methods@manchester

What is Structural Equation Modelling?

Nick Shryane

Institute for Social Change

University of Manchester

Topics

- Where SEM fits in the families of statistical models
- Causality
  - SEM is useful for representing causal models, but can't demonstrate causality on its own
- Useful applications:
  - measurement error, missing data, mediation models, group differences

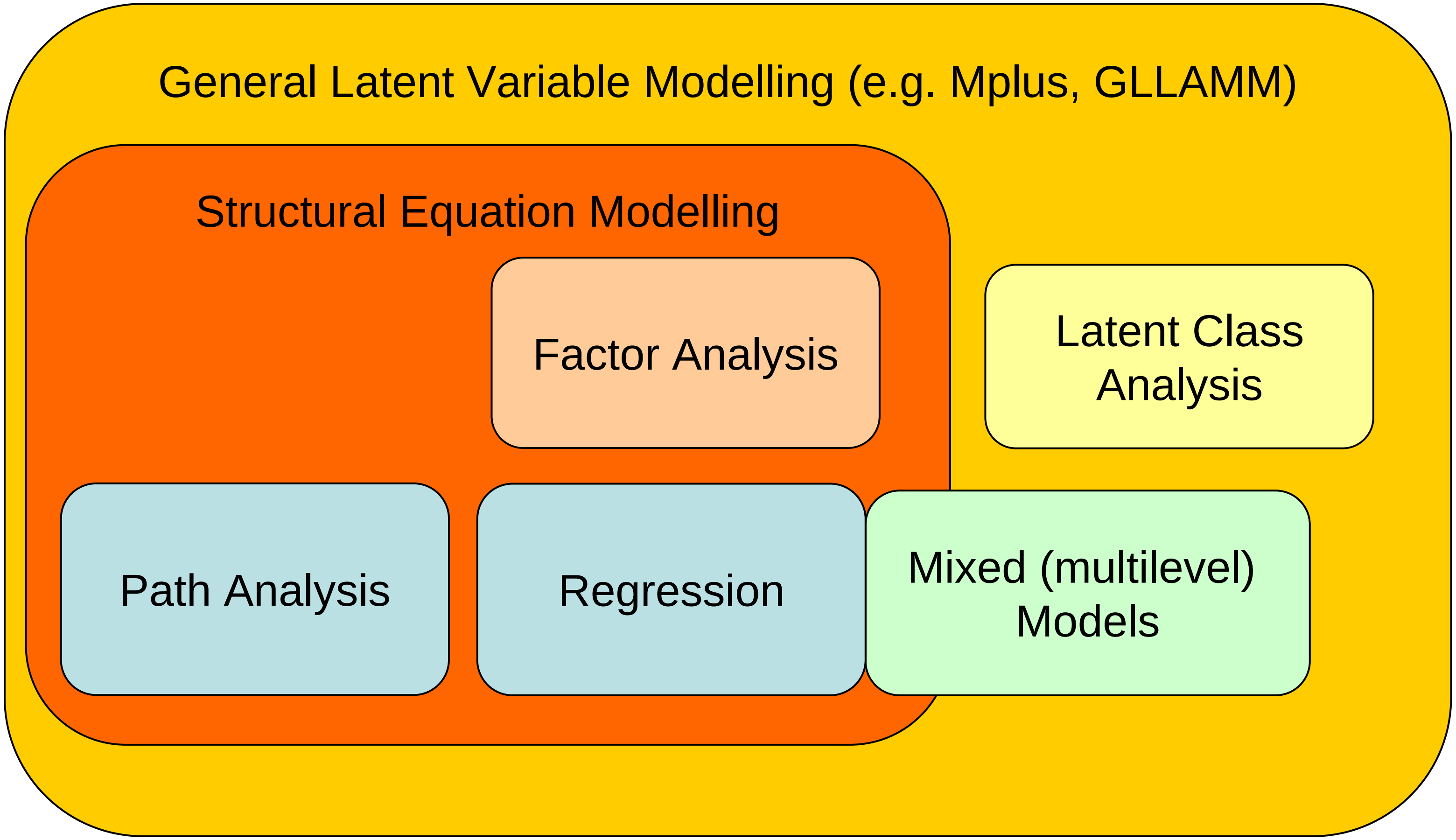
MANCHESTER 1824

The Institute for Social Change

History

- Structural Equation Modelling
  - Was cobbled together out of
    - Regression
    - Path Analysis (Wright, 1921)
    - Confirmatory Factor Analysis (Jöreskog, 1969)

Families of Statistical Models



History

- Structural Equation Modelling
  - Used to be known as:
    - Covariance Structure Modelling
    - Covariance Structure Analysis
    - Causal Modelling

Causality

$$\begin{tikzcd} X \arrow[rr, bend left, leftrightarrow] & & Y \end{tikzcd}$$

We observe a correlation between two variables. Why?

1. X causes Y ?

$$X \longrightarrow Y$$

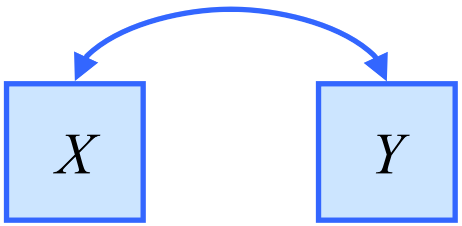
2. Y causes X ?

$$X \longleftarrow Y$$

3. Reciprocal causation ?

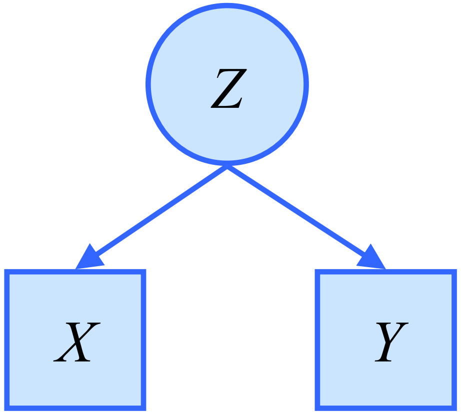
$$X \longleftrightarrow Y$$

Causality



We observe a correlation between two variables. Why?

4. A third, unmeasured variable ?



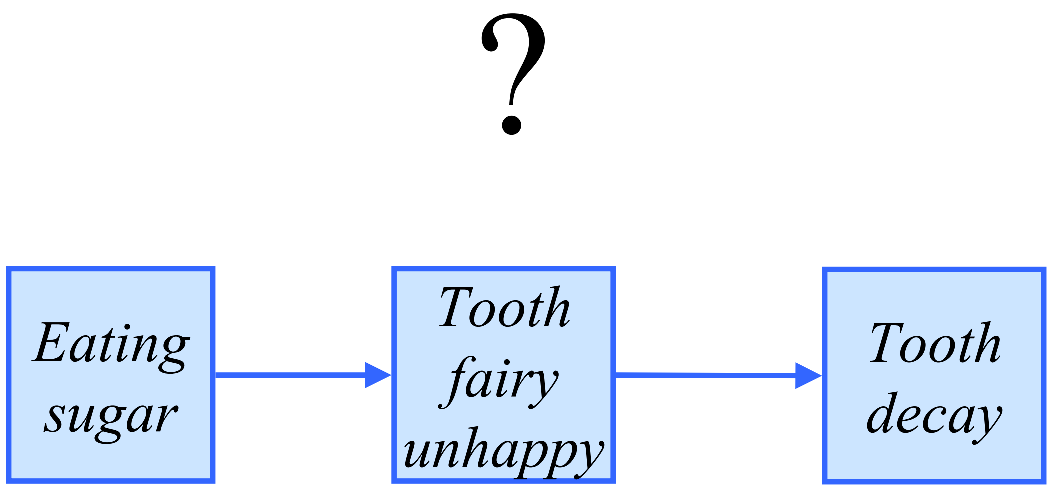
Whichever is ‘true’, there will be a whole causal chain of mechanisms between the supposed cause and effect variables.

