

Statistika ve středoškolské matematice: počítání s daty nebo přemýšlení o datech?

Tomáš Měkota



RVP

G

- diskutuje a kriticky zhodnotí statistické informace a daná statistická sdělení
- volí a užívá vhodné statistické metody k analýze a zpracování dat (využívá výpočetní techniku)
- reprezentuje graficky soubory dat, čte a interpretuje tabulky, diagramy a grafy, rozlišuje rozdíly v zobrazení obdobných souborů vzhledem k jejich odlišným charakteristikám
- **práce s daty** – *analýza a zpracování dat v různých reprezentacích, statistický soubor a jeho charakteristiky (vážený aritmetický průměr, medián, modus, percentil, kvartil, směrodatná odchylka, mezikvartilová odchylka)*

SOV

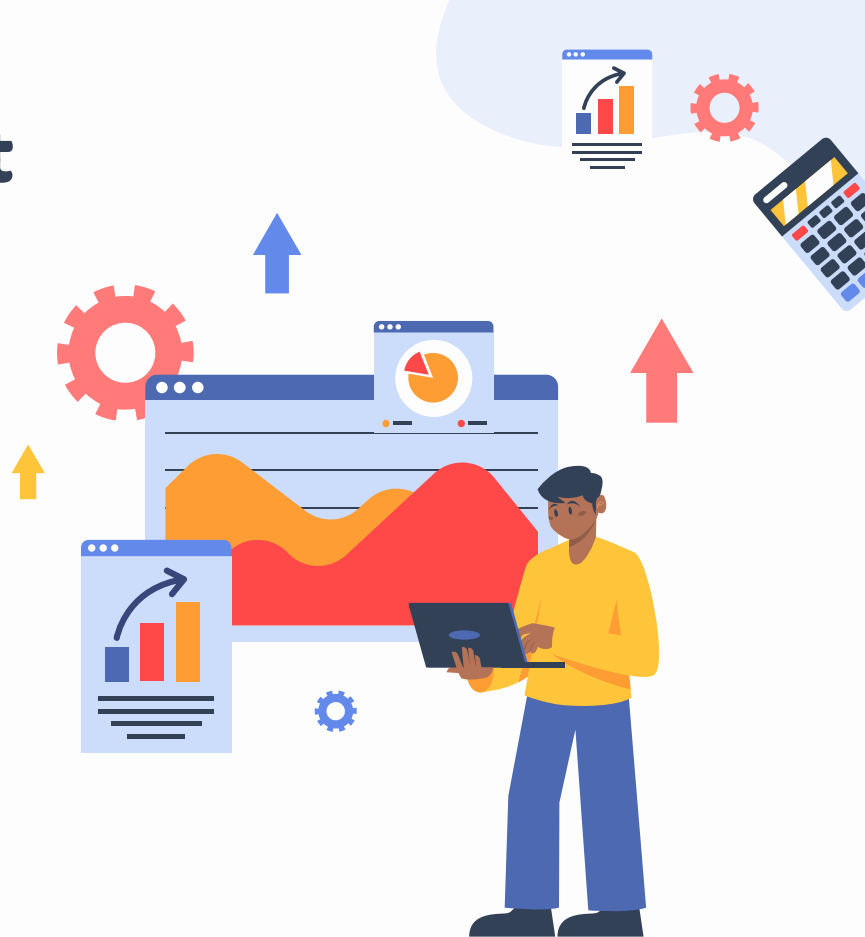
- užívá pojmy: statistický soubor, absolutní a relativní četnost, variační rozpětí
- čte, vyhodnotí a sestaví tabulky, diagramy a grafy se statistickými údaji
- **základy statistiky**

Statistická gramotnost

Galův model

- Poznatková složka
 - Jazyková gramotnost
 - Statistické znalosti
 - Matematické znalosti
 - Kontextové znalosti
 - Kritické otázky
- Dispoziční složka
 - Kritický postoj
 - Víry a postoje

aktivace ve čtenářských kontextech
společná aktivace všech složek



Statistická gramotnost



Watson & Callingham (2003)	Sharma et al. (2011)	GAISE (2007)	
Idiosynkratický	Neformální, idiosynkratická	A	Náhodné, intuitivní odpovědi, základní čtení grafů a tabulek
Neformální			Práce s jazykem a základními významy bez hlubšího chápání variability
Nekonzistentní	Konzistentní nekritická	B	Nekonzistentní práce s kontextem, nekritické používání základních pojmů
Konzistentní nekritická			Konzistentní používání základních dovedností, ale bez kritiky a reflexe
Kritická	Raně kritická	C	Kritické zapojení ve známých kontextech, začátek chápání variability a kontextu
Kritická matematická	Pokročile kritická		Kritické zhodnocení validity, schopnost zobecnění a predikce, plná integrace statistických a kontextových znalostí





Implikace do výuky

Statistická gramotnost je komplexní konstrukt.

nejen čtení grafů a tabulek,
ale i interpretace, kritika,
predikce

Schopnosti nejsou přirozené.

nutnost explicitní výuky

Jděte nad rámec procedurálních dovedností.

důraz na interpretaci, hodnocení
cizích studií, mediálních zpráv

Dlouhodobý proces

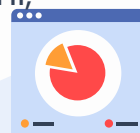
začít v nižších ročnících,
systematický postup

Praktické úkoly a promyšlené otázky

úkoly zasazené do reálného
kontextu (např. z médií),
graduující otázky a aktivity

Vhodné metody

podpora diskuze a
skupinové práce, důraz na
postoje: kritické myšlení,
skepticismus



Statistická gramotnost a RVP

RVP G

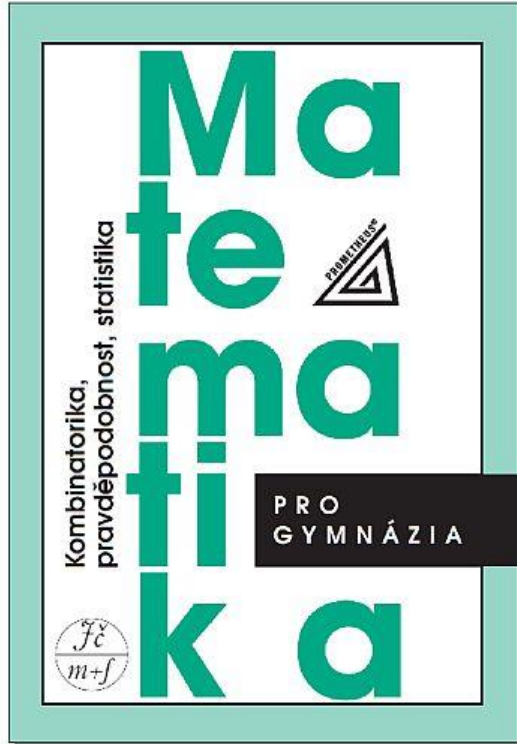
- vysoké požadavky (kritické hodnocení, rozdíly v obdobných souborech)
- vyžaduje aktivaci více složek dle Galova modelu
- dostává se na kritickou, resp. raně kritickou úroveň

