Metodologie pro Informační studia a knihovnictví 2

Modul 3: GIGO. Popis dat. Kontrola a čištění dat.

Co se dozvíte v tomto modulu?

- Proč je potřeba dbát na kvalitu dat na vstupu
- Jak popsat výběrový soubor a na jaké hodnoty proměnných dávat pozor při kontrole?
- Jak vybrat jen určité případy (nový dataset)
- Jak postupovat v Excelu a v Google Spreadsheets?

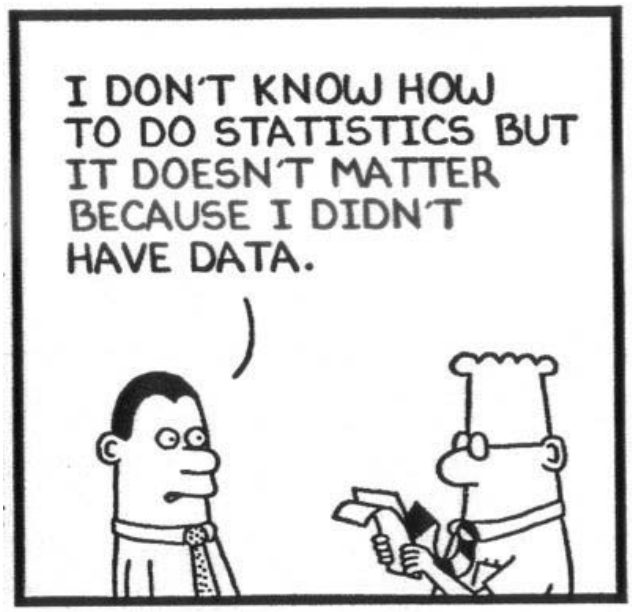
V tomto modulu si připravíme dataset k samotné analýze. To, zda budete mít na konci analýzy smysluplné výsledky, do značné míry záleží právě na tom, jakou míru pozornosti budete věnovat počáteční kontrole dat.

Obsah

1 Pravidlo GIGO: „Garbage in – garbage out!“

Úvodní příprava dat na samotnou analýzu není sice nejzajímavější a nejzábavnější částí analytické práce, ale pro kvalitní výsledek je naprosto nezbytná. V této souvislosti se používá pořekadlo „***garbage in – garage out“*** – pokud jsou na vstupu nekvalitní data, nekvalitní bude i výstup. Proto je v první řadě potřeba věnovat se právě kontrole a čištění dat.

Charles Wheelan (2013) říká, že za každou důležitou výzkumnou studií jsou data, která umožňují analýzu a naopak špatné výzkumy bývají i založené na špatných datech.



Wheelan (2013) identifikuje několik obecných příkladů GIGO:

Zkreslení výsledků kvůli výběru

Možná jste v roce 2008 byli mezi těmi, kdo si byli jistí postupem Strany zelených do krajských zastupitelstev. Stranu zelených totiž volilo hodně lidí z vašeho okolí, tehdejší předseda Bursík předpokládal několikanásobný nárůst počtu zastupitelů, výzkumy veřejného mínění slibovaly Straně zelených zisk 7-9 % voličských hlasů. Po volbách však strana nezískala zastoupení ani v jednom kraji. Podobná situace se opakovala v roce 2010 při volbách do Poslanecké sněmovny. Pokud by řada vysokoškoláků mohla odhadovat výsledky voleb podle statusů a profilových fotografií na Facebooku, Zelení by byli jasnými favority. Přesto ani v těchto volbách nezískali potřebný počet hlasů. Jak je možné, že se tolik lišily odhady (výzkumy) a realita?

Jeden z nejznámějších podobných případů, kde byly špatné výsledky výzkumů ovlivněny špatným výběrem respondentů, a tedy od počátku špatnými daty, byl příklad předvolebního průzkumu, který v roce 1936 realizoval časopis *Literary Digest*. Časopis oslovil před volbami 10 milionů amerických voličů s otázkou, zda budou volit republikána Alfa Landona či demokrata Franklina Roosevelta. 10 milionů je obrovský vzorek, a tak se výsledkům šetření, které přiřkly 57 procent

Landonovi, přikládala velká váha. Velký problém byl však ve výběru respondentů. Literary Digest totiž oslovil své předplatitele a také majitele telefonních přístrojů a automobilů (jejich adresy totiž byly veřejně dohledatelné). Pro autory šetření byly potom velkým překvapením výsledky voleb, kde se 60 % zvítězil Franklin Roosevelt. Ukázalo se, že vzorek, byť velký, nebyl v žádném případě reprezentativní vzhledem k celé americké populaci – předplatitelé časopisu Literary Digest, stejně jako majitelé telefonů a vozů, patřili mezi bohatší část společnosti a nebyli rozhodně obrázkem „průměrného Američana". Wheelan dodává: *„Čím větší jsou dobře sestavené vzorky, tím lépe, protože se zmenšuje riziko chyby. Čím větší jsou špatně sestavené vzorky, hromada smetí pouze narůstá a čím dál více zapáchá"* (s. 119).

Speciálním případem zkreslení výsledků kvůli nesprávně vybranému vzorku jsou **ankety** a tzv. **samovýběry**, například dobrovolnické studie. Dobrovolníci, kteří jsou ochotni se přihlásit např. do výzkumu sexuálního chování, nemusí reprezentovat sexuální chování celé populace. Riziko špatného výběru při samosběru se může ještě zvýšit, pokud nabízíme za zapojení do výzkumu odměnu.

Zkreslení výsledků pro publikování

Wheelan (ibid) upozorňuje, že pozitivní výsledky studií mají větší šanci na opublikování, protože nejsou tak zajímavé. *„Pokud si vezmete 100 statistických šetření, je pravděpodobné, že jedno z nich bude mít naprosto nesmyslné výsledky – například statistickou asociaci mezi hraním videoher a výskytem rakoviny střev. A tady je ten problém: zatímco 99 studií, které dokázaly nulovou závislost mezi hraním her a rakovinou střev, nebude nikdy publikováno, protože výsledky nejsou dost zajímavé, jediná studie s pozitivními výsledky půjde do tisku a bude se jí věnovat další pozornost"* (s. 121).

Tento efekt byl popsán například u publikování výsledků studií účinnosti léků na depresi – u studií, které dokazovaly účinnost léku, byla publikována velká část, zatímco studie s nepozitivními výsledky vydávány nebyly.

Zkreslení výsledků kvůli paměti

Velká část šetření je založena na zjišťování reálních zážitků a chování respondentů. Ukazuje se ale, že paměť je velmi složitý mechanismus. Wheelan (ibid) zmiňuje harvardskou studii, ve které se vědci dotazovali žen s rakovinou prsu na jejich stravovací návyky. Ukázalo se, že ženy, které onemocněly rakovinou prsu, vykazovaly ve studii větší sklon k předchozí konzumaci tučných jídel oproti zdravým ženám. Ve skutečnosti se však nejednalo o studii závislosti konzumace tuku a výskytu rakoviny prsu, ale o výzkum toho, jaký vliv má onemocnění rakovinou prsu na paměť. Všechny ženy podstoupily dotazování na stravovací návyky léta předtím, než jim byla rakovina diagnostikována. Srovnání výsledků prvního dotazování založeného na měření reálného aktuálního chování a druhého šetření zjišťujícího stejné chování v minulosti, ukázalo, že fakt onemocnění má vliv na to, jak si ženy „převyprávěly" svou minulost vlivem hledání příčin onemocnění.

