Katedra
aplikované matematiky

matematika s radostí

Jak nelhat se statistikou?

Martina Litschmannová
Katedra aplikované matematiky, FEI, VŠB-TU Ostrava

evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

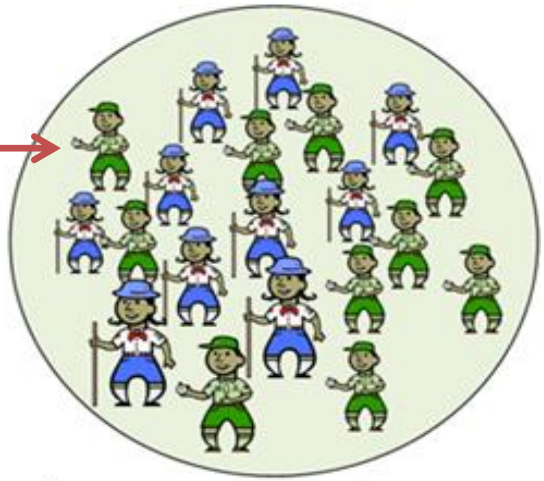
Co je to statistika?

- teoretická disciplína, která se zabývá metodami sběru a analýzy dat

Jak získat data?

úplné šetření

statistická jednotka



POPULACE = ZÁKLADNÍ SOUBOR

statistické znaky – údaje, které u statistických znaků sledujeme (např. váha, výška, IQ, …)

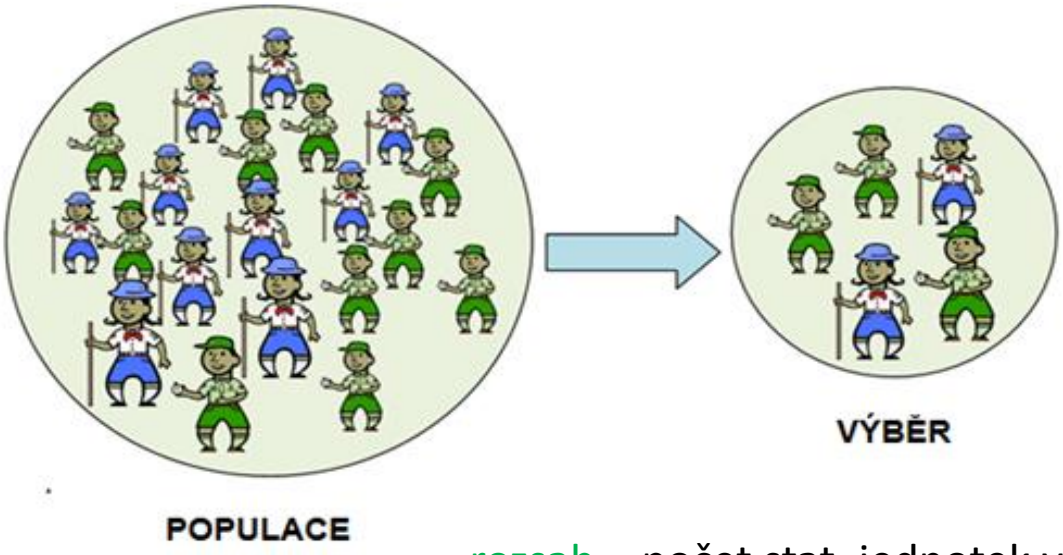
Co je to statistika?

- teoretická disciplína, která se zabývá metodami sběru a analýzy dat

Jak získat data?

úplné šetření

výběrové šetření



REPREZENTATIVNÍ výběr

rozsah – počet stat. jednotek ve výběru (v populaci)

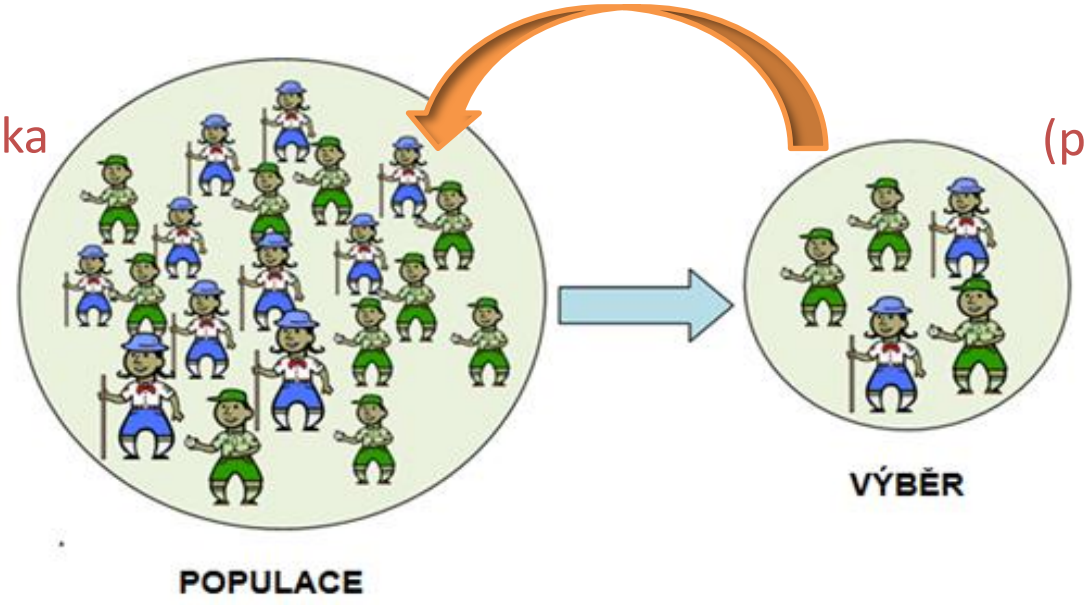
Co je to statistika?

- teoretická disciplína, která se zabývá metodami sběru a analýzy dat

Jak analyzovat data?

Statistická indukce

Exploratorní (popisná) statistika

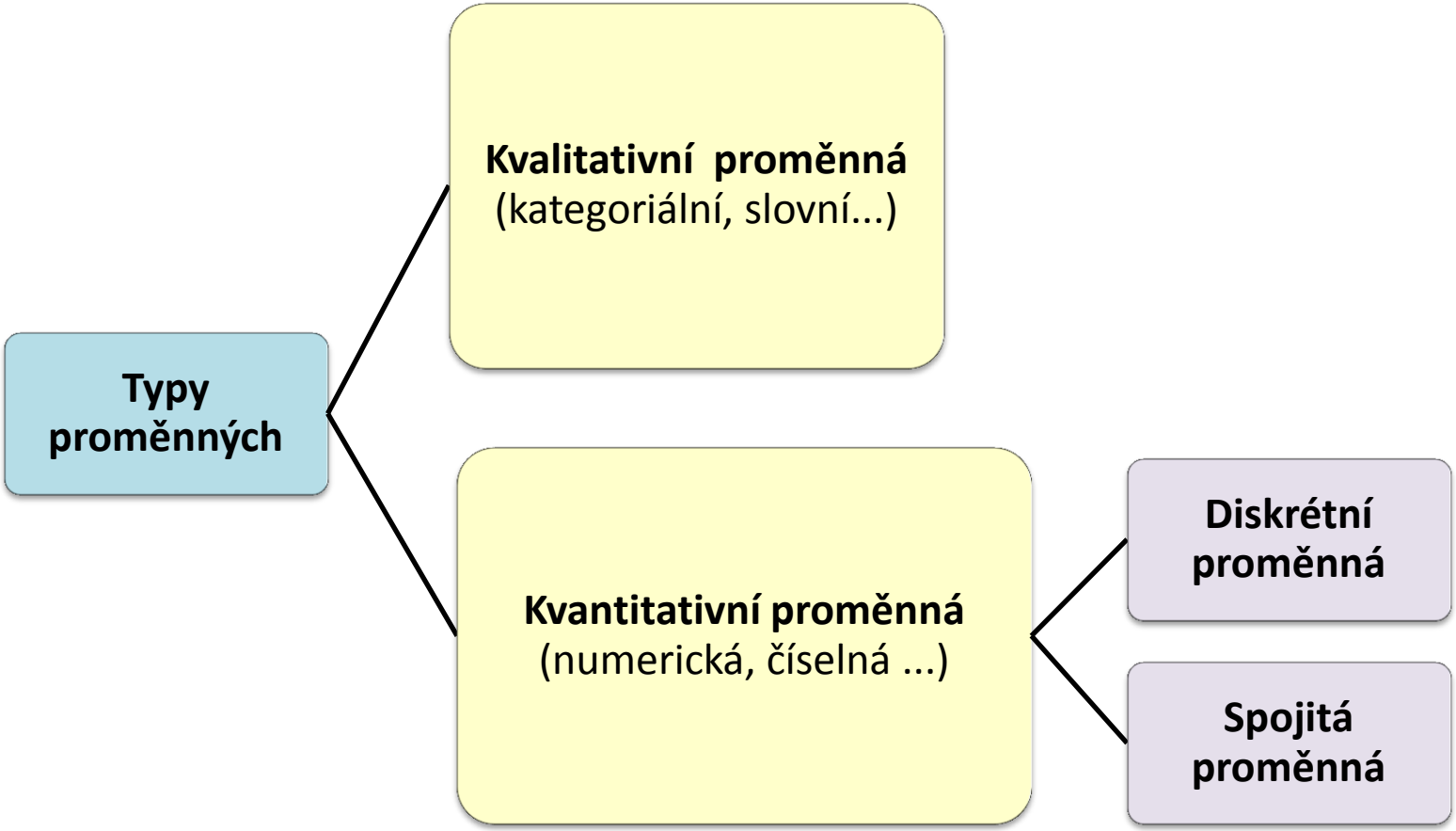


Exploratorní (popisná) statistika

Exploratorní analýza dat

= **Grafická prezentace** a uspořádání dat do názornější formy a jejich **popis několika málo hodnotami**, které by obsahovaly co největší množství informací obsažených v původním souboru.

Typy statistických znaků (proměnných)



EDA pro kategoriální veličinu

Číselné charakteristiky

TABULKA ROZDĚLENÍ ČETNOSTI		
Varianty x_i	**Absolutní četnosti** n_i	**Relativní četnosti** p_i
x_1	n_1	$p_1=n_1/n$
x_2	n_2	$p_2=n_2/n$
x_k	n_k	$p_k=n_k/n$
Celkem:	$n_1+n_2+\ldots+n_k=n$	1

+ Modus (název nejčetnější varianty)

Číselné charakteristiky

TABULKA ROZDĚLENÍ ČETNOSTI		
„Pohlaví“	**Absolutní četnosti**	**Relativní četnosti (%)**
Muž	77	37,37864
Žena	85	41,26214
Dítě	44	21,35922
Celkem:	206	100,00000

1% … 2,06 osob
0,00001% … 0,0000206 osob
0,1% … 0,206 osob

Jak zaokrouhlovat relativní četnost?

Číselné charakteristiky

TABULKA ROZDĚLENÍ ČETNOSTI		
„Pohlaví“	**Absolutní četnosti**	**Relativní četnosti (%)**
Muž	77	37,4
Žena	85	41,3
Dítě	44	21,4
Celkem:	206	100,1

POZOR na zaokrouhlovací chybu!

Číselné charakteristiky

TABULKA ROZDĚLENÍ ČETNOSTI		
„Pohlaví“	**Absolutní četnosti**	**Relativní četnosti (%)**
Muž	77	37,4
Žena	85	41,3
Dítě	44	21,3
Celkem:	206	100,0

Dopočet do 100%!

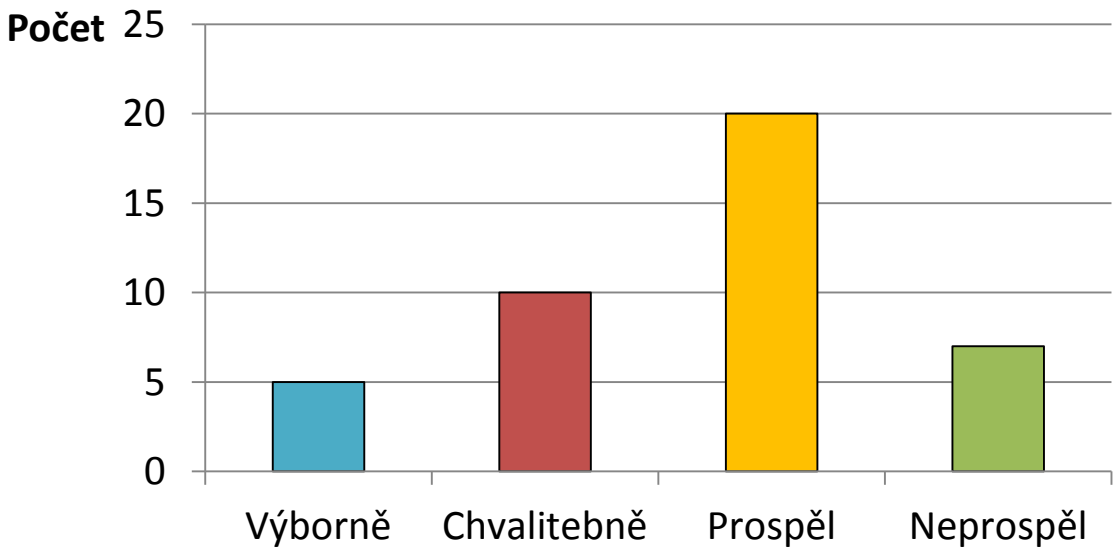
Číselné charakteristiky

TABULKA ROZDĚLENÍ ČETNOSTI		
„Pohlaví“	**Absolutní četnosti**	**Relativní četnosti (%)**
Muž	?	37,4
Žena	?	41,3
Dítě	?	21,3
Celkem:	206	100,0

Relativní četnosti uvádějme vždy pouze jako doplněk absolutních četností, nikoliv samostatně!

