Intro to Structural Equation Modeling in R

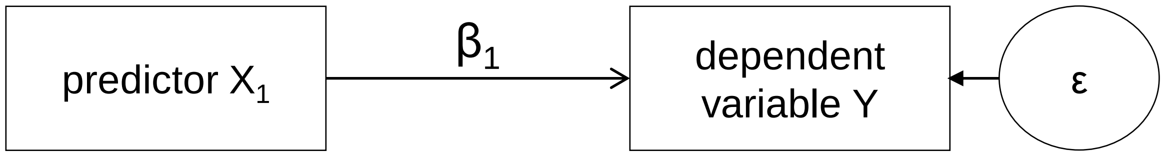
Lara Kobilke & Magdalena Obermaier
Department of Media and Communication
LMU (aka Lavaan Masters University) Munich

Computational Exchange Sessions | 17 July 2024

INTRO I: THEORETICAL ASSUMPTIONS

Intro I: Theoretical Assumptions

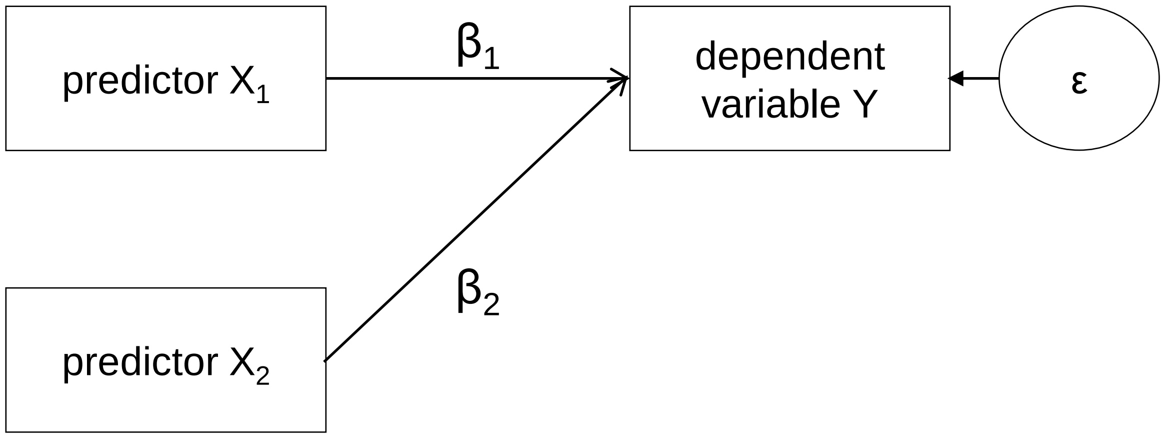
Simple linear regression



$$Y = \alpha + \beta_1 \cdot X_1 + \varepsilon$$

Intro I: Theoretical Assumptions

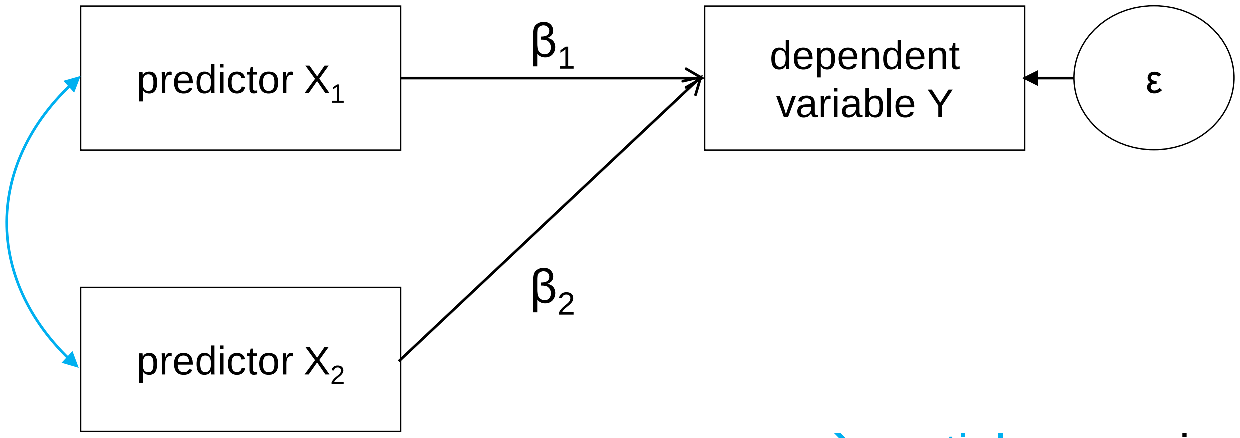
Multiple linear regression



$$Y = \alpha + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \varepsilon$$

Intro I: Theoretical Assumptions

Multiple linear regression



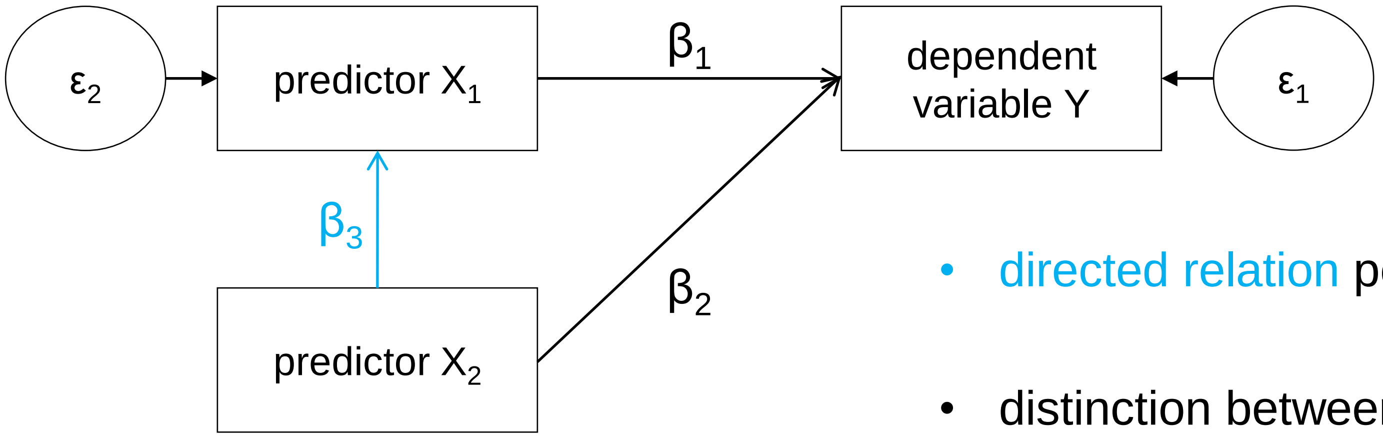
$$Y = \alpha + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \varepsilon$$

