Virginia Commonwealth University

VCU

The Measurement and Analysis of Complex Traits

Everything you didn't want to know about measuring behavioral and psychological constructs

Leuven Workshop
August 2008

Overview

- SEM factor model basics
- Group differences: - practical
- Relative merits of factor scores & sum scores
- Test for normal distribution of factor
- Alternatives to the factor model
- Extensions for multivariate linkage & association

VCU

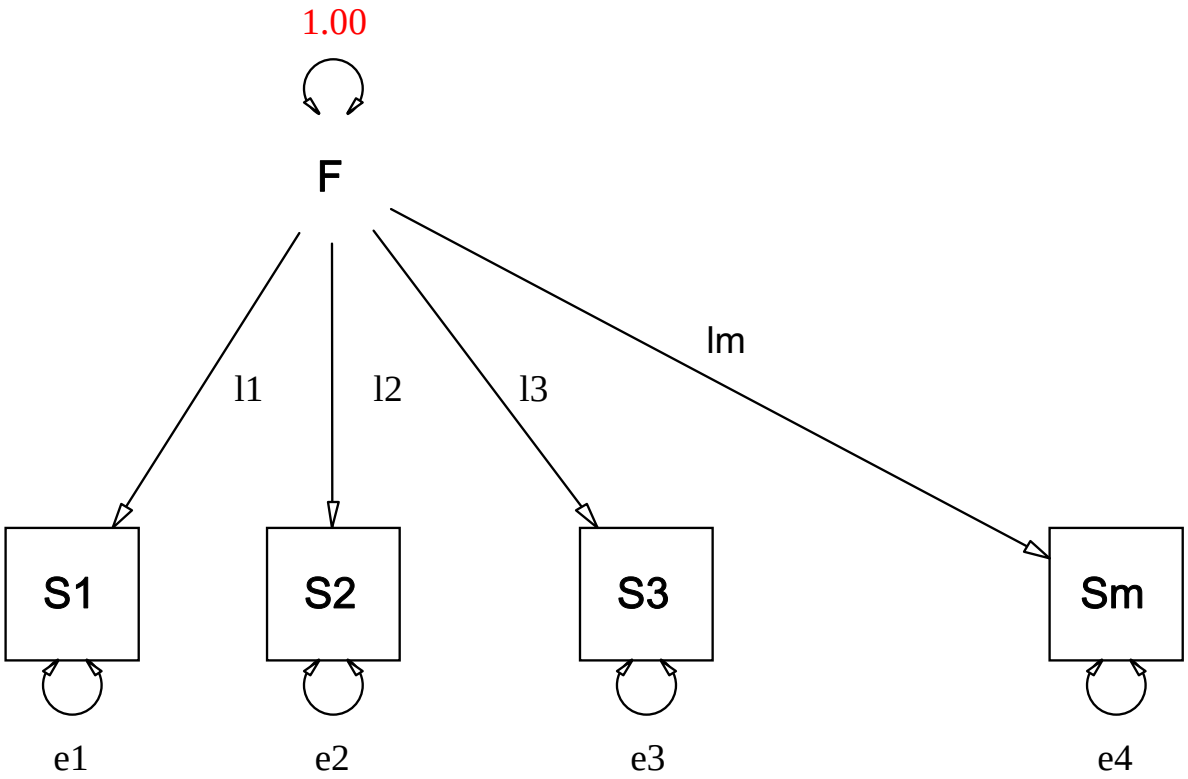
Structural Equation Model basics

- Two kinds of relationships
  - Linear regression X -> Y single-headed
  - Unspecified Covariance X<->Y double-headed

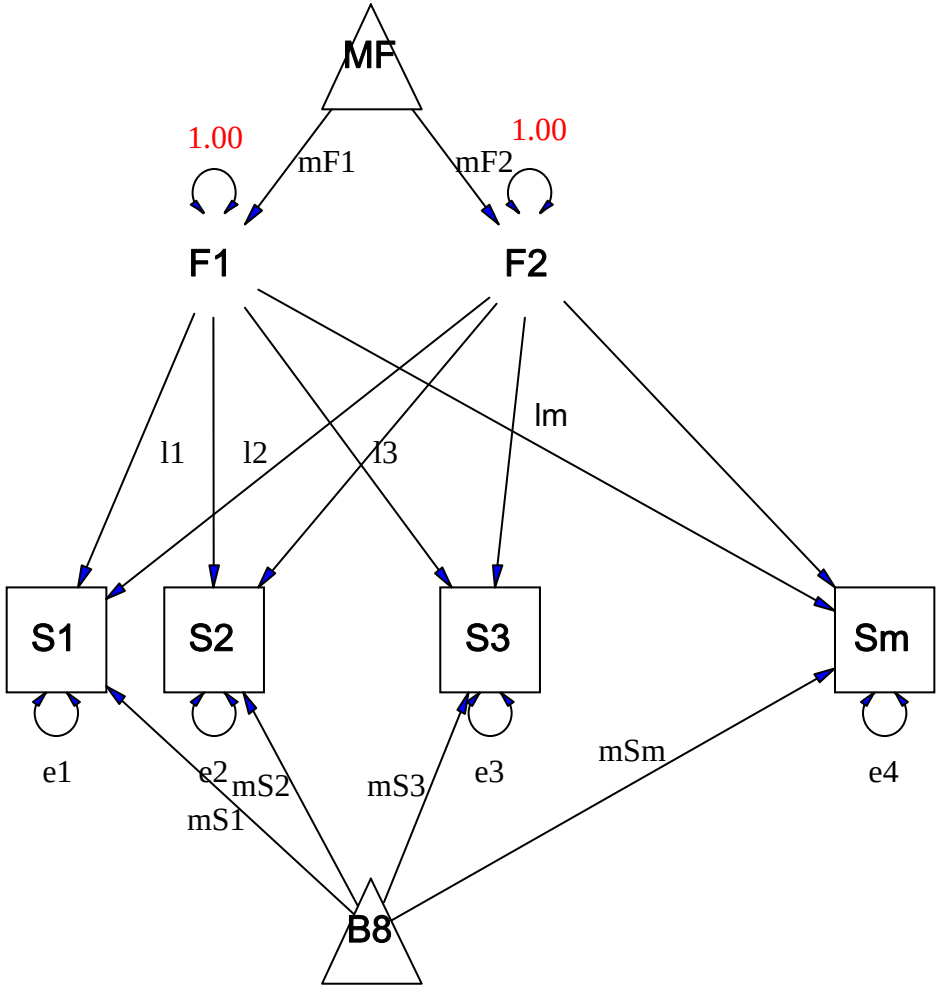
- Four kinds of variable
  - Squares – observed variables
  - Circles – latent, not observed variables

  - Triangles – constant (zero variance) for specifying means
  - Diamonds -- observed variables used as moderators (on paths)

Single Factor Model



Factor Model with Means



Factor model essentials

- Diagram translates directly to algebraic formulae
- Factor typically assumed to be normally distributed: SEM
- Error variance is typically assumed to be normal as well
- May be applied to binary or ordinal data
  - Threshold model

VCU

What is the best way to measure factors?

- Use a sum score
- Use a factor score
- Use neither - model-fit

Factor Score Estimation

- Formulae for continuous case
  - Thompson 1951 (Regression method)
  - $C = LL' + V$
  - $f = (I+J)^{-1}L'V^{-1}x$
  - Where $J = L'V^{-1}L$

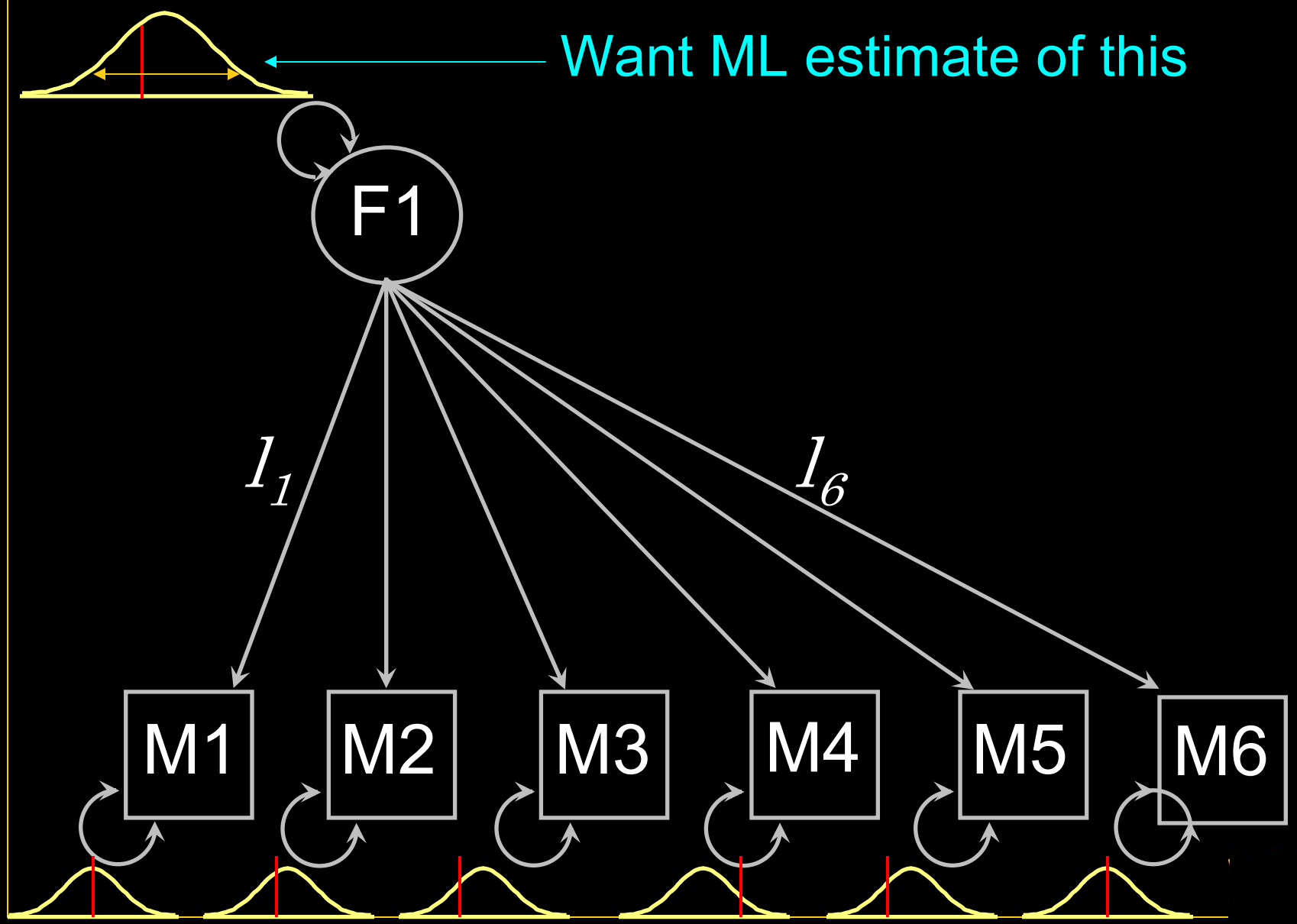
Factor Score Estimation

- Formulae for continuous case
  - Bartlett 1938
  - $C = LL' + V$
  - $f_b = J^{-1}L'V^{-1}x$
  - where $J = L'V^{-1}L$

- Neither is suitable for ordinal data

VCU

Estimate factor score by ML



ML Factor Score Estimation

- Marginal approach

- L(f&x) = L(f)L(x|f) (1)
- L(f) = pdf(f)
- L(x|f) = pdf(x*)
- x* ~ N(V,Lf)

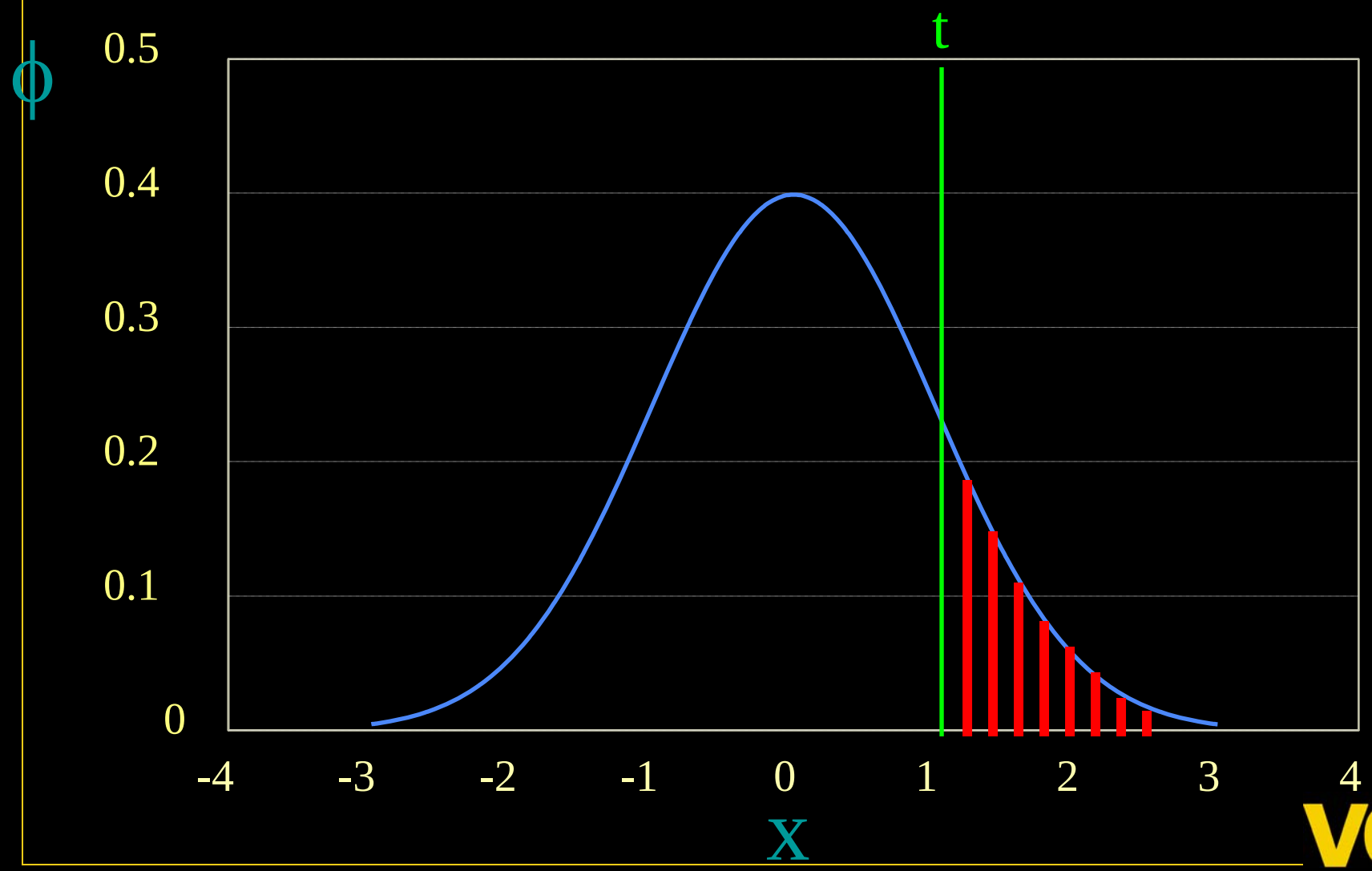
- Maximize (1) with respect to f
- Repeat for all subjects in sample

  – Works for ordinal data too!

