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 - pojetí

- pojem statistika – běžně ve dvou významech:
  - **1. praktická činnost** (zaznamenání, třídění, shrnování číselných údajů o skutečnostech)
  - **2. teoretická disciplína** , předmět zkoumání : stav a vývoj číselně vyjádřených hromadných jevů
- statistika se zabývá **hromadnými jevy** tj. jevy, které se vyskytují u souboru lidí, věcí, událostí buď v kvantitativní formě nebo i kvalitativní formě převoditelné na číselnou
- hromadné jevy – příklady:
  - věk osob, ……studenti dopíší další příklady….

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

Jsou tři druhy lži: lež prostá,
lež odsouzeníhodná a statistika.

Statistika je zvlášť rafinovanou formou lži.

Co je typické pro statistiku

- Zkoumá **hromadné** jevy.
- Zabývá se proměnlivými - **variabilními** - vlastnostmi.
- Pracuje s čísly a vyjadřuje se pomocí čísel - zajímá se především o **kvantitativní stránku** reality.
- Používá výpočetní techniku k vytváření a správě statistických **databází**, k provádění hromadného **zpracování** a **analýzy** dat a ke **komunikaci.**

Co statistika „umí“

- **Zjišťování** (počet domácností ČR, počet pracovníků v odvětví XY)
- Popis **struktury** (věková struktura obyvatel ČR, roční chod hodnot meteorologických prvků)
- **Shrnování** dílčích ukazatelů v čase a prostoru (průměrná nezaměstnanost v regionu)
- **Srovnávání** agregovaných ukazatelů v čase nebo prostoru (trend vývoje počtu obyvatelstva, teploty vzduchu dvou lokalit)
- Měření **závislosti** (závislost mezd na HDP, závislost met. prvku na nadmořské výšce).

… a co statistika „neumí“:

Statistika selhává, pokud:

- Nemá k dispozici adekvátní číselné údaje
- Nemá-li k dispozici dostatečně rozsáhlý soubor případů
- Není-li v datech přítomna proměnlivost (variabilita).

II.STATISTIKA jako vědní disciplína

- **statistika popisná**
  - popisuje jev statistickými charakteristikami
  - takto zpřehledňuje velké množství dat – shrnují je do kategorií (průměr, nejčetnější hodnota, grafické znázornění dat)
  - využívá numerické a grafické metody
- **statistika dynamická**
  - hledá pravidelnosti, souvislosti, vývoj, usuzuje z části na celek
  - **matematická statistika**
  - usuzování na závěry o sledovaném jevu z malého vzorku (zkoumání veřejného mínění, namátkový test), tj. z chování části usuzujeme na chování celku zobecňuje výsledky (odhad a testování hypotéz) - používá **počtu pravděpodobnosti**

Významy pojmu STATISTIKA

I. Statistika jako praktická činnost

- Statistická evidence (např. sběr údajů, třídění, sumarizace apod.),
- Instituce, která tuto evidenci provádí (např. ČSÚ, ministerstva aj.)
- Souhrn údajů o nějaké skutečnosti (statistika nezaměstnanosti, ročenka meteorologických pozorování atd.)

Základní etapy statistického zpracování dat

- 1. **Zjišťování/ Sběr údajů**

- shromáždění a zaznamenání údajů, jejich kontrola aj.,

  - periodicita sběru:
    - a) periodické (např. 1* ročně)
    - b) běžné – krátké, pravidelné lhůty
    - c) jednorázové

- **2., 3., 4.**

Základní etapy statistického zpracování dat

- **2. <u>Zpracování</u> -** uspořádání, seskupení, shrnování, sumarizace,
- **3. <u>Analýza</u>** - výpočet charakteristik, měření závislostí, srovnávání, měření dynamiky
- **4. <u>Prezentace</u>** <u>výsledků</u> - tabulkové či grafické vyjádření a slovní zhodnocení výsledků předcházejících etap.

- **Druhy statistického zjišťování:**
  - **výkaznictví** - nejběžnější
  - **soupisy** – rozsáhlá zjišťování na rozsáhlých souborech k určitému okamžiku – např.sčítání obyvatelstva
  - **statistický odhad** - subjektivní hodnocení
  - **anketa** – šetření určité vrstvy lidí na urč. problematiku

Základní dělení statistických údajů

- podle zdroje — **primární** a **sekundární**,
- podle reálnosti situace — **skutečné** a **simulované**,
- podle periodicity zjišťování — **průběžné**, **periodické** a **jednorázové**,
- podle časového hlediska — **okamžikové** a **intervalové**.
- podle použité škály měření – **nominální, ordinální, intervalové, poměrové**

1.2 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:

Velikost datových souborů:

Snadný přenos dat:

Jakmile se jednou data dostala do počítače, lze je snadno přenést elektronicky (například pomocí Internetu) na jiné místo.

Nevýhody počítačového zpracování

- **Chyby v softwaru.**
- **Univerzálnost.**

Může vést k výběru nevhodné metody zpracování. Je velmi důležité, aby každý, kdo používá statistický software, si byl <u>vědom úrovně svých statistických znalostí a užíval pouze ty metody, kterým rozumí.</u> Pozor na používání neznámých statistických metod.

- **Černá skříňka.**

Počítač vzdaluje uživatele od dat i metody zpracování..

- **Špatná data plodí špatné závěry.**

Vymezení základních statistických pojmů

- **hromadné jevy** tj. jevy, které se vyskytují u souboru lidí, věcí, událostí buď v kvantitativní formě nebo i kvalitativní formě převoditelné na číselnou

Hromadné jevy: Jevy, které jsou výsledkem působení velkého množství příčin

Některé jevy, které v geografii studujeme pomocí statistických metod mají povahu jevů náhodných (hydrologické jevy či meteorologické jevy).

Statistická jednotka: je to určitý jev či prvek, který je předmětem statistického šetření a pro který se zjišťují údaje

Statistická jednotka musí být přesně vymezena na počátku vlastního šetření a to z hlediska **věcného, časového, prostorového.**

Příklady:

Statistický znak: je to určitá vlastnost statistické jednotky, kterou se snažíme postihnut. Tzv. **shodné (společné) znaky** vymezují příslušnost statistické jednotky k určitému statistickému souboru. Ostatní jsou znaky **proměnlivé (variabilní)**.

Příklady:

Statistický soubor: skupina statistických jednotek stejného druhu (věcně, prostorově a časově vymezených)

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. **identifikační znaky** - umožňující určit, zda prvek do daného statistického souboru patří nebo nepatří. Identifikační znaky tedy statistický soubor vymezují.

Z hlediska cílů statistického zkoumání sledujeme na prvcích statistického souboru jednu nebo více vlastností - **sledované znaky**.

Je-li vlastnost měřitelná v nějakých jednotkách, jde o kvantitativní znak, jinak jde o kvalitativní znak.

Statistický znak: je to určitá vlastnost statistické jednotky, kterou se snažíme postihnut.

Tzv. **shodné (společné) znaky** vymezují příslušnost statistické jednotky k určitému statistickému souboru.

Ostatní jsou znaky **proměnlivé (variabilní).**

Příklady:

Statistické znaky lze dělit na znaky

- A) **prostorové**
- **B) časové**
- **C) věcné:**

1. **kvalitativní:**
   - **alternativní**
   - **možné**

2. **kvantitativní:**
   - **spojité**
   - **diskrétní/nespojité**

stat. jednotka

novorozenec

místo narození: Brno

datum: 2.3. 2006

pohlaví:muž

národnost:česká

délka v cm: 55

Doplňte další příklady

Statistické znaky můžeme získat :

přímo – (např. měřením) – **primární data**

nepřímo (výpočtem). (znaky odvozené) – **sekundární data**

Statistický soubor

Statistický soubor: **skupina statistických jednotek stejného druhu** (věcně, prostorově a časově vymezených)

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. **identifikační znaky**

- **sledované znaky –** statisticky šetříme .

Příklad:
statistický soubor Novorozenci v ČR
identifikační znak: novorozenost
sledovaný znak: váha, živý, pohlaví

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 podmnožinou základního souboru. Je vytvořen ze statistických jednotek, vybraných podle určitého hlediska.

Př. Novorozenci v Jihomoravském kraji

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

Př. šetření průzkum volebních výsledků, peoplemetry

Rozsah statistického souboru:

počet statistických jednotek v souboru:

N – rozsah základního souboru

n – rozsah výběrového souboru

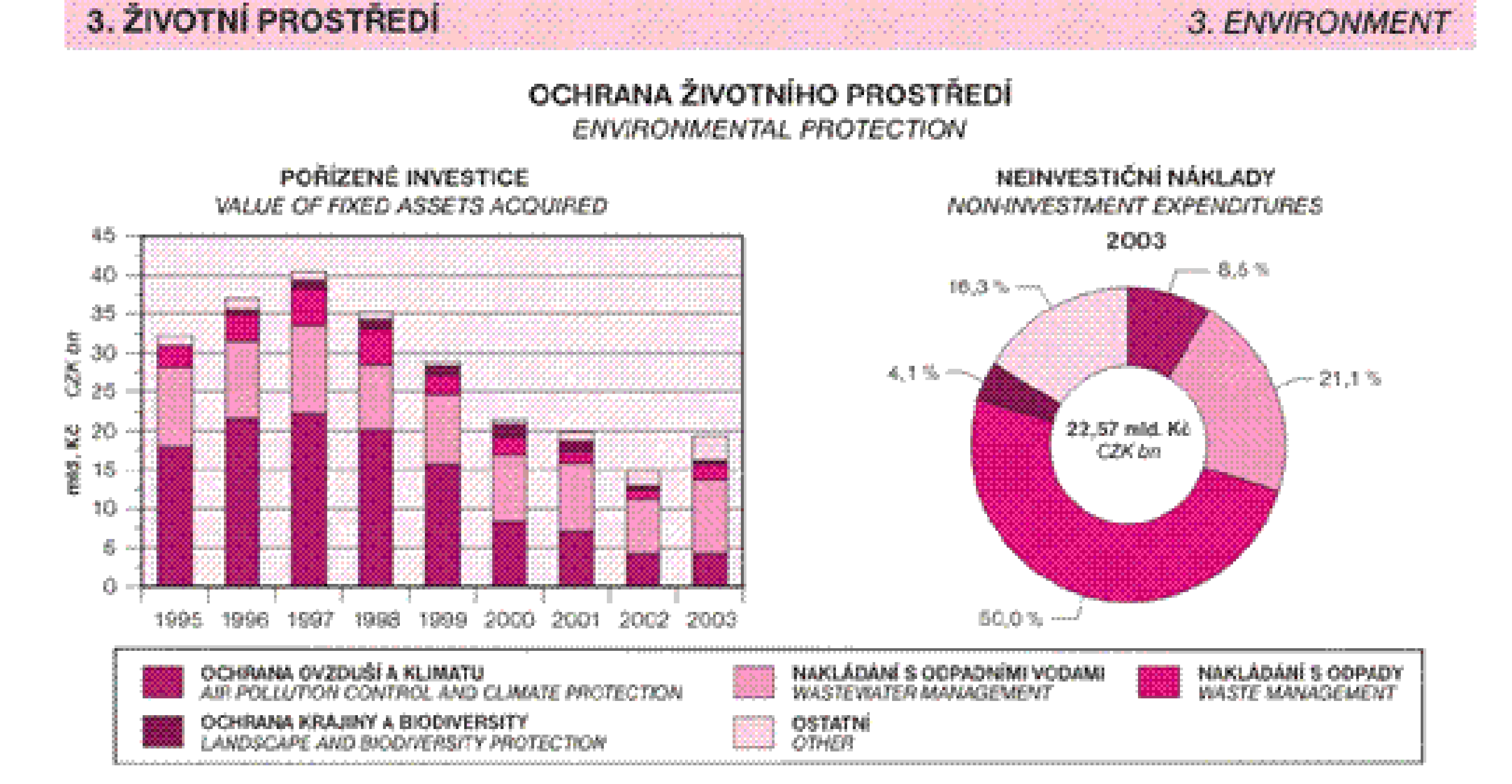
Grafické znázornění jevů

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.legenda/klíč
- 5.zdroj údajů
- vysvětlivky, poznámky,

