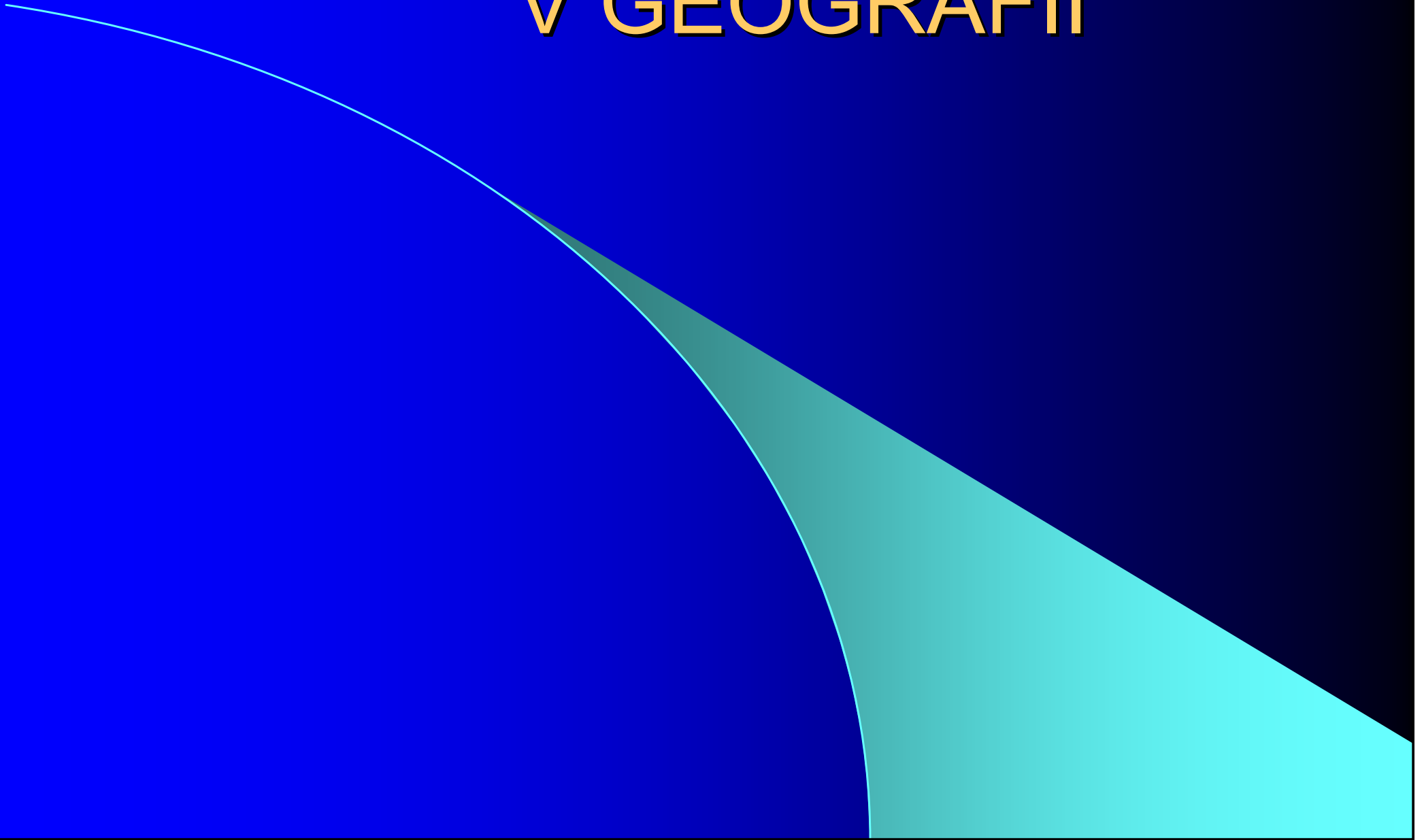


STATISTICKÉ METODY V GEOGRAFII





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
- 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 % ($\mu \pm 1,960\sigma$),
- pro 99 % ($\mu \pm 2,576\sigma$),

