Structural Equation Modelling

Short course in Applied Psychometrics

Peterhouse College,
Cambridge,
27-29 March 2012

UNIVERSITY OF CAMBRIDGE

This course

The course is funded by the ESRC RDI and hosted by The Psychometrics Centre

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

RDI Researcher Development Initiative

Tutors

Jon Heron, PhD (Bristol) jon.heron@bristol.ac.uk

Anna Brown, PhD (Cambridge) ab936@medschl.cam.ac.uk

Tim Croudace, PhD (Cambridge) tjc39@cam.ac.uk

	Day 1	Day 2	Day 3	
9:00-	Coffee on arrival	Lec-6 – Special issues in CFA	Lec-9 – SEM	9:00-
9:20-	Introductions + Aims of course	Correlated errors	Incorporating latent traits	9:20-
9:40-	Lec-1	Bi-factor modelling	into path models.	9:40-
10:00-	Mplus modelling framework	Method factors		10:00-
10:20-		Multi-group CFA		10:20-
10:40-				10:40-
11:00-	Coffee	Coffee	Coffee	11:00-
11:20-	Lec-2 – Regression models	Lec-7 – Path models 1	Examples 5 – SEM	11:20-
11:40-		The basics / figures /	EAS - SEM	11:40-
12:00-	Examples 1	Identification/ model fit/	Wrapping up, further reading and	12:00-
12:20-	EAS - regression models	equivalent models	questions	12:20-
12:40-		Examples 3: SZ paper.	Lunch and depart	12:40-
13:00-	Lunch	Lunch		13:00-
13:20-				13:20-
13:40-				13:40-
14:00-	Lec-3 - CFA with continuous	Lec-8 – Path models 2		14:00-
14:20-	variables	Model refinement		14:20-
14:40-		Direct and indirect effects		14:40-
15:00-	Lec-4 – EFA with continuous	Binary mediators - logit/probit		15:00-
15:20-	variables			15:20-
15:40-				15:40-
16:00-	Coffee	Coffee		16:00-
16:20-	Lec-5 - CFA and EFA with	Examples 4		16:20-
16:40-	categorical variables	Path model using EAS		16:40-
17:00-	Examples 2			17:00-
17:20-	EAS – CFA/EFA			17:20-
17:40-				17:40-

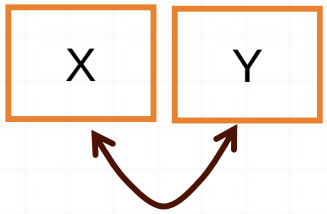
UNIVERSITY OF CAMBRIDGE

Mplus model syntax

A quick refresher

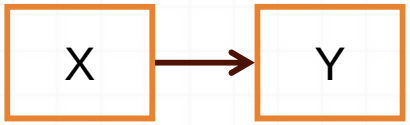
UNIVERSITY OF CAMBRIDGE

Model statements: BY / WITH / ON



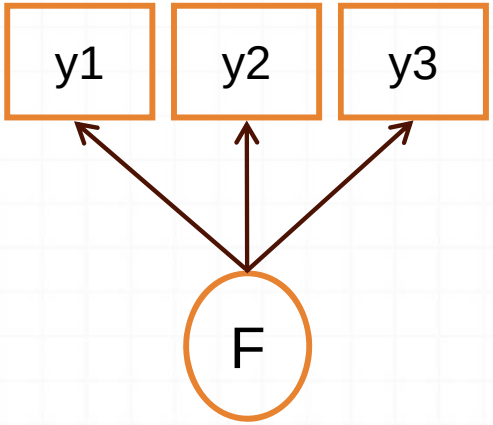
- X is correlated WITH Y

  X with Y;



- Y (outcome) is regressed on X (predictor)

  Y on X;



- F (the factor) is measured BY Y1 Y2 Y3

  F by Y1 Y2 Y3;

UNIVERSITY OF CAMBRIDGE

Variable means

- Stuff in a square bracket is a mean/intercept:

  [wt_7 wt_9 wt_11];

- It's just the same to say:

