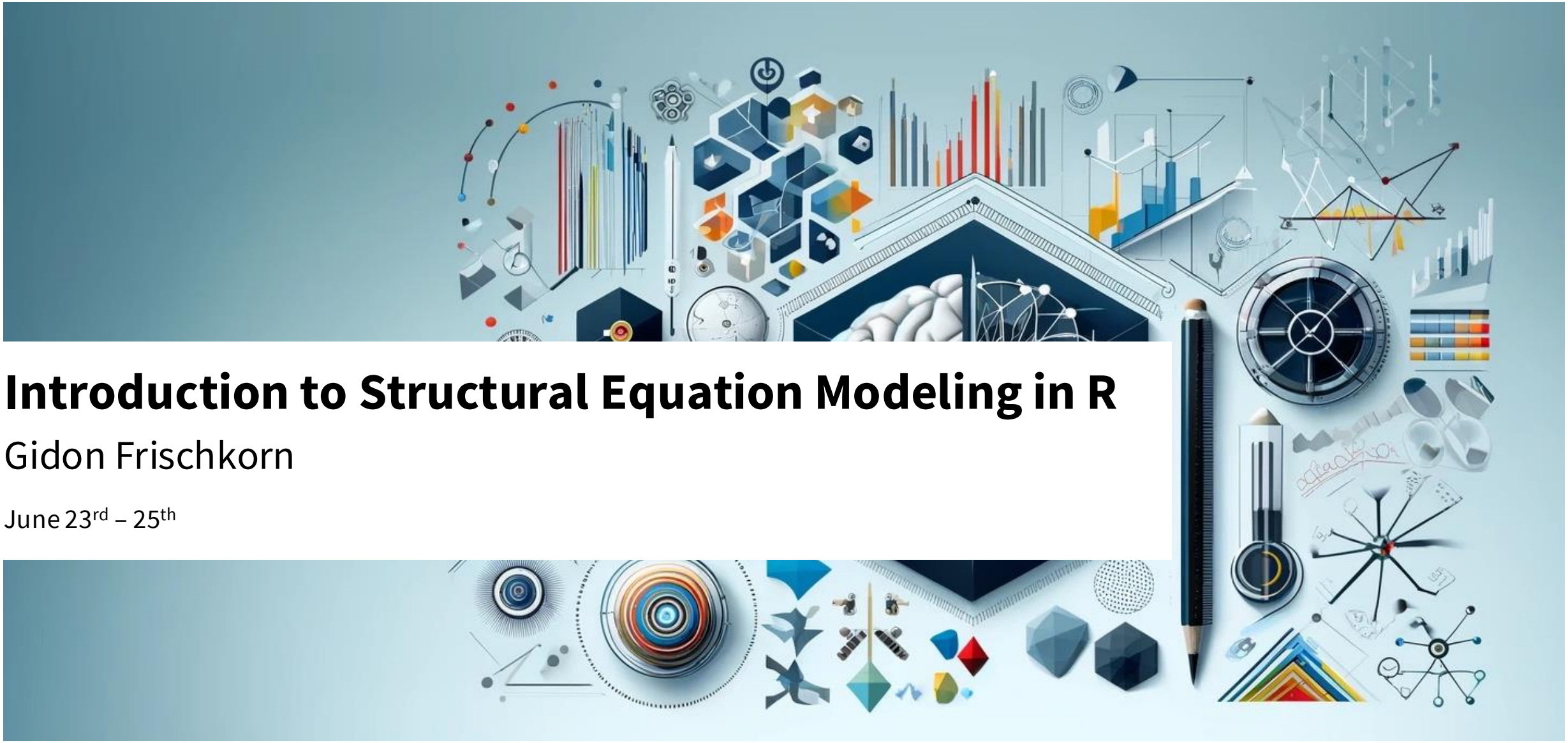
Universität Zürich[UZH]

