

## Analýza dat z dotazníkových šetření

### Cvičení 2.

- Jednorozměrné třídění

Zdrojová data: dotazník <http://www.vyplnto.cz/realizovane-pruzkumy/konzumace-ryb-a-rybich-vyrob/>

- Seznamte se s dotazníkem a strukturou otázek, zamyslete se nad vhodností jednotlivých odpovědí a škál
- Stáhněte si data z dotazníku ve formátu \*.csv a otevřete je v aplikaci Excel, zamyslete se, která data by bylo vhodnější upravit přímo v Excelu -> uložte dotazník ve formátu \*.xls a otevřete datový soubor v prostředí SPSS

### Výstupy podle typu proměnné:

kategoriální proměnné nominální	tabulka a graf četností (sloupcový, výsečový) modus
kategoriální proměnné ordinální	tabulka a graf četností (sloupcový, spojnicový) modus a medián
diskrétní číselné proměnné a škály	aritmetický průměr, modus a medián směrodatná odchylka intervaly spolehlivosti, error bar
spojité číselné proměnné	graf intervalových četností (histogram) aritmetický průměr, směrodatná odchylka kvartilové charakteristiky, box-plot intervaly spolehlivosti, error bar graf normality rozdělení (P-P graf)
alternativní proměnné	tabulka a graf četností (sloupcový, výsečový) aritmetický průměr, modus směrodatná odchylka intervaly spolehlivosti, error bar
baterie škál (alternativních proměnných)	tabulka a graf průměrů (sloupcový, spojnicový) intervaly spolehlivosti, error bar

Modus = hodnota vyskytující se nejčastěji s největší četností

Medián = dělí řadu vzestupně seřazených výsledků na dvě stejné početné poloviny

### 1. Kategoriální proměnná nominální: (Tabulka a graf četností)

- řádky tabulky představují jednotlivé třídy (kategorie)
- sloupce tabulky vyjadřují četnosti (počty jednotek)

**Př.:** sloupec (PokudanoJakčasto) -> *Analyze* -> *Descriptive statistics* -> *Frequencies*

- Výběr proměnné (proměnných) – Pokud ano, jak často? – dále výběr co chceme zobrazit tlačítka: *Statistics* (střední hodnoty, odchylky, kvartily, medián, atd...), *Charts* (grafy – teď vybereme Bar Chart – histogram), *Format, Style, Bootstrap*.
- Provedli jsme výběr - výstup tabulka + graf (co je špatně, podezřelé???)

Ve sloupci je několikrát uvedena prázdná hodnota (pokud respondent v předchozí otázce odpověděl ne, tato otázka nebyla vůbec položena -> proč jsou tato data vyhodnocena?)

Tyto prázdné buňky musí z analýzy odstranit:

- Vymažeme celé řádky (velmi nevhodné znehodnotíme část dotazníku a můžeme přijít o cenná data, případně o cenné souvislosti pro další šetření)
- Upravím data už v Excelu (formátování přes **KDYŽ** nebo **IF**)
- Nadefinuji **Missing Values** přímo v SPSS (v tuto chvíli musím data „natvrdo“ přepsat a do prázdných buněk napsat „0“), pak přes *Variable View*, sloupec *Missing* -> *Discrete missing values* -> 0

Znovu provedeme vyhodnocení: *Analyze* -> *Descriptive statistics* -> *Frequencies* –  
 - Pokud ano, jak často?

Co znamenají jednotlivé sloupce tabulky?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	denně	1	,5	,5	,5
	jednou týdně	68	34,0	36,8	37,3
	jednou za měsíc	45	22,5	24,3	61,6
	několikrát týdně	33	16,5	17,8	79,5
	příležitostně	38	19,0	20,5	100,0
	Total	185	92,5	100,0	
Missing	0	15	7,5		
Total		200	100,0		

- **Missing** – počet nadefinovaných chybějících hodnot!
- **Frequency** – absolutní četnost jednotlivých proměnných
- **Percent** - procentuální zastoupení vzhledem ke VŠEM hodnotám i z neplatných
- **Valid Percent** – Relativní četnost z platných (Valid) hodnot
- **Cumulative percent** – Kumulativní četnost

Grafickým výstupem 2 typy grafů histogram (**Bar chart**) a koláčový graf (**Pie chart**).

**Pozor!!!** U Pie chart, vždy mě zajímá, z kolika dat byl graf tvořen a kolik dat (procent) je zastoupeno v jednotlivých výsečích!!!

