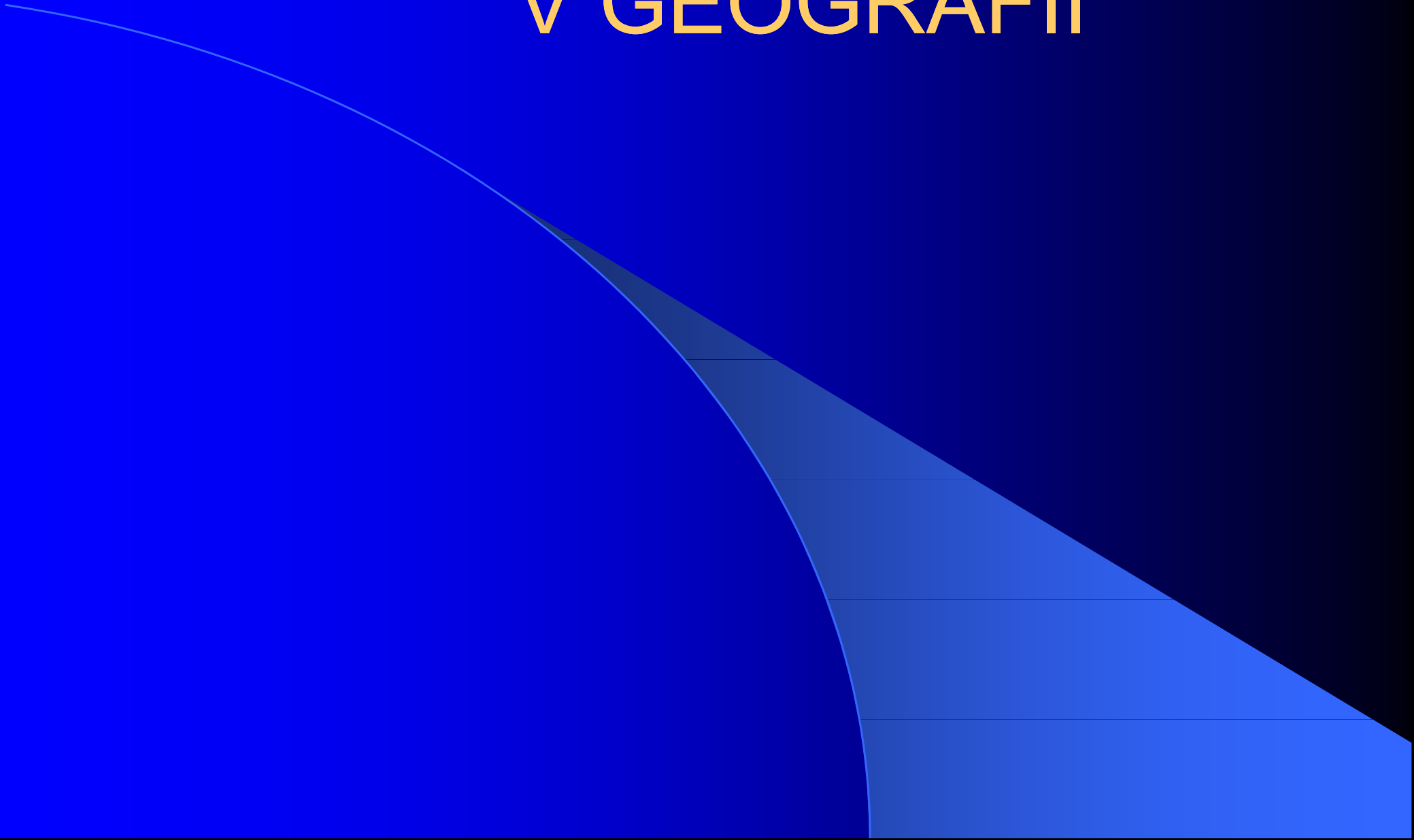


STATISTICKÉ METODY V GEOGRAFII





Odhady parametrů intervaly spolehlivosti

Základní pojmy

- základní soubor,
- statistický soubor
- výběrový soubor
- náhodný výběr
- k základnímu jednomu souboru lze získat více výběrových, různé charakteristiky
- U dobré výběrové metody - dílčí směrodatné odchylky se kompenzují

