

# STATISTICKÉ METODY V GEOGRAFII

Kdybych měl poslední den života, chtěl bych ho strávit na přednášce ze statistiky –  
- je tak nekonečně dlouhá .....

# statistika

- definice
- pojetí

# statistika - definice

***Statistika* je vědní obor zabývající se zkoumáním jevů, které mají hromadný charakter.**

***Statistika* je v určitém smyslu jazykem pro shromažďování, zpracování, rozbor, hodnocení a interpretaci hromadných jevů**

# Hromadný jev ve statistice

Statistika se zabývá **hromadnými jevy**

tj.:.....

jevy, které se vyskytují u souboru lidí, věcí,  
událostí bud'

.....formě nebo i

.....převoditelné na .....

hromadné jevy –...**studenti dopíší další příklady ( např. viz. statistické ročenky)....**

**Co je typické pro statistiku :**

# statistika – dva významy - dvě pojetí

— .....

- předmět zkoumání : stav a vývoj číselně vyjádřených hromadných jevů

— .....

- ( zaznamenání, třídění, shrnování číselných údajů o skutečnostech, *udělám si statistiku* .....)

# STATISTIKA jako vědní disciplína

- statistika popisná:.....

- matematická statistika:

.....



# Statistika jako praktická činnost

- **Činnost** .....
- **Instituce**, .....
- **výsledek** .....

# Základní etapy statistického zpracování dat

**Zjišťování – zpracování – analýza - prezentace**

- 1. Zjišťování/ Sběr údajů.....
- 2. Zpracování – .....
- 3. Analýza - .....
- 4. Prezentace výsledků .....

# Základní dělení statistických údajů

- podle zdroje ODKUD?— **primární a sekundární,**
- podle periodicity zjišťování JAK ČASTO? — **průběžné, periodické a jednorázové,**
- podle časového hlediska KDY? ZA JAK DLOUHO?— **okamžikové a intervalové**
- podle reálnosti situace OPRAVDU?— **skutečné a simulované,**
-

**Co statistika „umí“:**

**... a co statistika „neumí“:**

Statistika **selhává**, pokud:

# Statistika a výpočetní technika

- Výpočetní technika zasahuje do všech etap statistického zpracování dat.
- Exploze výpočetní techniky umožňuje provádět výpočty, které byly dříve nerealizovatelné (z důvodů velkého objemu dat, pracnosti, ...).
- Na druhou stranu však roste nebezpečí výběru nesprávného postupu.

# Výhody počítačového zpracování I.

- **Přesnost a rychlost**

- 

- **Univerzálnost**

- 

- **Grafika**

- **Flexibilita**

- **Nové veličiny:** *Snadno lze vytvářet nové veličiny pomocí požadovaných transformací.*

- **Velikost datových souborů:**

- **Snadný přenos dat:**

...ale

# Nevýhody počítačového zpracování I.

## **Chyby v softwaru.**

Ne všechny statistické programy jsou spolehlivé.

## **Univerzálnost.**

Může vést k výběru nevhodné metody zpracování.

*Je velmi důležité, aby každý, kdo používá statistický software, si byl vědom úrovně svých statistických znalostí a užíval pouze ty metody, kterým rozumí. Pozor na používání neznámých statistických metod.*

## **Černá skříňka.**

.

## **Špatná data plodí špatné závěry.**

.



# Vymezení základních statistických pojmů

# statistika - definice

***Statistika* je vědní obor zabývající se zkoumáním jevů, které mají hromadný charakter.**

## Hromadné jevy:

jevy, které

které se vyskytují u .....

A jsou

výsledkem .....

Příklady:

## Statistická jednotka:.....

Statistická jednotka musí být přesně vymezena na počátku vlastního šetření a to z hlediska

.....  
.....  
.....)

Příklady:

stat. jednotka – občan, novorozenec, byt, rodinný dům,

Co:

Kde:

Kdy:

## Statistický znak:

je to .....

Tzv. **shodné (společné) znaky** vymezují

.....  
.....





Ostatní jsou znaky

.....

Příklady: **shodný znak** – .....

.....znaky – váha, délka,  
jméno, národnost.....

Statistické znaky lze dělit na znaky

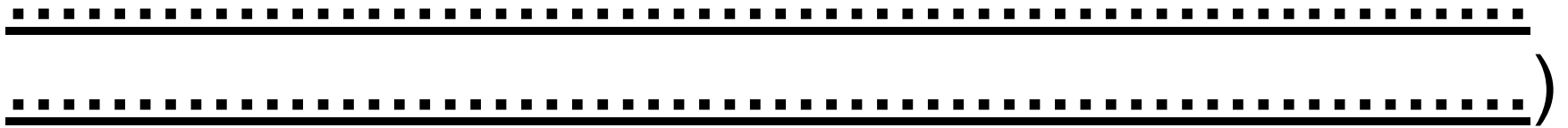
- **A) prostorové** 
- **B) časové** 
- **C) věcné:**
  - 1. kvalitativní:**
    - **alternativní**
    - **možné** 
  - 2. kvantitativní:** 
    - **spojité**
    - **diskrétní/nespojité**

  
Doplňte další příklady

Statistické znaky můžeme získat :

- **přímo** – (např. měřením, zvážením) tj.
- .....**data**
- **nepřímo** (výpočtem), (znaky odvozené) tj.  
.....**data**
- **Příklady:**
- **Nastudovat z Thinking Geography – uč. materiály**

## Statistický soubor:



Je to množina všech prvků, které jsou  
předmětem daného statistického zkoumání.  
**Každý z prvků je statistickou jednotkou.**

.



Prvky tvořící statistický soubor mají:

určité společné vlastnosti - tzv.

.....znaky

-.....znaky –

**tyto znaky** statisticky šetříme

Příklad:

statistický soubor

Shodný - :identifikační znak

sledovaný znak:

Statistický soubor:Občané v produktivním věku

Shodný - identifikační znak:

Sledovaný znak:

Statistický soubor můžeme podle různých hledisek dále dělit:

## Statistický soubor

- **jednorozměrný**
- **vícerozměrný**

Příklady

(váha dítěte), 1 –rozm.: 3650, 2100, 1200, 3500, 4100, 2800

dvourozm. (váha; délka), 3650, 55;

! jako dvojice! 2100, 47;

1200, 36,

3500, 50

# Statistický soubor **základní** a **výběrový**

## Výběrový soubor

je ..... základního souboru. Je vytvořen ze....., vybraných podle určitého hlediska.

Př. Novorozenci v .....

## Reprezentativní výběr:

Pokud zkoumaný výběr dobře odráží ..... celého zkoumaného souboru, nazýváme jej reprezentativním výběrem.

Př. ....

## **Rozsah statistického souboru:**

počet statistických jednotek v souboru:

$N$  – rozsah základního souboru

$n$  – rozsah výběrového souboru

# Grafické znázornění jevů

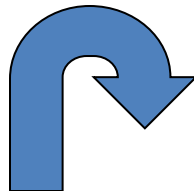
# Grafické znázornění jevů

- **Graf – definice**
- – kresba podle pravidel znázorňující kvalitativní a kvantitativní informace
  
- **Základní prvky** grafického znázornění:
  - 1.Název, příp. podnázev
  - 2.vlastní kresba
  - 3.stupnice a její popis (rovnoměrná, nerovnoměrná)
  - 4.legendy/klíč
  - 5.zdroj údajů
  - vysvětlivky, poznámky,

# Graf – ukázka

# Typy grafů

- **schéma** – znázorňuje strukturu a vztahy jevu či procesu
- **Příklad**
- **diagram** – znázorňuje kvantitativní údaje o souboru
  - sloupcový, bodový, plošný atd.
- **příklad**
- **statistická mapa** – prostorové rozložení prvku v podkladové mapě





schéma

# Diagram

**Český statistický úřad, 1994**

# Diagram\_ - věkové složení obyv., tzv.pyramida života

# Odchylka od průměrné teploty na Zemi

# Statistická mapa

# Použití grafických papírů při studiu geografických jevů

**Grafický papír** usnadňuje vynášení prvků do grafu.

- Milimetrový papír – rovnoměrné stupnice, čáry se jeví v původní, nezkrácené podobě
- Polologaritmický papír – kombinace dvou sítí – rovnoměrné a logaritmické
- Pravděpodobnostní papír – kombinace rovnoměrné a pravděpodobnostní stupnice