Tento druh zkreslení je tedy velkým rizikem studií, které zjišťují minulé chování.

Survivorship bias – „klam přeživších"

Tzv. klam přeživších je chybou, která je založena na vyšší viditelnosti těch, kteří „přežili“ určitý proces. Například pokud bychom zjišťovali spokojenost se studiem na KISKu na absolventech našeho oboru, dobrali bychom se pravděpodobně jiných čísel, než kdybychom zjišťovali spokojenost se studiem mezi všemi studenty, tedy i těmi, kteří z nějakého důvodu studium nedokončili. Klam přeživších tedy může často vést k optimističtějším závěrům.

Klam zdravého uživatele

Tzv. klam zdravého uživatele byl popsán v epidemiologii.

- Do výzkumných studií o zdraví se například hlásí obecně zdravější lidé – prostě proto, že se více zajímají o zdraví.
- Lidé, kteří berou vitamíny, jsou zdravější. Prostě proto, že je to *ten druh lidí*, kteří berou pravidelně vitamíny (tito lidé také pravděpodobnější pravidelně sportují, sledují své zdraví a věnují se prevenci).

Do vztahů, které mezi proměnnými sledujeme, zkrátka vstupují ještě další proměnné, a ty je potřeba hlídat. Jinak se nemůžeme vyvarovat omylů, které se dají shrnout pod heslo ***„garbage in – garbage out“***.

Vliv nepozorovaných proměnných

Disman (2002) ukazuje, že do analýzy mohou vstupovat další proměnné s rizikem ovlivnění výsledků. Tato rizika je potřeba hlídat:

1. **Nepravá korelace**. Ačkoliv se může zdát, že proměnná A ovlivňuje proměnnou B, může existovat ještě třetí nepozorovaná či neanalyzovaná proměnná C, která ovlivňuje A i B.
   (C → A ^ C → B)
2. **Vývojová sekvence.** V tomto případě se nám opět zdá, že proměnná A ovlivňuje proměnnou B a může tomu skutečně tak být. Co však nepozorujeme, je proměnná 0, která ovlivňuje proměnnou A.
   (0 → A → B)
3. **Chybějící střední člen.** Tato situace nastává, pokud jsme do analýzy nezařadili proměnnou, která je ovlivňována proměnnou A a dále ovlivňuje proměnnou B.
   (A → X → B)
4. **Dvojí příčina.** Závislá proměnná B může mít více příčin, ale ne všechny jsou zahrnuty do výzkumu.
   (A+X+Y→ B)

Zdroje chybných dat při zápisu

Chyby v datech mohou vznikat i při zápisu do datového souboru. Obvykle se jedná o posuny desetinných čárek, záměnu znaků či další chyby při přepisování (například záměna „O“ a „0“).

Pokud vás téma chyb v analýze zaujalo, přečtěte si třeba článek ***Why Most Published Research Findings Are False?***

2 Popis a kontrola dat

Prvním úkolem výzkumníka je popis výběrového souboru. Charakteristikou vzorku by měla začít každá analýza i analytická kapitola v bakalářské či diplomové práci. Zajímá nás například:

- Kolik je ve výběrovém souboru jednotek?
- Kolik je v souboru mužů a žen?
- Kolik je v souboru lidí se ZŠ/SŠ/VŠ vzděláním?
- Jak je v souboru distribuován věk?

Toto rozložení může být vyjádřeno v **absolutních**, **relativních**, či **kumulativních relativních četnostech**.

- **Absolutní četnost** udává absolutní číslo – hodnotu četnosti varianty proměnné v souboru.
  Například: V souboru je 1456 mužů a 1201 žen.
- **Relativní četnost** udává **podíl** četnosti varianty proměnné v souboru.
  Například: V souboru je 24 % osob se základním vzděláním.
- **Kumulativní relativní četnost** udává kumulativní podíly variant proměnné v souboru (nejsou použitelné pro nominální proměnné).
  Například: V souboru je 36 % respondentů, kteří mají alespoň maturitu (tedy nejen úspěšní středoškoláci s maturitou, ale také vysokoškoláci se všemi variantami diplomů).

Popis a kontrola kategorizovaných dat

Tabulky četností

Pro zobrazení základních hodnot popisu rozložení hodnot kategorizovaných proměnných (tedy proměnných nominálních a ordinálních s menším počtem variant odpovědí) se používá tzv. **tabulka četností**. Ta obsahuje jak absolutní, tak relativní četnosti hodnot proměnných. Takto vypadá správná a kompletní tabulka četností:

Jaké je Vaše vzdělání?				
		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 nástavba, VOŠ	40	6,5 %	6,6 %
	Vysokoškolské	153	25,0 %	25,2 %
	Celkem validní hodnoty	608	99,2 %	100,0 %
Chybějící hodnoty (neví, neodpověděl/a)	Chybějící hodnoty	5	0,8 %	
Celkem		**613**	**100,0 %**	

V praxi se často používá jen zkrácená verze tabulky obsahující pouze validní četnosti:

Jaké je Vaše vzdělání?	**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 %**

Před počítáním četností je ale potřeba zkontrolovat data. Kontrolujeme, zda se nachází v platném intervalu (například proměnná pohlaví nabývá v našem souboru pouze hodnot 1 a 2, všechny ostatní varianty by měly být omyly).

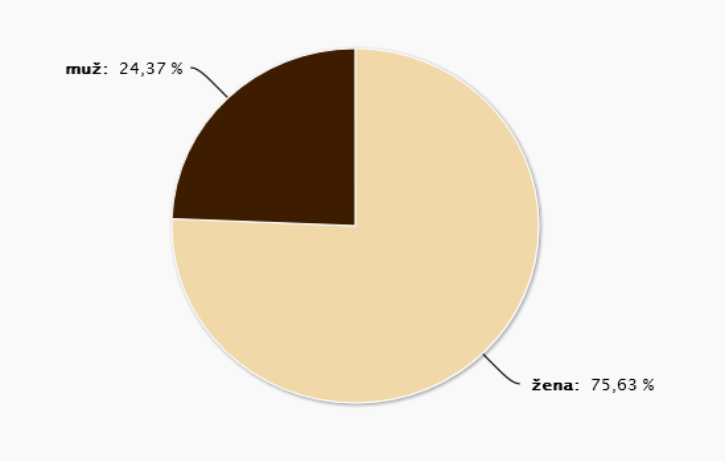
Grafy četností

Pro znázornění rozložení četností se využívají i grafy znázorňující četnosti hodnot proměnných. Nejznámějšími variantami jsou koláčový a sloupcový graf.

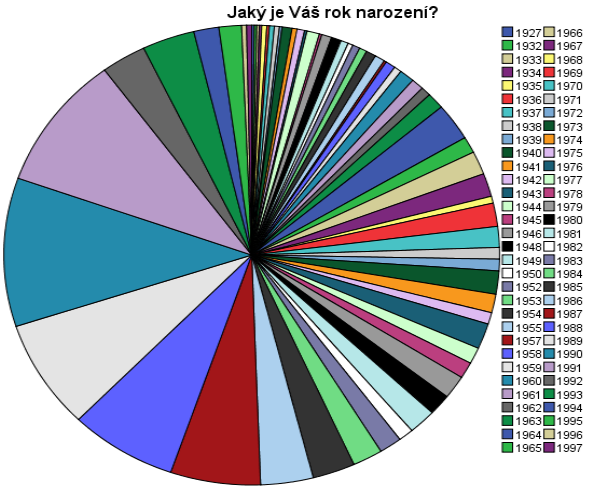
Koláčový graf je vhodný:

- pro třídění prvního stupně (jedna datová řada),
- pro porovnání četností u nominálních proměnných, které nemají příliš mnoho hodnot (méně než 7),
- pokud hodnoty, které chcete vykreslit, nejsou nulové,
- pokud hodnoty představují část celku.

