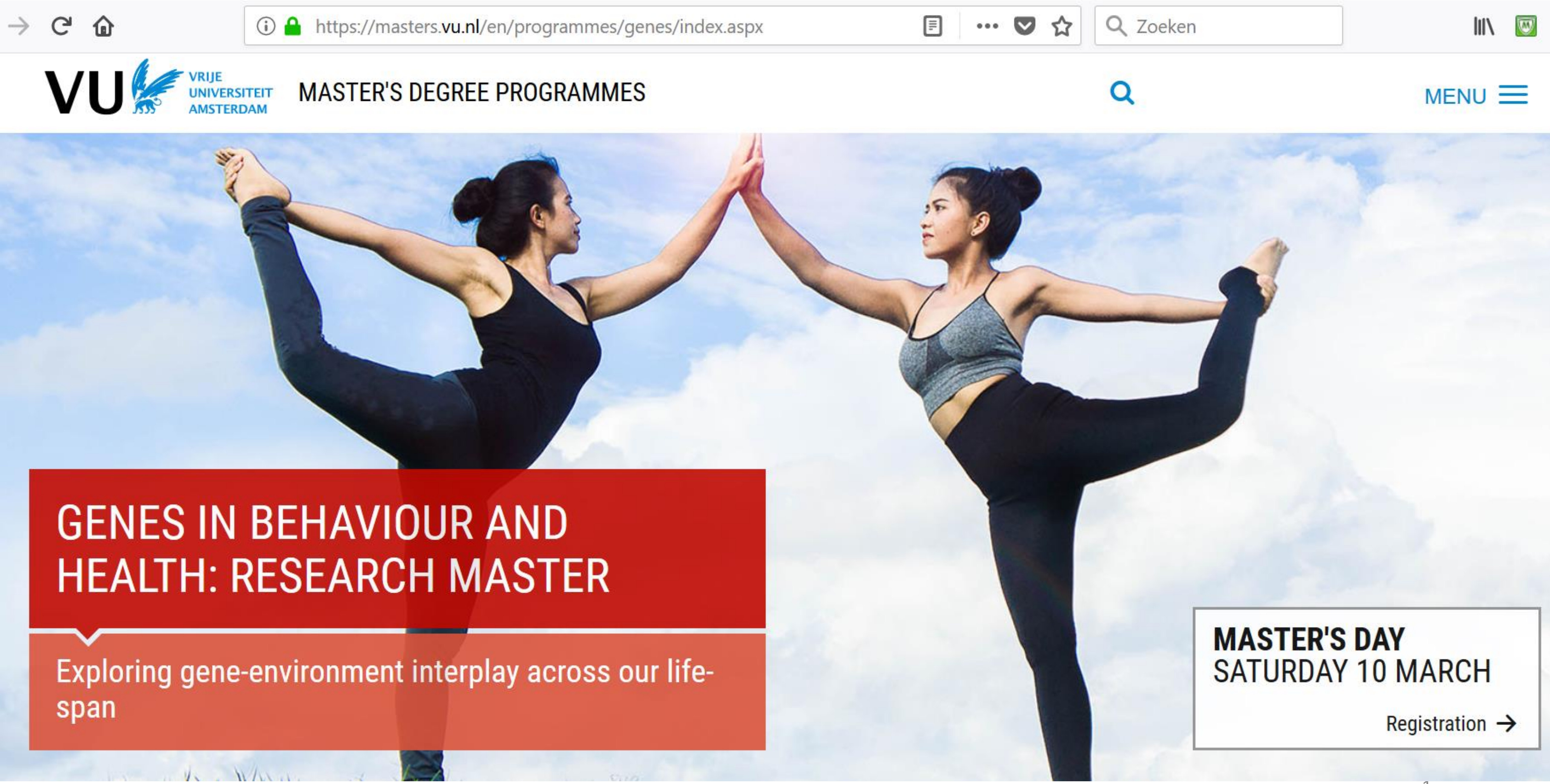

https://masters.vu.nl/en/programmes/genes/index.aspx
Zoeken
VU VRIJE UNIVERSITEIT AMSTERDAM
MASTER'S DEGREE PROGRAMMES
MENU
GENES IN BEHAVIOUR AND HEALTH: RESEARCH MASTER
Exploring gene-environment interplay across our life-span
MASTER'S DAY
SATURDAY 10 MARCH
Registration

Phenotypic factor analysis

Conor V. Dolan & Michel Nivard
VU, Amsterdam

Boulder Workshop - March 2018

Phenotypic factor analysis

A statistical technique to investigate the dimensionality of correlated variables in terms of common *latent* variables (a.k.a. common factors).

Applications in psychometrics (measurement), biometrical genetics, important in differential psychology (IQ, personality).

Psychometric perspective (not the only one): FA as a measurement model.

Questionnaire items are formulated to measure
a latent – unobservable – trait, such as

Perceptual speed
Working memory
Verbal intelligence
Depression
Disinhibition
Extroversion

latent variables, not observable, hypothetical
latent, unobservable....
so how can we measure these?

measure these by considering observable variables – questionnaire items – that are dependent on these latent variables. items as **indicators**.

8 depression items

1. Little interest or pleasure in doing things?
2. Feeling down, depressed, or hopeless?
3. Trouble falling or staying asleep, or sleeping too much?
4. Feeling tired or having little energy?
5. Feeling bad about yourself - or that you are a failure or have let yourself or your family down?
6. Trouble concentrating on things, such as reading the newspaper or watching television?
7. Moving or speaking so slowly that other people could have noticed?
8. Thoughts that you would be better off dead, or of hurting yourself in some way?

A psychometric analysis:

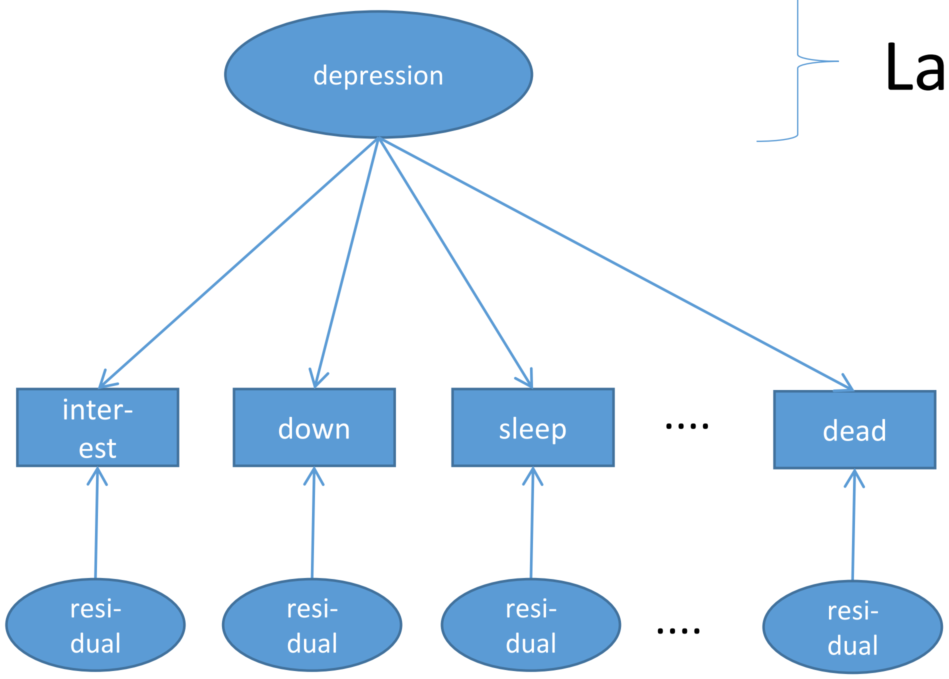
Investigate the dimensionality of the item responses in terms of substantive latent variables.

A psychometric causal perspective:

An <u>implicit causal</u> hypothesis: the latent variable ("depression") causes the item response.

Your theoretical point of departure!

what we expect (theory)

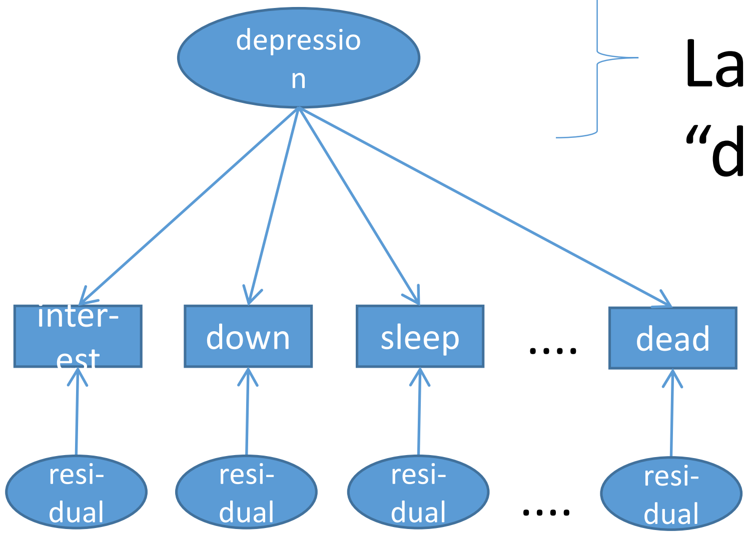


Latent variable

observed variables
(indicators)

The items share a common cause (depression):
depression is a source of shared variance in the items,
gives rise to covariance / correlation among the item scores.

what we expect (theory)



Latent variable
“depression”

what we observe

correlation matrix of 8 items scores (general pop sample N=1000).

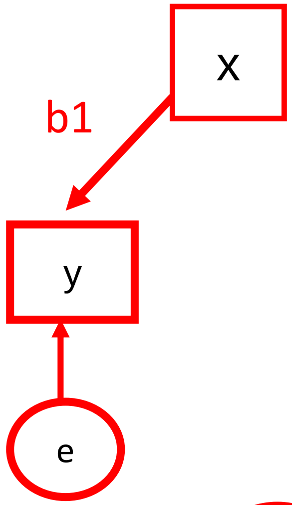
1.00
0.24 1.00
0.20 0.19 1.00
0.26 0.20 0.20 1.00
0.25 0.18 0.15 0.26 1.00
0.23 0.19 0.17 0.24 0.22 1.00
0.16 0.16 0.13 0.22 0.14 0.19 1.00
0.16 0.09 0.17 0.16 0.18 0.18 0.16 1.00

Is the observed correlation matrix (right) compatible with the model (left?).

Single common factor model: A set of linear regression equations

$$y_i = b0 + b1*X_i + e_i$$

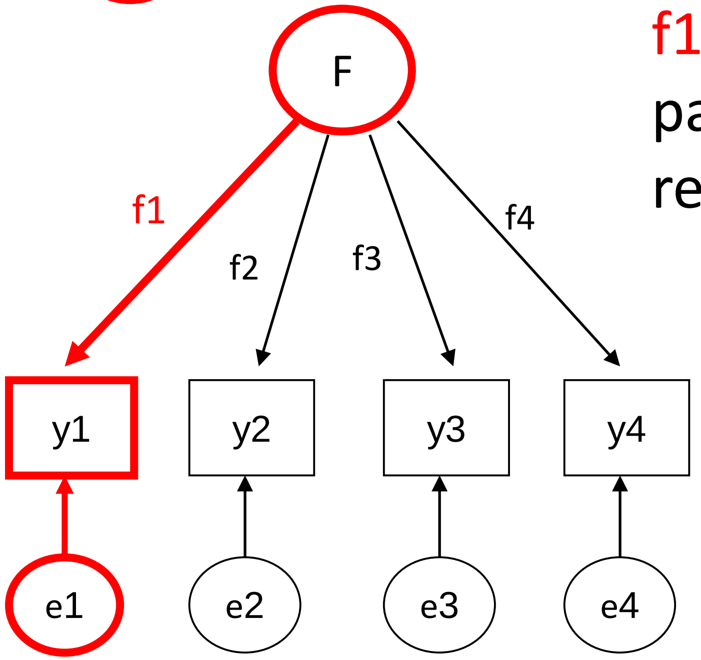
intercept regression coefficients



b1 is a regression coefficient (slope parameter)

$$\left.\begin{aligned} y1_i &= t1 + f1*F_i + e1_i \\ y2_i &= t2 + f2*F_i + e2_i \\ y3_i &= t3 + f3*F_i + e3_i \\ y4_i &= t4 + f4*F_i + e4_i \end{aligned}\right\}$$

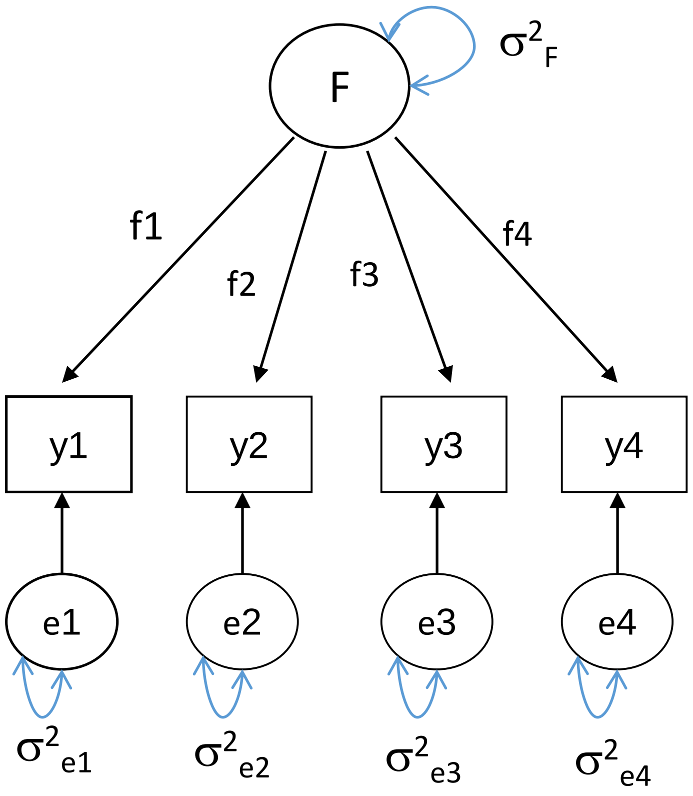
intercepts factor loadings



f1 is a factor loading path diagram: linear regression.