# Sítě

- Trojúhelníková síť – znázorňování jevů o třech prvcích, které mají vždy konstantní součet
- např. půdní druhy
- půda A:: 50 % jílu, 25% hlíny, 25%, písku

písek

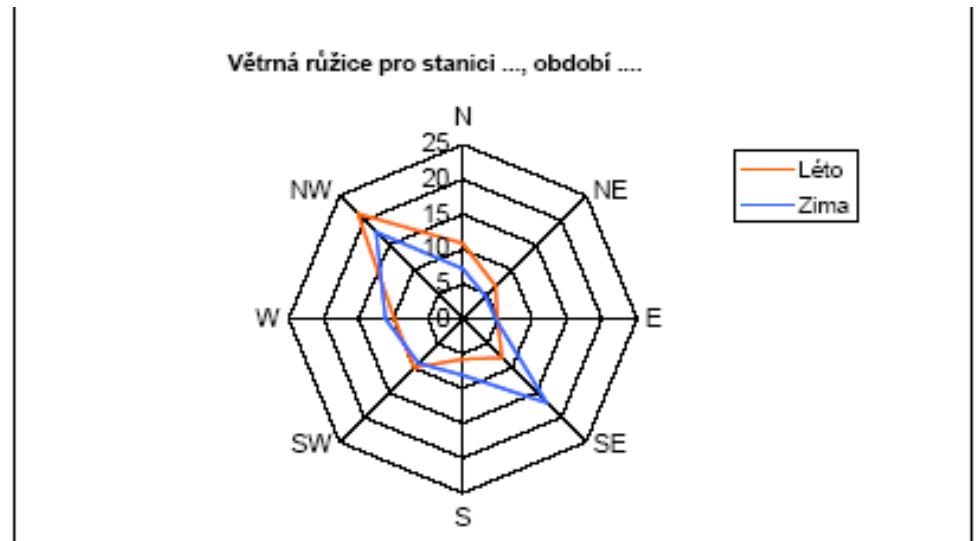
hlína

0

jíl

# Sítě

- Kruhová (radiální) síť – kombinace soustředných kružnic a přímek procházejících středem kružnice
- pro grafické znázorňování opakujících se jevů, struktury jevů
- Příklad
- roční chod teploty
- směry větru





statistická mapa:

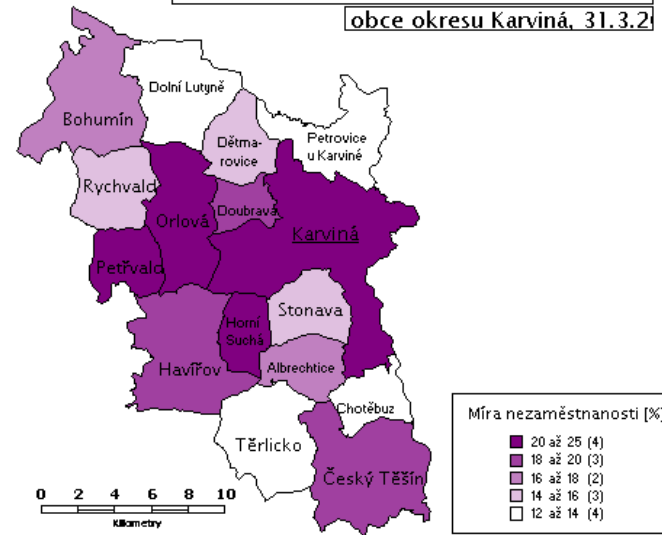
kartogram

kartodiagram

# Míra nezaměstnanosti

obce okresu Karviná, 31.3.21

## kartogram

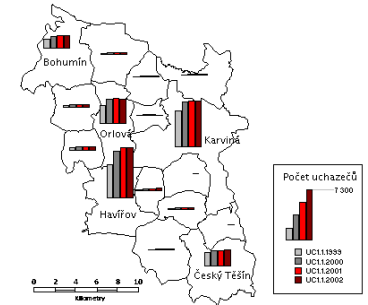


**Kartogram** je obrysová kartografická kresba územních celků, ve kterých jsou grafickým způsobem (barevný odstín, rast) plošně znázorněna statistická data týkající se různých geografických jevů (lidnatost, využívání ploch apod.)

Kartogramy lze rozdělit podle územního dělení na:

- kartogramy s ..... hranicemi
- kartogramy s ..... hranicemi

# Kartodiagram

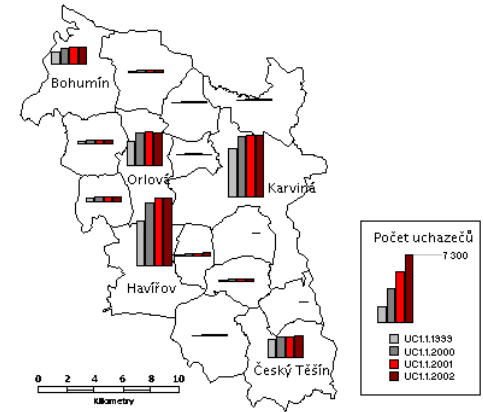


Kartodiagramy jsou diagramy vložené c  
mapové kostry, kterou tvoří dílčí územní celky.

Jejich údaje se vztahují na celé území jednotky,  
kde leží

( rozdíl od metody lokalizovaných diagramu –  
údaj vztahující se k urč. bodu – např. chod  
roční srážek na meteorolog. stanici)

# Kartodiagramy



Vkládanými diagramy mohou být:

- ..... diagramy pro vyjadřování časových řad
- .....diagramy (sloupce, věkové pyramidy apod.)
- různě dělené ..... značky

# Grafické metody analýzy geografických jevů

- 1. znázornění intenzity jevu v prostoru
- a) absolutními metodami
- \* značková metoda  
(.....)
- \* bodová metoda  
(.....)
- b) relativními metodami (např.  
.....)

- 2.znázornění **struktury** jevu v prostoru
- využití ..... grafů
- \*pouze strukturu vyjádříme výsečovými grafy se stejným .....
- \*strukturu a velikost celku (  
.....  
.....)

# Náležitosti statistické mapy

Obsah mapy tvoří všechny objekty, jevy a jejich vztahy, které jsou v mapě kartograficky znázorněny

Základní údaje tvoří

- **Název mapy** - stručně a výstižně charakterizuje zobrazené území, druh mapy  
lze i název hlavní a vedlejší)
- **Mapový rámeček** – „vlastní mapa“
- **Měřítko** v číselné, grafické nebo slovní formě
- **Legenda** (vysvětlivky) – podávají výklad použitých mapových značek a ostatních kartografických vyjadřovacích prostředků včetně barevných a velikostních stupnic, legenda musí být:
  - Úplná, logicky uspořádaná, přehledná a zapamatovatelná, **POZOR na intervaly, na barevnou škálu**
- **Autoři**

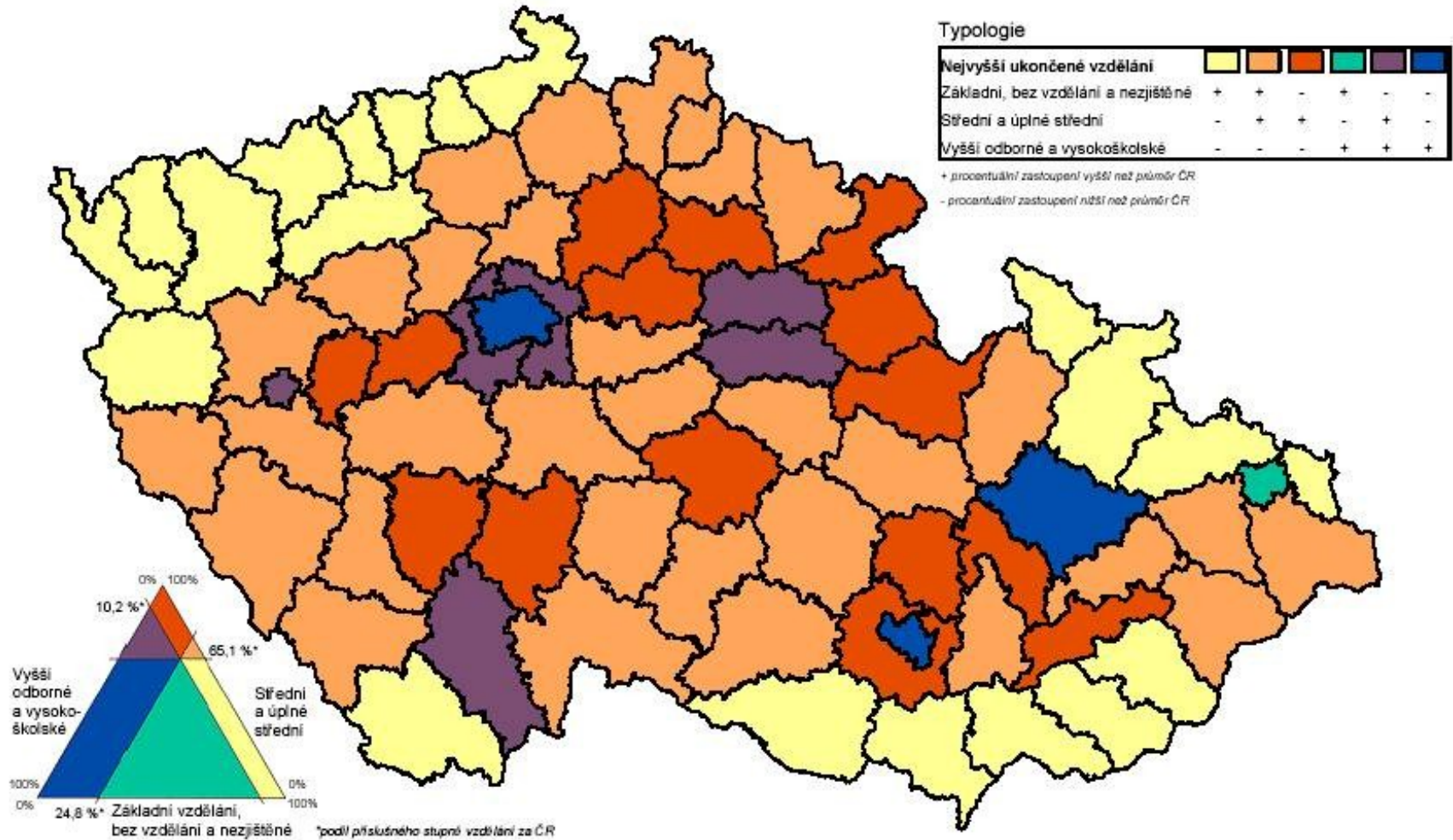
Dalšími údaji mohou být :