Psychologisches Institut

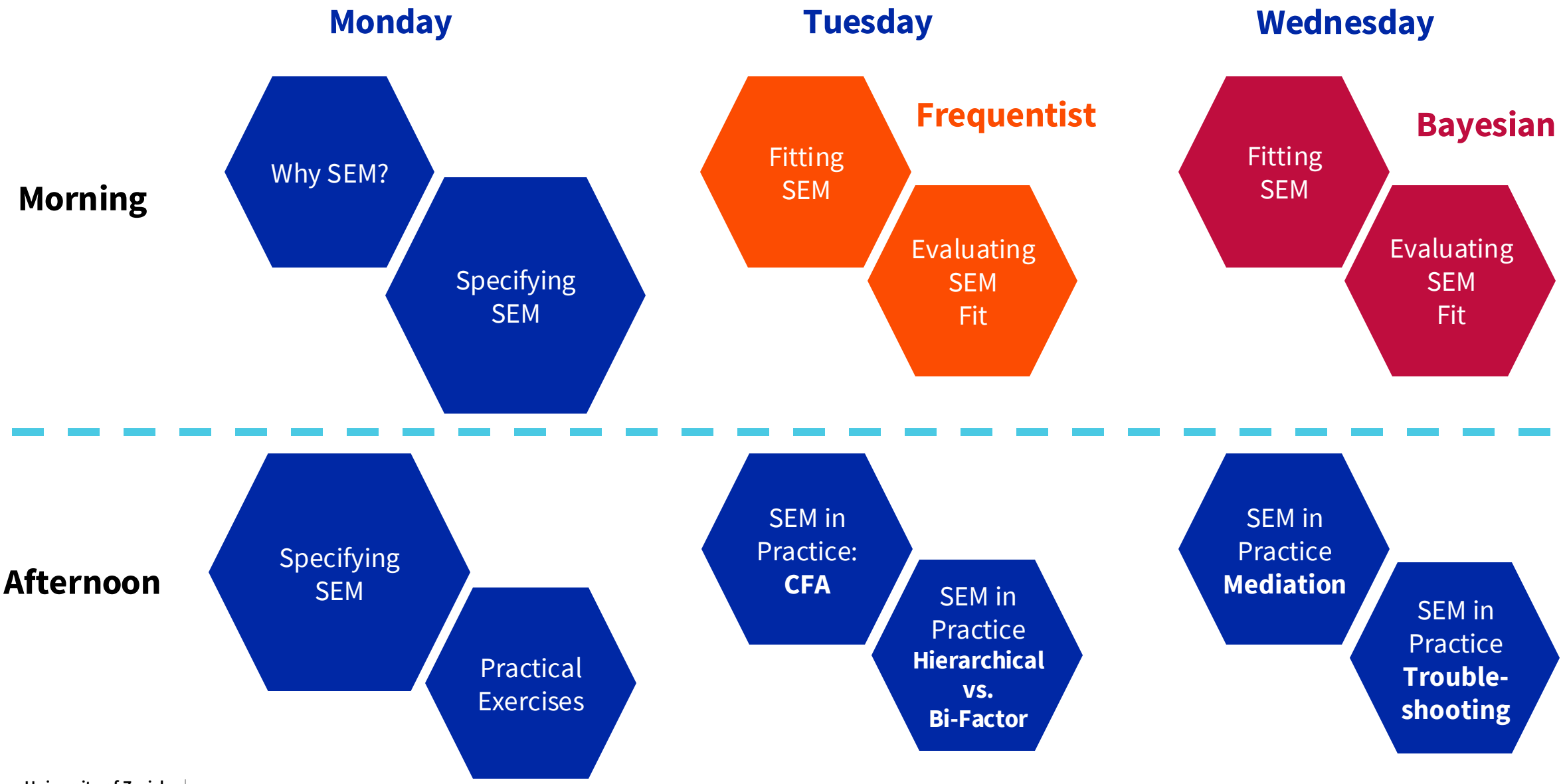


Introduction to Structural Equation Modeling in R

Gidon Frischkorn

June 23rd – 25th

SEM Workshop: Overview



How to fit SEM to data?

Fitting SEM

Specified SEM

Implies covariance structure for observed vars.

Given a set of parameters θ → Implied Covariance Matrix

$\Sigma(\theta)$

Empirical Data

Observed Covariance Matrix

Model Fitting

Find parameter set that minimizes the difference between these covariance matrices

$F[\hat{\Sigma}, \Sigma(\theta)]$

Discrepancy Function →summarize difference into one value

Unweighted / Ordinary Least Squares ULS

Generalized Least Squares GLS

Maximum Likelihood ML

Repetition: To be estimated parameters?

Variances of exogeneous variables

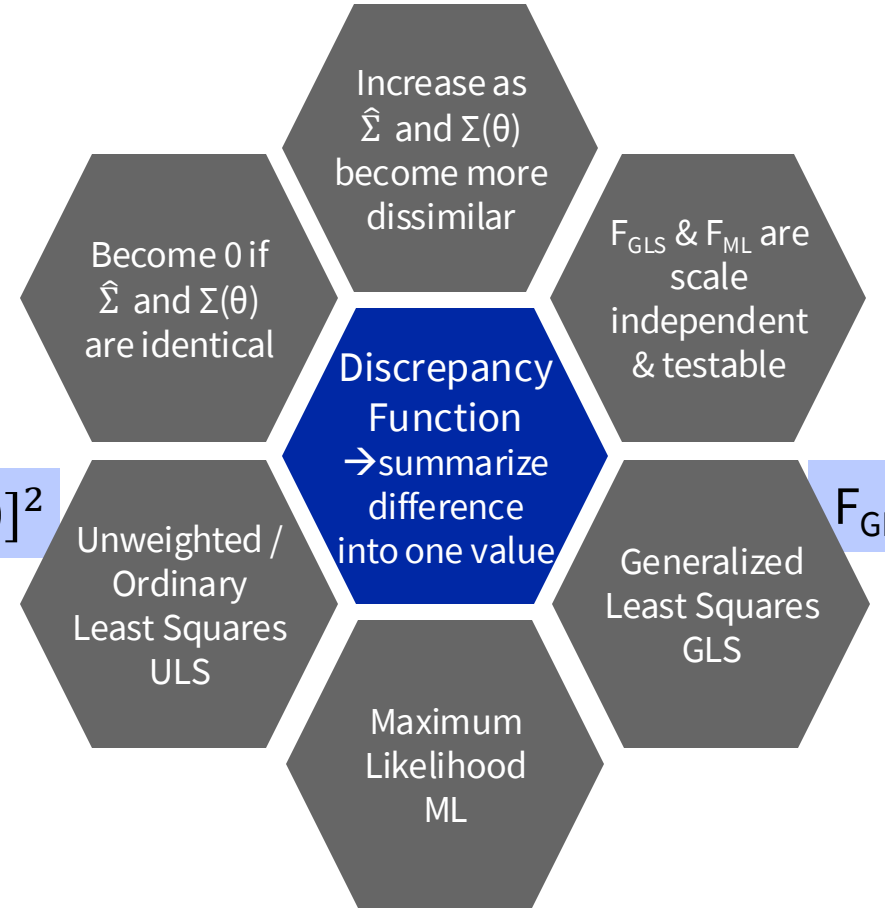
Path Coefficients Loadings / Regression Weights / Covariances

$\mathbf{S}$ = sample covariance matrix

$\hat{\mathbf{\Sigma}}$ = estimated population covariance matrix

How to fit SEM to data?

Fitting SEM



$$F_{ULS}[\hat{\Sigma}, \Sigma(\theta)] = \frac{1}{2} tr[\hat{\Sigma} - \Sigma(\theta)]^2$$

$$F_{GLS}[\hat{\Sigma}, \Sigma(\theta)] = \frac{1}{2} tr[\hat{\Sigma}^{-1}(\hat{\Sigma} - \Sigma(\theta))]^2$$

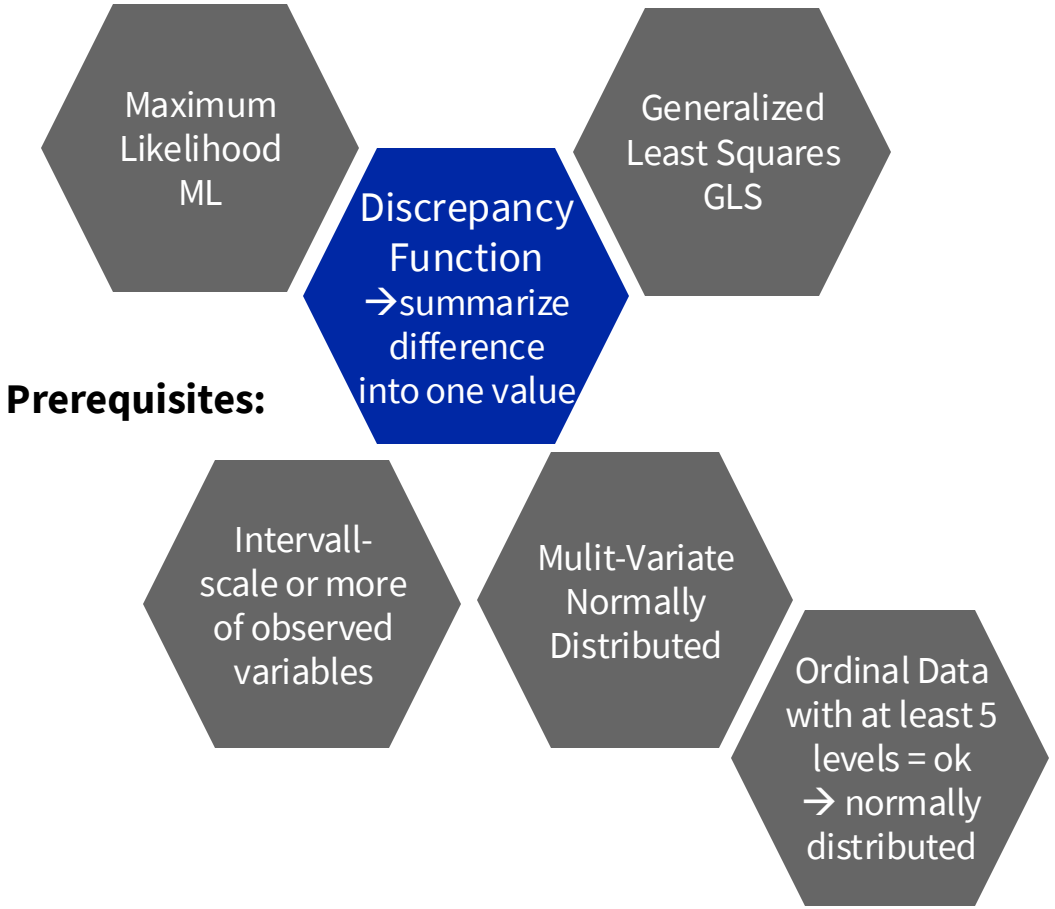
$$F_{ML}[\hat{\Sigma}, \Sigma(\theta)] = \log|\Sigma(\theta)| + tr\left[\Sigma(\theta)^{-1}\hat{\Sigma}\right] - \log|\hat{\Sigma}| - p$$

