Introduction to structural equation modeling and mixed models in R

Day 3 – Part 2: SEM

Oksana Buzhdygan

oksana.buzh@fu-berlin.de

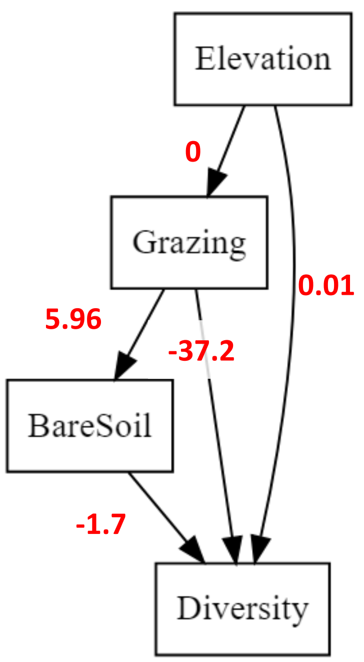
Outline

- Understanding path coefficients
  - ✓ Variance, covariance, correlation, regression coefficients
  - ✓ Indirect effects
  - ✓ Unexplained variances

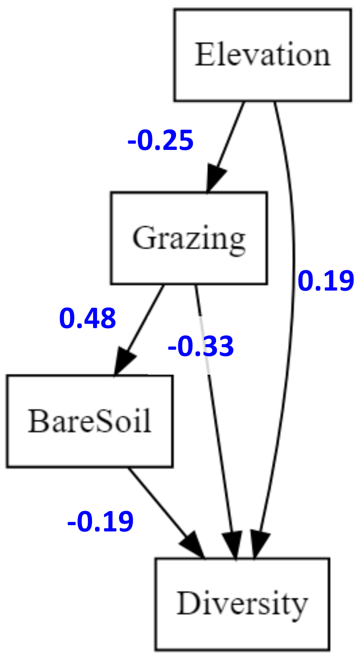
Path Coefficients

```
# Results from SEM model

> summary(sem.fit , standardize = T)

Parameter Estimates:
Regressions:
                   Estimate   Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  Grazing ~
    Elevation        -0.000     0.000   -2.475    0.013   -0.000   -0.252
  BareSoil ~
    Grazing           5.963     1.161    5.136    0.000    5.963    0.476
  Diversity ~
    Elevation         0.011     0.005    2.062    0.039    0.011    0.190
    Grazing         -37.259    11.739   -3.174    0.002  -37.259   -0.331
    BareSoil         -1.696     0.913   -1.856    0.063   -1.696   -0.189
```

Unstandardized



Standardized

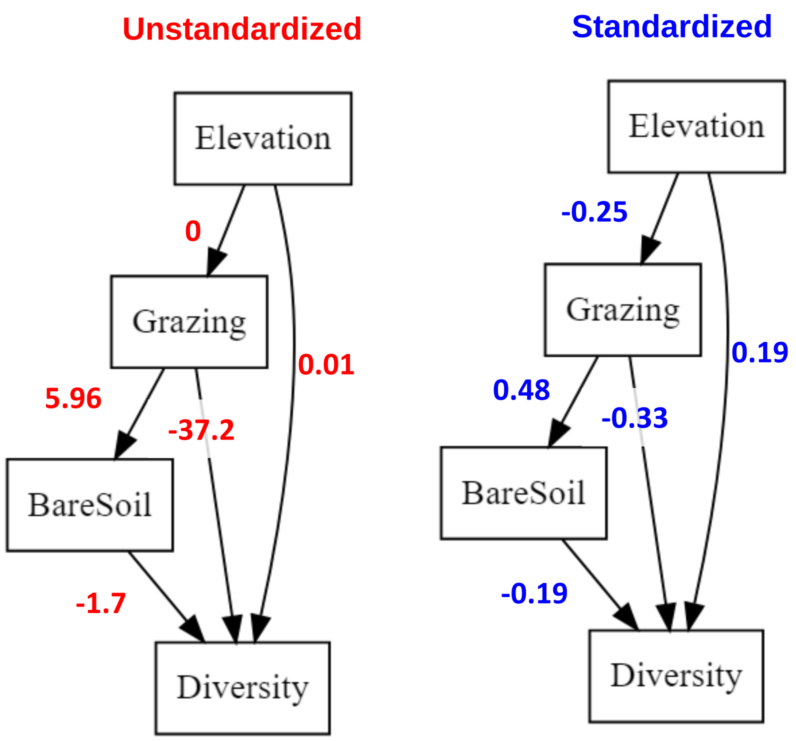


Path Coefficients

```
# Results from SEM model

> summary(sem.fit , standardize = T)

Parameter Estimates:
Regressions:
                   Estimate   Std.all
  Grazing ~
    Elevation        -0.000    -0.252
  BareSoil ~
    Grazing           5.963     0.476
  Diversity ~
    Elevation         0.011     0.190
    Grazing         -37.259    -0.331
    BareSoil         -1.696    -0.189
```

The building blocks of path coefficients

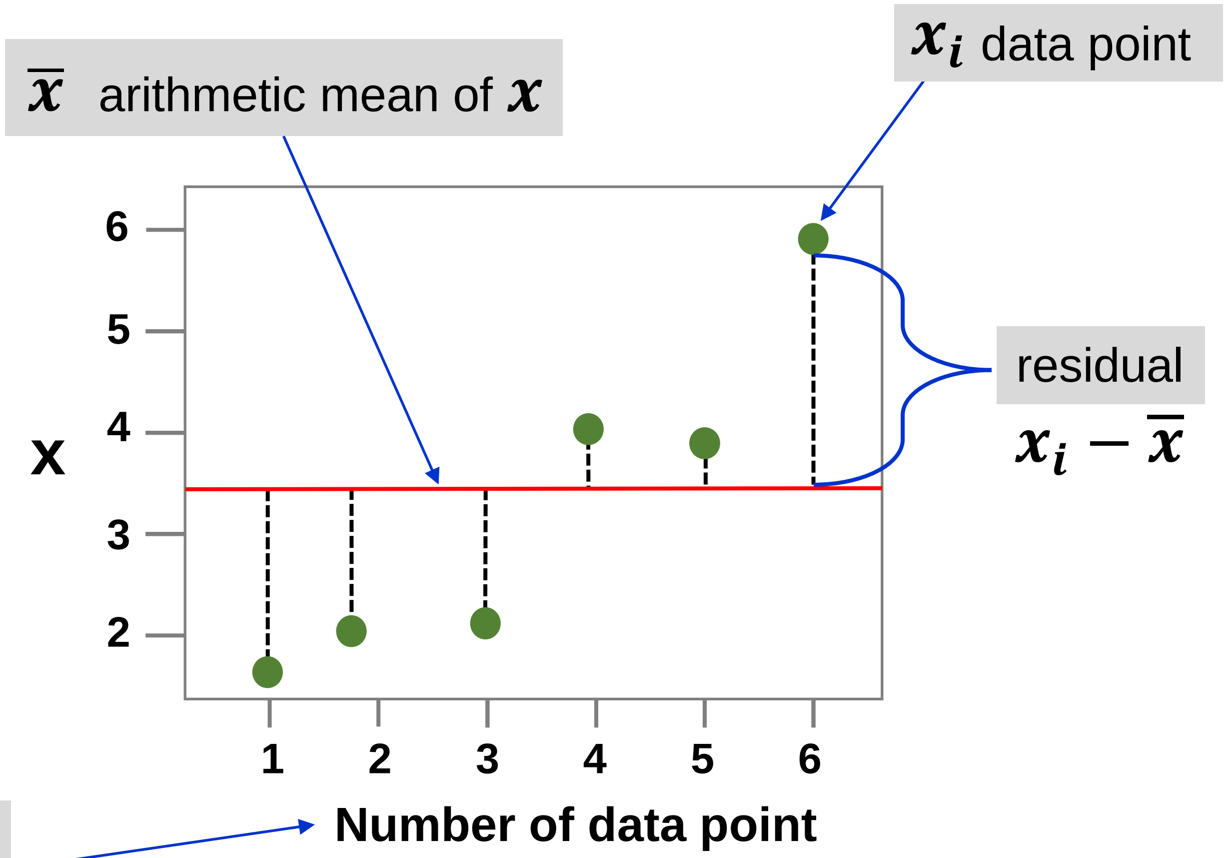
- variances,
- covariances,
- correlations,
- regression coefficients