Grafické znázornění

A) Sloupcový graf (bar chart)

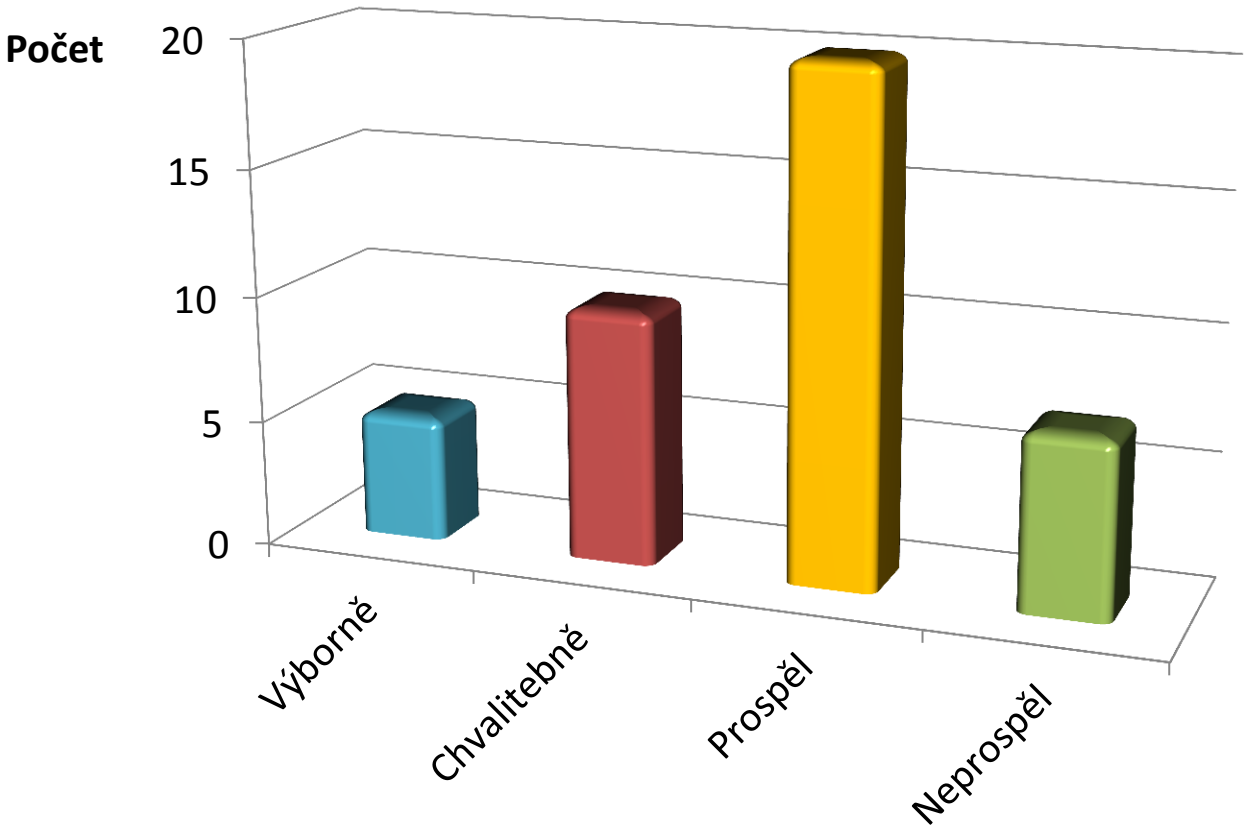


„…můžete vytvořit sloupcový graf a dodat mu zcela nový a přitažlivý vzhled“

http://office.microsoft.com/cs-cz/excel-help/prezentace-dat-ve-sloupcovem-grafu-HA010218663.aspx

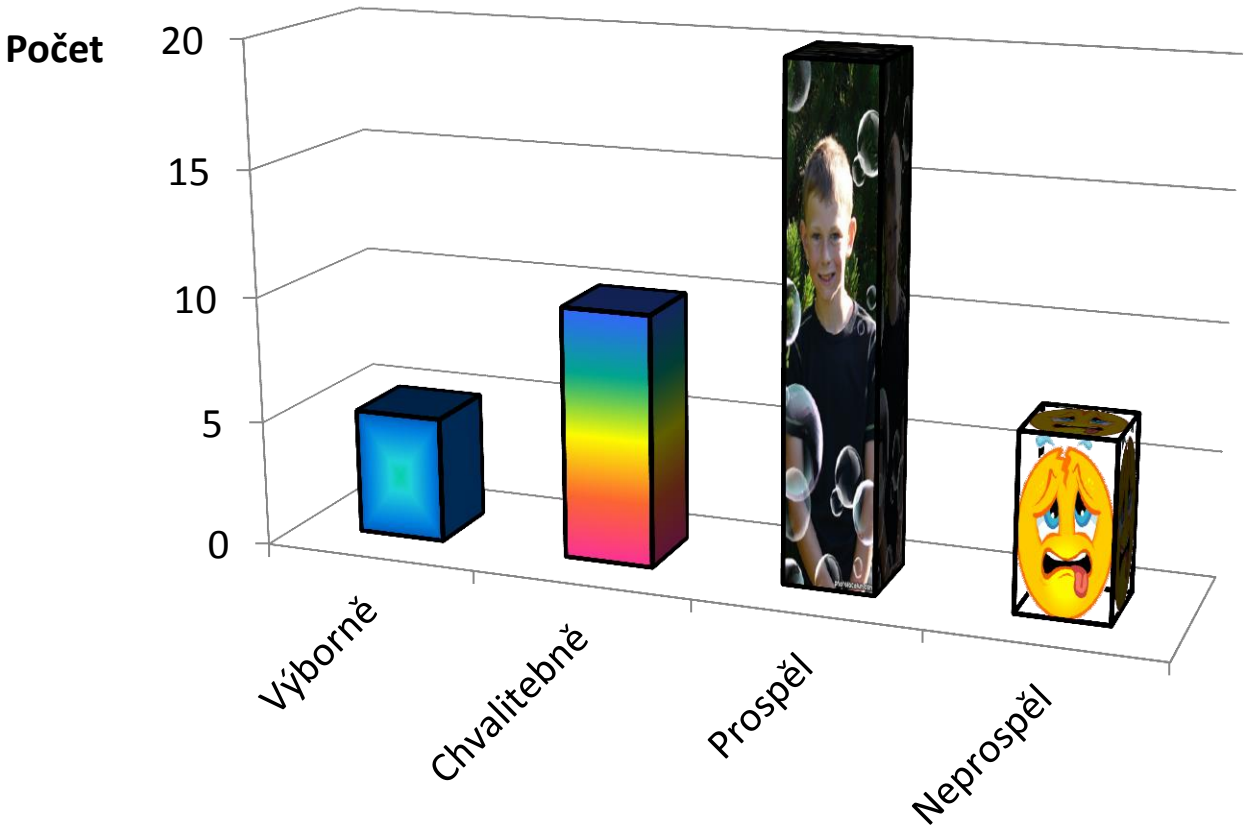
Grafické znázornění

A) Sloupcový graf (bar chart)



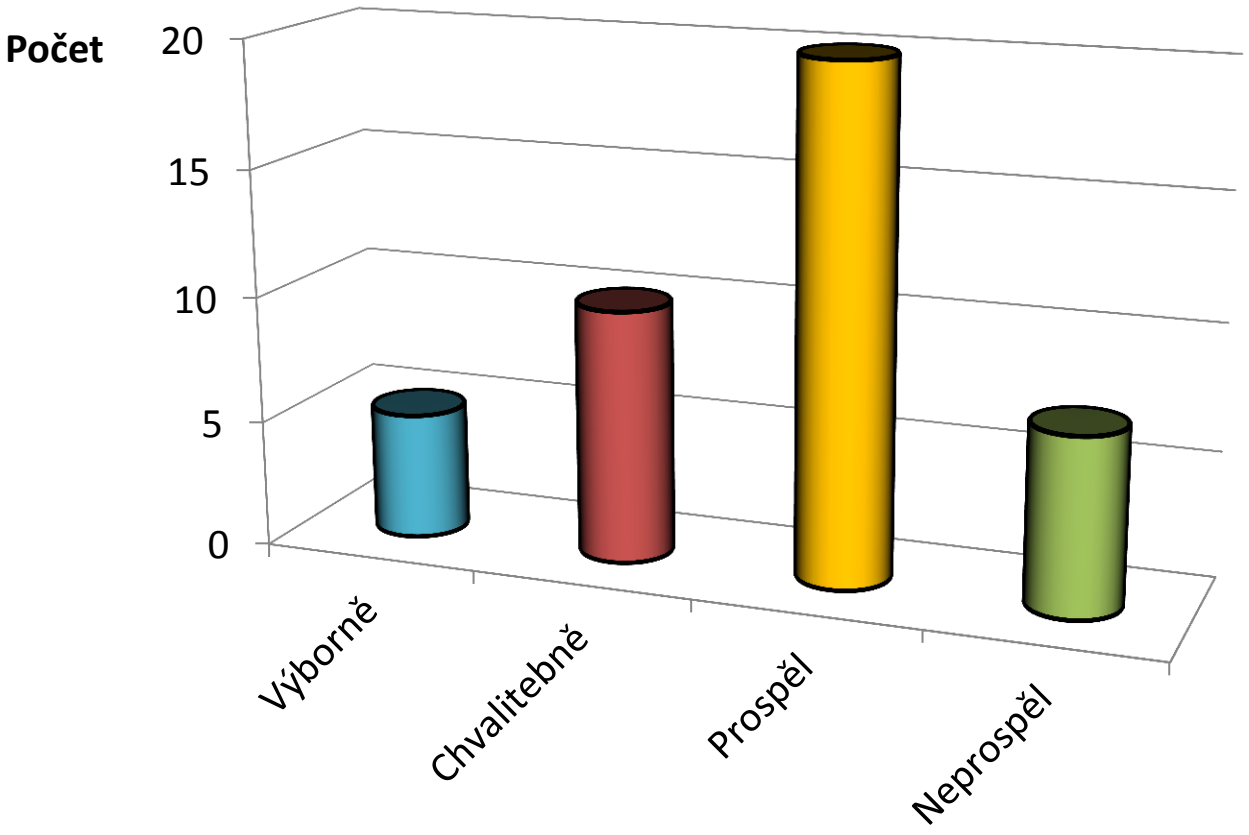
Grafické znázornění

A) Sloupcový graf (bar chart)



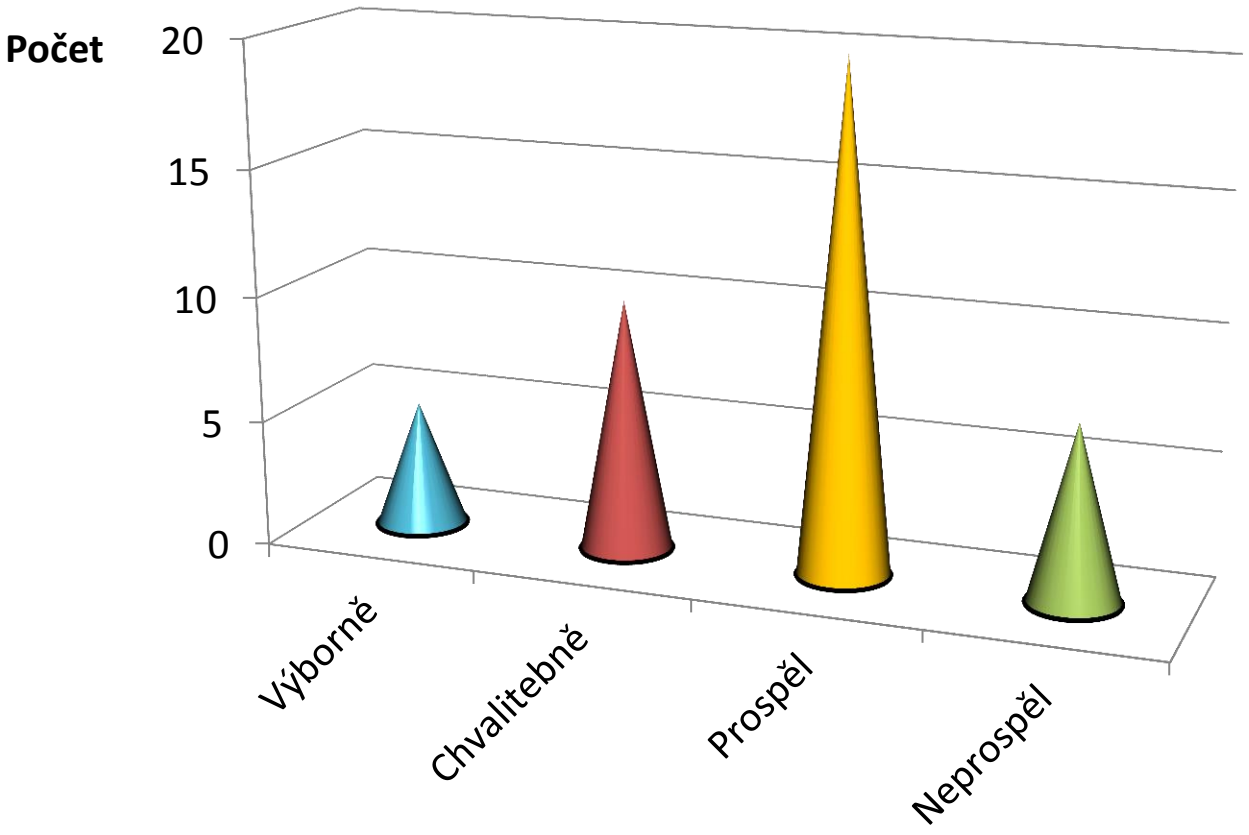
Grafické znázornění

A) Sloupcový graf (bar chart)



