

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**

Vymezení základních statistických pojmů

Hromadný jev ve statistice

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:

u souboru lidí hromadný jev: věk osob, váha, dosažené vzdělání **studenti dopíší další příklady (např. viz. statistické ročenky)....**

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. (CO, KDY, KDE)**

Příklady:

stat. jednotka – občan, novorozenec, rodina, dům,, obec, výrobní podnik,

Co:

Kde:

Kdy:

Statistický znak:

je to určitá vlastnost statistické jednotky, kterou se snažíme postihnout.

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

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

Příklady: **shodný znak – novorozenost**

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

stat. jednotka
novorozenec

Statistické znaky lze dělit na znaky

- A) **prostorové**

místo narození: Brno

- B) **časové**

datum: 30.9.2011

- C) **věcné:**

- 1. **kvalitativní:**

pohlaví: muž

- **alternativní**

- **možné**

národnost: česká

- 2. **kvantitativní:**

- **spojité**

délka v cm: 55

- **diskrétní/nespojité**

Doplňte další příklady

Statistické znaky můžeme získat :

- **přímo** – (např. měřením, zvážením) tj.
.....**data**
- **nepřímo** (výpočtem), (znaky odvozené) tj.
.....**data**

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. **shodné - identifikační znaky**

- **sledované znaky** – tyto znaky statisticky šetříme

Příklad:

statistický soubor Novorozenci v ČR

Shodný - identifikační znak: novorozenost

sledovaný znak: váha, živý, pohlaví

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 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

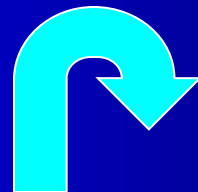
Grafické znázornění jevů

- **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.legendy/klíč
- 5.zdroj údajů
- vysvětlivky, poznámky,