VCU

Multifactorial Threshold Model

Normal distribution of liability x. ‘Yes’ when liability x > t



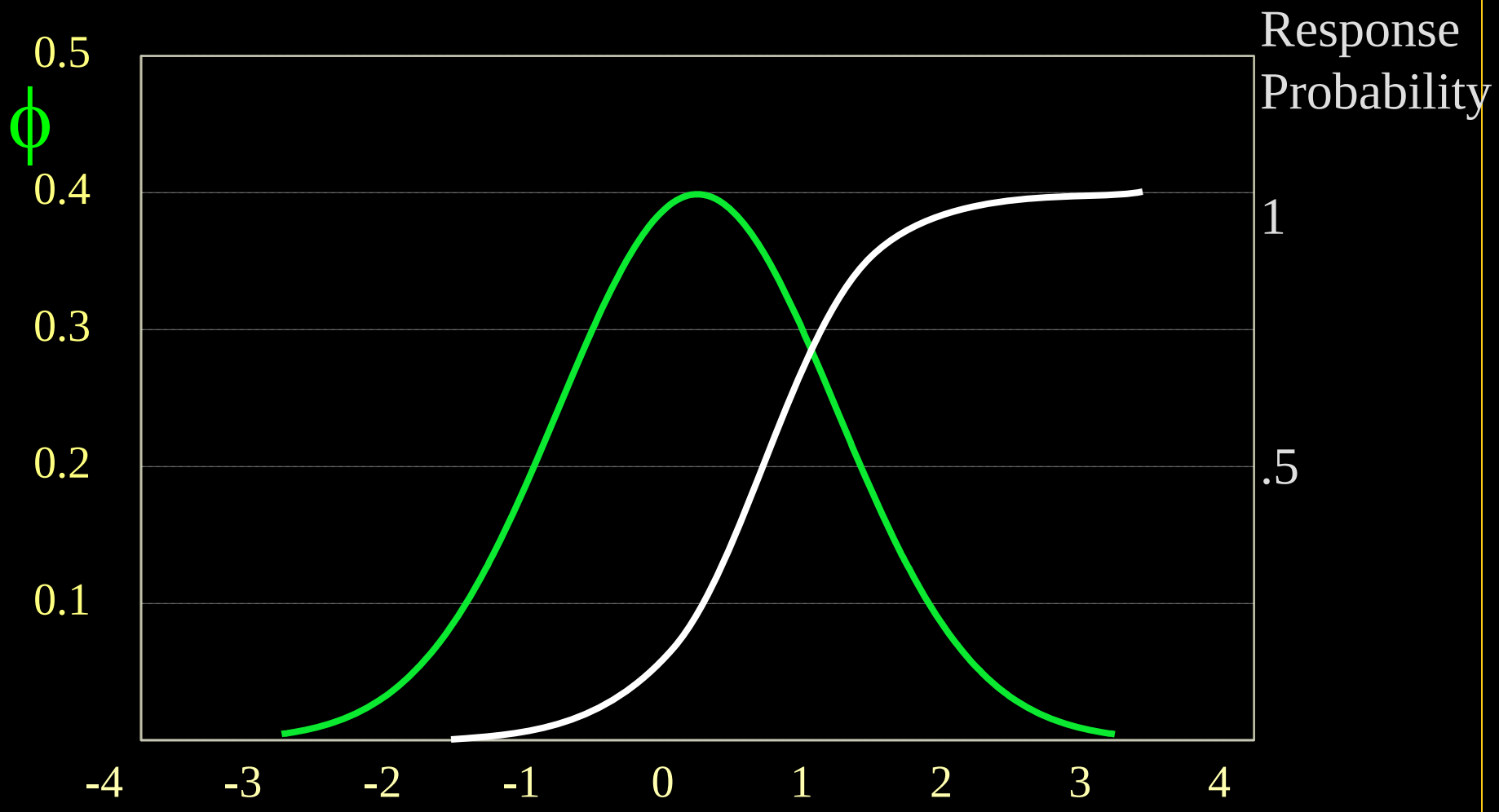
Item Response Theory - Factor model equivalence

- Normal Ogive IRT Model
- Normal Theory Threshold Factor Model
- Takane & DeLeeuw (1987 Psychometrika)
  - Same fit
  - Can transform parameters from one to the other

VCU

Item Response Probability

Example item response probability shown in white

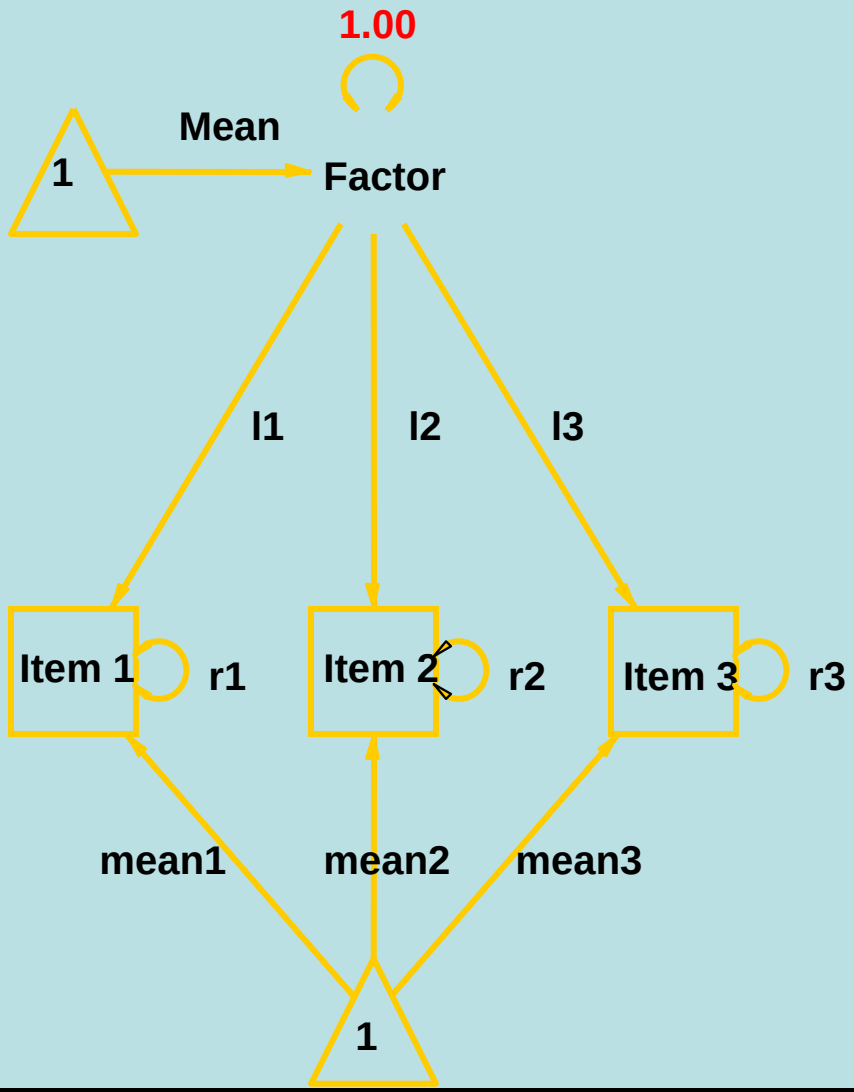


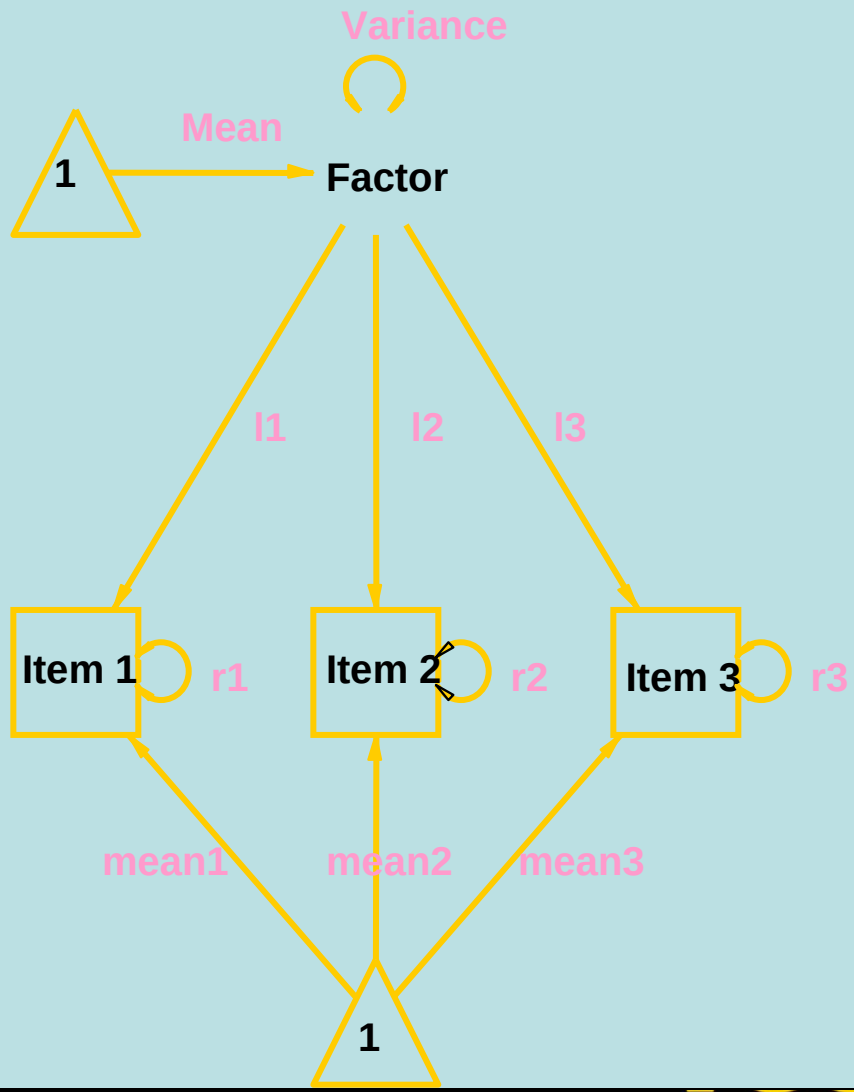
Do groups differ on a measure?

- Observed
  - Function of observed categorical variable (sex)
  - Function of observed continuous variable (age)
- Latent
  - Function of unobserved variable
  - Usually categorical
  - Estimate of class membership probability
    - Has statistical issues with LRT

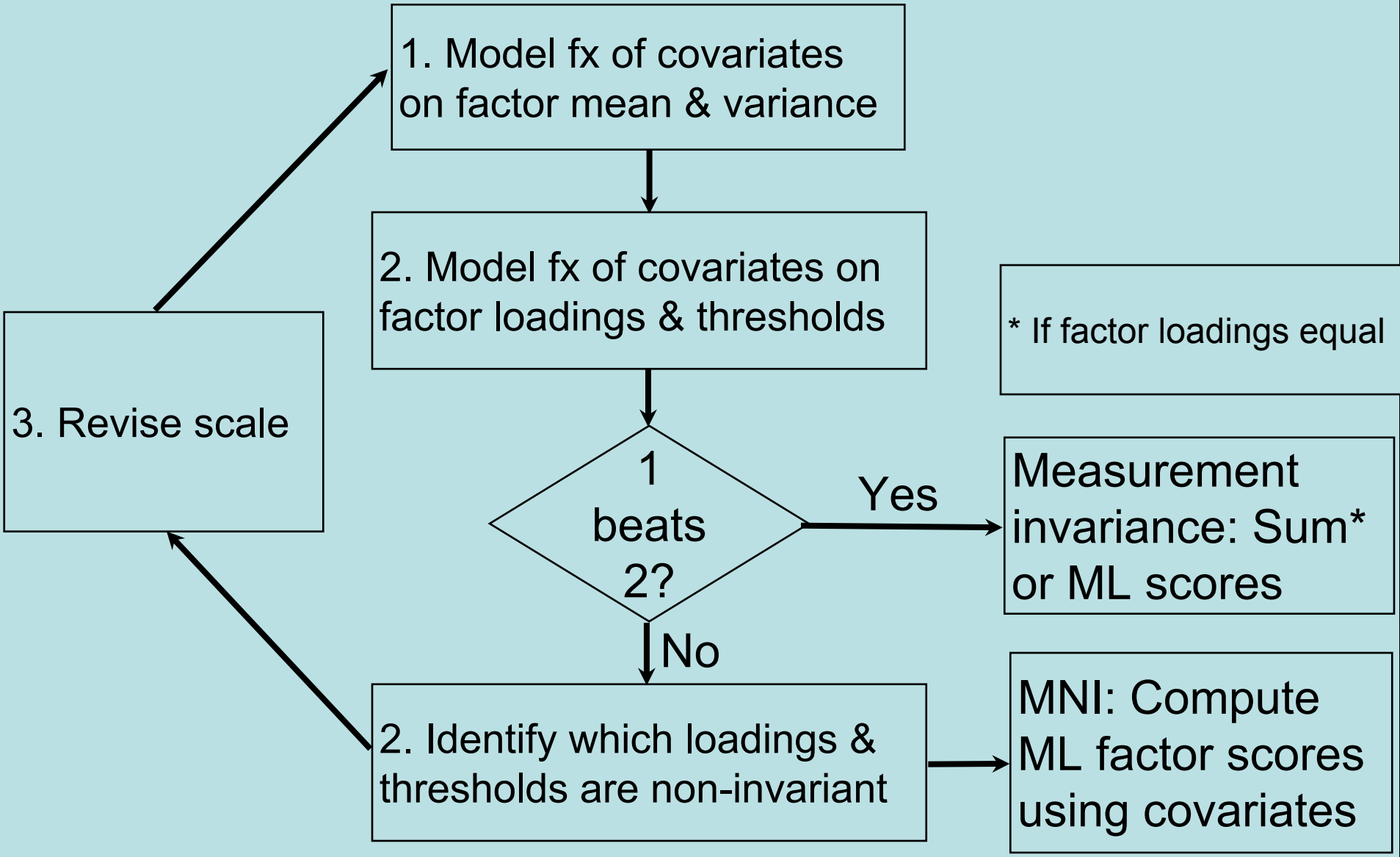
VCU

Practical: Find the Difference(s)

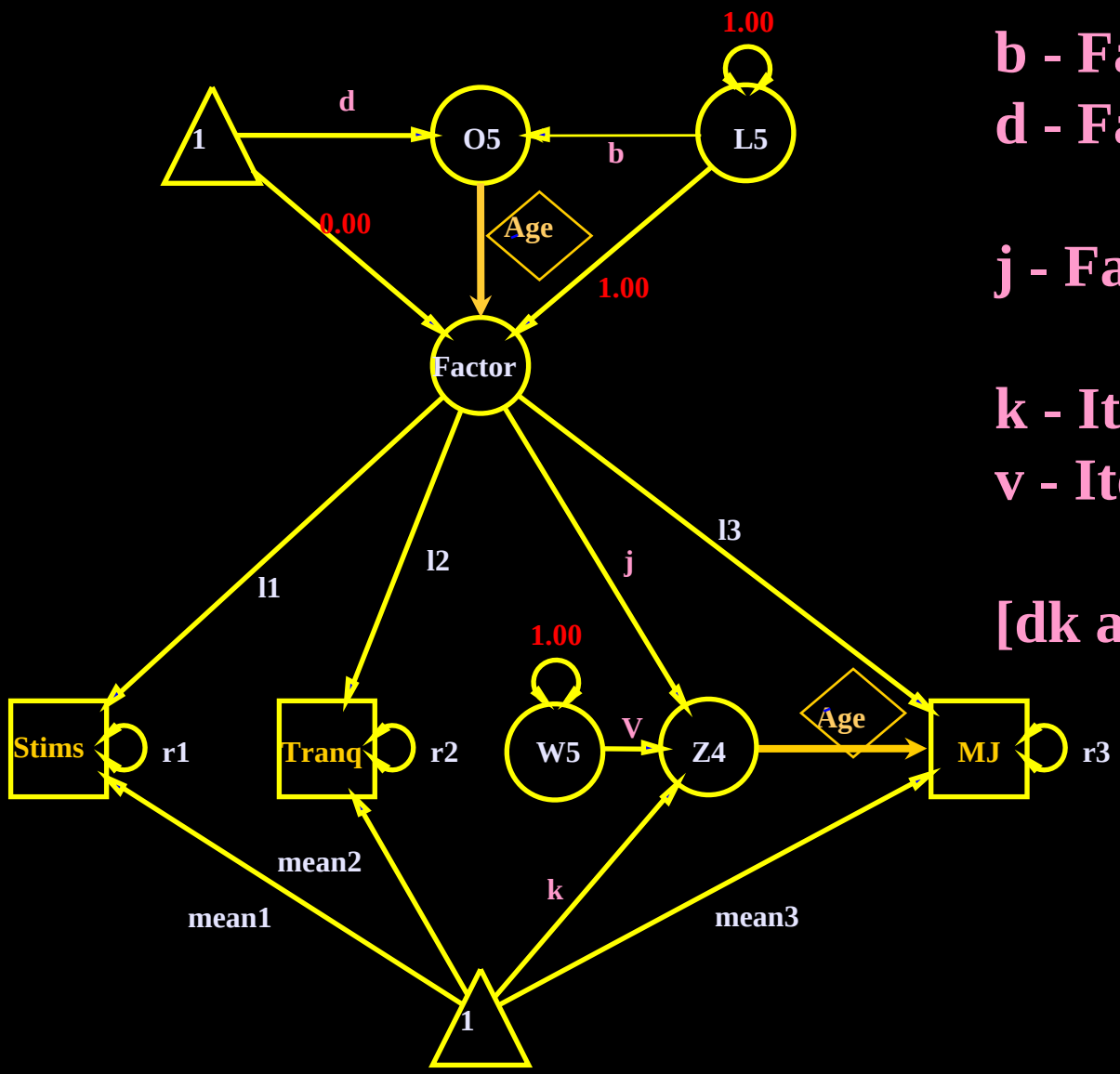




Sequence of MNI testing



Continuous Age as a *Moderator* in the Factor Model



b - Factor Variance
d - Factor Mean

j - Factor Loadings

k - Item means
v - Item variances

[dk and bj confound]

What is the best way to measure and model variation in my trait?

