Phenotypic Factor Analysis

Marleen de Moor & Meike Bartels

Department of Biological Psychology, VU University Amsterdam

mhm.de.moor@psy.vu.nl / m.bartels@psy.vu.nl

Outline

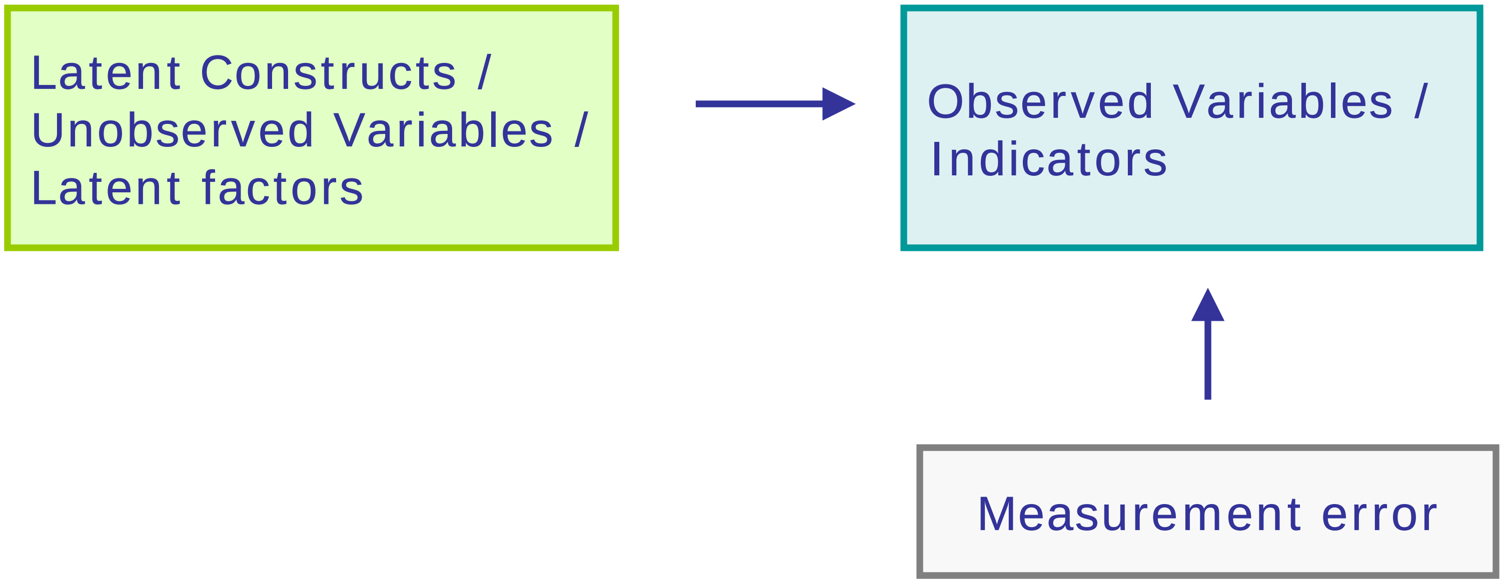
- Introduction to factor analysis
  - What is factor analysis
  - Relationship with regression and SEM
  - Types of factor analysis
- Phenotypic factor analysis
  - 1 factor model
  - 2 factor model
- More advanced models
  - Factor models for categorical data
  - Multigroup factor models and measurement invariance
- From phenotypic to genetic factor analysis...

Outline

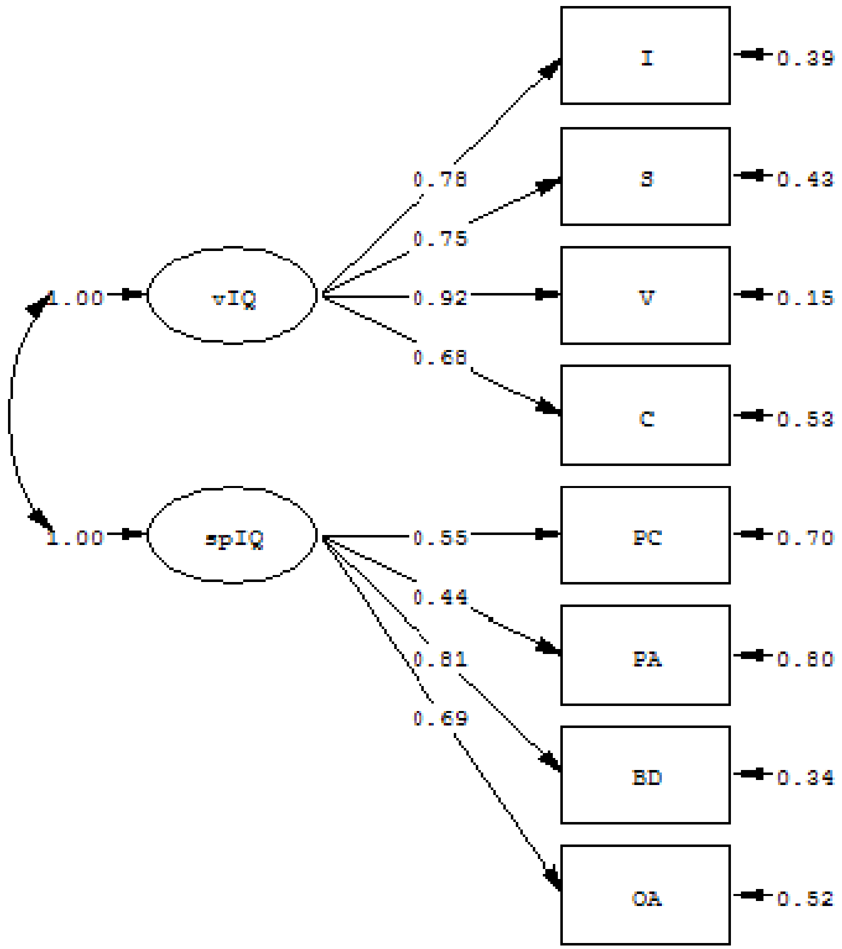
- Introduction to factor analysis
  - What is factor analysis
  - Relationship with regression and SEM
  - Types of factor analysis
- Phenotypic factor analysis
- More advanced models
- From phenotypic to genetic factor analysis...

Factor analysis

- Collection of methods
- Measurement model
- Describe/explain pattern of observed correlations

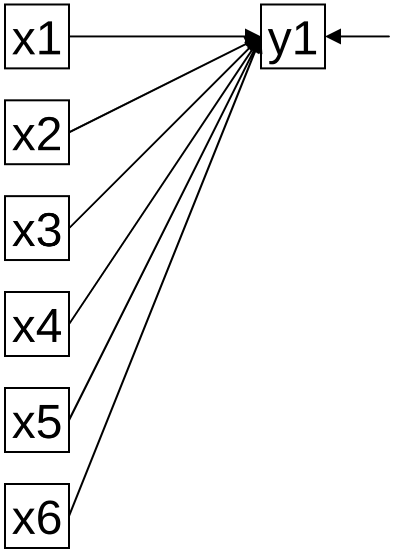


Classic example: IQ

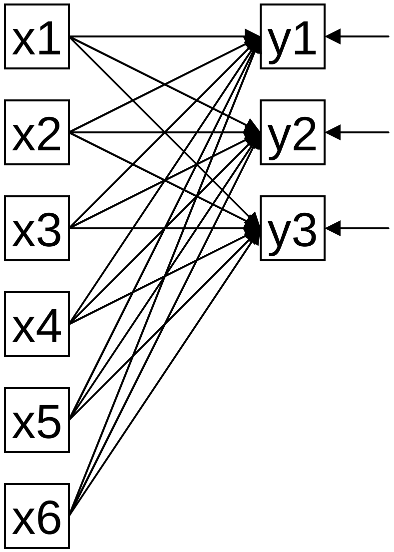


Relationship with regression analysis

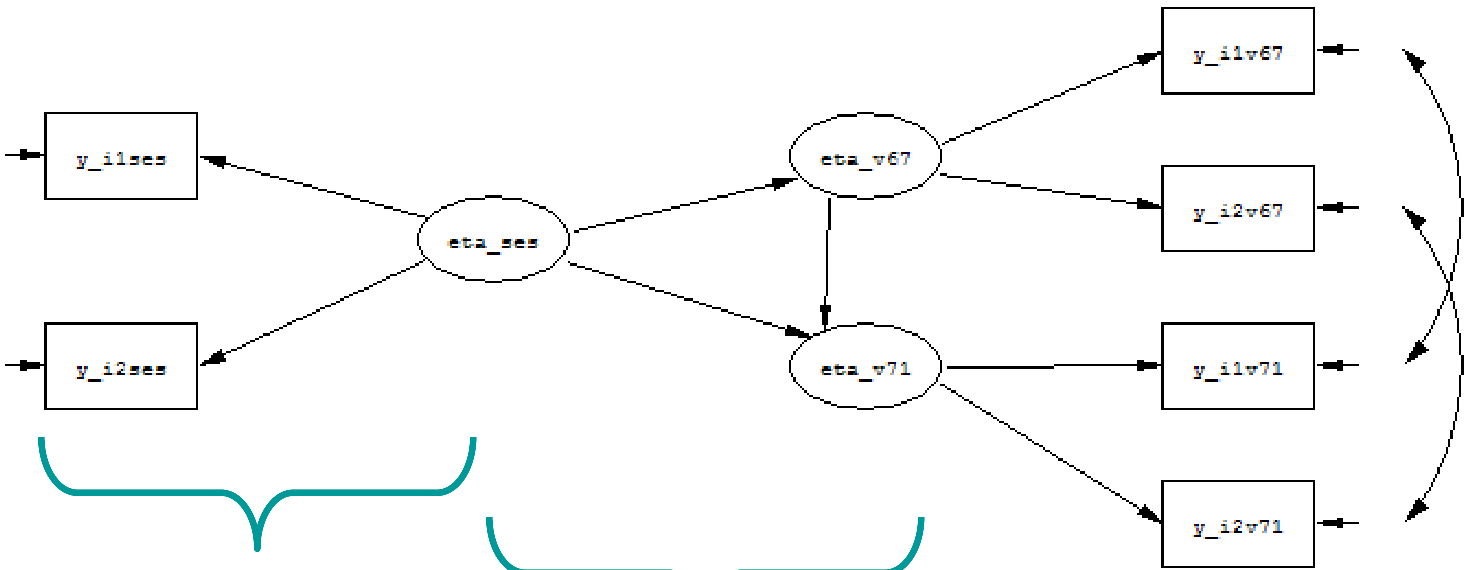
Multiple regression:



Multivariate multiple regression:



Relationship with SEM



Measurement model

(latent variables measured by observed variables)

Structural model

(regression model among latent variables)

Measurement model

(latent variables measured by observed variables)

Types of factor analysis

- Principal component analysis (PCA)
- Exploratory factor analysis (EFA)
- Confirmatory factor analysis (CFA)

PCA

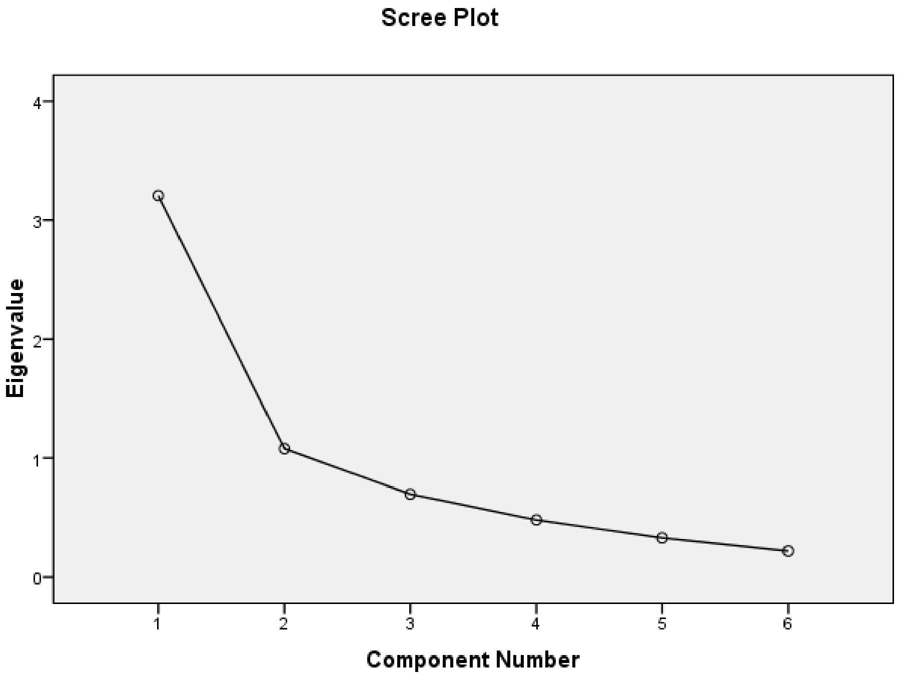
- Data reduction technique
- Linear transformation of the data
- Summarize the observed pattern of correlations among variables with a smaller number of principal components

PCA

- First component explains as much variance as possible
- Different rotations possible: orthogonal or oblique
- Principal components contain both common and residual variance!