Lavaan Default: ML

GLS is also fine

How to fit SEM to data?

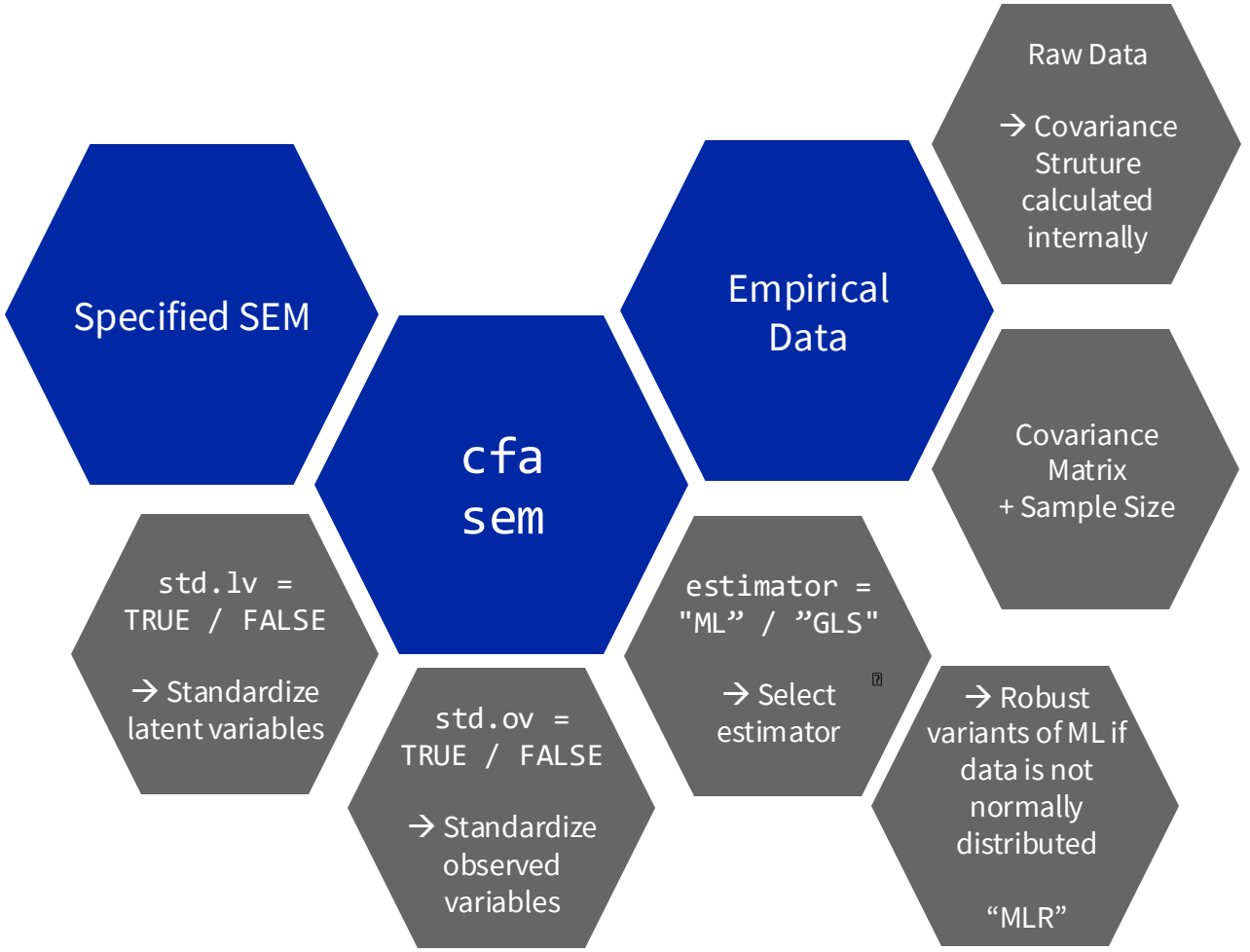
Fitting SEM



How to fit SEM to data?

→ lavaan

Fitting SEM



How to fit SEM to data?

→ Sample Size

Fitting SEM

Reliability of parameter estimation

Minimal Sample Size = 100 (Boomsma, 1982)

Better N > 200 (Hoyle, 1995)