What is Variance?

```
# in R

x <- c(1, 2, 3, 4)

var(x) # Variance

[1] 1.666667
```

$$VAR_x = \frac{\sum(x_i - \bar{x})^2}{n-1}$$



sample size
n = 6

Path Coefficients

```
# in R
x <- c(1, 2, 3, 4)
var(x) # Variance
[1] 1.666667

y <- c(70, 30, 10, 90)
var(y) # Variance
[1] 1333.333

cov(x,y) # Covariance
[1] 6.666667


> mean(x)
[1] 2.5
> mean(y)
[1] 50
```

What is Covariance?

- Dependency between two variables
- Scaled to the raw values

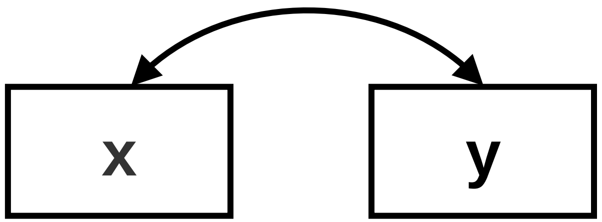
$$VAR_x = \frac{\sum(x_i - \overline{x})^2}{n-1}$$

Covariance Matrix

	x	y
x	1.66	
y	**6.66**	1333.3

$$VAR_y = \frac{\sum(y_i - \overline{y})^2}{n-1}$$

$$COV_{xy} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{n-1}$$



Path Coefficients

What is Correlation?

```
# Covariance
cov(x,y)
[1] 6.666667

# Correlation
cor(x, y)
[1] 0.14
# calculate by hand
 cov(x, y)/(sd(x)*sd(y))
[1] 0.14
```

Covariance Matrix

	x	y
x	1.66	
y	6.66	1333.3

Raw Covariance Matrix

$$COV_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Correlation Matrix

	x	y
x	1	
y	0.14	1

Standardised Covariance Matrix

$$r_{xy} = \frac{COV_{xy}}{SD_x \times SD_y}$$

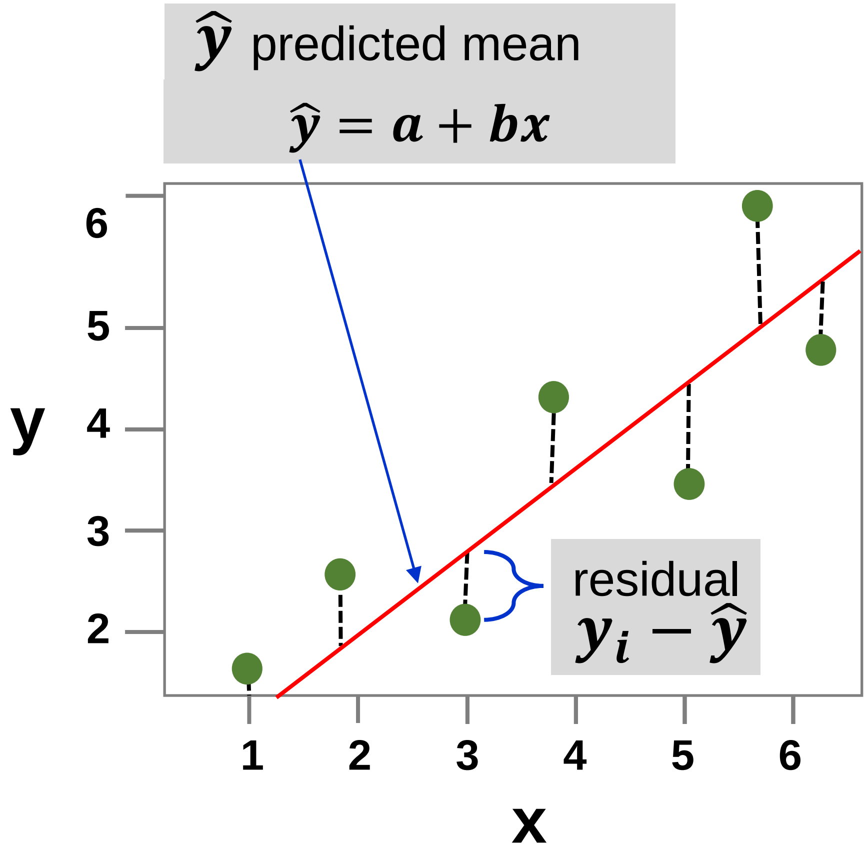
standard deviation of the mean
(the square-root of the variance)

Path Coefficients

What is Regression Coefficient?

$$y = a + bx$$

$$x \xrightarrow{b_{xy}} y$$



$$b_{xy} = \frac{COV_{xy}}{VAR_x}$$

Unstandardized regression coefficient

$$b_{xy} = r_{xy} = \frac{COV_{xy}}{SD_x \times SD_y}$$

Standardized regression coefficient

Path Coefficients

What is Regression Coefficient?

$$y = a + \boldsymbol{b}x$$

$$\boxed{x} \xrightarrow{b_{xy}} \boxed{y}$$

When two variables are connected by a single path, the coefficient of that path is the correlation coefficient

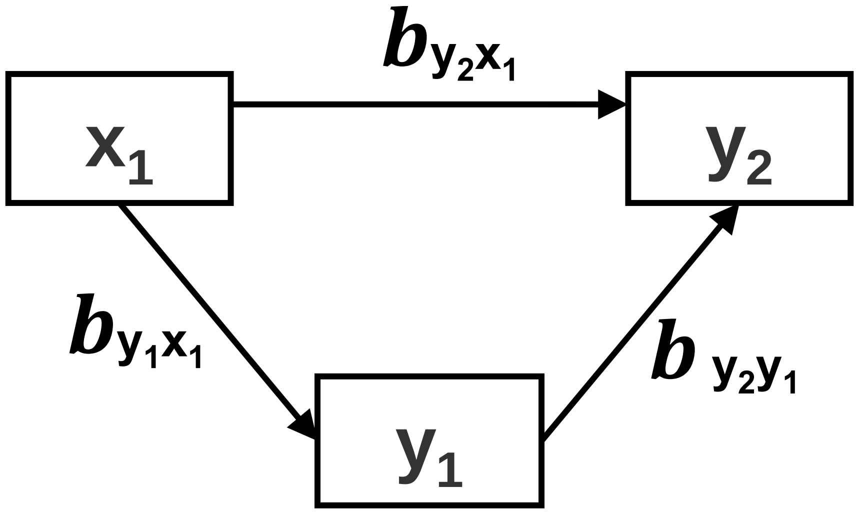
$$b_{xy} = \frac{COV_{xy}}{VAR_x}$$

Unstandardized regression coefficient

$$\boxed{b_{xy} = r_{xy}} = \frac{COV_{xy}}{SD_x \times SD_y}$$

Standardized regression coefficient

Path Coefficients



When variables are connected by more than one path, each path coefficient is the <u>'partial'</u> regression coefficient.

