

Statistické metody a zpracování dat

I. Úvod, základní pojmy

Petr Dobrovolný 

Obsah přednášky

1. Úvod, základní pojmy
2. Základní vyjadřovací prostředky ve statistice
3. Základní popisné statistické charakteristiky
4. Úvod do počtu pravděpodobnosti, teoretická rozdělení
5. Odhady parametrů a intervaly spolehlivosti
6. Testování statistických hypotéz
7. Měření závislosti náhodných veličin
8. Analýza kategoriálních dat
9. Úvod do analýzy rozptylu
10. Úvod do analýzy časových řad
11. Úvod do vícerozměrných statistických metod

Základní literatura

Brázdil a kol. (1995): Statistické metody v geografii. MU Brno, 177 s.

Prezentace z přednášek – doplňky

Rogerson, P. A (2001): *Statistical methods for Geography*. Sage Publications, London., 236 s.

Hendl, J. (2004): Přehled statistických metod zpracování dat. Portál, Praha, 583 s.

IS – Studijní materiály

Program STATISTICA

Cvičení – zadání, podkladová data - přes IS

Multilicence MU

<https://inet.muni.cz/app/soft/licence>

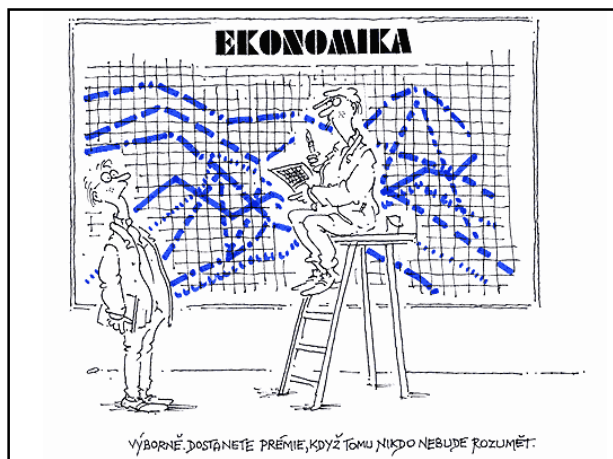
Základy práce s programem – 1. cvičení. Další informace, elektronická učebnice:

<http://www.statsoft.cz>



Zakončení předmětu

- účast na cvičeních, včasné odevzdání protokolů
- 1. písemný test (polovina semestru) – řešení praktické úlohy
- 2. písemný test (konec semestru) – odpřednášená látka
- hodnocení: minimálně 50% bodů v součtu obou testů, žádný jeden by neměl být horší než 25%
- termíny testů budou včas oznámeny



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**

Co je typické pro statistiku

- 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.

Významy pojmu STATISTIKA

I. Statistika jako **praktická činnost** - statistická evidence, instituce, ročenky meteorologických pozorování atd.

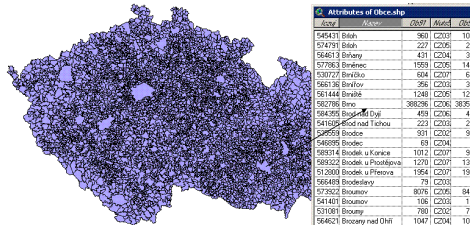
II. Statistika jako **vědní disciplína** - popisná a matematická (induktivní) statistika, aplikované vědy (ekonometrie, chemometrie atd.), vědy se silným statistickým základem: klimatologie, hydrologie, sociologie, psychologie, demografie aj.

Statistika se těší pochybnému vyznamenání tím, že je nejvíce nepochopeným vědním oborem

(H. Levinson)

Vymezení základních pojmů I

Hromadné jevy: přírodní či společenské jevy, které jsou výsledkem působení velkého množství příčin, jejich vlastnosti se neprojevují v jednotlivých jevech, ale jen v souboru a to prostřednictvím řady náhod.



Řada jevů, které v geografii studujeme pomocí statistických metod, má povahu jevů náhodných – tzv. stochastických (hydrologické jevy či meteorologické jevy).

Vymezení základních pojmů II

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**.

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

Statistický soubor: skupina statistických jednotek stejného druhu (věcně, prostorově a časově vymezených), které jsou předmětem 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 (**vymezují** statistický soubor).

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

Vymezení základních pojmů III

Statistické znaky lze dělit na znaky **prostorové, časové a věcné**.

Věcné znaky se dělí na znaky **kvantitativní** a **kvalitativní**

Kvalitativní znaky mohou být **alternativní** a **možné**

Kvantitativní znaky dělíme nejčastěji na znaky **spojité** a **diskrétní**.