Causality

Causal statement:
Eating sugary foods causes tooth decay

Causality

If the causal statement is true, then there must be a causal mechanism…

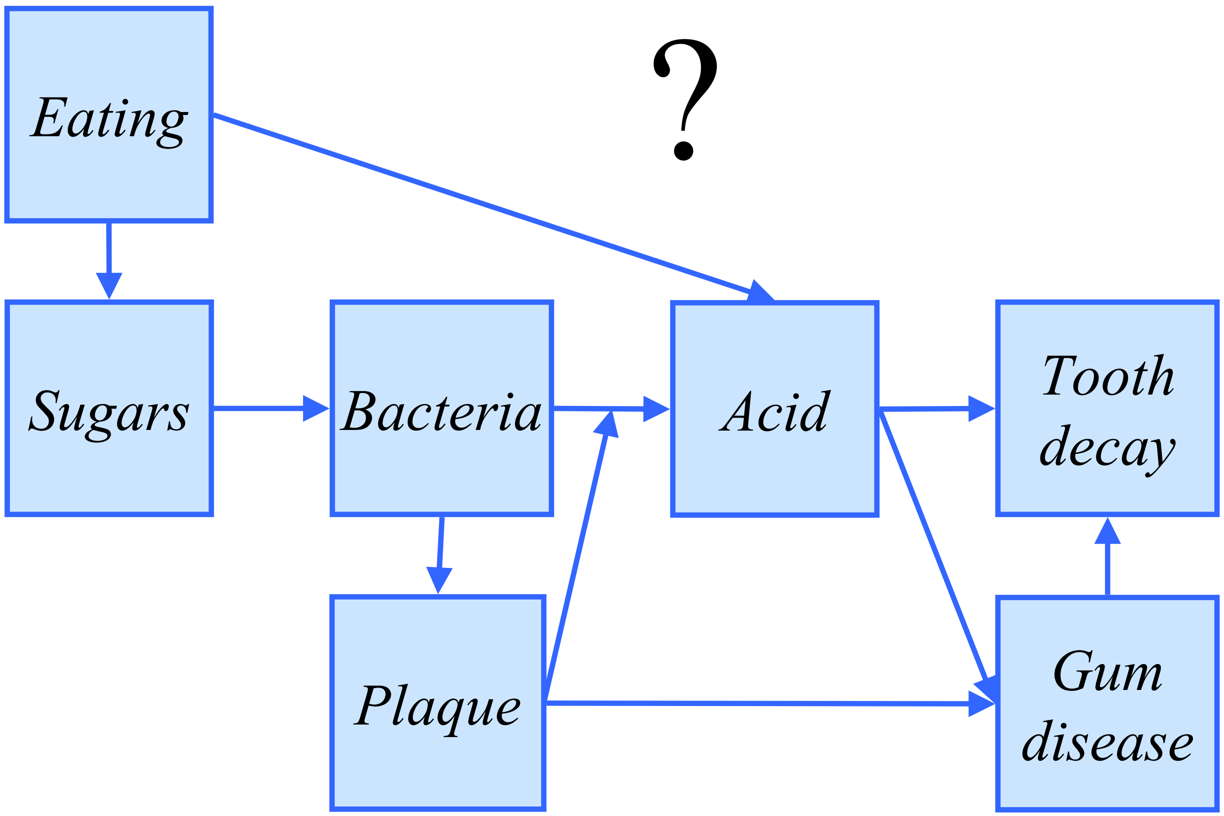


MANCHESTER 1824

The Institute for Social Change

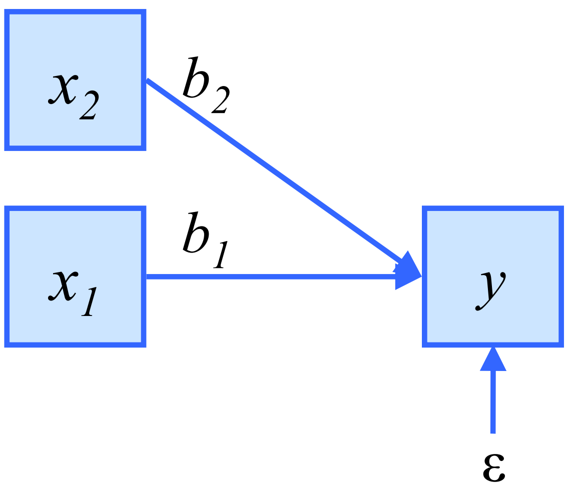
Causality

If the causal statement is true, then there must be a causal mechanism…



Regression Models

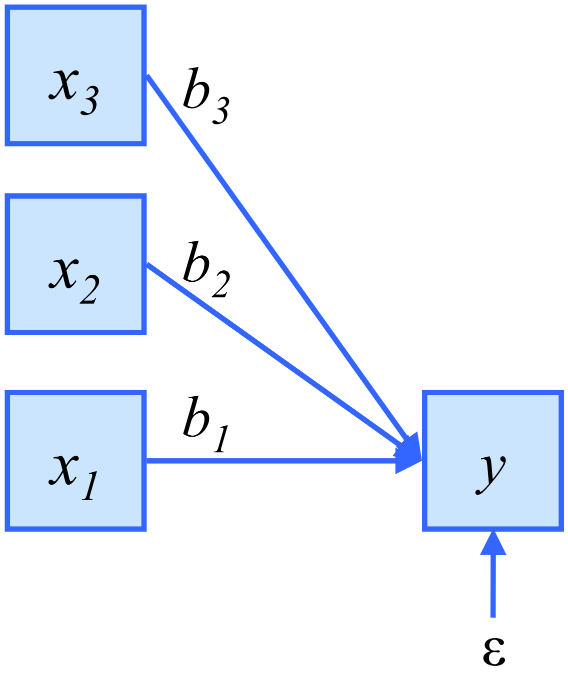
$$y = x_1 b_1 + x_2 b_2 + e$$



Regression is not well suited to describing these causal sequences

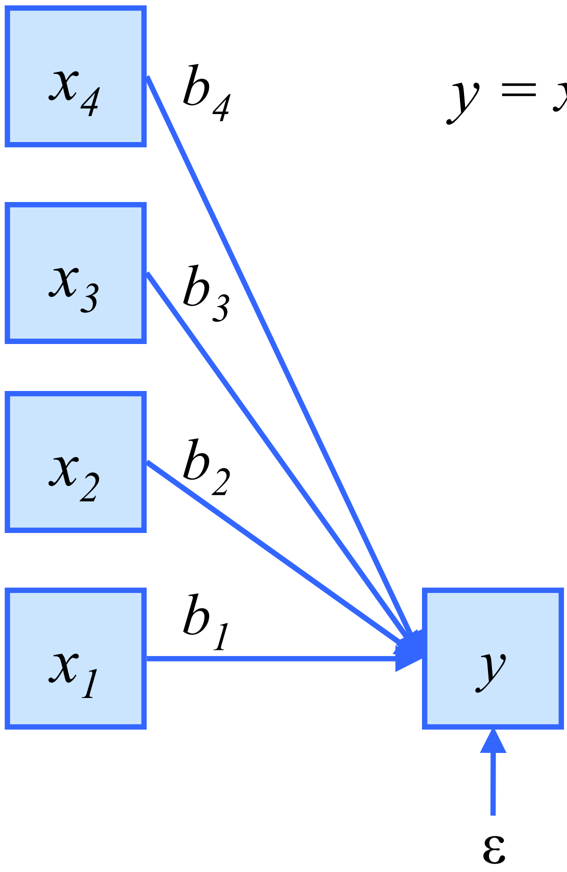
Regression Models

$$y = x_1b_1 + x_2b_2 + x_3b_3 + e$$



Adding new explanatory variables makes the model more comprehensive and complicated…

Regression Models



$$y = x_1b_1 + x_2b_2 + x_3b_3 + x_4b_4 + e$$

…but doesn't easily admit causal sequences or unmeasured causes

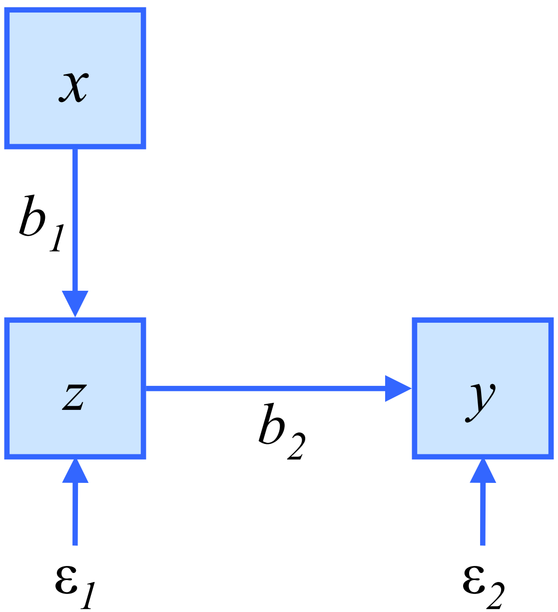
MANCHESTER 1824

The Institute for Social Change

Path Models

$$z = xb_1 + e_1$$ (z predicted by x)

$$y = zb_2 + e_2$$ (y predicted by z)



Path models allow us to fit chains of conditional relationships (here a **mediation** hypotheses, i.e. that *z* mediates the relationship between *x* and *y*)

MANCHESTER 1824

The Institute for Social Change

Path Models

- Regression analysis - does x affect/predict y?
- Path/mediation analysis – **how/why** does x affect/predict y? Via the action of some intervening variable Z?
- Brushing your teeth (*x*) reduces tooth decay (*y*) by removing bacteria (*z*)
