Introduction to structural equation modeling and mixed models in R

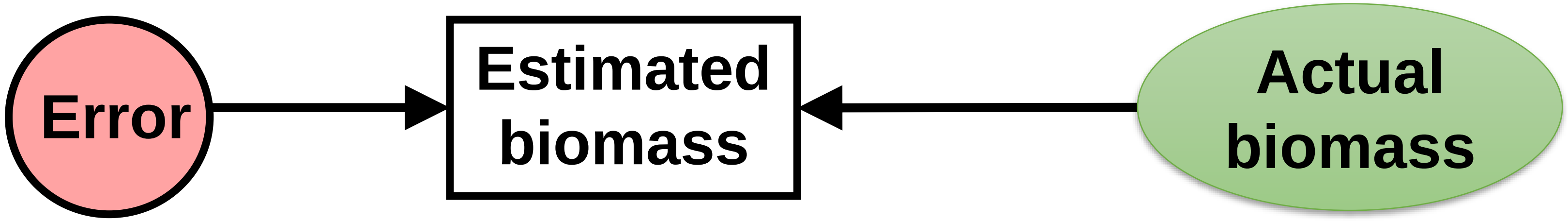
Day 5 – Part 2: SEM

Oksana Buzhdygan

oksana.buzh@fu-berlin.de

Outline

- Latent Variables in SEM



Outline

Latent Variables in SEM

- What are Latent Variables? Why to use them?
- Multi-indicator Latent Variables
- Fitting Latent Variables

  (Confirmatory Factor Analysis)

Outline

Latent Variables in SEM

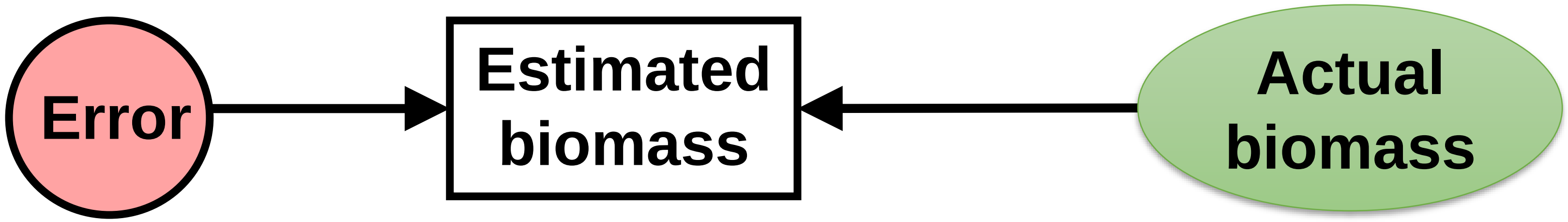
- **What are Latent Variables? Why to use them?**
- Multi-indicator Latent Variables
- Fitting Latent Variables

  (Confirmatory Factor Analysis)

What is Latent Variable?

Latent – hypothetical, hidden

- a variable that is **unmeasured**, but is **hypothesized to exist**
- scientific concept that is **not directly observed**, but is hypothetical **construct**
- can be **approximated using observable indicators**

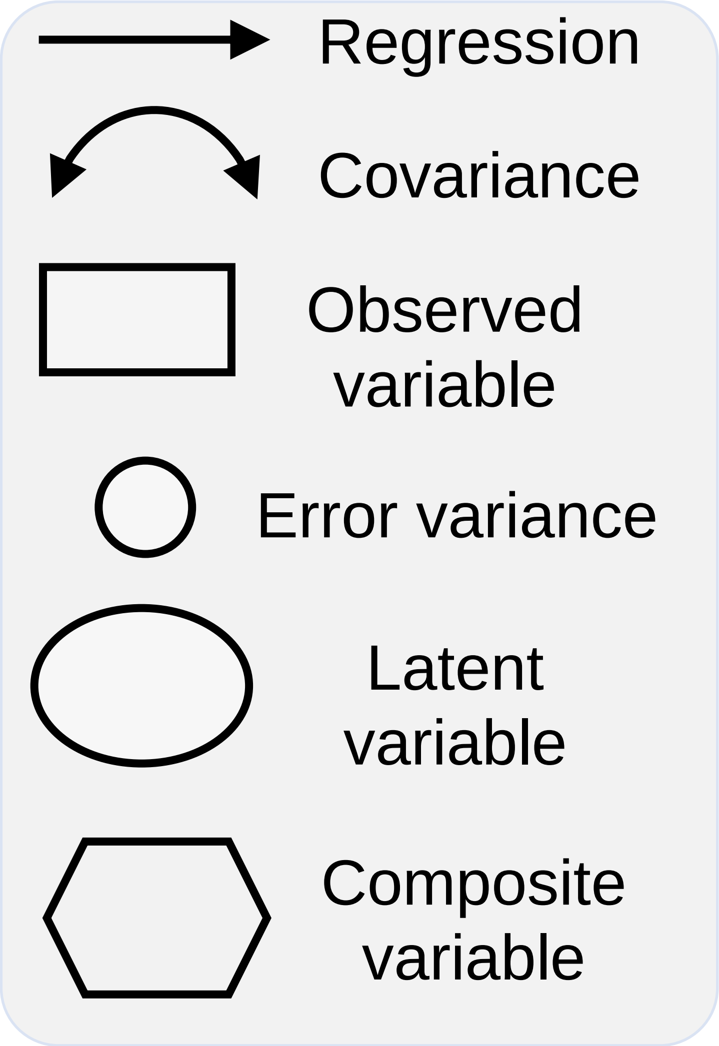


What is Latent Variable?

Specification operators in ‘lavaan’

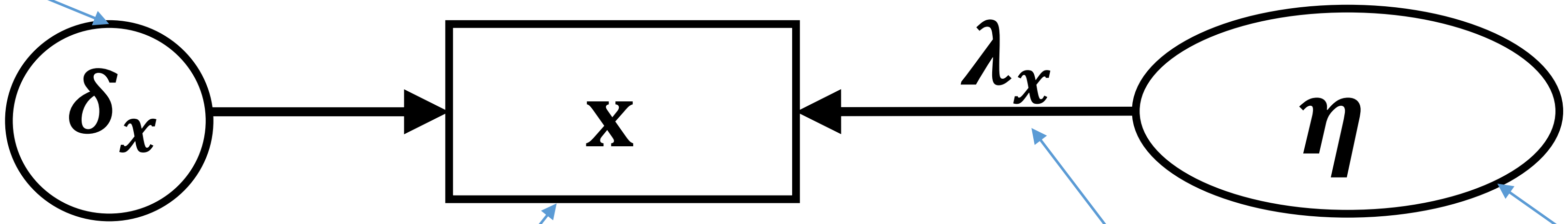
formula type	operator	meaning
Regression	~	“regressed on”
Correlation	~~	"correlated with"
Intercept	~ 1	"estimates intercept"
Latent variable	=~	"is measured by"
Composite	<~	"is caused by"

Path Diagram Notations:



What is Latent Variable?

The error in the measurement of **x** b η



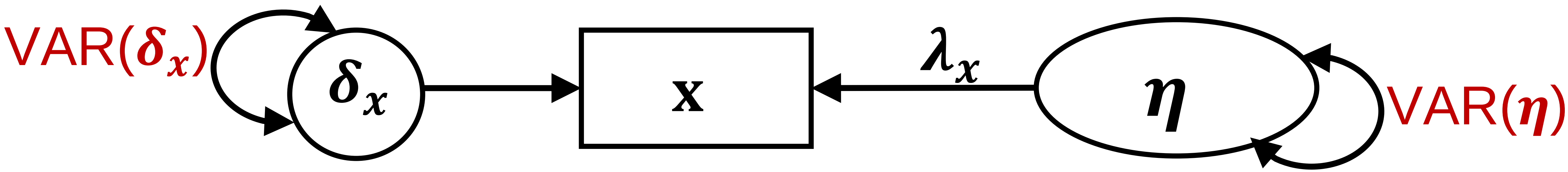
Observed variable
"**manifest indicator**"

" **factors**" or "**latent traits**"

The relationship between a latent variable and its observed indicator

Latent variable

What is Latent Variable?



$$\mathbf{x} = \boldsymbol{\lambda_x}\boldsymbol{\eta} + \boldsymbol{\delta_x}$$

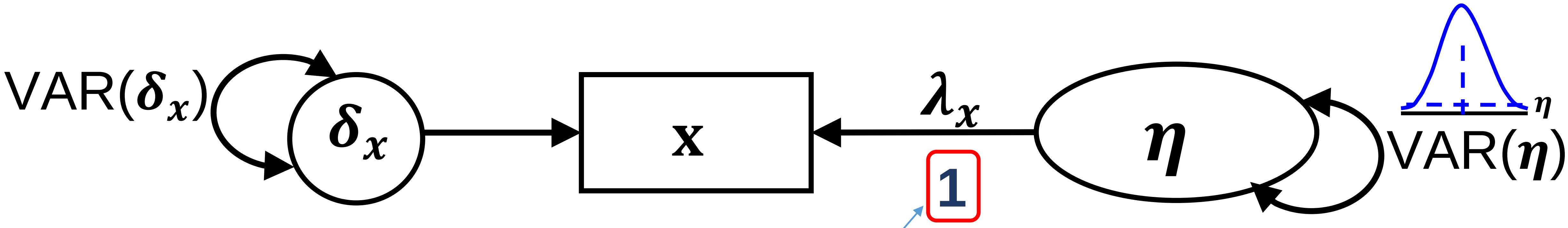
$$\boldsymbol{\eta} \sim \mathrm{N}(0, \mathrm{SD}(\boldsymbol{\eta}))$$
$$\boldsymbol{\delta} \sim \mathrm{N}(0, \mathrm{SD}(\boldsymbol{\delta}))$$

$$\mathrm{VAR}(\mathrm{x}) = \boldsymbol{\lambda_x}^2\mathrm{VAR}(\boldsymbol{\eta}) + \mathrm{VAR}(\boldsymbol{\delta})$$

How much variance does the LV explain?

$$\frac{\boldsymbol{\lambda_x}^2\mathrm{VAR}(\boldsymbol{\eta})}{\boldsymbol{\lambda_x}^2\mathrm{VAR}(\boldsymbol{\eta}) + \mathrm{VAR}(\boldsymbol{\delta})}$$

What is Latent Variable?



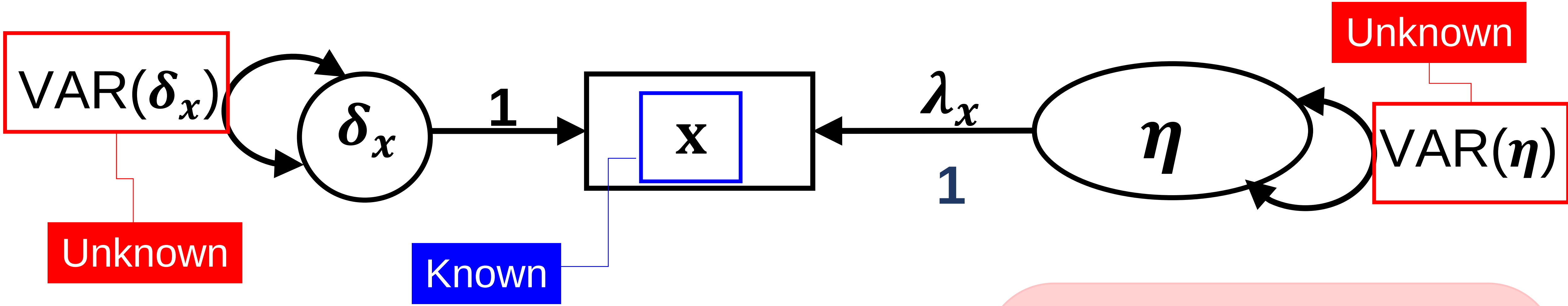
Raw scale coefficient: matches observed (co)variances to parameters VAR(δ) and VAR(η)

We explain the data well if:

VAR(x) = VAR(η) + VAR(δ)

- What is the scale/unit of our LV?
  It needs to be defined to get the regression weights.

