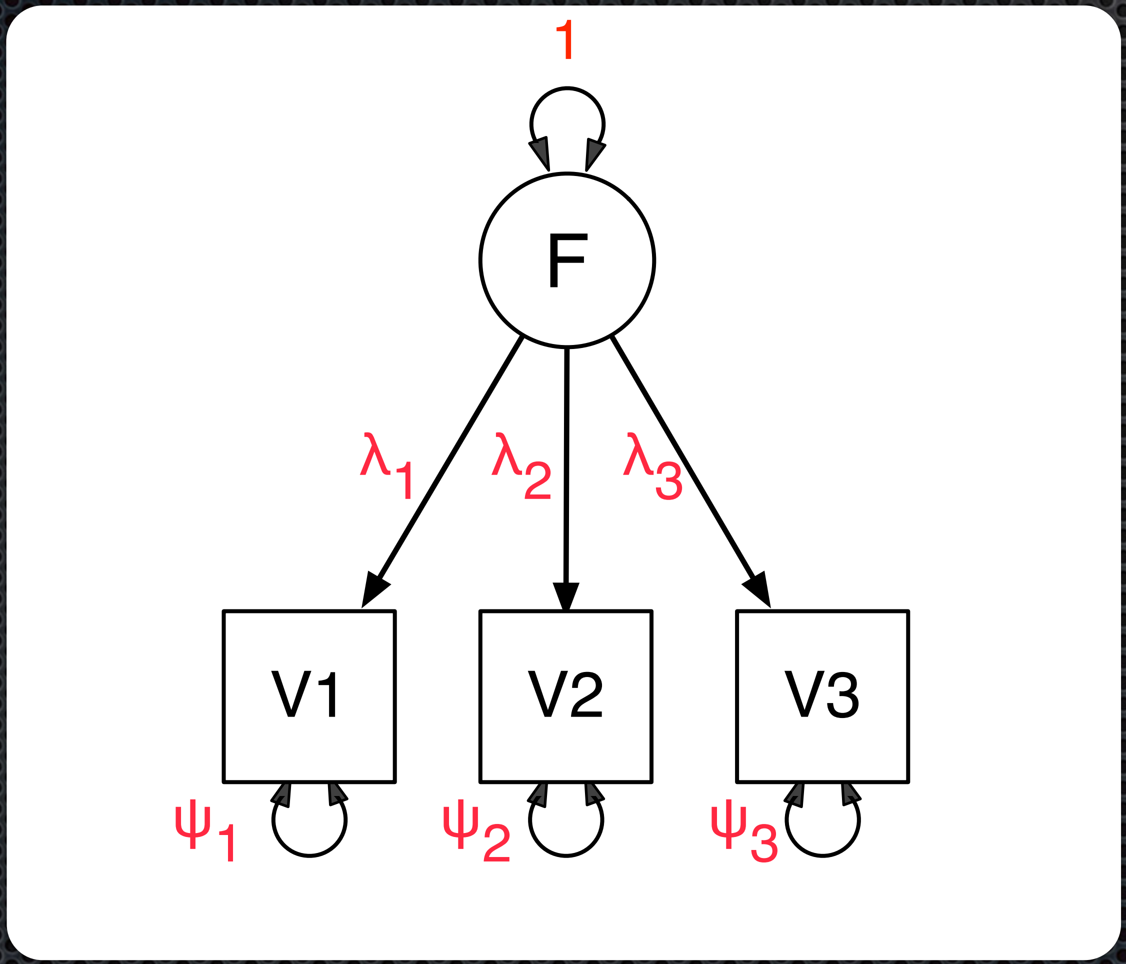
Phenomics

Michael C. Neale
Virginia Institute for Psychiatric & Behavioral Genetics
Virginia Commonwealth University

Boulder Workshop
March 6 2015

Table of Contents

Measurement Invariance: Factor Model



We *usually* want to know about the factor!

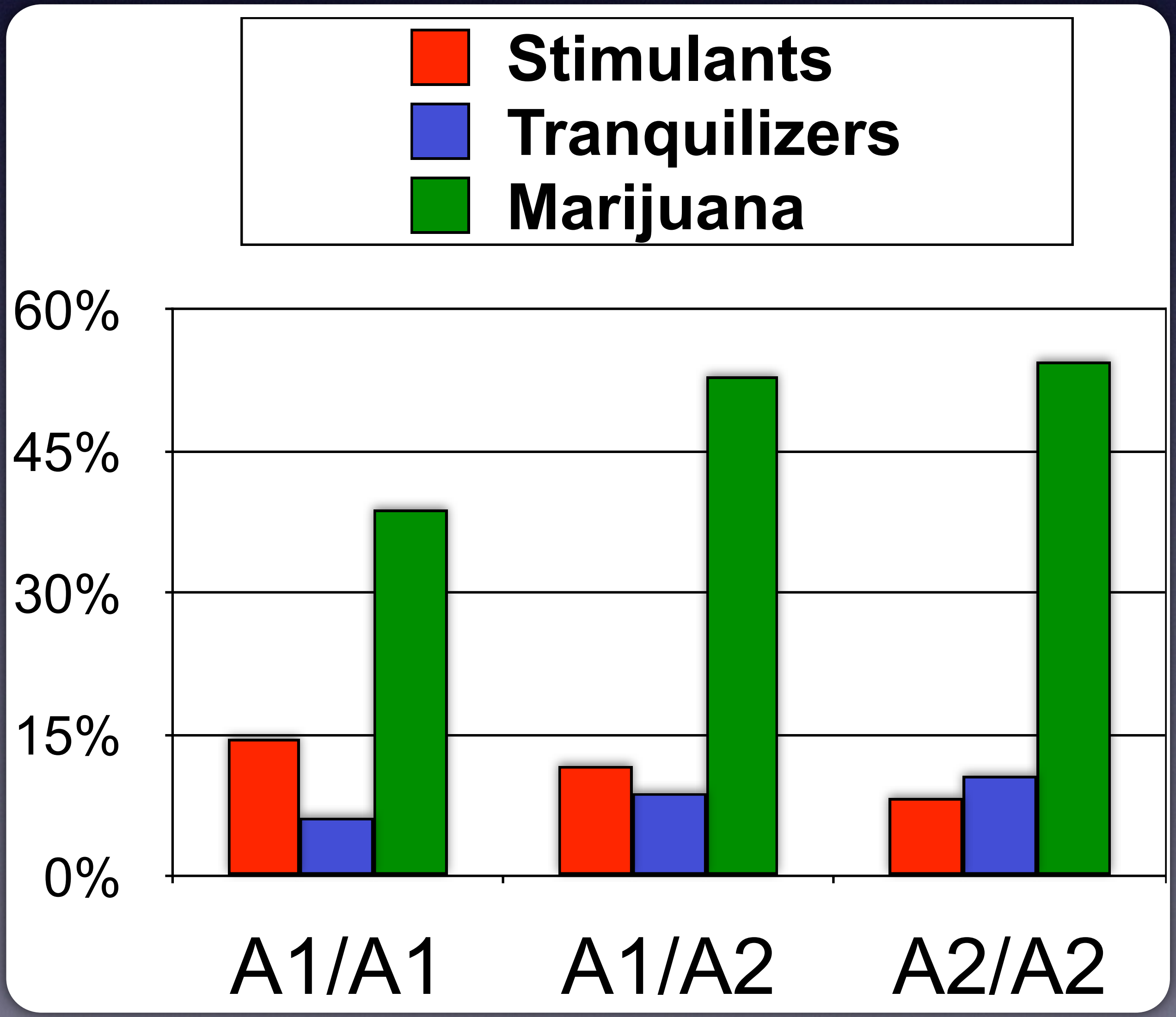
Correlations across Substances: Add Health

	Stimulants	Tranquilizers	Marijuana
Stimulants	1		
Tranquilizers	0.74	1	
Marijuana	0.63	0.66	1
Factor Loadings	0.84	0.87	0.75

Medland & Neale (2010) An integrated phenomic approach to multivariate allelic association
European Journal of Human Genetics 18, 233–239

DRD2 Association Results

- Univariate associations
  - Stimulants: χ^2=3.88, β= -.18, $p < .05$
  - Tranquilizers: χ^2=1.65, β= .13, NS
  - Marijuana: χ^2=2.60, β= .11, NS
- Factor level association
  - χ^2=0.65, kF= .06, NS
- Multivariate association
  - χ^2=13.91 (3df; $p < 0.005$)
  - $k_{Stimulants}$ = -0.19
  - $k_{Tranquilizers}$ = 0.14
  - $\beta_{Marijuana}$ = 0.11



Measurement Invariance Classic Papers

- Meredith, W. (1993) Measurement invariance, factor analysis, and factorial invariance. *Psychometrika* 58:525–543

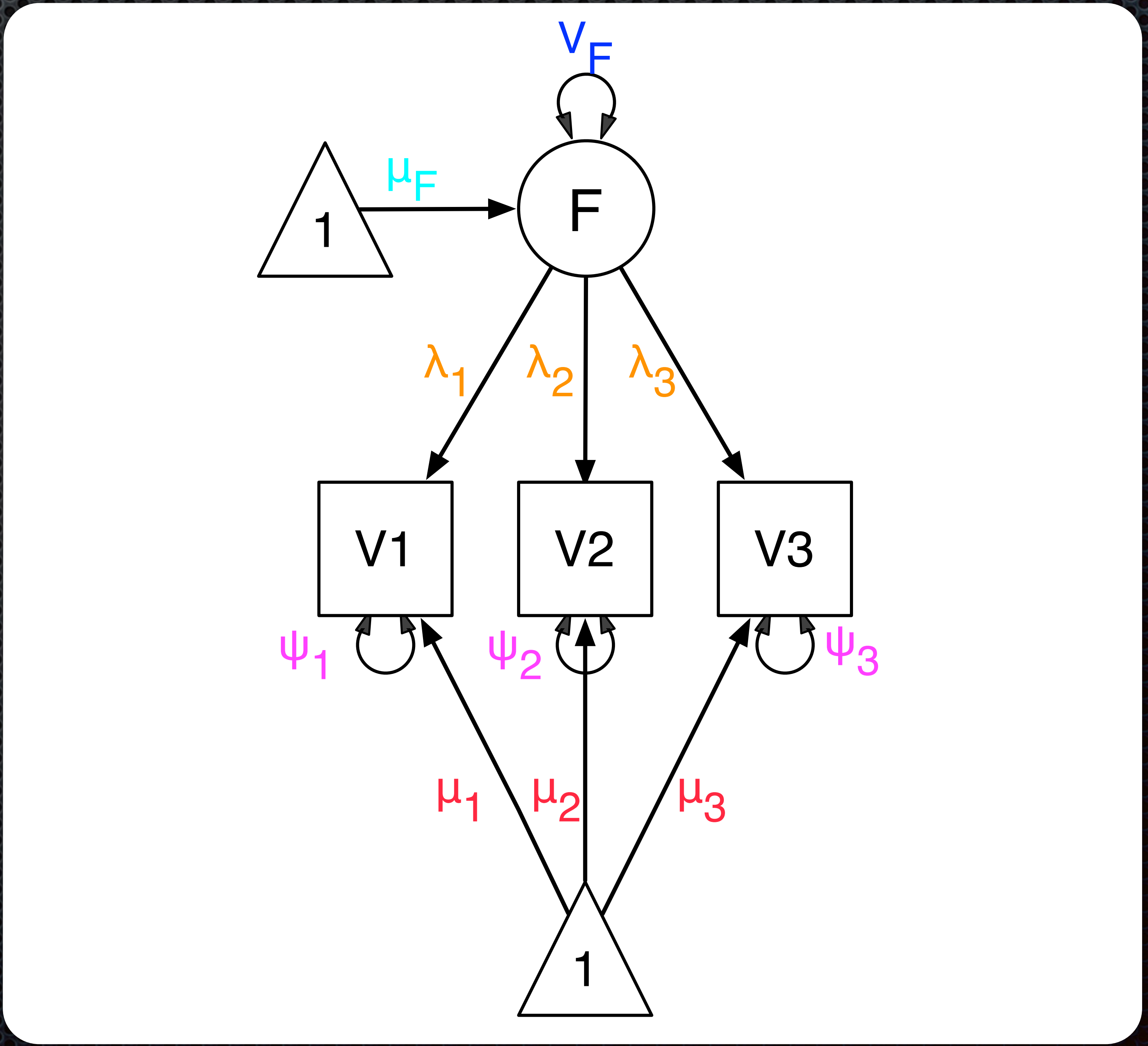
- Millsap RE & Jenn Y-T (2004) Assessing Factorial Invariance in Ordered-Categorical Measures. *Multivariate Behavioral Research* 39:479-515

