Structural Equation Modeling

M&S Lecture 1, 2017:

Confirmatory Factor Analysis

Ellen Hamaker (E.L.Hamaker@uu.nl)
Methodology and Statistics
Faculty of Social and Behavioural Sciences

Outline

- One-factor model:

  - covariance structure

  - representation: path diagram, set of equations, matrices

  - scaling

- Mplus
- Nested models

Structural Equation Modeling

SEM consists of analyzing:

- **covariance structure**
- **mean structure** (only when there are multiple groups or longitudinal data)

SEM in a nutshell:

- based on **substantive theory** we expect **certain relationships** between the observed variables
- these expectations can be translated into a **model**
- this model can be **fit** to the **data**
- through **evaluating model fit**, we determine whether our theory makes sense

Example: Negative affect

525 psychology students indicated on a 7-point scale how much the following words described them in general:

- afraid
- distress
- upset
- scared
- nervous

Correlations

		afraid	distress	upset	scared	nervous
afraid	Pearson Correlation	1.000	.485**	.550**	.742**	.528**
	Sig. (2-tailed)		.000	.000	.000	.000
	N	525	524	524	524	524
distress	Pearson Correlation	.485**	1.000	.584**	.542**	.494**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	524	524	524	524	524
upset	Pearson Correlation	.550**	.584**	1.000	.612**	.510**
	Sig. (2-tailed)	.000	.000		.000	.000
	N	524	524	524	524	524
scared	Pearson Correlation	.742**	.542**	.612**	1.000	.555**
	Sig. (2-tailed)	.000	.000	.000		.000
	N	524	524	524	524	524
nervous	Pearson Correlation	.528**	.494**	.510**	.555**	1.000
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	524	524	524	524	524

**. Correlation is significant at the 0.01 level (2-tailed).

NOTE: **Correlations** are **standardized covariances**, that is

$$\mathrm{Cor}(x, y) = \frac{\mathrm{Cov}(x, y)}{\sqrt{\mathrm{var}(x)}\sqrt{\mathrm{var}(y)}}$$

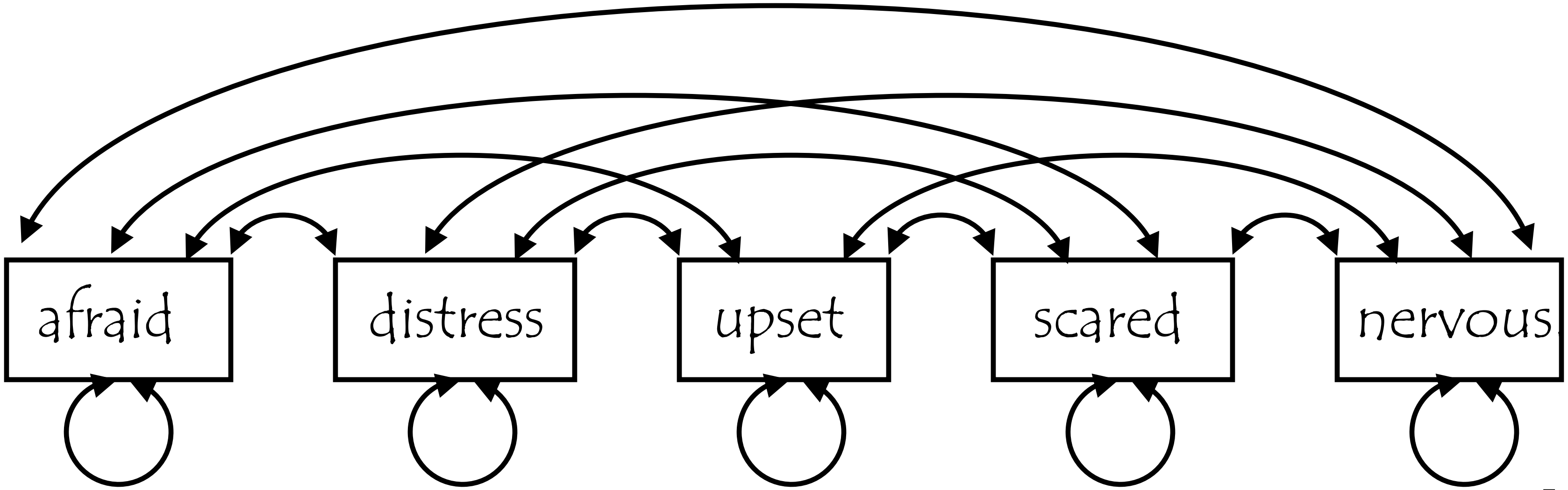
Observed covariance structure

If we have **5 observed variables**, we have (5*6)/2 = **15 unique elements** in the covariance matrix: 5 variances and (5*4)/2=10 covariances.

Our theory is that these observed variables measure a **single underlying construct**: Negative Affect (NA).

