

Ukázky jak prezentovat tabulky a interpretovat vztahy proměnných.

Jiří Šafr, FHS UK, KŘS & HiSo
(poslední aktualizace 21.8.12)

Základní pravidla pro prezentaci tabulek

Tabulka má vždy **název**, v němž musí být patrné **co jsou proměnné**, k jaké **populaci/výběru** se výsledky vztahují a jaké **míry (statistiky)** jsou použity.

U tabulky, nejčastěji pod ní, uvádíme **zdroj** (tj. datový soubor na němž je analýza provedena) a také **počet platných případů**, tj. absolutní četnost za celý soubor (případně podsoubory), který se vztahuje k aktuální tabulce či grafu. Vhodné je také přímo uvést, jaké je procento chybějících hodnot. To uvádějte raději u všech tabulek, i kdyby se vám zázrakem stalo, že všichni odpověděli na všechny otázky.

Aktivní čtenář si může najít informaci o celkové velikosti vašeho výběru (populaci), případně velikosti podsouboru, který analyzujete v daném textu. Tu uvádíte někde v úvodní části textu (na to nikdy nezapomeňte), tam kde popisujete výzkum resp. datový soubor. A může si z aktuálního platného počtu spočítat, kolik procent tvoří u té které tabulky chybějící hodnoty. To dělají poučení čtenáři proto, aby si ověřili, zda náhodou nepřekračuje cca 5 %, což většinou považujeme za hranici, kdy už by chybějící hodnoty (odmítnutí odpovědět, „nevím“ atd.) mohlo být nenáhodné, tj. vázat se na určité skupiny respondentů. Pro čtenáře je ale příjemnější, když mu podíl chybějících hodnot pod tabulkou přímo sami uvedete.

Zejména u malých výběrů/populací uvádějte raději všechny tyto údaje a především na ně myslte při interpretaci výsledků. Připomínám, že u kontingenční tabulky pro dvě proměnné je počet platných případů průnikem platných případů u obou proměnných (může se tedy lišit oproti třídění prvního stupně).

U některých proměnných – otázek je třeba zmínit, pokud byly vynechány některé speciální kategorie odpovědí (např. „nevím“ nebo „odmítl odpovědět“). Vždy musí být patrné, z čeho je počítán procentní základ (zda z „čistých“ odpovědí, kupříkladu u kategorií nějaké postojové škály, nebo zda je to včetně oněch kategorií „mimo rámec“ a/nebo chybějících hodnot).

Pokud ukazujete **procenta**, **tak vždy uveďte o jaký typ jde**, zda se jedná o sloupcová (column) nebo řádková (row) procenta (případně celková (total)). Orientaci také naznačí uvedení součtu 100 % (a znaku „%“ v příslušném řádku či sloupci s celkem). Připomínám, že orientace procent v kontingenční tabulce je klíčová pro smysl a cíl analýzy.

Nezapomeňte, že samotná procenta říkají málo nebo nic. Nezamlčujte proto absolutní četnosti, zejména u malých souborů a při třídění více znaků (to aby 2 respondenti z celkového počtu 9 nebyli interpretováni jako 22,22 procent). Absolutní počty pro marginální četnosti, tj. třídění 1. stupně pro závislou proměnnou můžeme uvést ve sloupci/ řádku Celkem.

Čísla v tabulce zaokrouhľujte (třemi desetinnými místy na přesnosti nic nezískáte, jen znepřehledníte text), u procent nejčastěji na celá čísla. Více jak jedno desetinné místo, pokud nejde o čísla menší než 1, v podstatě nepoužíváme (využijte výhody editace tabulky v Excelu → formát buněk).

V textu uvádějte, jak byly proměnné operacionalizovány. Zejména pokud se jedná o postojové otázky, vždy zveřejněte jejich přesné znění (včetně kategorií odpovědí), například v poznámce pod čarou či v příloze (u kratších otázek je také někdy vkládáme do nadpisu tabulky).

K tomu, jak tabulku zkonstruovat viz presentace kurzu AKD, např. Kontingenční tabulky a analýza kategoriálních dat, AKD I. na <http://metodykv.wz.cz> .

Následuje ukázka úpravy kontingenční tabulky pro třídění druhého stupně a grafu třídění 3. stupně.

