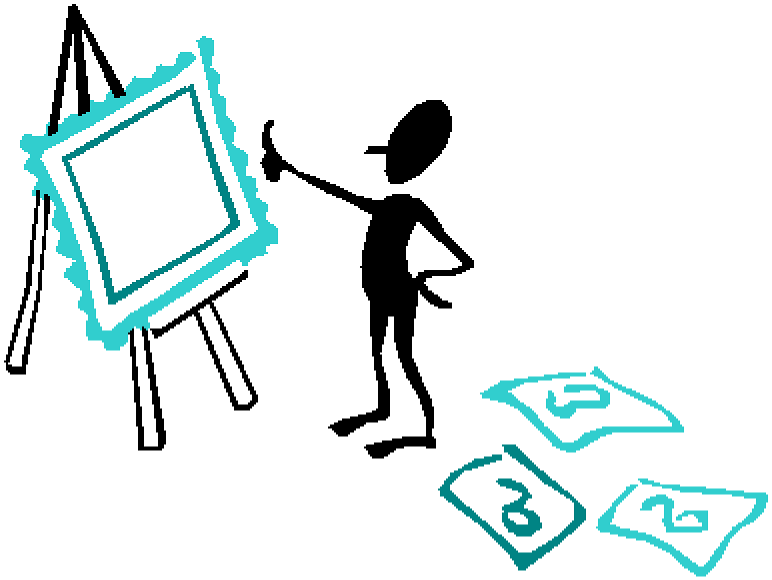
CORRELATION
PARTIAL CORRELATION
MULTIPLE REGRESSION

Modeling latent variables

Outline

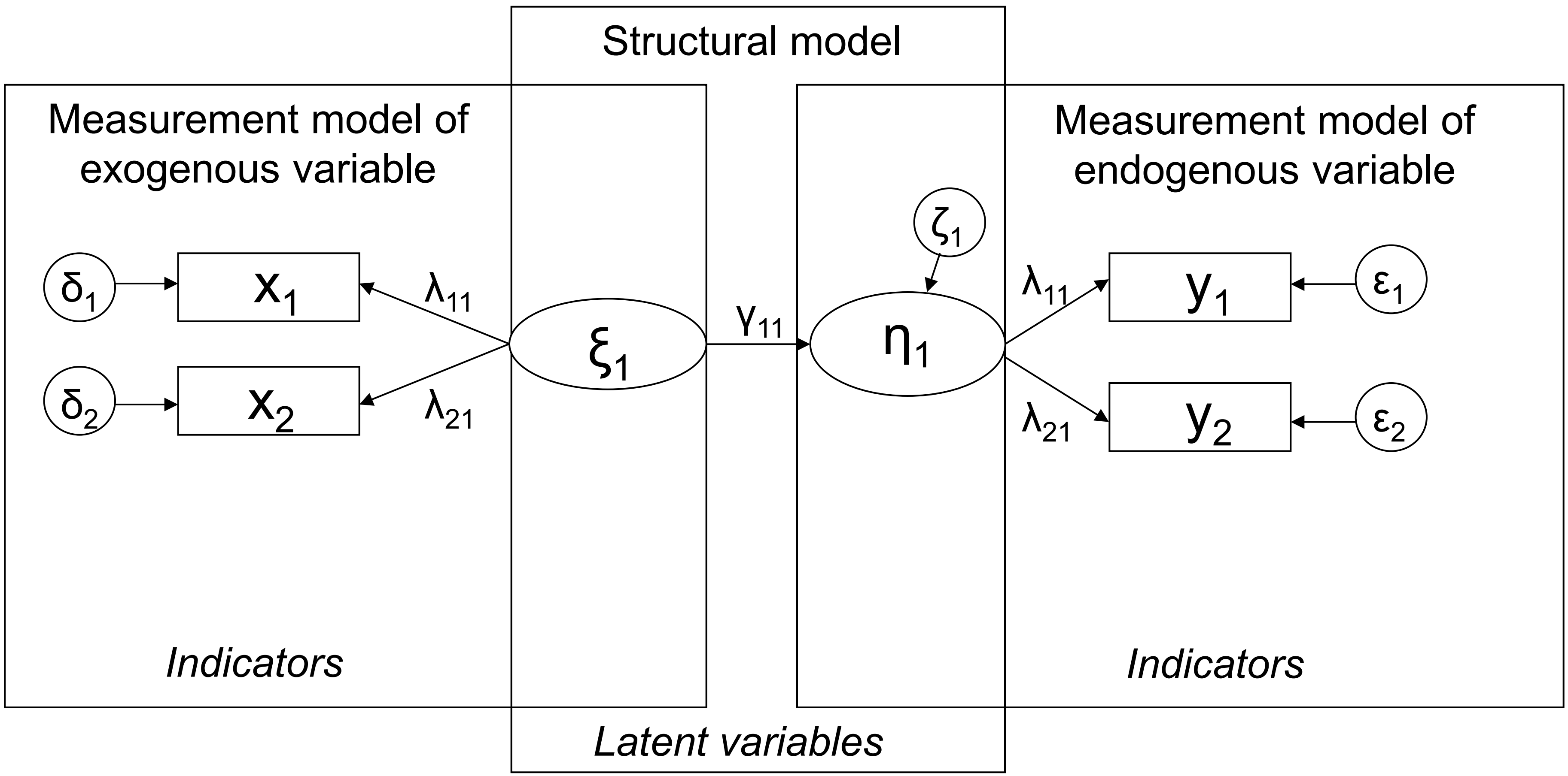
- 1. Introduction
- 2. Correlation
- 3. Partial correlation
- 4. Multiple Regression



