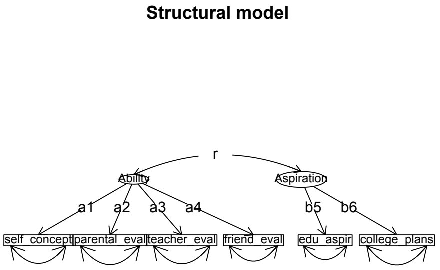
```
fx <- structure.list(6,list(c(1:4),c(5:6)),item.labels = rownames(ability),f.lab
mod.edu <- structure.diagram(fx,"r",title="Lisrel example 3.2",errors=TRUE,lr=FA
fx
```

```
fx
              Ability Aspiration
self_concept  "a1"    "0"
parental_eval "a2"    "0"
teacher_eval  "a3"    "0"
friend_eval   "a4"    "0"
edu_aspir     "0"     "b5"
college_plans "0"     "b6"
```

The sem commands are in the mod.edu object

```
mod.edu
      Path                               Parameter Value
 [1,] "Ability->self_concept"            "a1"      NA
 [2,] "Ability->parental_eval"           "a2"      NA
 [3,] "Ability->teacher_eval"            "a3"      NA
 [4,] "Ability->friend_eval"             "a4"      NA
 [5,] "Aspiration->edu_aspir"            "b5"      NA
 [6,] "Aspiration->college_plans"        "b6"      NA
 [7,] "self_concept<->self_concept"      "x1e"     NA
 [8,] "parental_eval<->parental_eval"    "x2e"     NA
 [9,] "teacher_eval<->teacher_eval"      "x3e"     NA
[10,] "friend_eval<->friend_eval"        "x4e"     NA
[11,] "edu_aspir<->edu_aspir"            "x5e"     NA
[12,] "college_plans<->college_plans"    "x6e"     NA
[13,] "Aspiration<->Ability"             "rF2F1"   NA
[14,] "Ability<->Ability"                NA        "1"
[15,] "Aspiration<->Aspiration"          NA        "1"
```

A model of the Caslyn-Kenny (1997) data set



ability <- as.matrix(ability) sem requires matrix input sem.edu <- sem(mod=mod.edu,S=ability,N=556)
summary(sem.edu)

```
Model Chisquare =  9.2557   Df =  8 Pr(>Chisq) = 0.32118
 Chisquare (null model) =  1832   Df =  15
 Goodness-of-fit index =  0.99443
 Adjusted goodness-of-fit index =  0.98537
 RMSEA index =  0.016817   90% CI: (NA, 0.054321)
 Bentler-Bonnett NFI =  0.99495
 Tucker-Lewis NNFI =  0.9987
 Bentler CFI =  0.99931
 SRMR =  0.012011
 AIC =  35.256
 AICc =  9.9273
 BIC =  91.426
 CAIC =  -49.31

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.4410 -0.1870  0.0000 -0.0131  0.2110  0.5330

 R-square for Endogenous Variables
 self_concept parental_eval  teacher_eval   friend_eval     edu_aspir college_plans
       0.7451        0.7213        0.6482        0.4834        0.6008        0.8629
```

With parameter estimates

```
 Parameter Estimates
      Estimate Std Error z value Pr(>|z|)
a1    0.86320  0.035145  24.5612 3.2846e-133 self_concept <--- Ability
a2    0.84932  0.035450  23.9582 7.5937e-127 parental_eval <--- Ability
a3    0.80509  0.036405  22.1149 2.2725e-108 teacher_eval <--- Ability
a4    0.69527  0.038634  17.9964  2.0795e-72 friend_eval <--- Ability
b5    0.77508  0.040357  19.2058  3.3077e-82 edu_aspir <--- Aspiration
b6    0.92893  0.039410  23.5712 7.6153e-123 college_plans <--- Aspiration
x1e   0.25488  0.023367  10.9075  1.0617e-27 self_concept <--> self_concept
x2e   0.27865  0.024128  11.5491  7.4600e-31 parental_eval <--> parental_eval
x3e   0.35184  0.026919  13.0703  4.8660e-39 teacher_eval <--> teacher_eval
x4e   0.51660  0.034725  14.8768  4.6594e-50 friend_eval <--> friend_eval
x5e   0.39924  0.038196  10.4525  1.4266e-25 edu_aspir <--> edu_aspir
x6e   0.13709  0.043505   3.1511  1.6264e-03 college_plans <--> college_plans
rF2F1 0.66637  0.030954  21.5276 8.5783e-103 Ability <--> Aspiration
```

