SEM

Professor Timothy Bates

tim.bates@ed.ac.uk

http://timbates.wikidot.com/mv-stats

Outline

- Week 1: Factor analysis
  - What is a factor analysis?
  - Factor extraction
  - Factor rotation
  - Factor interpretation
  - Factor Scores
- Week 2: Confirmatory Factor Analysis
- **Week 3: Path Analysis and SEM**
- Week 4: Complex causal modelling
  - Twins, multiple groups…

SEM is highly general

- Almost every other technique can be viewed as a subset of SEM
- Regression? ✓
- IRT/Measurement theory? ✓
- Latent variables? ✓
- Path analysis? ✓
- Multi-level models
- Time-series
- Matrix algebra and simultaneous equations?✓

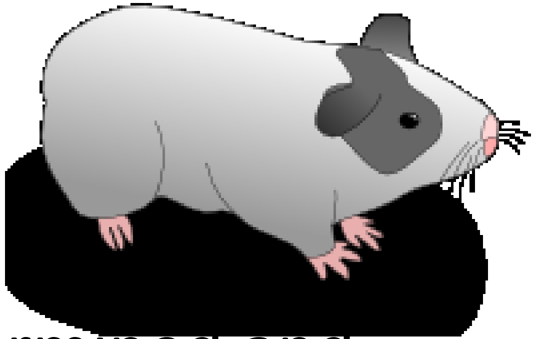
Once you know SEM, you wonder why we ever use anything else

- Measured more than one predictor? SEM will help
- Measured a construct with error? You need SEM
- Measured multiple DVs? Likely need SEM
- Measured variables that affect each other? You need SEM
  - Mediation & Moderation come for free
- Want to be explicit about causes and effects? You want SEM
- Want to draw these explicit models? You need SEM

SEM

- Can implement anything you have done in linear modeling
- Plus..
- You can create true scores (latent variables)
- Deal with multiple DVs simultaneously
- Constrain coefficients within and across groups
- Estimate effects over time (longitudinal models)
- etc. …

History



- Spearman, C.(1904). General intelligence objectively determined and measured. *American Journal of Psychology*, **15**, 201–293.
- Wright, S.(1920). The relative importance of heredity and environment in determining the piebald pattern of guinea–pigs. *Proceedings of the National Academy of Sciences*, **6**, 320–332.
- Duncan, O. D.(1966). Path analysis: Sociological examples. *The American Journal of Sociology*, **72**, 1–16.
- Also, there's a nice history of SEM in: Wolfle, L. M. (1999). Sewall wright on the method of path coefficients: An annotated bibliography. *Structural Equation Modeling*, **6**, 280–291.

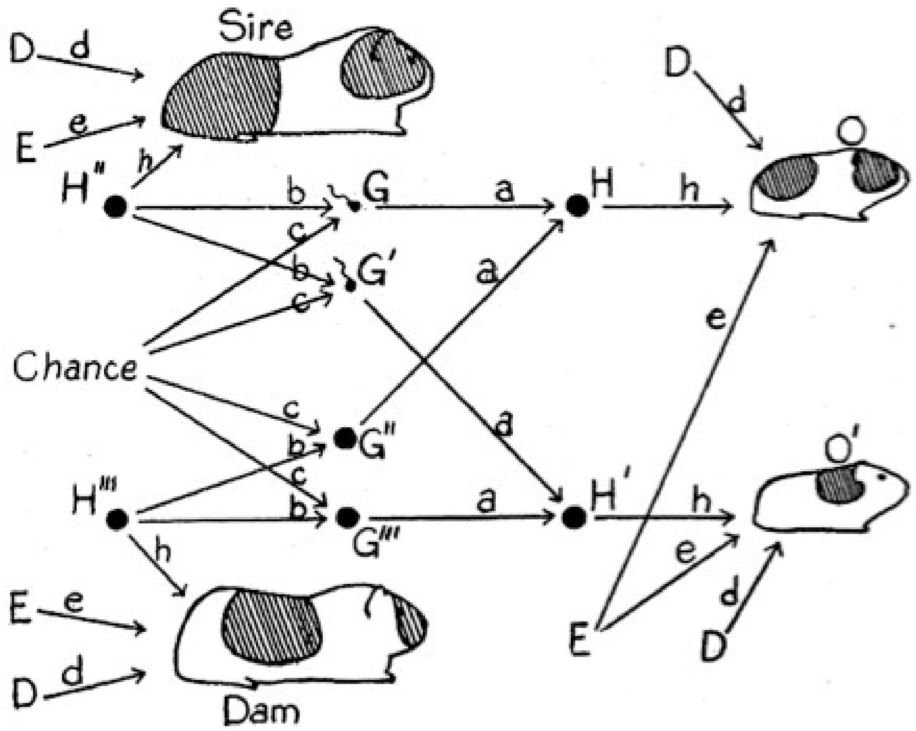
First steps

- Correlation (Pearson, 1896)
- Factor analysis (Spearman, 1904)
- Path analysis & SEM (Wright, 1918, 1920, 1934)

"The correlation between two variables can be shown to equal the sum of the products of the chains of path coefficients along all of the paths by which they are connected."

(Wright, 1920)

The first Path Diagram: Guinea-pig color

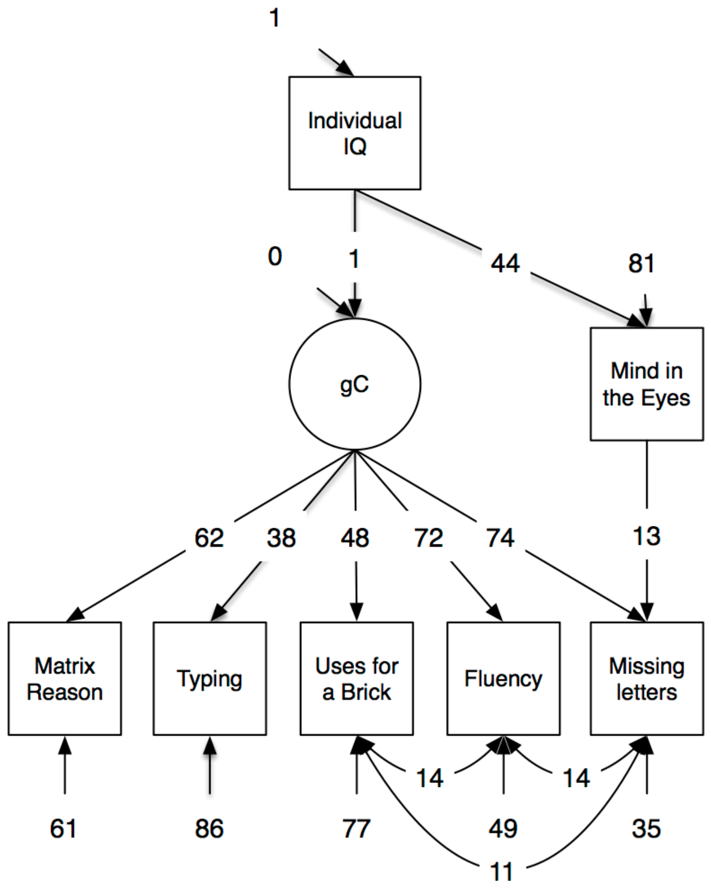


- Wright, S.(1920). The relative importance of heredity and environment in determining the piebald pattern of guinea–pigs. Proceedings of the National Academy of Sciences, 6, 320–332.

A modern diagram

Failure to replicate Woolley, etal. (2010). Collective intelligence factor

Science. doi:10.1126/science.1193147



Bates, T. C., & Gupta, S. (2017). Smart groups of smart people: Evidence for IQ as the origin of collective intelligence in the performance of human groups. *Intelligence, 60*, 46-56. doi:10.1016/j.intell.2016.11.004

