

# Zpracování a interpretace údajů

přednáška k Metodologii, K.Vlčková, JS 2003

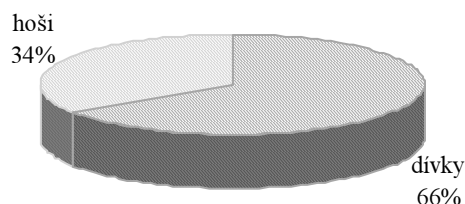
## 1. Zpracování údajů

- po všech těch přípravných fázích formulace problému, studia literatury, formulování hypotéz, vymezení základních pojmů, tvorby výzkumného nástroje, volby výzkumného vzorku, ověřování konstrukce a vlastností výzkumného nástroje aj. v předvýzkumu, po sběru dat konečně nastává naprosto odlišná fáze, dosti mechanická, a to fáze zpracování dat
- o tom, jak bude vypadat zpracování dat rozhoduje to, zda jsme dělali kvalitativní či kvantitativní výzkum
- u kvalitativního budeme získaná data třídít, kategorizovat, kvalitativní analýza (typické, reprezentativní příklady X atypické), interpretovat, vysvětlovat
- u kvantit. výzkumu bylo o způsobu zpracování dat bylo už rozhodnuto předem, než se začala sbírat data!!!
- způsob zprac. byl také ozkoušen v předvýzkumu
- nejprve se data předzpracovávají, připravují pro zpracování, zvažují se možné kategorizace dat, zadávají se proměnné a jejich hodnoty do hlavičky databáze, pak se zadávají data převedená většinou do čísel
- vše už rovnou psát do počítače, do (nejlépe) statistického programu nebo alespoň do Excelu, z něj se to dá převést do SPSS, Statistica aj.
- databáze je uspořádaný soubor dat (uzavřený soubor dat X u longitudiálního výzkumu – otevřený), utřídění dat do databáze lépe umožňuje výsledky statisticky zpracovat, databázi zkontrolovat, bude tam dost chyb (např. vytisknout) atd.
- je důležité znát, o jaký typ proměnné se jedná, zda je kategoriální či měřitelná, ve statistických programech se to zadává, ovlivňuje to naše možnosti jaké stat. metody můžeme použít

### NOMINÁLNÍ

- nejméně kvalitní, počítají se jen četnosti
- nejjednodušší forma přiřazení čísla charakteristikám proměnných => nejméně inf. nám to přinese, je to jen kategorizace – jednoduchá třídění do vzájemně se vylučujících kategorií
- př. **typ školy, pohlaví, věk, vzdělání**, druh motivace,
- číselné označení kategorií muž 2, žena 1 – neznamena určitou měřitelnou hodnotu, napomáhá pouze klasifikaci dat, místo 1,2 jsi lze dát do databáze Ž,M
- v dosti omezené míře lze statisticky zpracovávat – lze spočítat absolutní četnosti a relativní četnosti (% zastoupení jednotl. kategorií v celku)
- nelze počítat průměry, SD, korelace
- lze použít analýzy rozptylu, výpočtů chí-kvadrátu

### Poměr dívek a hochů ve výzkumu



### Frekvenční tabulka rozložení pohlaví

pohlaví	absolutní četnosti	kumulativní četnosti	relativní četnosti (%)
dívky	150	150	65,5
hoši	79	229	34,5
celkem	229	229	100,0

### ORDINÁLNÍ

- pořadí prvků je známo (př. pořadí v běhu)
- nevím ale nic o rozdílech v jednotlivcích
- nelze sčítat, odčítat, lze jen porovnávat rozdíl, co je víc a co míň
- v pedagogice často nelze určit přesné hodnoty (píle žájů, snaha, míra spolupráce)  
=> jen se relativním způsobem posoudí daná hodnota ve srovnání s jejich výskytem u jiných jedinců => nejde o měření absolutních hodnot
- př. snahu 1 žáka srovnáme se snahou dalších, dostaneme škálu od min. po max. a seřadíme žáky do pořadí, známe jejich pořadovou hodnotu, intervaly mohou být různé, ale to pořadové měření nezjistí
- př. známky, výkon sportovců,
- více možností pro stat. zprac. dat
- Me, směrodat. odchylka, kvartil
- můžeme zjišťovat těsnost vztahů mezi růz. proměnnými – používá se postupů adaptovaných pro tuto úroveň měření – Spearmanův pořadový korelační koeficient