But how does this work if the common factor (**the independent variable, F**) is not observed? How can we estimates the regression coefficients (factor loadings)?

$$
\left.
\begin{aligned}
y1_i - t1 &= f1*F_i + e1_i \\
y2_i - t2 &= f2*F_i + e2_i \\
y3_i - t3 &= f3*F_i + e3_i \\
y4_i - t4 &= f4*F_i + e4_i
\end{aligned}
\right\}
$$

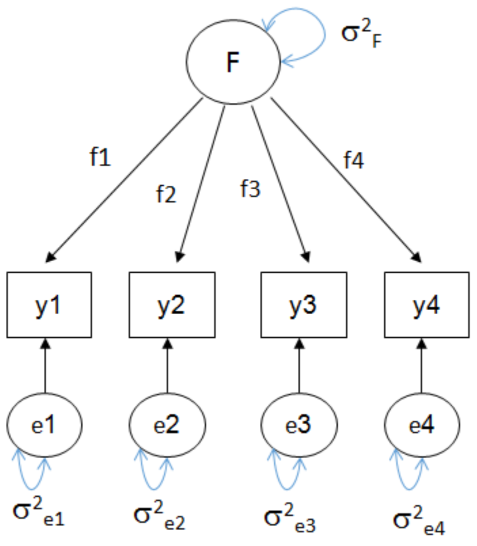


Consider the implied covariance matrix – the covariance matrix expressed in terms of the parameters in the model

Implied covariance matrix among y1 to y4 (call it Σ).

$$\begin{pmatrix} f_1^2*\sigma^2_F + \sigma^2_{e1} & & & \\ f_2*f_1*\sigma^2_F & f_2^2*\sigma^2_F + \sigma^2_{e2} & & \\ f_3*f_1*\sigma^2_F & f_3*f_2*\sigma^2_F & f_3^2*\sigma^2_F + \sigma^2_{e3} & \\ f_4*f_1*\sigma^2_F & f_4*f_2*\sigma^2_F & f_4*f_3*\sigma^2_F & f_4^2*\sigma^2_F + \sigma^2_{e4} \end{pmatrix}$$

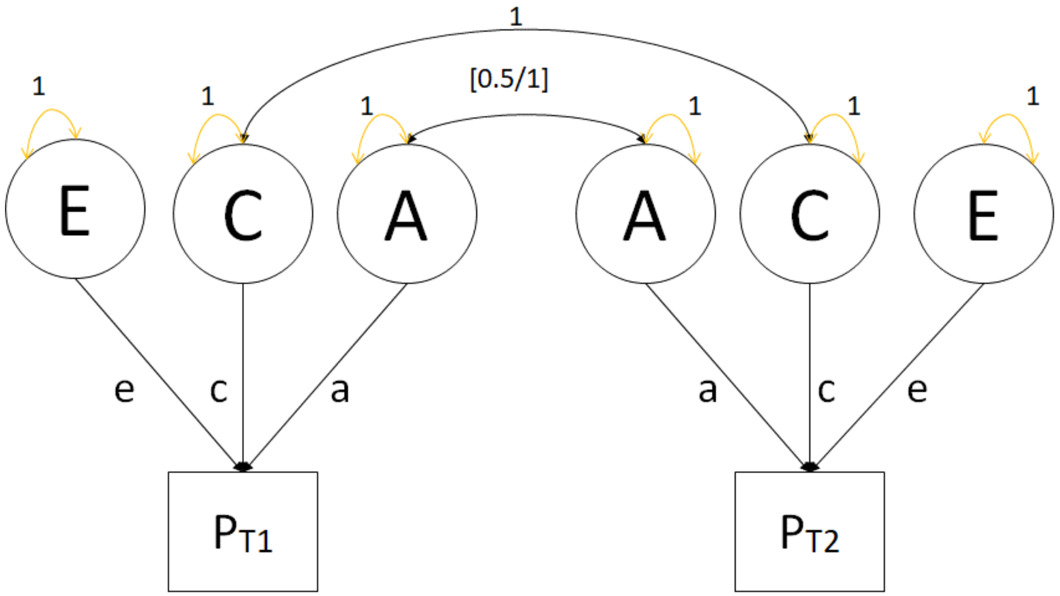
in next slides, I am going to drop "*", e.g., $f_1^2*\sigma^2_F + \sigma^2_{e1} = f_1^2\sigma^2_F + \sigma^2_{e1}$

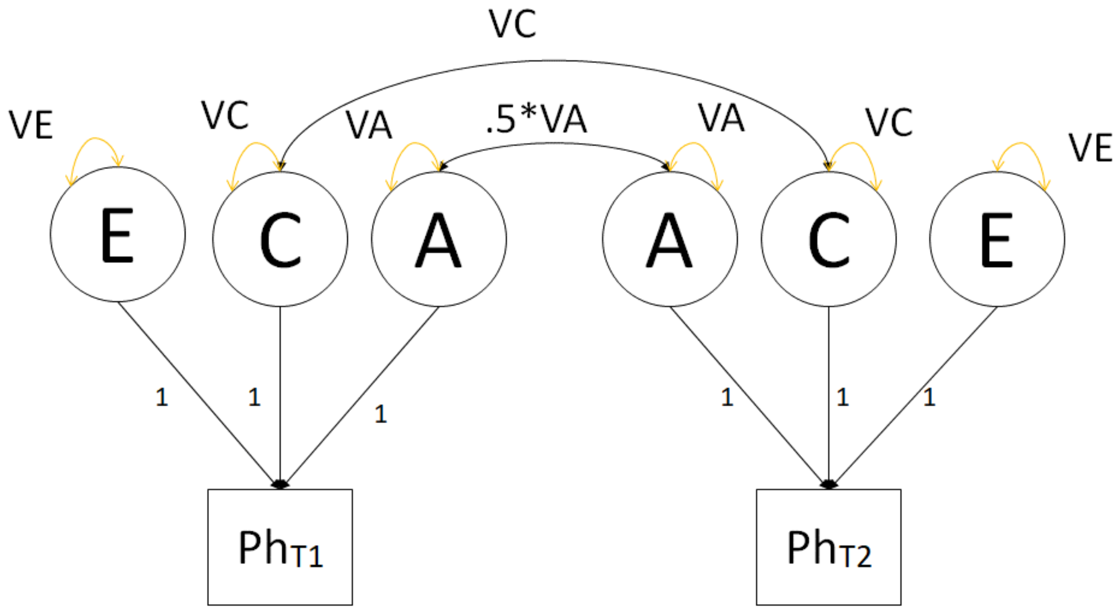


Scaling of the common factor (latent variable) –
how can be estimate variance of F, is F is not observed?

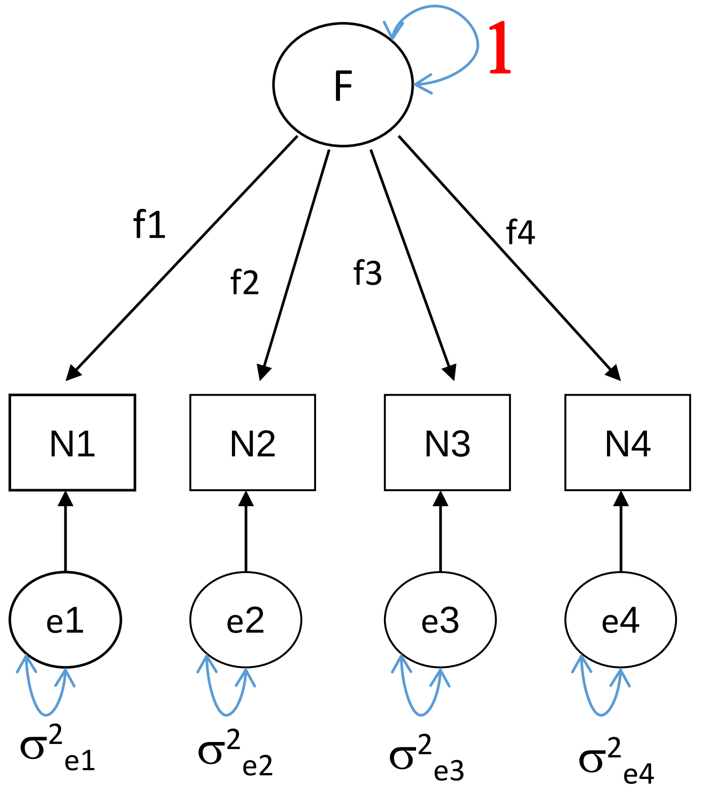
1) standardize F so that $\sigma^2_F = 1$ or
2) fixed a factor loading to 1 so that the variance of F depends directly on the scale of the indicator

Actually you already know about scaling



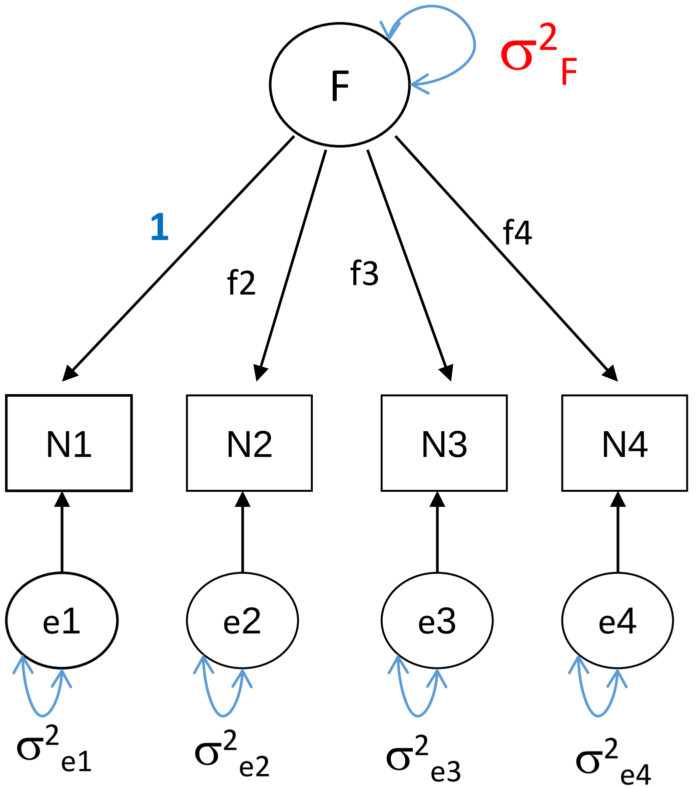


A, C and E are statistically latent variale: in the twin model, we do not observe them directly ....



$$
\begin{pmatrix}
f_1^2 1 + \sigma^2_{e1} & & & \\
f_2 f_1 1 & f_2^2 1 + \sigma^2_{e2} & & \\
f_3 f_1 1 & f_3 f_2 1 & f_3^2 1 + \sigma^2_{e3} & \\
f_4 f_1 1 & f_4 f_2 1 & f_4 f_3 1 & f_4^2 1 + \sigma^2_{e4}
\end{pmatrix}
=
\begin{pmatrix}
f_1^2 + \sigma^2_{e1} & & & \\
f_2 f_1 & f_2^2 + \sigma^2_{e2} & & \\
f_3 f_1 & f_3 f_2 & f_3^2 + \sigma^2_{e3} & \\
f_4 f_1 & f_4 f_2 & f_4 f_3 & f_4^2 + \sigma^2_{e4}
\end{pmatrix}
$$

Latent variance scaled by fixed its variance to 1 (standardization)



$$\begin{pmatrix} 1^2\sigma^2_F + \sigma^2_{e1} & & & \\ f_2 1\sigma^2_F & f_2^{\,2}\sigma^2_F + \sigma^2_{e2} & & \\ f_3 1\sigma^2_F & f_3 f_2\sigma^2_F & f_3^{\,2}\sigma^2_F + \sigma^2_{e3} & \\ f_4 1\sigma^2_F & f_4 f_2\sigma^2_F & f_4 f_3\sigma^2_F & f_4^{\,2}\sigma^2_F + \sigma^2_{e4} \end{pmatrix}$$

$$\begin{pmatrix} \sigma^2_F + \sigma^2_{e1} & & & \\ f_2\sigma^2_F & f_2^{\,2}\sigma^2_F + \sigma^2_{e2} & & \\ f_3\sigma^2_F & f_3 f_2\sigma^2_F & f_3^{\,2}\sigma^2_F + \sigma^2_{e3} & \\ f_4\sigma^2_F & f_4 f_2\sigma^2_F & f_4 f_3\sigma^2_F & f_4^{\,2}\sigma^2_F + \sigma^2_{e4} \end{pmatrix}$$

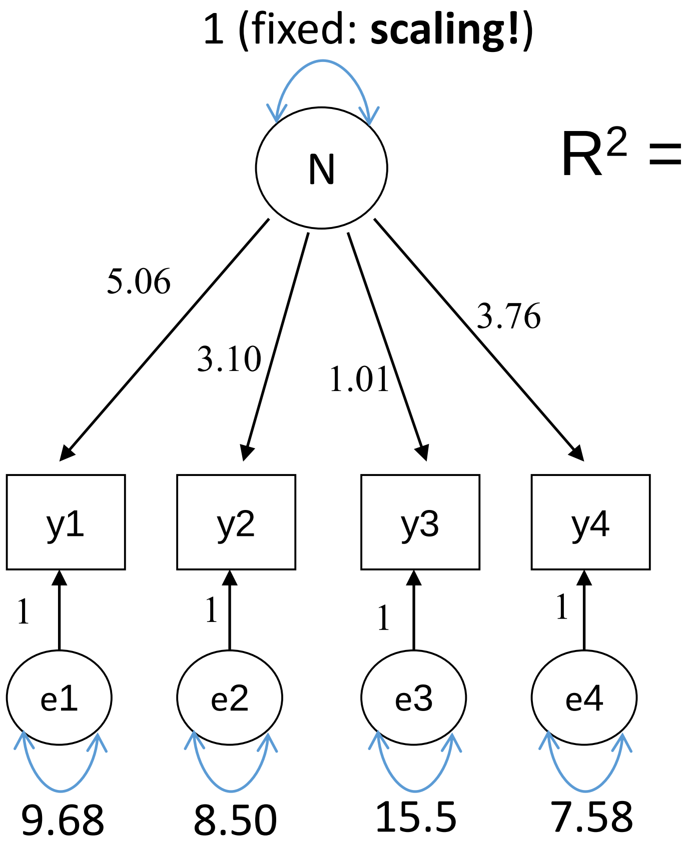
Latent variance scaled by fixing $f_1 = 1$ (or fix f_2, f_3, or f_4 to 1).

Observed covariance matrix (N=361)

35.278
15.763 18.109
 4.942 2.661   16.594
18.970 11.622   4.262 21.709

Expected covariance matrix (Σ)

35.278
15.682 18.109
5.085   3.115   16.594
19.011 11.649 3.777 21.709



$$R^2 = (f_1^{\,2} * \sigma^2_N) / (f_1^{\,2} * \sigma^2_N + \sigma^2_{e1})$$

var(n1) = 5.06^2*1 + 9.68 =35.27
rel(n1) = 5.06^2*1 / 35.27 = .725
(R^2 in regression of y1 on N)

how do we get Σ ? see previous slides!