Graf – ukázka

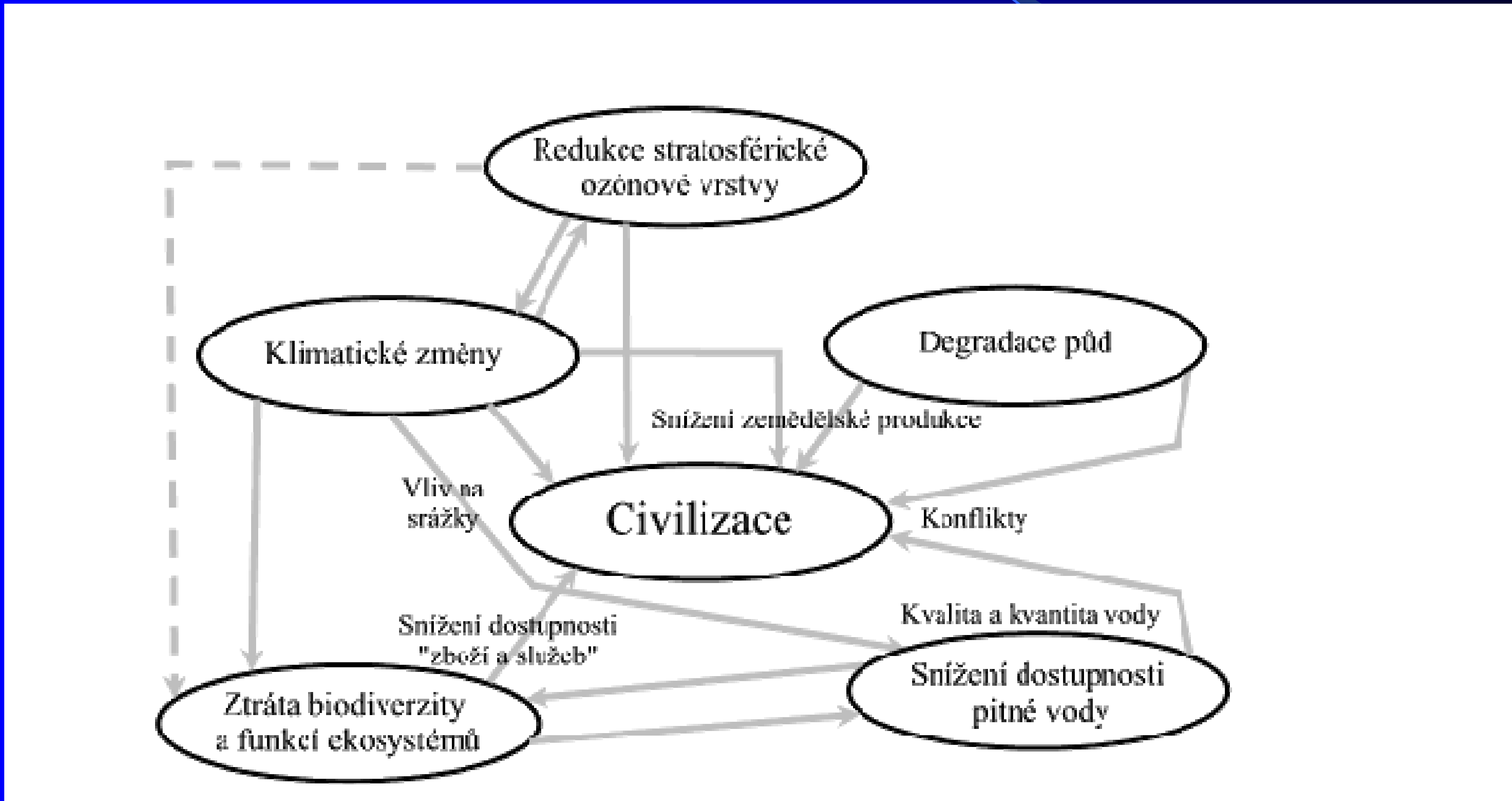


Český statistický úřad, 2003

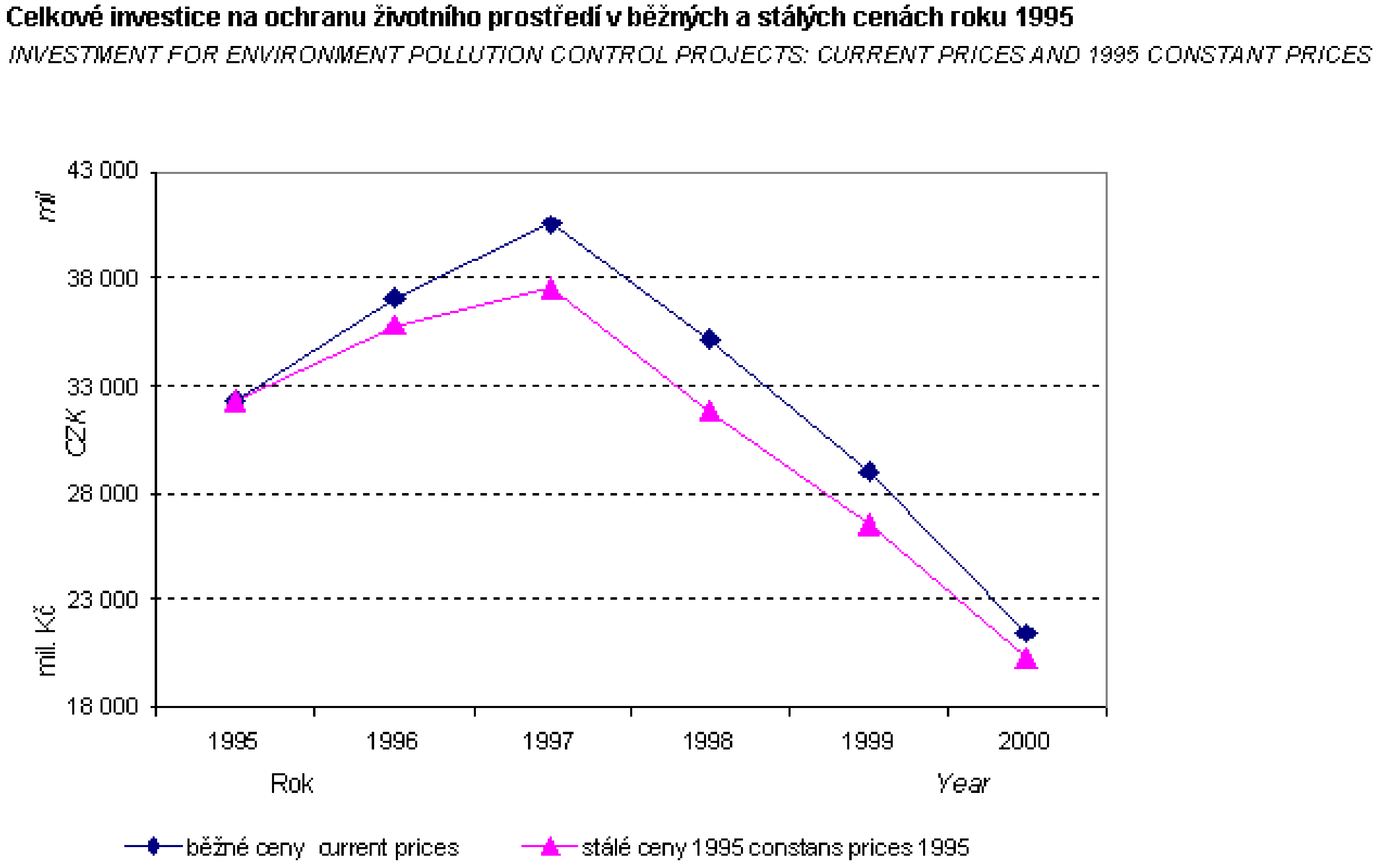
Typy grafů

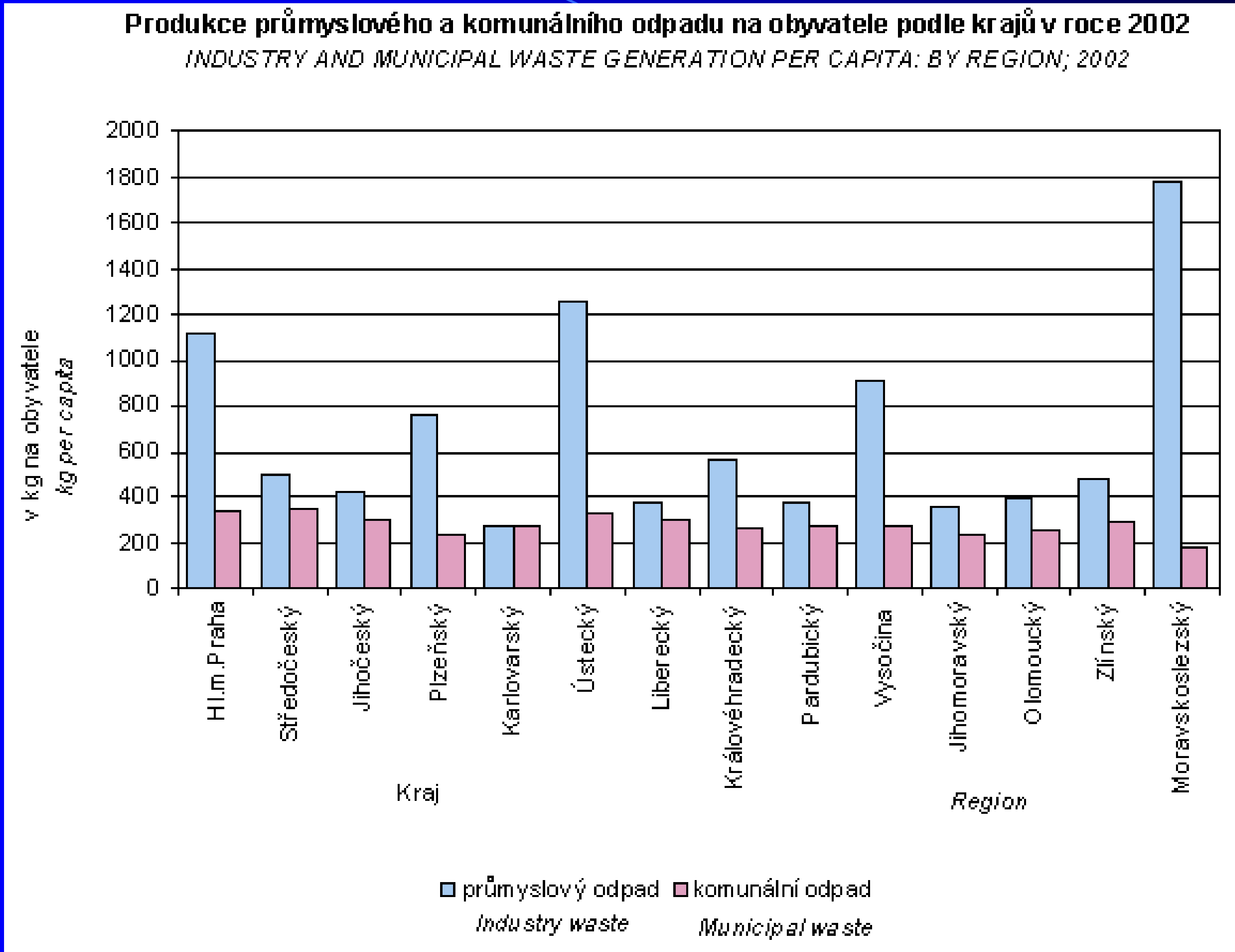
- **schéma** – znázorňuje strukturu a vztahy jevu či procesu
- Příklad
- **diagram** – znázorňuje kvantitativní údaje o souboru
  - sloupcový, bodový, plošný atd.
- příklad
- **statistická mapa** – prostorové rozložení prvku v podkladové mapě

schéma


Redukce stratosférické ozónové vrstvy
Klimatické změny
Degradace půd
Snížení zemědělské produkce
Vliv na srážky
Civilizace
Konflikty
Snížení dostupnosti "zboží a služeb"
Kvalita a kvantita vody
Ztráta biodiverzity a funkcí ekosystémů
Snížení dostupnosti pitné vody

Diagram


Celkové investice na ochranu životního prostředí v běžných a stálých cenách roku 1995
INVESTMENT FOR ENVIRONMENT POLLUTION CONTROL PROJECTS: CURRENT PRICES AND 1995 CONSTANT PRICES
43 000
38 000
33 000
28 000
23 000
18 000
mil
CZK
mil. Kč
1995
1996
1997
1998
1999
2000
Rok
Year
běžné ceny current prices
stálé ceny 1995 constans prices 1995



Český statistický úřad, 2001

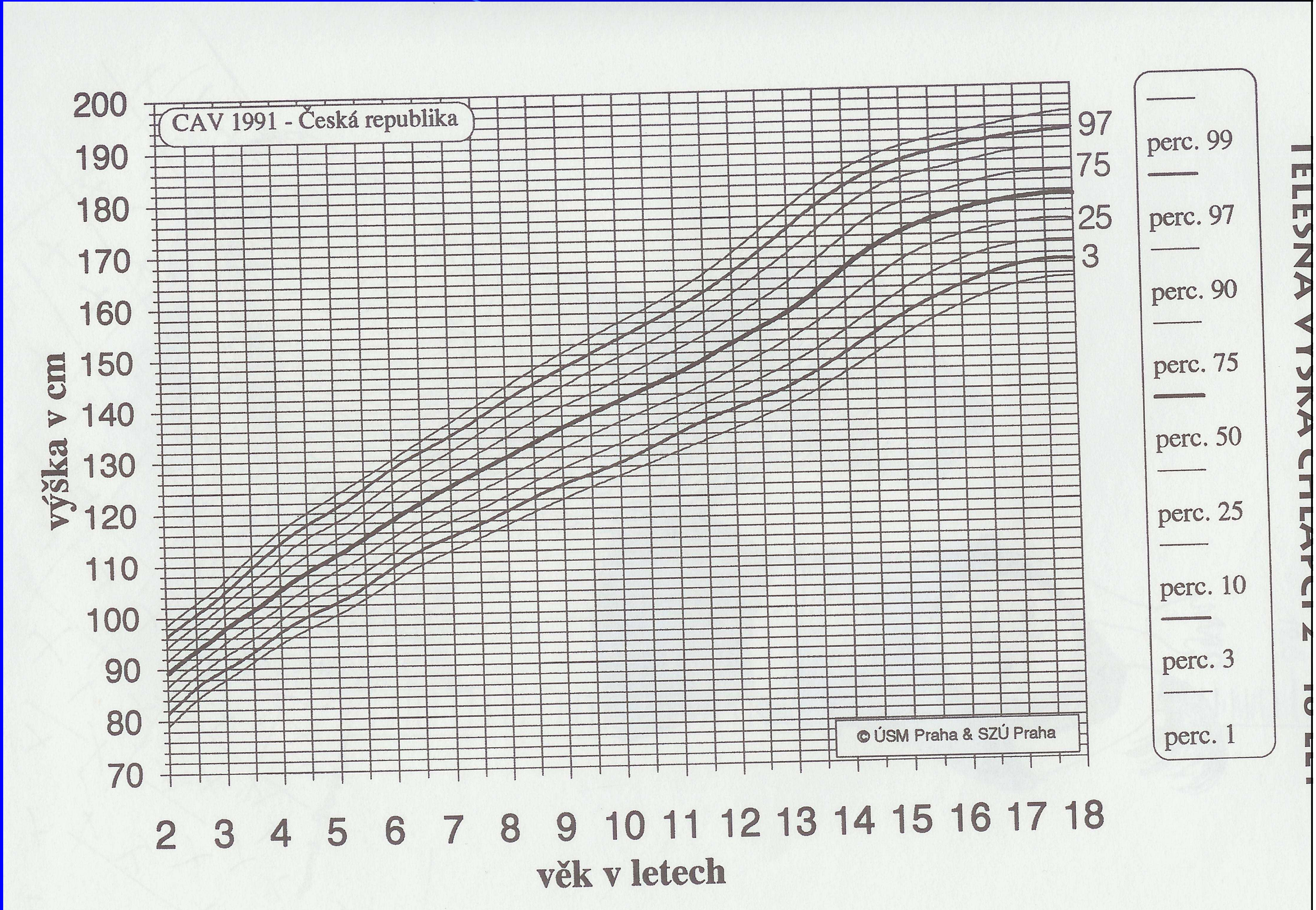
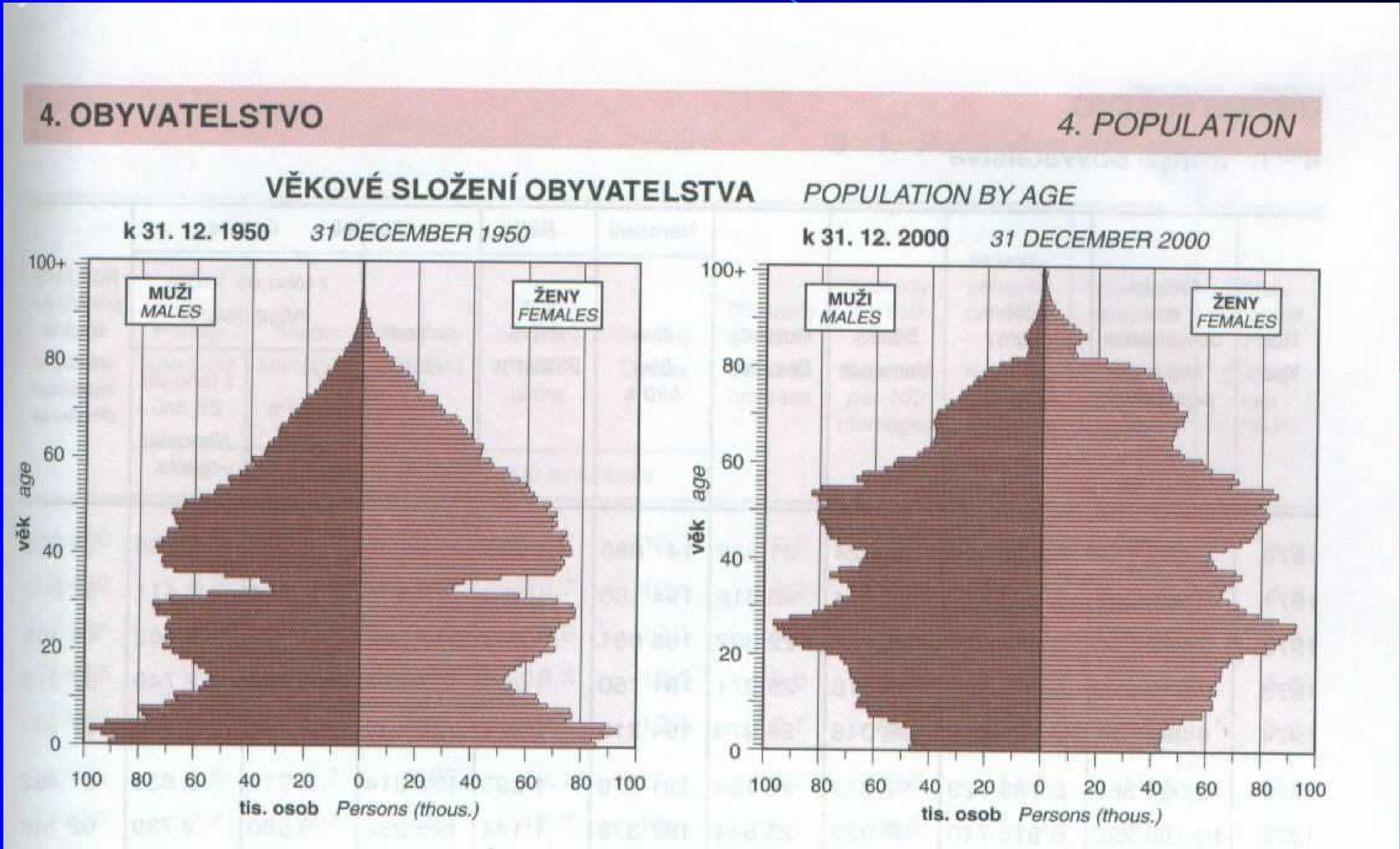
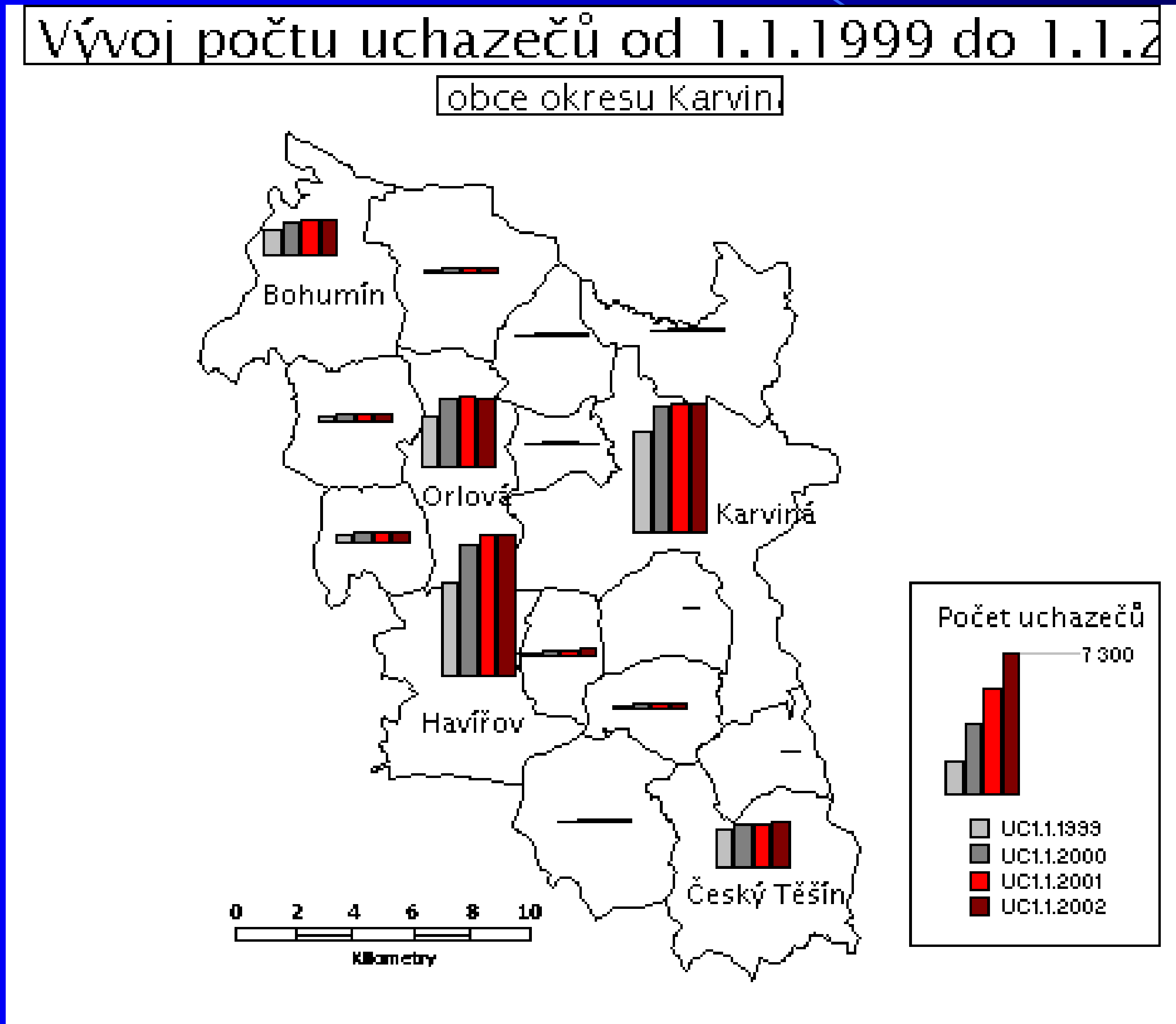