Three competing models

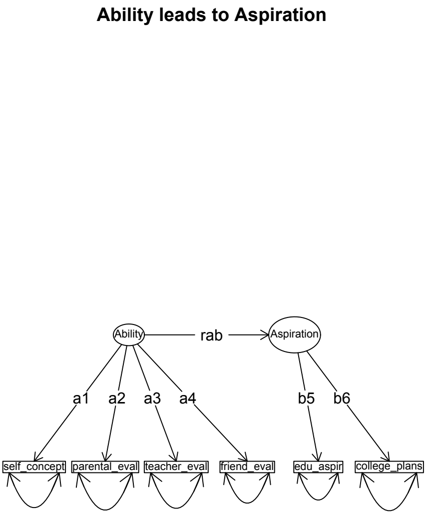
1. Ability and aspirations are correlated
    - r = .66
2. Ability causes aspirations
    - beta = .89
3. Aspirations cause ability
    - beta = .89

Create the model where ability lead to aspirations

```
phi <- phi.list(2,c(2))
phi
mod.edu <- structure.diagram(fx,phi,title="Aspiration leads to ability",errors=T
```

```
mod.edu
      Path                              Parameter Value
 [1,] "Ability->self_concept"           "a1"      NA
 [2,] "Ability->parental_eval"          "a2"      NA
 [3,] "Ability->teacher_eval"           "a3"      NA
 [4,] "Ability->friend_eval"            "a4"      NA
 [5,] "Aspiration->edu_aspir"           "b5"      NA
 [6,] "Aspiration->college_plans"       "b6"      NA
 [7,] "self_concept<->self_concept"     "x1e"     NA
 [8,] "parental_eval<->parental_eval"   "x2e"     NA
 [9,] "teacher_eval<->teacher_eval"     "x3e"     NA
[10,] "friend_eval<->friend_eval"       "x4e"     NA
[11,] "edu_aspir<->edu_aspir"           "x5e"     NA
[12,] "college_plans<->college_plans"   "x6e"     NA
[13,] "Ability ->Aspiration"            "rF1F2"   NA
[14,] "Ability<->Ability"               NA        "1"
[15,] "Aspiration<->Aspiration"         NA        "1"
```

Ability causes aspiration



Fit statistics are identical

```
> sem.edu <- sem(mod=mod.edu,S=ability,N=556)
> summary(sem.edu)

summary(sem.edu)

 Model Chisquare =  9.2557   Df =  8 Pr(>Chisq) = 0.32118
 Chisquare (null model) =  1832   Df =  15
 Goodness-of-fit index =  0.99443
 Adjusted goodness-of-fit index =  0.98537
 RMSEA index =  0.016817   90% CI: (NA, 0.054321)
 Bentler-Bonnett NFI =  0.99495
 Tucker-Lewis NNFI =  0.9987
 Bentler CFI =  0.99931
 SRMR =  0.012011
 AIC =  35.256
 AICc =  9.9273
 BIC =  91.426
 CAIC =  -49.31

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.4410 -0.1870  0.0000 -0.0131  0.2110  0.5330

 R-square for Endogenous Variables
 self_concept parental_eval  teacher_eval    friend_eval     Aspiration      edu_aspir college_plans
       0.7451        0.7213        0.6482         0.4834         0.4440         0.6008        0.8629
```