RVP SOV

- mnohem nižší požadavky (pouze vyhodnocení tabulek, grafů, blíže nespecifikované)
- dle Galova modelu matematické a statistické znalosti
- dle úrovní nejvýše konzistentní nekritická



Učebnice a statistická gramotnost



- nutná aktualizace
 - pro usnadnění výpočtu PC, ne vlastnosti charakteristik
- chybí interpretace
 - žádné úlohy s grafy
 - porovnání průměru a mediánu
 - význam variačního koeficientu
 - úlohy pouze na výpočty
- obecné vysvětlování

Proč je v příkladu 1 koeficient korelace kladný a v příkladu 2 záporný? Odpovíme na tuto otázku obecněji:

Je-li statistická závislost mezi znaky x a y taková, že nadprůměrným hodnotám x zpravidla odpovídají nadprůměrné hodnoty y a podprůměrným podprůměrné, pak bude v čitateli zlomku ve vzorci (1) většina součinů *kladných*, a tedy i koeficient korelace bude kladný. Např. v souboru zaměstnanců podniku: počet let vzdělávání a roční příjem zaměstnance; v souboru evropských zemí: roční spotřeba cigaret na 1 obyvatele a roční míra úmrtnosti na plicní rakovinu na 100 000 obyvatel.



Učebnice a statistická gramotnost

- důraz na postupy
 - jednoduché řešené úlohy hned u definic
- místy vhodné poznámky k interpretaci nebo k vhodnosti užití
 - ale mají cha

Aritmetický průměr lze spočítat pro „jakákoli“ čísla, ale výsledek, který získáme, má praktický smysl, pokud čísla mají stejnou měřítkovou jednotku. Jinak je výsledek bezvýznamný.

- Můžeme vypočítat průměr z množství čaje v různých obchodech, protože se nejedná o p
- Můžeme vypočítat průměr naměřené teploty v různých místech, protože toto číslo nebude mít výpovědní hodnotu, pro ně odlišné oblasti.

Hasiči

$$s_x = \sqrt{\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2}$$

$$s_x = \sqrt{\frac{841 + 225 + 2704 + 169 + 25}{5}}$$

$$s_x \doteq 28,2$$

$$v_x = \frac{s_x}{\bar{x}} \cdot 100 \%$$

$$v_x = \frac{28,2}{46} \cdot 100 \%$$

$$v_x \doteq 61,3 \%$$

Záchraná služba

$$s_x = \sqrt{\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2}$$

$$s_x = \sqrt{\frac{2304 + 676 + 3844 + 49 + 361}{5}}$$

$$s_x \doteq 38,0$$

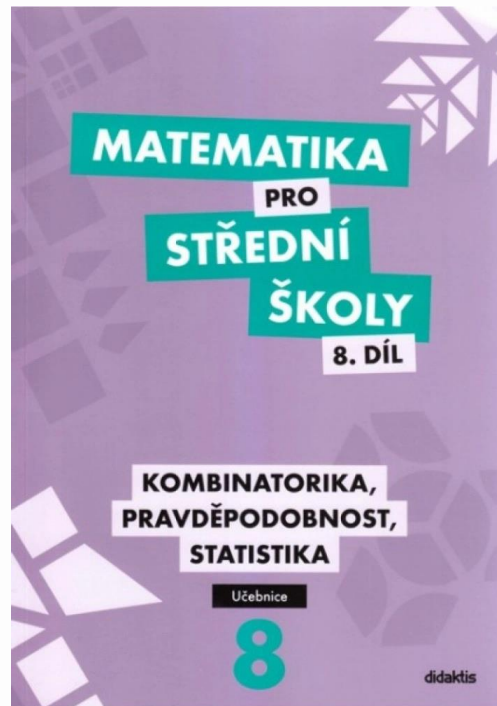
$$v_x = \frac{s_x}{\bar{x}} \cdot 100 \%$$

$$v_x = \frac{38}{169} \cdot 100 \%$$

$$v_x \doteq 22,5 \%$$

Variační koeficient pro hodnoty týkající se výjezdů hasičů je téměř trojnásobně vyšší než variační koeficient pro počet výjezdů záchrané služby. Počty výjezdů záchrané služby během vybraného týdne byly mnohem vyrovnanější než počty výjezdů hasičů.

To, že variační koeficient je bezrozměrné číslo, oceníme, pokud porovnáme více souborů, v nichž jsou hodnoty znaků vyjádřeny v různých jednotkách.





Sbírky úloh a statistická gramotnost



- úlohy pouze na výpočty, nikoli na interpretaci
- u grafů v zadání specifikovaný typ grafu

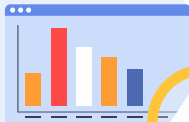


- jen 5 úloh ze statistiky
- pouze výpočty

69 Při kontrole hmotnosti sušenek bylo zkontrolováno 10 krabic se sušenkami a zjistili se následující hodnoty: 250 g, 247 g, 251 g, 249 g, 252 g, 248 g, 251 g, 250 g, 251 g, 248 g. Vypočítejte průměrnou hmotnost krabice sušenek, směrodatnou odchylku a variační koeficient.

4.3 Tabulka uvádí rozdělení denní dojivosti krav v litrech. Vypočítejte průměrnou dojivost a směrodatnou odchylku.

Dojivost za 1 den	0-2	2-4	4-6	6-8	8-10	10-12
Počet krav	5	8	15	30	25	17



Možná není tak zle...

900 respondentů se vyjádřilo ke vstupu ČR do Evropské unie. 660 se vyslovilo pro, 195 proti a 45 nemělo vyhraněné stanovisko. Vytvořte podklady pro kruhový diagram a výsledek znázorněte graficky.