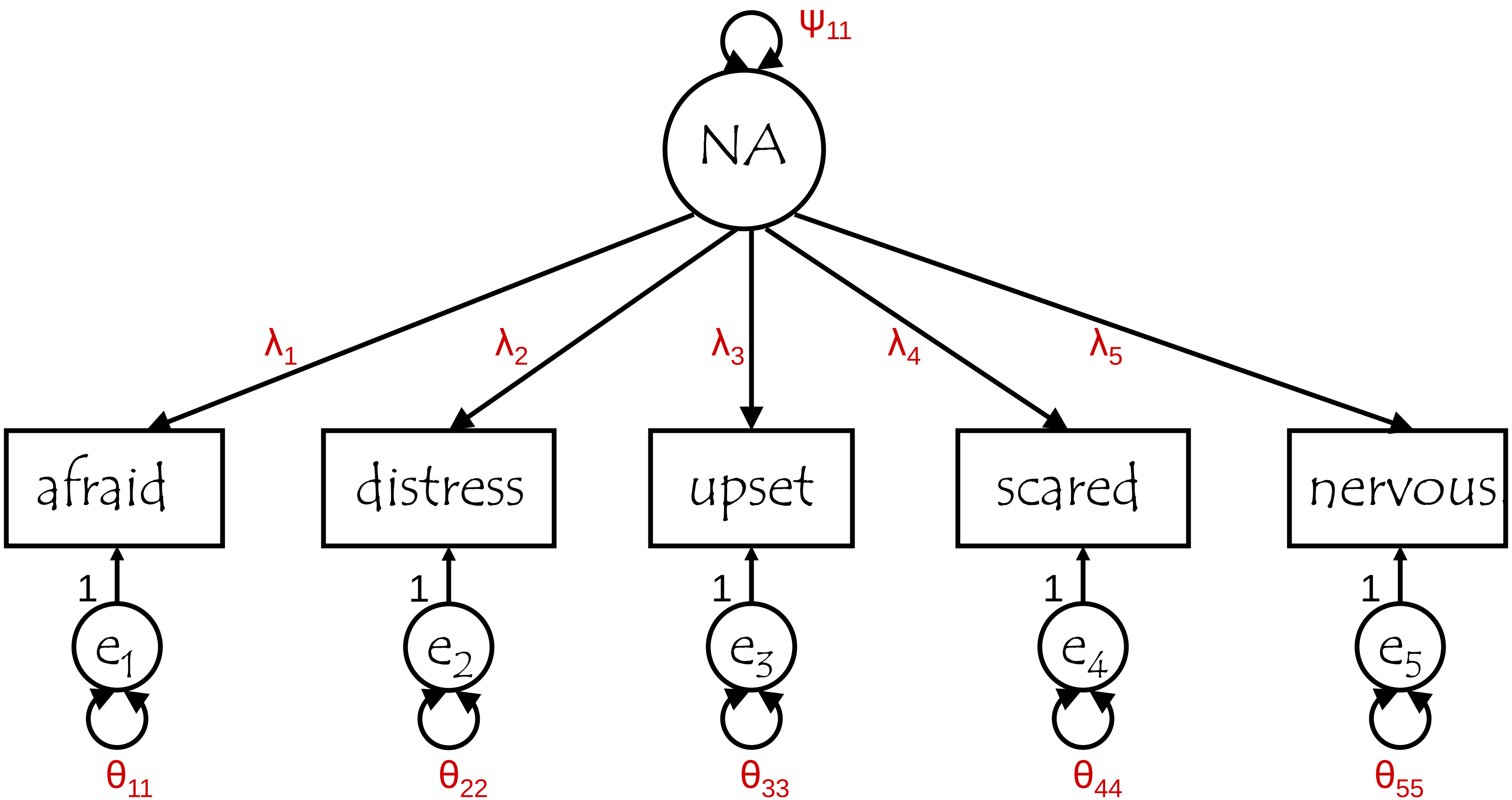
Hence, we are interested in fitting a **one factor model**.

We compare this to the **saturated model**, where everything covaries with everything:



Model representation 1: Path diagram

The one factor model can be represented as the following path diagram:



Model representation 2: Set of equations

The path diagram is **the same** as the set of equations:

$$\text{afraid}_i = \lambda_1 NA_i + e_{1i}$$

$$\text{distress}_i = \lambda_2 NA_i + e_{2i}$$

$$\ldots$$

$$\text{nervous}_i = \lambda_5 NA_i + e_{5i}$$

Model representation 3: Matrix notation

…which can **also** be expressed using matrix notation as:

$$\begin{bmatrix} afraid_i \\ distress_i \\ upset_i \\ scared_i \\ nervous_i \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{bmatrix} NA_i + \begin{bmatrix} e_{1i} \\ e_{2i} \\ e_{3i} \\ e_{4i} \\ e_{5i} \end{bmatrix}$$

A **general matrix notation** for factor models is the **measurement equation**:

$$\mathbf{y}_i = \Lambda \eta_i + \varepsilon_i$$

where Λ is the **matrix with factor loadings**

and η is a vector with **latent variables** (=factors).

Covariance matrix

We do not need to estimate all the individual η's and ε's.

$$\Sigma = E\left[(y_i - \mu)(y_i - \mu)^T\right] = E\left[y_i y_i^T\right]$$

$$= E\left[(\Lambda\eta_i + \varepsilon_i)(\eta_i^T\Lambda^T + \varepsilon_i^T)\right]$$

$$= E\left[\Lambda\eta_i\eta_i^T\Lambda^T + \Lambda\eta_i\varepsilon_i^T + \varepsilon_i\eta_i^T\Lambda^T + \varepsilon_i\varepsilon_i^T\right]$$

$$= E\left[\Lambda\eta_i\eta_i^T\Lambda^T\right] + E\left[\Lambda\eta_i\varepsilon_i^T\right] + E\left[\varepsilon_i\eta_i^T\Lambda^T\right] + E\left[\varepsilon_i\varepsilon_i^T\right]$$

$$= \Lambda E\left[\eta_i\eta_i^T\right]\Lambda^T + \Lambda E\left[\eta_i\varepsilon_i^T\right] + E\left[\varepsilon_i\eta_i^T\right]\Lambda^T + E\left[\varepsilon_i\varepsilon_i^T\right]$$

$$= \Lambda\Psi\Lambda^T + 0 + 0 + \Theta$$

Modeled covariance structure

From the general notation in matrices, a **general expression of the covariance structure** can be derived.

Σ is the **modeled covariance matrix** for the observed variables, that is:

$$\Sigma = \Lambda \Psi \Lambda^T + \Theta$$

where

- **Λ** is the matrix with **factor loadings**
- **Ψ** is the **covariance matrix of the latent variables** (i.e., the factors)
- **Θ** is the **covariance matrix of the residuals** (i.e., measurement errors)

Note that these three matrices contain **unknown parameters** we need to estimate.

