



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



Statistika pro střední školu

Autoři:	Jiří Břehovský, Daniela Bímová
Vzdělávací oblast:	Matematika a její aplikace
Tematický okruh:	Práce s daty
Typy úloh:	Analýza a zpracování statistických dat s využitím programu Excel
Rozvíjené oblasti digitálních kompetencí:	1.3. Správa dat, informací a obsahu; 3.1. Vytváření digitálního obsahu
Klíčová slova:	statistický soubor, rozdělení četností (tabulky četností, grafická znázornění četností), charakteristiky polohy a variability (aritmetický průměr, medián, módus, rozptyl, směrodatná odchylka), korelace, Excel

Předkládaný DVZ lze využít při výuce statistiky na středních školách. Primárně je určen učitelům matematiky a může sloužit jako námět a jistý návod, jak lze při výuce statistiky využít výpočetní techniku.

Studenti se s nejrůznějšími druhy dat, jejich prezentací a interpretací setkávají neustále nejen ve sdělovacích prostředcích. Schopnost rozumět tomu, jak se data zpracovávají, je nezbytnou podmínkou pro kritické posouzení informačních zdrojů, které podávají interpretaci statistického zpracování těchto dat. RVP pro gymnázia tyto skutečnosti reflektuje a mezi očekávanými výstupy se objevuje schopnost pracovat s daty. Úlohy ze statistiky (s výjimkou jednoduchých ukázkových příkladů) ovšem nemá smysl řešit s křídou u tabule, ev. s kalkulačkou. Chceme-li žákům ukázat skutečně hromadné zpracování reálných dat (nikoli jen umělých učebnicových příkladů, ve kterých datový soubor čítá několik položek), je nezbytné použít počítač.

Digitální vzdělávací zdroj Statistika pro SŠ se zaměřuje na rozvíjení a zvyšování digitální gramotnosti studentů prostřednictvím rozvoje klíčových kompetencí ve vzdělávací oblasti Matematika a její aplikace, konkrétně v okruhu Práce s daty. Tematicky je zaměřen na zpracovávání a analyzování předem zadaných reálných dat a statistických souborů. K těmto popisům a analýzám je využíváno výhradně prostředí programu Excel. Úlohy, které studenti mohou řešit, svým zaměřením nepřevyšují vzdělávací oblast popisné statistiky podle její charakteristiky v RVP pro Gymnázia. Celý soubor je připraven tak, aby

při řešení úloh umožňoval studentům přirozeně rozvíjet i digitální gramotnost. Primárně jde o rozvíjení digitálních kompetencí správy dat, informací a digitálního obsahu.

DVZ Statistika pro SŠ se skládá ze dvou souborů: Soubor se zadáním úloh a s návodem, jak s DVZ pracovat (DVZ_Statistika_pro_SS) a soubor s tabulkami v programu Excel (DVZ_Statistika_SS_Datove_tabulky). Druhý uvedený soubor je rozdělen do několika sešitů, které obsahují statistické tabulky se zadávanými daty a možným řešením úloh. Řešení úloh je primárně určeno vyučujícím a slouží k rychlejší kontrole případných studentských řešení. Listy mají stejný název jako příslušná zadávaná úloha. V textu je také uveden stručný návod, jak pomocí Excelu dané úlohy řešit.

Předkládané úlohy je vhodné zařadit na závěr teoretického výkladu jako ukázkou komplexní práce s daty, na které si studenti ověří a prohloubí nově získané poznatky ze statistiky. Mohou tak využít své poznatky z informatiky o práci s tabulkovým editorem a seznámit se s novými statistickými funkcemi programu Excel.

DVZ Statistika pro SŠ si neklade za cíl výuku statistiky, pouze předkládá možné využití programu Excel při zpracování dat. Cílem DVZ je ukázat, že využití výpočetní techniky při výuce matematiky je možné, účelné a v některých případech nezbytné. Nabízí učitelům připravené a řešené úlohy, které lze využít přímo při výuce statistiky a snaží se tak vyučujícímu usnadnit práci s jejich přípravou. Data jsou vybrána tak, aby jednotlivé úkoly měly smysl a abychom ukázali význam jednotlivých statistických charakteristik.

V závěru dokumentu jsou v seznamu literatury uvedeny, užitečné webové odkazy, které obsahují zajímavá statistická data z mnoha oblastí, i některé statistické analýzy.

V následujícím textu předkládáme souhrn základních statistických pojmů tak, jak jsou vymezeny v základní středoškolské literatuře, např. [1] a které budeme při řešení úloh využívat. Neklademe si za cíl tyto pojmy blíže vysvětlovat, uvádíme je pouze jako výčet základních pojmů, které jsou k řešení předkládaných úloh nutné, a jen a pouze z důvodu úplnosti předkládaného textu.

ZÁKLADNÍ STATISTICKÉ POJMY

Výsledkem statistického šetření je zpravidla seznam jednotek s udáním hodnoty znaku u každé z nich. Pokud jednotky v seznamu očíslováme $1, 2, 3, \dots, n$, pak jim odpovídající hodnoty znaku x označujeme x_1, x_2, \dots, x_n . V některých případech může znak x nabývat jen určitého počtu r různých hodnot, ty pak označujeme $x_1^*, x_2^*, \dots, x_r^*$. Pro každou možnou hodnotu x_j^* pak zjistíme, kolikrát se vyskytla mezi x_1, x_2, \dots, x_n . Tomuto počtu n_j říkáme **četnost hodnoty** x_j^* . Součet četností všech možných hodnot znaku se rovná počtu všech jednotek souboru:

$$\sum_{j=1}^r n_j = n. \quad (1)$$

Relativní četnost v_j značí, jaká část souboru má hodnotu znaku x_j^* , lze ji vypočítat ze vzorce

$$v_j = \frac{n_j}{n}, \quad (2)$$

kde n_j je četnost příslušného znaku a n je celková četnost souboru. Součet relativních četností se rovná jedné. Relativní četnost lze také vyjádřit v procentech jako

$$v_{j[\%]} = \frac{n_j}{n} \cdot 100, \quad (3)$$

v tomto případě je součet relativních četností roven 100 %.

Nejčastěji používanou charakteristikou polohy je **aritmetický průměr** \bar{x} , který je definován jako:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (4)$$

kde $x_1, x_2, x_3 \dots x_n$ jsou naměřené číselné hodnoty a n je celková četnost všech hodnot. Ve většině případů je nezbytné vypočítat průměr z dat, která jsou obsažena v tabulce četností. V takových případech můžeme pro výpočet aritmetického průměru využít vzorec

$$\bar{x} = \frac{1}{n} \sum_{j=1}^r x_j^* \cdot n_j, \quad (5)$$

kde n je celková četnost všech hodnot, x_j^* je určitá hodnota, n_j je četnost hodnoty x_j^* .

Medián \tilde{x} je prostřední hodnota z řady naměřených hodnot seřazených podle velikosti.