Příklad proměnné, kde je vhodné využít koláčový graf:



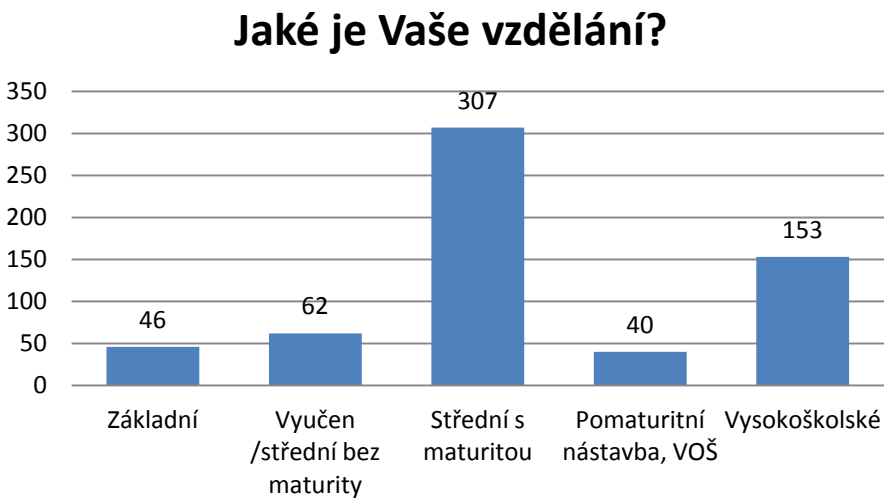
Příklad proměnné, kde NENÍ vhodné využít koláčový graf:



Sloupcový graf je vhodný pro:

- porovnání položek,
- ordinální proměnné a kardinální proměnné s menším počtem kategorií,
- znázornění změn za časové období (třídění druhého stupně).

Příklad sloupcového grafu:



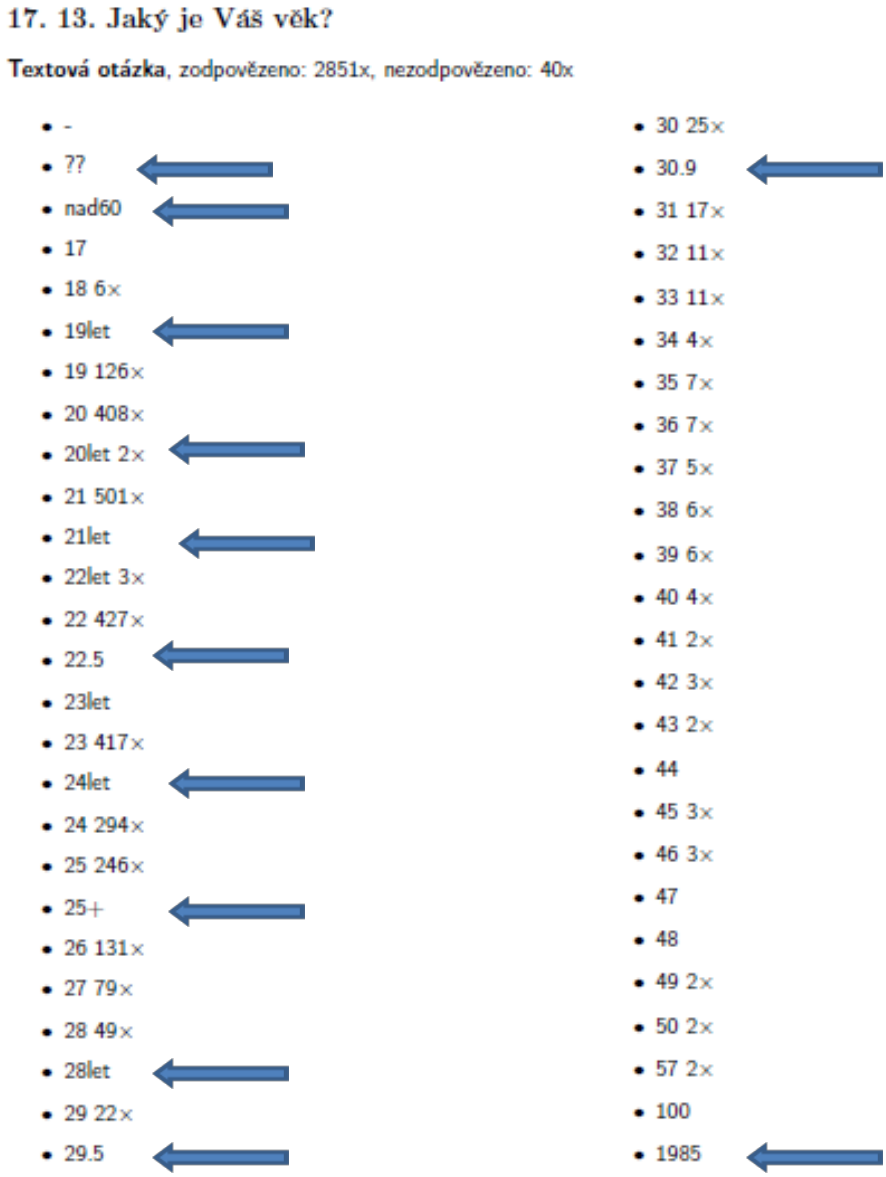
Grafy se v Excelu vkládají pomocí funkce „**Grafy**“ na listu „**Vložení**“.

Popis a kontrola nekategorizovaných dat

Pro první kontrolu nekategorizovaných dat nám bude stačit podívat se na **minimální** a **maximální** hodnoty dat. Například u proměnné „rok narození“ by naši respondenti neměli být narozeni později než v roce 1995 (máme rok 2013 a respondenti měli být starší 18 let). Dřívější datum narození není jasné, ale nejstarší občance ČR je momentálně 109 let, držme se tedy limitu 1904 jako

nejmenšího možného roku narození. U hodnot 1904–1995 tedy máme důvod domnívat se, že jsou v pořádku. Často se však mohou vyskytnout chyby vzniklé při zápisu (např. rok 11982 či naopak vynechání číslice – rok 198). Tato data je potřeba opravit.

Někdy se může stát, že respondenti nevědí, jak odpovědět. Potom můžete na jednoduchou otázku („*Kolik je vám let*“) získat velmi různé formáty odpovědí:



17. 13. Jaký je Váš věk?

Textová otázka, zodpovězeno: 2851x, nezodpovězeno: 40x

- -
- ??
- nad60
- 17
- 18 6×
- 19let
- 19 126×
- 20 408×
- 20let 2×
- 21 501×
- 21let
- 22let 3×
- 22 427×
- 22.5
- 23let
- 23 417×
- 24let
- 24 294×
- 25 246×
- 25+
- 26 131×
- 27 79×
- 28 49×
- 28let
- 29 22×
- 29.5
- 30 25×
- 30.9
- 31 17×
- 32 11×
- 33 11×
- 34 4×
- 35 7×
- 36 7×
- 37 5×
- 38 6×
- 39 6×
- 40 4×
- 41 2×
- 42 3×
- 43 2×
- 44
- 45 3×
- 46 3×
- 47
- 48
- 49 2×
- 50 2×
- 57 2×
- 100
- 1985

Co s chybnými daty?