The modern era

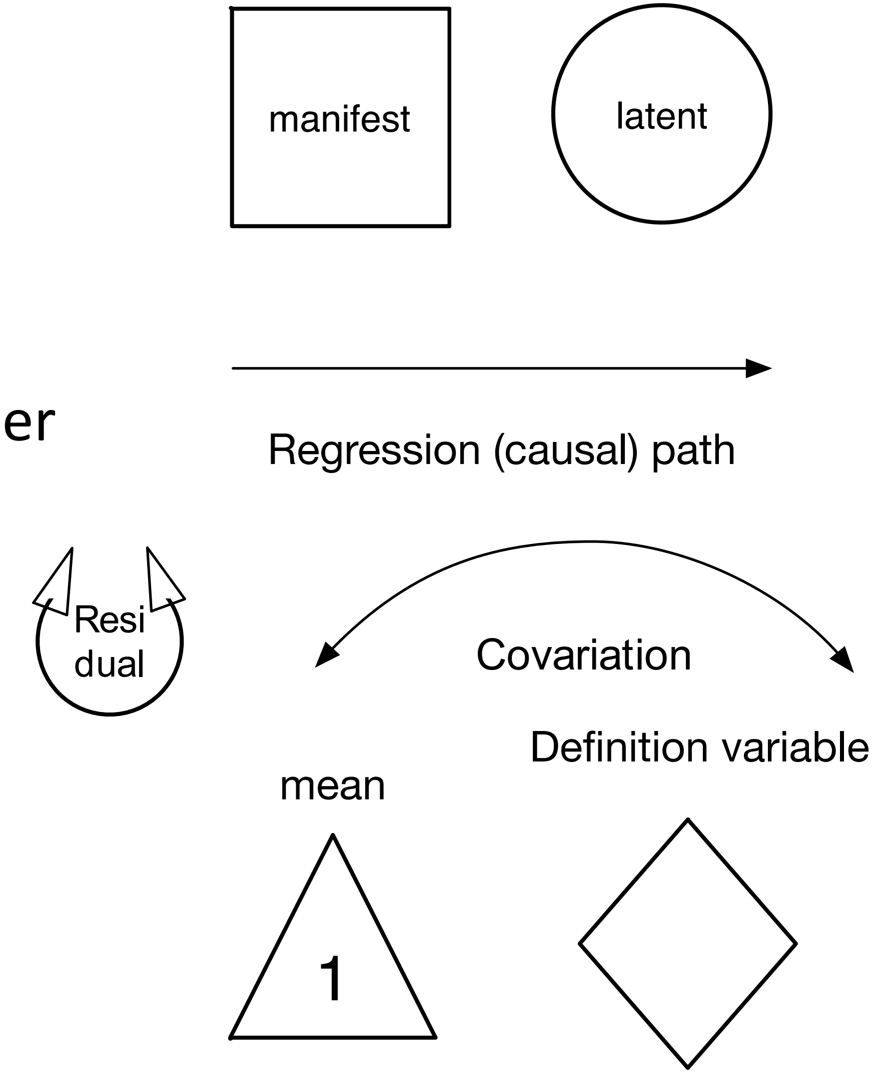
- Wright brought to social science (Duncan, 1966; Goldberger, 1972)
- Practical software (Jöreskog, 1970)
- Proliferation of packages: RAMpath, Mx, Amos, EQS, Mplus, OpenMx
- Flexible features: FIML, nonlinear constraints, multiple-groups

Tutorial-guide Paper on umx

- *umx: Twin and Path-based Structural Equation Modeling in R* (under revision) Timothy C. Bates, Hermine Maes, and Michael C. Neale, J Statistical Software.
  - https://www.dropbox.com/s/9w23ra06lcf0140/bates.pdf?dl=0

Building Blocks of any model

- Squares: manifestVars
- Circles: latentVars
- 1-headed arrows
  - Regression path from one variable to another
- 2-headed arrows
  - Variance or covariance of variables
- Triangle: constant
  - “one” is a special built-in constant
- Diamond: Definition variable
  - measured for each case

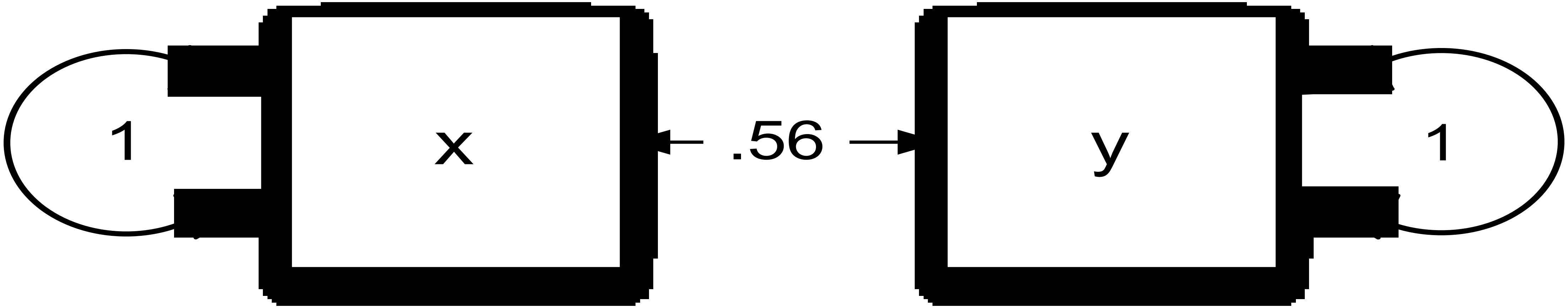


Some (simple) examples to link SEM to what you know from `cor.test()` and `Anova(lm())`

Correlation

- **cor.test**(~ x + y, data = xy)
- *r* = 0.58 CI95[0.43, 0.70] (t(98) = 7.00, p = 3.212e-10)

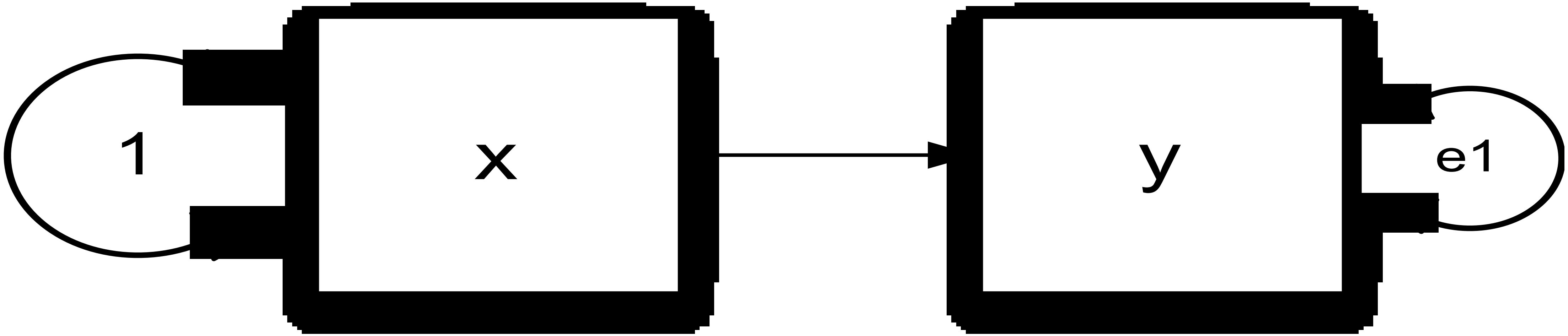
Correlation in SEM



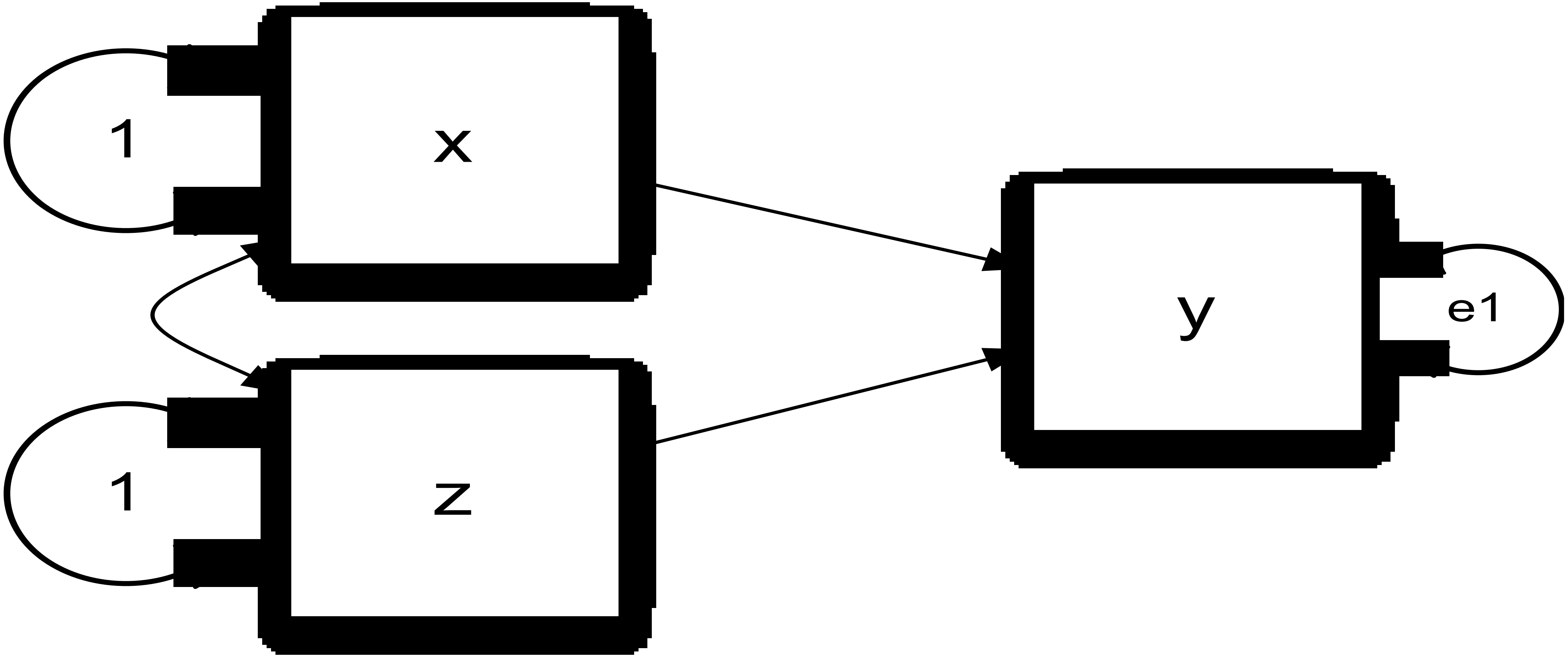
Correlation output

- umxSummary(m1)
- name matrix row col Std.Estimate Std.SE
- 1 x_with_x S x x 1.00 0.14
- 2 y_with_x S x y 0.58 0.12
- 3 y_with_y S y y 1.00 0.14

