Psychology 454: Latent Variable Modeling

Putting it all together:
Taking latent variables seriously

Department of Psychology
Northwestern University
Evanston, Illinois USA

March 9, 2011

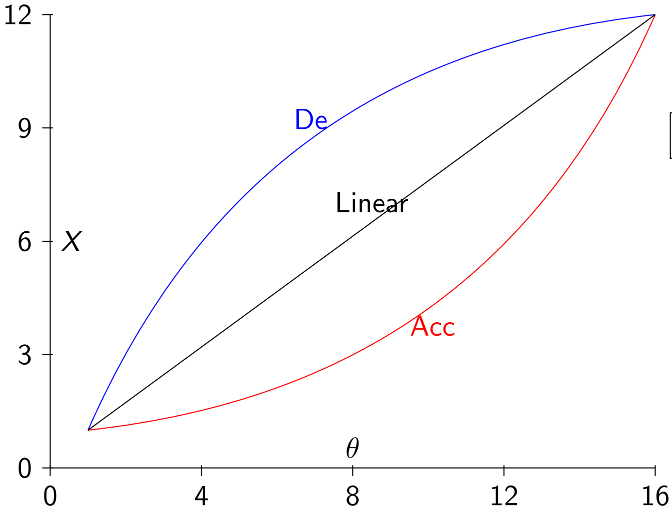
Outline

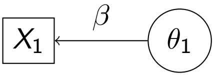
1. Data = Model + Residual
   - Types of data and types of scales
   - Correlation and regression
   - Multivariate Regression and Partial Correlation
   - Wright's rules: Applying path models to regression
2. Measurement models
   - Reliability models (Including Item Response Theory models)
   - Exploratory models
   - Confirmatory models
3. Structural models
   - Path diagram approach to structure using regression
   - Measurement + structure
   - Multiple Indicators, Multiple Causes
4. More complicated designs
   - STARTS models as a general framework
   - simplex cases as special change models
   - Applications in genetics

Data = Model + Residual

1. What are the data?
   - Comparisons of rank order $(O_i < S_j, O_i < O_j, S_i < S_j)$
   - Comparisons of proximity $(|O_i - S_j| < \delta, |O_i - S_j| < |O_k - S_l|)$
2. How are they modeled?
   - As observed variables
     - Correlation of OVs
     - Regression of OVs
     - Principal Components
   - As latent variables
     - Correlation of LVs with OVs
     - Correlation of LVs
     - Regression of LVs
3. How large are the residuals?

Mapping Observed Variables to Latent Variables

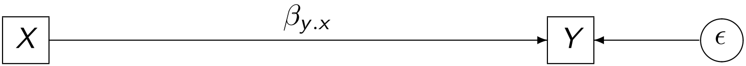




- Modeling the data
- Types of scales
  - Ordinal
  - Interval
  - Ratio

Bivariate Regression

δ X Y ϵ



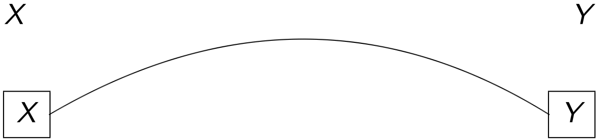
$$\hat{y} = \beta_{y.x} x + \epsilon$$

$$\beta_{y.x} = \frac{\sigma_{xy}}{\sigma_x^2}$$

$$\hat{x} = \beta_{x.y} y + \delta$$

$$\beta_{y.x} = \frac{\sigma_{xy}}{\sigma_y^2}$$

Bivariate Correlation



$\hat{x} = \beta_{x.y} y + \delta$

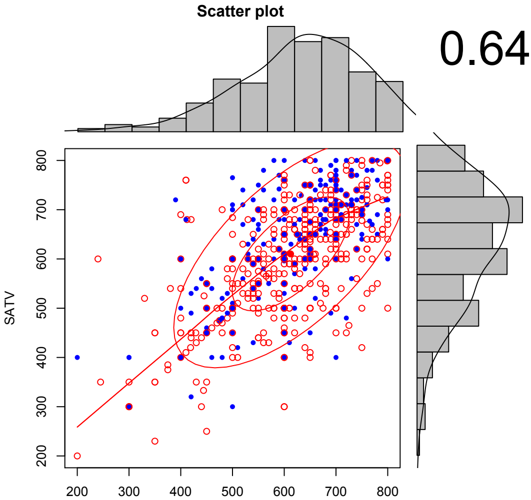
$\hat{y} = \beta_{y.x} x + \epsilon$

$\beta_{y.x} = \frac{\sigma_{xy}}{\sigma_y^2}$

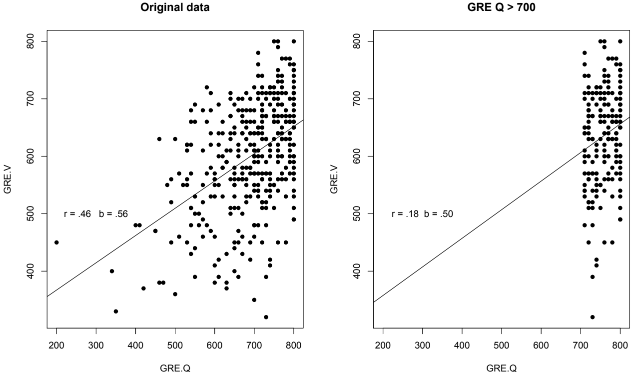
$\beta_{y.x} = \frac{\sigma_{xy}}{\sigma_x^2}$

$$r_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma^{2x} \sigma_y^2}}$$

The correlation as a linear model



Regression and restriction of range



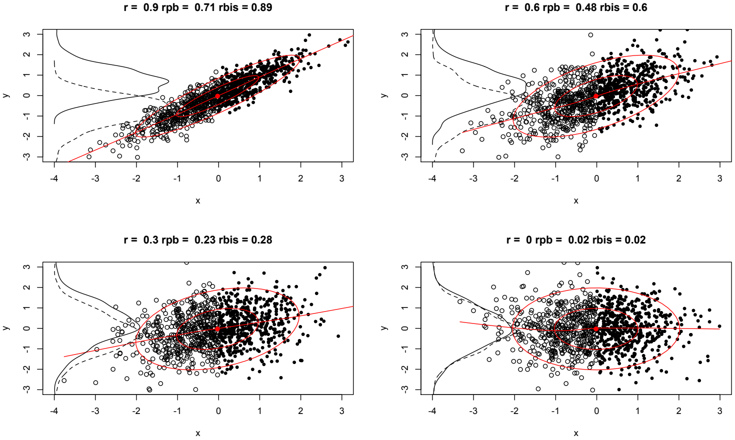
Although the correlation is very sensitive, regression slopes are relatively insensitive to restriction of range.

Alternative versions of the correlation coefficient

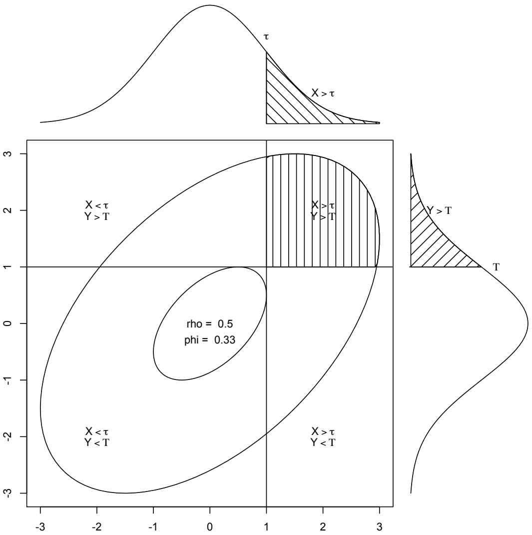
Table: A number of correlations are Pearson r in different forms, or with particular assumptions. If $r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$, then depending upon the type of data being analyzed, a variety of correlations are found.