- Widaman KF, Ferrer E, & Conger RD (2010). Factorial Invariance within Longitudinal Structural Equation Models: Measuring the Same Construct across Time. *Child Dev Perspect* 4:10–18

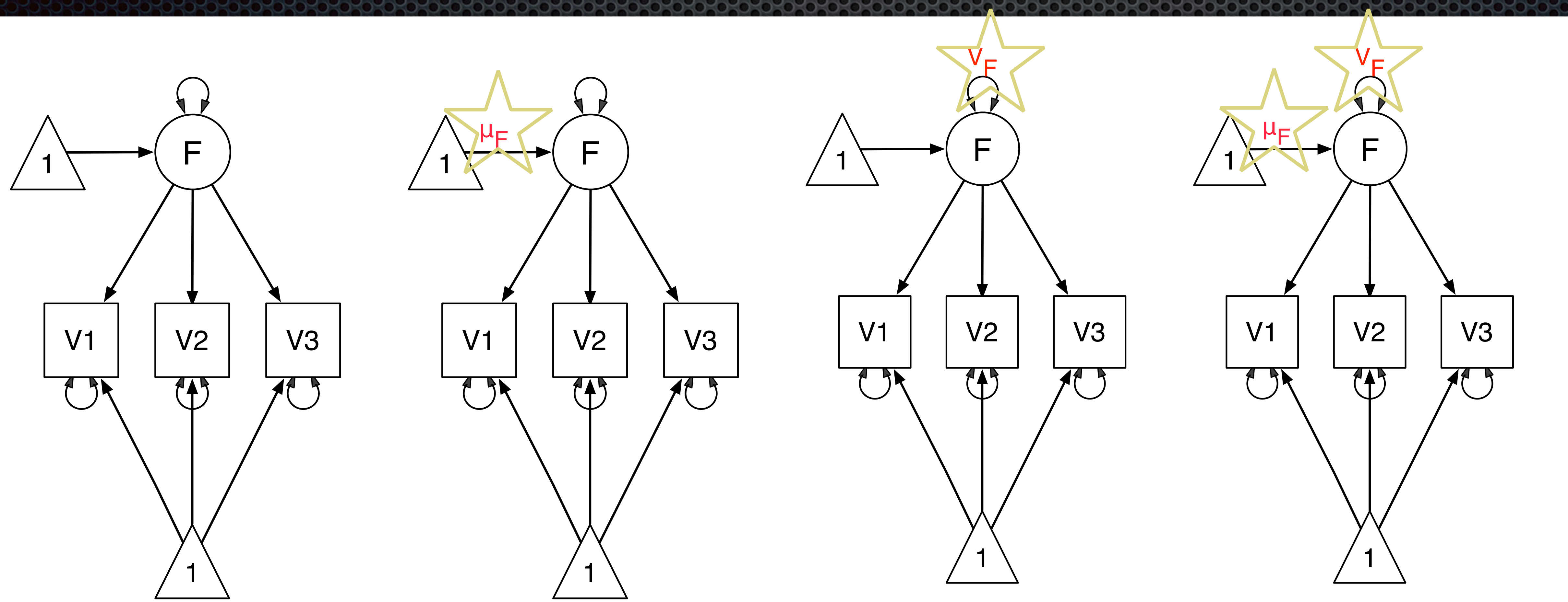
- Vandenberg, RJ & Lance, CE (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods* 3:4–70

Invariance: Five Potential Types of Difference

- Factor Variances
- Factor Means
- Factor Loadings
- Item Variances
- Item Means



Invariance Models of Factor-Level Effects wrt Sex and Age



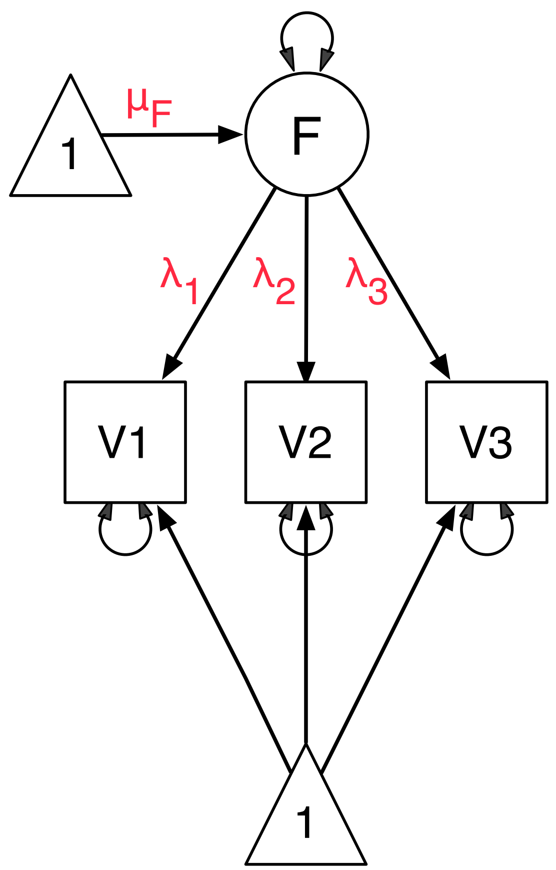
1. No Covariates

2. Age/Sex on Factor Mean

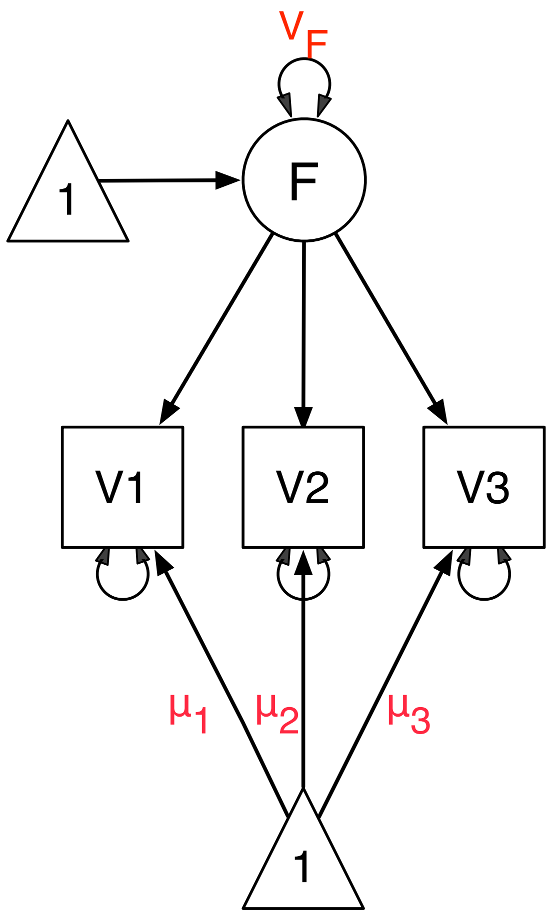
3. Age/Sex on Factor Variance

4. Age/Sex on Factor Mean and Variance

Non-Invariance Models of Item-Level Effects wrt Sex and Age



5. Age/Sex on Mean and Loadings



6. Age/Sex on Thresholds and Factor Variance

Application: National Survey of Drug Use in Households (NSDUH)

- Substance Abuse and Mental Health Services Administration (SAMSA) regular data collection
- ~50,000 persons per assessment
- Face-to-face Interviews(!)
- Audio-Computer-Assisted Testing

Map Items to DSM-IV Substance Abuse and Dependence Criteria

- A1 During the past 12 months, did using marijuana or hashish cause you to have serious problems like this either at home, work, or school?
- A2 During the past 12 months, did you regularly use marijuana or hashish and then do something where using marijuana or hashish might have put you in physical danger?
- A3 During the past 12 months, did using marijuana or hashish cause you to do things that repeatedly got you in trouble with the law?
- A4 Did you continue to use marijuana or hashish even though you thought it caused problems with family or friends?

DSM-IV Dependence Criteria

- D1 During the past 12 months, did you need to use more marijuana or hashish than you used to in order to get the effect you wanted?
- D3 Were you able to keep to the limits you set, or did you often use marijuana or hashish more than you intended to?
- D4 During the past 12 months, did you want to or try to cut down or stop using marijuana or hashish?
- D5 During the past 12 months, was there a month or more when you spent a lot of your time getting or using marijuana or hashish?
- D6 This question is about important activities such as working, going to school, taking care of children, doing fun things such as hobbies and sports, and spending time with friends and family.
  - During the past 12 months, did using marijuana or hashish cause you to give up or spend less time doing these types of important activities?
- D7 Did you continue to use marijuana or hashish even though you thought it was causing you to have physical problems?

OpenMx Function for MNI Testing (FIML)

```
#function definition:

nonInvar <- function(data, variableNames, moderatorNames, nFactors,
 testFactorMeans=NULL, testFactorVariances=NULL, testLoadings=NULL,
 testItemMeans=NULL, testItemVariances=NULL, useDeviations=T)

#example use:

vars <-  c( 'ALCA1', 'ALCA2', 'ALCA3', 'ALCA4', 'ALCD1', 'ALCD3', 'ALCD4',
 'ALCD5', 'ALCD6', 'ALCD7')
mods <- c('sex')

nsduhALC <- nonInvar(nsduh[,c(vars,mods)] , vars, mods, nFactors=1)
nsduhALCFM <- nonInvar(nsduh[,c(vars,mods)] , vars, mods,
 nFactors=1,testFactorMeans=c(T) )
```

Test of Factor Loading Invariance: Cannabis in NSDUH

Model	Npar	Comparison Model	Likelihood	df	AIC	diffLL	diffdf	p	Age Effect	Sex Effect
1. No Covariates (1f model)	20	NA	62514	78204	-93894	NA	NA	NA	NA	NA
2. Age/Sex on Factor Mean	22	1	62009	78202	-94395	505	2	<.0001	-3.85567	-0.09112
3. Age/Sex on Factor Variance	22	1	62480	78202	-93924	33.84	2	<.0001	0.70624	0.40302
4. Age/Sex on Factor Mean and Variance	24	1	61893	78200	-94507	620.4	4	<.0001		
5. Age/Sex on Mean and Loadings	42	4	61801	78182	-94563	92.34	18	<.0001		
6. Age/Sex on Thresholds and Factor Variance	42	4	61802	78182	-94562	91.2	18	<.0001		