  [wt_7];
  [wt_9];
  [wt_11];

UNIVERSITY OF CAMBRIDGE

Variances

- No bracket, then it's a variance / residual variance:

  wt_7;

  wt_9;

  wt_11;

- Or

  wt_7 wt_9 wt_11;

UNIVERSITY OF CAMBRIDGE

Parameter equality constraints

- Three residual variances constrained to be equal:

  wt_7 (1);
  wt_9 (1);
  wt_11 (1);

- Three intercept constrained to be equal:

  [wt_7] (2);
  [wt_9] (2);
  [wt_11] (2);

UNIVERSITY OF CAMBRIDGE

Parameter equality constraints

- Three residual variances constrained to be equal:

  wt_7 (fixvar);
  wt_9 (fixvar);
  wt_11 (fixvar);

- Three intercept constrained to be equal:

  [wt_7] (fixmean);
  [wt_9] (fixmean);
  [wt_11] (fixmean);

UNIVERSITY OF CAMBRIDGE

Fixing parameters

- Constraining a covariance to be zero:

  X with Y@0;

- Constraining a mean to be zero:

  [wt_7@0];

- Constraining a variance to be zero:

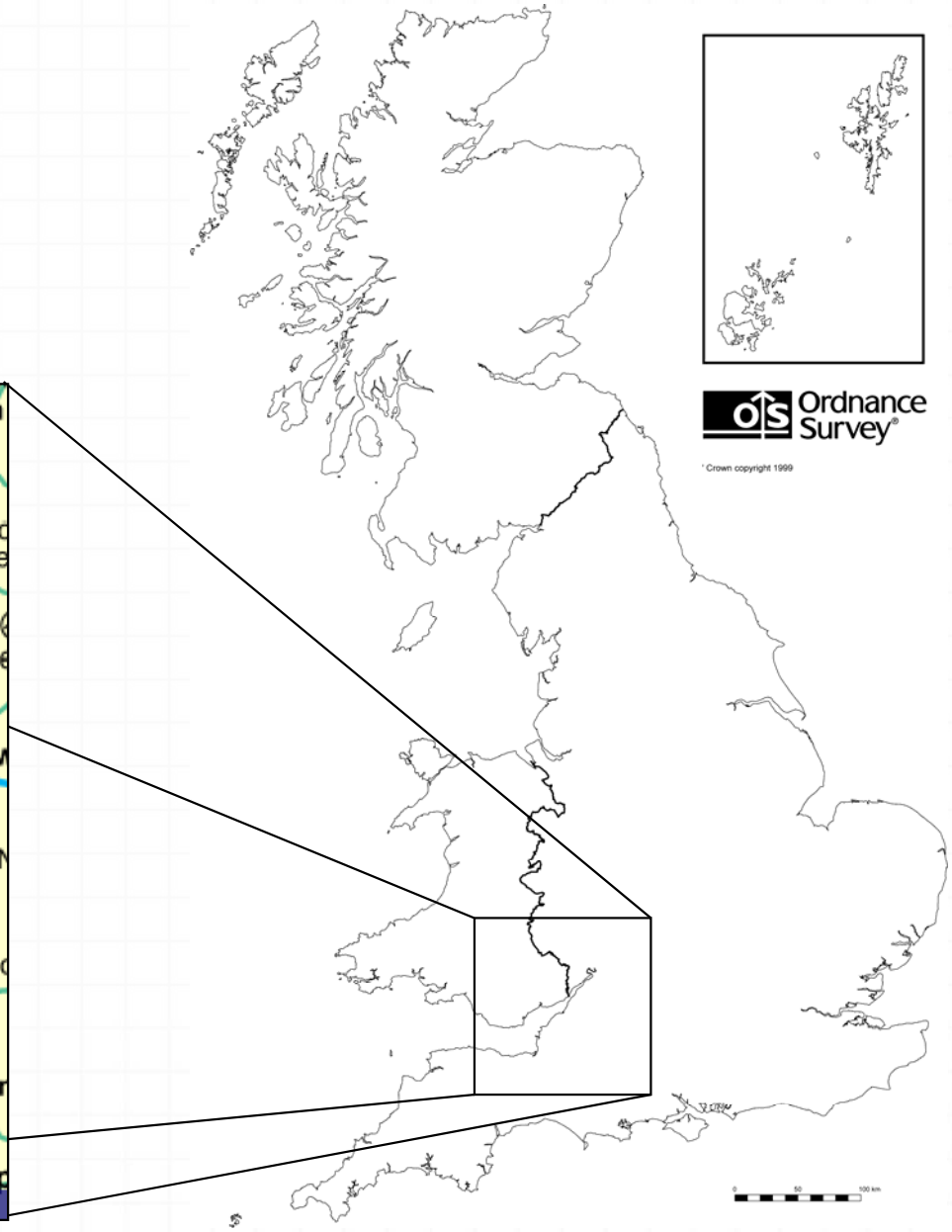
  i@0;