Matrix algebraic representation of the model for Σ,
given p observed variables, and m latent variables

$$\Sigma = L_f * \Sigma_F * L_f^t + \Sigma_R$$

Σ is the pxp symmetric expected covariance matrix
L_f is the pxm matrix of factor loading
Σ_F is the mxm covariance (correlation) matrix of the common factors
Σ_R is the pxp covariance matrix of the residuals.

given p observed variables, and m latent variables

$$\Sigma = L_f * \Sigma_F * L_f^t + \Sigma_R$$

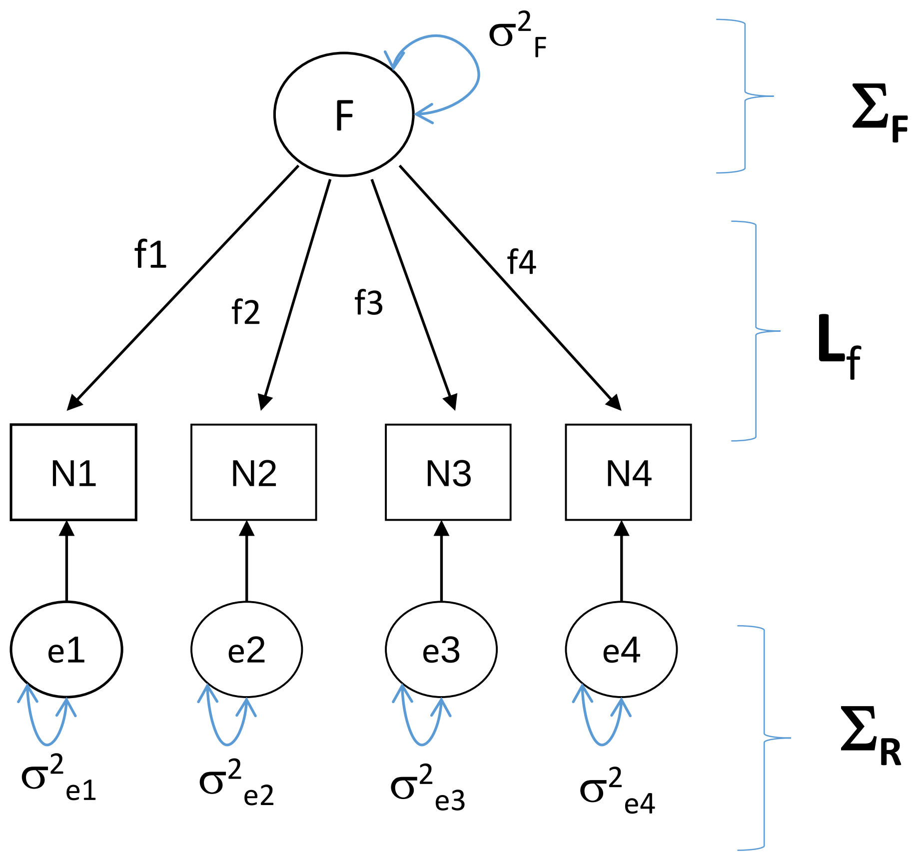
Given P=4, m=1

$\mathbf{L}_f = \begin{bmatrix} f1 \\ f2 \\ f3 \\ f4 \end{bmatrix}$ 4 x 1

$\mathbf{L}_f^{\,t} = \begin{bmatrix} f1 & f2 & f3 & f4 \end{bmatrix}$ 1 x 4

$\Sigma_{\mathbf{F}} = \begin{bmatrix} \sigma^2{}_F \end{bmatrix}$ 1 x 1

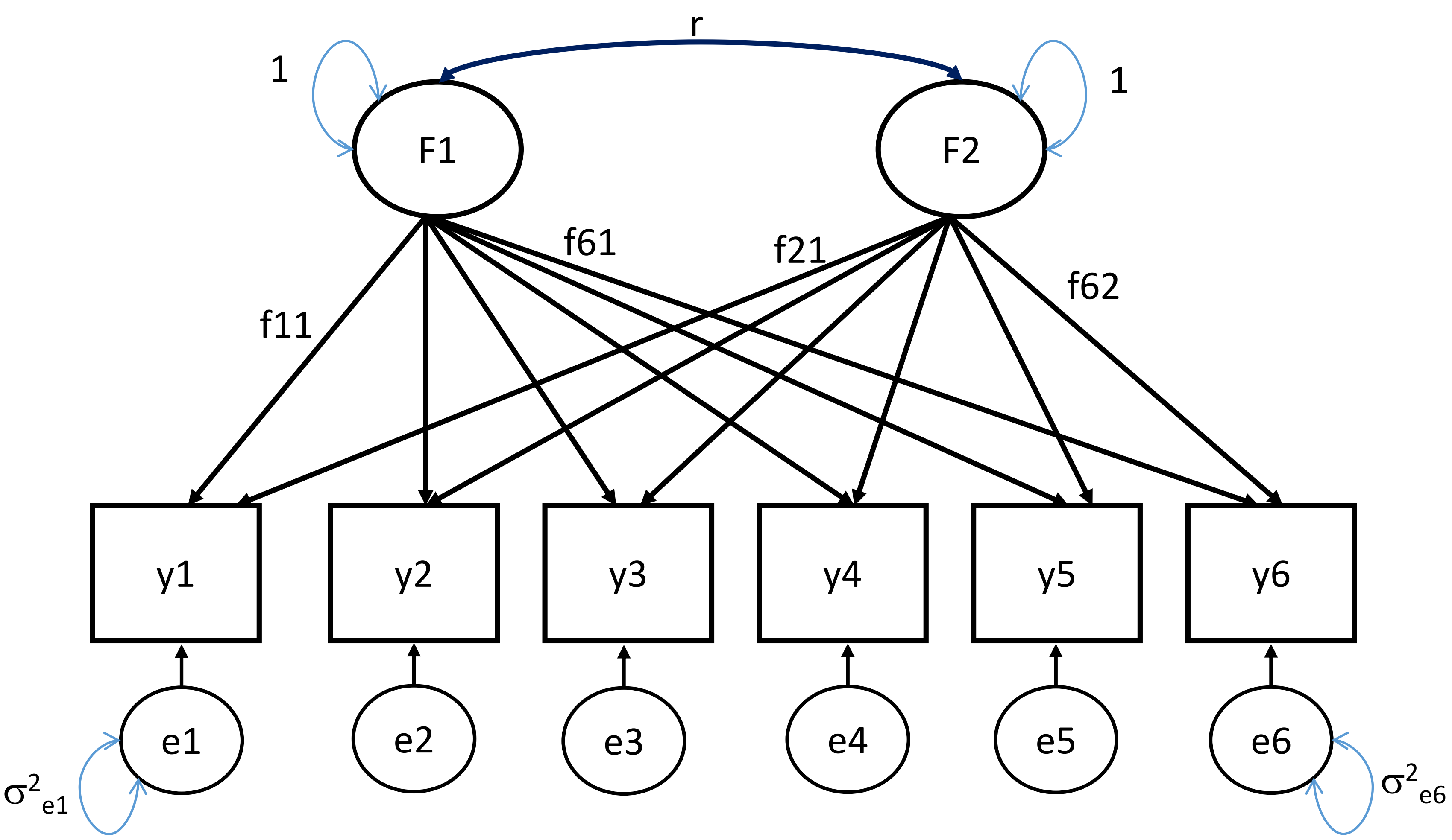
$\Sigma_{\mathbf{R}} = \begin{bmatrix} \sigma^2{}_{e1} & 0 & 0 & 0 \\ 0 & \sigma^2{}_{e2} & 0 & 0 \\ 0 & 0 & \sigma^2{}_{e3} & 0 \\ 0 & 0 & 0 & \sigma^2{}_{e4} \end{bmatrix}$ 4 x 4



Multiple common factors: Confirmatory vs. Exploratory Factor Analysis (CFA vs EFA). EFA **Aim**: determine dimensionality and derive meaning of factors from factor loadings

Exploratory approach: How many common factor? What is the pattern of factor loadings? Can we derive the meaning of the common factor from the pattern of factor loadings ($\mathbf{L}_f$)? Low on prior theory, but still involves choices. How many common factors: Screeplot, Eigenvalue > 1 rule, Goodness of fit measures (RMSEA, NNFI), info criteria (BIC, AIC).

EFA (two) factor model as it is fitted in standard programs: all indicators (p=6) load on all common factors (m=2). Note: scaling (σ^2_{F1}=1, σ^2_{F2}=1)



$y_1 = f_{11} F_1 + f_{12} F_2 + e_1$
$y_2 = f_{21} F_1 + f_{22} F_2 + e_2$
$y_3 = f_{31} F_1 + f_{32} F_2 + e_3$
$y_4 = f_{41} F_1 + f_{42} F_2 + e_4$
$y_5 = f_{51} F_1 + f_{52} F_2 + e_5$
$y_6 = f_{61} F_1 + f_{62} F_2 + e_6$

$\mathbf{L_f}$ (6x2)= $\begin{matrix} f_{11} & f_{12} \\ f_{21} & f_{22} \\ \dots & \dots \\ f_{51} & f_{52} \\ f_{61} & f_{62} \end{matrix}$

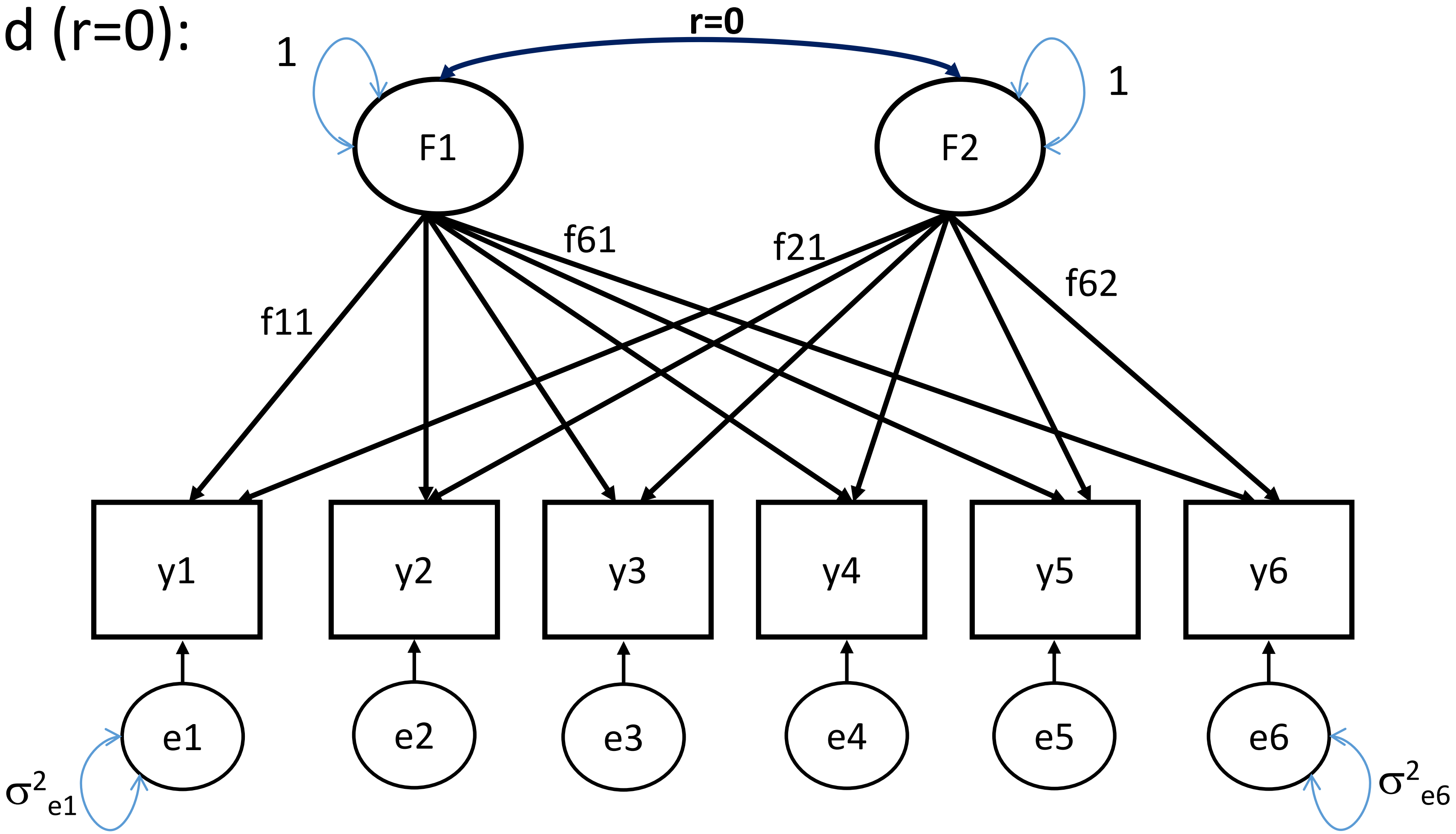
expected covariance matrix:

$$\underset{(p \times p)}{\Sigma} = \underset{(p \times m)}{L_f} * \underset{(pxp)}{\Sigma_F} * \underset{(p \times m)}{L_f^t} + \underset{(p \times p)}{\Sigma_R}$$

$$\Sigma_F \text{ (2x2)} = \begin{matrix} 1 & r \\ r & 1 \end{matrix}$$

$$\Sigma_R \text{ (6x6)} = \mathrm{diag}(\sigma^2_{e1} \quad \sigma^2_{e2} \quad \sigma^2_{e3} \quad \sigma^2_{e4} \quad \sigma^2_{e5} \quad \sigma^2_{e6})$$

EFA as fitted (r=0):



$\mathbf{L_f}$ (6x2) is not necessarily interpretable and **r=0** is not necessarily desirable.
not 6x2 = 12 free loadings, actually 12 – 1 loadings (indetification)

example

N=300 (o1, o2, o3, o4 openness to experience; a1, a2, a4, a5 agreeableness)

Correlation Matrix

		o1	o2	o3	o4	a1	a2	a4	a5
Correlation	o1	1.000	.258	.325	.130	.095	.062	.096	.051
	o2	.258	1.000	.503	.246	.093	.138	-.037	.063
	o3	.325	.503	1.000	.202	.211	.189	-.010	.109
	o4	.130	.246	.202	1.000	.108	.102	.080	.059
	a1	.095	.093	.211	.108	1.000	.441	.427	.281
	a2	.062	.138	.189	.102	.441	1.000	.415	.473
	a4	.096	-.037	-.010	.080	.427	.415	1.000	.431
	a5	.051	.063	.109	.059	.281	.473	.431	1.000

L_f (6x2)

Factor Matrix [a]

	Factor	
	1	2
o1	.295	.268
o2	.415	.514
o3	.539	.557
o4	.254	.169
a1	.564	-.214
a2	.643	-.280
a4	.505	-.471
a5	.525	-.323

$$\Sigma_F \text{ (2x2)} = \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$$

$$\Sigma = L_f * \Sigma_F * L_f^t + \Sigma_R$$

Unrotated factor loading matrix: not necessarily interpretable. Transform $\mathbf{L_f}$ by 'factor rotation" to increase interpretability not interpretable

Factor Matrix [a]

	Factor	
	1	2
o1	.295	.268
o2	.415	.514
o3	.539	.557
o4	.254	.169
a1	.564	-.214
a2	.643	-.280
a4	.505	-.471
a5	.525	-.323

not rotated r=0

interpretable ...?

Rotated Factor Matrix [a]

	Factor	
	1	2
o1	.065	.394
o2	.007	.661
o3	.076	.771
o4	.094	.291
a1	.575	.182
a2	.678	.179
a4	.688	-.056
a5	.612	.073

varimax r=0

interpretable ...?

