

Úvod do programu STATA

Týden 1 – Zopakování základních statistických pojmů



Vyučující

- Tomáš Katrňák
 - katrnak@fss.muni.cz
 - konzultační hodiny: středa 11.15-12.30
- Tomáš Doseděl
 - dotomas@mail.muni.cz
 - konzultační hodiny: po domluvě e-mailem kdykoliv

Studijní povinnosti

- Odevzdání seminární práce: do konce zkouškového
 - Návrh tématu, hypotézy, data: do 3. týdne
 - Deskriptivní analýza: do 8. týdne
 - Multivariační analýza: do 11. týdne
- Veřejná prezentace seminární práce: 13. týden
- Docházka

Hodnocení

- Odevzdání seminární práce: 26 bodů
 - Návrh tématu, hypotézy, data: 10 bodů
 - Deskriptivní analýza: 10 bodů
 - Multivariační analýza: 10 bodů
- Veřejná prezentace seminární práce: 20 bodů
- Docházka: 24 bodů

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

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

Osnova

- Základy práce v programu STATA
- Transformace proměnných
- Deskriptivní analýza dat
- Multivariační analýza dat
- Pokročilé techniky analýzy dat

Zopakování základních statistických pojmů

Proměnné a jejich typy

- Co je to proměnná?

Proměnné a jejich typy

- Co je to proměnná?

věk pohlaví výška příjem

„zastupuje libovolný myslitelný objekt dané třídy“

Proměnné a jejich typy

- Co je to proměnná?
- Typy proměnných

Proměnné a jejich typy

- Co je to proměnná?
- Typy proměnných
 - nominální

Proměnné a jejich typy

- Co je to proměnná?

- Typy proměnných

- nominální

jméno

barva

pohlaví

povolání

„pojmenovává různé vlastnosti bez možnosti seřazení“

Proměnné a jejich typy

- Co je to proměnná?
- Typy proměnných
 - nominální
 - ordinální

Proměnné a jejich typy

- Co je to proměnná?

- Typy proměnných

 - nominální

 - ordinální

stupeň vzdělání umístění sportovce pořadí ve frontě

„hodnoty můžeme seřadit, nedokážeme ale rozhodnout, jak jsou od sebe vzdálené“

Proměnné a jejich typy

- Co je to proměnná?
- Typy proměnných
 - nominální
 - ordinální
 - kardinální/spojité

Proměnné a jejich typy

- Co je to proměnná?