- vyznačení severu nebo směrová růžice, souřadnicový systém, přehled použitých mapových podkladů, datum, ke kterému se obsah mapy vztahuje
- obrázky, grafy, tabulky, text



# Vysvětlete

Typologie okresů ČR dle nejvyššího ukončeného vzdělání obyvatel ve věku 15 a více let - SLDB 2001



# Izolinie – konstrukce a vlastnosti

- Izolinie – .....
- získávají se metodou prostorové interpolace hodnot vynesných do grafu
- plynulé čáry
- Izo....., izo....., v.....atd.
- Konstrukce izolinie - příklad

# Rozdělení četností

# Absolutní, relativní kumulované četnosti

- četnost – počet .....určité hodnoty v souboru, frekvence hodnoty
- rozdělení četností – ..... prvků s určitými hodnotami statistického znaku, obvykle pro nespojitě hodnoty
- skupinové rozdělení četností - počty prvků s hodnotami statistického znaku, které patří do určitého ....., obvykle pro spojitě hodnoty

# skupinové rozdělení četností

- roztřídíme statistické jednotky podle velikosti jejich statistického znaku do .....
- interval – hranice, dolní a horní mez, šířka (délka)

## zásady:

- vymezené hranice pro jednoznačné zařazení prvků
- Obvykle .....šířka
- přiměřený ..... intervalů

# Četnosti

- absolutní četnost –
- relativní četnost –
- kumulovaná četnost –
- příklad

Tab.S Skupinové rozdělení četností,  
ukázka – příklad váha 1 novorozenců v JMK

Interval	střed	abs. č.	relativ. č.	kumul. abs.	kumul. relat.
500 - 1000	750				
1001 - 1500	1250				
1501 - 2000	1750				
atd.					

# Grafické znázornění rozdělení četností

- histogram
- polygon
- čára kumulovaných četností

čára kumulovaných četností – součtová čára,  
graf kumulované četnosti, vždy k horní hranici intervalu



# Histogram

Histogram – .....

šířka sloupce – šířka intervalu,  
výška sloupce - četnost

# Polygon

Polygon – spojnicový diagram,  
hodnoty četnosti se vynáší ke středům intervalu

# Čára kumulovaných četností

čára kumulovaných četností –

.....  
.....k horní hranici intervalu

histogram – věkové složení  
obyvatelstva, věková struktura,  
pyramida života

# Základní statistické charakteristiky

# Základní statistické charakteristiky

- základní statistické charakteristiky „popisují“ statistický soubor
- a) charakteristiky úrovně – tzv. střední hodnoty
- b) charakteristiky variability
- c) charakteristiky asymetrie a špičatosti

# Střední hodnoty

- Místo .....u  
jednorozměrného statistického souboru  
používáme často .....
- Střední hodnoty umožňují porovnávání  
souborů

# Střední hodnoty

- aritmetický průměr (+ vážený aritm. průměr, geometrický průměr, harmonický průměr)
- modus
- aritmetický střed
- medián a kvantily
- geografický medián



# Aritmetický průměr

- nejčastěji používaná st. charakteristika
- typický a netypický průměr
- (jedno a více vrcholová rozdělení četností)
- typický aritm. průměr – .....
- Netypický.....

Obr.

# Vážený aritmetický průměr

- při výpočtu množství srážek v povodí – váha – plocha území
- v klimatologii – výpočet denního průměru teplot ze tří měření

Př. výpočtu

# Modus

- modus - ..... hodnota kvantitativního znaku ve studovaném souboru
- významný především u souboru nespojitých veličin
- modální interval – interval zahrnující ..... počet jednotek, závisí však na stanovení hranic intervalů
- rozdělení s více mody – polymodální rozdělení

příklad

# Aritmetický střed

- Aritm. střed je polovina součtu min. a max. hodnoty znaku v souboru
- pokud soubor obsahuje extrémní hodnoty, je aritmetický střed značně zkreslující charakteristika

příklad

# Medián

- **Medián** – tzv. **prostřední hodnota**,
- je to prvek řady uspořádané v neklesajícím pořadí .....í), který ji dělí na dvě poloviny, které mají menší a větší hodnotu znaku
- POZOR: soubor je třeba .....
- vzorce určují .pořadí prvku (kolikátý prvek to je, hodnota prvku je medián!)
- pro řadu s lichým počtem prvků .....
- pro řadu o sudém počtu je medián průměr z hodnot mezi prvkem na (.....) místě
- **Příklad**

# Kvantily

- Medián je kvantil dělící soubor na dvě poloviny dle předch. pravidel

obdobně

- kvartily –  $x_{25}$ ,  $x_{50}$ ,  $x_{75}$ ,
- decily
- percentily

kvantily obecně široké použití ve statistice a v geografii

příklad

# Geografický medián

- Geografický medián je čára dělicí  
.....kde se jev vyskytuje tak, aby  
hodnota jevu byla v obou plochách stejná

# Charakteristiky variability

- variační rozpětí
- kvantilové odchytky
- průměrné odchytky
- rozptyl
- směrodatná odchytky
- variační koeficient



# Variační rozpětí

- rozdíl největší a nejmenší hodnoty sledovaného statist. znaku
- $R = x_{\max} - x_{\min}$
- jednoduchá charakteristika
- podléhá ..... hodnotám, které mohou být i chybami

příklad

# Průměrné odchyly

- průměrná odchyly je definována jako

.....

.....od vybrané střední hodnoty

(tj. od aritmetického průměru, mediánu, modu apod.)

# Kvantilové odchyly

- Založeny na kladných odchylnkách jednotlivých sousedních kvantilů
- např. kvartilová odchylnka
- decilová odchylnka
- percentilová odchylnka

# Střední diference

- je def. jako aritmetický průměr absolutních hodnot

.....  
.....