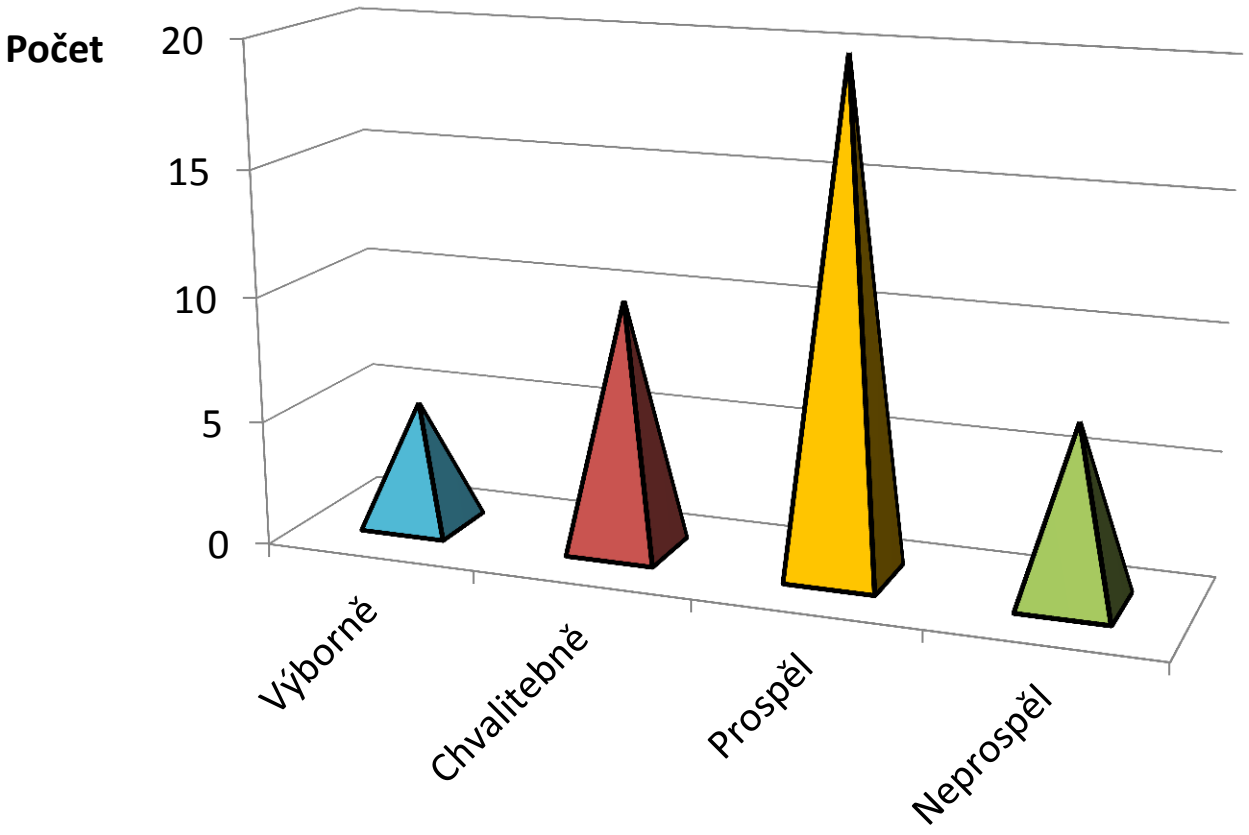
Grafické znázornění

A) Sloupcový graf (bar chart)



Grafické znázornění

A) Sloupcový graf (bar chart)



Grafické znázornění

A) Sloupcový graf (bar chart)

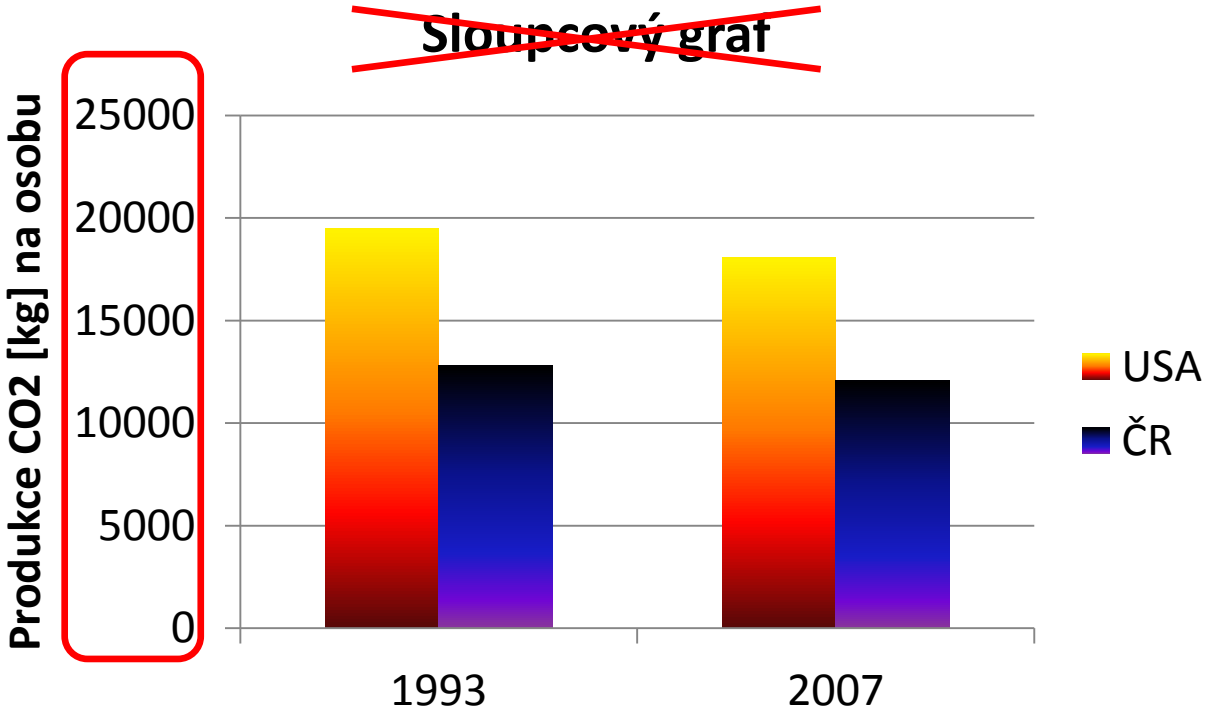
Na co si dát pozor?

- Subjektivně vnímáme plochu (objem), nikoliv výšku jednotlivých „sloupců“.

Grafické znázornění

A) Sloupcový graf (bar chart)

Na co si dát pozor?



zdroj dat:
http://en.wikipedia.org/wiki/List_of_countries_by_carbon_dioxide_emissions_per_capita

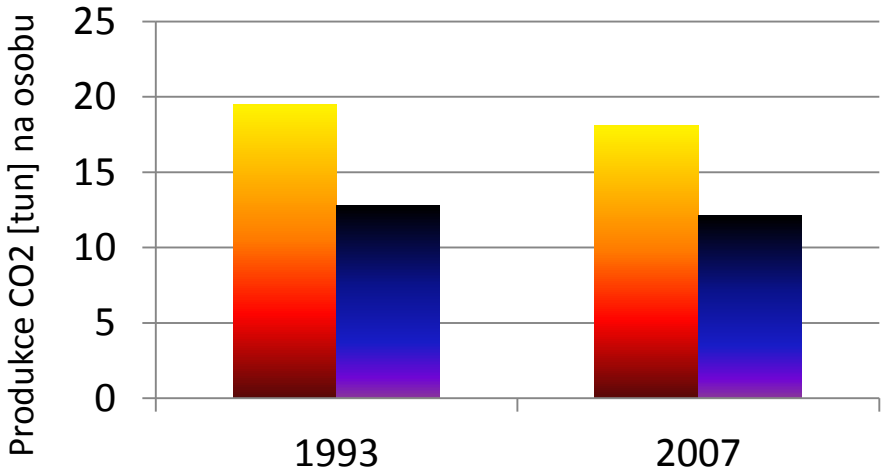
Grafické znázornění

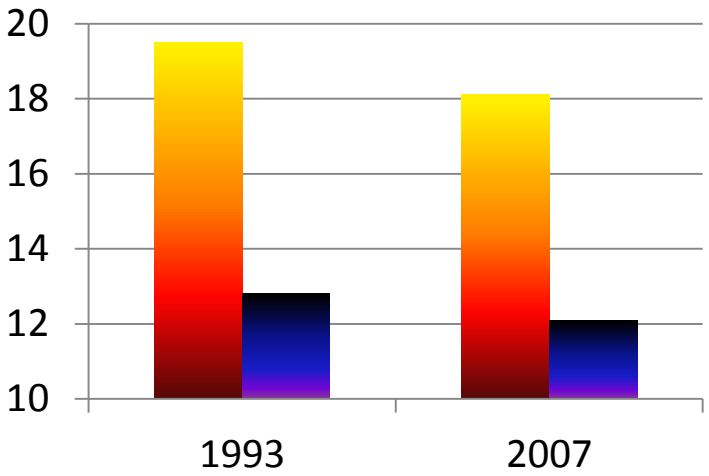
A) Sloupcový graf (bar chart)

Na co si dát pozor?

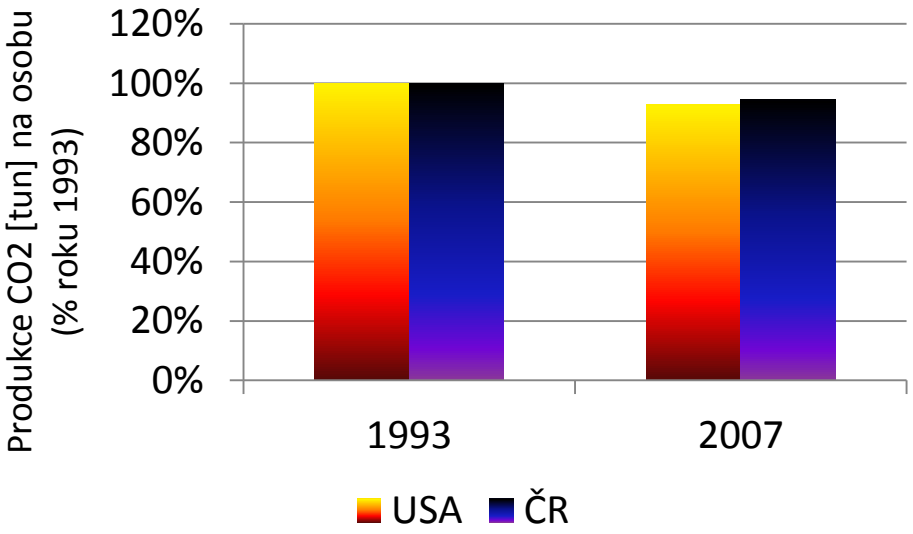
- Subjektivně vnímáme plochu (objem), nikoliv výšku jednotlivých „sloupců“.
- Nadbytečné názvy grafu, legendy, …
- Neefektivní nuly

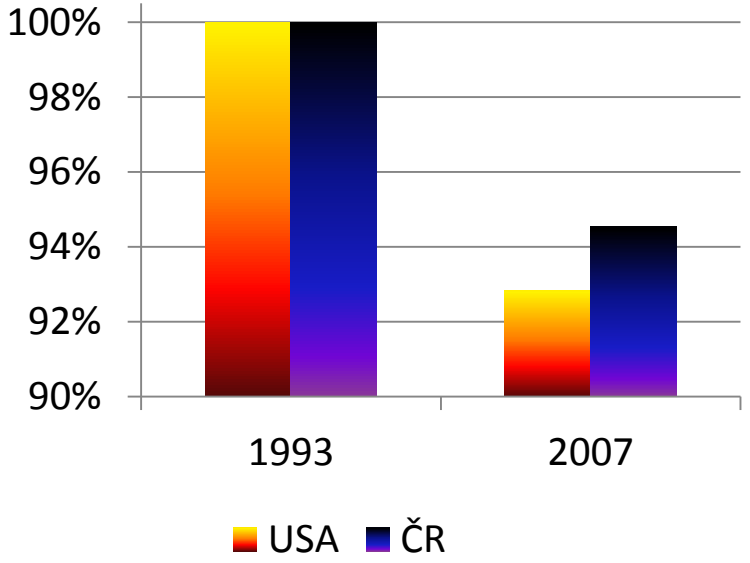
A na co ještě?