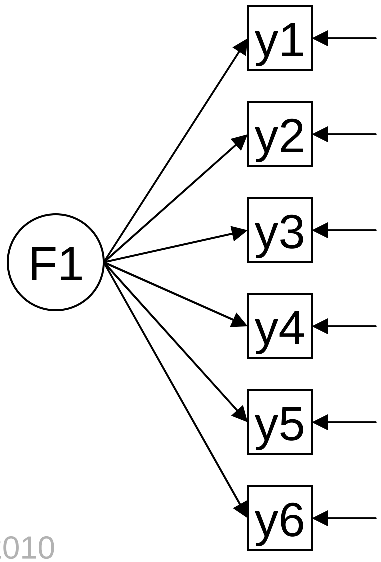
Adolescent data on:

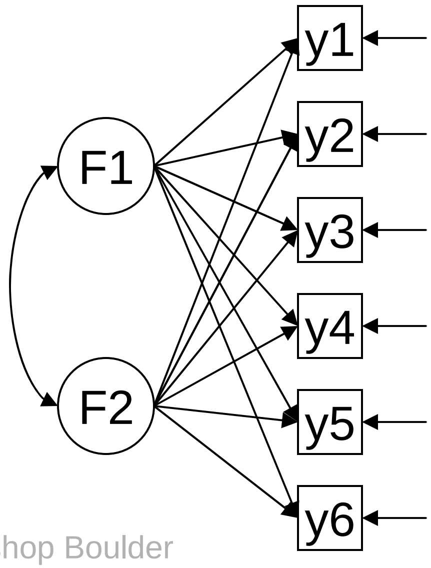
Quality of life	Anxious depression
Happiness	Somatic complaints
Life satisfaction	Social problems



EFA

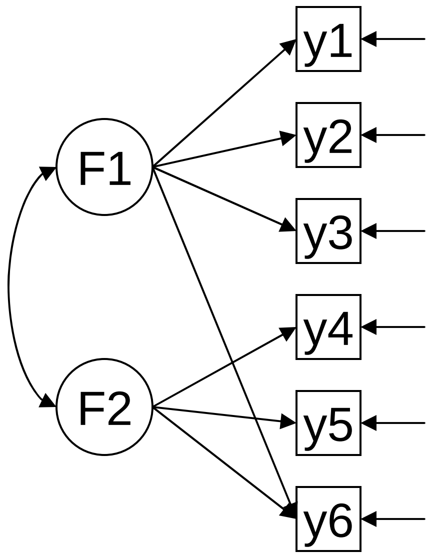
- Atheoretical
- ***Discover*** the underlying constructs
- Determine number of latent factors
- Again, different rotations possible





CFA

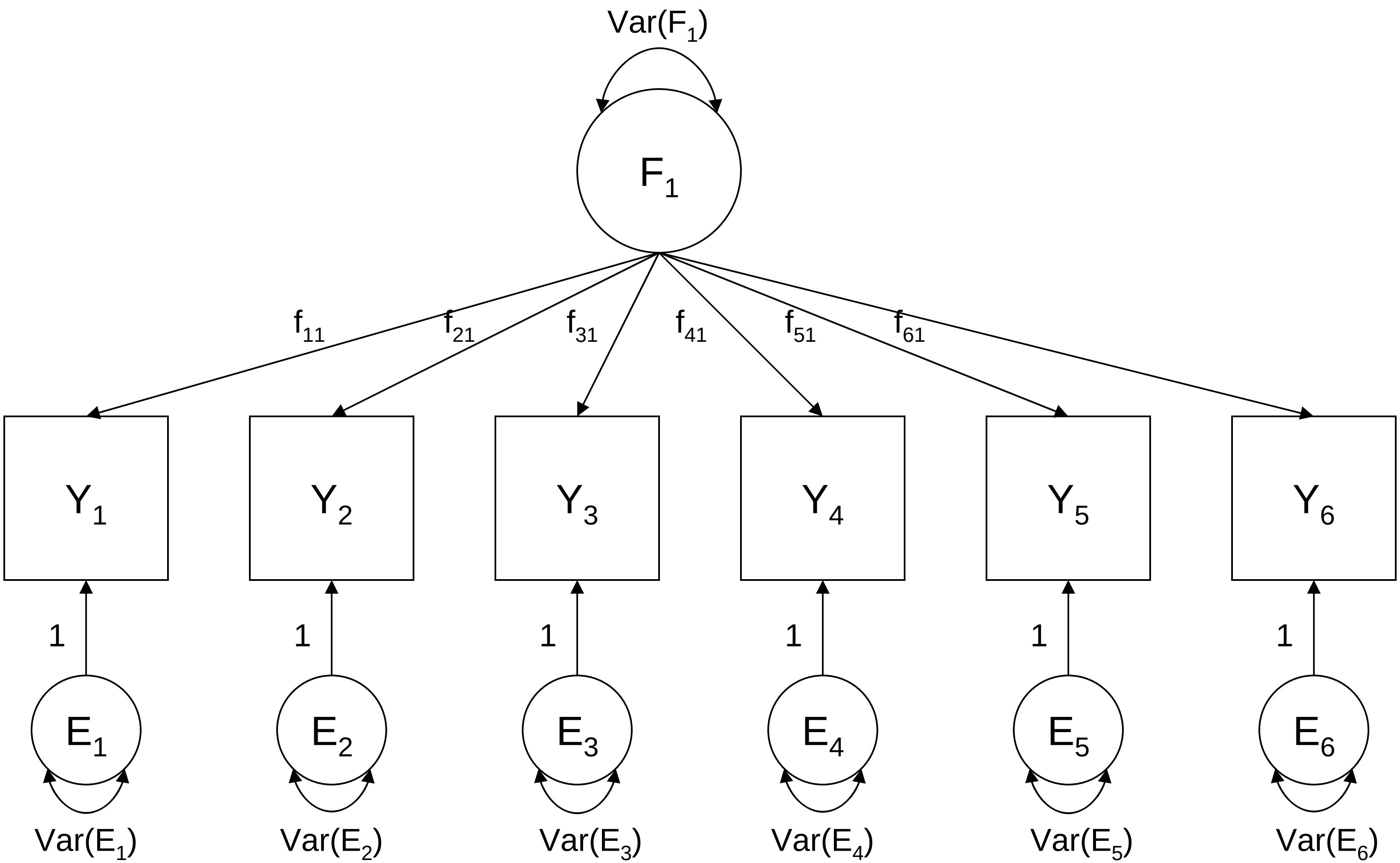
- Theoretical (model=hypothesis)
- ***Test*** hypothesis about underlying constructs



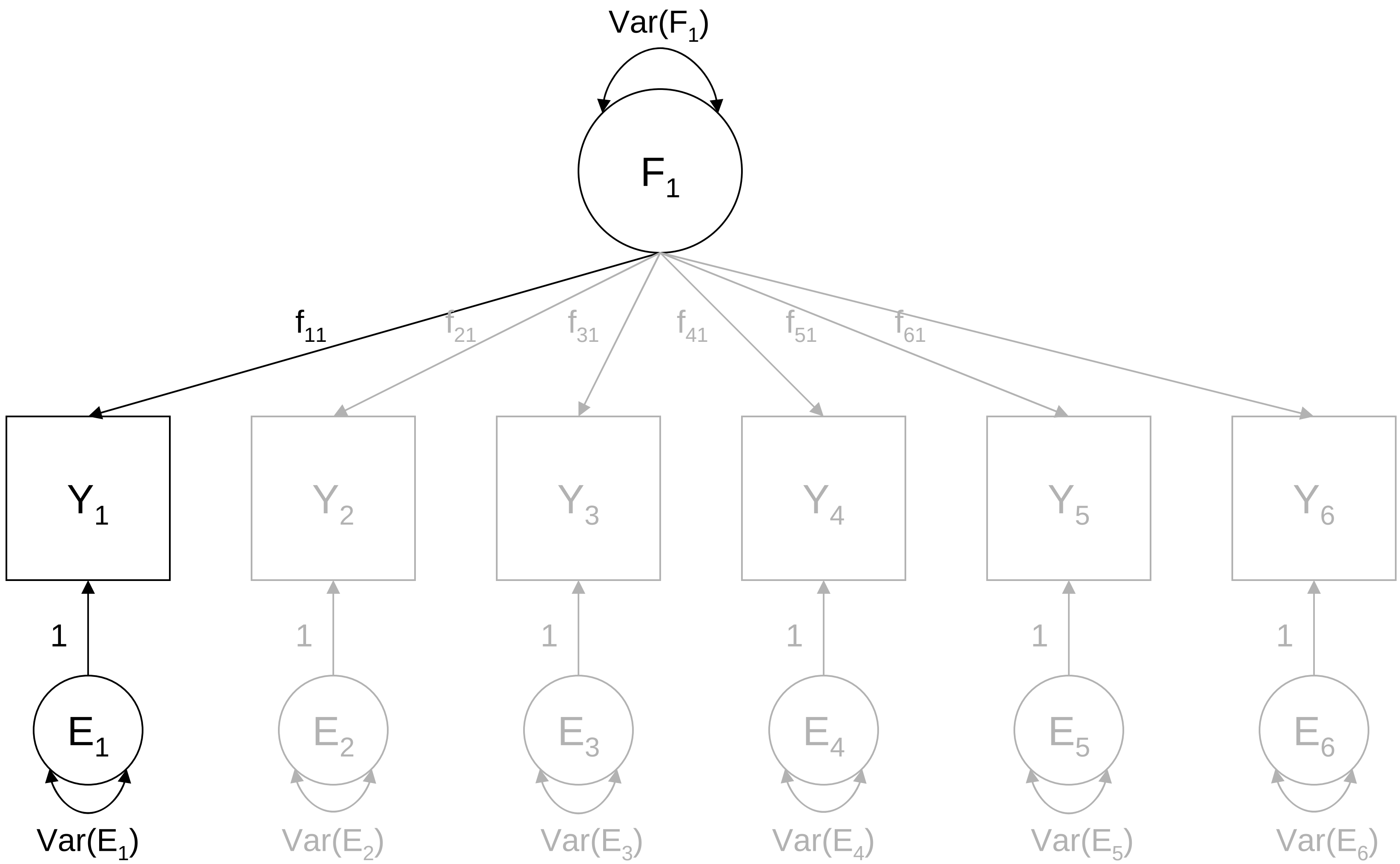
Outline

- Introduction to factor analysis
- Phenotypic factor analysis
  - 1 factor model
  - 2 factor model
- More advanced models
- From phenotypic to genetic factor analysis...

The 1 factor model



Y is influenced by F and E



More formally...

- $Y_{i1} = f_{11} * F_{i1} + E_{i1}$

Random variables (varies across individuals i=1...N)

More formally...