Coefficient	symbol	X	Y	Assumptions
Pearson	r	continuous	continuous	
Spearman	rho (ρ)	ranks	ranks	
Point bi-serial	r_{pb}	dichotomous	continuous	
Phi	ϕ	dichotomous	dichotomous	
Bi-serial	r_{bis}	dichotomous	continuous	normality
Tetrachoric	r_{tet}	dichotomous	dichotomous	bivariate normality
Polychoric	r_{pc}	categorical	categorical	bivariate normality

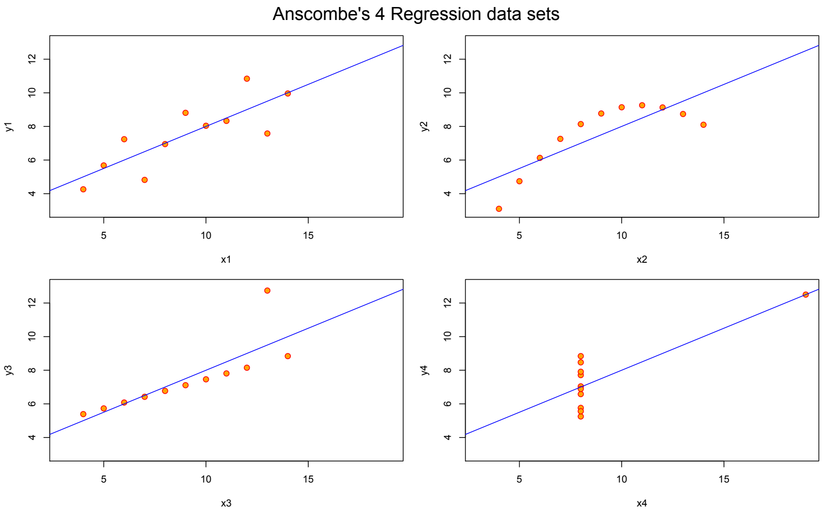
The biserial correlation estimates the latent correlation



The tetrachoric correlation estimates the latent correlation



Cautions about correlations: Anscombe data set

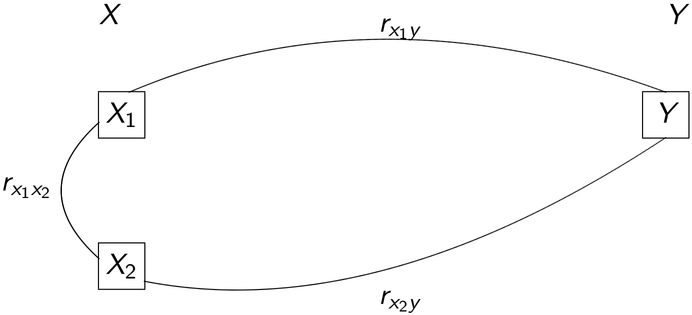


The ubiquitous correlation coefficient

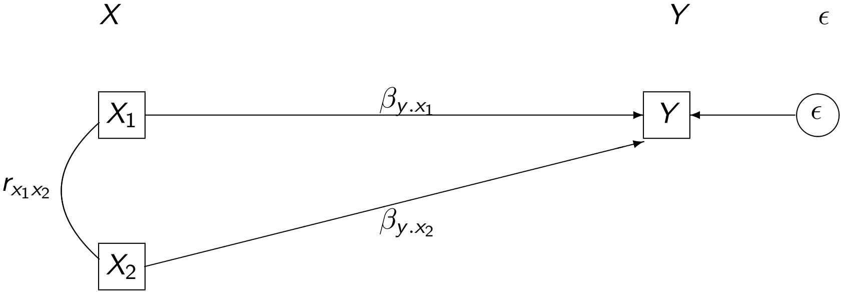
Table: Alternative Estimates of effect size. Using the correlation as a scale free estimate of effect size allows for combining experimental and correlational data in a metric that is directly interpretable as the effect of a standardized unit change in x leads to r change in standardized y.

Statistic	Estimate	r equivalent	as a function of r
Pearson correlation	$r_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}$	r_{xy}	
Regression	$b_{y.x} = \frac{C_{xy}}{\sigma_x^2}$	$r = b_{y.x}\frac{\sigma_y}{\sigma_x}$	$b_{y.x} = r\frac{\sigma_x}{\sigma_y}$
Cohen's d	$d = \frac{X_1 - X_2}{\sigma_x}$	$r = \frac{d}{\sqrt{d^2+4}}$	$d = \frac{2r}{\sqrt{1-r^2}}$
Hedge's g	$g = \frac{X_1 - X_2}{s_x}$	$r = \frac{g}{\sqrt{g^2+4(df/N)}}$	$g = \frac{2r\sqrt{df/N}}{\sqrt{1-r^2}}$
t - test	$t = \frac{d\sqrt{df}}{2}$	$r = \sqrt{t^2/(t^2+df)}$	$t = \sqrt{\frac{r^2 df}{1-r^2}}$
F-test	$F = \frac{d^2 df}{4}$	$r = \sqrt{F/(F+df)}$	$F = \frac{r^2 df}{1-r^2}$
Chi Square		$r = \sqrt{\chi^2/n}$	$\chi^2 = r^2 n$
Odds ratio	$d = \frac{ln(OR)}{1.81}$	$r = \frac{ln(OR)}{1.81\sqrt{(ln(OR)/1.81)^2+4}}$	$ln(OR) = \frac{3.62r}{\sqrt{1-r^2}}$
$r_{equivalent}$	r with probability p	$r = r_{equivalent}$	

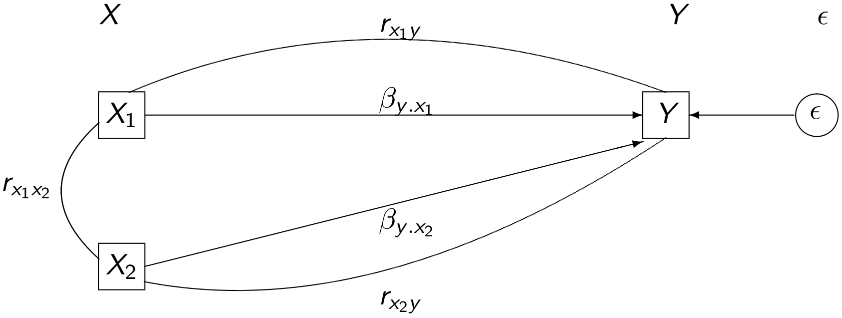
Multiple correlations



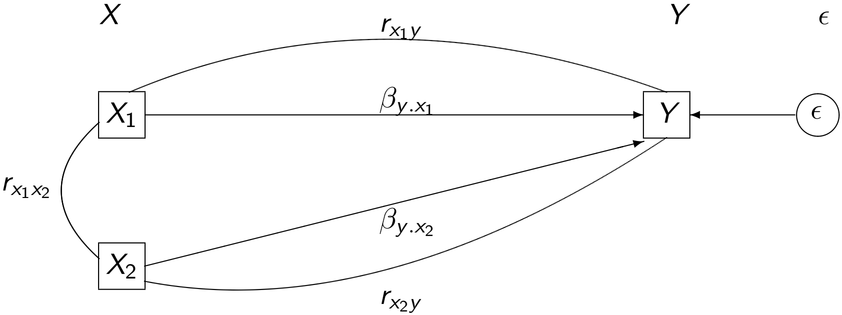
Multiple Regression



Multiple Regression: decomposing correlations



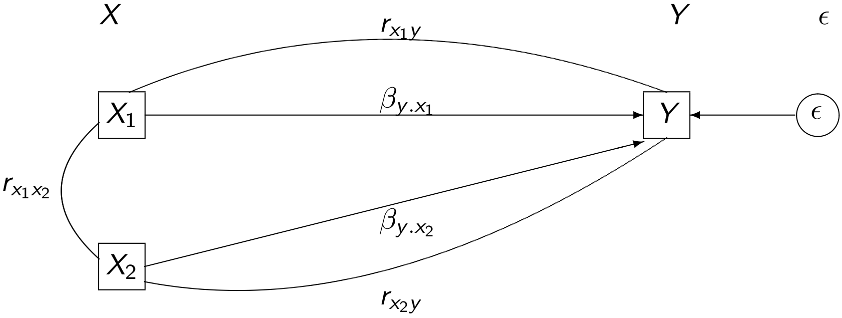
Multiple Regression: decomposing correlations



$$r_{x_1y} = \overbrace{\beta_{y.x_1}}^{\color{blue}{direct}} + \overbrace{r_{x_1x_2}\beta_{y.x_2}}^{\color{red}{indirect}}$$

$$r_{x_2y} = \underbrace{\beta_{y.x_2}}_{\color{blue}{direct}} + \underbrace{r_{x_1x_2}\beta_{y.x_1}}_{\color{red}{indirect}}$$