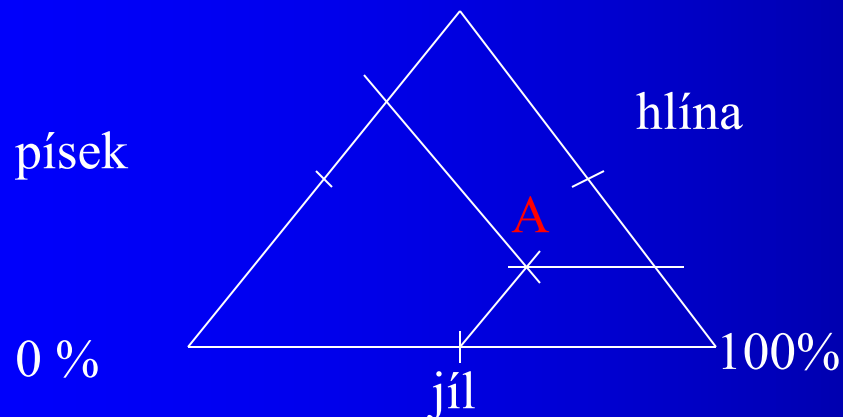
Typy grafů

- **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



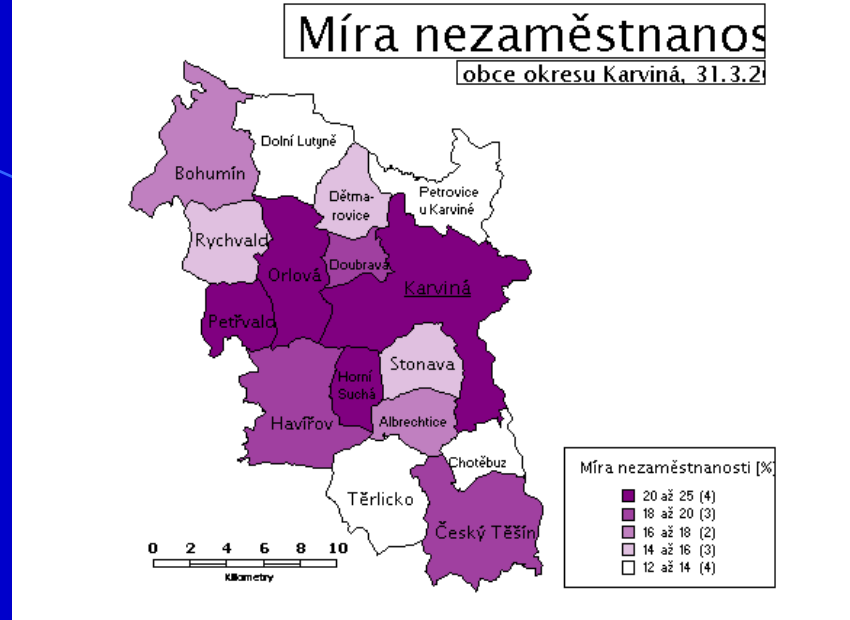
Sítě

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



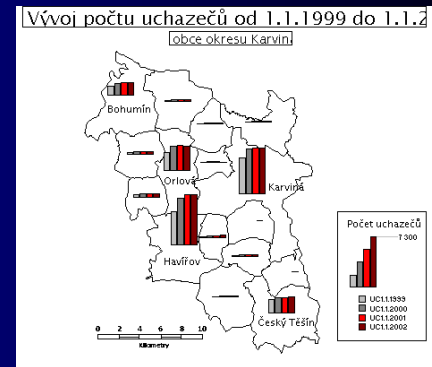
statistická mapa:
kartogram
kartodiagram

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.)

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)

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. šrafovaní-hustota obyv.)

př

Grafické metody analýzy geografických jevů

- 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)

Izolinie – konstrukce a vlastnosti

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

Rozdělení četností

A decorative graphic consisting of a large, light blue arc that starts from the top left and curves towards the bottom right. A smaller, darker blue triangle is positioned on the right side, pointing towards the center of the arc. The background is solid black.

