



Metodologie pro ISK 2
Úvod do práce s daty

Základní termíny

- **Proměnná (znak)** – měřitelná dimenze pojmu
- **Hodnota proměnné** – vyjádření stavu zkoumané proměnné
- 3 podmínky:
 - Rozlišitelnost
 - Úplnost
 - Jednoznačnost

Základní termíny – druhy proměnných

- **Nominální proměnné**
(nelze je uspořádat hierarchicky či podle velikosti)
- **Ordinální proměnné**
(jedna hodnota je vyšší než druhá, nemůžeme však s jistotou tvrdit, o kolik je vyšší)
- **Kardinální proměnné**
(dokážeme říci, o kolik přesně je jedna hodnota větší, než jiná)

Základní termíny – druhy proměnných

- **Kategorizovaná data:** nominální, ordinální i kardinální proměnné s malým počtem variant (pohlaví, vzdělání, počet dětí)
- **Nekategorizovaná data:** kardinální proměnné s velkým počtem variant (plat)

• **2. Považujete obor Informační studia a knihovnictví za perspektivní?**

- | | |
|---------------------------|----|
| • velmi perspektivní | 1 |
| • spíše perspektivní | 2 |
| • spíše neperspektivní | 3 |
| • zcela neperspektivní | 4 |
| • nevím, nemohu odpovědět | -1 |
| • neodpověděl/a | -2 |

Chybějící hodnoty
(missing values)

Hodnoty
proměnné
okódované

Kontrola dat: GIGO

- „Garbage in, garbage out!“



Kontrola dat: GIGO

- **Chyby při zpracování**
 - Posuny desetinných čárek
 - Záměna znaků (nepoužívat „0“ pro missing values)
 - Chyby při přepisování
- **Kontrola dat**
 - Kontrola okem („vytisknout a řádek po řádku zkontrolovat“ *Hendl, 2009*)
 - Předběžné grafické zobrazení dat

Chyby v zápisu

Jméno tazatele				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	10	1,9	1,9	1,9
A: Pangráčová	1	,2	,2	2,1
A. Halfarová	10	1,9	1,9	4,0
A. Pangráčová	9	1,7	1,7	5,7
B. Turek	10	1,9	1,9	7,5
D. Tomančáková	16	3,0	3,0	10,6
E. Sedláková	10	1,9	1,9	12,5
I. Burdová	10	1,9	1,9	14,3
I. Michálková	10	1,9	1,9	16,2
J, Hlavinková	1	,2	,2	16,4
J. Foral	10	1,9	1,9	18,3
J. Fryš	10	1,9	1,9	20,2
J. Galuszka	10	1,9	1,9	22,1
J. Hlavinková	9	1,7	1,7	23,8

Co s missing values?

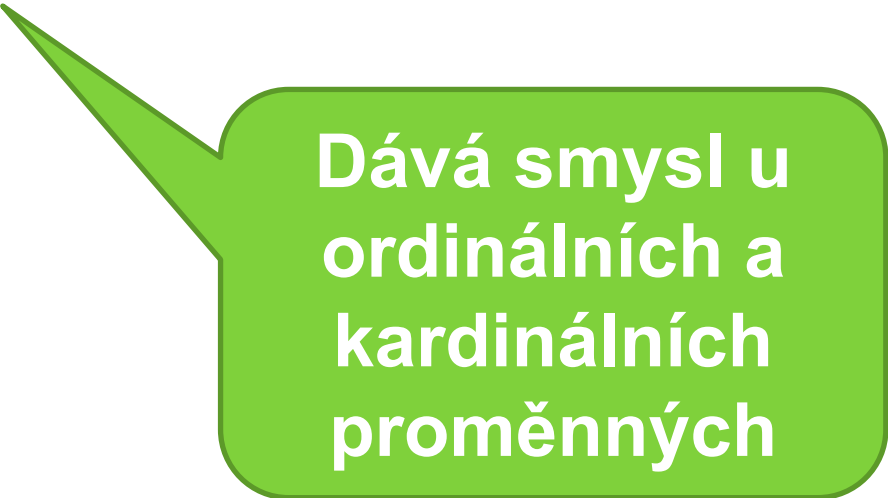
- **Hodnota „neodpověď/a“**
- Jsou výskyty rozděleny náhodně?
- Způsoby kontroly:
 - rozdělit soubor na skupiny záznamů s missing values a bez nich, porovnat charakteristiky obou souborů
 - Kontrola korelací vyplnění/nevyplnění s jinou proměnou

Co s missing values?

