

# GEOINFORMATIKA A KARTOGRAFIA I

MGR. ADRIANA ZLACKÁ  
Katedra geografie a geoekológie, FHPV PU v Prešove  
1. poschodie, č.miestnosti 204

## ZÁKLADY KVANTITATÍVNYCH METÓD

### ODPORÚČANÁ LITERATÚRA:

- NIŽNANSKÝ, B. (2000): *Základy geoinformatiky*. Vysokoškolské učebné texty. Prešov, FHPV PU. 232. p.
- BAKYTOVÁ, H. a kol. (1975): *Základy štatistiky*. Bratislava, Alfa. 392 p.
- PAVLÍK, Z., KŮHNL, K. (1981): *Úvod do kvantitatívnych metod pro geografy*. Učebné texty. Praha: SPN.
- BRABEC, F. a kol. (1971): *Matematické a grafické metody v geografii*. Praha, SPN. 176p.
- GROFÍK, R. a kol. (1987): *Štatistika*. Bratislava, Príroda. 520 p.
- KOVAČKA, M., KONTŠEKOVA, O. (1962): *Štatistické metody*. Bratislava, Alfa. 266 p.
- CYHELSKÝ, L. (1981): *Úvod do teorie statistiky*. Praha, SNTL-Alfa. 352 p.
- CYHELSKÝ, L. (1985): *Statistika v příkladech*. Praha, SNTL-Alfa. 308 p.
- CYHELSKÝ, L., HUSTOPECKÝ, J., ZÁVODSKÝ, P. (1988): *Příklady k základům statistiky*. Praha, SNTL-Alfa. 416 p.

## ÚVOD DO KVANTITATÍVNYCH METÓD

### DEFINÍCIA POJMU ŠTATISTIKA

Pojem štatistika sa nechápe jednotne. Najčastejšie sa používa v nasledujúcich troch významoch:

- **praktická činnosť** – získavanie, spracovanie a vyhodnocovanie štatistických dát (t.j. údajov o hromadných javoch)
- **štatistické údaje** – údaje o hromadných javoch bezprostredne získané pozorovaním alebo z nich vypočítané charakteristiky
- **vedná disciplína** – zaoberajúca sa metódami skúmania a vyhodnocovania štatistických údajov (t.j. údajov o hromadných javov)

### PREDMET ŠTATISTICKÉHO VÝSKUMU

Keď povieme, že životné podmienky obyvateľstva sa u nás zlepšili, že návštevnosť kín stále klesá, že vlaky ešte stále meškajú, sú to všetko výroky, ktoré majú niečo spoločné. Zovšeobecňujú výsledok veľkého počtu pozorovaní. Pritom však výsledok takéhoto hromadného zovšeobecnenia nemusí platiť vo všetkých konkrétnych prípadoch. No len pomocou takýchto zovšeobecnení je možné zachytiť hlavné tendencie a vzťahy v prírodných a spoločenských javoch. Keby sme skúmali tieto javy u jednotlivých prvkov, **individuálne javy**, neboli by sme schopní určiť čo je pre daný jav podstatné, čo nepodstatné, čo je náhodné.

- predmetom štatistického skúmania je **hromadný jav**.

**Hromadný jav** – každý prírodný alebo spoločenský jav, ktorý sa skúma nie jednotlivo, ale na veľkom počte prípadov (hromadne).

Skladá sa z mnohých **individuálnych javov**.

Prírodné a spoločenské javy sa skúmajú ako hromadné javy, preto lebo takýto spôsob umožňuje **poznať ich podstatu a vlastnosti, povahu činiteľov, ktoré ich podmieňujú a ovplyvňujú v ich vzájomných vzťahoch a súvislostiach**.

Bez hromadného pozorovania by sme nemohli o príslušnom jave robiť **zovšeobecňujúce závery**.

- štatistika pracuje s číselnými hodnotami, aj kvalitatívne javy transformuje do čísel a potom ich vyhodnocuje, preto metódy, ktoré používa sa označujú ako kvantitatívne.
- Pojmom **štatistické údaje** sa preto označujú **číselné (kvantitatívne) údaje o hromadných javoch**

## **ZÁKLADNÉ ŠTATISTICKÉ POJMY**

**ŠTATISTICKÝ SÚBOR** – množina prvkov, objektov (osôb, vecí, javov), ktoré sú predmetom štatistického skúmania, sú zhromaždené na základe istých spoločných vlastností.