What is Latent Variable?



$$t \leq t_{max}$$

$$t_{max} = \frac{s(s+1)}{2} = 1$$

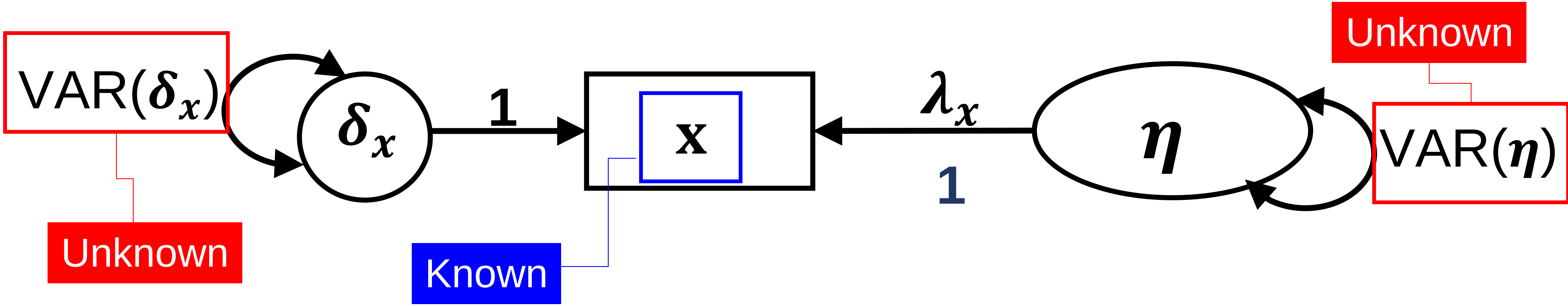
$s = 1$ known

$t = 2$ unknowns

$$\mathbf{DF} = t_{max} - t = -1$$

- Model is not identified

What is Latent Variable?



Rules for LV models:

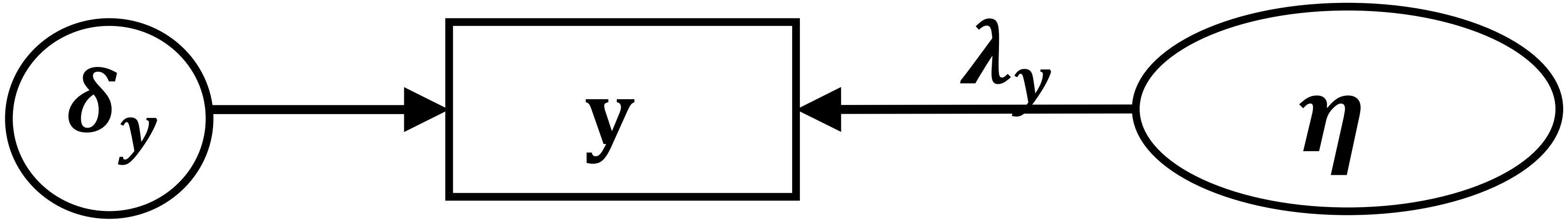
- Scaling of LV
- Non-negative DF

We need at least:

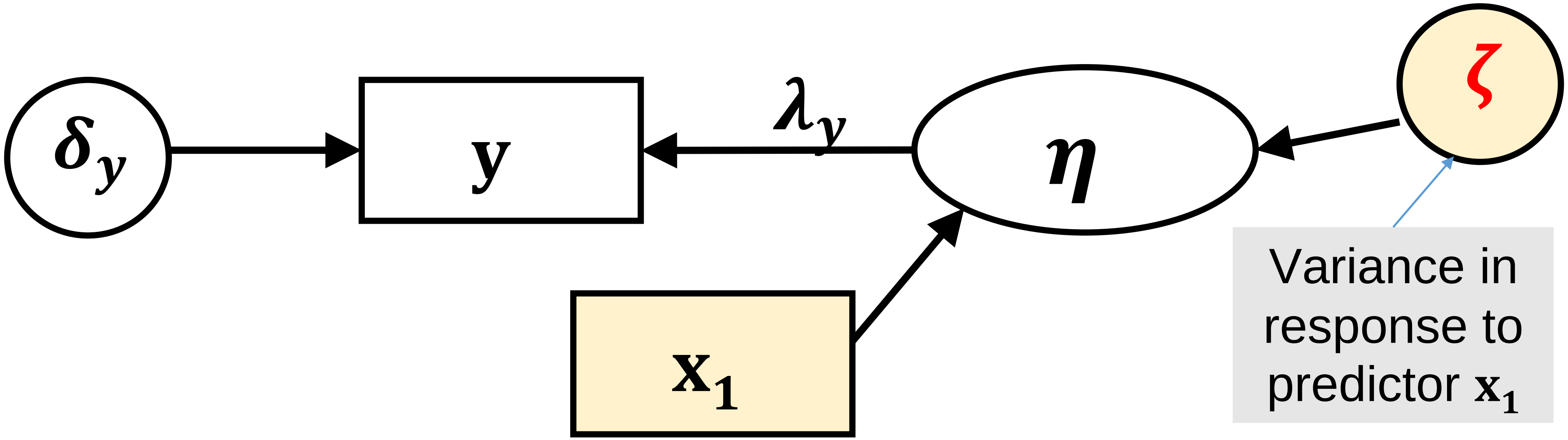
- 3 indicators for a single LV
- 2 indicators per LV for models with multiple (correlated) LVs

What is Latent Variable?

Latent **Exogenous** Variable



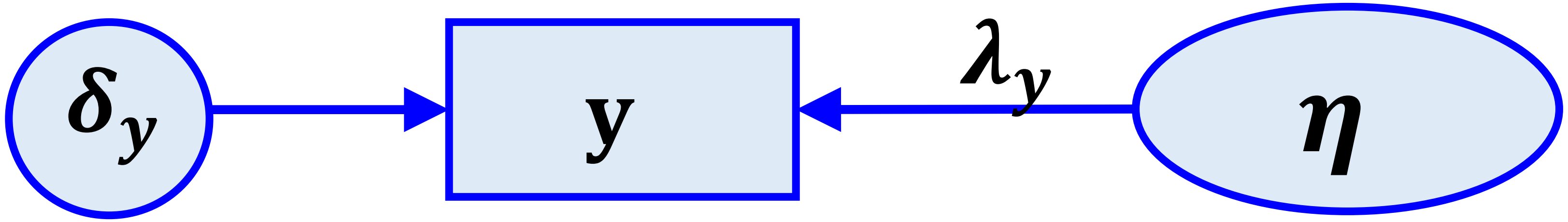
Latent **Endogenous** Variable

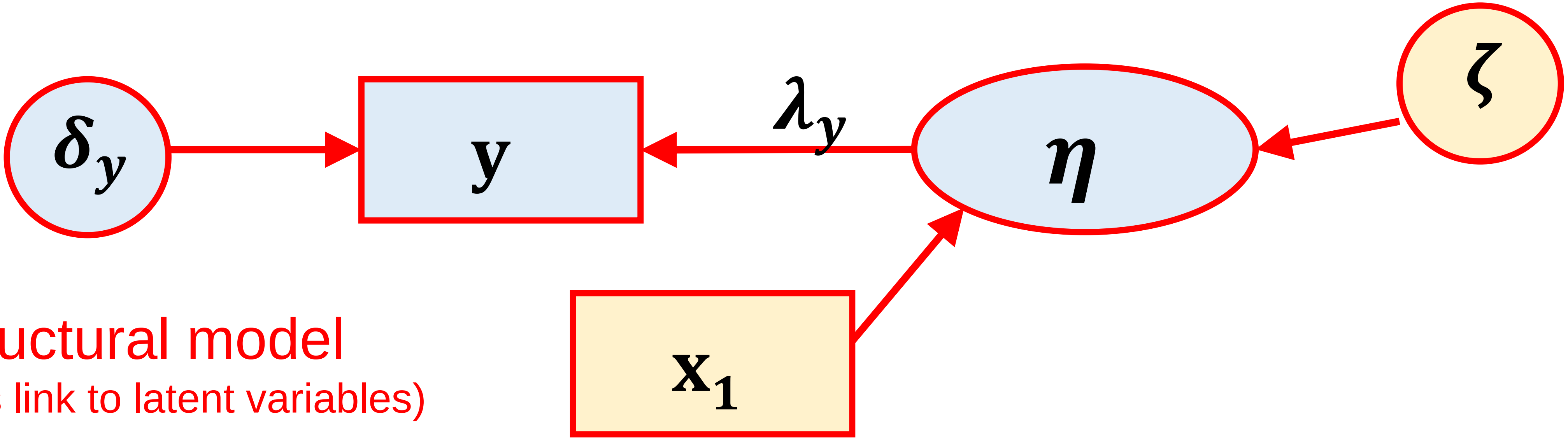


What is Latent Variable?

Measurement model

(solely relates indicators to latent variables)

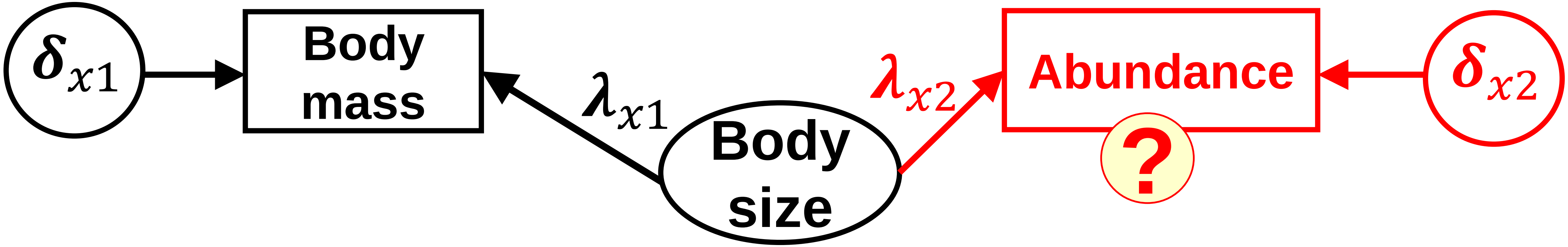




Structural model

(has link to latent variables)

Latent Variables



- Be sure that the latent variable reflects the actual properties captured by the indicator variables!

Why use Latent Variables?

- Allows estimating complex and **multifaceted concepts**
- **Reduces random error** in construct (latent variable)

random error in dependent variables
$\rightarrow$ less precisely measured estimates

random error in independent variables
$\rightarrow$ underestimated regression coefficients

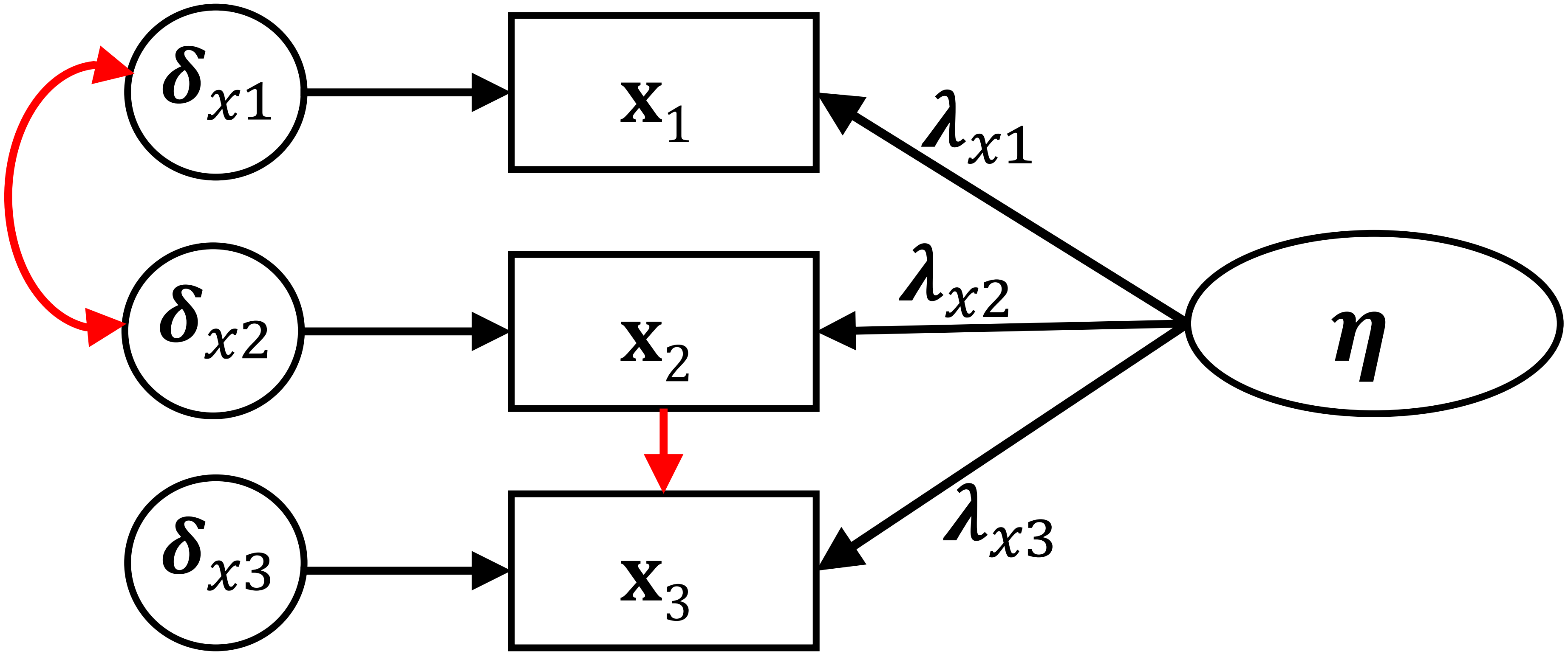
Outline

Latent Variables in SEM

- What are Latent Variables? Why to use them?
- **Multi-indicator Latent Variables**
- Fitting Latent Variables