Sample Size

Larger sample = higher reliability

Parameter Values

Esp. path coefficients

Higher values = higher reliability

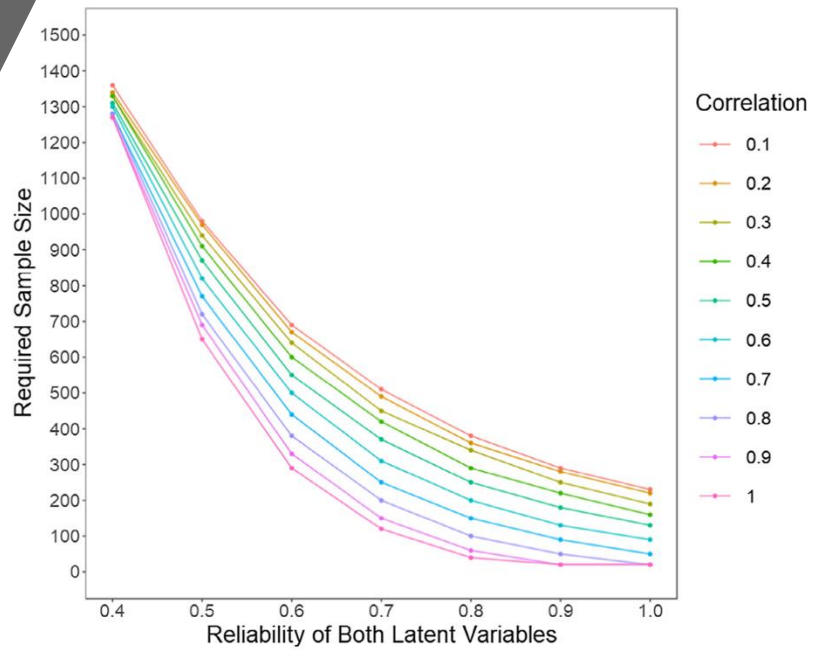
Number of Model Parameters

Less parameters = higher reliability

At least 5 subjects per parameter (Bentler & Chou, 1987)

Ideally: Sample Size planning with simulation or software

For an extensive intro, see the `semPower2` Manual



What are your questions so far?

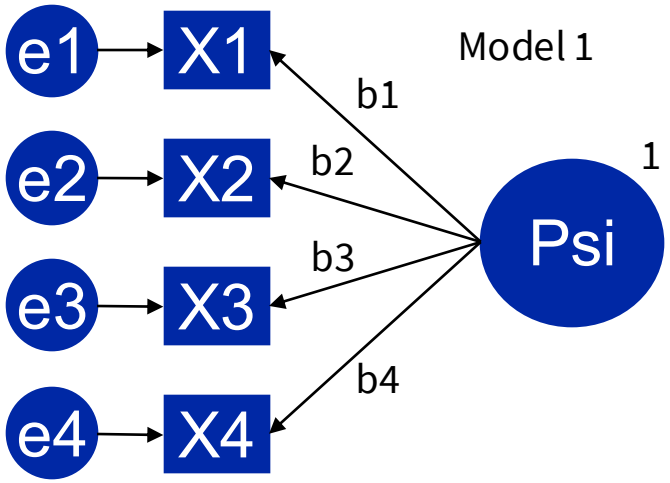
How to fit SEM to data?

Fitting SEM

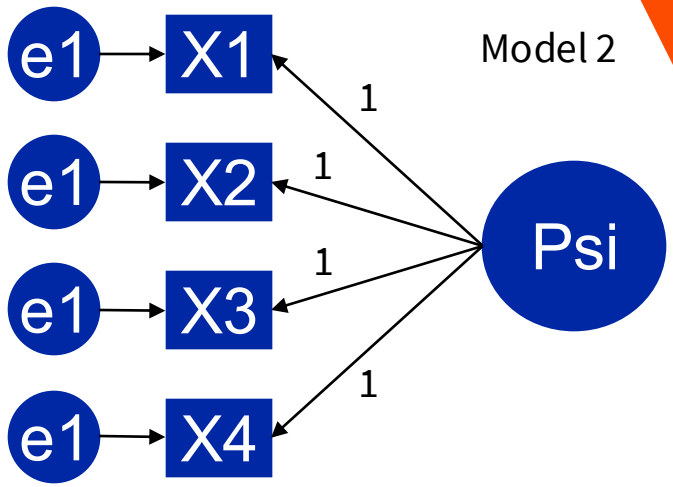
Exercise: Fit these models to data you simulated

```
# fit SEM
my_fit <- sem(model, data)
summary(my_fit)
```