Narazíme-li na chybnou hodnotu, máme v zásadě několik možností:

- **Zjistit chybu a nahradit chybný zápis správnou hodnotou.** Například pokud chyba vznikla při přepisu papírového dotazníku do elektronické tabulky, je možné dotazník dohledat a chybu opravit. Stejně postupujeme i v případě, že respondenti nevyplnili pole tak, jak jsme chtěli (např. hodnotu „23let“ si překódujeme jen na „23“).

- Pokud není možné zjistit chybu, můžeme **prohlásit odpověď za chybějící** a nakládat s ní, jako by nebyla otázka vůbec zodpovězena. Variantně můžeme respondenta úplně vyřadit ze souboru.

Co s chybějícími daty?

Kromě chybných dat je potřeba zkoumat i **chybějící hodnoty**. Vyplatí se před samotnou analýzou zkontrolovat, kolikrát se vyskytly v odpovědích varianty „nevím / nemohu odpovědět".

Jsou odpovědi rozděleny náhodně? Nemá výskyt nevím souvislost s nějakou jinou proměnnou?

Pro kontrolu můžeme rozdělit soubor na skupiny záznamů s chybějícími hodnotami a bez nich, porovnat charakteristiky obou souborů, nebo nechat korelovat vyplnění/nevyplnění s jinou proměnnou (o korelacích bude řeč v dalších modulech).

3 Práce s datovým souborem

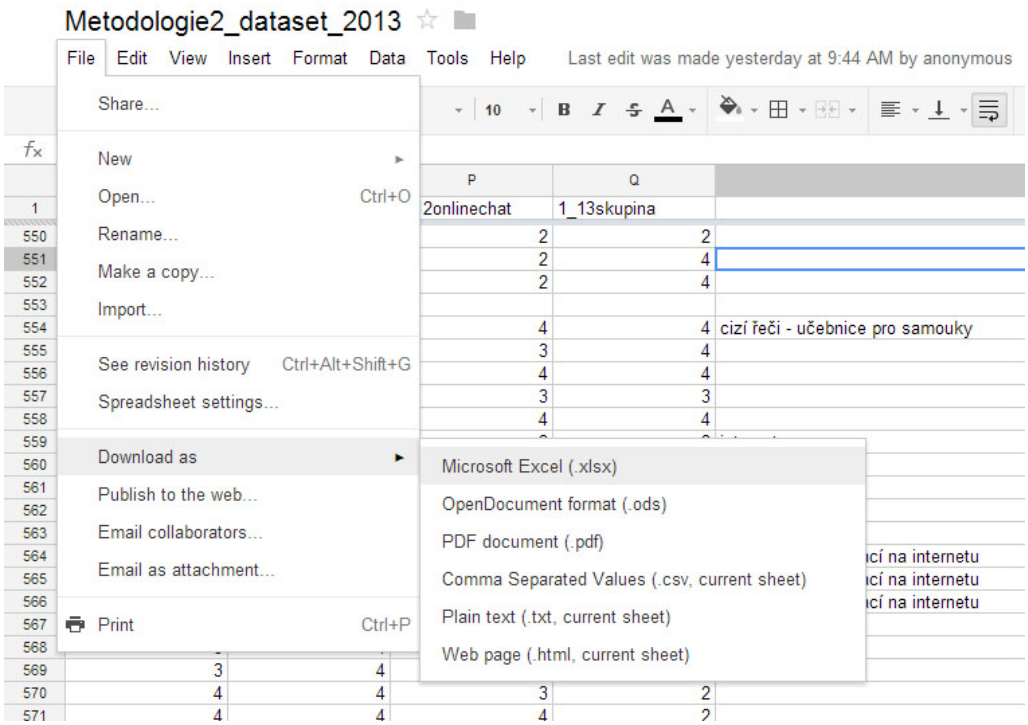
Dřív než začneme pracovat s datovým souborem, je potřeba zmínit několik zásad.

1. Ať už pracujeme v jakémkoliv programu, je vždy důležité pravidelně **zálohovat data**. Ponechte si zálohovaný původní datový soubor, ať se k němu v případě nejistot můžete vrátit. Zálohujte si také průběžnou práci – při analýze často vytváříte nové proměnné, o které byste mohli bez zálohování přijít. Při nepozornosti si také můžete přemazat některá data, proto je vhodné mít zazálohovaných několik posledních verzí souborů s daty.
2. Pokud pracujete ve **sdíleném souboru**, dbejte na to, aby byly kroky jednotlivých výzkumníků odlišitelné a zpětně dohledatelné. Pokud to prostředí neumožňuje, zvažte jinou variantu způsobu práce s daty.
3. Než začnete analyzovat, data **zkontrolujte** a **pečlivě popište**.

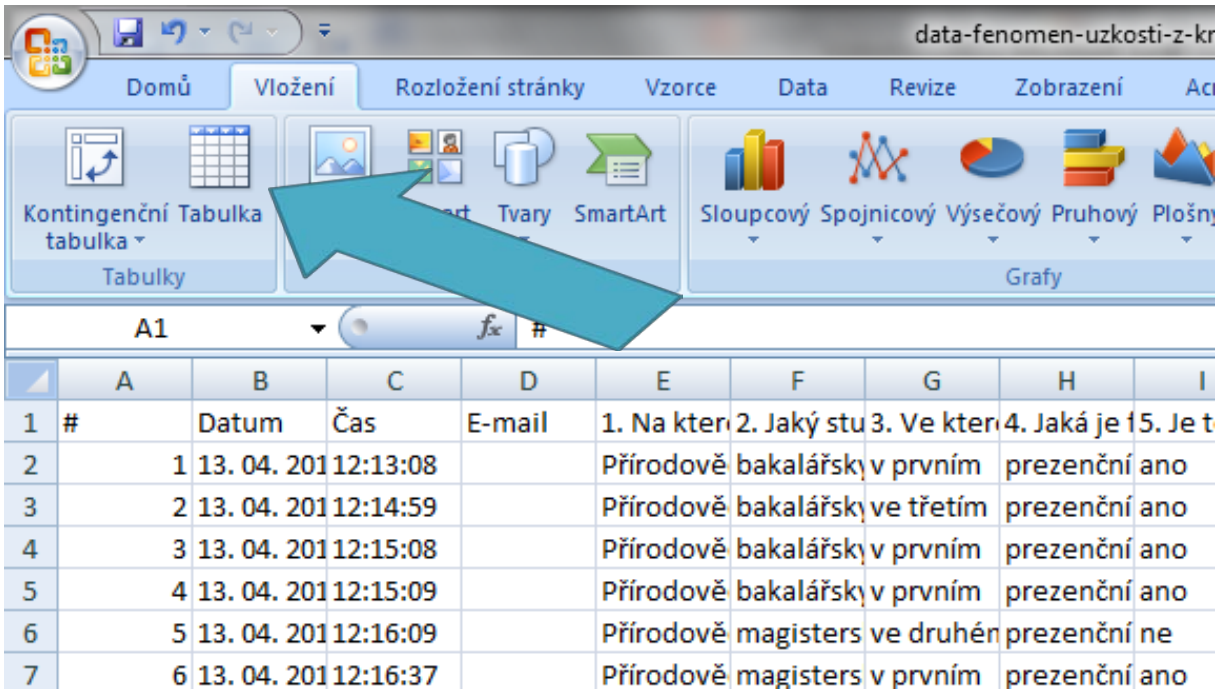
Stažení tabulky

V tomto semestru budeme pracovat se souborem, který jsme si společně vytvořili v Google dokumentech. Většinu operací, které budeme používat, lze provádět přímo v Google Spreadsheets. Pro práci v Excelu je možné si stáhnout tabulku z Google dokumentů pomocí funkce „**Download as**".