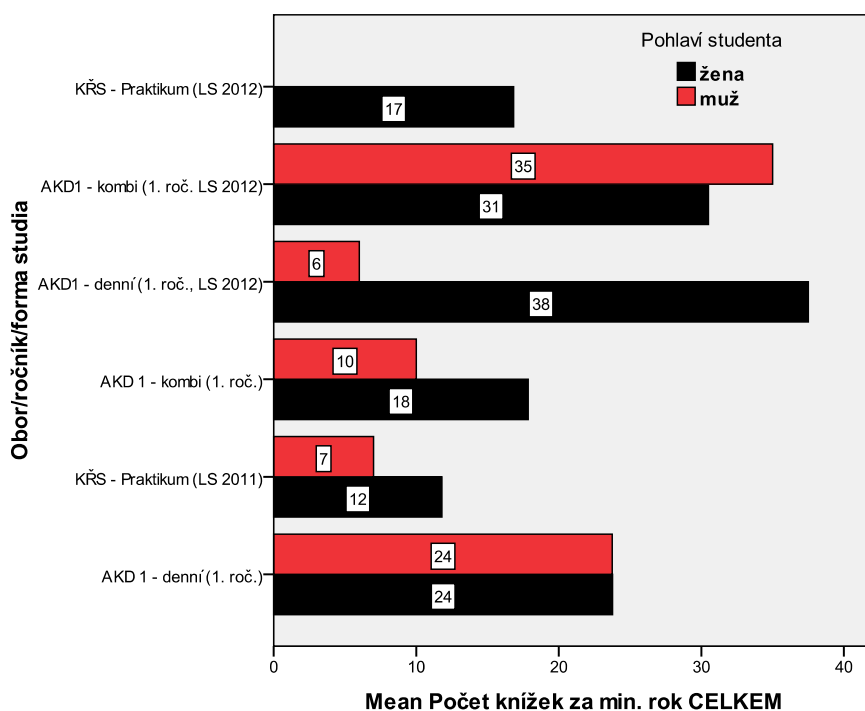
Tabulka 1. Počet knížek přečtený za rok celkem podle oboru studia, průměry a směrodatné odchylky, studenti FHS předmětů J. Šafra v letech 2011 až 2012.

studium	Obor/ročník/forma studia	Průměr	N	Std.D.
	AKD 1 - denní (1. roč.)	24	17	(8,19)
	KŘS - Praktikum (LS 2011)	10	7	4,9
	AKD 1 - kombi (1. roč.)	17	14	9,1
	SHV - (bc.)	15	3	3,0
	AKD1 - kombi (2. roč. LS 2011)	20	1	
	AKD1 - denní (1. roč., LS 2012)	34	9	19,9
	AKD1 - kombi (1. roč. LS 2012)	31	12	13,3
	KŘS - Praktikum (LS 2012)	17	6	5,1
	Celkem	23	69	12,9

Zdroj: FHS TV a knihy 2011-2012

N (validní) = 69 (chybějící hodnoty 4,2 %).

Graf 1. Počet knížek přečtený za rok celkem podle oboru studia a pohlaví, průměry a směrodatné odchylky, studenti FHS předmětů J. Šafra v letech 2011 až 2012.



Zdroj: FHS TV a knihy 2011-2012

N (validní) = 65 (chybějící hodnoty 9,7 %).

Prezentace a interpretace kontingenční tabulky (výzkumná otázka a hypotéza)

V zásadě k tabulkám můžeme přistoupit dvojím způsobem. Buď postupujeme jen víceméně deskriptivně – explorativně, kdy nemáme žádnou předem danou otázku/ hypotézu a jen systematicky popisujeme rozdíly v závislé (vysvětlované) proměnné z hlediska třídění jinými znaky. A nebo máme předem definované výzkumné otázky (obecné i ty konkrétní, nejčastěji formulované již ve fázi přípravy výzkumu) a z nich odvozené hypotézy (tj. vlastně odpovědi na tyto otázky na základě teorie) a postupujeme tak, že tabulku či graf konstruuje podle těchto otázek/hypotéz, tak abychom přímo ověřili předpokládané vztahy. Viz následující příklad ze cvičného výzkumu studentů na FHS s názvem „TV a knihy“.

Výzkumná otázka (RQ): Souvisí počet přečtených knih s dobou sledování TV?

Hypotéza „sociologická“: Počet přečtených knih roste s dobou strávenou sledováním TV. → předpokládáme *pozitivní souvislost*

Nulová „statistická“ hypotéza (H₀): Počet přečtených knih se neliší v závislosti na době strávené u TV. → Statistickým postupem je očekávat a ověřovat *žádný vztah*.

K ní lze formulovat tzv. alternativní hypotézu (H_A) – kdy H₀ neplatí, tj. vztah existuje.

Dodejme ještě, že vztahy nemusí být takto jednoduché, vyjádřené jako lineární souvislosti, navíc se mohou projevovat třeba jen v určitých podskupinách (tj. jen za určitých podmínek). Takže se zamítnutím jednoduše formulované hypotézy bychom se neměli spokojit. Navíc mnoho vztahů, kde chceme usuzovat na kauzální působení, je zprostředkovaných a tudíž bivariátní vztah dvou znaků může být způsoben působením třetí proměnné. Práci výzkumníka je právě v dalším kroku odhalovat i tyto komplikovanější vztahy (o podobě vztahu hodně napoví kontingenční tabulka a nebo pro spojité znaky bodový – scatterplot graf a pak třídění třetího stupně).