- Behavioral / Psychological characteristics usually Likert
  - Might use ipsative?

- What if Measurement Invariance does not hold?
  - How do we judge:
    - Development
    - GxE interaction
    - Sex limitation

- Start simple: Finding group differences in mean

VCU

Simulation Study (MK)

- Generate True factor score f ~ N(0,1)
- Generate Item Errors e_j ~ N(0,1)
- Obtain vector of j item scores $s_j = L*f_j + e_j$
- Repeat N times to obtain sample
- Compute sum score
- Estimate factor score by ML

VCU

Two measures of performance

- Reliability

Reliability refers to how consistent individual scores or summary statistics of those scores remain across repeated tests under identical conditions. Because the mean of factor scores is often the statistic of primary interest, we restrict ourselves here to discussion of the reliability of the mean, which can be quantified as one minus the error of the mean, $1 - E([E(\hat{\xi}_i^{ML}) - E(\xi)]^2)$, or equivalently to $1 - E(E[\hat{\xi}_i^{ML}]^2)$ when $\xi \sim N(0,1)$.

Two measures of performance

- Validity

The validity of factor scores refers to the degree to which they measure what they are supposed to measure. In simulation studies, this can be quantified as the correlation between the factor score estimates and true factor scores: $cor(\hat{\xi}_i^{ML}, \xi_i)$ {Penev & Raykov, 2006}.

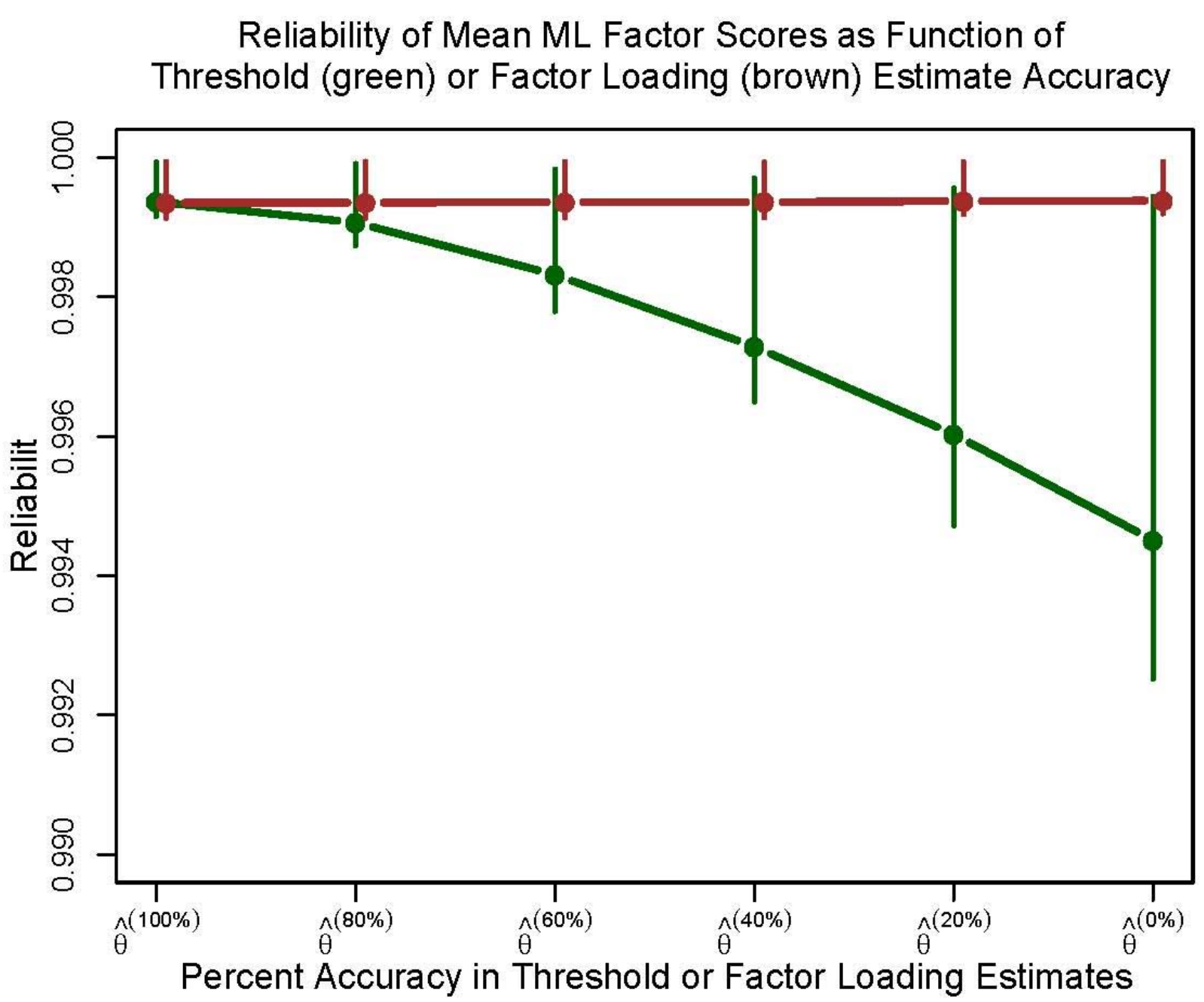
Simulation parameters

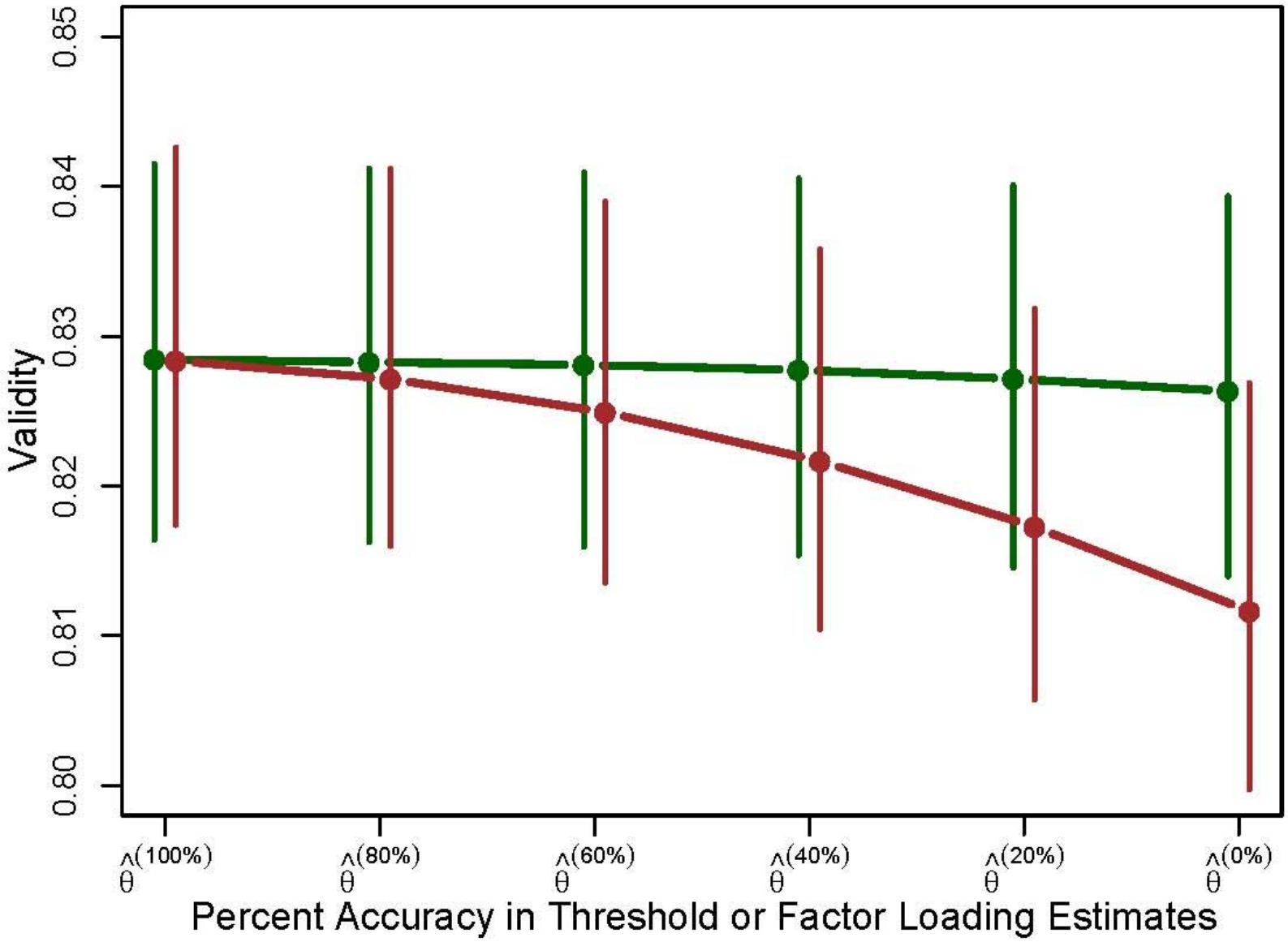
- 10 binary item scale
- Thresholds
  - [-1.8 -1.35 -0.9 -0.45 0.0 0.45 0.9 1.35 1.8]
- Factor Loadings
  - [.30 .80 .43 .74 .55 .68 .36 .61 .49]

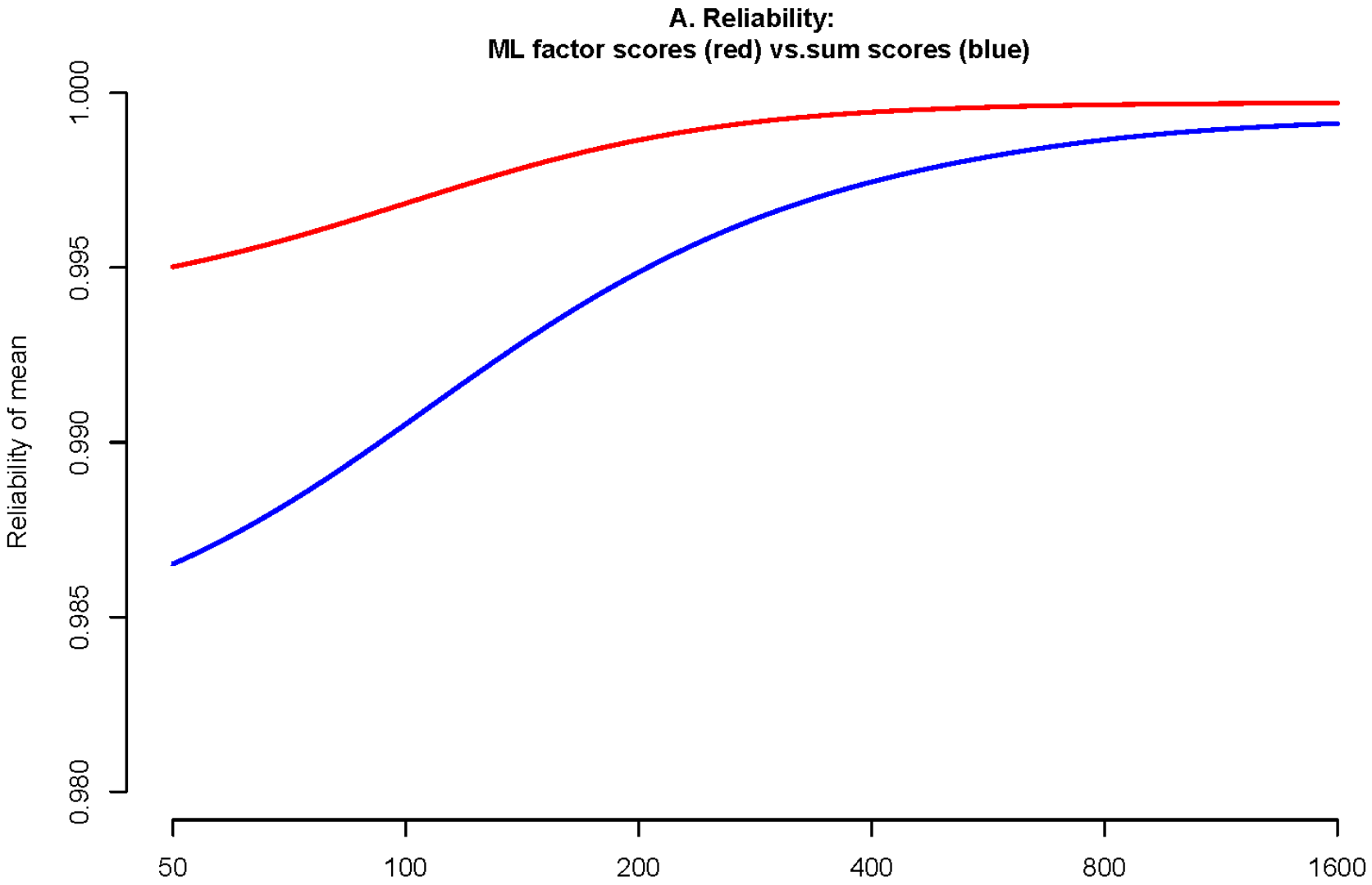
VCU

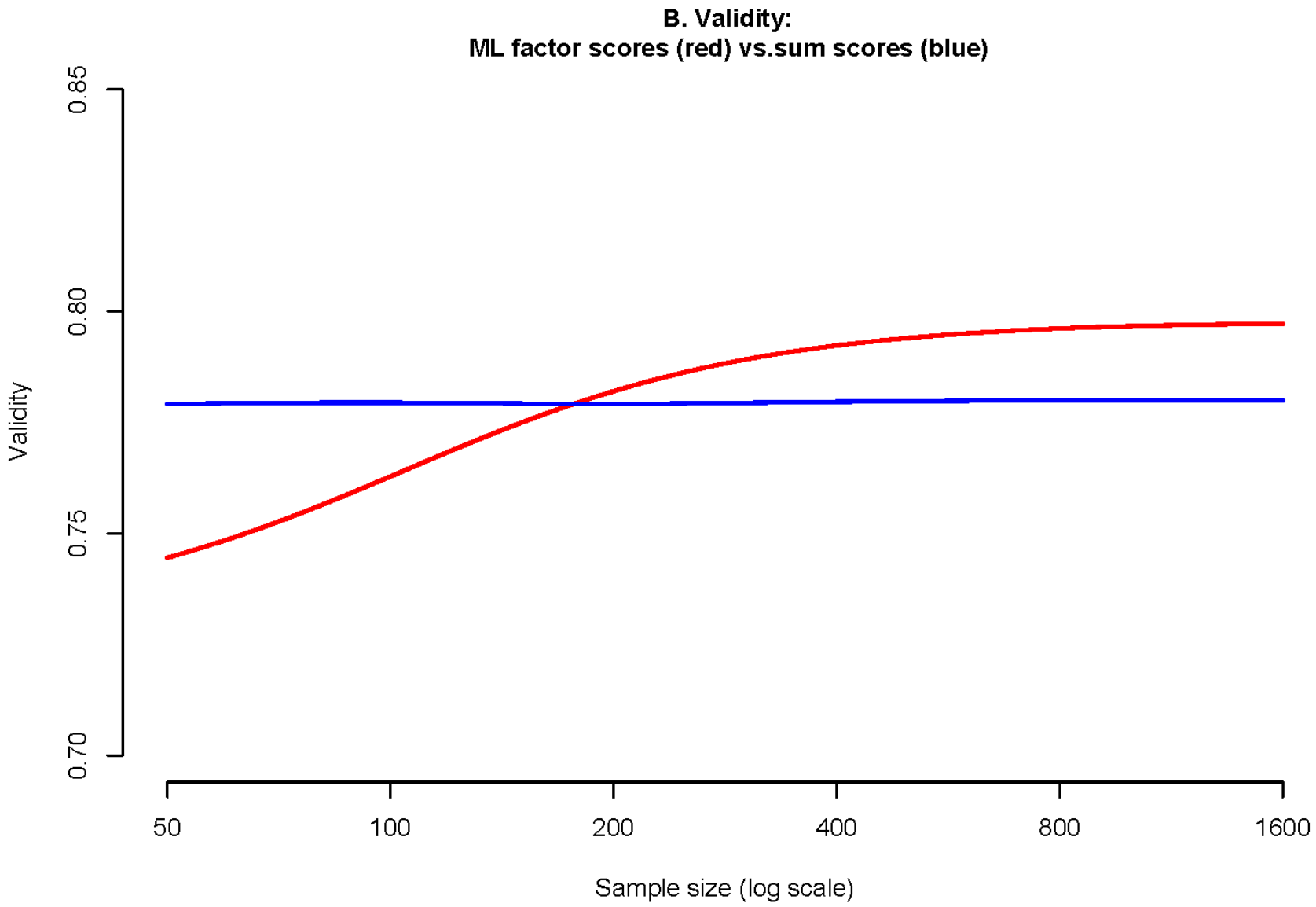
Mess up measurement parameters

- Randomly reorder thresholds
- Randomly reorder factor loadings
- Blend reordered estimates with originals 0% - 100% ‘doses’