Parameters in our model matrices

$$\Lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{bmatrix}$$

5 factor loadings

$$\Theta = \begin{bmatrix} \theta_{11} & & & & \\ 0 & \theta_{22} & & & \\ 0 & 0 & \theta_{33} & & \\ 0 & 0 & 0 & \theta_{44} & \\ 0 & 0 & 0 & 0 & \theta_{55} \end{bmatrix}$$

5 residual variances (i.e., variances of measurement errors)

$$\Psi = [\psi_{11}]$$

1 factor variance

DIY

Write down Σ for this model using $\Sigma = \Lambda\Psi\Lambda^T + \Theta$

Obtaining analytical expressions

Instead of using matrix algebra, you can also determine the elements of $\mathbf{\Sigma}$ separately.

For instance, the variance of y_1 can be obtained through:

$$Var(y_{1i}) = E\left[y_{1i}^2\right] = E\left[\{\lambda_1\eta_i + \varepsilon_{1i}\}^2\right]$$

$$= E\left[\{\lambda_1\eta_i\}^2\right] + E\left[\varepsilon_{1i}^2\right] + 2E\left[\lambda_1\eta_i\varepsilon_{1i}\right]$$

$$= \lambda_1^2 E\left[\eta_i^2\right] + E\left[\varepsilon_{1i}^2\right] + 2\lambda_1 E\left[\eta_i\varepsilon_{1i}\right]$$

$$= \lambda_1^2\psi_{11} + \theta_{11}$$

DIY

Obtain an analytical expression for the covariance between y_1 and y_2 in the same way.

Means

SEM also allows for the modeling of the **mean structure**.

Here we have 5 observed means: **m**.

If we do not specify a mean structure (as is the case here), Mplus will:

– estimated 5 intercepts for the 5 observed variables: $\boldsymbol{\nu}$

– fix the latent mean to zero: E(η) = α = 0

Then we have: $\mathbf{y}_i = \nu + \Lambda \eta_i + \varepsilon_i$

with covariance matrix: $\Sigma = \Lambda \Psi \Lambda^T + \Theta$

and mean vector: $\mu = \nu$

Likelihood estimation

The unknown parameters in the model matrices Λ, Ψ, Θ, ν, and α, are collected in the parameter vector $\boldsymbol{\omega}$.

To **estimate** $\boldsymbol{\omega}$ and to determine the **model fit**, we make use of the likelihood function (based on multivariate normal density of the data):

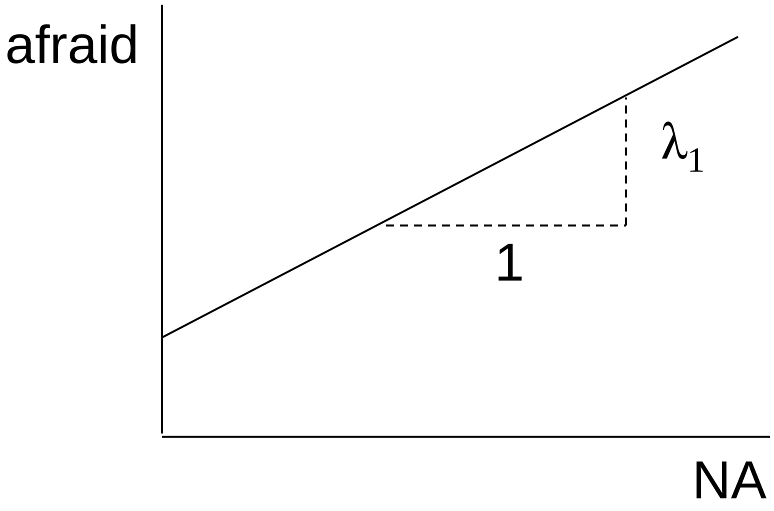
$$\log L = c - \frac{N}{2}\log|\Sigma_{\omega}| - \frac{N}{2} tr\left(S\Sigma_{\omega}^{-1}\right) - \frac{N}{2}(\mathrm{m} - \mu_{\omega})^{\mathrm{T}}\Sigma_{\omega}^{-1}(\mathrm{m} - \mu_{\omega})$$

Hence, in the likelihood function the **modeled covariance matrix $\boldsymbol{\Sigma_{\omega}}$** and **modeled means $\boldsymbol{\mu_{\omega}}$** are compared to the **observed covariance matrix S** and **observed means m**.

Scaling in a factor model

The model defined above is **not identified**, i.e., there is **no unique solution**.

The problem is that the latent variable is not observed, and therefore its **scale** is not known.



$$afraid_i = \nu_1 + \lambda_1 NA_i + e_{1i}$$
$$= \nu_1 + (k \times \lambda_1)\frac{NA_i}{k} + e_{1i}$$

To ensure the model is identified, we need to **scale the factor**: we either fix one of the factor loadings to 1, or the variance of the factor to 1.

Outline

- One-factor model
- **Mplus:**

  - input

  - model fit

  - modification indices

  - parameter estimates

  - Tech1 output
- Nested models

Specifying the model in Mplus

PANASmodel1.inp

```
TITLE:  NA trait short version

DATA:   FILE IS PANAS.dat;

VARIABLE:

NAMES ARE ashamed afraid inspired
        nervous enthusia strong
        distress excited interest irritabl
        jittery determin upset alert proud
        guilty scared active hostile attentiv;

USEVARIABLES ARE afraid distress upset scared nervous;

MODEL:  NA BY   afraid distress upset scared nervous;

OUTPUT: MODINDICES TECH1;
```

Sample statistics, free parameters and df

Data:

- observed covariance matrix **S** has 5*6/2=15 unique elements
- observed mean vector **m** has 5 elements

TOTAL: 20 sample statistics

Parameter to be estimated:

- 5 factor loadings (in $\boldsymbol{\Lambda}$)
- 5 residual variance (in $\boldsymbol{\Theta}$)
- 1 variance of the factor (in $\boldsymbol{\Psi}$)
- 5 means (in $\boldsymbol{\nu}$, as we are not modeling the mean structure)