- $Y_{i1} = f_{11} * F_{i1} + E_{i1}$

Random variables (varies across individuals i=1...N)

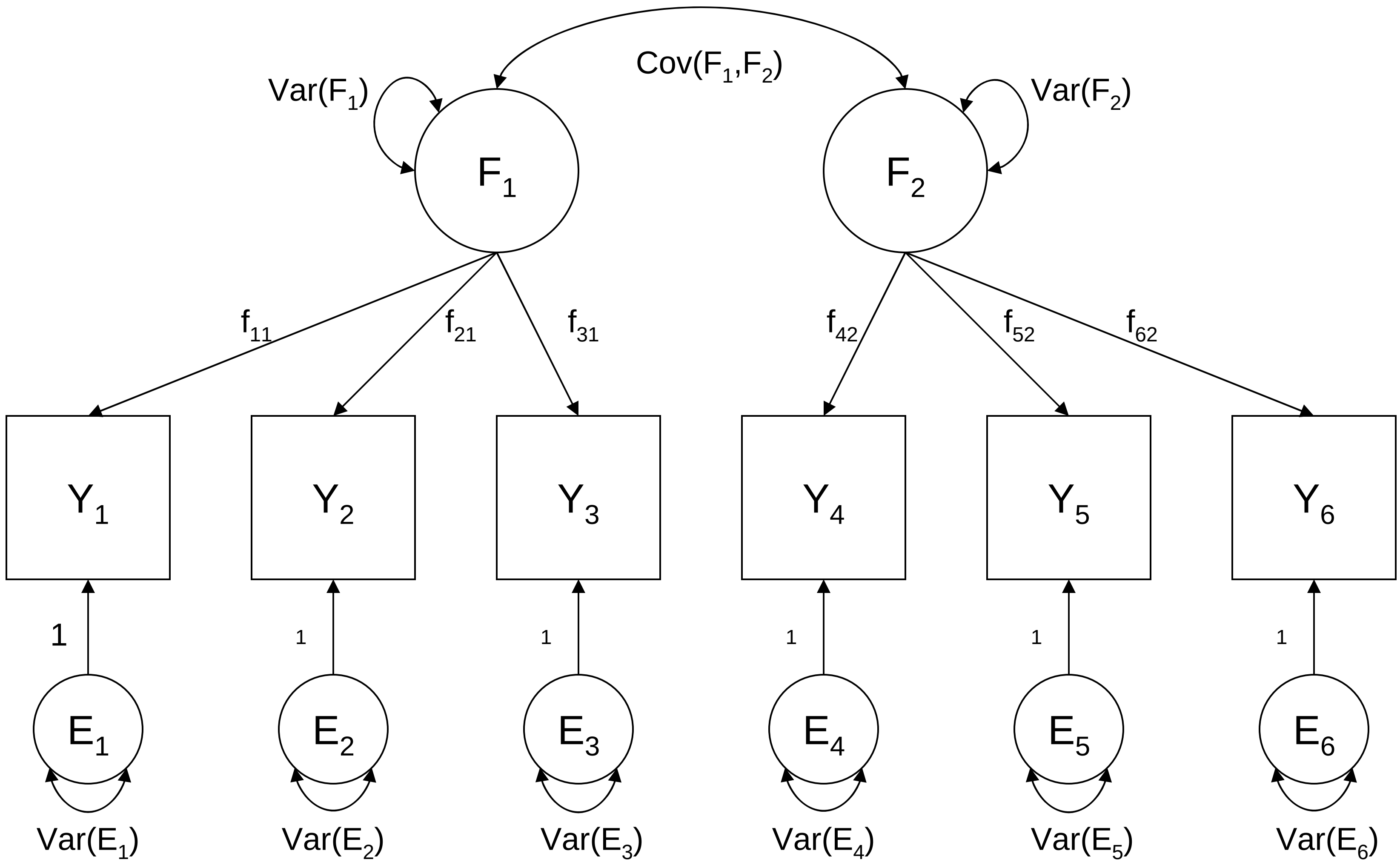
Fixed parameter (constant across individuals)

More formally...

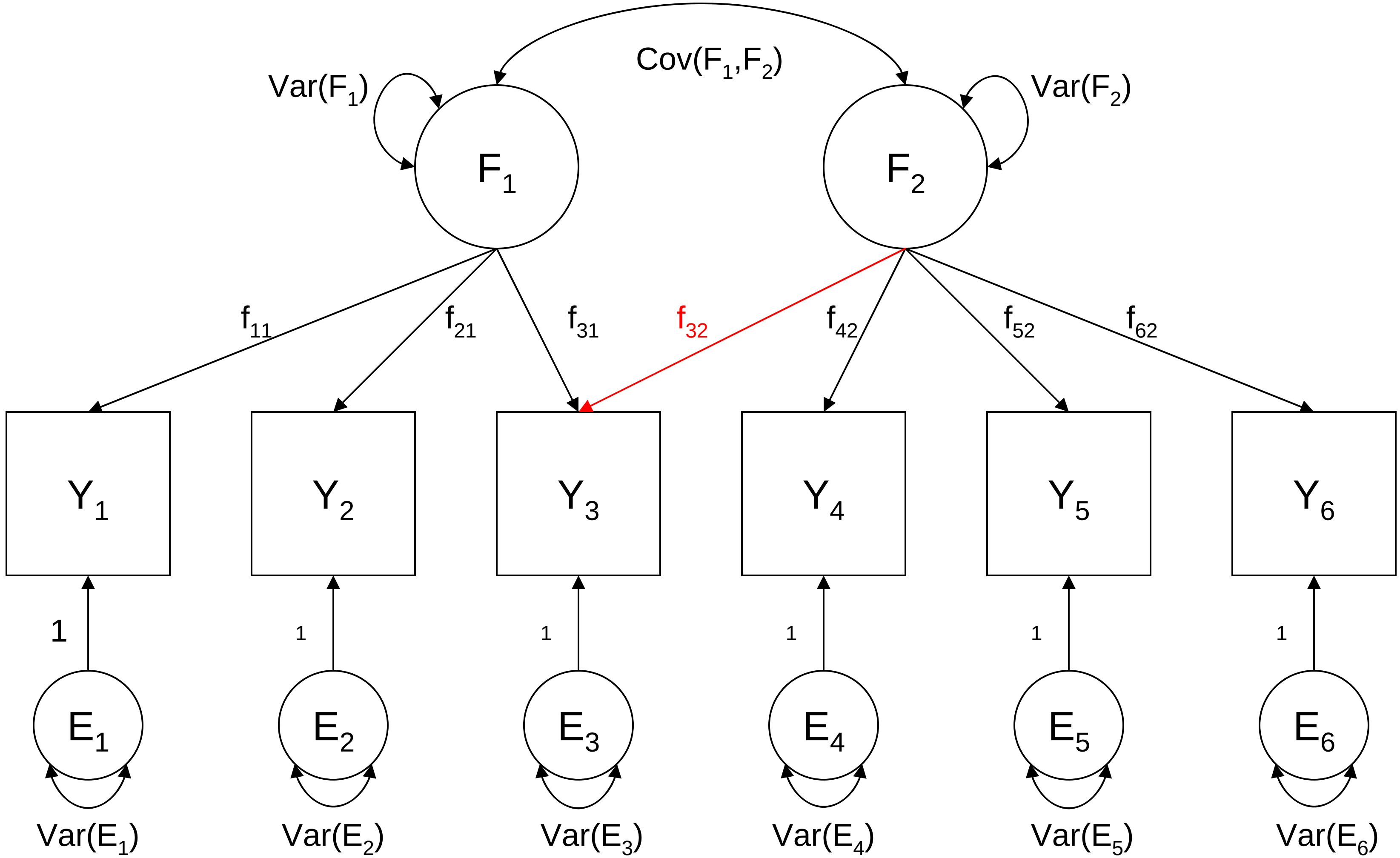
- $Y_{i1} = f_{11} * F_{i1} + E_{i1}$
- $Y_{i2} = f_{21} * F_{i1} + E_{i2}$
- $Y_{i3} = f_{31} * F_{i1} + E_{i3}$