CAV 1991 - Česká republika
výška v cm
věk v letech
200
190
180
170
160
150
140
130
120
110
100
90
80
70
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
97
75
25
3
perc. 99
perc. 97
perc. 90
perc. 75
perc. 50
perc. 25
perc. 10
perc. 3
perc. 1
© ÚSM Praha & SZÚ Praha

Diagram_ - věkové složení obyv., tzv.pyramida života



Český statistický úřad, 2003

Statistická mapa



okresní úřad Karviná, 2003

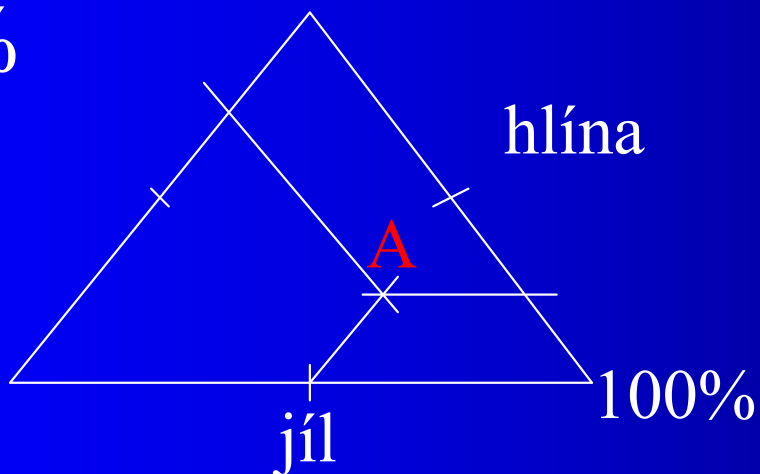
Použití grafických papírů při studiu geografických jevů

Grafický papír usnadňuje vynášení prvků do grafu.

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

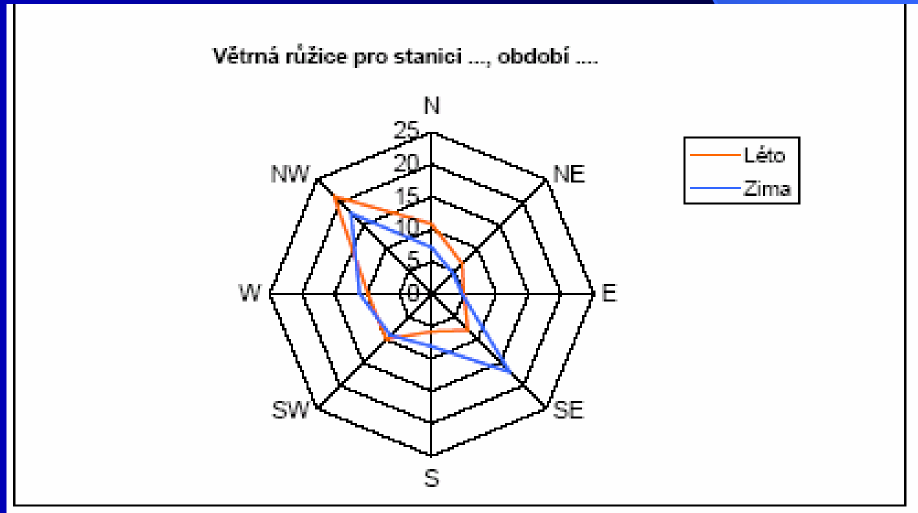
Sítě

- <u>Trojúhelníková síť</u> – znázorňování jevů o třech prvcích, které mají vždy konstantní součet
- např. půdní druhy
- půda A:: % jílu 30, % hlíny, 50%, písku 20%



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

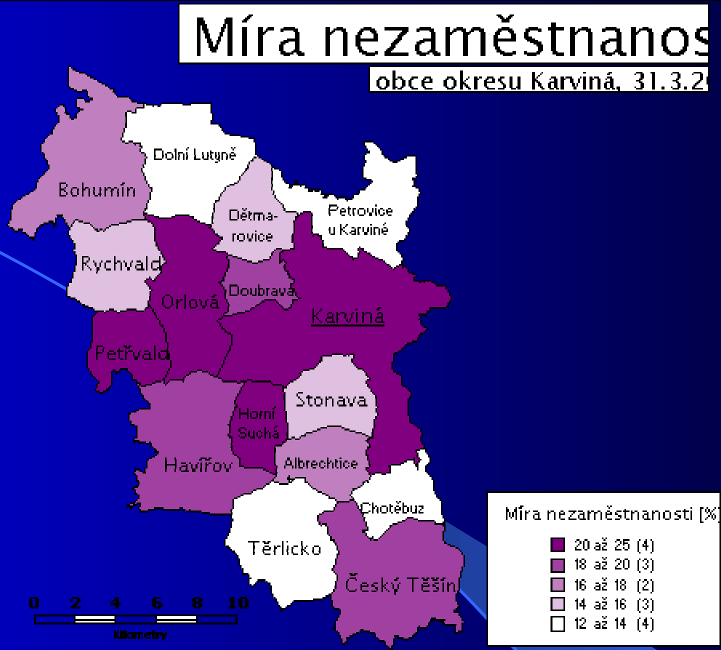


statistická mapa:

kartogram

kartodiagram

kartogram

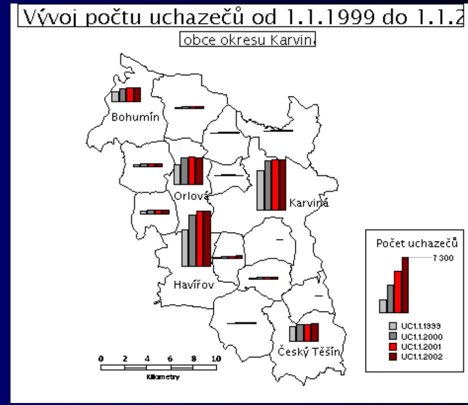


Kartogram je <u>obrysová kartografická kresba územních celků</u>, ve kterých jsou grafickým způsobem (barevný odstín, rast) <u>plošně znázorněna statistická data</u> týkající se různých geografických jevů (lidnatost, využívání ploch apod.)

Kartogramy lze rozdělit podle územního dělení na:

- kartogramy s geografickými hranicemi
- kartogramy s geometrickými hranicemi

Kartodiagram

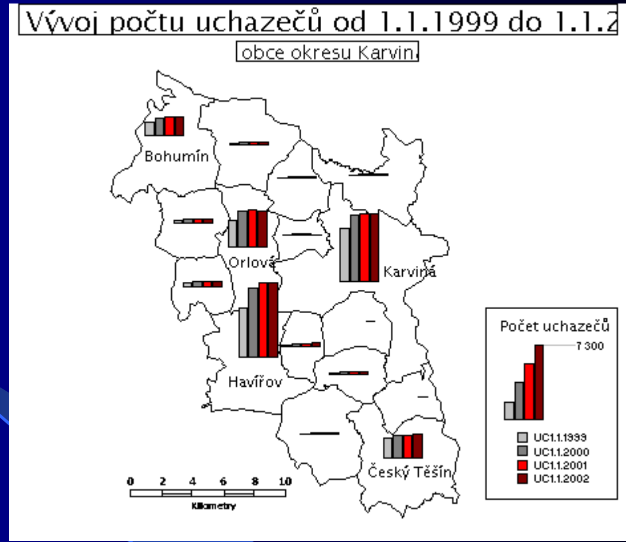


Kartodiagramy jsou diagramy vložené do mapové kostry, kterou tvoří dílčí územní celky.

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:

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

Grafické metody analýzy geografických jevů

- 1.znázornění intenzity jevu v prostoru
- a) absolutními metodami
- *značková metoda (velikost značky odpovídá velikosti jevu)
- * bodová metoda (počet prvků….velikost jevů)
- b) relativními metodami (např. šrafování-hustota obyv.)

- 2.znázornění struktury jevu v prostoru
- využití výsečových grafů
- *pouze strukturu vyjádříme výsečovými grafy se stejným poloměrem
- *strukturu a velikost celku (výsečový graf + velikost poloměru odp. velikosti jevu)

Náležitosti statistické mapy

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

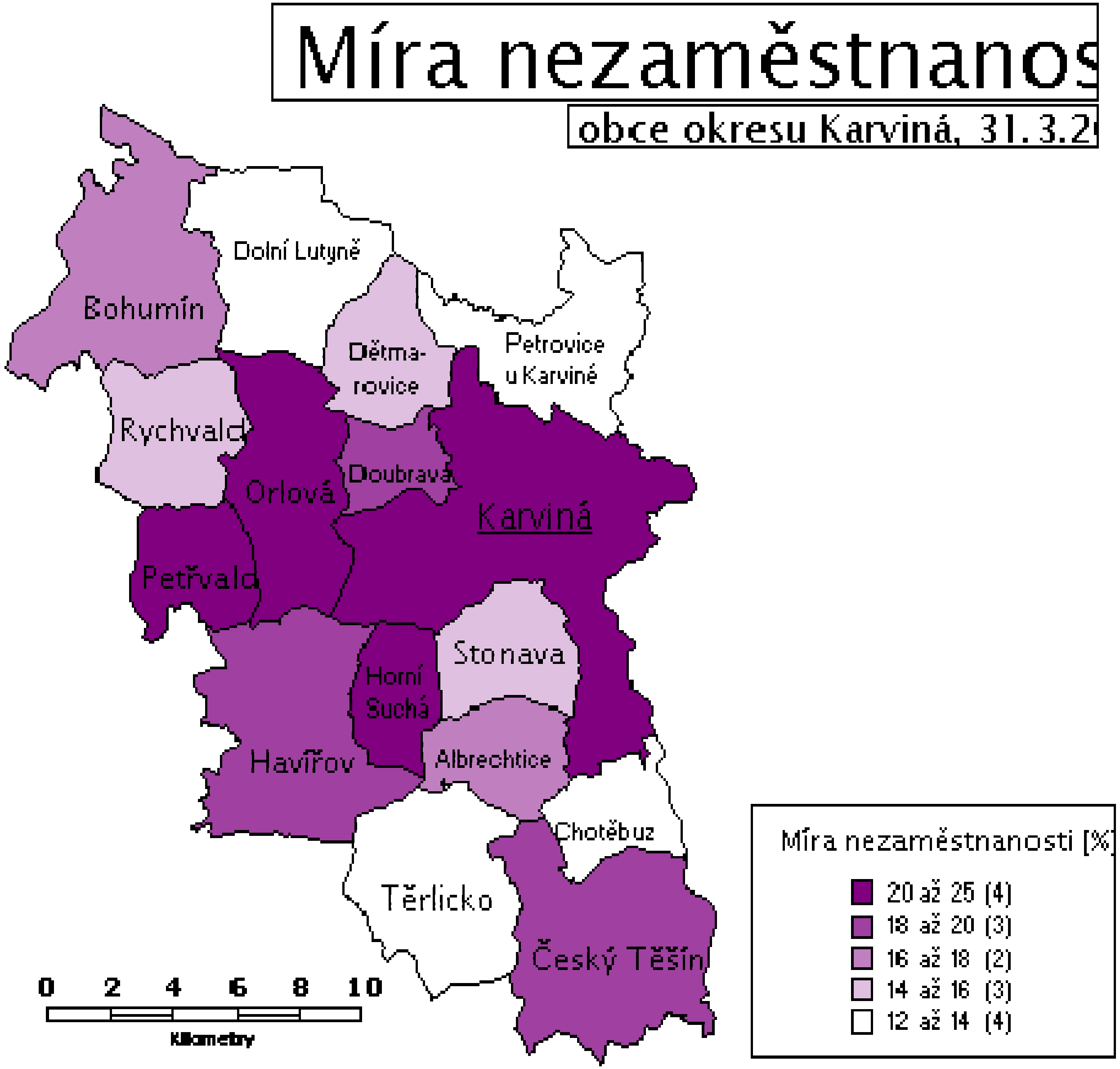
Základní údaje tvoří

- **Název mapy** - stručně a výstižně charakterizuje zobrazené území, druh mapy
  lze i název hlavní a vedlejší)
- **Mapový rámec – „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 -příklady
- **Autoři**