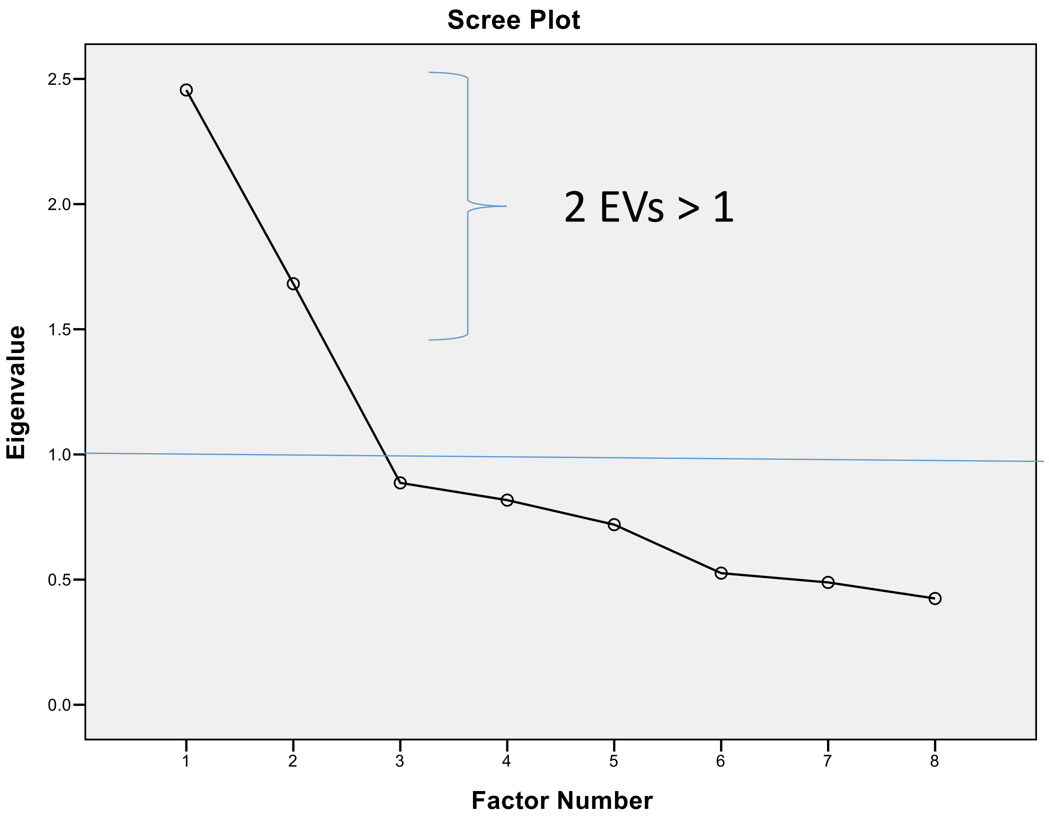
Pattern Matrix [a]

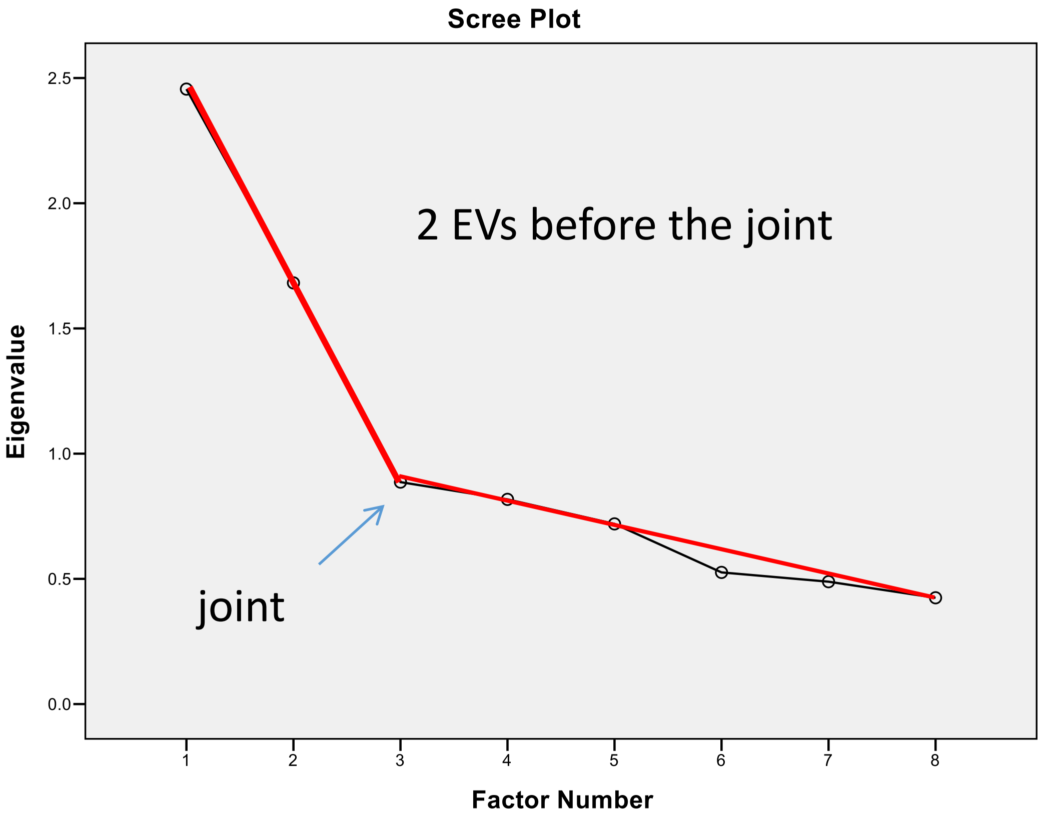
	Factor	
	1	2
o1	.016	.395
o2	-.079	.676
o3	-.022	.780
o4	.059	.285
a1	.565	.112
a2	.671	.096
a4	.713	-.147
a5	.618	-.005

oblimin r=.25

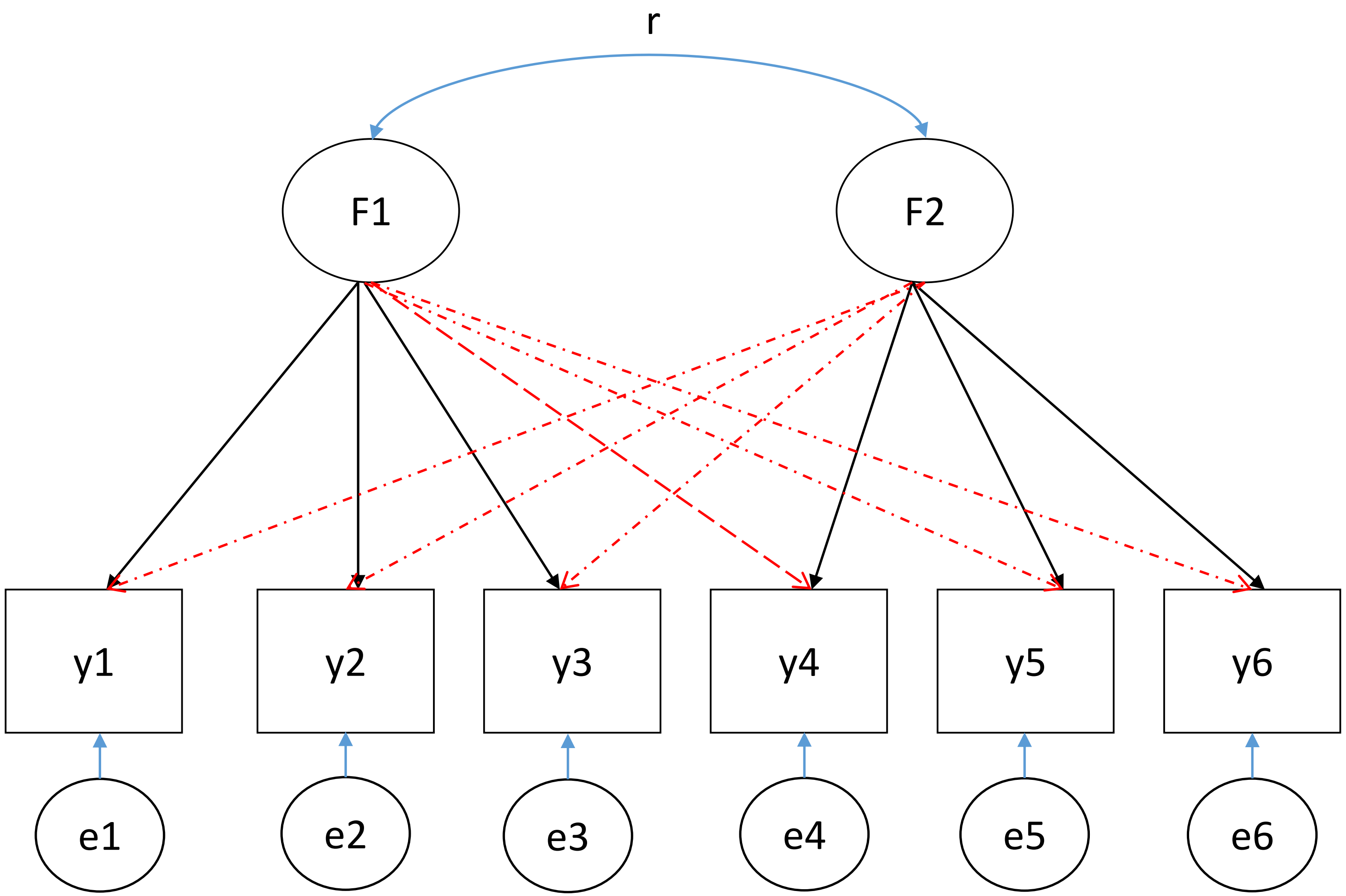
There is not statistical test here of r=0!

Determining the number of common factors in a EFA. Prior theory, or rules of thumb.
Eigenvalues > 1 rule (number of eigenvalues > 1 = ~ number of factors)
Elbow joint in the plot of the Eigenvalue (number of Eigenvalues before the elbow joint = ~ number of factors)





Confirmatory factor model: impose a pattern of loadings based on theory , define the common factors based on prior knowledge .



$y_1 = f_{11} F_1 + 0 F_2 + e_1$
$y_2 = f_{21} F_1 + 0 F_2 + e_2$
$y_3 = f_{31} F_1 + 0 F_2 + e_3$
$y_4 = 0 F_1 + f_{42} F_2 + e_4$
$y_5 = 0 F_1 + f_{52} F_2 + e_5$
$y_6 = 0 F_1 + f_{62} F_2 + e_6$

expected covariance matrix:

$$\underset{(p \times p)}{\Sigma} = \underset{(p \times m)}{L_f} * \underset{(pxp)}{\Sigma_F} * \underset{(p \times m)}{L_f^t} + \underset{(p \times p)}{\Sigma_R}$$

$$\mathbf{L_f}\ (6x2) = \begin{matrix} f_{11} & 0 \\ f_{21} & 0 \\ \dots & \dots \\ 0 & f_{52} \\ 0 & f_{62} \end{matrix}$$

$$\Sigma_F\ (2x2) = \begin{matrix} 1 & r \\ r & 1 \end{matrix}$$

$$\Sigma_R\ (6x6) = \mathrm{diag}(\sigma^2_{e1}\ \ \sigma^2_{e2}\ \ \sigma^2_{e3}\ \ \sigma^2_{e4}\ \ \sigma^2_{e5}\ \ \sigma^2_{e6})$$

CFA

$$o_1 = .416\,F_1 + 0\,F_2 + e_1$$
$$o_2 = .663\,F_1 + 0\,F_2 + e_2$$
$$o_3 = .756\,F_1 + 0\,F_2 + e_3$$
$$o_4 = .756\,F_1 + 0\,F_2 + e_4$$

$$a_1 = 0\,F_1 + .594\,F_2 + e_5$$
$$a_2 = 0\,F_1 + .726\,F_2 + e_6$$
$$a_4 = 0\,F_1 + .630\,F_2 + e_6$$
$$a_5 = 0\,F_1 + .617\,F_2 + e_4$$

$$\Sigma_F\ (2x2) = \begin{matrix} 1 & .24 \\ .24 & 1 \end{matrix}$$

Statistical test of r=0 can be done in CFA

EFA

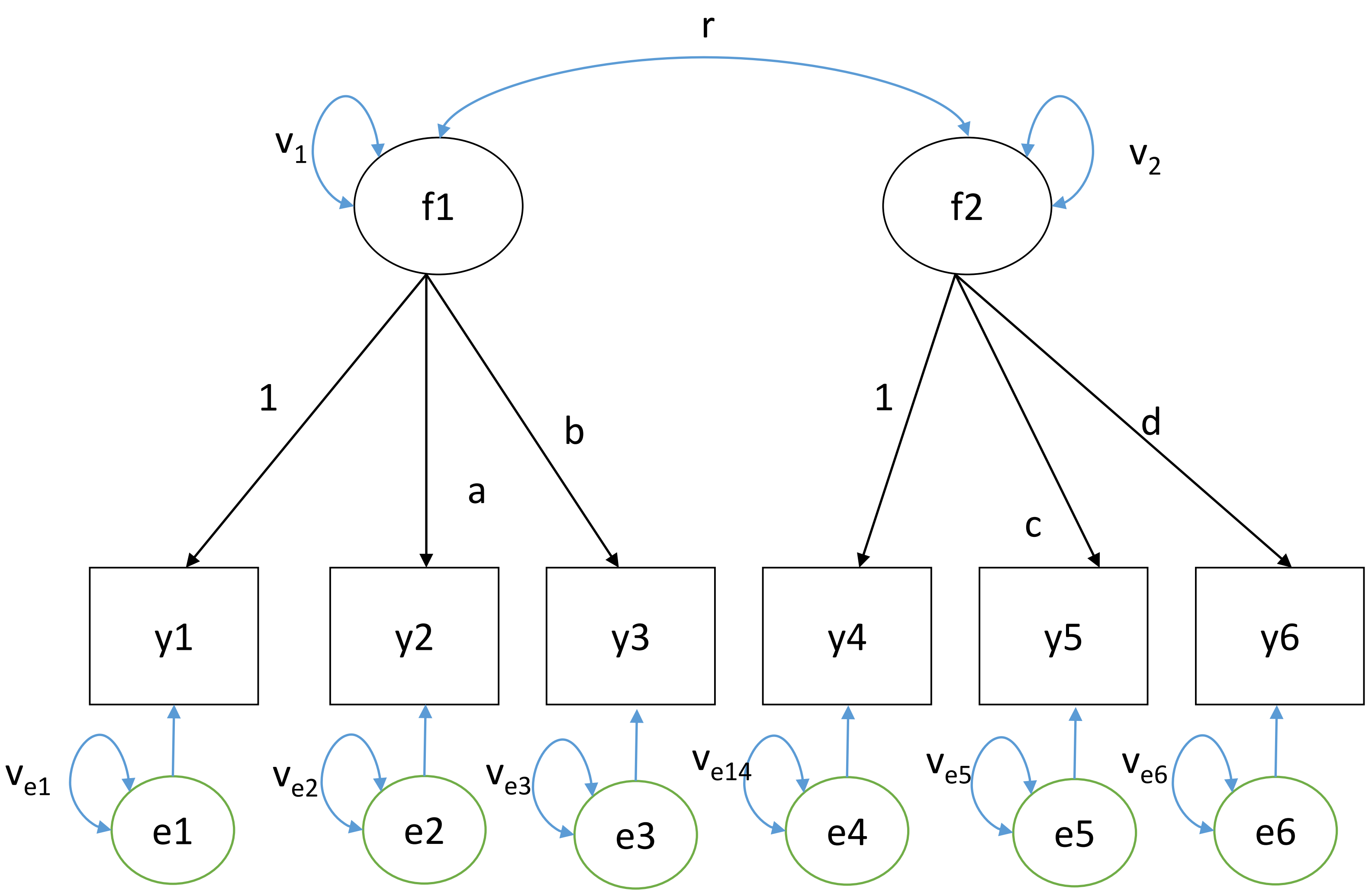
Pattern Matrix [a]

	Factor	
	1	2
o1	.016	.395
o2	-.079	.676
o3	-.022	.780
o4	.059	.285
a1	.565	.112
a2	.671	.096
a4	.713	-.147
a5	.618	-.005