TAB:

názor (j)	A	N	?
četnost (n_j)	660	195	45
relativní četnost (v_j) [%]	$73, \bar{3}$	21,6	5
výseč (x_j) [$^\circ$]	264	78	18

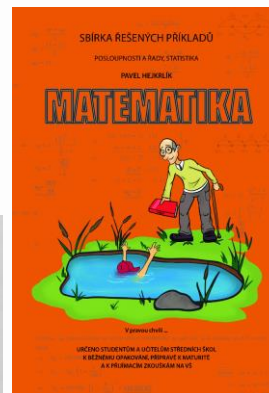
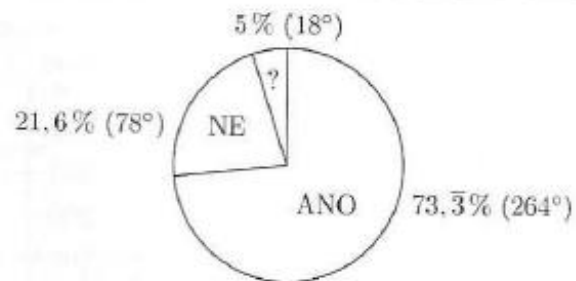
$n = 900$

$$v_j = \frac{n_j}{n} \cdot 100 \Rightarrow$$

$$x_j = \frac{n_j}{n} \cdot 360^\circ \Rightarrow$$

Kruhový diagram:

Vstup do EU





Možná není tak zle...

2.1.7: Statistici zaznamenali do tabulky údaje o rozložení žáků na SOU, gymnáziích a SOŠ v letech 1982/83 a 1983/84. Určete rozložení relativních četností v daných letech.

TAB:

období	1982/83	1983/84
SOU	421 029	423 510
gymnázia	150 630	144 948
SOŠ	310 856	290 038

⇒ TAB 1:
(1982/83)

$v_j = \frac{x_j}{n} \cdot 100 \Rightarrow$

škola (j)	SOU	gymnázia	SOŠ
četnost (x_j)	421 029	150 630	310 856
relativní četnost (x_j) [%]	47,7	17,1	35,2

$n = 882\,515$

⇒ TAB 2:
(1983/84)

$v_j = \frac{x_j}{n} \cdot 100 \Rightarrow$

škola (j)	SOU	gymnázia	SOŠ
četnost (x_j)	423 510	144 984	290 038
relativní četnost (x_j) [%]	49,3	16,9	33,8

$n = 858\,532$

No a co?



Možná není tak zle...

2.1.9: Celkový počet rozvodů v r. 2006 byl 31 415. Tabulka zobrazuje počet rozvodů podle délky trvání manželství. Doplňte ji a nakreslete spojnicový, sloupcový a kruhový diagram četností.

délka trvání manželství (x_j) [roky]	$\langle 0; 1 \rangle$	$\langle 2; 5 \rangle$	$\langle 6; 9 \rangle$	$\langle 10; 14 \rangle$	$\langle 15; 19 \rangle$	$\langle 20; 70 \rangle$	$n = 31\,415$
počet rozvodů (n_j) [ks]	1 280	n_2	n_3	n_4	n_5	7 331	
míra rozvodovosti (v_j) [%]	v_1	v_2	17,8	18,1	17,4	n_6	

$$\Rightarrow v_1 = \frac{n_1}{n} = \frac{1\,280}{31\,415} \approx 0,041 \Rightarrow \underline{\underline{4,1\%}}$$

$$v_6 = \frac{n_6}{n} = \frac{7\,331}{31\,415} \approx 0,233 \Rightarrow \underline{\underline{23,3\%}}$$

$$n_3 = n \cdot v_3 = 31\,415 \cdot 0,178 \approx \underline{\underline{5\,592}}$$

$$n_4 = n \cdot v_4 = 31\,415 \cdot 0,181 \approx \underline{\underline{5\,686}}$$

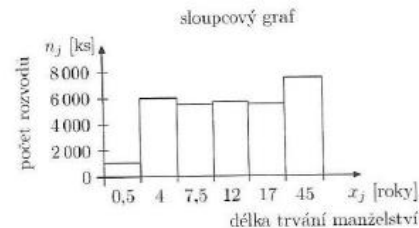
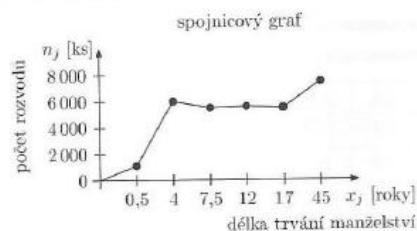
$$n_5 = n \cdot v_5 = 31\,415 \cdot 0,174 \approx \underline{\underline{5\,466}}$$

$$n_2 = n - (n_1 + n_3 + n_4 + n_5 + n_6) = 31\,415 - 25\,355 = \underline{\underline{6\,060}}$$

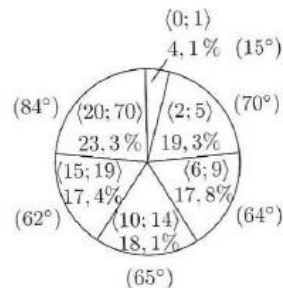
$$v_2 = 100 - (v_1 + v_3 + v_4 + v_5 + v_6) = 100 - 80,7 = \underline{\underline{19,3\%}}$$

$\Rightarrow$ TAB:

délka trvání manželství (x_j) [roky]	$\langle 0; 1 \rangle$	$\langle 2; 5 \rangle$	$\langle 6; 9 \rangle$	$\langle 10; 14 \rangle$	$\langle 15; 19 \rangle$	$\langle 20; 70 \rangle$	$n = 31\,415$
počet rozvodů (n_j) [ks]	1 280	6 060	5 592	5 686	5 466	7 331	
míra rozvodovosti (v_j) [%]	4,1	19,3	17,8	18,1	17,4	23,3	



kruhový graf
délka manželství





Možná není tak zle...

2.1.53: Učitel zapsal výsledky testu 100 studentů do tabulky. Určete průměrnou známku testu, modus a medián.