### INTERVALOVÉ

- **lze sčítat i odčítat**, víme, zda je to větší či menší a o kolik, známe rozdíly mezi sousedními body + jsou konstantní
- př. teplota na Celsiově stupnici, má 0 (ale jen *dohodnutou* – teplota 0°C neznamená, že není teplota X 0 bodů v testu neznamená, že nemá žádné vědomosti, jen to, že nestačily na ten test
- nesmí se dělit (př. byl 2x lepší), násobit
- lze počítat průměry, směrodat. odchylky, parametrické testy rozdílů, Pearsonův korelační koeficient, regrese atd., pak i faktorová analýza ad.

### Deskriptivní tabulka strategií

průměr	Me	Mo	četnost	min.	max.	dolní	horní	percentil	percentil	rozptyl	std.	std.
--------	----	----	---------	------	------	-------	-------	-----------	-----------	---------	------	------

			<i>modu</i>			<i>kvartil</i>	<i>kvartil</i>	<i>10</i>	<i>90</i>		<i>odchylka</i>	<i>chyba</i>
2,82	2,86	3,07	10	1,49	4,14	2,53	3,09	2,26	3,37	0,19	0,44	0,03

### *Korelační matice věku a používání nepřímých strategií učení*

	<i>nepřímé strategie</i>	<i>metakognitivní</i>	<i>afektivní</i>	<i>sociální</i>
<i>Pearsonův korelační koeficient</i>	0,108	0,110	0,085	0,063
<i>hladina významnosti</i>	<i>p = 0,105</i>	<i>p = 0,097</i>	<i>p = 0,200</i>	<i>p = 0,341</i>

### POMĚROVÉ

- číslo, které jev zastupuje se mu blíží tak dokonale, že vyjadřuje míru vlastnosti, kt. měří
- poměr intervalů mezi 2 sousedními body škály je stejný jako mezi dalšími
- **lze násobit i dělit**
- u ped. jevů téměř nikdy (max. jen měření výšky, váhy)
- má to reálnou nulu, byť by v praxi byla nedosažitelná (el. odpor, teplota)
- geometrický průměr, variační koeficient atd. - všechny stat. m. pro intervalové měření

Známky ve škole jsou v podstatě nominální, max. ordinální – mezi stupni není stejná vzdálenost (NE intervalové)

### Zpracování údajů statistickými postupy

- zpracování utříděných dat sám nebo se statistikem – statistická analýza
- **1/ primární zpracování dat (třídění 1. stupně)** – zpracování skupin dat, zjišťujeme *absolutní a relativní četnosti, průměr, Me, směrodatné odchylky* u jednotl. proměnných
- **2/ sekundární zpracování dat, třídění 2. stupně** – zjišťují se vazby mezi jednotlivými proměnnými, příp. jejich skupinami => výpočty *korelací, regresí, použití různých variant neparametrických výpočtů, faktorovou analýzu, trsovou analýzu atd.*,
- testují se rozdíly mezi proměnnými, skupinami apod. (Studentův t-test nebo testem chí-kvadrát), uvede se, zda výsledky jsou nebo nejsou statisticky významné
- viz statistické příručky
- programy: Excel (je v balíku Microsoft Office), statistické softwary – SPSS, Statistica, Statgraphic, Origin aj. => vypočítají výsledky a umožňují i grafické znázornění výsledků, po zacvičení je práce s nimi velmi jednoduchá a rychlá, umožňuje zkoušet různé možnosti výpočtů a vytěžit z údajů maximum
- pozor: počítač nebude protestovat, když ho budete nutit zpracovat údaje pro daný účel naprosto nevhodnou statistickou metodou, za výběr metod a interpretaci ručí výzkumník, lze požádat o pomoc matematika, statistika, sociology aj.
- nepsát si údaje prve na papír, ale rovnou např. do Excelu