UNIVERSITY OF CAMBRIDGE

Dataset for exercises

EAS-ing you into SEM

Where ALSPAC is based



What is ALSPAC?

- Avon Longitudinal Study of Parents and Children AKA "Children of the Nineties"
- Cohort study of ~14,000 children and their parents, based in South-West England
- Eligibility criteria: Mothers had to be resident in Avon and ha an expected date of delivery between April 1st 1991 and December 31st 1992

- Population-Based Prospective Birth-Cohort

UNIVERSITY OF CAMBRIDGE

What data does ALSPAC have?

- Self completion questionnaires
  - Mothers, Partners, Children, Teachers
- Hands on assessments
  - 10% sample tested regularly since birth
  - Yearly clinics for all since age 7
- Data from external sources
  - SATS from LEA, Child Health database
- Biological samples
  - DNA / cell lines

SECTION F: TEMPERAMENT AND BEHAVIOUR

TY OF
IDGE

How often is your child's behaviour like that given below:

		Never	Rarely	Some-times	Often	Always
F1.	He tends to be shy	1	2	3	4	5
F2.	He cries easily	1	2	3	4	5
F3.	He likes to be with people	1	2	3	4	5
F4.	He is always on the go	1	2	3	4	5
F5.	He prefers playing with others rather than alone	1	2	3	4	5
F6.	He is somewhat emotional	1	2	3	4	5
F7.	When he moves about he moves slowly	1	2	3	4	5

UNIVERSITY OF CAMBRIDGE

Full set of ordinal EAS items

act_t1_1	Always on the go (+)	**shy_t1_1**	Shy (-)
act_t1_2	Moves about slowly (-)	**shy_t1_2**	Makes friends (+)
act_t1_3	Active on waking (+)	**shy_t1_3**	Sociable (+)
act_t1_4	Very energetic (+)	**shy_t1_4**	Takes time warming to strangers (-)
act_t1_5	Prefers quiet games (-)	**shy_t1_5**	Friendly with strangers (+)
emo_t1_1	Cries easily (-)	**soc_t1_1**	Likes being with people (+)
emo_t1_2	Emotional (-)	**soc_t1_2**	Prefers playing with others (+)
emo_t1_3	Often fusses and cries (-)	**soc_t1_3**	Finds people stimulating (+)
emo_t1_4	Gets upset easily (-)	**soc_t1_4**	Something of a loner (-)
emo_t1_5	Reacts intensely when upset (-)	**soc_t1_5**	Isolated when alone (+)

UNIVERSITY OF CAMBRIDGE

Possible predictors

Sex	Gender (1=Male, 2=Female)
mumage	Maternal age at delivery (1=<25, 2=25-29, 3=30-34, 4=35+)
tenure	Housing tenure (0 = mortgaged, 1 = private rented, 2 = subsidized rented)
crowding	Home overcrowding (> 1 person per room; 0=no, 1=yes)
parity	Parity (0=1st born, 1=2nd born, 2 = 3rd born+)
mumed	Maternal educational attainment (0 = A-level+, 1 = O-level, 2 = <O-level)
income	Household income (0 = bottom 20%, 1 = middle 60%, 2 = top 20%)
social	Social class (0 = I/II, 1 = III non-manual or lower)
mumalc	Regular maternal alcohol use in the early postnatal period (0=no, 1=yes)
mumsmk	Maternal cigarette use in the early postnatal period (0=none, 1=low, 2=high)
mdep_pn	Mother exceeding threshold for EPDS in early postnatal period (0=no, 1=yes)