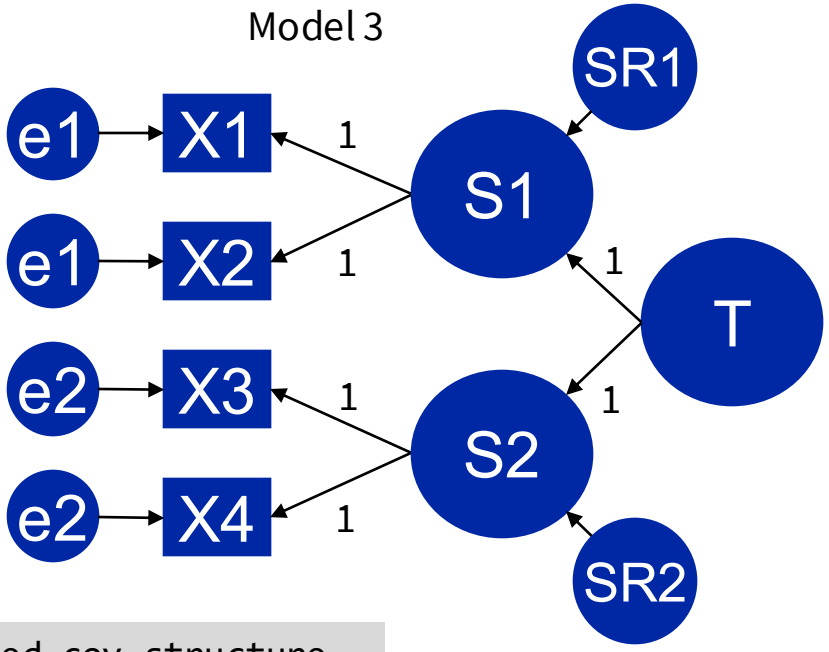
Model 1



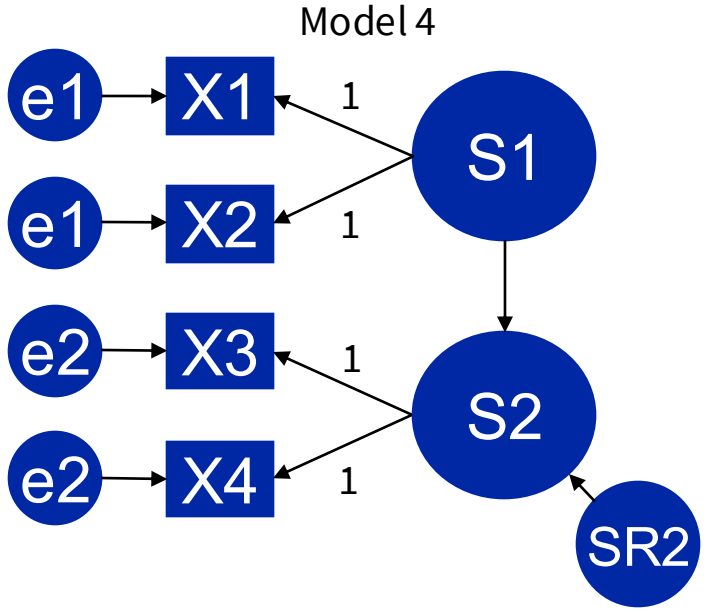
Model 2



Model 3



Model 4



Try fitting different models to the same data

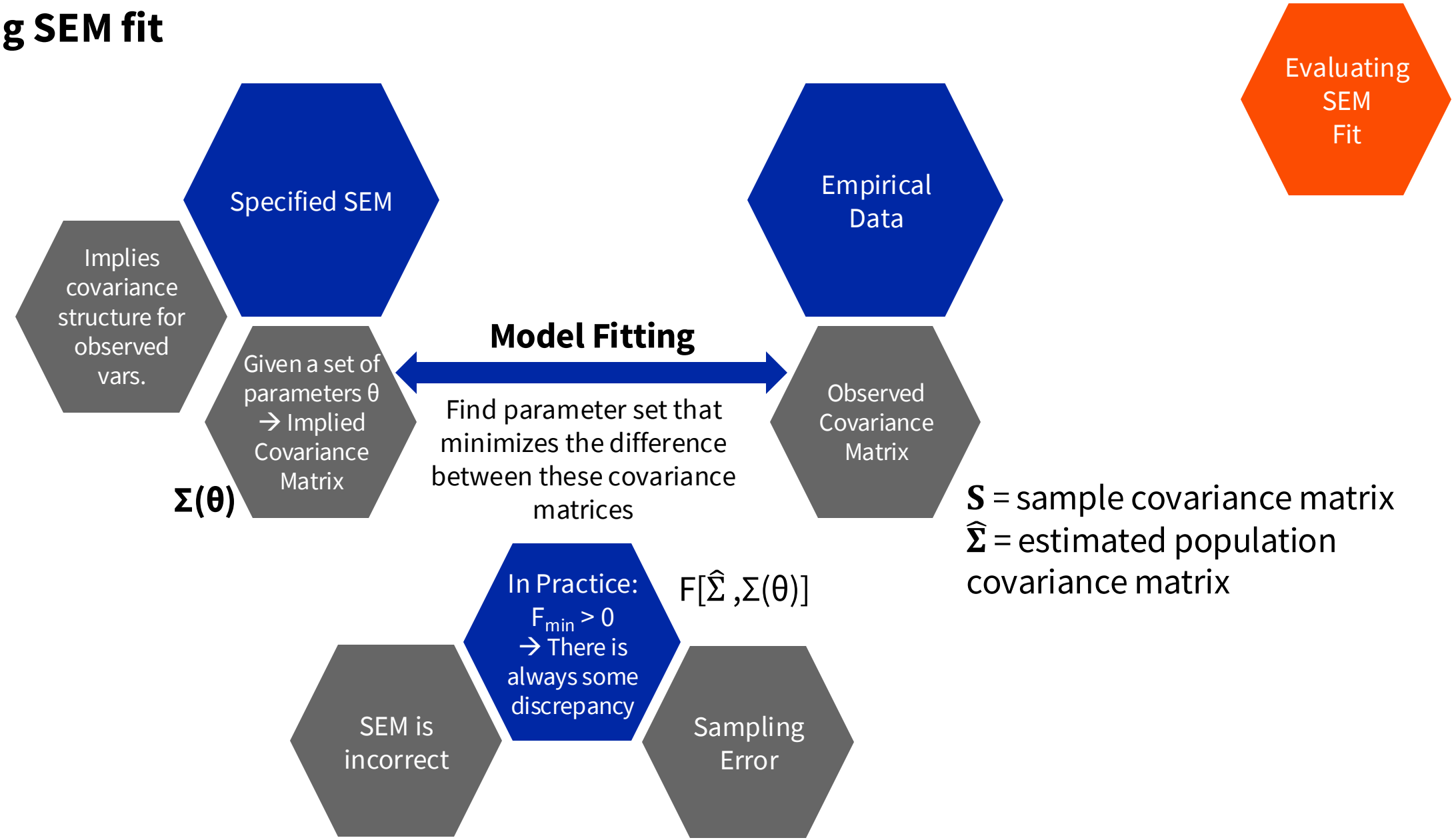
Be aware: change the model syntax compared to simulation

Vary the sample size used to simulate data

Check the implied covariance structure

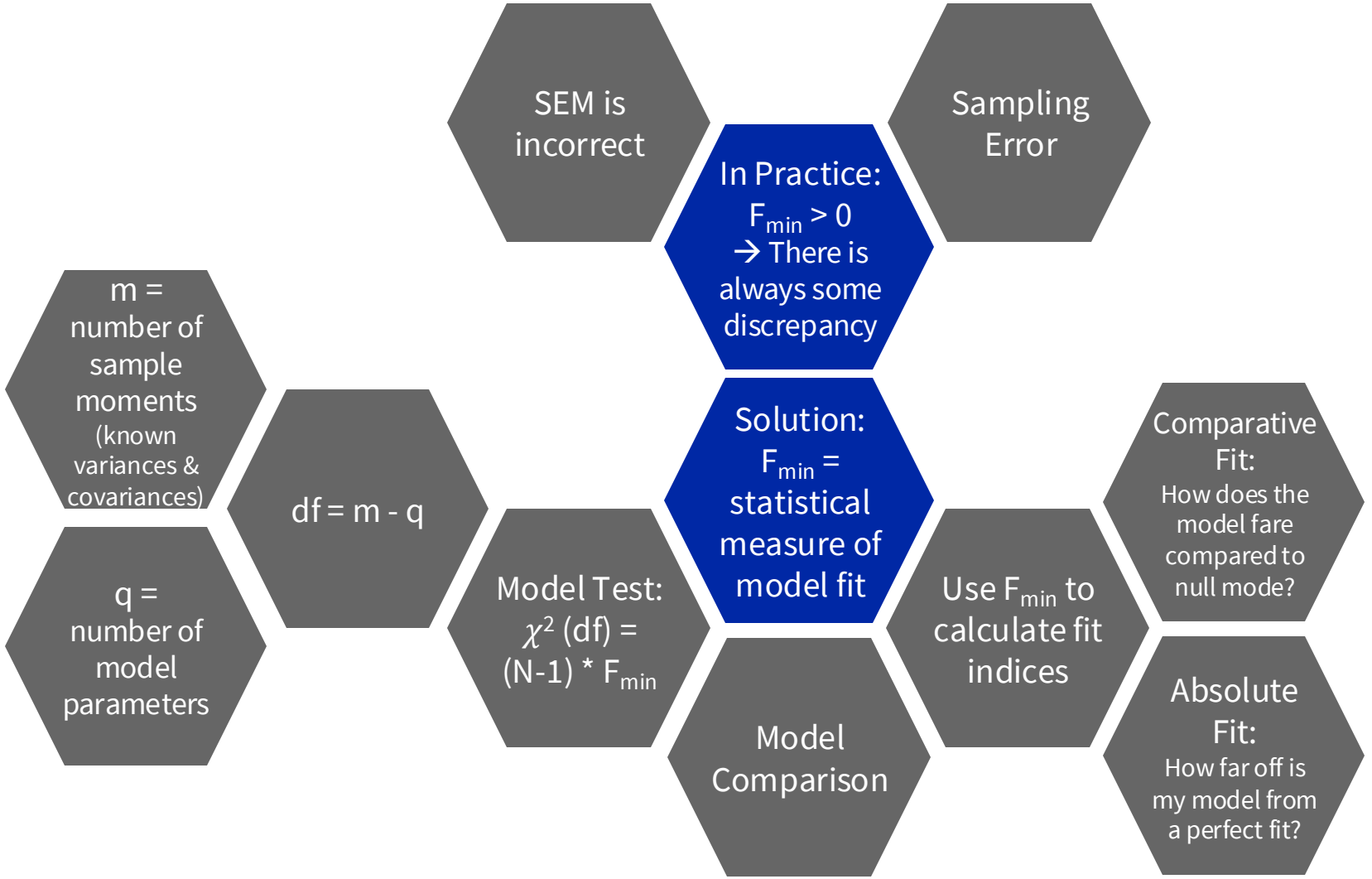
```
# implied cov.structure
fitted(my_fit)
```