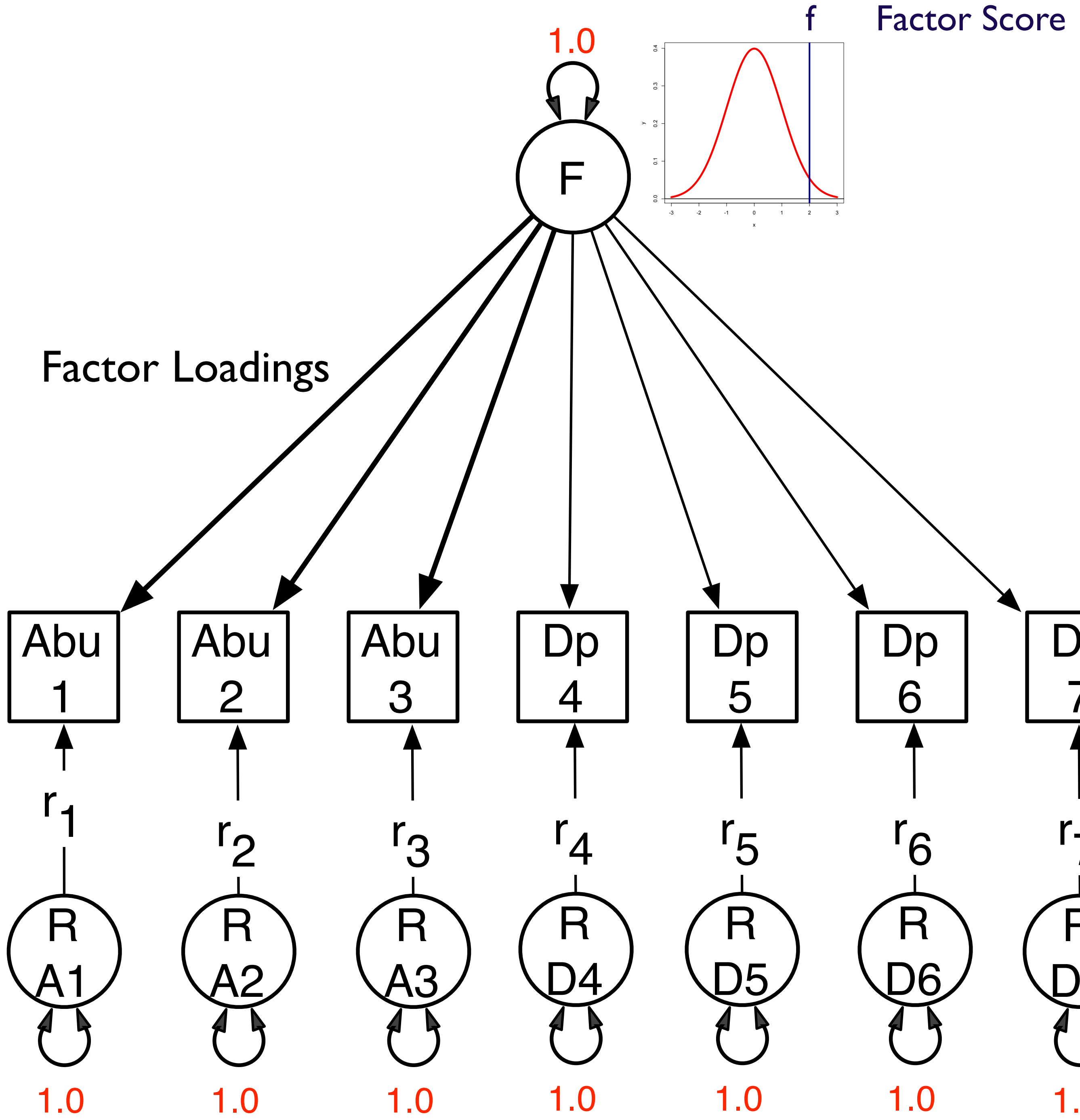
Test of Item Mean Invariance: Cannabis in NSDUH

Model	Npar	Comparison Model	Likelihood	df	AIC	diffLL	diffdf	p	Age Effect	Sex Effect
1. No Covariates (1f model)	20	NA	62514	78204	-93894	NA	NA	NA	NA	NA
2. Age/Sex on Factor Mean	22	1	62009	78202	-94395	505	2	<.0001	-3.85567	-0.09112
3. Age/Sex on Factor Variance	22	1	62480	78202	-93924	33.84	2	<.0001	0.70624	0.40302
4. Age/Sex on Factor Mean and Variance	24	1	61893	78200	-94507	620.4	4	<.0001		
5. Age/Sex on Mean and Loadings	42	4	61801	78182	-94563	92.34	18	<.0001		
6. Age/Sex on Thresholds and Factor Variance	42	4	61802	78182	-94562	91.2	18	<.0001		

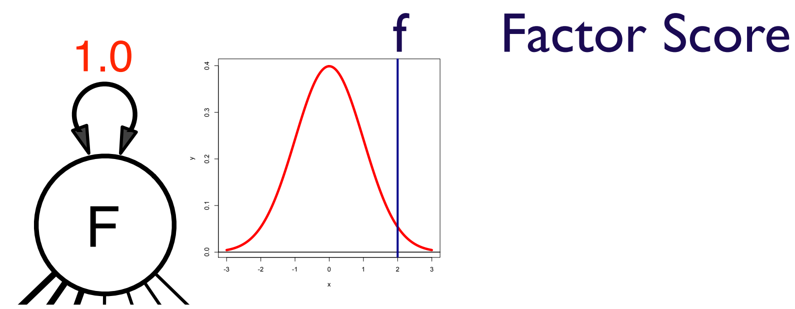
-2lnL Likelihood Ratio Test Statistics: Marijuana Item Means & Factor Loadings

Item		Sex: IM_sex	Sex: FL_sex	Age: IM_age	Age: FL_age
MRJA1	Work	2.51	0.46	18.48	2.7
MRJA2	Danger	0.36	1.04	1.64	0.31
MRJA3	Law	26.06	−0.07	7.26	2.47
MRJA4	Friends	0.3	1.47	13.05	0.31
MRJD1	Tol	1.5	−0.46	10.57	10.3
MRJD3	>Intend	−0.06	0.23	23.26	3.88
MRJD4	TryCut	10.41	10.79	6.33	−0.06
MRJD5	TimeGet	25.49	25.05	0.42	0.05
MRJD6	TimeOther<	0.32	−0.12	13.21	18.91
MRJD7	PhysProb	30.86	24.59	22.85	7.66