- v praxi vhodná pouze pro malé soubory

Příklad

# Rozptyl a směrodatná odchylka

- nejdůležitější charakteristiky variability
- **Rozptyl  $s^2$**  z  $n$  hodnot znaku  $x$  je průměr druhých mocnin odchylek jednotlivých hodnot znaku od aritmetického průměru
- **směrodatná odchylka  $s$**  je mírou měnlivosti hodnot souboru kolem aritmetického průměru
- je druhou odmocnina rozptylu

# Variační koeficient

- je častou používanou relativní mírou variability
- je definován jako poměr směrodatné odchylky k aritmetickému průměru

# Charakteristiky asymetrie

- Charakteristiky asymetrie ( míry šikmosti) jsou čísla dávající představu

o **souměrnosti** .....**tvaru**  
rozdělení četností

- míra šikmosti pro souměrné rozdělení je nula
- pro nesouměrné je kladná nebo záporná

# Charakteristiky asymetrie

Symetrické

Záporně sešikmené

Kladně  
sešikmené

ar. průměr, medián, modus



# charakteristiky špičatosti

- Charakteristiky špičatosti ( míry špičatosti) jsou čísla charakterizující ..... prvků souboru v .....určité hodnoty znaku

Obr. Špičaté, normální a ploché rozdělení

# charakteristiky špičatosti

1 – špičaté

2 – normální

3 – ploché

rozdělení

# STATISTICKÉ METODY V GEOGRAFII



Karl Friedrich  
Gauss  
1777-1855

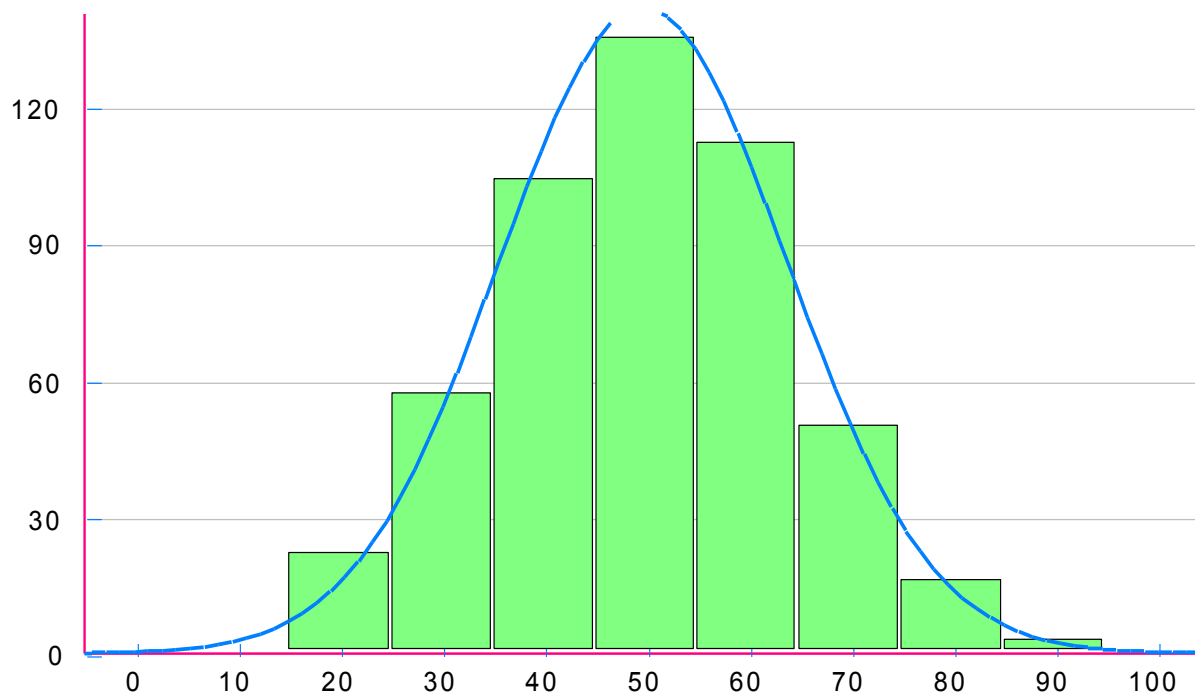
# Teoretická rozdělení

## Základní pojmy

- náhodná veličina .....
- Může teoreticky nabývat nekonečného množství hodnot z určitého intervalu např. teplota)
  
- náhodná veličina .....
- Nabývá jen konečného množství hodnot urč. intervalu. Např. počet měsíců s teplotou nad...)
- Každé hodnotě je možno přiřadit pravděpodobnost jejího výskytu, součet všech dílčích pravděpodobností je 1

# Teoretická rozdělení

- histogram – grafické znázornění četností
- rozsah souboru se blíží k nekonečnu + náhodná veličina je spojitá
- – frekvenční funkce / hustota pravděpodobnosti



- kumulativní relativní četnost tj. součtová čára  
-
- distribuční funkce
- obr.

# Normální rozdělení / Gaussovo, Laplaceovo-Gaussovo

- Normální rozdělení se univerzálně používá k aproximaci (k přibližnému vyjádření) rozdělení pravděpodobnosti velkého množství náhodných veličin (v biologii, technice, ekonomii atd.)

Hustota pravděpodobnosti normálního rozdělení je symetrická zvonovitá **Gaussova křivka**.

# Normální rozdělení

- Zvonovitý tvar
- Souměrný
- Šikmost 0, špičatost 0
- Asymptoticky se blíží 0

Obr.

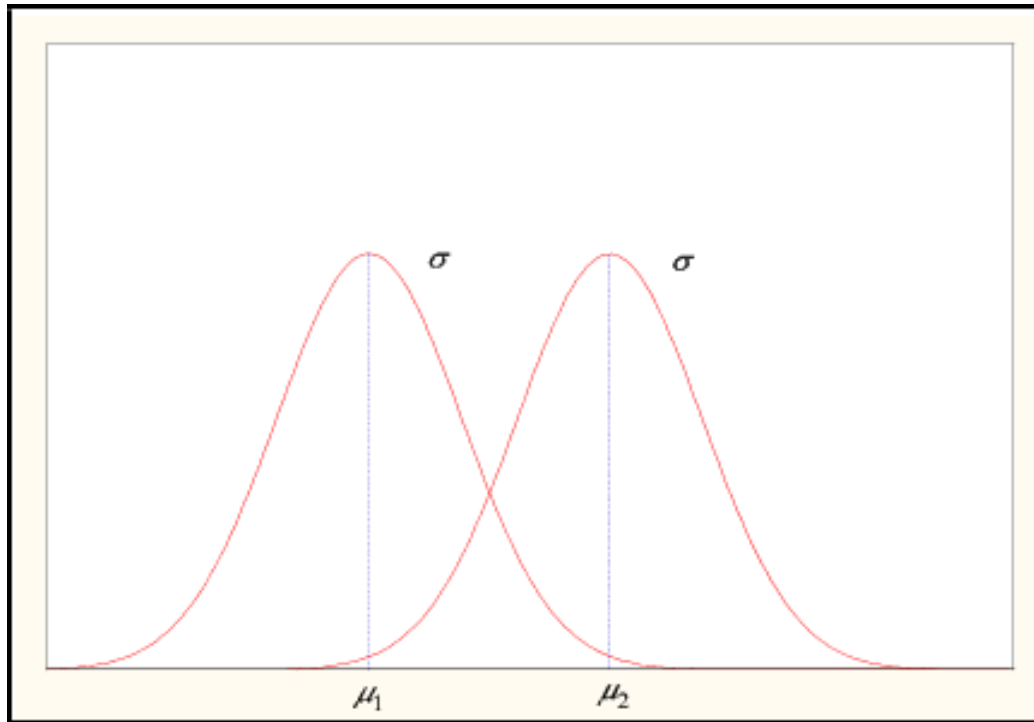


- Normální rozdělení s parametry:
- stejný průměr, různé směrodatné odchylky
- čím větší odchylka , tím „plošší tvar rozdělení

Načrtni obr s oběma křivkami

- Normální rozdělení s parametry:
- stejný průměr, různé směrodatné odchylky
- čím větší odchylka , tím „plošší tvar rozdělení

Načrtni obr



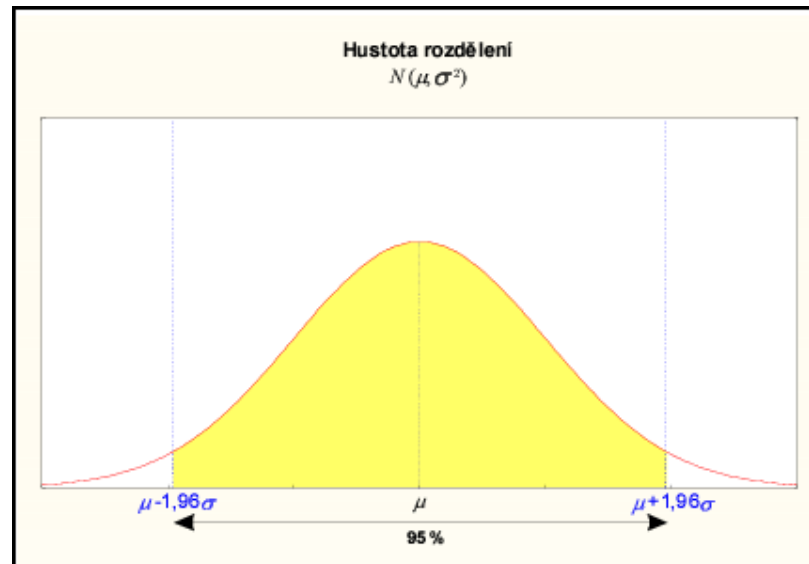
- Normální rozdělení
- různé průměry, stejná směrodatná odchylka

- Normální rozdělení
- různé průměry, stejná směrodatná odchylka

Načrtni obr s oběma křivkami

# Normální rozdělení / Gaussovo pokračování

- Normální křivka a osa x vymezují plochu 100%,
- tj. lze stanovit pravděpodobnosti, s nimiž leží hodnoty v určitém intervalu,
- hranice intervalu tvoří průměr a násobky směrodatné odchylky
- obr.