Pr1. pri sčítaní ľudu je to množina všetkých obyvateľov štátu, alebo všetkých domácností.

Pr2. pri skúmaní produkcie priemyslového podniku tvorí štatistický súbor množina všetkých podnikov konkrétneho rezortu.

- **Základný štatistický súbor** – všetky štatistické jednotky, ktoré patria do štatistického súboru. Štatistický výskum zameraný na všetky jednotky súboru označujeme ako **úplný (vyčerpávajúci) štatistický výskum (zistovanie)**.
- Základné súbory môžu byť veľmi rozsiahle, takže skúmanie všetkých jednotiek by bolo príliš zdĺhavé, nákladné, alebo niekedy aj nerealizovateľné. V takomto prípade sa robí **výberové štatistické zistovanie**.  
**Výberový štatistický súbor** – súbor, ktorý tvoria jednotky vybrané na základe určitých kritérií zo základného súboru. Výber, ktorého charakteristiky sú **takmer zhodné** s charakteristikami základného súboru je **reprezentatívny výber**.
- ♦ **Jednorozmerný štatistický súbor** – o takomto súbore hovoríme vtedy, ak pri štatistickom výskume súboru sa zaoberáme jednotlivými štatistickými znakmi samostatne, bez toho, žeby sme zohľadňovali ich vzájomný vzťah
- ♦ **Dvojrozmerný štatistický súbor** – o takomto súbore hovoríme vtedy, ak u štatistického súboru sledujeme vzájomný vzťah dvoch znakov
- ♦ **Viacrozmerný štatistický súbor** - o takomto súbore hovoríme vtedy, ak u štatistického súboru sledujeme vzájomný vzťah viacerých znakov

**ŠTATISTICKÁ JEDNOTKA** – jednotlivé prvky, objekty, resp. elementy štatistického súboru

Štatistické jednotky musia mať určité spoločné základné vlastnosti, ktoré určujú ich príslušnosť k štatistickému súboru. Pri každom štatistickom skúmaní sa musia štatistické jednotky presne zdefinovať. Vymedzujú sa z troch hľadísk:

**Priestorové (miestne) vymedzenie** – určuje územie, na ktoré sa bude vzťahovať štatistický výskum

**Časové vymedzenie** – určuje časové obdobie alebo časový moment, pre ktorý bola štatistická jednotka zaradená do štatistického výskumu

**Vecné vymedzenie** – vo všeobecnosti určuje, čo pre daný prípad je alebo nie je štatistickou jednotkou. Je to najvýznamnejšie a niekedy aj najnáročnejšie vymedzenie. Niekedy je potrebné k definícii štatistickej jednotky pripojiť aj podrobné vysvetlenie ako postupovať v jednotlivých prípadoch.

**ROZSAH SÚBORU (*n*)** – počet všetkých prvkov (štatistických jednotiek) štatistického súboru

- **Malé súbory** – obsahujú do 30 štatistických jednotiek

- **Veľké súbory** – častejšie sa pracuje s rozsiahlejšími súbormi, majú niekoľko sto, tisíc, či miliónov jednotiek
- **Nekonečne veľké súbory** – majú nekonečný počet jednotiek (čas trvania nejakého pracovného postupu (činnosti), môžeme teoreticky zisťovať jeho trvanie v neobmedzenom počte prípadov)

**ŠTATISTICKÝ ATRIBÚT (štatistický znak, štatistická premenná) (X)** – vlastnosť štatistickej jednotky.

Pr1. pri sčítaní ľudu, to môže byť vek, pohlavie, zamestnanie obyvateľov...

Pr2. pri skúmaní produkcie podniku, to môže byť počet pracovníkov, veľkosť mzdových fondov, plánovaná produkcia, skutočná produkcia...

Štatistické znaky sa delia z rôznych hľadísk:

### **1. podľa toho, či charakterizujú rovnako všetky jednotky**

**Spoločné (zhodné) znaky** – sú spoločné pre všetky jednotky súboru, slúžia na vymedzenie štatistického súboru

**Variabilné (premenlivé) znaky** – u jednotlivých štatistických jednotiek sa vyskytujú rôzne, sú vlastným predmetom štatistického skúmania.