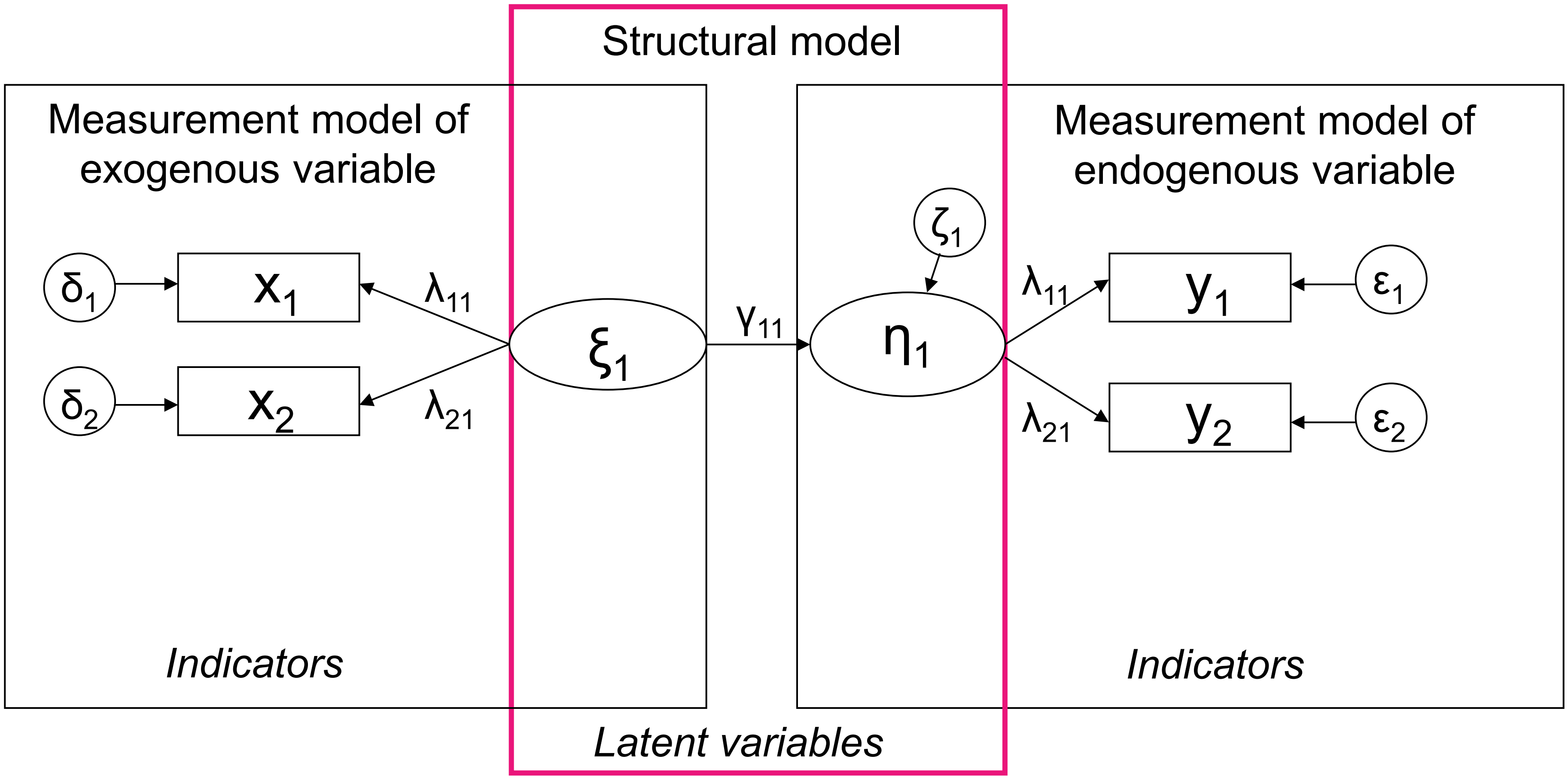
1. Introduction

- This course unit prepares for the modeling of the relationships between latent variables at the latent level.

1. Introduction



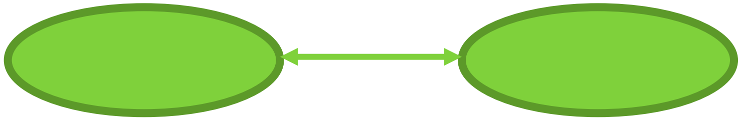
1. Introduction



2. Correlation

2. The correlation

- The correlation serves the description of relationships between random variables



  - It provides information on the degree of the relationship
  - It provides **no** information on the direction of the relationship (same direction vs contrary)

2. The correlation

- There are different types of correlation:

	Data scales:
Pearson correlation / PM correlation	interval – interval
tetrachoric correlation	binary – binary
polychoric correlation	ordered cat. – ord. cat.
biserial correlation	binary – ordinal
Phi coefficient	dichotomous – dich.
rank correlation	ordinal – ordinal
point-biserial correlation	binary - interval

2. The correlation

- There are different types of correlation:

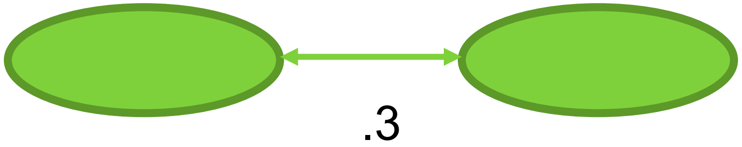
  … of importance:

  - <u>Pearson correlation / PM correlation</u>
  - <u>tetrachoric correlation</u>
  - <u>polychoric correlation</u>

  - biserial correlation
  - Phi coefficient
  - rank correlation
  - point-biserial correlation

2. The correlation

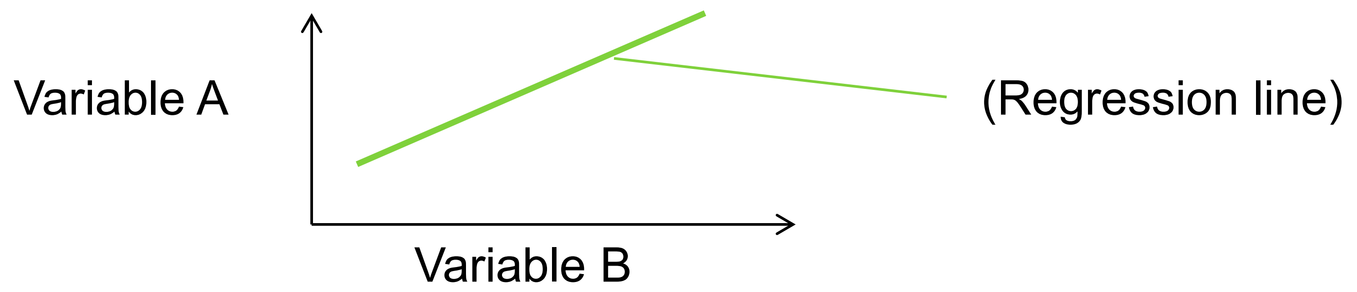
- An example: the correlation of „logical thinking“ and „creativity“:



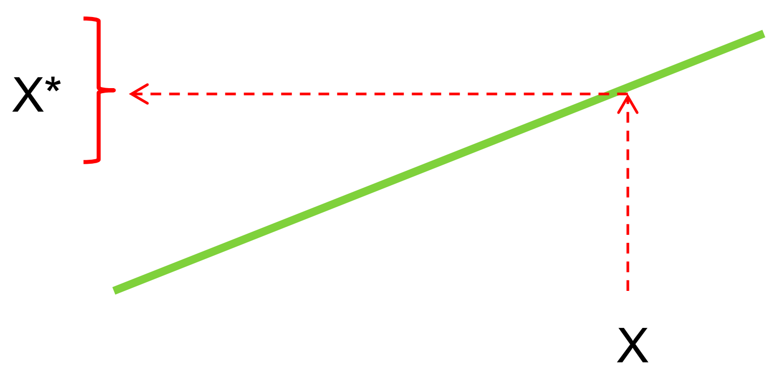
  - There is information on the size of the correlation but not on the direction of an influence

2. The correlation

- Note. Correlation coefficients typically assume that there is a linear relationship

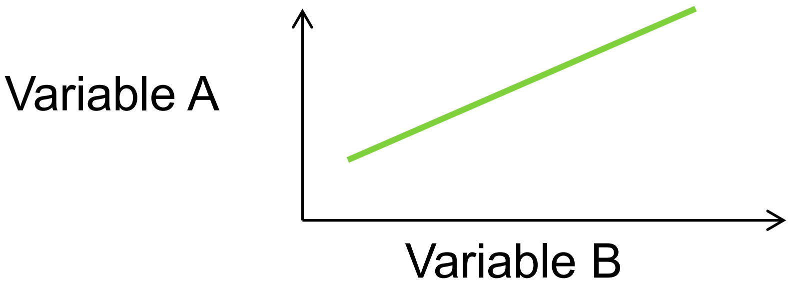


… in a model it can enables basic prediction:

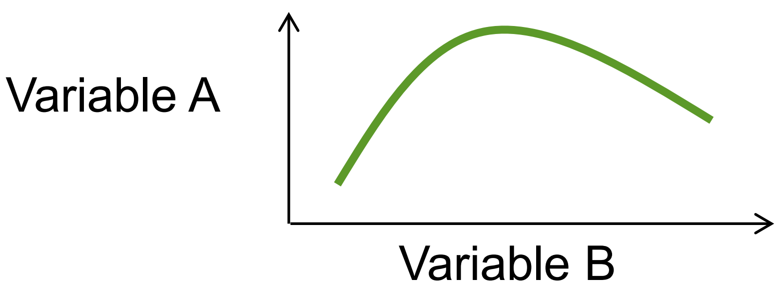


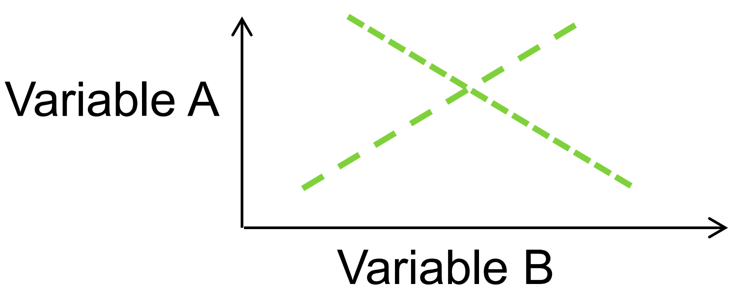
2. The correlation

- Note. Correlation coefficients typically assume that there is a linear relationship



  - But there may be a non-linear relationship or different relationships in sub-groups

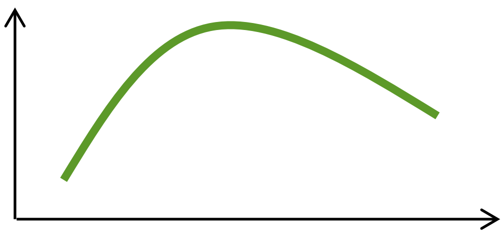


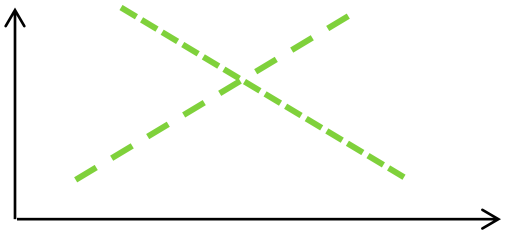


2. The correlation

- Note. Correlation coefficients typically assume that there is a linear relationship
  - But there may be a non-linear relationship or different relationships in sub-groups

… such cases should be avoided; search for reasons.



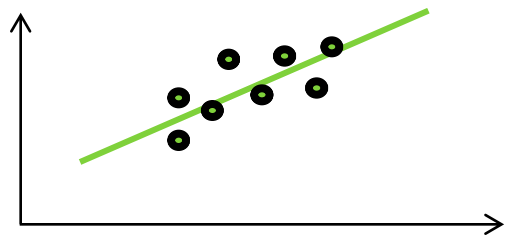


2. The correlation

- Note. Correlation coefficients typically assume that there is a linear relationship

How to find out? Use the scatterplot

Variable A

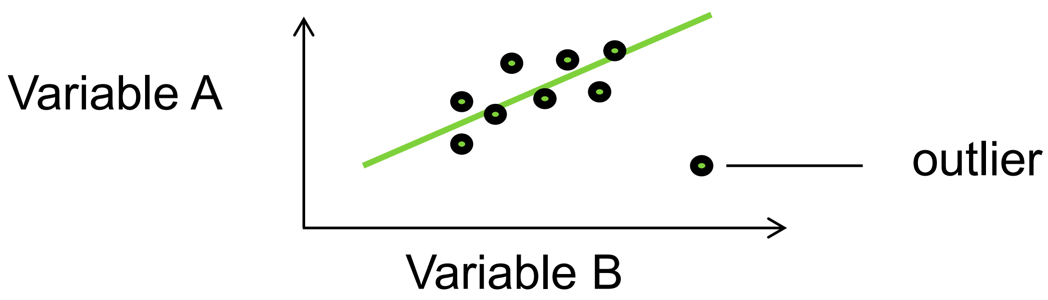


Variable B

2. The correlation

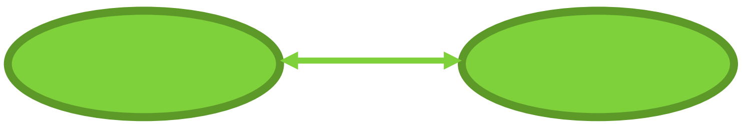
- Note. Correlation coefficients typically assume that there is a linear relationship

It also helps to detect outliers!



2. The correlation

- The correlation is important because it can provide the basis for investigating hypotheses

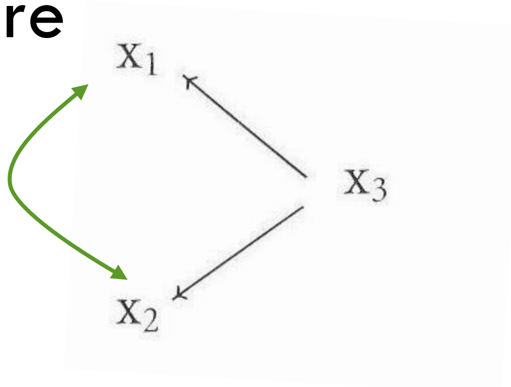


  - … it can be assumed to include effects
  - … it can provide the basis for estimating effects

3. Partial correlation

3. Partial correlation: aspects

- *Idea*: an observed correlation may not reflect the assumed structure



- Possibility: x1 and x2 are influenced by x3 // x3 brings about the correlation between x1 and x2
- A high correlation between x1 and x2 may be a **fake correlation**

3. Partial correlation: aspects

- **„fake correlation“** = correlation between two variables that is actually due to another variable

  - **an example** from southern Sweden

    … allegedly the return of storks from their Winter journey coincides with an increase of house burns

    (… *are storks responsible for the house burns?*)

    … the presumably true reason is **spring** with all kinds of spring activities including festivities!