→ partial regression coefficient

β_i takes predictors‘ covariances into account

Intro I: Theoretical Assumptions

Path analysis



- directed relation postulation between predictors possible

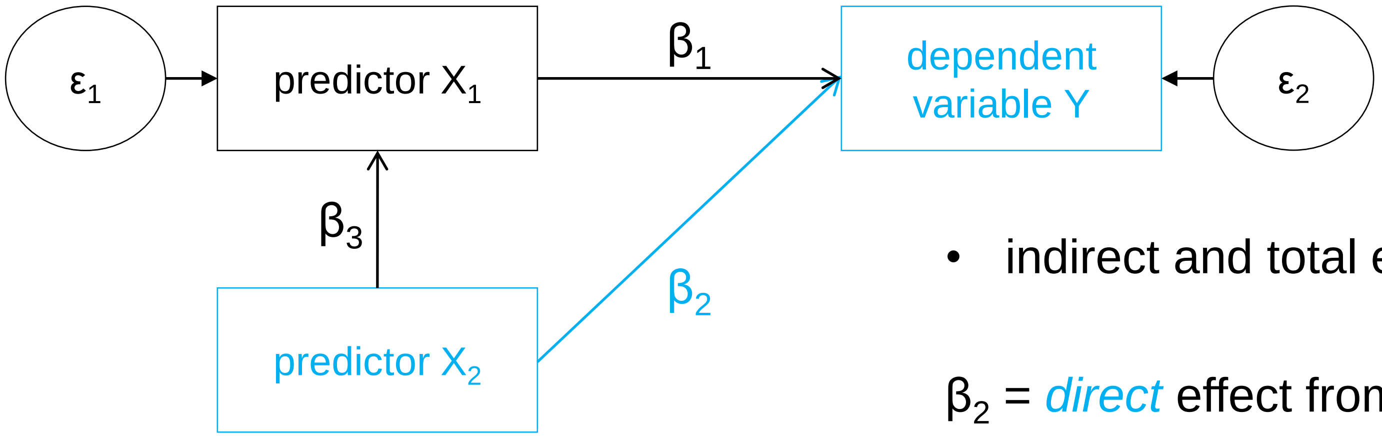
- distinction between *exogenous* (X_2) und *endogenous* (X_1, Y) variables (additionally: X_1 = *intervening* variable)

$$Y = \alpha_1 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \varepsilon_1$$

$$X_1 = \alpha_2 + \beta_3 \cdot X_2 + \varepsilon_2$$

Intro I: Theoretical Assumptions

Path analysis



- indirect and total effects

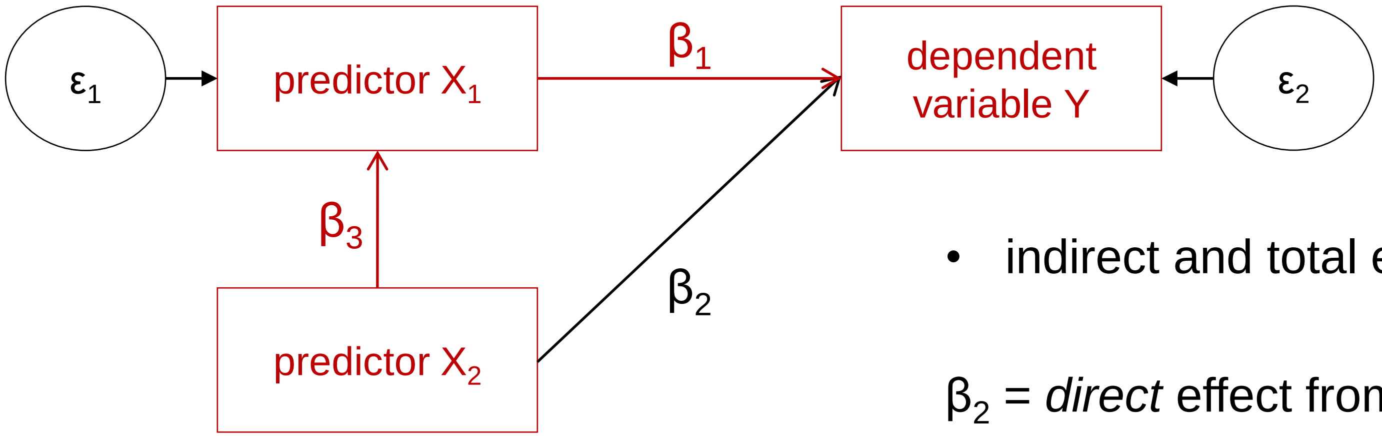
β_2 = *direct* effect from X_2 on Y
$\beta_3 \cdot \beta_1$ = *indirect* effect from X_2 on Y via mediating X_1
$\beta_2 + \beta_3 \cdot \beta_1$ = *total* effect from X_2 on Y

$$Y = \alpha_1 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \varepsilon_1$$

$$X_1 = \alpha_2 + \beta_3 \cdot X_2 + \varepsilon_2$$

Intro I: Theoretical Assumptions

Path analysis



- indirect and total effects

β_2 = *direct* effect from X_2 on Y

$\beta_3 \cdot \beta_1$ = *indirect* effect from X_2 on Y via mediating X_1

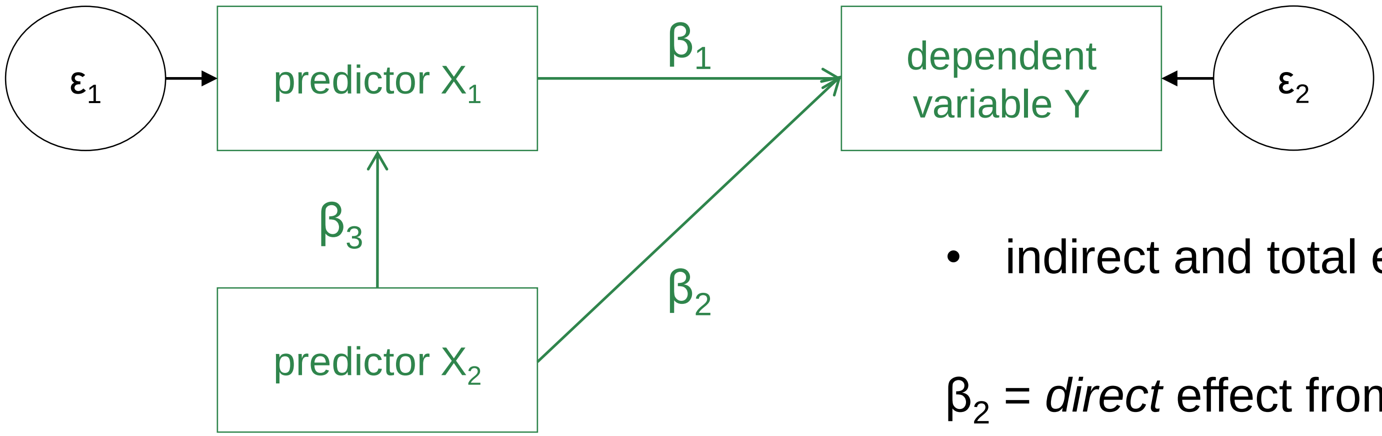
$\beta_2 + \beta_3 \cdot \beta_1$ = *total* effect from X_2 on Y