UNIVERSITY OF CAMBRIDGE

Basic Input file (courtesy of "Stata2mplus")

```
Data:
  File is H:\Courses\SEM_2012\data\eas_1500.dta.dat ;

Variable:
  Names are id
     sex
     act_t1_1 act_t1_2 act_t1_3 act_t1_4 act_t1_5
     emo_t1_1 emo_t1_2 emo_t1_3 emo_t1_4 emo_t1_5
     <snip>
     shy_t3_1 shy_t3_2 shy_t3_3 shy_t3_4 shy_t3_5
     soc_t3_1 soc_t3_2 soc_t3_3 soc_t3_4 soc_t3_5
     mumage tenure crowding parity mumed income social mumalc mumsmk mdep_pn
     mfq10_01 mfq10_02 mfq10_03 mfq10_04 mfq10_05 mfq10_06
     mfq10_07 mfq10_08 mfq10_09 mfq10_10 mfq10_11 mfq10_12 mfq10_13
     mfq18_01 mfq18_02 mfq18_03 mfq18_04 mfq18_05 mfq18_06
     mfq18_07 mfq18_08 mfq18_09 mfq18_10 mfq18_11 mfq18_12 mfq18_13
     emotott1 emotott2 emotott3 acttott1 acttott2 acttott3
     shytott1 shytott2 shytott3 soctott1 soctott2 soctott3;
  Missing are all (9999) ;

Analysis:
  Type = basic ;
```

UNIVERSITY OF CAMBRIDGE

Logit and Probit models

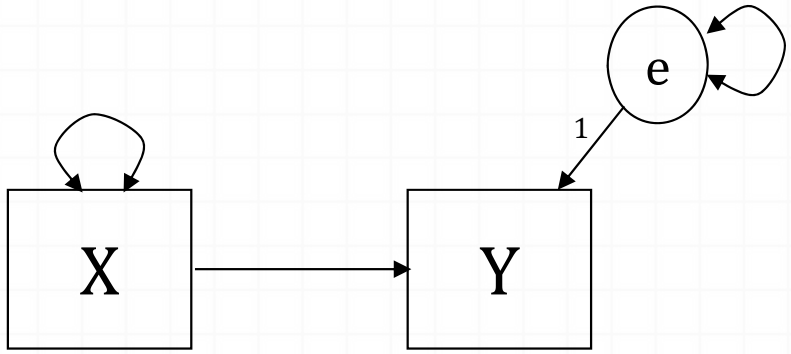
UNIVERSITY OF CAMBRIDGE

Logit and Probit models

- Both are latent variable models of sorts
- Observed binary variable Y
- Assumed underlying continuous variable Y*
- Variance of Y* is unknown