Estimating Factor Scores

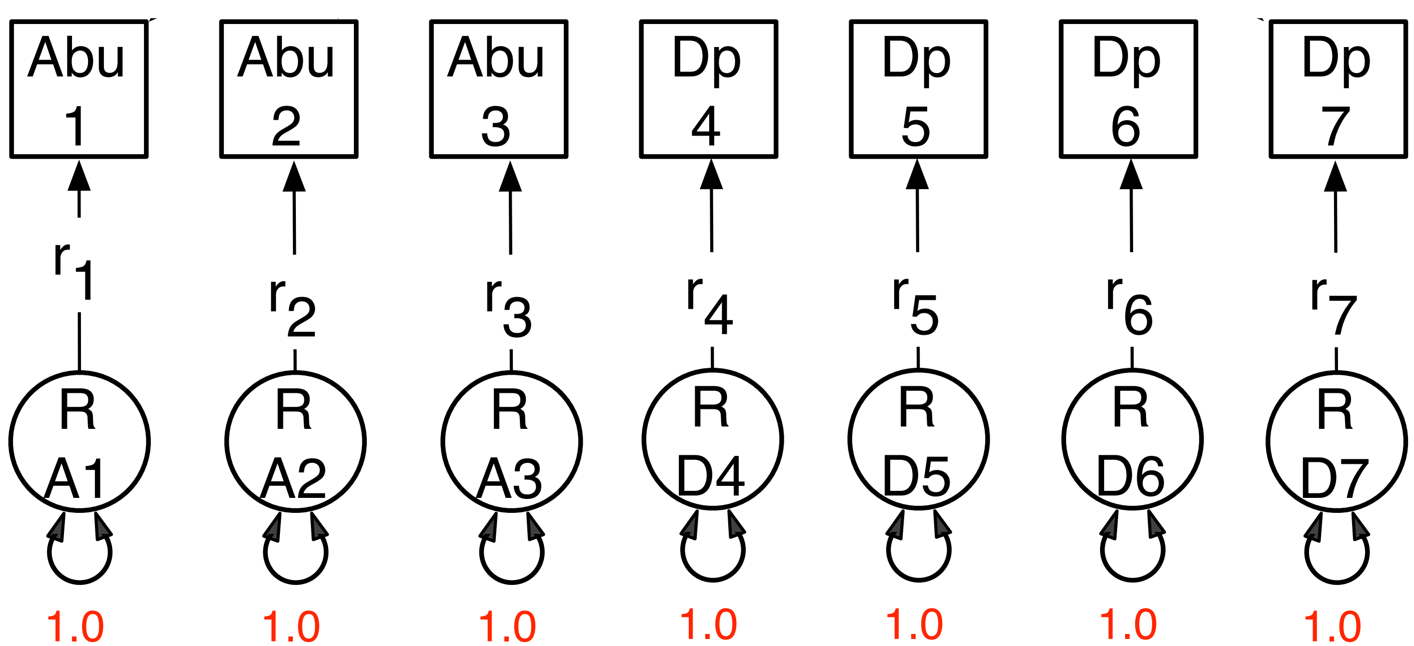


ML Estimation of Factor Scores

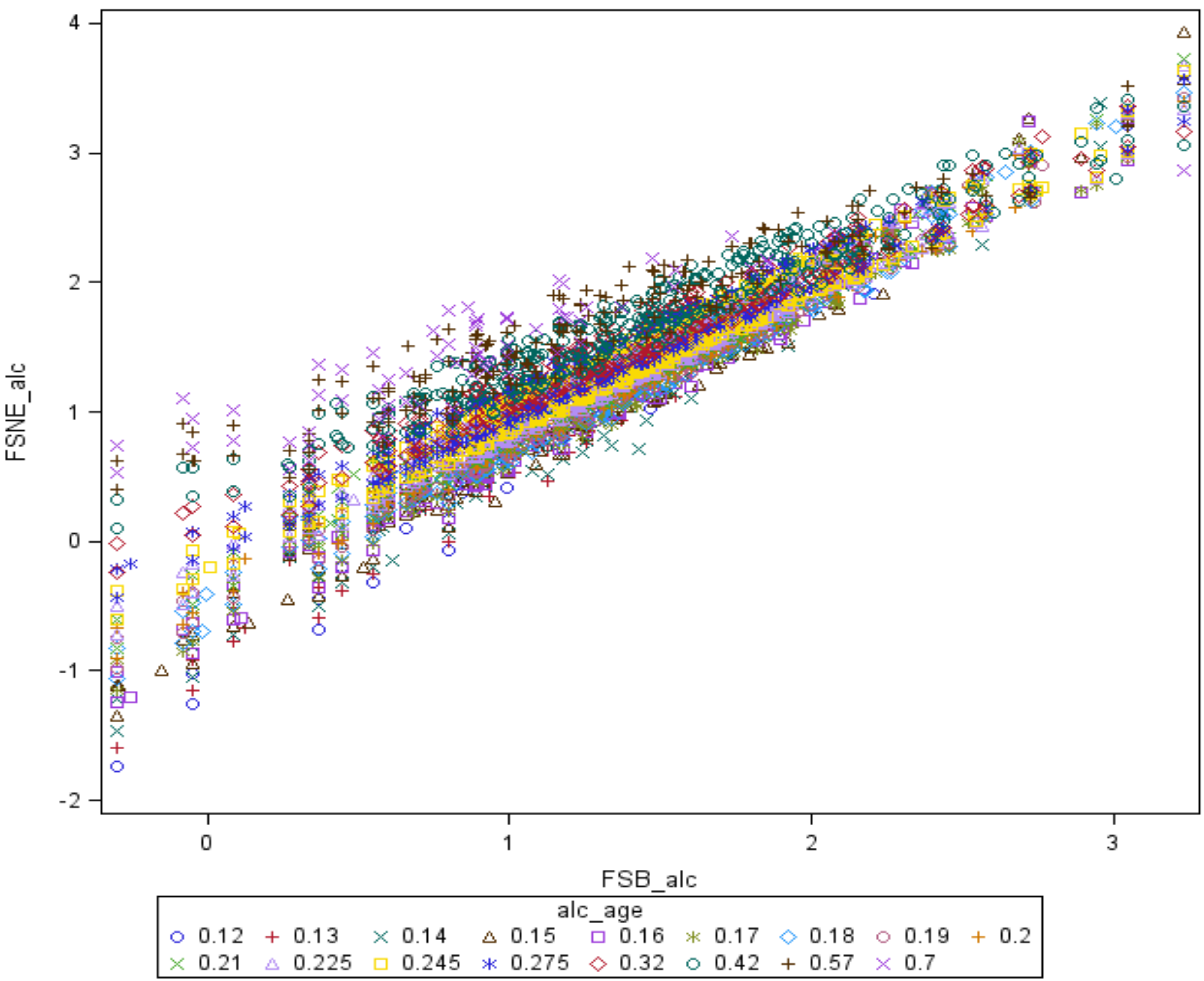


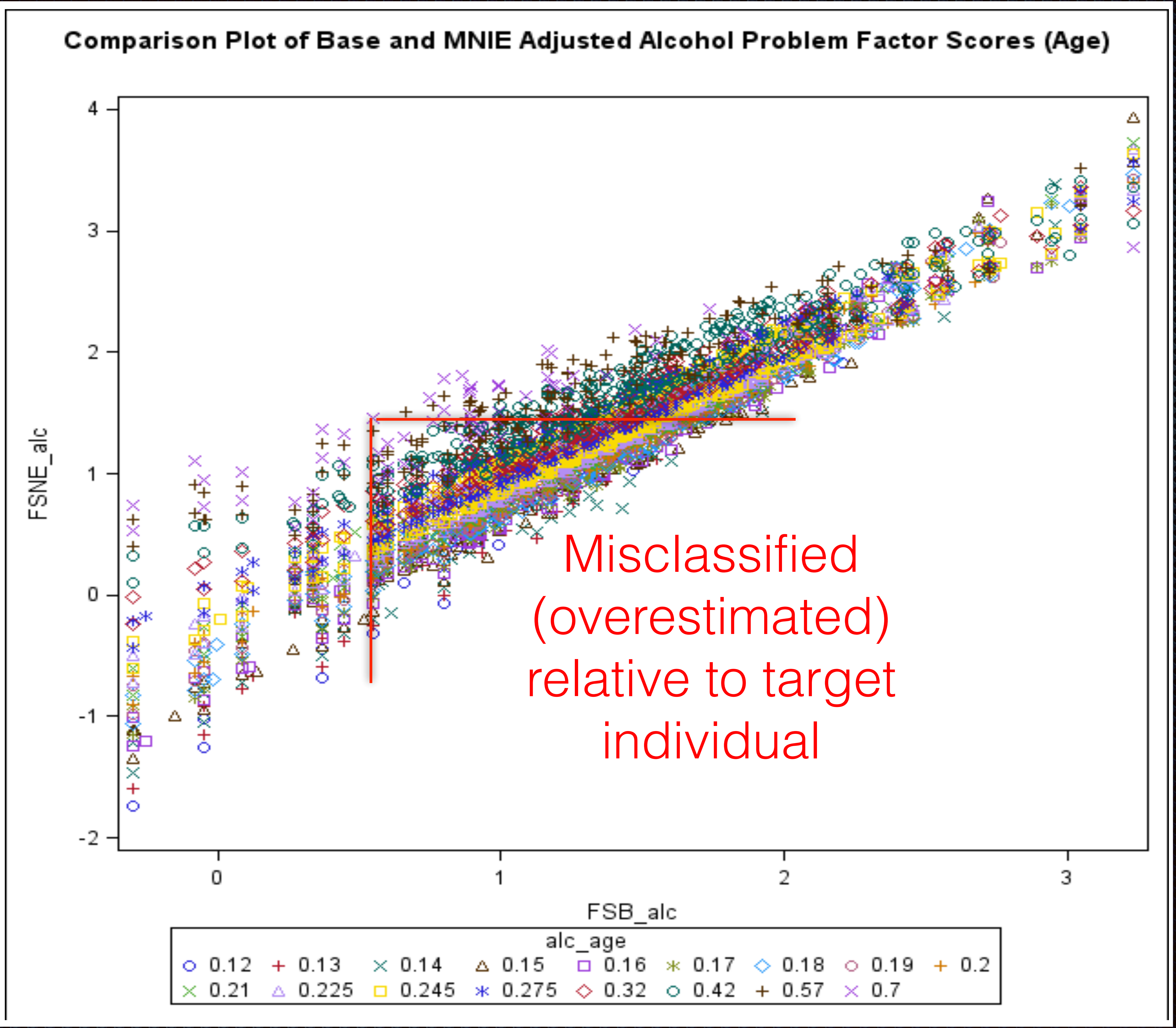
Factor Score * Likelihood of items conditional on factor score

Items independent conditional on factor score:
Means and variances change according to size of factor loadings

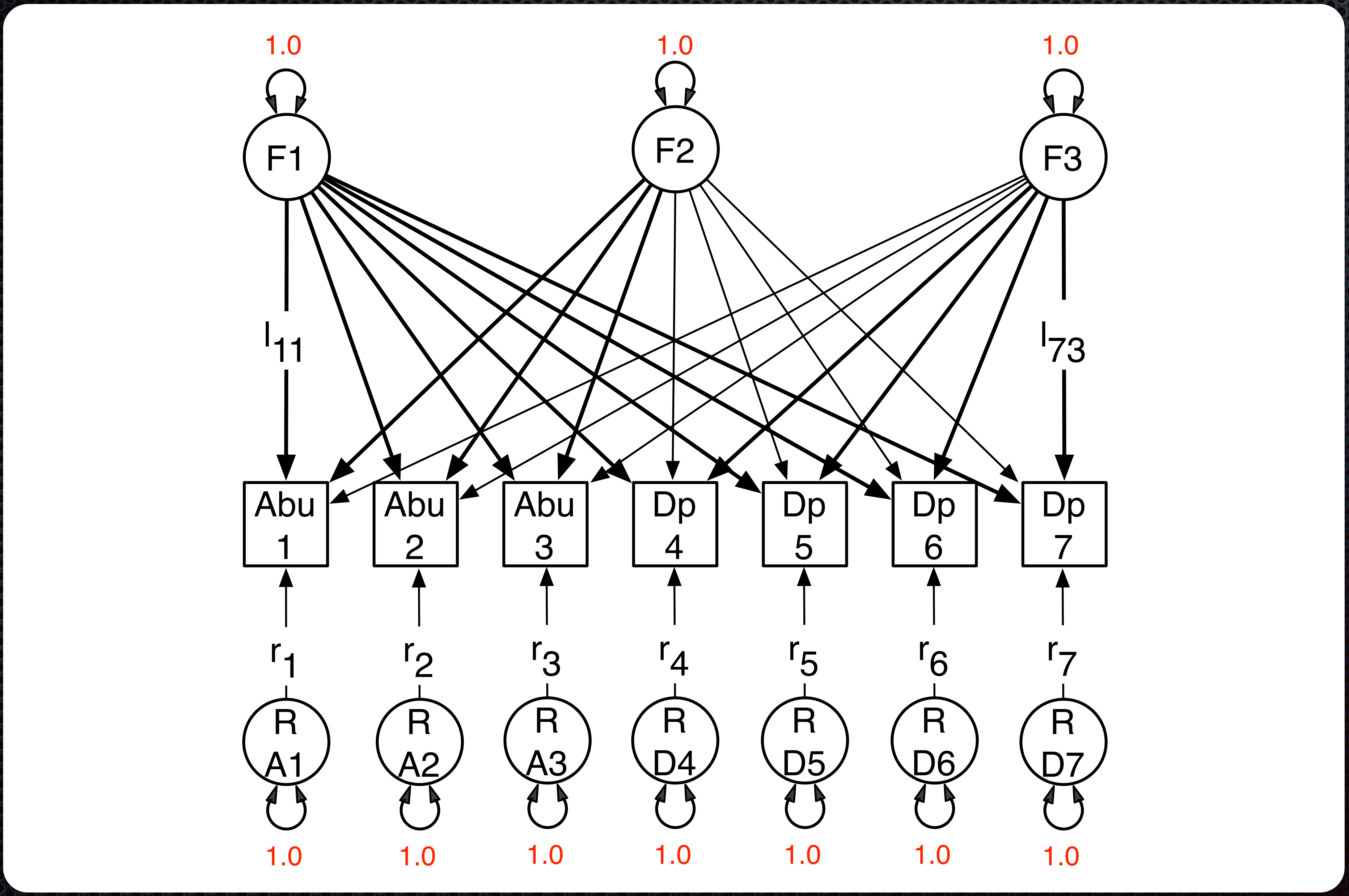


Comparison Plot of Base and MNIE Adjusted Alcohol Problem Factor Scores (Age)

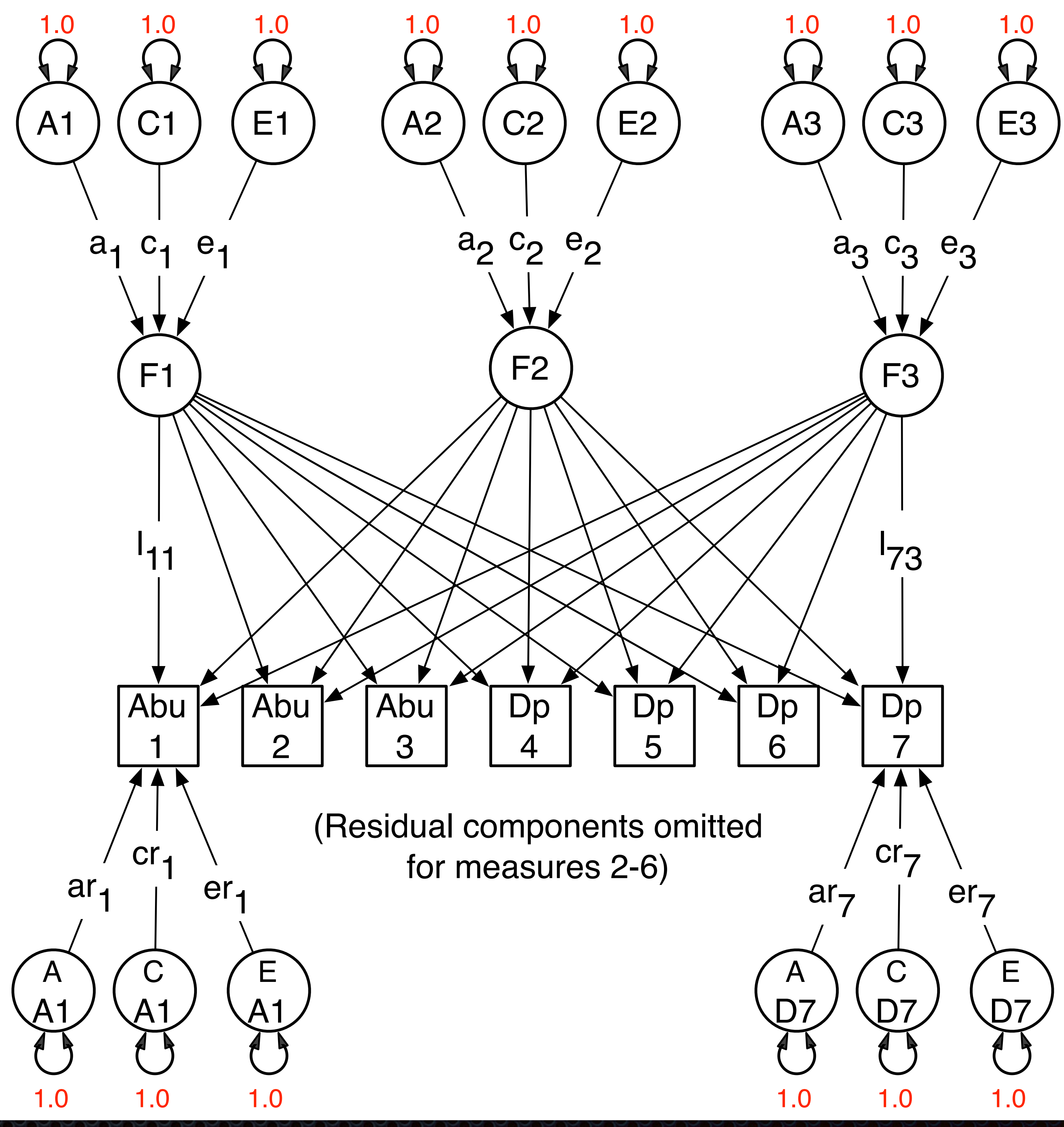



Comparison Plot of Base and MNIE Adjusted Alcohol Problem Factor Scores (Age)
FSNE_alc
FSB_alc
alc_age
0.12
0.13
0.14
0.15
0.16
0.17
0.18
0.19
0.2
0.21
0.225
0.245
0.275
0.32
0.42
0.57
0.7
Misclassified (overestimated) relative to target individual