Reliability of Mean ML Factor Scores as Function of
Threshold (green) or Factor Loading (brown) Estimate Accuracy
Reliabilit
1.000
0.998
0.996
0.994
0.992
0.990
θ^(100%)
θ^(80%)
θ^(60%)
θ^(40%)
θ^(20%)
θ^(0%)
Percent Accuracy in Threshold or Factor Loading Estimates


Validity of ML Factor Scores as Function of
Threshold (green) or Factor Loading (brown) Estimate Accuracy
Validity
0.80
0.81
0.82
0.83
0.84
0.85
θ̂(100%)
θ̂(80%)
θ̂(60%)
θ̂(40%)
θ̂(20%)
θ̂(0%)
Percent Accuracy in Threshold or Factor Loading Estimates

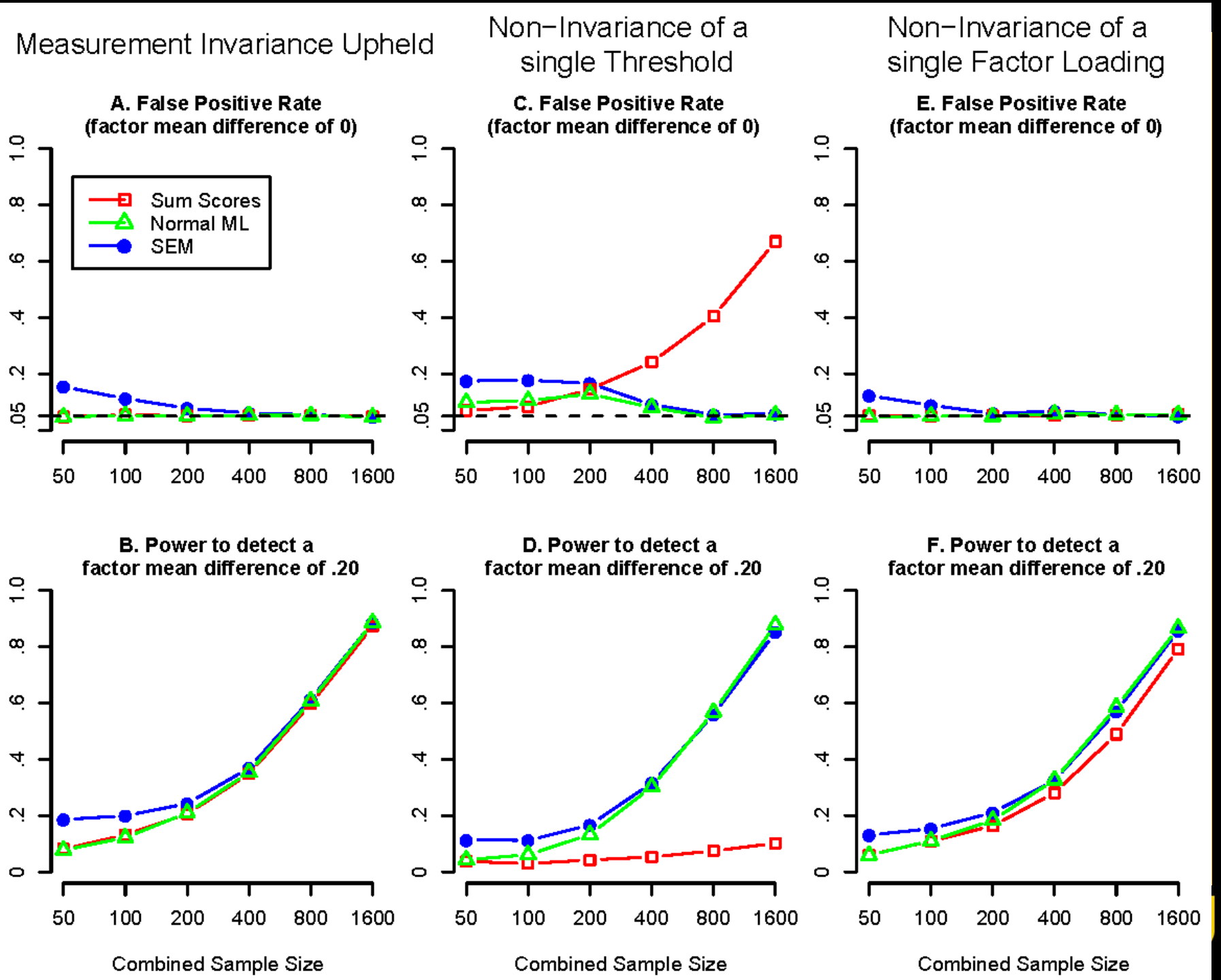

A. Reliability:
ML factor scores (red) vs.sum scores (blue)
Reliability of mean
1.000
0.995
0.990
0.985
0.980
50
100
200
400
800
1600


B. Validity:
ML factor scores (red) vs.sum scores (blue)
Validity
0.70
0.75
0.80
0.85
50
100
200
400
800
1600
Sample size (log scale)

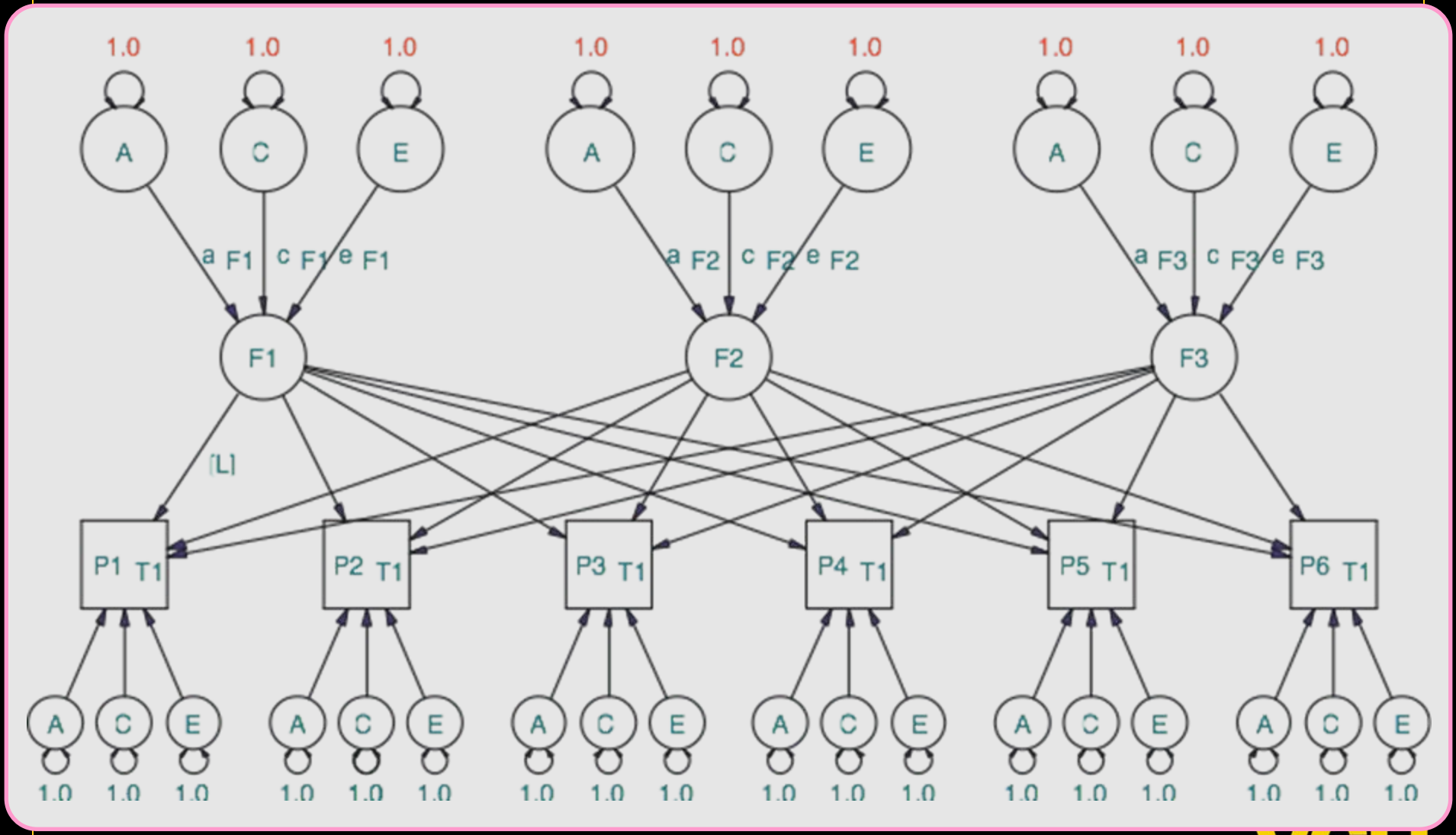
Measurement non-invariance

- Which works better: ML or Sum score?

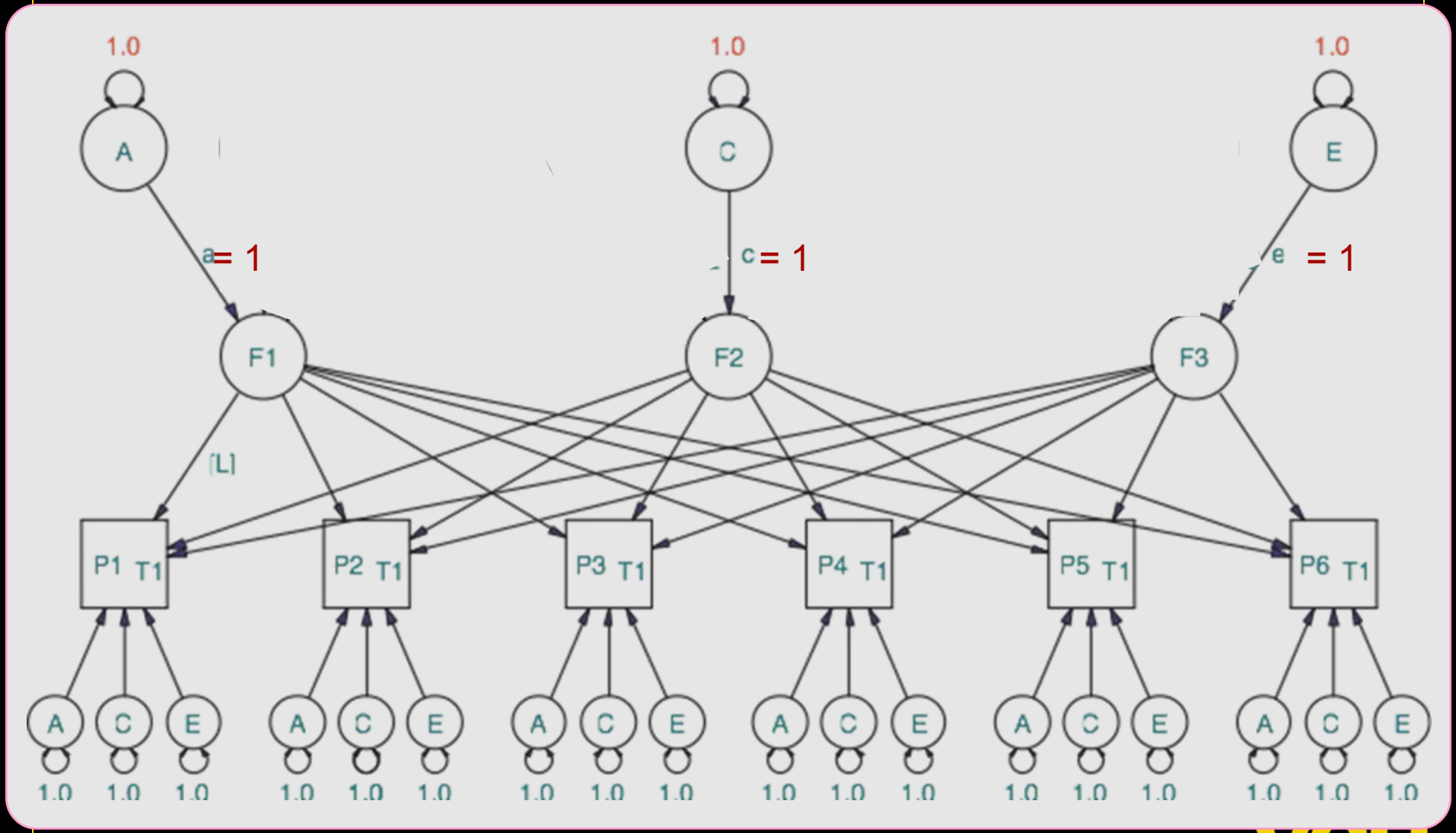
- Three tests:
  - SEM - Likelihood ratio test difference in latent factor mean
  - ML Factor score t-test
  - Sum score t-test


Measurement Invariance Upheld
Non−Invariance of a single Threshold
Non−Invariance of a single Factor Loading
A. False Positive Rate (factor mean difference of 0)
C. False Positive Rate (factor mean difference of 0)
E. False Positive Rate (factor mean difference of 0)
Sum Scores
Normal ML
SEM
B. Power to detect a factor mean difference of .20
D. Power to detect a factor mean difference of .20
F. Power to detect a factor mean difference of .20
0
.05
.2
.4
.6
.8
1.0
50
100
200
400
800
1600
Combined Sample Size