UNIVERSITY OF CAMBRIDGE

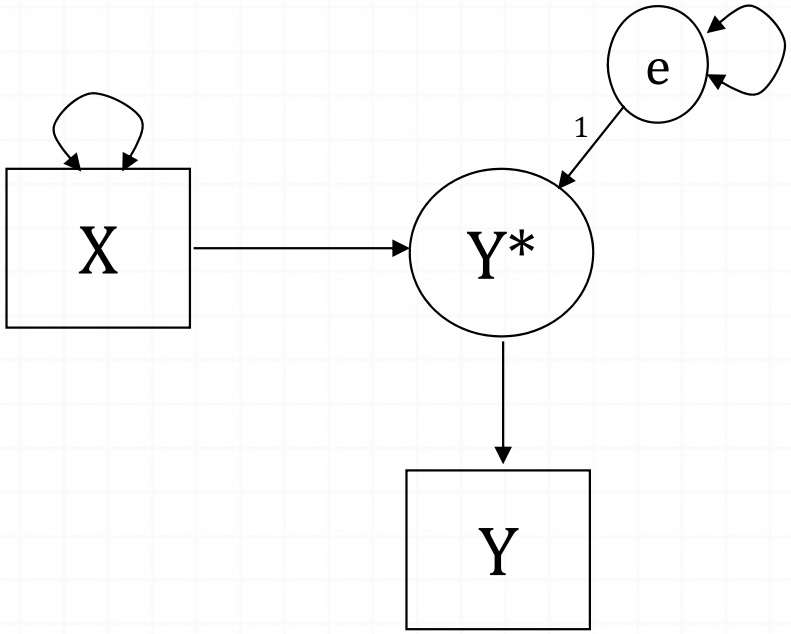
OLS Regression model



Variance of Y is known
Residuals are assumed to be standard normal
Aim is to use covariates to explain the variance in Y
Residual variance will typically reduce

UNIVERSITY OF CAMBRIDGE

Logit and Probit models



Variance of Y is unknown
Residuals assumed to be standard normal / logistic
Residual variance is usually FIXED at one or $\pi^2/3$
Otherwise the scaling is arbitrary