Pr. Ako predmet štatistického výskumu zvolíme poslucháčov 1. ročníka FHPV PU v Prešove v šk. roku 2001/2002. Základným prvkom hromadného javu, a teda štatistickou jednotkou je poslucháč 1. ročníka FHPV. Všetky prvky (jednotky) štatistických súborov musia mať niektoré vlastnosti **spoločné** a ktoré sú zároveň presne definované z hľadiska vecného, priestorového, časového.

V našom prípade má každá jednotka tieto **spoločné vlastnosti (znakmi)**:

z hľadiska vecného: poslucháč 1. ročníka FHPV

z hľadiska priestorového: PU v Prešove

z hľadiska časového: školský rok 2001/2002

Takto vymedzený súbor môže skúmať štatisticky, teda realizovať vlastné štatistické výskumy. Tu sa už zaoberáme premenlivými (variabilnými) znakmi, napr. prospech poslucháčov, vek, pohlavie, záľuby ...

### **2. podľa toho, z akého hľadiska charakterizujú znaky štatistickú jednotku triedia sa tiež na:**

**Priestorové znaky** – sú dané územím, na ktorom vznikli alebo existujú štatistické jednotky. Pričom tieto môžu byť spoločné alebo variabilné. Ak je priestorový znak spoločný prispieva k vymedzeniu súboru, ak je variabilný (premenlivý, triediaci) umožňuje rozdeliť základný súbor na čiastkové súbory, ktoré sa vzájomne priestorovo odlišujú..

**Časové znaky** – sú dané časom vzniku alebo existencie štatistických jednotiek. Ako znaky spoločné vymedzujú časovo štatistický súbor. Ako znaky variabilné sú podkladom pre rozdelenie základného súboru na parciálne súbory, ktoré sa vzájomne líšia rôznym časom vzniku alebo existencie jednotiek.

**Vecné znaky** – charakterizujú najrôznejšie vecné vlastnosti jednotiek. Ako znaky spoločné slúžia k vecnému vymedzeniu štatistického súboru. Ako znaky variabilné sú podkladom pre triedenie základného súboru podľa obmien hodnôt daného vecného znaku.

Vecné znaky sa *podľa spôsobu vyjadrenia hodnôt znaku delia na:*

- **Kvalitatívne (slovné) znaky** – hodnoty znaku sú vyjadrené slovne (popisne) (národnosť, povolanie)

*Podľa počtu variantov (obmien), ktoré môžu nadobúdať hodnoty kvalitatívnych znakov rozlišujeme:*

- **Alternatívne (dichotomické) znaky** – hodnoty sa vyskytujú iba v dvoch variantoch (pohlavie, prosper-neprosper, vojak-nevojak)

- **Množné (polytomické) znaky** – hodnoty, ktoré môžu nadobúdať mnoho variantov (najmenej 3) (národnosť, zamestnanie, mzda)









- **Kvantitatívne (číselné) znaky** – hodnoty znaku sú vyjadrené číslami (vek, mzda, počet obyvateľov)

- **Priame znaky** – kvantitatívne znaky, hodnoty ktorých sú získané priamo, t.j. meraním, vážením a lebo sčítaním (počet oviec, vek v rokoch, výška v cm), výsledok je vyjadrený absolútnymi číslami.
  - **Nepriame (odvodené) znaky** – hodnoty sú získané odvodením z dvoch absolútnych čísel (hustota obyvateľstva).
  - **Extenzívne (objemové) znaky** – charakterizujú objemovú stránku štatistických jednotiek, ich veľkosť, rozsah, a pod. Údaje o nich získavame priamym meraním, vážením, spočítaním, ....
  - **Intenzívne (úrovňové) znaky** - charakterizujú úroveň (intenzitu) vlastnosti štatistickej jednotky. Zväčša sú odvodené z objemových znakov. Sú vždy pomernými (relatívnymi) číslami, ktoré získame delením dvoch absolútnych veličín. Pr. Pri skúmaní výsledkov činnosti poľnohospodárskeho podniku v Prešovskom kraji za rok 23002 boli použité tieto kvantitatívne ukazovatele: produkcia obilia, produkcia mlieka, hektárová úroda zemiakov. Prvé dva sú extenzívne znaky. Posledný je intenzívny, pretože sme ho získali z porovnania objemových veličín; celkovej úrody zemiakov a počtu hektárov, na ktorých sa zemiaky pestovali. **Základný rozdiel medzi extenzívnymi a intenzívnymi znakmi** je pri ich zoskupovaní. Extenzívne veličiny je možné sčítať, avšak súčet intenzívnych veličín nemá zmysel a preto sa sumarizujú priemerovaním.
- Po formálnej stránke sa členia kvantitatívne znaky na:*
- **Spojité znaky** – môžu nadobúdať ľubovoľné reálne hodnoty v určitom intervale, tieto znaky sú získavané meraním, môžu byť intenzívne i extenzívne znaky (spotreba auta v l)
  - **Nespojité (diskrétne) znaky** – sú to iba konkrétne celé hodnoty, tieto znaky sú získavané sčítaním (počet zamestnancov v podniku), zväčša sú extenzívne
  - **Alternatívne (dichotomické) znaky** – hodnoty sa vyskytujú iba v dvoch variantoch
  - **Množné (polytomické) znaky** – hodnoty, ktoré môžu nadobúdať mnoho variantov (najmenej 3)