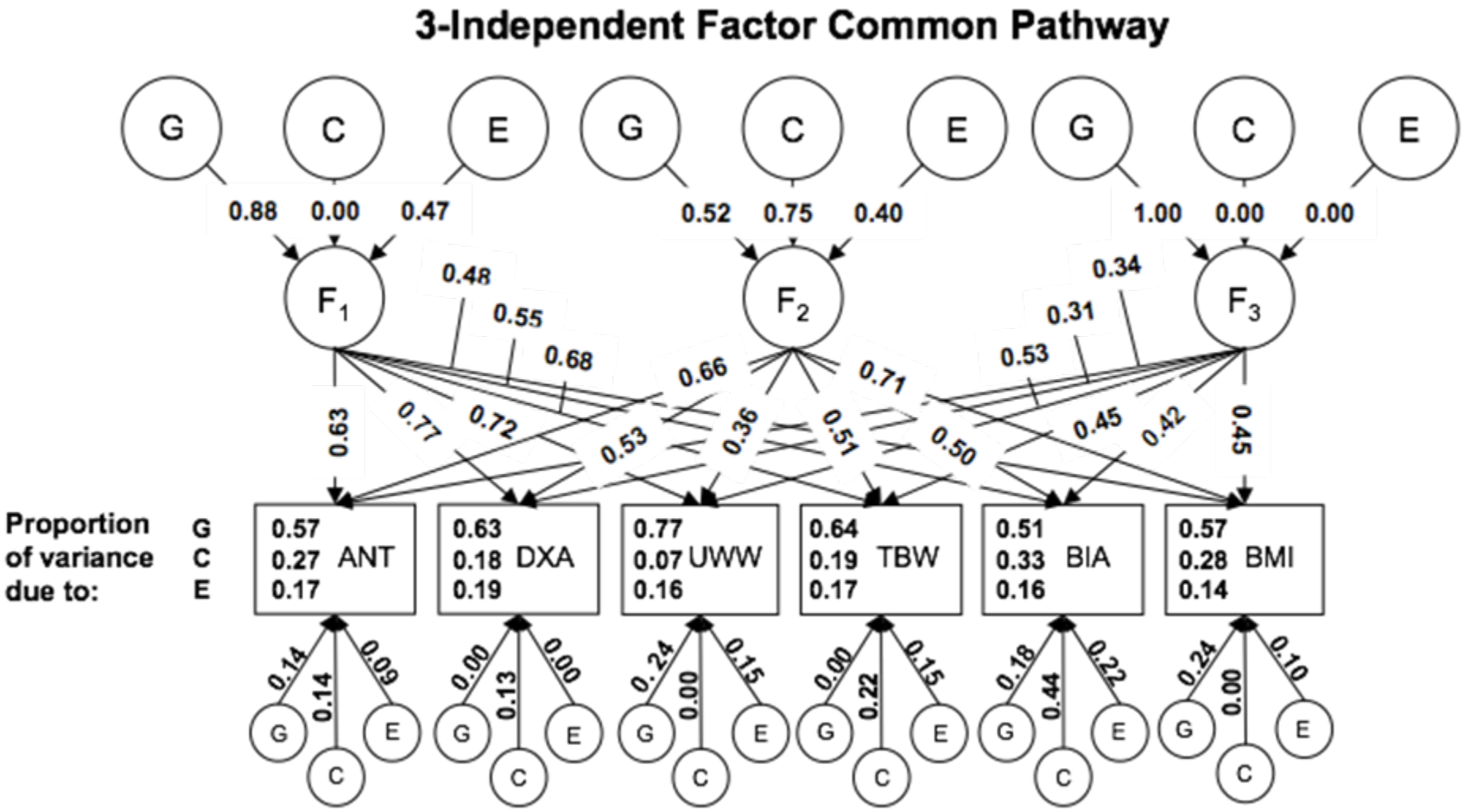
More Factors: Common Pathway Model



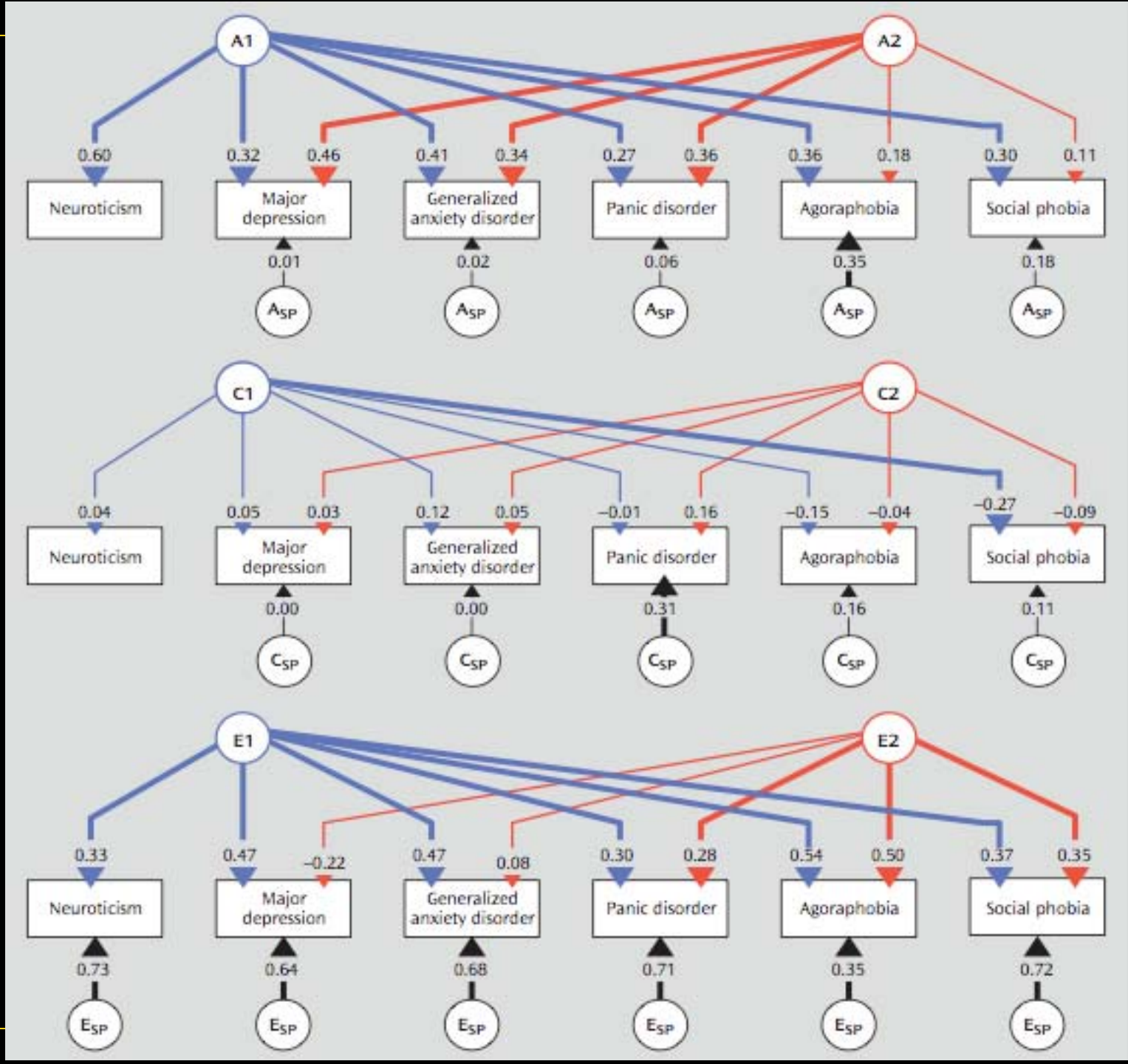
More Factors: Independent Pathway Model



Example: Fat MZT MZA



Results of fitting twin model



ML A, C, E or P Factor Scores

- Compute joint likelihood of data and factor scores
  - p(FS,Items) = p(Items|FS)*p(FS)
  - works for non-normal FS distribution
- Step 1: Estimate parameters of (CP/IP) (Moderated) Factor Model
- Step 2: Maximize likelihood of factor scores for each (family's) vector of observed scores
  - Plug in estimates from Step 1

VCU

Business end of FS script

```
!-----------------------------------------------------------------
#define $Tupper .432882        ! Test significance of difference of FS from this value
#define $Tlower -.095536       ! Test significance of difference of FS from this value

#define MZ 1      ! Set to 1 for MZ, 0 for DZ
#define $nvar 12 ! Number of variables altogether, before selection
#define nafac 2  ! Number of A common factors
#define ncfac 2  ! Number of C common factors
#define nefac 2  ! Number of E common factors
#define nafac2 nafac * 2 ! Twice number of A common factors (for DZ's only)
#define nefac2 nefac * 2 ! Twice number of E common factors
#if MZ = 1
#define DZ 0
#define nfac = nafac+ncfac+nefac2  ! Total factors must equal nafac+ncfac+nefac2 for MZ
#define $faclabels A1 A2 C1 C2 E1 E2 E3 E4 ! A1 A2...Anafac (MZ) C1 C2...Cnafac E1 E2...Enafac2
#else
#define DZ 1
#define nfac 10 ! Total factors must equal nafac2+ncfac+nefac2 for DZ
#define $faclabels A1 A2 A3 A4 C1 C2 E1 E2 E3 E4 ! A1 A2...Anafac2 (DZ) C1 C2...Cnafac E1 E2...Enafac2
#endif
#define nv 6      ! Number of observed variables per twin
#define ncov 0    ! number of covariates
#define $variable1 MDD GAD PAN AGO SOC NEU
#define $variable2 MDD1 GAD1 PAN1 AGO1 SOC1 NEU1 MDD2 GAD2 PAN2 AGO2 SOC2 NEU2
#define maxcat 12  ! Maximum score of any item
#define $highests 1 1 1 1 1 12 1 1 1 1 1 12  ! Highest value of each item
```

The guts of it

```
! Labels should be A1 A2 etc up to nafac
 Labels Col X A1 A2
 Matrix X
0.55 0.15
0.48 0.25
0.31 0.23
0.19 0.39
0.11 0.33
0.24 0.54
! Labels should be C1 C2 etc up to ncfac
 Labels Col Y C1 C2
 Matrix Y
0.08 0.01
0.03 -.08
0.40 -.06
0.08 0.23
0.04 0.29
-.10 -.07
! Labels should be E1 E2 etc up to nefac
 Labels Col Z E1 E2
 Matrix Z
0.09 0.46
0.36 0.36
0.40 0.12
0.66 0.21
0.45 0.15
0.14 0.29
!
```

! Residuals only

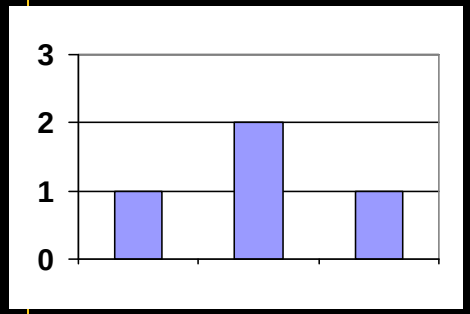
```
  A = P.P;
  C = Q.Q;
  E = R.R;
  M = (L*F')';   ! Means conditional on factor scores
 End Algebra;

 Thresholds   (T|T) - V@M ;
#if MZ = 1
 Covariance   (A+C+E|A+C_A+C|A+C+E);
#else
 Covariance   (A+C+E|h@A+C_h@A+C|A+C+E);
#endif
 Weight       \pdfnor(F_0_S);
```

Shell script to FS everyone

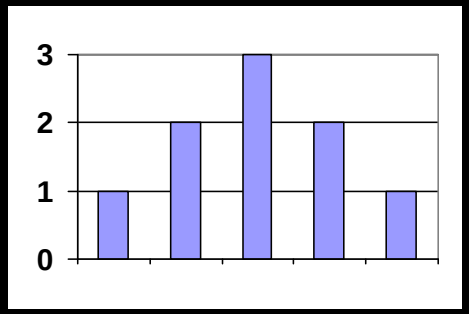
```
#!/bin/ksh
#PBS -q serial
#PBS -N Mx-facscore
#
echo "******STARTING*****************************"
#
# cd to the directory from which I submitted the
# job.  Otherwise it will execute in my home directory.
#
echo "Working directory of this job is: " $PBS_O_WORKDIR
#
cd $PBS_O_WORKDIR
# shell script to run bat.mx several times
#MZ zyg group 1 first
awk '{if ($1==1) print $2, $3, $4, $5, $6, $7, $8, $9, $10, $11, $12, $13}' twindata.rec > tmp
i=`wc tmp| awk '{print $1}'`
while (( 0 < i )) ; do
tail -$i tmp | head -1 > pattern1
cat pattern1 >> f.mat
/usr/local/bin/mxt156f < fzyg1.mx >>fzyg1.mxo
        (( i = i - 1 ))
done
!mv f.mat fzyg1.mat
grep 'Difference Chi' fzyg1.mxo > chidiffs.txt
#
# Repeat above for other zygosities replacing $1==1  with $1==2 etc
# and make mv f.mat fzyg1.mat fzyg2.mat etc...
#
```

Central Limit Theorem
Additive effects of many small factors

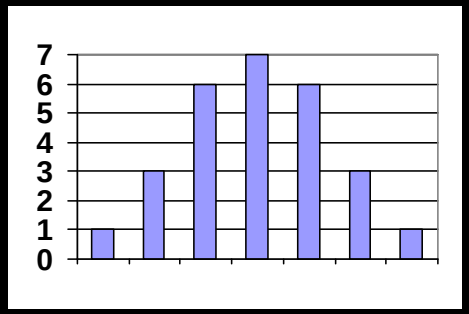
1 Gene
→ 3 Genotypes
→ 3 Phenotypes



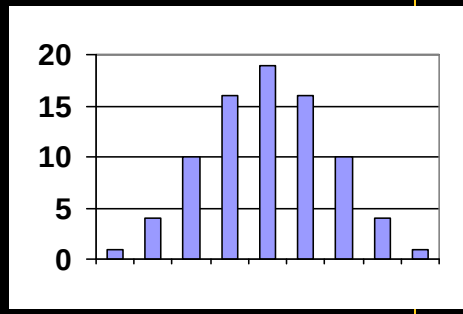
2 Genes
→ 9 Genotypes
→ 5 Phenotypes



3 Genes
→ 27 Genotypes
→ 7 Phenotypes



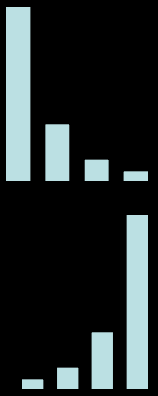
4 Genes
→ 81 Genotypes
→ 9 Phenotypes



VCU

Measurement artifacts

- Few binary items
- Most items rarely endorsed (floor effect)
- Most items usually endorsed (ceiling effect)



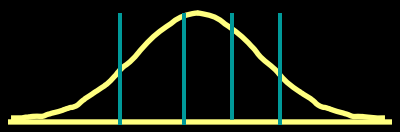
- Items more sensitive at some parts of distribution
- Non-linear models of item-trait relationship