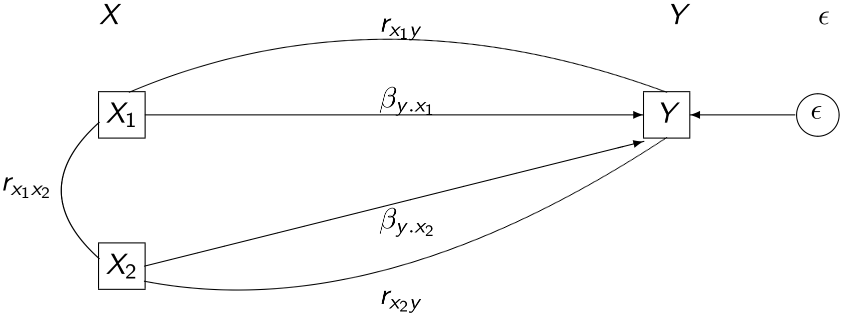
Multiple Regression: decomposing correlations



$$r_{x_1y} = \overbrace{\beta_{y.x_1}}^{\textit{direct}} + \overbrace{r_{x_1x_2}\beta_{y.x_2}}^{\textit{indirect}} \qquad \beta_{y.x_1} = \frac{r_{x_1y} - r_{x_1x_2}r_{x_2y}}{1 - r^2_{x_1x_2}}$$

$$r_{x_2y} = \underbrace{\beta_{y.x_2}}_{\textit{direct}} + \underbrace{r_{x_1x_2}\beta_{y.x_1}}_{\textit{indirect}} \qquad \beta_{y.x_2} = \frac{r_{x_2y} - r_{x_1x_2}r_{x_1y}}{1 - r^2_{x_1x_2}}$$

Multiple Regression: decomposing correlations



$$r_{x_1y} = \overbrace{\beta_{y.x_1}}^{\textit{direct}} + \overbrace{r_{x_1x_2}\beta_{y.x_2}}^{\textit{indirect}}$$

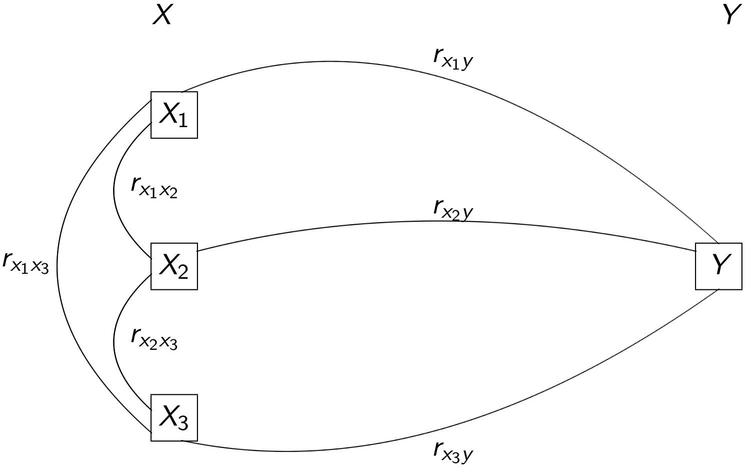
$$r_{x_2y} = \underbrace{\beta_{y.x_2}}_{\textit{direct}} + \underbrace{r_{x_1x_2}\beta_{y.x_1}}_{\textit{indirect}}$$

$$\beta_{y.x_1} = \frac{r_{x_1y} - r_{x_1x_2}r_{x_2y}}{1 - r^2_{x_1x_2}}$$

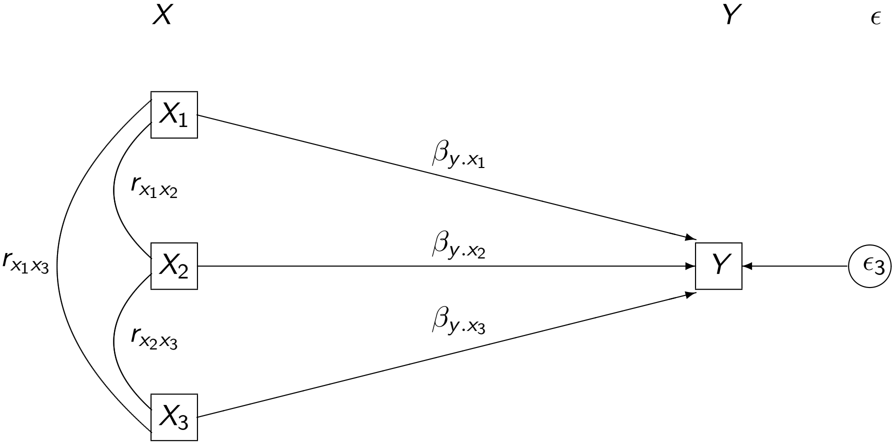
$$\beta_{y.x_2} = \frac{r_{x_2y} - r_{x_1x_2}r_{x_1y}}{1 - r^2_{x_1x_2}}$$

$$R^2 = r_{x_1y}\beta_{y.x_1} + r_{x_2y}\beta_{y.x_2}$$

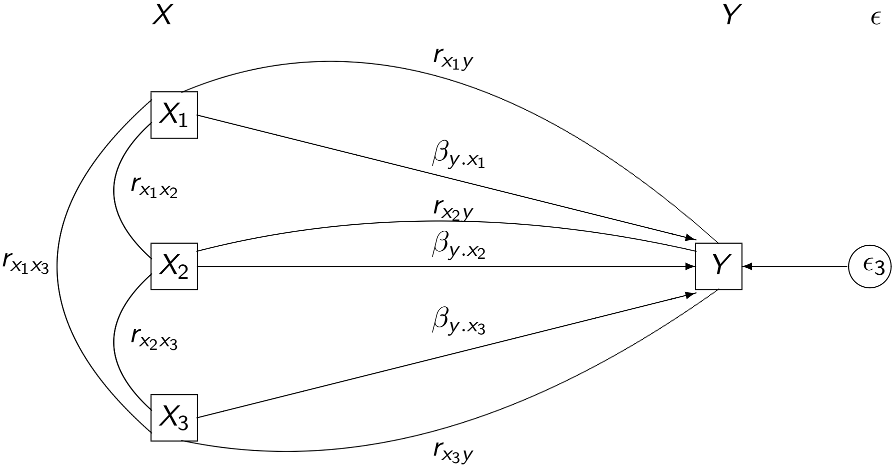
What happens with 3 predictors? The correlations



What happens with 3 predictors? β weights



What happens with 3 predictors?



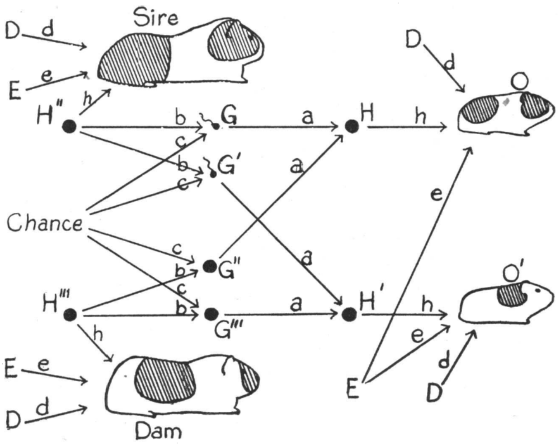
$$r_{x_1y} = \underbrace{\beta_{y.x_1}}_{\text{direct}} + \underbrace{r_{x_1x_2}\beta_{y.x_1} + r_{x_1x_3}\beta_{y.x_3}}_{\text{indirect}} \qquad r_{x_2y} = \dots \qquad r_{x_3y} = \dots$$

The math gets tedious

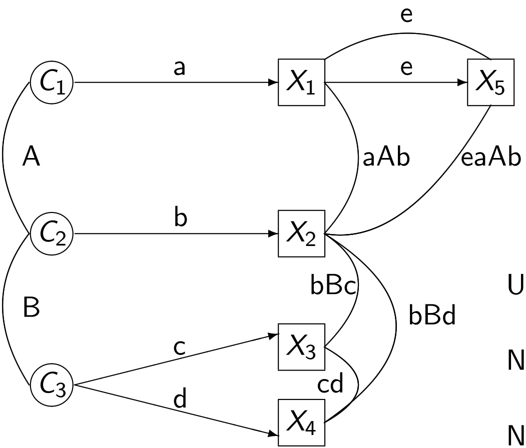
Multiple regression and matrix algebra

- Multiple regression requires solving multiple, simultaneous equations to estimate the direct and indirect effects.
  - Each equation is expressed as a $r_{x_i y}$ in terms of direct and indirect effects.
  - Direct effect is $\beta_{y.x_i}$
  - Indirect effect is $\sum_{j \neq i} beta_{y.x_j} r_{x_j y}$
- How to solve these equations?
- Tediously, or just use matrix algebra

Wright's Path model of inheritance in the Guinea Pig (Wright, 1921)



The basic rules of path analysis–think genetics



Up ... and over and down ...