NOTE: Mplus will scale the factor by fixing the first factor loading to 1

TOTAL: 15 free parameters

Thus: 20 – 15 = 5 degrees of freedom for this model

Model fit: Log likelihood

```
MODEL FIT INFORMATION

Number of Free Parameters                              15

Loglikelihood

          H0 Value                              -3704.358
          H1 Value                              -3676.246
```

H0 is the model we specified (the one factor model);

H1 is the saturated model (allowing everything to correlate with everything);

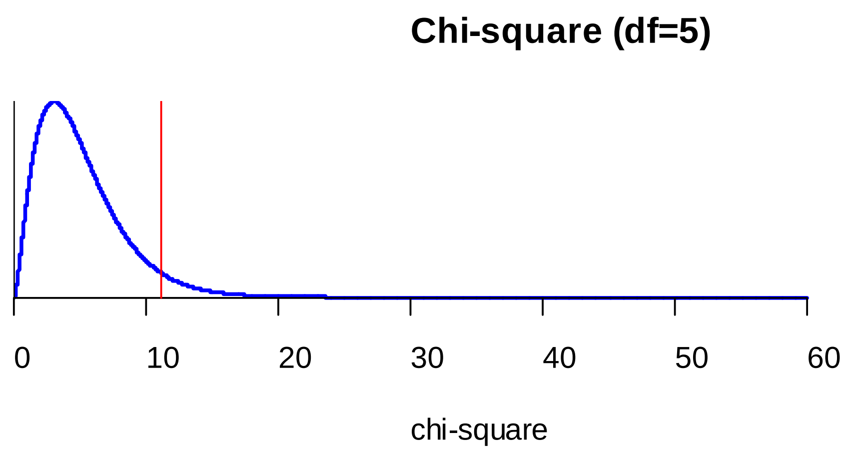
NOTE: All models are a special case of the saturated model, i.e., every model is nested under H1.

Model fit: Chi-square

```
Chi-Square Test of Model Fit

          Value                                  56.224
          Degrees of Freedom                          5
          P-Value                                0.0000
```

Chi-square statistic is:

$-2\log L_0 + 2\log L_1 = 2*3704.358 - 2*3676.246 = \mathbf{56.224}$



Hence, H0 (our model) needs to be **rejected** in favor of H1 (saturated model).

Thus, our model does not fit the data well.

Model fit: Chi-square

While the chi-square test is the traditional test in SEM, and is most frequently used, it is also criticized for being (too) dependent on **sample size**:

- small samples: model fits well (no power to reject H0)
- large samples: small deviations from truth result in significant test, and thus rejection of H0 (our model)

Therefore, many alternative measures of model fit have been developed.

Model fit: RMSEA

```
RMSEA (Root Mean Square Error Of Approximation)

          Estimate                                   0.140
          90 Percent C.I.                            0.109  0.174
          Probability RMSEA <= .05                   0.000
```

RMSEA is based on the chi-square, the df and sample size.

Rules of thumb: should be smaller than .08; <.05 implies good model fit.

NOTE: can be misleading when the df are small and sample size is not large

Our model has an RMSEA of .14, which implies the model does not fit well.

Model fit: CFI andTLI

```
CFI/TLI

          CFI                                    0.957
          TLI                                    0.914
```

Comparative Fit Index (CFI) and **Tucker Lewis index (TLI)** compare chi-square of fitted model to chi-square of baseline model.

Rules of thumb: should be larger than .90; larger than .95 implies good model fit.

NOTE: if the observations have low correlations, both CFI and TLI will be low too.

Our model has TLI>.9 and CFI>.95, which indicates good model fit.

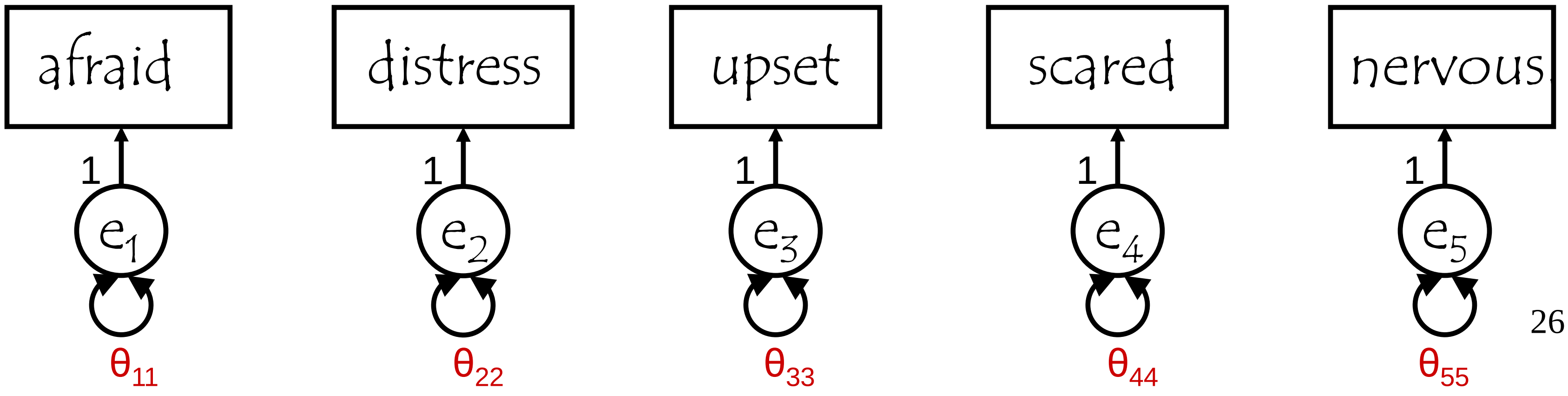
Model fit: CFI and TLI

```
Chi-Square Test of Model Fit for the Baseline Model

          Value                          1196.676
          Degrees of Freedom                   10
          P-Value                          0.0000
```

Baseline model or independence model is the model in which the variables are modeled as independent of each other.

This model is **nested under** our specified model.