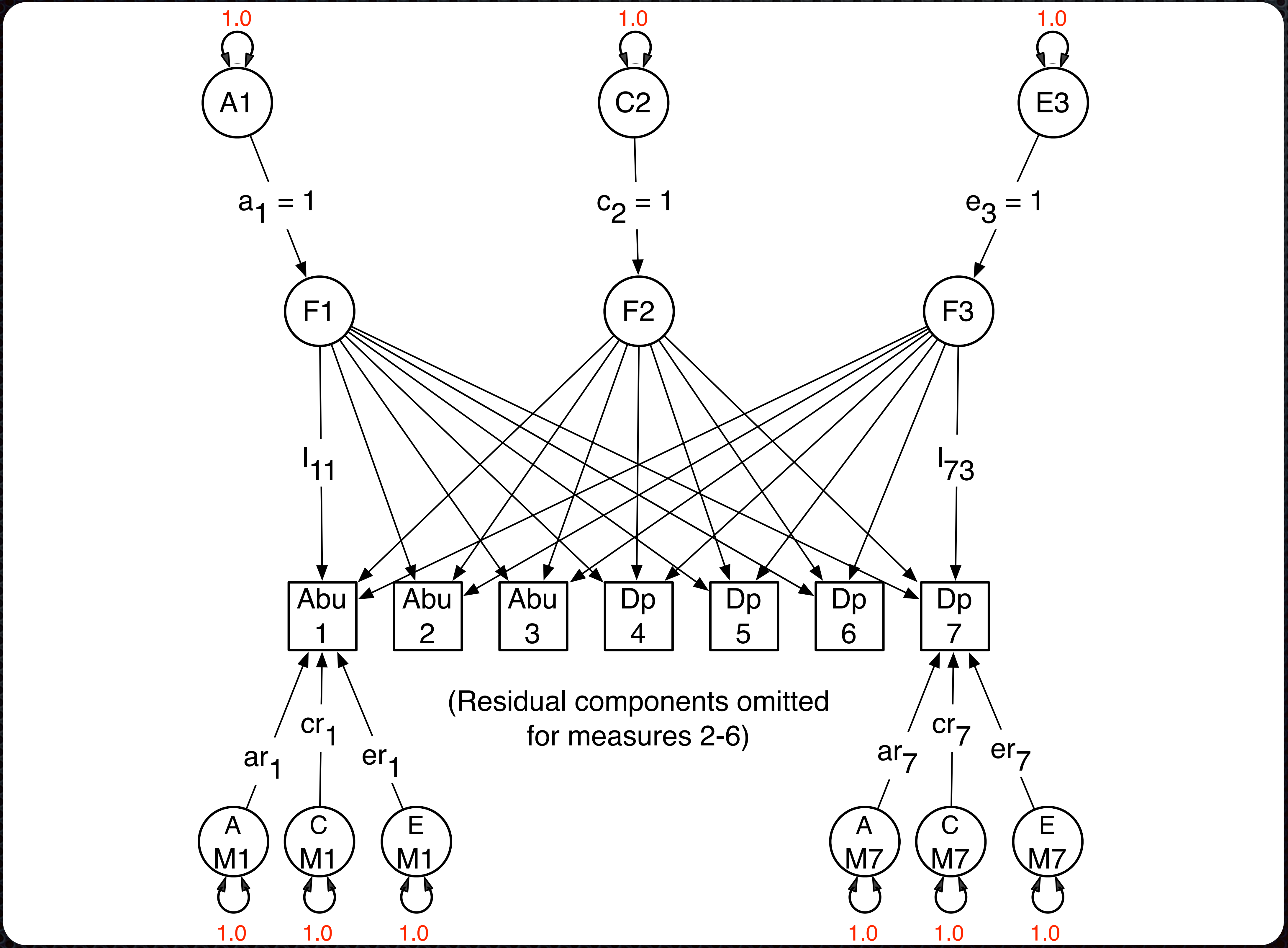
Multiple Factor Model Beware Rotation



Genetic and Environmental Factors: Common Pathway Model



Genetic and Environmental Factors: Independent Pathway Model



Application

Molecular Psychiatry (2006) 11, 752–762

www.nature.com/mp

ORIGINAL ARTICLE

Association between glutamic acid decarboxylase genes and anxiety disorders, major depression, and neuroticism

JM Hettema, SS An, MC Neale, J Bukszar, EJCG van den Oord, KS Kendler and X Chen

Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA

- Used genetic factor scores to select extreme groups
- Found significant association
- Step right up - everyone’s a winner

Factor Score Notes

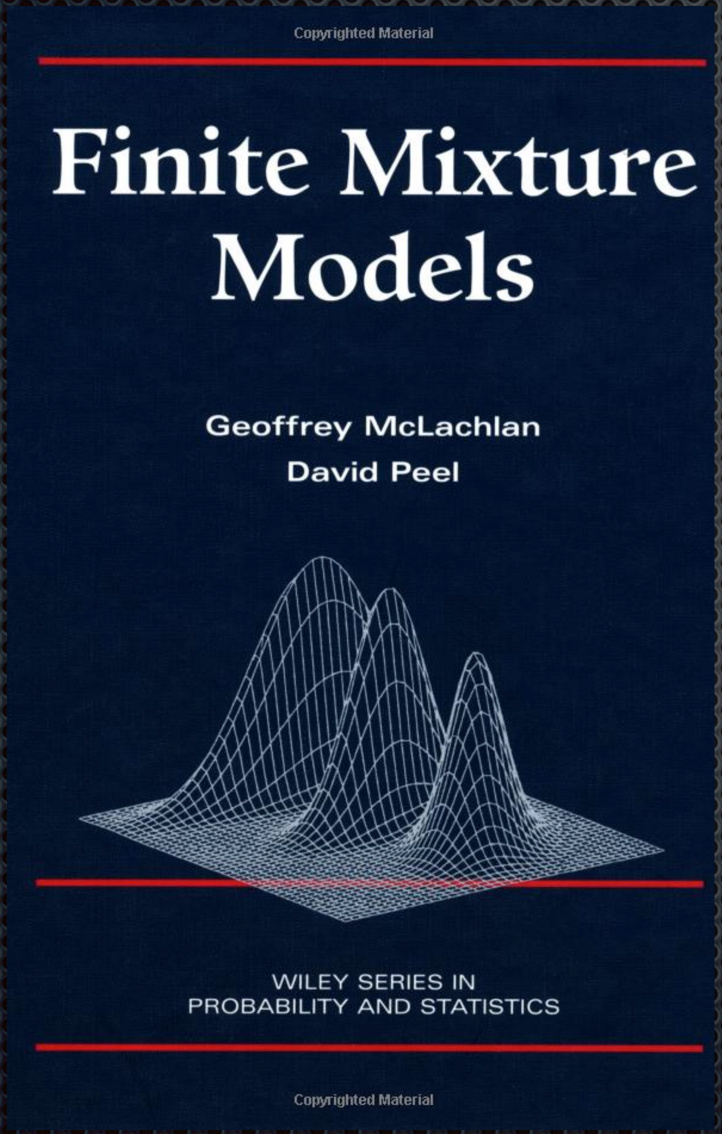
- Factor scores do not all have same error variance
- Factor scores of A, C & E components may correlate highly
- Latent trait may be non-normal (Schmitt et al 2006 Multiv Behav Res)
- Factor loadings (precision) may vary across the distribution and give spurious GxE results
- Variation may be discrete not continuous

What if Variation is Discrete?

- Latent Class and Latent Profile Models
- Factor Mixture Models
- Latent Growth Curve Mixture Models
- Regime Switching

Mixture Distributions

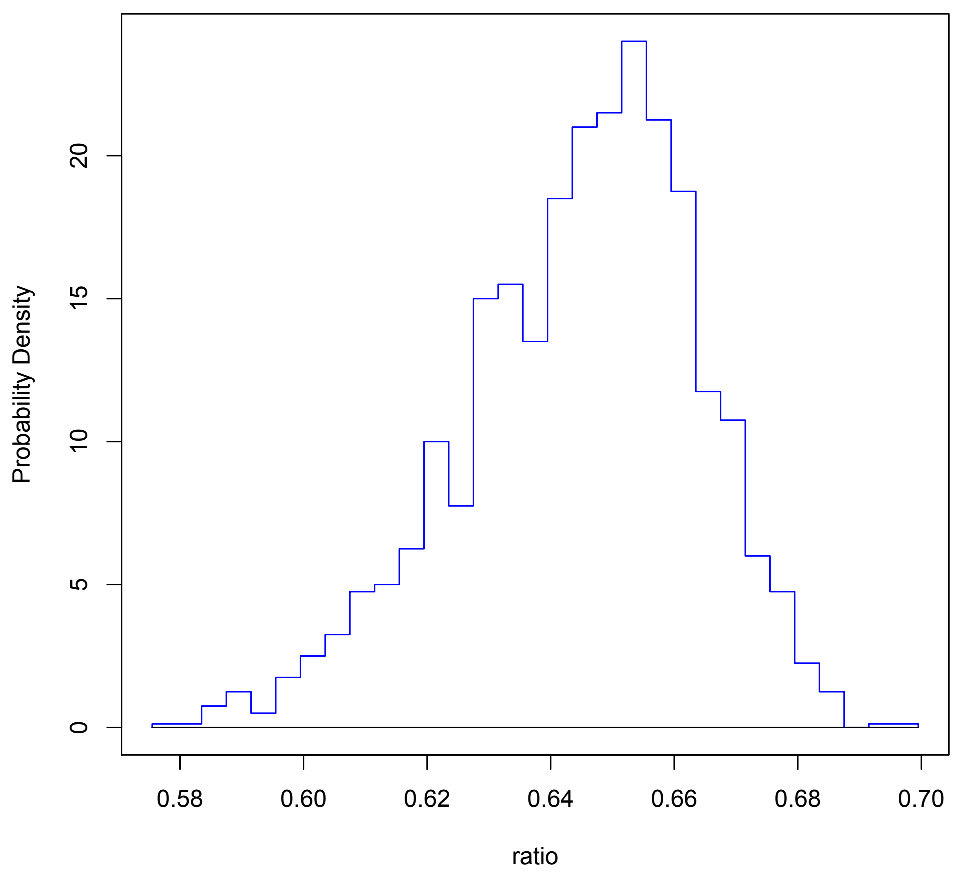
Pearson, K. (1894). Contributions to the mathematical theory of evolution. II. skew variation in homogeneous material. Philosophical Transactions of the Royal Society of London A, 186, 343-414.