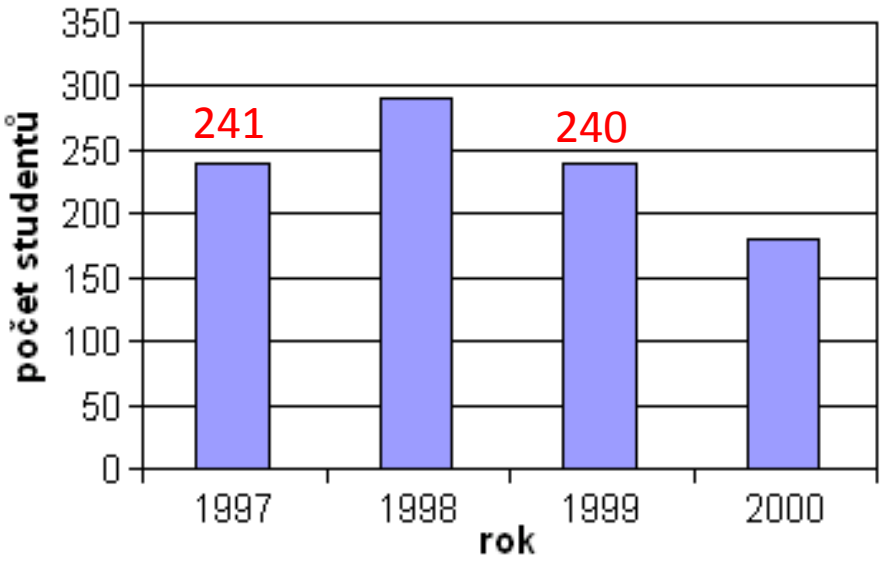




Který z grafů je „správný“?







?

Určete pravdivost tvrzení:
V žádných dvou letech nebyl počet studentů stejný.

Zdroj: Testové příklady určené žákům 9. tříd.

Grafické znázornění

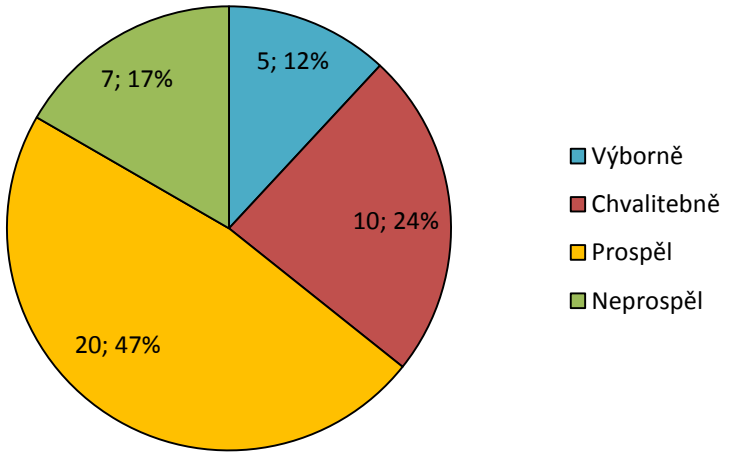
A) Sloupcový graf (bar chart)

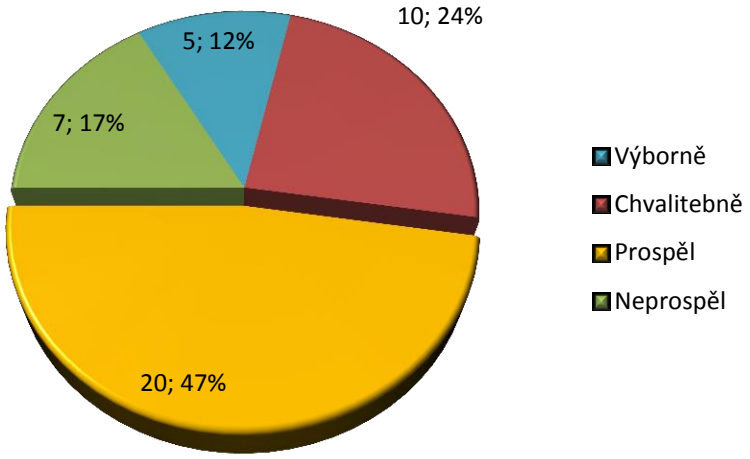
Na co si dát pozor?

- Subjektivně vnímáme plochu (objem), nikoliv výšku jednotlivých „sloupců“.
- Nadbytečné názvy grafu, legendy, …
- Neefektivní nuly
- Informativní hodnota grafu

Grafické znázornění

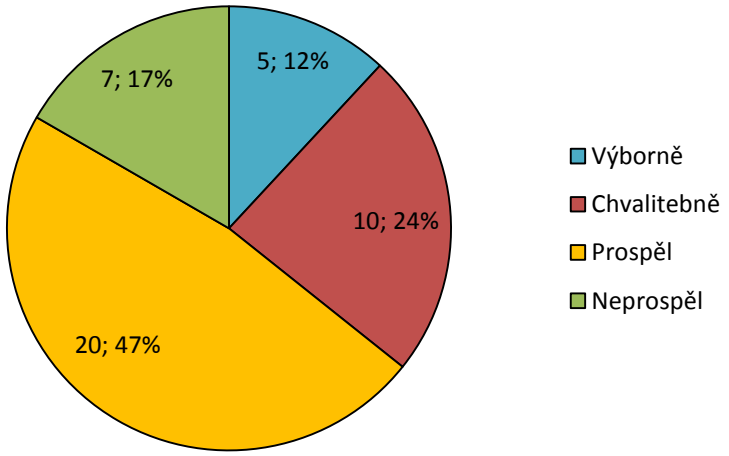
B) Výsečový graf – koláčový graf (pie chart)

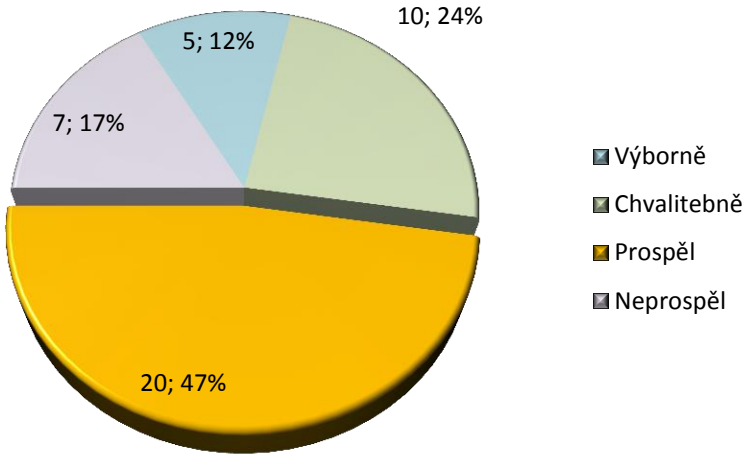




Grafické znázornění

B) Výsečový graf – koláčový graf (pie chart)





Grafické znázornění

B) Výsečový graf – koláčový graf (pie chart)

Na co si dát pozor?

Anketa

Jste pro navýšení hodinové dotace Matematiky na SŠ?

TAKHLE NE!!!

Grafické znázornění

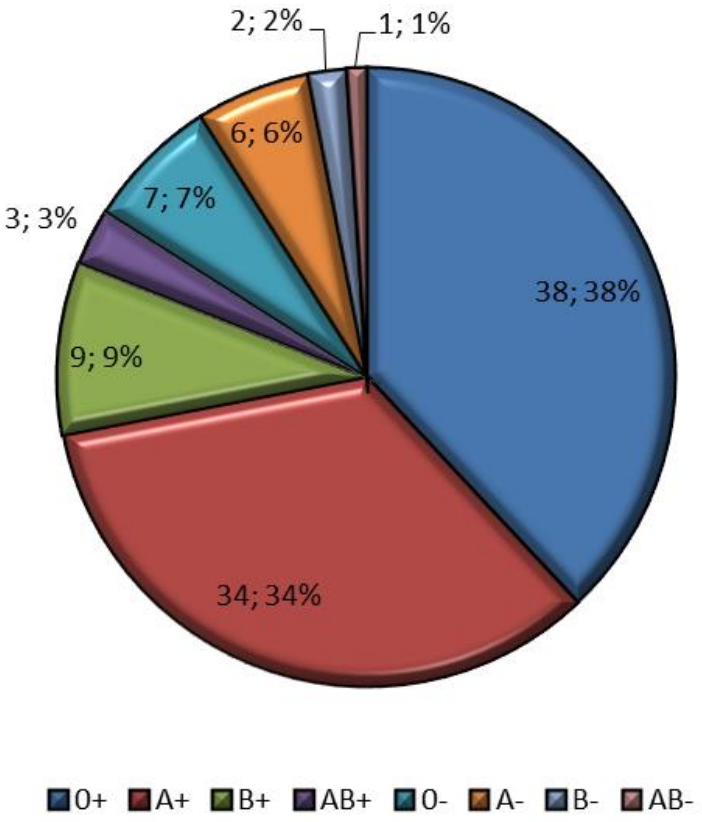
B) Výsečový graf – koláčový graf (pie chart)

Na co si dát pozor?

- Neuvádění absolutních četností, resp. celkového počtu respondentů v „blízkosti“ grafu

- Nadbytečné názvy grafu

Výskyt krevních skupin a Rh faktoru v USA



Krevní skupina	Rh faktor		Celkem
	Rh+	Rh-	
0	38	7	45
A	34	6	40
B	9	2	11
AB	3	1	4
Celkem	84	16	100

Procentuální zastoupení krevních skupin v populaci USA

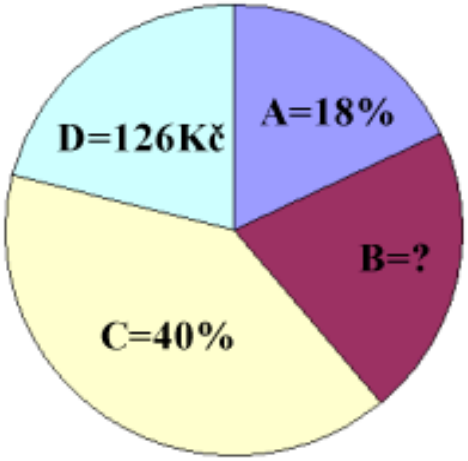
Grafické znázornění

B) Výsečový graf – koláčový graf (pie chart)

Na co si dát pozor?

- Neuvádění absolutních četností, resp. celkového počtu respondentů v „blízkosti“ grafu
- Nadbytečné názvy grafu, legendy, …
- Ne vždy je graf přehlednější než tabulka

Zdroj: Testové příklady určené žákům 9. tříd.



Co je to A, B, C, D?

Jsou výseče odpovídající variantám B a D stejně velké?

Lze velikosti jednotlivých výsečí charakterizovat v absolutních číslech i v procentech?

Určete pravdivost tvrzení:

a) Místo otazníku patří 20%.
b) Místo otazníku patří 126 Kč.
c) Část C je dvojnásobkem části D.

Rozdělení četností kvalitativního znaku se znázorňuje kruhovým diagramem, kde různým hodnotám znaku odpovídají kruhové výseče, jejichž plošné obsahy jsou úměrné četnostem. (Prometheus)

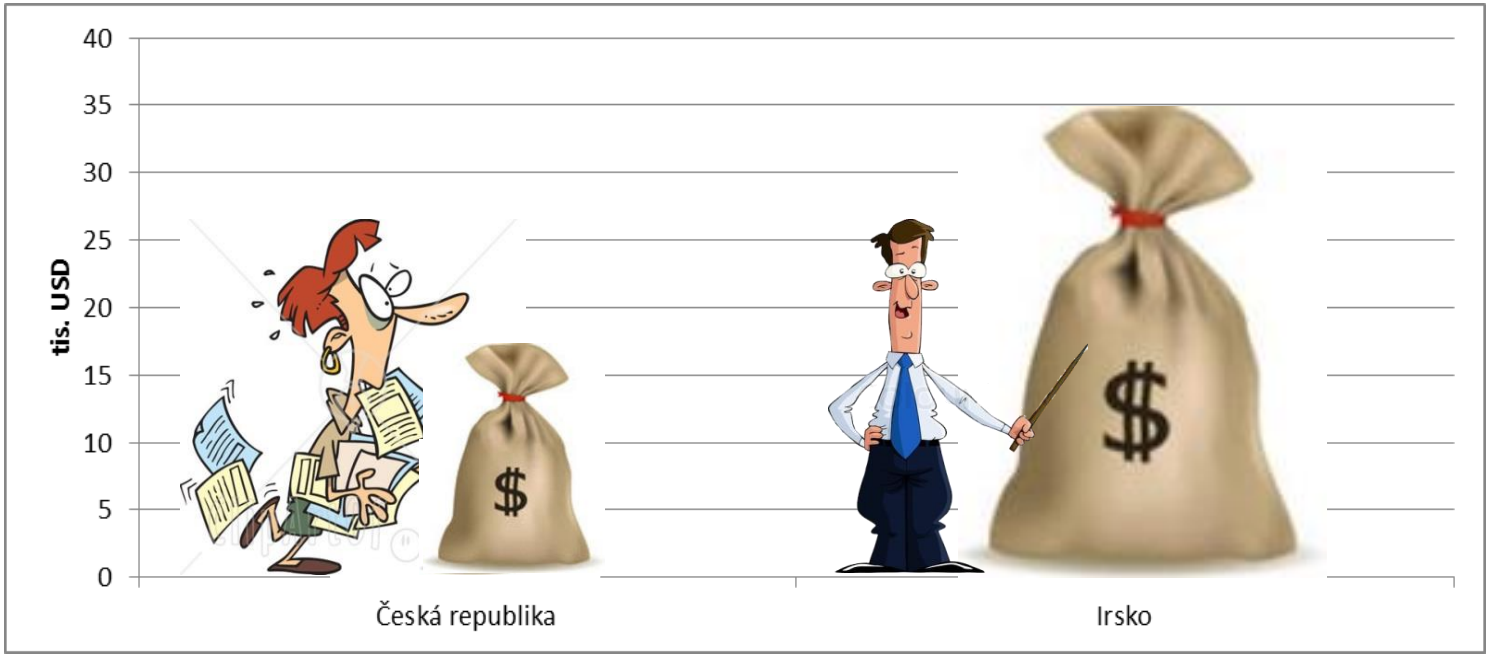
Grafické znázornění

A) Sloupcový graf (bar chart)

B) Výsečový graf – koláčový graf (pie chart)

C) Obrázkové grafy

Obrázkové grafy – užiteční pomocníci?