$$Y = \alpha_1 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \varepsilon_1$$

$$X_1 = \alpha_2 + \beta_3 \cdot X_2 + \varepsilon_2$$

Intro I: Theoretical Assumptions

Path analysis



- indirect and total effects

β_2 = *direct* effect from X_2 on Y

$\beta_3 \cdot \beta_1$ = *indirect* effect from X_2 on Y via mediating X_1

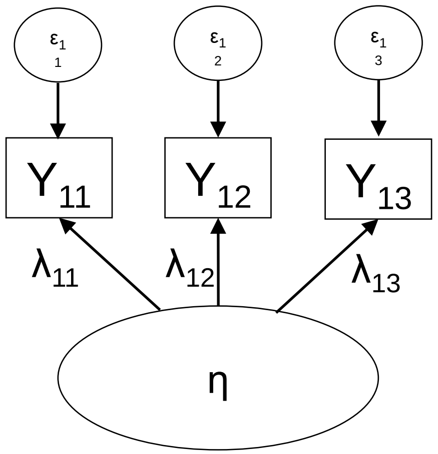
$\beta_2 + \beta_3 \cdot \beta_1$ = *total* effect from X_2 on Y

$$Y = \alpha_1 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \varepsilon_1$$

$$X_1 = \alpha_2 + \beta_3 \cdot X_2 + \varepsilon_2$$

Intro I: Theoretical Assumptions

Factor analysis



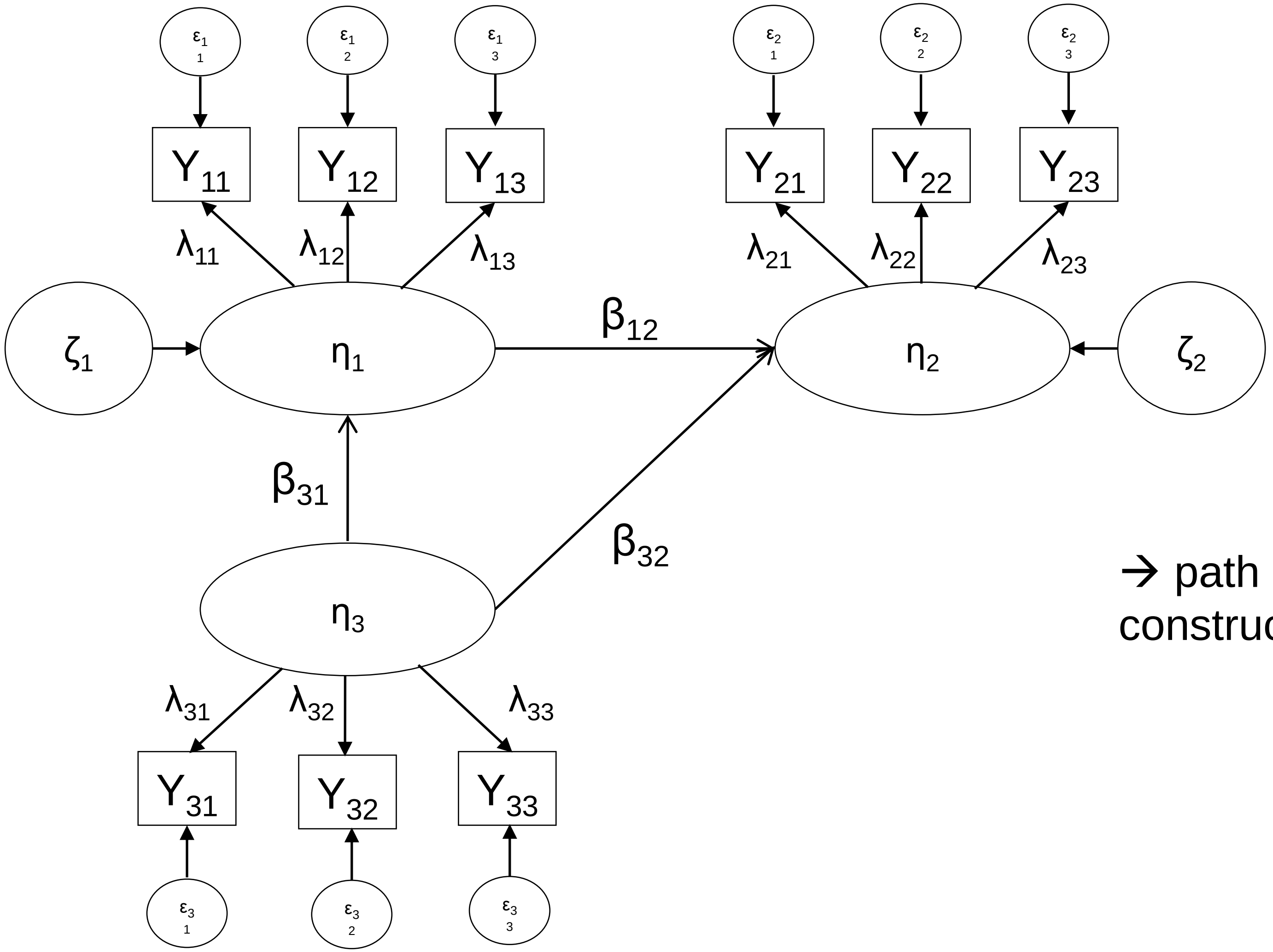
- relationships between *manifest variables* are understood as an expression of a mutual *latent variable* (factors)

- exploratory (EFA) vs
  confirmatory factor analysis (CFA)

ε_{ij} = residuals
Y_{ij} = manifest variables
λ_{ij} = factor loadings
η_i = latent variable(s)