  – Measuring and testing mediators can help in evaluating a causal hypothesis

Factor Models

Variables may be correlated due to the action of unobserved influences.

Sometimes these are confounding variables, but many constructs of interest are not directly observed (or even observable)

Unobserved Construct	Observed Measures
Social Capital	Bowling club membership
	Local newspaper reading
Ethnic prejudice	Housing segregation
	Ethnic intermarriage

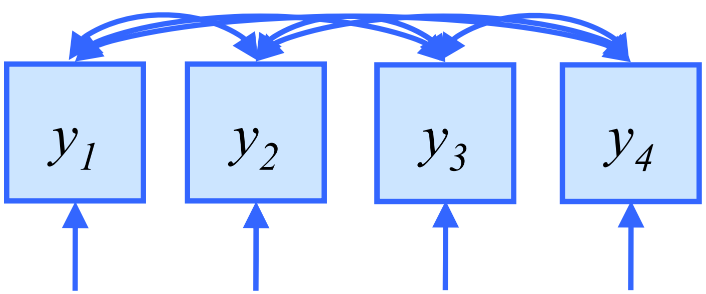
MANCHESTER 1824

The Institute for Social Change

Factor Models

Correlations may not be due to causal relations among the observed variables at all, but due to these unmeasured, latent influences - factors

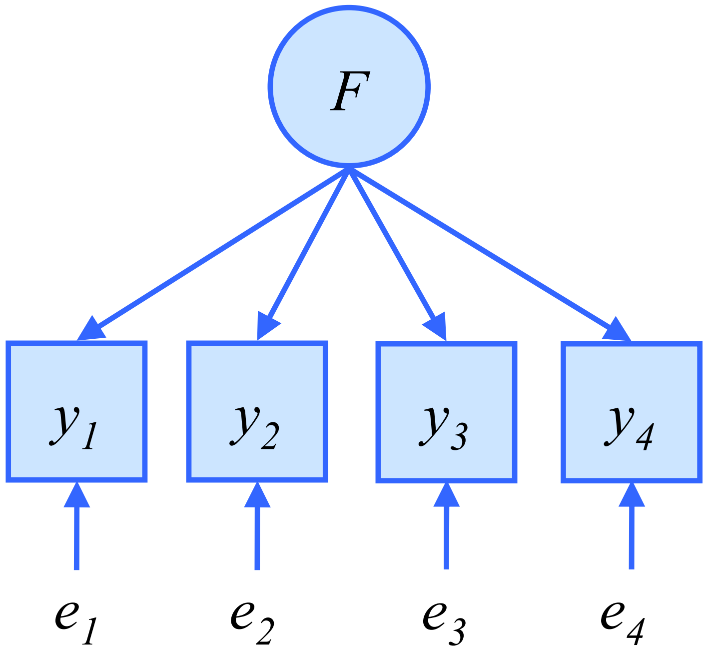
	y_1	y_2	y_3	y_4
y_1	1.0			
y_2	0.6	1.0		
y_3	0.7	0.6	1.0	
y_4	0.5	0.6	0.8	1.0



Factor Models

The observed correlations may due to each observed measure sharing an unobserved component (F)

$$y_1 = F + e_1$$
$$y_2 = F + e_2$$
$$y_3 = F + e_3$$
$$y_4 = F + e_4$$



Factor Models

1	F	e1	e2	e3	e4		y1	y2	y3	y4
2	1.2	-0.4	0.2	-1.5	-1.4		0.8	1.4	-0.3	-0.2
3	3.3	0.8	-0.2	-0.1	0.9		4.1	3.1	3.2	4.2
4	2.2	0.8	-1.8	0.0	1.5		3.0	0.4	2.2	3.7
5	1.3	0.6	-1.9	0.3	1.0		1.9	-0.6	1.6	2.3
6	1.5	-0.9	0.1	1.6	1.0		0.6	1.6	3.1	2.5
7	1.6	-1.5	1.0	0.5	-0.4		0.1	2.6	2.1	1.2
8	2.2	1.5	1.2	-0.7	0.7		3.7	3.4	1.5	2.9
9	2.1	-0.6	0.7	0.1	0.2		1.5	2.8	2.2	2.3
10	0.7	0.3	0.2	-0.4	1.5		1.0	0.9	0.3	2.2
11	1.9	0.5	-1.3	0.2	-0.1		2.4	0.6	2.1	1.8