Statistické znaky můžeme získat přímo – (např. **měřením**) a nebo **nepřímě** (výpočtem). Tyto potom nazýváme znaky odvozenými.

Podle škály, na které znaky zjišťujeme je dělíme na znaky **nominální, ordinální, poměrové, intervalové**

Vymezení základních pojmů IV

Základní statistický soubor - populace

Výběrový statistický soubor je podmnožinou základního souboru.

Je vytvořen ze statistických jednotek, vybraných podle určitého hlediska.

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

Statistický soubor **jednorozměrný, vícerozměrný**

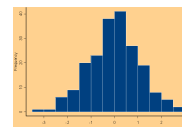
Rozsah statistického souboru:

N – rozsah základního souboru
n – rozsah výběrového souboru



Popisná statistika

Popisná (deskriptivní) statistika se zabývá uspořádáním souborů, jejich popisem a účelnou sumarizací.



$$\bar{x} = 0,5$$

$$x_{\min} = -3,6$$

$$x_{\max} = 3,0$$

Jak mohou být tyto jevy jednoduše popsány (charakterizovány, sumarizovány)?

Existují dvě základní možnosti, které se vzájemně doplňují:

- Numerické metody** – jedním nebo několika málo čísly lze vystihnout určité vlastnosti jevu. Jsou přesnější a objektivnější

- Grafické metody** – sestavení vhodného typu grafu. Jsou názornější a umožňují vystihnout vztahy.

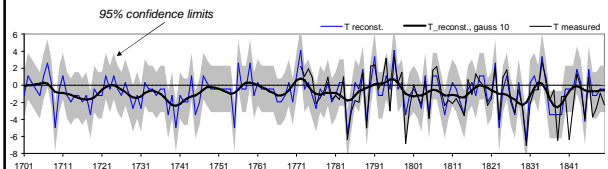
Induktivní statistika

Induktivní (matematická) statistika se vyvinula z popisné statistiky a jejím základem je **teorie pravděpodobnosti**.

Matematická statistika zkoumá soubory nepřímo prostřednictvím výběrů

Induktivní statistika se zabývá metodami jak poznatky **přenášet** a umožňuje z pozorovaných dat vytvářet **obecné závěry** s udáním *stupně jejich spolehlivosti*.

Výpočet stupně spolehlivosti závěrů je však objektivní, neboť je založen na poznacích teorie pravděpodobnosti a nezávisí na subjektivním názoru hodnotitele.



Geografická data a jejich specifika

- **Zdroje geografických dat** – primární, sekundární
- **Prostorový aspekt** – statistika prostorově lokalizovaných dat (geostatistika)
- **Časový aspekt**

Dělení geografických dat podle použité škály měření

- **Nominální** (kategorie využití země)
- **Ordinální** (řád vodního toku, stupnice síly větru)
- **Intervalová** (teplota vzduchu) nula = data
- **Poměrová** (množství srážek, délka vodního toku) nula = neexistence jevu

Typy geografických dat

Nominální data – hodnota představuje konkrétní kategorii či třídu a vyjadřuje její označení (jméno), kategorie se nesmějí překrývat – jsou disjunktní. Každý objekt je zařaditelný alespoň do jedné kategorie, žádný nespadá do více než jedné. Čísla, která označují kategorie jsou pouze symboly a nelze s nimi provádět aritmetické operace. V nejjednodušší podobě mají binární charakter a lze je pouze porovnávat.

Ordinální data – data, která lze seřadit do uspořádané posloupnosti podle určitého kritéria. Je známé pořadí kategorií, rozdíl však nemá smysl. Např. řád vodního toku, třída silnice, bonita půdy atd.

Typy geografických dat

Intervalová data – umožňují provádět i odečítání mezi kategoriemi definovat rozdíl mezi kategoriemi. Teplota vzduchu. Stupnice většinou nezačíná nulou. Poměr dat závisí na zvolených jednotkách.

Poměrová data – vedle rovnosti, uspořádání a odčítání umožňují také dělení. Nula vyjadřuje neexistenci jevu – objem, délka ...