Modus \hat{x} je ta hodnota, která se v daném souboru vyskytuje nejčastěji (která má největší četnost).

Je-li charakteristikou polohy, jako bude i v našem případě, aritmetický průměr, pak za charakteristiku variability volíme zpravidla **rozptyl**, respektive **směrodatnou odchylku**. Výpočet rozptylu ze základního souboru lze provést podle vzorce

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (6)$$

kde x_i je určitá naměřená číselná hodnota, n je celková četnost všech hodnot a \bar{x} je aritmetický průměr. V případě, že rozptyl počítáme z tabulky četností, využijeme vzorec

$$s^2 = \frac{1}{n} \sum_{j=1}^r n_j (x_j^* - \bar{x})^2, \quad (7)$$

kde x_j^* je určitá naměřená číselná hodnota, n_j je četnost hodnoty x_j^* , n je celková četnost všech hodnot a \bar{x} je aritmetický průměr. Pro výpočet je výhodnější použít upravený vzorec

$$s^2 = \frac{1}{n} \sum_{j=1}^r n_j \cdot x_j^{*2} - \bar{x}^2. \quad (8)$$

Výpočet podle tohoto upraveného vzorce je podstatně jednodušší a navíc bývá i přesnější s ohledem na možné chyby způsobené zaokrouhlováním. **Směrodatnou odchylku** s vypočítáme jako druhou odmocninu z rozptylu, tedy

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}. \quad (9)$$

Chceme-li charakterizovat variabilitu znaku bezrozměrným číslem, použijeme tzv. **variální koeficient** v_x , který je v procentech definován jako

$$v_x = \frac{s}{\bar{x}} \cdot 100\%, \quad (10)$$

kde s je směrodatná odchylka a \bar{x} je aritmetický průměr. Variační koeficient vyjadřuje, kolik procent z průměrné hodnoty směrodatná odchylka činí.

Je-li medián „prostřední“ hodnota, pak **první kvartil** (dolní kvartil) Q_1 je hodnota „čtvrtinová“ a **třetí kvartil** (horní kvartil) Q_3 hodnota „tříčtvrtinová“.

Jestliže popisujeme dvojice znaků (x, y) , mezi kterými může být nějaká statistická závislost, pak výsledkem měření v souboru dat o rozsahu n jsou data (x_1, y_1) , (x_2, y_2) , ..., (x_n, y_n) . Použijeme-li za charakteristiky polohy aritmetické průměry \bar{x} a \bar{y} a za charakteristiky variability směrodatné odchylky s_x a s_y , pak za charakteristiku závislosti volíme zpravidla **koeficient korelace** r_{xy} . Je definován vzorcem

$$r_{xy} = \frac{\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x \cdot s_y}. \quad (11)$$

Po úpravě čitatele získáme vzorec, který je výhodnější pro ruční výpočet, tj.

$$r_{xy} = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}}{s_x \cdot s_y}. \quad (12)$$

Pomocí koeficientu korelace zjišťujeme, zda je mezi daty lineární vztah. Koeficient korelace je vždy číslo z intervalu $\langle -1; 1 \rangle$. Přičemž krajních hodnot 1 a -1 nabývá tehdy, je-li mezi znaky x a y lineární závislost (tedy nejen statistická). Znaménko korelačního koeficientu závisí na čitateli zlomku. Je-li statistická závislost mezi znaky taková, že nadprůměrným hodnotám x zpravidla odpovídají nadprůměrné hodnoty y , pak bude v čitateli zlomku ve vzorci (12) většina součinů kladných, a tedy i koeficient korelace bude kladný (např. procentuální zastoupení obézních lidí v souboru zemí a údaje o ročním počtu úmrtí v těchto zemích souvisejících s obezitou). Naproti tomu jestliže nadprůměrným hodnotám x zpravidla odpovídají podprůměrné hodnoty y , pak bude v čitateli zlomku ve vzorci (12) většina součinů záporných, a tedy i koeficient korelace bude záporný. Konkrétněji přiblížíme pojem korelace v řešení úlohy s názvem Dopravní nehody.

ZADÁNÍ ÚLOH

V následujícím textu předkládáme zadání čtyř úloh, jejichž datové tabulky jsou uvedeny v listech excelovského souboru. Data byla čerpána na stránkách ČSÚ. U první úlohy (Dopravní nehody) uvádíme i komentář k řešení a návod, jak lze s DVZ pracovat. Při řešení využíváme terminologii a vzorce uvedené v [1]. Při řešení dalších úloh předpokládáme obdobný postup jako v úloze Dopravní nehody. V listech se zadáním dat jsou ponechány čísla tabulek tak, jak je udává ČSÚ.

Dopravní nehody

Tabulka v souboru Datové tabulky v listu Dopravní nehody udává statistická data týkající se dopravních nehod v České republice v letech 1989 až 2017.

(zdroj ČSÚ)

- a) Popište, co tvoří statistický soubor, statistické jednotky a z hlediska jakého znaku jednotky vyšetřujeme. Rozhodněte, zda jde o kvantitativní nebo kvalitativní znak.
- b) Sestrojte polygony četností dopravních nehod a věcných škod v předkládaných obdobích.
- c) Proveďte statistickou analýzu dat s ohledem na počet nehod v jednotlivých letech (Určete aritmetický průměr, medián, příslušné charakteristiky variability).
- d) Proveďte statistickou analýzu dat s ohledem na výši věcné škody v jednotlivých letech.
- e) Zjistěte, jaká je korelace:
 - i. mezi počty nehod a počty úmrtí,
 - ii. mezi počty nehod a počty zraněných,
 - iii. mezi počty nehod a výší věcné škody.

KOMENTÁŘ K ŘEŠENÍ

ad a)

Před vlastním řešením musí studenti provést základní analýzu předkládaných dat, to je samozřejmě nezbytné k jejich dalšímu zpracování. Kromě jiných skutečností (popisu toho, co tvoří statistický soubor, statistické jednotky, z hlediska jakého znaku jednotky vyšetřujeme, musí dále rozhodnout, zda jde o kvantitativní nebo kvalitativní znaky) si studenti musí uvědomit, že tabulka obsahuje data dvojího druhu, protože se v roce 2009 změnila legislativa pro hlášení dopravních nehod. To je uvedeno pod tabulkou s daty a v tabulce zvýrazněno červeným podtržením řádku pro rok 2008 (viz výřez tabulky).