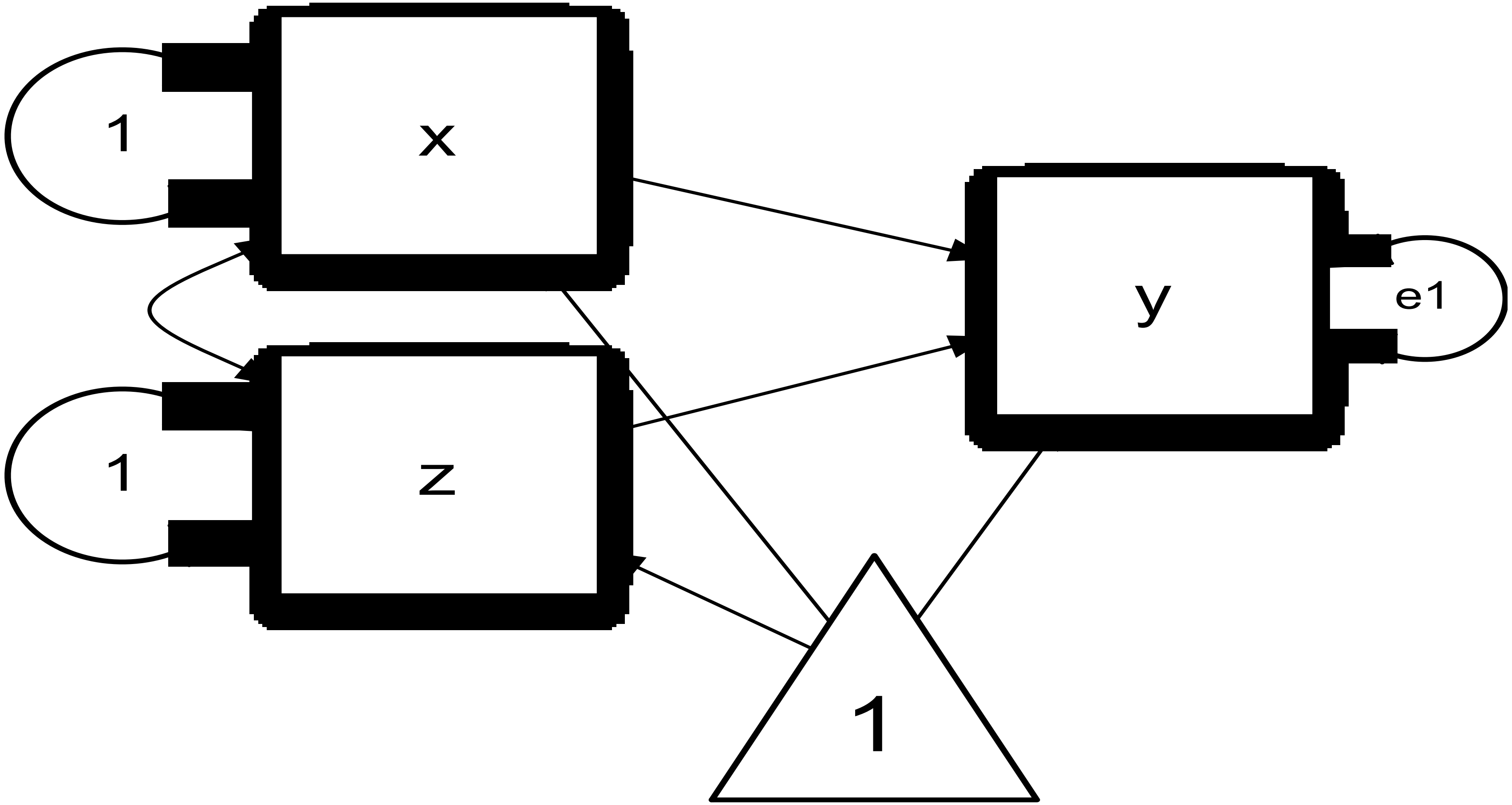
Regression



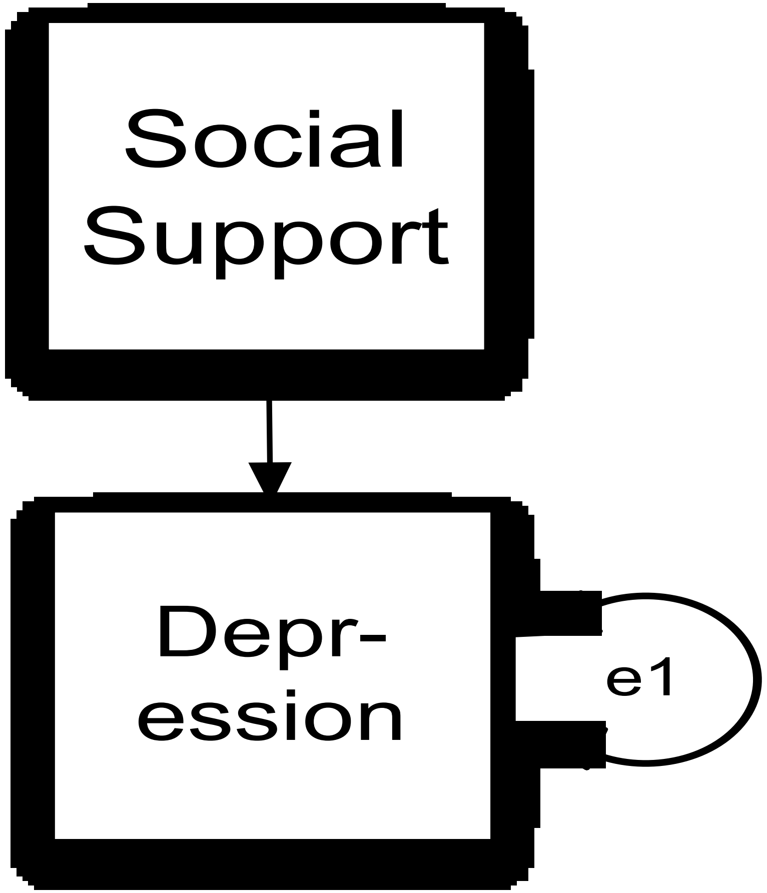
Multiple Regression



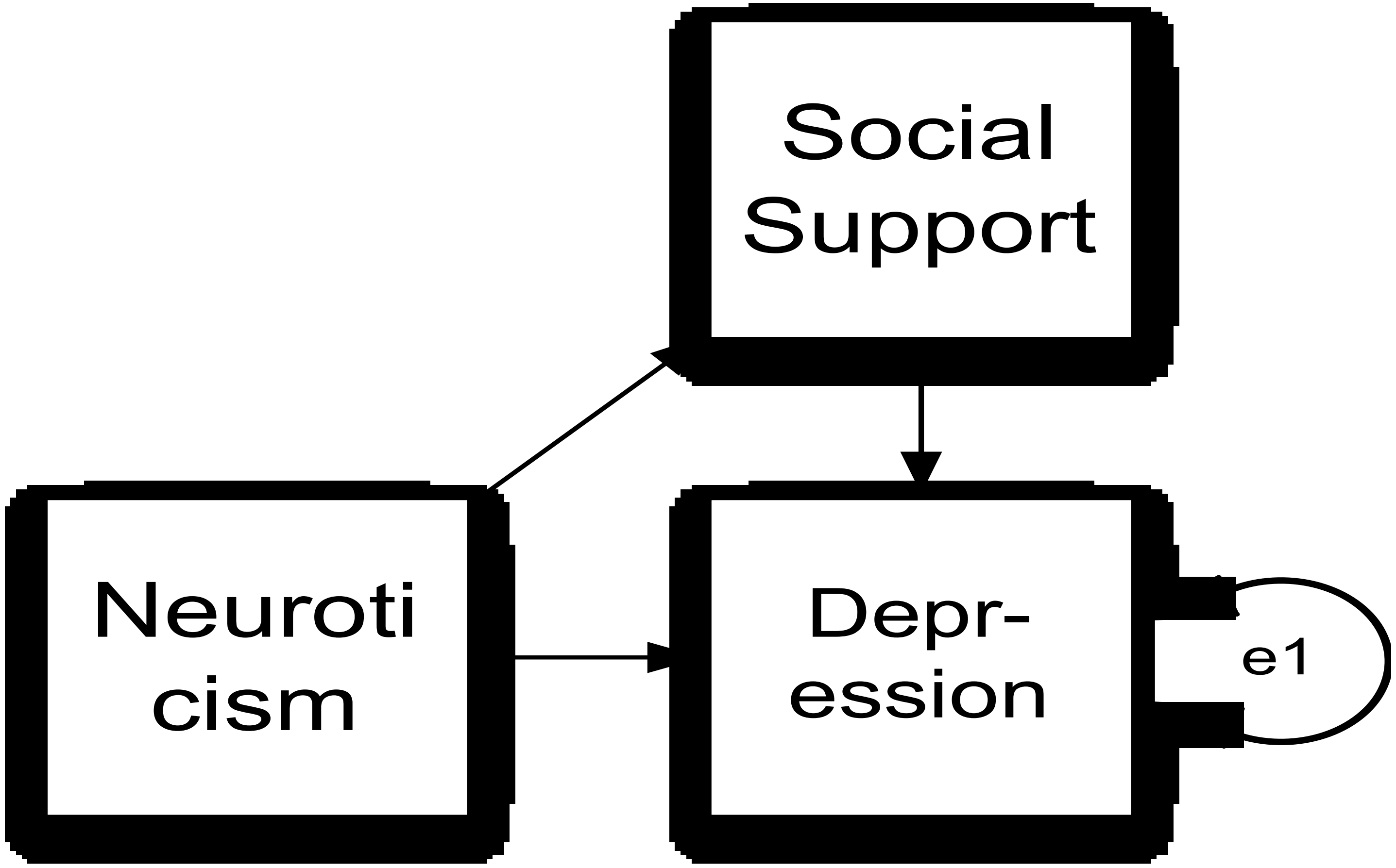
Multiple Regression (raw data)