  ...
- $Y_{i6} = f_{61} * F_{i1} + E_{i6}$

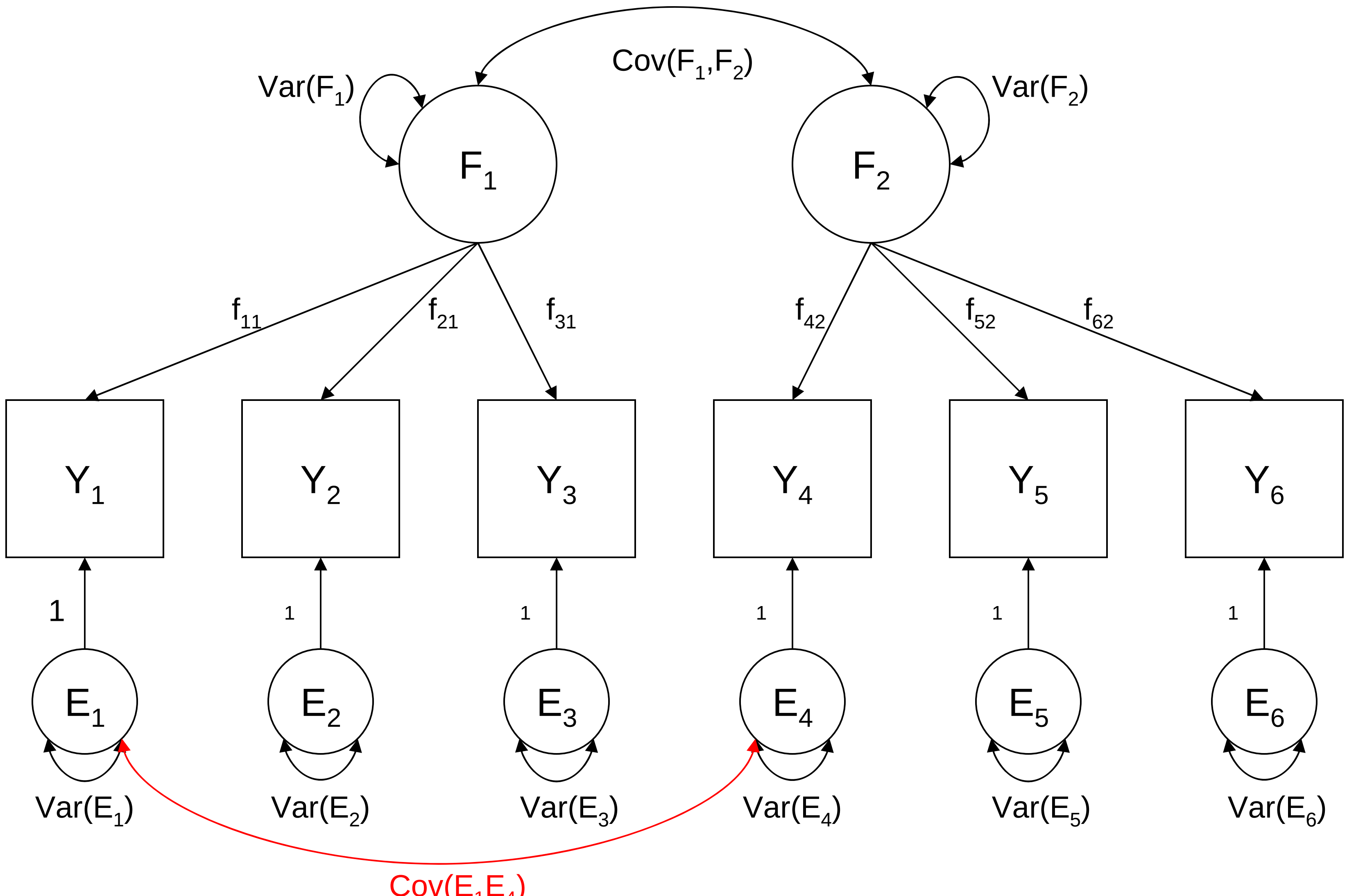
The 2 factor model



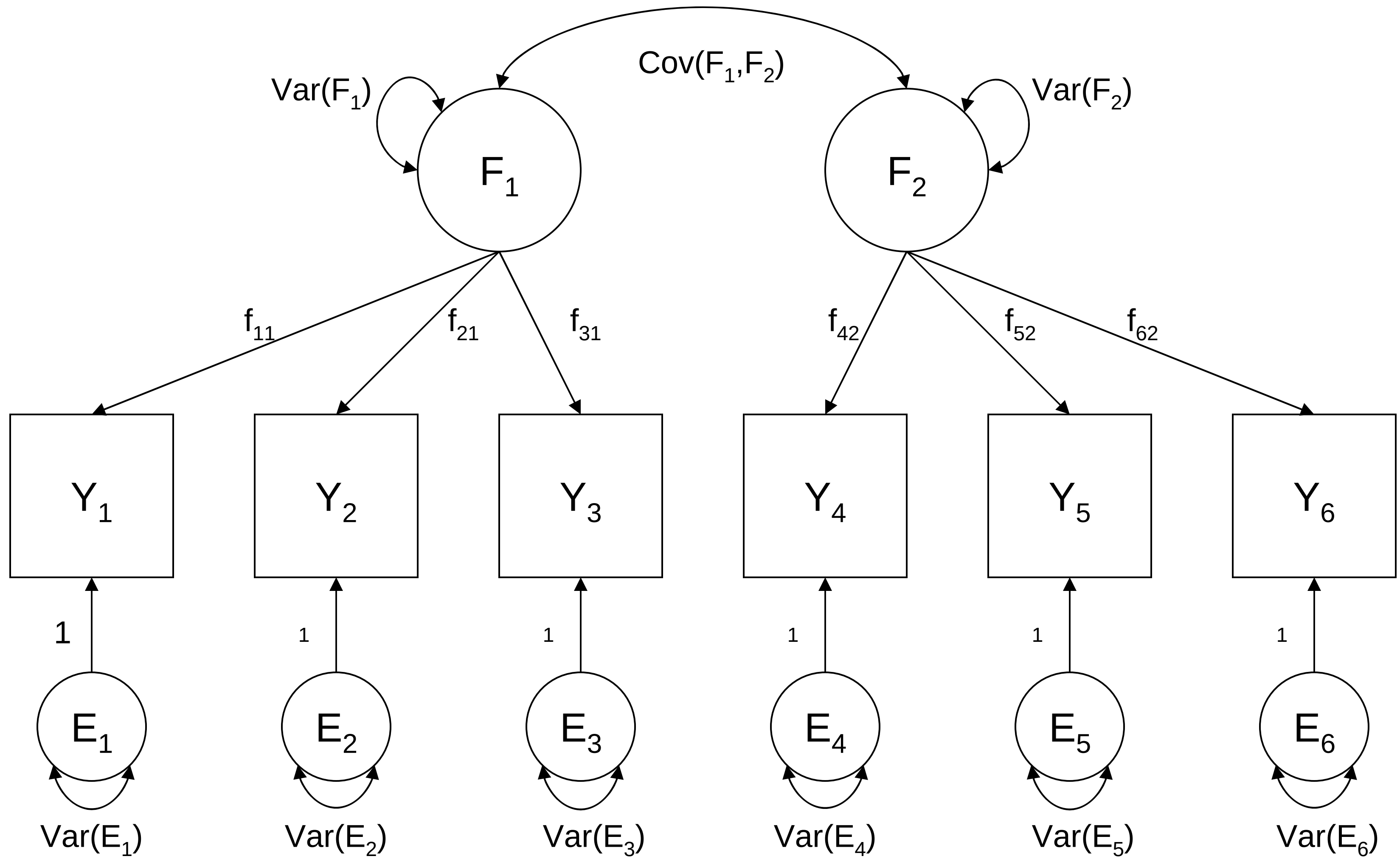
The 2 factor model



The 2 factor model



The 2 factor model



From equations to matrices

- $Y_{i1} = f_{11} * F_{i1} + E_{i1}$
- $Y_{i2} = f_{21} * F_{i1} + E_{i2}$
- $Y_{i3} = f_{31} * F_{i1} + E_{i3}$
- $Y_{i4} = f_{42} * F_{i2} + E_{i4}$
- $Y_{i5} = f_{52} * F_{i2} + E_{i5}$
- $Y_{i6} = f_{62} * F_{i2} + E_{i6}$

$$\begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} \\ Y_{i5} \\ Y_{i6} \end{bmatrix} = \begin{bmatrix} f_{11} & 0 \\ f_{21} & 0 \\ f_{31} & 0 \\ 0 & f_{41} \\ 0 & f_{51} \\ 0 & f_{61} \end{bmatrix} * \begin{bmatrix} F_{i1} \\ F_{i2} \end{bmatrix} + \begin{bmatrix} E_{i1} \\ E_{i2} \\ E_{i3} \\ E_{i4} \\ E_{i5} \\ E_{i6} \end{bmatrix}$$

i=1...N number of individuals
j=1...J number of observed variables
k=1...K number of factors

Assumption: Data follow a multivariate normal distribution

Expected (co)variances

Can be obtained in 2 ways:

- Using path diagram (Wright's rules)
- Using equations (algebraic derivation)

Expected (co)variances – path diagram

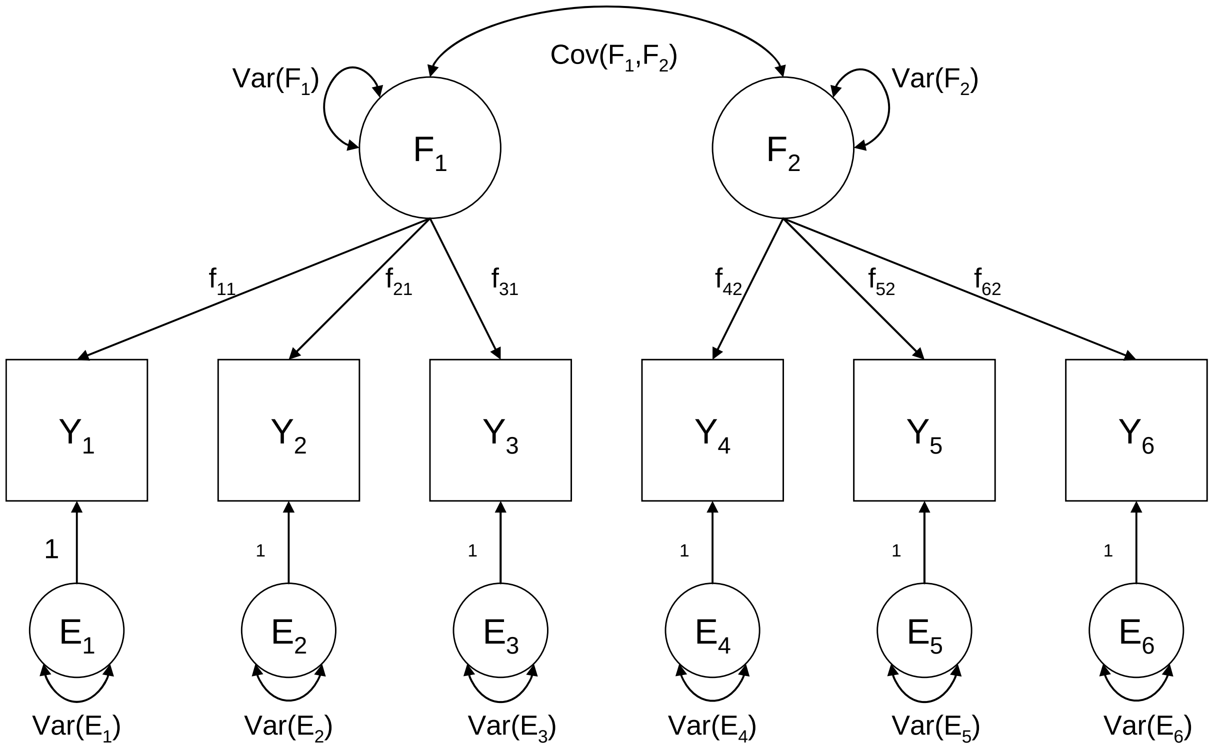
EXERCISE:

Write down the expectations for:

$Var(Y_1)=??$

$Cov(Y_1,Y_2)=??$

$Cov(Y_1,Y_4)=??$



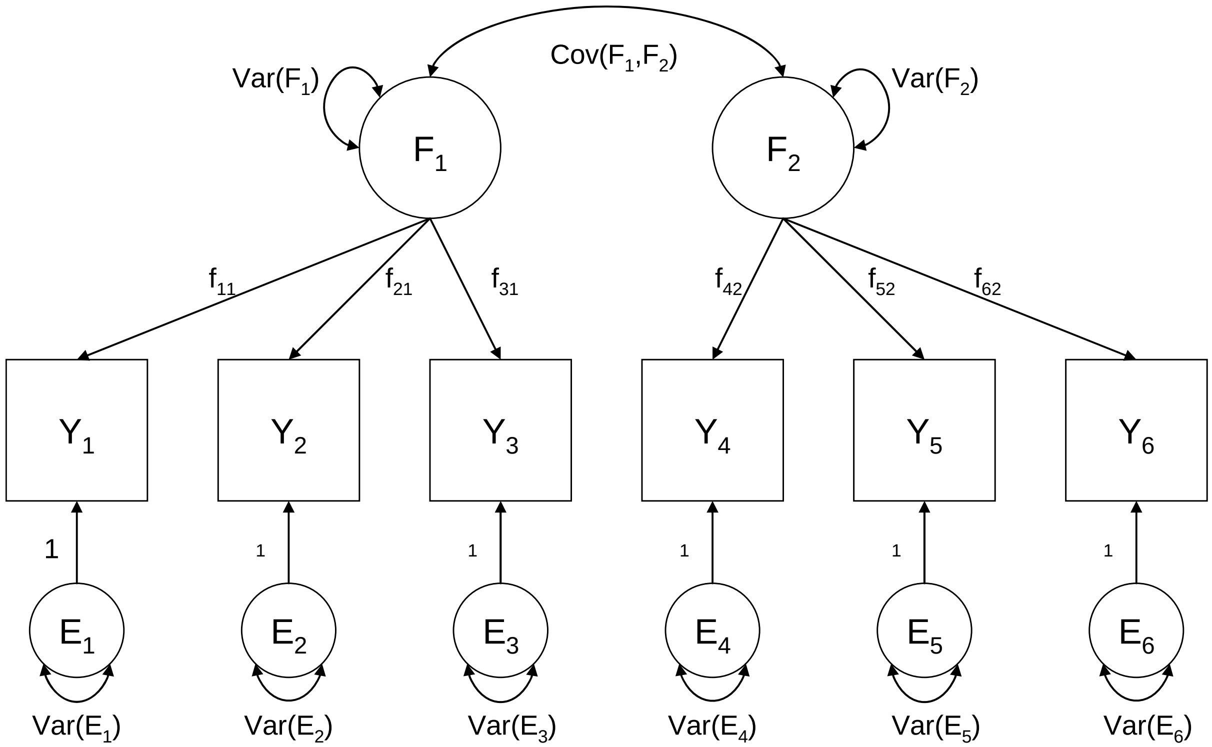
Expected (co)variances – path diagram

ANSWER:

$Var(Y_1) = f_{11}^2 * var(F_1) + var(E_1)$

$Cov(Y_1, Y_2) = ??$

$Cov(Y_1, Y_4) = ??$



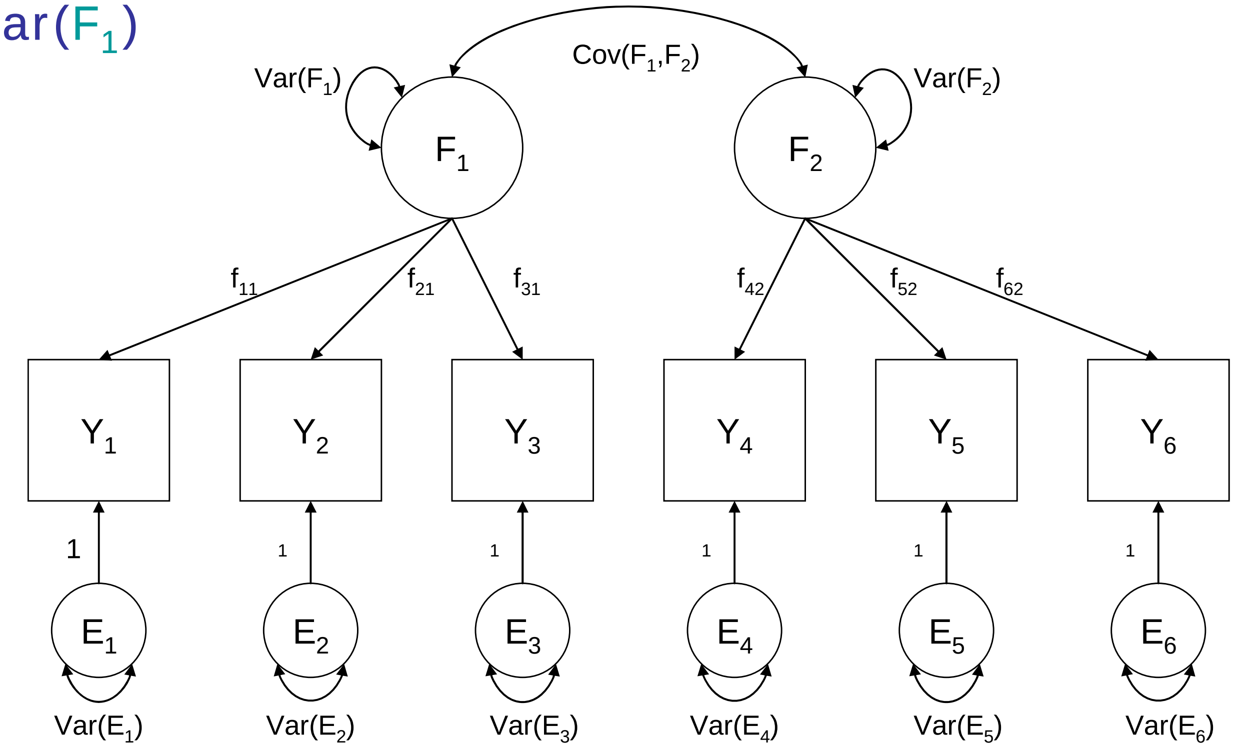
Expected (co)variances – path diagram

ANSWER:

$Var(Y_1) = f_{11}^2 * var(F_1) + var(E_1)$

$Cov(Y_1,Y_2) = f_{11} * f_{21} * var(F_1)$

$Cov(Y_1,Y_4) =$??