	Dopravní nehody	Usmrcené osoby ¹⁾	z toho zemřelí do 24 hodin	Zraněné osoby	v tom		Věcná škoda (mil. Kč)
					těžce	lehce	
∴	∴	∴	∴	∴	∴	∴	∴
2006	187965	1063	956	13106	3990	24231	9116
2007	182736	1222	1123	12427	3960	25382	8467
2008	160376	1076	992	28585	3809	24776	7741
2009	74815	901	832	27313	3536	23777	4981
2010	75522	802	753	24433	2823	21610	4925
2011	75137	773	707	25611	3092	22519	4628
∴	∴	∴	∴	∴	∴	∴	∴

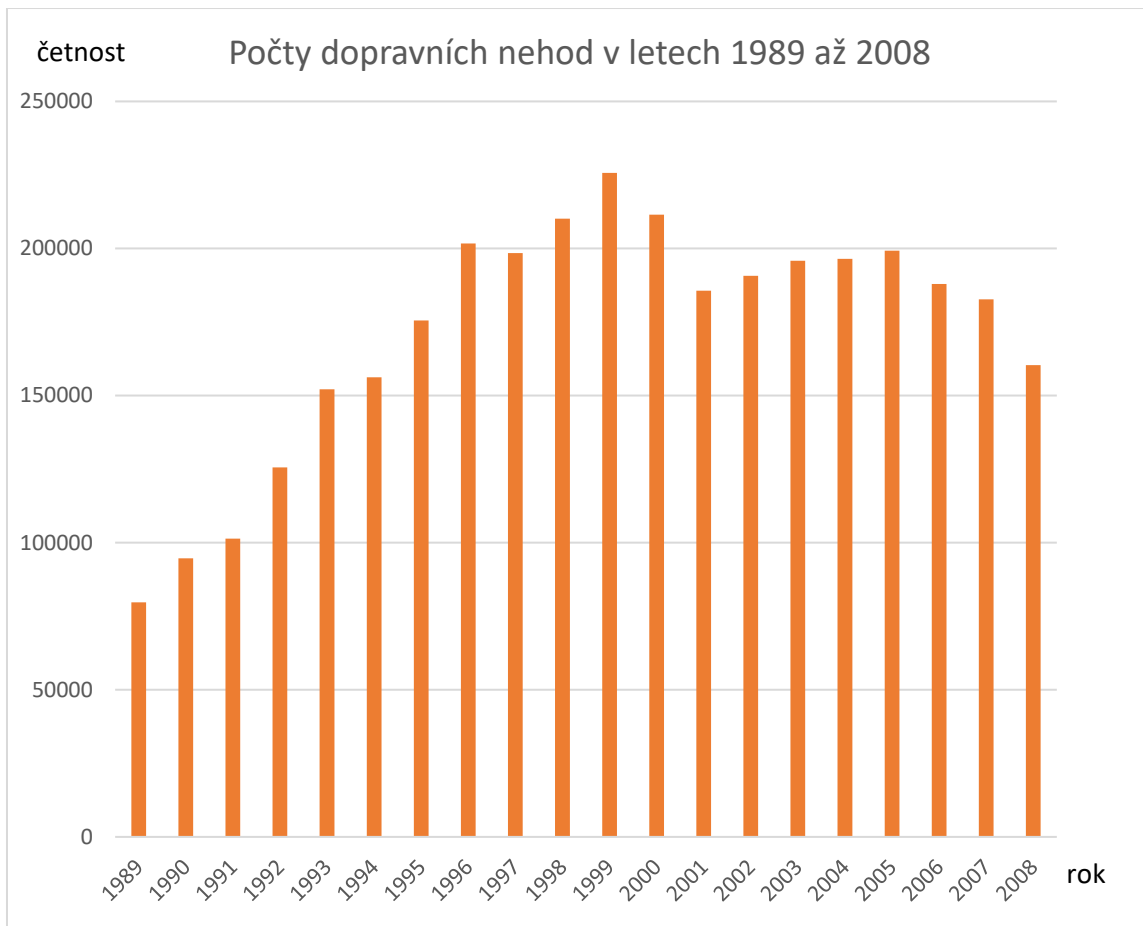
Pozn.: od roku 2009 změna pravidel pro hlášení dopravních nehod Policií ČR

Před vlastní analýzou by měli studenti sami (nebo s pomocí učitele) navrhnout správný postup řešení. Tedy rozdělení dat do dvou disjunktních skupin. Analýzu je tedy nutné provádět zvlášť pro roky 1989 až 2008 a zvlášť pro roky 2009 až 2017 (viz list Dopravní nehody řešení).

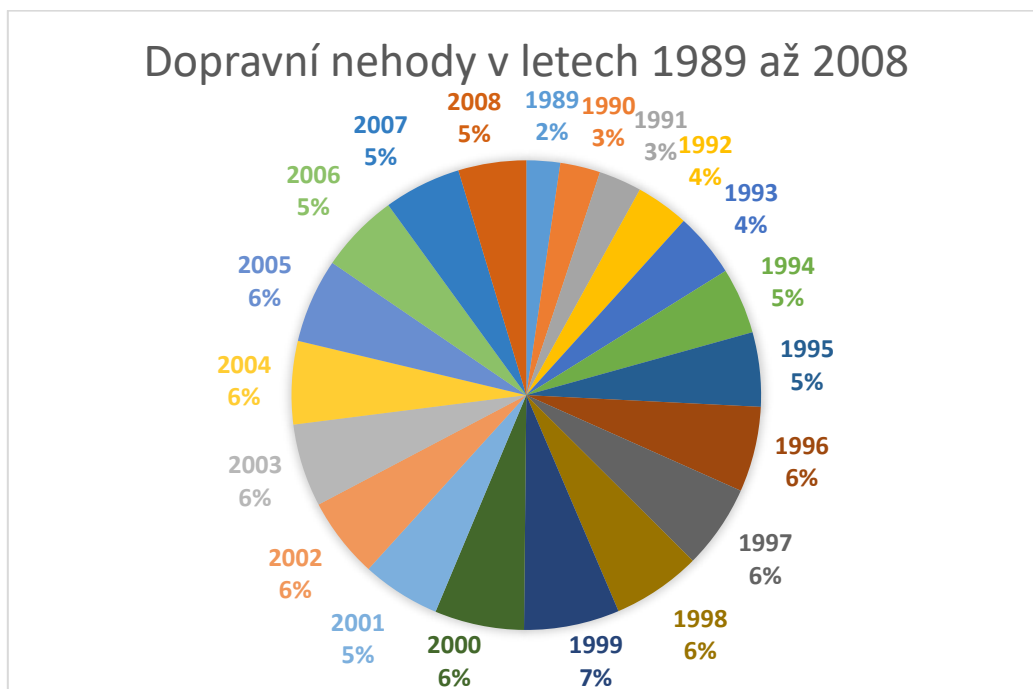
ad b)

Pro grafická znázornění lze použít i jiné grafy, než které uvádíme v řešení. Studenti by měli být schopni zdůvodnit, které grafy jsou k interpretaci analyzovaných dat vhodné a které nikoliv.

Jde vždy o to, jaká data chceme graficky znázornit. V tomto případě lze kromě polygonu (viz list Dopravní nehody řešení) použít například sloupcový diagram.