takes the bivariate correlation between **x_1** and **y_2**

removes the joint influence of **x_1** and **y_1** on **y_2**

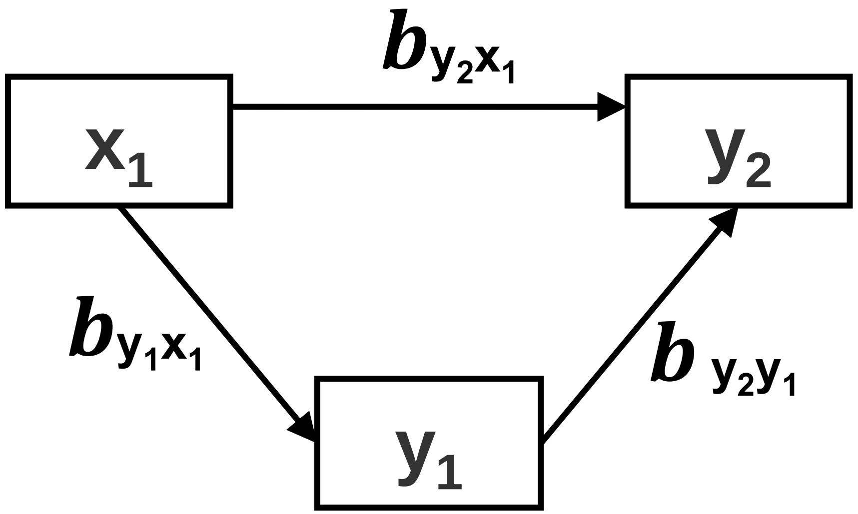
$$b_{y2x1} = \frac{r_{x1y2} - (r_{x1y1} \times r_{y1y2})}{1 - r_{x1y1}^2}$$

scales this effect by the shared variance between **x_1** and **y_1**

Corresponding equations:

$$y_1 = b_1 x_1$$
$$y_2 = b_2 x_1 + b_3 y_1$$

Path Coefficients

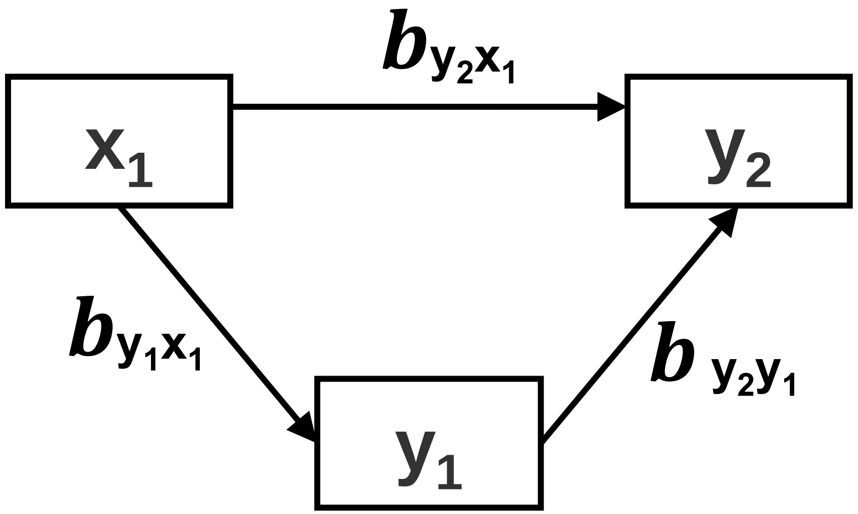


When variables are connected by more than one path, each path coefficient is the <u>'partial'</u> regression coefficient.

Standardized $$b_{y2x1} = \frac{r_{x1y2} - (r_{x1y1} \times r_{y1y2})}{1 - r^2_{x1y1}}$$

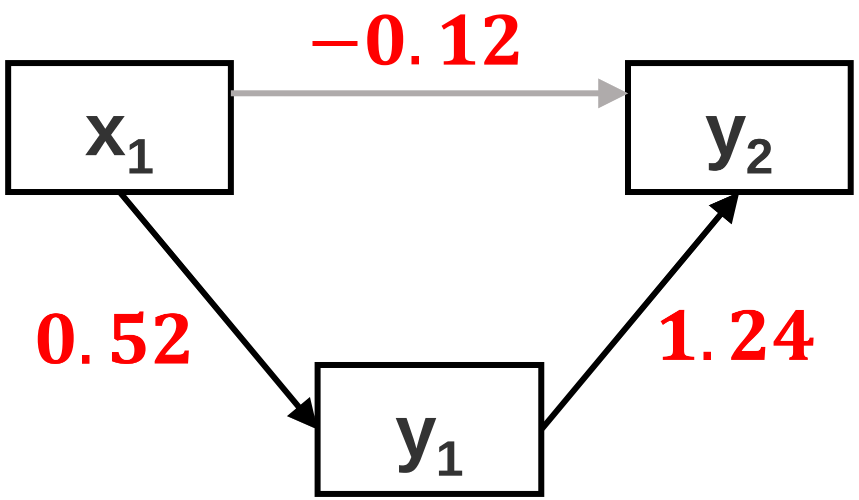
Unstandardized $$b_{y2x1} = \frac{SD_{y2}}{SD_{x1}} \times \frac{r_{x1y2} - (r_{x1y1} \times r_{y1y2})}{1 - r^2_{x1y1}}$$

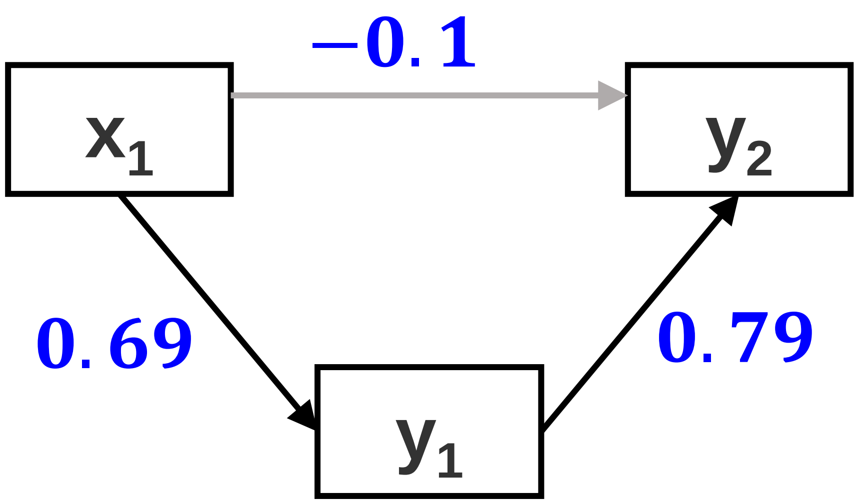
Path Coefficients




```
data1 <- read.table("Data/SEMdata1.txt",
                                          header = T)


# Specify the model in lavaan
sem_mod1 <- ` y1 ~ x1
              y2 ~  x1 + y1
`
# Fit the model
sem.fit1 <- sem(sem_mod1, data=data1)

# Extract results
summary(sem.fit1, standardize = T)
```