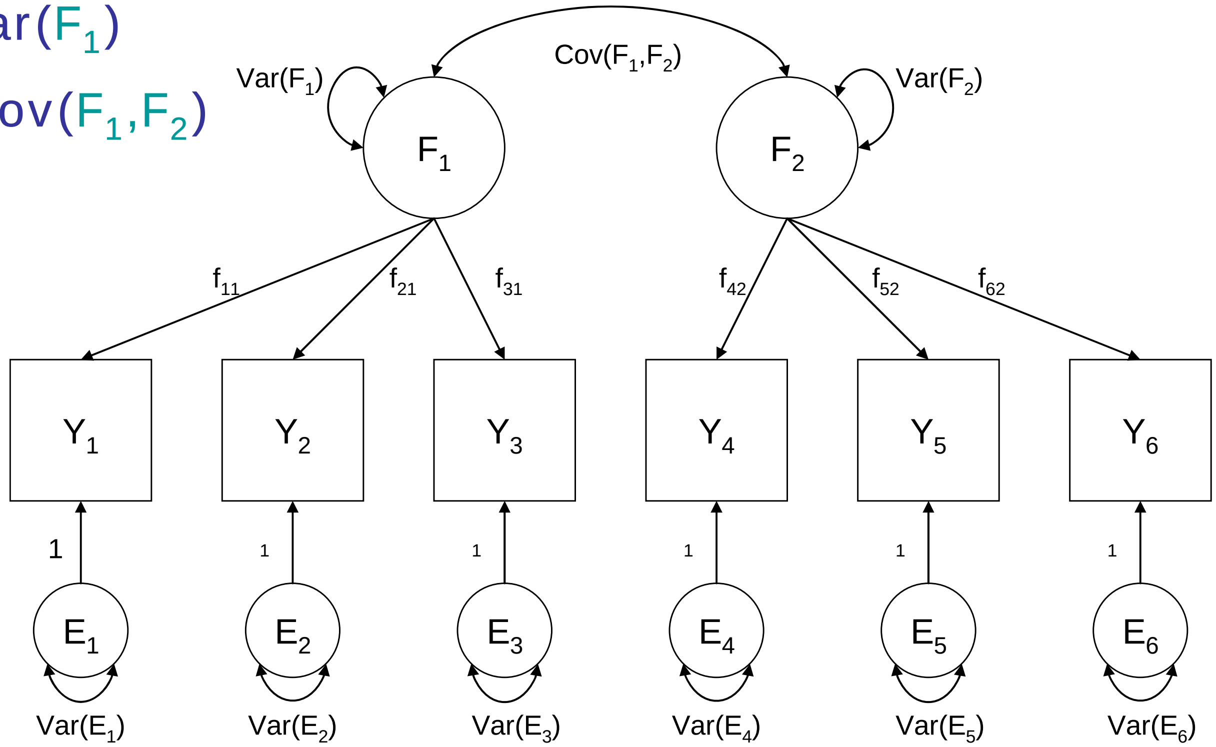
Expected (co)variances – path diagram

ANSWER:

$Var(Y_1) = f_{11}^2 * var(F_1) + var(E_1)$

$Cov(Y_1,Y_2) = f_{11} * f_{21} * var(F_1)$

$Cov(Y_1,Y_4) = f_{11} * f_{42} * cov(F_1,F_2)$



Expected (co)variances - equations

$$\text{Var}(Y_1) = E[(f_{11}*F_{i1}+E_{i1})*(f_{11}*F_{i1}+E_{i1})]$$
$$= E[(f_{11}*F_{i1})^2 + 2*f_{11}*F_{i1}*E_{i1} + (E_{i1})^2]$$
$$= f_{11}^2 * \text{var}(F_1) + \text{var}(E_1)$$

$$\text{Cov}(Y_1,Y_2) = E[(f_{11}*F_{i1}+E_{i1})*(f_{21}*F_{i1}+E_{i2})]$$
$$= E[f_{11}*F_{i1}*f_{21}*F_{i1} + f_{11}*F_{i1}*E_{i2} + E_{i1}*f_{21}*F_{i1} + E_{i1}*E_{i2}]$$
$$= f_{11}*f_{21} * \text{var}(F_1)$$

$$\text{Cov}(Y_1,Y_4) = f_{11}*f_{42} * \text{cov}(F_1,F_2)$$

Expected (co)variances - equations

$$\begin{bmatrix} var(Y_1) & cov(Y_1Y_2) & cov(Y_1Y_3) & cov(Y_1Y_4) & cov(Y_1Y_5) & cov(Y_1Y_6) \\ cov(Y_2Y_1) & var(Y_2) & cov(Y_2Y_3) & cov(Y_2Y_4) & cov(Y_2Y_5) & cov(Y_2Y_6) \\ cov(Y_3Y_1) & cov(Y_3Y_2) & var(Y_3) & cov(Y_3Y_4) & cov(Y_3Y_5) & cov(Y_3Y_6) \\ cov(Y_4Y_1) & cov(Y_4Y_2) & cov(Y_4Y_3) & var(Y_4) & cov(Y_4Y_5) & cov(Y_4Y_6) \\ cov(Y_5Y_1) & cov(Y_5Y_2) & cov(Y_5Y_3) & cov(Y_5Y_4) & var(Y_5) & cov(Y_5Y_6) \\ cov(Y_6Y_1) & cov(Y_6Y_2) & cov(Y_6Y_3) & cov(Y_6Y_4) & cov(Y_6Y_5) & var(Y_6) \end{bmatrix} =$$

j=1...J number of observed variables
k=1...K number of factors

JxJ symmetric

$$\begin{bmatrix} f_{11} & 0 \\ f_{21} & 0 \\ f_{31} & 0 \\ 0 & f_{42} \\ 0 & f_{52} \\ 0 & f_{62} \end{bmatrix} * \begin{bmatrix} var(F_1) & cov(F_1,F_2) \\ cov(F_1,F_2) & var(F_2) \end{bmatrix} * \begin{bmatrix} f_{11} & f_{21} & f_{31} & 0 & 0 & 0 \\ 0 & 0 & 0 & f_{42} & f_{52} & f_{62} \end{bmatrix} + \begin{bmatrix} var(E_1) & 0 & 0 & 0 & 0 & 0 \\ 0 & var(E_2) & 0 & 0 & 0 & 0 \\ 0 & 0 & var(E_3) & 0 & 0 & 0 \\ 0 & 0 & 0 & var(E_4) & 0 & 0 \\ 0 & 0 & 0 & 0 & var(E_5) & 0 \\ 0 & 0 & 0 & 0 & 0 & var(E_6) \end{bmatrix}$$

JxK full *KxK symm* *KxJ full* *JxJ diag*

$$\Sigma = \Lambda\Psi\Lambda' + \Theta$$

(LISREL notation)

$$\exp Cov = L\%*\%P\%*\%t(L) + T$$

(OpenMx notation)

Identification

- Latent factors have no scale: means, variances?

For each latent factor:

- Mean: fix to zero
- Variance: two most commonly used options
  - Fix to one, estimate all factor loadings
  - Estimate variance, fix first factor loading to one

Identification of (co)variances

$$\begin{bmatrix} \mathrm{var}(Y_1) & \mathrm{cov}(Y_1Y_2) & \mathrm{cov}(Y_1Y_3) & \mathrm{cov}(Y_1Y_4) & \mathrm{cov}(Y_1Y_5) & \mathrm{cov}(Y_1Y_6) \\ \mathrm{cov}(Y_2Y_1) & \mathrm{var}(Y_2) & \mathrm{cov}(Y_2Y_3) & \mathrm{cov}(Y_2Y_4) & \mathrm{cov}(Y_2Y_5) & \mathrm{cov}(Y_2Y_6) \\ \mathrm{cov}(Y_3Y_1) & \mathrm{cov}(Y_3Y_2) & \mathrm{var}(Y_3) & \mathrm{cov}(Y_3Y_4) & \mathrm{cov}(Y_3Y_5) & \mathrm{cov}(Y_3Y_6) \\ \mathrm{cov}(Y_4Y_1) & \mathrm{cov}(Y_4Y_2) & \mathrm{cov}(Y_4Y_3) & \mathrm{var}(Y_4) & \mathrm{cov}(Y_4Y_5) & \mathrm{cov}(Y_4Y_6) \\ \mathrm{cov}(Y_5Y_1) & \mathrm{cov}(Y_5Y_2) & \mathrm{cov}(Y_5Y_3) & \mathrm{cov}(Y_5Y_4) & \mathrm{var}(Y_5) & \mathrm{cov}(Y_5Y_6) \\ \mathrm{cov}(Y_6Y_1) & \mathrm{cov}(Y_6Y_2) & \mathrm{cov}(Y_6Y_3) & \mathrm{cov}(Y_6Y_4) & \mathrm{cov}(Y_6Y_5) & \mathrm{var}(Y_6) \end{bmatrix} =$$

$$\begin{bmatrix} f_{11} & 0 \\ f_{21} & 0 \\ f_{31} & 0 \\ 0 & f_{42} \\ 0 & f_{52} \\ 0 & f_{62} \end{bmatrix} * \begin{bmatrix} 1 & \mathrm{cov}(F_1, F_2) \\ \mathrm{cov}(F_1, F_2) & 1 \end{bmatrix} * \begin{bmatrix} f_{11} & f_{21} & f_{31} & 0 & 0 & 0 \\ 0 & 0 & 0 & f_{42} & f_{52} & f_{62} \end{bmatrix} + \begin{bmatrix} \mathrm{var}(E_1) & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathrm{var}(E_2) & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathrm{var}(E_3) & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathrm{var}(E_4) & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathrm{var}(E_5) & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathrm{var}(E_6) \end{bmatrix}$$

Identification of (co)variances

$$\begin{bmatrix} \text{var}(Y_1) & \text{cov}(Y_1Y_2) & \text{cov}(Y_1Y_3) & \text{cov}(Y_1Y_4) & \text{cov}(Y_1Y_5) & \text{cov}(Y_1Y_6) \\ \text{cov}(Y_2Y_1) & \text{var}(Y_2) & \text{cov}(Y_2Y_3) & \text{cov}(Y_2Y_4) & \text{cov}(Y_2Y_5) & \text{cov}(Y_2Y_6) \\ \text{cov}(Y_3Y_1) & \text{cov}(Y_3Y_2) & \text{var}(Y_3) & \text{cov}(Y_3Y_4) & \text{cov}(Y_3Y_5) & \text{cov}(Y_3Y_6) \\ \text{cov}(Y_4Y_1) & \text{cov}(Y_4Y_2) & \text{cov}(Y_4Y_3) & \text{var}(Y_4) & \text{cov}(Y_4Y_5) & \text{cov}(Y_4Y_6) \\ \text{cov}(Y_5Y_1) & \text{cov}(Y_5Y_2) & \text{cov}(Y_5Y_3) & \text{cov}(Y_5Y_4) & \text{var}(Y_5) & \text{cov}(Y_5Y_6) \\ \text{cov}(Y_6Y_1) & \text{cov}(Y_6Y_2) & \text{cov}(Y_6Y_3) & \text{cov}(Y_6Y_4) & \text{cov}(Y_6Y_5) & \text{var}(Y_6) \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 0 \\ f_{21} & 0 \\ f_{31} & 0 \\ 0 & 1 \\ 0 & f_{52} \\ 0 & f_{62} \end{bmatrix} * \begin{bmatrix} \text{var}(F_1) & \text{cov}(F_1,F_2) \\ \text{cov}(F_1,F_2) & \text{var}(F_2) \end{bmatrix} * \begin{bmatrix} 1 & f_{21} & f_{31} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & f_{52} & f_{62} \end{bmatrix} + \begin{bmatrix} \text{var}(E_1) & 0 & 0 & 0 & 0 & 0 \\ 0 & \text{var}(E_2) & 0 & 0 & 0 & 0 \\ 0 & 0 & \text{var}(E_3) & 0 & 0 & 0 \\ 0 & 0 & 0 & \text{var}(E_4) & 0 & 0 \\ 0 & 0 & 0 & 0 & \text{var}(E_5) & 0 \\ 0 & 0 & 0 & 0 & 0 & \text{var}(E_6) \end{bmatrix}$$

Identification

- Count number of observed statistics
- Count number of free parameters

- If #obs. stat. < #free par. → Model unidentified
- If #obs. stat. = #free par. → Model justidentified
- If #obs. stat. > #free par. → Model identified

Identification of 1 factor model

Observed statistics:

#obs. var. = J = 6

#obs. cov. = J(J-1)/2 = 6*5/2 = 15

#obs. var/cov. = J(J+1)/2 = 6*7/2 = 21

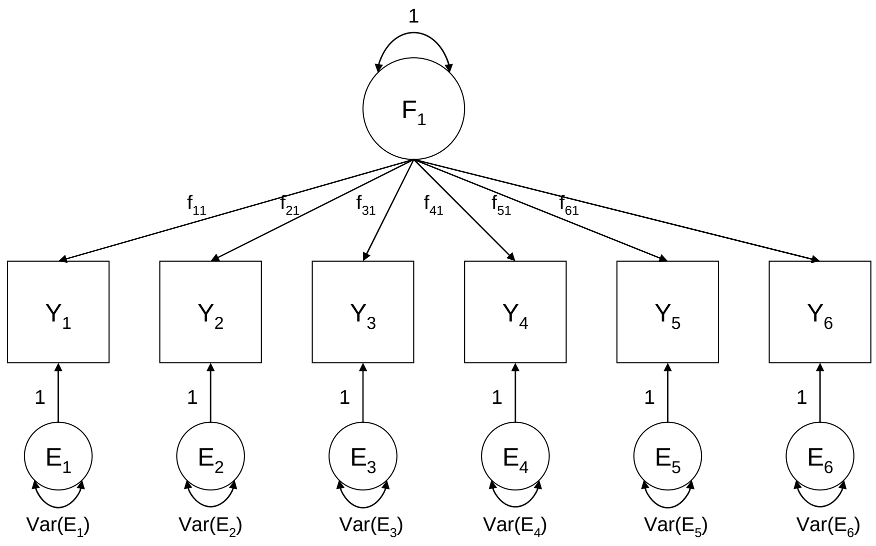
Free parameters:

residual variances = 6

factor loadings = 6

Degrees of freedom:

df = 21-12 = 9



Identification of 2 factor model

Observed statistics:

#obs. var. = J = 6

#obs. cov. = J(J-1)/2 = 6*5/2 = 15

#obs. var/cov. = J(J+1)/2 = 6*7/2 = 21

Free parameters:

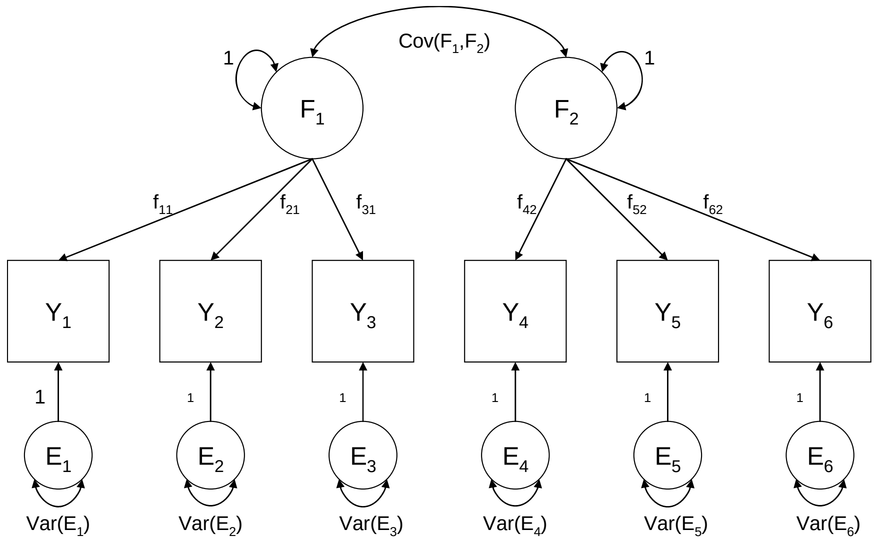
residual variances = 6

factor loadings = 6

covariances among factors = 1

Degrees of freedom:

df = 21-13 = 8



Practical – Description of data

DATASET

- Netherlands Twin Register (www.tweelingenregister.org)
- Dutch Health and Behavior Questionnaire (DHBQ)
- Adolescent Twins and non-twin Siblings
- Aged 14 and 16 (siblings between 12 and 25)
- Online & Paper and Pencil

CONTENT DHBQ

- Psychopathology
- Exercise
- Self –esteem
- Optimism
- Life Events
- Loneliness
- Number of peers and peer relation
- General Health and Illnesses (astma, migraine, etc)
- Pubertal Development
- Hours sleep
- Personality (age 16)
- Family Functioning (Family Functioning, Family Conflict)
- Life style (smoking, alcohol use, marihuna use)
- Educational Achievement (incl truancy)
- Wellbeing (Happiness, Satisfaction with Life, Quality of Life)
- Leisure time activities
- Family Size (no of sibs)
- Family situation (divorce)
- Zygosity
- Height, Weight
- Eating Disorders

Sample overview

	N of individuals	N of families	1 twin	2 twins	1 twin + sib	2 twins + sib
MZM	1061	474	28	290	15	141
DZM	917	425	56	232	14	123
MZF	1540	697	54	432	11	200
DZF	1116	512	49	309	13	141
DOS	2061	999	169	566	32	232
Sibs only	78	78	--	--	--	--
Total	6773	3185	356	1829	85	837

Today's Focus

- Youth Self Report (YSR)
- Subjective Wellbeing
    * subjective happiness
    * satisfaction with life
    * quality of life
- General Family Functioning

Practical 1: Single group factor models

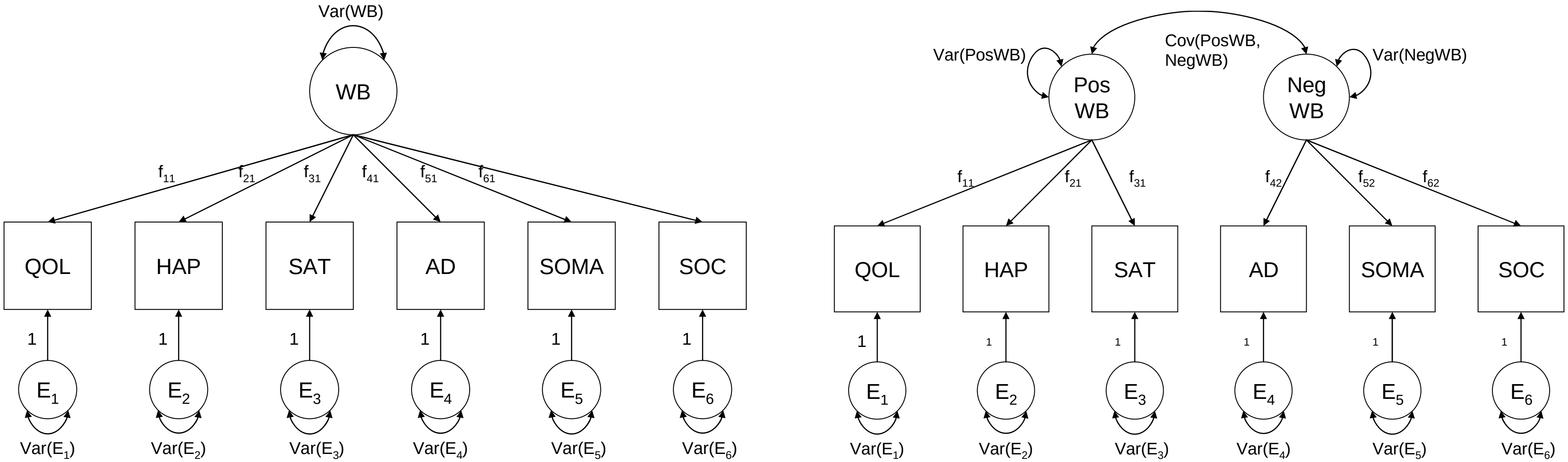
Data: CFA_family_wellbeing.dat

- 1000 adolescent twins (one twin per family)
- Observed variables:
  - Quality of life
  - Happiness
  - Satisfaction with life
  - Anxious depression scale (YSR)
  - Somatic complaints scale (YSR)
  - Social problems scale (YSR)

Files are on F:\marleen\Boulder2010\CFA

Practical 1: Single group factor models

1 factor model *vs.* 2 factor model



Practical 1a: Single group 1 factor model

OpenMx script OneFactorModelMatrix_WELLBEING.R

```
require(OpenMx)

# Prepare Data
# -----------------------------------------------------------------------
allData<-read.table("CFA_family_wellbeing.dat", header=TRUE, na.strings=-999)

cfaData<-allData[, c('qol','hap','sat','ad','soma','soc')]

colMeans(cfaData, na.rm=TRUE)
cov(cfaData[,c('qol','hap','sat','ad','soma','soc')],use="pairwise.complete.obs")
cor(cfaData[,c('qol','hap','sat','ad','soma','soc')],use="pairwise.complete.obs")

nvar<-6
nfac<-1
```

→ Read in data, -999 are treated as missing NA

→ Select variables to use in CFA

→ Compute descriptives of variables

→ Specify number of variables and factors

Practical 1a: Single group 1 factor model

OpenMx script OneFactorModelMatrix_WELLBEING.R

```
# Run single group 1 factor model - cov data input
# -----------------------------------------------------------------
observedVars <- names(cfaData)

oneFactorModelcov <- mxModel("One Factor",
   mxMatrix(type="Full", nrow=nvar, ncol=nfac, values=0.2, free=TRUE, name="L"),
   mxMatrix(type="Symm", nrow=nfac, ncol=nfac, values=1, free=TRUE, name="P"),
   mxMatrix(type="Diag", nrow=nvar, ncol=nvar, values=1, free=TRUE, name="T"),
   mxAlgebra(expression=L %*% P %*% t(L) + T, name="expCov"),
   mxData(cov(cfaData, use="pairwise.complete.obs"),type="cov", numObs=1000),
   mxMLObjective(covariance="expCov", dimnames = observedVars))

oneFactorFitcov<-mxRun(oneFactorModelcov)

summary(oneFactorFitcov)
```

Save variable names

Specify factor model

Run factor model

Demand summary output

Practical 1a: Single group 1 factor model

- Copy files from F:\marleen\Boulder2010\CFA to own directory
- Check whether your own directory is your working directory!

1. Open script OneFactorModelMatrix_WELLBEING.R
2. Identify the factor model by constraining Var(WB)=1
3. Run the 1 factor model
4. Write down the following information:

Model:	#obs. stat.	#free par.	Chi^2	df	AIC	BIC	RMSEA
1 factor							

Practical 1a: Single group 1 factor model

```
oneFactorModelcov <- mxModel("One Factor",
   mxMatrix(type="Full", nrow=nvar, ncol=nfac, values=0.2, free=TRUE, name="L"),
   mxMatrix(type="Symm", nrow=nfac, ncol=nfac, values=1, free=FALSE, name="P"),
   mxMatrix(type="Diag", nrow=nvar, ncol=nvar, values=1, free=TRUE, name="T"),
   mxAlgebra(expression=L %*% P %*% t(L) + T, name="expCov"),
   mxData(cov(cfaData, use="pairwise.complete.obs"),type="cov", numObs=1000),
   mxMLObjective(covariance="expCov", dimnames = observedVars))
```

Model:	#obs. stat.	#free par.	Chi^2	df	AIC	BIC	RMSEA
1 factor	21	12	508.6	9	490.6	223.2	0.24

Practical 1b: Single group 2 factor model

OpenMx script TwoFactorModelMatrix_WELLBEING.R

```
require(OpenMx)

# Prepare Data
# -----------------------------------------------------------------
allData<-read.table("CFA_family_wellbeing.dat", header=TRUE, na.strings=-999)
```

→ Read in data, -999 are treated as missing NA

```
cfaData<-allData[, c('qol','hap','sat','ad','soma','soc')]
```

→ Select variables to use in CFA

```
colMeans(cfaData, na.rm=TRUE)
cov(cfaData[,c('qol','hap','sat','ad','soma','soc')],use="pairwise.complete.obs")
cor(cfaData[,c('qol','hap','sat','ad','soma','soc')],use="pairwise.complete.obs")
```

→ Compute descriptives of variables

```
nvar<-6
nfac<-2
```

→ Specify number of variables and factors

Practical 1b: Single group 2 factor model

OpenMx script TwoFactorModelMatrix_WELLBEING.R

```
# Run single group 2 factor model - cov data input
# -----------------------------------------------------------------
observedVars <- names(cfaData)

twoFactorModelCov <- mxModel("Two Factor",
   mxMatrix(type="Full", nrow=nvar, ncol=nfac,
     values=c(rep(0.3,3),rep(0,6),rep(0.3,3)),
     free=c(rep(TRUE,3),rep(FALSE,6),rep(TRUE,3)), name="L"),
   mxMatrix(type="Symm", nrow=nfac, ncol=nfac, values=c(0.9,0.5,0.9),
     free=c(TRUE,TRUE,TRUE), name="P"),
   mxMatrix(type="Diag", nrow=nvar, ncol=nvar, values=1, free=TRUE, name="T"),
   mxAlgebra(expression=L %*% P %*% t(L) + T, name="expCov"),
   mxData(cov(cfaData, use="pairwise.complete.obs"), type="cov", numObs=1000),
   mxMLObjective(covariance="expCov", dimnames = observedVars))

twoFactorFitCov<-mxRun(twoFactorModelCov)

summary(twoFactorFitCov)
```