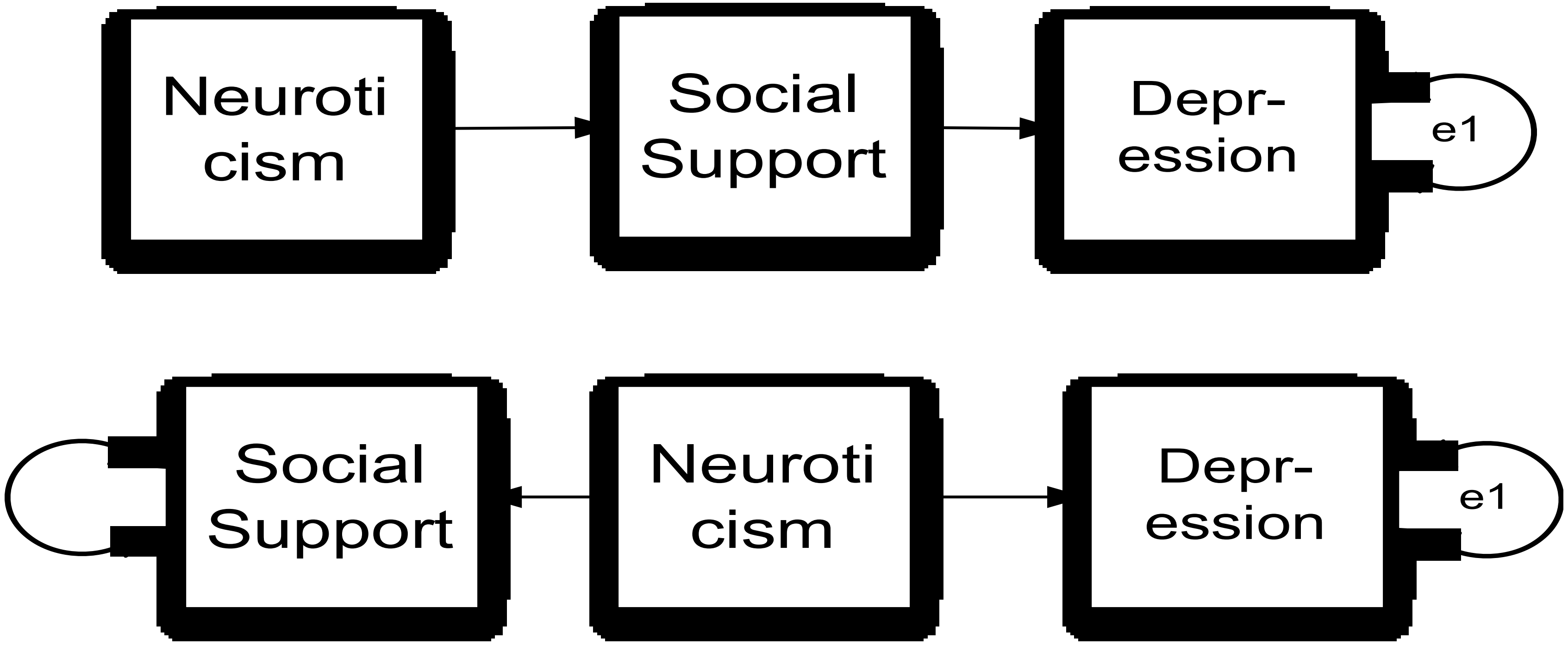
Case Study



What causes what?

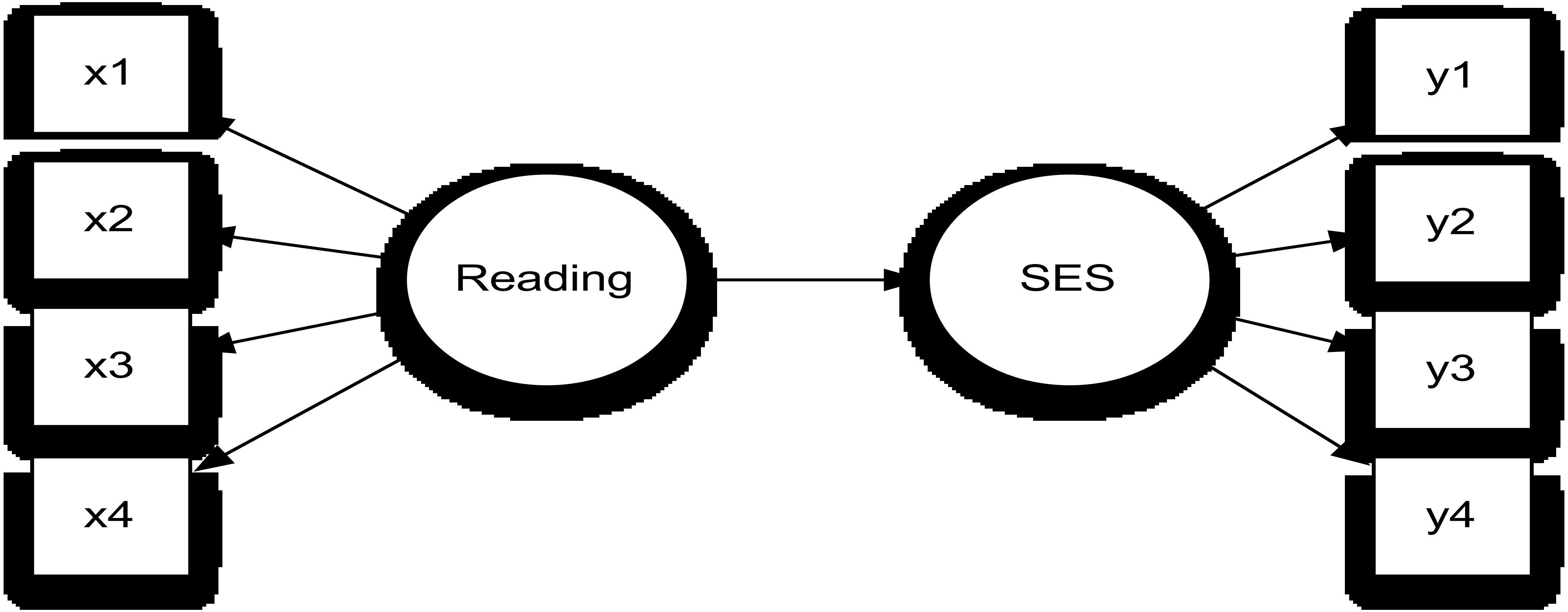


Two very different models (claims about what causes depression)