Path Coefficients

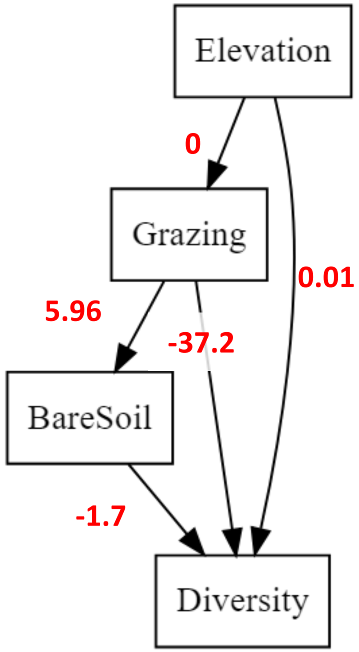


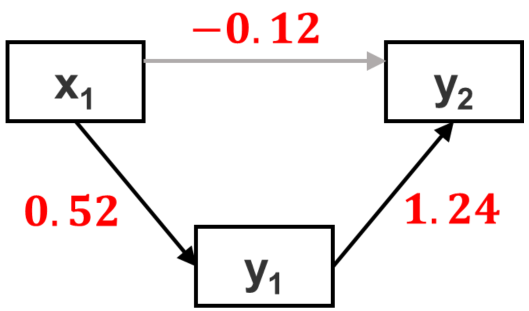


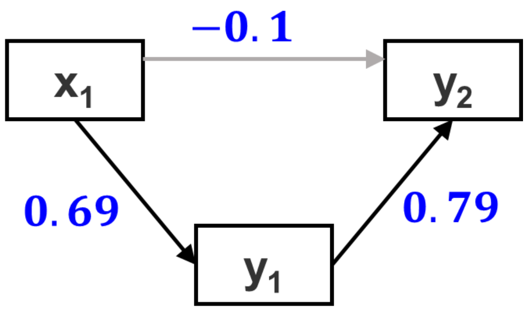

```
# Results
...
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  y1 ~
    x1                0.517    0.054    9.525    0.000    0.517    0.690
  y2 ~
    x1               -0.116    0.113   -1.034    0.301   -0.116   -0.099
    y1                1.239    0.150    8.248    0.000    1.239    0.787
```

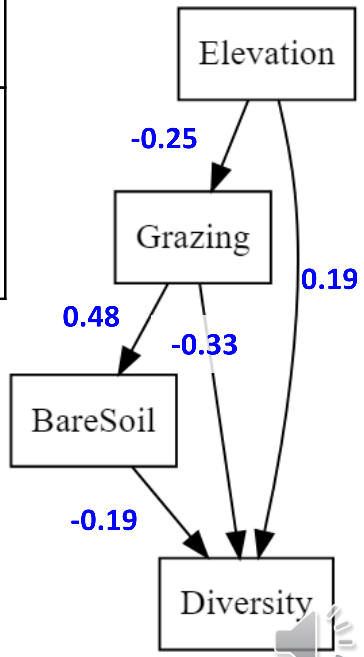
Path Coefficients

Unstandardized path coefficients	Standardized path coefficients
• Good for prediction: coefficients are in raw units	• Good for ranking: coefficients are in equivalent units
• Has direct real world meaning	• Less clear real world meaning
• Can be compared across pathways or models that have identical units	• Can be compared across all pathways in all models







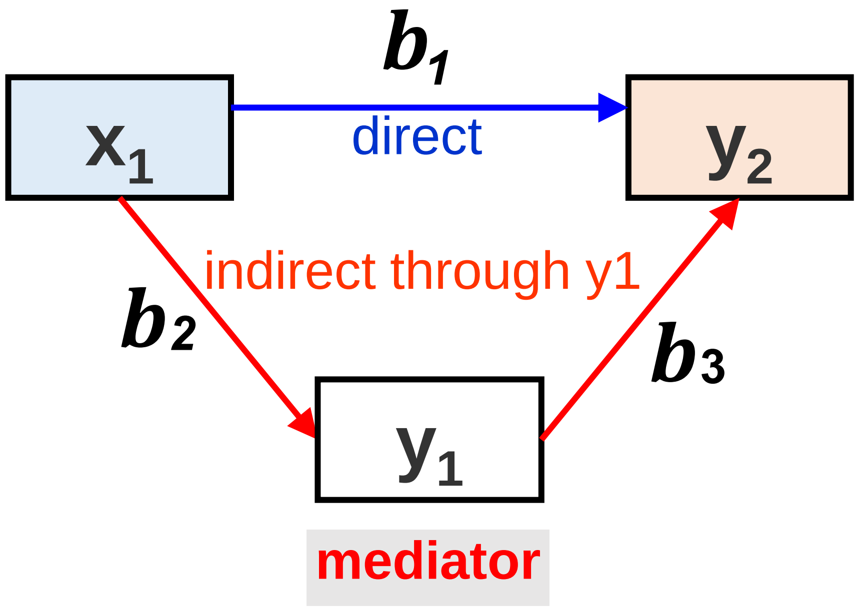


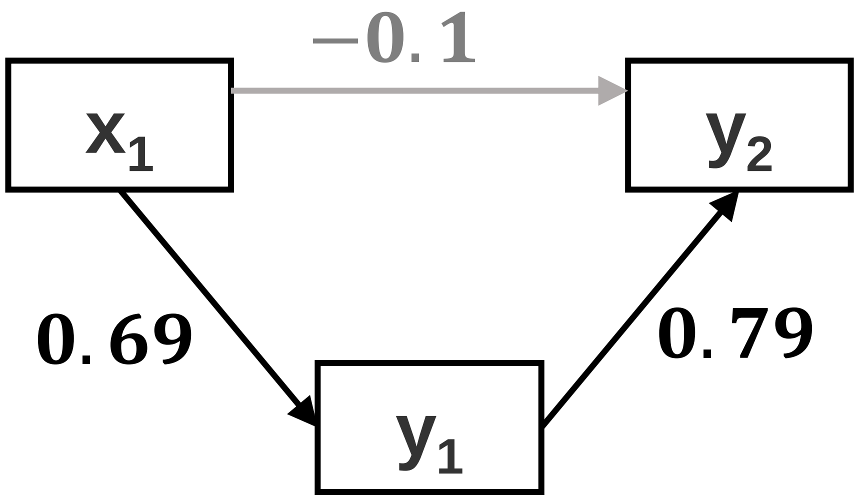
Outline

- Understanding path coefficients
  - ✓ Variance, covariance, correlation, regression coefficients
  - ✓ **Indirect effects**
  - ✓ Unexplained variances

Indirect effects

Effects of x_1 on y_2





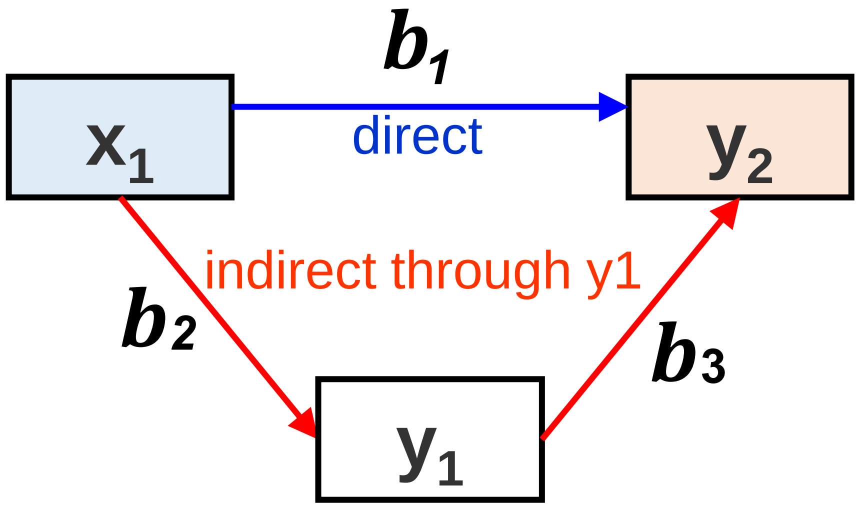
direct b_1

indirect $b_2 \times b_3$

Total effect = direct + indirect

direct	**-0.1**
indirect	$0.69 \times 0.79 = \mathbf{0.54}$
total	$-0.1 + 0.55 = \mathbf{0.44}$