Absolutní, relativní kumulované četnosti

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

skupinové rozdělení četností

- roztrídíme 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

Četnosti

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

Grafické znázornění rozdělení četností

- histogram
- polygon
- čára kumulovaných četností

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

Základní statistické charakteristiky

Základní statistické charakteristiky

Popisná statistika

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

Střední hodnoty

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

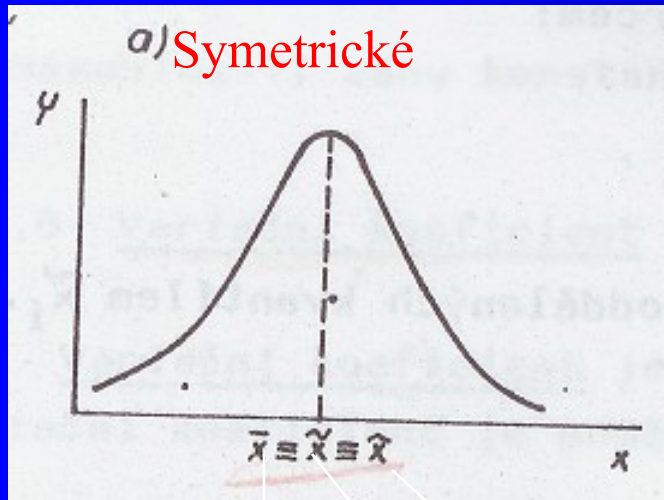
Charakteristiky variability

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

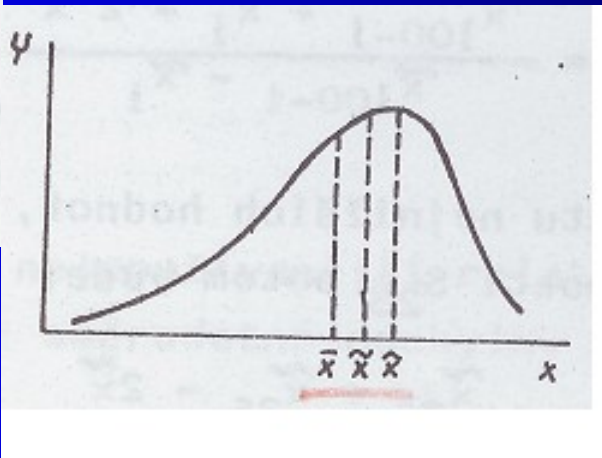
Charakteristiky asymetrie

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

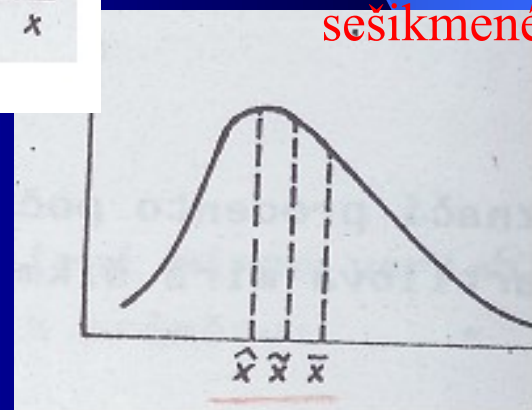
Charakteristiky asymetrie



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

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

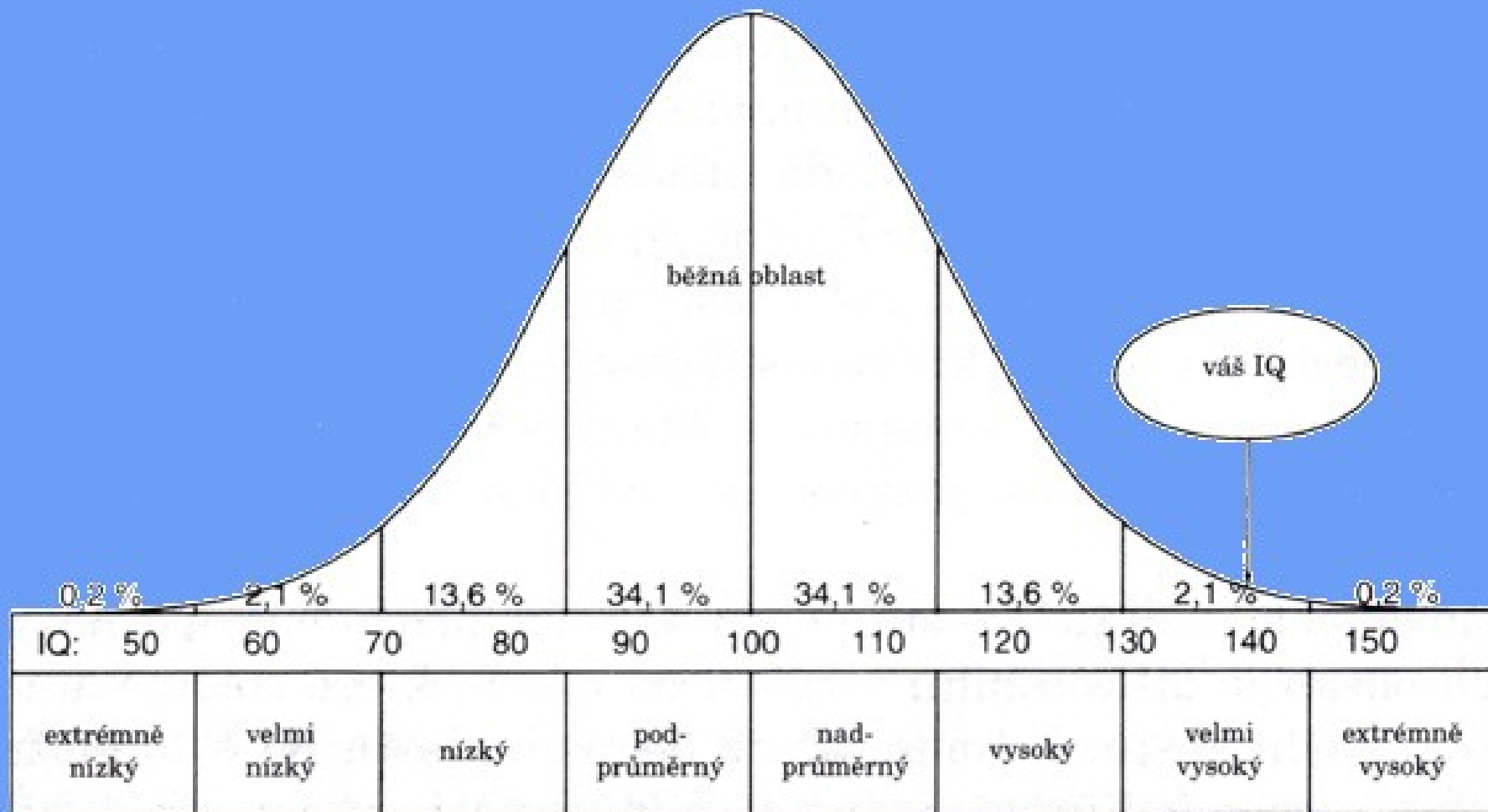
V normálním rozdělení:

- **68, 27% leží v intervalu:**
- **(průměr + - směr. odchylka)**

- **95% leží v intervalu:**
- **(ar. průměr +/- 1,96 směr. odchylky)**

- **99% leží v intervalu:**
- **(ar. průměr +/- 2,576 směr. odchylky)**

Normální rozdělení pro IQ



imbecilita

debilita

Lehká d.

průměr

vynikající

genialita

idiocie

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

Poisson - příklad

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 .