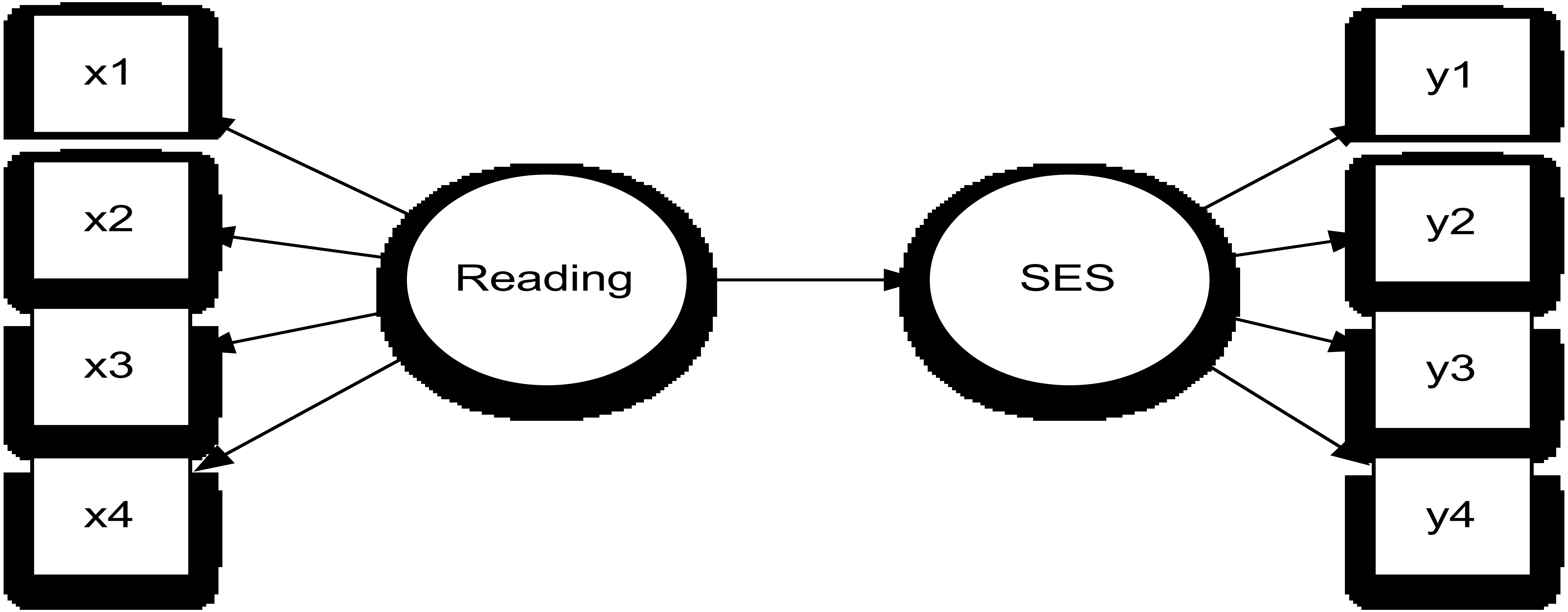


Lewis, G. J., Bates, T. C., Posthuma, D., & Polderman, T. J. C. (2012). No mediating role for perceived social support in the effects of stable personality traits on symptoms of depression and anxiety. *Psychological Medicine*.

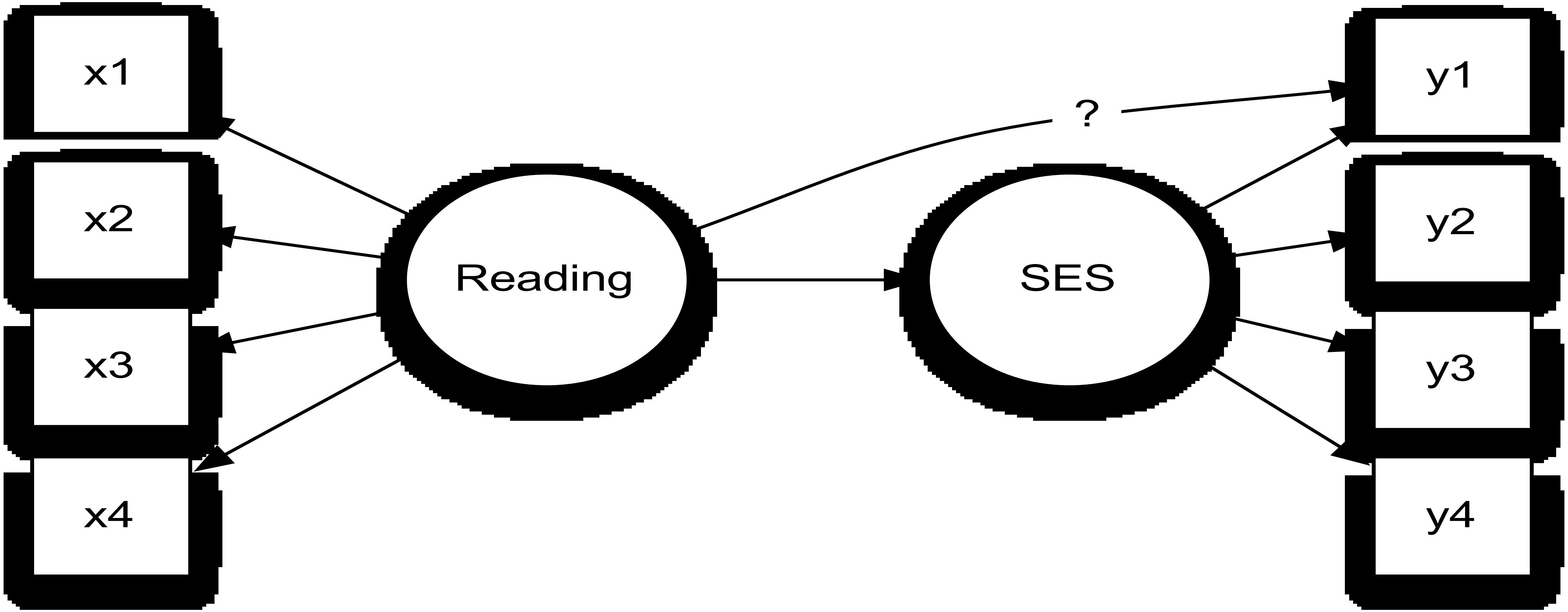
Structural Regression



Does reading raise income?



Specific influences?



A workflow

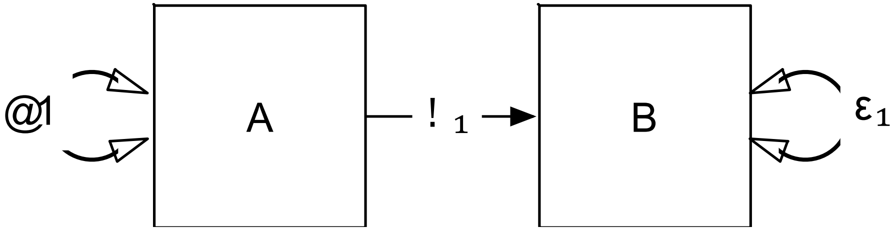
- Get data containing the variables you need
- Build the model
- **umxRAM**(
  - Add regression paths: `“a”, to = “b”`
  - Add variance paths. : `var = “a”`
  - Add covariances. : `“a”, with = “b”` or `cov= c(“a”, “b”)`
  - Add means paths. : `“one”, to = “a”`
  - Add data : data = mtcars
  - *Note: Any variables not in the data are assumed to be latent*
- m1 **= mxRun**(m1) the model if necessary
- View **summary**()

What are circles and boxes and paths?