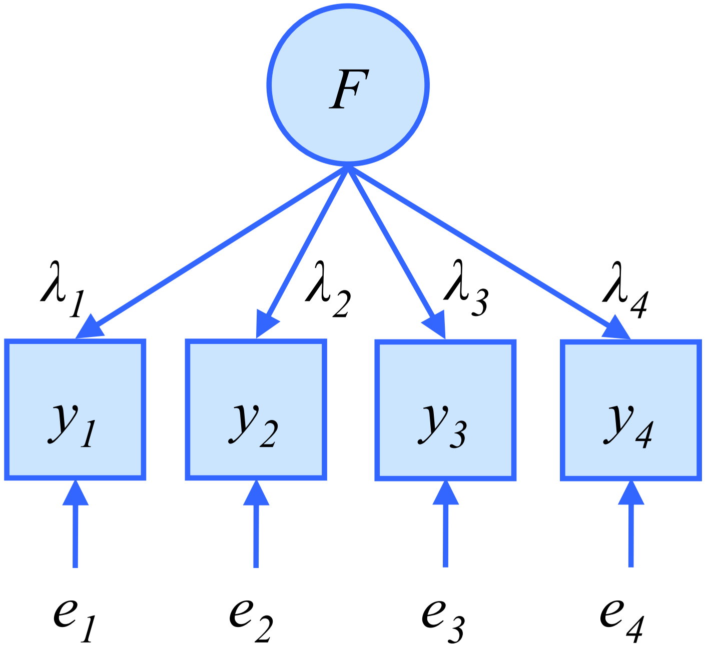
Hypothesised Factor Model

Observed data

Factor Models

The full factor model allows the strength of relationship between F and the observed 'indicators' *y* to vary – different loadings (λ).

$$y_1 = F\lambda_1 + e_1$$
$$y_2 = F\lambda_2 + e_2$$
$$y_3 = F\lambda_3 + e_3$$
$$y_4 = F\lambda_4 + e_4$$



Structural Equation Models

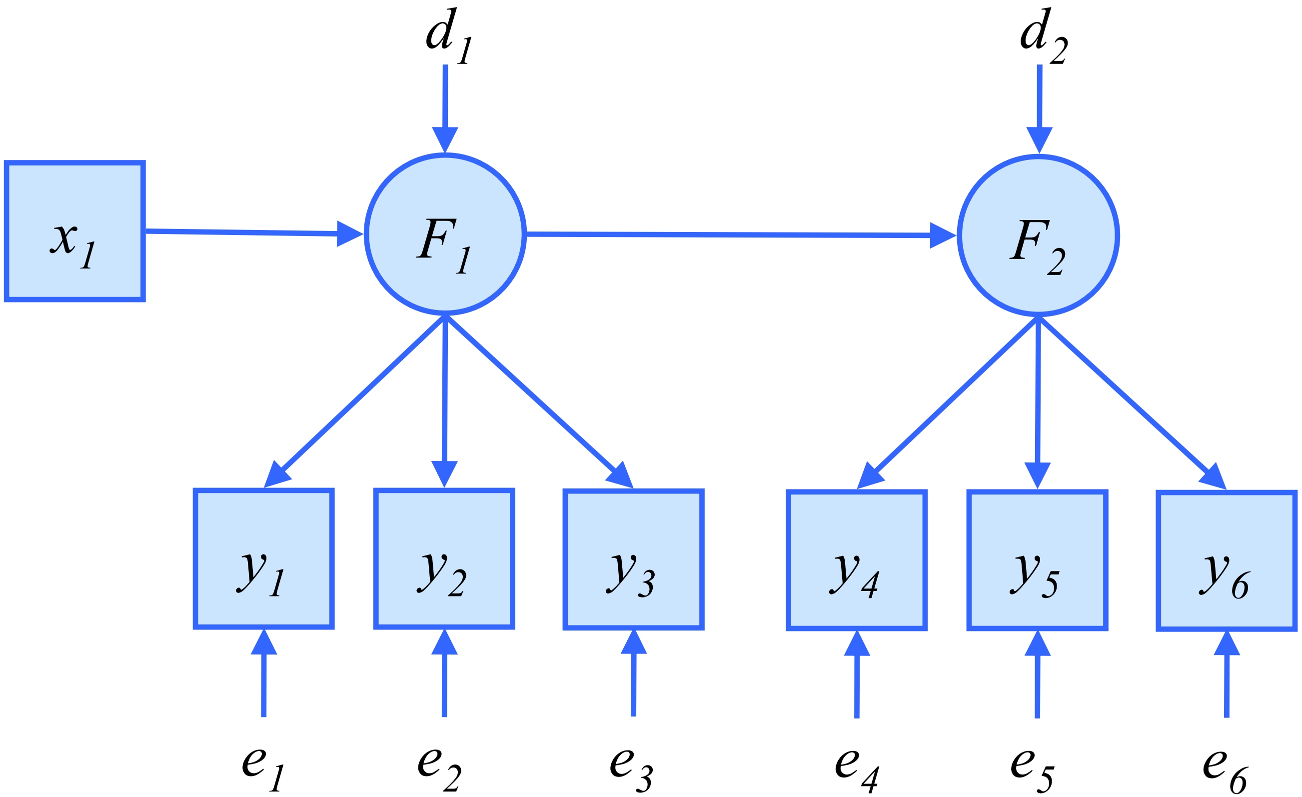
The full Structural Equation model is a combination of some or all of these elements:

- Regression model
- Path model
- Factor model

MANCHESTER 1824

The Institute for Social Change

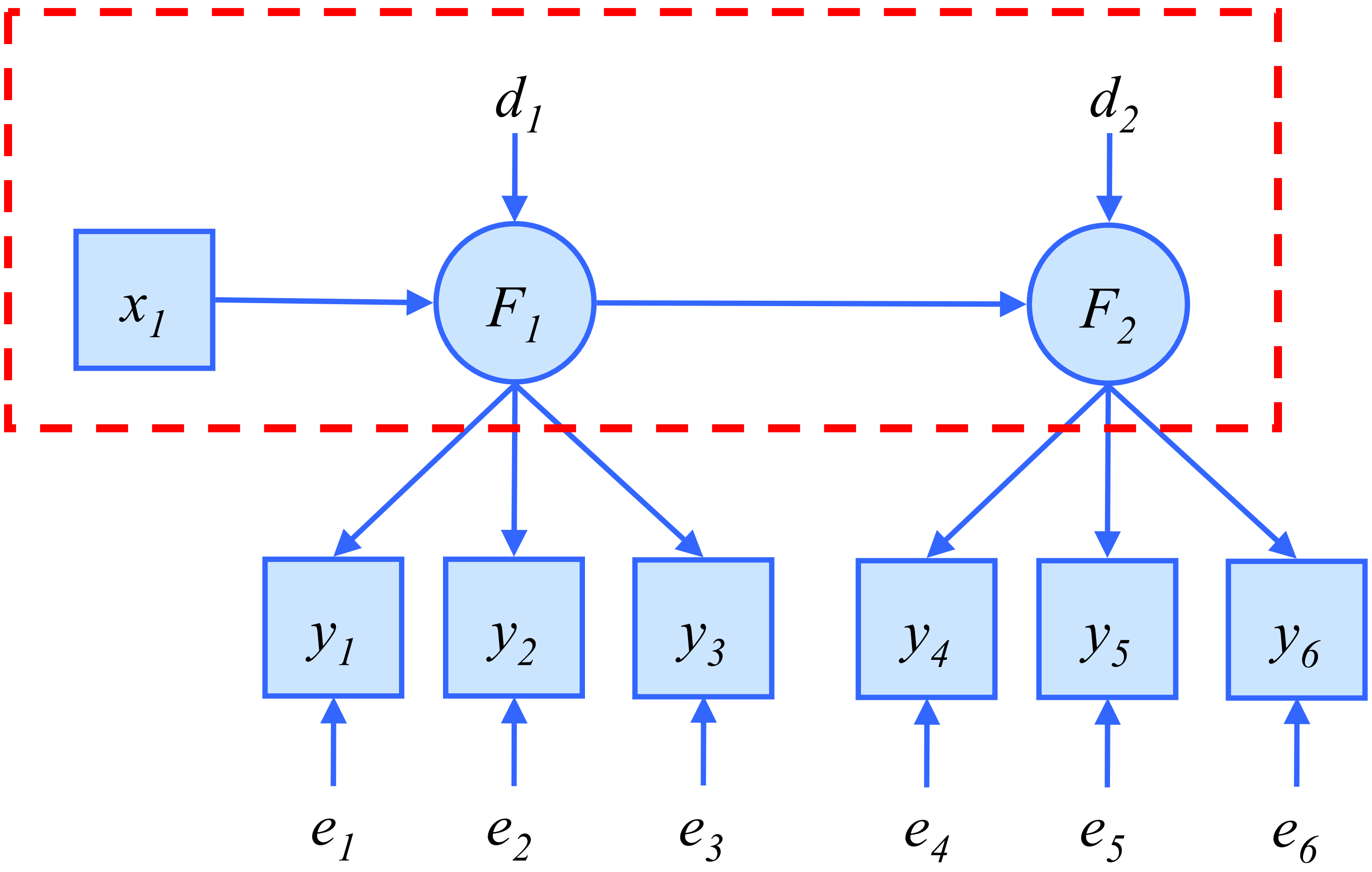
Structural Equation Models



MANCHESTER 1824

The Institute for Social Change

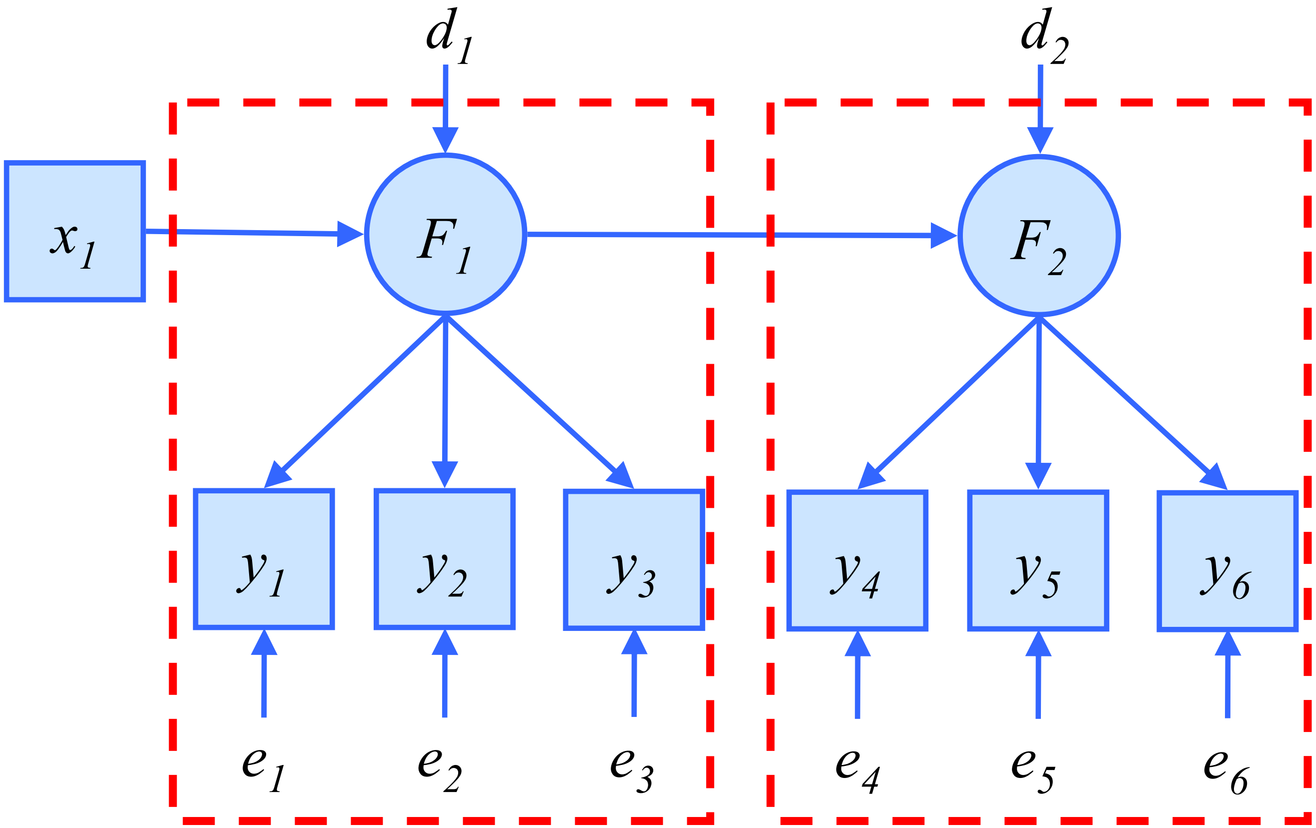
Structural Equation Models



Structural part

(Regression / Path Analysis)

Structural Equation Models



Measurement part

(Factor Analysis)

Model Fit

- Often the goal of the analysis is to assess the plausibility of the model as a whole
- Some aspects of plausibility are nothing to do with statistics
    - If I claim X →Y, does that make sense? Does eating turkey cause Christmas…?
- Plausibility is often assessed by the ability of the model to reproduce or 'account for' the observed variances and covariances.