Intro I: Theoretical Assumptions

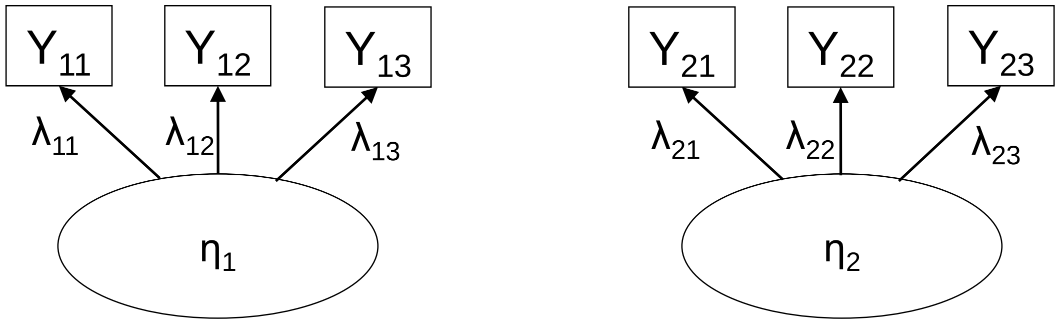
Linear structural equation models

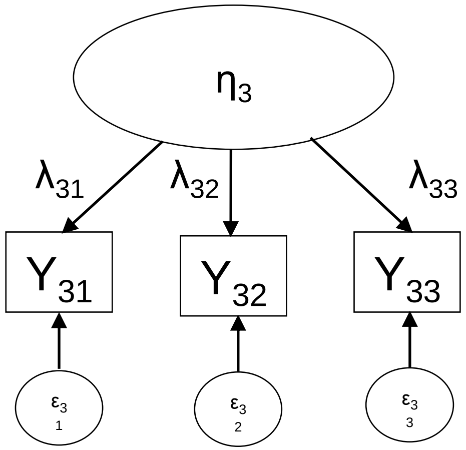


→ path analysis extension using latent constructs

Intro I: Theoretical Assumptions

Linear structural equation models: Measurement model



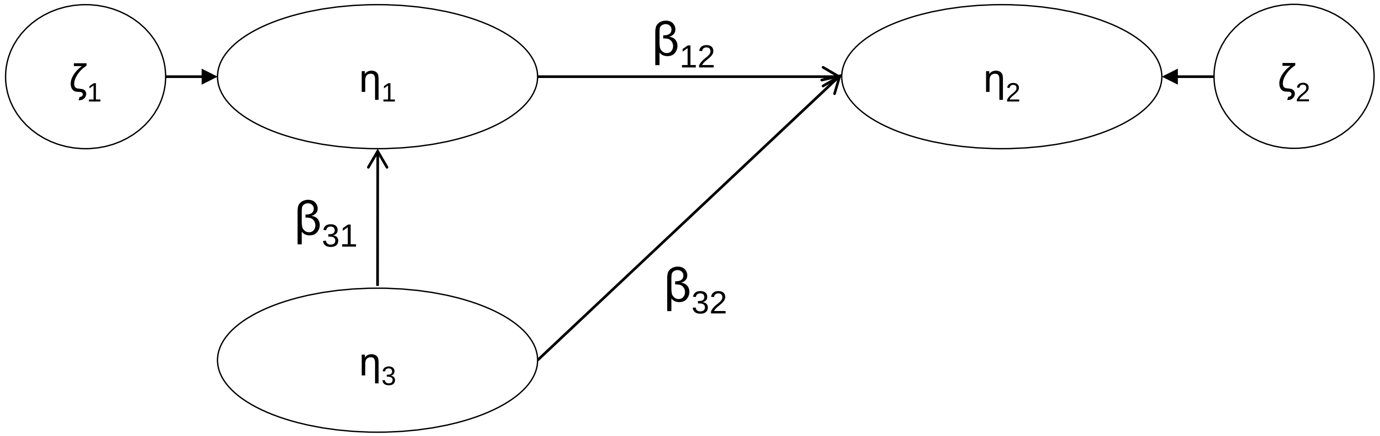


- *latent variables* are measured through manifest variables
- assumption: latent variable causes covariance

- connection of manifest/latent variables using regression (λ_{ij} = regression coefficient or factor loadings)
- dependent variables: indicators,
  independent variables: factors (*reflective meansurement model*)

Intro I: Theoretical Assumptions

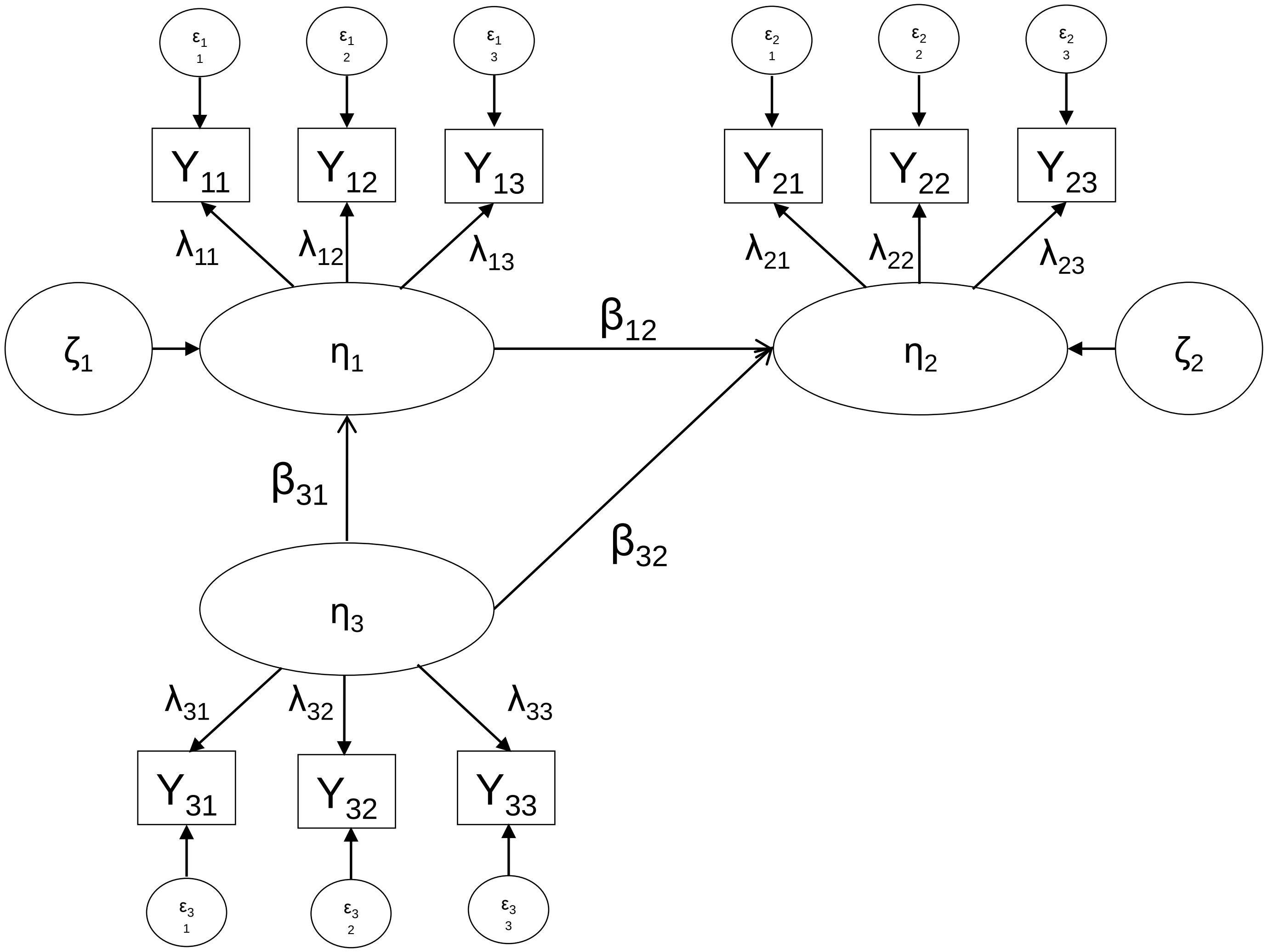
Linear structural equation models: Structural model