- _________ latent (unmeasured) variables
- _________ are measured variables
- _________ headed arrows imply a causal path
- _________ headed arrows (drawn curved) allow covariation
- _________ = mean
- _________ = definition variable

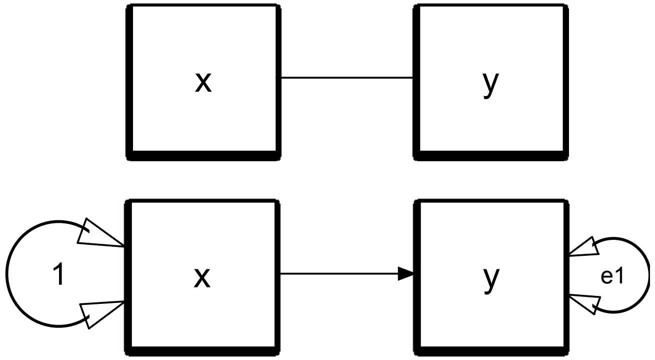
A diagram specifies EVERYTHING needed to estimate the model

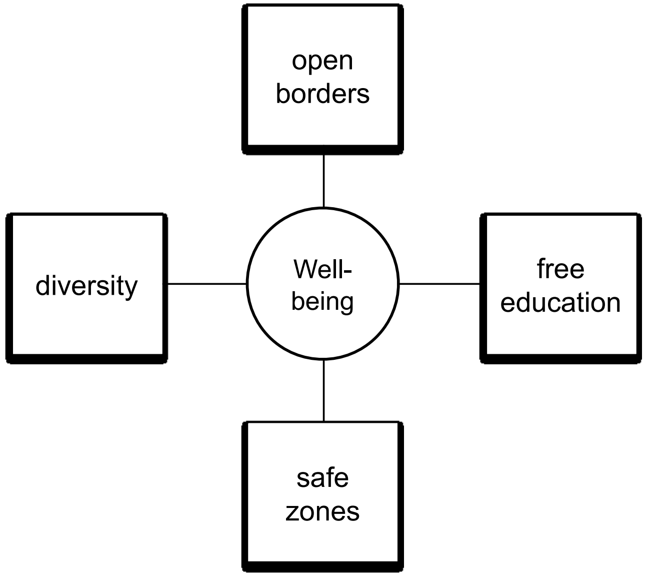
- m1 = lm(B ~ A)



Important notes on what is **not** a structural diagrams

- Paths are not beams or rafters
  - They are causes and they HAVE to have at least 1 head
  - Have to draw residuals (or they are claimed not to be there
- Boxes are not rooms
  - Connecting things you like is not a model, it's just a dream house





Is a structure
Is not a structural model ☺

Endogenous and Exogenous

- Exogenous variables have no incoming paths from other variables.
    - Have variance
- Endogenous: Has arrow pointed at it from another variable
    - Has residual variance
        - Represents unmeasured/unknown variable(s) which influence the endogenous variable…

What does it mean to “run” a model? Optimization

- In SEM, we always have 2 things: data and a model
  - Data gives us observed statistics: Usually means, variances, and covariances
  - The model and any given set of values assigned to its parameters implies a set of statistics about the data: “model implied statistics”

What does it mean to “run” a model?
Optimization

- Using the the observed statistics, and our model (and our ability to move its free parameters), we wish to determine the following three things:

1. What set of values for our parameters that make the model implied statistics most similar to the observed statistics (the data)
2. At this best-set of values, how likely is it that our model-implied statistics are drawn from the same population that generated the model observed statistics?
3. Between any two models, which is the most likely?

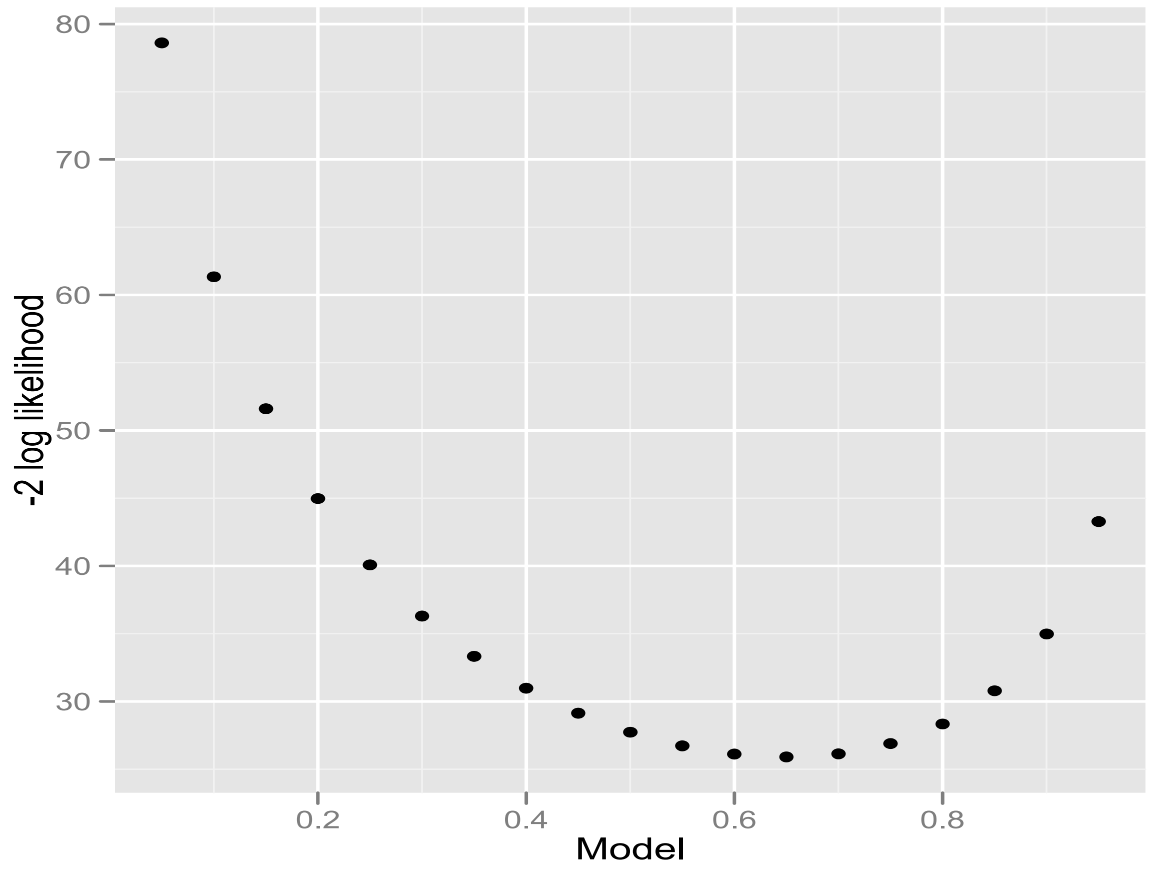
What does it mean to "run" a model? Optimization

- How would you decide where to put parameters?
- Imagine "*hand optimizing*" the following model
  - "Variance of X"
- Start values
- Optimization process
  - Value updating: Value guessing rules, step sizes
  - Assess fit at the new values
  - Decision about whether fit is "as good as it is going to get"
    - Compare to fit on previous estimates
    - Stopping rules, precision settings

Maximum Likelihood estimation (MLE)

- Estimates the parameters of a model given observations
- Parameter values are selected which maximize the likelihood of the data being observed
- We are interested in the population – that's what our model represents
- MLE relies on an optimizer to seek our parameter values that make the observed results the most probable given the model.
- Model returns a likelihood
  - log(likelihood) is distributed χ^2(df)

-2 log(likelihood)



Log Likelihood Example

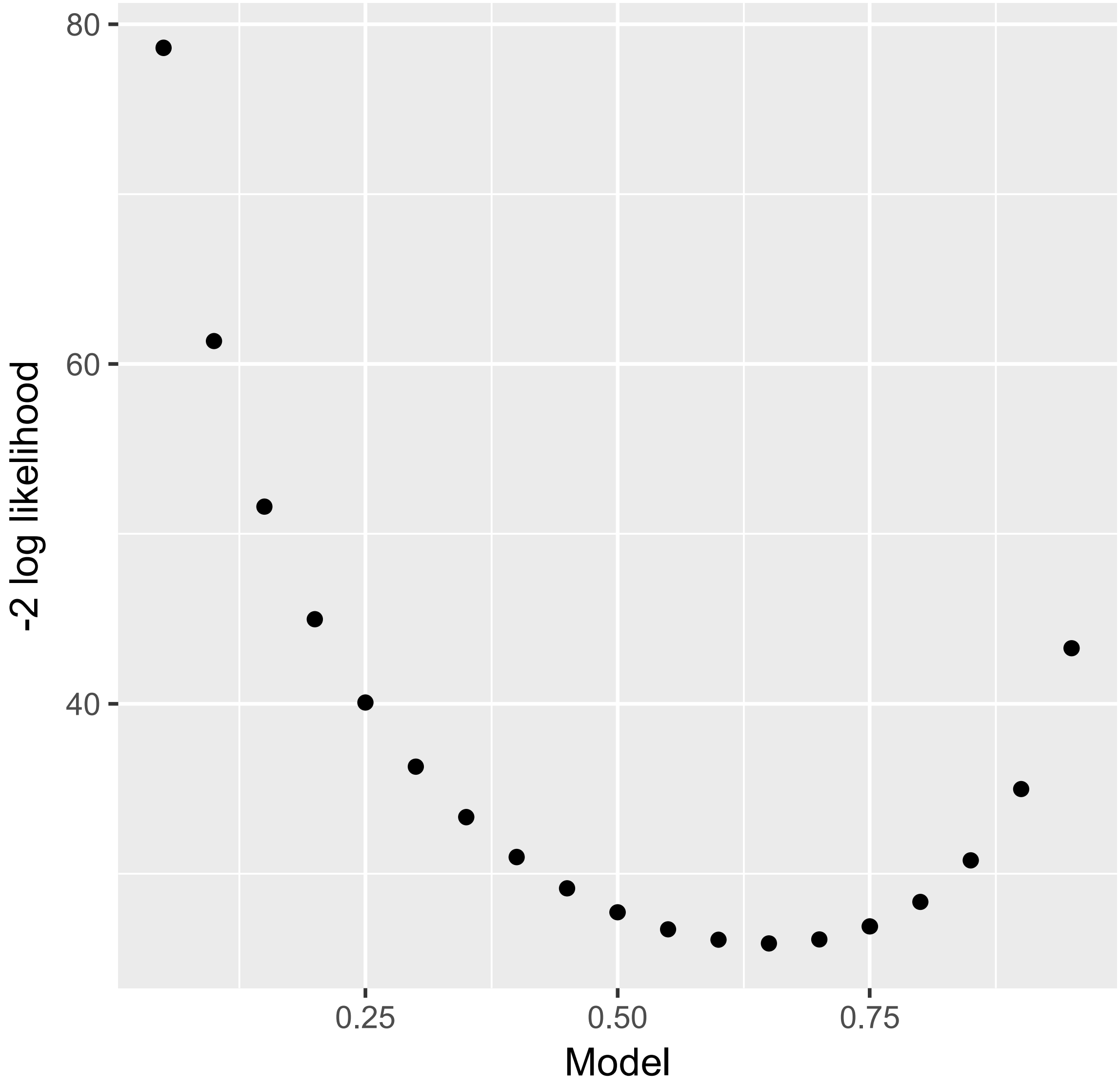
- Imagine we have tossed a coin 20 times
  - Observed 13 heads.
- People propose three competing models of the gambling hall
  - A friend says he wins all the time on heads: p(heads) = 1.0
  - The casino says they are fair : p(heads) = 0.5
  - A competitor says their coins are rigged : p(heads) = 0.0
- The closer the -2LL is to zero, the more likely are the data given the chosen model.
- Let’s use R to calculate the -2LL for all the claimed p(heads), from 0.5 to .95 stepping by .05