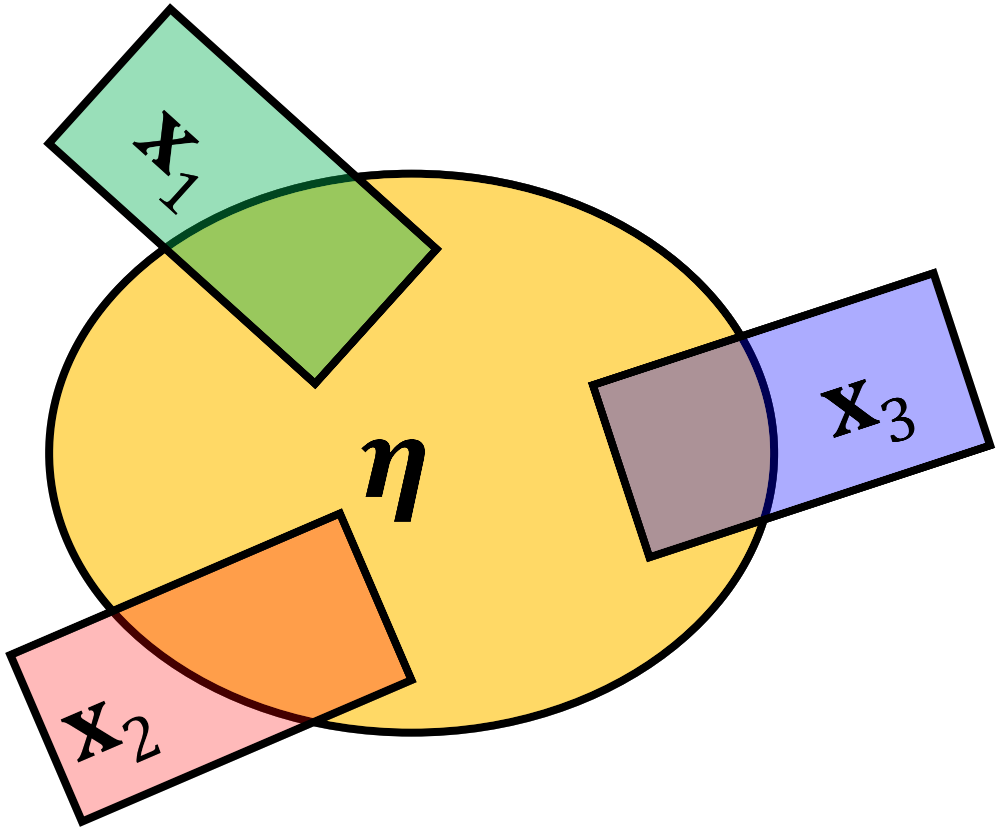
  (Confirmatory Factor Analysis)

Multi-indicator Latent Variables



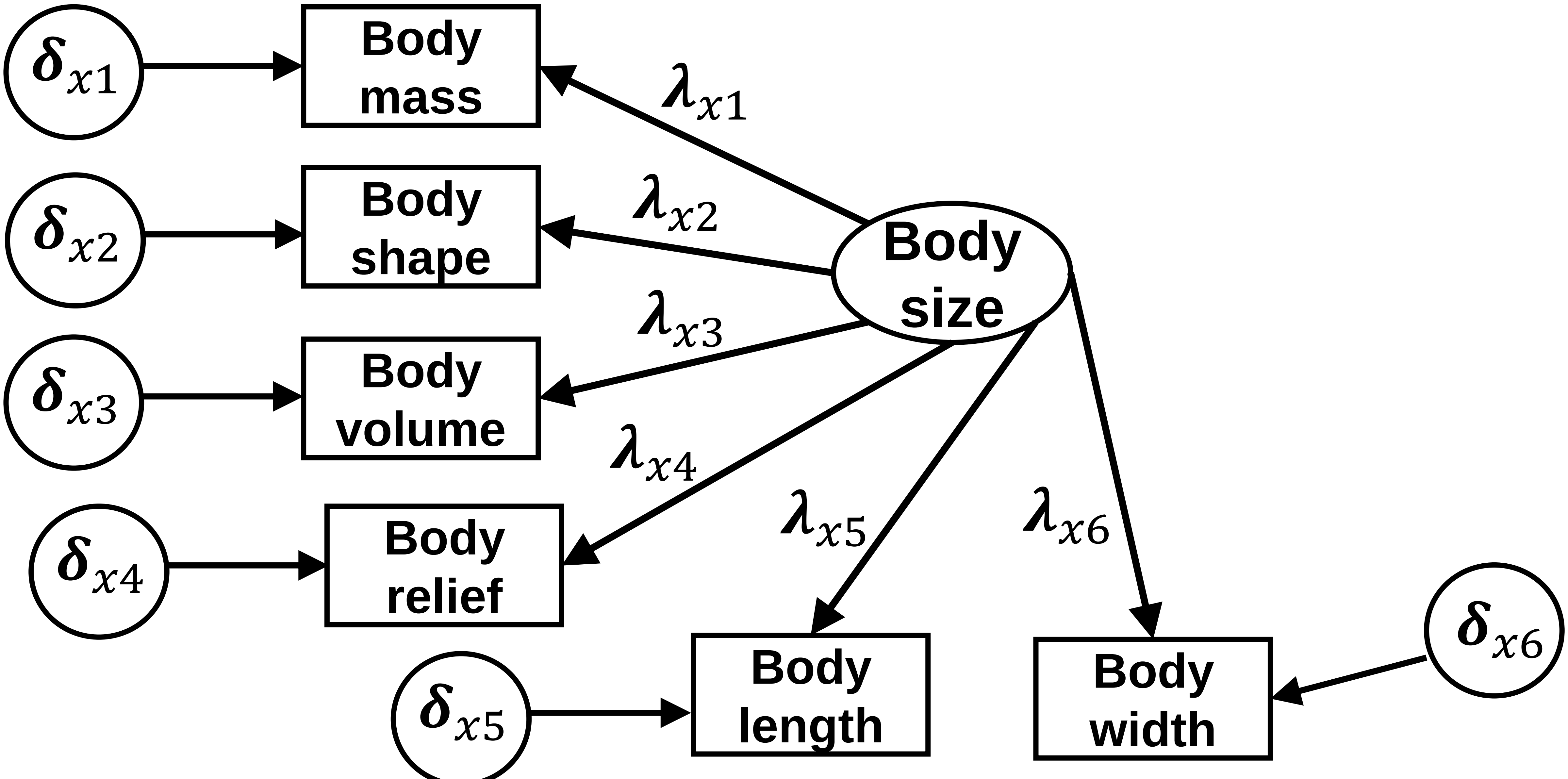
- Indicators may have causal links
- Indicators may covary for other reasons

Multi-indicator Latent Variables



Latent variable $\boldsymbol{\eta}$ represents shared information of observed indicators $\mathbf{x}$

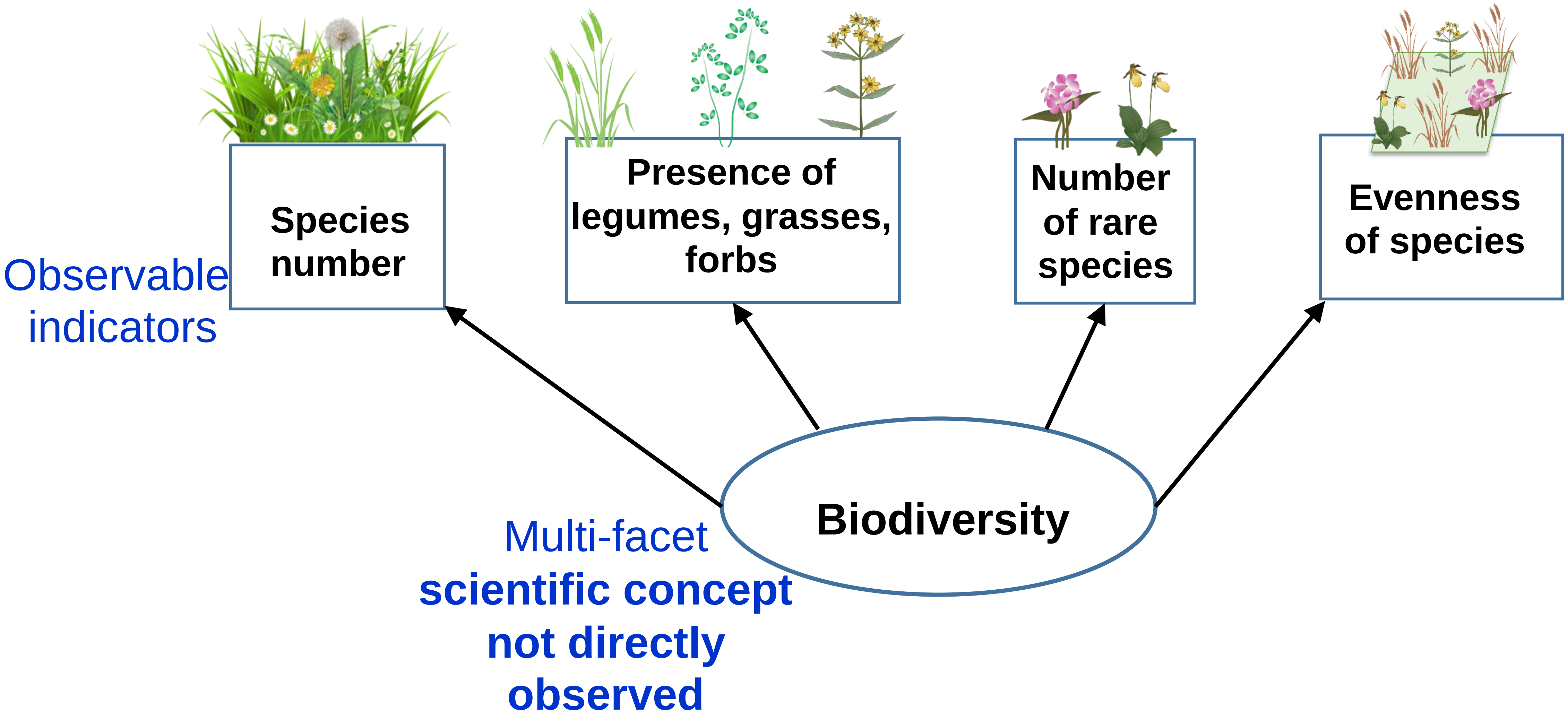
Multi-indicator Latent Variables



What is Latent Variable?