Ritomský, A. 1998, Ritomský, A., Hanks, L. 1994 – příručky pro používání SPSS

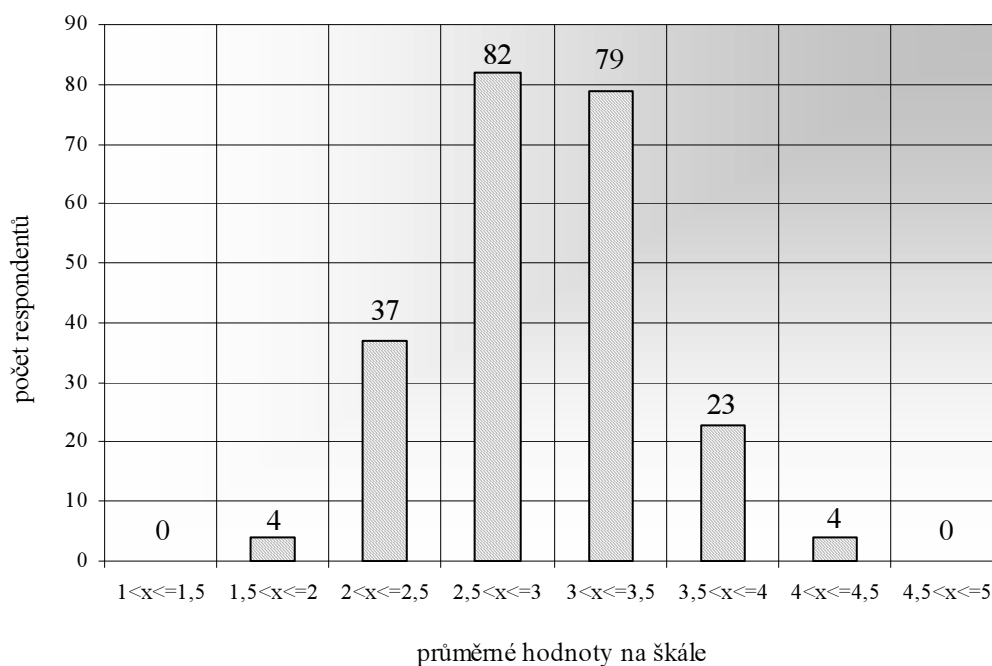
Kerlinger, Chráska, Disman

## Prezentace dat v práci

zpracování dat = uspořádání a shrnutí dat, jejich transformace do grafů a tabulek (= přehledná, úsporná forma prezentování údajů, je třeba zdůraznit důležitá zjištění (ta, kt. podporují očekávané trendy nebo naopak údaje, kt. nebyly očekávány