Kruhový diagram na tomto místě uvádíme jako **příklad nevhodného typu** grafu pro zobrazení zmíněných četností. Na tomto místě by bylo vhodné prodiskutovat se studenty důvody, proč právě kruhový diagram, či jiné typy grafů, nejsou pro zobrazení těchto dat vhodné.



ad c), d)

Při vlastních výpočtech by studenti měli být schopni využívat vzorce, se kterými se již při výuce statistiky seznámili (viz základní statistické pojmy nebo [1]). Tyto vzorce by také měli správně používat v prostředí programu Excel pro tzv. ruční výpočet daných hodnot. Účelné je k výpočtům využívat i statistické funkce programu Excel. V předkládaných řešeních úloh uvádíme vždy oba postupy. Po provedení analýz by měli studenti vždy získané hodnoty interpretovat s ohledem na analyzovaný soubor. Studenti by se měli zaměřit na dílčí vypovídající hodnotu jednotlivých charakteristik s ohledem na celý soubor. Tedy vysvětlit, do jaké míry charakterizuje daný soubor dat jeho průměr, co o souboru vypovídají první a druhý kvartil, rozptyl a směrodatná odchylka. Při vlastním vysvětlování by měli využívat své teoretické poznatky z předchozího učiva i využívat definice jednotlivých charakteristik.

ad e)

Před vlastním výpočtem statistické závislosti zadaných znaků (korelace) považujeme za žádoucí, aby se studenti nejprve nad možnými závislostmi zamysleli a pokusili se počítané korelace nejprve odhadnout. Neméně důležité je, aby své odhady zdůvodnili. Tedy, aby vysvětlili, podle čeho usuzují na jimi odhadnuté korelace. Poté odhady ověřili výpočtem a pokusili se vysvětlit případné rozdíly mezi odhady a skutečností. Studenti by vždy měli popsat a vysvětlit význam vypočtených hodnot.

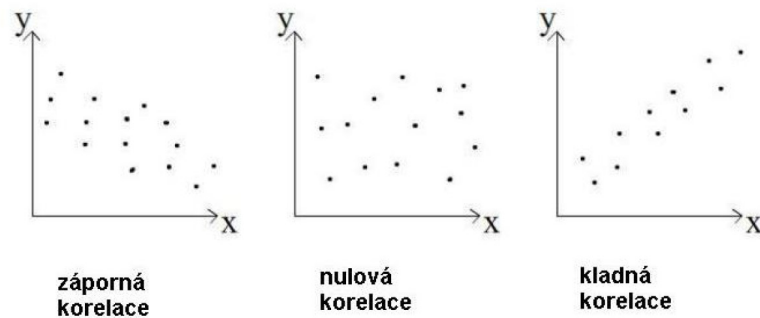
Tedy například pro výpočet korelace mezi počty nehod a výší věcné škody, která mezi lety 2009 až 2017 vychází $r_{xy} = 0,924$, je třeba vysvětlit, zda jde o velkou či malou závislost, proč je daná závislost kladná a proč je korelace mezi počty nehod a počty úmrtí v letech 2009 až 2017 záporná ($r_{xy} = -0,8473$). K tomu můžeme využít následující tabulku přibližné interpretace hodnot korelačního koeficientu, uvedenou v [3] (Míru závislosti zde posuzujeme podle absolutní hodnoty korelačního koeficientu).

Koeficient korelace	Interpretace
$ r_{xy} = 1$	naprostá závislost (funkční závislost)
$1,00 > r_{xy} \geq 0,90$	velmi vysoká závislost
$0,90 > r_{xy} \geq 0,70$	vysoká závislost
$0,70 > r_{xy} \geq 0,40$	střední (značná) závislost
$0,40 > r_{xy} \geq 0,20$	nízká závislost
$0,20 > r_{xy} > 0,00$	velmi slabá závislost
$ r_{xy} = 0$	naprostá nezávislost

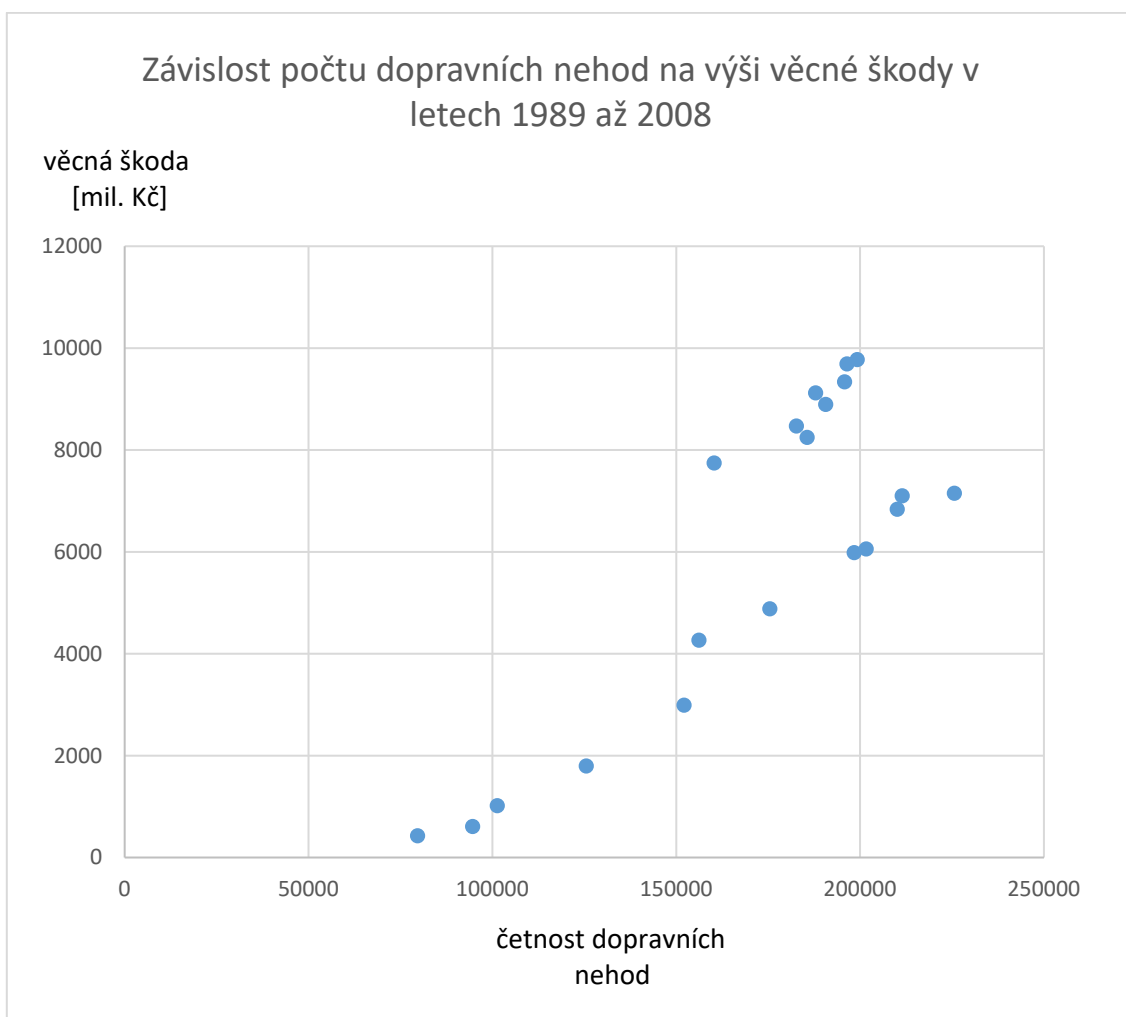
Ve výzkumech se většinou pracuje s koeficienty korelace, jejichž absolutní hodnota je minimálně 0,40. Uvedená tabulka slouží jen k orientačnímu posouzení vypočítané hodnoty

korelace. Určitá hodnota korelačního koeficientu může být hodnocena v různých situacích různě.