Example

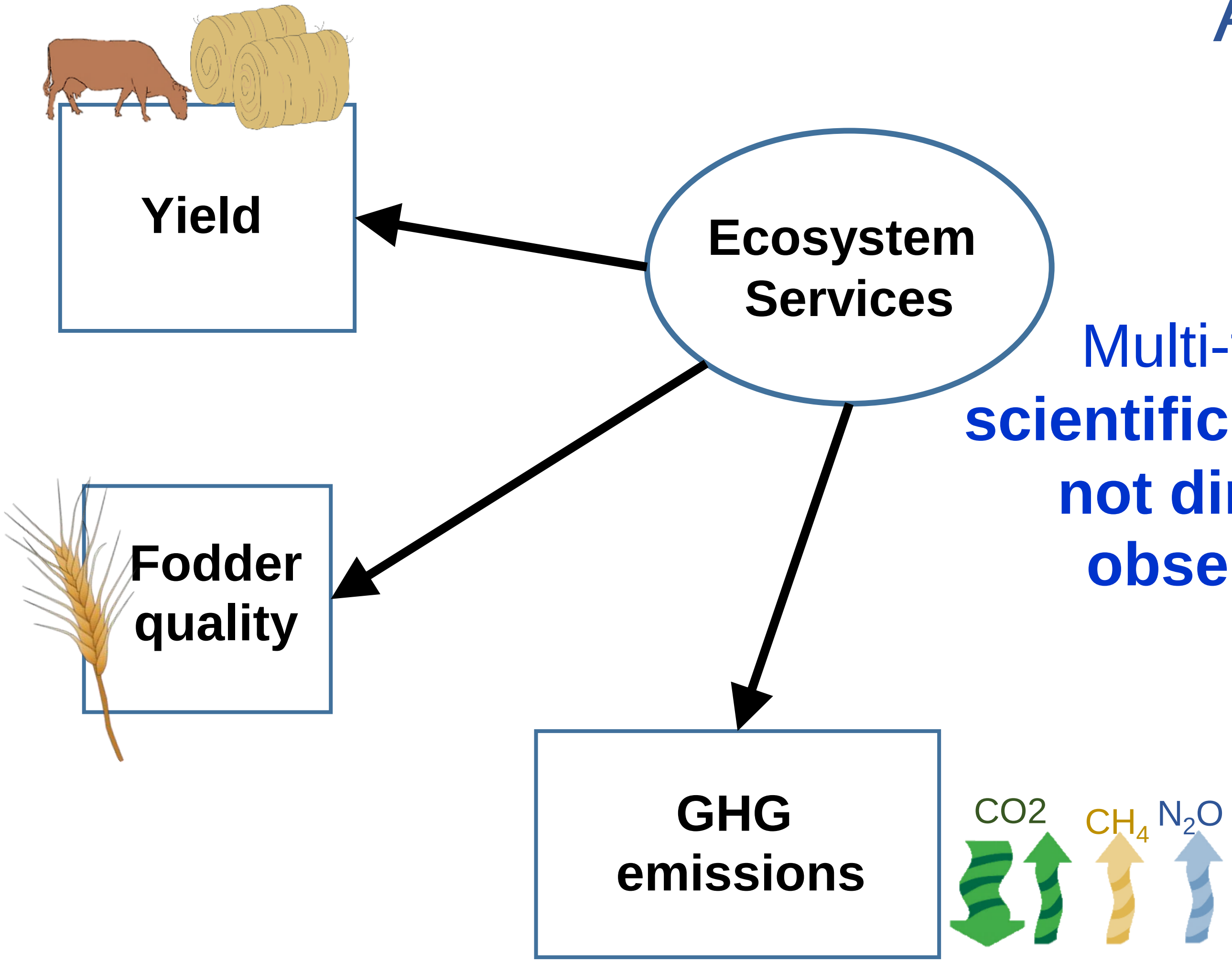
As Theoretical Constructs



What is Latent Variable?

Example

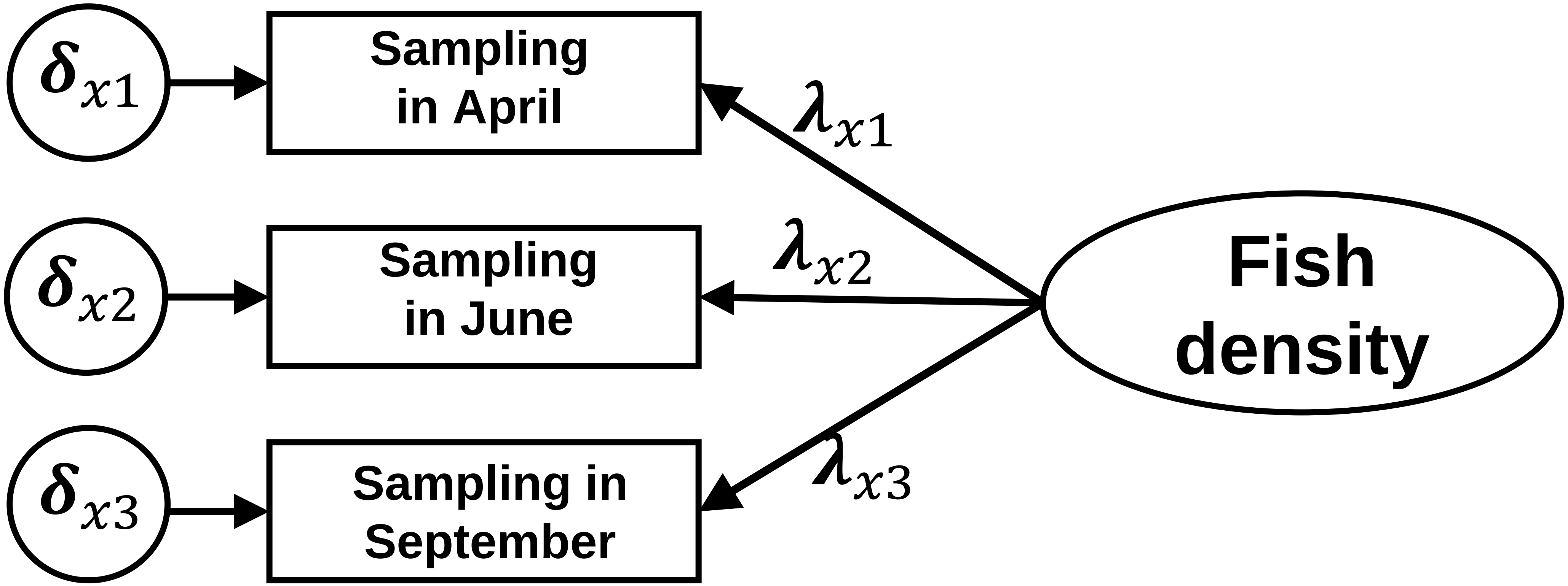
As Theoretical Constructs



Multi-facet
scientific concept not directly observed

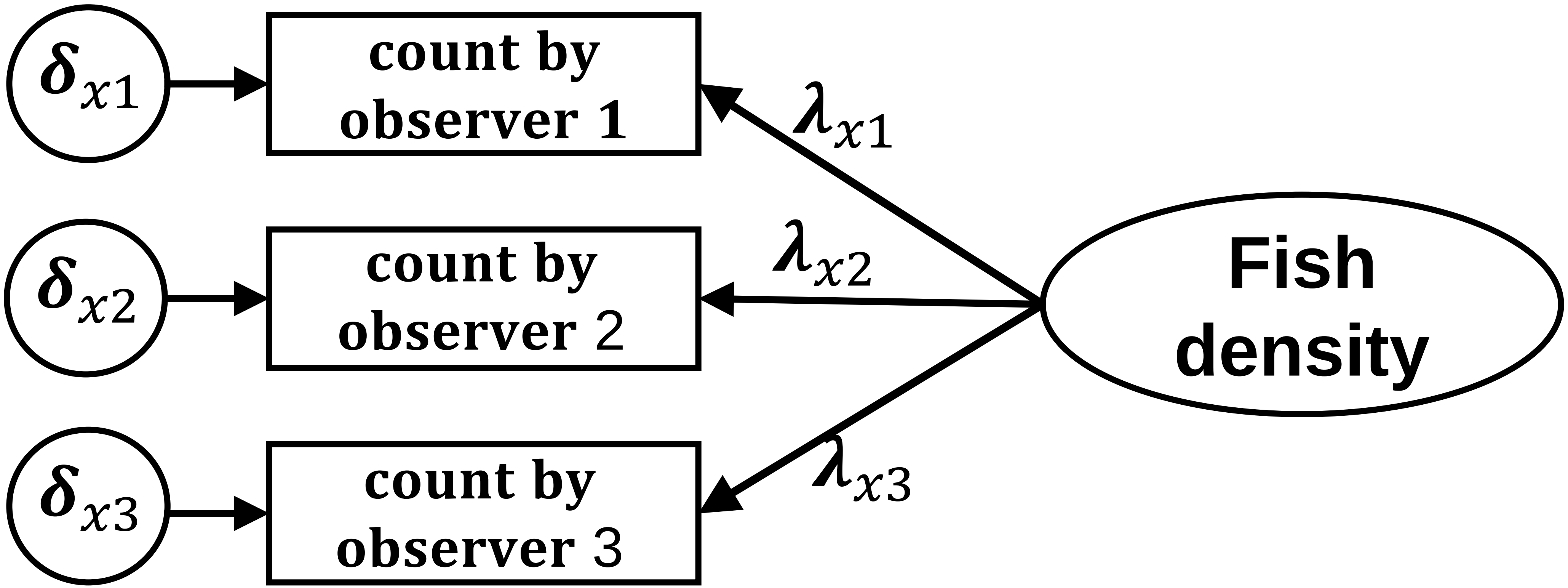
Multi-indicator Latent Variables

Repeated Measurements



Multi-indicator Latent Variables

Multi-sampling



Why use Latent Variables with Multiple Indicators?

- Allows estimating complex and multifaceted concepts
- Reduces random error in construct (latent variable)
- Better accuracy in measurement of relationships due to shared variation between observed indicators.

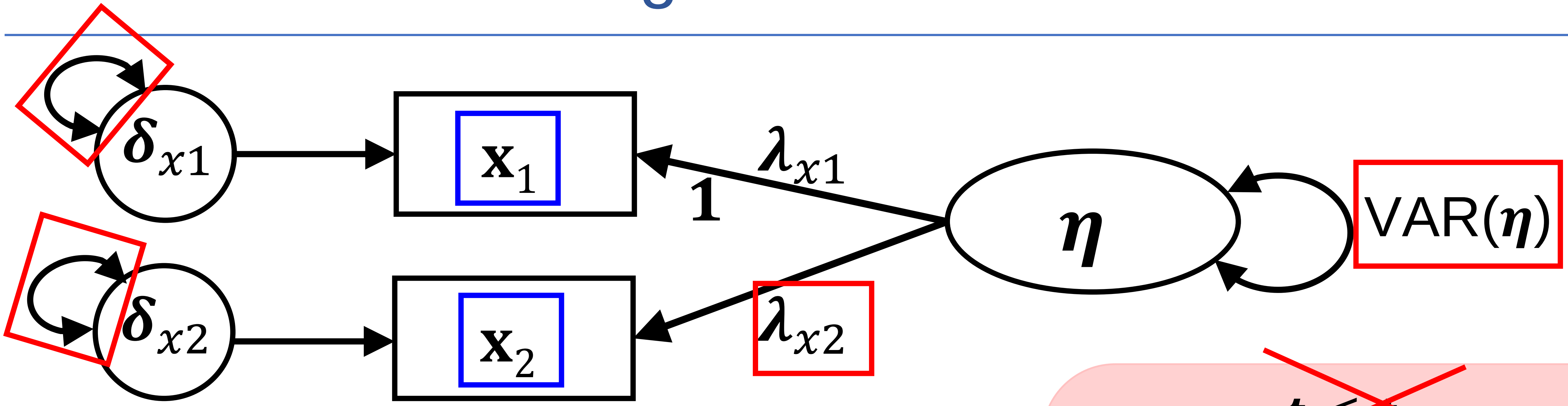
Outline

Latent Variables in SEM

- What are Latent Variables? Why to use them?
- Multi-indicator Latent Variables
- **Fitting Latent Variables**

  (Confirmatory Factor Analysis)

Fitting Latent Variables



Rules for LV models:

- Scaling of LV
- Non-negative DF

$$\mathbf{DF = -1}$$

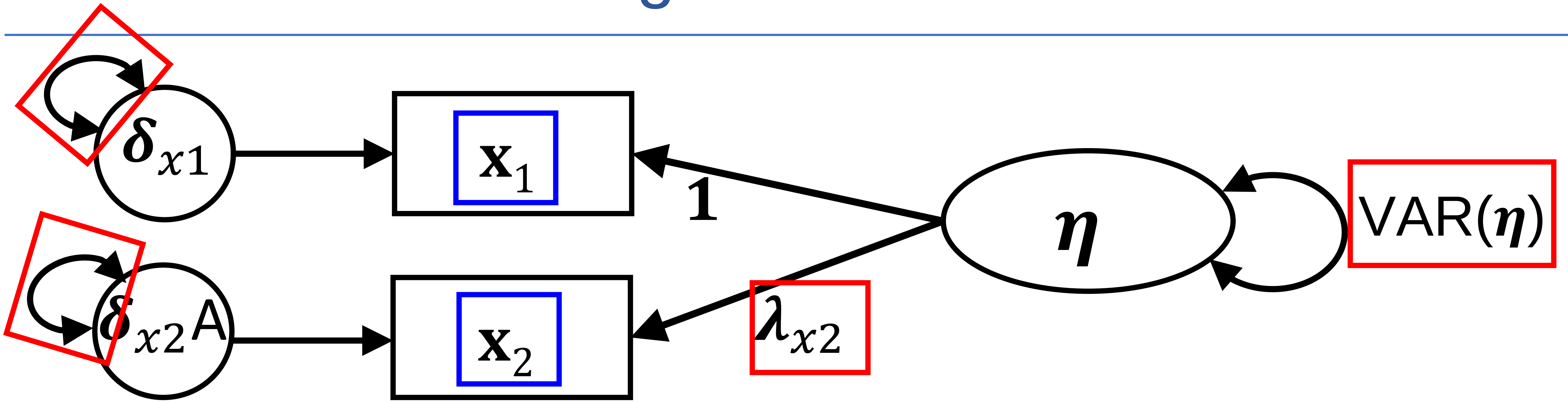
$$\cancel{t \leq t_{max}}$$

$$t_{max} = \frac{s(s+1)}{2} = 3$$

s = 2 knowns

t = 4 unknowns

Fitting Latent Variables



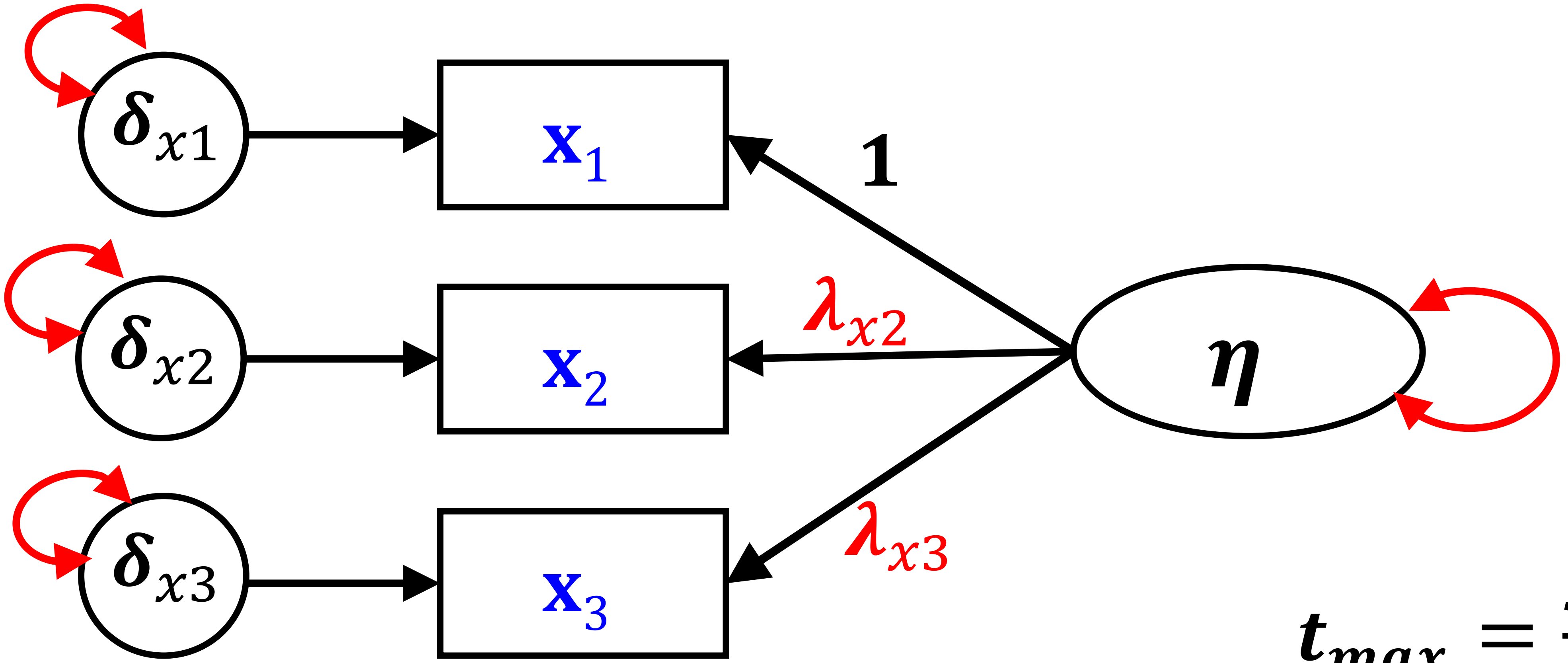
Rules for LV models:

- Scaling of LV
- Non-negative DF

We need at least:

- 3 indicators for a single LV
- 2 indicators per LV for models with multiple (correlated) LVs

Fitting Latent Variables



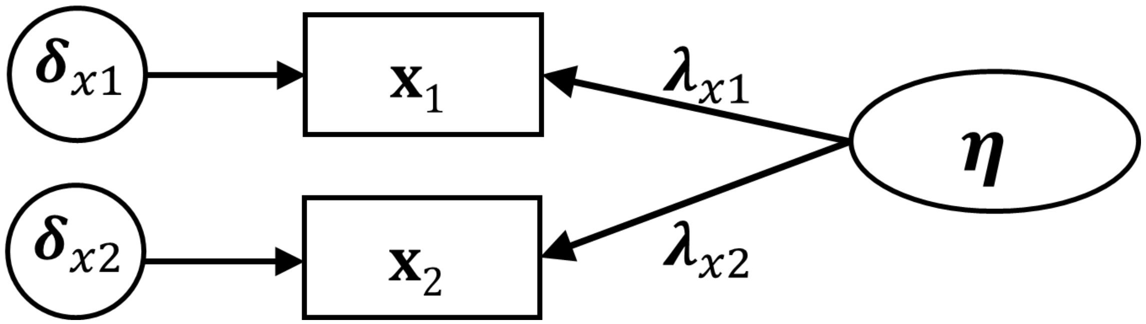
$$t_{max} = \frac{s(s+1)}{2} = 6$$

$$\mathrm{DF} = t_{max} - t = 0$$

$s = 3$ knowns

$t = 6$ unknowns

Fitting Latent Variables



Two indicators of LV

Solution:

1) To set $\lambda_{x1} = \lambda_{x2}$, assuming that $\mathbf{x}_1$ and $\mathbf{x}_2$ have equal weight in the estimation of the LV

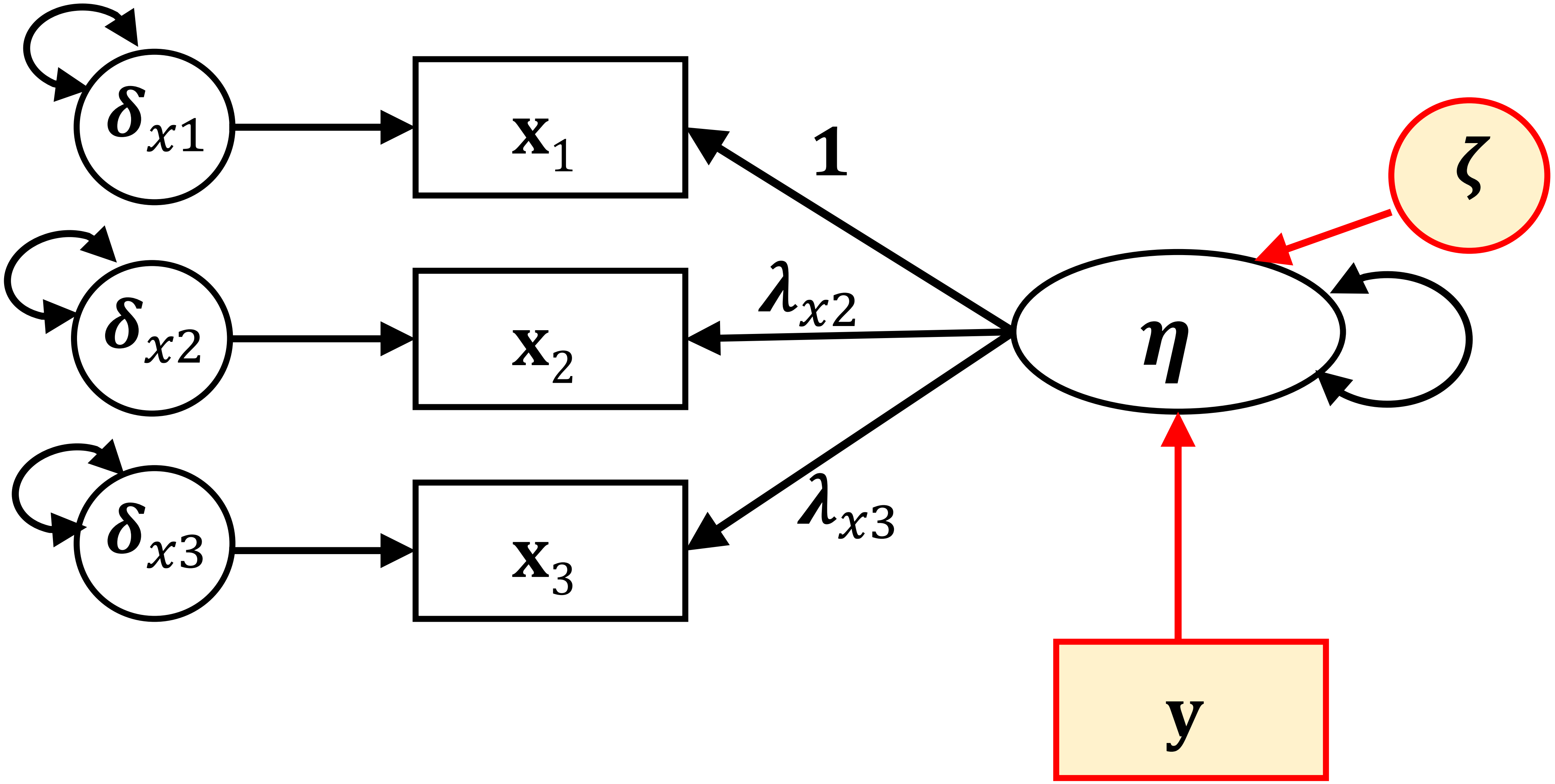
2) To measure errors experimentally.

```
# specify latent variable
Sem.mod <- '
           n =~ a1*x1 + a2*x2
              a1==a2 # λx1= λx2
             x1 ~~ 0.213 * x1 # fix error variance
'
```

Knowing your measurement error:

- Increases accuracy in estimating relationship between variables.
- Increases explanatory power of measured variables.

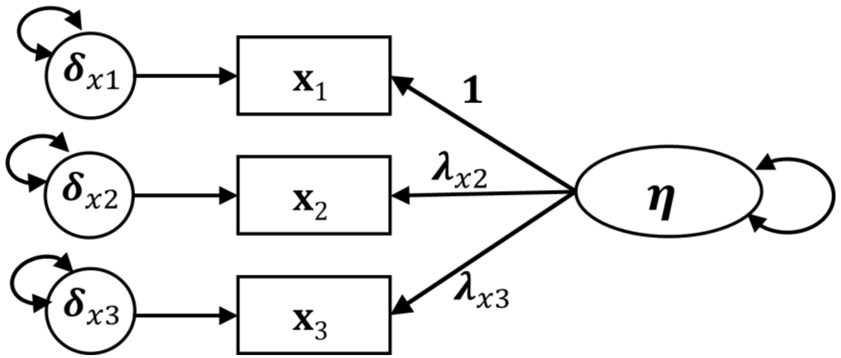
Fitting Latent Variables



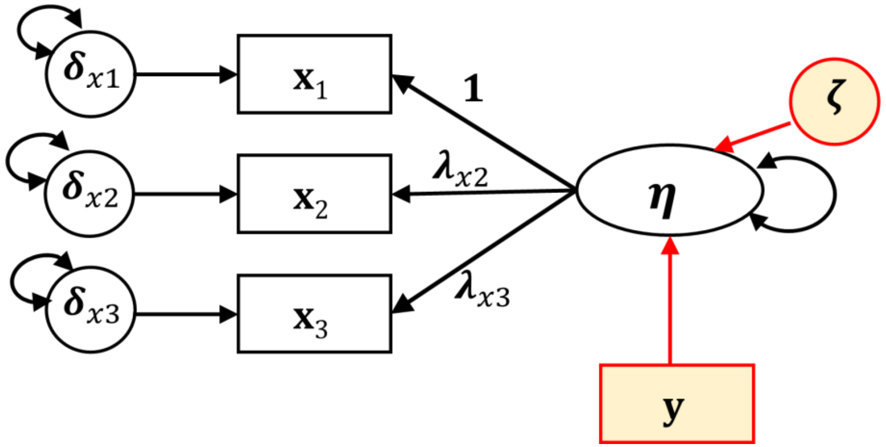
Fitting Latent Variables

1) Evaluate the latent relationships among variables (**Confirmatory Factor Analysis**).

- Do our indicators make a Good Latent Variable?