Statistický software

1. Programové vybavení založené na využití vlastního programovacího jazyka (R, Splus, SAS)
2. Interaktivní zpracování v „oknech“ MINITAB, SPSS, STATGRAPHICS, Statistica
3. Programové vybavení s knihovnou statistických, matematických a grafických funkcí (EXCEL)

Základní vyjadřovací prostředky ve statistice

- Statistické tabulky
- Statistické grafy

Tabulky – složené z buněk, přehledné, nezávislé na textu

Tab. 1 Základní statistické charakteristiky teploty vzduchu [°C] na vybraných stanicích za období 1961-2000

Charakteristika	Stanice	
	Praha- Klementinum	Strání
průměr [°C]	9,205	7,673
rozptyl ¹⁾	0,679	0,403
rozsah souborů	120	30
směrodatná odchylka [°C]	0,821	0,624
F	1,667	-
F (kritické)	1,699	-

Pramen: ČHMÚ
Vysvětlivky: 1)

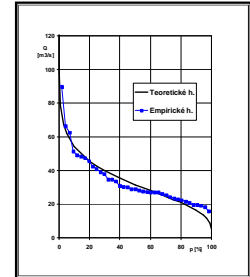
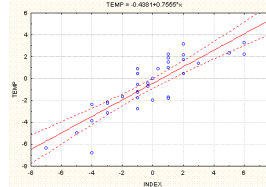
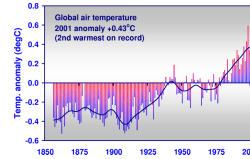
Statistické tabulky

- Záhlaví a legenda mají obsahovat měrné jednotky
- Tabulka má vyplněna všechna políčka
- Smluvené znaky pro políčka bez číselného údaje
 - – údaj se nevyskytuje
 - x – údaj není možný z logických důvodů
 - 0 – hodnota je menší než polovina nejmenší měrné jednotky
 - . – údaj nelze spolehlivě zjistit

Metody grafického znázornění geografických jevů

Motto:

Jeden obrázek je za tisíce slov

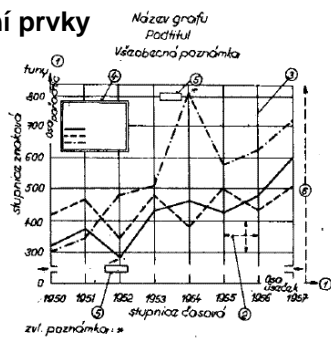


Cílem grafického znázornění je podat rychlou a srozumitelnou informaci o studovaném jevu či o vzájemném vztahu více jevů.

Graf a jeho základní prvky

Graf – kresba provedená podle předem dohodnutých pravidel, která znázorňuje kvalitativní či kvantitativní znaky.

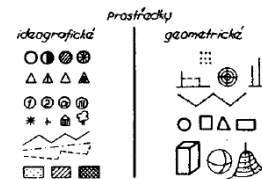
1. stupnice
2. grafický interval
3. síť
4. klíč
5. vysvětlivka
6. délka stupnice



Grafický obraz – soubor grafických prostředků, pomocí kterých na základě dohodnutého výkladu jejich smluveného významu sestavujeme graf

Grafický výklad – soubor zásad, podle kterých interpretujeme (čteme) příslušný graf).

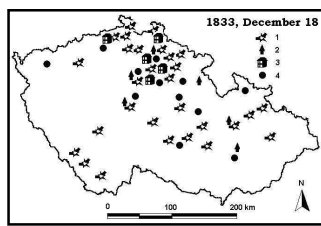
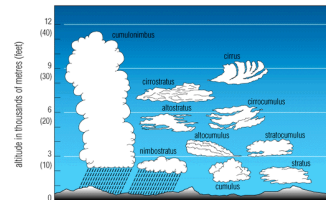
Dělení grafických prostředků podle významu



Ideaografické – mají kvalitativní význam a v grafu fungují jako znaky (klasifikační, identifikační). Jejich tvar a rozměry slouží pouze k jejich odlišení, nemají kvantitativní význam (písmena, číslice, symboly, geometrické obrazce, šraflura, barva, druhy čar apod.).

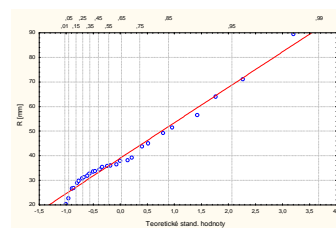
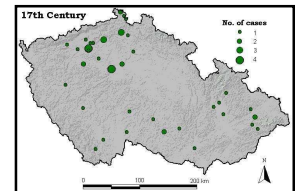
Geometrické – mají vždy kvantitativní význam, často však také slouží ke kvalitativnímu odlišení statistických jednotek (body, úsečky, obrazce).