Stažení souboru ve formátu .xls:

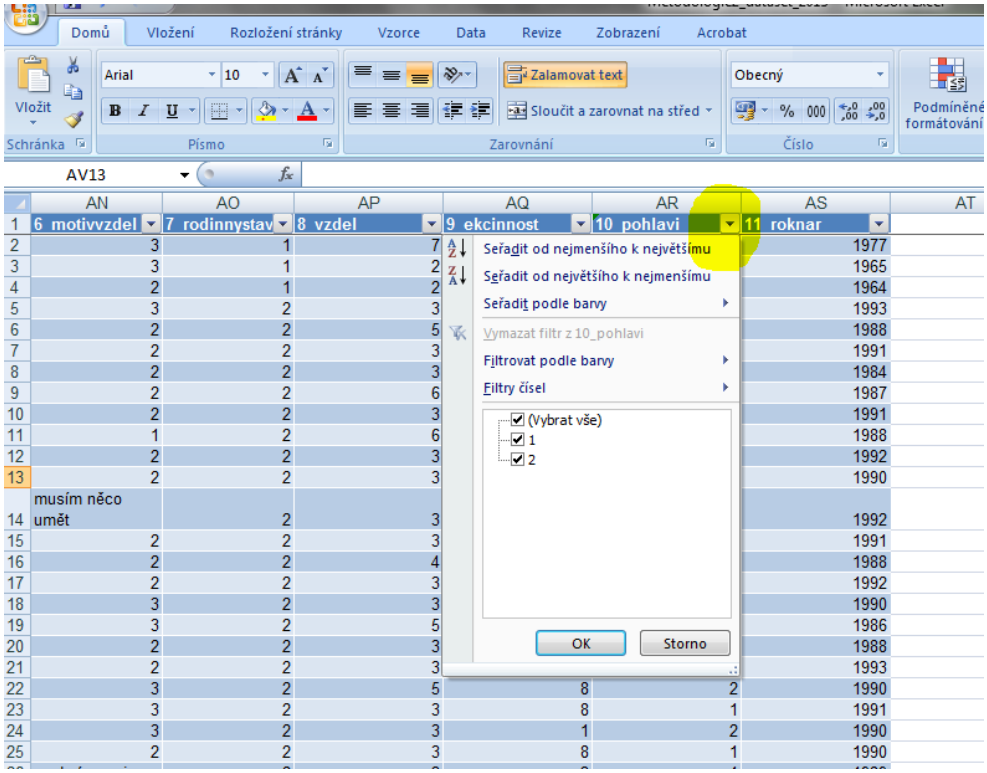


V Excelu je poté pro práci s daty vhodné data převést na inteligentní tabulku pomocí funkce „**Tabulka**“ v listu “**Vložení**”:



Excel rozpozná záhlaví a převede data na přehlednější tabulku.

Někdy nechceme pracovat s celým datovým souborem, ale zajímají nás například pouze ženy. V Excelu si můžeme jednoduše vyfiltrovat rozkliknutím položky v záhlaví:



	AN	AO	AP	AQ	AR	AS	AT
1	6 motivvzdel	7 rodinnystav	8 vzdel	9 ekcinnost	10 pohlavi	11 roknar	
2	3	1	7			1977	
3	3	1	2			1965	
4	2	1	2			1964	
5	3	2	3			1993	
6	2	2	5			1988	
7	2	2	3			1991	
8	2	2	3			1984	
9	2	2	6			1987	
10	2	2	3			1991	
11	1	2	6			1988	
12	2	2	3			1992	
13	2	2	3			1990	
14	musím něco umět	2	3			1992	
15	2	2	3			1991	
16	2	2	4			1988	
17	2	2	3			1992	
18	3	2	3			1990	
19	3	2	5			1986	
20	2	2	3			1988	
21	2	2	3			1993	
22	3	2	5	8	2	1990	
23	3	2	3	8	1	1991	
24	3	2	3	1	2	1990	
25	2	2	3	8	1	1990	

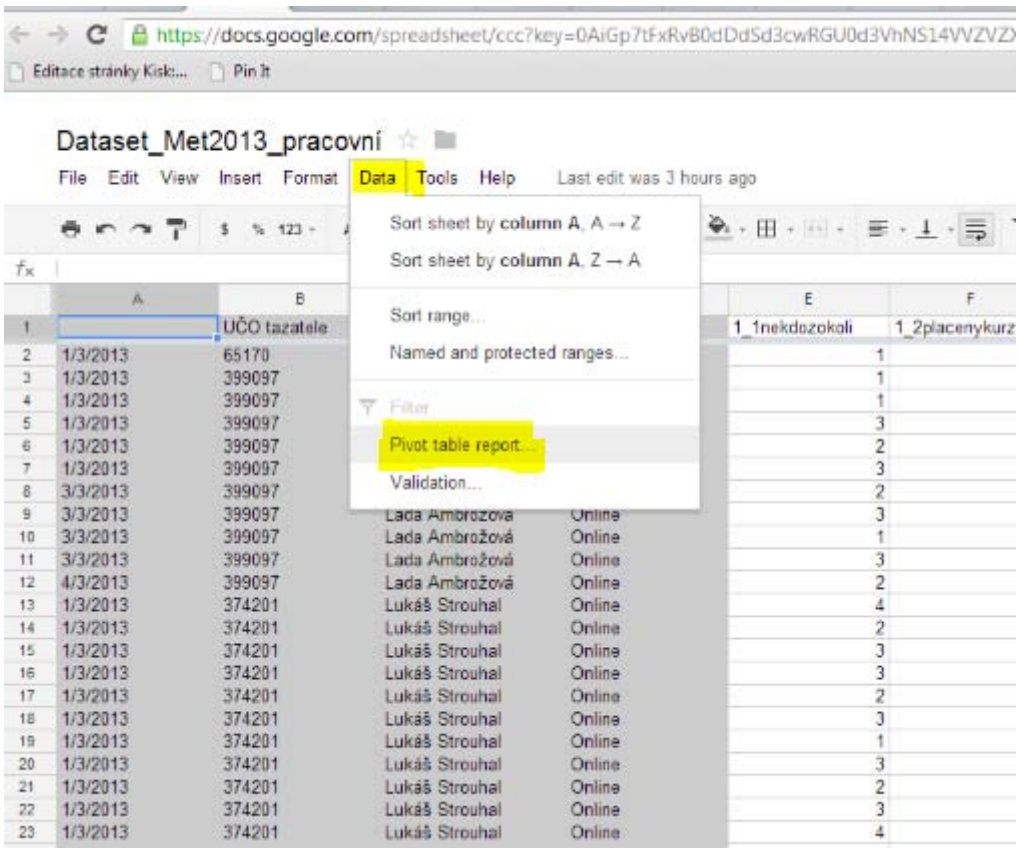
Popis rozložení hodnot proměnných

Pro počítání absolutních četností v Excelu slouží příkaz **COUNTIF**.