No down and up

No double overs

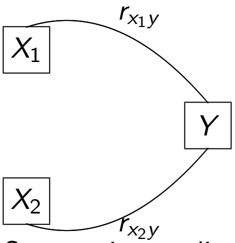
Up ... and down ...

Parents cause children

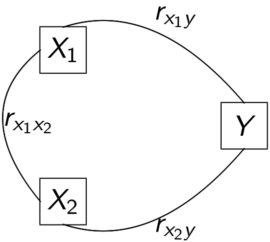
children do not cause parents

3 special cases of regression

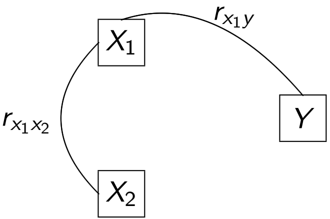
Orthogonal predictors



Correlated predictors

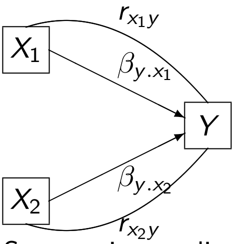


Suppressive predictors

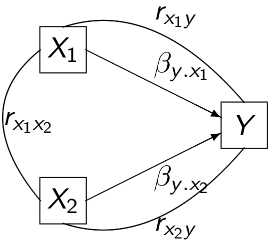


3 special cases of regression

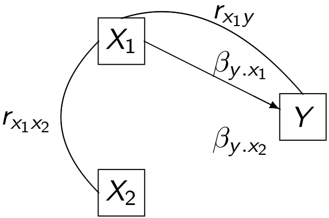
Orthogonal predictors



Correlated predictors



Suppressive predictors

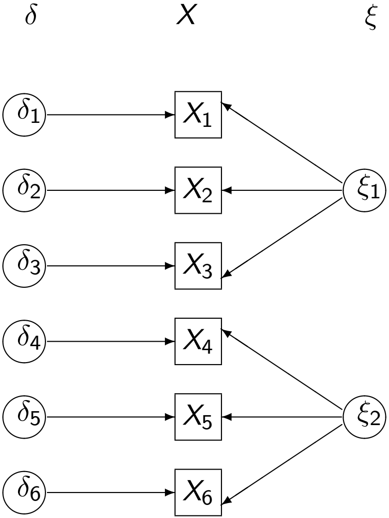


$$\beta_{y.x_1} = \frac{r_{x_1y} - r_{x_1x_2} r_{x_2y}}{1 - r^2_{x_1x_2}}$$

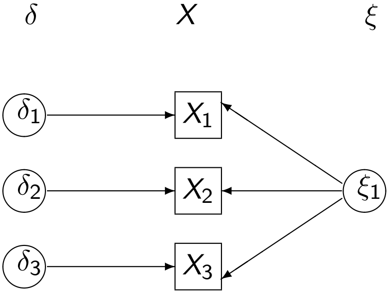
$$\beta_{y.x_2} = \frac{r_{x_2y} - r_{x_1x_2} r_{x_1y}}{1 - r^2_{x_1x_2}}$$

$$R^2 = r_{x_1y}\beta_{y.x_1} + r_{x_2y}\beta_{y.x_2}$$

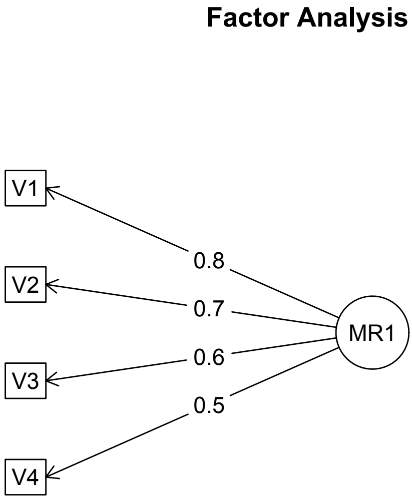
A measurement model for X



Congeneric Reliability

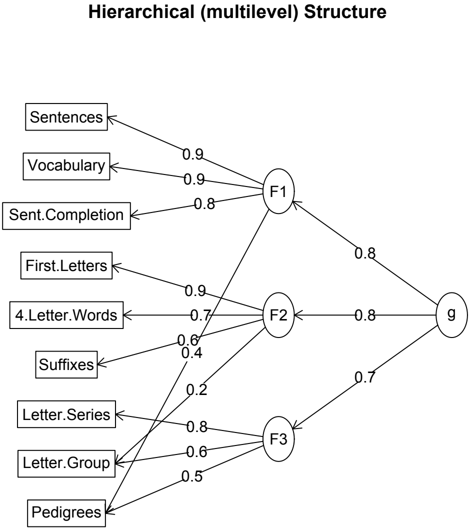


Factor analysis to show a congeneric structure

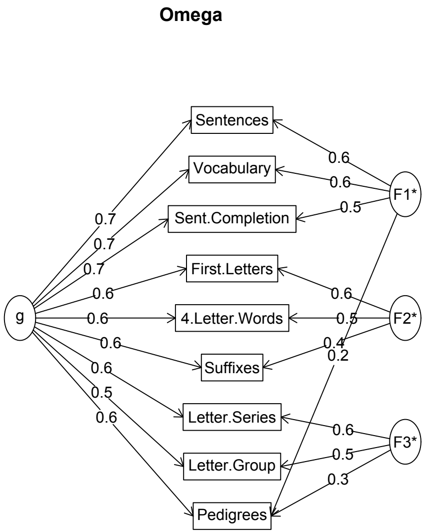


Hierarcical reliability models

Tests can have lower level and higher order reliability



Schmid Leiman transformation estimates general factor saturation



Multiple estimates of reliabilty

- Generic: $\frac{\sigma_t^2}{\sigma_x^2} = \frac{\sigma_t^2}{\sigma_t^2+\sigma_e^2} = 1 - \frac{\sigma_e^2}{\sigma_x^2}$
  - $\sigma_t^2 = \sigma_g^2 + \Sigma(\sigma_{g_i}^2) + \Sigma(\sigma_{s_j}^2)$
  - How to estimate σ_e^2?
  - α uses $1 - \bar{r}$ and is $\alpha = \frac{\sigma_x^2 - \Sigma(\sigma_j^2)}{\sigma_x^2} * \frac{n}{n-1}$
- $\omega_h = \frac{\sigma_g^2}{\sigma_x^2}$
- $\omega_t = \frac{\sigma_g^2 + \Sigma(\sigma_{g_i}^2)}{\sigma_x^2}$
- $\beta = \frac{2r_w}{1+r_w}$ where r_w is the correlation between the two worst split halves.
- $\beta \approx \omega_h \leq \alpha \leq \omega_t$

Guttman (1945); McDonald (1999); Revelle & Zinbarg (2009); Zinbarg, Revelle, Yovel & Li (2005)

Item Response Theory

- Item Response Theory attempts to model the response to an item in terms of two latent variables
  - Subject ability (trait value) θ
  - Item difficulty (item location) δ
  - Item discrimination (item sensitivity) α
  - $p(R|\theta, \delta) = \frac{1}{1+e^{\alpha(\delta-\theta)}}$
- Item Response Theory may be either exploratory or confirmatory
  - Rasch Model is a confirmatory model because it constrains all factor loadings to be equal.
  - Many multidimensional IRT models constrain cross loadings to 0.
- Factor analysis of Tetrachoric/Polychoric correlations based upon normal theory yields estimates that are equivalent of IRT parameters.
  - $\delta = \frac{\tau}{\sqrt{1-\lambda^2}}$
  - $\alpha = \frac{\lambda}{\sqrt{1-\lambda^2}}$