Ideaografické prostředky



- 1 – windbreakage
- 2 – damage on buildings of lesser extent
- 3 – destroyed buildings
- 4 – damage without specification

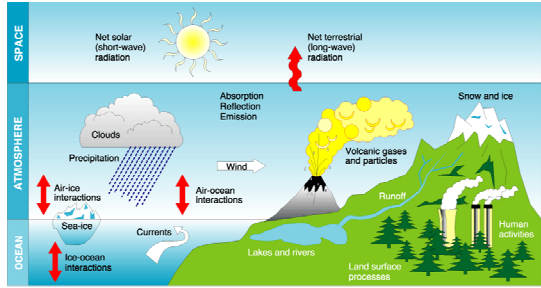
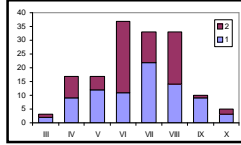
Geometrické prostředky



Základní typy grafů

Z hlediska předmětu grafu:

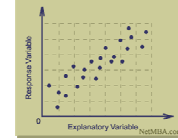
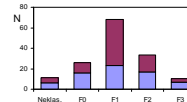
- schémata (struktura, vztahy, ...)
- diagramy (kvantita, četnost, ...)



Základní typy grafů

Z hlediska způsobu použití geometrických prostředků:

- rozměrové grafy
- souřadnicové grafy



Speciální typy grafů využívané v geografii:

- ternární graf
- větrná růžice, klimadiagram, ...

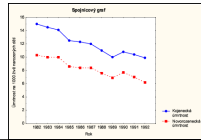
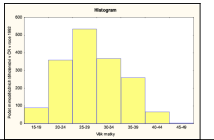
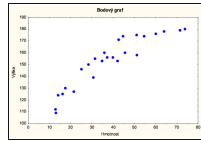
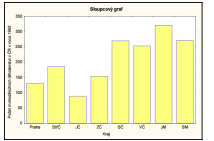
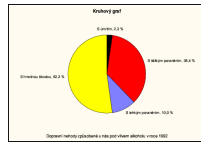
Statistické mapy

- kartogramy
- kartodiagramy

Základní typy grafů

Grafy pro vyjádření jedné proměnné

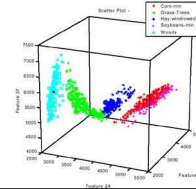
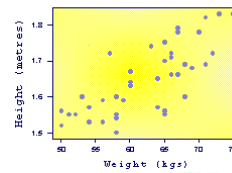
- sloupkový diagram
- histogram
- kruhový diagram, výšečový graf
- bodový graf
- spojnicový graf



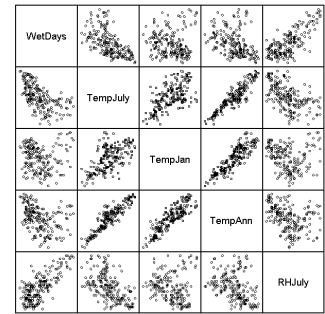
Základní typy grafů

Grafy pro vyjádření dvou a více proměnných - korelogram

Scatterplot



Climatic predictors



Speciální typy grafů

- krabičkový graf
- graf stonku a listů (stem-and-leaf-plot)
- piktoqram



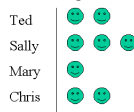
Stem-and-leaf plot of "Skinfold Thickness"

N = 40
Leaf Unit = 0.10

```

5  02468
6  245889
7  0000226
8  2444666
9  44666
10 488
11 088
12
13 22
14 0
15
16
17
18 2
    
```

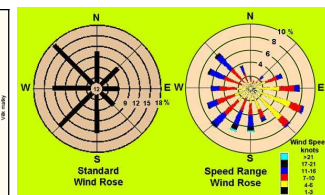
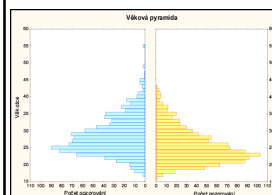
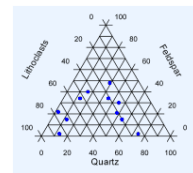
Good grades on spelling test



KEY: Represents a month of 80%+ scores

Speciální typy grafů využívané v geografii:

- ternární graf
- „strom života“
- větrná růžice
- klimadiagram



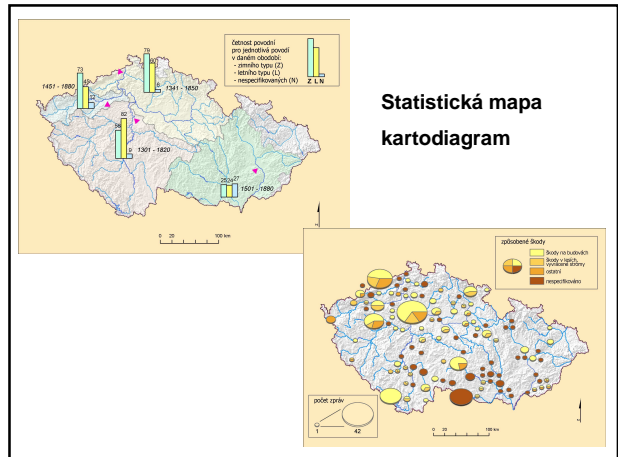
Analyza grafů

Všimáme si základního tvaru a také odchylek od něho

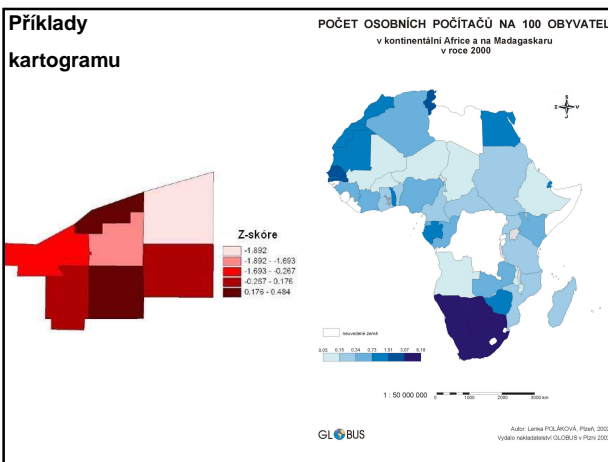
U tvaru grafu hodnotíme:

- zhuštění – místa největší četnosti hodnot
- shluky – existence jednoho či více shluků hodnot
- mezery – existence intervalů či oblastí bez hodnot
- odlehle hodnoty – existence údajů podstatně rozdílných od ostatních hodnot
- extrémní hodnoty – poloha min a max hodnot v grafu
- tvar rozdělení – jak ho lze popsat – symetrie, počet vrcholů

Volba vhodného typu grafu musí zohledňovat typ zobrazované proměnné (spojitá či diskrétní)



Statistická mapa kartodiagram

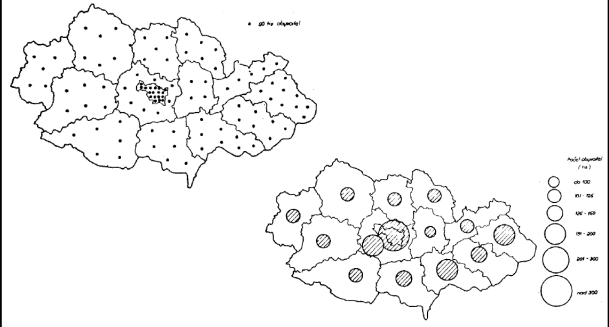


Grafické znázornění prostorové diferenciace intenzity jevů:

- Absolutní metoda:

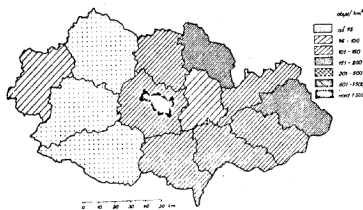
bodová – počet grafických prvků je úměrný velikosti jevu

značková – velikost grafického prvku je úměrná velikosti jevu

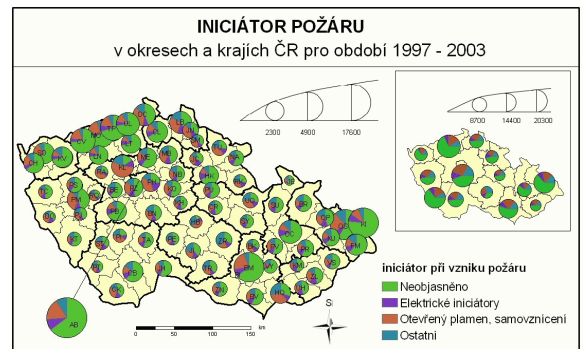


Grafické znázornění prostorové diferenciace intenzity jevů:

Relativní metoda: barva, šrafování



Grafické znázornění prostorové diferenciace intenzity a struktury jevů



Konstrukce izolinií - interpolace:

Konstrukce spojitých polí

(viz. Geostatistika)