The baseline model and the saturated models form **two extremes**.



Model fit: SRMR

```
SRMR (Standardized Root Mean Square Residual)

          Value                                  0.035
```

SRMR is based on difference between the observed covariance matrix and the modeled covariance matrix (i.e., $\mathbf{S}$-$\mathbf{\Sigma_0}$).

Rule of thumb: should be smaller than .08.

NOTE: This measure tends to be smaller as sample size increases and as the number of parameters in the model increases.

Our model has an SRMR<.08, which indicates good model fit.

What now?

Some measures (CFI, TLI, SRMR) indicate the model fits well, while other measure (chi-square test and RMSEA) indicate the model does not fit well.

If we are willing to take a **more exploratory approach**, we can look at the **modification indices** to see how our model can be improved.

Researchers do this all the time, to improve their model fit.

But…. **BEWARE!!!**

Once you go from confirmatory research to an exploratory approach, you allow **coincidence** to play a much bigger role in your results.

Also: don’t forget to use your brain!

Modification indices

```
Minimum M.I. value for printing the modification index    10.000

                                M.I.     E.P.C.  Std E.P.C.  StdYX E.P.C.

WITH Statements

DISTRESS WITH AFRAID          12.473    -0.142      -0.142        -0.210
UPSET    WITH AFRAID          12.718    -0.138      -0.138        -0.236
UPSET    WITH DISTRESS        35.595     0.244       0.244         0.317
SCARED   WITH AFRAID          51.526     0.330       0.330         0.689
SCARED   WITH DISTRESS        11.708    -0.145      -0.145        -0.230
```

Modification indices indicate how much the **chi-square statistic** will **go down**, if you included this parameter in the model.

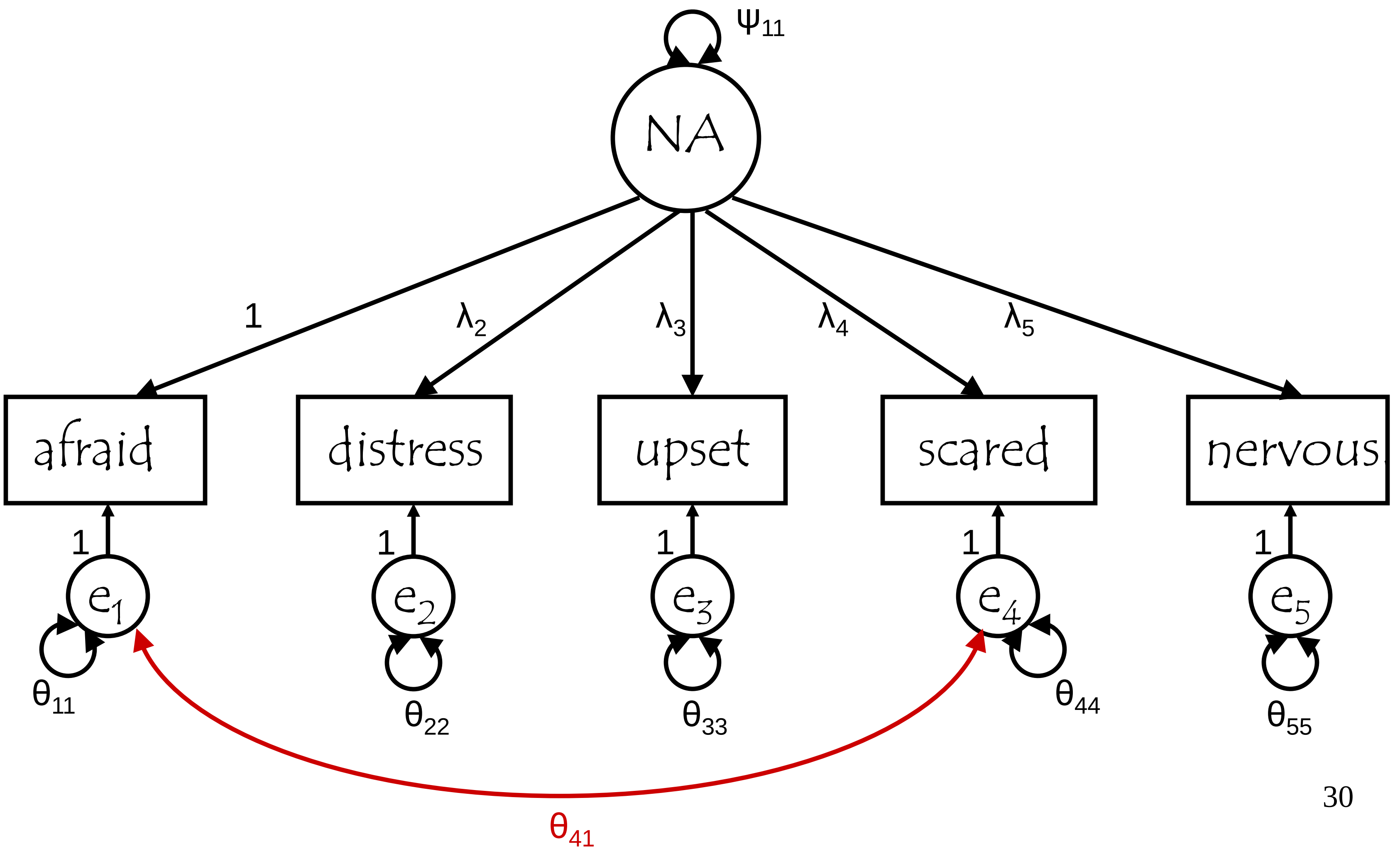
Here, we see **WITH statements**, which imply **covariances**.

The largest MI is for ***scared* and *afraid***: 51.526.

This makes sense from a **substantive perspective**.

scared WITH afraid (note it is the residuals!!)

