

Introduction to STATA

Week 1 – Revision of basic statistical concepts



Teachers

- Tomáš Katrňák
 - katrnak@fss.muni.cz
 - Office hours: Wednesday 11.15-12.30
- Tomáš Doseděl
 - dotomas@mail.muni.cz
 - Office hours: upon e-mail request

Course outline

- How to deal with STATA
- Variable transformation
- Descriptive data analysis
- Multivariate data analysis

How to succeed

- Submit a short paper: till the end of semester
 - Project proposal, hypotheses: till the end of 3rd week
 - Descriptive analysis: till the end of 7th week
 - Multivariation analysis: till the end of 10th week
- Final paper presentation
- Active participation

Assessment

- Submit a short paper: up to 23 points
 - Project proposal, hypotheses: up to 10 point
 - Descriptive analysis: up to 10 points
 - Multivariation analysis: up to 10 points
- Final paper presentation: up to 23 points
- Active participation: up to 24 points (2 pts/week)

90-100 points: A | 80-89 points: B | 70-79 points: C

60-69 points: D | 50-59 points: E | 0-49 points: F

Revision of basic statistical concepts

Variables

- What is a variable?

Variables

- What is a variable?

Age

Gender

Height

Income

„an abstraction of any possible object of the given class“

Variables

- What is a variable?
- Variable types

Variables

- What is a variable ?
- Variable types
 - nominal

Variables

- What is a variable ?
- Variable types

- nominal

name color gender occupation

„nominate different attributes without the possibility to order them “

Variables

- What is a variable ?
- Variable types
 - Nominal
 - Ordinal

Variables

- What is a variable ?
- Variable types

- Nominal
- Ordinal

Education level

rank in a queue

„we can order the values but are unable to decide their distance“

Variables

- What is a variable ?
- Variable types
 - Nominal
 - Ordinal
 - Interval / Continuous

Variables

- What is a variable ?
- Variable types
 - Nominal
 - Ordinal
 - Interval / Continuous

Height Income Years spent in education

„we can both order the values and decide about their distance “

Data matrix

Data matrix

Id	Gender	Age	Education	...
1	Male	19	Tertiary	
2	Female	27	Secondary	
3	Male	17	Primary	
4	Male	23	Tertiary	

Matrice dat

Id	Gender	Age	Education	...
1	Male	19	Tertiary	
2	Female	27	Secondary	
3	Male	17	Primary	
4	Male	23	Tertiary	