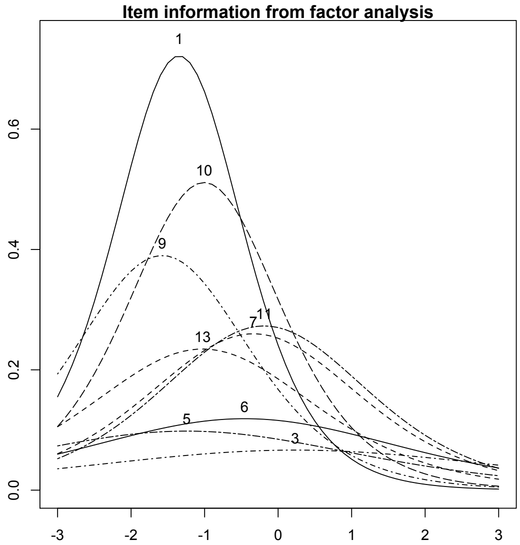
Factor analysis and IRT analysis of 14 iq items

```
Call: fa(r = iq.tf)
Standardized loadings based
upon correlation matrix
      MR1      h2   u2
iq1  0.51 2.7e-01 0.73
iq8  0.17 2.8e-02 0.97
iq10 0.21 4.5e-02 0.96
iq15 0.08 6.3e-03 0.99
iq20 0.22 4.6e-02 0.95
iq44 0.30 9.0e-02 0.91
iq47 0.41 1.7e-01 0.83
iq2  0.07 4.3e-03 1.00
iq11 0.38 1.4e-01 0.86
iq16 0.49 2.4e-01 0.76
iq32 0.39 1.6e-01 0.84
iq37 0.00 4.9e-06 1.00
iq43 0.38 1.4e-01 0.86
iq49 0.05 2.2e-03 1.00

                MR1
SS loadings    1.34
Proportion Var 0.10
```

```
      MR1      h2   u2
iq1  0.71 5.0e-01 0.50
iq8  0.25 6.4e-02 0.94
iq10 0.29 8.4e-02 0.92
iq15 0.12 1.6e-02 0.98
iq20 0.35 1.2e-01 0.88
iq44 0.38 1.4e-01 0.86
iq47 0.51 2.6e-01 0.74
iq2  0.11 1.1e-02 0.99
iq11 0.59 3.5e-01 0.65
iq16 0.64 4.1e-01 0.59
iq32 0.52 2.7e-01 0.73
iq37 0.00 1.3e-05 1.00
iq43 0.49 2.4e-01 0.76
iq49 0.07 4.6e-03 1.00

                MR1
SS loadings    2.48
Proportion Var 0.18
```

```
Item Response Analysis using
    Factor Analysis =
Call: irt.fa(x = iq.tf)

Item discrimination and location
      for factor  MR1
     discrimination location
iq1            1.00    -1.35
iq8            0.26    -0.95
iq10           0.30     0.23
iq15           0.13    -0.45
iq20           0.37    -1.24
iq44           0.41    -0.46
iq47           0.60    -0.34
iq2            0.11     0.42
iq11           0.73    -1.58
iq16           0.84    -1.01
iq32           0.61    -0.19
iq37           0.00     0.60
iq43           0.57    -1.03
iq49           0.07     0.42
```

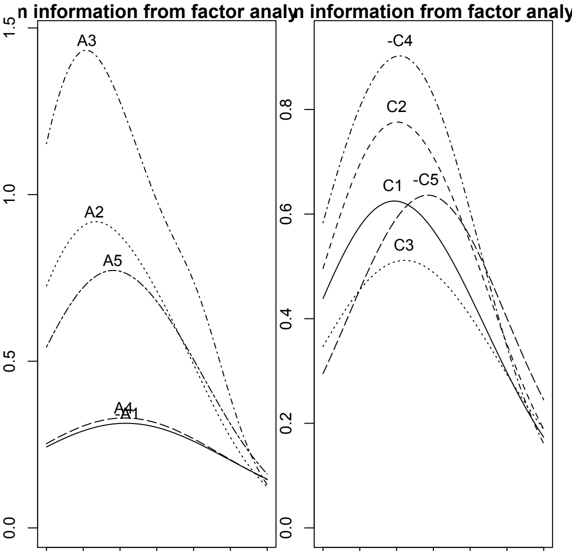
Item information function shows location and discrimination



IRT of multidimensional and polytomous items

- The generalization to ordered categorical items is relatively straightforward (Samejima, 1969).
- Much easier to see this in terms of factor of polychoric correlations with conversion to IRT parameters.
  - Example is from two scales of the Big Five.
  - Calculating polychorics is tedious
- `irt.fa` will find either tetrachoric or polychoric correlation, depending upon the data.
- Graphic output includes Item Information Curves, Item Characteristic Curves, and Test Information Curves.

Item information curves for factors 1 and 2 of BFI



Factor versus component models

- Factors model the correlations: $R \approx FF' + U^2$
  - Factors are latent (unobserved) entities that are *estimated*
  - $\hat{F}_s = XW = XR^{-1}F$
  - Factor scores are are indeterminate and need to be estimated, but the structure is determinant, given the model.
- Components model the observed scores
  - $C_s = XC$
  - This will model the error in X as well as X.
- The question of what is optimal number of factors/components to extract remains an open question. Good alternatives include:
  - Parallel analysis
  - Scree Test
  - Very Simple Structure
  - Minimum Average Partial

Exploratory versus confirmatory models

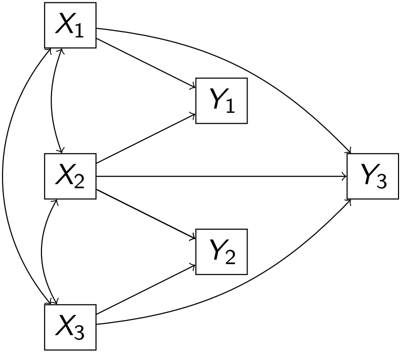
1. Exploratory models allow all cross loadings to be estimated.
   - Rotation to simple structure makes for more interpretable solution, but the cross loadings still are not forced to zero.
   - By using more parameters than confirmatory analysis, EFA will necessarily provide better fits, but at the cost of more complexity.
   - Alternative rotations will produce different appearing solutions, but all fit equally well.
2. Confirmatory analysis specifies some parameters to be 0.
   - Usually not allowing many if any cross loadings.
   - This will lead to lower levels of fit, but also more parsimonious models.
3. The supposed advantage of confirmatory models is to allow model testing.
   - This is particularly useful when examining structural similarity across groups and across ages.
   - Knowing that the measures are the same across groups is important when making comparisons.

Structural models

- Originally just path models as an approach to regression.
  - Possible to summarize many different regressions in one figure.
  - Regressions are done on observed variables.
- When combined with a measurement model to estimate factors, it is possible to do regression on latent variables.
- Do not need to estimate the factor scores, but rather fit the models to the covariances themselves.
  - In some sense, this is just doing regression on disattenuated correlations.
  - Needs to have good measures to estimate factors, poor measures will inflate structural values (just like a reliability correction).
- Goodness of fit tests allow for evaluation of fit of overall model as well as the parts of the model.
- Like many regression models, the direction of causality is not determined by the statistics, but by the design.