- Typy proměnných

 - nominální

 - ordinální

 - kardinální/spojité

výška příjem počet let strávených ve škole

„hodnoty můžeme seřadit a současně rozhodnout, jak jsou od sebe vzdálené“

intervalové/poměrové

Matice dat

Matice dat

Id	Pohlaví	Věk	Vzdělání	...
1	Muž	19	VŠ	
2	Žena	27	SŠ	
3	Muž	17	ZŠ	
4	Žena	23	VŠ	

Matice dat

Id	Pohlaví	Věk	Vzdělání	...
1	Muž	19	VŠ	
2	Žena	27	SŠ	
3	Muž	17	ZŠ	
4	Žena	23	VŠ	

Id	Pohlaví	Věk	Vzdělání	...
1	1	19	3	
2	2	27	2	
3	1	17	1	
4	2	23	3	

Korelace

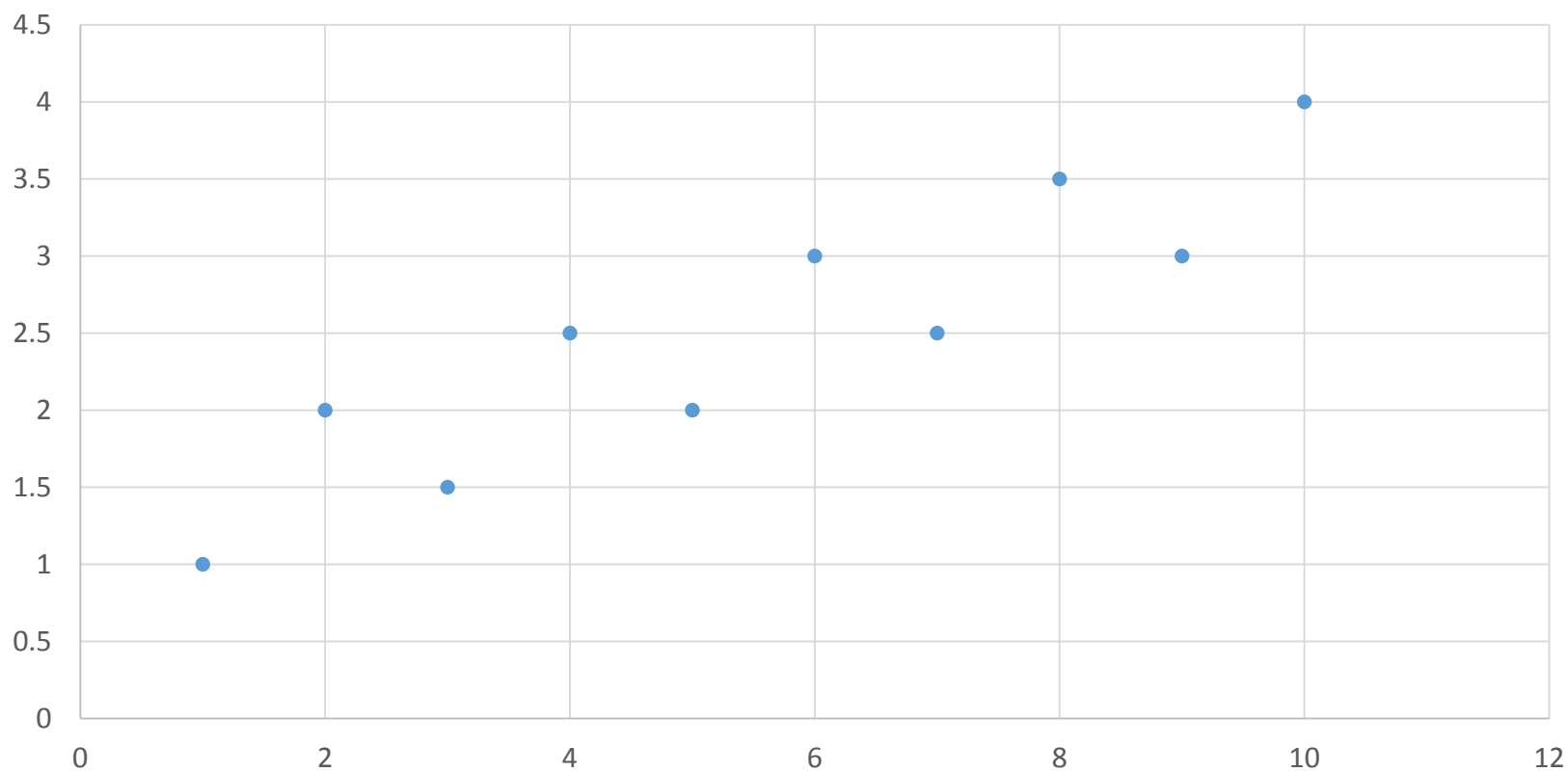
Korelace

- Síla vztahu mezi dvěma proměnnými
- Obě proměnné jsou v relaci (souvisí) + když se změní jedna, změní se i druhá (ko)
- Neznamená kauzalitu