3. Partial correlation: aspects

- „**fake correlation**“ = correlation between two variables that is actually due to another variable
- Basic idea of partial correlation:
  - There is a correlation between x_1 and x_2 that is due to x_3
  - Variables x_1 and x_2 must be „cleaned“ from the influence of x_3 → the influence of x_3 must be eliminated
  - Correlation „cleaned“ of influence of x_3 on x_1 and x_2 → **partial correlation**

3. Partial correlation: definition

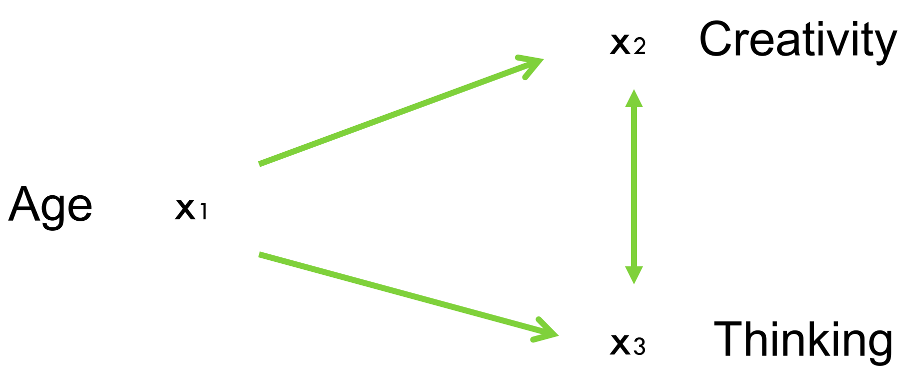
- Formula:

$$r_{x_1x_2 \cdot x_3} = \frac{r_{x_1,x_2} - r_{x_1,x_3} * r_{x_2,x_3}}{\sqrt{1 - r^2_{x1,x3}} * \sqrt{1 - r^2_{x2,x3}}}$$

- computation requires 3 *product-moment correlations* between x_1, x_2 and x_3
- partial correlation is zero? → Indication that the correlation is actually a fake correlation

3. Partial correlation: example

- Example: 2 traits (x_2, x_3) plus age (x_1) are correlated (r_{12}=.8; r_{13}=.7; r_{23}=.6)



- Research question: is the relationship of creativity and thinking a „fake correlation“?

3. Partial correlation: example

- Preparation for the application of the formula:
  - re-assignment of subscripts

age – creativity (r_{12}) → [r_{23}]

age – thinking (r_{13}) → [r_{13}]

creativity – thinking (r_{23}) → [r_{12}]

What it should be!

3. Partial correlation: example

- **Result:**

(using original assignment)

$$r_{23\cdot 1} = \frac{r_{23} - r_{12} \cdot r_{13}}{\sqrt{1-r^2_{12}} \cdot \sqrt{1-r^2_{13}}} = \frac{0.6 - 0.8 \times 0.7}{\sqrt{1-0.8^2} \times \sqrt{1-0.7^2}} = 0.09$$

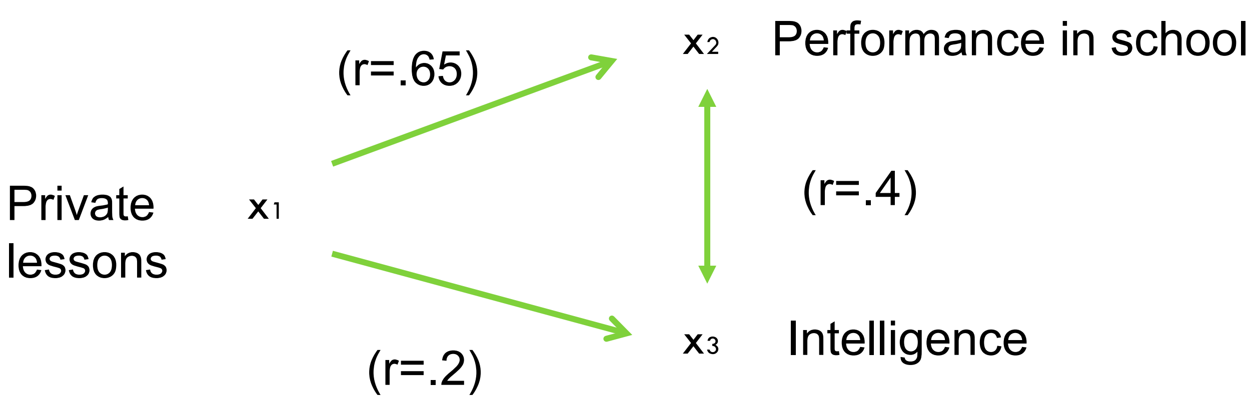
- Is the influence of the variable *age* eliminated from the variables creativity and thinking by partialling out, the remaining partial correlation is virtually zero!
- In sum: the relationship of x2 and x3 is almost completely due to the influence of x1

3. Partial correlation: *practice*

- Find out whether the positive correlation between performance in school and intelligence (.4) is due private lessens

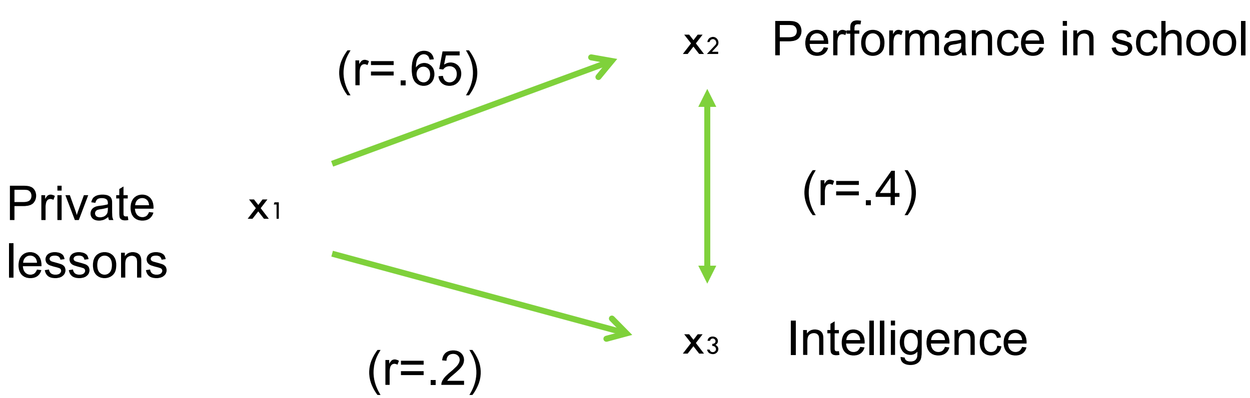
3. Partial correlation: *practice*

- Find out whether the positive correlation between performance in school and intelligence (.4) is due private lessens



3. Partial correlation: *practice*

Is the correlation of „Performance in school“ and „Intelligence“ a fake correlation?



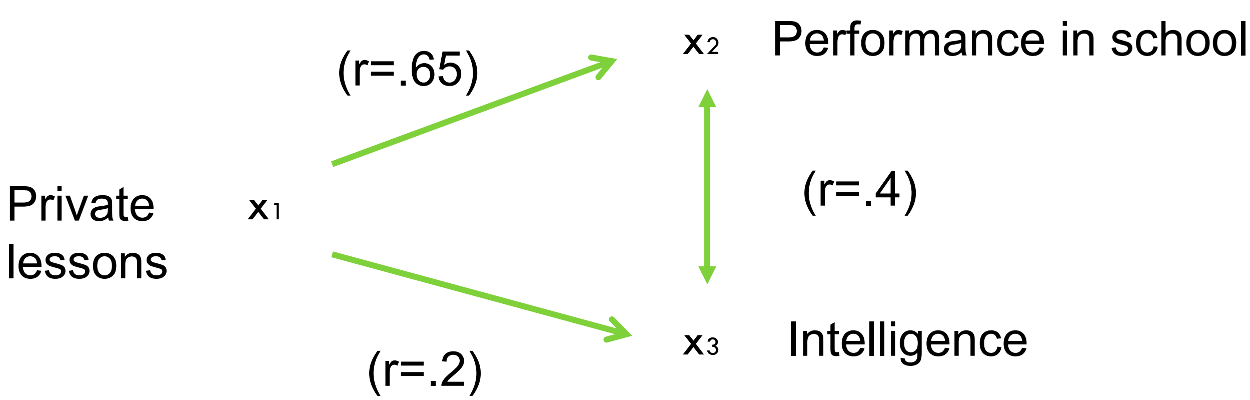
$$\boxed{r_{x_1x_2\cdot x_3} = \frac{r_{x_1,x_2} - r_{x_1,x_3} * r_{x_2,x_3}}{\sqrt{1-r^2_{x_1,x_3}} * \sqrt{1-r^2_{x_2,x_3}}}}$$

$$r_{23\cdot 1} = \frac{r_{23} - r_{12} \cdot r_{13}}{\sqrt{1-r^2_{12}} \cdot \sqrt{1-r^2_{13}}}$$