Indirect effects



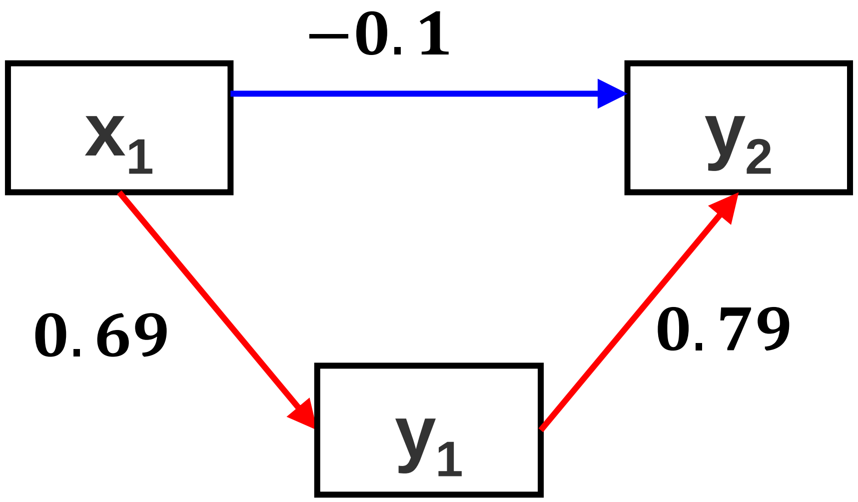
direct b_1

indirect $b_2 \times b_3$

Total effect = direct + indirect

```
# Naming the coefficients in lavaan
sem_mod1 <- '
     y2 ~ b1*x1 + b3*y1
     y1 ~ b2*x1
     # define direct, indirect and total effects
     direct   := b1
     indirect := b2*b3
     total    := b1 + (b2*b3)
     # or
     # total   := direct + indirect

'
sem.fit1 <- sem(sem_mod1, data=data1)
summary(sem.fit1, standardize = T)
```

Indirect effects



direct **-0.1**

indirect $0.69 \times 0.79 =$ **0.54**

total $-0.1 + 0.55 =$ **0.44**

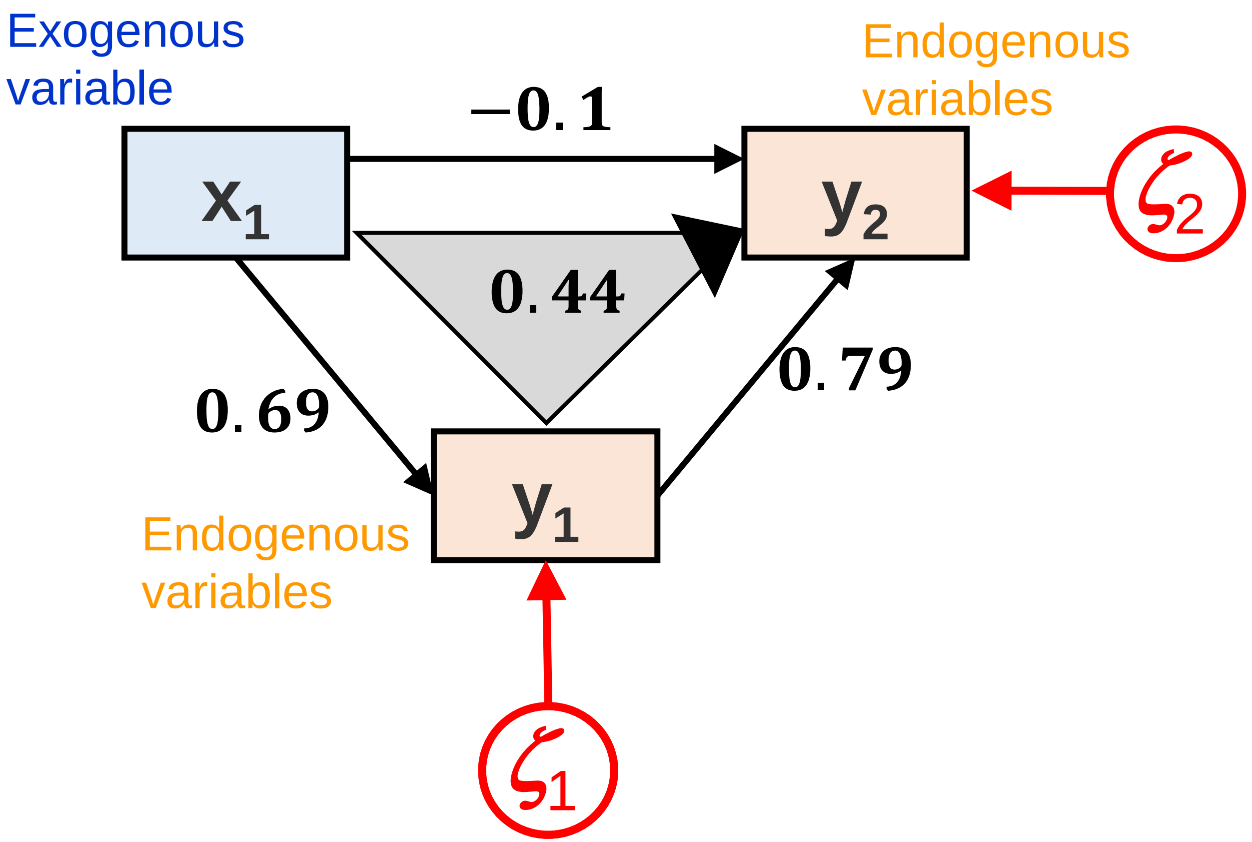
The total effect is equivalent to the total correlation