2) Use Latent Variables as a Response or a Predictor



Population-based ecological restoration

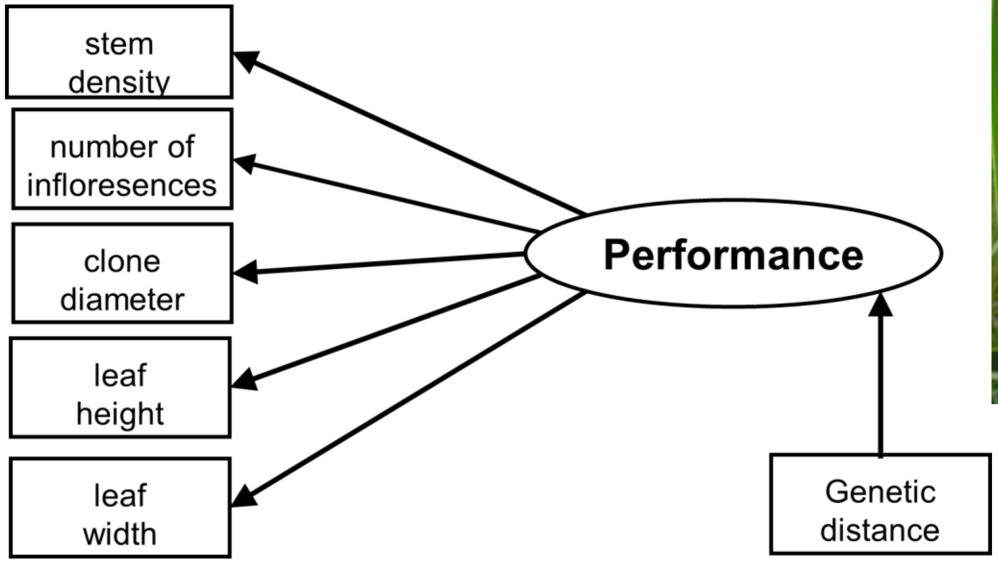
Aim: understand the performance of transplanted plants as a function of their dissimilarity to local conditions


```
# Read and check the data

travis <- read.csv(" Travis_data.csv")
```

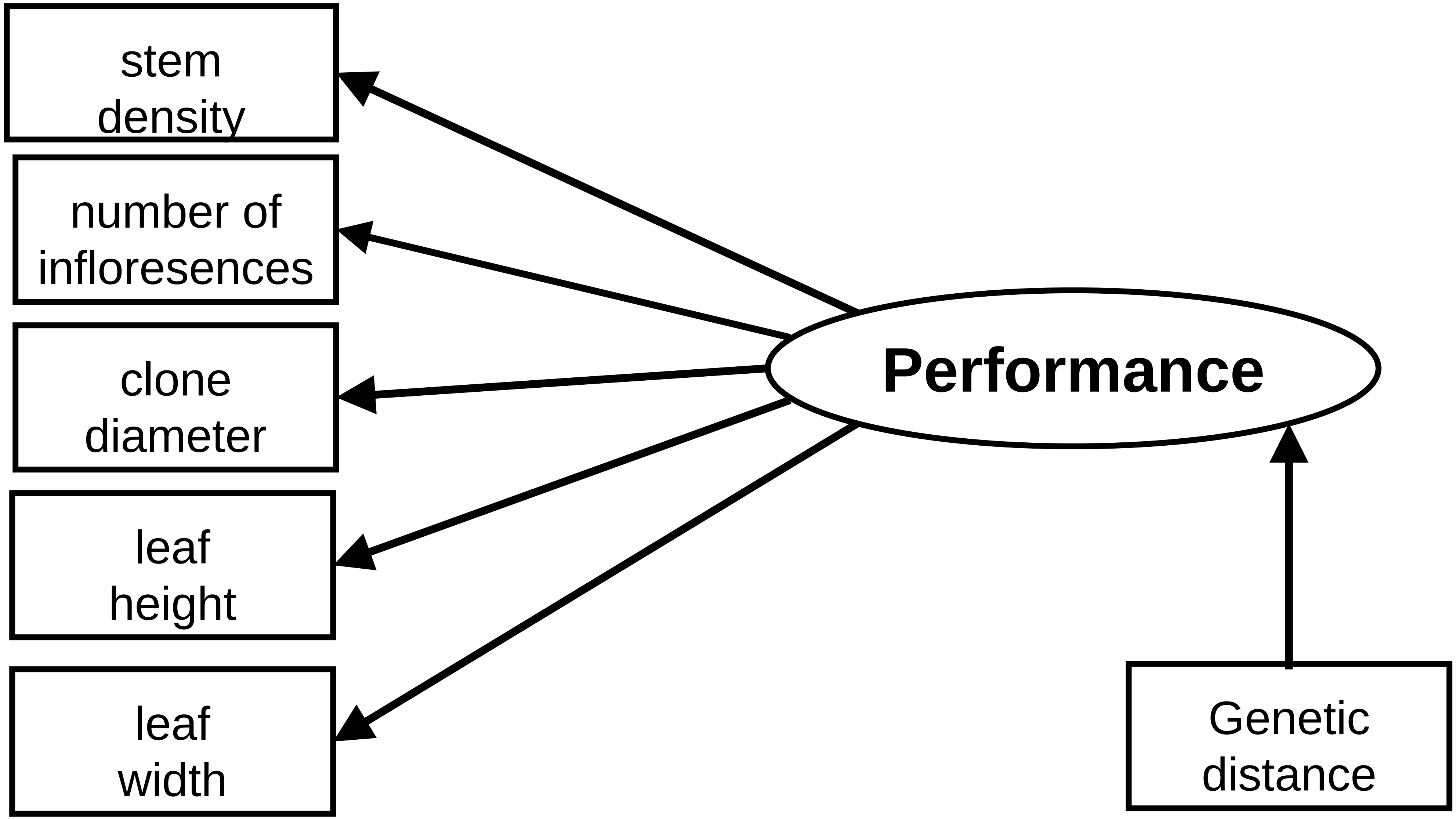
Travis, S. E., & Grace, J. B. (2010). Predicting performance for ecological restoration: a case study using Spartina alterniflora. Ecological Applications, 20(1), 192-204.

Confirmatory Factor Analysis

- Collected individuals of the salt marsh plant *Spartina alterniflora* eight clones each from 23 populations
- Transplanted individuals and measured their performance relative to local populations.
- Performance was approximated with stem density, the number of infloresences, clone diameter, leaf height, and leaf width



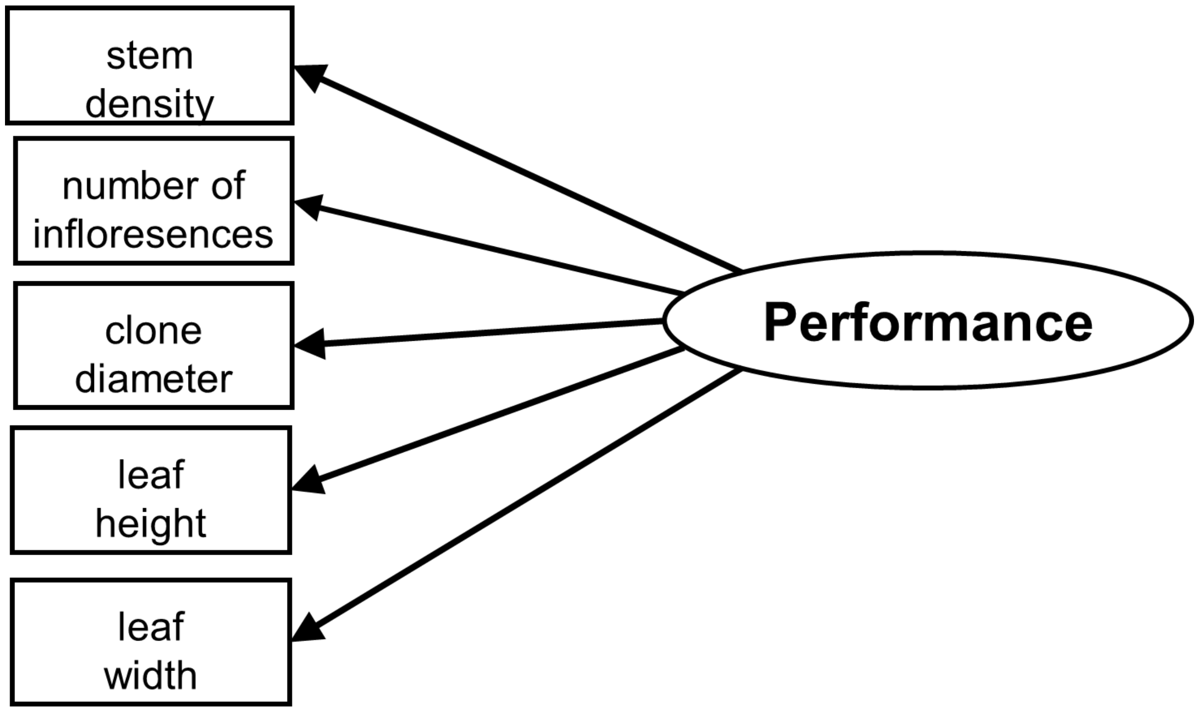
Confirmatory Factor Analysis



Confirmatory Factor Analysis

1) Evaluate the latent relationships among variables (**Confirmatory Factor Analysis**).

- Do our indicators make a Good Latent Variable?

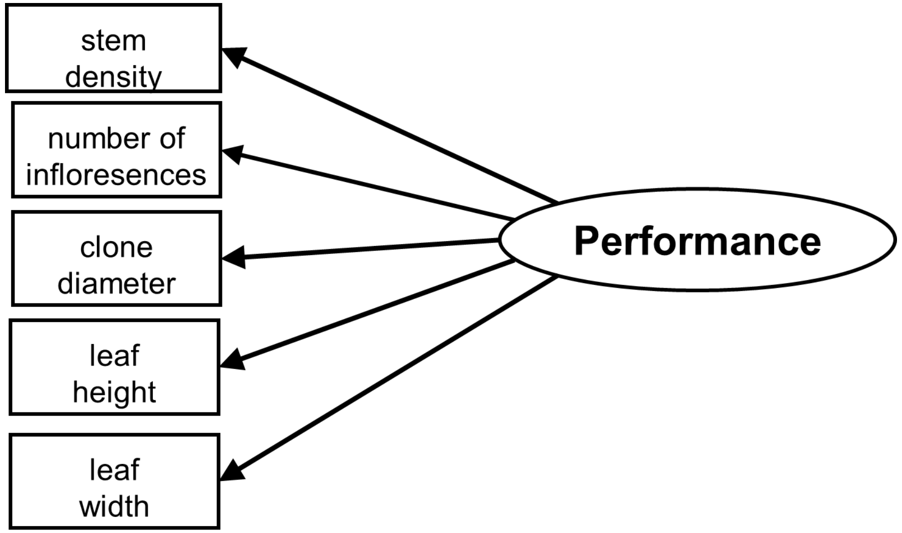


A first step is to analyze the “measurement model” using CFA.