(using original assignment)

3. Partial correlation: *practice*

- Find out whether the positive correlation between performance in school and intelligence (.4) is due private lessens



Possible Results:

.09 / .21 / .28 / .36 / .41

3. Partial correlation

A comment: the partial correlation

- … made apparent – probably for the first time – that a correlation can originate from different sources

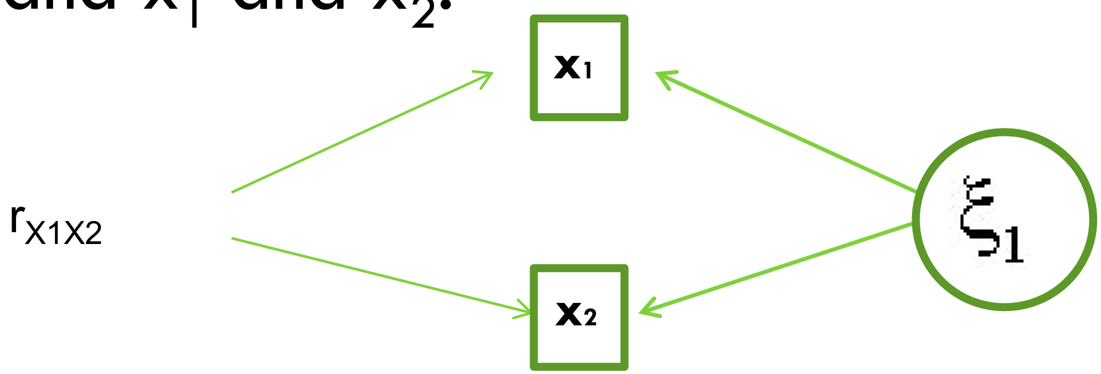
- … is a coefficient for evaluating a hypothesis

  (e.g., the hypothesis is that an external variable is the true source ....)

3. Partial correlation

- The idea of the partial correlation is related to the idea of the *models of measurement* with one latent variable:

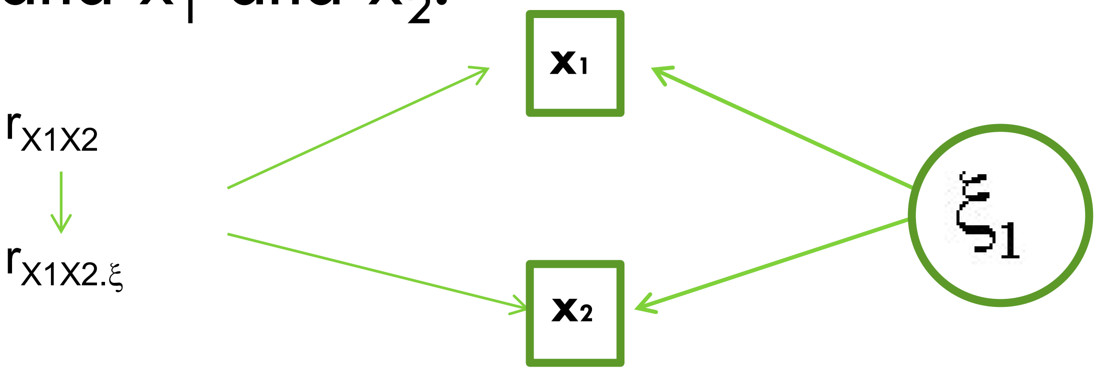
  … it is variable ξ_1 that establishes the relation of and x_1 and x_2:



3. Partial correlation

- The idea of the partial correlation is related to the idea of the *models of measurement* with one latent variable:

… it is variable ξ_1 that establishes the relation of and x_1 and x_2:



4. Multiple regression

4. Multiple regression: introduction

- multiple regression = prediction of a **criterion variable** (*y*) using a linear equation model including **several predictors** (*x*)

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \varepsilon$$

criterion predictor error component

4. Multiple regression: introduction

- multiple regression = prediction of a **criterion variable** (*y*) using a linear equation model including **several predictors** (*x*)

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \varepsilon$$

criterion   predictor   regression weight   error component

4. Multiple regression: introduction

- multiple regression = prediction of a **criterion variable** (*y*) using a linear equation model including **several predictors**

- one aim of multiple regression: <u>estimation of **β weights** (or b)</u>

- β (instead of b) weights are also addressed as **standard partial regression coefficients**

4. Multiple regression: introduction

- … serves the estimation of the effects of **several** predictors on the criterion
- … estimation of β respectively **b** is conducted according to the ordinary least-square (OLS) criterion:

$$\min = \varepsilon'\varepsilon$$

… the estimation method （OLS）：

$$\mathbf{b} = (\mathbf{X'X})^{-1}\ \mathbf{X'y}$$

4. Multiple regression: introduction

- An example regarding OLS

$$\min = \varepsilon'\varepsilon$$

... the estimation method:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{y}$$

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 2 \\ 5 \\ 6 \end{pmatrix}$$

4. Multiple regression: introduction

- An example regarding OLS

$$\min = \varepsilon'\varepsilon$$

… the estimation method for one predictor:

$$\mathbf{b} = (\mathbf{x}'\mathbf{x})^{-1}\,\mathbf{x}'\mathbf{y}$$

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 2 \\ 5 \\ 6 \end{pmatrix}$$

$\mathbf{x}'\mathbf{x} = 21$ $\quad (\mathbf{x}'\mathbf{x})^{-1} = 1/21$ $\quad \mathbf{x}'\mathbf{y} = 36$

$b = (1/21) \times 36 = 1.71$

4. Multiple regression: introduction

- ... serves the estimation of the effects of **several** predictors on the criterion

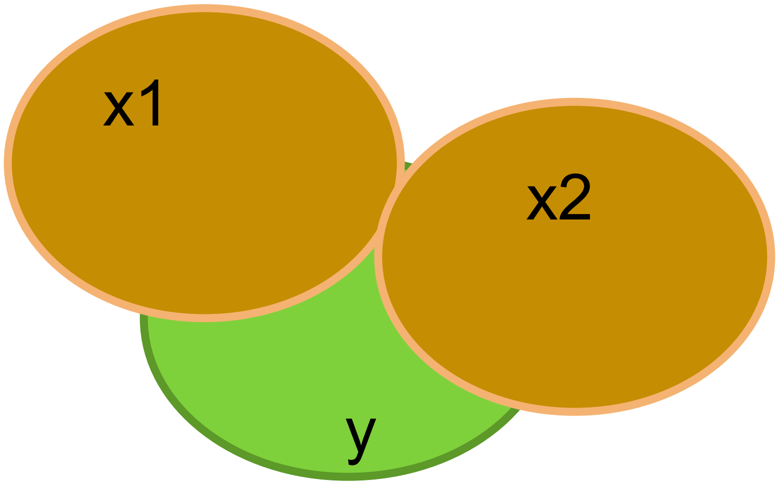
Problem:

- ... the method must takes the correlations among the predictors into consideration because ...
- ... **overlapping predictions** are a *major problem*:

4. Multiple regression: introduction

- … two predictor variables (x1, x2) predict a criterion variable (y)

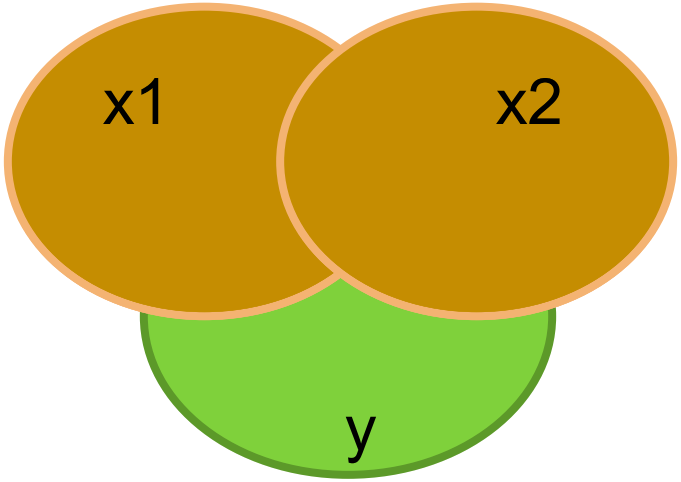
… no overlap!



The ellipses represent variance of variables

4. Multiple regression: introduction

- ... two predictor variables (x1, x2) predict a criterion variable (y)

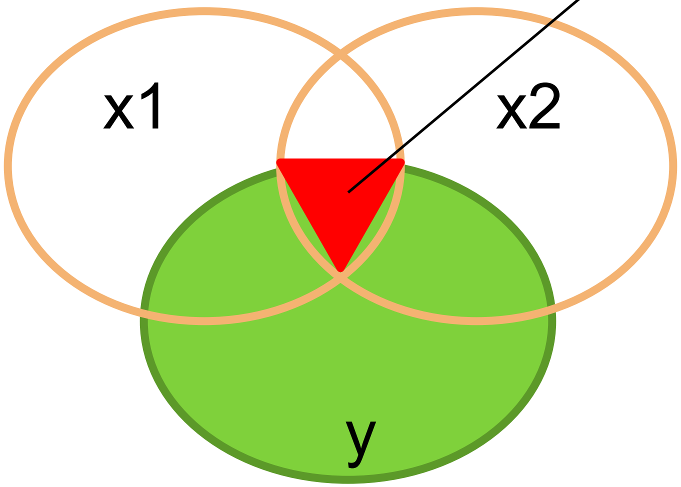


The ellipses represent variance of variables

4. Multiple regression: introduction

- … two predictor variables (x1, x2) predict a criterion variable (y)

Area predicted by both predictors



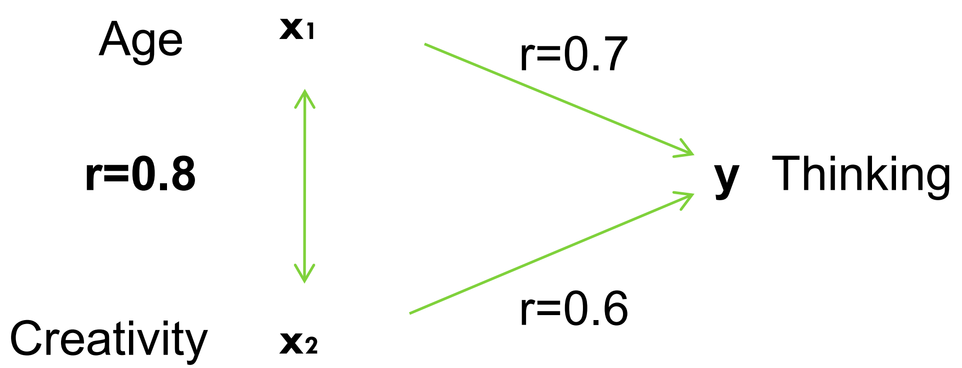
Formula for the standard partial regression coefficient takes the overlap into consideration

4. Multiple regression: example

- an example: *overlapping predictors*

Predictors

Criterion



$r_{x1,x2} = .8$

$r_{x1,y} = .7$

$r_{x2,y} = .6$

Question: do the correlations reflect the influences of the predictors correctly?

4. The alternative: partial regression

There is an alternative way of estimating regressions weights

… if correlations are available

.. the partial regression coefficient

4. Partial regression

The formula for the computation of the standard partial regression coefficient β is ...

$$\beta_{Y1,X1\cdot X2} = \frac{r_{Y1,X1} - r_{Y1,X2} \cdot r_{X1,X2}}{1 - r^2_{X1,X2}}$$

Criterion variable

Predictor 1

Predictor 2 (= additionally considered predictor)

4. Partial regression

The formula for the computation of the standard partial regression coefficient β is ...

$$\beta_{Y1,X1\cdot X2} = \frac{r_{Y1,X1} - r_{Y1,X2} \cdot r_{X1,X2}}{1 - r^2_{X1,X2}}$$

… compare denominator of partial correlation:

$$\sqrt{1 - r^2_{X1X3}}\sqrt{1 - r^2_{X2X3}}$$

4. Partial regression: example

Formulas:

Criterion: thinking
Predictor: age

$$\beta_{Y1,X1 \cdot X2} = \frac{r_{Y1,X1} - r_{Y1,X2} \cdot r_{X1,X2}}{1 - r^2_{X1,X2}}$$

Criterion: thinking
Predictor: creativity

$$\beta_{Y1,X2 \cdot X1} = \frac{r_{Y1,X2} - r_{Y1,X1} \cdot r_{X1,X2}}{1 - r^2_{X1,X2}}$$

4. Partial regression: example

- Equations:

$$\beta_{Y1,X1\cdot X2} = \frac{r_{Y1,X1} - r_{Y1,X2} \cdot r_{X1,X2}}{1 - r^2_{X1,X2}}$$

Predictor: age

$$\beta_{Y1,X2\cdot X1} = \frac{r_{Y1,X2} - r_{Y1,X1} \cdot r_{X1,X2}}{1 - r^2_{X1,X2}}$$

Predictor: creativity

- Results:

$$\beta_{Y1,X1\cdot X2} = \frac{0.70 - 0.60 \cdot 0.80}{1 - 0.80^2} = \frac{0.22}{0.36} = 0.61$$