TAB:

známka (x_j)	1	2	3	4	5
počet studentů (n_j)	15	35	30	15	5

$n = 100$

$\uparrow \quad \uparrow$
 $x_{50} \quad x_{51}$

$$\Rightarrow \text{a) } \bar{x} = \frac{15 \cdot 1 + 35 \cdot 2 + 30 \cdot 3 + 15 \cdot 4 + 5 \cdot 5}{100} = \frac{260}{100} = \underline{\underline{2,6}}$$

$$\text{b) } \text{Mod}(x) = \underline{\underline{2}}$$

$$\text{c) } \text{Med}(x) = \frac{x_{50} + x_{51}}{2} = \frac{2 + 3}{2} = \frac{5}{2} = \underline{\underline{2,5}}$$

No a co?





Možná není tak zle...

2.1.44: Kontrolor výroby zjišťoval vlhkost mouky v procentech, přičemž bylo provedeno postupně celkem 9 měření s těmito výsledky: 13,2; 12,3; 12,3; 13,4; 12,6; 16,8; 13,2; 12,5; 12,7. Určete aritmetický průměr, modus a medián. Určete, který výsledek kontrolu charakterizuje nejvíce objektivně.

TAB:

vlhkost (x_j) [%]	12,3	12,5	12,6	12,7	13,2	13,4	16,8
četnost měření (n_j)	2	1	1	1	2	1	1

$n = 9$

↑
 x_5

$$\Rightarrow \text{a) } \bar{x} = \frac{2 \cdot 12,3 + 1 \cdot 12,5 + 1 \cdot 12,6 + 1 \cdot 12,7 + 2 \cdot 13,2 + 1 \cdot 13,4 + 1 \cdot 16,8}{9} = \underline{\underline{13,2}}$$

$$\text{b) } \text{Mod}(x) - \text{dvojnásobný} \left\{ \begin{array}{l} \underline{\underline{12,3}} \\ \underline{\underline{13,2}} \end{array} \right.$$

$$\text{c) } \text{Med}(x) = x_5 = \underline{\underline{12,7}}$$

Závěr: Nejvíce objektivně charakterizuje kontrolu Med(x), průměr je ovlivněn vysokou hodnotou x_9 a $\text{Mod}(x)$ je nejednoznačný.

Možná není tak zle...



2.1.55: Dva závodníci ve skoku dalekém měli tyto série výsledků v centimetrech:

1. závodník: 724, 733, 728, 710, 712, 726.

2. závodník: 741, 721, 728, 720, 706, 716.

Rozhodněte podle variačního rozpětí, který závodník podal vyrovnanější výkon.

$$R_1 = x_{\max} - x_{\min} = 733 - 710 = 23 \text{ cm}$$

$$R_2 = x_{\max} - x_{\min} = 741 - 706 = 35 \text{ cm}$$

$\Rightarrow$ Vyrovnanější výkon podal 1. závodník.

!! Nevhodná charakteristika !!





Možná není tak zle...

Závěr: $a = (2, \dots)$

2.1.59: Statistické sledovali ve dvou krajích A, B výkup železa ve sběrných surovinách za rok 2010. Porovnejte variabilitu výkupu v obou krajích.

TAB A:

železo (x_j) [$\cdot 10^3$ t]	20	50	60	65	70
počet sběrů (n_j)	3	5	10	2	5

$n = 25$

TAB B:

železo (x_j) [$\cdot 10^3$ t]	15	20	30	45	50	60
počet sběrů (n_j)	8	2	3	5	1	1

$n = 20$

$$\Rightarrow \text{a) } \bar{x}_A = \frac{1390}{25} = 55,6 [\cdot 10^3] \text{ t}$$

$$\bar{x}_B = \frac{585}{20} = 29,25 [\cdot 10^3] \text{ t}$$

$$\Rightarrow \text{b) } S_{xA} = \sqrt{\frac{1}{25} \cdot (3 \cdot 20^2 + 5 \cdot 50^2 + 10 \cdot 60^2 + 2 \cdot 65^2 + 5 \cdot 70^2) - 55,6^2} = \sqrt{214,64} \approx 14,65 \text{ t}$$

$$S_{xB} = \sqrt{\frac{1}{20} \cdot (8 \cdot 15^2 + 2 \cdot 20^2 + 3 \cdot 30^2 + 5 \cdot 45^2 + 50^2 + 60^2) - 29,25^2} = \sqrt{220,69} \approx 14,86 \text{ t}$$

$$\Rightarrow \text{c) } v_{xA} = \frac{14,65}{55,6} \approx 0,263 \%$$

$$v_{xB} = \frac{14,86}{29,25} \approx 0,508 \%$$

$$\Rightarrow \frac{v_{xB}}{v_{xA}} \approx \frac{0,508}{0,263} \approx \underline{\underline{1,93}}$$

Závěr: Variabilita výkupu železa je ve městě B 1,93 krát větší než v kraji A.





Možná není tak zle...

2.1.60: Sledovaný statistický znak – příjem zaměstnanců v roce 1990 je vyjádřen ve dvou firmách v různých jednotkách (měsíční a hodinová mzda). Porovnejte variabilitu mezd v těchto firmách.

TAB 1:

měsíční mzda (x_j) [Kč]	2 000	2 500	3 000	3 500	4 000
počet pracovníků (n_j)	20	30	30	15	5

$n = 100$

TAB 2:

měsíční mzda (x_j) [Kč]	10	15	20	25	30
počet pracovníků (n_j)	30	30	15	10	5

$n = 90$

$$\Rightarrow \text{a) } \bar{x} = \frac{227\,500}{100} = 2\,275 \text{ Kč}$$

$$\bar{y} = \frac{1\,450}{90} = 16,1 \text{ Kč}$$

$$\Rightarrow \text{b) } S_x = \sqrt{\frac{1}{100} \cdot (20 \cdot 2\,000^2 + 30 \cdot 2\,500^2 + 30 \cdot 3\,000^2 + 15 \cdot 3\,500^2 + 5 \cdot 4\,000^2) - 2\,275^2} \doteq \sqrt{311\,877,57} \doteq 558,46 \text{ Kč}$$

$$S_y = \sqrt{\frac{1}{90} \cdot (30 \cdot 10^2 + 30 \cdot 15^2 + 15 \cdot 20^2 + 10 \cdot 25^2 + 5 \cdot 30^2) - 16,1^2} \doteq \sqrt{34,876} \doteq 5,9056 \text{ Kč}$$

$$\Rightarrow \text{c) } v_x = \frac{558,46}{2\,275} \doteq 0,201 \%$$

$$v_y = \frac{5,9056}{16,1} \doteq 0,367 \%$$

$$\Rightarrow \frac{v_y}{v_x} \doteq \frac{0,367}{0,201} \doteq \underline{\underline{1,83}}$$

Závěr: Ve druhé firmě je asi 1,83 krát větší variabilita mezd než v první firmě.