**HODNOTA ATRIBÚTU ( $x_i$ )** – jednotlivé údaje štatistického znaku štatistickej jednotky

### **TECHNIKA ŠTATISTICKÉHO SPRACOVANIA**

Vo všeobecnosti je možné technické postupy (metódy) rozdeliť do dvoch veľkých skupín na metódy ručného spracovania a metódy mechanizovaného počítačového spracovania.

- ◆ **RUČNÉ SPRACOVANIE** – pomocou bežnej kancelárskej techniky a bezprostrednej osobnej účasti spracovateľa
  - **Čiarková metóda** – najjednoduchšia metóda ručného spracovania. Čiarky sa zapisujú tak, že každú piatu jednotku v skupine označíme preškrtnutím predchádzajúcich štyroch (     || ). Taksa dajú ľahšie spočítať. Zápis pri viac ako 4-6 čiarkach je nepriehľadný (     ).
  - **Metóda ručných štítkov** – veľmi práca, dnes už neefektívna metóda (napr. evidencia údajov o zamestnancoch). Pre každú štatistickú jednotku sa vytvorí samostatný štítok (kartička), na okraj ktorého sú postupne uvedú údaje o jednotlivých štatistických znakoch. Podľa nich sa rozčleňujú štítky na potrebné skupiny a spočítavajú sa.
  - **Metóda sústredovacích (koncentračných) tabuliek** – štatistický súbor sa rozdelil na niekoľko skupín podľa určitého znaku, či znakov a pre každú takto definovanú skupinu sa vytvorila samostatná tabuľka, do ktorej sa zapísali ďalšie údaje o štatistických jednotkách, patriacich do danej skupiny.
- ◆ **AUTOMATIZOVANÉ SPRACOVANIE** – pomocou počítačovej techniky
  - **počítačové programy** pre štatistické spracovanie údajov napr.: MS Excel (z balíka MS Office), StatGraphic.

## ETAPY ŠTATISTICKÉHO VÝSKUMU

Štatistický výskum sa skladá z troch etáp:

**1. ŠTATISTICKÉ ZISŤOVANIE** – proces získavania a zhromažďovania údajov o skúmaných javoch.

### 2. ŠTATISTICKÉ SPRACOVANIE

➤ **Triedenie údajov** – prvý krok v spracovaní štatistických údajov

⇒ Pod triedením rozumieme **usporiadanie** štatistického súboru (rozdelenie jeho jednotiek) do **skupín** (tried), podľa určitého štatistického znaku alebo znakov (**triediaci atribút**).

→ Ak sa pri triedení použije **len jeden triediaci znak**, ide o **jednostupňové triedenie**.

→ Ak sa štatistický súbor triedi podľa **dvoch alebo viacerých znakov**, ide o **dvojstupňové**, resp. **viacstupňové triedenie**.

⇒ V širšom slova zmysle sa pod ním chápe **usporiadanie** štatistického súboru (rozdelenie jeho jednotiek) z **časového, priestorového alebo vecného hľadiska**.