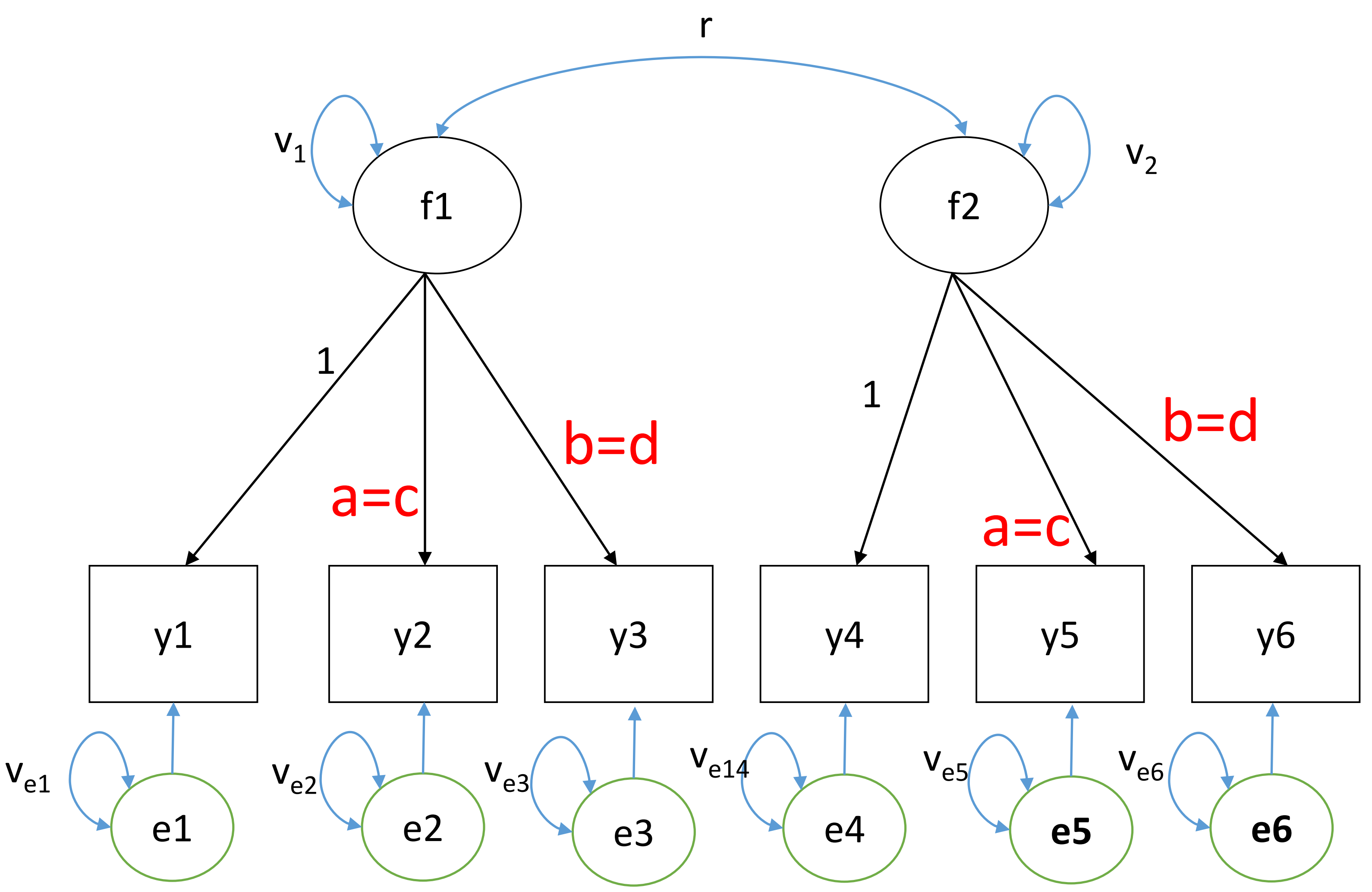
oblimin rotation

$$\Sigma_F\ (2x2) = \begin{matrix} 1 & .25 \\ .25 & 1 \end{matrix}$$

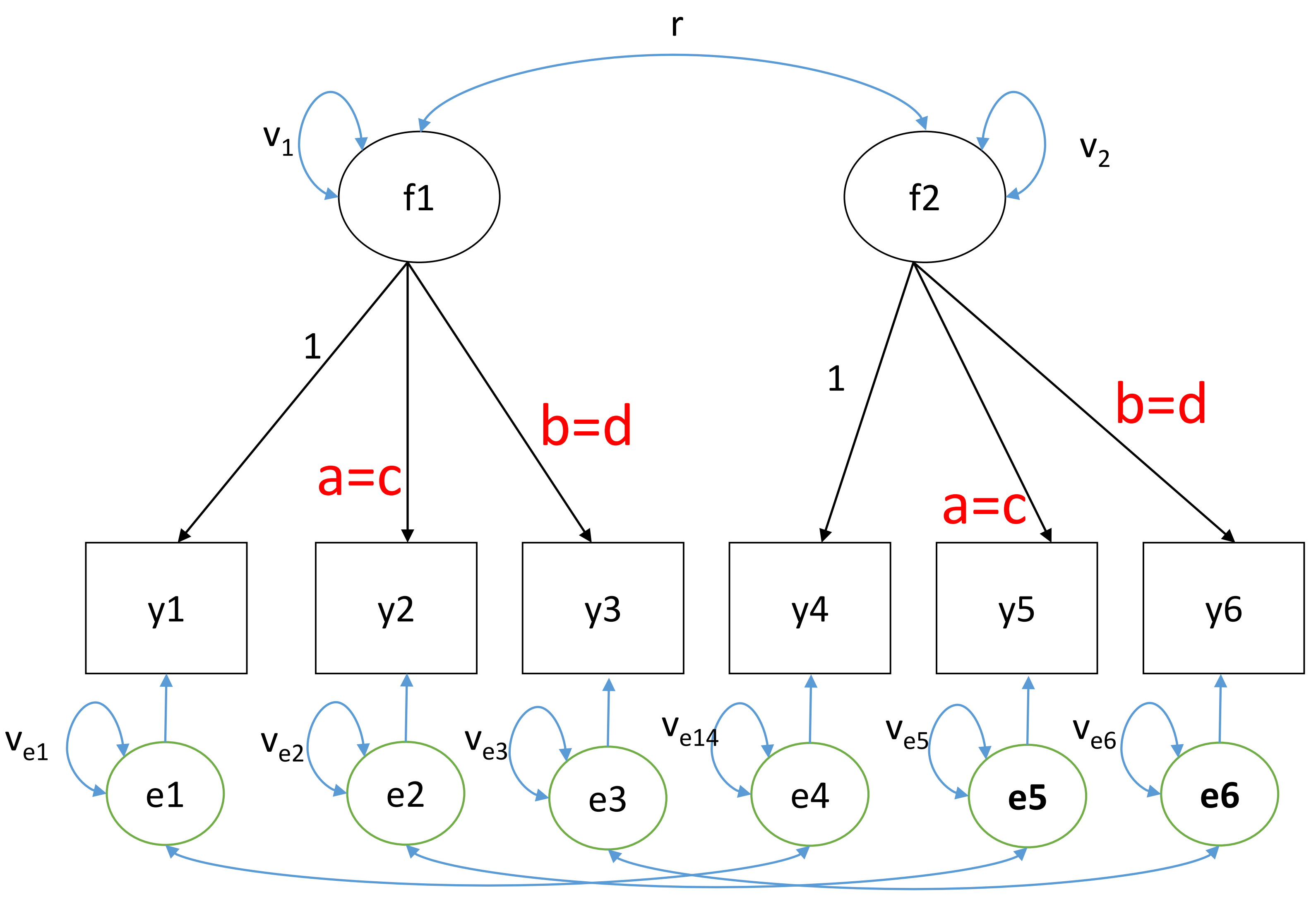
Suppose 3 indicators at 2 time points



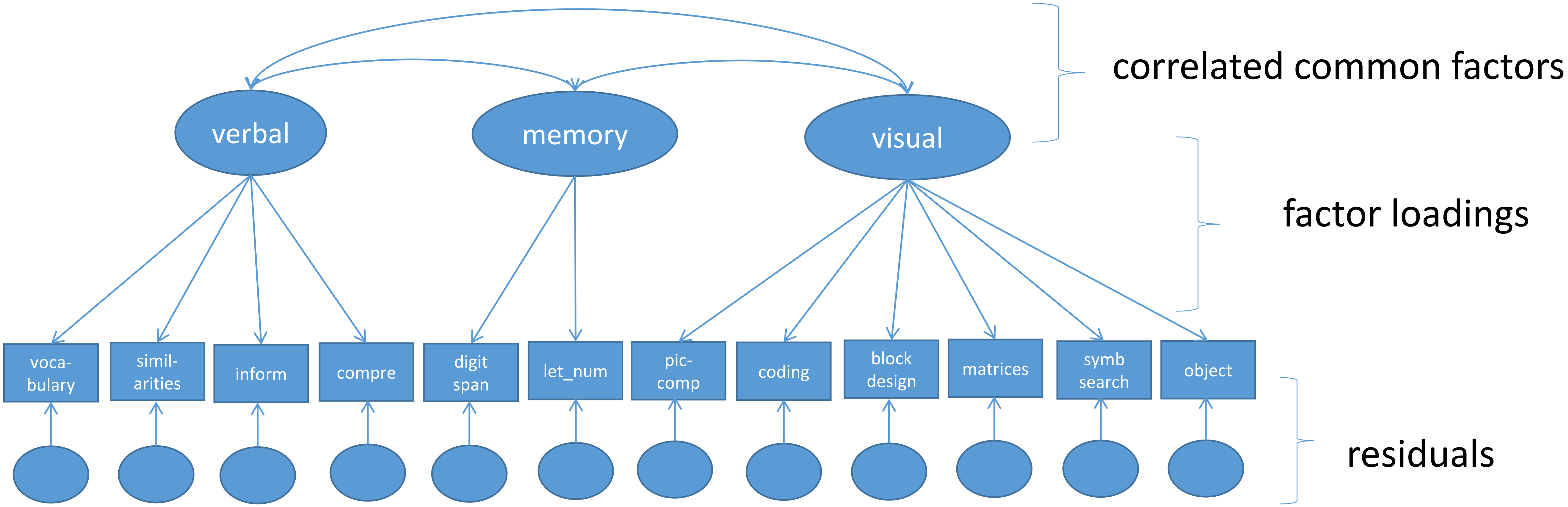
Suppose 3 indicators at 2 time points

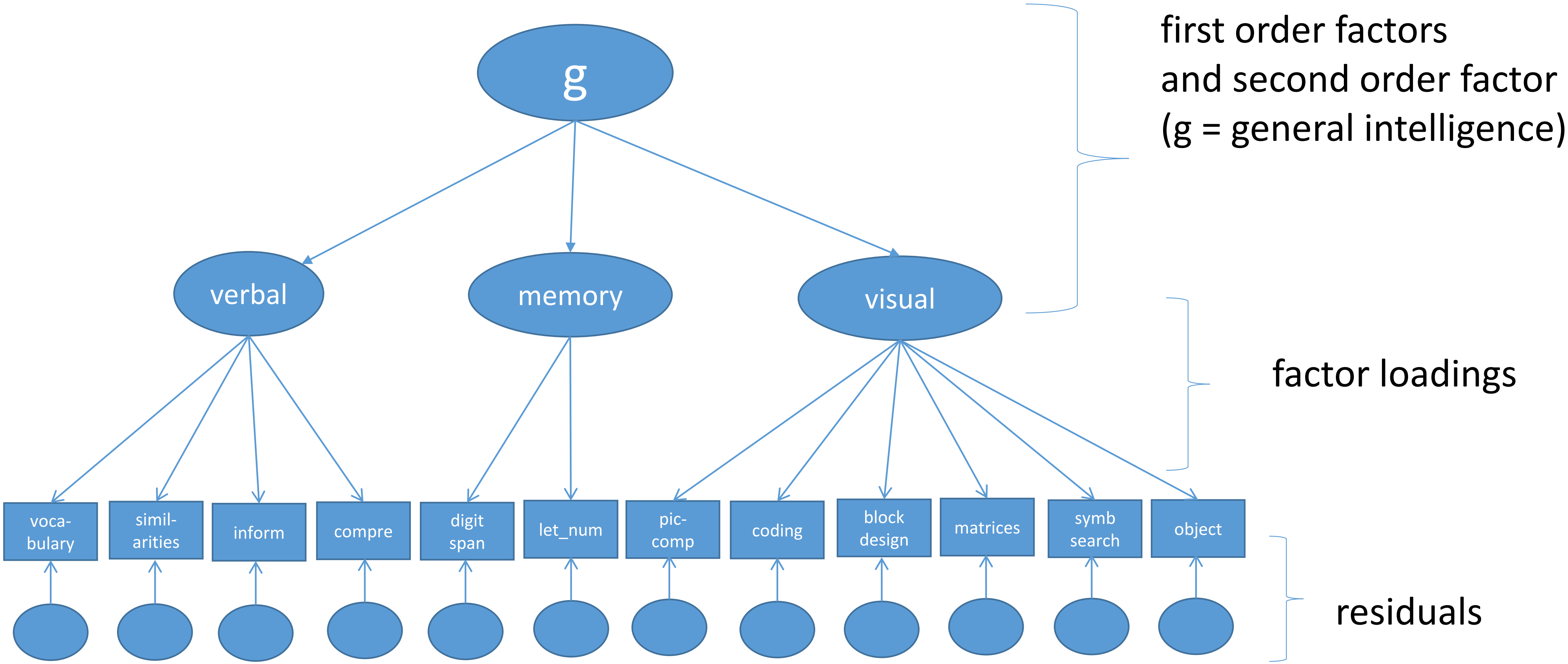


Suppose 3 indicators at 2 time points

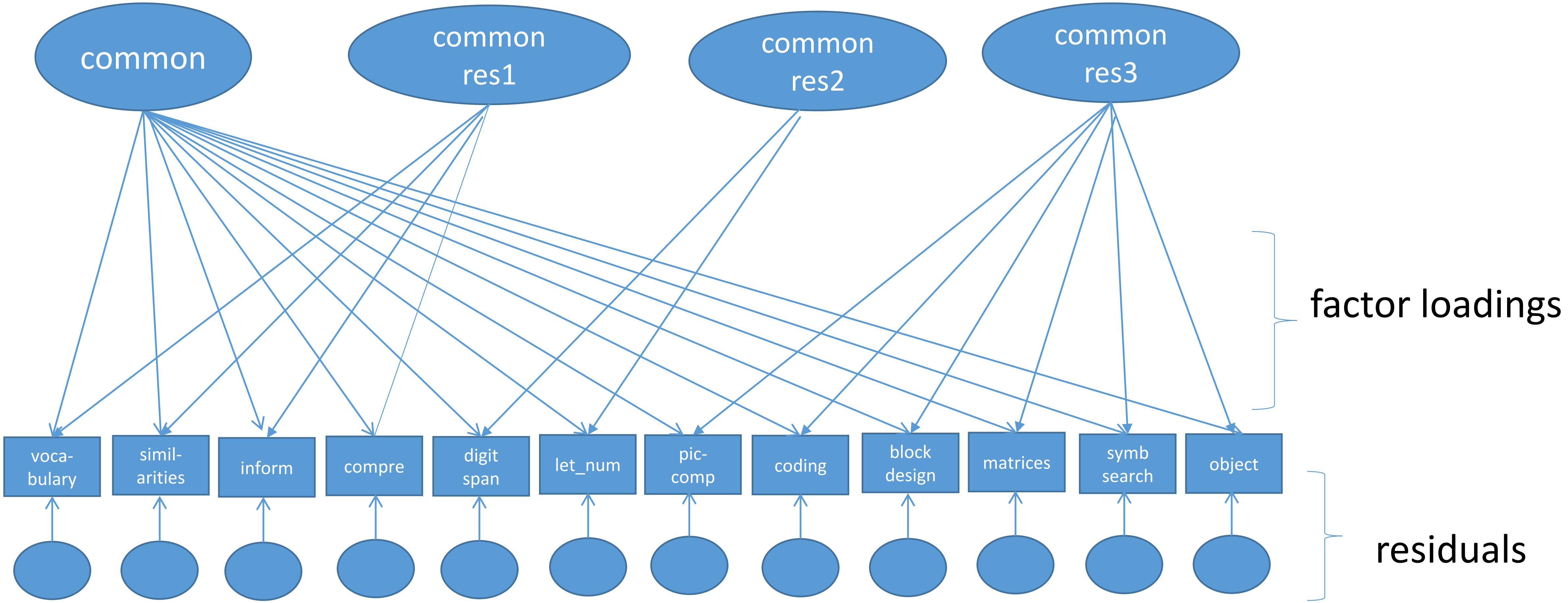


CFA applied alot to cognitive ability test scores. WAIS (Wechsler)




g
verbal
memory
visual
voca-
bulary
simil-
arities
inform
compre
digit
span
let_num
pic-
comp
coding
block
design
matrices
symb
search
object
first order factors
and second order factor
(g = general intelligence)
factor loadings
residuals

Bifactor model: alternative. Includes 1st order general factor.