Základní pojmy

- **reprezentativnost výběru** – kvalita výběru
- **prostý náhodný výběr** (s opakováním a bez opakování)
- **oblastní náhodný výběr** (výběr z každé dílčí části)
- **systematický náhodný výběr** (podle pravidla, které nesouvisí se sledovaným znakem, např. sledovaný znak - počet obyvatel obce, seřadit obce podle abecedy a vybrat vždy každou pátou obec)

Intervaly spolehlivosti

- normální rozdělení,
- interval spolehlivosti hranice ($\mu + - 2\sigma$),
- hodnoty, které leží mimo interval, v tzv. kritickém oboru se považují za nepřipustné, jejich odchylky od průměru za významné
- lze použít i jiné intervaly spolehlivosti
- např. pro 95 % ($\mu + - 1,960\sigma$),
- pro 99 % ($\mu + - 2,576\sigma$),

Testování statistických hypotéz

- jak ověřit předpoklady o charakteristikách statistických souborů?
- Je soubor A výběrem ze souboru B?
- Do jaké míry se soubory shodují v rozdělení četností, podle aritm. Průměru, podle směrodatné odchylky apod.

Příklad

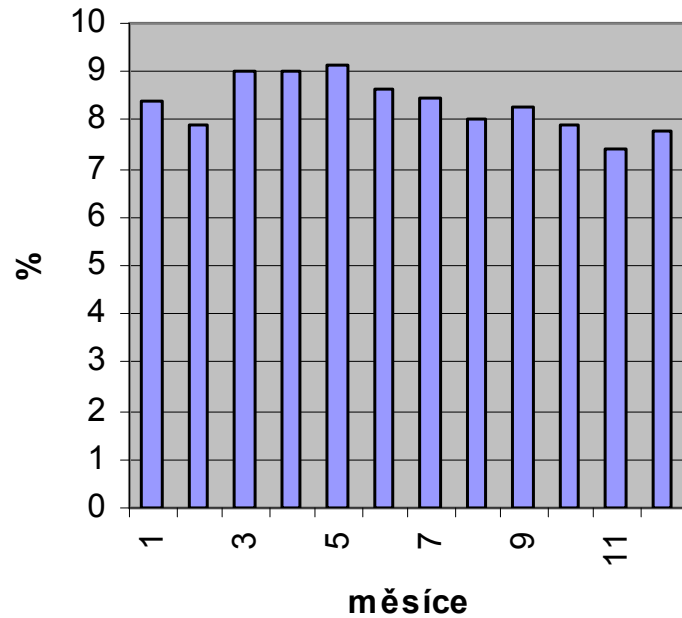
Soubor A

Soubor a

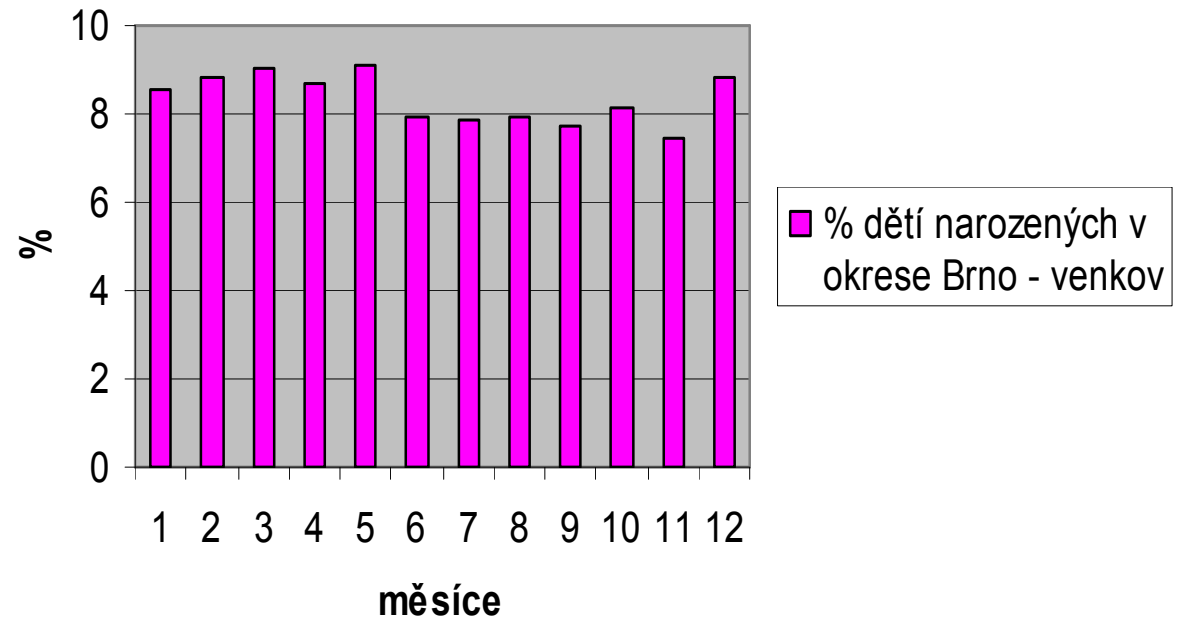
měsíc	% dětí narozených v ČR	% dětí narozených v okrese Brno - venkov
1	8,39	8,52
2	7,91	8,81
3	9,02	9,01
4	9,03	8,72
5	9,15	9,12
6	8,64	7,94
7	8,45	7,84
8	8,04	7,93
9	8,28	7,74
10	7,93	8,13
11	7,41	7,44
12	7,75	8,81

Rozdělení četností souborů A , a

% dětí narozených v ČR



% dětí narozených v okrese Brno - venkov



průměr

8,333333333

směrodatná odchylka