VCU

Assessing the distribution of latent trait

- Schmitt et al 2006 MBR method
- N-variate binary item data have 2^N possible patterns
- Normal theory factor model predicts pattern frequencies
  - E.g., high factor loadings but different thresholds
  - 0 0 0 0
  - 0 0 0 1
  - 0 0 1 1
  - 0 1 1 1
  - 1 1 1 1

but 0 0 1 0 would be uncommon

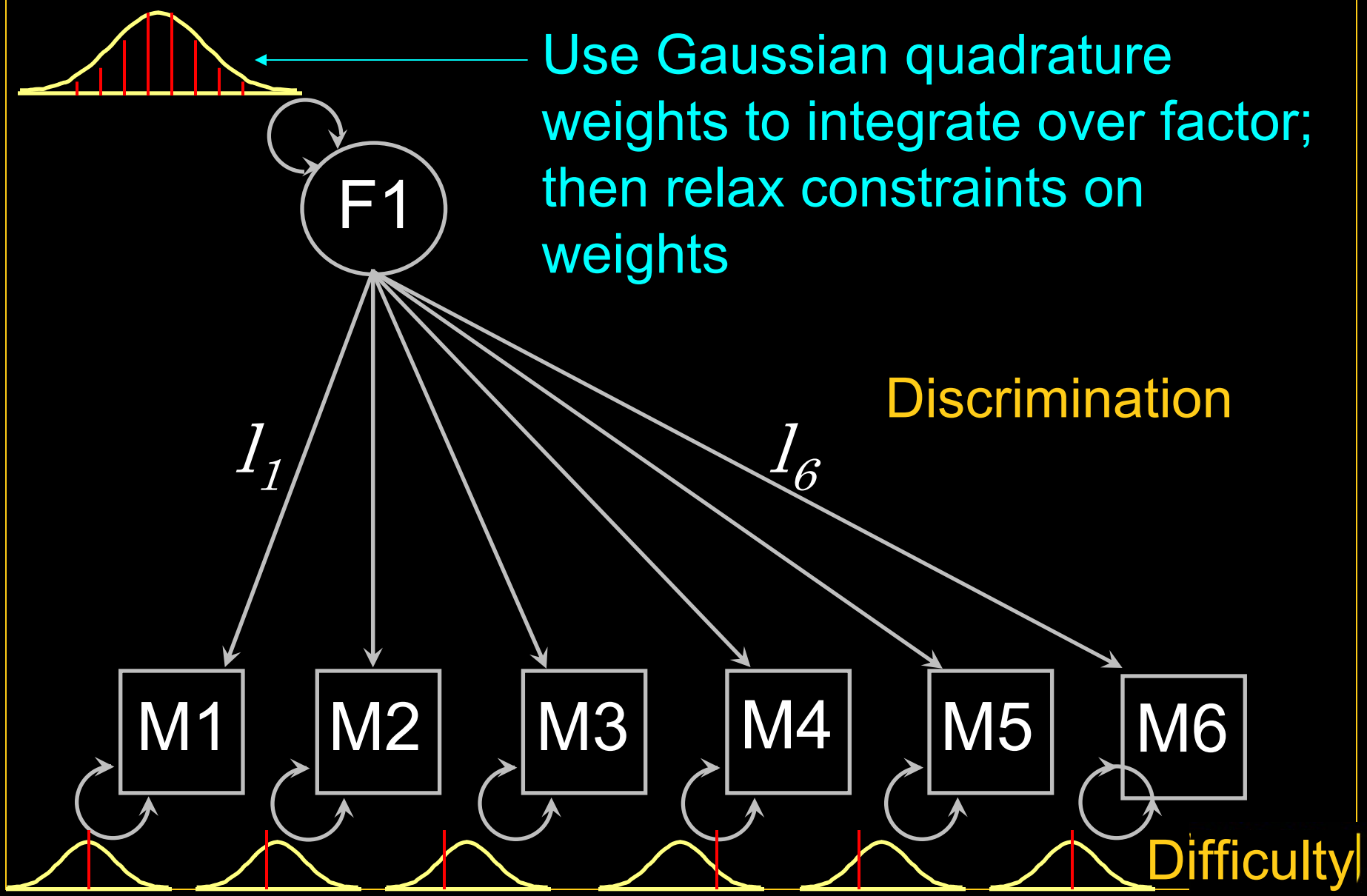


1 2 3 4

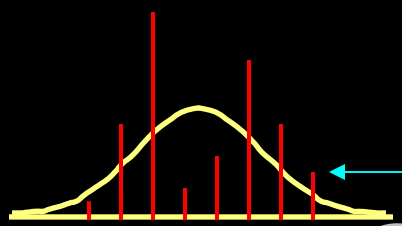
item threshold

VCU

Latent Trait (Factor) Model



Latent Trait (Factor) Model



Difference in model fit:
LRT~ χ^2

F1

Discrimination

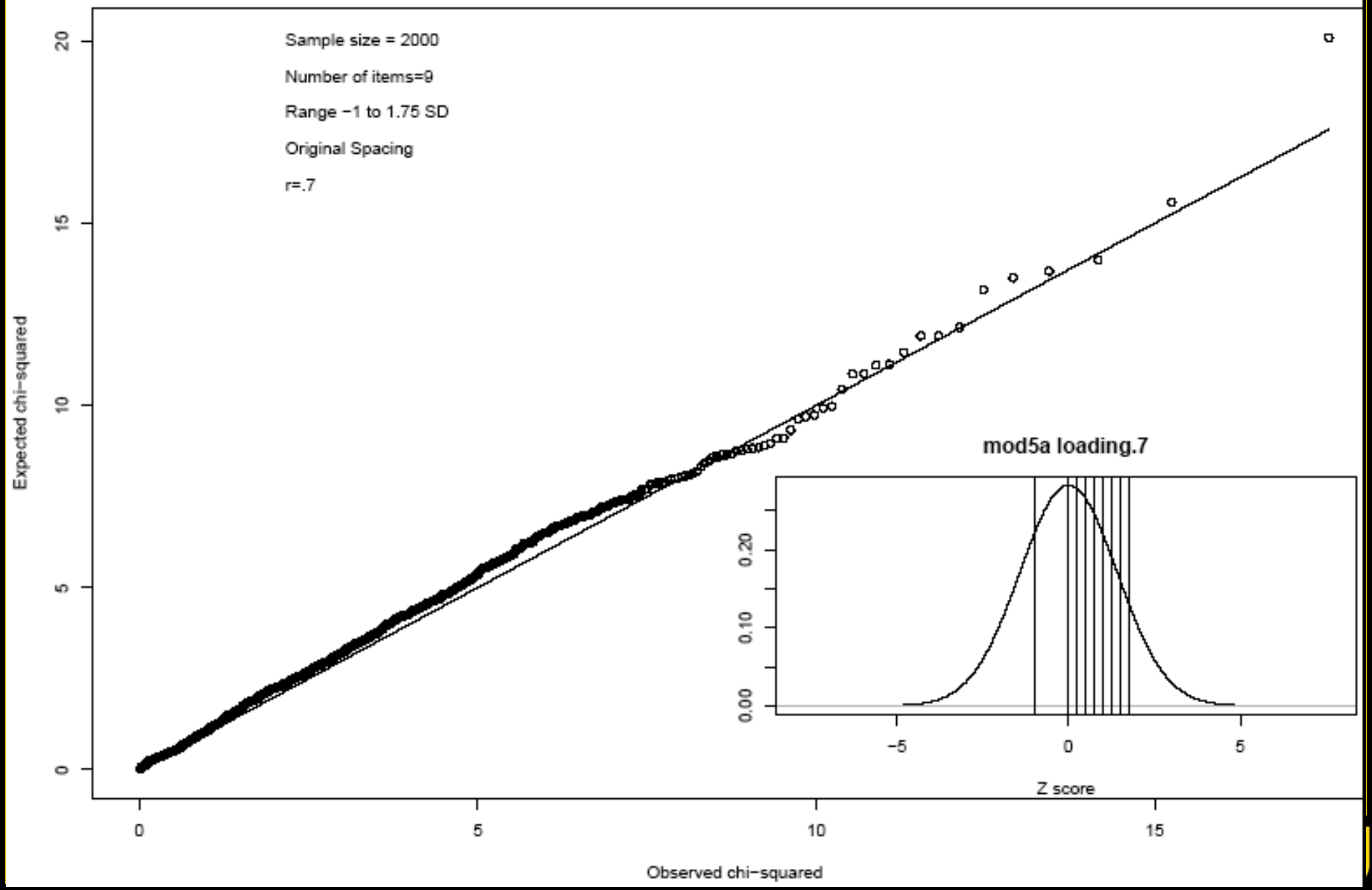
l_1

l_6

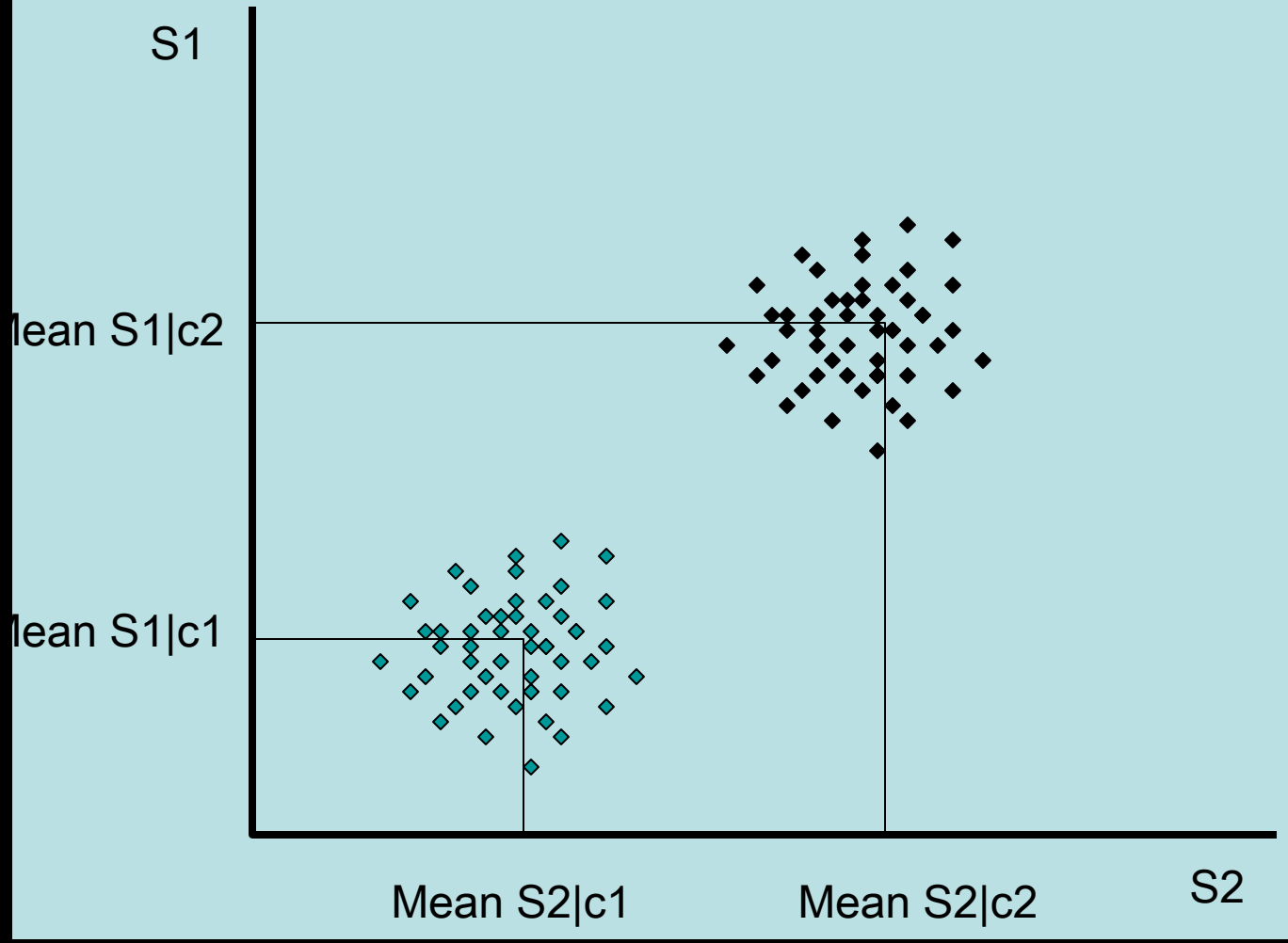
M1 M2 M3 M4 M5 M6

Difficulty

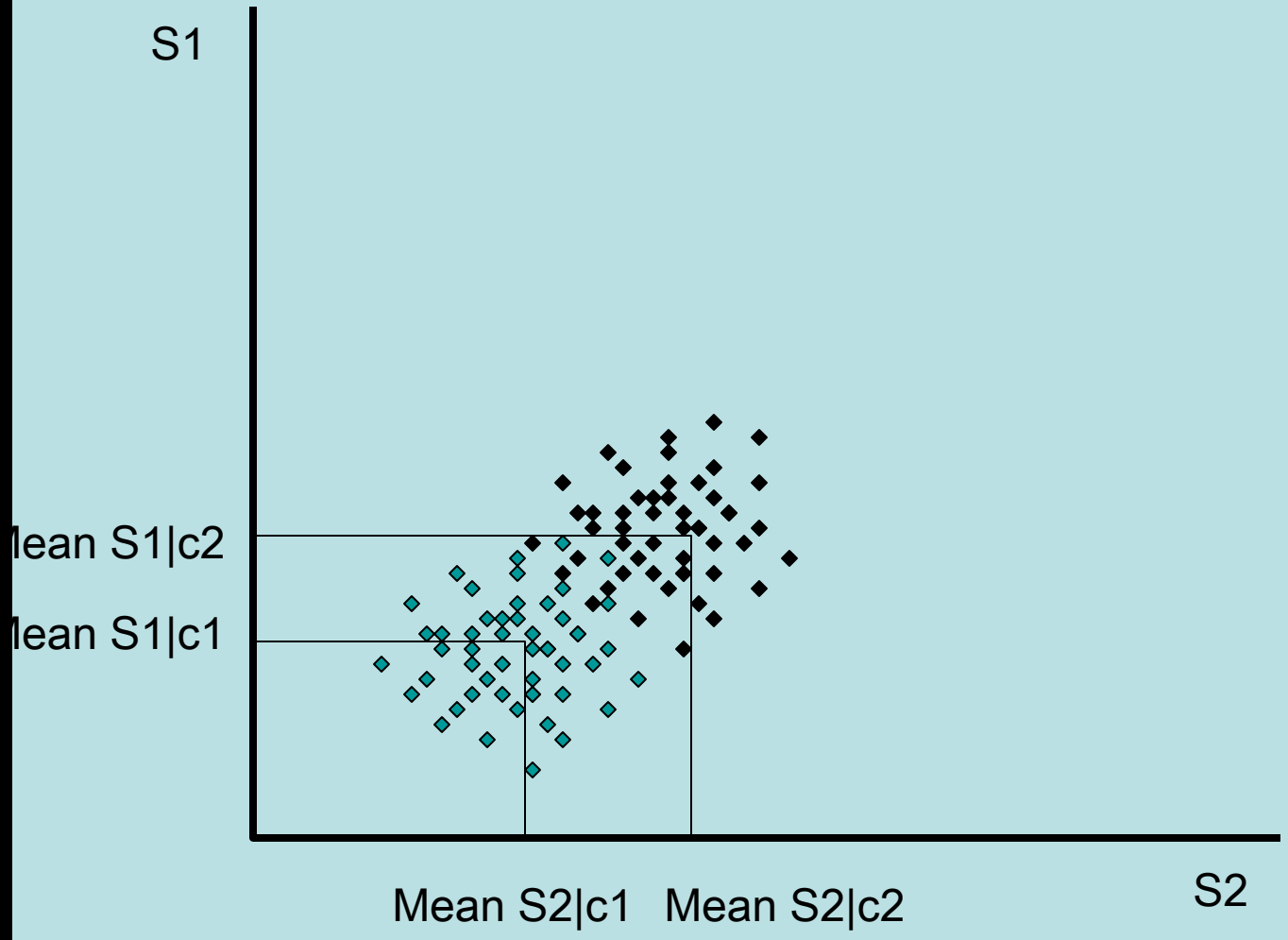
Chi-squared test for non-normality performs well



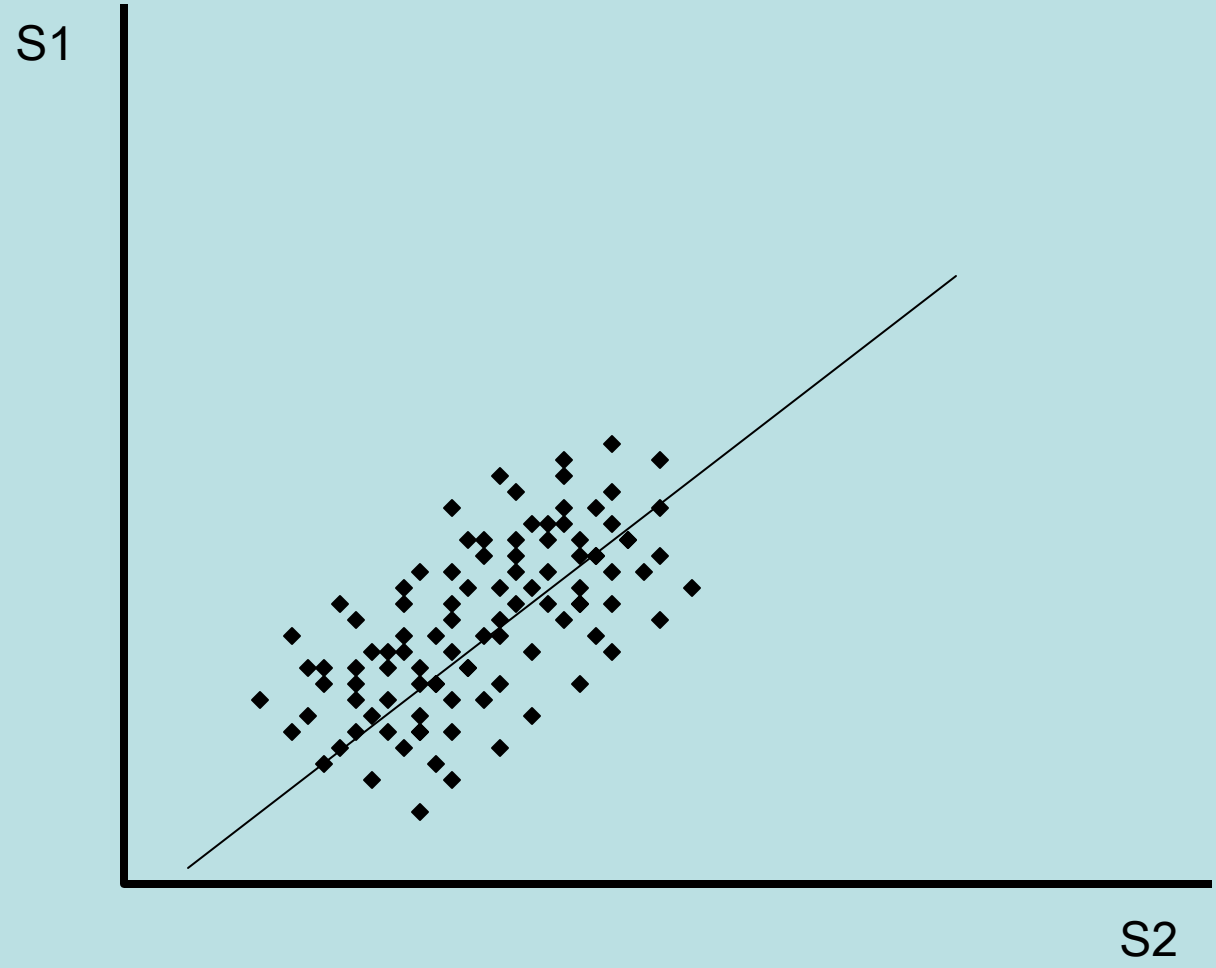
Detecting latent heterogeneity
Scatterplot of 2 classes