Confirmatory Factor Analysis

Exercise

```
# Read and check the data
travis <- read.csv("Travis_data.csv")

str(travis)
# correlations
cor(travis[, 4:8])
```

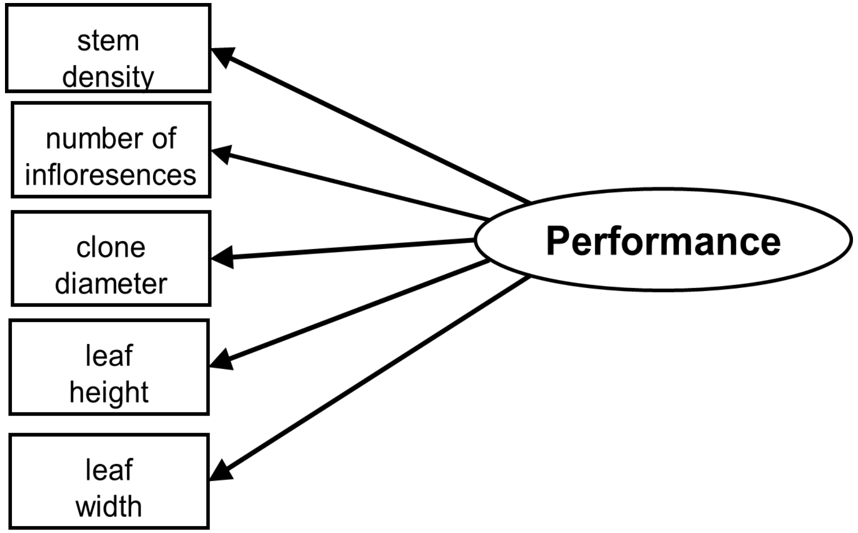



```
> cor(travis[, 4:8])
              stems     infls clonediam    leafht  leafwdth
stems     1.0000000 0.8339227 0.9333150 0.7275625 0.6457378
infls     0.8339227 1.0000000 0.8126388 0.6925888 0.6026302
clonediam 0.9333150 0.8126388 1.0000000 0.7729843 0.7296621
leafht    0.7275625 0.6925888 0.7729843 1.0000000 0.9687725
leafwdth  0.6457378 0.6026302 0.7296621 0.9687725 1.0000000
```

Confirmatory Factor Analysis

Exercise

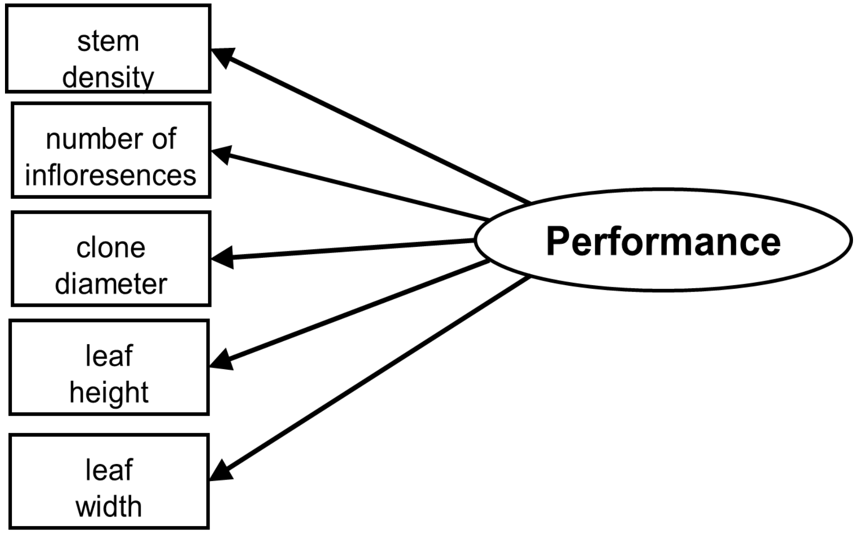
```
# specify the model
cfa_mod <- `
performance =~ stems + infls + clonediam + leafht + leafwdth
`
# fit the model
cfa_fit <- sem(cfa_mod, travis)
```




```
Warning message:
In lav_object_post_check(object) :
  lavaan WARNING: some estimated ov variances are negative
```

Confirmatory Factor Analysis

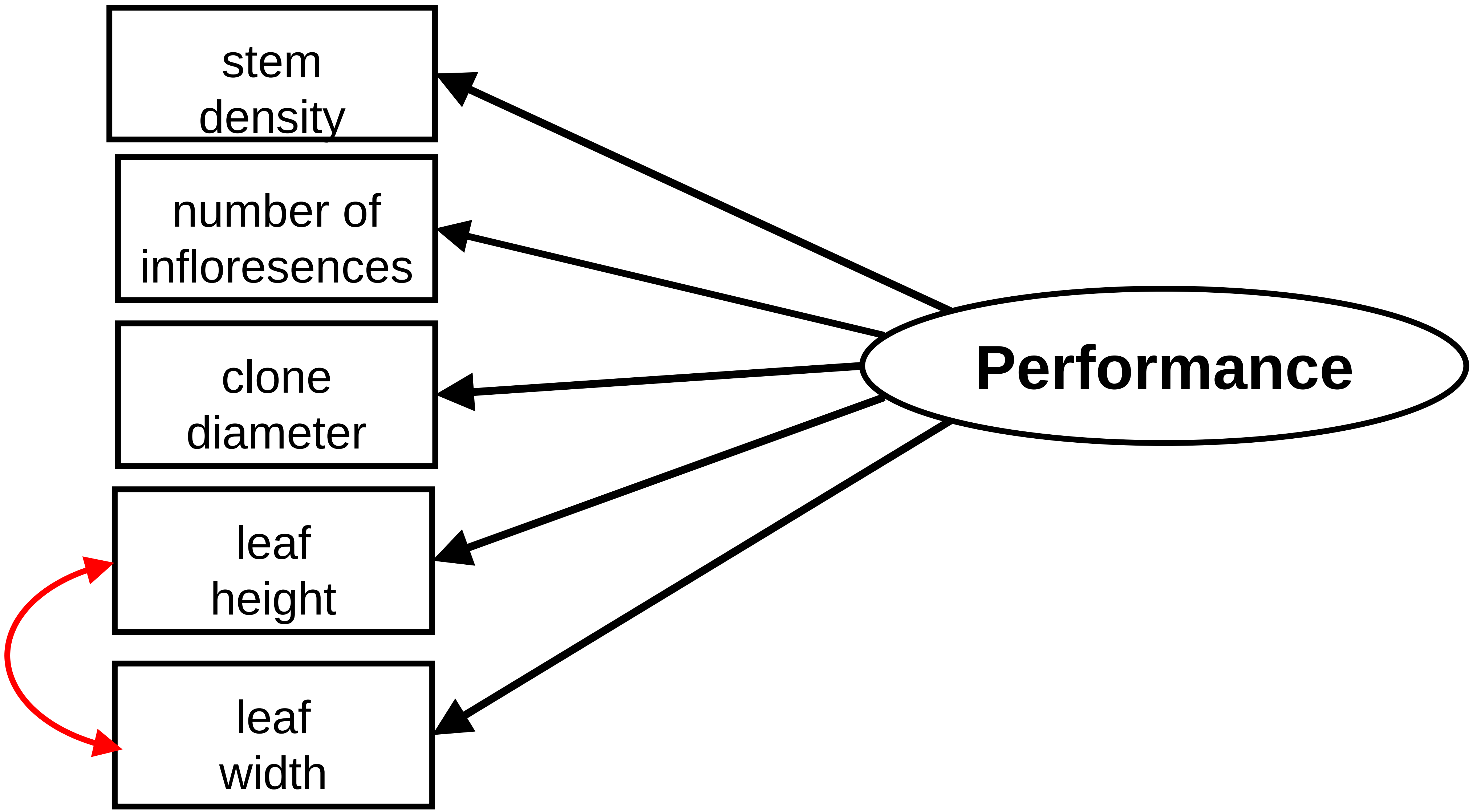

```
> summary(cfa_fit)

lavaan 0.6-9 ended normally after 82 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        10

  Number of observations                            23
Model Test User Model:

  Test statistic                                51.106
  Degrees of freedom                                 5
  P-value (Chi-square)                           0.000
```



Confirmatory Factor Analysis

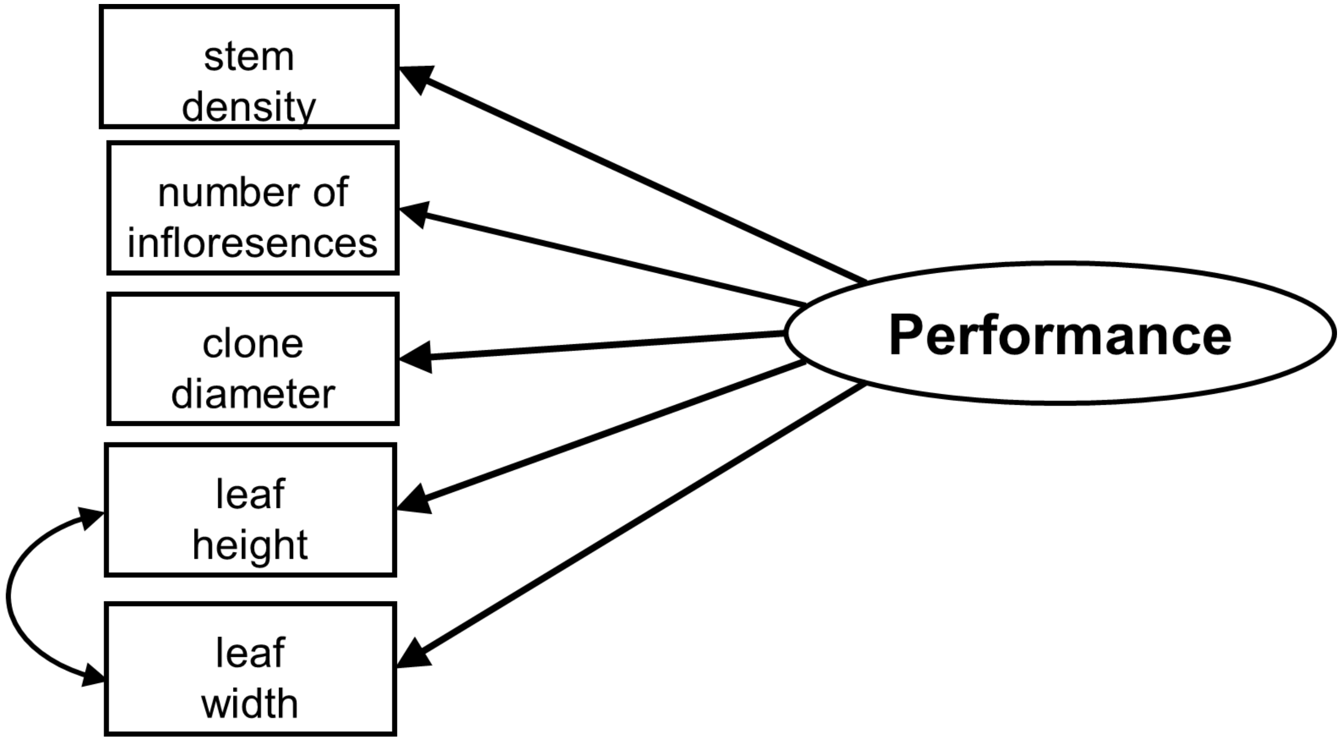

```
> modindices(cfa_fit)
         lhs op       rhs     mi     epc sepc.lv sepc.all sepc.nox
12     stems ~~     infls 10.470  11.784  11.784    0.677    0.677
13     stems ~~ clonediam 17.152 112.521 112.521    0.871    0.871
14     stems ~~    leafht  0.693  -7.889  -7.889   -0.517   -0.517
15     stems ~~  leafwdth  2.214  -1.836  -1.836   -0.346   -0.346
16     infls ~~ clonediam  8.773  11.092  11.092    0.621    0.621
17     infls ~~    leafht  0.062  -0.312  -0.312   -0.148   -0.148
18     infls ~~  leafwdth  2.906  -0.281  -0.281   -0.383   -0.383
19 clonediam ~~    leafht  4.028 -21.233 -21.233   -1.357   -1.357
20 clonediam ~~  leafwdth  0.037  -0.261  -0.261   -0.048   -0.048
21    leafht ~~  leafwdth 37.862  17.177  17.177   26.752   26.752
```

Confirmatory Factor Analysis



Confirmatory Factor Analysis


```
cfa_mod2 <- '
performance =~ stems + infls + clonediam + leafht + leafwdth
leafht ~~ leafwdth
'
cfa_fit2 <- sem(cfa_mod2, travis)
summary(cfa_fit2)
```



Confirmatory Factor Analysis


```
Estimator                                          ML
  Optimization method                          NLMINB
  Number of model parameters                       11

  Number of observations                           23

Model Test User Model:

  Test statistic                                7.410
  Degrees of freedom                                4
  P-value (Chi-square)                          0.116
```