UNIVERSITY OF CAMBRIDGE

Normal and Logistic distributions

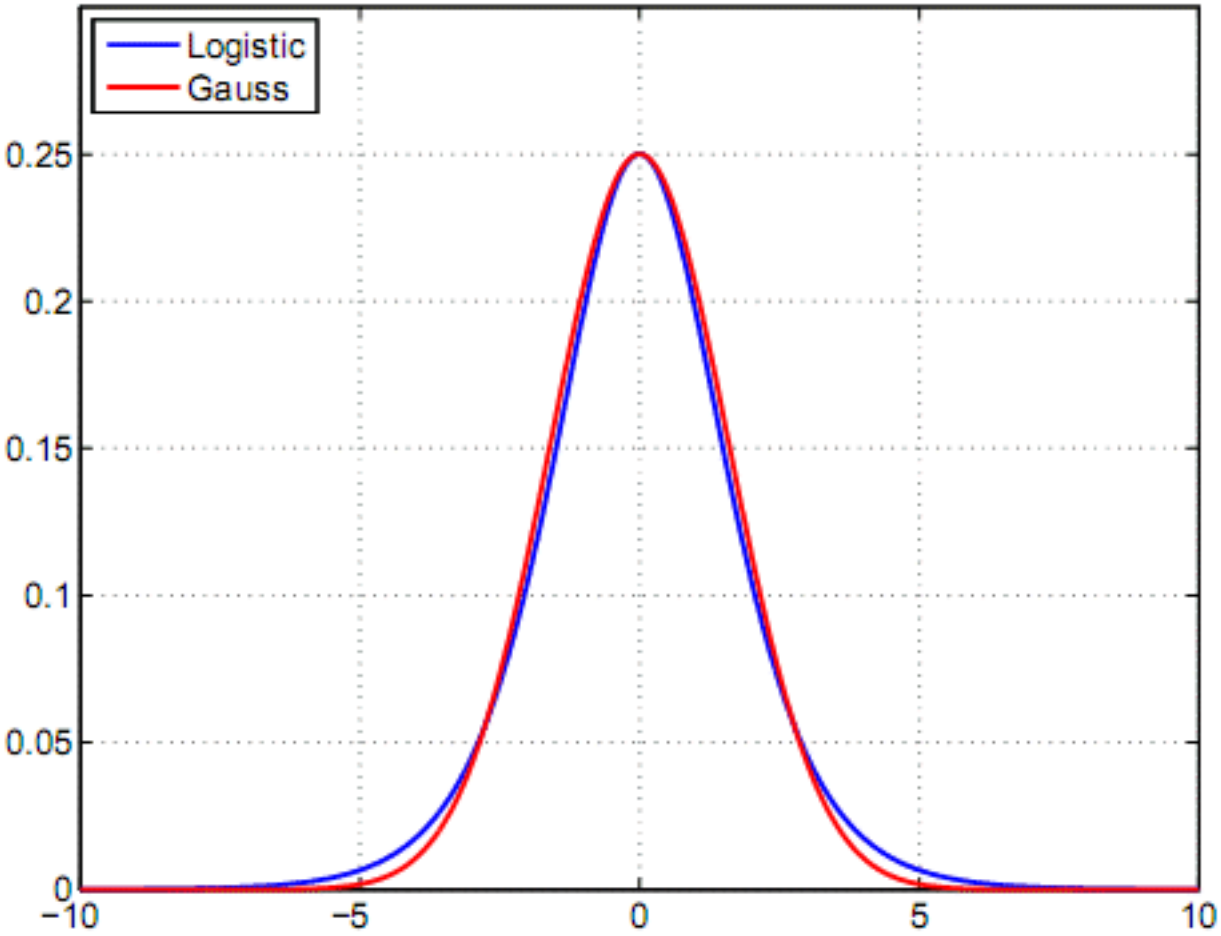


Fig. 25. Logistic versus normal distribution with matched peaks, for $r = 1$, $t^* = 0$, $\dot{m}^* = 1/4$.

T.W. Patzek, G.D. Croft / Energy 35 (2010) 3109–3122

UNIVERSITY OF CAMBRIDGE

Quick example

```
Data:
  File is "C:\Work\SEM Course\eas_1500.dta.dat" ;

Define:
   mfqsum18 = mfq18_01 + mfq18_02 + mfq18_03 + mfq18_04 + mfq18_05
            + mfq18_06 + mfq18_07 + mfq18_08 + mfq18_09 + mfq18_10
            + mfq18_11 + mfq18_12 + mfq18_13;
   mfqcase = (mfqsum18>11);
   emomean = (emotott1+emotott2+emotott3)/3;

Variable:
  Names are id etc....
  Usevariables = mfqcase emomean;
  Categorical = mfqcase;
```

UNIVERSITY OF CAMBRIDGE

ML Logit model

```
Analysis:
   estimator = ML;

Model:
   mfqcase on emomean;

Output:
    cint;
```

UNIVERSITY OF CAMBRIDGE

Logit model results

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
MFQCASE ON EMOMEAN	0.954	0.222	4.300	0.000
Thresholds MFQCASE$1	2.403	0.195	12.293	0.000

LOGISTIC REGRESSION ODDS RATIO RESULTS

MFQCASE ON EMOMEAN	2.595

CONFIDENCE INTERVALS OF MODEL RESULTS

	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%
MFQCASE ON EMOMEAN	0.519	0.589	0.954	1.319	1.389

UNIVERSITY OF CAMBRIDGE

WLSMV Probit model

```
Analysis:
   estimator = WLSMV;

Model:
   mfqcase on emomean;
Output:
    cint;
```

UNIVERSITY OF CAMBRIDGE

ML Probit model

```
Analysis:
   estimator = ML;
   link = probit;
Model:
   mfqcase on emomean;
Output:
    cint;
```

UNIVERSITY OF CAMBRIDGE

Comparison of results

- ML Probit

	Estimate	S.E.	Est./S.E.	P-Value
MFQCASE ON EMOMEAN	0.532	0.124	4.279	0.000
Thresholds - MFQCASE$1	1.406	0.107	13.198	0.000

- WLSMV Probit

	Estimate	S.E.	Est./S.E.	P-Value
MFQCASE ON EMOMEAN	0.531	0.124	4.272	0.000
Thresholds - MFQCASE$1	1.406	0.107	13.185	0.000

- ML Logit

	Estimate	S.E.	Est./S.E.	P-Value
MFQCASE ON EMOMEAN	0.954	0.222	4.300	0.000
Thresholds MFQCASE$1	2.403	0.195	12.293	0.000

UNIVERSITY OF CAMBRIDGE

Logit and probit models

- Probit – ML and WLSMV give almost identical results
- Logit and Probit params have different interpretation however models are very similar
- Not statistical criteria for choosing between probit/logit
- Down to preference and research discipline
- Fixed scales (important when we come on to mediation)

UNIVERSITY OF CAMBRIDGE

Time for practical #1

Regression models in Mplus