0,529013757

rozptyl

0,279855556

průměr

8,334166667

směrodatná odchylka

0,537563304

rozptyl

0,288974306

- STATISTICKÁ HYPOTÉZA:
- předpoklad: průměrná výška studentek PdF MU je **shodná** s průměrnou výškou žen ve věku 20 - 25 let v ČR
- **NULOVÁ HYPOTÉZA**
- **Průměry obou souborů jsou shodné**
- zvolíme hladinu významnosti
- např. 5% , tj. $p=0,05$, tj. shoda je s pravděpodobností 95 %
- aplikace testovacího kritéria
- je výsledek testování významný ?

Závislost náhodných veličin

Závislost náhodných veličin

- Do jaké míry závisí změna prvku jednoho statistického souboru změnu prvku druhého statistického souboru?
- Jak podmiňuje změna prvku x změnu prvku y ?
- Jak těsně na sobě závisí prvky dvourozměrného statistického souboru?
- Např.
 - vztahy teplota a nadm. výška,
 - srážky a odtok v povodí
 - váha a výška člověka,

Vztahy náhodných veličin

- Jednostranné (nezávislá hodnota x jednoho stat. souboru podmiňuje hodnotu y druhého stat. Souboru)
- Vzájemné (nelze rozlišit závislou a nezávislou proměnou)

Vztahy náhodných veličin

- Podle stupně závislosti
- Funkční (pevnou)
- (určité hodnotě x odpovídá jediná hodnota y , vztah x a y lze tedy vyjádřit mat. funkcí),
- *např.*
- *Konkrétní teplotě odpovídá jedna hodnota stupně nasycení vodní párou*

Vztahy náhodných veličin

- Statistická
- (jedné hodnotě x odpovídá více hodnot y , hodnoty y mají své rozdělení s průměrem, tento průměr hodnot y je i pro různá x shodný)