Compute log(likelihood)

```r
model_prob = seq(.05, .95, by =.05)
minus2loglikelihood = rep(NA, length(model_prob))

i = 1
for(p in model_prob){
  # -2 Ã— log(likelihood) at a given value of p is:
  # (13 Ã— log(p)) + (17 Ã— log(1 - p))
  ll_p =  13 * log(p)
  ll_q = (20 - 13) * log(1 - p)
  minus2loglikelihood[i] = -2 * (ll_p + ll_q)
  i = i + 1
}

ggplot2::qplot(x = model_prob, y = minus2loglikelihood)
```

Compute -2*log(likelihood)

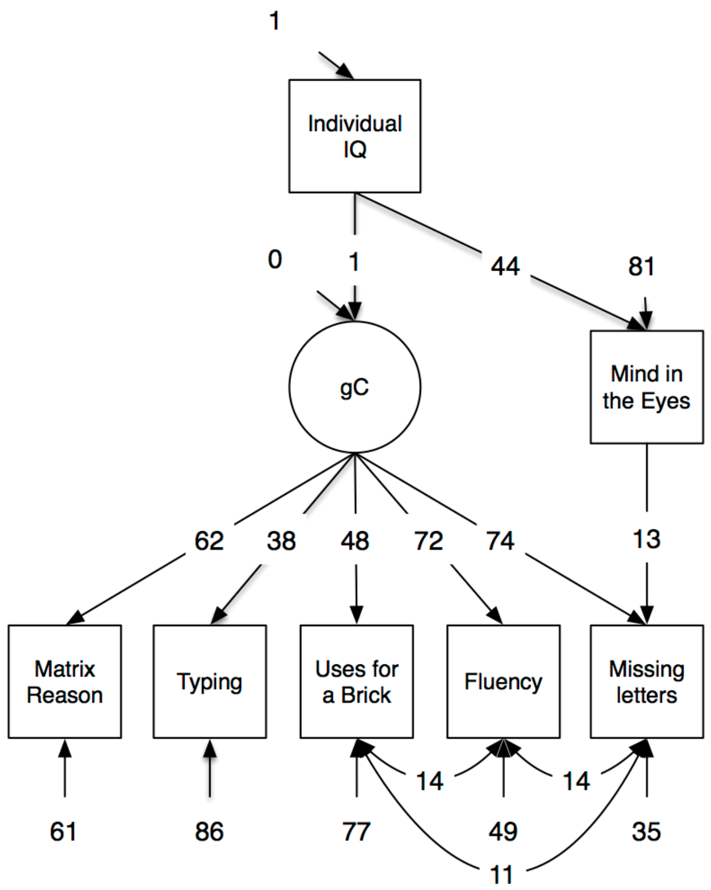
```
model_prob = seq(.05, .95, by =.05)
minus2ll   = rep(NA, length(model_prob))

i = 1
for(p in model_prob){
  ll_p =  13 * log(p)
  ll_q = (20 - 13) * log(1 - p)
  minus2ll[i] = -2 * (ll_p + ll_q)
  i = i + 1
}

qplot(model_prob, minus2ll)
```



Model Fit: How do we assess if a model fits “*well*” or “*poorly*”, or “*better*”?

- Model fit of this specification can be evaluated by how closely the covariance of the data is recreated (likelihood), penalised by how many parameters we used.
- Key insight: Bad models don’t allow very likely solutions…

If the theory that empathy causes collective IQ was true, this model would fit badly



Model Fit: How do we assess if a model fits "*well*" or "*poorly*", or "*better*"?

- Key insight: Bad models don't permit very likely solutions…
- We can compute the likely penalized fit of a good model, and thus set out criteria for good fit
    - Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6, 1-55.
    - Yu, C.Y. (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. University of California, Los Angeles, Los Angeles. Retrieved from http://www.statmodel.com/download/Yudissertation.pdf

The Chi Square Test: χ2

- -2xLL is distributed as χ^2 with df degrees of freedom
- X^2 alone generates an excess of Type 1 errors due to
  - non-normality
  - small sample size
- With large sample size, it is too conservative (Type II)

Building on χ^2

- χ^2/df takes into account these limitations, but fails to adjudicate on good fit
- TLI and RMSEA build on this, with criteria for goodness (and inadequacy) of fit

Concept of worst and best fitting possible models

- Imagine we only model the mean and variance of each variable
  - *Independence* model AKA *null* model
- Imagine we add covariances between all possible variable pairings
  - *Saturated* model

- **Incremental Fit** (AKA ***relative fit***)
  - Fit relative to explaining nothing
  - Proportion of variance from 0 to 1
    - (null – myModel) /(null – Saturated)

Tucker-lewis Index (AKA NNFI)

$$\frac{\chi^2/\mathrm{df}(\text{Null Model}) - \chi^2/\mathrm{df}(\text{Proposed Model})}{\chi^2/\mathrm{df}(\text{Null Model}) - 1}$$

- Along with CFI, RMSEA, this is a core fit index
- X^2, penalized for degrees of freedom
- Good fit = TLI > .95

Root Mean Square Error of Approximation (RMSEA)

$$\frac{\sqrt{(\chi^2 - df)}}{\sqrt{[df(N - 1)]}}$$

- N = sample size
- *df* = model df
- Lower Bounded at 0.
- Good fit = RMSEA < .06

Root-mean square error of approximation

- RMSEA (Steiger & Lind, 1980)
  - (relatively) independent of sample size.
  - an index of being “close” to fitting the data.

```
RMSEA = sqrt((-2*ln(likelihood)-df)/(N*df) )
    # < 0.06 is “good” fit and > 0.10 is ”poor” fit.

AIC = -2 * ln(likelihood) + 2(p + 1)
    # lower better
```

Akaike Information Criterion (AIC)

- Comparative measure of fit
  - Only meaningful when comparing models.
  - Model with the lower AIC is preferred
  - Several version exist: as long as you use the difference, all are the same
- $\chi^2 + k(k + 1) - 2df$
- k = number of variables
- AIC penalizes likelihood by 2 for each parameter added

```
AIC = -2 * ln(likelihood) + 2(p + 1)
    # lower better
```

Effects: Path tracing rules

The expected correlation due to a chain traced between two variables is the product of the standardized path coefficients in the chain

The total expected correlation between two variables is the sum of these contributing path-chains.