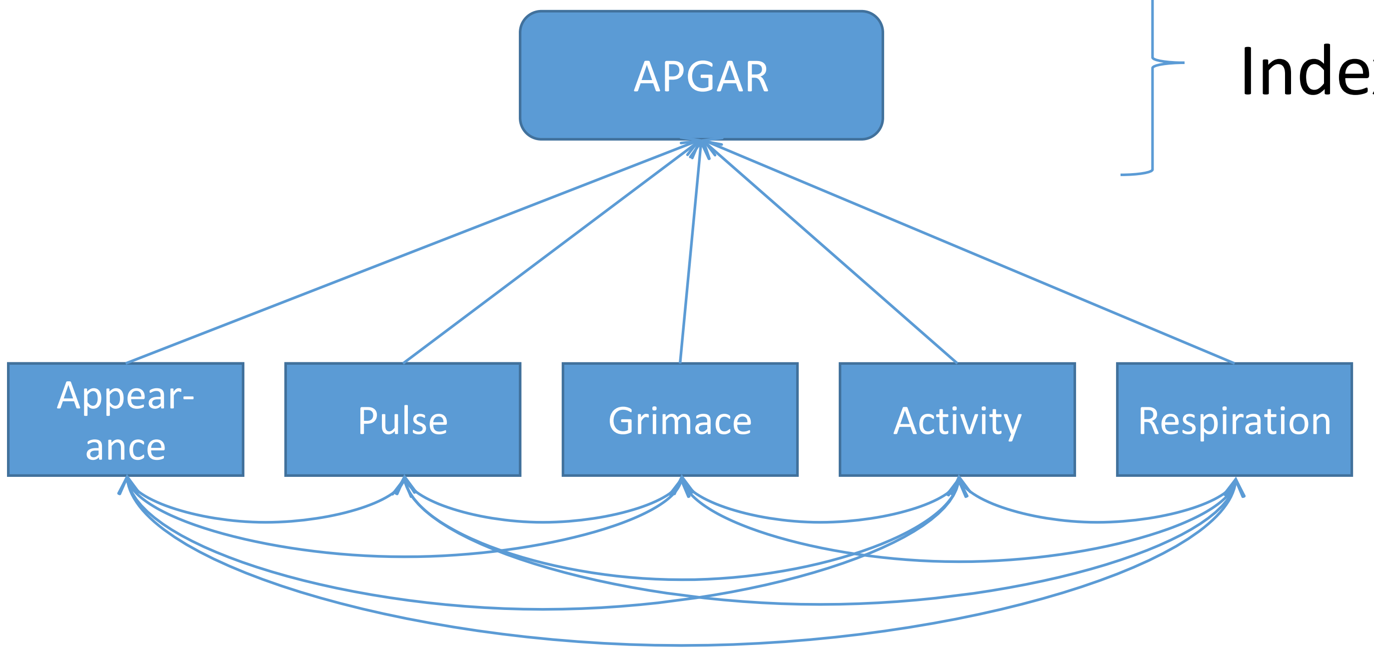


Caveat: A factor model implies phenotypic correlation, but phenotypic correlations do not necessarily imply a factor model

Apgar Scoring System

Indicator		0 Points	1 Point	2 Points
A	Activity (muscle tone)	Absent	Flexed arms and legs	Active
P	Pulse	Absent	Below 100 bpm	Over 100 bpm
G	Grimace (reflex irritability)	Floppy	Minimal response to stimulation	Prompt response to stimulation
A	Appearance (skin color)	Blue; pale	Pink body, Blue extremities	Pink
R	Respiration	Absent	Slow and irregular	Vigorous cry

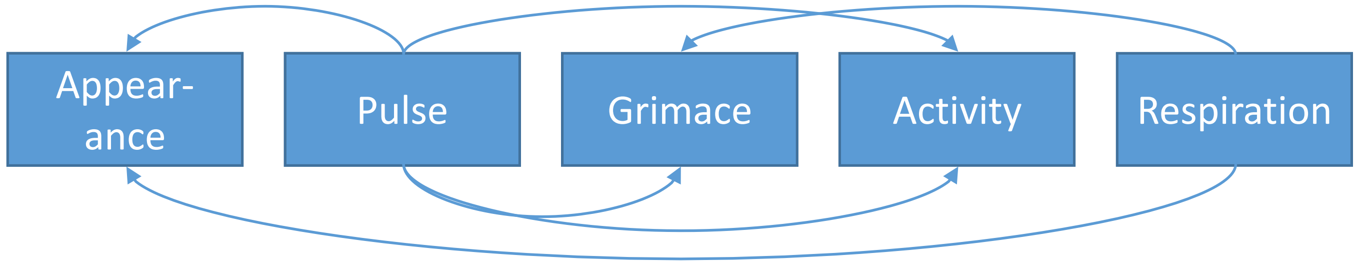
APGAR



Index of neonatal health

based on
5 formative indicators

Items are **formative**: itemscores form the APGAR score
Index variable = defined by formative items. The APGAR is dependent on the formative items. APGAR does not determine or cause the scores on the APGAR items



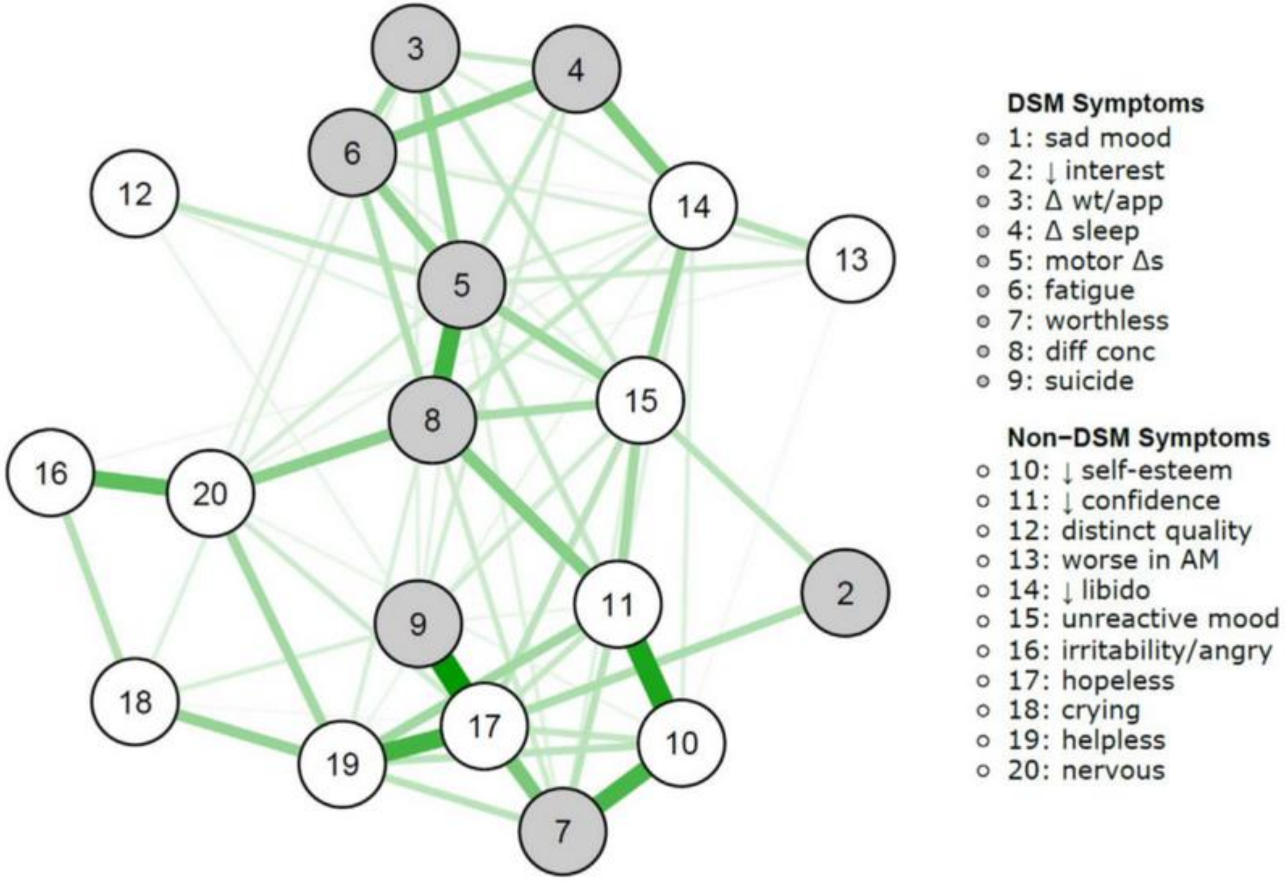
They could be a network of mutualistic direct causal effect....gives rise to correlations, which is consistent with factor model, but the generating model is **a network model**, not **the factor model**

The APGAR score is useful in diagnosis and prediction

The Centrality of DSM and non-DSM Depressive Symptoms in Han Chinese Women with Major Depression (2017). Kendler, K. S., et al. *Journal of Affective Disorders*.

Psychometric:
Depression symptoms are correlated because indicators of latent variable depression ....

Network:
Depression symptoms are correlation because they are directly interdependent
in a network



What if I want to carry out a phenotypic factor analysis given twin data?
N pairs, but N*2 individual...

1) Ignore family relatedness treat N twin pairs as 2*N individuals ? OK does not effect estimate of the covariance matrix, but renders statistical tests invalid (eigenvalues and scree plots are ok)

2) Ignore family relatedness treat N twin pairs as 2*N individuals use a correction for family clustering. OK and convenient. Requires suitable software

3) Do the factor analysis in N twins and replicate the model in the other N twins? Ok, but not true replication (call it pseudo replication)

4) Do the factor analysis in twins separately and simultaneously, but include the twin 1 – twin 2 phenotypic covariances. Ok, but possibly unwieldy (especially is you have extended pedigrees).

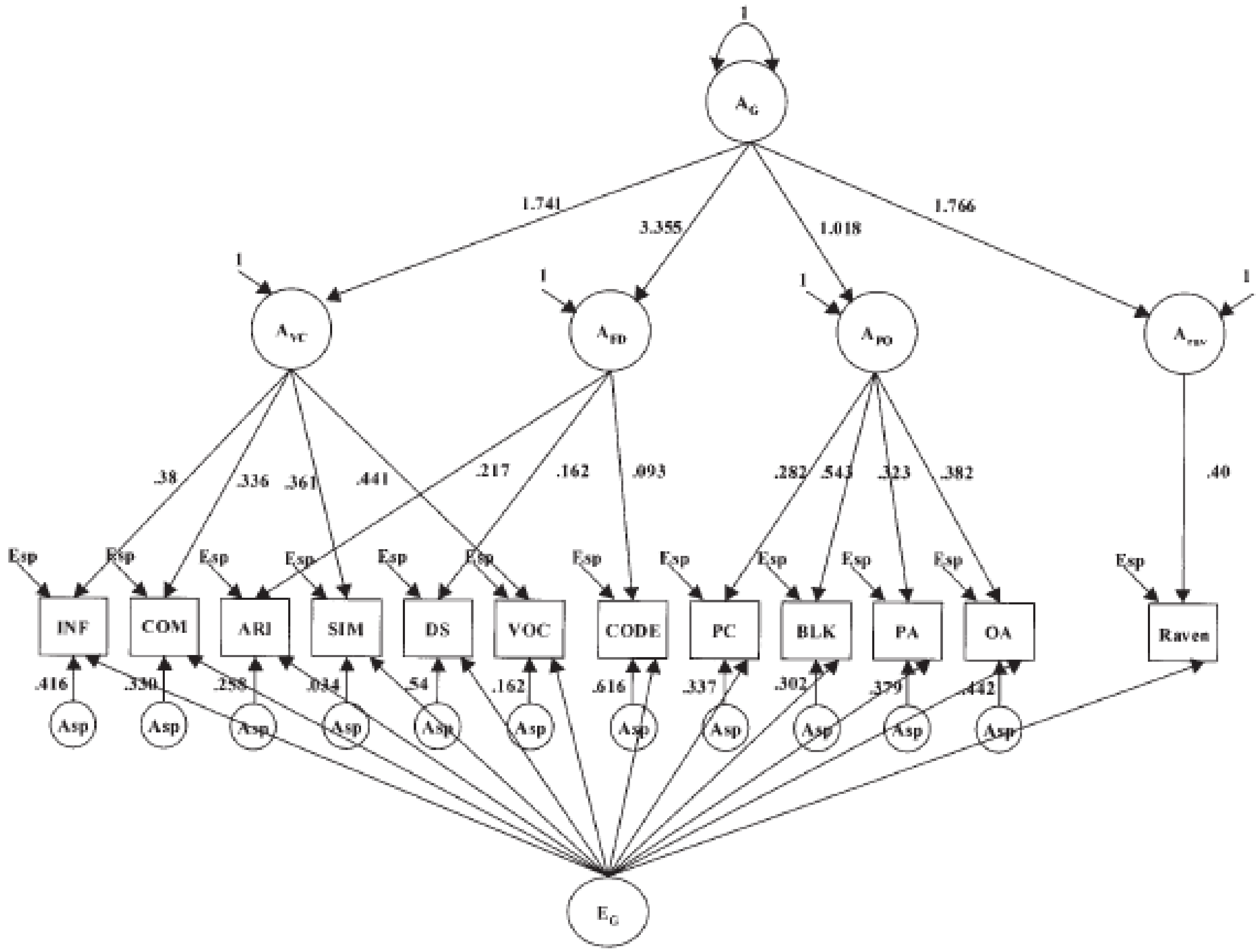
Relevance of factor analysis to twin studies genetic studies (GWAS)
1) understanding phenotypic covariance in terms of sources of A, C (D), E covariance

Decomposition of a 12x12 phenotypic covariance matrix
into 12x12 A, C, and E covariance matrices

$$\Sigma_{ph} = \Sigma_A + \Sigma_C + \Sigma_E$$

Subsequent factor modelling of Σ_A , Σ_C , Σ_E to understand the covariance structures, get a parsimonious representation

Rijsdijk FV, Vernon PA, Boomsma DI. . *Behavior Genetics, 32, 199-210,* 2002



12 cognitive ability test
(raven + WAIS)

Σ_A factor model (4 factors)

Σ_E , no common factor

Σ_C ,factor model (1 factor)