→ Save variable names

→ Specify factor model

→ Run factor model

→ Demand summary output

Factor loading matrix "L"

```
mxMatrix(type="Full", nrow=nvar, ncol=nfac,
      values=c(rep(0.3,3),rep(0,6),rep(0.3,3)),
      free=c(rep(TRUE,3),rep(FALSE,6),rep(TRUE,3)), name="L"),
```

Factor loading matrix L:

$$\begin{bmatrix} f_{11} & 0 \\ f_{21} & 0 \\ f_{31} & 0 \\ 0 & f_{42} \\ 0 & f_{52} \\ 0 & f_{62} \end{bmatrix}$$

Starting values for elements in this matrix:

$$\begin{bmatrix} 0.3 & 0 \\ 0.3 & 0 \\ 0.3 & 0 \\ 0 & 0.3 \\ 0 & 0.3 \\ 0 & 0.3 \end{bmatrix}$$

Free parameters in this matrix:

$$\begin{bmatrix} TRUE & FALSE \\ TRUE & FALSE \\ TRUE & FALSE \\ FALSE & TRUE \\ FALSE & TRUE \\ FALSE & TRUE \end{bmatrix}$$

Covariance matrix latent factors "P"

```
mxMatrix("Symm", nfac, nfac, values=c(0.9,0.5,0.9),
   free=c(TRUE,TRUE,TRUE), name="P"),
```

Covariance matrix P:

$$\begin{bmatrix} \mathrm{var}(F_1) & \mathrm{cov}(F_1F_2) \\ \mathrm{cov}(F_1F_2) & \mathrm{var}(F_2) \end{bmatrix}$$

Starting values for elements in this matrix:

$$\begin{bmatrix} 0.9 & 0.5 \\ 0.5 & 0.9 \end{bmatrix}$$

Free parameters in this matrix:

$$\begin{bmatrix} TRUE & TRUE \\ TRUE & TRUE \end{bmatrix}$$

Practical 1b: Single group 2 factor model

1. Open script TwoFactorModelMatrix_WELLBEING.R
2. Identify the factor model by constraining Var(PosWB)=1 and Var(NegWB)=1
3. Run the 2 factor model
4. Write down the following information:

Model:	#obs. stat.	#free par.	Chi^2	df	AIC	BIC	RMSEA
1 factor	21	12	508.6	9	490.6	223.2	0.24
2 factor							

5. How do the models fit? Which model fits best?

Files are on F:\marleen\Boulder2010\CFA

Practical 1b: Single group 2 factor model

```
twoFactorModelCov <- mxModel("Two Factor",
 mxMatrix(type="Full", nrow=nvar, ncol=nfac,
   values=c(rep(0.3,3),rep(0,6),rep(0.3,3)),
   free=c(rep(TRUE,3),rep(FALSE,6),rep(TRUE,3)), name="L"),
  mxMatrix(type="Symm", nrow=nfac, ncol=nfac, values=c(1,0.5,1),
   free=c(FALSE,TRUE,FALSE), name="P"),
  mxMatrix(type="Diag", nrow=nvar, ncol=nvar, values=1, free=TRUE, name="T"),
  mxAlgebra(expression=L %*% P %*% t(L) + T, name="expCov"),
  mxData(cov(cfaData, use="pairwise.complete.obs"), type="cov", numObs=1000),
  mxMLObjective(covariance="expCov", dimnames = observedVars))
```

Model:	#obs. stat.	#free par.	Chi^2	df	AIC	BIC	RMSEA
1 factor	21	12	508.6	9	490.6	223.2	0.24
2 factor	21	13	69.9	8	53.9	7.3	0.09

Outline

- Introduction to factor analysis
- Phenotypic factor analysis
- More advanced models
  - Factor models for categorical data
  - Multigroup factor models and measurement invariance
- From phenotypic to genetic factor analysis...

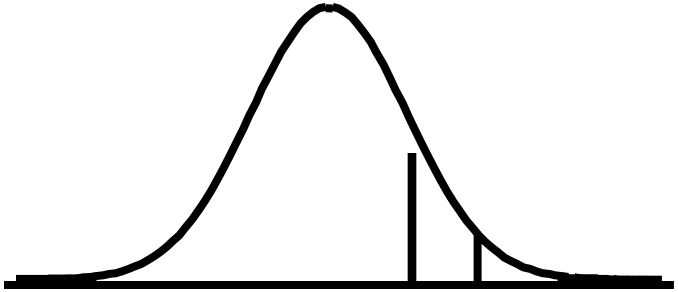
Factor models for categorical data

- What if my observed data are categorical?
- For example, multiple items of one scale

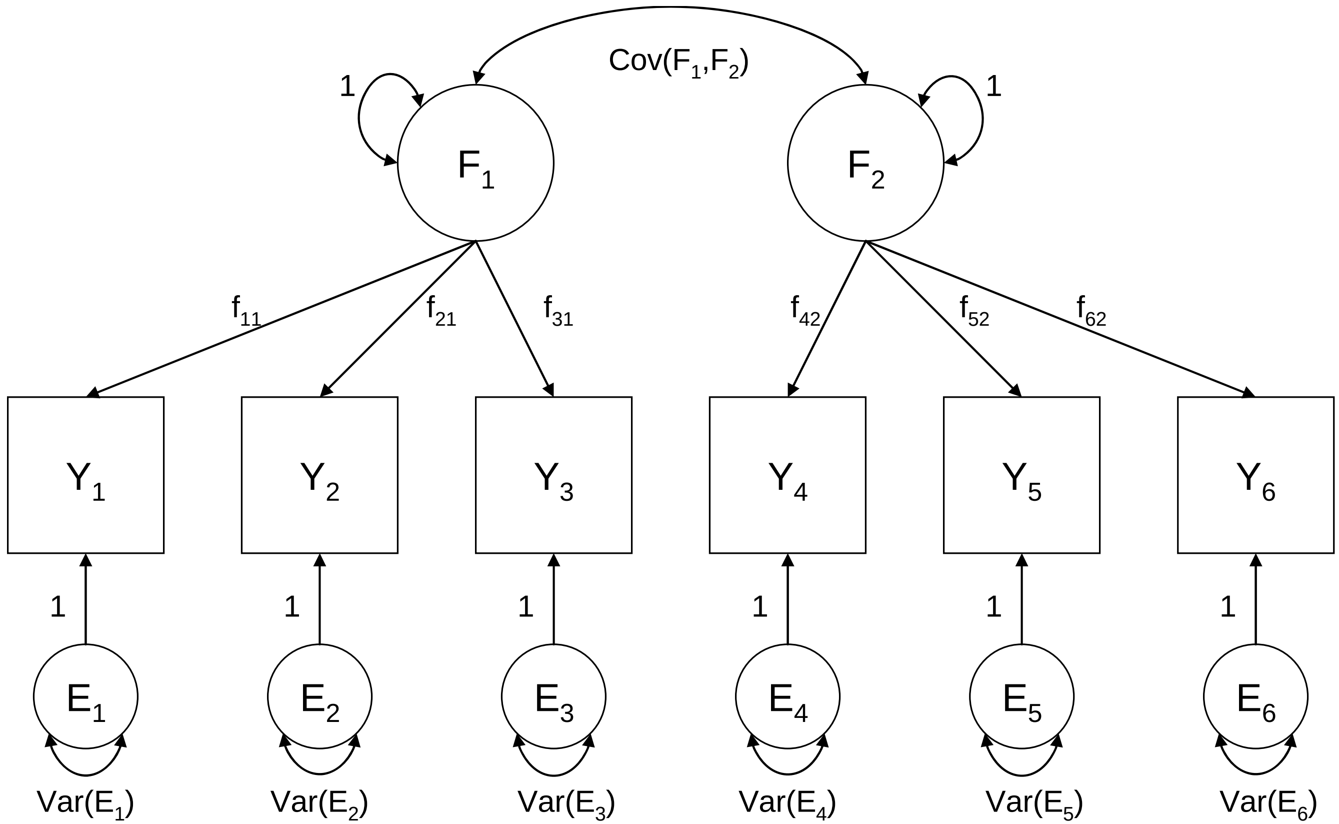
→ Threshold models

=

Latent response variable models



The 2 factor model – continuous data

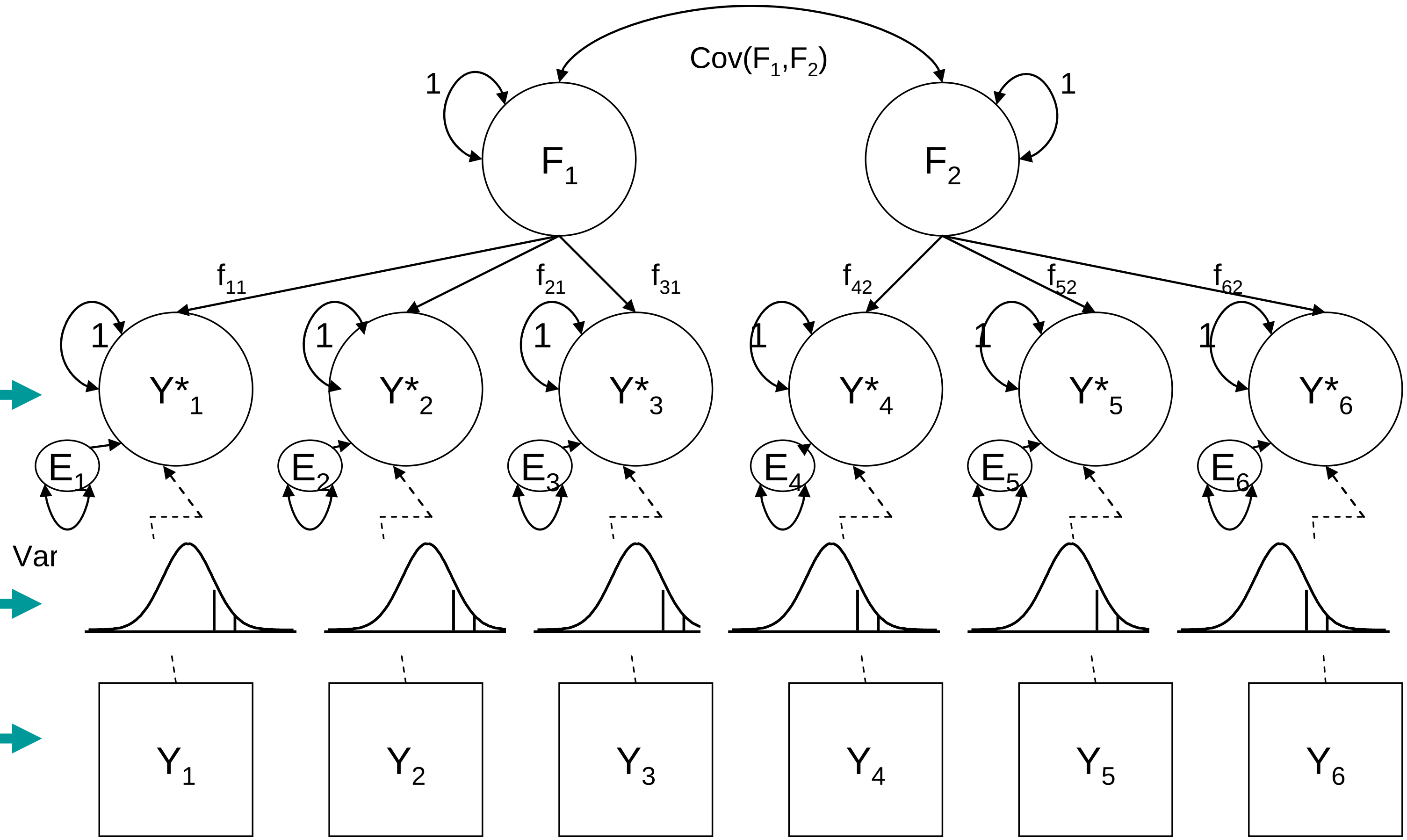


The 2 factor model – categorical data

Latent response variables (continuous)