- Caveats
  - Wright's rules assume a model without feedback loops: the directed graph of the model must contain no cycles.
- See also Boker 2005 (pdf)
  - BG handbook

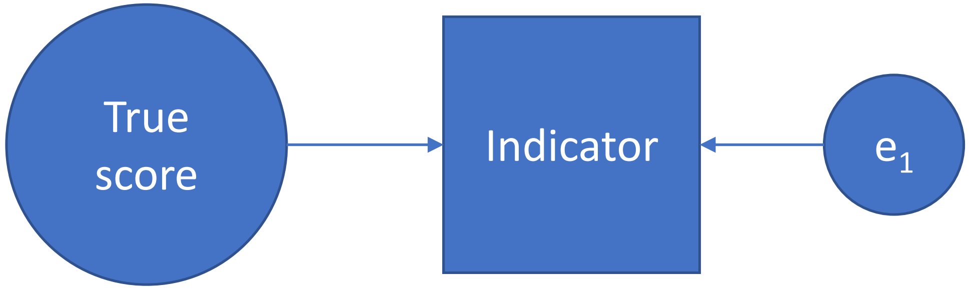
Effects: Path tracing rules

- The rules for path tracing are:

1. You can trace backward up an arrow and then forward along the next, or forwards from one variable to the other, but never forward and then back.
   - You can never pass out of one arrow head and into another arrowhead
   - Heads-tails, or tails-heads, not heads-heads.
2. You can pass through each variable only once in a given chain of paths.
3. No more than one bi-directional arrow can be included in each path-chain.

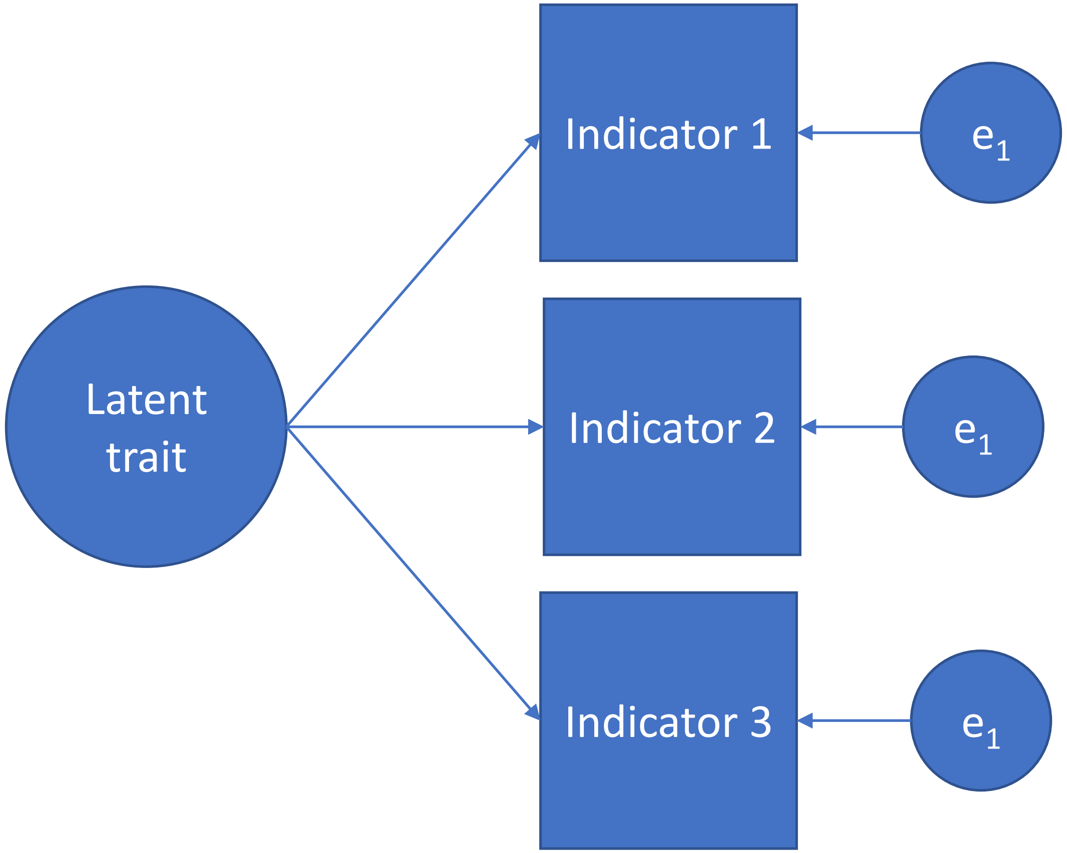
How can we measure error?

- We only ever measure manifestations of a theorized cause
- Things we measure are caused by what we want to measure (the true score) and “error”



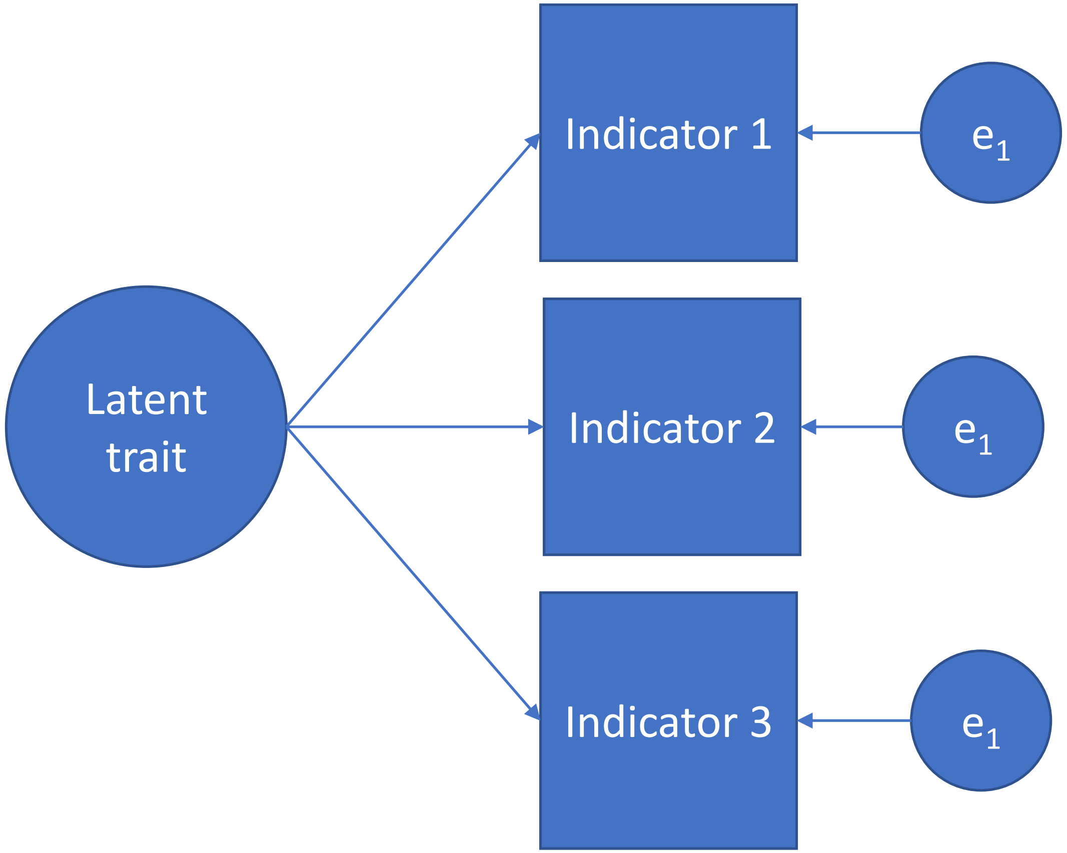
How can we measure error?

- We are interested in variables that we measure via indicators, each of which has errors
- Multiple indicators allow us to separate our trait of interest from the error in our indicators



How can we measure without error?

- We are interested in variables that we measure via indicators, each of which has errors
- Multiple indicators allow us to separate our trait of interest from the error in our indicators



Next week: Missing data, constraints, equating, algebras, matrices and twins …

- Often some data are missing.
- lm() and friends drop the entire row ☹
- Full information maximum likelihood modeling in SEM uses everything
    - Calculates the fit of the model for each row with its particular missingness (or rows with a shared pattern of missingness) using raw data

References

- Duncan, O. D.(1966). Path analysis: Sociological examples. *The American Journal of Sociology*, **72**, 1–16.
- Goldberger, A. S.(1972). Structural equation methods in the social sciences. *Econometrica*, **40**, 979–1001.
- Jöreskog, K. G.(1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85–112). New York: Seminar.
- Spearman, C.(1904a). General intelligence objectively determined and measured. *American Journal of Psychology*, 15, 201–293.
- Spearman, C.(1904b). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72–101.
- Wright, S.(1918). On the nature of size factors. *The Annals of Mathematical Statistics*, **3**, 367–374.
- Wright, S.(1920). The relative importance of heredity and environment in determining the piebald pattern of guinea–pigs. *Proceedings of the National Academy of Sciences*, **6**, 320–332.
- Wright, S.(1934). The method of path coefficients. *The Annals of Mathematical Statistics*, **5**, 161–215.