Absolutní a relativní četnost:

### rozdělení četností

– absolutní četnosti (count)  $n_i$

$$p_i = \frac{n_i}{N}$$

– relativní četnosti (percentage)  $p_i$

*ve výběrových šetřeních se uvádějí relativní četnosti*

### kumulativní četnosti

*pouze u uspořádaných znaků (ordinální, metrické)*

$$F_i = p_1 + p_2 + \dots + p_i$$

*např. „kolik procent respondentů konzumuje ryby alespoň jednou za měsíc?“ (61,6)*

## Transformace vstupních dat

filtrace souboru (*Select Cases*):

- zkoumáme pouze jednotky s danou vlastností
- filtrační proměnná (podle ní vybíráme, např. věk)
- filtrační podmínka (např. věk  $\geq 18$ )

*příklad: analyzujeme pouze dospělé respondenty*

rozdělení souboru (*Split File*):

- rozdělení na podsoubory – dvourozměrné třídění
- porovnání vlastností podsouborů
- rozdělující proměnná (např. pohlaví)

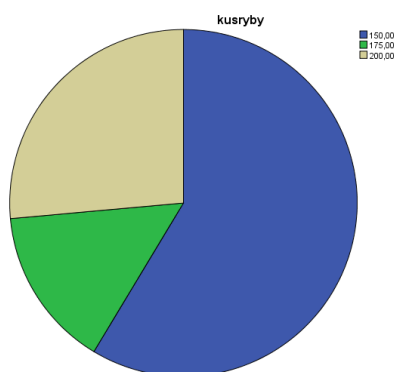
*příklad: srovnání odpovědí mužů a žen*

**Př.:** Filtrace souboru (*Select Cases*)

Sloupec *Jakou částku jste ochotni utratit za jeden kus ryby* -> nejprve nutnost překódovat na numerickou proměnnou (*Recode into Different Variables*).

Zajímají nás respondenti, kteří jsou ochotni utratit za jeden kus ryby více než 100 Kč. Kolik jich je? Kolik z nich je ochotno utratit i více peněz za jeden kus ryby?

*Data* -> *Select cases* -> **If condition is satisfied** -> If.. -> *kusryby* > 100



Filtr zobrazí pouze ty, kteří jsou ochotni dát za rybu více jak 100 Kč.

**Př.:** rozdělení souboru (*Split File*)...

Sloupec *Dávají podle vašeho názoru přednost spíše kaprům nebo řízkům...* nejprve provedeme analýzu bez rozdělení souboru (53% řízky; 47% kapr)

Analýza v závislosti na pohlaví: *Data* -> *Split file* -> **Compare groups** -> Vaše pohlaví  
Opět *Analyze* -> *Frequencies* ->

*Dávají podle vašeho názoru přednost spíše kaprům nebo řízkům...*

(muž: 55% kapr, 43% řízek; žena: 44% kapr, 55% řízek)

Vyzkoušejte na výsledné tabulce „*pivotování*“ výměna řádků a sloupců přehození jednotlivých sloupců 😊

➔ po ukončení analýz nezapomenout *Split file* vypnout nebo budou veškeré analýzy s tímto rozdělením

## Vyvážení souboru

### kvótní výběr

- známe rozdělení populace podle vybraných znaků (např. podle věku – muži : ženy = 1 : 1)
- snažíme se, aby výběr dodržoval tyto kvóty – někdy se to nepovede -> vyvažování

### vyvážení souboru (*Weight Cases*)

- vytvoříme váhovou proměnnou – např. W
- každé jednotce (řádku) souboru přiřadíme váhy

příklad: vyvážení odpovědí mužů a žen

Vyvážení podle konkrétní proměnné

- nerovnoměrné rozdělení -> rovnoměrné  $w_i = \frac{1}{k \cdot p_i}$
- k- počet kategorií proměnné

Příklad – vyvážení podle pohlaví:  $w_i = \frac{1}{2p_i}$

- v tomto souboru je 21% mužů a 71% žen
- vyvážení pomocí SPSS

- 1) -> **Transform** -> **Recode into Different Variables** (*překódujeme muže a ženy; muž=1, žena=2*) -> nová proměnná **W** (*jako Weight = váha*)
- 2) **Data** -> **Weight Cases** (*poslední položka*) -> **W**

Příklad: Vyzkoušejme vyvažování na sloupci *Jakou částku jste ochoten/ochotna za jeden kus ryby utratit?*  
Máme již nastaveno vyvážení, takže provedeme jen analýzu

*Analyze* -> *Descriptive statistics* -> *Frequencies*

Jakou částku jste ochoten/ochotna za jeden kus ryby utratit?					Jakou částku jste ochoten/ochotna za jeden kus ryby utratit?				
(původní data, bez weight)					(s využitím vyvažování)				
	Frequenc		Valid	Cumulative		Frequenc		Valid	Cumulative
	y	Percent	Percent	Percent		y	Percent	Percent	Percent
Valid do 100 Kč	78	39,0	39,0	39,0	Valid do 100 Kč	141	39,4	39,4	39,4
do 150 Kč	51	25,5	25,5	64,5	do 150 Kč	93	26,0	26,0	65,4
do 175 Kč	13	6,5	6,5	71,0	do 175 Kč	23	6,4	6,4	71,8
do 50 Kč	19	9,5	9,5	80,5	do 50 Kč	34	9,5	9,5	81,3
do 75 Kč	16	8,0	8,0	88,5	do 75 Kč	29	8,1	8,1	89,4
více než 200 Kč	23	11,5	11,5	100,0	více než 200 Kč	38	10,6	10,6	100,0