Vztahy náhodných veličin

- **Korelační**
- Se změnou hodnot x se mění soubory hodnot y , které mají své rozdělení a různých průměrech
- *např. pro určitou těl výšku existuje více hodnot hmotnosti, které budou mít normální rozdělení,*
- *různým výškám odpovídají hmotnosti s normálním rozdělením, ale s různým průměrem*
- Př. Pro 170 cm existuje norm. rozdělení hmotností o průměru 68 kg, pro 180 cm opět normální rozdělení hmotností s průměrem 76 kg

Korelační závislost

- Určení těsnosti korelační závislosti
- (jak těsný je vztah mezi výškou a hmotností člověka)
- Korelační počet – snaha vyjádřit **tendenci** změny hodnoty závislé proměnné na nezávislé proměnné pomocí matematické funkce
- Tuto regresní funkci lze graficky znázornit **regresní čarou**

- **Korelace** je druh závislosti mezi prvky dvou souborů
- **Regresní čára** znázorňuje graficky tuto korelační závislost

Určení korelační závislosti

- 1. Korelační závislost vyjádřená lineární regresní přímkou (lineární regrese)
- Jedna nezávislá proměnná x a jedna závislá proměnná y' (ta je průměrem možných hodnot – viz. definice korelace)
- $X = 170$ cm a $y' = 68$ kg (68 kg zastupuje možné hodnoty hmotnosti pro 170cm)
- Regresní přímkou lze analyticky vyjádřit jako
- $y' = bx + a$, kde b je koeficient regrese a
- a dopočítáme po pomocném výpočtu průměrů souborů a dosazením jedné dvojice hodnot do rovnice
- $y' - \bar{y} = b(x - \bar{x}) + a$

Intervaly a pásy spolehlivosti pro lineární regresní závislost

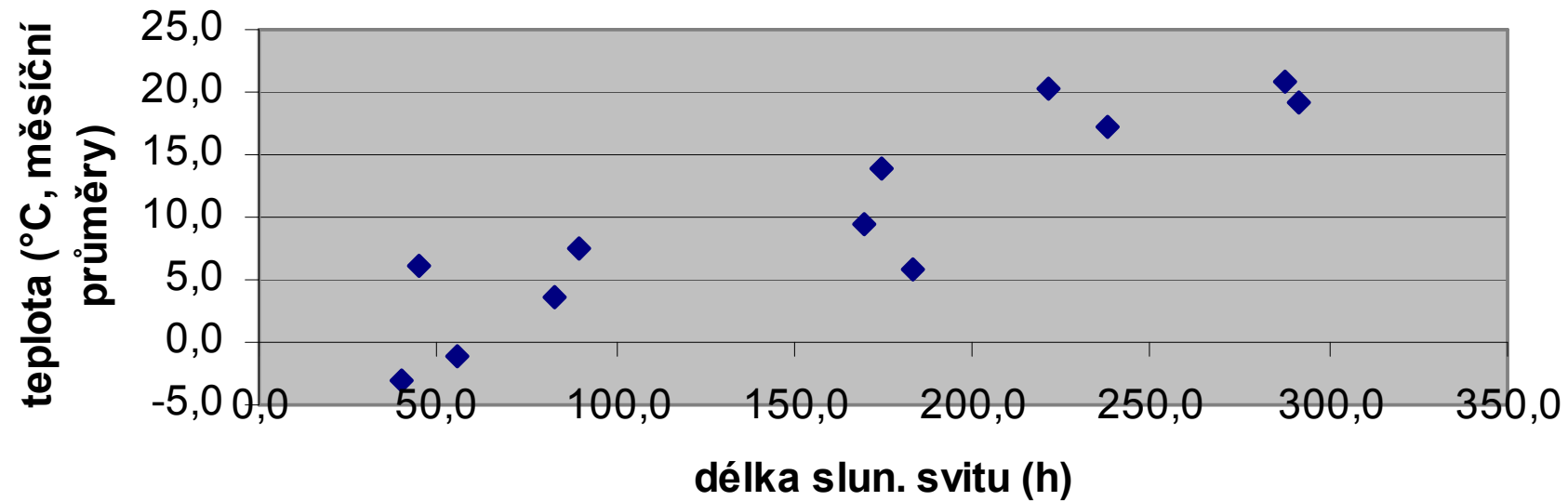
- Kolem regresní přímky lze sestavit
- **interval spolehlivosti,**
- který určuje pro vybrané x
- **interval, ve kterém se budou s určitou pravděpodobností nacházet hodnoty y**