But the paths are different

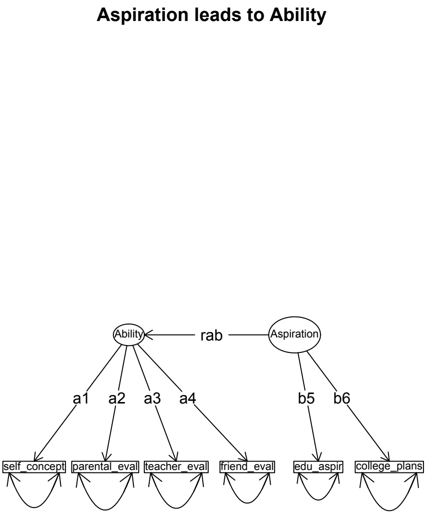
```
 Parameter Estimates
      Estimate Std Error z value Pr(>|z|)
a1    0.86320  0.035145  24.5612 3.2848e-133 self_concept <--- Ability
a2    0.84932  0.035450  23.9582 7.5919e-127 parental_eval <--- Ability
a3    0.80509  0.036405  22.1149 2.2732e-108 teacher_eval <--- Ability
a4    0.69527  0.038634  17.9964  2.0794e-72 friend_eval <--- Ability
b5    0.57792  0.030630  18.8678  2.0977e-79 edu_aspir <--- Aspiration
b6    0.69263  0.037979  18.2370  2.6257e-74 college_plans <--- Aspiration
x1e   0.25488  0.023367  10.9075  1.0617e-27 self_concept <--> self_concept
x2e   0.27865  0.024128  11.5491  7.4610e-31 parental_eval <--> parental_eval
x3e   0.35184  0.026919  13.0703  4.8654e-39 teacher_eval <--> teacher_eval
x4e   0.51660  0.034725  14.8768  4.6595e-50 friend_eval <--> friend_eval
x5e   0.39924  0.038196  10.4525  1.4266e-25 edu_aspir <--> edu_aspir
x6e   0.13709  0.043505   3.1511  1.6264e-03 college_plans <--> college_plans
rF1F2 0.89371  0.074673  11.9683  5.2068e-33 Aspiration <--- Ability

 Iterations =  30
```

Let aspirations cause ability

```
> phi[1,2] <- phi[2,1]
> phi[2,1] <- "0"
> mod.edu <- structure.diagram(fx,phi,main="Aspiration leads to Ability",errors=
>mod.edu
```

```
> mod.edu
      Path                              Parameter Value
 [1,] "Ability->self_concept"           "a1"      NA
 [2,] "Ability->parental_eval"          "a2"      NA
 [3,] "Ability->teacher_eval"           "a3"      NA
 [4,] "Ability->friend_eval"            "a4"      NA
 [5,] "Aspiration->edu_aspir"           "b5"      NA
 [6,] "Aspiration->college_plans"       "b6"      NA
 [7,] "self_concept<->self_concept"     "x1e"     NA
 [8,] "parental_eval<->parental_eval"   "x2e"     NA
 [9,] "teacher_eval<->teacher_eval"     "x3e"     NA
[10,] "friend_eval<->friend_eval"       "x4e"     NA
[11,] "edu_aspir<->edu_aspir"           "x5e"     NA
[12,] "college_plans<->college_plans"   "x6e"     NA
[13,] "Aspiration<-Ability"             "rF2F1"   NA
[14,] "Ability<->Ability"               NA        "1"
[15,] "Aspiration<->Aspiration"         NA        "1"
```

Aspiration leads to ability



Fits are still identical

```
> sem.edu <- sem(mod=mod.edu,S=ability,N=556)
> summary(sem.edu)
 Model Chisquare =  9.2557   Df =  8 Pr(>Chisq) = 0.32118
 Chisquare (null model) =  1832   Df =  15
 Goodness-of-fit index =  0.99443
 Adjusted goodness-of-fit index =  0.98537
 RMSEA index =  0.016817   90% CI: (NA, 0.054321)
 Bentler-Bonnett NFI =  0.99495
 Tucker-Lewis NNFI =  0.9987
 Bentler CFI =  0.99931
 SRMR =  0.012011
 AIC =  35.256
 AICc =  9.9273
 BIC =  91.426
 CAIC =  -49.31

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.4410 -0.1870  0.0000 -0.0131  0.2110  0.5330

 R-square for Endogenous Variables
 self_concept parental_eval  teacher_eval   friend_eval    Aspiration      edu_a
       0.7451        0.7213        0.6482        0.4834        0.4440      0.
```