Id	Gender	Age	Education	...
1	1	19	3	
2	2	27	2	
3	1	17	1	
4	1	23	3	

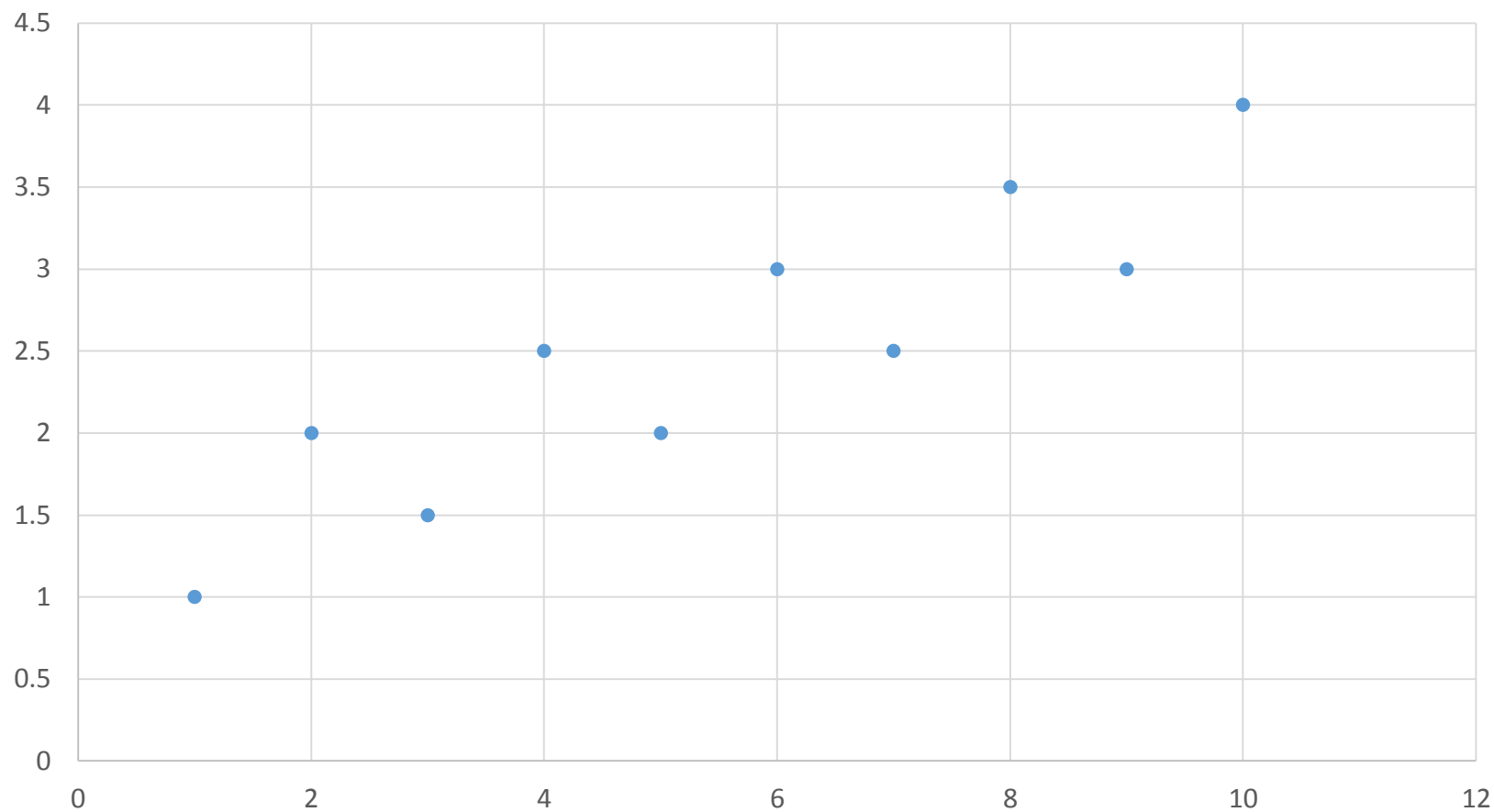
Correlation

Correlation

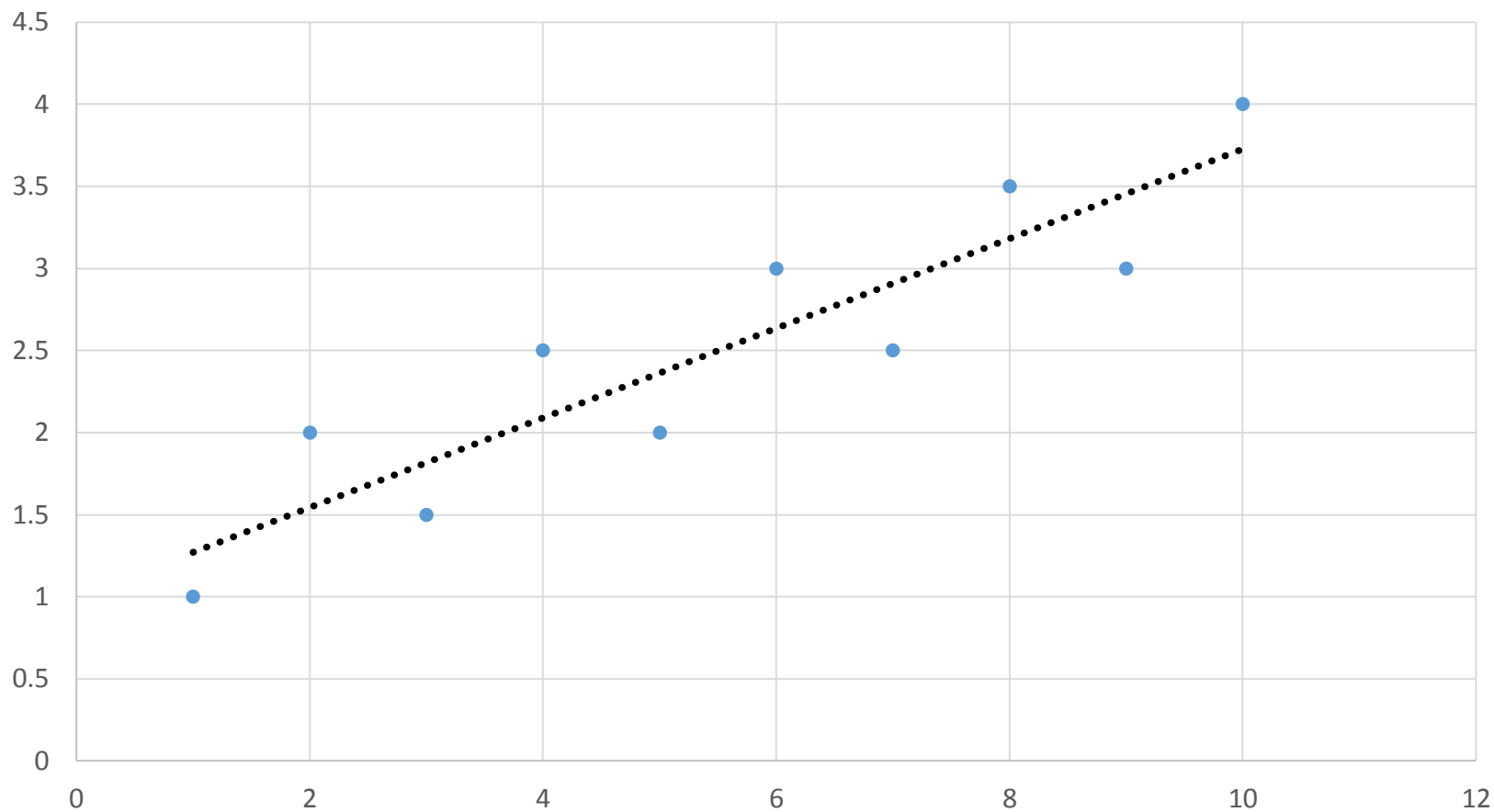
- The strength of a relation between two variables
- Both variables are in a relation (co-relate), if one changes, change the other
- Doesn't necessary mean causality!

Linear regression (OLS)

Linear regression (OLS)



Linear regression (OLS)



Linear regression (OLS)

$$y = a + bx$$

a – intercept

b – slope

Linear regression (OLS)

Ordinary least square – tries to find a solution for which the sum of the squares is minimal. Squares are defined by the distance of respective point from the line.