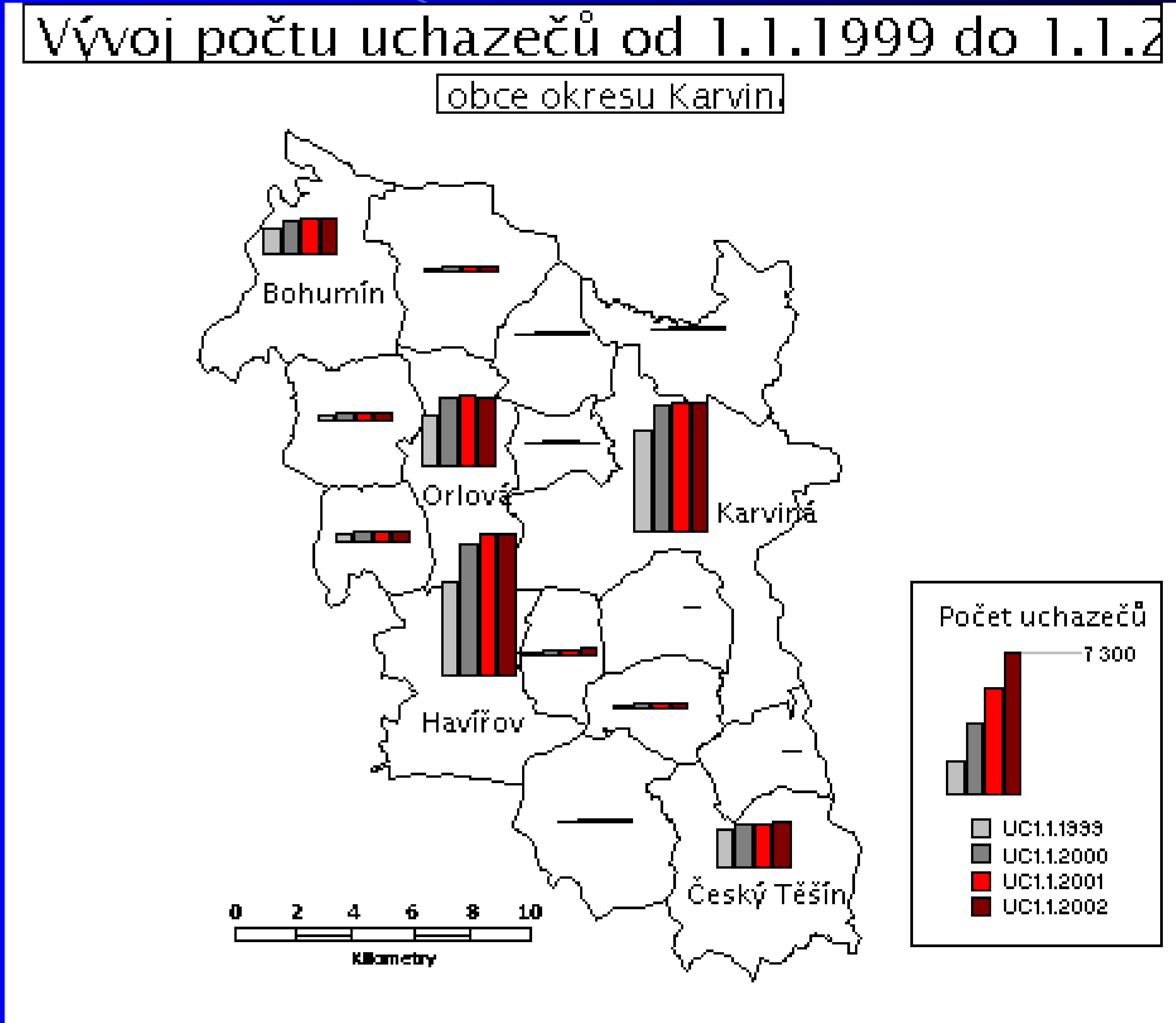
Dalšími údaji mohou být :

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

Hledejme chyby



Hledejme chyby



Izolinie – konstrukce a vlastnosti

- Izolinie – čáry, které v grafu spojují body se stejnou intenzitou (velikostí, hodnotou) jevu
- získávají se metodou prostorové interpolace hodnot vynesených do grafu
- plynulé čáry
- izobary, izotermy, vrstevnice atd.
- Konstrukce izolinie - příklad

Četnosti

Rozdělení četností

Absolutní, relativní kumulované četnosti

- četnost – počet výskytu určité hodnoty v souboru, frekvence hodnoty
- rozdělení četností – počty prvků s určitými hodnotami statistického znaku, obvykle pro nespojité hodnoty
- skupinové rozdělení četností - počty prvků s hodnotami statistického znaku, které patří do určitého intervalu, obvykle pro spojité hodnoty

skupinové rozdělení četností

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

zásady:

- vymezené hranice pro jednoznačné zařazení prvků
- obvykle stejná šířka
- přiměřený počet intervalů

- <u>absolutní četnost</u> – počet jednotek v intervalu
- <u>relativní četnost</u> – podíl četností na rozsahu souboru
- <u>kumulovaná četnost</u> – počet jednotek s hodnotami menšími nebo rovny horní hranici intervalu
- příklad

Tab.S Skupinové rozdělení četností,
ukázka – příklad váha 100 novorozenců v JMK

Interval	střed	abs. č.	relativ. č.	kumul. abs.	kumul. relat.
500 - 1000	750	1	1%	1	1%
1001 - 1500	1250	5	5%	6	6%
1501 - 2000	1750	15	15%	21	21%
atd.					
		100	100%		

Grafické znázornění rozdělení četností

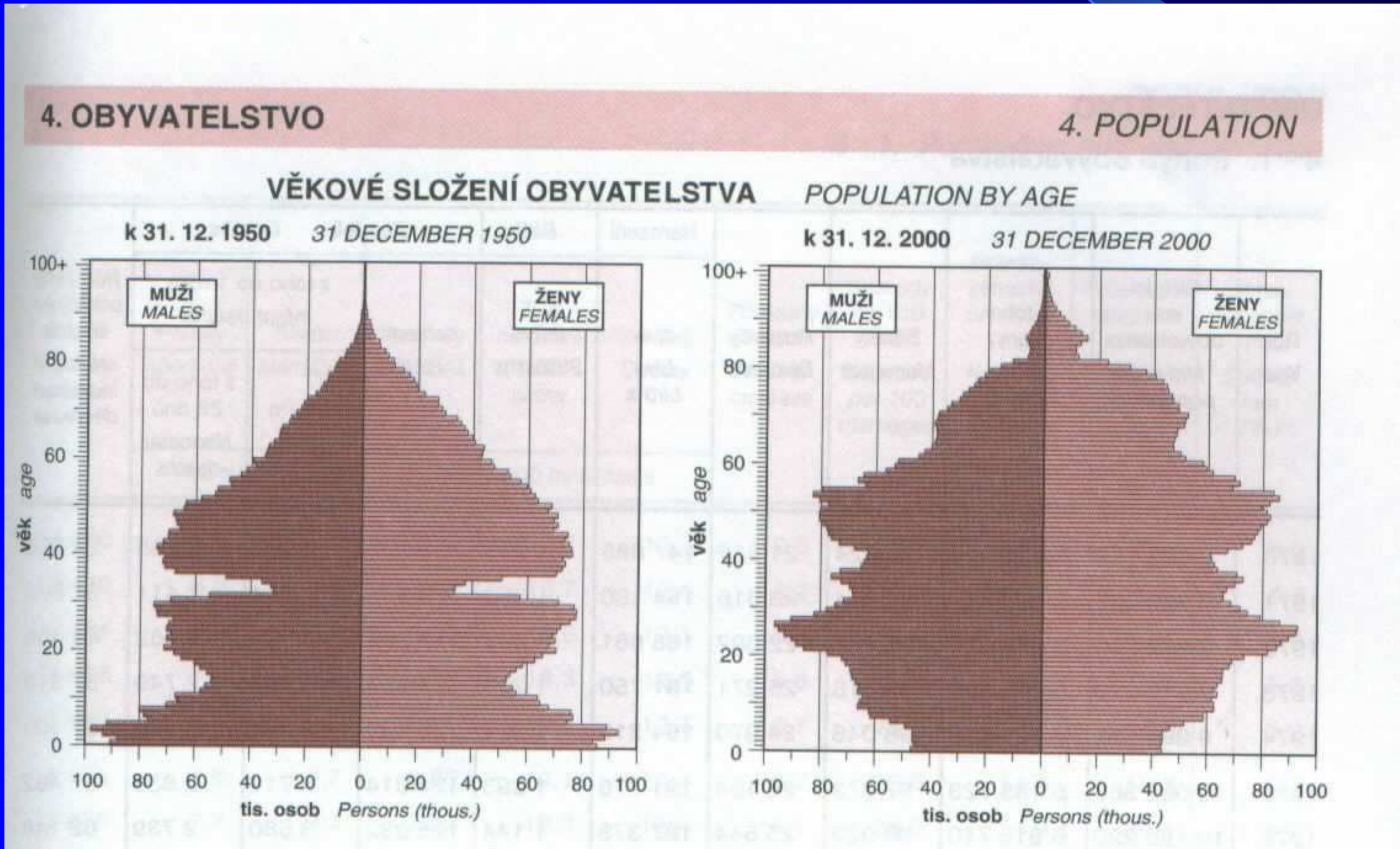
- histogram
- polygon
- čára kumulovaných četností

<u>Histogram</u> – sloupcový diagram,
šířka sloupce – šířka intervalu, výška sloupce - četnost

<u>Polygon</u> – spojnicový diagram,
hodnoty četnosti se vynáší ke středům intervalu

<u>čára kumulovaných četností</u> – součtová čára,
graf kumulované četnosti, vždy 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 jednotlivých hodnot u jednorozměrného statistického souboru používáme často střední hodnoty
- 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 – jednovrcholové rozdělení četností + blízký nejčetnější hodnotě

Vážený aritmetický průměr

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

Modus

- modus - nejčetnější hodnota kvantitativního znaku ve studovaném souboru
- významný především u souboru nespojitých veličin

- modální interval – interval zahrnující největší počet jednotek, závisí však na stanovení hranic intervalů
- rozdělení s více mody – polymodální rozdělení

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

Medián a kvantily

- Medián – tzv. prostřední hodnota, je to prvek řady uspořádané v neklesajícím pořadí (od nejm. po největší), který ji dělí na dvě poloviny, které mají menší a větší hodnotu znaku
- POZOR: soubor je třeba vždy uspořádat
- pořadí prvku (kolikátý prvek to je, hodnota prvku je medián!) určují vzorce :
- pro řadu s lichým počtem prvků (n+1)/2,
- pro řadu o sudém počtu je medián průměr z hodnot mezi prvkem na (n/2) a (n/2+1) místě
- Příklad

Kvantily

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

obdobně

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

kvantily obecně široké použití ve statistice a v geografii

Geografický medián

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

Cvičení

- Název: Výpočet středních hodnot jednorozměrného statistického souboru
- Cíl cvičení: naučit se vypočítat základní charakteristiky úrovně – střední hodnoty st. souboru
- Úkoly:

1. 1.Vytvořte jednorozměrný statistický soubor,
2. 2.definujte konkrétní hromadný jev, statistickou jednotku a její určení a statistický znak
3. 3.Zpracujte tabulky skupinového rozdělení četností pro soubory
4. 4.Zpracujte odpovídající histogramy včetně všech základních prvků (viz. přednáška!)
5. 5.Zpracujte součtovou čáru pro oba soubory
6. 6.Vypočítejte (ručně) včetně uvedeného postupu střední hodnoty pro oba soubory – tj. aritmetický průměr, modus, aritmetický střed, medián a kvartily
7. 7.Oba soubory podle charakteristik střední úrovně porovnejte a vyslovte závěry

Charakteristiky variability

- variační rozpětí
- kvantilové odchylky
- průměrné odchylky
- rozptyl
- směrodatná odchylka
- variační koeficient

Variační rozpětí

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

Kvantilové odchylky

- Založeny na kladných odchylkách jednotlivých sousedních kvantilů

- např. kvartilová odchylka
- vzorec
- decilová odchylka
- vzorec
- percentilová odchylka

Průměrné odchylky

- průměrná odchylka je definována jako aritmetický průměr jednotlivých hodnot znaku od vybrané střední hodnoty (tj. od aritmetického průměru, mediánu, modu apod.)

Střední diference

- je def. jako aritmetický průměr absolutních hodnot všech možných rozdílů jednotlivých hodnot sledovaného znaku
- v praxi vhodná pouze pro malé soubory

Rozptyl a směrodatná odchylka

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

Variační koeficient

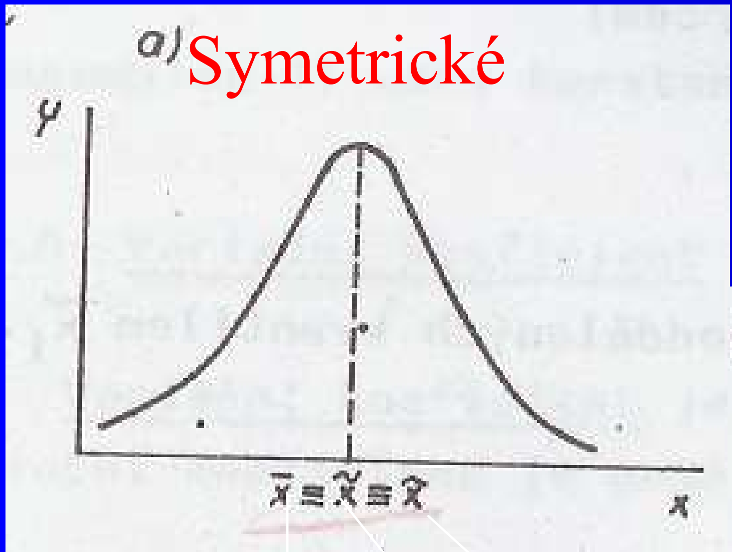
- je častou používanou relativní mírou variability
- je definován jako poměr směrodatné odchylky k aritmetickému průměru

Charakteristiky asymetrie

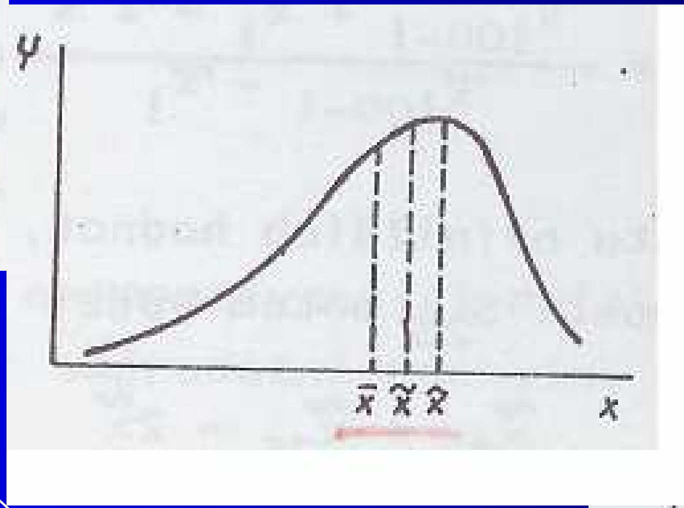
- Charakteristiky asymetrie (míry šikmosti) jsou čísla dávající představu o souměrnosti tvaru rozdělení četností
- míra šikmosti pro souměrné rozdělení je nula
- pro nesouměrné je kladná nebo záporná