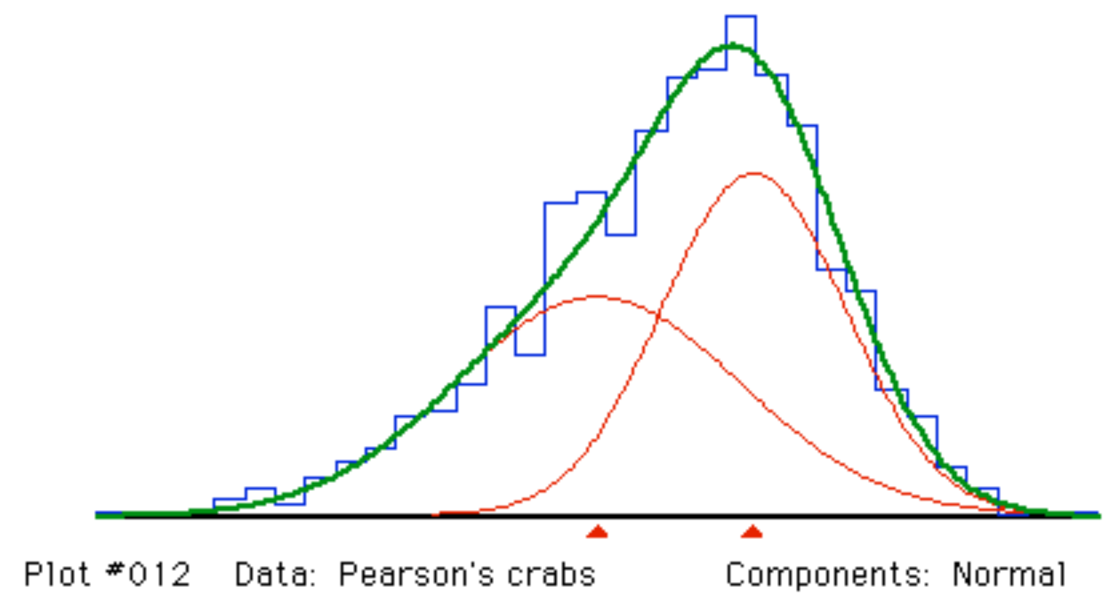


- Skewness in a set of measurements of the ratio of forehead to body length of crabs

- Two species or one?

Data & Model

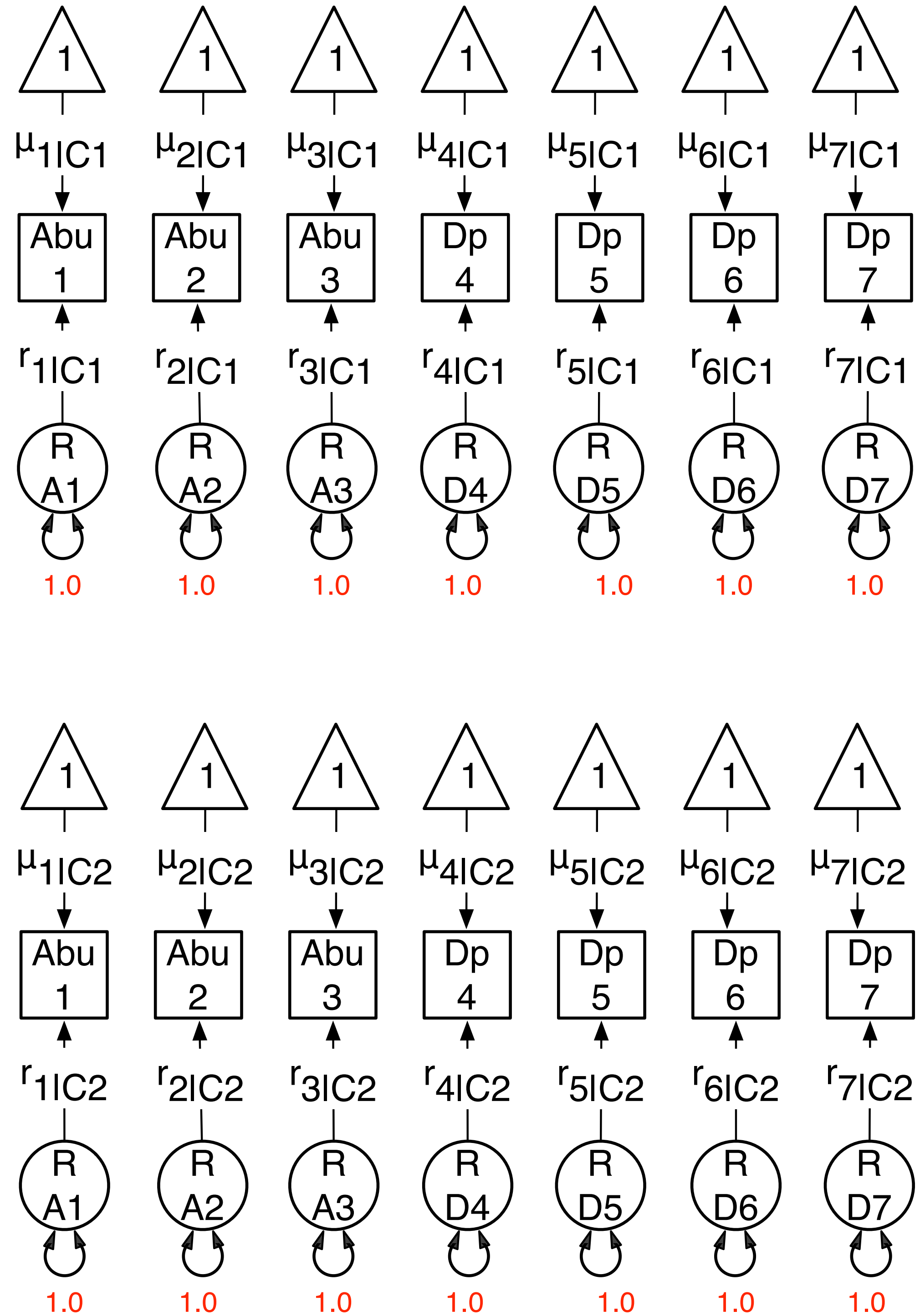

Probability Density
0
5
10
15
20
0.58
0.60
0.62
0.64
0.66
0.68
0.70
ratio


Plot #012 Data: Pearson's crabs Components: Normal

Latent Class (Subgroup)

Class 1 probability *p*

Class 2 probability (1-*p*)



Conditionally Independent?!

Expensive!

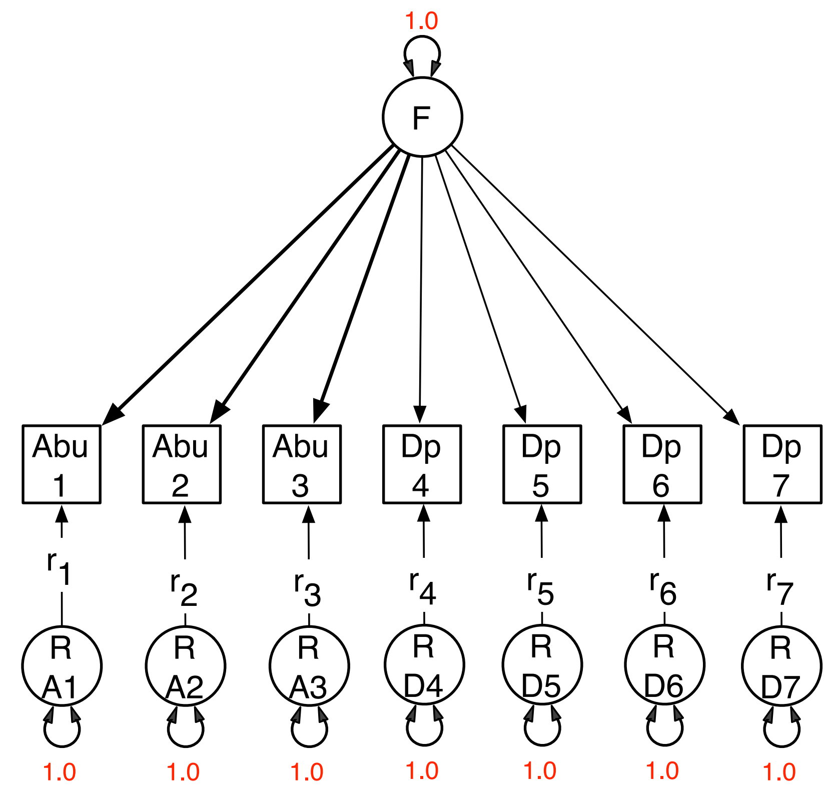
Published in final edited form as:
Int J Methods Psychiatr Res. 2010 June ; 19(2): 63–73. doi:10.1002/mpr.301.

Searching For Valid Psychiatric Phenotypes: Discrete Latent Variable Models

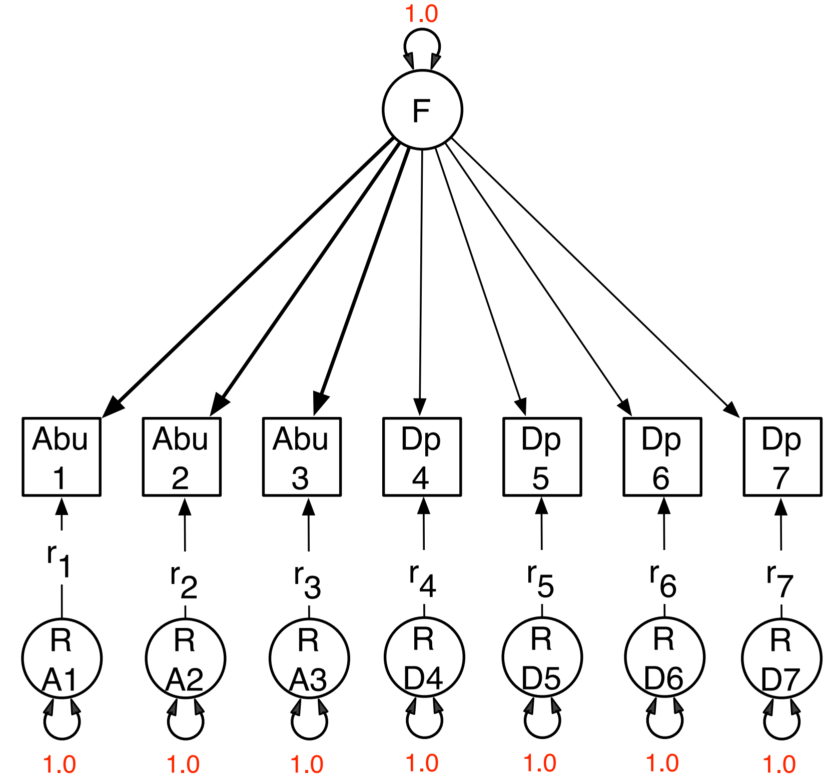
Jeannie-Marie S. Leoutsakos, PhD, MHS[1], Peter P. Zandi, PhD, MHS[2], Karen Bandeen-Roche, PhD[3], and Constantine G. Lyketsos, MD, MHS[1,2]