Tabulka 2. Počet knížek přečtený za rok (tercily) podle sledování TV v hodinách za den (tercilu), sloupcová procenta, studenti FHS předmětů J. Šafra v letech 2011 až 2012.

		Sledování TV (tercily)			Celkem
		I. tercil (-)	II. tercil (0)	III. tercil (+)	
Knihy přečtené celkem za rok (tercily)	I. tercil (-)	15,8	23,1	38,9	25,4%
	II. tercil (0)	42,1	46,2	44,4	44,4%
	III. tercil (+)	42,1	30,8	16,7	30,2%
Celkem		100 %	100 %	100 %	100 %
N		(19)	(26)	(18)	(63)

Zdroj: FHS TV a knihy 2011-2012

N (validní) = 63 (chybějící hodnoty 12,5 %).

Tabulka 2 ukazuje relativní podíl přečtených knih za rok (soubor jsme rozdělili na tři stejně velké skupiny čtenářů podle tercilů) ve skupinách definovaných podle doby sledování televize v běžném dni (rovněž kategorizováno na tercily). V tabulce ověřujeme **hypotézu, podle níž počet přečtených knih roste s dobou strávenou s TV**. Zatímco v kategorii podprůměrného množství přečtených knih (I. tercil) je pouze 16 %, těch kdo televizi sledují málo (I. tercil),

tak studentů, kteří se na televizi dívají nadprůměrně často (III. tercil) je v této kategorii zhruba 2,5 krát více (39 %). Obdobně v kategorii nejvíce přečtených knih (III. tercil) je 42 % těch, kdo se na televizi v podstatě nedívají a zároveň jen 17% těch, kteří se na ní dívají velmi často (III. tercil). Vidíme tak, že mezi čtením knih a sledováním televize existuje negativní souvislost (vyjádřeno pomocí kontingenčního koeficientu souvislost je tato souvislost středně silná, $CC = 0,25$). Naši hypotézu o pozitivní souvislosti mezi počtem přečtených knih a sledováním TV nám tedy nezbyvá než **zamítnout, vztah je totiž přesně obrácený: čím více studenti čtou, tím méně se dívají na televizi.** **To byl ale pouze popis vztahů v tabulce, následovat musí věcné vysvětlení, většinou ho odvozujeme z určité teorie. Zde alespoň krátký náznak:**

Jedním z možných vysvětlení by mohl být vzájemný konkurenční vztah konzumace médií, den má jen 24 hodin a člověk se musí rozhodnout, jaké médium zvolí... Můžeme také uvažovat o tom, že tato média reprezentují poněkud odlišný kulturní svět (knihy má blízko k intelektuální – vysoké kultuře, zatímco televize spíše k masové – zábavné kultuře) a tak v pozadí výběru jednoho nebo druhého média může působit odlišný kulturní vkus. Nicméně k detailnějšímu posouzení těchto vysvětlení bude třeba přistoupit k dalším analýzám, které zhodnotí souvislosti čtení s jinými aktivitami volného času a zejména ověří, zda tento vztah je platný v rámci různých sociálních skupin (např. dle pohlaví, věku apod.).

Při interpretaci dat nebud'te pouze popisní, mechanicky nekopírujte, co čtenář vidí v tabulce. Vyberte pouze podstatné vztahy z hlediska vaší hypotézy – ty které ji potvrzují a zejména ty, které ji vyvrací (!). Interpretujte jen skutečně věcně významné rozdíly – zamyslete se, jak velké jsou rozdíly v závislé–vysvětlované proměnné mezi kategoriemi nezávislé proměnné, pokud vezmete v úvahu metriku (počty či vzdálenosti kategorií) závislé proměnné. Nabízejte vysvětlení. Pište čtivý text, který nebude čtenáře nudit.

Pro zajímavost, předchozí tabulku bylo nejprve třeba upravit, původně vypadala ve výstupu SPSS takto:

Case Processing Summary						
Cases						
Valid						
Missing						
Total						
	N	Percent	N	Percent	N	Percent
knihy_celk3t * TV3t Sledování TV (tercily)	63	87,5%	9	12,5%	72	100,0%

knihy_celk3t * TV3t Sledování TV (tercily) Crosstabulation						
TV3t Sledování TV (tercily)						
1. I. tercil						
2. II. tercil						
3. III. tercil						
Total						
knihy_celk3t		Count	1. I. tercil	2. II. tercil	3. III. tercil	Total
1. I. tercil	Count		3	6	7	16
	% within TV3t Sledování TV (tercily)		15,8%	23,1%	38,9%	25,4%
2. II. tercil	Count		8	12	8	28
	% within TV3t Sledování TV (tercily)		42,1%	46,2%	44,4%	44,4%
3. III. tercil	Count		8	8	3	19
	% within TV3t Sledování TV (tercily)		42,1%	30,8%	16,7%	30,2%
Total	Count		19	26	18	63
	% within TV3t Sledování TV (tercily)		100,0%	100,0%	100,0%	100,0%