V normálním rozdělení:

- **68, 27% hodnot leží v intervalu:**

- (.....)

- **95% leží v intervalu:**

- (.....)

- **99% leží v intervalu:**

- (.....)

# Příklady

# Př.1

- Populace má v daném testu průměr 100, směrodatnou odchylku 15.
- Vypočítejte hranice intervalů, v kterém se nachází 68 % populace.



## Příklad 2

- Výška v populaci chlapců ve věku 3,5 - 4 roky má normální rozdělení s průměrem 102 cm a směrodatnou odchylkou 4,5 cm.
- Vypočítejte hranice intervalu hodnot výšky , ve kterých se nachází
- A)68%
- B) 95%
- C)99%
- příslušné populace

# Příklad 3

- zadání:
- Výška v populaci chlapců ve věku 3,5 - 4 roky má normální rozdělení s průměrem 102 cm a směrodatnou odchylkou 4,5 cm.
- **Spočtěte, jaké procento chlapců v uvedeném věku má výšku menší nebo rovnou 93 cm.**



# Příklad 4

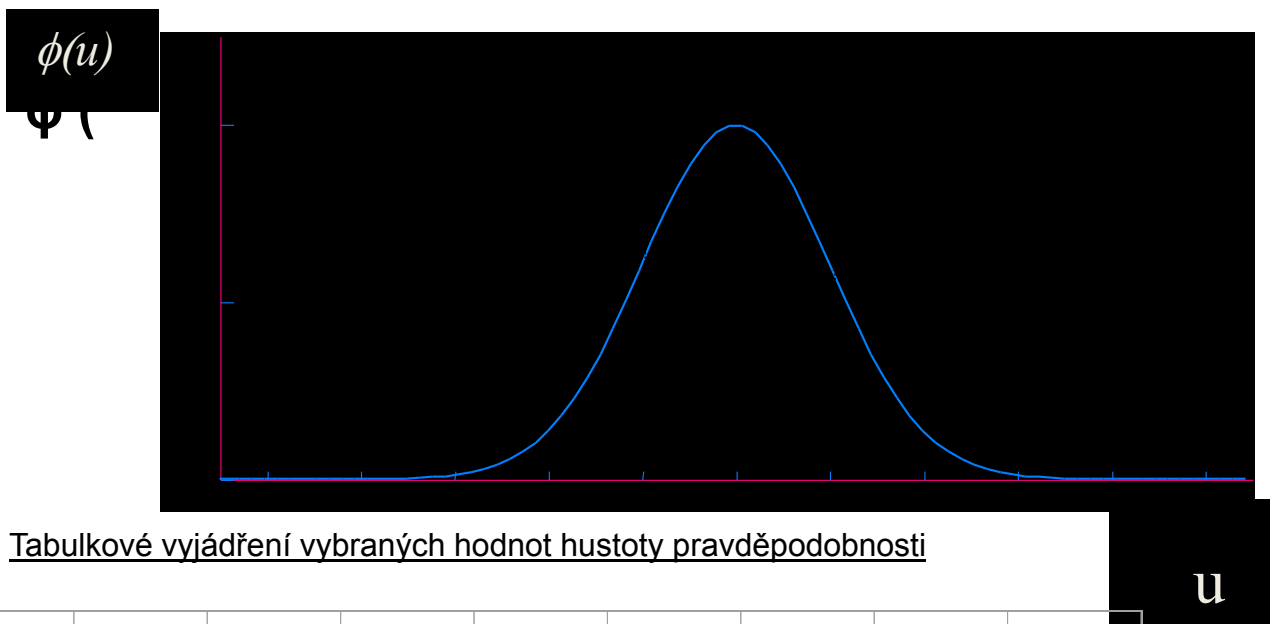
- Psychologickými testy bylo zjištěno, že hodnota IQ populace je náhodnou veličinou s normálním rozdělením, jehož střední hodnota je **104 a směrodatná odchylka 8.**
- Určete hodnotu IQ, kterou podle uvedených pravděpodobnostních předpokladů:
  - **meze, ve kterých bude 50% populace,**
  -



- Pro normované normální rozdělení zavedeme označení  $N(0, 1)$ .

Normování hodnoty: od hodnoty se odečte aritmetický průměr, výsledek (tj. odchylka) se dělí směr. odchylkou

Hustota pravděpodobnosti normovaného normálního rozdělení:



Tabulkové vyjádření vybraných hodnot hustoty pravděpodobnosti

$u$	0,0	0,5	1,0	1,5	2,0	2,5	3,0	3,5
$\phi(u)$	0,399	0,352	0,242	0,130	0,054	0,018	0,004	0,001

Tabulkové vyjádření vybraných hodnot distribuční funkce

$u$	0,0	0,5	1,0	1,5	2,0	2,5	3,0	3,5
$\Phi(u)$	0,500	0,691	0,841	0,933	0,977	0,994	0,999	0,999



# Binomické rozdělení



# Binomické rozdělení

- pro diskrétní náhodné proměnné,
- které mohou nabývat pouze dvou hodnot ( např. ano, ne)
- pravděpodobnost, že nastane alternativa ANO označme  $\pi$
- pravděpodobnost, že nastane NE ... $q = 1 - \pi$ ), protože
- platí ..... (100 %)
- k výpočtu se používá **binomický rozvoj**

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$P(X = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}, \quad \text{pro } k = 0, 1, 2, \dots, n.$$

# Příklad 1 – binomické rozdělení

- Předpokládejme, že pravděpodobnost narození dívky je 0,49.
- Jaká je pravděpodobnost toho, že mezi třemi dětmi v rodině je právě jedna dívka?

# Řešení 1

**Tabulka3:** Parametry binomického rozdělení v příkladu

Pokus	Úspěch	Neúspěch	Pravděpodobnost úspěchu	Počet pokusů	Počet úspěchů
				$n$	$k$
narození dítěte	dívka	chlapec	0,49	počet dětí	počet dívek

# Řešení 1

Jak je vidět z tabulky, počet narozených dívek v rodině je náhodná veličina s binomickým rozdělením. Pravděpodobnost, že mezi třemi dětmi je právě jedna dívka tedy vypočteme jako

$$P(X = 1) = \binom{3}{1} 0,49^1 \cdot 0,51^2 = 3 \cdot 0,127 = 0,38. \diamond$$

$$\binom{3}{1} = \frac{3!}{1!2!} = \frac{3 \cdot 2 \cdot 1}{1 \cdot 2 \cdot 1} = 3.$$

Pravděpodobnost, že ze tří dětí bude jedna dívka, je 38%.

Microsoft Excel

Binomická funkce: **BINOMDIST**

**Úspěch** 1 = 1

**Pokusy** 3 = 3

**Prst\_úspěchu** 0,49 = 0,49

**Počet** NEPRAVDA = NEPRAVDA

Vrátí hodnotu binomického rozdělení pravděpodobnosti jednotlivých veličin.

**Počet** je logická hodnota: součtová distribuční funkce = PRAVDA, hromadná pravděpodobnostní funkce = NEPRAVDA.

Výsledek = 0,382347

OK Storno

# Příklad 2

Jaká je pravděpodobnost, že v rodině s 8 dětmi jsou právě 3 dívky? Pravděpodobnost narození dívky je 0,49.