Možná není tak zle...



09

V roce 2013 se v ČR konaly prezidentské volby. Uvádíme počty voličů ve vybraných krajích v 1. kole voleb. V Praze bylo celkem 935 085 voličů, z nichž se voleb zúčastnilo 608 749. V Královéhradeckém kraji bylo celkově 449 235 právoplatných voličů a voleb se zúčastnilo 290 443. Kraj Vysočina měl v té době 415 554 voličů a k volbám jich přišlo 277 541. **Vhodným způsobem vyhodnoťte a porovnejte účast ve volbách v těchto krajích. (Zdroj: ČSÚ)**

Kraj	Celkový počet voličů	Počet zúčastněných	Relativní četnost	Relativní četnost (v procentech)
Praha	935 085	608 749	$\frac{608\,749}{935\,085} \doteq 0,651\,0$	65,10 %
Královéhradecký kraj	449 235	290 443	$\frac{290\,443}{449\,235} \doteq 0,646\,5$	64,65 %
Kraj Vysočina	415 554	277 541	$\frac{277\,541}{415\,554} \doteq 0,667\,9$	66,79 %

Nejvyšší účast ve volbách ve vybraných krajích byla v Kraji Vysočina.



Možná není tak zle...

11

V grafu jsou zaznačeny počty pacientů ošetřených na lékařské službě první pomoci v průběhu jednoho týdne. Na základě grafu doplňte následující věty.



Den v týdnu	Po	Út	St	Čt	Pá	So	Ne	Celkem
Počet ošetřených	7	13	9	11	24	42	31	137
Počet ošetřených	64					73		137

c) $\frac{73}{137} \div 0,5328$ d) $\frac{73}{64} \div 1,1406$ f) $\frac{9}{13} - 1 \div -0,3077$

- a) Počet ošetřených na pohotovosti byl nejvyšší v **sobotu** a nejnižší v **pondělí**.
- b) Ve všední dny bylo na pohotovosti ošetřeno celkem **64** lidí a o víkendu celkem **73** lidí.
- c) O víkendu bylo na pohotovosti ošetřeno **53,28** % lidí z celého týdne.
- d) O víkendu bylo na pohotovosti ošetřeno o **14,06** % lidí více než ve všedních dnech.
- e) Největší absolutní nárůst počtu ošetřených na LSPP oproti předchozímu dni byl v **sobotu**.
- f) Největší procentuální pokles počtu ošetřených na LSPP oproti předchozímu dni byl **30,77** %.

Možná není tak zle...



16

Na úřadu práce byla v jednom dni zaregistrována skupina osob bez zaměstnání tohoto věku: 37, 32, 21, 33, 22, 53, 28, 23, 56, 25, 49, 26, 28, 29, 30, 31, 35, 30, 36, 32, 38, 27, 38, 21, 39, 24, 50, 43, 23, 44, 25, 46, 36, 22, 49, 35, 22, 51, 58, 26. Proveďte statistické vyhodnocení věku této skupiny nezaměstnaných.

a) Stanovte rozsah souboru n .

$$n = 40$$

b) Seřadte vzestupně věk nezaměstnaných: 21, 21, 22, 22, 22, 23, 23, 24, 25, 25, 26, 26, 27, 28, 28, 29,

30, 30, 31, 32, 32, 33, 35, 35, 36, 36, 37, 38, 38, 39, 43, 44, 46, 49, 49, 50, 51, 53, 56, 58

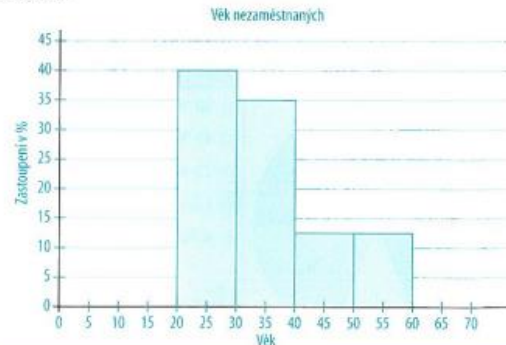
c) Proveďte intervalové třídění, tj. zvolte vhodnou délku intervalů a jejich počet.

Volíme 4 intervaly s délkou 10 jednotek (let).

$\langle 20; 30 \rangle$, $\langle 30; 40 \rangle$, $\langle 40; 50 \rangle$, $\langle 50; 60 \rangle$

d) Sestavte tabulku rozdělení četností. V tabulce u každého intervalu запиšte jeho střed, absolutní četnosti a relativní četnosti v desetinných číslech i v procentech.

Věk	20-30	30-40	40-50	50-60
Střed intervalu	25	35	45	55
Absolutní četnost	16	14	5	5
Relativní četnost	0,400	0,350	0,125	0,125
Relativní četnost [%]	40 %	35 %	12,5 %	12,5 %



e) Zakreslete histogram četností.

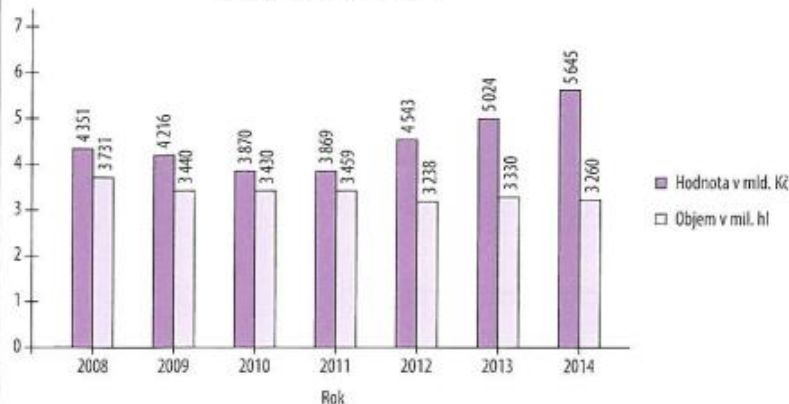


Možná není tak zle...