Př. lineární regrese

- Vypočítejte parametry lineární regrese pro vztah délky slunečního svitu a teploty na datech meteorol. stanice Tuřany, 2002

Délka slun. svitu (h)	55,6	82,7	183,4	169,5	238,3	291,4	288,0	221,2	174,5	89,4	44,7	40,3
Teplota (°C)	-1,2	3,6	5,8	9,4	17,1	19,1	20,9	20,4	14,0	7,6	6,0	-3,1

Závislost teploty na délce slunečního svitu, Brno, 2002



Výpočet koeficientu regrese b :
 Excel, funkce CORREL, POLE1 - hodnoty délka slun. Svitů,
 Pole2 - hodnoty teploty

Microsoft Excel

CorREL = =CORREL(C17:N17;C18:N18)

CORREL

Pole1 C17:N17 = {55,6;82,7;183,4;169,5;238,3;291,4;288,0;221,2;174,5;89,4;44,7;40,3}

Pole2 C18:N18 = {-1,2;3,6;5,8;9,4;17,1;19,1;20,9;20,4;14,0;7,6;6,0;-3,1}

= 0,903991059

Vrátí korelační koeficient mezi dvěma množinami dat.

Pole2 je druhá oblast buněk s hodnotami. Hodnoty mohou být čísla, názvy, matice nebo odkazy obsahující čísla.

Výsledek = 0,903991059

OK Storno

10													
11													
12													
13	teplota	-1,2	3,6	5,8	9,4	17,1	19,1	20,9	20,4	14,0	7,6	6,0	-3,1
14	úhrn srážě	8,1	21,3	21,0	28,6	45,8	81,7	58,0	91,2	39,2	71,9	48,2	46,0
15													
16													
17	délka slun.	55,6	82,7	183,4	169,5	238,3	291,4	288,0	221,2	174,5	89,4	44,7	40,3
18	teplota	-1,2	3,6	5,8	9,4	17,1	19,1	20,9	20,4	14,0	7,6	6,0	-3,1
19													
20	korelace t/s	0,656547											
21													
22	korelace t/d	0,903991											
23													
24	korelace s/d	0,461355											
25													
26													
27	regresní přímka												
28													
29													
30													