Evaluating SEM fit



$\mathbf{S}$ = sample covariance matrix
$\hat{\boldsymbol{\Sigma}}$ = estimated population covariance matrix

Evaluating SEM fit



Evaluating SEM fit
Model test

Evaluating SEM Fit

m = number of sample moments (known variances & covariances)

q = number of model parameters

df = m - q

χ^2 (df) = (N-1) * F_{min}

Test Null Hypothesis: There is no difference between $\hat{\Sigma}$ and $\Sigma(\theta)$

p > .05 ☺ accept H_0 **but:** unknown power

p < .05 ☹ reject H_0

But: any model without perfect fit ($F_{min} > 0$) will be rejected in large samples

All models are simplified → Perfect fit is unlikely

Model Test = Heuristic

Evaluating SEM fit
Model Fit Indices

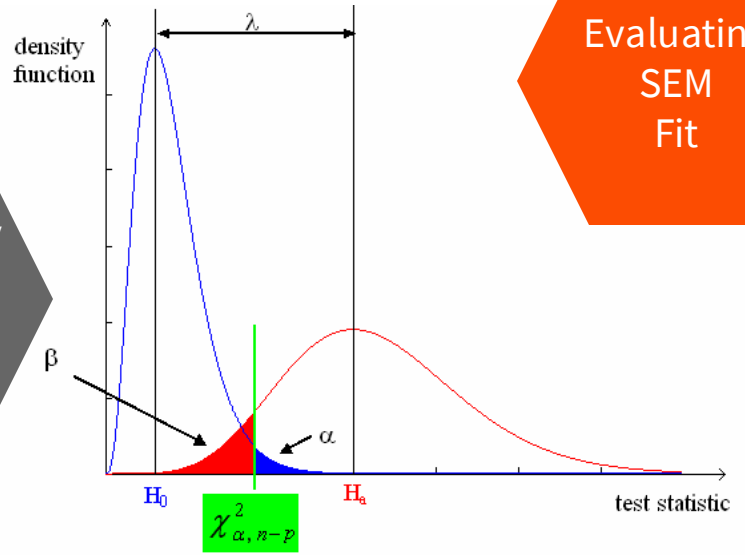
Evaluating SEM Fit

	X1	X2	X3	X4
X1	Var(X1)	0	0	0
X2		Var(X2)	0	0
X3			Var(X3)	0
X4				Var(X4)

Baseline Model: All observed variables are uncorrelated

Standardize F_{min} to make it interpretable

Noncentrality Parameter: $\lambda = \frac{\chi^2 - df}{N}$



Goodness of Fit Indices

Comparative Fit: How does the model fare compared to null mode?

Compare non-centrality of target Model Z (λ_Z) to Baseline B (λ_β)

Comparative Fit Index (CFI) $CFI = 1 - \frac{\widetilde{\lambda_Z}}{\widetilde{\lambda_\beta}}$

Non-centrality is trimmed: $\widetilde{\lambda_\beta} = \max(0, \lambda_\beta)$ $\widetilde{\lambda_Z} = \max(0, \lambda_Z)$

Interpretation acceptable: CFI > .90 good: CFI > .95

Absolute Fit: How far off is my model from a perfect fit?

Root Mean Square Error of Approximation (RMSEA) $RMSEA = \sqrt{\frac{\widetilde{\lambda_Z}}{df}}$

Interpretation acceptable: RMSEA < .08 good: RMSEA < .05

Penalizes more complex models → Parsimoniy is rewarded

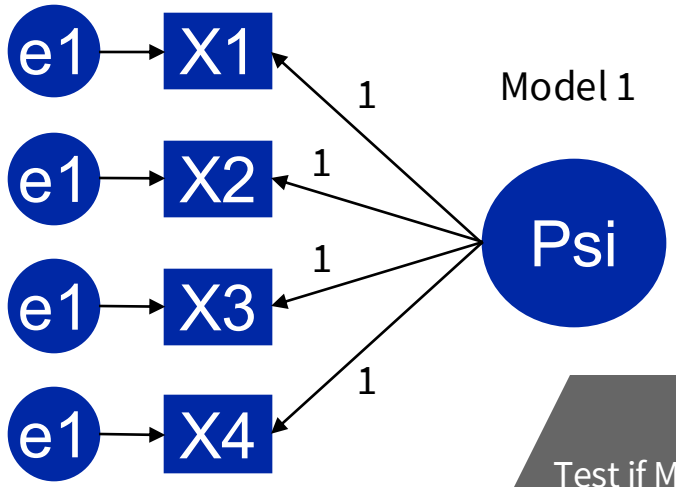
Can be used for sample size planning

```
# get Model fit indices
my_fit <- sem(model, data)
fitMeasures(my_fit)
```

Evaluating SEM fit
Model Fit Indices

Evaluating SEM Fit

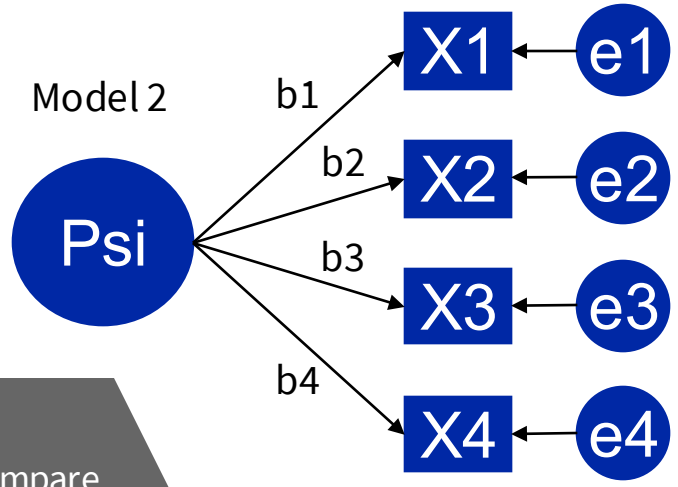
Nested Models
One model is a constrained version of the other