Jako nadstavbu lze využít grafického znázornění posuzované korelace pomocí bodového grafu. V takovém případě je nejprve nutné studentům vysvětlit, jak se korelace v takovém grafu projevuje. Využít lze například tyto obrázky:



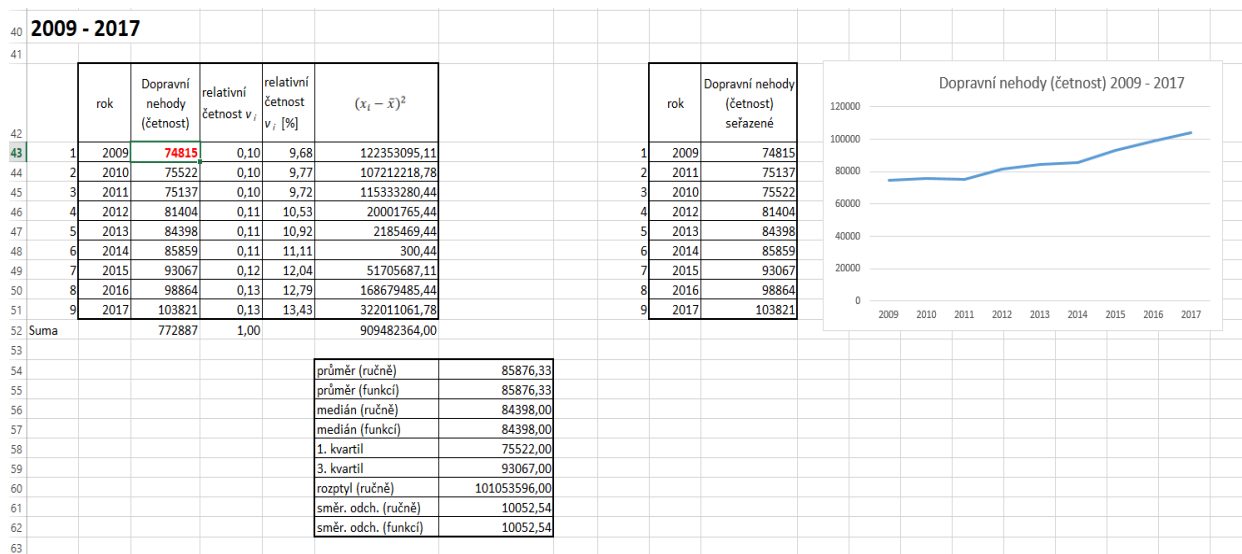
Vždy je také dobré vést diskuzi nad významem vypočtených hodnot s ohledem na počet korelovaných dat. Graf závislosti mezi počty dopravních nehod a výšemi věcné škody v letech 1989 až 2008 v našem řešeném příkladu potom vypadá následovně.



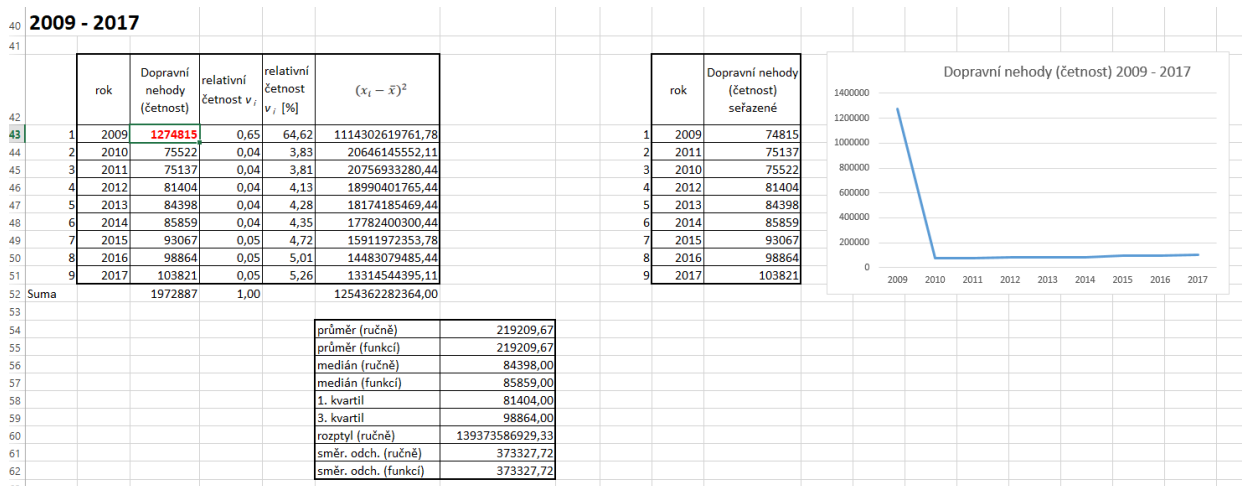
Při využívání korelačního koeficientu je nutné studenty upozornit, že pomocí korelace můžeme prokázat vzájemnou souvislost dvou znaků. Ze vzájemné souvislosti však nijak

nevyplývá příčinný vztah. Například je zřejmé, že výška platu koreluje s cenou soukromého automobilu. To znamená, že lidé s vyššími příjmy mají dražší automobily. Tvrdit však, že si musíme koupit drahé auto, aby nám zvýšili mzdu, by bylo velmi odvážné. Každý cítí, že příčinná souvislost je zřejmě opačná.

Využitím programu Excel při řešení statistických úloh získáváme také jistou míru interaktivnosti analyzovaných dat. V programu lze přímo sledovat vliv provedené změny vstupních dat na výsledných hodnotách. Nabízí se zde možnost využít této skutečnosti pro další práci se studenty, kteří takto mohou přímo sledovat a vyhodnocovat, do jaké míry změny na vstupu ovlivní výsledné hodnoty. Například pro dopravní nehody v letech 2009 až 2017 vypadá správné řešení takto:



Po změně hodnoty v roce 2009 (podbarvená červeně) se výsledky změny následovně:



FUNKCE A PŘÍKAZY VYUŽÍVANÉ V PROGRAMU EXCEL

- V prostředí programu Excel by studenti měli v buňkách správně zadávat a používat vzorce, využívat a chápat rozdíl mezi relativním a **absolutním odkazem** na buňku (absolutní odkaz na buňku provedeme pomocí vložení symbolu „\$“ před a za písmeno tuto buňku označující, např. \$G\$30).