Charakteristiky asymetrie

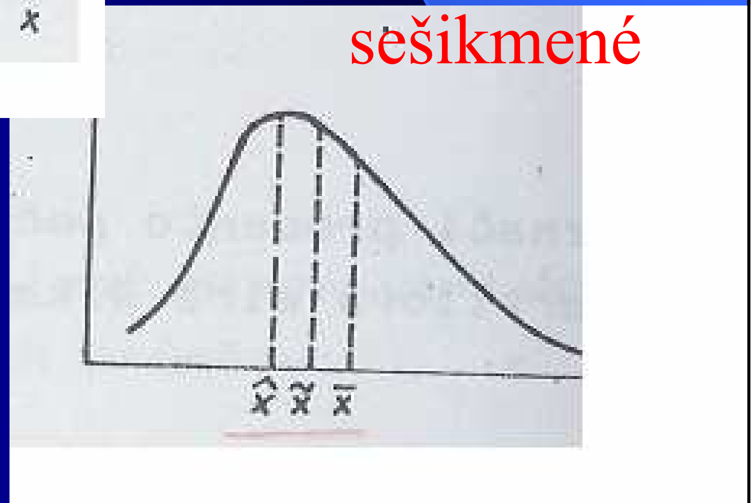
Symetrické



Záporně sešikmené



Kladně sešikmené



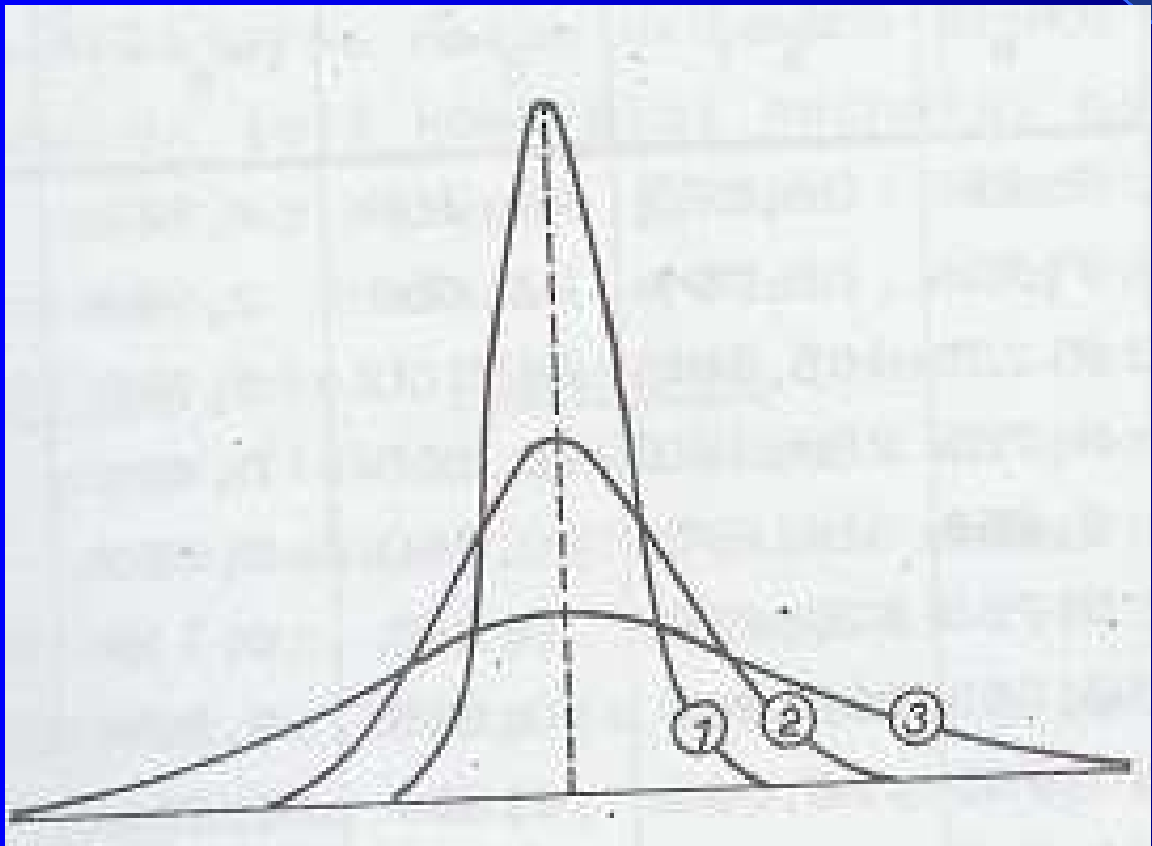
ar. průměr, medián, modus

charakteristiky špičatosti

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

Obr. Špičaté, normální a ploché rozdělení

charakteristiky špičatosti



1 – špičaté

2 – normální

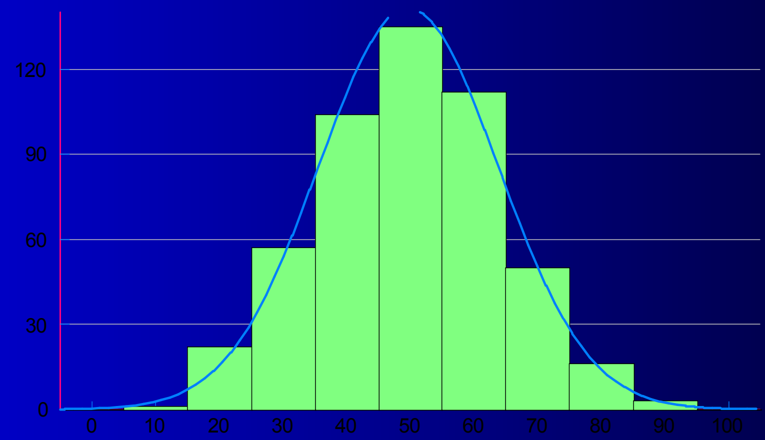
3 – ploché

rozdělení

Teoretická rozdělení

Teoretická rozdělení

- Základní pojmy
- náhodná veličina spojitá (teplota) a nespojitá (počet měsíců s teplotou nad…)
- 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.



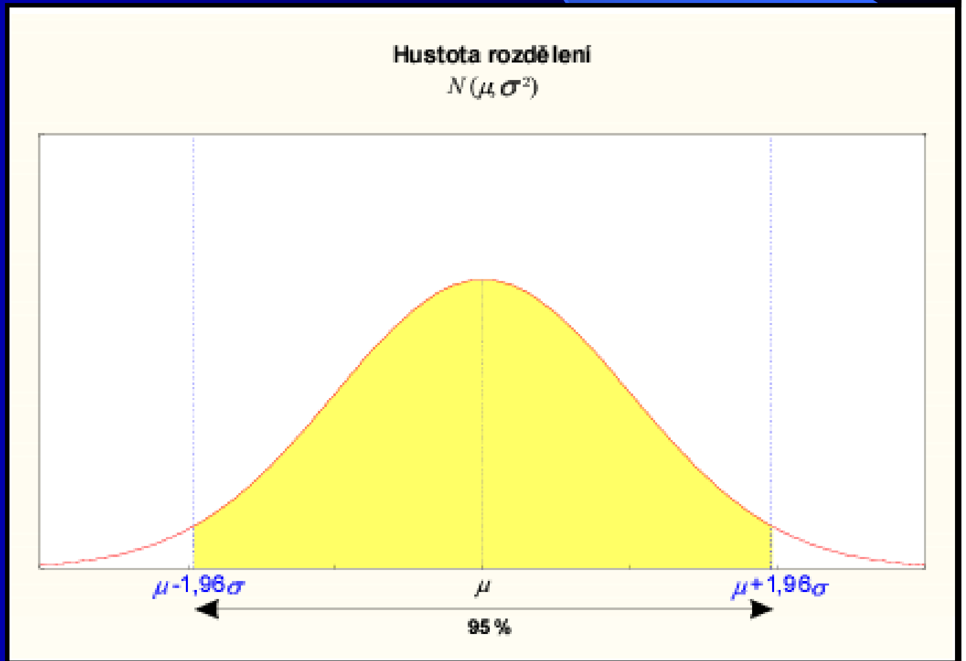
Normální rozdělení / Gaussovo, Laplaceovo- Gaussovo

- **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 / Gaussovo, Laplaceovo- Gaussovo

- Normální křivka a osa x vymezují plochu 100%, tj. 1
- lze stanovit pravděpodobnosti, s nimiž leží hodnoty v určitém intervalu, hranice intervalu tvoří průměr a násobky směrodatné odchylky
- obr.



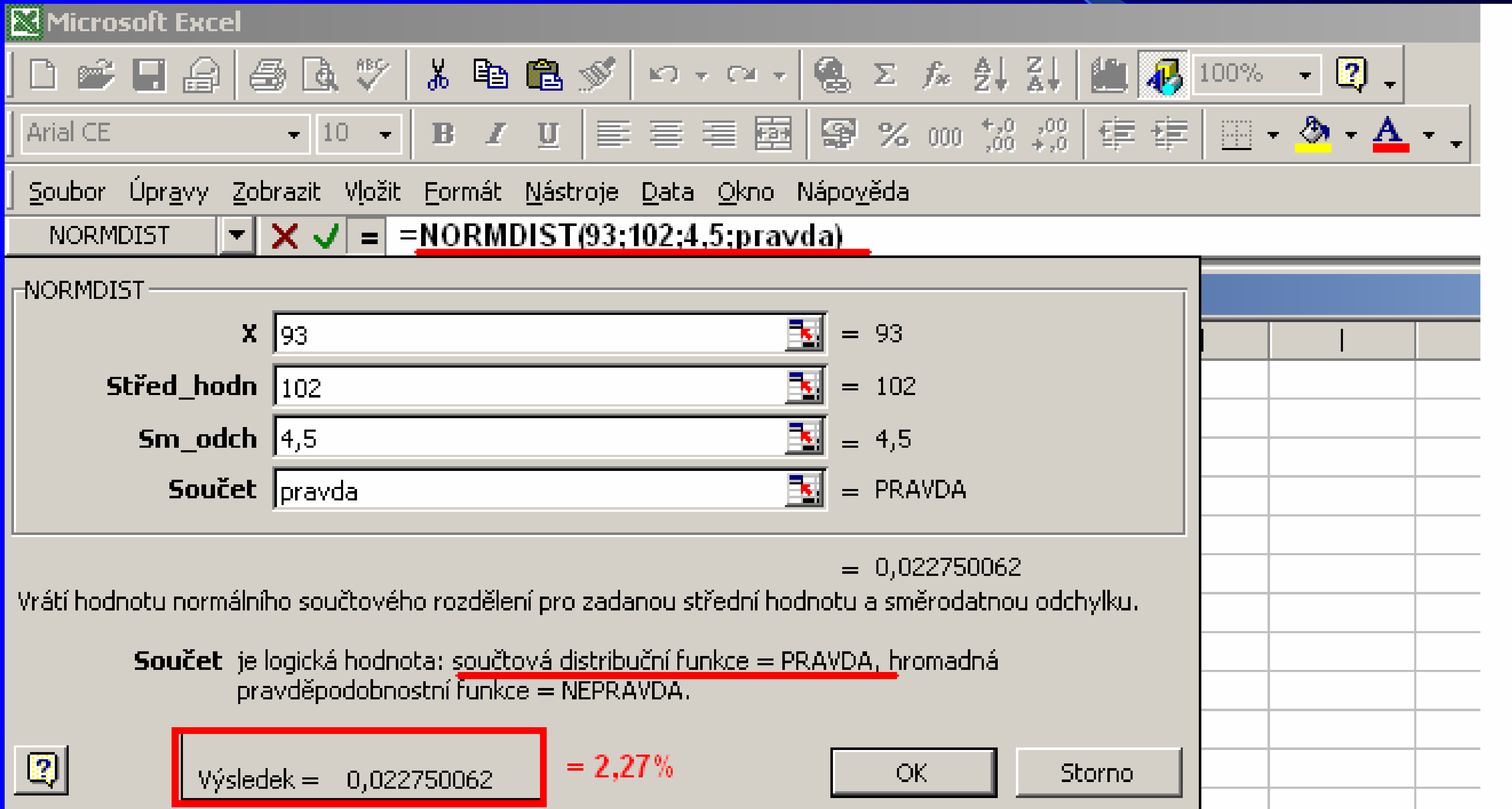
příklady

Příklad 1

- 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.

Řešení 1

- Pravděpodobnost, že výška nabude hodnoty menší nebo rovné 93 cm, je vyjádřena hodnotou distribuční funkce F (93) pro parametry normálního rozdělení 102;4,5



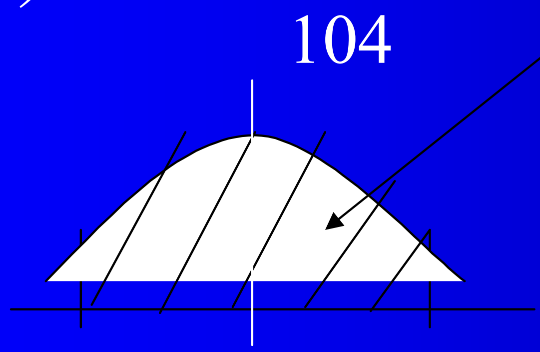
Odpověď: 2,27 % chlapců ve věku 3,5 – 4 roky je menších než 93 cm

Příklad 2

- Psychologickými testy bylo zjištěno, že hodnota IQ mužské populace je náhodnou veličinou s normálním rozdělením, jehož střední hodnota je 104 a směrodatná odchylka 8.
- Určete hodnotu IQ, kterou podle uvedených pravděpodobnostních předpokladů:
- a) dosáhnou 3 / 4 mužské populace,
- b) nepřesáhne 5% mužské populace,
- c) překročí 5% mužské populace.
- d) Odhadněte v jakých mezích se pohybuje IQ 99.9 % mužské populace.

Řešení 2a)