23

V následujících úlohách vycházejte z uvedeného grafu (Zdroj: deník MF Dnes, 12. 02. 2015). U každé úlohy vyberte z pěti nabízených možností právě jednu tak, aby vzniklé tvrzení bylo pravdivé.

Vývoz piva z ČR v letech 2008–2014



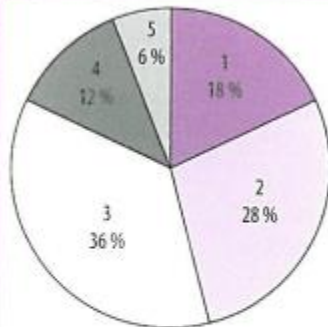
- A) V letech 2008–2014 objem exportovaného piva:
- a) stále rostl.
 - b) stále klesal.
 - c) rostl pouze v posledních čtyřech letech.
 - d) klesal pouze v posledních čtyřech letech.
 - ☒ e) se stále pohyboval mezi 3 až 4 miliony hektolitřů.
- B) V letech 2008–2014 hodnota vyvezeného piva vyjádřená v Kč:
- a) stále rostla.
 - b) stále klesala.
 - c) rostla pouze v posledních čtyřech letech.
 - ☒ d) rostla pouze v posledních třech letech.
 - e) přesáhla hodnotu 6 mld. Kč v jednom ze sledovaných roků.
- C) Z grafu lze usoudit, že:
- a) cena vyváženého piva přepočtená na 1 hl je stále stejná.
 - b) cena vyváženého piva přepočtená na 1 hl stále roste.
 - c) cena vyváženého piva přepočtená na 1 hl stále klesá.
 - ☒ d) v posledních třech letech objem exportovaného piva sice lehce klesl oproti předchozím letům, ale vyvážel se za vyšší ceny.
 - e) v roce 2015 cena exportovaného piva určitě přesáhne hranici 6 mld. Kč.
- D) Cena půllitru vyváženého piva:
- a) byla nejnižší v roce 2014.
 - b) byla nejvyšší v roce 2011.
 - c) byla v roce 2008 minimálně 7 Kč.
 - ☒ d) byla v roce 2014 v rozmezí 8,50–9,00 Kč.
 - e) nelze z grafu určit.

Možná není tak zle...



06

Pololetní hodnocení studentů třídy 3. A z matematiky znázorňuje kruhový diagram. Určete průměr známek z matematiky v této třídě.



$$\bar{x} = \frac{18 \cdot 1 + 28 \cdot 2 + 36 \cdot 3 + 12 \cdot 4 + 6 \cdot 5}{100}$$

$$\bar{x} = 2,6$$

Průměr známek z matematiky ve třídě 3. A je 2,6.





Možná není tak zle...

13

Na chmelové brigádě byli pracovníci rozděleni do dvou skupin. V první skupině si každý za brigádu vydělal průměrně 14 000 Kč, ve druhé skupině 8 500 Kč. Celkový průměrný výdělek byl 11 580 Kč. Kolik procent brigádníků bylo zařazeno do druhé skupiny?

celkový počet brigádníků n

počet brigádníků ve druhé skupině n_2

počet brigádníků v první skupině $n - n_2$

$$\frac{(n - n_2) \cdot 14\,000 + n_2 \cdot 8\,500}{n} = 11\,580$$

$$-14\,000n_2 + 8\,500n_2 = 11\,580n - 14\,000n$$

$$2\,420n = 5\,500n_2$$

$$\frac{n_2}{n} = \frac{2\,420}{5\,500}$$

$$\frac{n_2}{n} = 0,44$$

Ve druhé skupině bylo 44 % ze všech brigádníků.



Možná není tak zle...



15

Průměrná výška stromků v lesní školce je 120 cm, směrodatná odchylka je 18 cm. Vyberte pravdivé tvrzení.

- a) Nejvýše tři čtvrtiny stromků v lesní školce mají výšku, která se pohybuje v intervalu (102; 138) cm.
- b) Alespoň tři čtvrtiny stromků v lesní školce mají výšku, která se pohybuje v intervalu (102; 138) cm.
- c) Nejvýše dvě třetiny stromků v lesní školce mají výšku, která se pohybuje v intervalu (84; 156) cm.
- ☒ d) Alespoň dvě třetiny stromků v lesní školce mají výšku, která se pohybuje v intervalu (102; 138) cm.
- e) Nejvýše dvě třetiny stromků v lesní školce mají výšku, která se pohybuje v intervalu (102; 138) cm.

Za předpokladu, že...





Možná není tak zle...

16

V laboratorním cvičení z chemie měli studenti stanovit procentuální koncentraci roztoku manganistanu draselného. Pracovali ve dvou pětičlenných skupinách, přičemž každá z nich dostala roztok s jinou koncentrací. Každý člen skupiny provedl jedno měření a zaznamenal získanou hodnotu do tabulky. Pomocí variačního koeficientu charakterizujte přesnost měření obou skupin. Která skupina studentů pracovala přesněji?

	Získané hodnoty koncentrace roztoku				
	1.	2.	3.	4.	5.
1. skupina	12,15 %	11,80 %	11,70 %	12,35 %	13,00 %
2. skupina	5,25 %	4,95 %	4,70 %	4,80 %	5,20 %

$$\bar{x}_1 = \frac{12,15 + 11,80 + 11,70 + 12,35 + 13,00}{5}$$

$$\bar{x}_1 = 12,2$$

$$s_1^2 = \overline{x^2} - \bar{x}_1^2$$

$$s_1^2 = \frac{12,15^2 + 11,80^2 + 11,70^2 + 12,35^2 + 13,00^2}{5} - 12,2^2$$

$$s_1^2 = 0,215$$

$$s_1 = \sqrt{s_1^2}$$

$$s_1 = \sqrt{0,215}$$

$$s_1 \approx 0,464$$

$$v_1 = \frac{s_1}{\bar{x}}$$

$$v_1 = \frac{0,464}{12,2} = 0,038$$

$$v_1 = 3,8 \%$$

$$\bar{x}_2 = \frac{5,25 + 4,95 + 4,70 + 4,80 + 5,20}{5}$$

$$\bar{x}_2 = 4,98$$

$$s_2^2 = \overline{x^2} - \bar{x}_2^2$$

$$s_2^2 = \frac{5,25^2 + 4,95^2 + 4,70^2 + 4,80^2 + 5,20^2}{5} - 4,98^2$$

$$s_2^2 = 0,0466$$

$$s_2 = \sqrt{s_2^2}$$

$$s_2 = \sqrt{0,0466}$$

$$s_2 \approx 0,216$$

$$v_2 = \frac{s_2}{\bar{x}}$$

$$v_2 = \frac{0,216}{4,98} = 0,043$$

$$v_2 = 4,3 \%$$

Přesněji pracovala první skupina.