Factor Mixture Model

Class 1
probability
p



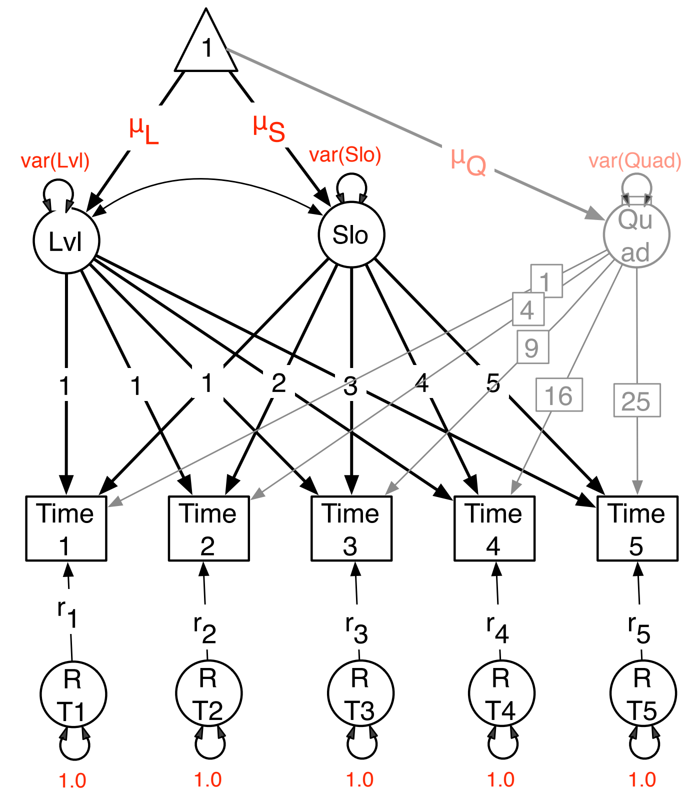
Class 2
probability
(1-*p*)



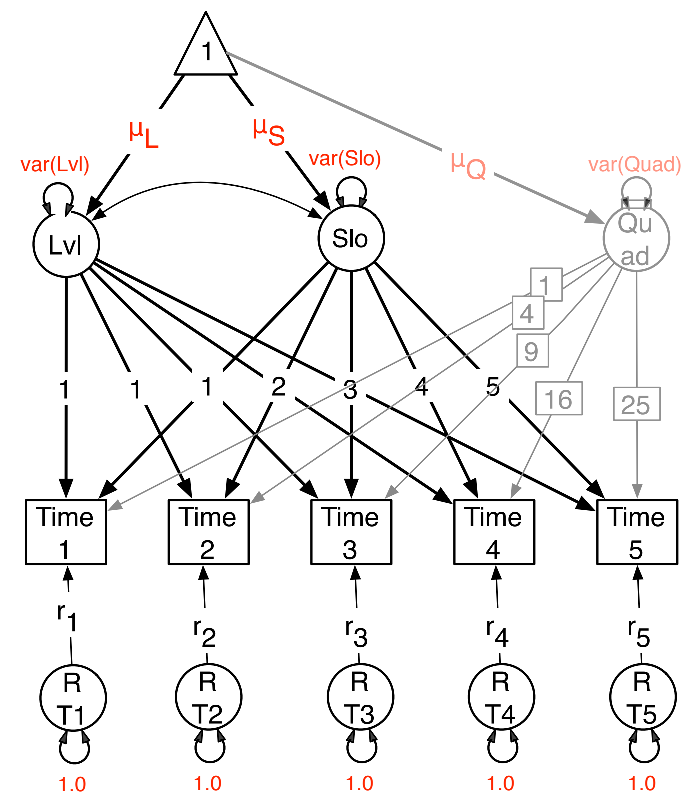
Very
Expensive!

Growth Curve Mixture

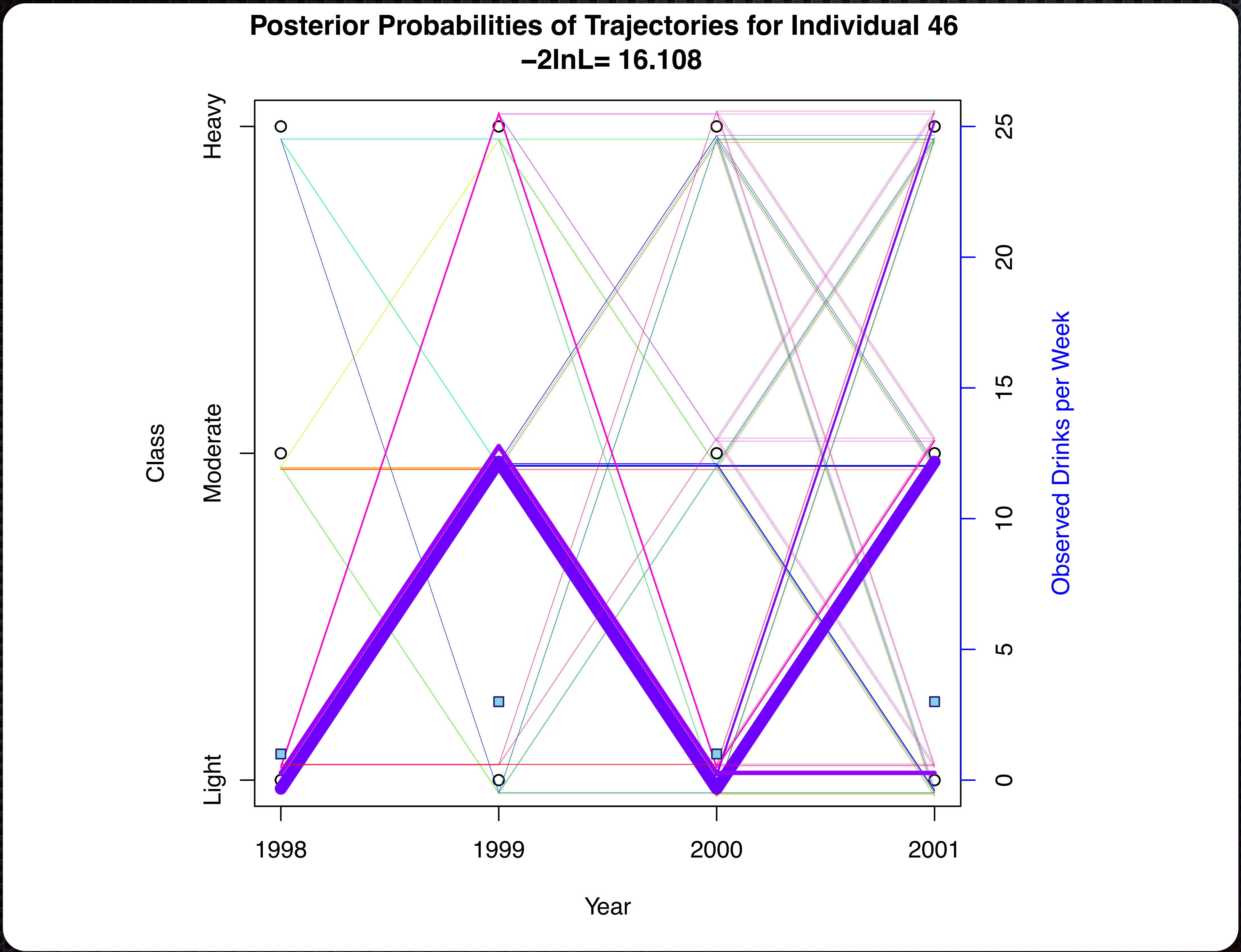
Class 1
probability
p



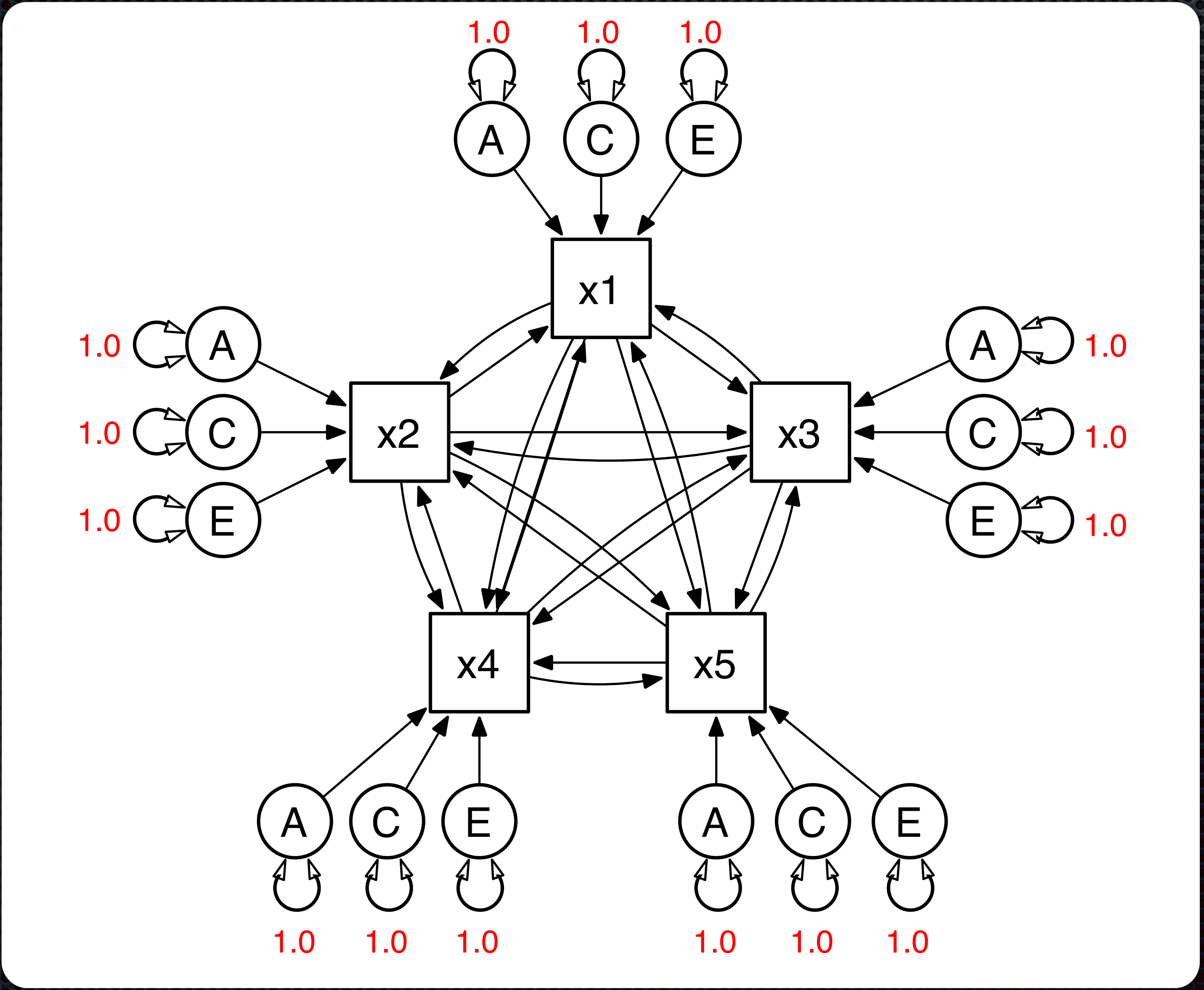
Class 2
probability
(1-p)