Srovnání průměrných ročních nástupních platů učitelů středních škol v ČR (17 244 $) a Irsku (34 604 $)

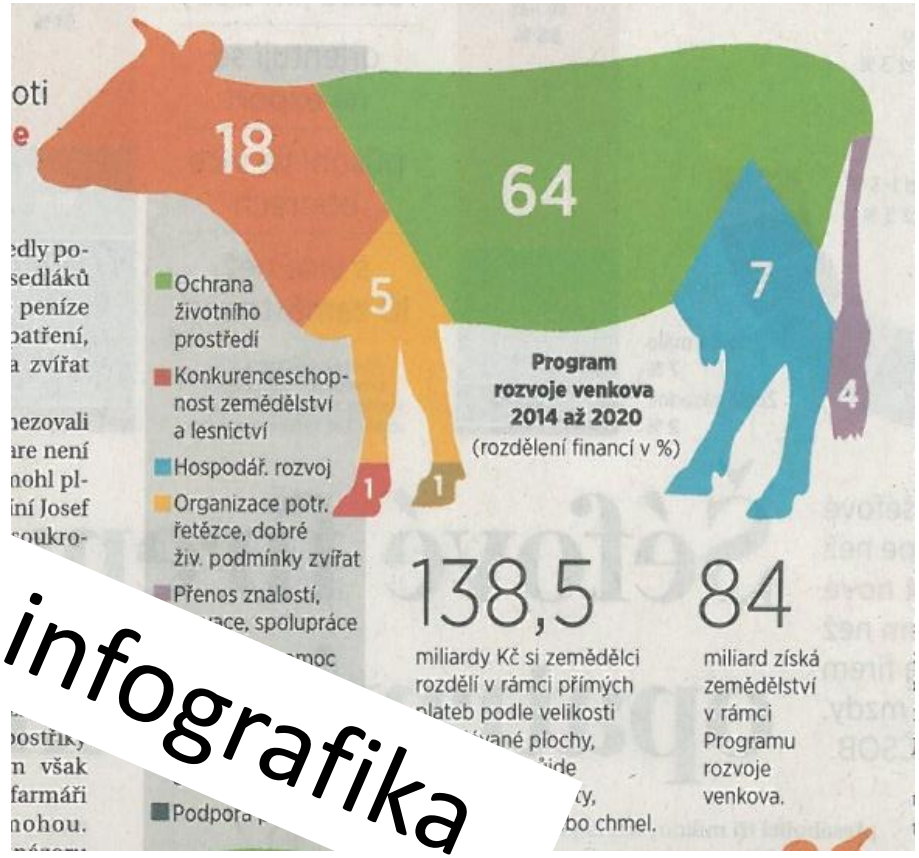
Obrázkové grafy – užiteční pomocníci?

Srovnání průměrných ročních nástupních platů učitelů středních škol

v ČR (17 244 $) a Irsku (34 604 $)

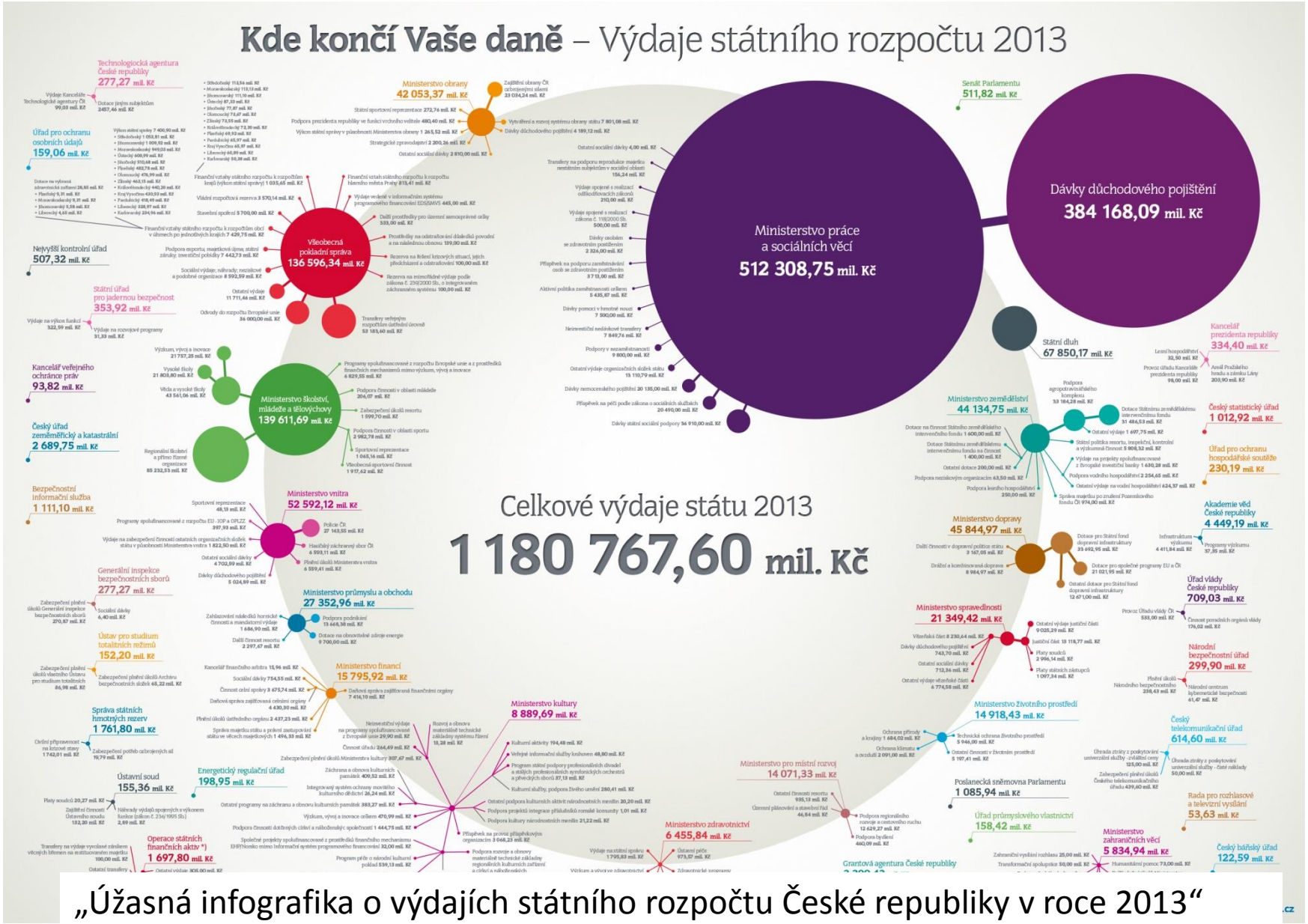
Několik praktických příkladů aneb „To přece bylo v novinách…“

Obrázkové grafy – užiteční pomocníci?



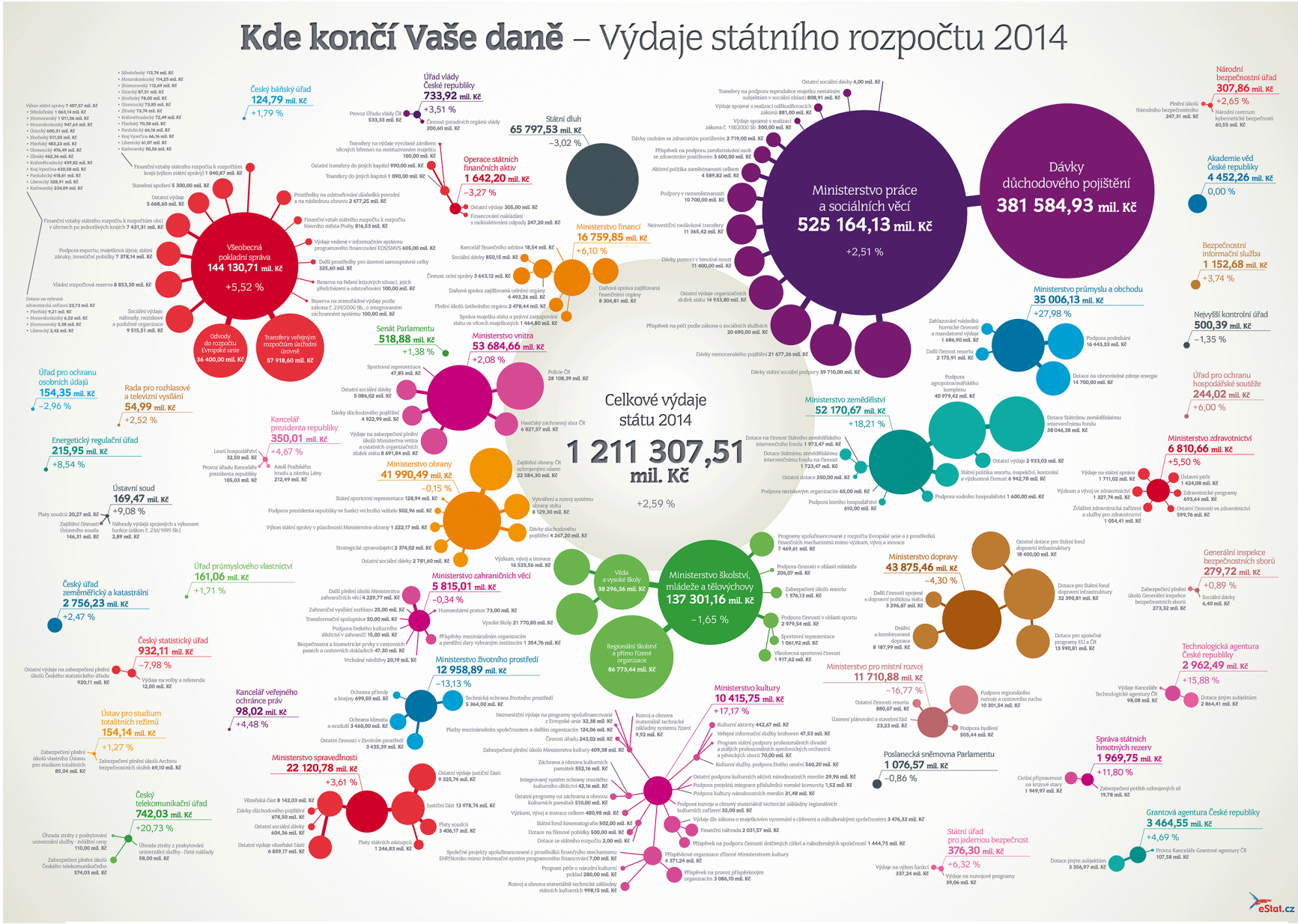
(Zdroj: Mf Dnes, 10. 7. 2014:

Zemědělci si rozdělí miliardy. Krávy a vepři se budou mít lépe.



„Úžasná infografika o výdajích státního rozpočtu České republiky v roce 2013“

Zdroj: http://www.estat.cz/zpravy/informace-k-projektum/kde-konci-vase-dane/



Zdroj: http://www.estat.cz/zpravy/informace-k-projektum/kde-konci-vase-dane/

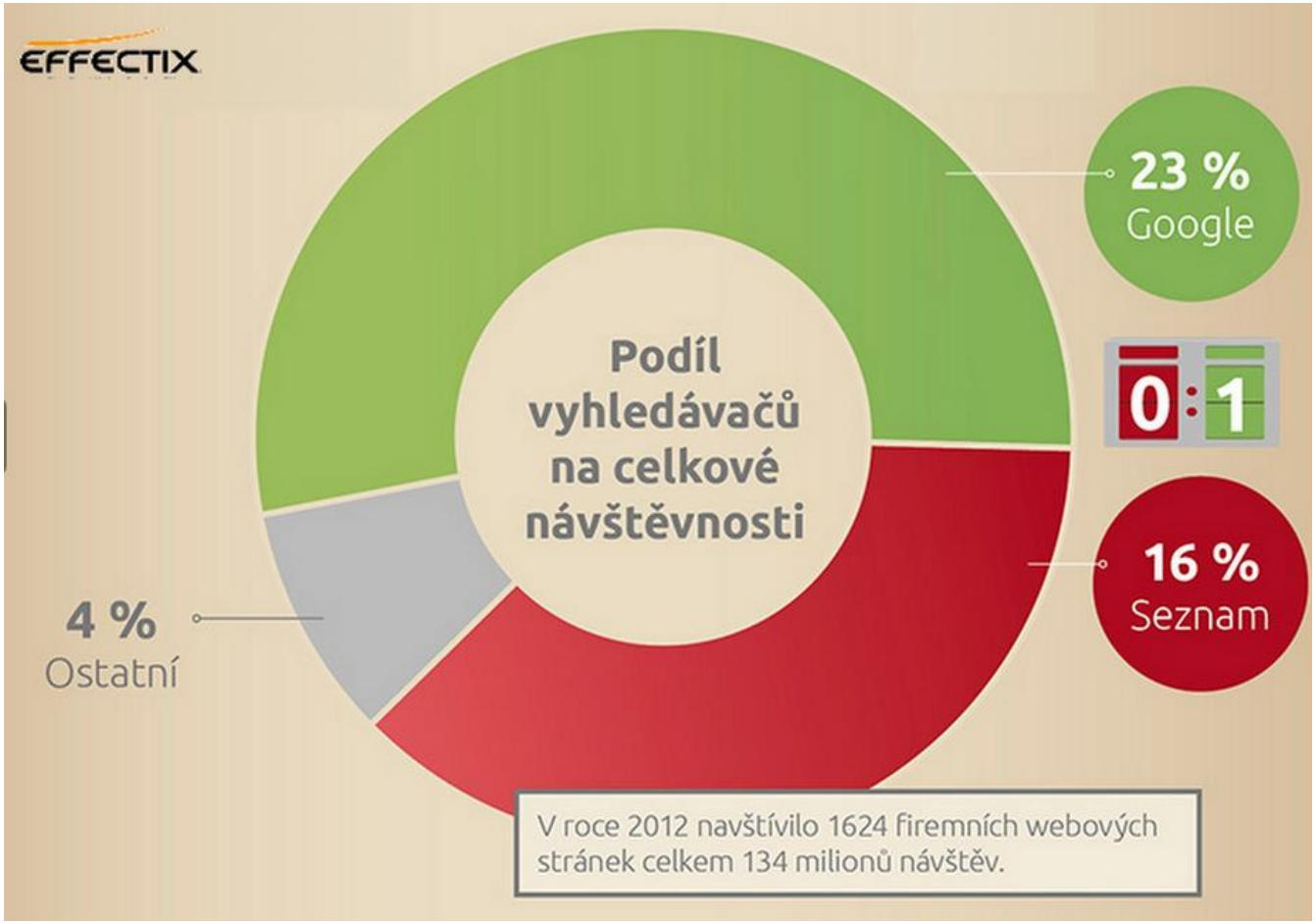
Příklad s klobásou

Příklad s klobásou

2. místo: klobása z trhů na náměstí Míru (45 Kč)