Although Mplus uses the names of the **observed variables**, she actually means to say the **measurement errors**, that is:



Specifying the modified model in Mplus

PANASmodel1a.inp

```
TITLE:   NA trait short version

DATA:    FILE IS PANAS.dat;

VARIABLE:

NAMES ARE ashamed afraid inspired
        nervous enthusia strong
        distress excited interest irritabl
        jittery determin upset alert proud
        guilty scared active hostile attentiv;

USEVARIABLES ARE afraid distress upset scared nervous;

MODEL:   NA BY   afraid distress upset scared nervous;
        afraid WITH scared;

OUTPUT: MODINDICES TECH1;
```

Model fit modified model: Part 1

```
MODEL FIT INFORMATION

Number of Free Parameters                              16

Loglikelihood

          H0 Value                              -3681.196
          H1 Value                              -3676.246

Information Criteria

          Akaike (AIC)                           7394.393
          Bayesian (BIC)                         7462.515
          Sample-Size Adjusted BIC               7411.728
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                                     9.900
          Degrees of Freedom                            4
          P-Value                                  0.0421
```

Model fit modified model: Part 2

```
RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.053
          90 Percent C.I.                    0.009  0.096
          Probability RMSEA <= .05           0.385

CFI/TLI

          CFI                                0.995
          TLI                                0.988

Chi-Square Test of Model Fit for the Baseline Model

          Value                           1196.676
          Degrees of Freedom                    10
          P-Value                           0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                              0.016
```

Model results: Λ

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
NA BY				
AFRAID	1.000	0.000	999.000	999.000
DISTRESS	1.040	0.076	13.759	0.000
UPSET	1.097	0.074	14.854	0.000
SCARED	1.133	0.056	20.353	0.000
NERVOUS	0.996	0.072	13.856	0.000

Model results: θ_{41} and means

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
AFRAID WITH				
SCARED	0.304	0.050	6.104	0.000
Intercepts				
AFRAID	2.400	0.054	44.826	0.000
DISTRESS	2.971	0.055	54.013	0.000
UPSET	2.374	0.053	45.117	0.000
SCARED	2.448	0.056	43.551	0.000
NERVOUS	2.851	0.055	51.763	0.000

Model results Ψ and Θ

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Variances				
NA	0.752	0.089	8.420	0.000
Residual Variances				
AFRAID	0.745	0.060	12.341	0.000
DISTRESS	0.767	0.060	12.701	0.000
UPSET	0.541	0.051	10.673	0.000
SCARED	0.685	0.061	11.290	0.000
NERVOUS	0.837	0.063	13.269	0.000

TECH1 output

```
PARAMETER SPECIFICATION


           NU
              AFRAID        DISTRESS      UPSET         SCARED        NERVOUS
              ________      ________      ________      ________      ________
      1          1             2             3             4             5


           LAMBDA
              NA
              ________
AFRAID             0
DISTRESS           6
UPSET              7
SCARED             8
NERVOUS            9
```

NOTE: Mplus does not estimate the first factor loading; this was used for **scaling**.

TECH1 output

THETA

	AFRAID	DISTRESS	UPSET	SCARED	NERVOUS
AFRAID	10				
DISTRESS	0	11			
UPSET	0	0	12		
SCARED	13	0	0	14	
NERVOUS	0	0	0	0	15

ALPHA

	NA
1	0

BETA

	NA
NA	0

PSI

	NA
NA	16

Outline

- One-factor model
- Mplus
- **Nested models**

Comparing nested models

Often when using SEM, we will be **comparing multiple models**.

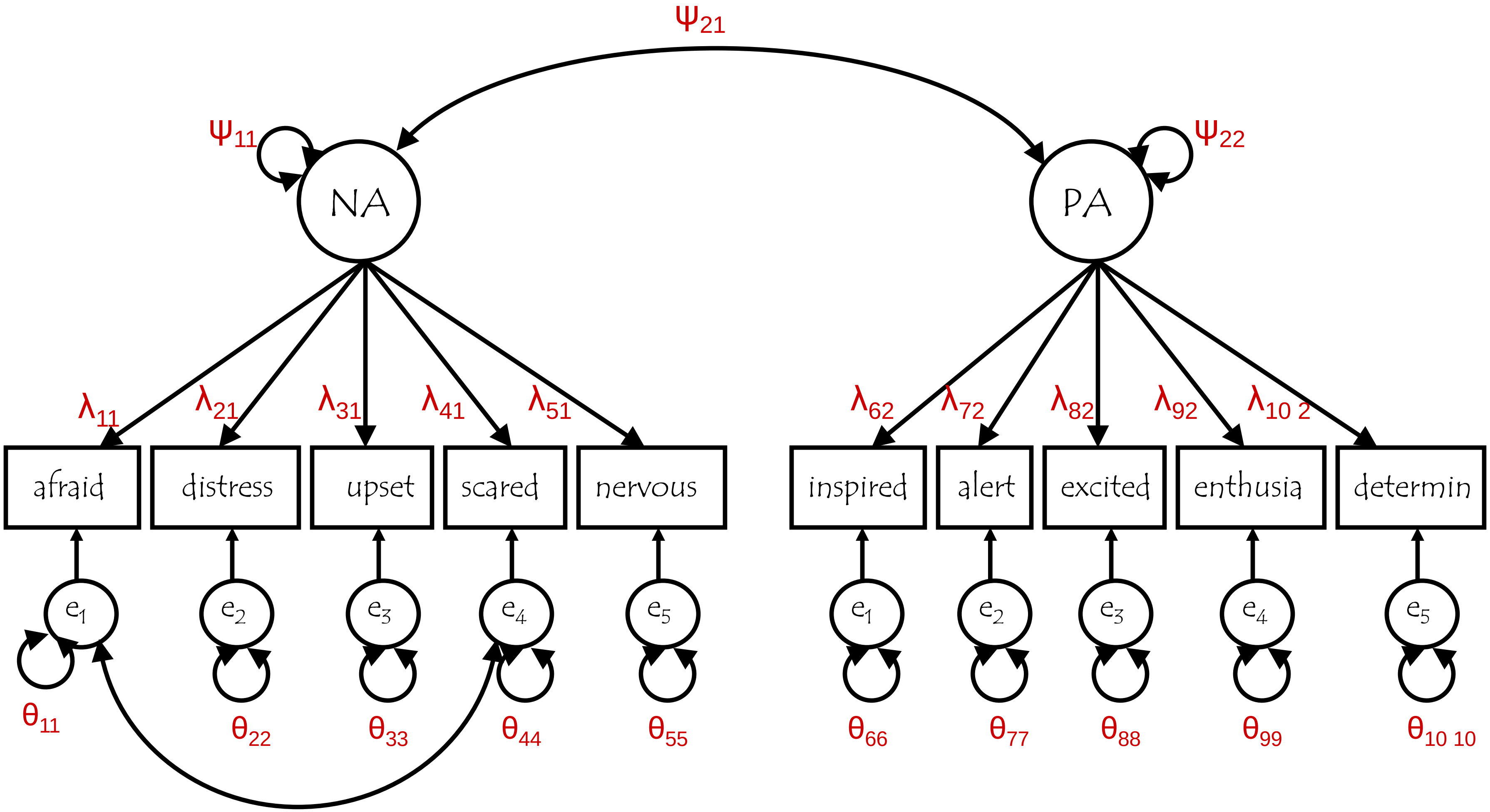
These models represent **different theories**.