Confirmatory Factor Analysis

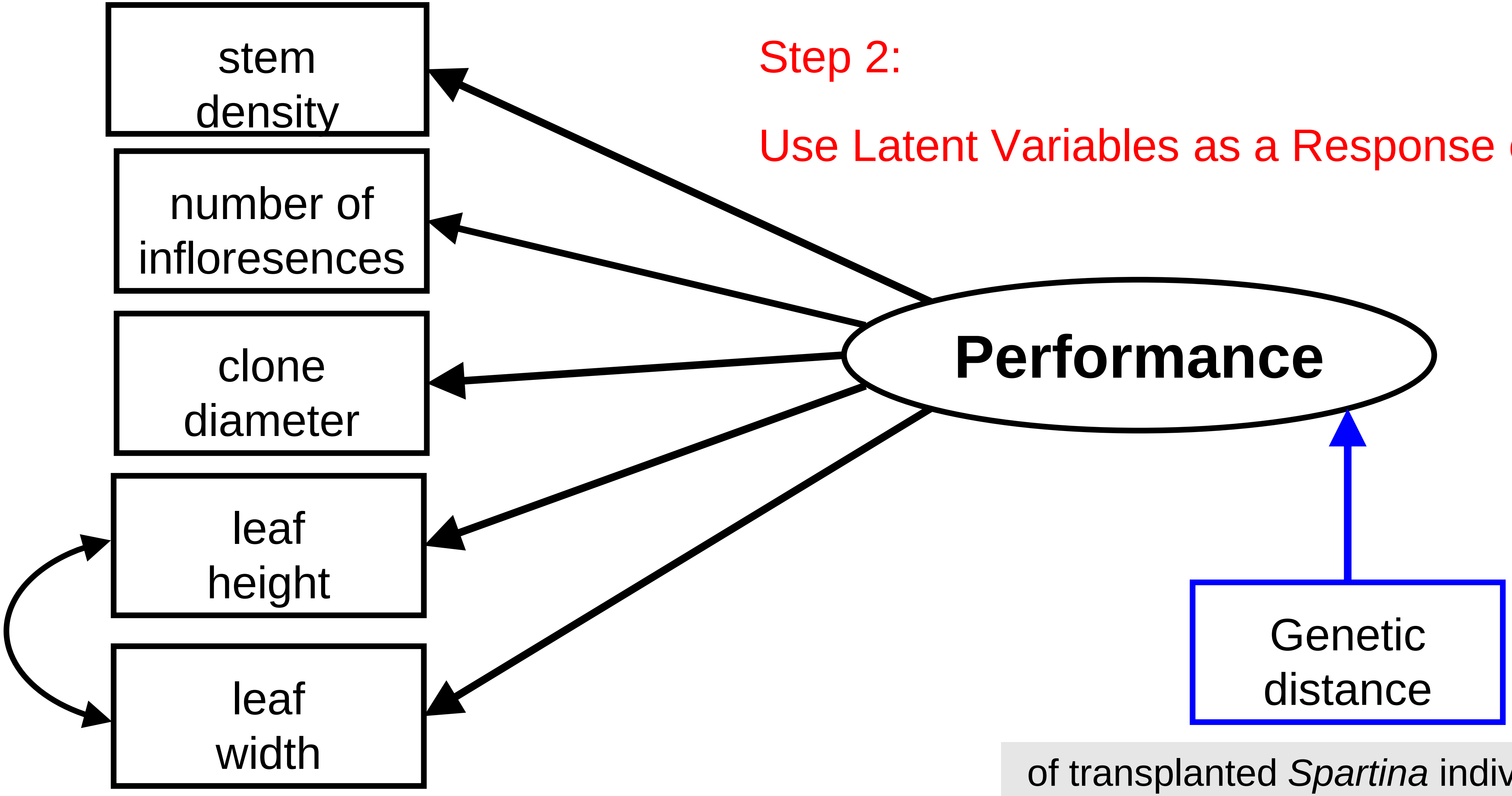

```
...

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  performance =~
    stems             1.000
    infls             0.117    0.016    7.173    0.000
    clonediam         1.086    0.096   11.319    0.000
    leafht            0.697    0.127    5.509    0.000
    leafwdth          0.082    0.018    4.529    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
 .leafht ~~
   .leafwdth         10.831    3.432    3.156    0.002
```

CFA as a part of structural model

Exercise

Step 2:

Use Latent Variables as a Response or a Predictor



of transplanted *Spartina* individuals from the local population

CFA as a part of structural model

Exercise

```
SEM_latent_mod <- `
            # latent
performance =~ stems + infls + clonediam + leafht + leafwdth

            # structural paths
performance ~ geneticdist

             # correlated errors
leafht ~~ leafwdth
'

SEM_latent_fit <- sem(SEM_latent_mod , travis)

summary(SEM_latent_fit, standardize = T, rsq = T, fit.measures=T)
```

CFA as a part of structural model


```
Estimator                                          ML
  Optimization method                          NLMINB
  Number of model parameters                       12

  Number of observations                           23

Model Test User Model:

  Test statistic                               12.237
  Degrees of freedom                                8
  P-value (Chi-square)                          0.141
```

CFA as a part of structural model

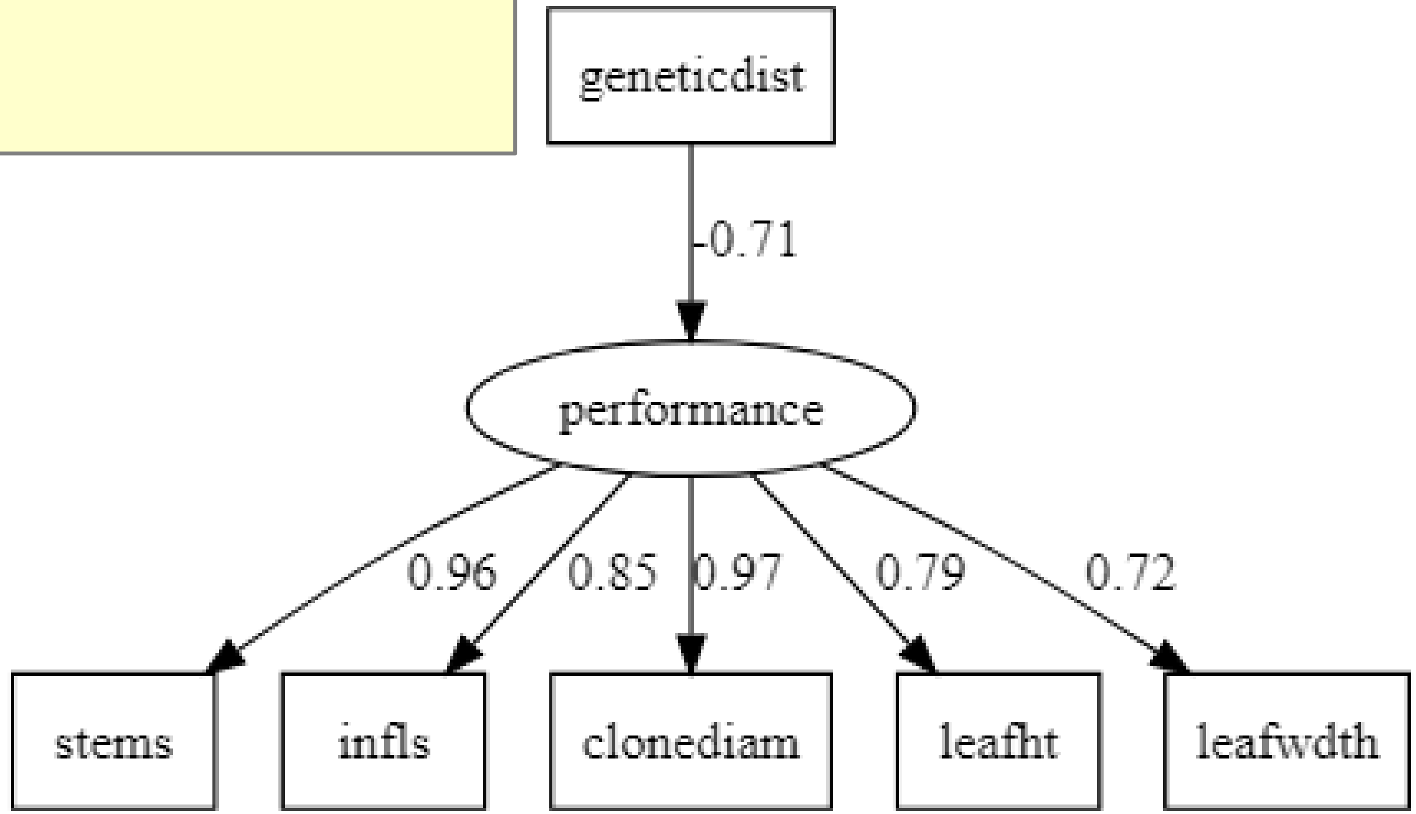
Exercise

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  performance =~
    stems             1.000                               15.555    0.962
    infls             0.117    0.017    6.929    0.000     1.822    0.853
    clonediam         1.106    0.096   11.508    0.000    17.199    0.969
    leafht            0.711    0.127    5.601    0.000    11.066    0.785
    leafwdth          0.084    0.018    4.650    0.000     1.308    0.718

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  performance ~
    geneticdist     -51.673   11.365   -4.547    0.000    -3.322   -0.708

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
 .leafht ~~
   .leafwdth         10.416    3.312    3.145    0.002    10.416    0.940
```

CFA as a part of structural model

Exercise

```
library(lavaanPlot)
lavaanPlot(model = SEM_latent_fit,
           coefs = TRUE, stand=TRUE,
           # graph_options = list(layout = "circo"),
           # stars = 'regress', # shows stars for regr coef
           digits = 2)
```



Day 5 Task 2

Human Impact Intensity

Macroinvertebrate body size

Body size traits

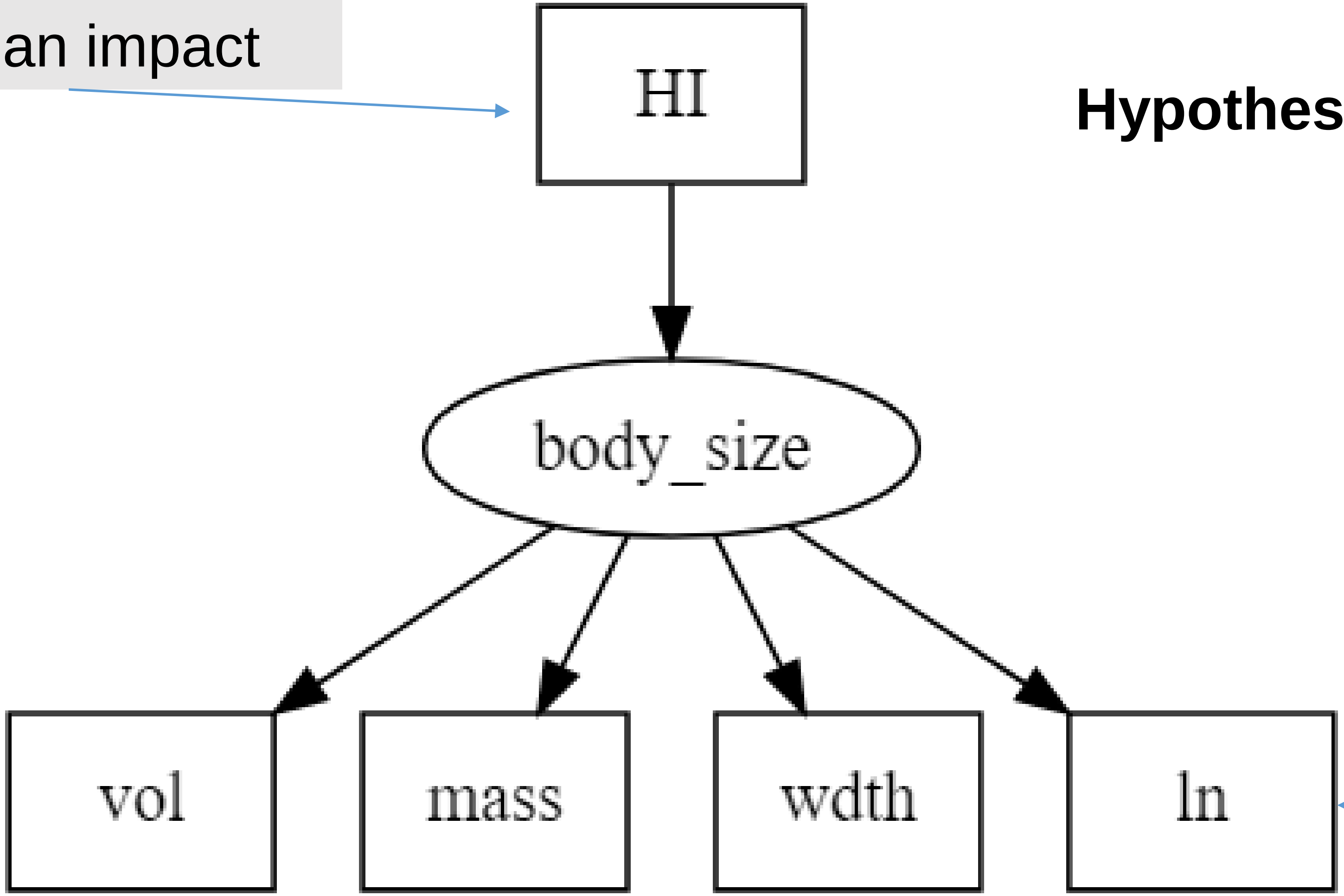
Body mass
Body volume
Body length
Body width

```
# Read and check the data

read.csv(" Bodysize_data.csv")
```

Day 5 Task 2

Human impact

Hypothesized model



Body volume
Body mass
Body width
Body length

Day 5 Task 2

1. Perform the confirmatory factor analysis for the latent variable “body size”
2. Use the results from step 1 and perform the SEM by adding human impact variable

3. Fill in Standardized Coeficients and R^2 for the model, add the fit indices

4. Think about how to interpret the results