Obsah masa:	79,73 %
Obsah tuku:	24,94 %
Obsah bílkovin:	12,4 %
Obsah vody:	55,56 %
Obsah kolagenu:	2,6 %
Sójová bílkovina:	méně než 0,7 %

Souboj vyhledávačů

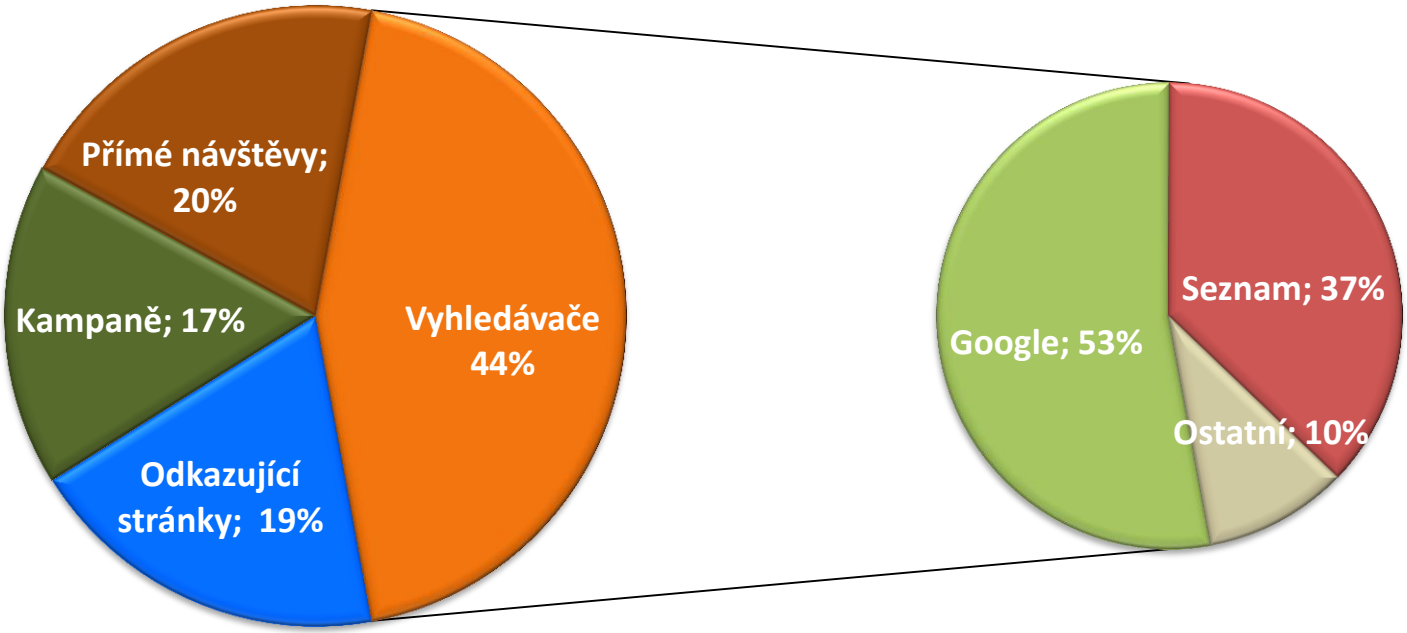


Zdroj: http://www.zive.cz/clanky/infografika-souboj-vyhledavacu-seznamcz-a-google/sc-3-a-167776/default.aspx

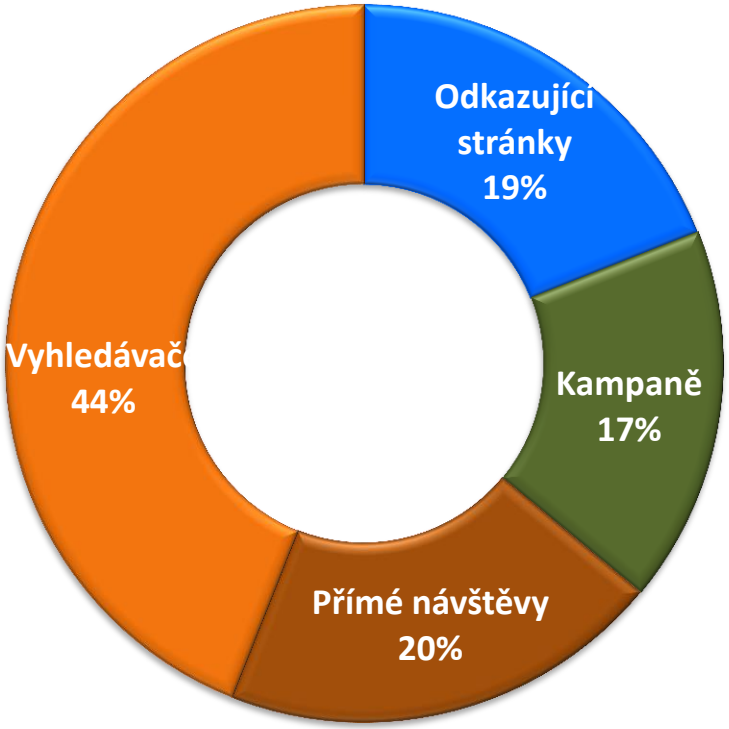
Souboj vyhledávačů

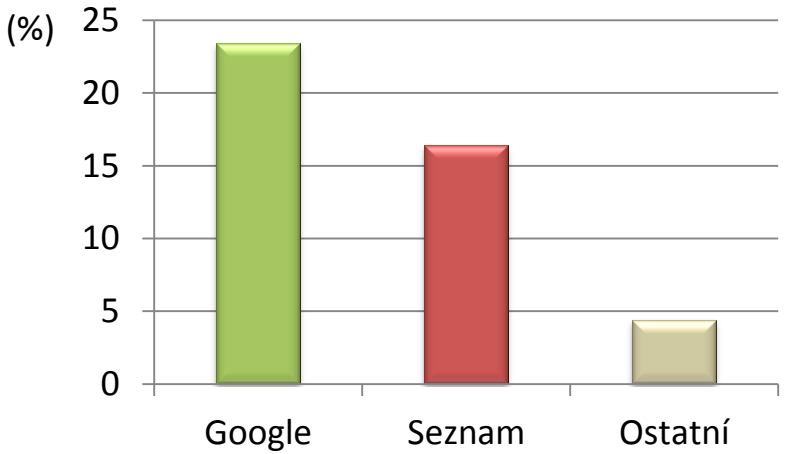
Zdroj: http://www.zive.cz/clanky/infografika-souboj-vyhledavacu-seznamcz-a-google/sc-3-a-167776/default.aspx

Jak výsledky šetření zobrazit správně?



Jak výsledky šetření zobrazit správně?





Průzkum o představách studentů o budoucím zaměstnání

DNES MIMOŘÁDNÁ PŘÍLOHA

RADY Životopis a přijímací pohovor
Jak zaujmout budoucího zaměstnavatele a vyhnout se obvyklým chybám Str. C2 a C5

INTERNET Hledání práce přes LinkedIn
Pět tipů, jak mít atraktivní on-line profil a dostávat nabídky od personalistů Str. C3

ČTVRTEK 27. 3. 2014 WWW.IDNES.CZ C

Velký přehled trainee programů, které firmy připravily pro absolventy, najdete na str. C2 a C7

Ideální místo? Mezinárodní firma, nástup aspoň 18 tisíc

Podle kterých kritérií si absolventi škol vybírají práci, jaký by si přáli nástupní plat a nakolik jsou ochotni ustoupit ze svých požadavků, to zjišťoval nedávný průzkum. Některá čísla překvapila.

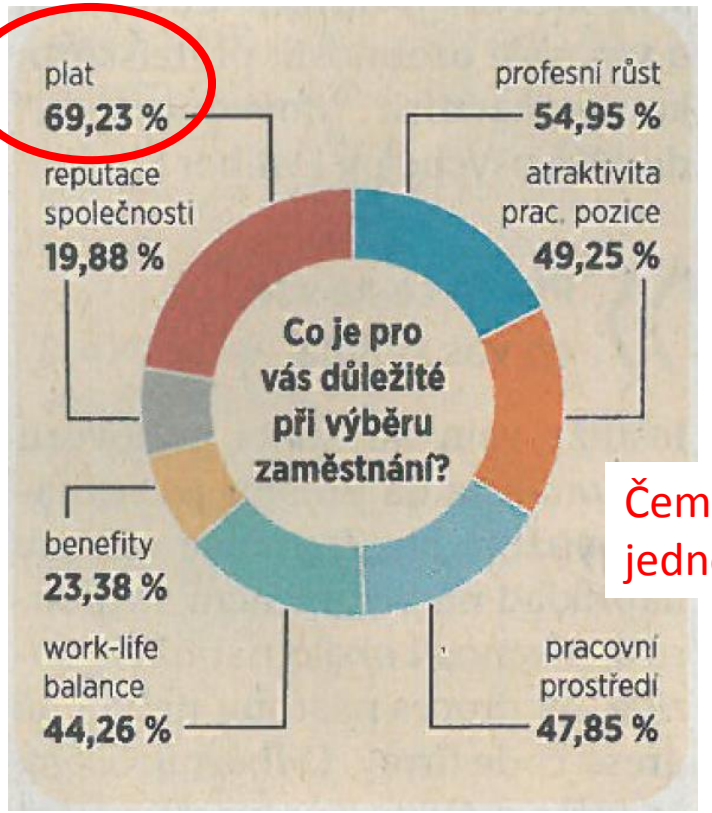
Jaké představy mají vysokoškoláci o své budoucí práci? To mapoval on-line průzkum

Mimořádná příloha Mf Dnes, 27. 3. 2014 – výsledky šetření spol. Studenta Media (typ šetření: online dotazování, specifikace výběru: „přes tisíc vysokoškoláků ze všech ročníků po celé republice“)

Průzkum o představách studentů o budoucím zaměstnání

S přesností na setinu procenta…

1000 studentů … 100%
10 studentů … 1%
0,1 studentů … 0,01%



Proč není součet 100%?

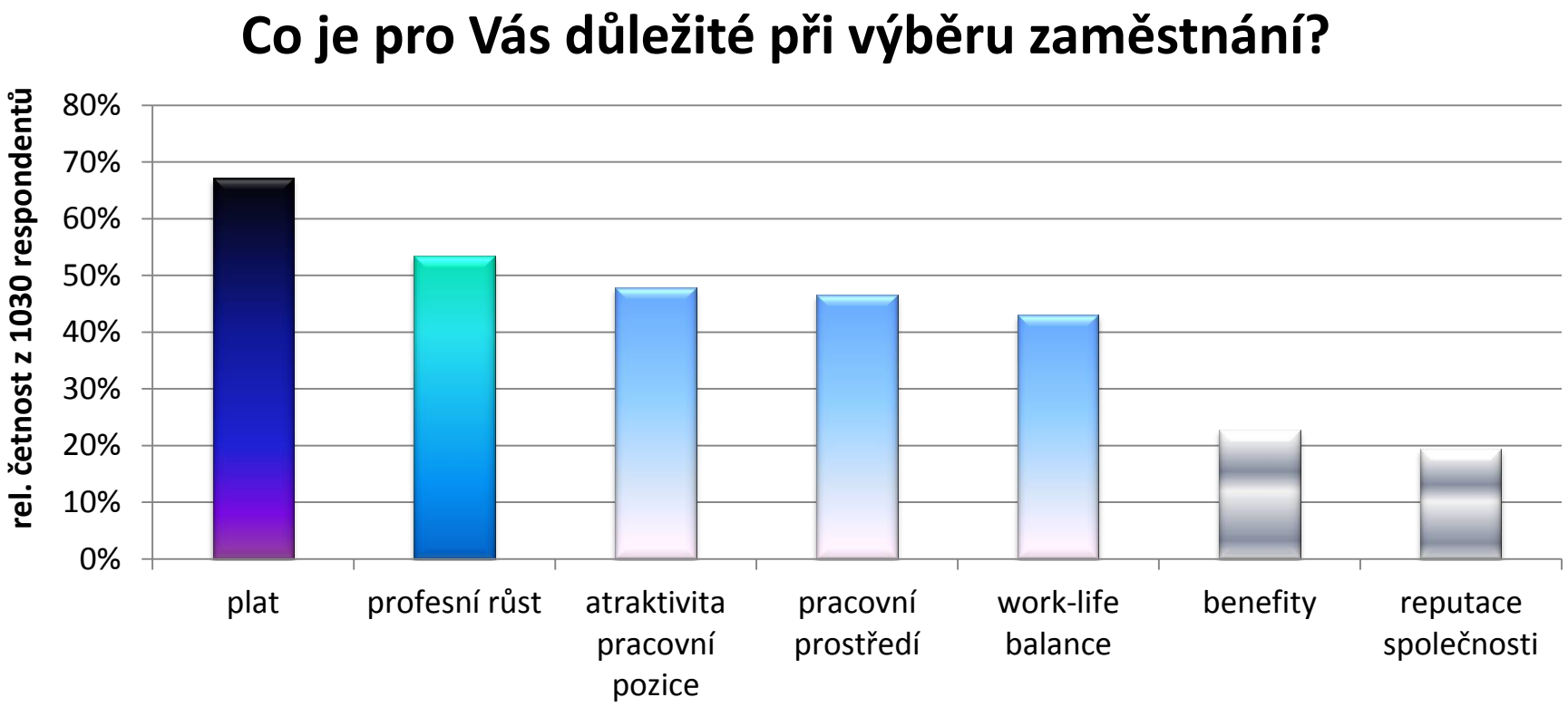
Čemu odpovídá velikost jednotlivých částí prstence?