If the models are **nested**, we can use the **chi-square difference test** to compare them (otherwise use AIC or BIC).

Models are nested if one is a **special case** of the other.

For instance, we may compare a **two-factor** model with a **one-factor model**: The one factor model is nested under the two-factor model, because it implies the correlation between the two factors is 1.

Two-factor model for NA and PA



DIY

We have: $\mathbf{y}_i = \nu + \Lambda \eta_i + \varepsilon_i$ Write down Λ, Ψ, Θ and ν.

$$y_i = \begin{bmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \\ y_{4i} \\ y_{5i} \\ y_{6i} \\ y_{7i} \\ y_{8i} \\ y_{9i} \\ y_{10i} \end{bmatrix}$$

$$\eta_i = \begin{bmatrix} \eta_{1i} \\ \eta_{2i} \end{bmatrix}$$

Sample statistics, free parameters and df

Data:

- observed covariance matrix **S** has 10*11/2=55 unique elements
- observed mean vector **m** has 10 elements

TOTAL: 65 sample statistics

Parameter to be estimated:

- 10 factor loadings (in $\boldsymbol{\Lambda}$); 2 fixed to 1 for scaling, so 8 factor loadings to be estimated
- 10 residual variances (in $\boldsymbol{\Theta}$)
- 1 covariance between the residuals (in $\boldsymbol{\Theta}$)
- 2 variances of the factors (in $\boldsymbol{\Psi}$)
- 1 covariance between the factors (in $\boldsymbol{\Psi}$)
- 10 means (in $\boldsymbol{\nu}$, as we are not modeling the mean structure)

TOTAL: 32 free parameters

Thus: 65 – 32 = 33 degrees of freedom for this model

Mplus input for 2-factor model

PANASmodel2.inp

```
TITLE:   PANAS traits, short versions
DATA:    FILE IS PANAS.dat;
VARIABLE:

NAMES ARE ashamed afraid inspired
         nervous enthusia strong
         distress excited interest irritabl
         jittery determin upset alert proud
         guilty scared active hostile attentiv;

USEVARIABLES ARE afraid distress upset scared nervous
             inspired alert excited enthusia determin;

MODEL:   NA BY afraid distress upset scared nervous;
         scared WITH afraid;
         PA BY inspired alert excited enthusia determin;

OUTPUT:  MODINDICES TECH1;
```

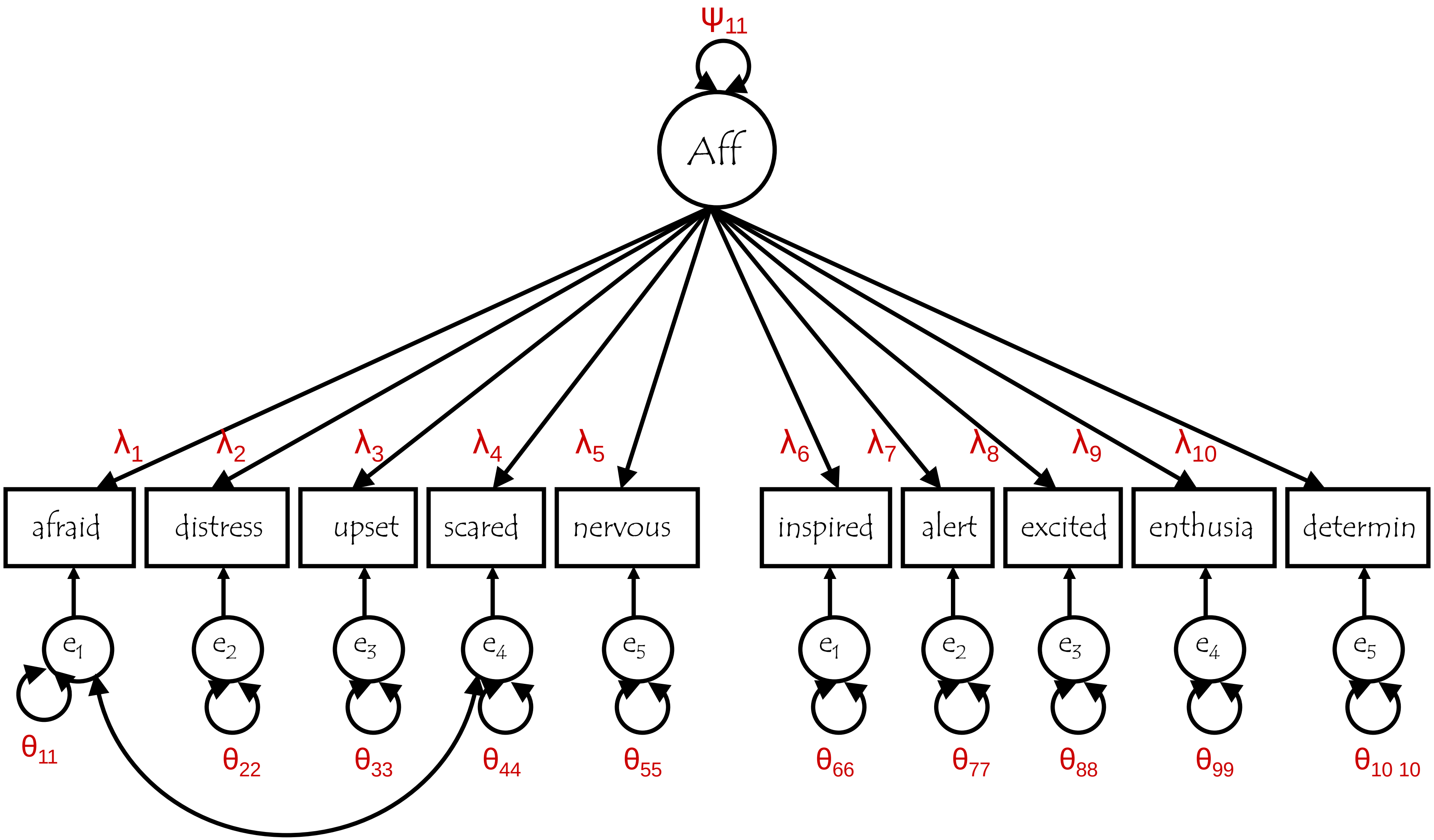
Model fit: Two-factor model

```
MODEL FIT INFORMATION

Number of Free Parameters                         32

Loglikelihood

          H0 Value                         -7323.654
          H1 Value                         -7275.520

Information Criteria

          Akaike (AIC)                     14711.308
          Bayesian (BIC)                   14847.554
          Sample-Size Adjusted BIC         14745.978
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                               96.269
          Degrees of Freedom                      33
          P-Value                             0.0000
```