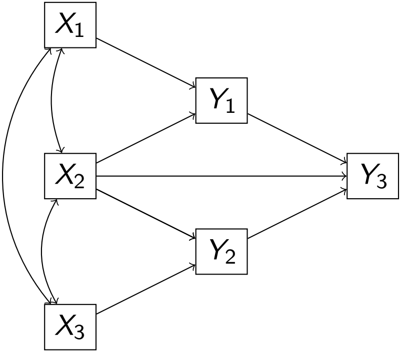
More complicated regression - 3 dependent variables

```
# ex3.11
Data <- read.table("ex3.11.dat")
names(Data) <- c("y1","y2","y3",
"x1","x2","x3")
model.ex3.11 <- ' y1 + y2  + y3 ~ x1 + x2 + x3 '
fit <- sem(model.ex3.11, data=Data)
summary(fit, standardized=TRUE)
```



More complicated regression - a path model

```
# ex3.11
Data <- read.table("ex3.11.dat")
names(Data) <- c("y1","y2","y3",
"x1","x2","x3")
model.ex3.11 <- ' y1 + y2 ~ x1 + x2 + x3
y3 ~ y1 + y2 + x2 '
fit <- sem(model.ex3.11, data=Data)
summary(fit, standardized=TRUE, fit.measures=TRUE)
```



The LISREL model of the Bollen data set – note the correlated residuals

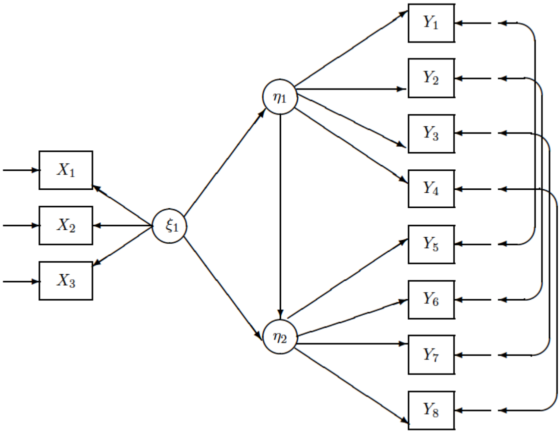
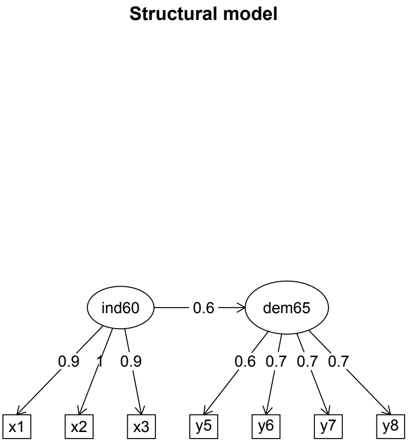


Figure 2: Panel Model of Democracy and Industrialization

http://www.ssicentral.com/lisrel/techdocs/sbcres.pdf

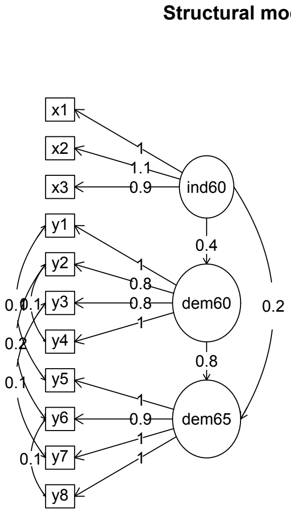
The most simple model: Industrialization leads to Democracy

lavaan.diagram(sem.fit,lr=FALSE,e.size=.2)



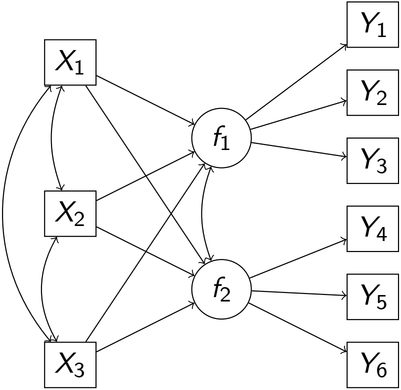
The very modified model – another way

lavaan.diagram(fit3,cut=0)



A MIMIC model

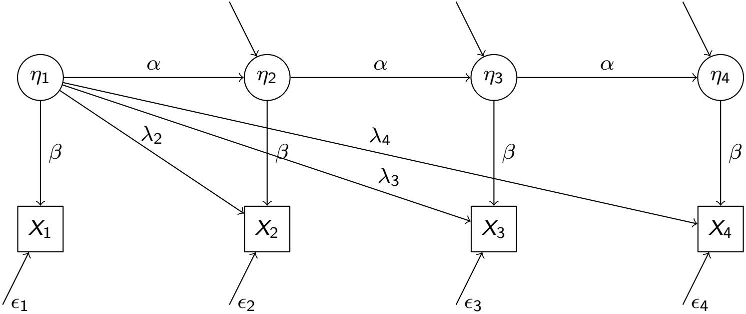
Consider the case of 6 variables with two latent factors that have 3 covariates (MPlus Example 5.8)



Stable Traits, Auto Regressive Trait, State Components

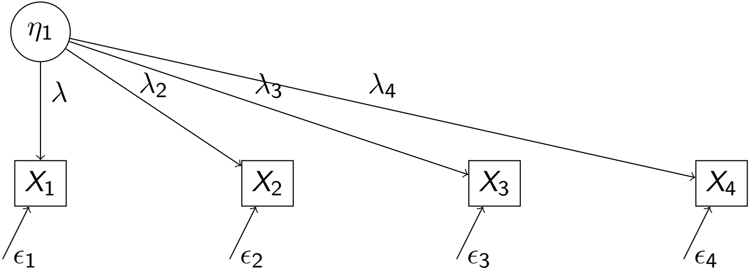
1. Test retest reliabilities reflect three different components and we need multiple measures to estimate them.
2. Stable Traits should lead to same score over time (λ)
3. Auto regressive traits suggest slow drop off in correlation over time (α)
4. State leads to internal consistency ($\frac{\beta^2}{\beta^2+\epsilon^2}$) but not stability

(Kenny & Zautra (2001); Lucas & Donnellan (2007))



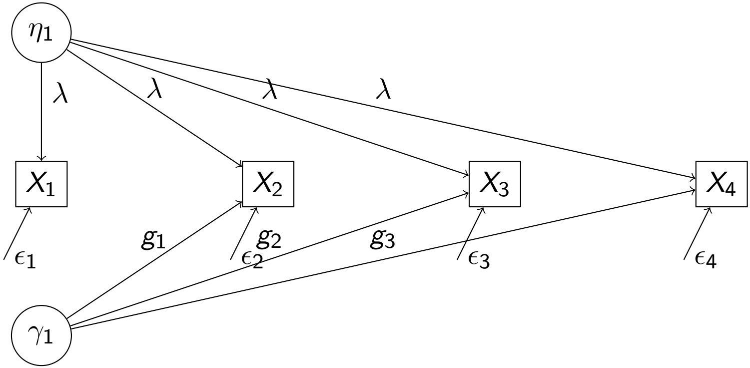
Stability across time

If $\alpha = 0$ and $\lambda_i = k$ then this is just congeneric measurement with equal loadings.



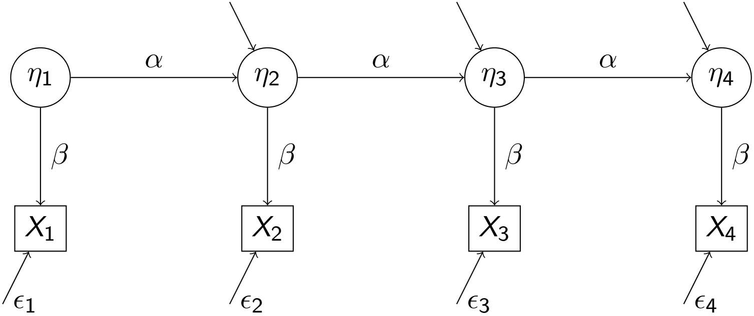
Growth across time

If $\alpha = 0$ and $\lambda = k$ then this is just congeneric measurement, but we can also model growth.



Simplex change over time

Predicts diminishing correlations across time.