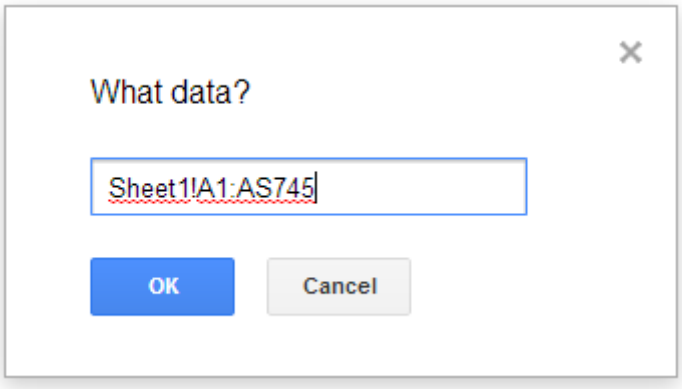
A	B
Prodejce	**Faktura**
Novák	15 000
Novák	9 000
Horák	8 000
Horák	20 000
Novák	5 000
Veselý	22 500
Vzorec	**Popis (výsledek)**
=COUNTIF(A2:A7;"Novák")	Počet faktur od Nováka (3)
=COUNTIF(A2:A7;A4)	Počet faktur od Horáka (2)
=COUNTIF(B2:B7,"< 20000")	Počet faktur s hodnotou nižší než 20 000 (4)
=COUNTIF(B2:B7,">="&B5)	Počet faktur s hodnotou vyšší nebo rovnou 20 000 (2)

Zdroj: http://office.microsoft.com

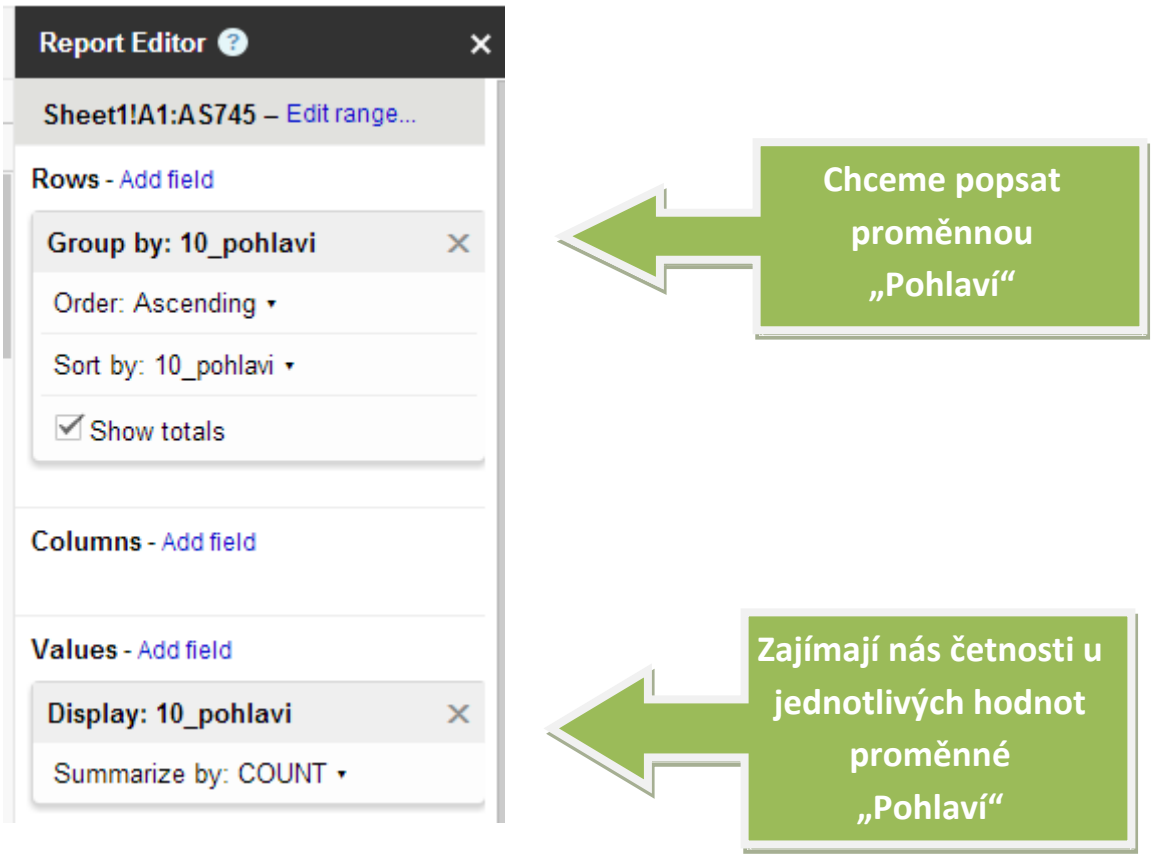
Příkaz COUNTIF nám spočítá výskyt konkrétní varianty hodnoty proměnné. Pro vytvoření tabulky četností je však užitečnější funkce „pivot tables“. Najdete ji v sekci „Data“.



Aplikace se vás nejprve zeptá na rozsah dat. Dávejte si pozor, abyste zahrnuli celou tabulku.



Nová tabulka se vám objeví na novém listu. Tabulku četností vytvoříte tak, že v položce „Řádky“ / „Rows“ specifikujete proměnnou, kterou chcete popsat a proces výpočtu hodnot. Pro tabulku četností budeme nejčastěji používat příkaz „COUNT“.

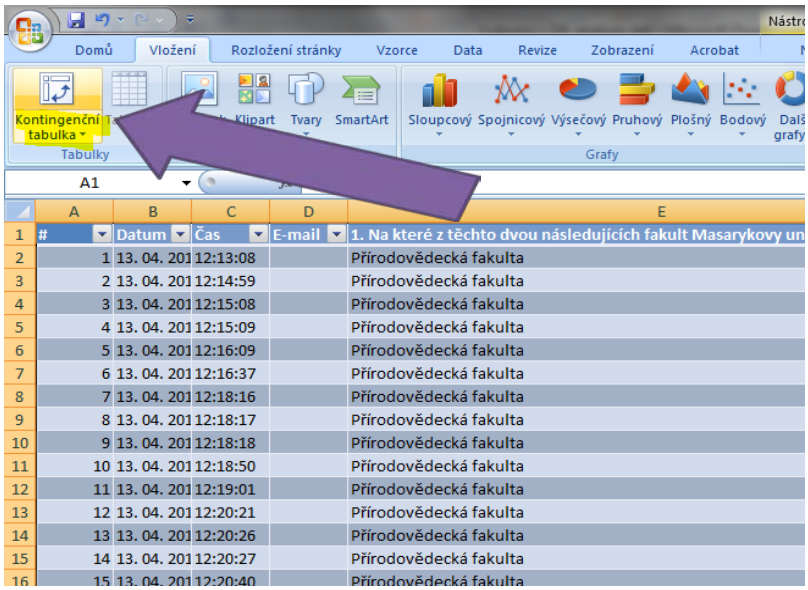


Zpracování v Google Spreadsheets může chvilku trvat, proto buďte trpěliví, pokud tabulka nebude hned reagovat na zadané změny.

Pokud jste si nepřekódovali odpovědi předem, výsledná tabulka bude obsahovat naše kódy, před publikováním je tedy třeba ji ještě upravit – místo kódů (např. „1“) by výsledná tabulka měla obsahovat reálné hodnoty proměnných (např. „muž“).