- specified relationships between latent variables

Intro I: Theoretical Assumptions

Linear structural equation models: Hypothesized model



Intro I: Theoretical Assumptions

Linear structural equation models

What you get

- Test complex relationships (using latent constructs = less measurement errors)
- Test directed relations between multiple (in)dependent variables
- Take measurement errors into account → more precise than factor/regression analysis combination
- Use multiple indicators per construct
- Model comparisons possible

What you need

- Theory
- Approx. three manifest variables per latent construct
- According sample size (200+) (e.g., https://www.danielsoper.com/statcalc/calculator.aspx?id=89)
- Frustration Tolerance: multiple model re-specification necessary

INTRO II: VALIDATING MODEL FIT (USING LAVAAN)

Intro II: Validating Model Fit (LAVAAN)

A good linear SEM should …

- fit on the data
- be theoretically meaningful
- be as slim as possible
- explain variance
- be replicable

Intro II: Validating Model Fit (LAVAAN)

Tests of model fit

- describe how well a SEM fits our data

```
lavaan 0.6-18 ended normally after 71 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        47

  Number of observations                           604
  Number of missing patterns                         1

Model Test User Model:

  Test statistic                               137.469
  Degrees of freedom                                68
  P-value (Chi-square)                           0.000

Model Test Baseline Model:

  Test statistic                              4479.201
  Degrees of freedom                                95
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.984
  Tucker-Lewis Index (TLI)                       0.978

  Robust Comparative Fit Index (CFI)                NA
  Robust Tucker-Lewis Index (TLI)                   NA

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -7361.741
  Loglikelihood unrestricted model (H1)      -7293.006

  Akaike (AIC)                               14817.482
  Bayesian (BIC)                             15024.450
  Sample-size adjusted Bayesian (SABIC)      14875.236

Root Mean Square Error of Approximation:

  RMSEA                                          0.041
  90 Percent confidence interval - lower         0.031
  90 Percent confidence interval - upper         0.051
  P-value H_0: RMSEA <= 0.050                    0.929
  P-value H_0: RMSEA >= 0.080                    0.000

  Robust RMSEA                                   0.041
  90 Percent confidence interval - lower         0.031
  90 Percent confidence interval - upper         0.051
  P-value H_0: Robust RMSEA <= 0.050             0.929
  P-value H_0: Robust RMSEA >= 0.080             0.000

Standardized Root Mean Square Residual:

  SRMR                                           0.030
```

In it (LAVAAN)

```
lavaan 0.6-18 ended normally after 71 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        47

  Number of observations                           604
  Number of missing patterns                         1

Model Test User Model:

  Test statistic                               137.469
  Degrees of freedom                                68
  P-value (Chi-square)                           0.000

Model Test Baseline Model:

  Test statistic                              4479.201
  Degrees of freedom                                95
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.984
  Tucker-Lewis Index (TLI)                       0.978

  Robust Comparative Fit Index (CFI)                NA
  Robust Tucker-Lewis Index (TLI)                   NA

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -7361.741
  Loglikelihood unrestricted model (H1)      -7293.006

  Akaike (AIC)                               14817.482
  Bayesian (BIC)                             15024.450
  Sample-size adjusted Bayesian (SABIC)      14875.236

Root Mean Square Error of Approximation:

  RMSEA                                          0.041
  90 Percent confidence interval - lower         0.031
  90 Percent confidence interval - upper         0.051
  P-value H_0: RMSEA <= 0.050                    0.929
  P-value H_0: RMSEA >= 0.080                    0.000

  Robust RMSEA                                   0.041
  90 Percent confidence interval - lower         0.031
  90 Percent confidence interval - upper         0.051
  P-value H_0: Robust RMSEA <= 0.050             0.929
  P-value H_0: Robust RMSEA >= 0.080             0.000

Standardized Root Mean Square Residual:

  SRMR                                           0.030
```

- tests H_0 assuming the covariance matrix to be equal in the model and the population
- **Chi² ≤ .05** → low probability to find these results if H_0 applies → **reject H_0 (that model fits population)**

- tests for an exact fit

- df = difference between available information (*S²*, covariances, ggf. *M*) and amount of estimated model parameters

In it (LAVAAN)

```
lavaan 0.6-18 ended normally after 71 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        47

  Number of observations                           604
  Number of missing patterns                         1

Model Test User Model:

  Test statistic                               137.469
  Degrees of freedom                                68
  P-value (Chi-square)                           0.000

Model Test Baseline Model:

  Test statistic                              4479.201
  Degrees of freedom                                95
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.984
  Tucker-Lewis Index (TLI)                       0.978

  Robust Comparative Fit Index (CFI)                NA
  Robust Tucker-Lewis Index (TLI)                   NA

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -7361.741

  P-value H_0: Robust RMSEA <= 0.050             0.929
  P-value H_0: Robust RMSEA >= 0.080             0.000

Standardized Root Mean Square Residual:

  SRMR                                           0.030
```

- tests H_0 assuming the covariance matrix to be equal in the model and the population
- **Chi² ≤ .05** $\rightarrow$ low probability to find these results if H_0 applies $\rightarrow$ **reject H_0 (that model fits population)**

- tests for an exact fit

- df = difference between available information (*S*², covariances, ggf. *M*) and amount of estimated model parameters

CAUTION:

- Chi²-value of complex models (high number of model parameters) tends to be smaller due to the reduction in df
- larger samples lead to larger Chi²-values, which end up in discarding (probably useful) models

In it (LAVAAN)

```
lavaan 0.6-18 ended normally after 71 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        47

  Number of observations                           604
  Number of missing patterns                         1

Model Test User Model:

  Test statistic                               137.469
  Degrees of freedom                                68
  P-value (Chi-square)                           0.000

Model Test Baseline Model:

  Test statistic                              4479.201
  Degrees of freedom                                95
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.984
  Tucker-Lewis Index (TLI)                       0.978

  Robust Comparative Fit Index (CFI)                NA
  Robust Tucker-Lewis Index (TLI)                   NA

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -7361.741
  Loglikelihood unrestricted model (H1)      -7293.006

  Akaike (AIC)                               14817.482
  Bayesian (BIC)                             15024.450
  Sample-size adjusted Bayesian (SABIC)      14875.236

Root Mean Square Error of Approximation:

  RMSEA                                          0.041
  90 Percent confidence interval - lower         0.031
  90 Percent confidence interval - upper         0.051
  P-value H_0: RMSEA <= 0.050                    0.929
  P-value H_0: RMSEA >= 0.080                    0.000

  Robust RMSEA                                   0.041
  90 Percent confidence interval - lower         0.031
  90 Percent confidence interval - upper         0.051
  P-value H_0: Robust RMSEA <= 0.050             0.929
  P-value H_0: Robust RMSEA >= 0.080             0.000

Standardized Root Mean Square Residual:

  SRMR                                           0.030
```

Chi² difference test

- necessary for model comparisons

In it (LAVAAN)

```
lavaan 0.6-18 ended normally after 71 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        47

  Number of observations                           604
  Number of missing patterns                         1

Model Test User Model:

  Test statistic                               137.469
  Degrees of freedom                                68
  P-value (Chi-square)                           0.000

Model Test Baseline Model:

  Test statistic                              4479.201
  Degrees of freedom                                95
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.984
  Tucker-Lewis Index (TLI)                       0.978

  Robust Comparative Fit Index (CFI)                NA
  Robust Tucker-Lewis Index (TLI)                   NA

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -7361.741
  Loglikelihood unrestricted model (H1)      -7293.006

  Akaike (AIC)                               14817.482
  Bayesian (BIC)                             15024.450
  Sample-size adjusted Bayesian (SABIC)      14875.236

Root Mean Square Error of Approximation:

  RMSEA                                          0.041
  90 Percent confidence interval - lower         0.031
  90 Percent confidence interval - upper         0.051
  P-value H_0: RMSEA <= 0.050                    0.929
  P-value H_0: RMSEA >= 0.080                    0.000

  Robust RMSEA                                   0.041
  90 Percent confidence interval - lower         0.031
  90 Percent confidence interval - upper         0.051
  P-value H_0: Robust RMSEA <= 0.050             0.929
  P-value H_0: Robust RMSEA >= 0.080             0.000

Standardized Root Mean Square Residual:

  SRMR                                           0.030
```

Comparative Fit Index (CFI)

- compares current model fit with base model fit (base model → constructs with variance but without covariance)
- the higher, the better the current model fits the data

- **aim: CFI > .95**
  or stricter as well as more common: CFI > .97

In it (LAVAAN)

```
lavaan 0.6-18 ended normally after 71 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        47

  Number of observations                           604
  Number of missing patterns                         1

Model Test User Model:

  Test statistic                               137.469
  Degrees of freedom                                68
  P-value (Chi-square)                           0.000

Model Test Baseline Model:

  Test statistic                              4479.201
  Degrees of freedom                                95
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.984
  Tucker-Lewis Index (TLI)                       0.978

  Robust Comparative Fit Index (CFI)                NA
  Robust Tucker-Lewis Index (TLI)                   NA

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -7361.741
  Loglikelihood unrestricted model (H1)      -7293.006

  Akaike (AIC)                               14817.482
  Bayesian (BIC)                             15024.450
  Sample-size adjusted Bayesian (SABIC)      14875.236

Root Mean Square Error of Approximation:

  RMSEA                                          0.041
  90 Percent confidence interval - lower         0.031
  90 Percent confidence interval - upper         0.051
  P-value H_0: RMSEA <= 0.050                    0.929
  P-value H_0: RMSEA >= 0.080                    0.000

  Robust RMSEA                                   0.041
  90 Percent confidence interval - lower         0.031
  90 Percent confidence interval - upper         0.051
  P-value H_0: Robust RMSEA <= 0.050             0.929
  P-value H_0: Robust RMSEA >= 0.080             0.000

Standardized Root Mean Square Residual:

  SRMR                                           0.030
```

Approximate data fit index

- tests H_0 assuming RMSEA to be ≤ .05
- **aim: RMSEA < .06**
  or stricter (RMSEA ≤ .05)

In it (LAVAAN)

```
lavaan 0.6-18 ended normally after 71 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        47

  Number of observations                           604
  Number of missing patterns                         1

Model Test User Model:

  Test statistic                               137.469
  Degrees of freedom                                68
  P-value (Chi-square)                           0.000

Model Test Baseline Model:

  Test statistic                              4479.201
  Degrees of freedom                                95
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.984
  Tucker-Lewis Index (TLI)                       0.978

  Robust Comparative Fit Index (CFI)                NA
  Robust Tucker-Lewis Index (TLI)                   NA

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -7361.741
  Loglikelihood unrestricted model (H1)      -7293.006

  Akaike (AIC)                               14817.482
  Bayesian (BIC)                             15024.450
  Sample-size adjusted Bayesian (SABIC)      14875.236

Root Mean Square Error of Approximation:

  RMSEA                                          0.041
  90 Percent confidence interval - lower         0.031
  90 Percent confidence interval - upper         0.051
  P-value H_0: RMSEA <= 0.050                    0.929
  P-value H_0: RMSEA >= 0.080                    0.000

  Robust RMSEA                                   0.041
  90 Percent confidence interval - lower         0.031
  90 Percent confidence interval - upper         0.051
  P-value H_0: Robust RMSEA <= 0.050             0.929
  P-value H_0: Robust RMSEA >= 0.080             0.000

Standardized Root Mean Square Residual:

  SRMR                                           0.030
```

Standardized index for overall residual evaluation (difference sample parameters and model parameters)

- **aim: SRMR < .05**

Intro II: Validating Model Fit (LAVAAN)