MANCHESTER 1824

The Institute for Social Change

Model Fit

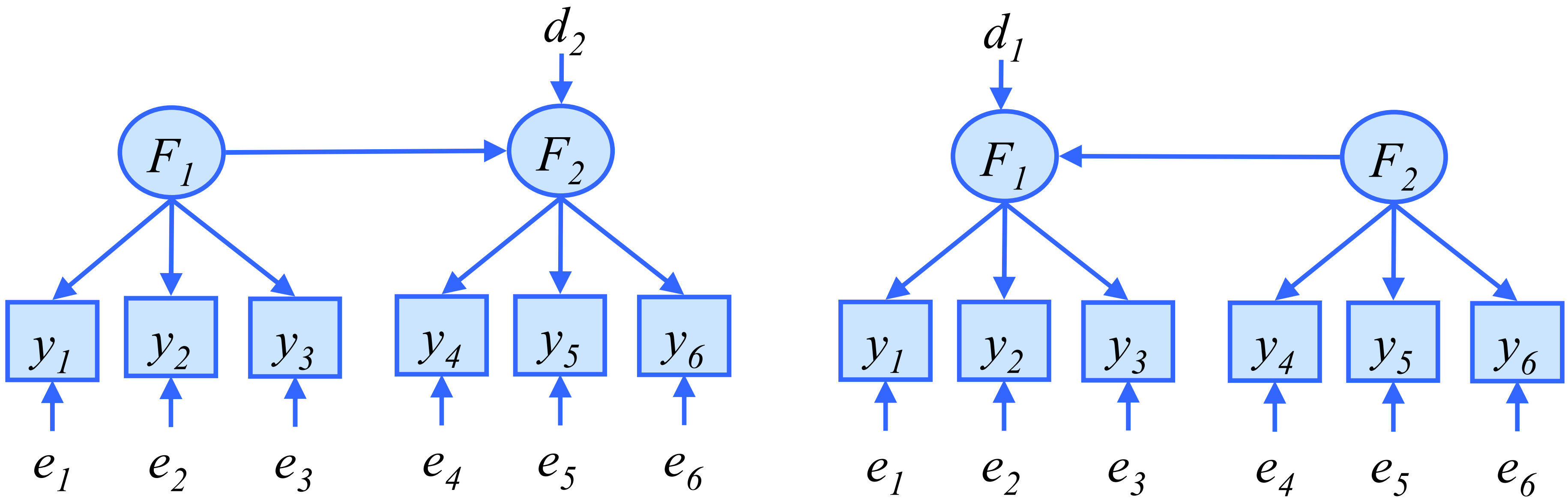
- Many indices have been suggested to assess global model fit to the observed data, e.g.
  - Comparative Fit Index (CFI; Bentler, 1990)
  - Root Mean Square Error of Approximation (RMSEA; Brown & Cudeck, 1993; Steiger, 1990)
- These suggest whether the model is statistically plausible, not whether it is 'true'

Causality again

- Ultimately, the SEM is a **hypothesis**, which is either supported or not by the data
  - Poor model fit
    - Some aspect of the model (structural/measurement) is not a plausible description of the 'data-generating mechanism'
  - Good model fit
    - The model is a plausible, candidate explanation
    - But so might be lots of equivalent or alternative models:
      - MacCallum, R. C., Wegener, D. T., Uchino, B. N. et al. (1993). The problem of equivalent models in applications of covariance structure analysis. Psychological Bulletin, 114(1), 185-199.

Equivalent Models

- These two models would give **identical** fit



- A well fitting model does not, on it's own, mean that the causal hypothesis is true

MANCHESTER 1824

The Institute for Social Change

Applications: Covariate Measurement Error

- Standard regression analysis assumes covariates/ predictors are measured without error
- Effects are biased (weaker); more error, greater bias
- With multiple indicators you can build a factor model to 'measure' your construct, excluding error

MANCHESTER 1824

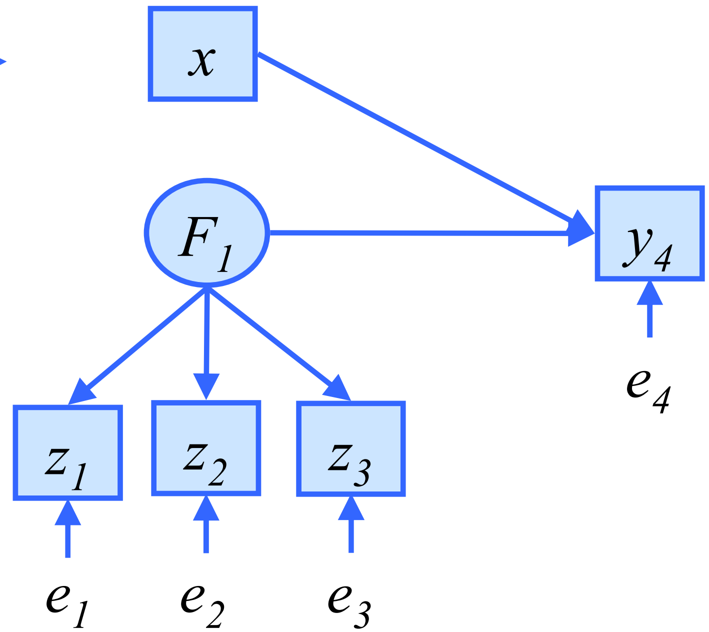
The Institute for Social Change

Applications: Covariate Measurement Error

Covariate (x) assumed error free (e.g. sex)

Factor represents shared z variance

Unique/error z variance tucked away here



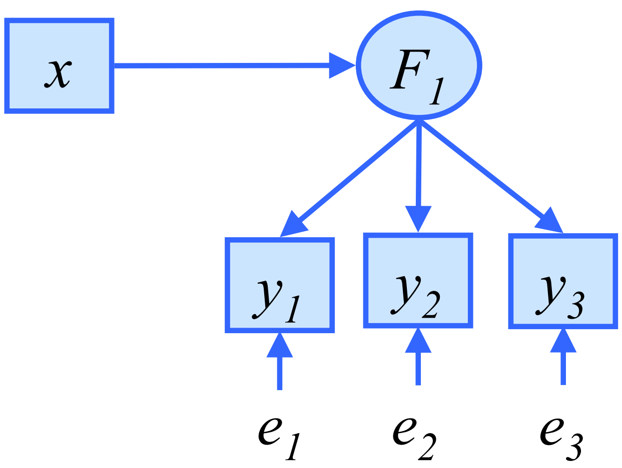
(This model suitable for multicolinearity too)

Applications: Correlated outcomes with missing data

- Three highly correlated observed outcomes (y) but with lots of missing data in each

Individuals with just one non-missing y measure can still be included

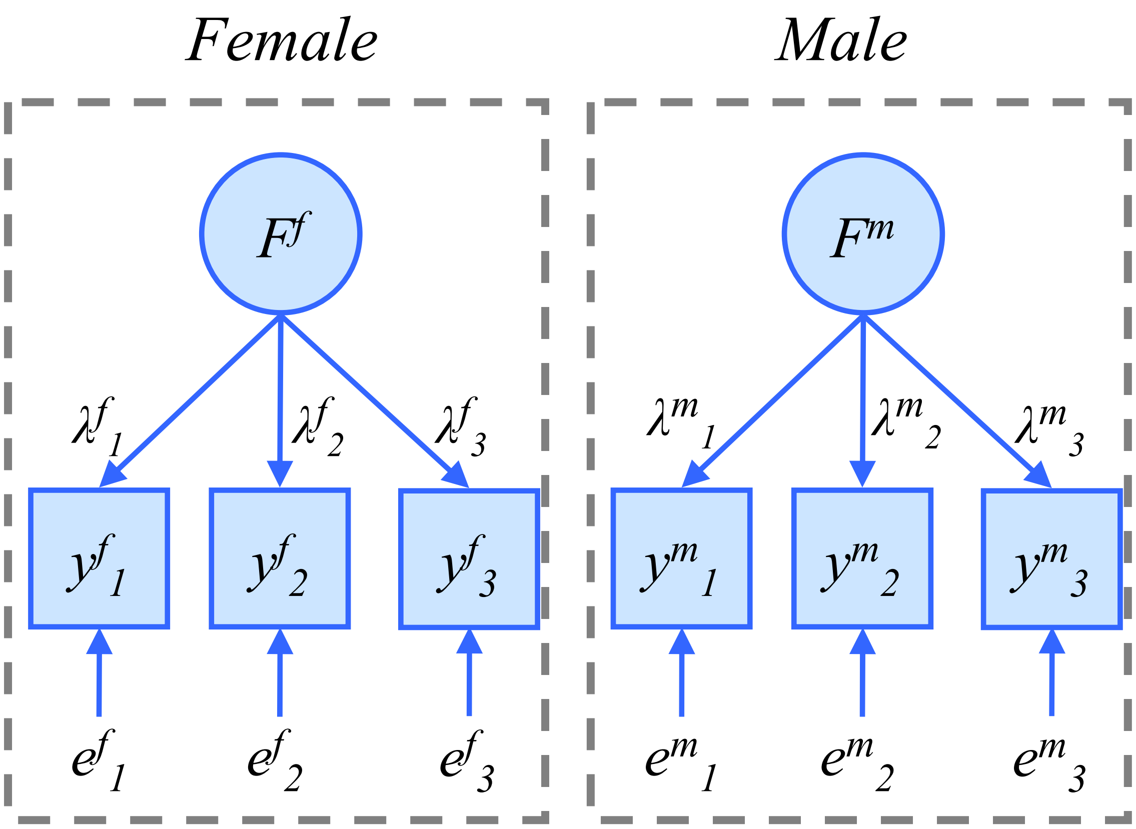
(if we make certain assumptions about the missing data)