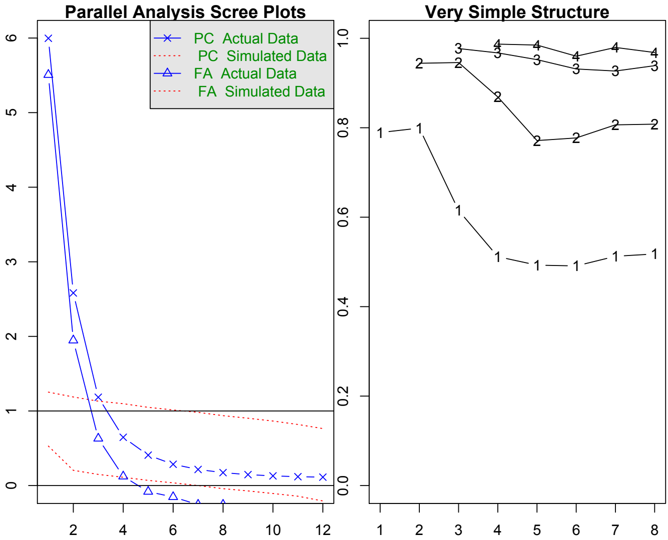


A growth set with simplex structure

```
> s12 <- sim.simplex(nvar =12, r=.8,mu=NULL, n=0)
> round(s12,2)
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
V1	1.00	0.80	0.64	0.51	0.41	0.33	0.26	0.21	0.17	0.13	0.11	0.09
V2	0.80	1.00	0.80	0.64	0.51	0.41	0.33	0.26	0.21	0.17	0.13	0.11
V3	0.64	0.80	1.00	0.80	0.64	0.51	0.41	0.33	0.26	0.21	0.17	0.13
V4	0.51	0.64	0.80	1.00	0.80	0.64	0.51	0.41	0.33	0.26	0.21	0.17
V5	0.41	0.51	0.64	0.80	1.00	0.80	0.64	0.51	0.41	0.33	0.26	0.21
V6	0.33	0.41	0.51	0.64	0.80	1.00	0.80	0.64	0.51	0.41	0.33	0.26
V7	0.26	0.33	0.41	0.51	0.64	0.80	1.00	0.80	0.64	0.51	0.41	0.33
V8	0.21	0.26	0.33	0.41	0.51	0.64	0.80	1.00	0.80	0.64	0.51	0.41
V9	0.17	0.21	0.26	0.33	0.41	0.51	0.64	0.80	1.00	0.80	0.64	0.51
V10	0.13	0.17	0.21	0.26	0.33	0.41	0.51	0.64	0.80	1.00	0.80	0.64
V11	0.11	0.13	0.17	0.21	0.26	0.33	0.41	0.51	0.64	0.80	1.00	0.80
V12	0.09	0.11	0.13	0.17	0.21	0.26	0.33	0.41	0.51	0.64	0.80	1.00

How many factors in a simplex?

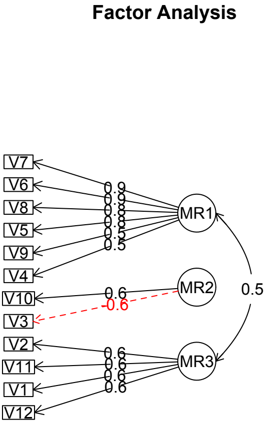


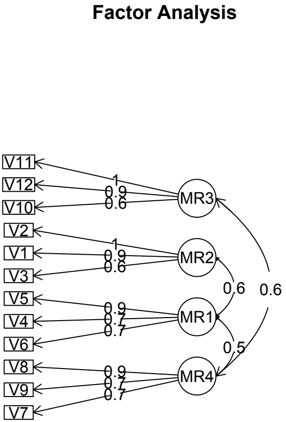
3 or 4 factor solution?

```
Factor Analysis using method =  minres
Call: fa(r = s12, nfactors = 3, n.obs = 500)
Standardized loadings based upon correlation matrix
     MR1   MR2   MR3   h2   u2
V1  0.05 -0.49  0.57 0.60 0.40
V2  0.11 -0.57  0.62 0.80 0.20
V3  0.29 -0.56  0.52 0.83 0.17
V4  0.54 -0.45  0.29 0.74 0.26
V5  0.77 -0.29  0.06 0.74 0.26
V6  0.94 -0.10 -0.10 0.81 0.19
V7  0.94  0.10 -0.10 0.81 0.19
V8  0.77  0.29  0.06 0.74 0.26
V9  0.54  0.45  0.29 0.74 0.26
V10 0.29  0.56  0.52 0.83 0.17
V11 0.11  0.57  0.62 0.80 0.20
V12 0.05  0.49  0.57 0.60 0.40

                MR1  MR2  MR3
SS loadings    4.14 2.33 2.54
Proportion Var 0.35 0.19 0.21
Cumulative Var 0.35 0.54 0.75

 With factor correlations of
     MR1 MR2  MR3
MR1 1.00   0 0.54
MR2 0.00   1 0.00
MR3 0.54   0 1.00
```

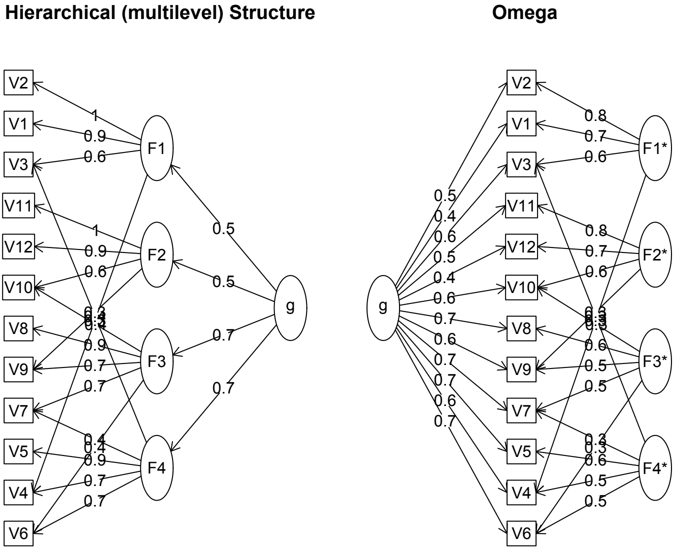
```
Standardized loadings based upon correlation matrix
      MR3   MR2   MR1   MR4   h2    u2
V1  -0.02  0.86 -0.07  0.04 0.68 0.321
V2  -0.01  0.96  0.00  0.03 0.92 0.082
V3   0.04  0.64  0.37 -0.05 0.78 0.218
V4   0.07  0.30  0.71 -0.07 0.79 0.214
V5   0.06  0.02  0.88  0.04 0.86 0.137
V6  -0.01 -0.04  0.69  0.36 0.79 0.209
V7  -0.04 -0.01  0.36  0.69 0.79 0.209
V8   0.02  0.06  0.04  0.88 0.86 0.137
V9   0.30  0.07 -0.07  0.71 0.79 0.214
V10  0.64  0.04 -0.05  0.37 0.78 0.218
V11  0.96 -0.01  0.03  0.00 0.92 0.082
V12  0.86 -0.02  0.04 -0.07 0.68 0.321

                MR3  MR2  MR1  MR4
SS loadings    2.38 2.38 2.44 2.44
Proportion Var 0.20 0.20 0.20 0.20
Cumulative Var 0.20 0.40 0.60 0.80