Řešení

binomický rozvoj:

# Příklad 2, binomické rozdělení

- Vypočítejte pravděpodobnost, se kterou se vyskytne určitý počet měsíců v roce hodnocených jako „suché“.
- Konkretizace:
  - oblast Oxford,
  - období 1851 – 1943, tj. 1116 měsíců
  - Suchý měsíc - tj. méně srážek v měsíci než je dlouhodobý průměr tohoto měsíce.
  - 617 měsíců hodnocených jako suché
  - 499 – vlhké měsíce

## Řešení 2

„úspěch“	„neúspěch“	Pravděpodobnost suchého měsíce	Pravděpodobnost vlhkého měsíce	Počet měsíců	Počet suchých měsíců
suchý	vlhký	$\pi = \dots$ $\pi = .$	$q = \dots$ $q = ..$ $(q = 1 - \pi)$	n = 12	k = 0 až 12

### Řešení

- Ručně pomocí binomického rozvoje
- s podporou např. Excel

Řešíme dílčí příklady, tj. jaká je pravděpodobnost, že v roce se vyskytne

- žádný suchý měsíc, tj-  $k = 0$
- Jeden suchý měsíc, tj.  $k = 1$
- Atd.
- všechny měsíce suché,  $k = 12$

## Řešení 2

Microsoft Excel

Toolbar: File, Edit, Format, Tools, Data, Window, Help

Menu: Soubor Úpravy Zobrazit Vložit Formát Nástroje Data Okno Nápověda

Formula Bar: BINOMDIST [X] [✓] [=] =BINOMDIST(5;12;0,553;npravda)

**BINOMDIST**

Úspěch	5	= 5
Pokusy	12	= 12
Prst_úspěchu	0,553	= 0,553
Počet	npravda	= NEPRAVDA

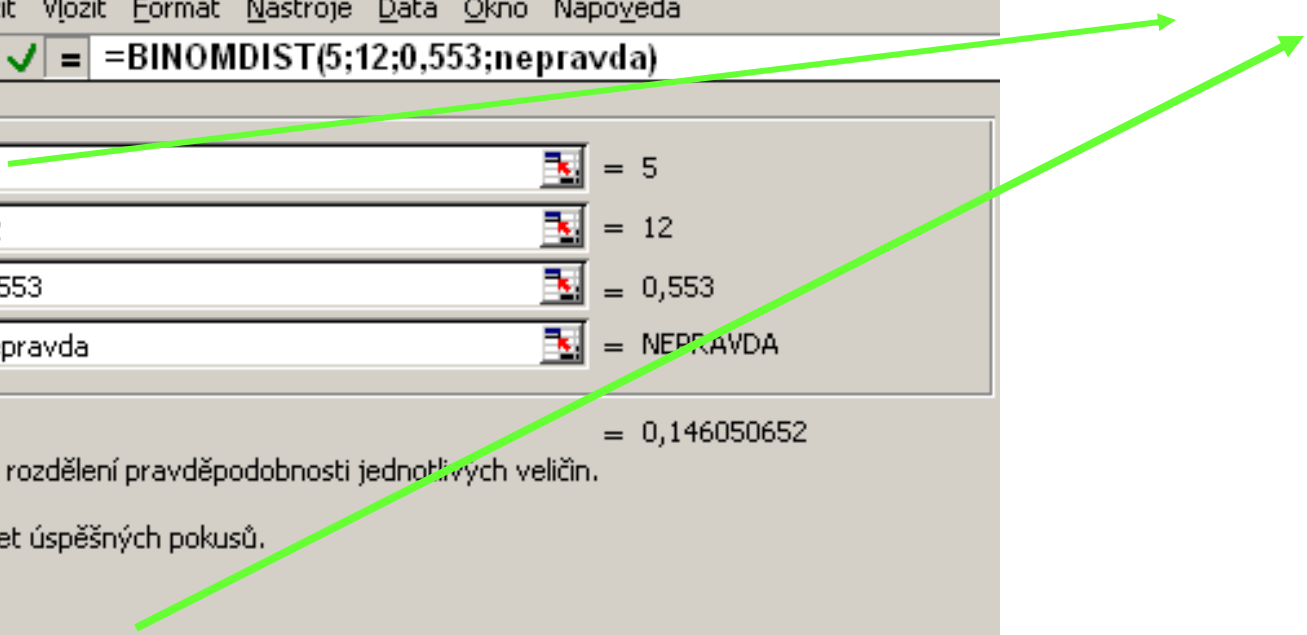
= 0,146050652

Vrátí hodnotu binomického rozdělení pravděpodobnosti jednotlivých veličin.

**Úspěch** je počet úspěšných pokusů.

Výsledek = 0,146050652

Buttons: OK, Storno







Jak bude vypadat situace pro „vlhké“ měsíce?

# Poisson - příklad

# Poissonovo rozdělení

- – pro rozdělení ..... případů
- (zimní bouřka, výskyt mutace apod.).
  
- Je-li pravděpodobnost nějaké výjimečné události (např. určité mutace genu) relativně malá a rozsah výběru poměrně velký, pak **Poissonovo rozdělení v podstatě splývá s binomickým**, ale je mnohem výhodnější pro počítání .

# Poisson - příklad

- Předpokládejme, že v určité populaci krys se vyskytuje albín s pravděpodobností
- $p = 0,001$  , ostatní krysy jsou normálně pigmentované.
- Ve vzorku 100 krys náhodně vybraných z této populace určete pravděpodobnost, že vzorek
- a) neobsahuje albína,
- b) obsahuje právě jednoho albína.

# Řešení

- určete pravděpodobnost, že vzorek
- neobsahuje albína,

Pravděpodobnost, že neobsahuje albína, je.....

# Řešení 3

**Pravděpodobnost, že 100 členná populace krys bude obsahovat albína, je .....%.**

# Další rozdělení



# Pearsonova křivka III. typu

- Na empirické rozdělení mnoha statistických souborů s nimiž v geografii pracujeme, **nelze aplikovat normální rozdělení.**
- Platí to například v těch případech, kdy studovaná náhodná veličina **nemá teoreticky zdůvodněnou možnost nabývat nekonečných hodnot** nebo je-li omezena konečnými čísly  
V takovýchto případech lze aplikovat na studovaný soubor některou ze dvanácti křivek Pearsonova systému.

# Pearsonova křivka III. typu

- Pearsonova křivka III. typu
- - obvykle pro veličiny s omezeným množstvím hodnot, které může nabývat
- - z křivky lze např. vyčíst.....  
.....
- v hydrologii se počítá Pearsonova křivka ve variantě součtová čára četností jako
- tzv. .....

- **příklad**
- Konstrukce čáry překročení z průměrných ročních průtoků vodního toku Lažánka za říjen 2002.

## načrtni

Křivka překročení průměrných ročních průtoků , Lažanka, říjen 2002

**m<sup>3</sup>/s**

**20**

**15**

**10**

**5**

**0**

**0**

**20**

**40**

**60**

**80**

**100 %**

# Odhady parametrů intervaly spolehlivosti

# Základní pojmy

- základní soubor,
- statistický soubor
- výběrový soubor
- náhodný výběr
- k základnímu jednomu souboru lze získat více výběrových, různé charakteristiky

# Základní pojmy

- reprezentativnost výběru –

.....

- prostý náhodný výběr (

.....

- oblastní náhodný výběr (

.....

- systematický náhodný výběr (

.....

.....

.....

# Intervaly spolehlivosti

- normální rozdělení,
- Statistický soubor s norm rozdělením  $(X, s)$
- Jeho výběrový soubor bude mít norm rozdělení s param  $(x, s/\sqrt{n})$ ,
- Interval spolehlivosti – pro zvolený koeficient spolehlivosti ( pravděpodobnost , že tam  $X$  padne) (např. 95 %)
- vypočítáme interval, ve kterém s touto pravděpodobností leží  $X$ .



- provedeme-li výběr o rozsahu  $n$  a spočteme  $\bar{x}$ , pak průměr  $\bar{X}$  leží s pravděpodobností 0,95 ve vzdálenosti menší než  $1,96 s / \sqrt{n}$  od  $\bar{x}$ ,
- tj. v intervalu s krajními body
- $(\bar{x} - 1,96 s / \sqrt{n}, \bar{x} + 1,96 s / \sqrt{n})$ ... *interval spolehlivosti pro průměr*.
- *koefficient spolehlivosti*  $P = 0,95$
- (tj. *hladinu významnosti*  $\alpha = 0,05$  )

- lze použít intervaly spolehlivosti např.
- pro 95 % ( $\mu \pm 1,960\sigma$ ),
- pro 99 % ( $\mu \pm 2,576\sigma$ ), tj. širší! interval
  
- hodnoty, které leží mimo interval, v tzv. **kritickém oboru** se považují za nepřípustné, jejich odchylky od průměru za významné

# Testování statistických hypotéz

- jak ověřit předpoklady o charakteristikách statistických souborů?
- Je soubor A reprezentativním výběrem ze souboru B?
- Do jaké míry se soubory shodují v rozdělení četností, podle aritm. průměru, podle směrodatné odchylky apod.
- Test nulové hypotézy tedy spočívá v tom, zda  $p$  překročí, nebo nepřekročí zvolenou mez