Compare model fit

„Hoyle and Panter [HOY 95] recommend using χ2 (or scaledχ2) … (not greater than df *2 or *3) You can carefully follow the recommendations of [HU 99] or [SCH 06] who suggest the following guidelines for judging a model goodness-of-fit (based on the hypothesis where the maximum likelihood method is the estimation method): 1) RMSEA value ≤ 0.06, with confidence interval at 90% values should be between 0 and 0.10; 2) SRMR value ≤ 0.08; and 3) CFI and TLI values ≥ 0.95.“ (Gana & Broc, 2019. *Structural equation modeling with lavaan*, p. 43)

Fit type	Index	Interpretation for guidance
Absolute	RMR/SRMR	≤ 0.08 = good fit
	WRMR	≤ 1.00 = good fit
Parsimonious	PRATIO	Between 0.00 (saturated model) and 1.00 (parsimonious model)
	RMSEA	≤ 0.05 = very good fit ≤ 0.06 and ≤ 0.08 = good fit
	AIC	Comparative index: the lower value of this index, the better the fit
	BIC	Comparative index: the lower value of this index, the better the fit
Incremental	CFI	≥ 0.90 and ≤ 0.94 = good fit ≥ 0.95 = very good fit
	TLI	≥ 0.90 and ≤ 0.94 = good fit ≥ 0.95 = very good fit

Buchempfehlung (p. 43):
https://www.wiley.com/en-us/Structural+Equation+Modeling+with+lavaan-p-9781786303691

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online*, *8*(2), 23-74.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.

INTRO III: BUILDING MODELS (USING LAVAAN)

Intro III: Building Models (LAVAAN)

Choose your estimator based on data distribution

Buchempfehlung (p. 33):
https://www.wiley.com/en-us/Structural+Equation+Modeling+with+lavaan-p-9781786303691

Data Type and Normality Assumption	Recommended Estimator
Continuous data	
1. Approximately normal	ML
2. Violation of normality	ML (in case of moderate violation) MLM, MLR, Bootstrap
Ordinal/categorical data	
1. Approximately normal	ML (if at least 6 response categories) MLM, MLR (if at least 4 response categories) WLSMV (binary response or 3 response categories)
2. Violation of normality	ML (if at least 6 response categories) MLM, MLR (if at least 4 response categories) WLSMV (in case of severe violation)

Intro III: Building Models (LAVAAN)

Major operators of lavaan syntax

Buchempfehlung (p. 65-66):
https://www.wiley.com/en-us/Structural+Equation+Modeling+with+lavaan-p-9781786303691

Command	Operator	Illustration	Significance
Estimate a covariance (cor)	~~	X ~~ Y	X is correlated with Y
Estimate a regression	~	Y ~ X	Y is regressed on X
Define a reflective latent variable	=~	F =~ item 1 + item 2 + item 3	The F factor is measured by indicators item 1, item 2, and item 3 over which it has effects
Define a formative latent variable	<~	F <~ item 1 + item 2 + item 3	The F factor is formed by items 1, 2, and 3
Estimate the intercept	~ 1	item 1 ~ 1 F ~ 1	Intercept of item 1 Intercept of latent variable F (factor)
Label/fix a parameter	*	F =~ 1item 1 + b1item 2 + b2*item 3	Item 1 is set to 1, item 2 is named “b1” and item 3 “b2”. The name must begin with a letter.
Constrain parameters to equality	= =	b1 = = b2	Factor loading of item 1 equals that of item 2 (giving the same name to both items is another way to force them to be equal: bitem 2 + bitem 3).
Create a new parameter	: =	b1b2 := b1*b2	Define a parameter that is not in the model (for example, indirect effect) from the existing parameters. Example: b1b2 = indirect effect of parameters b1 and b2
Insert a comment in the syntax	#	b1b2 := b1*b2 # indirect effect	Explain to the reader the meaning of a command (for example, that here b1b2 := b1*b2 is used to estimate an indirect effect

Intro III: Building Models (LAVAAN)

Hands-on: Let’s build our own SEM using the famous PoliticalDemocracy data set that comes pre-installed with lavaan. Run:

```
library(tidyverse)
library(lavaan)

# Get the data
data <- lavaan::PoliticalDemocracy
data <- data %>%  rename(
freedom_press_1960 = y1,
freedom_opposition_1960 = y2,
fair_elections_1960 = y3,
legislature_effectiveness_1960 = y4,
freedom_press_1965 = y5,
freedom_opposition_1965 = y6,
fair_elections_1965 = y7,
legislature_effectiveness_1965 = y8,
gnp_per_capita_1960 = x1,
energy_consumption_per_capita_1960 = x2,
labor_force_industry_1960 = x3
)
```

Intro III: Building Models (LAVAAN)

The data:

	freedom_press_1960	freedom_opposition_1960	fair_elections_1960	legislature_effectiveness_1960	freedom_press_1965	freedom_opposition_1965	fair_elections_1965	legislature_effectiveness_1965	gnp_per_capita_1960	energy_consumptio
1	2.50	0.000000	3.333333	0.000000	1.250000	0.000000	3.726360	3.333333	4.442651	
2	1.25	0.000000	3.333333	0.000000	6.250000	1.100000	6.666666	0.736999	5.384495	
3	7.50	8.800000	9.999998	9.199991	8.750000	8.094061	9.999998	8.211809	5.961005	
4	8.90	8.800000	9.999998	9.199991	8.907948	8.127979	9.999998	4.615086	6.285998	
5	10.00	3.333333	9.999998	6.666666	7.500000	3.333333	9.999998	6.666666	5.863631	
6	7.50	3.333333	6.666666	6.666666	6.250000	1.100000	6.666666	0.368500	5.533389	
7	7.50	3.333333	6.666666	6.666666	5.000000	2.233333	8.271257	1.485166	5.308268	
8	7.50	2.233333	9.999998	1.496333	6.250000	3.333333	9.999998	6.666666	5.347108	
9	2.50	3.333333	3.333333	3.333333	6.250000	3.333333	3.333333	3.333333	5.521461	
10	10.00	6.666666	9.999998	8.899991	8.750000	6.666666	9.999998	10.000000	5.828946	
11	7.50	3.333333	9.999998	6.666666	8.750000	3.333333	9.999998	6.666666	5.916202	
12	7.50	3.333333	6.666666	6.666666	8.750000	3.333333	6.666666	6.666666	5.398163	
13	7.50	3.333333	9.999998	6.666666	7.500000	3.333333	6.666666	10.000000	6.622736	
14	7.50	7.766664	9.999998	6.666666	7.500000	0.000000	9.999998	0.000000	5.204007	
15	7.50	9.999998	3.333333	10.000000	7.500000	6.666666	9.999998	10.000000	5.509388	
16	7.50	9.999998	9.999998	7.766666	7.500000	1.100000	6.666666	6.666666	5.262690	
17	2.50	3.333333	6.666666	6.666666	5.000000	1.100000	6.666666	0.368500	4.700480	