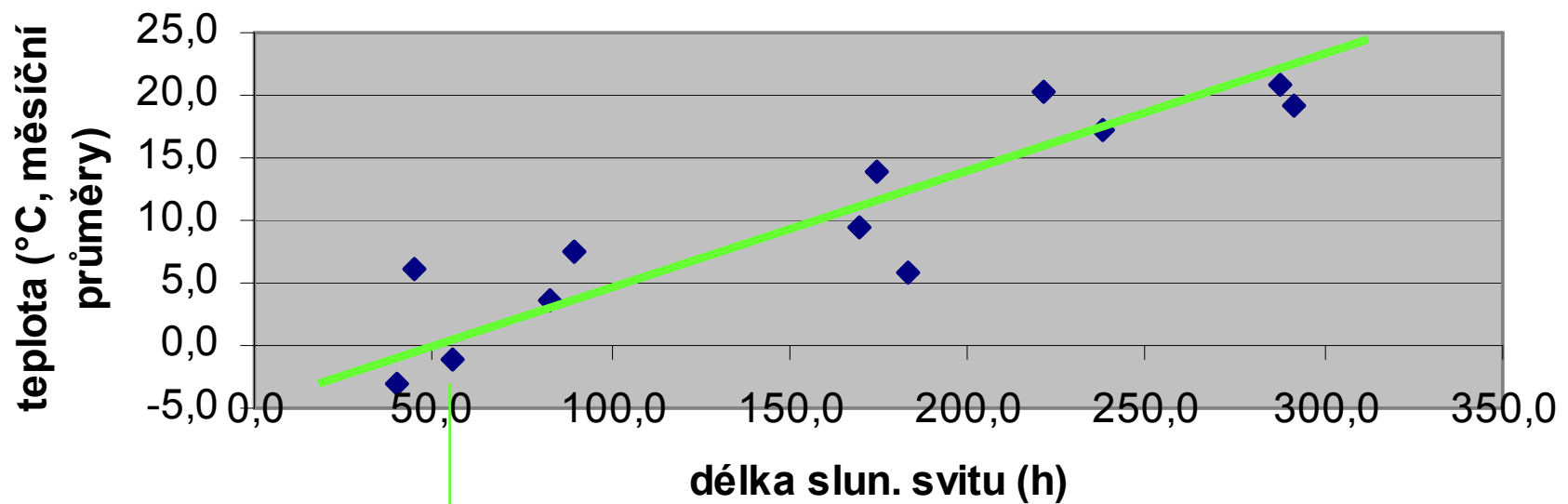
Závislost teploty na délce slunečního svitu, Brno, 2002

Y-axis: teplota (°C, měsíční průměry)

X-axis: délka slunečního svitu (h, měsíční průměry)

- Regresní parametr $b = 0,9$
- **Určení parametru a**
- Rovnice:
- **$y' - \bar{y} = b(x - \bar{x}) + a$**
- 1. Vypočítám aritm. průměr z hodnot x a y
- $\bar{x} = 156,6$ a $\bar{y} = 9,6$
- 2. Dosadíme z tabulky dvojici např. $(82,7 ; 3,6)$
- 3. řeším rovnici o jedné neznámé
- $3,6 - 9,6 = 0,9 * (82,7 - 156,6) + a$
- $a = -60$ —

Závislost teploty na délce slunečního svitu, Brno, 2002



60

Časové řady

Bazické a řetězové

Z - diagram

časová řady – základní pojmy

- **statistická řada**
- posloupnost hodnot znaku uspořádaných podle určitého hlediska
- časová řada
- statistická řada upořádaná podle času
- časová řada=dynamická=chronologická = vývojová

Sestavování časových řad

Cíl – získat porovnatelná čísla

- dodržovat zásady:
 - stejně dlouhá časová období
 - (přepočít na „standardizovaný“ měsíc se 30 dny,
přepočít na počet shodný počet pracovních dní v
měsíci p
 - stejně velká území, příp. stejná úroveň (shodná
rozloha, povodí řádu toku, administrativní jednotka)
 - stejně jednotky

● časová řada OKAMŽIKOVÁ

- sleduje se hodnoty znaku k určitému okamžiku
- např. počet obyvatel ČR k 31.12. 2000, 2001,

● časová řada INTERVALOVÁ

- sleduje se hodnota znaku v intervalu , období
- denní úhrn srážek, průměrná denní teplota, měsíční těžba...
 - pouze k této řadě se vztahuje **požadavek stejného intervalu** zvláště u sledování **ekonomických ukazatelů**

Klouzavé úhrny

- zvláštní typ součtové čáry
- vhodné pro porovnávání dvou či více řad hodnot za po sobě následující období
- např. kolísání ročního chodu srážek
- postup viz. např. skripta Brázdil. a kol. str. 147

měsíc	1	2	3	4	5	6	7	8	9	10	11	12
prům úhrn srážek;2002; mm	8,1	21,3	21	29	45,8	81,7	58	91,2	39,2	71,9	48,2	46
prům úhrn srážek;2003, mm	26,6	4,3	4,1	22	92,8	59,8	66,1	37	24,3	58,5	32,4	54, 3