MANCHESTER 1824

The Institute for Social Change

Applications: Multiple Group Model



3-item scale

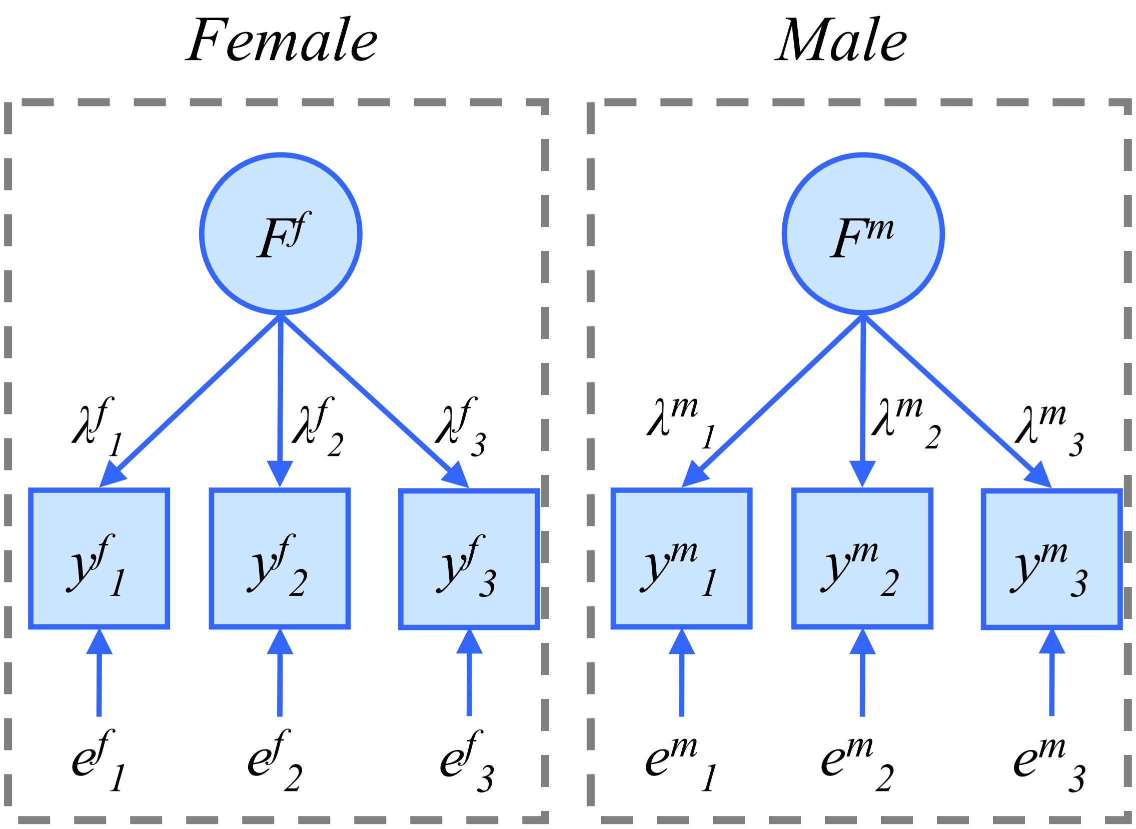
Hypothesis: Item 3 is sex-biased

Can test for Differential Item Functioning (DIF)

MANCHESTER 1824

The Institute for Social Change

Applications: Multiple Group Model



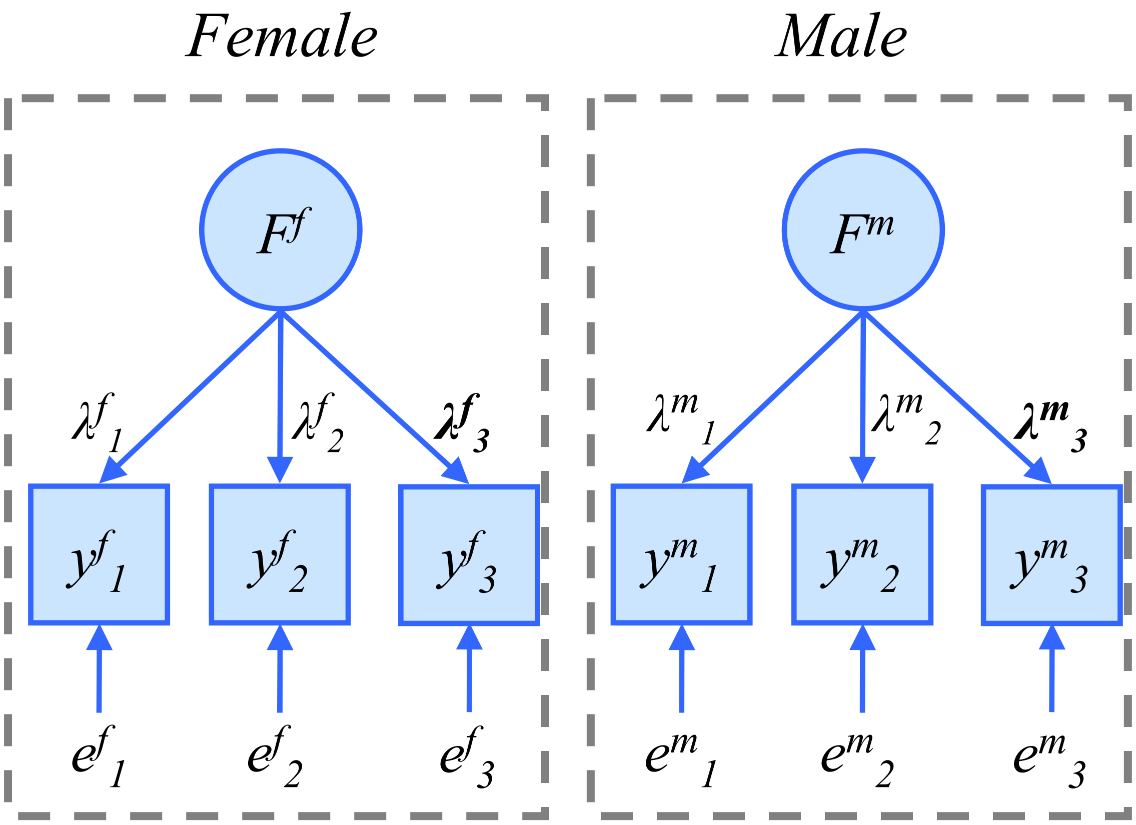
Model 1:
All item loadings constrained to be equal across groups