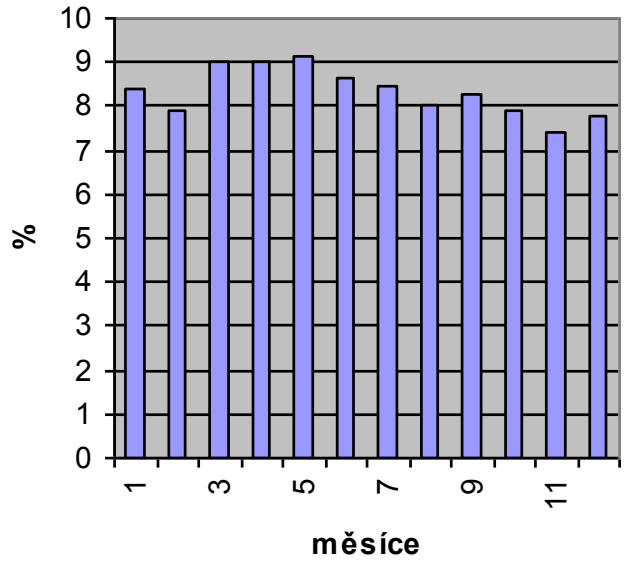
# Příklad

Soubor **A**

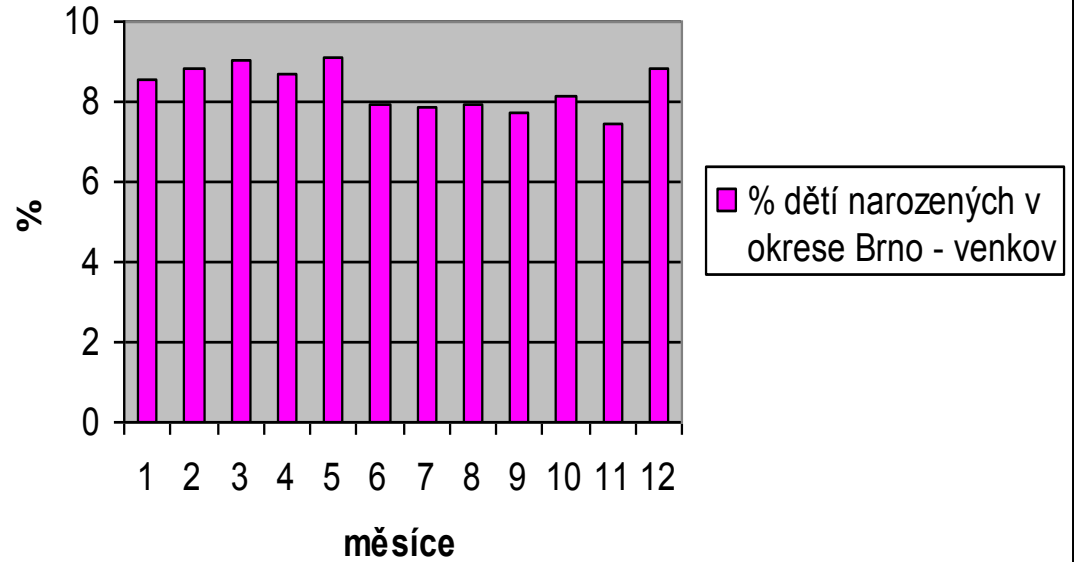
Soubor **a**

# Rozdělení četností souborů A, a

% dětí narozených v ČR



% dětí narozených v okrese Brno - venkov



# cvičení

- a. stanovte hladinu významnosti pro porovnání souborů na 95 %.
- b. formulujte nulovou hypotézu pro shodu souborů.
- c. porovnejte shodnost rozložení četností souboru pro Brno – venkov se souborem pro ČR pomocí testu CHÍ – kvadrát (CHITEST)
- d. Testujte významnost rozdílu rozptylů souborů pomocí F – testu (ftest)
- e. Testujte významnost rozdílu pomocí T - testu.( ttest)
- f. Přijměte nebo odmítněte nulovou hypotézu.
- chitest 0,99 ftest 0,958 t test 0,997

- Test nulové hypotézy tedy spočívá v tom, zda  $p$  překročí, nebo nepřekročí zvolenou mez
- STATISTICKÁ HYPOTÉZA:
- předpoklad: průměrná výška studentek PdF MŮ 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.  $\alpha = 0,05$ , tj. ( pravděpodobností 95 % )
- aplikace testovacího kritéria
- je výsledek testování významný ?
- podle výsledku přijmeme nebo odmítneme nulovou hypotézu

# 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 (

.....

.....

....

- Vzájemné

(.....

.....)

# 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.*  
.....  
.
- ..
- *Př.* .

# Korelační závislost

- Určení těsnosti korelační závislosti
- (jak těsný je vztah mezi výškou a hmotností člověka)

- **Korelace** je

.....

- **Regresní čára** znázorňuje

.....  
\_\_\_\_\_

# Intervaly a pásy spolehlivosti pro lineární regresní závislost

Kolem regresní přímky lze sestavit

.....,

který určuje pro vybrané  $x$

interval, ve kterém se budou s určitou  
pravděpodobností nacházet hodnoty  $y$

# Př. lineární regrese

- Vypočítejte koeficient korelace 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

<b>t</b>	<b>-1,2</b>	<b>3,6</b>	<b>5,8</b>	<b>9,4</b>	<b>17,1</b>
<b>h</b>	<b>55,6</b>	<b>82,7</b>	<b>183,4</b>	<b>169,5</b>	<b>238,3</b>

**corel h, t            0,920888**

**corel t, h            0,920888**





Výpočet koeficientu regrese  $b$  :

Excel, funkce CORREL, POLE1 - hodnoty délka slun. Svitů,

Pole2 - hodnoty teploty

lineární regresní čára - Přidat spojnici trendu

Časové řady  
Bazické a řetězové  
Z - diagram

# časová řady – základní pojmy

- statistická řada

- 



∴

- časová řada

- statistická řada upořádaná podle .....

- časová řada=.....

- 



# Sestavování časových řad

**Cíl – získat porovnatelná čísla**

- dodržovat zásady:

- stejně dlouhá časová období

- (

- .....

- .....

- stejně velká území, příp. stejná úroveň

- (.....

- .....

- stejné jednotky

- časová řada OKAMŽIKOVÁ

- sleduje se hodnoty znaku k určitému okamžiku
- např.

- časová řada INTERVALOVÁ

- sleduje se hodnota znaku v intervalu , období
- př.....

- pouze k této řadě se vztahuje **požadavek stejného intervalu** zvláště u sledování **ekonomických ukazatelů**

# 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 –

.....  
.....

- relativní přírůstek

- 

.....  
.....

# Řetězové a bazické indexy

- bazický index

- .....
- .....
- .....

- řetězový index (koeficient růstu )

- .....
- .....
- .....

# 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. skriptu 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



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

Klouzavý úhrn, vždy součet 12 měsíčních hodnot, tj. daný měsíc plus +11 předchozích

LEDNOVÁ HODNOTA – SOUČET „NOVÝ“ LEDEN + 11 předchozích měsíců

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

# Z - diagramy

- GRAFICKÉ ZNÁZORNĚNÍ

- Řada.....,
- .....čára,
- řada .....

- společné body Z - diagramu( tj. spol. hodnoty)

- .....
- .....
- .....
- .....

diagram, klouz.uhrn.xls

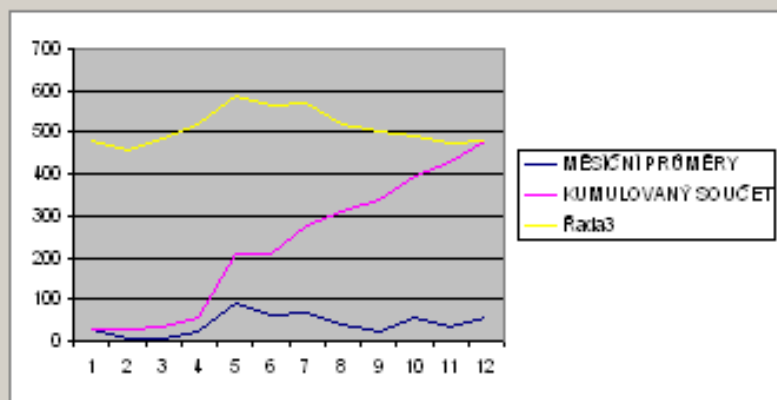
	A	B	C	D	E
			měsíc	1	2
			prům úhrn srážek;2002; mm	8,1	21,3
			prům úhrn srážek;2003, mm	26,6	4,3

průměrných úhrnů srážek Brno, 2003

	A	B	C	D	E
			měsíc	1	2
			MĚSÍČNÍ PRŮMĚRY	26,6	4,3
			KUMULOVANÝ SOUČET	26,6	30,9
			KLOUZAVÝ PRŮMĚR	482,6	454,9

### Zdrojová data

Oblast dat Řada



Řady

- MĚSÍČNÍ PRŮMĚRY
- KUMULOVANÝ SOUČET
- Řada3

Název:

Hodnoty:

Přidat

Odstranit

Popisky osy X (kategorie):