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

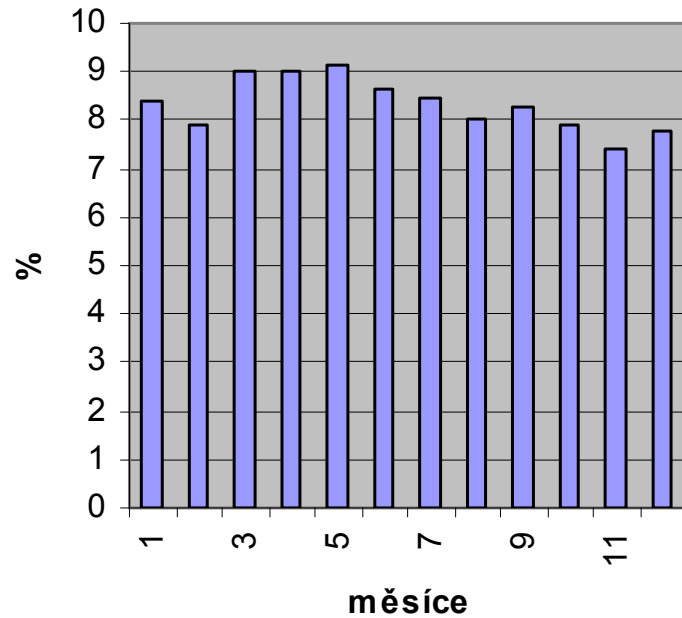
Soubor A

Soubor a

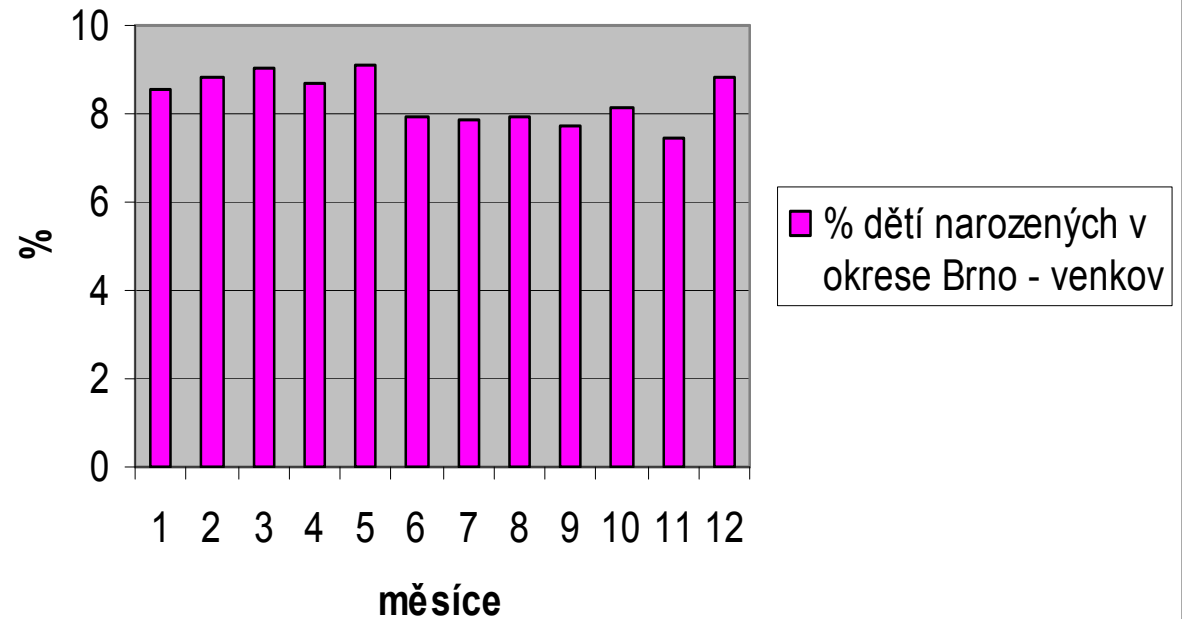
měsíc	% dětí narozených v ČR	% dětí narozených v okrese Brno - venkov
1	8,39	8,52
2	7,91	8,81
3	9,02	9,01
4	9,03	8,72
5	9,15	9,12
6	8,64	7,94
7	8,45	7,84
8	8,04	7,93
9	8,28	7,74
10	7,93	8,13
11	7,41	7,44
12	7,75	8,81

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:
- aritmetický průměr počtu narozených dětí v jednom měsíci je u obou souborů shodný
- **NULOVÁ HYPOTÉZA**
- **Ar. průměry četností obou souborů jsou shodné**
- zvolíme hladinu významnosti
- např. 5% , tj. $p=0,05$, tj. shoda je s pravděpodobností 95 %,
- **aplikace testovacího kritéria**

Testování

- Testovací kritérium – T – test
- Výsledek 0,995
- je výsledek testování významný ?
- podle výsledku přijmeme nebo odmítneme nulovou hypotézu
- **Přijmeme**

příklad

- Stejné soubory:
- Do jaké míry se liší rozložení četností v souboru pro Brno venkov od základního souboru pro ČR?
- Hladina významnosti: 99%
- Nulová hypotéza:
- Míra rozložení je shodná na hladině významnosti 99 %.

Test Chí - kvadrát

- Pro shodnost rozložení četností je vhodný test chí - kvadrát

The screenshot shows the Microsoft Excel interface with a CHITEST dialog box open. The dialog box has fields for 'Aktuální' (Actual) and 'Očekávané' (Expected) matrices, and a 'Výsledek =' field. Below the dialog box, a table displays statistical results for two data sets (List1 and List2).