Model fit: Two-factor model

```
RMSEA (Root Mean Square Error Of Approximation)

          Estimate                          0.061
          90 Percent C.I.                   0.047  0.075
          Probability RMSEA <= .05          0.101

CFI/TLI

          CFI                               0.962
          TLI                               0.948

Chi-Square Test of Model Fit for the Baseline Model

          Value                          1698.761
          Degrees of Freedom                   45
          P-Value                          0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                             0.054
```

One-factor model for NA and PA



DIY

Write down Λ, Ψ, Θ and ν for this one-factor model.

DIY

How can you show that this one-factor model is a special case of the previous two-factor model?

Sample statistics, free parameters and df

Data:

- observed covariance matrix **S** has 10*11/2=55 unique elements
- observed mean vector **m** has 10 elements

TOTAL: 65 sample statistics

Parameter to be estimated:

- 10 factor loadings (in $\mathbf{\Lambda}$); 1 fixed to 1 for scaling, so 9 factor loadings to be estimated
- 10 residual variances (in $\mathbf{\Theta}$)
- 1 covariance between the residuals (in $\mathbf{\Theta}$)
- 1 variance of the factor (in $\mathbf{\Psi}$)
- 10 means (in $\boldsymbol{\nu}$, as we are not modeling the mean structure)

TOTAL: 31 free parameters

Thus: 65 – 31 = 34 degrees of freedom for this model

Model fit: One-factor model

```
Number of Free Parameters                              31

Loglikelihood

          H0 Value                               -7488.452
          H1 Value                               -7275.520

Information Criteria

          Akaike (AIC)                           15038.905
          Bayesian (BIC)                         15170.892
          Sample-Size Adjusted BIC               15072.491
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                                    425.865
          Degrees of Freedom                            34
          P-Value                                   0.0000
```

Chi-square difference test

The one-factor model is **nested under** the two-factor model.

We can compare these models using their chi-squares and dfs:

Chi-square difference: 425.865 – 96.269 = 329.596

Df difference: 34 – 33 = 1

This test is significant. Hence, we have to **reject H0**, which is the **more parsimonious model** (i.e., the one-factor model) in favor of H1, the more general model (i.e., the two-factor model).

Nesting

Nested models can be compared using the **chi-square difference test**.

When models are **not nested**, you **cannot use** this test.

Other ways of comparing models (nested of non-nested) are by use of the **AIC** or **BIC**: lower values point to better models.

Note that it is **not always easy** to see that two models are nested.

Take home message

- **SEM** is concerned with analyzing the **covariance structure** (and sometimes the **mean structure** – Lecture 2).
- **Confirmatory factor analysis** is a special case of SEM.
- All CFA models can be expressed using the **measurement equation**, which relates **observed variables y** to **latent variables η**.
- **Latent variables** always need to be **scaled**.
- The measurement equation is defined using **Λ**, **Ψ** and **Θ**; using these, we obtain the **modeled covariance matrix Σ**.
- **Σ** is compared to the **observed covariance matrix S**.
- There are many **diverse fit indices**.
- Models are **nested** when one is a special case of another; they can be compared with the **chi-square difference test**.
- **All models** can be compared with **AIC or BIC** (if they are based on the **same dataset**! – Lecture 3)

Mplus

Mplus website:

http://www.statmodel.com/

Mplus demo version:

http://www.statmodel.com/demo.shtml

(restricted to 6 observed variables, and 2 latent variables)

User's guide:

http://www.statmodel.com/ugexcerpts.shtml

(chapters 1-14 on diverse modeling options in Mplus)

Mplus 7 manual as pdf:

https://www.statmodel.com/download/usersguide/MplusUserGuideVer_7.pdf

(also includes chapters 15-20, which treat the diverse Mplus modeling commands)

Online Chi-square calculator:

http://www.fourmilab.ch/rpkp/experiments/analysis/chiCalc.html