Storno

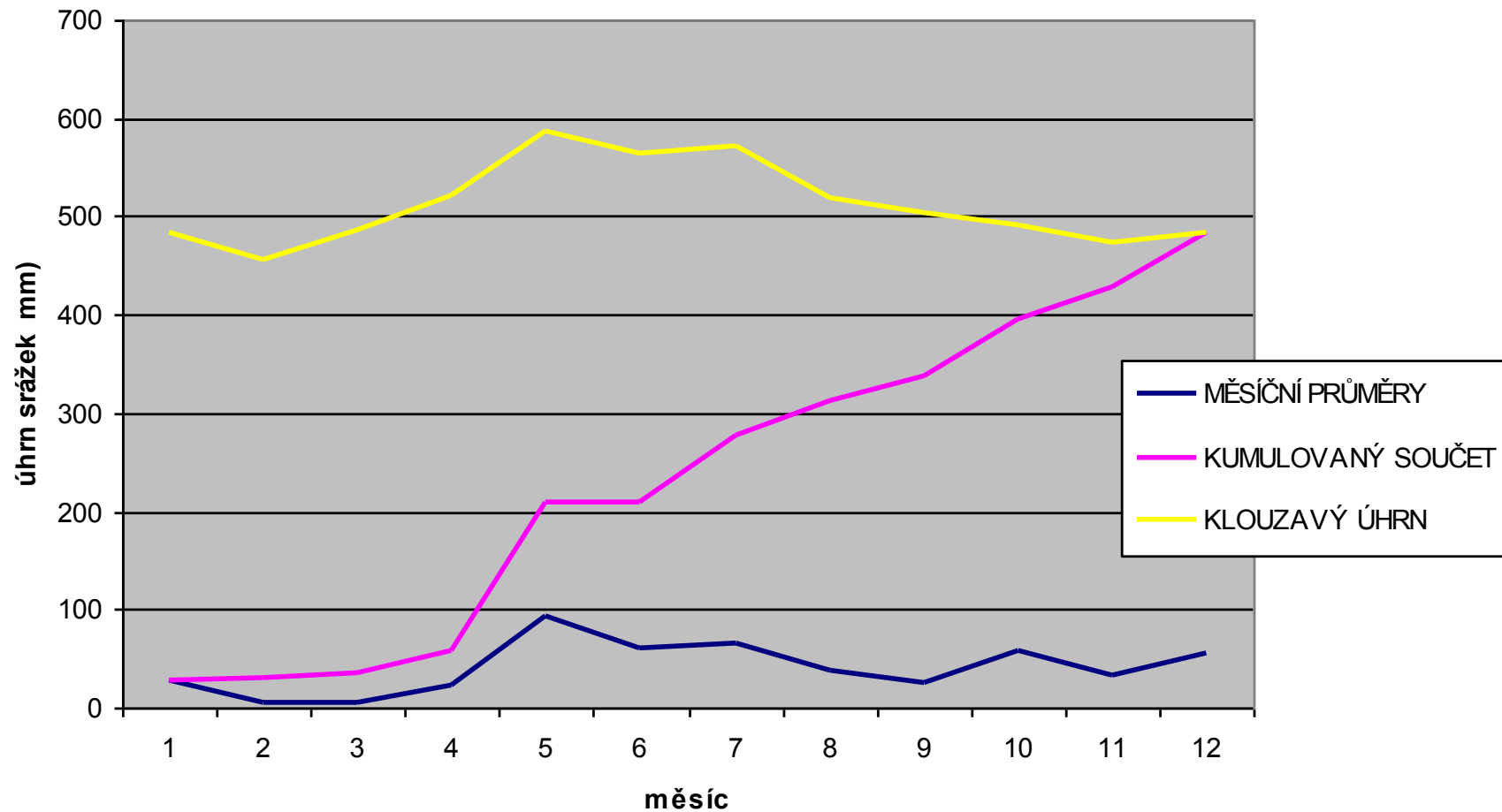
< Zpět

Další >

Dokončit

# Z - diagramy

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



# Testy a pís. zkoušky

- Příprava na test - pojmy viz. uč. Materiály na IS
- Pís. zkouška, vždy min jedna přípravná otázka ke st. zkoušce:



# Bc otázky

- Popište vývoj a pojetí statistiky. Objasněte základní kroky statistických šetření a vymezte následující pojmy: hromadný jev, statistický soubor, statistická jednotka, statistický znak.
- 16) Popište a vysvětlete rozdělení četností (absolutní, relativní, kumulované) statistických jevů jejich grafické znázornění a uveďte příklady geografických aplikací.
- 17) Uveďte příklady grafické znázornění statistických informací (grafy, sítě, statistické mapy) a na konkrétních příkladech doložte jejich náležitosti.
- 18) Přehledně představte základní míry úrovně (střední hodnoty), variability, asymetrie a špičatosti. Vysvětlete jejich význam a doložte nákresy.
- 19) Popište teoretická rozdělení náhodných veličin (Gaussovo, binomické a Pearsonovo rozdělení). S využitím nákresů popište a interpretujte význam normálního rozdělení. Uveďte další příklady jiných rozdělení.
- 20) Vysvětlete a popište statistické závislosti náhodných veličin (korelace regrese, korelační koeficient). Uveďte konkrétní příklady geografických aplikací.
- 21) Popište obecné principy statistického testování (základní kroky, hypotézy, teorie a druhy testů) a uveďte příklady geografických aplikací. Vysvětlete, proč se formuluje nulová hypotéza, co je hladina významnosti a jak se nulové hypotézy mohou ověřit.
- 22) Vysvětlete pojem časové řady a uveďte jejich hlavní rysy. S pomocí náčrtu objasněte tendence časových řad. Uveďte příklad využití řetězových a bazických indexů – doložte příkladem výpočtu.

# Ukázky části otázek k pís. zk., zdroje: Literatura viz. IS, učební texty na IS :historie a Thinking G.), zápisy z přednášek, cvičení

- Vysvětlete pojem primární data a napište hlavní zdroje primárních dat pro geografický výzkum
- Vysvětlete pojem sekundární data, napište hlavní zdroje sekundárních dat pro geografický výzkum
- Čím se geografové zabývají?( Na jakých pracovních pozicích a s jakým obsahem pracovní činnosti.)Udejte min dva konkrétní příklady a vysvětlete, jak využívají své odbornosti geografa a znalosti geografických konceptů.
- Jmenujte základná fáze vědeckého výzkumu a stručně je popišteDefinujte systém, objasněte s využitím náčrtu.
- Které dovednosti patří pod pojem „geografické myšlení“?
- Jmenujte typické geografické otázky, které jsou potřebné pro správnou výuku zeměpisu a rozvoj geografického myšlení.
- Definujte geografii
- Zařaďte geografii do souboru věd. Definujte objekt studia geografie a předmět studia geografie.
- Definujte a vymezte objekt studia geografie a předmět studia geografie. Dejte do souvislosti předmět studia geografie a typické geografické otázky.
- Systém geografických věd – napište strukturovaný výčet či nakreslete schéma.
- Stručně popište nejstarší etapu vývoje statistiky.
- Kdy se o statistice objevují první zprávy? Které činnosti se k ní vztahovaly?
- Co bylo obsahem politické aritmetiky?
- Definujte statistiku
- Definujte Hromadný jev, udejte min. dva konkrétní příklady
- Jmenujte a následně definujte či vysvětlete dvě základní pojetí statistiky
- Napište základní etapy statistického zpracování dat
- Definujte statistickou jednotku včetně vymezení. Uveďte jeden konkrétní příklad vymezení statistické jednotky
- Definujte statistický znak, napište základní členění st. znaků, uveďte vždy u každé dílčí skupiny min. jeden příklad
- Definujte statistický soubor, napište základní typy st. souborů, uveďte k nim vhodné příklady.
- Na konkrétním příkladu objasněte termíny: statistický soubor jednorozměrný a vícerozměrný.
- Na konkrétním příkladu objasněte termíny: statistický soubor základní a výběrový.
- Na konkrétním příkladu objasněte termíny: statistický soubor základní a reprezentativní.  
jmenujte základní typy grafů.
- Načrtněte trojúhelníkovou síť pro hodnocení nejvyššího dosaženého vzdělání a vyznačte v ní obec A s podíly obyvatel se základním vzděláním 30 %, se středním 50,% a vysokoškolským 20% a obec B s podíly (10 %, 60 %, 40%).
- Vysvětlete, pro které jevy lze použít kruhovou síť. Uveďte min. dva příklady.
- Načrtněte kruhovou síť pro znázornění průměrných srážek na met. stanici.
- Kartogram a kartodiagram, uveďte vhodné příklady pro jejich využití , objasněte, kdy je vhodné použít kartogram a kdy kartodiagram
- A další.....