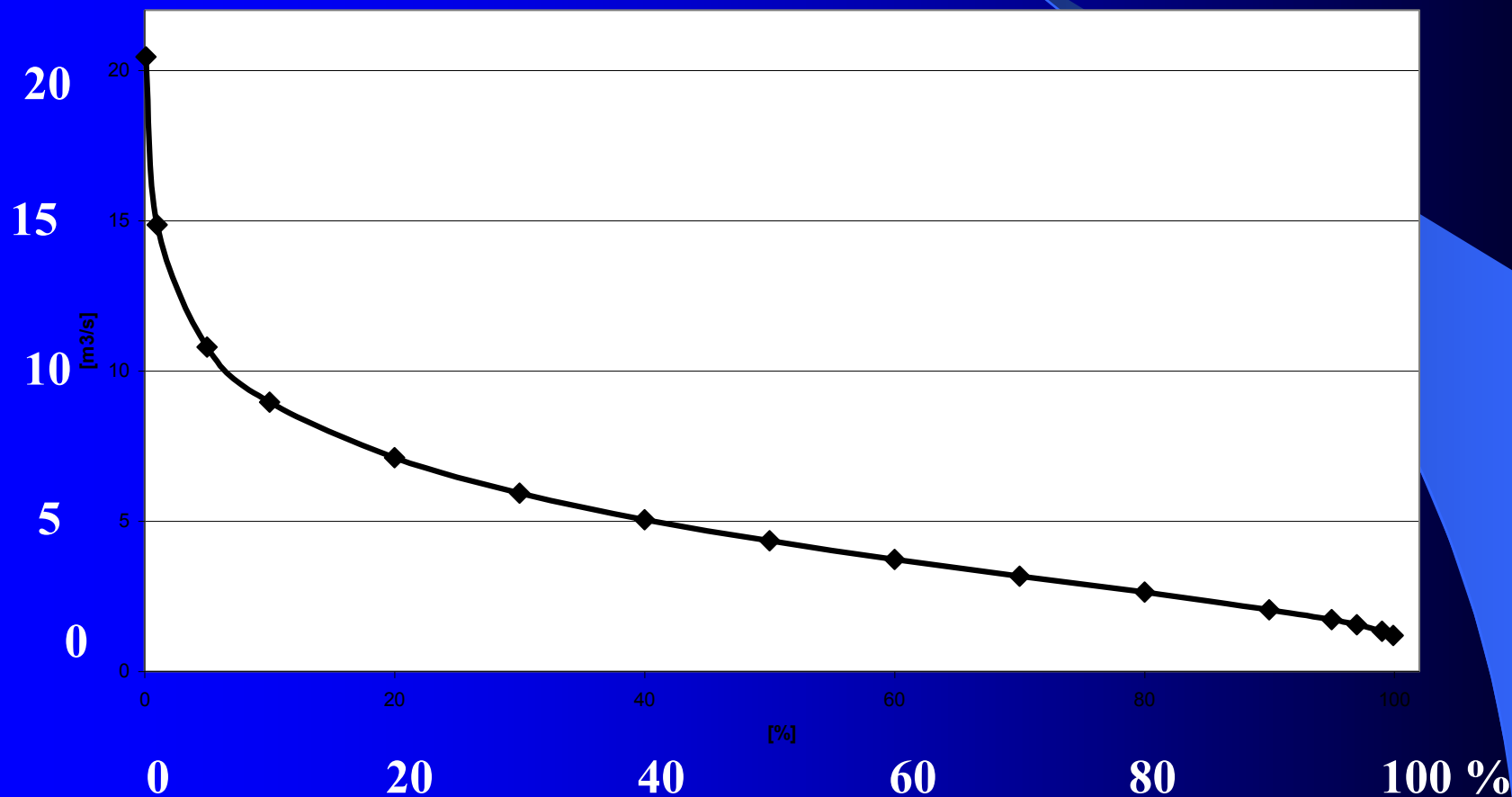
Pearsonova křivka III. typu

- 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

Křivka překročení průměrných ročních průtoků , Lažanka, říjen 2002

Křivka překročení průměrných ročních průtoků vodního toku Lažanka za říjen 2002

m³/s



Základní pojmy

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

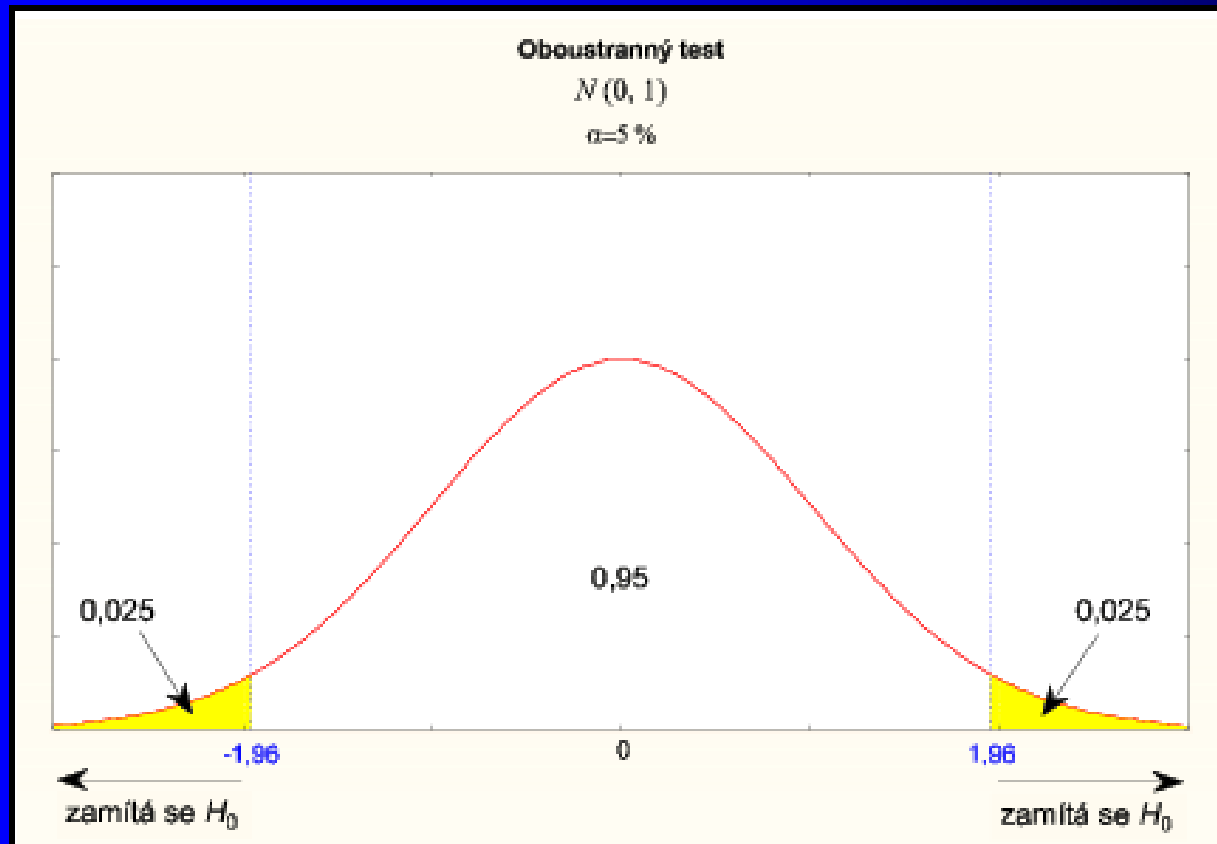
Základní pojmy

- **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,
- 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 .

- **Obrázek: Oboustranný test H_0**



- provedeme-li výběr o rozsahu n a spočteme \bar{x} , pak průměr \bar{X} leží s pravděpodobností 0,95 ve vzdálenosti menší než $1,96 \cdot s / \sqrt{n}$ od \bar{x} ,
- tj. v intervalu s krajními body
- $(\bar{x} - 1,96 \cdot s / \sqrt{n}, \bar{x} + 1,96 \cdot s / \sqrt{n})$... *interval spolehlivosti pro průměr.*
- *koeficient 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 nepřijatelné, jejich odchylky od průměru za významné

Závislost náhodných veličin

Vztahy náhodných veličin

- Jednostranné (nezávislá hodnota \underline{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
- (jediné 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

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

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)

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