The screenshot shows the Excel interface with the following data in the spreadsheet:

	A	B	C	D	E	F
1	DOPRAVNÍ NEHODY					
2						
3	1989 - 2008					
4						
5		rok	Dopravní nehody (četnost)	relativní četnost v_i	relativní četnost v_i [%]	$(x_i - \bar{x})^2$
6	1	1989	79717	0,02	2,32	$=(C6-\$G\$30)*(C6-\$G\$30)$
7	2	1990	94664	0,03	2,76	5917709480,22

- Pokud potřebujeme data v buňkách **seřadit** podle nějakého klíče (např. vzestupně), využijeme funkci Excelu [Data → Seřadit] a vybereme příslušnou skupinu buněk, v nichž chceme data seřadit.
- Pro výpočet **prvního kvartilu** lze také využít funkci [=quartil()], kterou vepíšeme do vybrané buňky. Do závorek vybereme buňky, ze kterých chceme kvartil určit, a za výběr buněk napíšeme středník a číslo 1, které určuje první kvartil. Například [=QUARTIL(C6:C25;1)]

STDEVPA : \times \checkmark f_x =QUARTIL(C6:C25;1)

	A	B	C	D	E	F	G	H
4								
		rok	Dopravní nehody (četnost)	relativní četnost v_i	relativní četnost v_i [%]	$(x_i - \bar{x})^2$		
5								
6	1	1989	79717	0,02	2,32	8440767564,32		
7	2	1990	94664	0,03	2,76	5917709480,22		
8	3	1991	101387	0,03	2,95	4928552473,32		
9	4	1992	125599	0,04	3,66	2115231869,72		
10	5	1993	152157	0,04	4,43	377666752,32		
11	6	1994	156242	0,05	4,55	235581056,82		
12	7	1995	175520	0,05	5,11	15439791,42		
13	8	1996	201697	0,06	5,88	906392310,32		
14	9	1997	198431	0,06	5,78	720404388,12		
15	10	1998	210137	0,06	6,12	1485821098,32		
16	11	1999	225690	0,07	6,58	2926739670,42		
17	12	2000	211516	0,06	6,16	1594033572,62		
18	13	2001	185664	0,05	5,41	198059180,22		
19	14	2002	190718	0,06	5,56	365855518,02		
20	15	2003	195851	0,06	5,71	588564582,12		
21	16	2004	196484	0,06	5,73	619678874,22		
22	17	2005	199262	0,06	5,81	765703610,82		
23	18	2006	187965	0,05	5,48	268119337,92		
24	19	2007	182736	0,05	5,32	124218826,62		
25	20	2008	160376	0,05	4,67	125768374,62		
26		suma	3431813	1,00		32720308332,55		
27								
28								
29								
30						průměr (ručně)	171590,65	
31						průměr (funkcí)	171590,65	
32						medián (ručně)	186814,50	
33						medián (funkcí)	186814,50	
34						1. kvartil	=QUARTIL(C6:C25;1)	
35						3. kvartil	QUARTIL(matice; kvartil)	
36						rozptyl (ručně)	1636015416,63	

- Pro výpočet **třetího kvartilu** lze také využít příkaz [=quartil()], který vepíšeme do vybrané buňky. Do závorek vybereme buňky, ze kterých chceme kvartil určit, a za výběr buněk napíšeme středník a číslo 3, které určuje třetí kvartil. Například [=QUARTIL(C6:C25;3)].
- Pro určení **mediánu** s využitím funkce vložíme do vybrané buňky příslušnou funkci programu Excel [Vložit funkci → Statistické → Median] a vybereme příslušnou skupinu buněk, ze které chceme medián určit.

	A	B	C	D	E	F	G	H
4								
		rok	Dopravní nehody (četnost)	relativní četnost v_i	relativní četnost v_i [%]	$(x_i - \bar{x})^2$		
5								
6	1	1989	79717	0,02	2,32	8440767564,32		
7	2	1990	94664	0,03	2,76	5917709480,22		
8	3	1991	101387	0,03	2,95	4928552473,32		
9	4	1992	125599	0,04	3,66	2115231869,72		
10	5	1993	152157	0,04	4,43	377666752,32		
11	6	1994	156242	0,05	4,55	235581056,82		
12	7	1995	175520	0,05	5,11	15439791,42		
13	8	1996	201697	0,06	5,88	906392310,32		
14	9	1997	198431	0,06	5,78	720404388,12		
15	10	1998	210137	0,06	6,12	1485821098,32		
16	11	1999	225690	0,07	6,58	2926739670,42		
17	12	2000	211516	0,06	6,16	1594033572,62		
18	13	2001	185664	0,05	5,41	198059180,22		
19	14	2002	190718	0,06	5,56	365855518,02		
20	15	2003	195851	0,06	5,71	588564582,12		
21	16	2004	196484	0,06	5,73	619678874,22		
22	17	2005	199262	0,06	5,81	765703610,82		
23	18	2006	187965	0,05	5,48	268119337,92		
24	19	2007	182736	0,05	5,32	124218826,62		
25	20	2008	160376	0,05	4,67	125768374,62		
26		suma	3431813	1,00		32720308332,55		
27								
28								
29								
30						průměr (ručně)	171590,65	
31						průměr (funkcí)	171590,65	
32						medián (ručně)	186814,50	
33						medián (funkcí)	=MEDIAN(C6:C25)	
34						1. kvartil	MEDIAN(číslo1; [číslo2]; ...)	
35						3. kvartil	198638,75	
36						rozptyl (ručně)	1636015416,63	

- Pro výpočet **aritmetického průměru** s využitím funkce vložíme do vybrané buňky příslušnou funkci programu Excel [Vložit funkci → Statistické → Průměr], poté vybereme příslušnou skupinu buněk, ze které chceme hodnotu aritmetického průměru spočítat.