Intro III: Building Models (LAVAAN)

```
# Define the SEM model
model <- `
# measurement model
industrialization60 =~ gnp_per_capita_1960 + energy_consumption_per_capita_1960 + labor_force_industry_1960
freedom60 =~ freedom_press_1960 + freedom_opposition_1960 + fair_elections_1960 + legislature_effectiveness_1960
freedom65 =~ freedom_press_1965 + freedom_opposition_1965 + fair_elections_1965 + legislature_effectiveness_1965

# regressions
freedom60 ~ industrialization60
freedom65 ~ industrialization60 + freedom60

# residual correlations
freedom_press_1960 ~~ freedom_press_1965
freedom_opposition_1960 ~~ freedom_opposition_1965
fair_elections_1960 ~~ fair_elections_1965
legislature_effectiveness_1960 ~~ legislature_effectiveness_1965
freedom_opposition_1960 ~~ legislature_effectiveness_1960
freedom_opposition_1965 ~~ legislature_effectiveness_1965
'

# Fit the model
fit <- sem(model, data = data)

# Summarize the results
summary(fit, fit.measures = TRUE, standardized = TRUE)
```

Intro III: Building Models (LAVAAN)

Measurement Model

Latent Variables:

1. **industrialization60**: This latent variable represents the level of industrialization in 1960, measured by:
    1. **gnp_per_capita_1960**: Gross national product per capita in 1960.
    2. **energy_consumption_per_capita_1960**: Inanimate energy consumption per capita in 1960.
    3. **labor_force_industry_1960**: Percentage of the labor force in industry in 1960.
2. **freedom60**: This latent variable represents the level of political freedom in 1960, measured by:
    1. **freedom_press_1960**: Expert ratings of the freedom of the press in 1960.
    2. **freedom_opposition_1960**: The freedom of political opposition in 1960.
    3. **fair_elections_1960**: The fairness of elections in 1960.
    4. **legislature_effectiveness_1960**: The effectiveness of the elected legislature in 1960.
3. **freedom65**: This latent variable represents the level of political freedom in 1965, measured by:
    1. **freedom_press_1965**: Expert ratings of the freedom of the press in 1965.
    2. **freedom_opposition_1965**: The freedom of political opposition in 1965.
    3. **fair_elections_1965**: The fairness of elections in 1965.
    4. **legislature_effectiveness_1965**: The effectiveness of the elected legislature in 1965.

Intro III: Building Models (LAVAAN)

Structural Model

1. Regression Paths:
    1. **freedom60 ~ industrialization60**: Hypothesizes that the level of industrialization in 1960 (industrialization60) influences the level of political freedom in 1960 (freedom60). *Rationale:* higher levels of industrialization contribute to greater economic development, which in turn supports political freedoms and democratic institutions.
    2. **freedom65 ~ industrialization60 + freedom60**: Hypothesizes that the level of political freedom in 1965 (freedom65) is influenced by both the level of industrialization in 1960 (industrialization60) and the level of political freedom in 1960 (freedom60). This implies that the impact of industrialization on political freedom persists over time and that the political freedoms established in 1960 continue to influence the political situation in 1965.

Intro III: Building Models (LAVAAN)

Structural Model

2. Residual Correlations:
   1. **…_1960 ~ ~ …_1965 :** Variables in 1960 and their counterparts in 1965 are correlated, suggesting that consistent unobserved variables influence these variables over time.
   2. **freedom_opposition_1960/65 ~~ legislature_effectiveness_1960/65:** Indicates that unobserved variables in 1960/1965 simultaneously affect both the freedom of political opposition and the effectiveness of the legislature (e.g., specific political events in that year). By correlating these residuals, the model accounts for the possibility that there are shared underlying influences affecting both the freedom of political opposition and the effectiveness of the legislature that are not captured by the latent variables in the model.

Intro III: Building Models (LAVAAN)

Model Fit Evaluation

Fit indices indicate a very good fit between the model and the data: $\chi^2(35) = 38.125$, $p = .329$; CFI = .995; TLI = .993; RMSEA = .035, 90% CI [0.000, 0.092]; SRMR = .044.

```
> summary(fit, fit.measures = TRUE, standardized = TRUE)
lavaan 0.6.17 ended normally after 68 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        31

  Number of observations                            75

Model Test User Model:

  Test statistic                                38.125
  Degrees of freedom                                35
  P-value (Chi-square)                           0.329

Model Test Baseline Model:

  Test statistic                               730.654
  Degrees of freedom                                55
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.995
  Tucker-Lewis Index (TLI)                       0.993

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -1547.791
  Loglikelihood unrestricted model (H1)      -1528.728

  Akaike (AIC)                                3157.582
  Bayesian (BIC)                              3229.424
  Sample-size adjusted Bayesian (SABIC)       3131.720

Root Mean Square Error of Approximation:

  RMSEA                                          0.035
  90 Percent confidence interval - lower         0.000
  90 Percent confidence interval - upper         0.092
  P-value H_0: RMSEA <= 0.050                    0.611
  P-value H_0: RMSEA >= 0.080                    0.114

Standardized Root Mean Square Residual:

  SRMR                                           0.044
```

Intro III: Building Models (LAVAAN)

Substantial interpretations

1. Higher levels of industrialization in 1960 are associated with higher levels of political freedom in 1960 ($\beta = 0.447$). This suggests that economic development, as measured by industrialization, has a positive impact on political freedom.
2. Industrialization in 1960 has a positive but smaller impact on political freedom in 1965 ($\beta = 0.182$) compared to its impact on freedom in 1960. This indicates that the influence of industrialization persists over time but diminishes.
3. Political freedom in 1960 strongly predicts political freedom in 1965 ($\beta = 0.885$). This suggests a high degree of stability in political freedom over time, where past levels of freedom significantly influence future levels.

To Take Home …

To Take Home …

Summary

Theory

regression + factor analysis

path v. full structural equation models

measured v. structured v. hypothesized model

Model fit

test if model fits population (χ^2)

test if baseline model fits population (χ^2)

test if model fits data better than baseline model (CFI/TLI)

test if model fits data at all (RMSEA)

test if model fits data if standardized (SRMR)

To Take Home …

Outlook

- compare multiple models (male v. female, various arousal states, ...)
- compare multiple levels (within- v. between-level variability)
- compare multiple points in time (latent growth models)
- include linear time-based data

Any Questions?

Thank you very much!

Lara Kobilke & Magdalena Obermaier
Department of Media and Communication
LMU (Lavaan Masters University) Munich

Computational Exchange Sessions | 17 July 2024