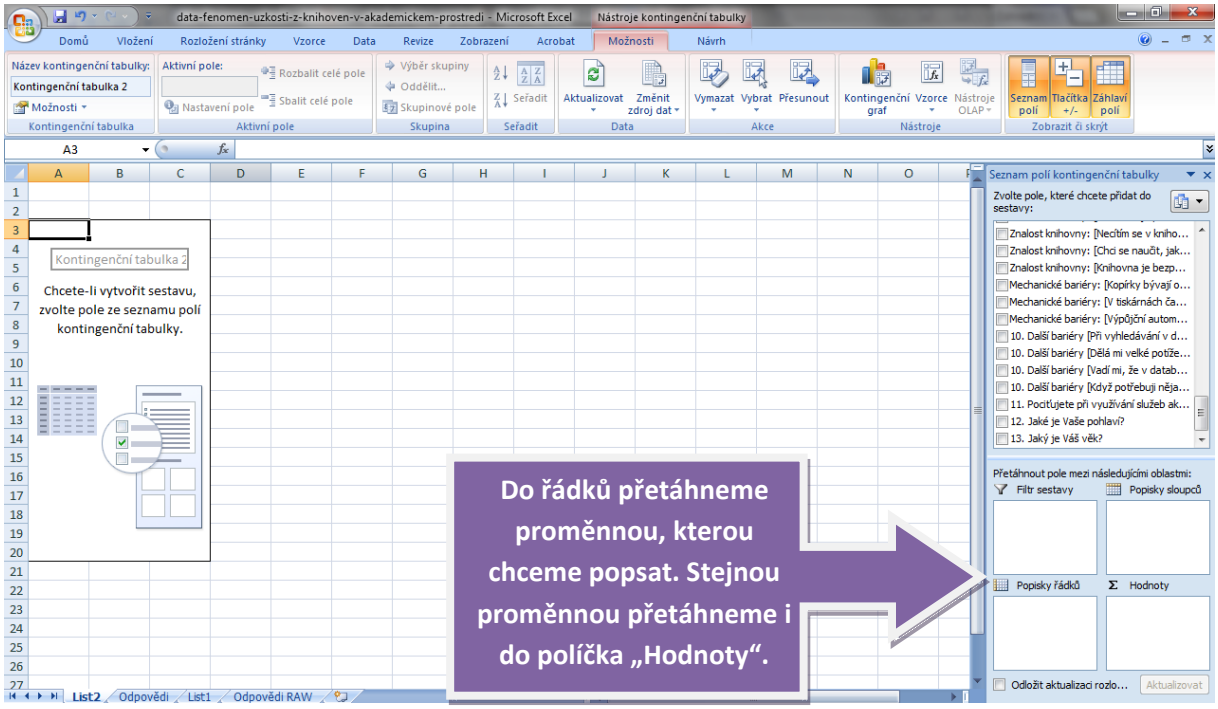
Jste:	**Četnost odpovědí**	**Validní relativní četnost**
Muž	80	40 %
Žena	120	60 %
Celkem	**200**	**100 %**

Pokud jste se rozhodli pracovat v Excelu, je postup velmi podobný. Tabulku vytvoříte tak, že označíte data, se kterými chcete pracovat, a zvolíte možnost „**Kontingenční tabulka**“ na kartě „**Vložení**“.



#	Datum	Čas	E-mail	1. Na které z těchto dvou následujících fakult Masarykovy un
1	13. 04. 201	12:13:08		Přírodovědecká fakulta
2	13. 04. 201	12:14:59		Přírodovědecká fakulta
3	13. 04. 201	12:15:08		Přírodovědecká fakulta
4	13. 04. 201	12:15:09		Přírodovědecká fakulta
5	13. 04. 201	12:16:09		Přírodovědecká fakulta
6	13. 04. 201	12:16:37		Přírodovědecká fakulta
7	13. 04. 201	12:18:16		Přírodovědecká fakulta
8	13. 04. 201	12:18:17		Přírodovědecká fakulta
9	13. 04. 201	12:18:18		Přírodovědecká fakulta
10	13. 04. 201	12:18:50		Přírodovědecká fakulta
11	13. 04. 201	12:19:01		Přírodovědecká fakulta
12	13. 04. 201	12:20:21		Přírodovědecká fakulta
13	13. 04. 201	12:20:26		Přírodovědecká fakulta
14	13. 04. 201	12:20:27		Přírodovědecká fakulta
15	13. 04. 201	12:20:40		Přírodovědecká fakulta

Na novém listu se objeví prostředí pro tvorbu kontingenčních tabulek. Pro tvorbu tabulek četností budeme využívat zatím jen možnosti popisů řádků:

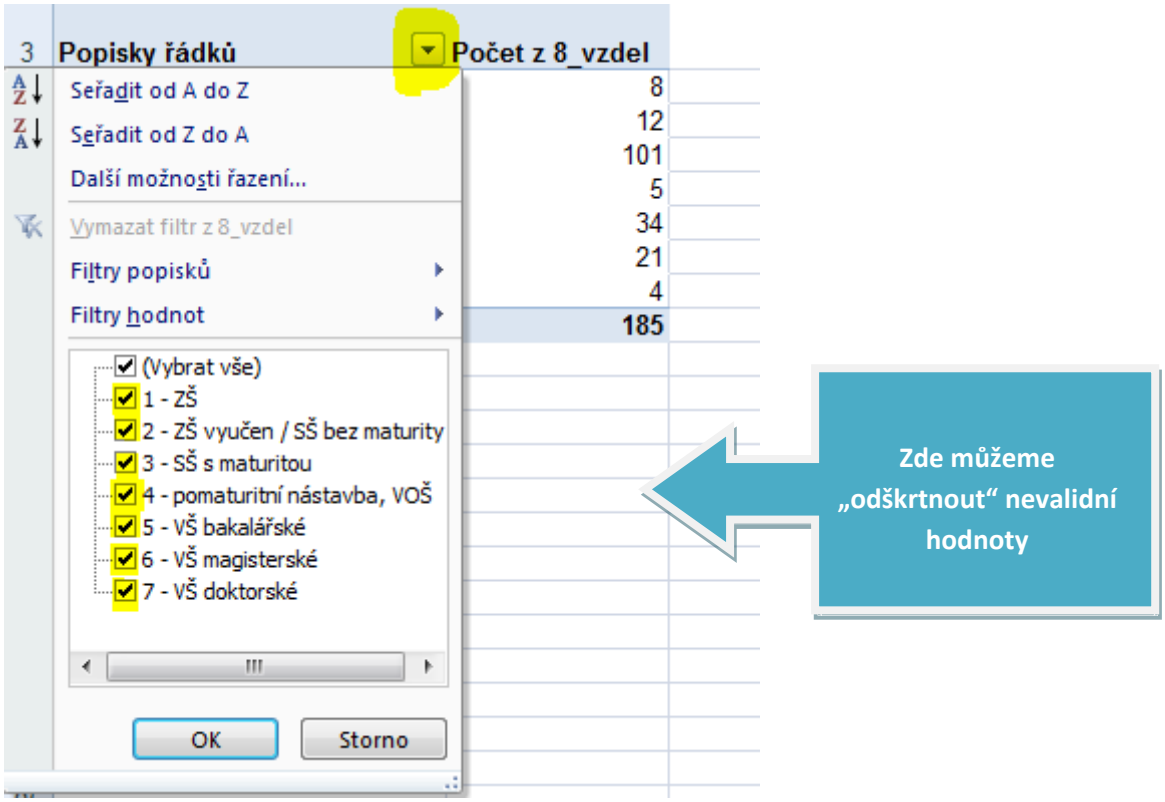


Pro ukázku si vytvořme tabulku se vzděláním:

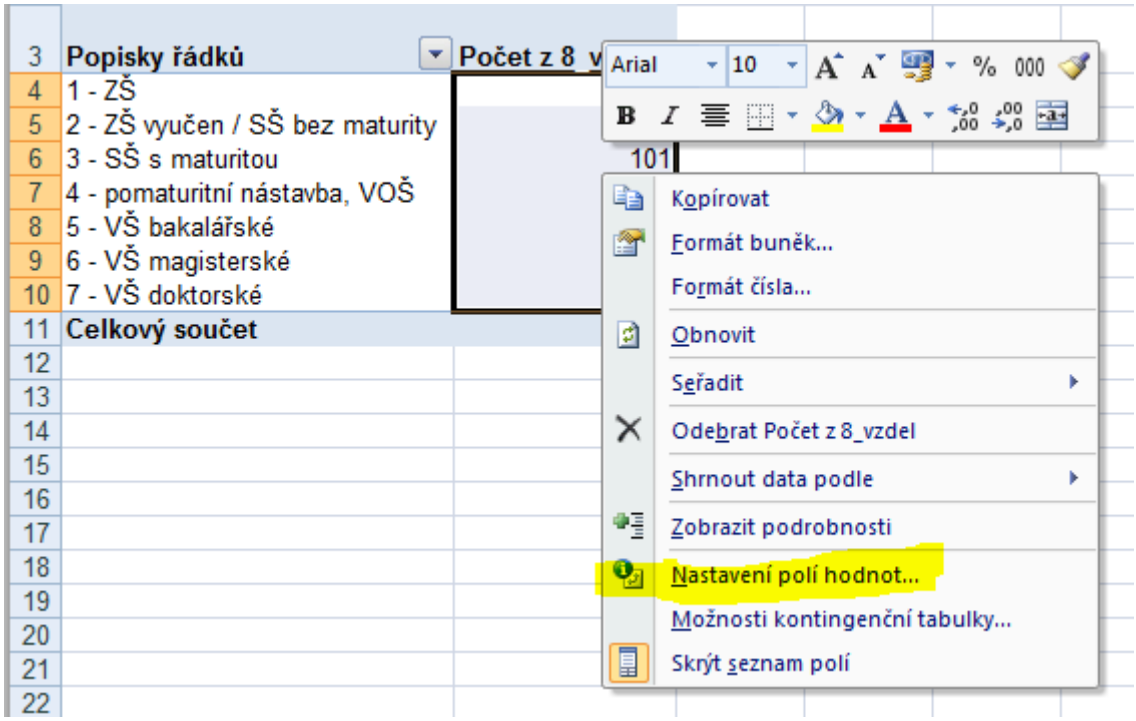
Popisky řádků	Počet z 8_vzdel
1 - ZŠ	8
2 - ZŠ vyučen / SŠ bez maturity	12
3 - SŠ s maturitou	101
4 - pomaturitní nástavba, VOŠ	5
5 - VŠ bakalářské	34
6 - VŠ magisterské	21
7 - VŠ doktorské	4
Celkový součet	**185**

Tato tabulka ukazuje hodnoty výskytu jednotlivých variant odpovědí u proměnné „Dokončené vzdělání“

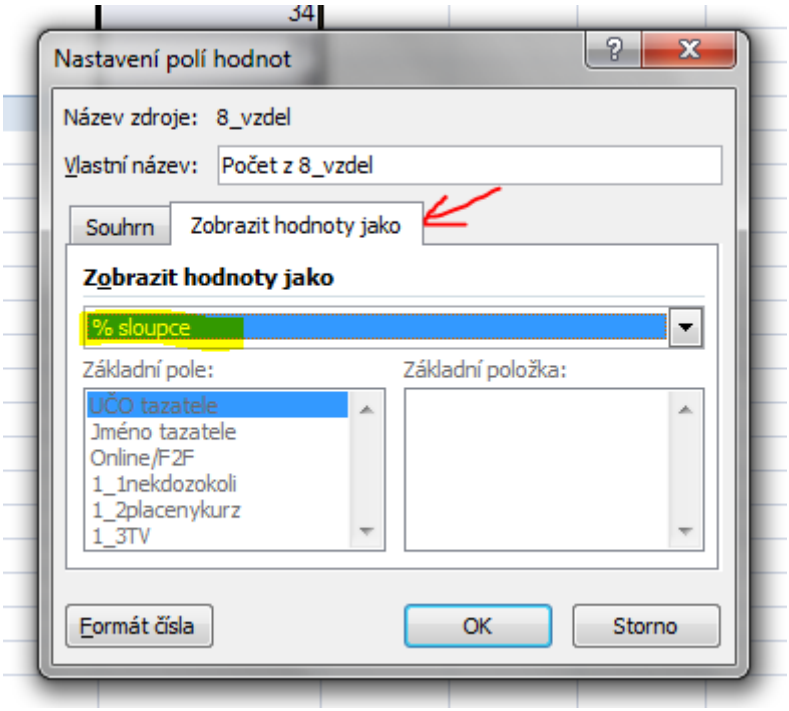
Pokud máme v otázce varianty odpovědí, které nechceme zahrnovat do analýzy (tzv. nevalidní odpovědi – tedy odpovědi typu „nevím“, „neodpověděl“), můžeme je odškrtnout v rozbalovacím menu:



Chceme-li přepočítat absolutní četnosti na relativní četnosti, klikneme na datovou oblast pravým tlačítkem myši a zvolíme možnost „**Nastavení polí hodnot**“:



Vybereme záložku „**Zobrazit hodnoty jako**“ a zvolíme „% sloupce“. Absolutní hodnoty se přepočítají na procenta:



Získáme tak **relativní četnosti**:

Popisky řádků	Počet z 8_vzdel
1 - ZŠ	4,32%
2 - ZŠ vyučen / SŠ bez maturity	6,49%
3 - SŠ s maturitou	54,59%
4 - pomaturitní nástavba, VOŠ	2,70%
5 - VŠ bakalářské	18,38%
6 - VŠ magisterské	11,35%
7 - VŠ doktorské	2,16%
Celkový součet	**100,00%**

Minimální a maximální hodnoty

Minimální a maximální hodnoty lze rozpoznat už z popisu rozloženíproměnných. U spojitých nekategorizovaných dat ale popis rozložení četností nepoužíváme, proto je výhodnější znát příkaz na rychlé zjištění minimálních a maximálních hodnot. V Excelu i v Google Spreadsheet se tyto hodnoty zjišťují pomocí funkce MIN a MAX. Zapisují se do políčka jako příkaz ve tvaru

„=MIN(datová oblast)“ či **„=MAX(datová oblast)“**

	A
1	Data
2	10
3	7
4	9
5	27
6	2

Vzorec	Popis (výsledek)
=MIN(A2:A6)	Nejmenší z výše uvedených čísel (2)
=MIN(A2:A6;0)	Nejmenší z výše uvedených čísel a čísla 0 (0)

	A
1	Data
2	10
3	7
4	9
5	27
6	2

Vzorec	Popis (výsledek)
=MAX(A2:A6)	Největší z výše uvedených čísel (27)
=MAX(A2:A6;30)	Největší z výše uvedených čísel a čísla 30 (30)

Využívejte podpory a nápovědy!

Pokud si nejste jistí provedením příkazu, využívejte podpory Microsoft Office i Google Spreadsheets. Na internetu lze najít také spoustu videotutorialů a návodů. V nejhorším případě pište na sucha@phil.muni.cz ☺.

Literatura

Disman, M. (2002) Jak se vyrábí sociologická znalost. Praha: Karolinum.

Ioannidis JPA (2005) Why Most Published Research Findings Are False. PLoS Med 2(8): e124.

Wheelan, Ch. (2013) Naked Statistics. New York: W. W. Norton & Company Ltd.