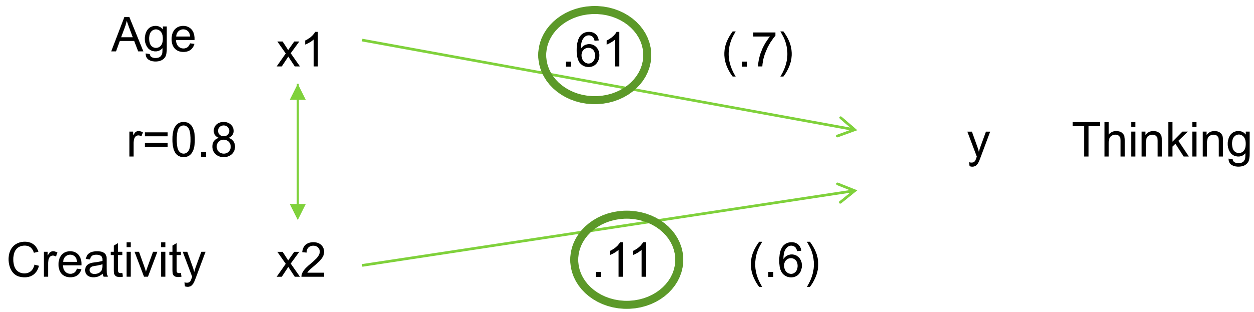
Age $\rightarrow$ Thinking

$$\beta_{Y1,X2\cdot X1} = \frac{0.60 - 0.70 \cdot 0.80}{1 - 0.80^2} = \frac{0.04}{0.36} = 0.11$$

Creativity $\rightarrow$ Thinking

4. Partial regression: example

- partial regression coefficients:



- the criterion *thinking* is influenced to a higher degree by *age* than by *creativity*

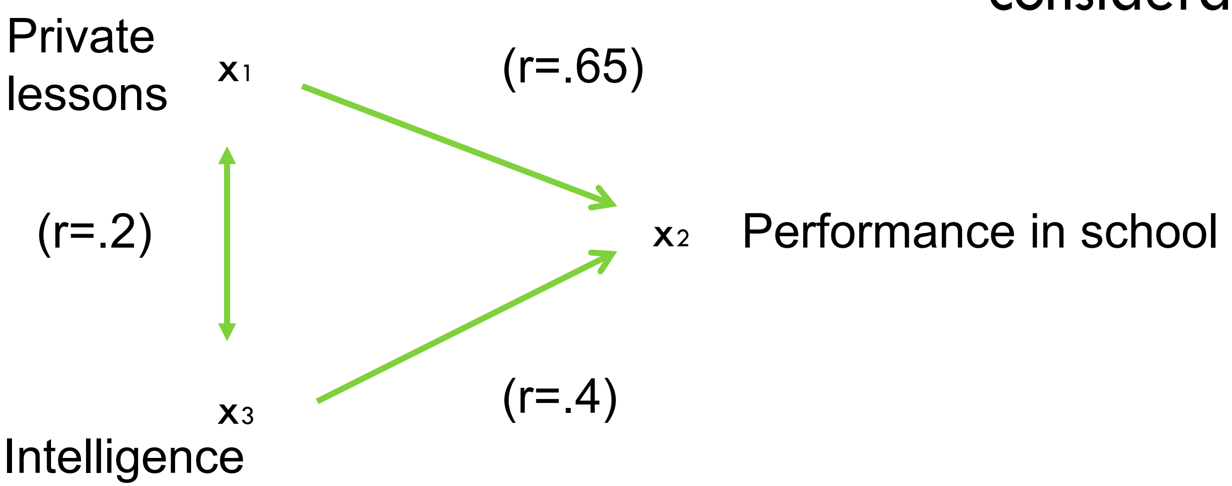
- the correlation of creativity & thinking: it appears to be to a high degree due to the correlation of the predictors that means due to an ***indirect effect***.

4. Partial regression: *practice*

- Find out how strong the effect of private lessons is on performance in school when taking *intelligence* into consideration!

4. Partial regression: *practice*

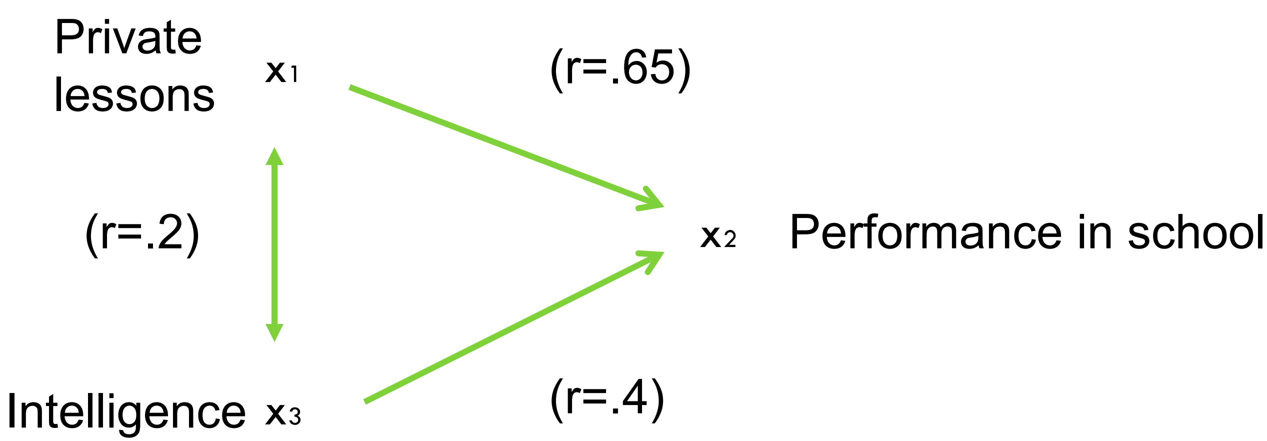
- Find out how strong the effect of <u>private lessons is</u> on <u>performance</u> in school when taking *intelligence* into consideration!



$$\beta_{Y1,X1\cdot X2} = \frac{r_{Y1,X1} - r_{Y1,X2} \cdot r_{X1,X2}}{1 - r_{X1,X2}^2}$$

4. Partial regression: *practice*

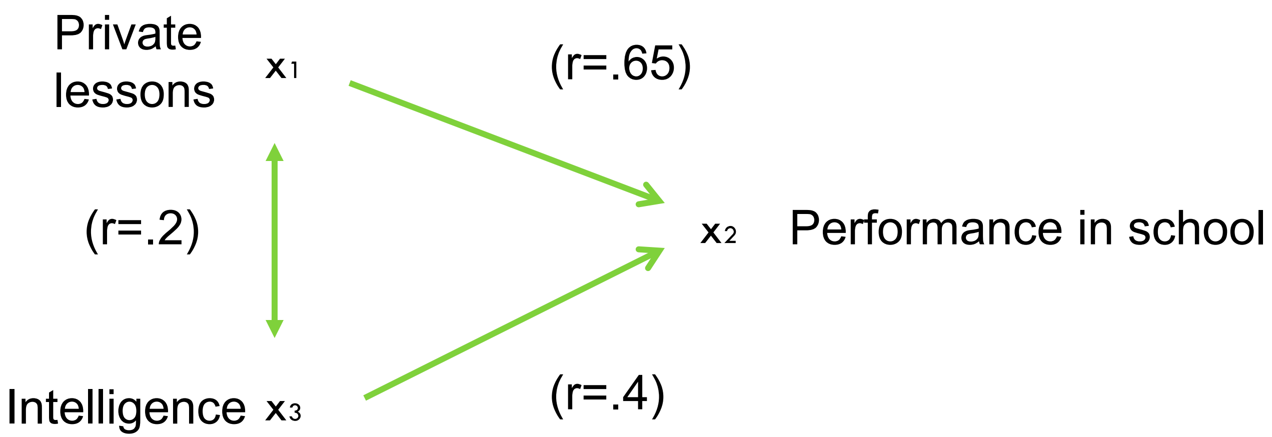
- Find out how strong ….