Možná není tak zle...

21

Pět spolužáků šlo spolu na oběd do školní jídelny. Toho dne byly k obědu ovocné knedlíky. V tabulce je uvedeno, kolik jich každý snědl. Po obědě šli kluci opět do vyučování. V hodině českého jazyka psali diktát. Jejich počet chyb v diktátu je rovněž uveden v tabulce. Určete korelační koeficient mezi těmito dvěma znaky a zhodnoťte, zda může počet knedlíků ovlivnit počet chyb v diktátu z českého jazyka.

	Počet snědených knedlíků	Počet chyb v diktátu	$x_j - \bar{x}$	$(x_j - \bar{x})^2$	$y_j - \bar{y}$	$(y_j - \bar{y})^2$	$(x_j - \bar{x}) \cdot (y_j - \bar{y})$
Pavel	5	1	0,4	0,16	-2,6	6,76	-1,04
Jan	5	3	0,4	0,16	-0,6	0,36	-0,24
David	4	2	-0,6	0,36	-1,6	2,56	0,96
Martin	3	5	-1,6	2,56	1,4	1,96	-2,24
Petr	6	7	1,4	1,96	3,4	11,56	4,76
Součet	23	18	0	5,20	0	23,20	2,20
Průměr	4,6	3,6					

$$r = \frac{\sum_{j=1}^5 (x_j - \bar{x}) \cdot (y_j - \bar{y})}{\sqrt{\sum_{j=1}^5 (x_j - \bar{x})^2 \cdot \sum_{j=1}^5 (y_j - \bar{y})^2}} = \frac{2,20}{\sqrt{5,20 \cdot 23,20}}$$
$$r = 0,2$$

Počet snědených knedlíků zřejmě neovlivňuje počet chyb v diktátu.

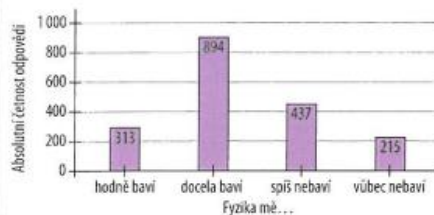


Možná není tak zle...

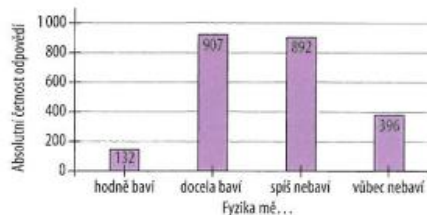
22

V letech 2006–2008 proběhl v ČR výzkum, který byl zaměřen na oblíbenost fyziky mezi žáky druhého stupně základních škol, žáky středních škol a gymnázií (včetně víceletých). Výsledky výzkumu jsou znázorněny graficky. Vypočítejte a porovnejte variační koeficienty pro jednotlivé soubory. (Návod: Jednotlivým odpovědím přiřadte vhodně zvolenou bodovou hodnotu.)

Žáci ZŠ a NG



Žáci SOŠ a VG



Volíme: hodně baví = 4, docela baví = 3, spíš nebaví = 2, vůbec nebaví = 1

x_j	n_j	$n_j \cdot x_j$	$n_j \cdot x_j^2$
4	313	1 252	5 008
3	894	2 682	8 046
2	437	874	1 748
1	215	215	215
Součet	1 859	5 023	15 017

$$\bar{x} = 2,7$$

$$s_x^2 = \frac{1}{n} \sum_{j=1}^4 n_j \cdot x_j^2 - \bar{x}^2$$

$$s_x^2 = \frac{5\,008 + 8\,046 + 1\,748 + 215}{1\,859} - 2,7^2 \approx 0,79$$

$$s_x \approx 0,9$$

$$v_x = \frac{s_x}{\bar{x}} = \frac{0,9}{2,7} = 0,33$$

$$v_x = 33 \%$$

x_j	n_j	$n_j \cdot x_j$	$n_j \cdot x_j^2$
4	132	528	2 112
3	907	2 721	8 163
2	892	1 784	3 568
1	396	396	396
Součet	2 327	5 429	14 239

$$\bar{x} = 2,3$$

$$s_x^2 = \frac{1}{n} \sum_{j=1}^4 n_j \cdot x_j^2 - \bar{x}^2$$

$$s_x^2 = \frac{2\,112 + 8\,163 + 3\,568 + 396}{2\,327} - 2,3^2 \approx 0,8$$

$$s_x \approx 0,9$$

$$v_x = \frac{s_x}{\bar{x}} = \frac{0,9}{2,3} = 0,39$$

$$v_x = 39 \%$$



Jedna úloha bez počítání

Jistý výzkumník testoval účinnost tří různých typů hnojiv rajčat. Sledoval hmotnost rajčat (v kg) sklizených z jednotlivých rostlin a pro jednotlivá hnojiva (a kontrolní nehnojené rostliny) spočítal následující statistiky:

	Bez hnojiva	Hnojivo 1	Hnojivo 2	Hnojivo 3
Průměr	5,4	7,2	4,8	6,9
Medián	5,7	7,4	4,9	6,6
Směrodatná odchylka	1,13	0,54	0,32	1,27
Variační koeficient	0,21	0,08	0,07	0,18
Počet rostlin	12	14	17	15

Rozhodněte podle výsledků o pravdivosti následujících tvrzení a své rozhodnutí vždy stručně zdůvodněte:

- a) Všechna hnojiva zvyšují výnosy rajčat.
- b) Všechna hnojiva snižují rozdíly mezi jednotlivými rostlinami rajčat z hlediska výnosů.
- c) V případě užití hnojiva 3 se pravděpodobně našly rostliny s extrémně vysokými výnosy.
- d) V případě užití hnojiva 2 se pravděpodobně našly rostliny s extrémně vysokými výnosy.