- údaje lze různě přeskupovat a kombinovat, lze vyrobit velké množství tabulek a grafů => vybrat jen rozumné množství, ve zprávě z výzkumu uvést jen podstatné výsledky vzhledem k cíli výzkumu
  - příliš velké množství tabulek ukazuje, že se výzkumník v datech ztratil, neumí najít správnou hierarchii a proto uvedl vše, co měl k dispozici
  - výzkumy z větším množstvím proměnných obvykle vyžadují větší počet tabulek než jednodušší výzkumy
  - úlohu hraje i žánr textu, v němž se výsledky publikují (do článku se vejde méně tabulek a grafů než do výzkumné práce)
  - disertační, diplomové práce – hlavní tabulky jsou v příslušné části o zpracování údajů, doplňující tabulky jsou v příloze
  - pořadí tabulek a grafů A/ nejprve ty, kt. obsahují hlavní a souhrnné informace => čtenář získá globální přehled o výsledcích, pak se hlavní výsledky přeměňují na drobné B/ tematické řazení – dle výzkumného problému a hypotéz, má-li výzkum 4 hypotézy, výsledky budou seřazeny do 4 okruhů
  - styl psaní – odborný, dosti suchý, neosobní, pro účely zábavného čtení jsou populárně-vědecké publikace, psát srozumitelně, s ohledem na čtenáře, nikoli komplikovaně, inspirovat se autory, kt. mají vhodný styl – i složité teoretické věci říkají jasně a jednoduše (Průcha, Gavora, Jan Slavík, Jiří Mareš), psát v první osobě mn.č. a v minulém čase (př. Mezi žáky nebyl žádný rozdíl X není = neomezená platnost, té ale nelze dosáhnout)
- ⇒ přehlednost grafů a tabulek
  - ⇒ srovnávání vhodných skupin v komentáři ke grafům
  - ⇒ komentář není převod čísel do slov, je třeba uplatnit nadhled
  - ⇒ vyjádřit se ke svým hypotézám (očekávám, předpokládám...)
  - ⇒ tematicky řadit údaje, tabulky a grafy
  - ⇒ rozlišit jasně samotné údaje a svou interpretaci údajů – jde o vhodné formulace
  - ⇒ srovnat své závěry s údaji z předcházejících výzkumů

### Používání metakognitivních strategií



Př.

### Úmrtnost zapříčiněná motorovými vozidly

věk	muži	ženy
1 - 4	10,5	8,0
5 - 14	10,4	5,4
15 - 19	54,2	16,4
20 - 24	76,3	12,7
25 - 44	35,6	9,1
45 - 64	33,1	12,9
65 a více	58,4	12,9
celková	32,9	11,1

pozn. (úmrtnost ze všech příčin je 100%)

Můžeme následující výroky prostřednictvím údajů z tabulky A) potvrdit, B) nemůžeme je potvrdit nebo C) je nemůžeme popřít ani potvrdit?

- Úmrtnost zapříčiněná motorovými vozidly je vyšší u mužů než u žen.
- Nehody zapříčiněné motorovými vozidly jsou hlavní příčinou úmrtí lidí ve věku 20 až 24 let.
- Muži nad 65 let jezdí bezpečněji než mládež ve věku 15 – 19 let.
- Největší počet úmrtí zapříčiněných motorovými vozidly je ve věku 65 let a starších.
- U celkového součtu jen asi 11% úmrtí žen zavinila motorová vozidla.