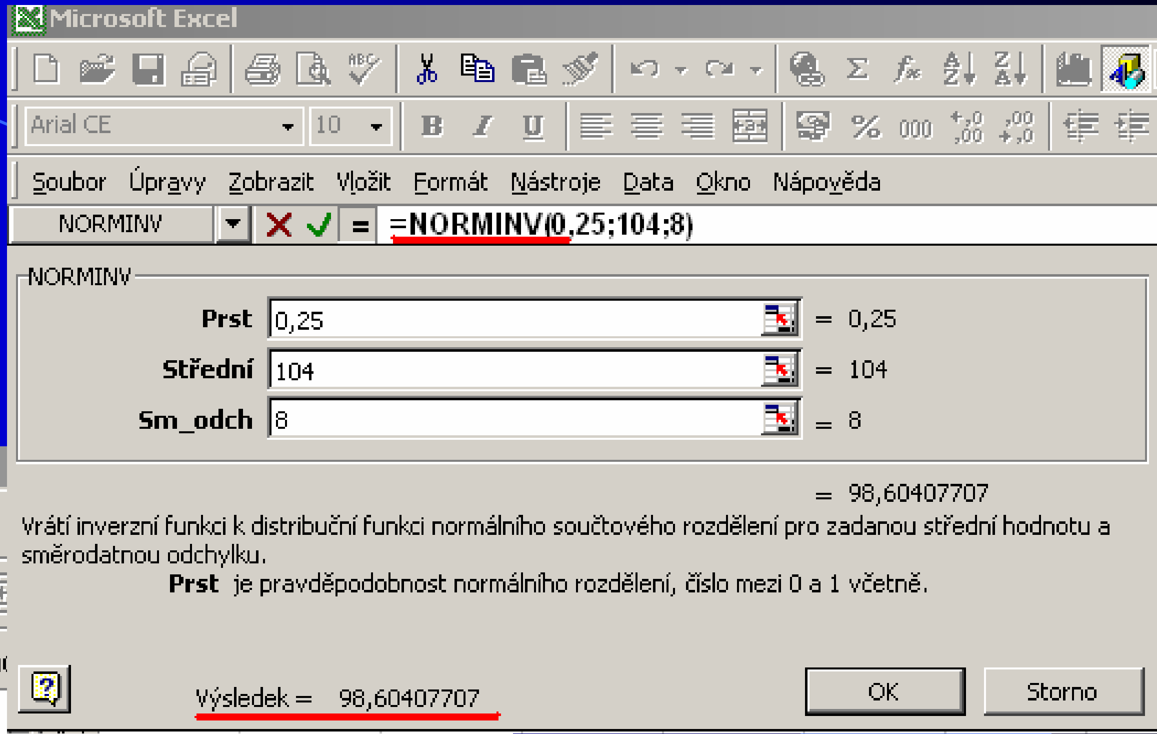
- *a) dosáhnou 3 / 4 mužské populace*

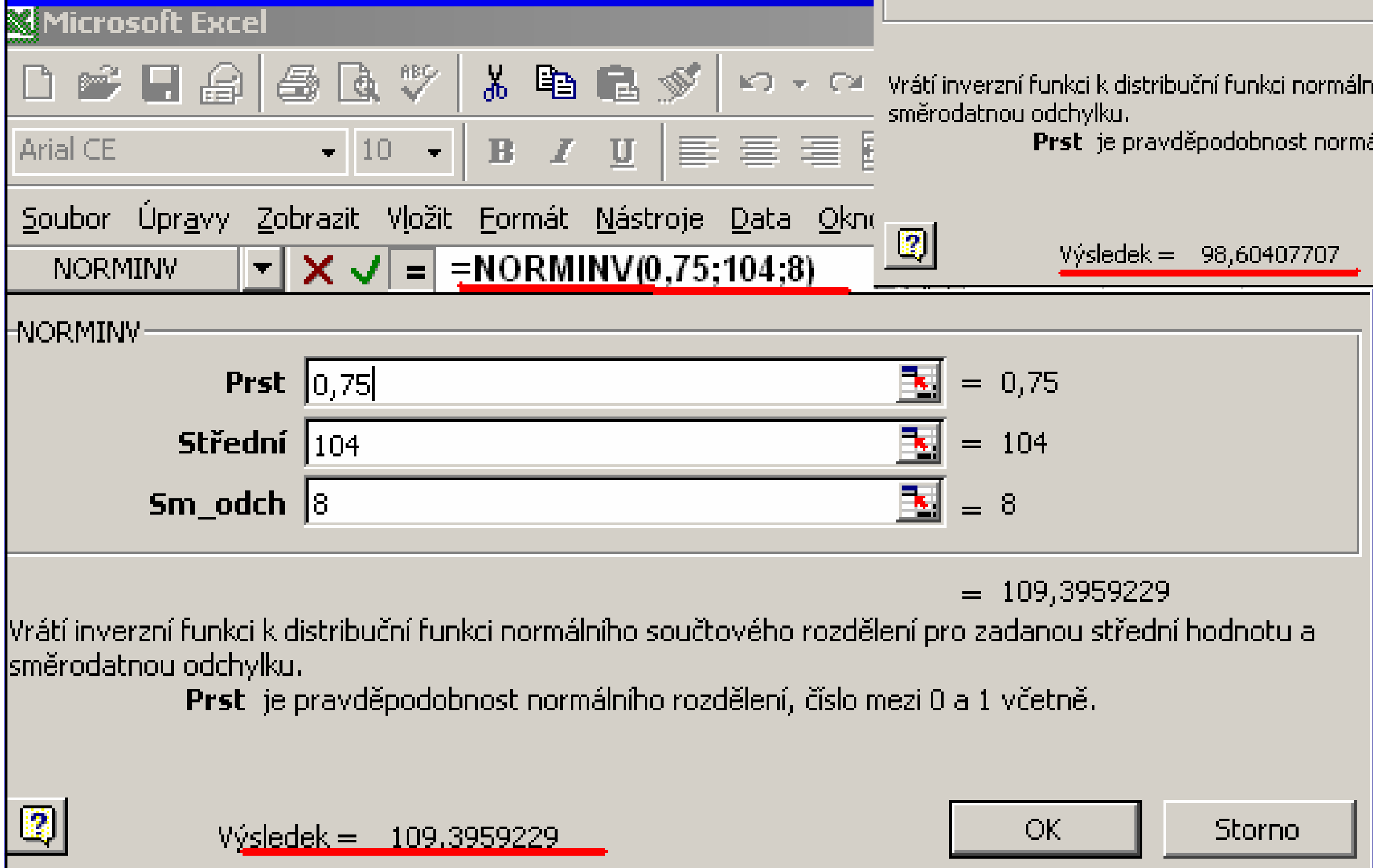


Hledáme dolní a horní meze intervalu (hodnot IQ),
ve které se bude nacházet 75% mužské populace, tj 1. a 3. kvartil

Řešení 2a)

Excel, statistická funkce inverzní k e Gauss. - NORMINV

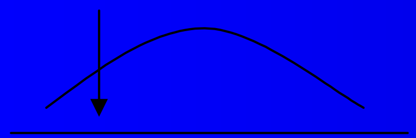




Podle parametrů daného normálního rozd 75 % mužské populace má IQ v intervalu 98,6 a 109,4.

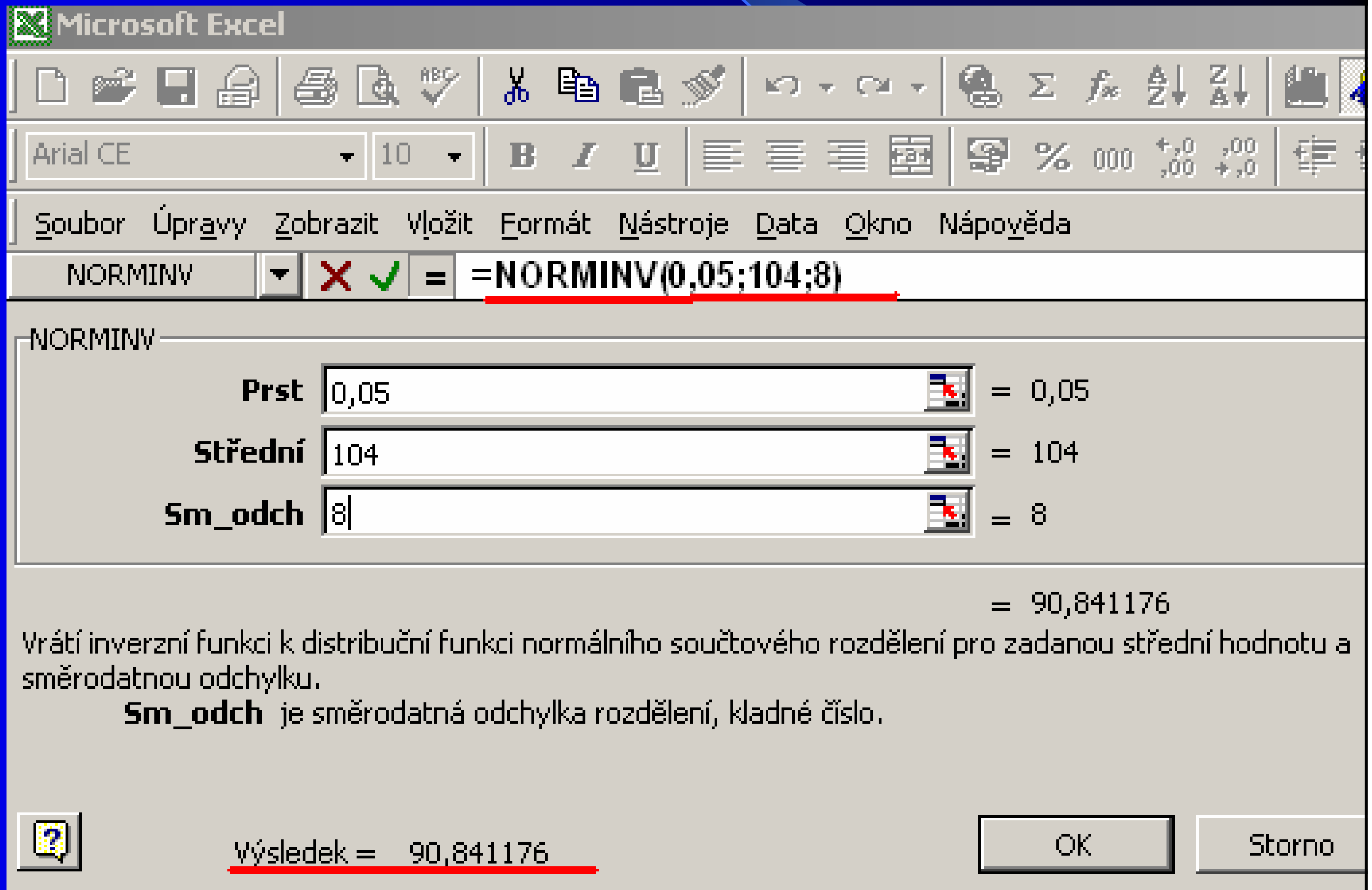
Řešení 2b)

- *b)   hodnotu IQ, pod níž je 5% mužské populace*   tj. 5% dosáhne max. ….IQ



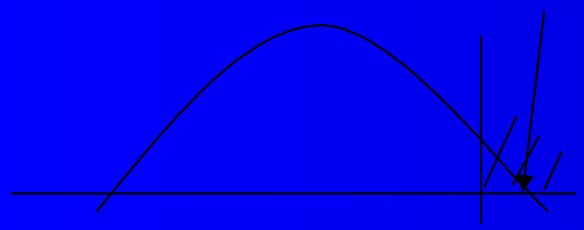
5 %

5 % mužské populace má IQ
(dle parametrů N)
nižší nebo rovno
90,84.



Řešení 2c)

- c) překročí <u>pouze </u>5% mužské populace



- analogicky s 2b) nebo využít symetrie normálního rozdělení a využít výsledku 2b)

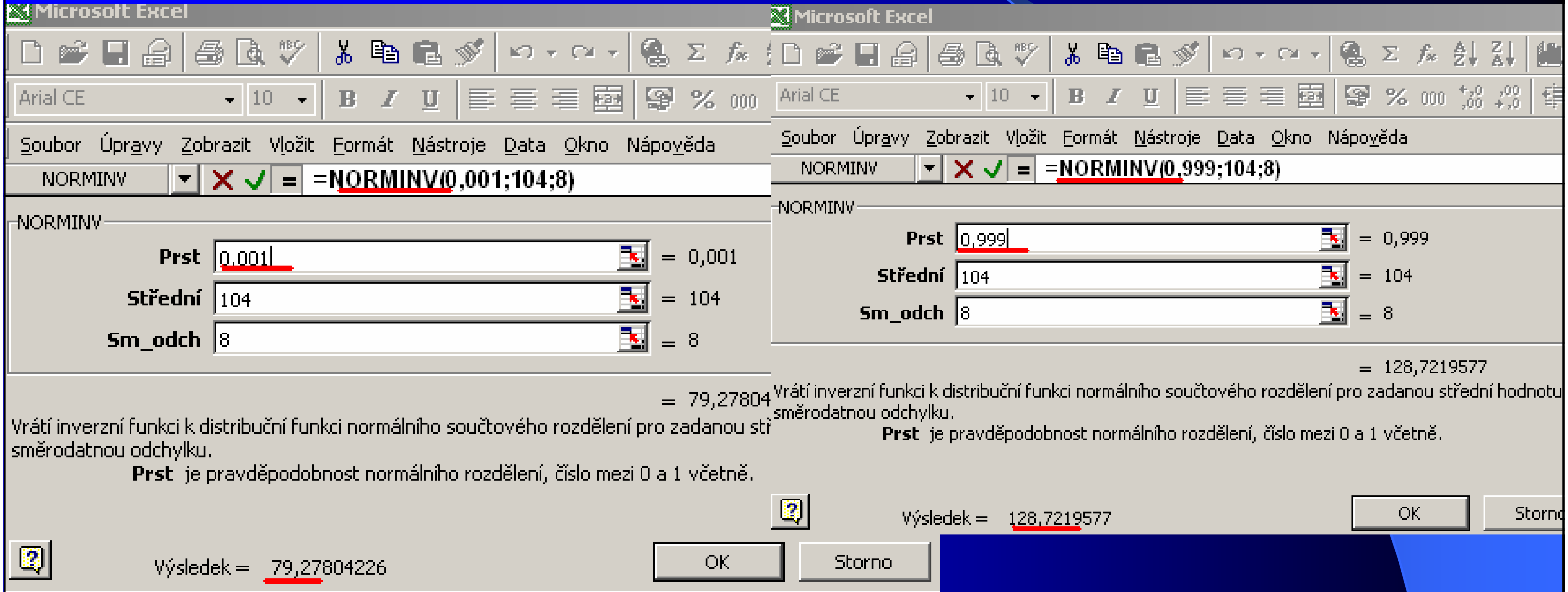
pak 104 – 90, 84 = 13,16

104 + 13,16 =117,16

5% mužské populace (dle N(104, 8)) má IQ rovno nebo vyšší 117,16

Řešení 2d

- *Odhadněte v jakých mezích se pohybuje IQ 99.9 %*
- *mužské populace.*



99,9 % mužské populace se pohybuje v mezích intervalu 79,3 až 128,7 jednotek IQ.

Binomické rozdělení

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

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

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

Příklad 3a – 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 3

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

Binomické rozdělení

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

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

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

Příklad 1a – 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 a

Tabulka3: Parametry binomického rozdělení v příkladu

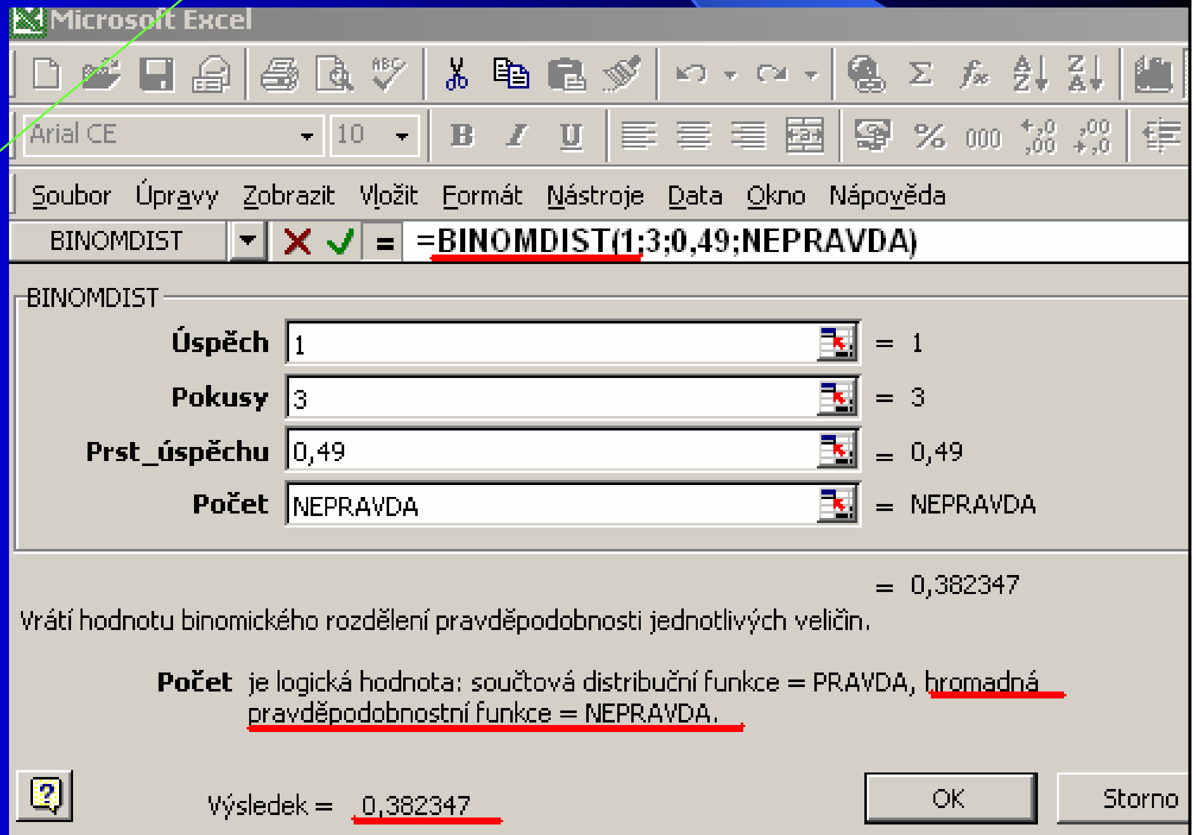
Pokus	Úspěch	Neúspěch	Pravděpodobnost úspěchu	Počet pokusů	Počet úspěchů
				n	k
narození dítěte	dívka	chlapec	0,49	počet dětí	počet dívek

Řešení 1a

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%.

Příklad 1b

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

Řešení

binomický rozvoj:

$$P(k=3)=\binom{8}{3}0{,}49^3\cdot 0{,}51^5=\frac{8\cdot 7\cdot 6}{3\cdot 2\cdot 1}\cdot 0{,}118\cdot 0{,}035=0{,}23.\ \diamond$$

Pravděpodobnost, že v rodině s 8 dětmi jsou tři dívky, je 0,23, tj. 23 %.