	A	B	C	D	E	F	G	H
4								
5		rok	Dopravní nehody (četnost)	relativní četnost v_i	relativní četnost v_i [%]	$(x_i - \bar{x})^2$		
6	1	1989	79717	0,02	2,32	8440767564,32		
7	2	1990	94664	0,03	2,76	5917709480,22		
8	3	1991	101387	0,03	2,95	4928552473,32		
9	4	1992	125599	0,04	3,66	2115231869,72		
10	5	1993	152157	0,04	4,43	377666752,32		
11	6	1994	156242	0,05	4,55	235581056,82		
12	7	1995	175520	0,05	5,11	15439791,42		
13	8	1996	201697	0,06	5,88	906392310,32		
14	9	1997	198431	0,06	5,78	720404388,12		
15	10	1998	210137	0,06	6,12	1485821098,32		
16	11	1999	225690	0,07	6,58	2926739670,42		
17	12	2000	211516	0,06	6,16	1594033572,62		
18	13	2001	185664	0,05	5,41	198059180,22		
19	14	2002	190718	0,06	5,56	365855518,02		
20	15	2003	195851	0,06	5,71	588564582,12		
21	16	2004	196484	0,06	5,73	619678874,22		
22	17	2005	199262	0,06	5,81	765703610,82		
23	18	2006	187965	0,05	5,48	268119337,92		
24	19	2007	182736	0,05	5,32	124218826,62		
25	20	2008	160376	0,05	4,67	125768374,62		
26		suma	3431813	1,00		32720308332,55		
27								
28								
29								
30						průměr (ručně)	171590,65	
31						průměr (funkcí)	=PRŮMĚR(C6:C25)	
32						medián (ručně)	PRŮMĚR(číslo1; [číslo2]; ...)	
33						medián (funkcí)	186814,50	

- Pro výpočet **směrodatné odchylky** s využitím funkce vložíme do vybrané buňky příslušnou funkci programu Excel [Vložit funkci → Statistické → STDEVPA] a vybereme příslušnou skupinu buněk, ze které chceme směrodatnou odchylku spočítat.

	A	B	C	D	E	F	G	H
		rok	Dopravní nehody (četnost)	relativní četnost v_i	relativní četnost v_i [%]	$(x_i - \bar{x})^2$		
5								
6	1	1989	79717	0,02	2,32	8440767564,32		
7	2	1990	94664	0,03	2,76	5917709480,22		
8	3	1991	101387	0,03	2,95	4928552473,32		
9	4	1992	125599	0,04	3,66	2115231869,72		
10	5	1993	152157	0,04	4,43	377666752,32		
11	6	1994	156242	0,05	4,55	235581056,82		
12	7	1995	175520	0,05	5,11	15439791,42		
13	8	1996	201697	0,06	5,88	906392310,32		
14	9	1997	198431	0,06	5,78	720404388,12		
15	10	1998	210137	0,06	6,12	1485821098,32		
16	11	1999	225690	0,07	6,58	2926739670,42		
17	12	2000	211516	0,06	6,16	1594033572,62		
18	13	2001	185664	0,05	5,41	198059180,22		
19	14	2002	190718	0,06	5,56	365855518,02		
20	15	2003	195851	0,06	5,71	588564582,12		
21	16	2004	196484	0,06	5,73	619678874,22		
22	17	2005	199262	0,06	5,81	765703610,82		
23	18	2006	187965	0,05	5,48	268119337,92		
24	19	2007	182736	0,05	5,32	124218826,62		
25	20	2008	160376	0,05	4,67	125768374,62		
26		suma	3431813	1,00		32720308332,55		
27								
28								
29								
30						průměr (ručně)	171590,65	
31						průměr (funkcí)	171590,65	
32						medián (ručně)	186814,50	
33						medián (funkcí)	186814,50	
34						1. kvartil	155220,75	
35						3. kvartil	198638,75	
36						rozptyl (ručně)	1636015416,63	
37						směr. odch. (ručně)	40447,69	
38						směr. odch. (funkcí)	=STDEVPA(C6:C25)	

- Pro výpočet **korelačního koeficientu** s využitím funkce vložíme do vybrané buňky příslušnou funkci programu Excel [Vložit funkci → Statistické → CORREL] a vybereme příslušná data.

SOUBOR DOMŮ VLOŽENÍ ROZLOŽENÍ STRÁNKY VZORCE DATA REVIZE ZOBRAZENÍ

Vymout Vložit Kopírovat Kopírovat formát Schránka

Calibri 12 A⁺ A⁻ Zamovat text Sloučit a zarovnat na střed Číslo % 000 000 000 Podmíněné formátování jako tabulku Normální Chybně Kontrolní bu... Poznám

STDEVPA =CORREL(C46:C54;F46:F54)

rook	Dopravní nehody (četnost) x	Usmrcené osoby y	Zraněné osoby z	Věcná škoda (mil. Kč) q	$x_i \cdot y_i$	$x_i \cdot z_i$	$x_i \cdot q_i$	
1	2009	74815	901	27313	4981	67408315	2043422095	372660323
2	2010	75522	802	24433	4925	60568644	1845229026	371944861
3	2011	75137	773	25611	4628	58080901	1924333707	347740107
4	2012	81404	742	25576	4875	60401768	2081988704	396878478
5	2013	84398	654	25359	4938	55196292	2140248882	416771959
6	2014	85859	688	26417	4933	59070992	2268137203	423562529
7	2015	93067	739	26966	5439	68776513	2509644722	506203028
8	2016	98864	611	27081	5804	60405904	2677335984	573826824
9	2017	103821	577	27079	6316	59904717	2811368859	655760087
Suma						549814046	20301709182	4065348196

	xy	xz	xq
průměr dopravní nehody	85876,33		
směrodatná odchylka dopravní nehody	10052,54		
průměr usmrcené osoby	720,78		
směrodatná odchylka usmrcené osoby	94,78		
průměr zraněné osoby		26203,89	
směrodatná odchylka zraněné osoby		942,05	
průměr věcná škoda			5204,51
směrodatná odchylka věcná škoda			512,49
$\frac{1}{n} \sum_{i=1}^n x_i \cdot y_i$	61090449,56	2255745464,67	451705355,09
$\frac{1}{n} \sum_{i=1}^n x_i \cdot y_i - \bar{x} \cdot \bar{y}$	-807303,15	5451567,81	4761320,83
r_{yz} (ručně)	-0,8473	0,5757	0,9242
r_{yz} (funkcí)	-0,8473	0,5757	=CORREL(C46:C54;F46:F54)

CORREL(matice1; matice2)

Dopravní nehody Dopravní nehody řešení Dopravní nehody korelace Odpady Odpady řešení Demografické údaje Demografické údaje řeše

ÚPRAVY

Odpady

Tabulka v souboru Datové tabulky v listu Odpady udává statistická data týkající se produkce odpadů v České republice v letech 2002 až 2017.

(zdroj ČSÚ)

- a) Popište, co tvoří statistický soubor, statistické jednotky a z hlediska jakého znaku jednotky vyšetřujeme. Rozhodněte, zda jde o kvantitativní nebo kvalitativní znak.
- b) Proveďte statistickou analýzu dat s ohledem na celkové množství produkováných odpadů v jednotlivých letech. Analýzy proveďte také zvlášť pro odpad produkováný domácnostmi a zvlášť pro odpad produkováný podniky (Určete aritmetický průměr, medián, příslušné charakteristiky variability).
- c) Do jednoho grafu znázorněte výši produkováných odpadů v daných letech pro domácnosti a pro průmysl.

Demografické údaje

Tabulka v souboru Datové tabulky v listu Demografické údaje udává statistická data týkající se vybraných demografických údajů o České republice v letech 1989 až 2017.