KLOUZAVÝ ÚHRN	482, 6	454, 9	48 6	52 1	58 6	56 5	57 3	51 8	50 4	49 0	474, 3	48 3
--------------------------	-----------	-----------	---------	---------	---------	---------	---------	---------	---------	---------	-----------	---------

LEDNOVÁ HODNOTA – SOUČET „NOVÝ“ LEDEN + STARÉ OSTATNÍ MĚSÍCE

ÚNOROVÁ HODNOTA – SOUČET „NOVÝ“ LEDEN + ÚNOR
+STARÉ OSTATNÍ MĚSÍCE

Z - diagramy

- GRAFICKÉ ZNÁZORNĚNÍ
 - řada běžných hodnot,
 - součtová čára,
 - řada klouzavých úhrnů
- společné body Z - diagramu(tj. spol. hodnoty)
 - výchozí bod součtové č. a řady běžných hodnot
 - poslední hodnota součtové čáry a poslední hodnota klouzavého úhrnu

Microsoft Word - Cvičení 12_07_casove_rady_indexy.doc

Microsoft Excel

Úpravy Zobrazit Vložit Formát Nástroje Data Okno Nápověda

D15

z diagram, klouz.uhrn.xls

	A	B	C	D	E
1					
2			měsíc	1	2
3			prům úhrn srážek;2002; mm	8,1	21,3
4			prům úhrn srážek;2003; mm	26,6	4,3
5					
6					
7			průměrných úhrnů srážek Brno, 2003		
8					
9			měsíc	1	2
10			MĚSÍČNÍ PRŮMĚRY	26,6	4,3
11			KUMULOVANÝ SOUČET	26,6	30,9
12			KLOUZAVÝ PRŮMĚR	482,6	454,9
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					

zdrojová data

Oblast dat Řada

Řady

- MĚSÍČNÍ PRŮMĚRY
- KUMULOVANÝ SOUČET
- Rada3

Název:

Hodnoty:

Popisky osy X (kategorie):