Mimořádná příloha Mf Dnes, 27. 3. 2014 – výsledky šetření spol. Studenta Media (typ šetření: online dotazování, specifikace výběru: „přes tisíc vysokoškoláků ze všech ročníků po celé republice“)

Jak výsledky šetření zobrazit správně?

Co je pro Vás důležité při výběru zaměstnání? (vyberte 3 pro Vás nejdůležitější faktory)			
	četnost	rel. četnost (%)	rel. četnost (%) vzhledem k počtu respondentů
plat	692	22	67
profesní růst	550	18	53
atraktivita pracovní pozice	493	16	48
pracovní prostředí	479	16	47
work-life balance	443	14	43
benefity	234	8	23
reputace společnosti	199	6	19
celkem	**3090**	**100%**	**---**

Jak výsledky šetření zobrazit správně?



EDA pro numerická data

Číselné charakteristiky

A) Míry polohy (úrovně)

B) Míry variability

Míry polohy

- Odhadují skutečnou populační střední hodnotu na základě výběrového souboru.

- Patří mezi ně: výběrový aritmetický průměr, výběrový geometrický průměr, výběrový medián a modus.

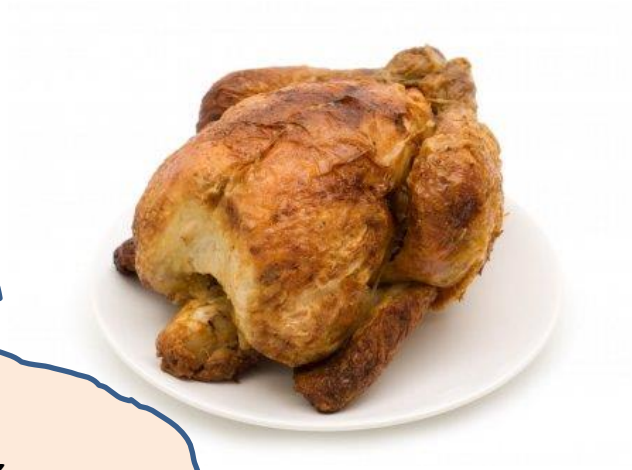
- Dalšími mírami polohy, které se týkají popisu i polohy jiných hodnot než středních, jsou kvantily.

Aritmetický průměr

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Pozor na ošidnost aritmetického průměru!

Ošidnost průměru



Průměrná produkce kuřat (na osobu):
0,5 (denně)

Soused snědl celé kuře, já jsem neměl nic…

Ošidnost průměru

- V malé vesnici někde v Americe žije 6 lidí, jejichž roční plat je uveden níže.

  $25 000 $27 000 $29 000

  $35 000 $37 000 $38 000

  Určete průměrný plat obyvatel této vesnice.

  ($31,830)

- Do vesnice se přistěhoval Bill Gates, jehož roční příjem je $40,000,000.

  $25 000 $27 000 $29 000

  $35 000 $37 000 $38 000 $40 000 000

  Určete průměrný plat obyvatel této vesnice.

  ($5 741 571)

Ošidnost průměru

Průměrná mzda přivádí Čechy k zuřivosti: Kdo z vás má 27 tisíc?

Pojem průměrná mzda přivádí Čechy k zuřivosti
Foto: Blesk

Zdroj: Blesk, 9.4.2013

Ošidnost průměru

5 důvodů, proč průměrný Čech nemá průměrnou mzdu

Na průměrnou mzdu dosáhne jen třetina Čechů. (Profimedia.cz)

SDÍLEJ! To se mi líbí 3 Sdílet 0 Tweet 0 12. března 2014 • 13:23

Průměrný Čech zůstal včera jako opařený: průměrná mzda v posledním čtvrtletí loňského roku činila 26 637 Kč hrubého. Co víc! Proti stejnému období z roku 2012 ještě navíc klesla o 484 Kč. Jenže kdo z »průměrných« Čechů ji má?

Zdroj: Blesk, 12.3.2014

Průměrné mzdy - 4. čtvrtletí 2013

Nízké mimořádné odměny srazily průměrnou mzdu

11.3. 2014

Ve 4. čtvrtletí 2013 se průměrná hrubá měsíční nominální mzda na přepočtené počty zaměstnanců v národním hospodářství snížila proti stejnému období předchozího roku o 1,8 %, reálně poklesla o 2,9 %. Medián mezd činil 22 288 Kč.

Ve 4. čtvrtletí 2013 činila průměrná hrubá měsíční nominální mzda*) (dále jen „průměrná mzda") na přepočtené počty zaměstnanců v národním hospodářství celkem 26 637 Kč, což je o 484 Kč (1,8 %) méně než ve stejném období roku 2012. Spotřebitelské ceny se zvýšily za uvedené období o 1,1 %, reálně se tak mzda snížila o 2,9 %. Objem mezd poklesl o 2,9 %, počet zaměstnanců o 1,2 %.

V podnikatelské sféře se průměrná mzda snížila nominálně o 2,4 %, reálně o 3,5 %, v nepodnikatelské sféře vzrostla nominálně o 0,8 %, reálně se snížila o 0,3 %.

Proti předchozímu čtvrtletí poklesla průměrná mzda ve 4. čtvrtletí 2013 po očištění od sezónních vlivů o 0,6 %.

Medián mezd (22 288 Kč) vzrostl proti stejnému období předchozího roku o 0,1 %, u mužů činil 24 330 Kč, u žen byl 20 029 Kč. Osmdesát procent zaměstnanců pobíralo mzdu mezi 10 546 Kč a 43 413 Kč.

V roce 2013 dosáhla průměrná mzda výše 25 128 Kč, v meziročním srovnání činil přírůstek 16 Kč (0,1 %). Spotřebitelské ceny se zvýšily za uvedené období o 1,4 %, reálně se mzda snížila o 1,3 %.

Zdroj: http://www.czso.cz/csu/csu.nsf/informace/cpmz031114.docx

Průměrné mzdy - 4. čtvrtletí 2013

Nízké mimořádné odměny srazily průměrnou mzdu

11.3. 2014

Ve 4. čtvrtletí 2013 se průměrná hrubá měsíční nominální mzda na přepočtené počty zaměstnanců v národním hospodářství snížila proti stejnému období předchozího roku o 1,8 %, reálně poklesla o 2,9 %. Medián mezd činil 22 288 Kč.

Ve 4. čtvrtletí 2013 činila průměrná hrubá měsíční nominální mzda[*)] (dále jen „průměrná mzda") na přepočtené počty zaměstnanců v národním hospodářství celkem 26 637 Kč, což je o 484 Kč (1,8 %) méně než ve stejném období roku 2012. Spotřebitelské ceny se zvýšily za uvedené období o 1,1 %, reálně se tak mzda snížila o 2,9 %. Objem mezd poklesl o 2,9 %, počet zaměstnanců o 1,2 %.

V podnikatelské sféře se průměrná mzda snížila nominálně o 2,4 %, reálně o 3,5 %, v nepodnikatelské sféře vzrostla nominálně o 0,8 %, reálně se snížila o 0,3 %.

Proti předchozímu čtvrtletí poklesla průměrná mzda ve 4. čtvrtletí 2013 po očištění od sezónních vlivů o 0,6 %.

Medián mezd (22 288 Kč) vzrostl proti stejnému období předchozího roku o 0,1 %, u mužů činil 24 330 Kč, u žen byl 20 029 Kč. Osmdesát procent zaměstnanců pobíralo mzdu mezi 10 546 Kč a 43 413 Kč.

V roce 2013 dosáhla průměrná mzda výše 25 128 Kč, v meziročním srovnání činil přírůstek 16 Kč (0,1 %). Spotřebitelské ceny se zvýšily za uvedené období o 1,4 %, reálně se mzda snížila o 1,3 %.

Zdroj: http://www.czso.cz/csu/csu.nsf/informace/cpmz031114.docx

Průměrné mzdy - 4. čtvrtletí 2013

Nízké mimořádné odměny srazily průměrnou mzdu

11.3. 2014

Ve 4. čtvrtletí 2013 se průměrná hrubá měsíční nominální mzda na přepočtené počty zaměstnanců v národním hospodářství snížila proti stejnému období předchozího roku o 1,8 %, reálně poklesla o 2,9 %. Medián mezd činil 22 288 Kč.

Ve 4. čtvrtletí 2013 činila průměrná hrubá měsíční nominální mzda*) (dále jen „průměrná mzda") na přepočtené počty zaměstnanců v národním hospodářství celkem 26 637 Kč, což je o 484 Kč (1,8 %) méně než ve stejném období roku 2012. Spotřebitelské ceny se zvýšily za uvedené období o 1,1 %, reálně se tak mzda snížila o 2,9 %. Objem mezd poklesl o 2,9 %, počet zaměstnanců o 1,2 %.

V podnikatelské sféře se průměrná mzda snížila nominálně o 2,4 %, reálně o 3,5 %, v nepodnikatelské sféře vzrostla nominálně o 0,8 %, reálně se snížila o 0,3 %.

Proti předchozímu čtvrtletí poklesla průměrná mzda ve 4. čtvrtletí 2013 po očištění od sezónních vlivů o 0,6 %.

Medián mezd (22 288 Kč) vzrostl proti stejnému období předchozího roku o 0,1 %, u mužů činil 24 330 Kč, u žen byl 20 029 Kč. Osmdesát procent zaměstnanců pobíralo mzdu mezi 10 546 Kč a 43 413 Kč.

V roce 2013 dosáhla průměrná mzda výše 25 128 Kč, v meziročním srovnání činil přírůstek 16 Kč (0,1 %). Spotřebitelské ceny se zvýšily za uvedené období o 1,4 %, reálně se mzda snížila o 1,3 %.

Zdroj: http://www.czso.cz/csu/csu.nsf/informace/cpmz031114.docx

Aritmetický průměr

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Na co si dát pozor?

- Průměr není rezistentní vůči odlehlým pozorováním!
- Harmonický průměr (proměnné vyjadřující čas na jednotku výkonu, poměrná čísla)
- Geometrický průměr (tempa růstu)
  - Vážený průměr
- Průměrování dat na cirkulární škále

Circular Statistics Toolbox

Příklad: *Cena jedné akcie energetické společnosti vzrostla na burze XY v období od 13. do 15. března téhož roku z 952,50 Kč na 982,00 Kč. Jaký byl průměrný denní relativní přírůstek ceny této akcie?*

	Cena akcie [Kč]	**Koeficient růstu**
13. března	952,50	
14. března	?	?/952,5
15. března	982,0	982,0/?

Průměrný denní relativní přírůstek ceny akcie byl 1,5%.

$$\bar{x} = \sqrt{\frac{?}{952{,}5} \cdot \frac{982{,}0}{?}} = \sqrt{\frac{982{,}0}{952{,}5}} = 1{,}015$$

Geometrický průměr

Výběrové kvantily

100p %-ní kvantil $\tilde{x}_p$

- odděluje 100p% menších hodnot od zbytku souboru

(100p% hodnot datového souboru je menších než toto číslo.)

Význačné výběrové kvantily

- Kvartily
  Dolní kvartil $\tilde{x}_{0,25}$
  Medián $\tilde{x}_{0,5}$
  Horní kvartil $\tilde{x}_{0,75}$

- Decily – $\tilde{x}_{0,1}; \tilde{x}_{0,2}; \ldots; \tilde{x}_{0,9}$

- Percentily – $\tilde{x}_{0,01}; \tilde{x}_{0,02}; \ldots; \tilde{x}_{0,03}$

- Minimum $\tilde{x}_{min}$ a Maximum $\tilde{x}_{max}$

Kde se s kvantily setkáme v praxi?