- Pokud potřebujeme pracovat jen s úplnými záznamy:
 - Vymazání jednotky
 - Vymazání proměnné
 - Imputace
 - Odhad scházející hodnoty (regresní analýza)
 - Nahrazení průměrnou hodnotou subpopulace

Základní popis kategorizovaných dat

- Absolutní četnost
- Relativní četnost
- Kumulativní relativní četnost

A green speech bubble with a white border and a tail pointing towards the top-left. It contains white text.

Dává smysl u
ordinálních a
kardinálních
proměnných

Základní popis kategorizovaných dat

- Tabulky četností – plná verze

<i>Jaké je Vaše vzdělání?</i>		Četnost odpovědí	Relativní četnost	Validní relativní četnost
Validní hodnoty	Základní	46	7,5 %	7,6 %
	Základní vyučen /střední bez maturity	62	10,1 %	10,2 %
	Střední s maturitou	307	50,1 %	50,5 %
	Pomaturitní nastavba, VOŠ	40	6,5 %	6,6 %
	Vysokoškolské	153	25,0 %	25,2 %
	Celkem validní hodnoty	608	99,2 %	100,0 %
Chybějící hodnoty (nevl, neodpověděl/a)	Chybějící hodnoty	5	0,8 %	
Celkem		613	100,0 %	

Základní popis kategorizovaných dat

- Tabulky četností – zkrácená verze

<i>Jaké je Vaše vzdělání?</i>	Četnost odpovědí	Validní relativní četnost
Základní	46	7,6 %
Základní vyučen /střední bez maturity	62	10,2 %
Střední s maturitou	307	50,5 %
Pomaturitní nástavba, VOŠ	40	6,6 %
Vysokoškolské	153	25,2 %
Celkem	608	100,0 %

Zobrazování kategorizovaných dat

- **Tabulka četností**
 - První přehled výsledků měření
 - Vhodné pro uvedení přesných čísel
 - Absolutní / relativní / validní / kumulativní četnosti
- **Graf četností**
 - Názornější
 - Používají se validní četnosti
 - Osa X: hodnoty proměnné
 - Osa Y: četnosti

Zobrazování kategorizovaných dat

Tipy pro výběr grafu:

- vždy si vyberte ten nejjednodušší graf
- vždy používejte popisek grafu
- popište obě osy
- vyvarujte se ilustrací a zbytečného používání grafiky na pozadí nebo okrajích grafu
- vyvarujte se používání módních piktogramů, které by mohly ztížit čitelnost dat
- vertikální osa by měla začínat nulou (pokud nezačíná negativními hodnotami)

Základní popis kategorizovaných dat

- Míry centrální tendence
- Nominální proměnná: **MODUS**
- Ordinální proměnná: **MODUS, MEDIÁN**

MODUS je hodnota, která se v datech vyskytuje nejčastěji.

MEDIÁN dělí řadu výsledků seřazených podle velikosti na dvě stejně početné poloviny

MEDIÁNOVÁ KATEGORIE je ta, ve které je dosaženo 50% všech údajů, postupujeme-li od první kategorie výše.

Základní popis nekategorizovaných dat

- Maximum
- Minimum
- MODUS
- MEDIAN
- ARITM. PRŮMĚR
- Rozptyl a směrodatná odchylka

Průměr vs. medián

- Průměrná mzda 2013: **26 637 Kč**
 - **Muži:**
 - **Ženy:**
- Medián mzdy: **22 288 Kč**
 - **Muži:** 24 330 Kč
 - **Ženy:** 20 029 Kč

Rozptyl a směrodatná odchylka

- Rozptyl – střední hodnota kvadrátů odchylek od střední hodnoty (průměru)
- Směrodatná odchylka – odmocnina z rozptylu

Rozptyl a směrodatná odchylka

- Směrodatná odchylka nám řekne, jak moc se hodnocení všech respondentů pohybovalo kolem průměru
- Příklad:

```
Zajímavost předmětu          není vůbec zajímavý .....(*)*X** je velmi zajímavý
Přínosnost předmětu          není vůbec přínosný .....(*)**X** je velmi přínosný
Obtížnost obsahu              velmi snadný .....(*)*X** velmi obtížný
Náročnost na přípravu        velmi snadný .....(.)**X** velmi obtížný
Dostupnost studijních zdrojů velmi špatně dostupné .....(.)**X** velmi dobře dostupné
Jak učitel učí                velmi špatný .....(*)**X** vynikající
Učitel jako odborník         není odborníkem .....(.)..*X* je odborníkem
```

Excel vs. SPSS

The SPSS logo is centered within a solid red square. The text "SPSS" is rendered in a bold, white, sans-serif font. A small registered trademark symbol (®) is positioned to the upper right of the final 'S'.

SPSS®