Total	200	100,0	100,0	Total	358	100,0	100,0
-------	-----	-------	-------	-------	-----	-------	-------

Všimněte si rozdílných hodnot ve sloupcích, a hlavně rozdílného **Total**. (je způsobeno vyvažováním, aby byly odpovědi mužů stejně zastoupeny jako odpovědi žen)

**Příklad:** Vyzkoušejme vyvažování na sloupci *PokudanoJakčasto?* Máme již nastaveno vyvážení, takže provedeme jen analýzu

*Analyze -> Descriptive statistics -> Frequencies*

Pokud ano, jak často? (původní data bez vyvažování)					Pokud ano, jak často? (s využitím vah)					
		Freque	Percen	Valid	Cumulativ		Frequenc	Percent	Valid	Cumulative
		ncy	t	Percent	e Percent		y	Percent	Percent	Percent
Valid	denně	1	,5	,5	,5	Valid	denně	1	,3	,3
	jednou týdně	68	34,0	36,8	37,3		jednou týdně	121	33,8	36,4
	jednou za měsíc	45	22,5	24,3	61,6		jednou za měsíc	80	22,3	24,1
	několikrát týdně	33	16,5	17,8	79,5		několikrát týdně	58	16,2	17,5
	příležitostně	38	19,0	20,5	100,0		příležitostně	72	20,1	21,7
	Total	185	92,5	100,0			Total	332	92,7	100,0
Missin	0	15	7,5			Missing	0	26	7,3	
Total		200	100,0			Total		358	100,0	

Vypneme vyvažování souboru: *Data -> Weight Cases -> Do not weight cases*

### Pivotování

- K úpravě tabulek do vhodného tvaru ke zpracování

**Příklad:** Vyhodnoťte (procentualně) sloupec *Jste* (zaměstnanost) v závislosti na pohlaví respondenta.

Řešení:

- 1) Vyhodnocují na základě znaku Pohlaví

*Data -> Split file -> Compare groups -> Vaše pohlaví*

- 2) Provedu analýzu zaměstnanosti

*Analyze -> Descriptive statistics -> Frequencies -> Jste*

Výsledek – velká nepřehledná tabulka, s nevhodnými sloupci a dalším nepotřebnými údaji

Pomocí menu Pivot a dalšími úpravami tabulky se pokuste dostat výslednou tabulku v tomto tvaru:

Jste:		Vaše pohlaví:	
		muž	žena
Valid	důchodce	7,1	

na mateřské dovolené	2,4	5,1
nezaměstnaný/á	14,3	3,8
OSVČ		5,1
student	40,5	58,9
zaměstnaný/á	35,7	27,2
Total	100,0	100,0

Příkaz *Pivot* je v horním menu okna *Output* (objeví se po vybrání, zvýraznění tabulky).

Další úpravy tabulky intuitivně po dvojkliku myši na tabulku odstranění zbytečných řádků a

sloupců.

Vizuální úpravy přes horní menu *Format* (např. odstranění nebo přidání dělicích řádků – *Table Properties* -> Horizontal category border)

Další úkoly a příklady k procvičení:

- Vyhodnoťte poslední otázku: *Podle vašeho názoru je na českém trhu poptávka po rybách a rybích výrobcích....*
  - Obecně podle názoru respondentů.
  - Podle pohlaví mají stejný názor muži i ženy?
  - Proveďte vyvážení souboru podle pohlaví a proveďte znovu analýzu
- Vyhodnoťte otázku: *Dáváte přednost spíše tuzemským nebo zahraničním produktům?*  
Co respondenti upřednostňují?  
Jak toto souvisí s cenou za 1 kus ryby? (vyhodnoťte cenu 1 kusu ryby, otázka: *Jakou částku jste ochotni/ochotna za jeden kus ryby utratit?*, v závislosti zda se jedná o tuzemskou nebo zahraniční rybu)
- Jsou respondenti s vyšším příjmem ochotni utratit za jeden kus ryby více peněz? *Zamyslete se nad vhodným vyjádřením odpovědi.*

U všech úkolů dbejte na správné vyjádření, přehlednou prezentaci a vždy si slovně zdůvodněte odpověď na co jsme se ptali jakou jsme dostali odpověď.....