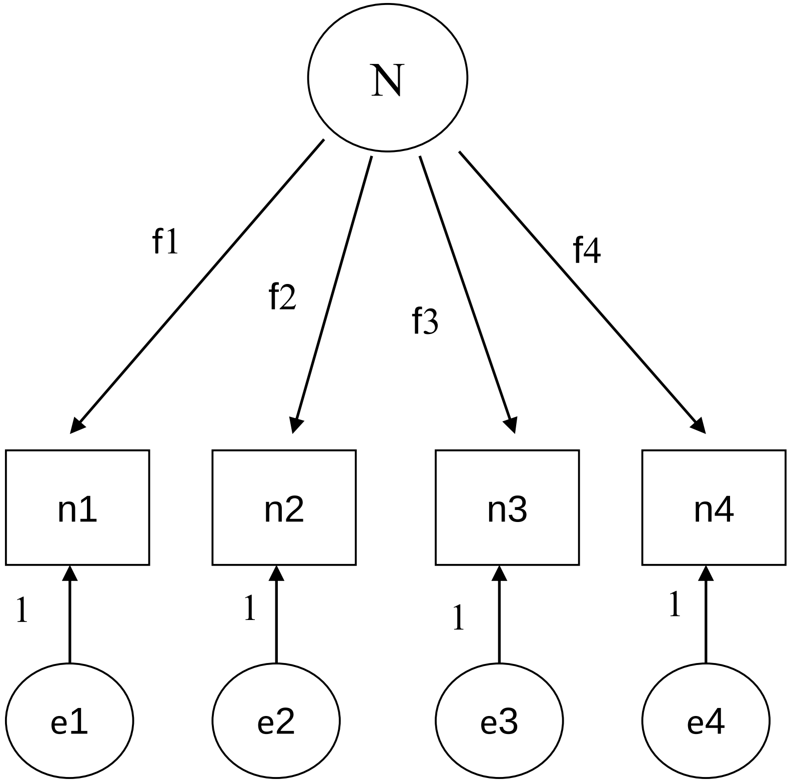
Relevance of factor analysis to twin studies genetic studies (GWAS)

2) understanding phenotypic covariance in terms of A, C (D), E covariance
Independent pathway model vs common pathway model

common refs: Kendler *et al.*, 1987, McArdle and Goldsmith, 1990.
However, Martin and Eaves presented the CP model in 1977

https://genepi.qimr.edu.au/staff/classicpapers/

This is were twin modeling meet psychometrics



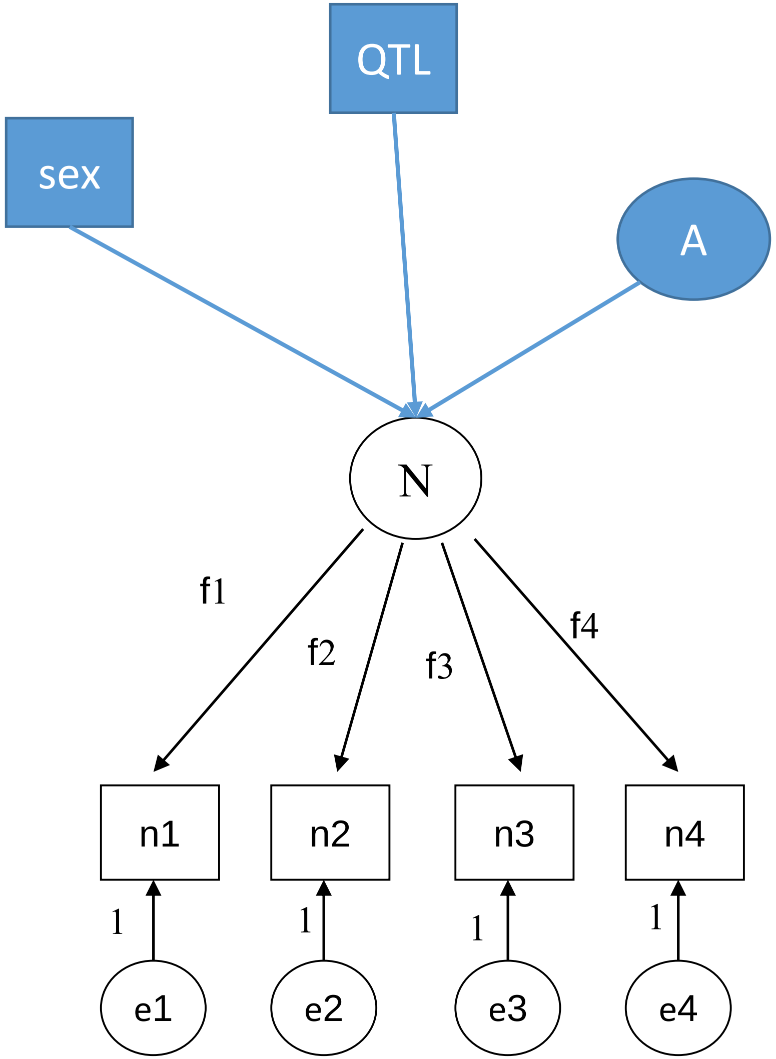
Reflective indicators: They reflect the causal action of the latent variable N

A substantive aspect of the common factor model: **interpretation** (that you bring to the model!)

Strong realistic view of the latent variable N:

N is **a real, causal, unidimensional** source of individual differences. It **exists beyond the realm of the indicator set**, and is not dependent on any given indicator set.

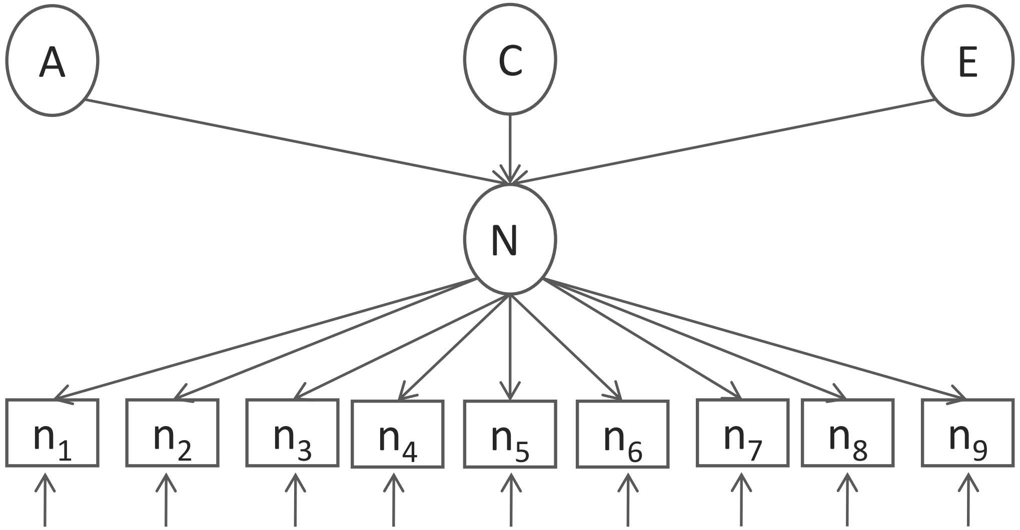
Causal - part I: **The position of N determines causally the response to the items. N is the only direct cause of systematic variation in the items.**



Causal part II: The relationship between any external variable (latent or observed) and the indicators is **mediated by the common factor N**: essence of "measurement invariance" and "differential item functioning".

If correct, the (weighted) sum of the items scores provide a proxy for N.

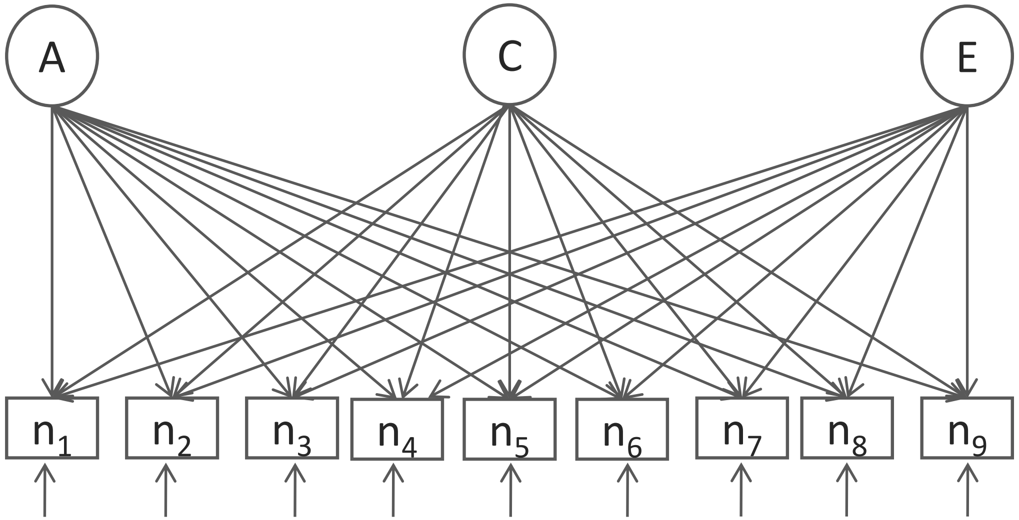
ACE modeling of (weighted) sum of items.
GWAS of (weighted) sum of items



Common pathway model

Psychometric model

Phenotypic unidimensionality N mediates all external sources of individual differences



Independent pathway model or Biometric model. Implies phenotypic multidimensionality….. What about N in the phenotypic analysis? The phenotypic (1 factor) model was incorrect?

If CP model holds, but you fit the IP, you will find that the A, C, and E factor loadings are approx. proportional (collinear): The plot the E and A loadings is a straight line (C, A; or C, E). IP model fits but CP more parsimonious option.

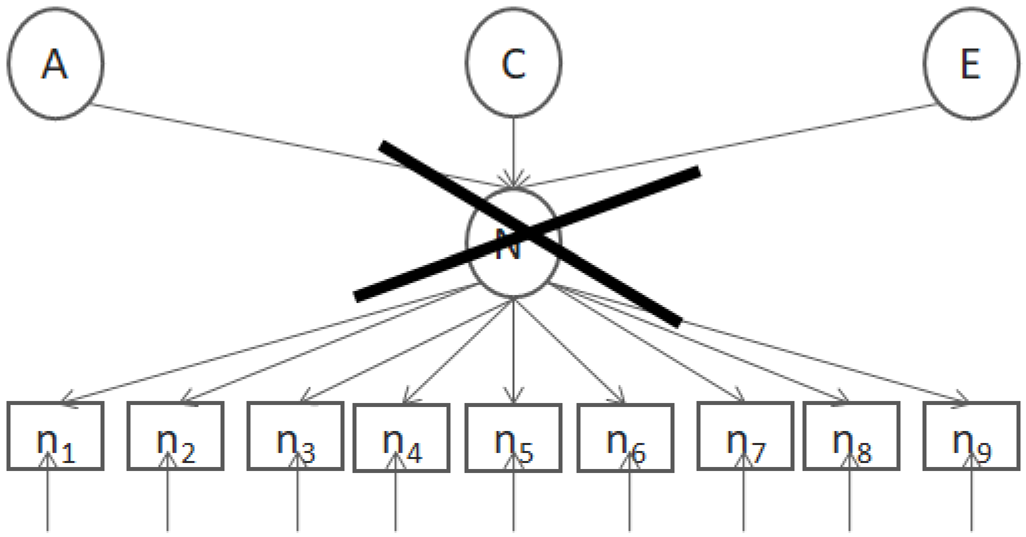
As noted by Martin and Eaves in **1977** (!)

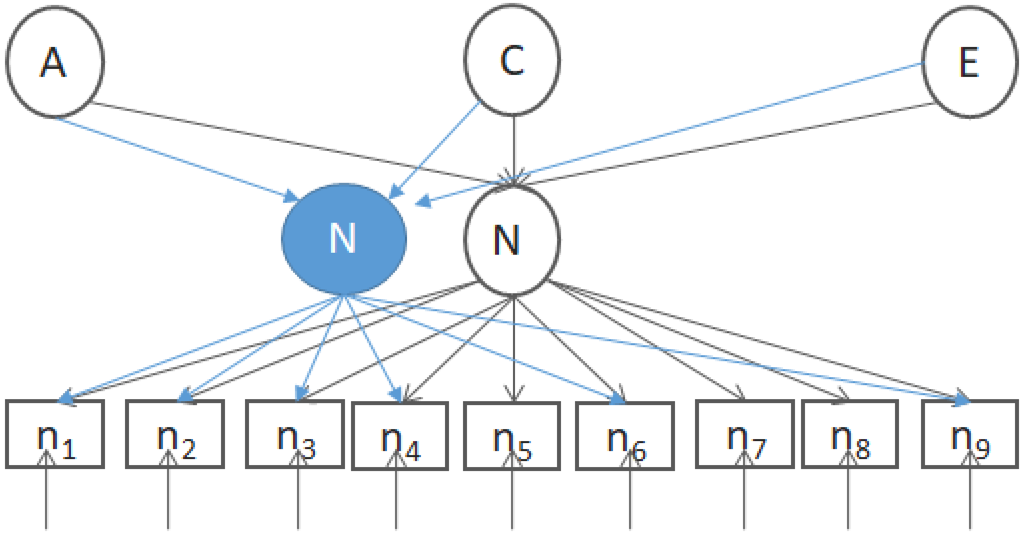
of Σ_{WMZ}. It is quite likely that we shall want to test the hypothesis that the genetical loadings (for example) are simply scaled versions of the environmental loadings. This would imply that the genetical and environmental structures are identical, apart from specific factors, and that genetical and environmental factors are affecting the same aspects of the organism in a consistent manner. Thus, to incorporate such a constraint in our model

Martin and Eaves 1977 (p 86)
https://genepi.qimr.edu.au/staff/classicpapers/

If IP model holds, but you fit the CP, you will find that the CP model does not fit.
This implies that the phenotypic factor model cannot be unidimensional.
This happens a lot.... why?

CP model is often based on a phenotypic factor model. Say single factor model...
If CP is rejected, we may conclude 1) there is not “psychometric” latent variable or 2) Mike Neale: the psychometric single factor was incorrect.