### 3. Interpretace dat

- zpracované údaje jsou jen holými čísly, sama o sobě moc neznamenaají
- interpretace = vysvětlení a vyhodnocení
- hlavním výstupem výzkumu nejsou údaje, ale jejich interpretace
- slovní popis není interpretace
- v kapitole „výsledky a interpretace“, „diskuse a závěry“
- po zpracování údajů je na chvíli odložit, interpretace vyžaduje nadhled, je to jiný druh činnosti než zpracovávání, vyžaduje jiné naladění, hluboké zažití výsledků, jinak se člověk mezi čísly ztrácí
- vyžaduje pochopení číselných údajů získaných z matematicko-statistických výpočtů a zároveň velký přehled a dobrou orientaci ve zkoumané problematice
- začátečníci mívají s interpretací velké problémy, často je to nejtěžší etapa výzkumu (někdy důsledek zanedbání studia problematiky na začátku výzkumu – jsou-li výzkumné hypotézy postaveny špatně, potom se i obtížně interpretují (jdou-li vůbec zpracovat) a naopak.
- interpretace – srovnávat údaje mezi sebou a ptát se např.
  - a) Vyplývají z tohoto srovnání nějaké souvislosti? Jsou v údajích nějaké diskrepance? Jak se dají vysvětlit?
  - b) Vyjadřují údaje nějaký trend, směřování, linii, anebo jsou spíše rozházené?
  - c) Jsou údaje v souladu s existující teorií o zkoumaném jevu?
  - d) Jsou údaje v souladu s údaji z jiných výzkumů? Nejsou-li, proč? Bylo to proto, že šlo o jinou populaci, jiné období, nebo proto, že výsledky byly zpracovány jiným způsobem? Anebo to bylo proto, že z údajů „vystoupily“ neznámé, nekontrolované proměnné?
- naše zjištění konfrontujeme se stanovenými hypotézami a komentujeme
- opíráme se o existující ped. teorii + své zkušenosti (viz hypotézy), na základě nových zjištění hypotézy přehodnocujeme a dále rozvíjíme
- vyjádřit se o podmínkách a rozsahu platnosti hypotéz
- vyjádřit se k tomu, zda se dají závěry široce zevšeobecňovat, nebo platí jen pro určitou omezenou populaci
- chyby u začínajících:
  - vedle legitimních, vytváří nelegitimní závěry – neopírající se o předcházející zjištění, přílišná zevšeobecnění (na základě zjištění o parciálních nedostatcích učitele, udělají urychlený závěr o celkově špatné práci učitele, školy), moralizování (dávají tam svůj světový názor, působí originálně X je to projev nedisciplinovanosti v myšlení)

<p><i>Zásady interpretace údajů:</i> Udělejte zřejmé zřejmým. Udělejte zřejmé pochybným. Udělejte skryté zřejmým.</p>
---

### **Zpracováno podle:**

GAVORA, P. *Úvod do pedagogického výzkumu*. Brno : Paido, 2000. ISBN 80-85931-79-6, s. 134 –139.

### **Doporučená literatura k vybraným okruhům (pro zpracovávání bakalářských a diplomových prací):**

#### **Validita a reliabilita výzkumné metody, korelační koeficient:**

GAVORA, P. *Výzkumné metody v pedagogice. Příručka pro studenty, učitele a výzkumné pracovníky*. Brno : Paido, 1996. ISBN 80-85931-15-X, s. 11- 15, 60 – 61 aj.

GAVORA, P. *Úvod do pedagogického výzkumu*. Brno : Paido, 2000. ISBN 80-85931-79-6, s. 71 – 75.

#### **Škálování a vyhodnocování škál:**

GAVORA, P. *Výzkumné metody v pedagogice. Příručka pro studenty, učitele a výzkumné pracovníky*. Brno : Paido, 1996. ISBN 80-85931-15-X, s. 42 – 53

GAVORA, P. *Úvod do pedagogického výzkumu*. Brno : Paido, 2000. ISBN 80-85931-79-6, s. 96 –98.

#### **Základní, srozumitelné vysvětlení podstaty statistiky (bez matematických vzorců) pro každého:**

DISMAN, MIROSLAV *Jak se vyrábí sociologická znalost*. Praha : Karolinum, 2000. ISBN 80-246-0139-7, s. 180 – 282.

deskriptivní (popisná) statistika:

- jaké střední hodnoty zjišťujeme u jednotlivých druhů proměnných (intervalová data – aritmetický průměr, pořadová data – medián, nominální data – modus)
- rozptyl, směrodatná odchylka

bivariační analýzy (zjišťování vztahů, souvislostí mezi proměnnými)

- souvislosti, korelace, regrese
- statistická významnost, míra omylu

multivariační analýzy (více proměnných)

- kontrola dalším faktorem, parciální (dílčí) korelace, parciální regresní koeficient, mnohonásobný korelační koeficient, Path analýza, diskriminační analýza, faktorová analýza,

#### **další autoři:**

Hendl, Jan *Přehled statistických metod zpracování dat*. Praha : Portál, 2004. ISBN 80-7178-820-1

F. Kerlinger, M. Chráska, či knihy o statistice pro pedagogy