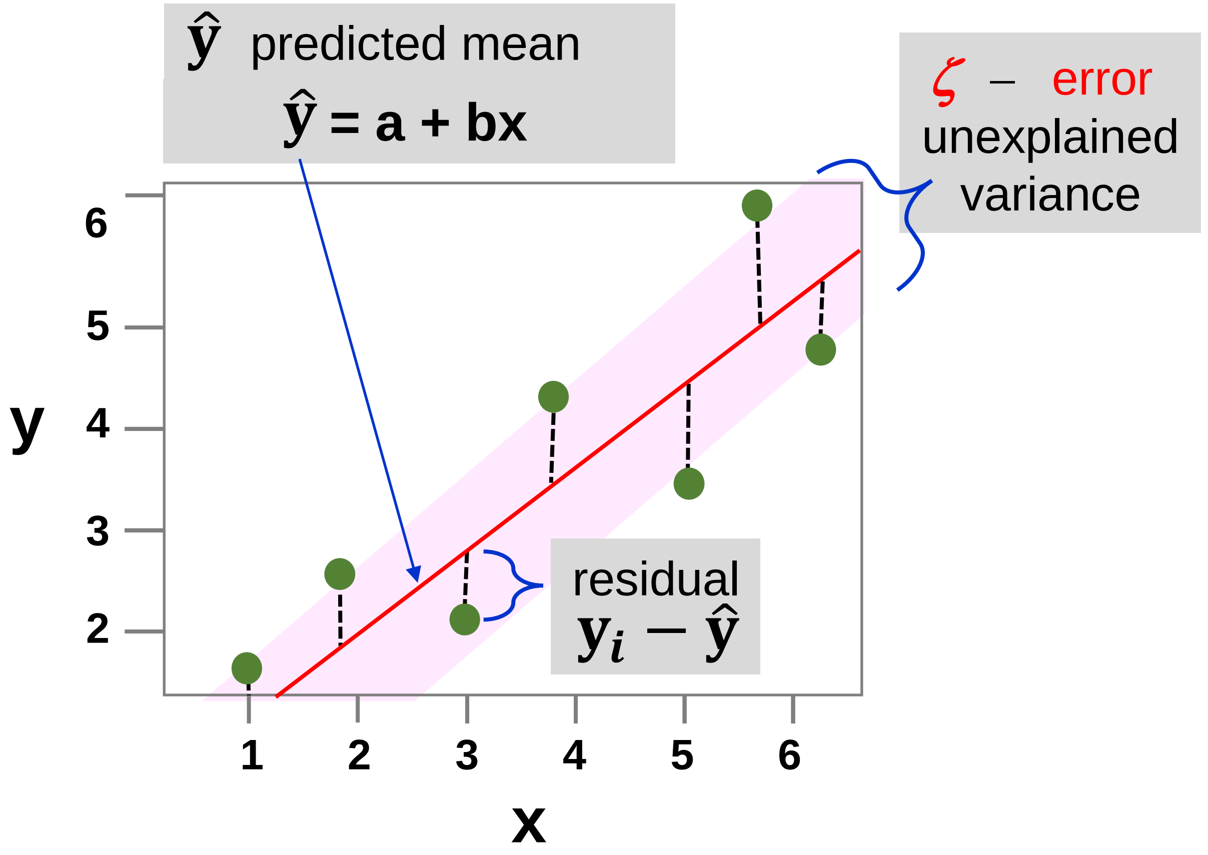
```
> summary(sem.fit1, standardize = T)
...
Defined Parameters:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
    direct           -0.116    0.113   -1.034    0.301   -0.116   -0.099
    indirect          0.640    0.103    6.235    0.000    0.640    0.543
    total             0.524    0.106    4.959    0.000    0.524    0.444
```

Outline

- Understanding path coefficients
  - ✓ Variance, covariance, correlation, regression coefficients
  - ✓ Indirect effects
  - ✓ **Unexplained variances**

Unexplained variances



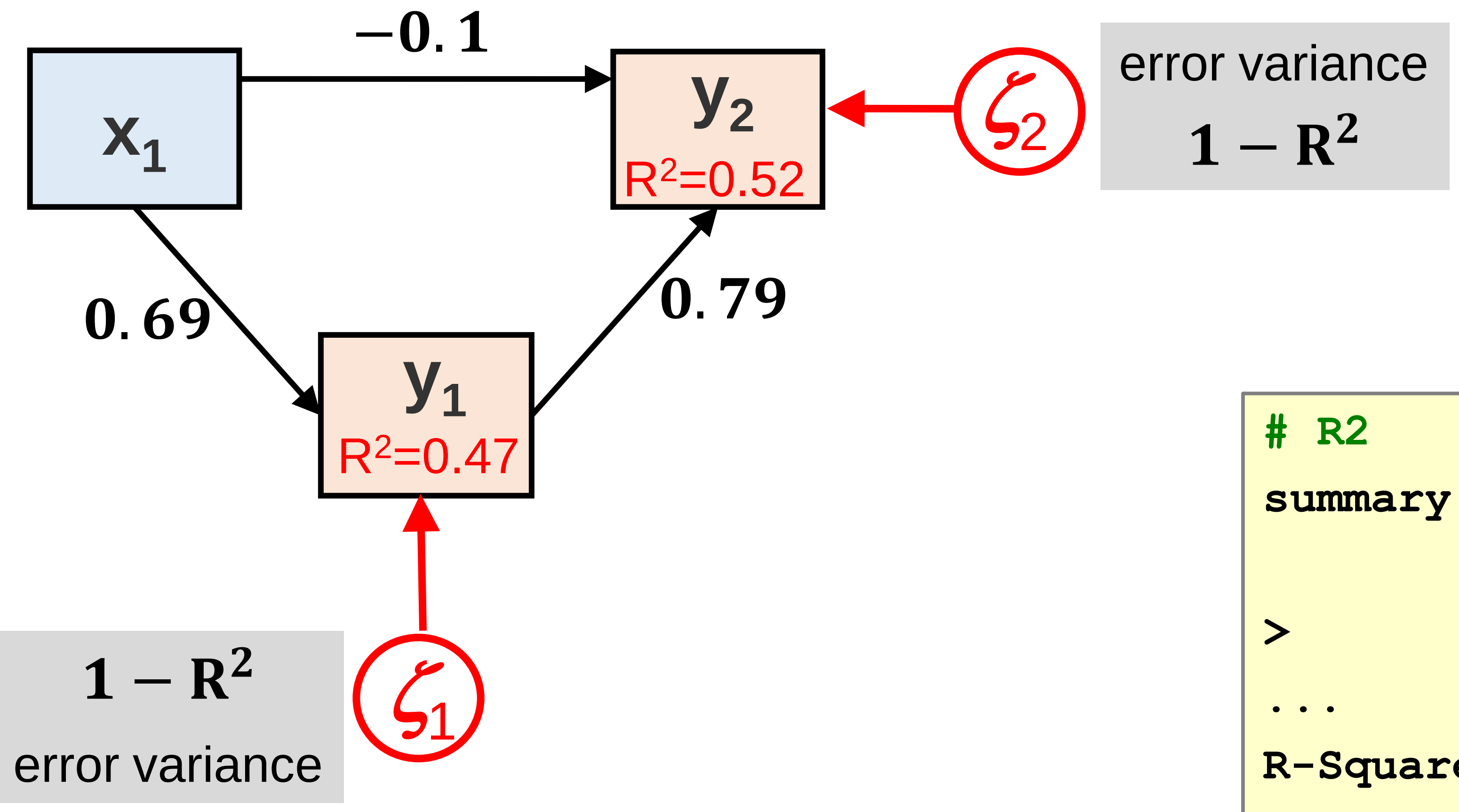


Equation form:

$$y_1 = a_1 + b_1 x_1 + \zeta_1$$

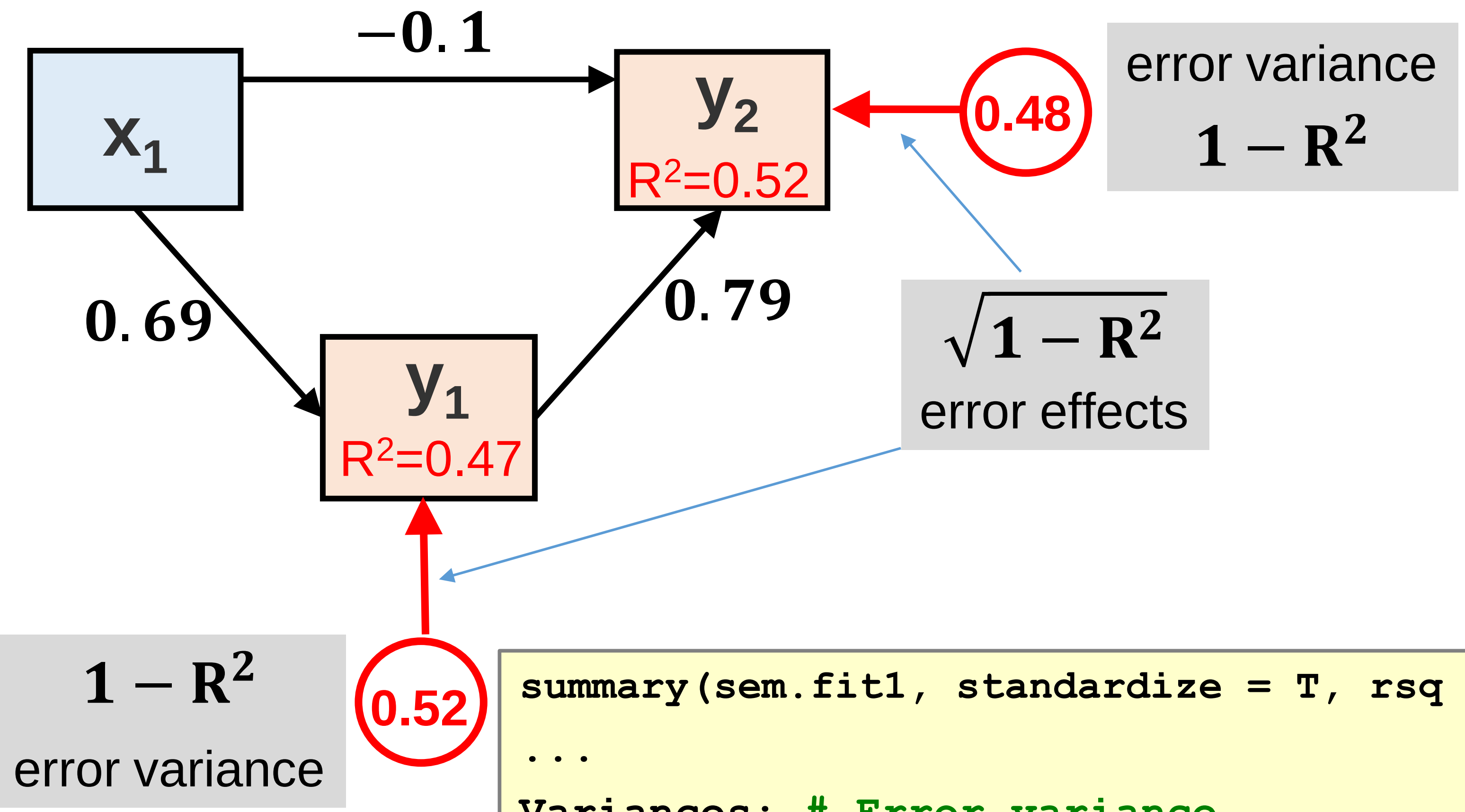
$$y_2 = a_2 + b_2 x_1 + b_3 y_1 + \zeta_2$$

Unexplained variances



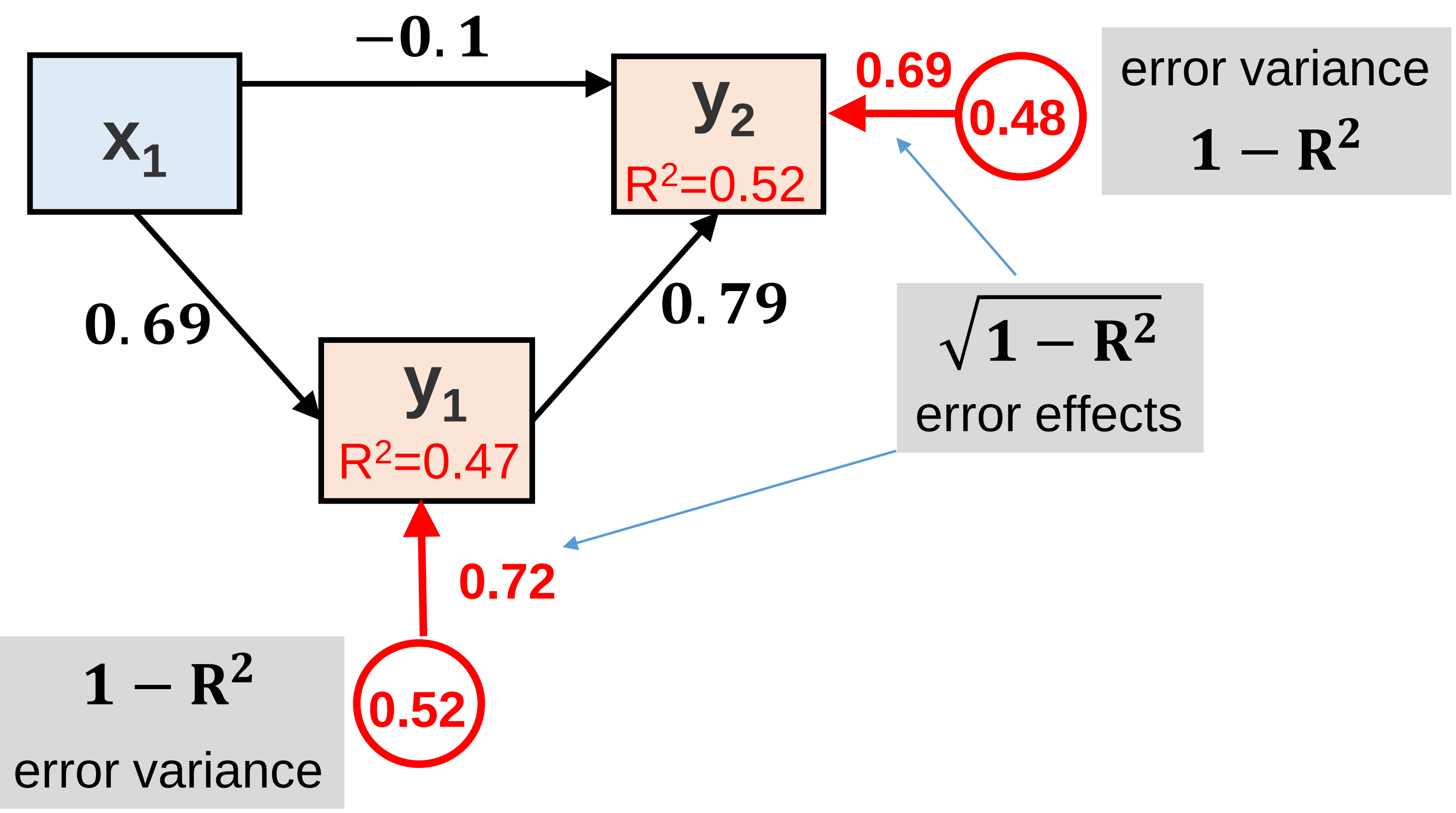

```
# R2
summary(sem.fit1, standardize = T,
                                rsq = T)
>
...
R-Square:
                         Estimate
    y1                      0.476
    y2                      0.517
```