Psychological Methods

1082-989X/13/$12.00 DOI: 10.1037/a0032755

Can Genetics Help Psychometrics? Improving Dimensionality Assessment Through Genetic Factor Modeling

Sanja Franić
Vrije Universiteit Amsterdam

Conor V. Dolan and Denny Borsboom
University of Amsterdam

James J. Hudziak
University of Vermont

Catherina E. M. van Beijsterveldt and Dorret I. Boomsma
Vrije Universiteit Amsterdam

Behav Genet
DOI 10.1007/s10519-013-9628-4

ORIGINAL RESEARCH

Three-and-a-Half-Factor Model? The Genetic and Environmental Structure of the CBCL/6–18 Internalizing Grouping

Sanja Franić · Conor V. Dolan · Denny Borsboom · Catherina E. M. van Beijsterveldt · Dorret I. Boomsma

Applications

Common pathway vs Independent pathway model.

nature COMMUNICATIONS

ARTICLE

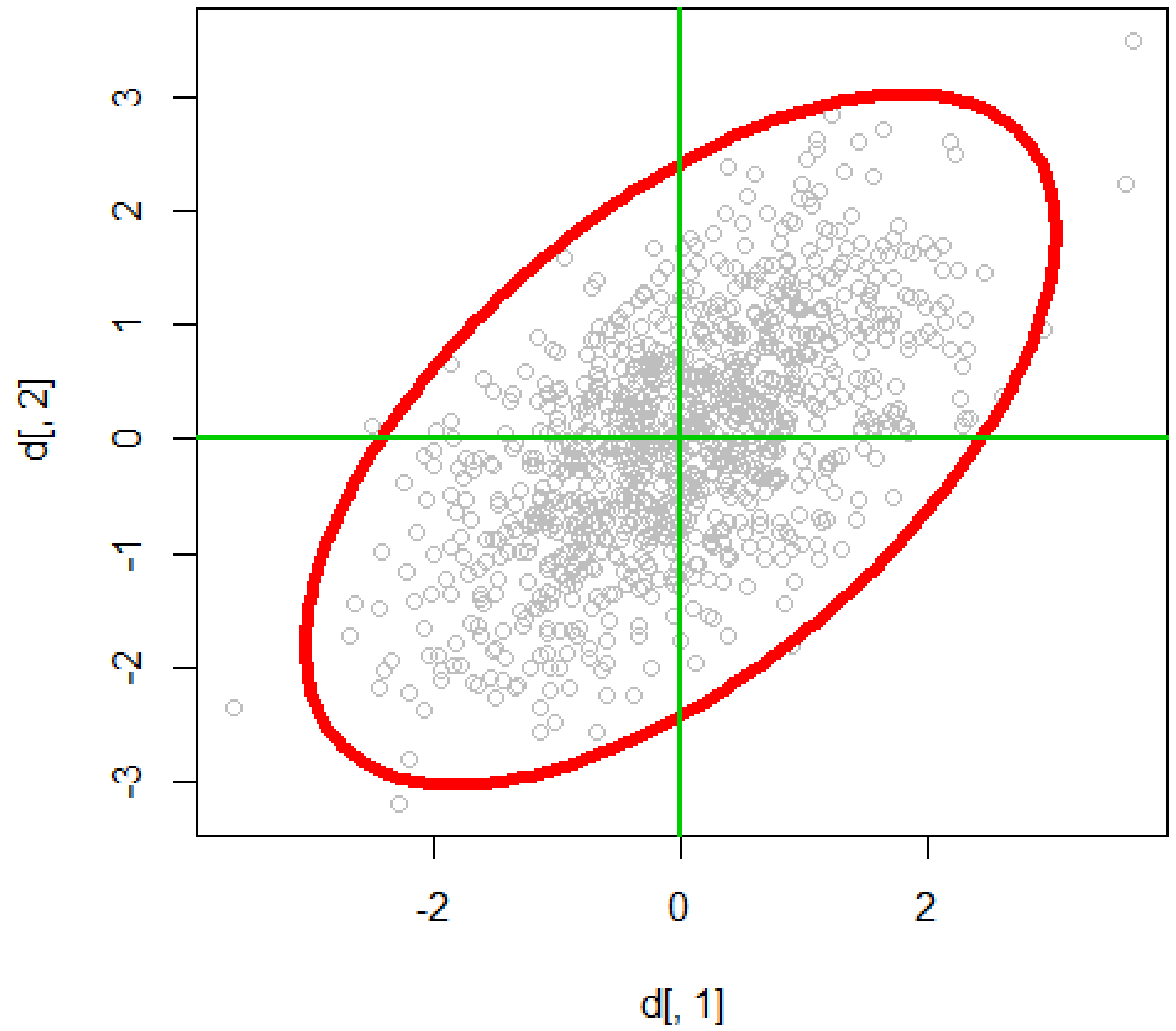
DOI: 10.1038/s41467-018-03242-8 OPEN

Item-level analyses reveal genetic heterogeneity in neuroticism

Mats Nagel[1], Kyoko Watanabe[2], Sven Stringer[2], Danielle Posthuma[1,2] & Sophie van der Sluis[1]

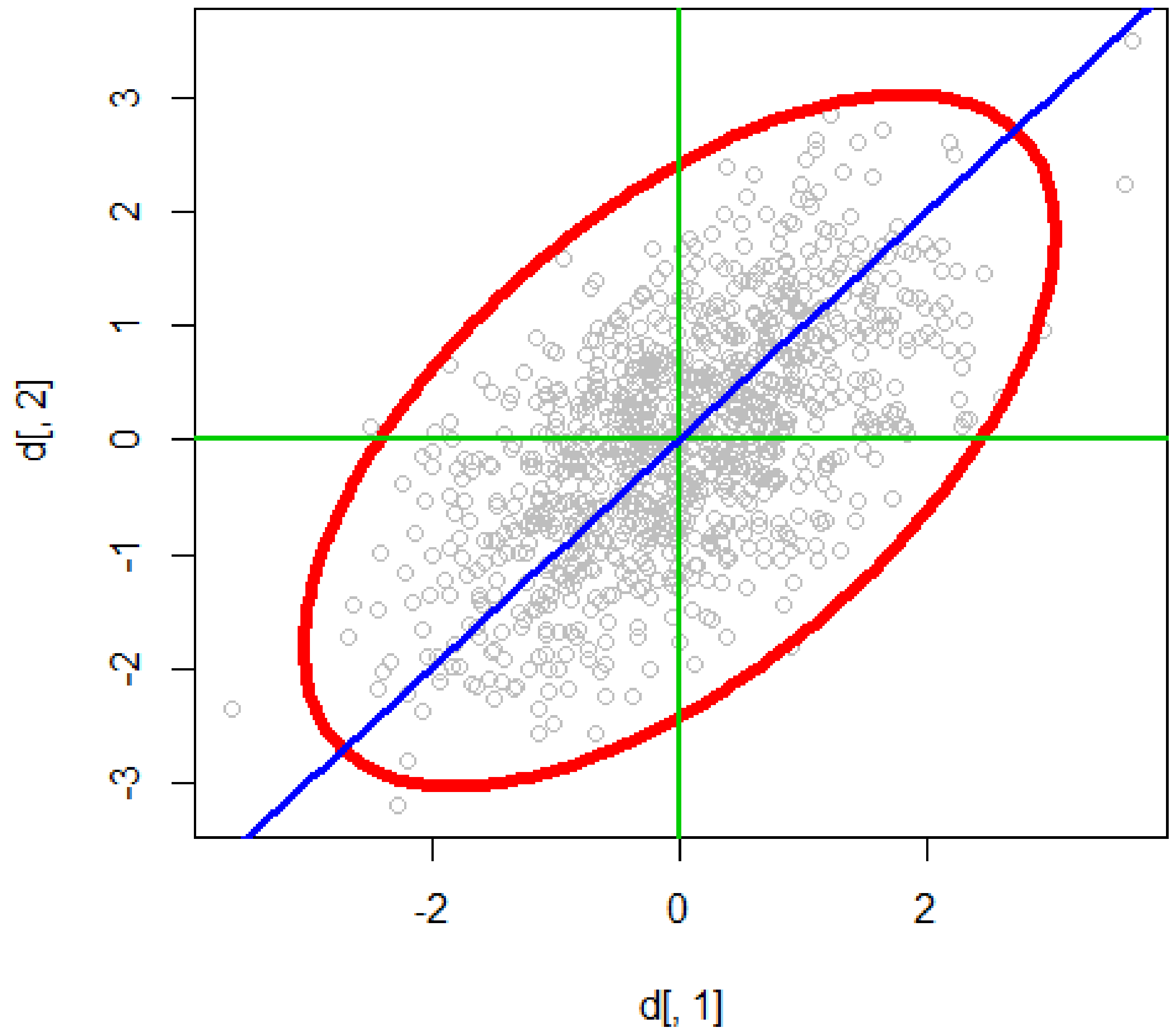
Practical:

Phenotypic factor analysis.



correlated data

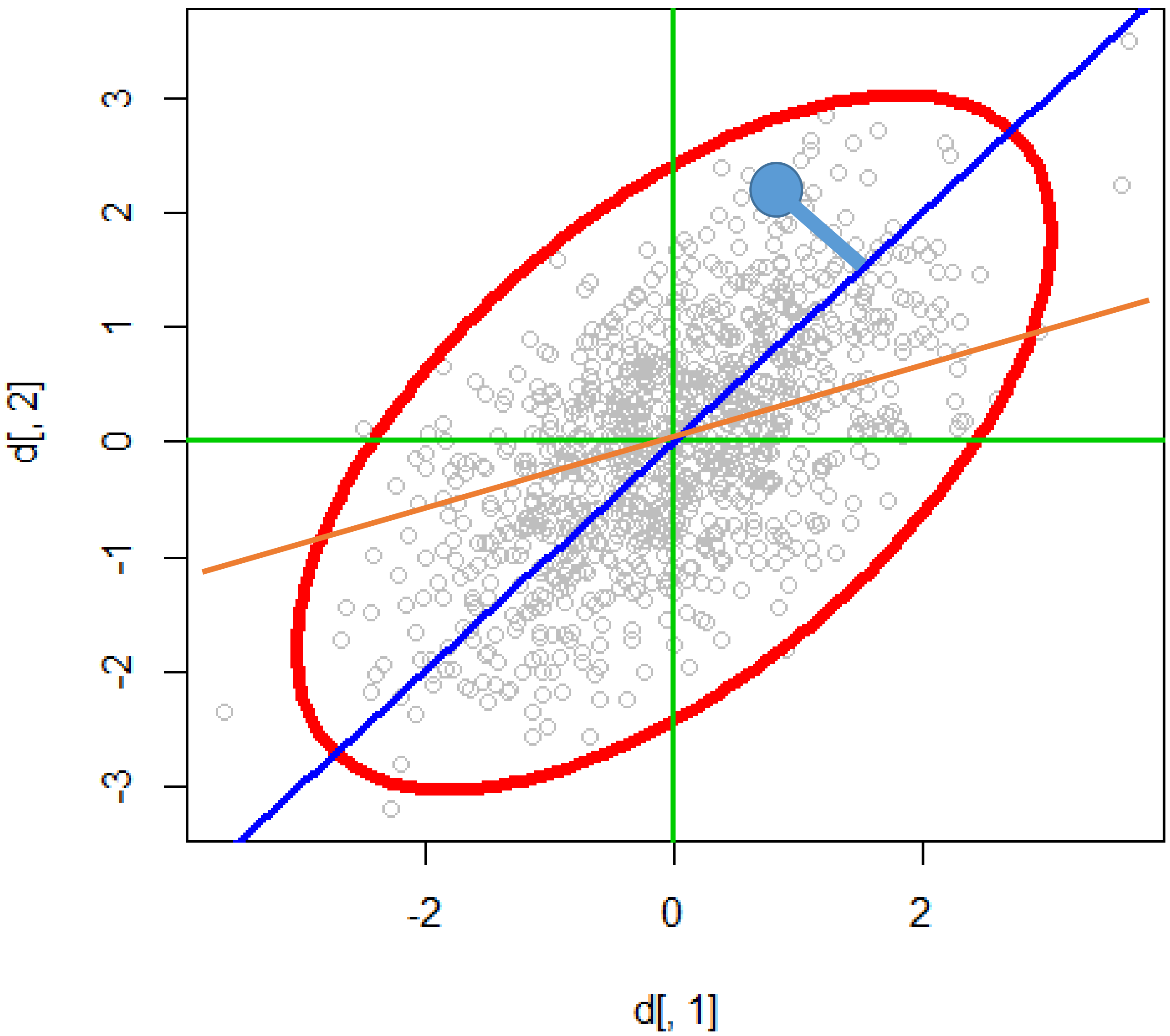
the correlation is about .60



Blue: 1st princpal compoent

the blue line draw through the ellips is special

why?



if you know the coordinates of the blue dot
(the X and Y values on the green dimensions)

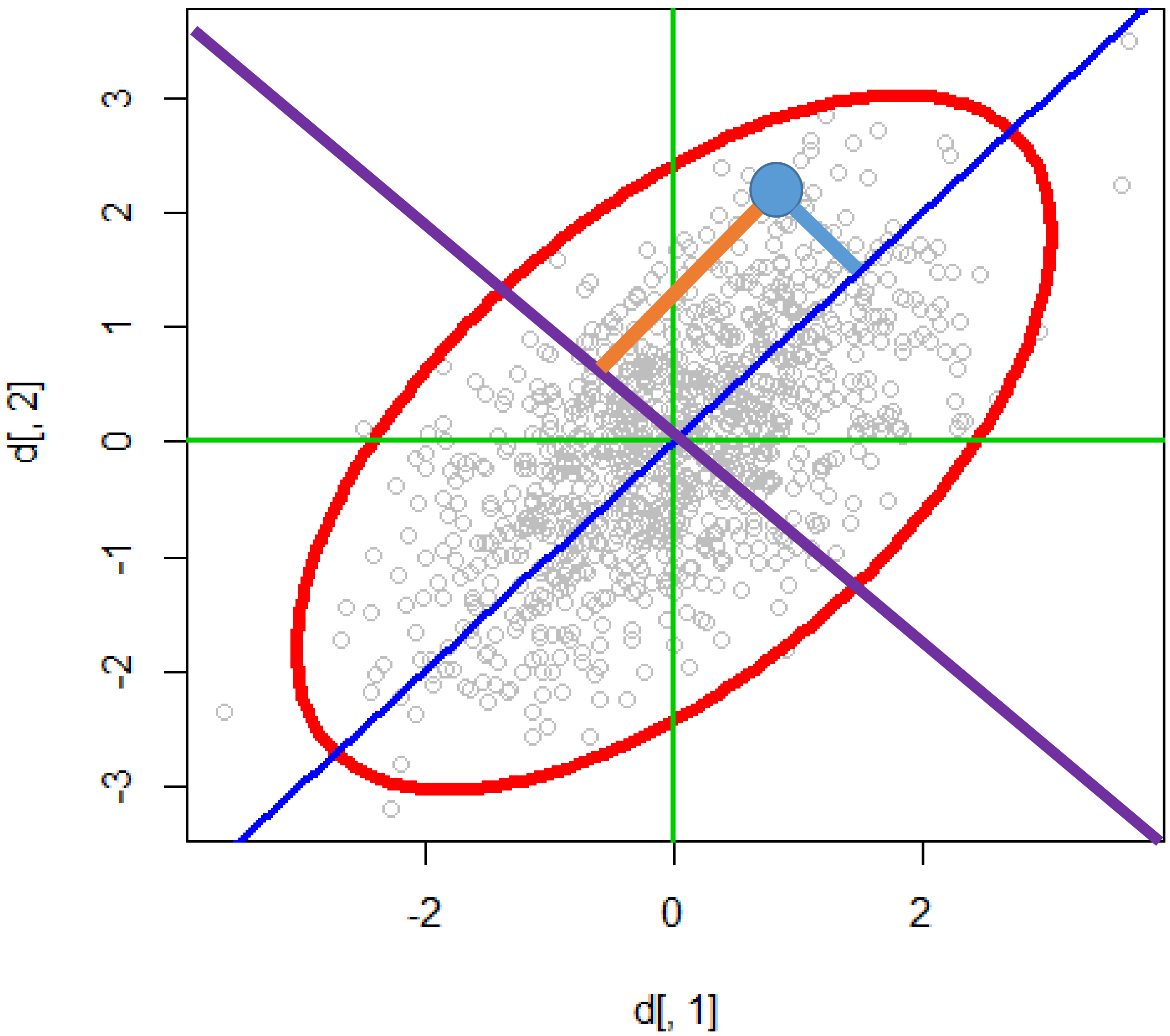
you can calculate the value on the blue dimension.
“project on to the blue dimension”

the variance of the projected values: var(p)

the blue line is chosen such that var(p) is maximal

you can project on the orange line, but the variance of the projected values will be smaller.

var(p) = the 1st eigenvalue



second line purple is perpendicular to the blue line

variance of the projections on the purple line

is the 2nd eigenvalue.

The eigenvalues of a covariance matrix should be positive. If so the matrix is called positive definite.

The eigen values of a 2x2 correlation matrix (r=.6) in R

```
R1=matrix(.6,2,2)
diag(R1)=1
evals=eigen(R!)$values
print(evals)
```

The eigen values of a 2x2 correlation matrix (r=.6) in R

```
#start
R1=matrix(.6,2,2)
diag(R1)=1
evals=eigen(R1)$values
print(evals)
# end
```

	x	y
x	1	.6
y	.6	1

[1] 1.6 0.4

Both positive, the matrix is positive definite!

What about this correlation matrix

1	0.75	0.10
0.75	1	0.75
0.10	0.75	1

```
R1=matrix(c(1,.75,.1,.75,1,.75,.1,.75,1),3,3,byrow=T)
evals=eigen(R1)$values
```

the matrix is not positive definite!