Lineární regrese

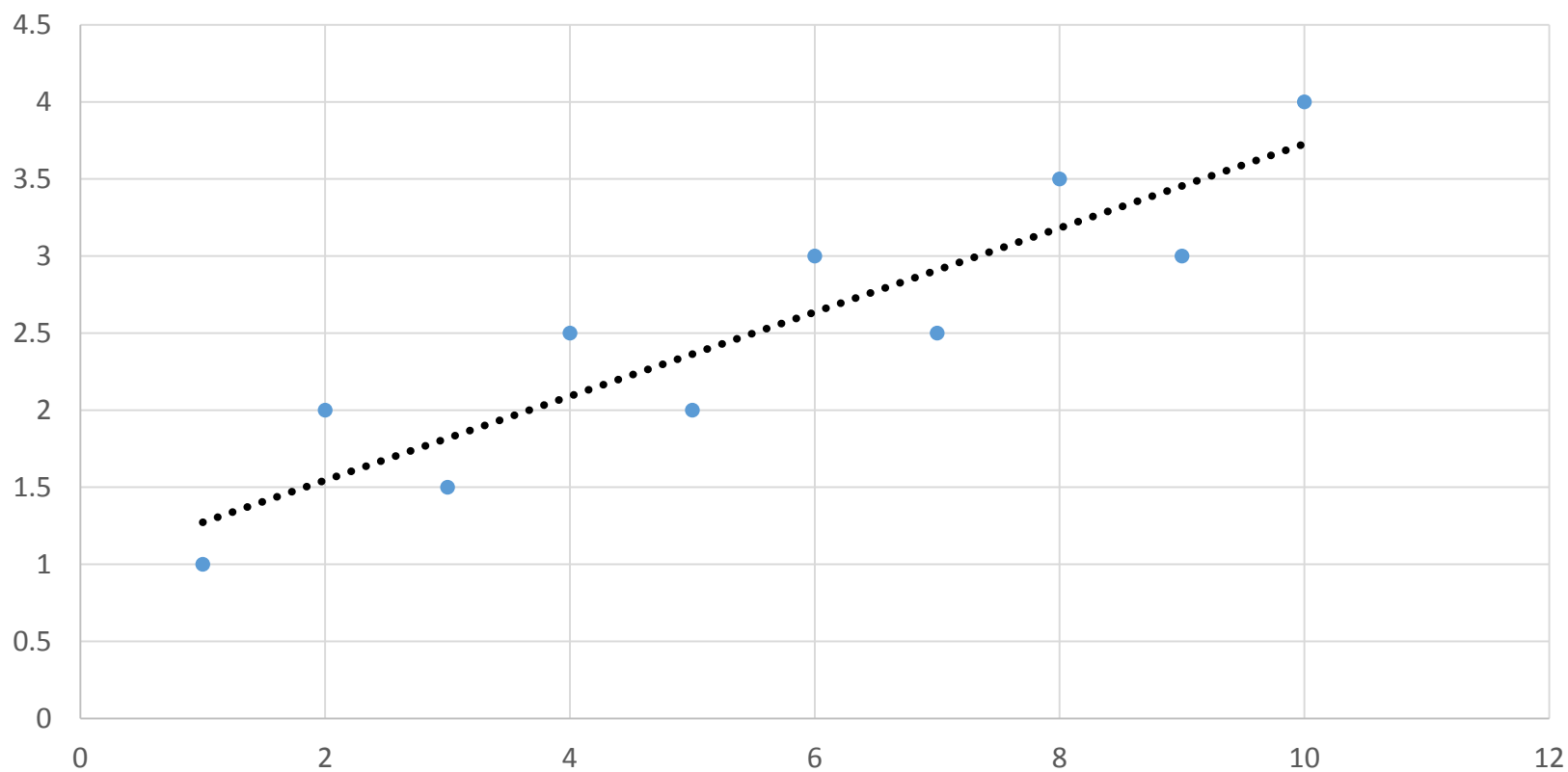
Lineární regrese

Hodnoty osy Y



Lineární regrese

Hodnoty osy Y



Lineární regrese

$$y = a + bx$$

a – průsečík, intercept

b – směrnice, sklon, tangens úhlu s osou x, slope

Lineární regrese

Metoda nejmenších čtverců, ordinary least square (OLS regression) – hledá takovou přímku, pro kterou platí, že součet druhých mocnin (čtverců) vzdáleností všech bodů od přímky je minimální