Model 1

Model Comparison

Model 2



Test if Models differ:
$\Delta_{\chi^2} = \chi^2_1 - \chi^2_2$

Δ_{χ^2} = chi-square distributed

$df_\Delta = df_1 - df_2$

Compare Likelihood based Indices

AIC = 2*q – 2LL

BIC = q * log(N) – 2LL

Model with smaller values wins

$\Delta^{AIC/BIC} > 10$ substantial

CFI / RMSEA not well suited for model comparison

```
# fit SEM
my_fit1 <- sem(model1, data)
my_fit2 <- sem(model2, data)

# compare models
lavTestLRT(my_fit1, my_fit2)
```

What are your questions so far?

Evaluating SEM fit

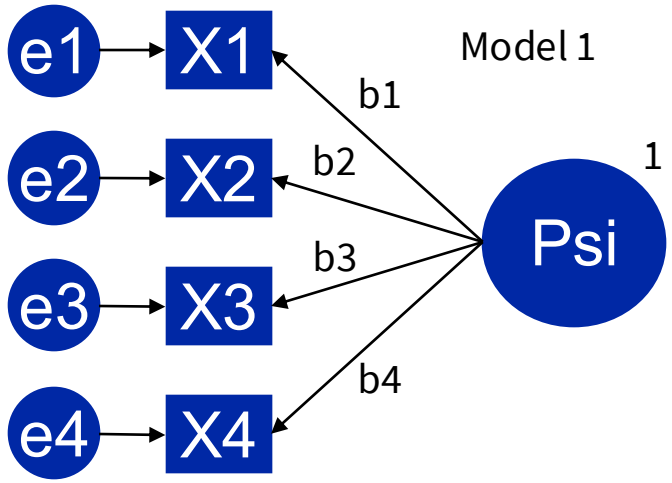
Exercise: Choose simulated data from one model

Fit all four models to the data set

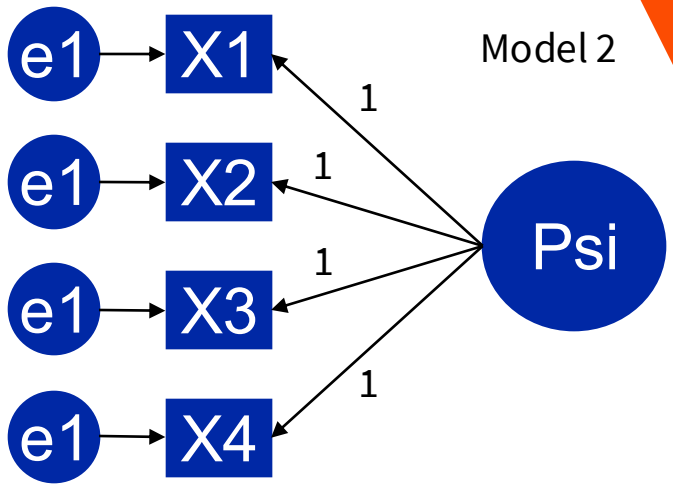
Evaluate fit of every model on its own

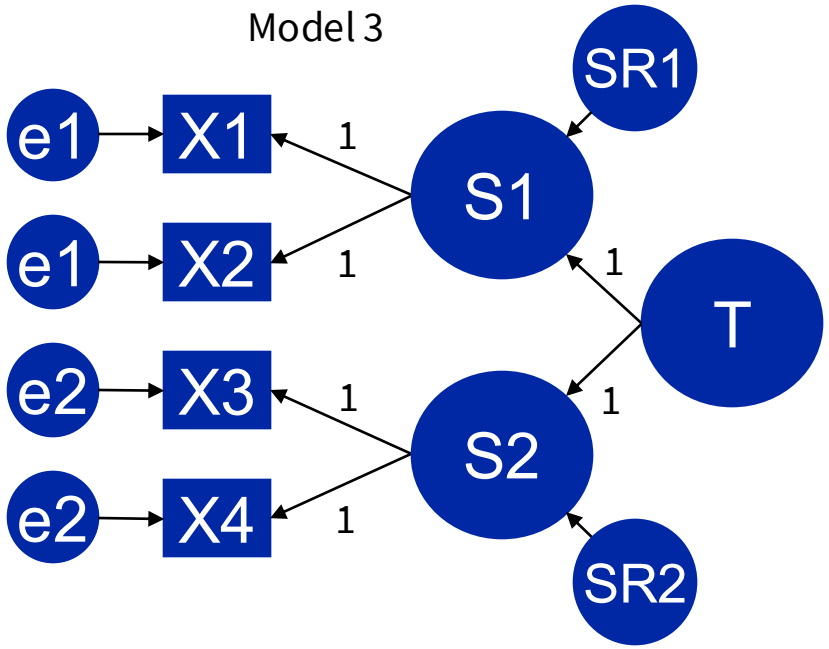
Compare the models to each other

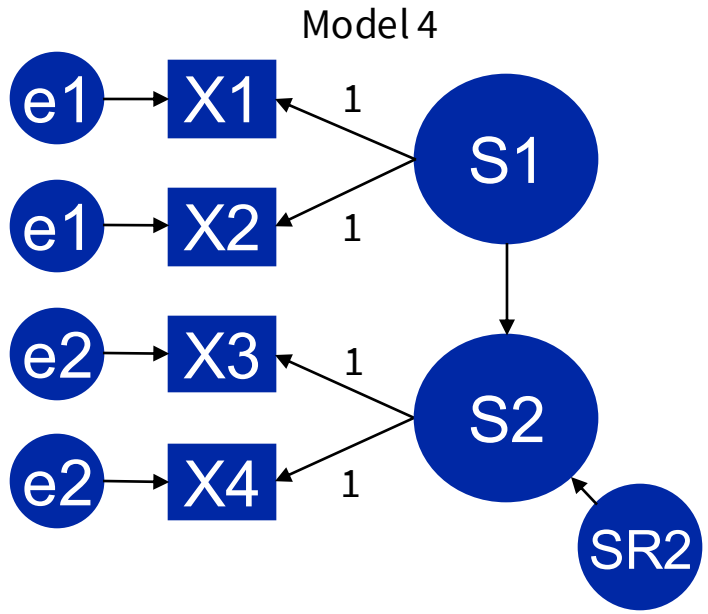
Does the correct model win?