⇒ **Hlavné pravidlá triedenia** (musia byť dodržané pri každom triedení):

→ **Pravidlo úplnosti** – skupiny vytvárame tak, aby každá štatistická jednotka zo súboru bola zaradená do niektorej z nich. Ak určitý prvok nemôžeme zaradiť do nijakej z vytvorených skupín, vytvoríme ďalšiu skupinu „Ostatné“ „Iné“ a pod.

→ **Pravidlo jednoznačnosti triedenia** - skupiny sa nesmú navzájom prekrývať (jeden prvok môže patriť len do jednej skupiny).

➤ **Opis štatistického súboru pomocou kvantitatívnych charakteristík**

➤ **Vyjadrenie podstatných znakov štatistického súboru prostredníctvom tabuliek, grafov**

### 3. ŠTATISTICKÝ ROZBOR – konečná fáza štatistického výskumu, ktorá zahŕňa:

➤ analýzu výsledkov štatistického spracovania

➤ formulovanie záverov, ktoré z nich vyplývajú.

## ŠTATISTICKÉ TABUĽKY

ŠTATISTICKÉ TABUĽKY - vhodný a najčastejšie používaný prostriedok na prehľadnú prezentáciu výsledkov štatistického výskumu.

### KOMPOZÍCIA TABUĽIEK

♦ **Názov** - obsahuje **vecné, časové a priestorové vymedzenie** štatistických jednotiek, o ktorých poskytuje informáciu

♦ **Textová časť** - tvorí ju **hlavička a legenda**

♦ **Číselná časť** - obsahuje **pole políčok** (buniek, cell), usporiadaných do **riadkov a stĺpcov**. Číselné údaje sa zapisujú tak, aby boli jednotky pod jednotkami, desiatky pod desiatkami, desatinné čiarky pod desatinnými čiarkami a pod. \_

➤ **Konvenčné (dohodnuté) označenia** v políčkach štatistických tabuliek:

→ - jav nejestvuje (nevyskytuje sa)

→ **0** hodnota je menšia ako polovica mernej jednotky

→ . údaj je neznámy (nespoľahlivý)

→ **x** logicky nemožný údaj

→ **!** oprava oproti predchádzajúcemu vydaniu

### ZÁKLADNÉ DRUHY TABUĽIEK

♦ **PODĽA OBSAHU** sa rozlišujú štatistické tabuľky:

➤ **jednoduché** – obsahujú údaje usporiadané podľa jedného znaku

Názov tabuľky

Názov legendy	Názov stĺpca	.....	
<i>a</i>	<i>1</i>	<i>2</i>	
<i>Legenda</i>	<i>políčko</i>		

Hlavička



- **kombinované (kombinačné)** – vyjadrujú údaje súčasne triedené podľa dvoch alebo viacerých znakov

<i>a</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>

→ **asociačné a kontingenčné** – vyjadrujú výsledky triedenia podľa dvoch kvalitatívnych znakov

→ **korelačné** – vyjadrujú výsledky triedenia súčasne podľa dvoch kvantitatívnych znakov

		Hodnoty $x_i$			
$y_i$					

- **skupinové** – prvky sú rozdelené do skupín

		1. skupina		2. skupina		
<i>a</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	
<i>Legenda</i>						

Hlavička



◆ **POĎĽA ZAČLENENIA V ŠTATISTICKOM PROCESE (ÚČELU)** sa rozlišujú štatistické tabuľky:

- **pomocné** (pracovné) – zhromažďujú sa v nich štatistické údaje v priebehu štatistického výskumu
- **koncentračné** – slúžia na sústredenie údajov z niekoľkých pracovných tabuliek
- **výsledné** (konečné) – obsahujú výsledky štatistického výskumu, tvoria súčasť záverečných správ
  - vyčerpávajúce
  - prehľadné
  - špeciálne

◆ **DATABÁZOVÉ TABUĽKY** – špecifický prípad tabuliek, ktoré sú súčasťou počítačových databázových informačných systémov (napr. formáty DBASE, Acces). Klasická tabuľka je pretransformovaná do digitálnej podoby, ktorá umožňuje na základe zvolených kritérií (požiadaviek), vyhľadávať konkrétne hodnoty, usporadúvať ich a triediť. Je to realizované pomocou jazyka **SQL** (Standard

Query Language). Údaje organizované v takýchto databázových tabuľkách sú označované ako **relačné databázy (relačné údajové bázy)**.