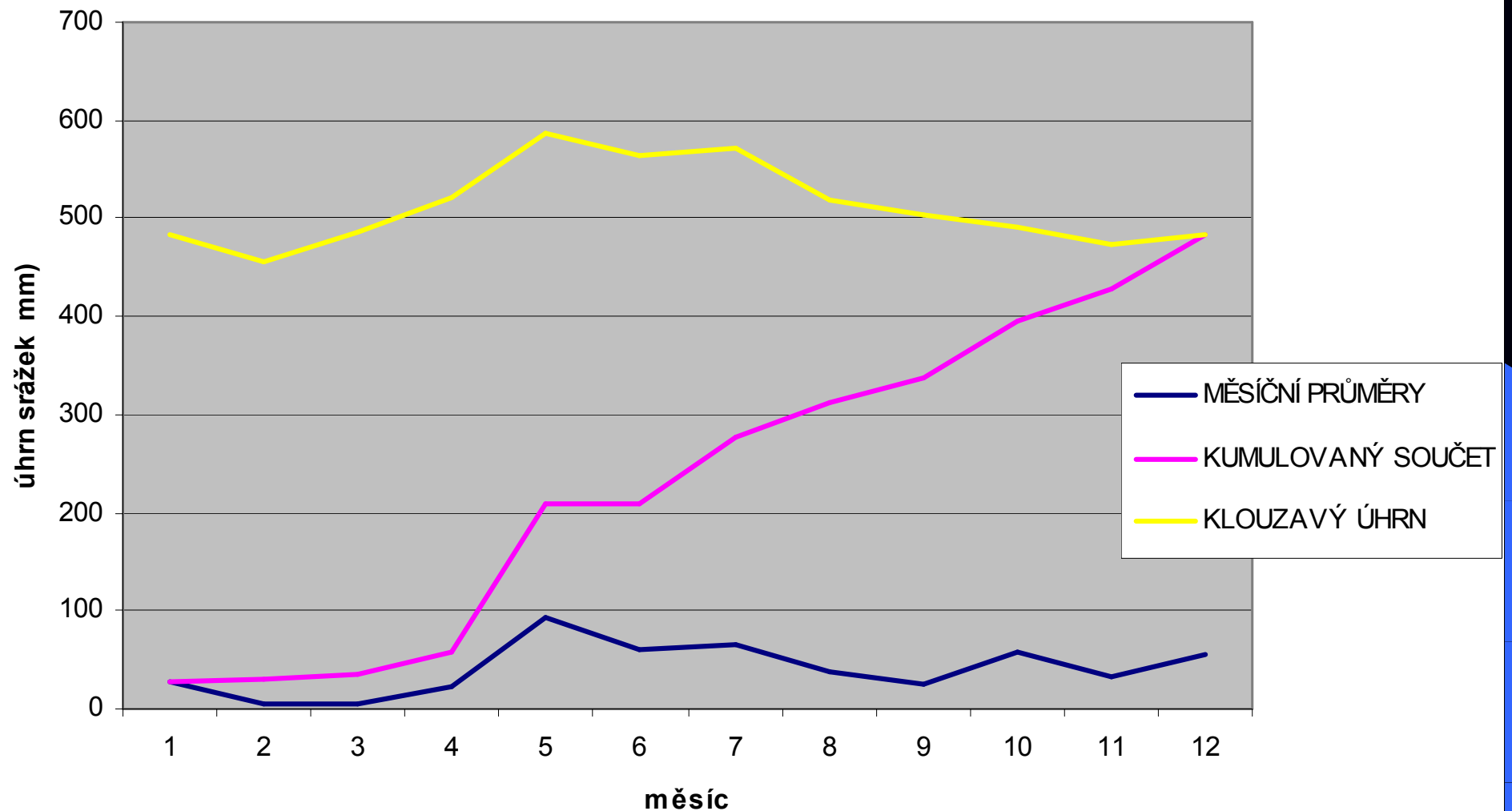
Storno < Zpět Další > Dokončit

List1 / List2 / List3 /

Start JARO_06 Cvičení_10_testo... Cvičení 12_07_ca... Microsoft Excel Microsoft Excel Darina Foltýnová... 10:59

Z - diagramy

Z - diagram průměrných úhrnů srážek (mm), Brno, 2003



Analýza časových řad

- cíle analýzy:
 - zjistit hlavní rysy průběhu časových řad a analyzovat je
- podle průběhu časové řady:
- stacionární nebo s trendem
- s periodickým opakováním výkyvů nebo bez výkyvů
- všechny možné kombinace

Charakteristiky časových řad

přírůstky a indexy

- přírůstky:
- **absolutní přírůstek** – rozdíl hodnot po sobě následujících („druhá“ – „první“)
- $X_i - X_{i-1}$
- **relativní přírůstek**
- podíl $X_i - X_{i-1} / X_{i-1}$

Řetězové a bazické indexy

- **bazický index**
- podíl $x_i / x_z * 100$,
- x_z - první „ základní „ hodnota časové řady
- změny k jedné základní (bazické) hodnotě

- **řetězový index** (koeficient růstu)
- podíl $x_i / x_{i-1} * 100$
- podíl v procentech po sobě následujících hodnot
- (změny např. z měsíce na měsíc“ – řetězení)

Témata přednášek k samostudiu

- Geografická metodologie
- Definice geografie
- Geografičnost studia
- Formy geogr. studia
- Obecný přístup k VŠ studiu
 - Literatura: skripta MEČIAR, J. *Úvod do studia geografie*, od. str. 107 do konce