Unexplained variances



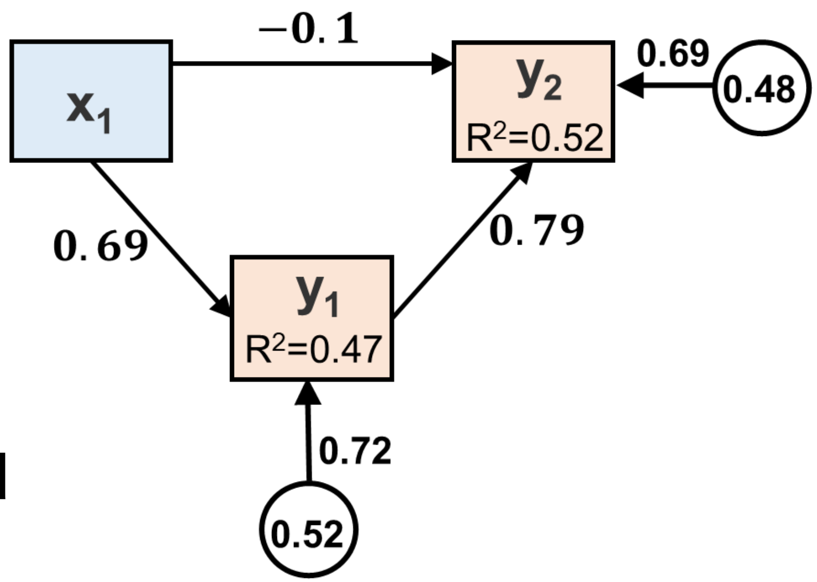

```
summary(sem.fit1, standardize = T, rsq = T)
...
Variances: # Error variance
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .y1                0.036    0.005    7.071    0.000    0.036    0.524
   .y2                0.082    0.012    7.071    0.000    0.082    0.483
```

Unexplained variances



Unexplained variances

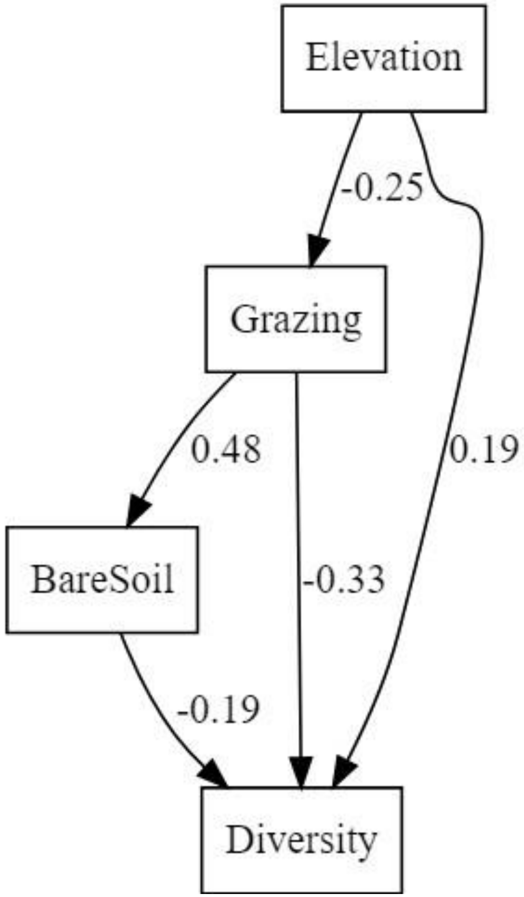
The major points to remember are:



- standardized coefficients reflect (partial) correlations;
- the indirect effect of one variable on another is obtained by multiplying the individual path coefficients (standardized or unstandardized);
- the total effect is the sum of direct and indirect paths;
- the bivariate correlation is the sum of the total effect plus any undirected paths.

Day 3 Task 1

Effects of grazing on plant diversity along elevation gradient




```
# data
data <- read.csv("Grassl_data.csv")
```

Day 3 Task 1

For the model on Fig. 1:

1. Calculate the standardised direct, indirect and total effects of ***grazing*** on ***diversity*** (do this in lavaan in R)
2. Define the exogenous and endogenous variables in the model
3. For each endogenous variable get the following:
   - the variance explained by the model
   - the error variance
   - the effect of the error (path coefficient with the error variance).

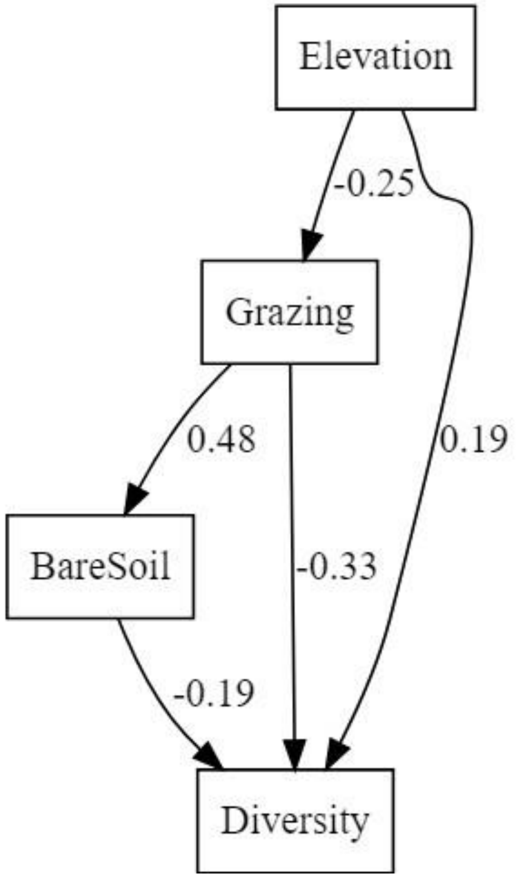


Fig. 1

Day 3 Task 1

For the model on Fig. 1:

1. Calculate the standardised direct, indirect and total effects of ***grazing*** on ***diversity*** (do this in lavaan in R)
2. Define the exogenous and endogenous variables in the model
3. For each endogenous variable get the following:
   - the variance explained by the model
   - the error variance
   - the effect of the error (path coefficient with the error variance).

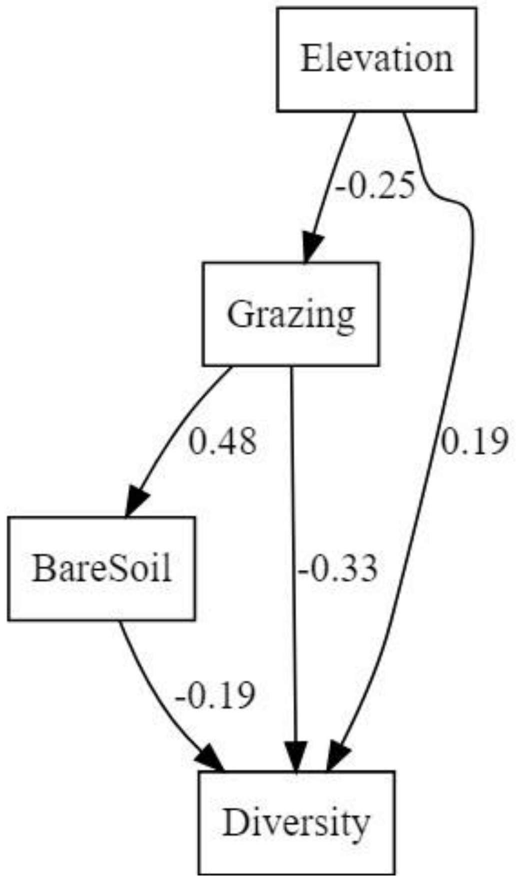


Fig. 1