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	$\pi = 617/1116$ $\pi = 0{,}553$	$q = 499/1116$ $q = 0{,}447$ $(q = 1 - \pi)$	n =12	k=0 až 12

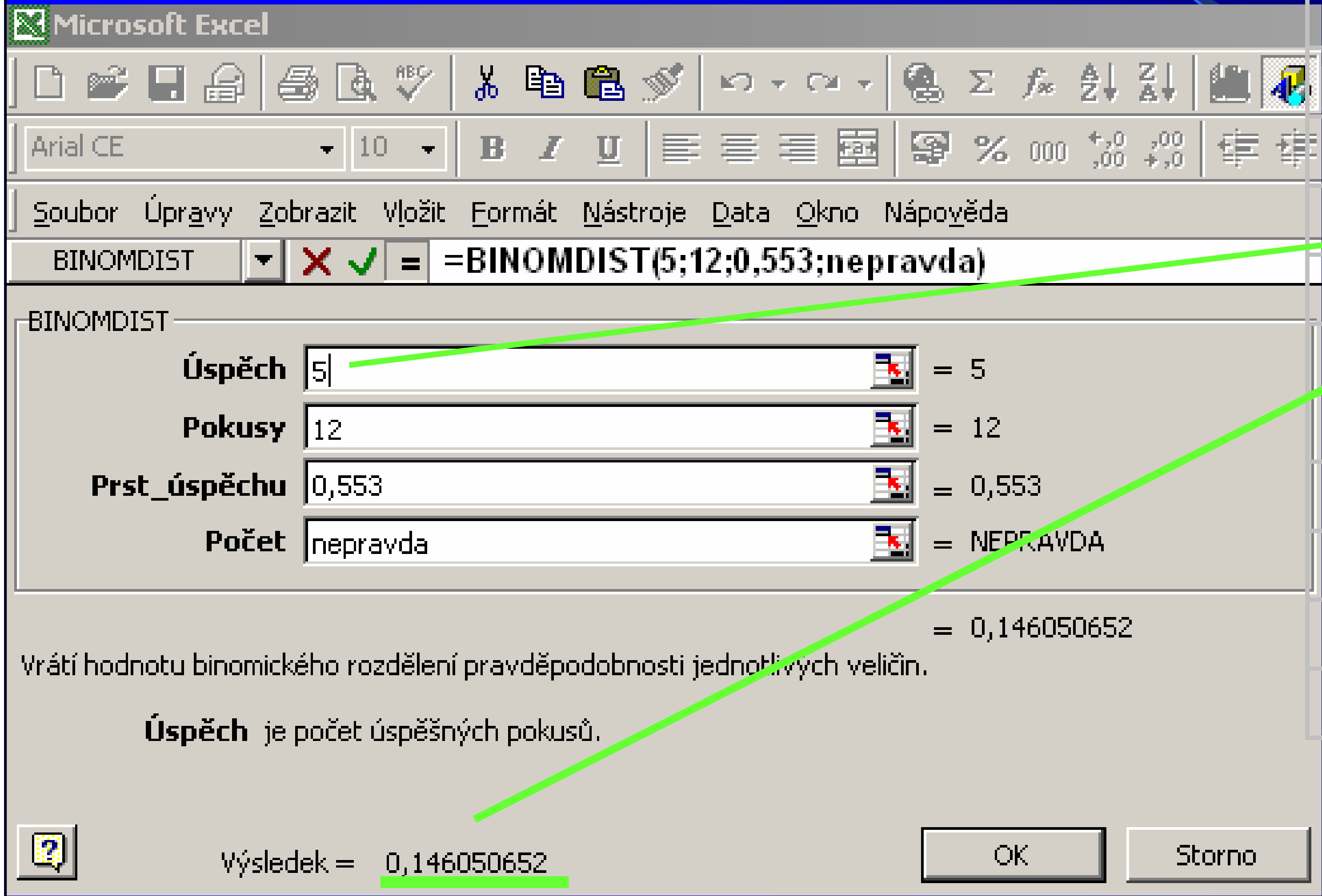
Řešení

a) Ručně pomocí binomického rozvoje
b) s podporou např. Excel

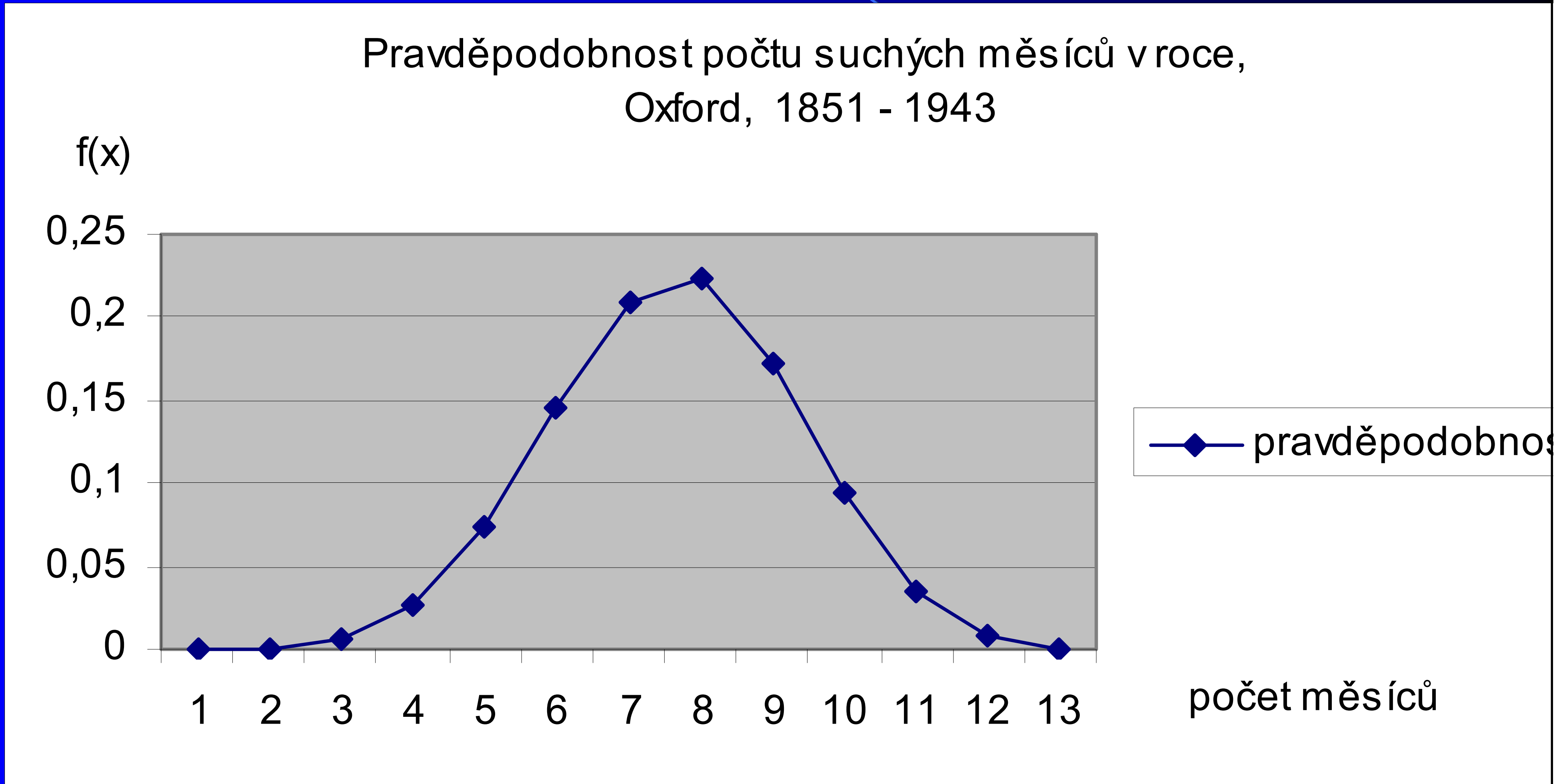
Řešíme dílčí příklady, tj. jaká je pravděpodobnost, že v roce se vyskytne

a) žádny suchý měsíc, tj- k = 0
b) Jeden suchý měsíc, tj. k = 1
c) Atd.
d) všechny měsíce suché, k= 12

Řešení 2



k	f(x)
0	0, 000
1	0,000945
2	0,006428
3	0,026507
4	0,073785
5	0,146051
6	0,21
7	0,223
8	0,172
9	0,095
10	0,035
11	0,0079
12	0,0008


Pravděpodobnost počtu suchých měsíců v roce,
Oxford, 1851 - 1943
f(x)
0,25
0,2
0,15
0,1
0,05
0
1
2
3
4
5
6
7
8
9
10
11
12
13
počet měsíců
pravděpodobnos

Poissonovo rozdělení

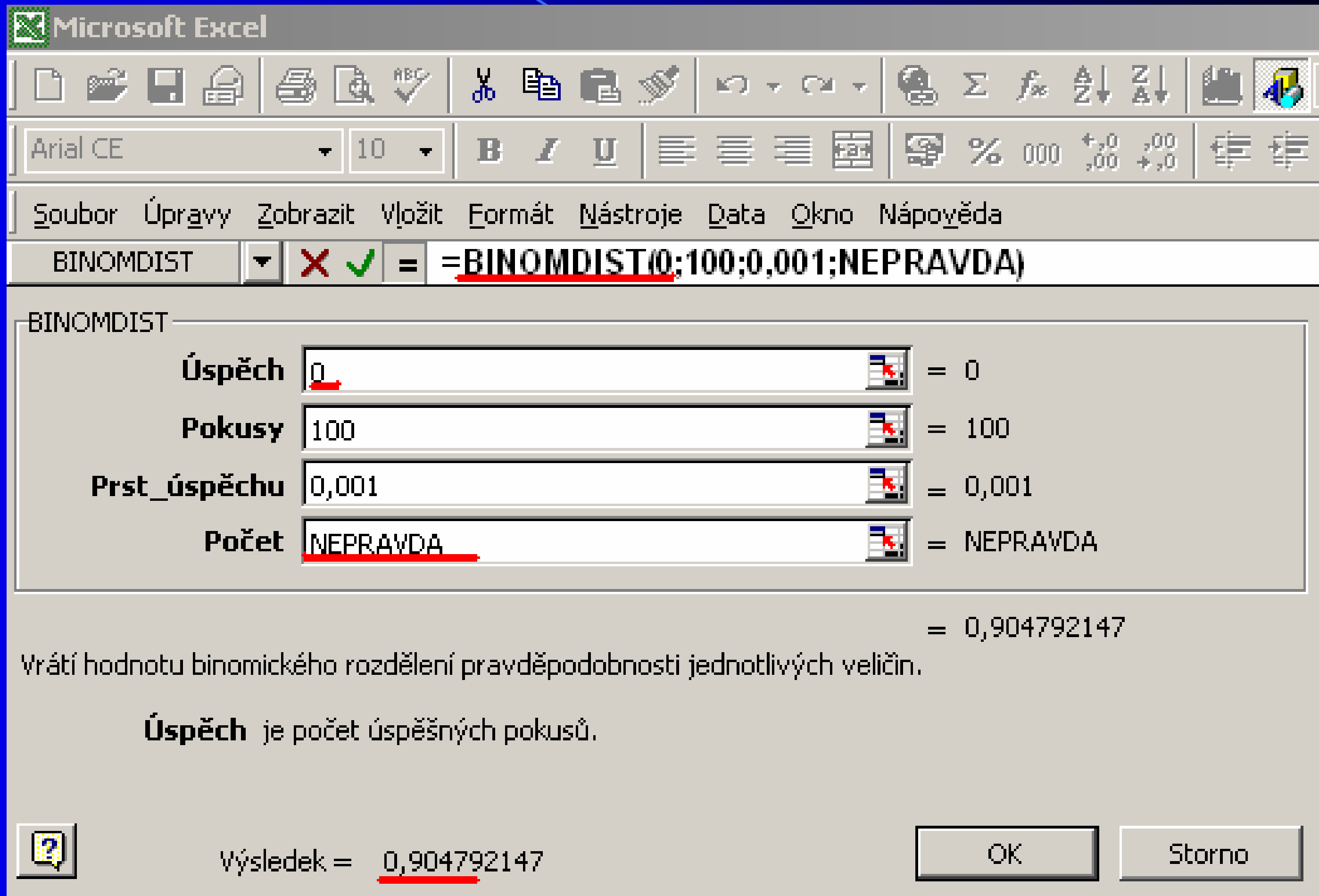
- – pro rozdělení vzácných případů (zimní bouřka, výskyt mutace apod.).
- Je-li pravděpodobnost nějaké výjimečné události (např. určité mutace genu) relativně malá a rozsah výběru poměrně velký, pak Poissonovo rozdělení v podstatě splývá s binomickým, ale je mnohem výhodnější pro počítání .

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) <u>neobsahuje albína,</u>
- <u>b) obsahuje právě jednoho albína.</u>

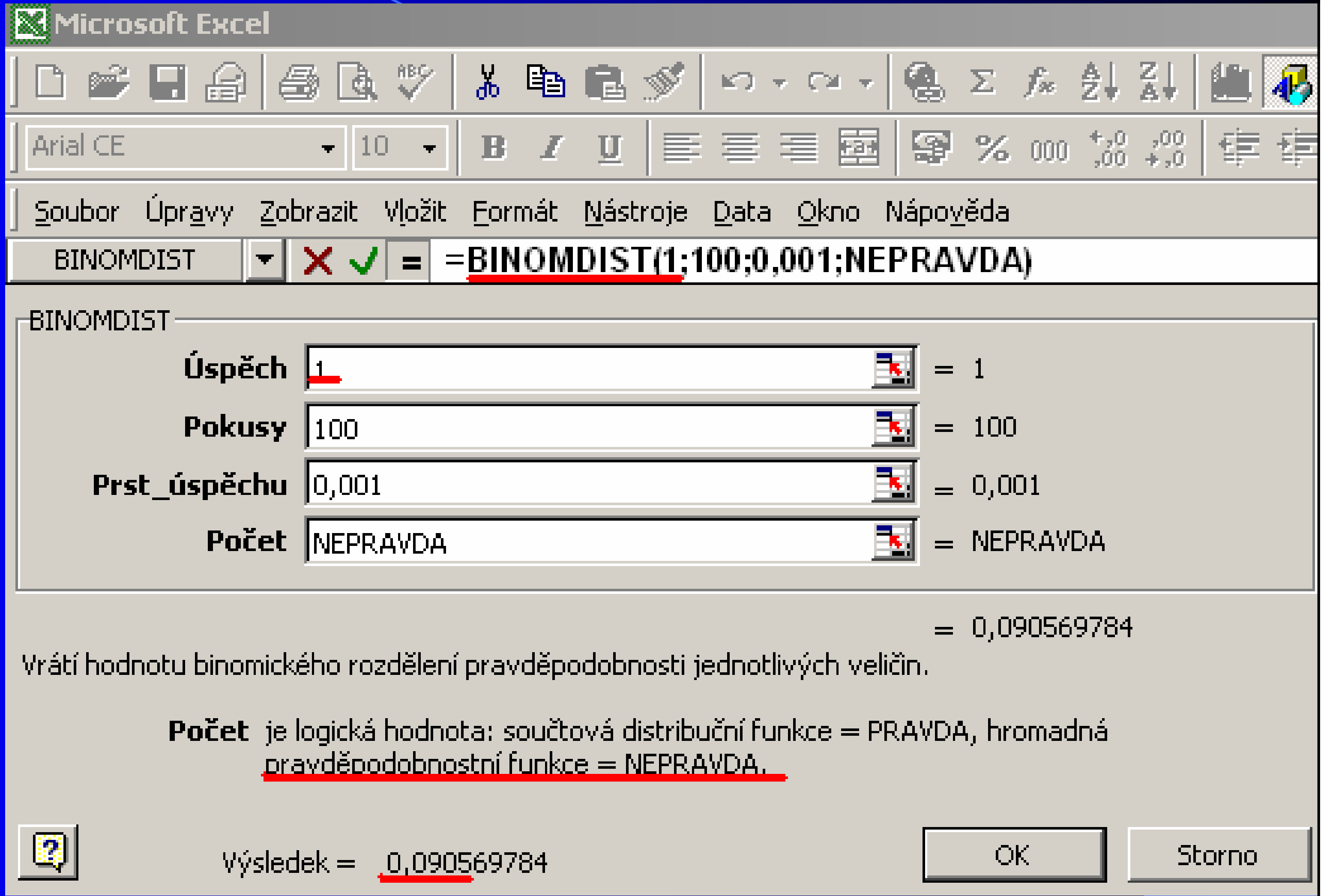
Řešení 3:

- určete pravděpodobnost, že vzorek
- neobsahuje albína,



Pravděpodobnost, že neobsahuje albína, je 90,47 %

Řešení 3



Pravděpodobnost, že 100 členná populace krys bude obsahovat albína, je 9 %.