Evaluating SEM Fit







Evaluating SEM fit
Model Fit Indices

Evaluating SEM Fit

We have a well fitting model ☺

→ Test if a model parameters *q* is significant

H_0 = the model parameter is zero

Critical Ratio Test:
$CR = \frac{\widehat{\theta_q}}{\widehat{SE_q}}$

p > .05
☹
Re-estimate model with parameter = 0

p < .05
☺
Model parameter is not zero

```
# get Model fit indices
my_fit <- sem(model, data)

# print model summary with fit indices
# and standardized parameter estimates
summary(my_fit,
        fit.measures = TRUE,
        standardized = TRUE)
```

Improper Solutions

Estimated Variance is negative

Correlation is larger than 1

If p > .05
Fix variance to zero and re-estimate model

If p < .05

Sometime correlation can be constrained to one

Bayesian Estimation Techniques can help ☺

Evaluating SEM fit

Exercise: Choose simulated data from one model

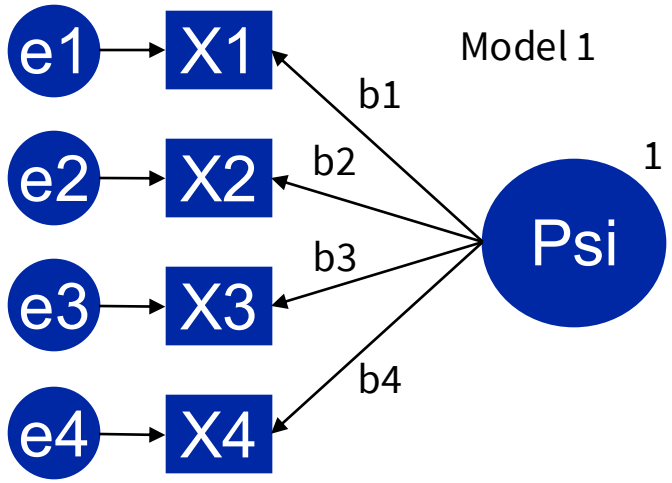
Fit all four models to the data set

Evaluate fit of every model on its own

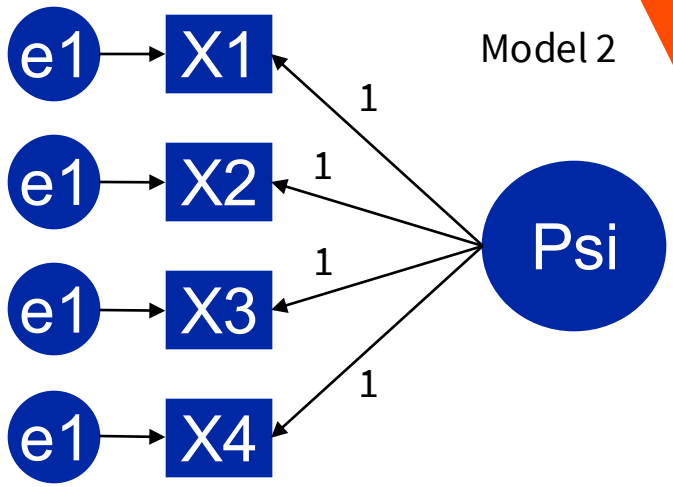
Compare the models to each other

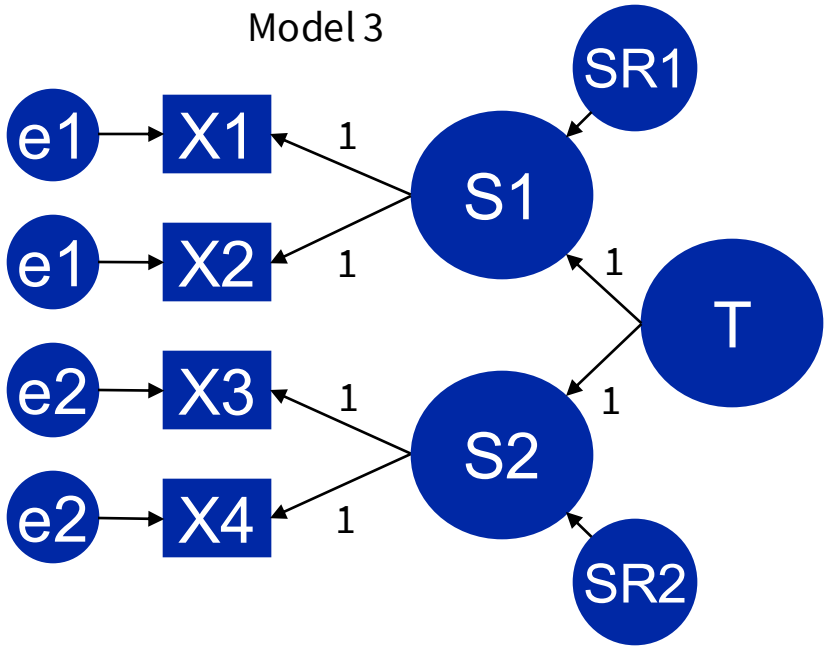
Does the correct model win?

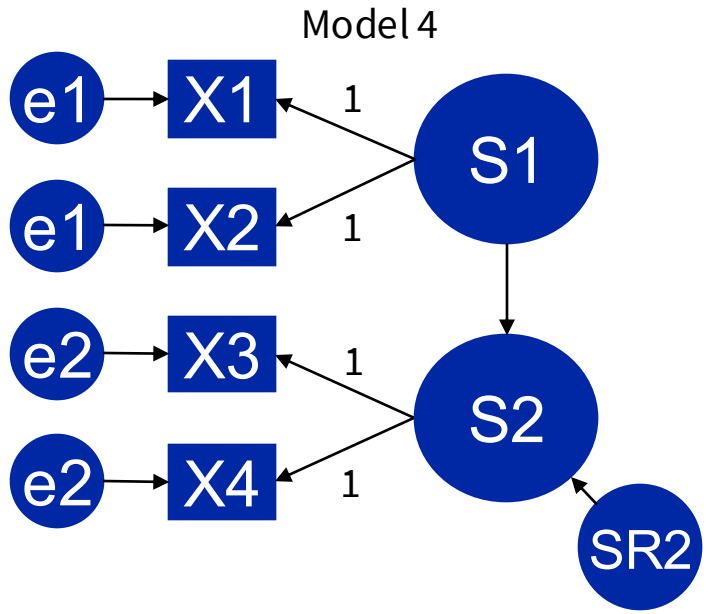
For winning model: check if parameters are significant



Evaluating SEM Fit







What are your questions so far?