	List1	List2			
	11	7,41			
	12	7,75			
	průměr	8,333333333	průměr	8,334166667	
	směrodatná odchylka	0,529013757	směrodatná odchylka	0,537563304	1%
	rozptyl	0,279855556	rozptyl	0,288974306	5%
	četnost	Chí kvadrát		0,99999653	
	rozptyl	F - test		0,958546132	
	dvě veličiny	t-test		0,997109132	
		CHITEST()		0,995860399	

Výsledek po dosazení
0,9999

Přijímáme nulovou
hypotézu, rozložení
četností obou souborů je
shodné.

F -test – test shody rozptylů

- Příklad
- Cvičné soubory A, a
- Hladina významnosti: 95%
- Nulová hypotéza: rozptyly obou souborů jsou shodné
- Aplikace testovacího kritéria: F test
- Výsledek: 0,958
- Přijmeme nul. hypotézu

Obecný postup testování vybr. charakteristik statistických souborů

1. Stanovení hladiny významnosti (95%, 99%)
2. Formulace nulové hypotézy
3. Aplikace testovacího kritéria:
 1. (Chí – kvadrát – četnost,
 2. F test - rozptyl
 3. T – test – dvě stat. charakteristiky
4. Výsledek testování
5. Přijetí * odmítnutí nulové hypotézy

Závislost náhodných veličin

Závislost náhodných veličin

- vztahy:
 - Počet obyvatel a název obce
 - teplota a nadm. výška,
 - Rok narození a jméno člověka
 - srážky a odtok v povodí
 - výška a váha člověka

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?

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 podle stupně závislosti

- 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 podle stupně závislosti

● 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,*
- *Př. Pro 170 cm existuje norm. rozdělení hmotností o průměru 68 kg, pro 180 cm opět normální rozdělení hmotností s průměrem 76 kg*

Obrázky vztah x a y

- Funkční závislost hodnoty y (tj. závislé) na x (nezávislé)
- Statistická dtto
- Korelační dtto

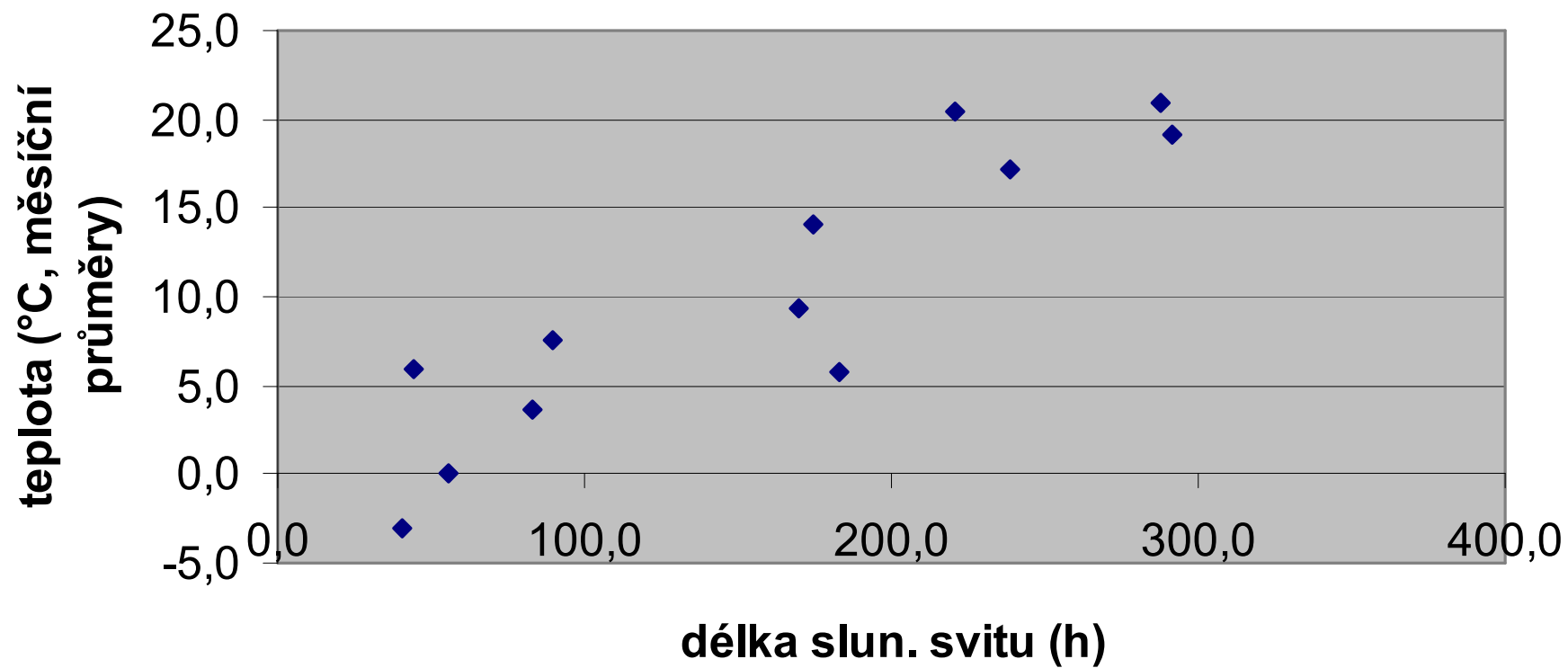
- **Korelace** je druh závislosti mezi prvky dvou souborů
- **Regresní čára** znázorňuje graficky tuto korelační závislost