Regime Switching Model



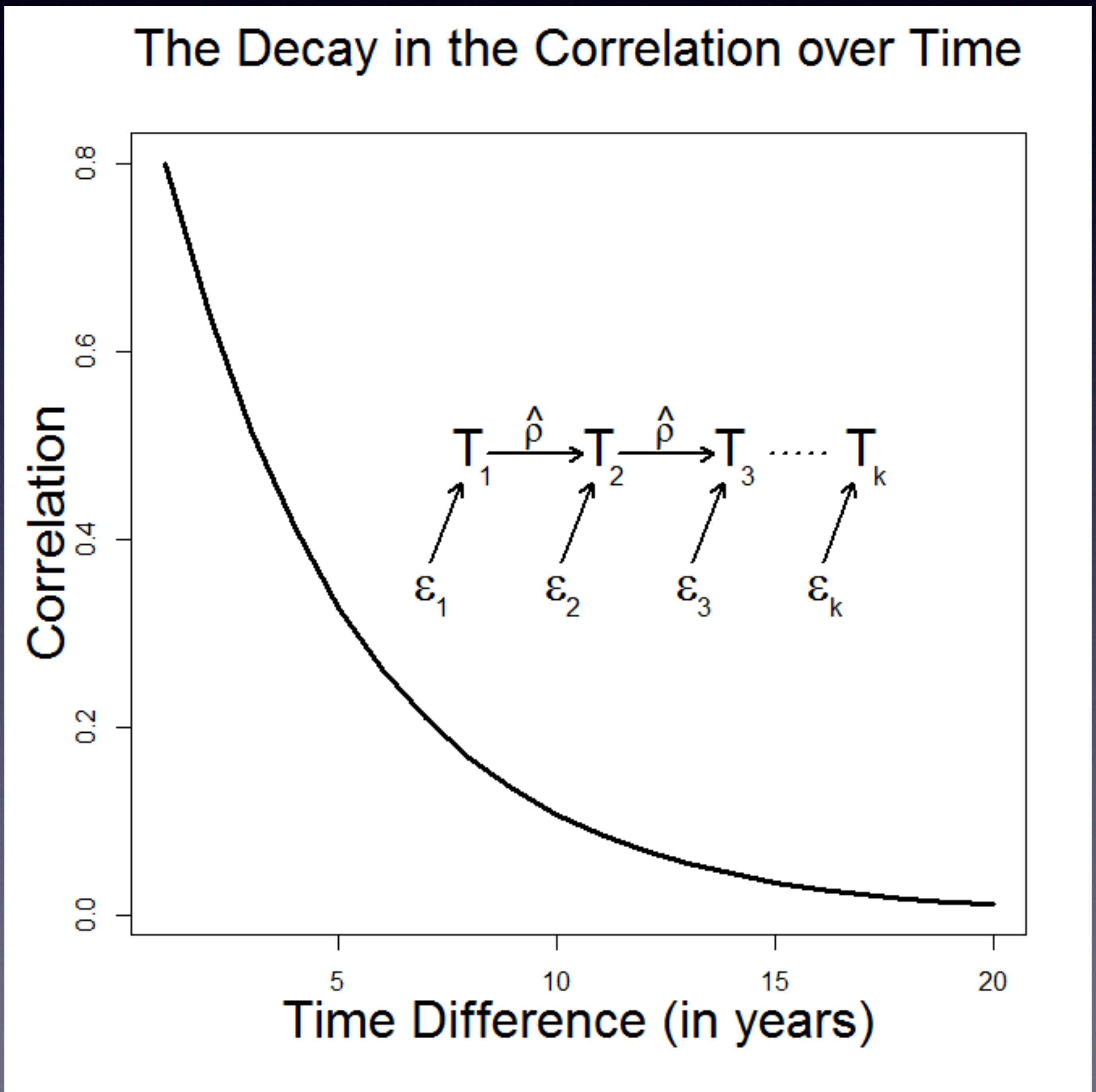
No Latent Variables Model (Mutualism)



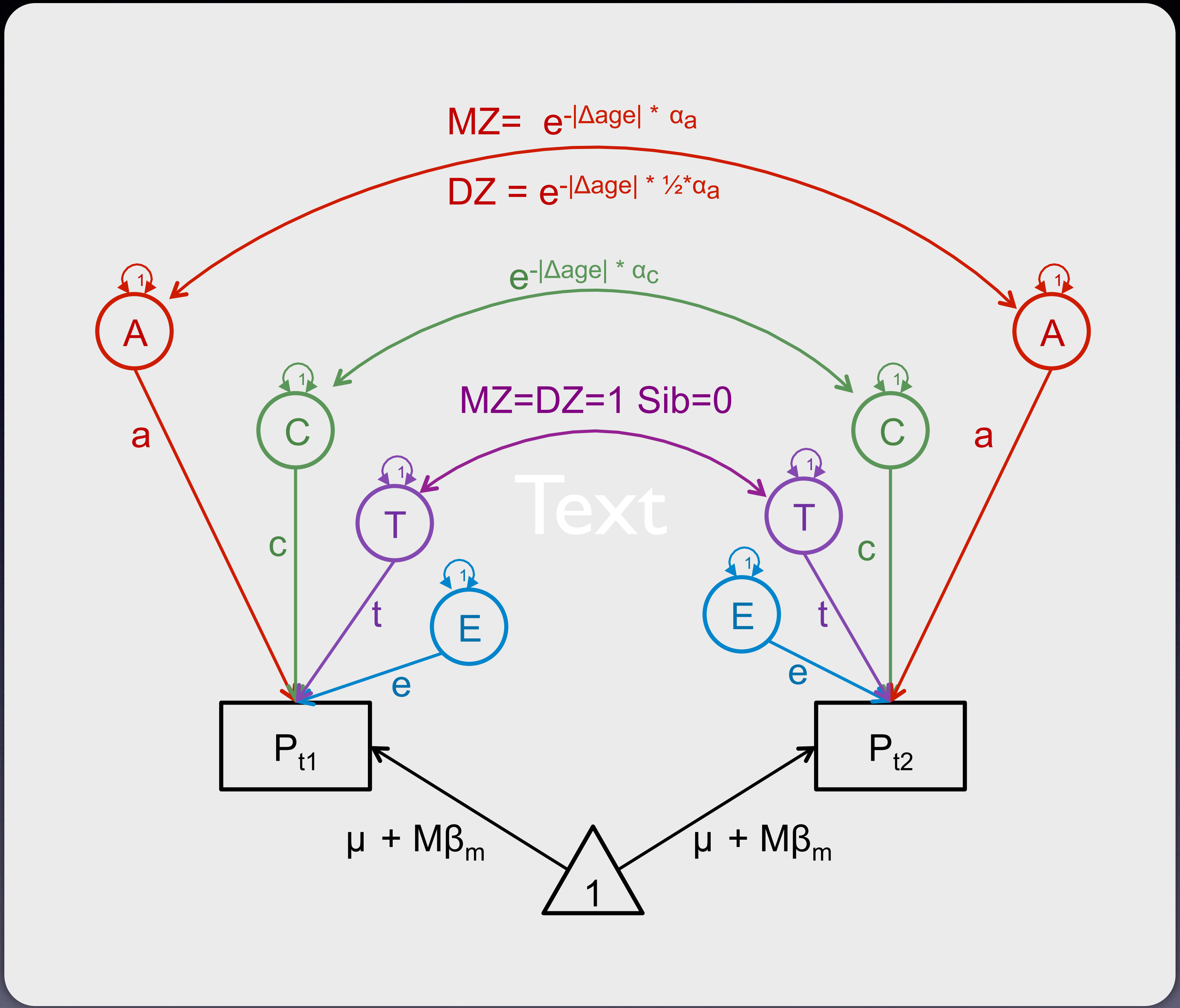
Genetic Heterogeneity

- Genetic factors change during development
- Height
- Neuroticism
- Detection

Different age, different genes?

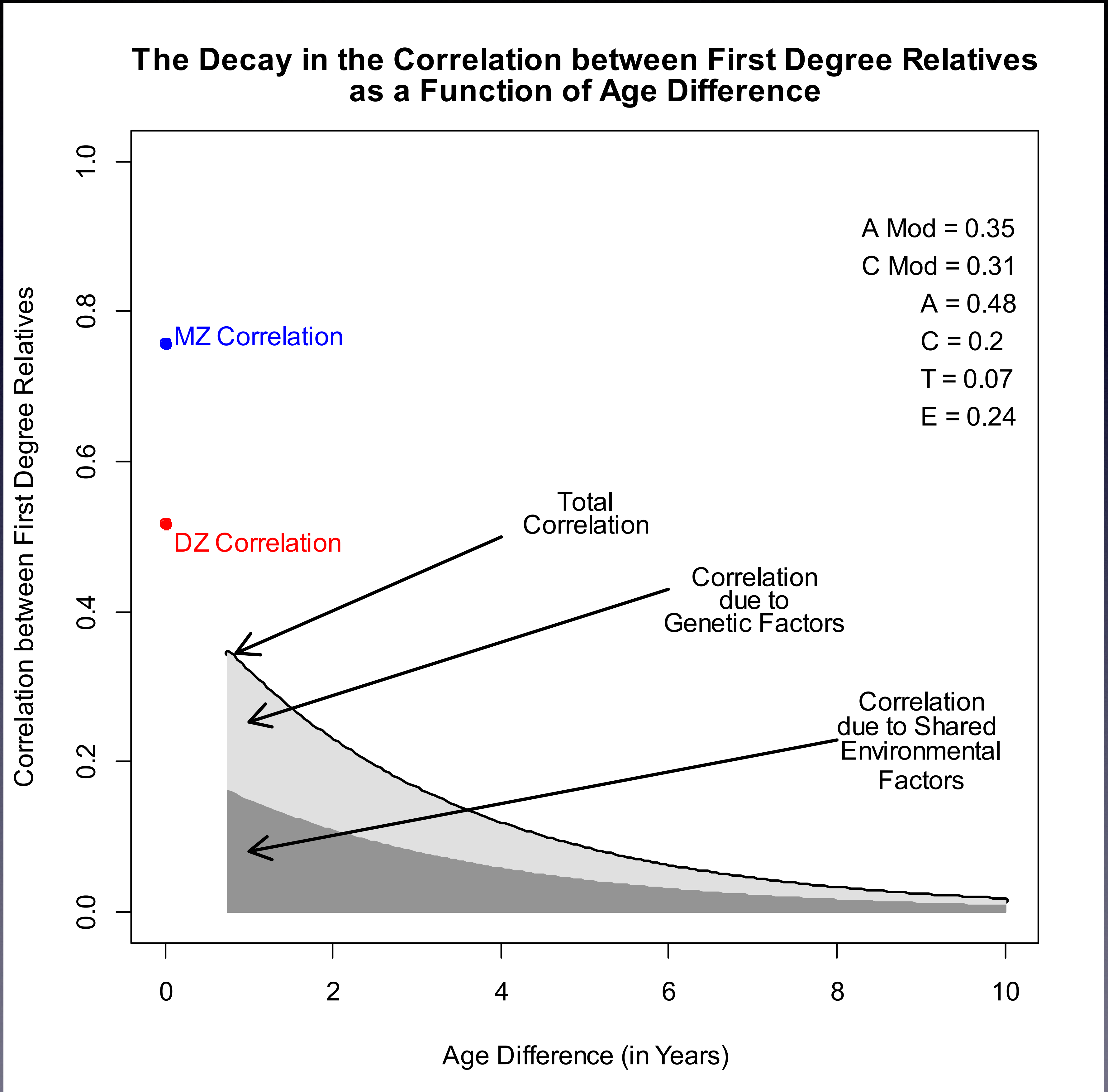


Verhulst, Eaves & Neale



$$Cov = Acov * e^{-|\Delta age| * \alpha_a} + Ccov * e^{-|\Delta age| * \alpha_c} + Tcov$$

Verhulst, Eaves & Neale



Care with Ascertainment

- Factor Analysis in Cases
- Latent Class Analysis in Cases
- Selection for Case Status
- Selection of (Super) Controls
- All the above can give very different results

Summary

- Measurement of complex traits is complex
- Measurement invariance desirable
- ML factor scores good start
- Mixture distribution models should be tested
- Choose your study participants carefully
- Analyze what you measure, and measure well what you analyze