(zdroj ČSÚ)

- a) Popište, co tvoří statistický soubor, statistické jednotky a z hlediska jakého znaku jednotky vyšetřujeme. Rozhodněte, zda jde o kvantitativní nebo kvalitativní znak.
- b) Sestrojte polygony četností zobrazující celkový počet obyvatel, celkový počet mužů a celkový počet žen v jednotlivých letech.
- c) Proveďte statistickou analýzu dat celkového počtu obyvatel, celkového počtu mužů a celkového počtu žen v jednotlivých letech.
- d) Proveďte statistickou analýzu dat celkového počtu sňatků a rozvodů v jednotlivých letech. Graficky data znázorněte a popište jejich vývoj v daném období.
- e) Zjistěte, jaká je korelace:
 - i. mezi počty sňatků a počty rozvodů,
 - ii. mezi počty sňatků a počtem obyvatel.

Obyvatelstvo a odpady

Jako vstupní data pro tento úkol využijte hodnoty uvedené v úlohách Odpady a Demografické údaje. Z obou tabulek vyberte podstatné informace a využijte pro řešení následujících úkolů.

- a) Popište, co tvoří statistický soubor, statistické jednotky a z hlediska jakého znaku jednotky vyšetřujeme. Rozhodněte, zda jde o kvantitativní nebo kvalitativní znak.
- b) Vytvořte tabulku četností, která udává počty obyvatel ČR a množství produkováného komunálního odpadu v letech 2002 až 2017. Vypočtěte hmotnost produkováného komunálního odpadu v kilogramech na jednoho obyvatele v jednotlivých letech a graficky tato data znázorněte.
- c) Zjistěte, jaká je v letech 2002 až 2017 korelace mezi počty obyvatel a množstvím produkováného komunálního odpadu.

ZÁVĚR

V předkládaných úlohách používáme k analýze reálná data, která jsou pravidelně zveřejňovaná Českým statistickým úřadem. Cílem využití uvedených úloh je mimo jiné také to, aby se studenti naučili pracovat s větším množstvím dat a lépe se orientovali ve zveřejňovaných statistických údajích. Byli schopni tato data třídit, zpracovávat a upravovat podle potřeby. Každou úlohu lze také využít jako samostatnou práci pro studenty, jejímž výsledkem je zpracovaná statistická analýza v podobě seminární práce se všemi náležitostmi včetně závěrečného shrnutí a prezentace.

Úlohy jsou koncipovány jako komplexní úlohy tak, aby umožňovaly i rozvoj digitálních kompetencí jejich řešitelů. Jde především o kompetence v oblasti prohlížení, vyhledávání a filtrování dat a informací, jejich hodnocení a správu, dále pak o kompetence v oblasti vytváření digitálního obsahu. V tomto ohledu je nezbytné, aby studenti získávali také zkušenosti s vyhledáváním potřebných a relevantních dat na internetu. Proto tam, kde je to účelné, by při zadávání úloh mohl vyučující některé potřebné informace záměrně zamlčet. Respektive upozornit studenty, že některé informace potřebné k řešení je třeba najít online. S tím souvisí také důraz na autorská práva a licence. Studenti by měli všechny své zdroje správně citovat dle normy ISO 690. Odkazy, kde lze data dohledat, spolu s dalšími, z našeho pohledu užitečnými odkazy uvádíme v seznamech v závěru textu.

Na internetových stránkách uvedených v užitečných odkazech lze také najít zpracované analýzy různých souborů dat i jejich následnou prezentaci a interpretaci. Považujeme za vhodné využít těchto odkazů k ukázkám toho, jak vypadá profesionální analýza dat a jejich následná prezentace či interpretace.

Všechny tabulky záměrně uvádíme bez jakýchkoliv grafických úprav (zvýraznění textu nebo podbarvení významných dat atp.), důvodů je několik. Studenti by měli procvičovat svou orientaci ve větším množství zdánlivě nepřehledných dat, zlepšit svou schopnost vybírat ze statistického souboru podstatné informace. Umožňujeme jim tím zvolit si sobě blízké zvýraznění dat, která považují za podstatná.

SEZNAM LITERATURY

- [1] CALDA, E., DUPAČ, V.: *Kombinatorika, pravděpodobnost, statistika*. Prometheus Praha 1993. ISBN978-80-7196-365-3.
- [2] DUBSKÁ, D., a kol.: *Školní statistická ročenka 2008*. Praha. Scientia 2008. Dostupné z: <https://www.czso.cz/csu/czso/skolni-statisticka-rocenka-n-japv8r6suy>.
- [3] CHRÁSKA, M.: *Metody pedagogického výzkumu*. Praha. Grada, 2007 ISBN 978-80-247-1369-4.
- [4] MOLNÁR, J., a kol.: *Matematika 9*. Olomouc. Prodos 2001. ISBN 80-7230-109-8.
- [5] Český statistický úřad [online]. ČSU: ©2019. Dostupné z: https://www.czso.cz/csu/czso/pro_studenty.
- [6] MATHS AND STATS SUPPORT CENTRE [online]. *Statistika 1*. MUNI ©2019. Dostupné z https://mathstat.econ.muni.cz/media/12657/pear_cor.pdf

UŽITEČNÉ ODKAZY

- Nejvíce dostupných materiálů lze nalézt na stránkách Českého statistického úřadu <https://www.czso.cz>, konkrétní zajímavé analýzy, články a odkazy například na:
 - <https://www.czso.cz/documents/10180/20564299/137007a4.pdf/1f6d0c29-1260-4e8f-8ec1-d1e882e6fe71?version=1.0>
 - <https://www.czso.cz/documents/10180/74414201/25011918.pdf/4c0c6000-4cbd-4452-8d54-15837865a4d9?version=1.1>
 - https://www.czso.cz/csu/czso/statistika_v_obrazcich
 - https://www.czso.cz/csu/czso/statisticke_zajimavosti
 - https://www.czso.cz/documents/10180/23181337/jak_dat_cislum_smysl.pdf/4f5c5aef-4ff9-408f-8429-4b3ef069cc1a?version=1.2
- Další zajímavý portál s velkým množstvím článků, grafů a analýz je <http://www.statistikaamy.cz/>, kde například v sekci archiv lze nalézt mnoho zajímavých článků.
- Následující odkazy jsou cizojazyčné a lze je tak využít i pro rozvíjení jazykových kompetencí studentů:
 - https://www.imf.org/external/datamapper/NGDP_RPCH@WEO/OEMDC/AD-VEC/WEOWORLD
 - https://ec.europa.eu/eurostat/cache/infographs/youineu/index_en.html
- Pro zájemce o hlubší studium statistiky uvádíme stránku Fakulty strojního inženýrství VUT v Brně:
 - <http://mathonline.fme.vutbr.cz/Popisna-statistika/sc-1146-sr-1-a-139/default.aspx> t