Thresholds

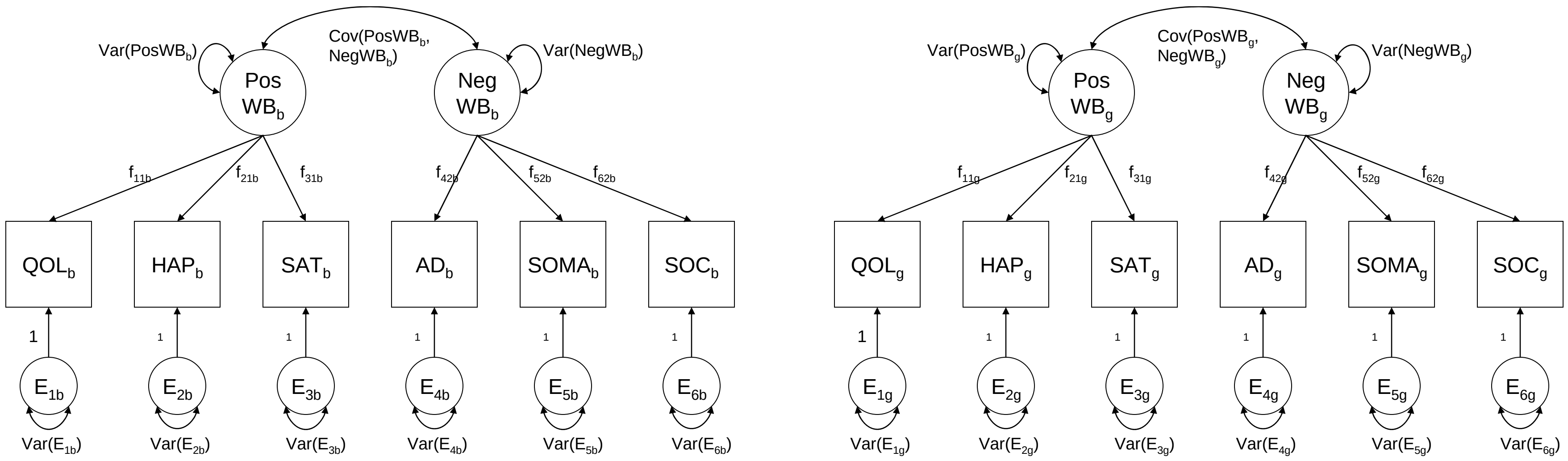
Observed variables (categorical)



Multigroup factor model

- Fit factor model in multiple groups

Group 1: Boys　　　　Group 2: Girls



Group comparisons

- Group comparisons of latent constructs:
  - Means

    For example:

    - IQ differences across ethnic groups
  - Covariance structure

    For example:

    - Covariance differences in negative and positive wellbeing in adolescent boys and girls
- Only meaningful if shown that same constructs are measured in all groups!

Modeling means and covariances

Means model:

$$E[Y_g] = \nu_g + \Lambda_g \alpha_g$$

Covariance model:

$$\Sigma_g = \Lambda_g \Psi_g \Lambda_g' + \Theta_g$$

Measurement invariance (MI)

= Absence of measurement bias

= Same measurement model holds in each group

→ Group differences in observed variables are only caused by group differences in latent factors, and not by other differences in the model, such as differences in factor loadings

Types of MI models

Four models (models 2-4 are nested under 1)

1. Configural Invariance model
2. Metric Invariance model
3. Strict Invariance model
4. Strong Invariance model

Most complex

Most parsimonious

Types of MI models

Four models (models 2-4 are nested under 1)

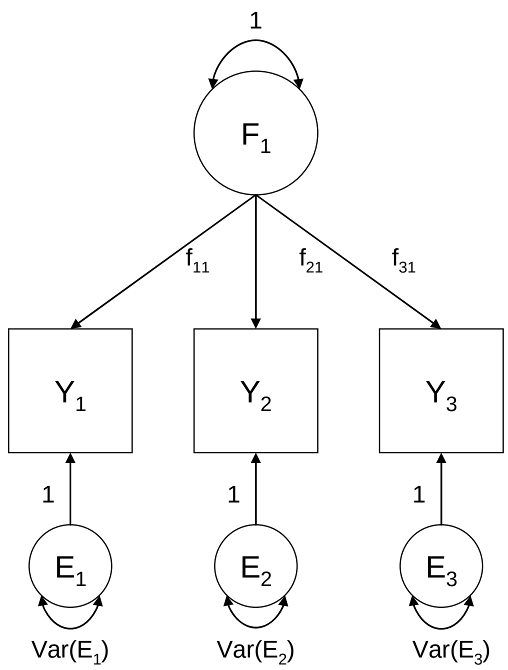
1. Configural Invariance model
   - Fit same factor model in each group
2. Metric Invariance model
   - Constrain factor loadings equal across groups
3. Strict Invariance model
   - Constrain factor loadings and intercepts equal across groups
4. Strong Invariance model
   - Constrain factor loadings, intercepts and residual variances equal across groups

Outline

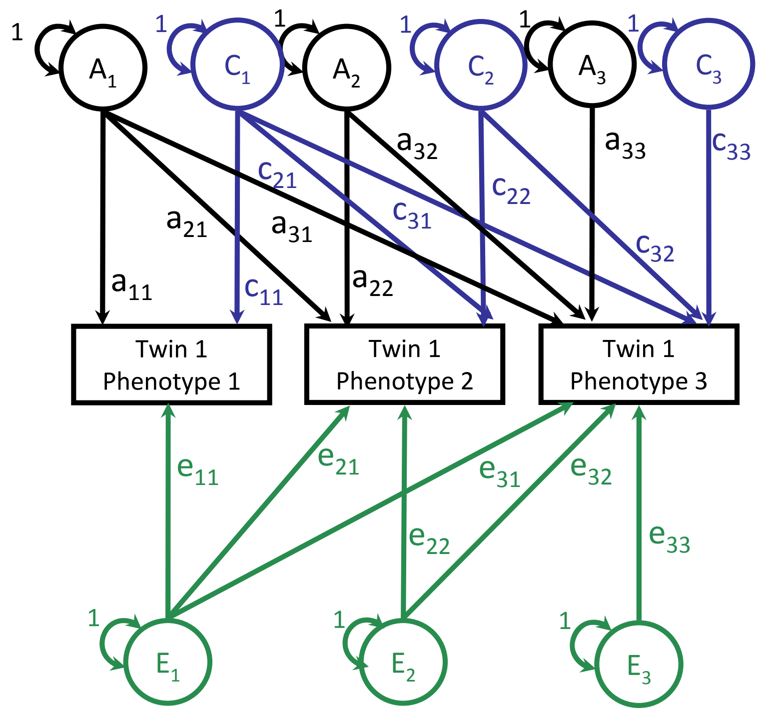
- Introduction to factor analysis
- Phenotypic factor analysis
- More advanced models
- From phenotypic to genetic factor analysis...

Phenotypic versus genetic models

Phenotypic factor model



Multivariate genetic models – Cholesky decomposition

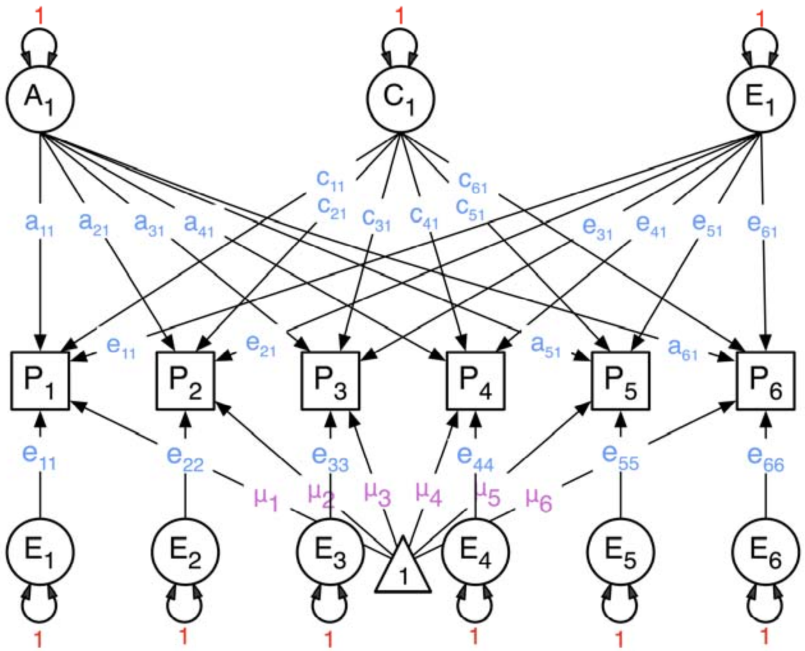


11.00-12.00 Danielle & Meike

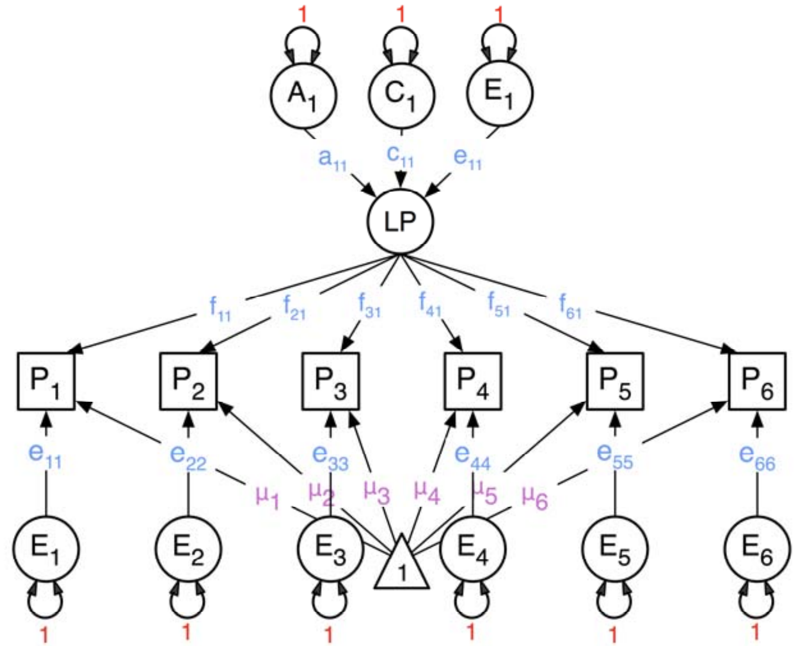
13.00-14.00 Meike & Danielle

Phenotypic versus genetic models

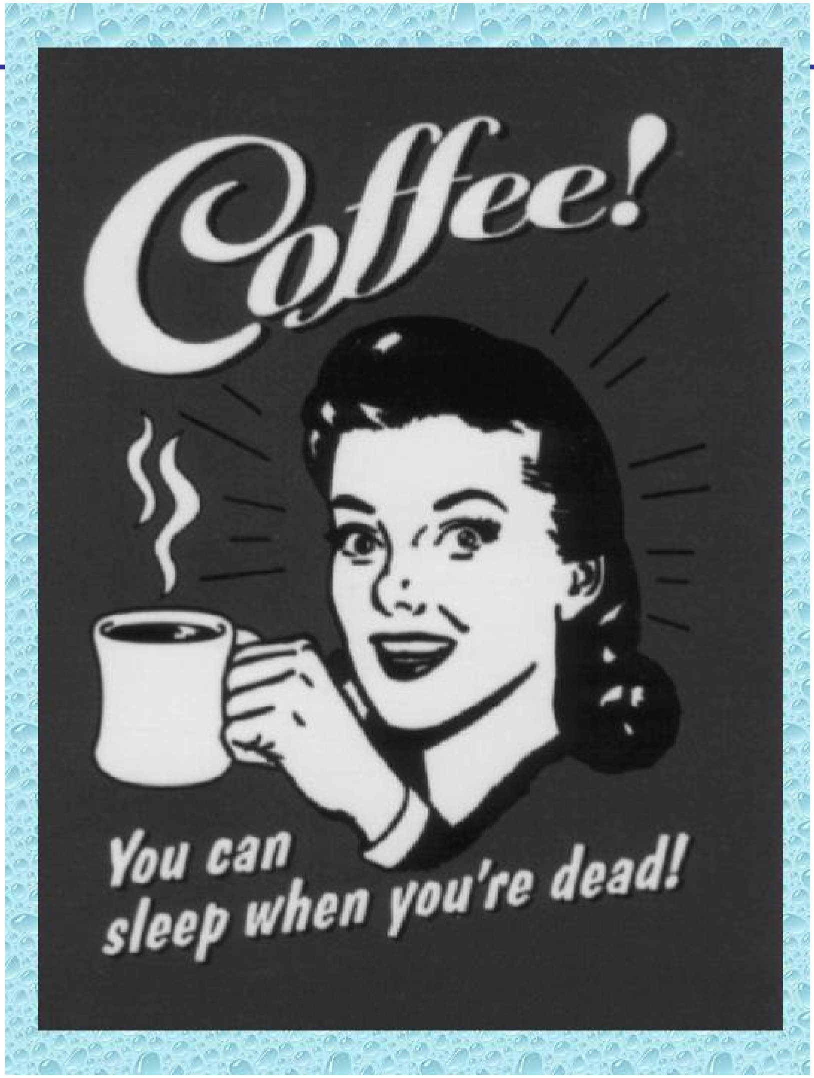
Multivariate genetic models – Independent pathway model



Multivariate genetic models – Common pathway model



14.30-16.45 Hermine & Nick


Coffee!
You can
sleep when you're dead!