Possible results: $\beta_{\text{Private lessons}}$ = .51 / .59 / .61 / .64

4. Partial regression: *practice*

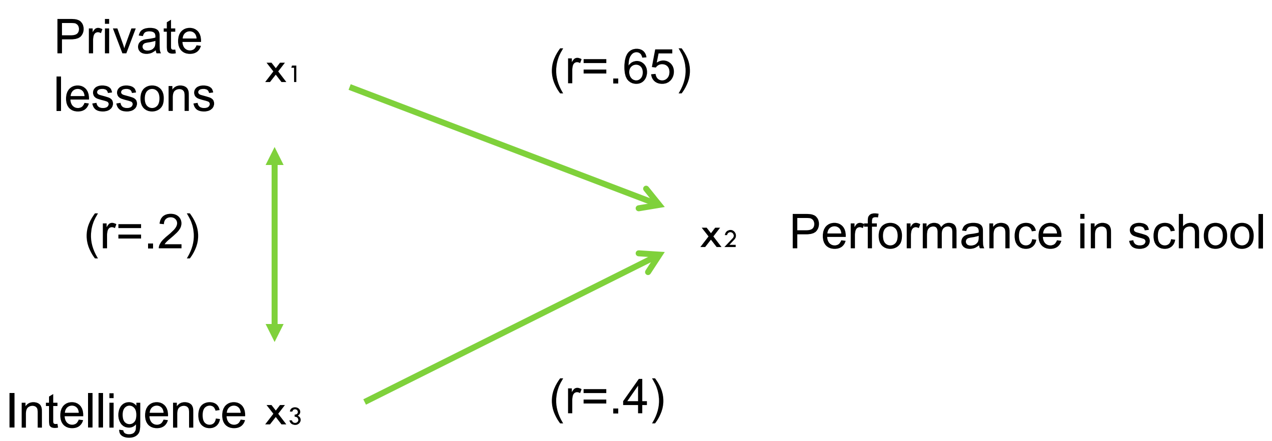
- Find out how strong the effect of <u>intelligence</u> on <u>performance</u> in school is in taking *private lessons* into consideration!



$$\beta_{Y1,X1 \cdot X2} = \frac{r_{Y1,X1} - r_{Y1,X2} \cdot r_{X1,X2}}{1 - r^2_{X1,X2}}$$

4. Partial regression: *practice*

- Find out …



Possible results: $\beta_{\text{Intellgience}}$ = .00 / .13 / .19 / .28

4. Partial regression: a note

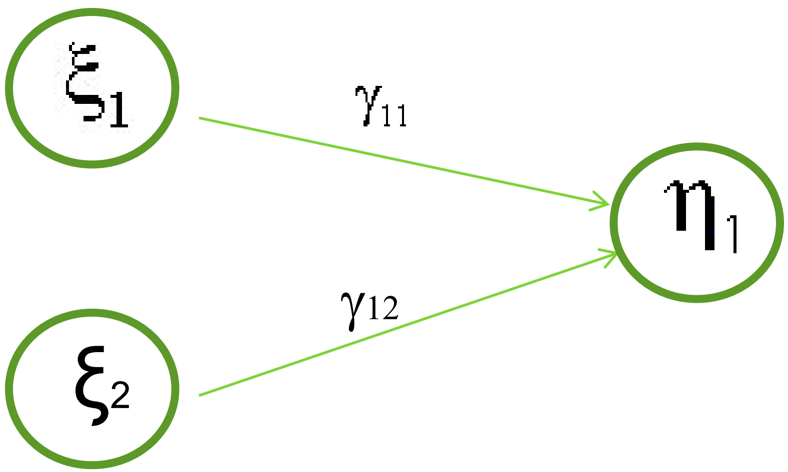
- in multiple regression the regression weights are computed on the basis of *raw data* and the computation is conducted according to another principle.

4. Partial regression: a note

- in multiple regression the regression weights are computed on the basis of *raw data* and the computation is conducted according to another principle.
- There are two way of determing regression coefficients:
  - computation using formulas
  - estimation using estimation methods
    - ordinary least square estimation （OLS）
    - maximum likelihood estimation （ML）

4. Multiple regression: *an analogy*

- Given the ***correlations*** between the latent variables ξ_1, ξ_2 and η, it is possible to compute the standardized partial regression coefficient for the following structural model:

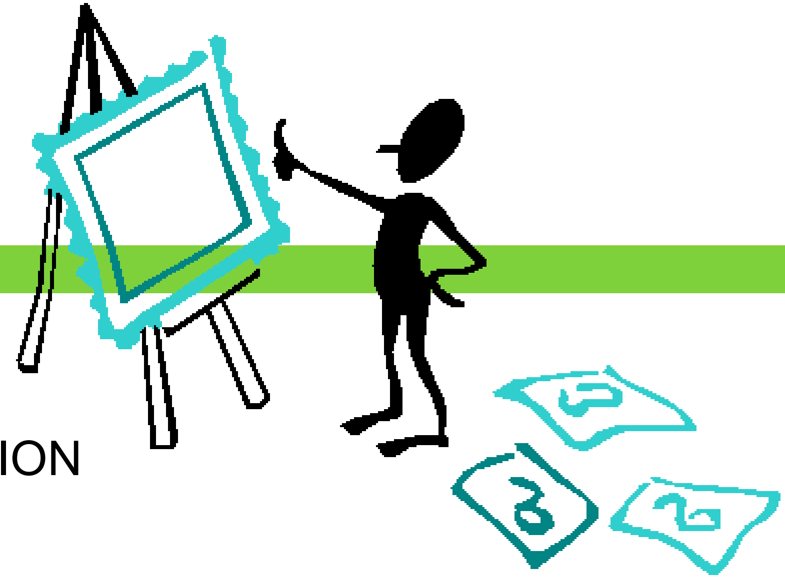


  - … with γ_{11} as β_1 and γ_{12} as β_2

4. Multiple regression: in sum

- multiple regression can be used for investigating effects of *several predictors* on a criterion
- under *appropriate preconditions* the standard partial regression coefficient can be used for estimating the influences of predictors in *the structural model*
- it can be conducted **on the basis of correlations**

Summary and brush up:



CORRELATION

PARTIAL CORRELATION

MULTIPLE REGRESSION

- 1. ...... **Concern with structure**
- 2. Correlation ... measure of the closeness of the relationship / but there may be a linearity problem
- 3. Partial correlation ... signifies what remains after removing the effect of another variable from the correlation
- 4. Multiple Regression ... prediction of criterion variable by means of several predictor variables

QUESTIONS REGARDING COURSE UNIT 7

- What is a fake correlation?
- What is a standardized partial regression coefficient?
- What is an indirect effect?
- Which correlations are suitable for CFA, SEM?

Literature

- Kline, R. B. (2011). *Principles and practices of structural equation modeling* (3rd edition) (Chapter 2: Basic statistical concepts: correlations and regression, Chapter 5: Introduction to path analysis). New York, NJ: The Guilford Press.
- Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (Partial regression, partial correlation, direct and indirect effects: S. 64-79). Mahwah, N.J.: Lawerence Erlbaum.