Př. lineární regrese

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

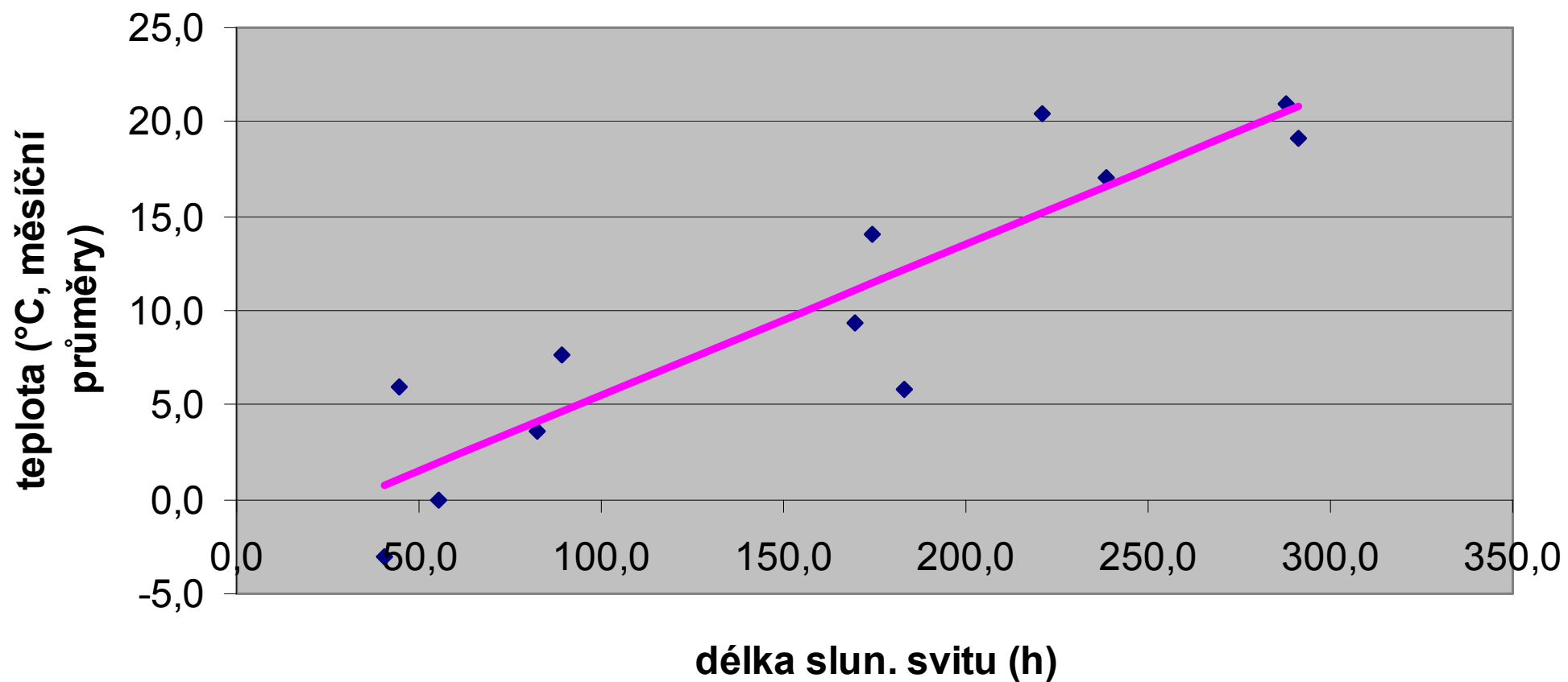
Délka slun. svitu (h)	55,6	82,7	183,4	169,5	238,3	291,4	288,0	221,2	174,5	89,4	44,7	40,3
Teplota (°C)	-1,2	3,6	5,8	9,4	17,1	19,1	20,9	20,4	14,0	7,6	6,0	-3,1

Závislost teploty na délce slunečního svitu, Brno, 2002



Závislost teploty na délce slunečního svitu, Brno, 2002

$$y = 0,0798x - 2,4271$$



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