$\lambda^f_1 = \lambda^m_1$
$\lambda^f_2 = \lambda^m_2$
$\lambda^f_3 = \lambda^m_3$

MANCHESTER 1824

The Institute for Social Change

Applications: Multiple Group Model



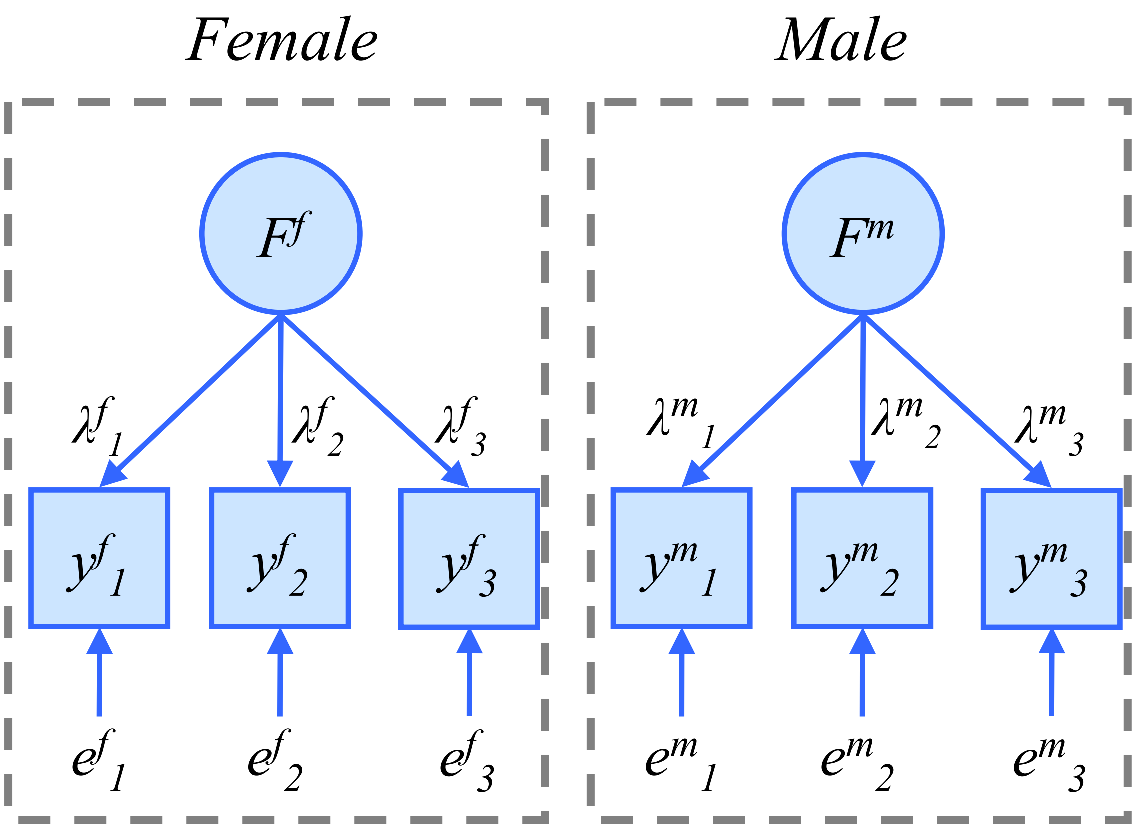
Model 2:
Item 3 loading can be different across groups

$\lambda^f_1 = \lambda^m_1$
$\lambda^f_2 = \lambda^m_2$
$\lambda^f_3 \neq \lambda^m_3$

MANCHESTER 1824

The Institute for Social Change

Applications: Multiple Group Model

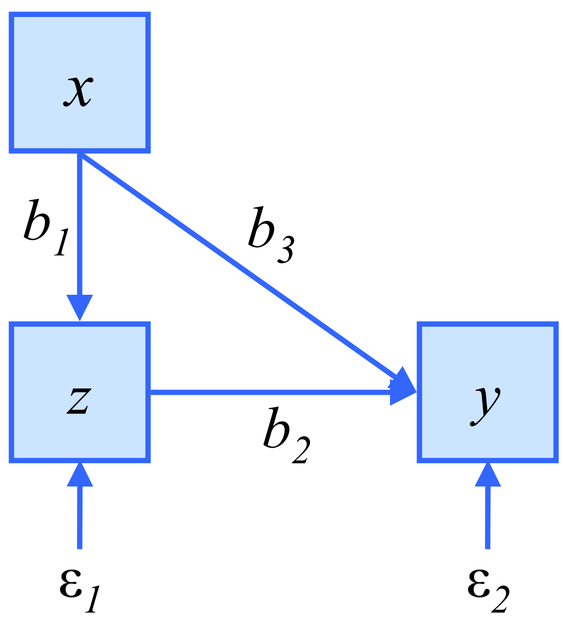


Does Model 2 fit the data significantly better than Model 1?

If so, Item 3 is sex-biased (its 'meaning' is different between the groups)

Application: Mediation Model

Test of partial/full mediation



Model 1: Fit full model

Model 2: Fit model with path b_3 fixed to zero

Does Model 2 fit significantly worse than Model 1? (Likelihood ratio test.)

Features / Limitations

- Need large-ish sample size, depending upon the strength of relationships
  - (E.g. simple factor model with strong loadings, N>200; complicated multilevel SEM with weak loadings, N>5000)

MANCHESTER 1824

The Institute for Social Change

Features / Limitations

- You need to think carefully about your theory
  - Each arrow is a hypothesis
  - The absence of an arrow is a hypothesis
  - It's easy to specify models, harder to interpret them
  - Where to stop? Tempting to always try and add a bit more

Features / Limitations

- Good model fit statistics DO NOT IMPLY THAT YOUR CAUSAL SEM MODEL IS TRUE!!!! Just that it may not be false.
  - Can you defend the assumptions underlying your model (E.g. Cross-sectional survey data? Why should one item predict another and not vice-versa?)
  - Have you considered equivalent models? (Different but statistically indistinguishable ones.)
  - Have you compared your model with alternative but similar model variations?

SEM software

- AMOS
  - Nice graphical interface (GUI); works with SPSS; available on campus network
- Mplus
  - Unparalleled range of models; expensive (but student version available); syntax, not GUI
- R
  - At least two libraries: OpenMX, *sem*; R is free!

MANCHESTER 1824

The Institute for Social Change

Further SEM

- ISC / CCSR 1-day course
  - "Introduction to SEM using Mplus"
  - March 22nd 2010. Book here: http://www.ccsr.ac.uk/courses/sem/

- Introductory Text
  - Kline (2004). Principles and Practice of SEM