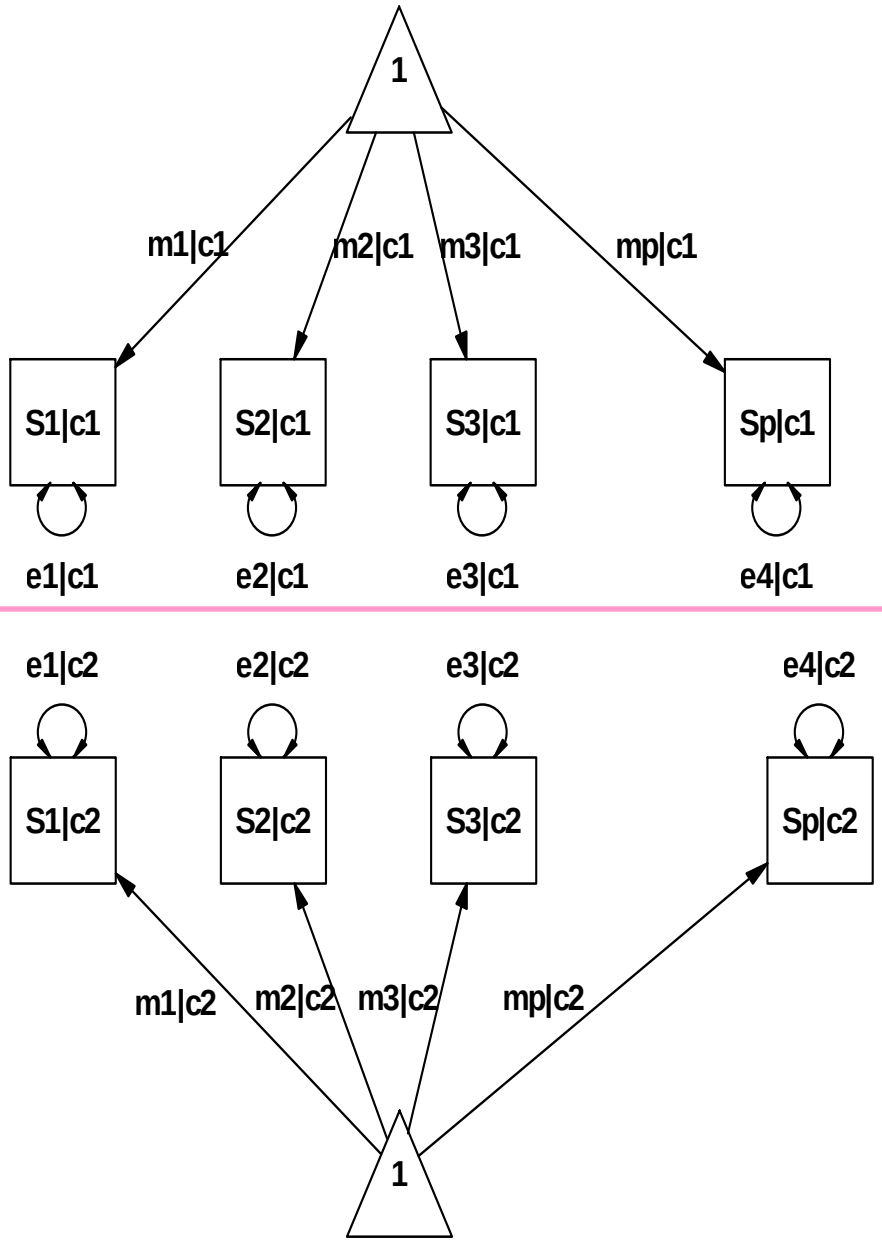


Scatterplot of 2 classes
Closer means



Scatterplot of 2 classes
Latent heterogeneity: Factors or classes?



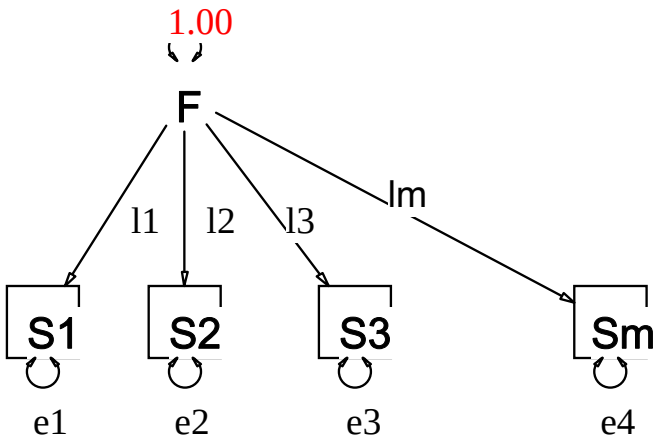


Class Membership probability

Class 1: p

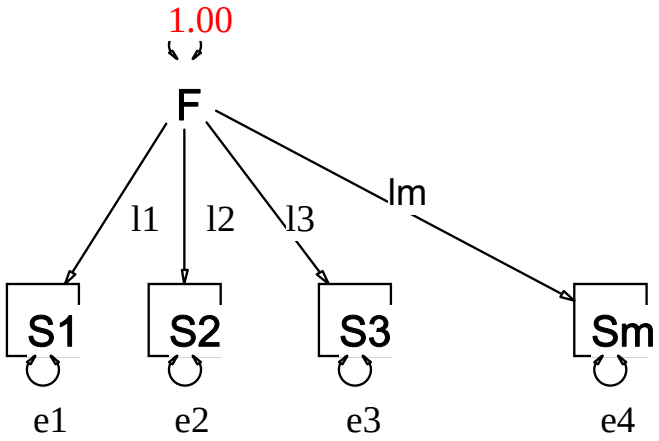
Class 2: (1-p)

Factor Mixture Model



Class Membership probability

Class 1: p



Class 2: (1- p)

NB means omitted

Classes or Traits?
A Simulation Study

- Generate data under:
  - Latent class models
  - Latent trait models
  - Factor mixture models
- Fit above 3 models to find best-fitting model
  - Vary number of factors
  - Vary number of classes

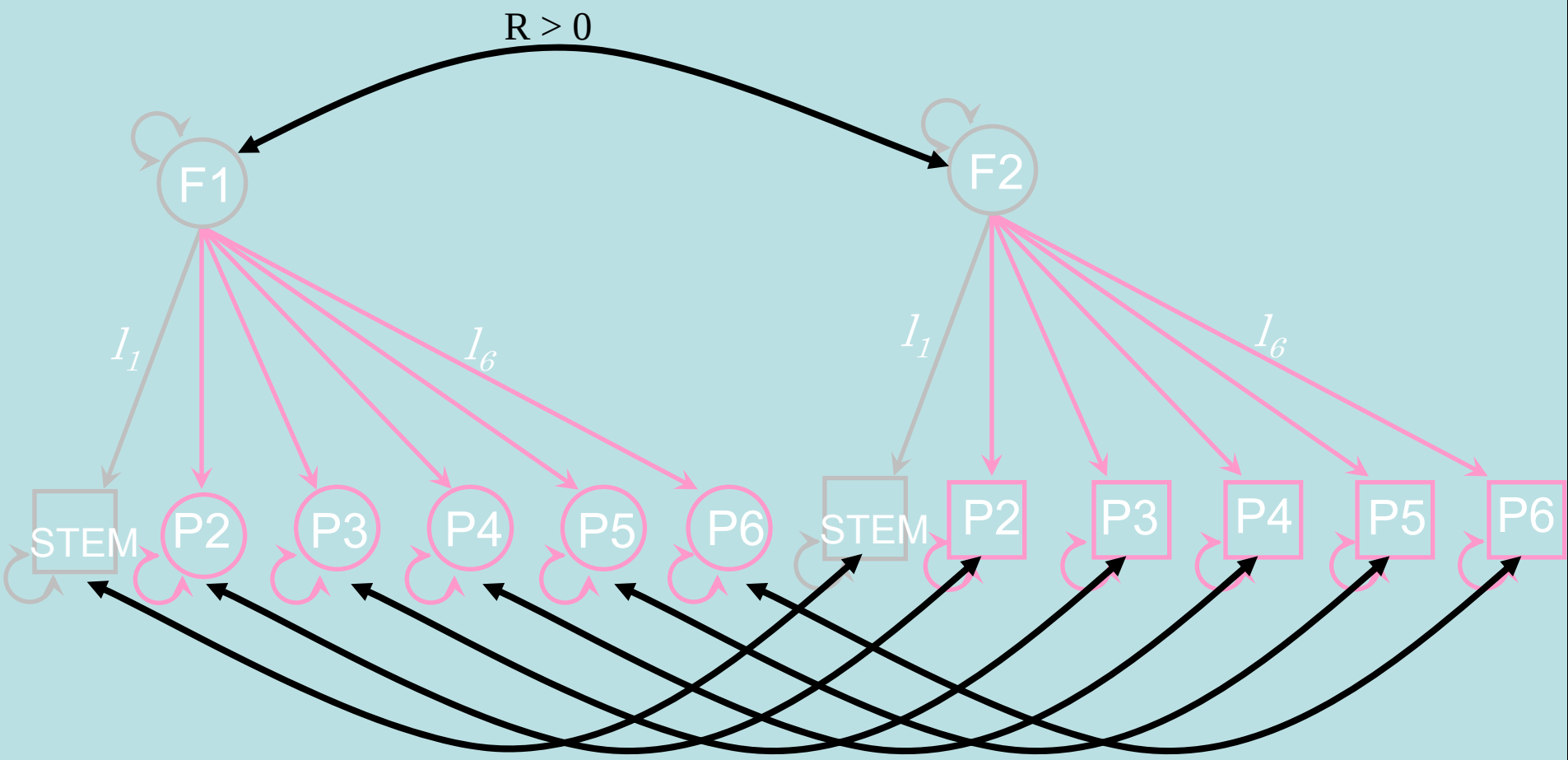
- See Lubke & Neale Multiv Behav Res (2007 & In press)

VCU

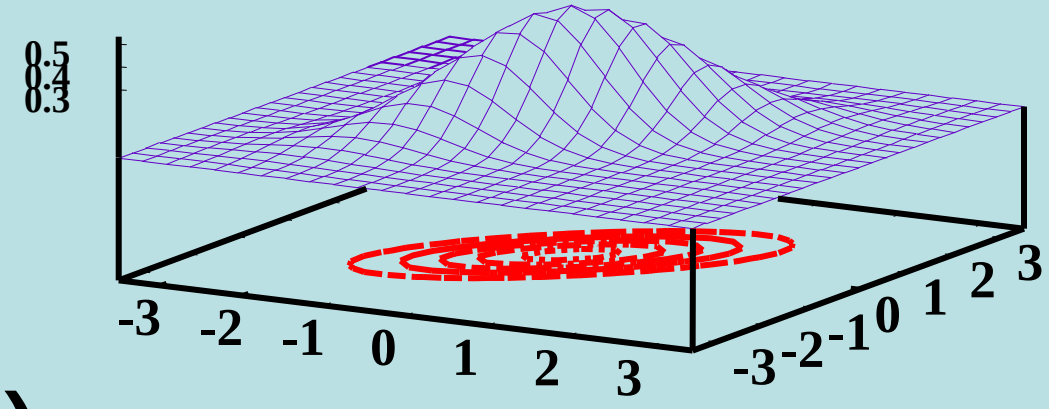
What to do about conditional data

- Two things
  - Different base rates of “Stem” item
  - Different correlation between Stem and “Probe” items
- Use data collected from relatives

Data from Relatives: Likely failure of conditional independence



Series of bivariate integrals



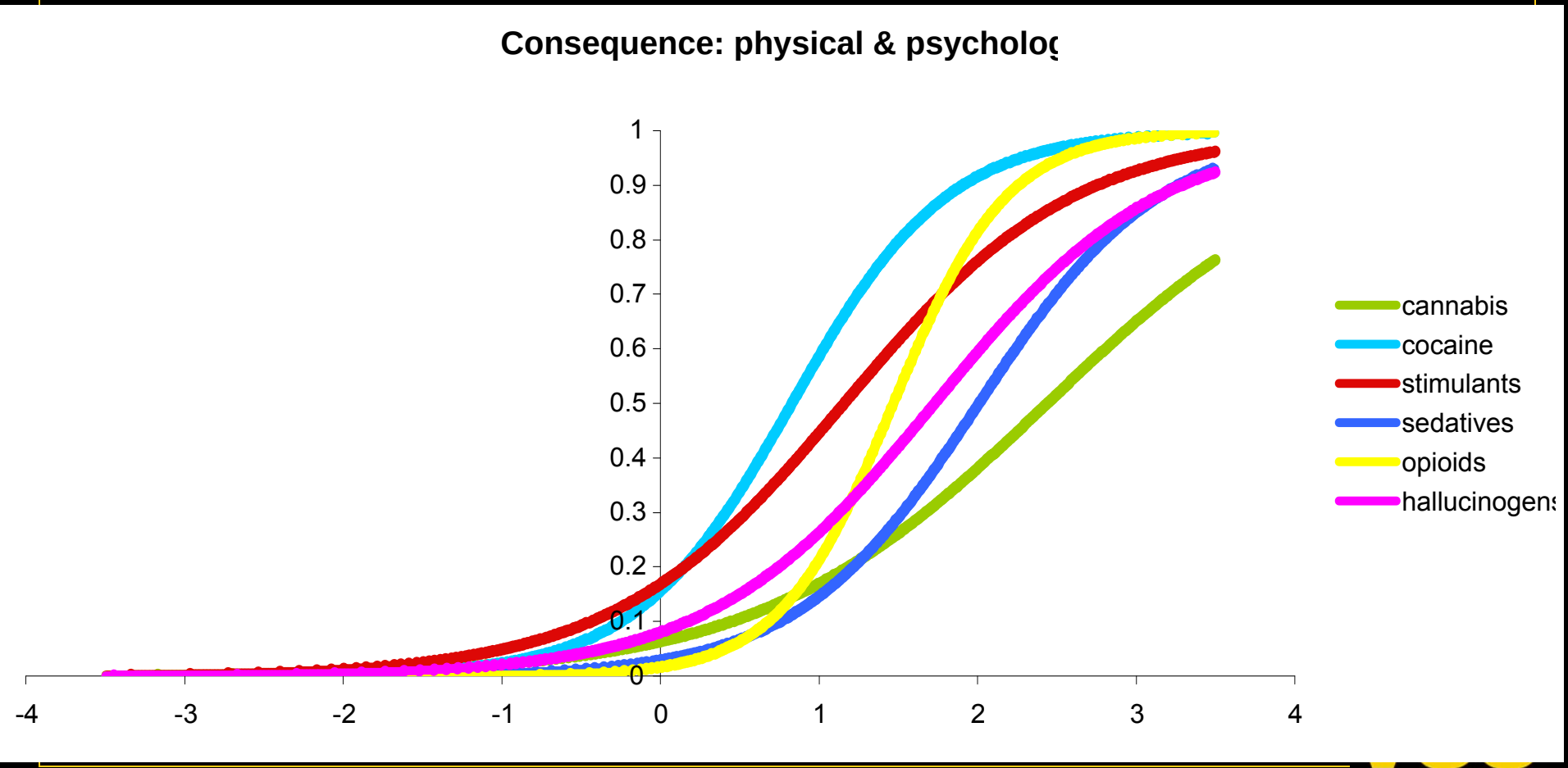
$$\prod_{j=1}^{m/2} \left(\int_{i-1}^{t1_i} \int_{i-1}^{t2_i} \phi(x_1, x_2)\, dx_1\, dx_2 \right)_j$$

Can work with p-variate integration, best if p<m ☺

"Generalized MML" built into Mx

Dependence 1

Did your use of it cause you physical problems or make you depressed or very nervous?