Další rozdělení : Pearsonova křivka III. typu

- **Pearsonova křivka III. typu**
- - obvykle pro veličiny s omezeným množstvím hodnot, které může nabývat
- - z křivky lze např. vyčíst pravděpodobnost se kterou bude hodnota sledovaného statistického znaku dosažena
- v hydrologii se počítá Pearsonova křivka ve variantě součtová čára četností jako
- tzv. čára překročení

Studentovo/t/ rozdělení rozdělení $\chi 2$

- **Studentovo/t/ rozdělení** – hodnocení odchylek aritmetického průměru základního souboru a výběrových souborů
- **rozdělení $\chi 2$** – výběr ze základního souboru, počet vybíraných prvků = počet stupňů volnosti,
- součtu druhých mocnin daného počtu vybraných prvků odpovídá určitá křivka,

Odhady parametrů
intervaly spolehlivosti

- základní soubor, výběrový soubor a jeho náhodný výběr
- reprezentativnost 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,
- **<u>interval spolehlivosti hranice</u>** (μ + - 2σ),
- hodnoty, které leží mimo interval, v tzv. **<u>kritickém oboru</u>** se považují za nepřípustné, jejich <u>odchylky od průměru za významné</u>
- lze použít i jiné intervaly spolehlivosti
- např. pro 95 % (μ + - 1,960σ),
- pro 99 % (μ + - 2,576σ),

Testování statistických hypotéz

- jak ověřit předpoklady o charakteristikách statistického soubor?
- 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 ?
- 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** (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 nasyceni 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 **<u>tendenci</u>** změny hodnoty závislé proměnné na nezávislé proměnné pomocí matematické funkce
- Tuto **<u>regresní funkci</u>** lze <u>graficky znázornit</u> <u>regresní čárou</u>

- Korelace je druh závislosti mezi prvky dvou souborů
- Regresní čára <u>znázorňuje graficky tuto korelační závislost</u>

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 přímku 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
- $y' - \bar{y} = b(x - \bar{x}) + a$

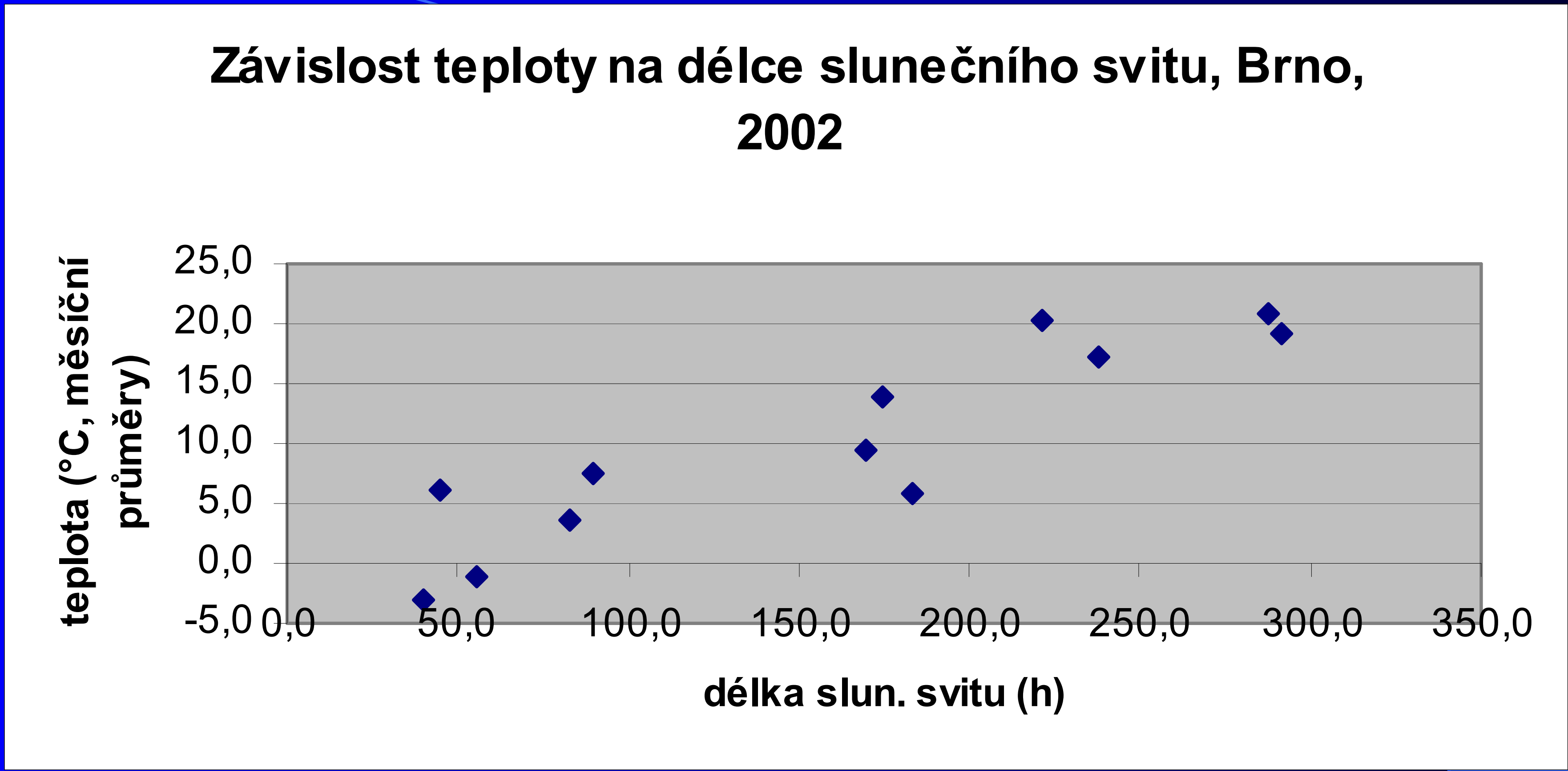
Intervaly a pásy spolehlivosti pro lineární regresní závislost

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

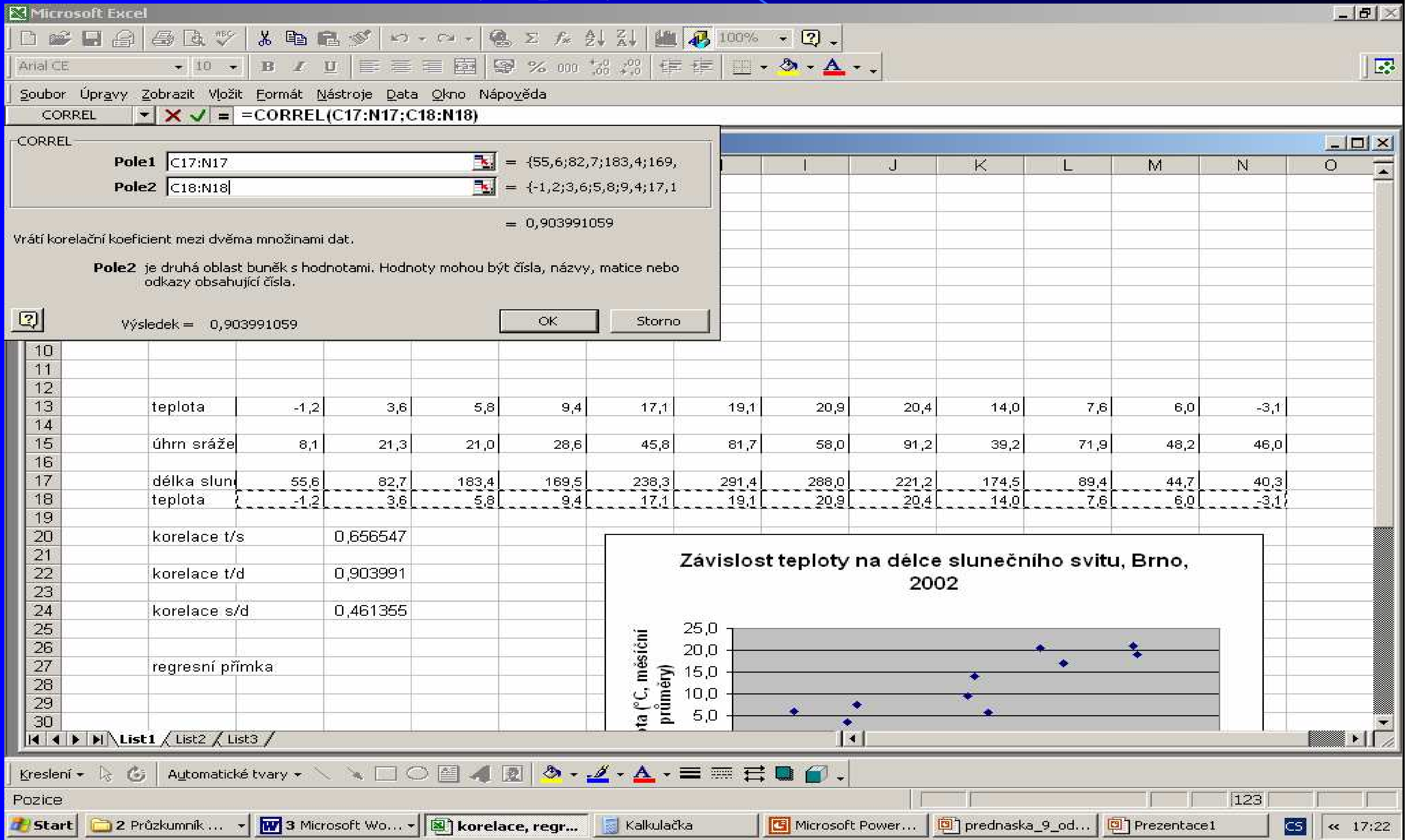
Př. lineární regrese

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

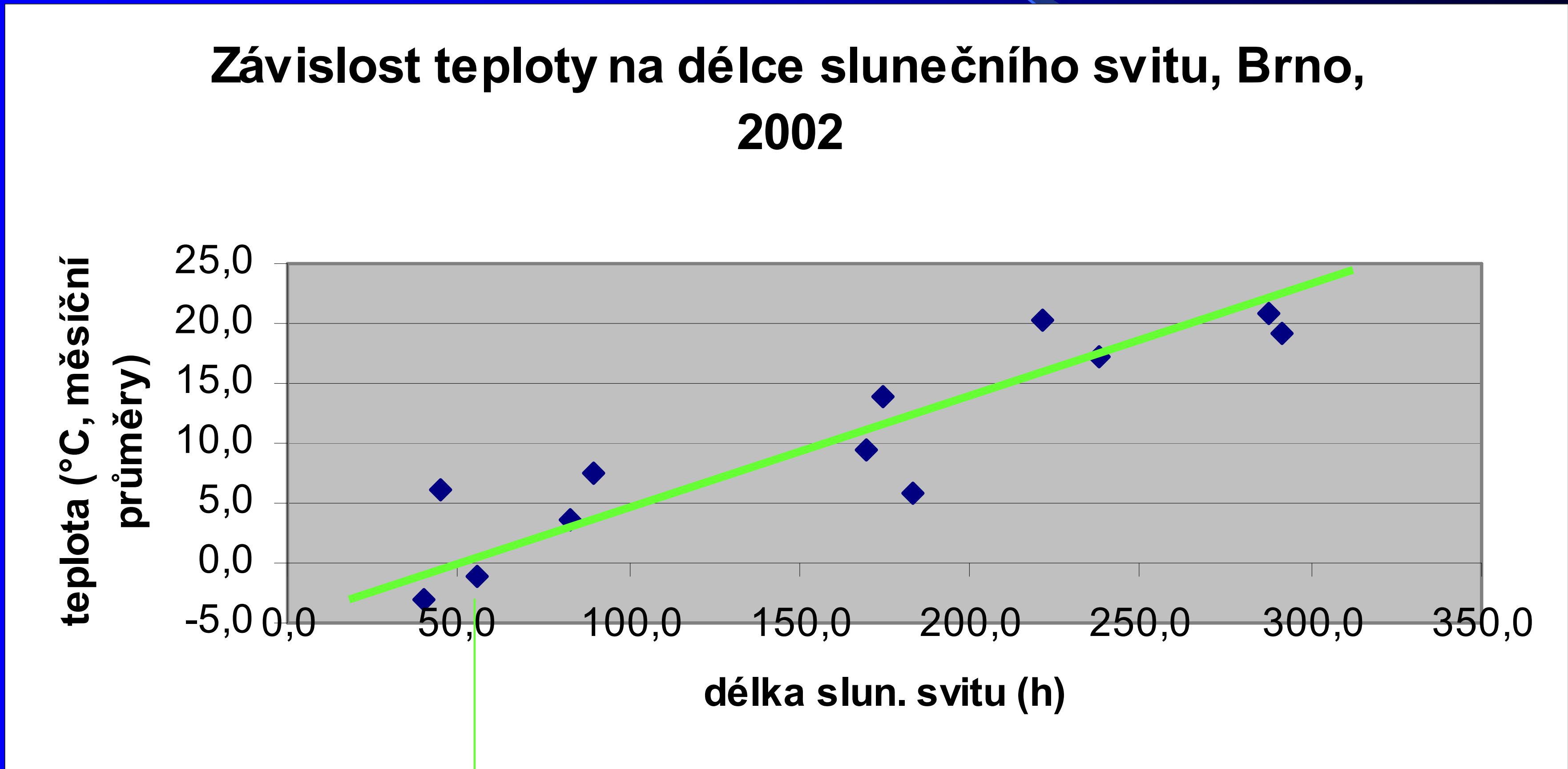
Délka slun. svitu (h)	55,6	82,7	183,4	169,5	238,3	291,4	288,0	221,2	174,5	89,4	44,7	40,3
Teplota (° C)	-1,2	3,6	5,8	9,4	17,1	19,1	20,9	20,4	14,0	7,6	6,0	-3,1


Závislost teploty na délce slunečního svitu, Brno, 2002
teplota (°C, měsíční průměry)
25,0
20,0
15,0
10,0
5,0
0,0
-5,0
0,0
50,0
100,0
150,0
200,0
250,0
300,0
350,0
délka slun. svitu (h)

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



- 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
teplota (°C, měsíční průměry)
25,0
20,0
15,0
10,0
5,0
0,0
-5,0
0,0
50,0
100,0
150,0
200,0
250,0
300,0
350,0
délka slun. svitu (h)

60

Časové řady

časová řady – základní pojmy

- **statistická řada** – posloupnost hodnot znaku uspořádaných podle určitého hlediska
- (např. viz výpočet mediánu – podle velikosti apod.)
- **časová řada** – statistická řada upořádaná podle času
- časová řada=dynamická=chronologická = vývojová

Sestavování časových řad

- dodržovat **<u>zásady</u>**:
  - **stejně dlouhá časová období** (přepočet na „standardizovaný“ <u>měsíc se 30</u> dny, přepočet na <u>stejně dlouhé roky</u> pokud se vyskytuje přestupný rok, přepočet na počet <u>shodný počet pracovních dní</u> v měsíci p
  - **stejně velká území** (shodná rozloha, povodí řádu toku, administrativní jednotka
  - **stejné jednotky**

- <u>časová řada OKAMŽIKOVÁ</u> – sledují se hodnoty znaku k určitému okamžiku
- např. počet obyvatel ČR k 31.12. 2000, 2001, 2002, 2003, 2004, 2005
- počet CHKO v ČR k………
- <u>časová řada INTERVALOVÁ</u> – 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

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

Z - diagramy

- intervalové řady typu
- řada běžných hodnot, součtová čára, řada klouzavých úhrnů lze znázornit v <u>Z - diagramu</u>
- společné body (tj. spol. hodnoty) jsou výchozí bod součtové č. a řady běžných hodnot a poslední hodnota součtové čáry a poslední hodnota klouzavého úhrnu (shodná hodnota)

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

- **<u>řetězový index</u>** (koeficient růstu)
- podíl $x_i / x_{i-1} * 100$
- podíl v procentech po sobě následujících hodnot
- (změny např. z měsíce na měsíc“ – řetězení)
- **<u>bazický index</u>**
- podíl $x_i / x_z * 100$,
- x_{z} - první „ základní „ hodnota časové řady
- změny k jedné základní (bazické) hodnotě

Témata přednášek k samostudiu

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