And the crucial coefficient is different

```
 Parameter Estimates
      Estimate Std Error z value Pr(>|z|)
a1    0.86320  0.035145  24.5612 3.2848e-133 self_concept <--- Ability
a2    0.84932  0.035450  23.9582 7.5919e-127 parental_eval <--- Ability
a3    0.80509  0.036405  22.1149 2.2732e-108 teacher_eval <--- Ability
a4    0.69527  0.038634  17.9964  2.0794e-72 friend_eval <--- Ability
b5    0.57792  0.030630  18.8678  2.0977e-79 edu_aspir <--- Aspiration
b6    0.69263  0.037979  18.2370  2.6257e-74 college_plans <--- Aspiration
x1e   0.25488  0.023367  10.9075  1.0617e-27 self_concept <--> self_concept
x2e   0.27865  0.024128  11.5491  7.4610e-31 parental_eval <--> parental_eval
x3e   0.35184  0.026919  13.0703  4.8654e-39 teacher_eval <--> teacher_eval
x4e   0.51660  0.034725  14.8768  4.6595e-50 friend_eval <--> friend_eval
x5e   0.39924  0.038196  10.4525  1.4266e-25 edu_aspir <--> edu_aspir
x6e   0.13709  0.043505   3.1511  1.6264e-03 college_plans <--> college_plans
rF2F1 0.89371  0.074673  11.9683  5.2068e-33 Aspiration <--- Ability
```

Compare the three models

	correlated	asp	abil
a1	0.86	0.64	0.86
a2	0.85	0.63	0.85
a3	0.81	0.60	0.81
a4	0.70	0.52	0.70
b5	0.78	0.78	0.58
b6	0.93	0.93	0.69
x1e	0.25	0.25	0.25
x2e	0.28	0.28	0.28
x3e	0.35	0.35	0.35
x4e	0.52	0.52	0.52
x5e	0.40	0.40	0.40
x6e	0.14	0.14	0.14
rF2F1	0.67	0.89	0.89

1. Although fits were identical
2. Paths differ as a function of presumed influence
3. Which solution is correct?
4. Is this even possible to answer?

Implications of arrows

1. Need to fit alternative models
    - Create alternative plausible models
    - Create alternative implausible models (they will fit also).
2. Need to consider alternative representations
    - Try reversing arrows
3. Are there external variables (e.g., time) that allow one to choose between models?
4. Confirmation that a model fits does not confirm theoretical adequacy of the model.

Two fundamentally different types of observed variables

1. Observed variables can be "reflective" of the latent variables. They are "effect indicators".
   - Variables are caused by the latent variables.
   - Covariation between the variables are explained by the latent variables
2. Observed variables can be "causal indicators" or formative indicators that can directly effect the latent variable
   - Variables cause the "latent" variable
   - Covariation of the the observed variables is not modeled

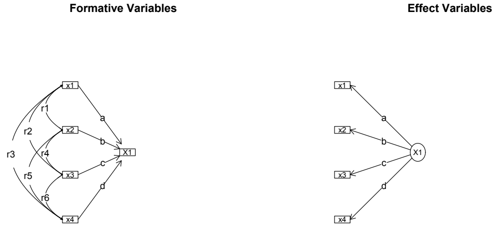
Formative indicators ?

1. The correlational structure of formative indicators is independent of the loadings on a factor.
   - They are not locally independent
2. Examples of formative indicators Time spent in social interaction
   - Time spent with family, time spent with friends, time spent with coworkers.
   - These might in fact be negatively correlated even though total score is important.

Effect (reflective) indicators

1. Test scores on various quantitative tests as effect indicators of trait
   - Feelings of self worth as effect indicators of self esteem
   - Ability items as indicators of ability
2. Correlational structure is a function of the path coefficients with latent variables
3. Variables are locally independent
   - (uncorrelated with each other when latent variable is partialled out)

Formative vs. Effect variables as regression versus factors



Bollen, K. A. (2002). *Latent variables in psychology and the social sciences*. US: Annual Reviews.

Fox, J., Nie, Z., & Byrnes, J. (2012). *sem: Structural Equation Models*.

Kerckhoff, A. C. (1974). *Ambition and Attainment: A Study of Four Samples of American Boys.* Washington, D. C.: American Sociological Association.