 With factor correlations of
     MR3  MR2  MR1  MR4
MR3 1.00 0.14 0.20 0.55
MR2 0.14 1.00 0.55 0.20
MR1 0.20 0.55 1.00 0.48
MR4 0.55 0.20 0.48 1.00
```

3 and 4 factor solutions

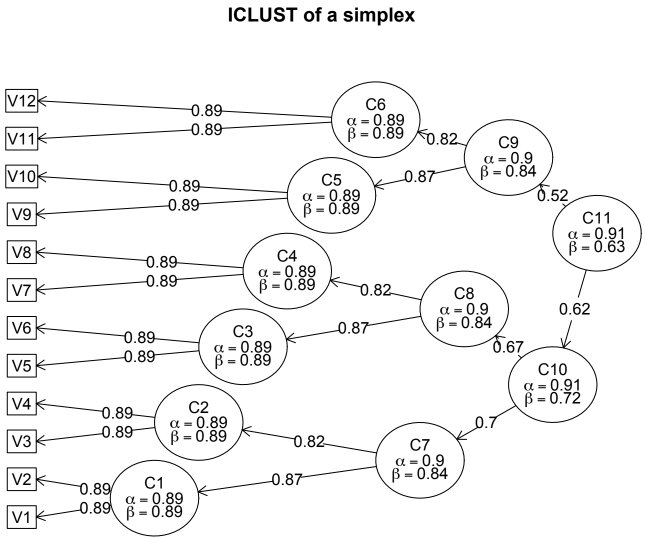




omega solutons



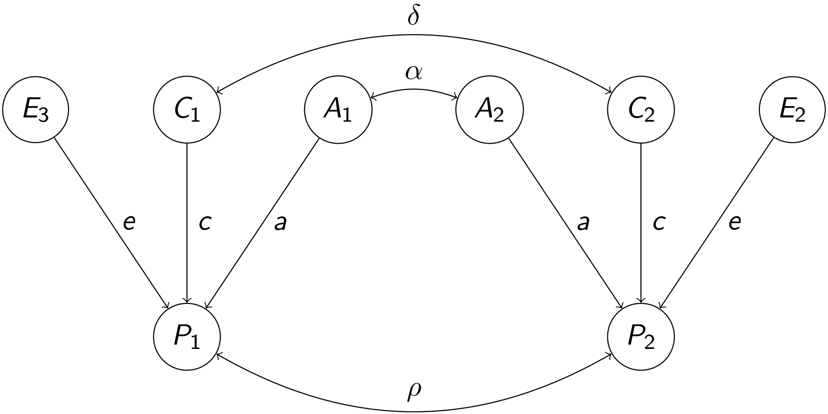
Clustering a simplex



ACE models: Additive genetic, Common environment, unique Environment

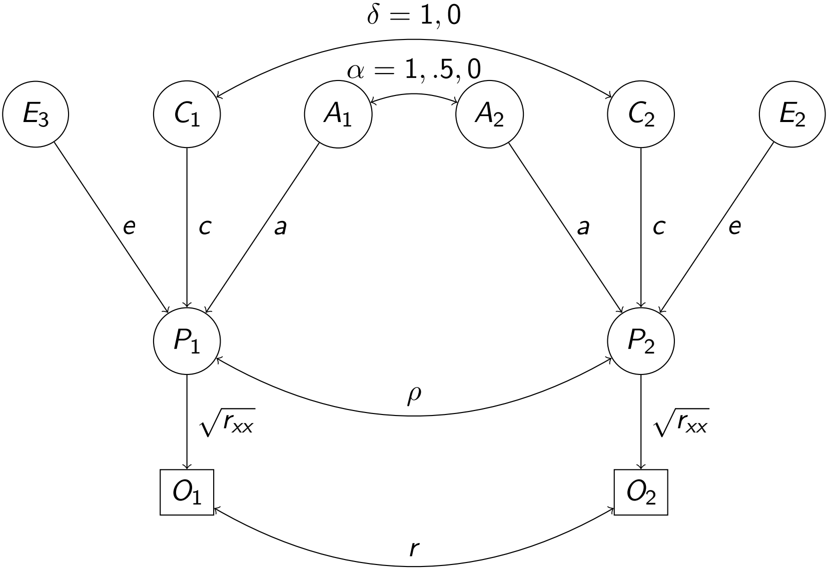
- Sources of variance in the common behavioral genetic models
  - Additive genetic variance
  - Shared family environmental variance
  - Unique environmental variance
- Modeled by the association between various family combinations
  - Monozygotic twins raised apart (sharing just genetic variance)
  - Monozygotic twins raised together (sharing genetic + environmental variance)
  - Dizygotic twins raised apart (sharing .5 genetic variance)
  - Dizygotic twins raised together (sharing .5 genetic + environmental variance)
  - Adopteds together (sharing just environmental variance)
  - unrelateds apart (sharing nothing)
- Estimate $\sigma_g^2, \sigma_c^2, \sigma_e^2$ and define $h^2 = \frac{\sigma_g^2}{\sigma_g^2+\sigma_c^2+\sigma_e^2}$

ACE model: no error correction



$\alpha = 1, .5, 0$ and $\delta = 1, 0$ depending upon family configuration.

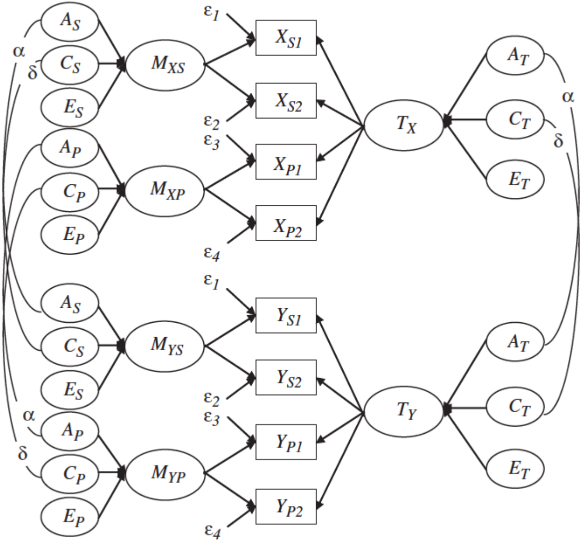
ACE model: with error correction. Model the observed correlations.



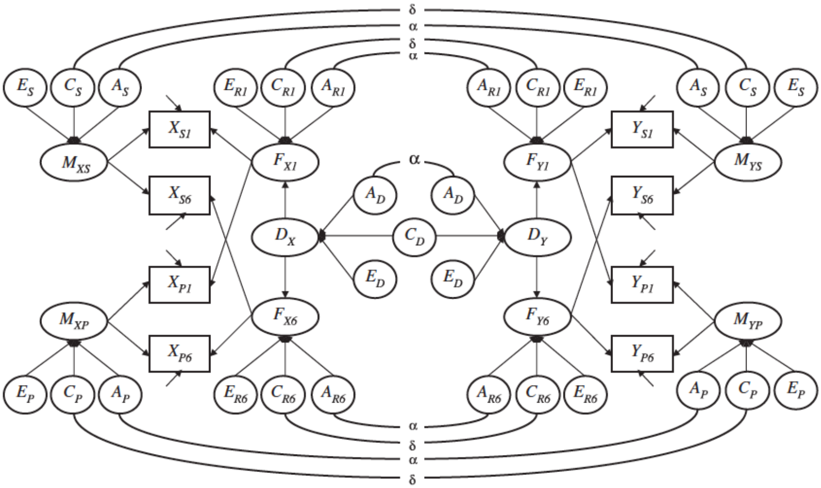
Genetic analysis in the German Observational Study of Twins

- Borkenau, Riemann, Angleitner & Spinath (2001a,b); Spinath, Angleitner, Borkenau, Riemann & Wolf (2002); Spinath & O'Connor (2003); Borkenau, Mauer, Riemann, Spinath & Angleitner (2004a,b); Spinath & Wolf (2006) report on the multi method analysis of personality and genetics.
  - Self report measures on the German NEO-PI-R.
  - Peer report measures on the peer version of the NEO.
- Analyzed the NEO at both the domain level (Big 5) as well as the facet level.
- Did this analysis for each pair of twins.

Multi method (self and peer) genetics analysis



Multi method (self and peer), multi-facets, genetics analysis



Modeling and data

- The power of analysis is no greater than the quality of the data
    - Data = Model + Residual
    - Model = Data - Residual
    - Model requires good data
- Multiple models will fit the data
    - Models should be compared to each other, not just the Null model
    - Is a model better than an alternative model?
    - How do you know?
    - What other models fit equally well?
- Be creative while being cautious and critical

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004a). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, *86*(4), 599–614.

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004b). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, *86*(4), 599–614.

Borkenau, P., Riemann, R., Angleitner, A., & Spinath, F. M. (2001a). Genetic and environmental influences on observed personality: Evidence from the german observational study of adult twins. *Journal of Personality and Social Psychology*, *80*(4), 655–668.

Borkenau, P., Riemann, R., Angleitner, A., & Spinath, F. M. (2001b). Genetic and environmental influences on observed personality: Evidence from the german observational study of adult twins. *Journal of Personality and Social Psychology*, *80*(4), 655–668.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255–282.

Kenny, D. & Zautra, A. (2001). Trait–state models for longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change, Decade of behavior* (pp. 243–263). Washington, DC, US: American Psychological Association.

Lucas, R. E. & Donnellan, M. B. (2007). How stable is happiness? using the starts model to estimate the stability of life satisfaction. *Journal of Research in Personality*, *41*(5), 1091 – 1098.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, N.J.: L. Erlbaum Associates.

Revelle, W. & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega and the glb: comments on Sijtsma. *Psychometrika*, *74*(1), 145–154.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric monographs No. 17.*, *34*(4 (Pt. 2)).

Spinath, F. M., Angleitner, A., Borkenau, P., Riemann, R., & Wolf, H. (2002). German observational study of adult twins (gosat): A multimodal investigation of personality, temperament and cognitive ability. *Twin Research*, *5*(5), 372–375.

Spinath, F. M. & O'Connor, T. G. (2003). A behavioral genetic study of the overlap between personality and parenting. *Journal of Personality*, *71*(5), 785–808.

Spinath, F. M. & Wolf, H. (2006). Cosmos and twinpaw: Initial report on two new german twin studies. *Twin Research and Human Genetics*, *9*(6), 787–790.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω_H): Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*(1), 123–133.