Jedna úloha s Excelem

Z listu 1 zkopírujte data o objednávkách, které odbavoval Břetislav, a zvlášť data o objednávkách, které odbavoval Emil.

Určete následující hodnoty pro objednávky odbavené Břetislavem a objednávky odbavené Emilem:

	Břetislav	Emil
průměrná hodnota	2986	3232,143
mediánová hodnota	2470	3315
průměrná absolutní odchylka	1646,667	876,4286
relativní odchylka	0,551462	0,27116
rozptyl	4436957	1394474
směrodatná odchylka	2106,409	1180,878
variační koeficient	0,705428	0,365355
rozpětí	6740	4680

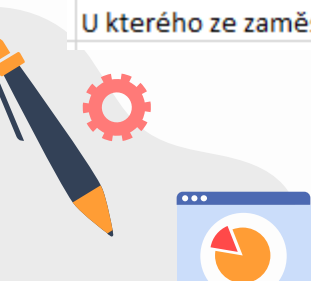
Který ze zaměstnanců odbavoval hodnotněji objednávky?

Emil, je to vidět z průměru i mediánu.

U Břetislava, hodnoty relativní odchylky i variačního koeficientu jsou u něj větší.

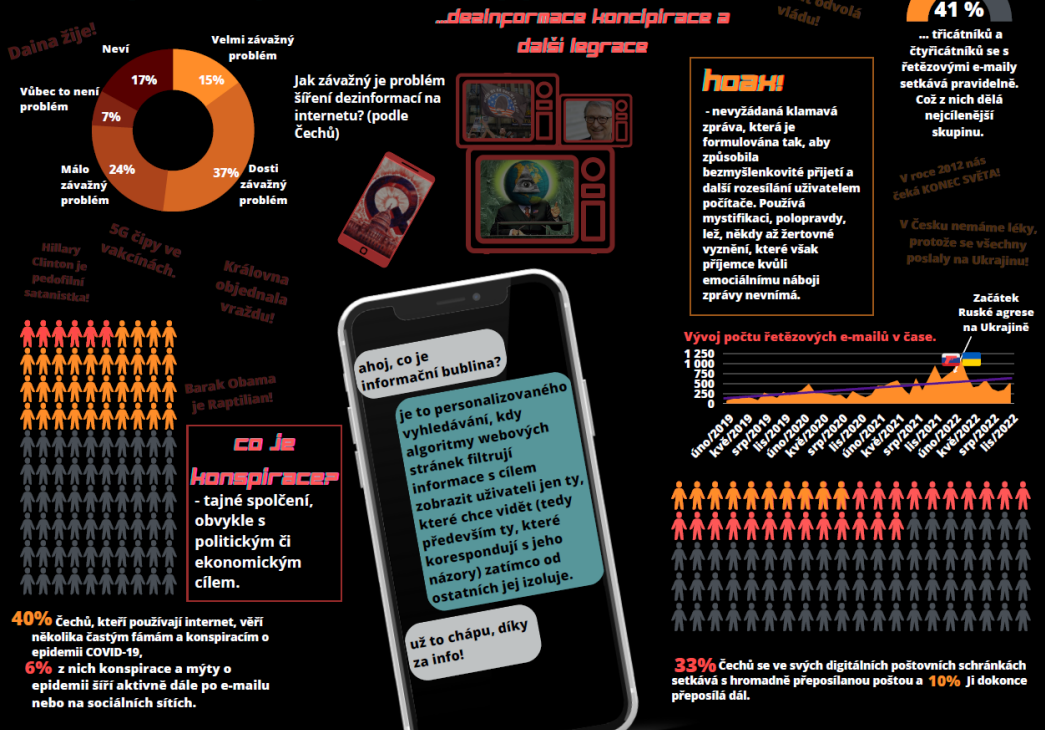
U kterého ze zaměstnanců se více lišily hodnoty objednávek?

Ostatní hodnoty jsou těžko porovnatelné, protože se liší průměry.



A jeden kreativní úkol

KDYŽ SE LEŽ STÁVÁ PRAVDOU...



Vášim úkolem je zpracovat statistický plakát, který splní následující kritéria:

- Využívá, interpretuje a komunikuje statistická data.
- Data mohou být převzatá, nebo je můžete sesbírat sami. V případě zobrazení vlastních dat je obsažena metodika jejich sběru.
- Musí být jednostranný.
- Je logicky a přehledně uspořádaný.
- Je originální po obsahové i grafické stránce. Nemůže být převzatý odjinud.
- Obsahuje citaci zdrojů dat.
- Je z něj jasné, co je jeho cílem.
- Závěry jsou podloženy daty.
- Tabulky a grafy jsou vhodně zvolené vzhledem k datům.
- Grafy obsahují všechny prvky potřebné pro interpretaci (typický legendu, popisky os).
- Tabulky jsou přehledné, mají popsání sloupce a řádky tam, kde je to potřeba.
- Každý graf nebo tabulka přináší novou informaci, kterou nelze získat z jiného prvku plakátu.
- Vypadá atraktivně a úhledně.

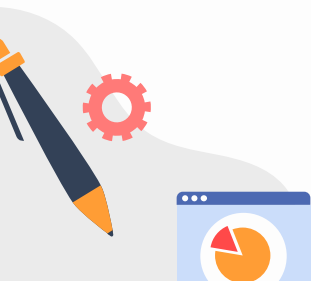
Téma plakátu je libovolné, důležité je si položit dostatečně konkrétní otázku, na kterou budete hledat odpověď, kterou budete plakátem prezentovat. Pracovat můžete sami nebo v nejvýše tříčlenných skupinách. Termín odevzdání plakátu je ve středu 10. ledna, prezentovat je budete na lekci statistiky ve čtvrtek 11. ledna. Na tvorbu plakátu budete mít vyhrazené lekce statistiky 21. 12. 2023 a 4. 1. 2024. Hotový plakát odevzdejte v Google Classroom ve formátu pdf. Stačí, když ho za skupinu odevzdá jeden její člen, do komentářů pak uvede jména zbývajících členů skupiny.

Závěrem



Výpočty popisných statistik nemají být cílem výuky statistiky.

Zařazujme úlohy na interpretaci statistik i úlohy podporující kritické uvažování o datech. Nikdo jiný to za nás na střední škole neudělá.



Děkuji za pozornost.