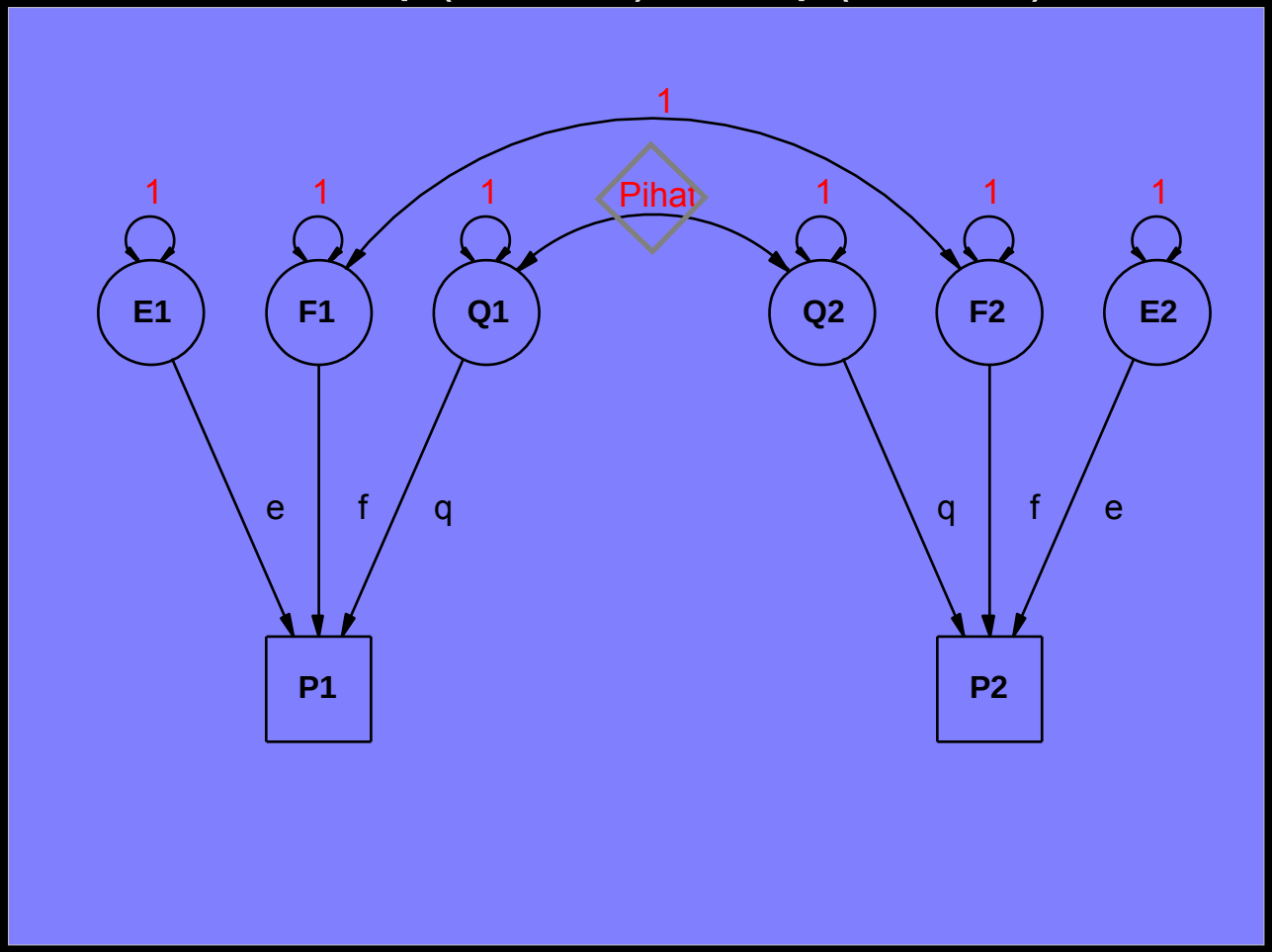


Extensions to More Complex Applications

- Endophenotypes
- Linkage Analysis
- Association Analysis

Basic Linkage (QTL) Model

$\hat{\pi}$ = p(IBD=2) + .5 p(IBD=1)



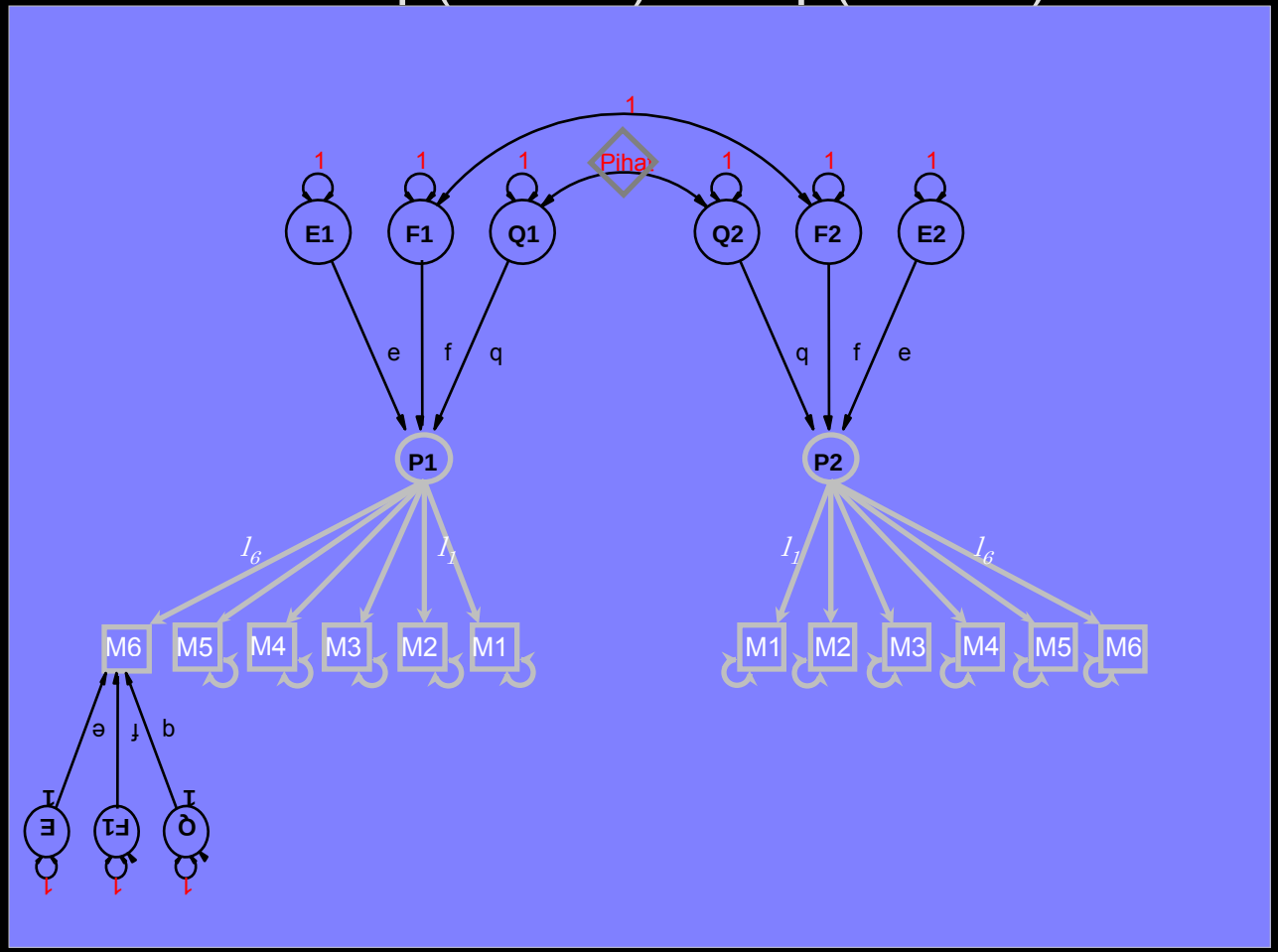
Q: QTL Additive Genetic F: Family Environment E: Random Environment

3 estimated parameters: q, f and e Every sibship may have different model

VCU

Measurement Linkage (QTL) Model

$\hat{\pi}$ = p(IBD=2) + .5 p(IBD=1)

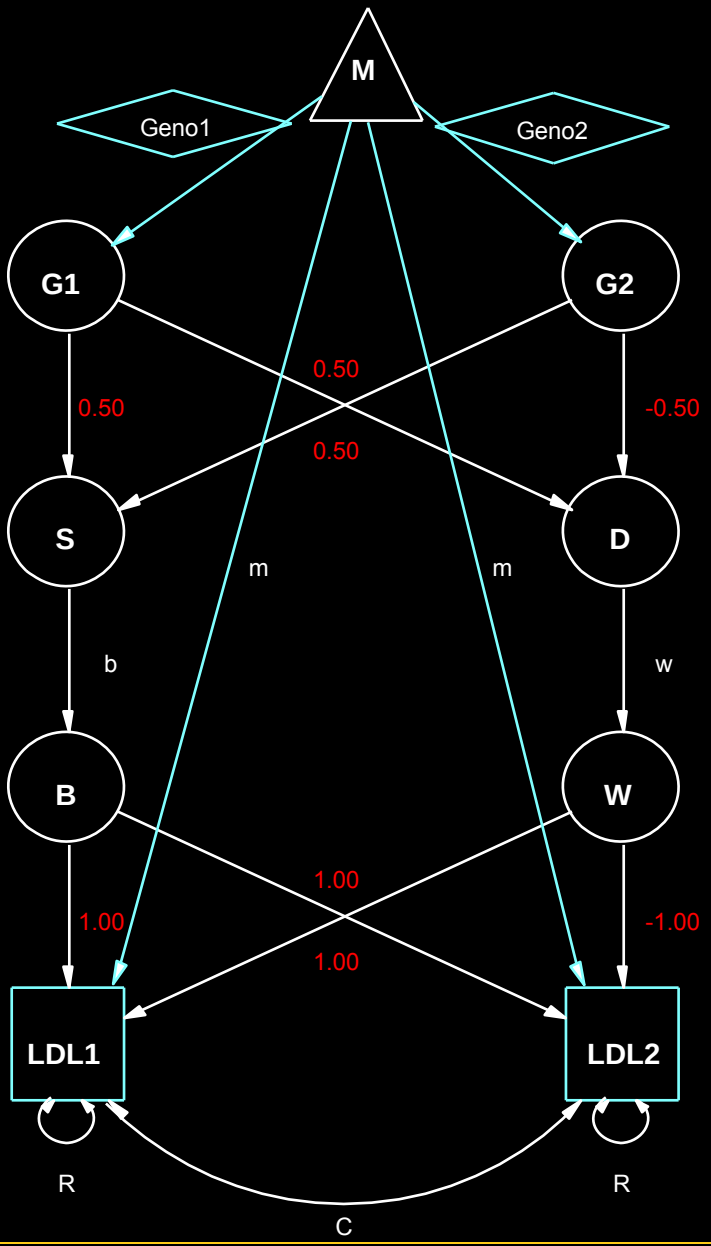


Q: QTL Additive Genetic F: Family Environment E: Random Environment

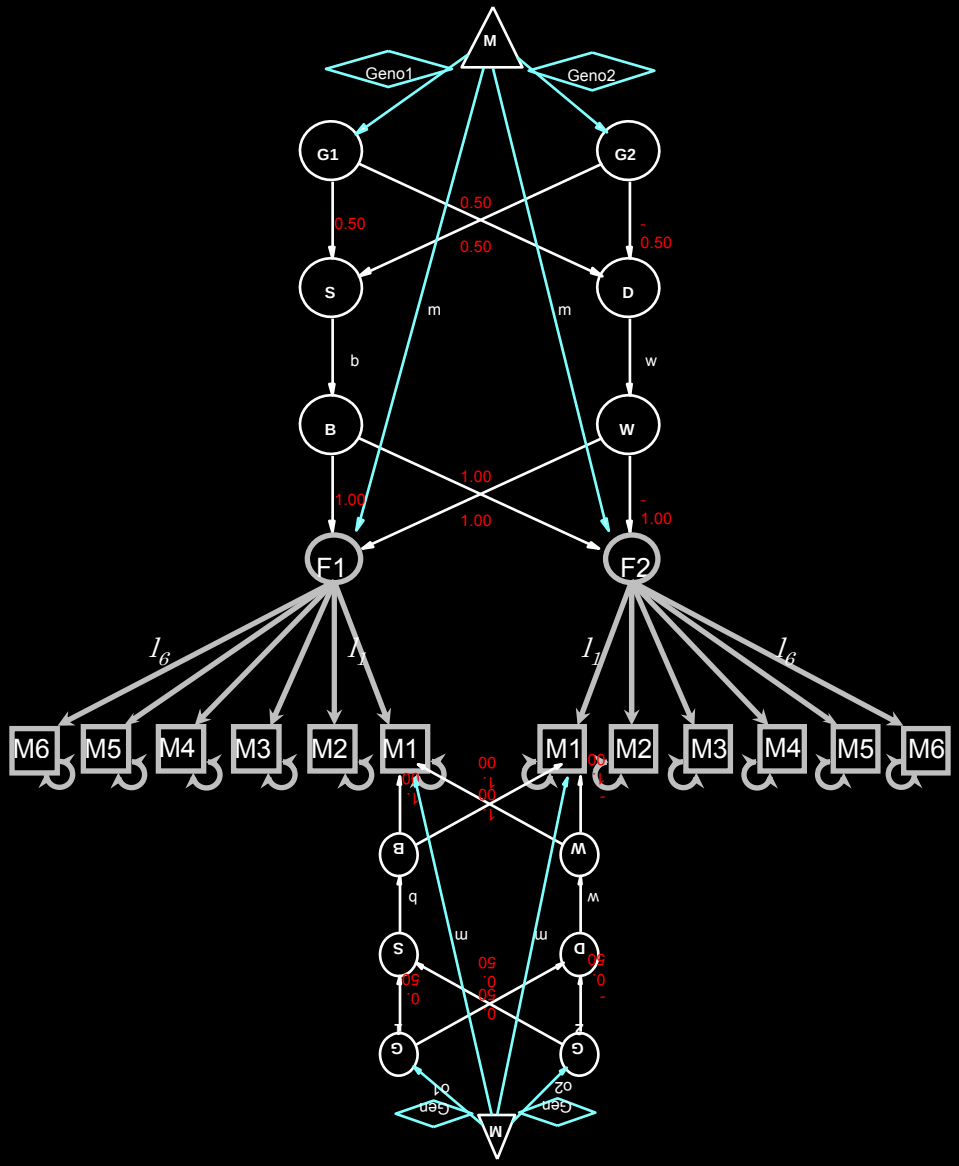
3 estimated parameters: q, f and e Every sibship may have different model

Fulker Association Model

Multilevel model for the means



Measurement Fulker Association Model (SM)



Multivariate Linkage & Association Analyses

- Computationally burdensome
- Distribution of test statistics questionable
- Permutation testing possible
  - Even heavier burden
- Potential to refine both assessment and genetic models
- Lots of long & wide datasets on the way
  - Dense repeated measures EMA
  - fMRI
  - Need to improve software! Open source Mx

VCU