- Vyhodnocení Národních srovnávacích zkoušek, …

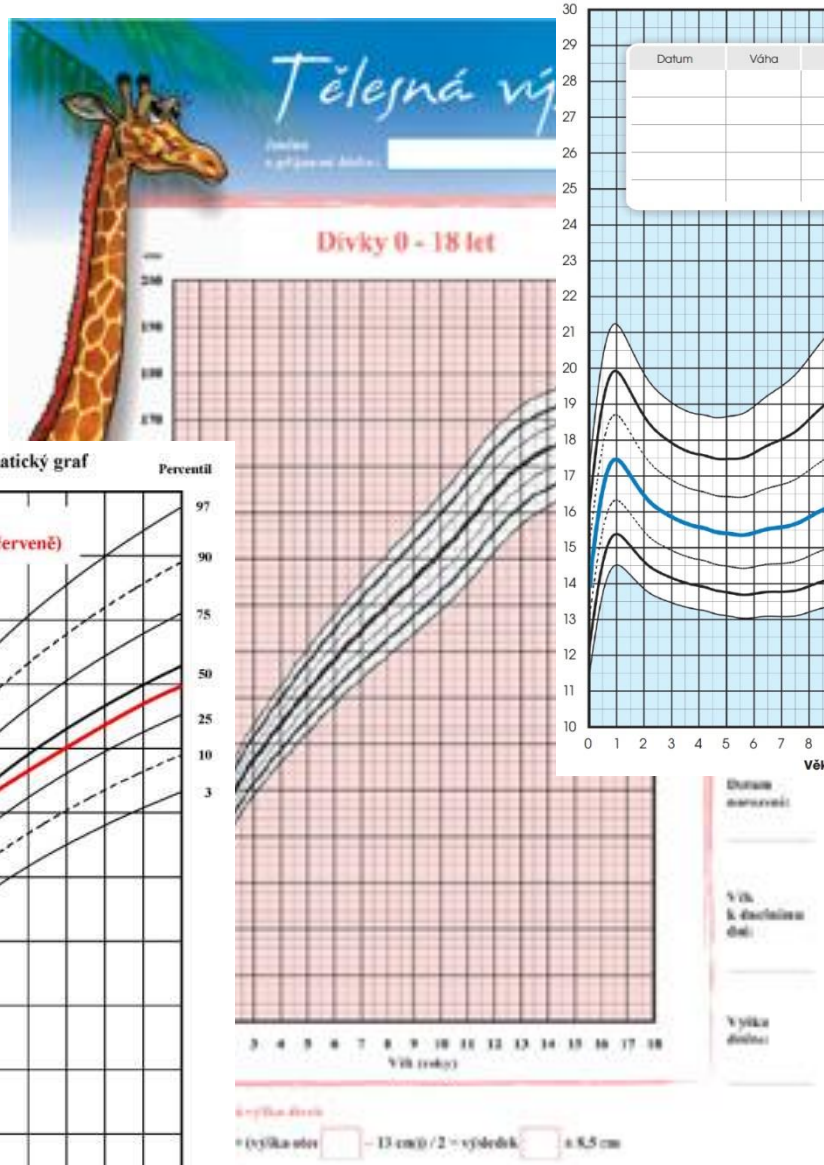
Percentil

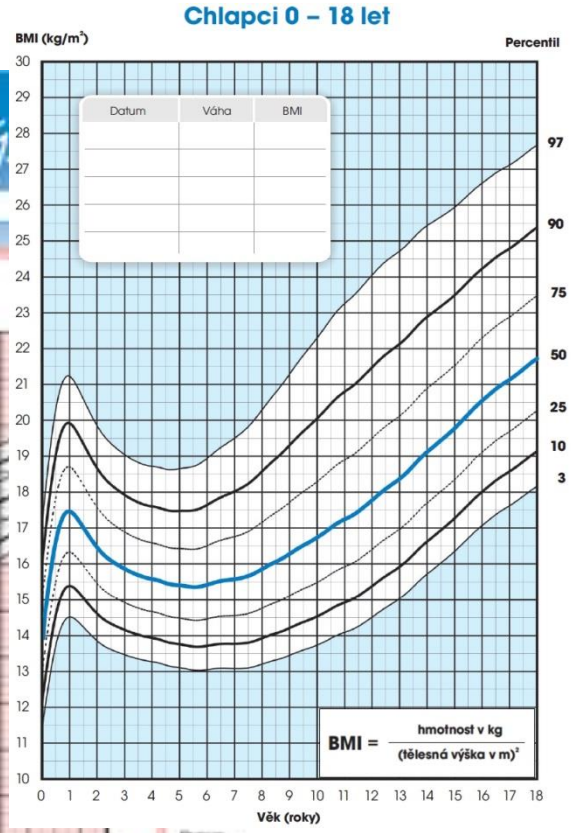
Skóre se převede na **percentil** - číslo udávající „kolik procent účastníků jste svým výsledkem v daném testu předstihli". Percentil 90 tedy znamená, že jste byli úspěšnější (nebo stejně úspěšní) než 90 % všech účastníků.

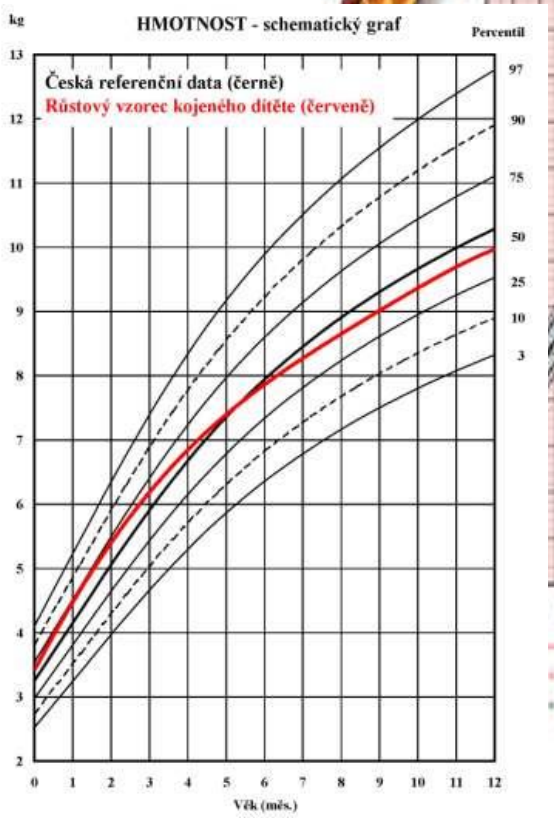
Zdroj: https://scio.cz/nsz/vyhodnoceni.asp

Kde se s kvantily setkáme v praxi?

- vyhodnocení Národních srovnávacích zkoušek, …
- růstové grafy


Tělesná
Dívky 0 - 18 let
Věk (roky)


Chlapci 0 – 18 let
BMI (kg/m²)
Percentil
Datum
Váha
BMI
97
90
75
50
25
10
3
BMI = hmotnost v kg / (tělesná výška v m)²
Věk (roky)


HMOTNOST - schematický graf
kg
Percentil
Česká referenční data (černě)
Růstový vzorec kojeného dítěte (červeně)
97
90
75
50
25
10
3
Věk (měs.)

Popis diskrétní proměnné

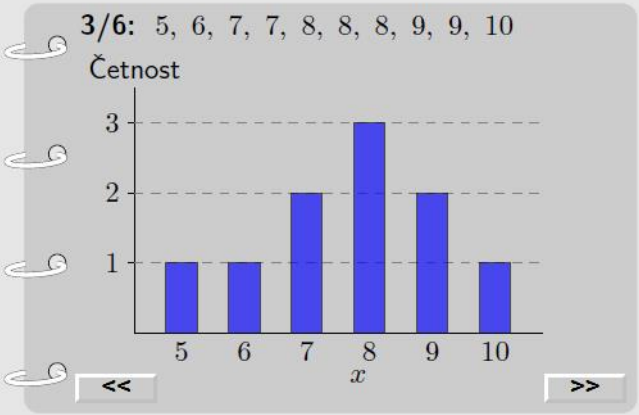
- V případě, že proměnná nabývá „několika málo“ hodnot, využíváme kombinace metod určených pro kvalitativní a kvantitativní proměnné.

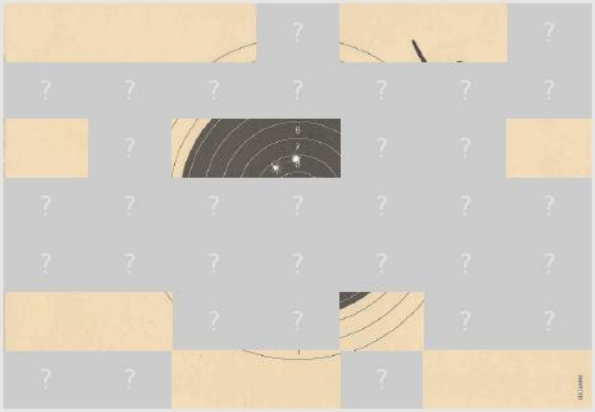
Hodnocení	Srozumitelnost výkladu	Srozumitelnost řešených příkladů	Praktické aplikace	Grafická úprava	Míra používání textu
1	61 (28%)	64 (29%)	34 (16%)	84 (38%)	73 (33%)
2	103 (47%)	106 (48%)	68 (31%)	92 (42%)	74 (34%)
3	45 (21%)	41 (19%)	80 (37%)	32 (15%)	51 (23%)
4	6 (3%)	4 (2%)	22 (10%)	8 (4%)	18 (8%)
5	2 (1%)	1 (0%)	2 (1%)	2 (1%)	2 (1%)
"X"	2 (1%)	3 (1%)	13 (6%)	1 (0%)	0 (0%)
Průměr	2,0	1,9	2,5	1,9	2,1
Medián	2,0	2,0	3,0	2,0	2,0

Jak určit medián a modus diskrétní proměnné?

Při sportovní střelbě ze vzduchovky bylo v jednotlivých sériích 10 střel dosaženo uvedených výsledků (znak x).
Přiřaďte jednotlivých sériím jejich modus $\mathrm{Mod}(x)$ a medián $\mathrm{Med}(x)$.

3/6: 5, 6, 7, 7, 8, 8, 8, 9, 9, 10





Medián a modus:

a $\mathrm{Med}(x) = 8,\ \mathrm{Mod}(x) = 8$

b $\mathrm{Med}(x) = 8,\ \mathrm{Mod}(x) = 7$

c $\mathrm{Med}(x) = 9,\ \mathrm{Mod}(x) = 10$

d $\mathrm{Med}(x) = 8,\ \mathrm{Mod}(x) = 9$

e $\mathrm{Med}(x) = 9,\ \mathrm{Mod}(x) = 8$

f $\mathrm{Med}(x) = 9,\ \mathrm{Mod}(x) = 9$

Míry variability

- Charakteristiky hodnotící rozptýlenost hodnot statistického souboru kolem nějaké míry polohy.

- Patří mezi ně: (variační) rozpětí, mezikvartilové (interkvartilové) rozpětí, rozptyl, směrodatná odchylka a variační koeficient.

Variační rozpětí

$$R = x_{max} - x_{min}$$

Interkvartilové rozpětí

$$IQR = \tilde{x}_{0,75} - \tilde{x}_{0,25}$$

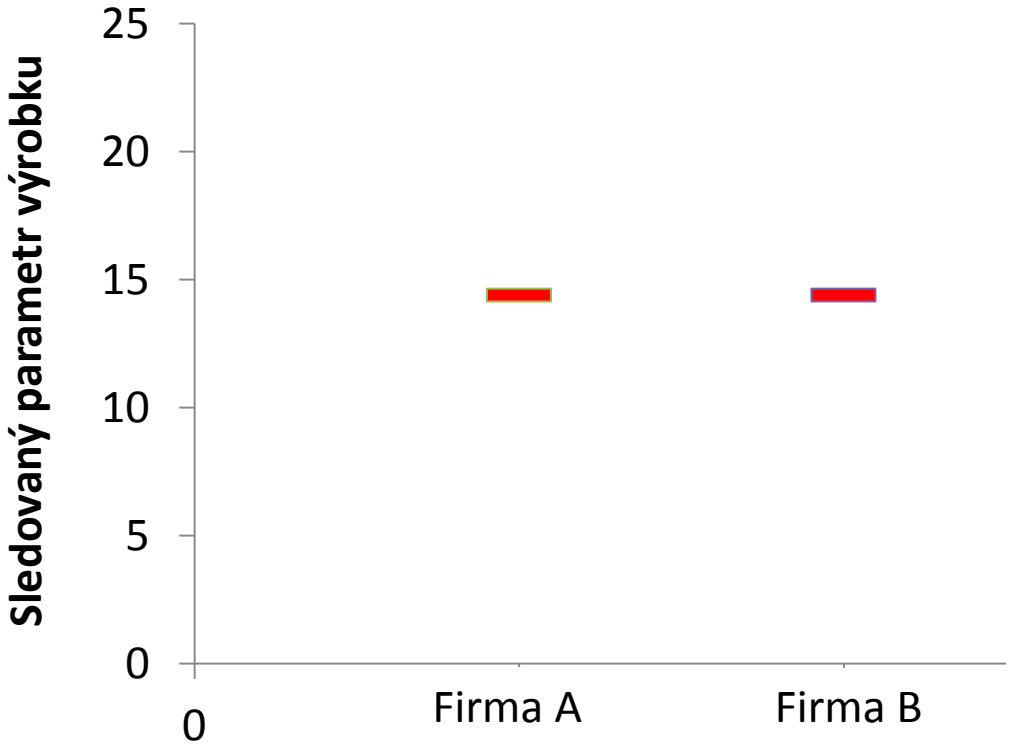
Interkvartilové rozpětí

$$IQR = \tilde{x}_{0,75} - \tilde{x}_{0,25}$$

Užití: např. při identifikaci odlehlých pozorování

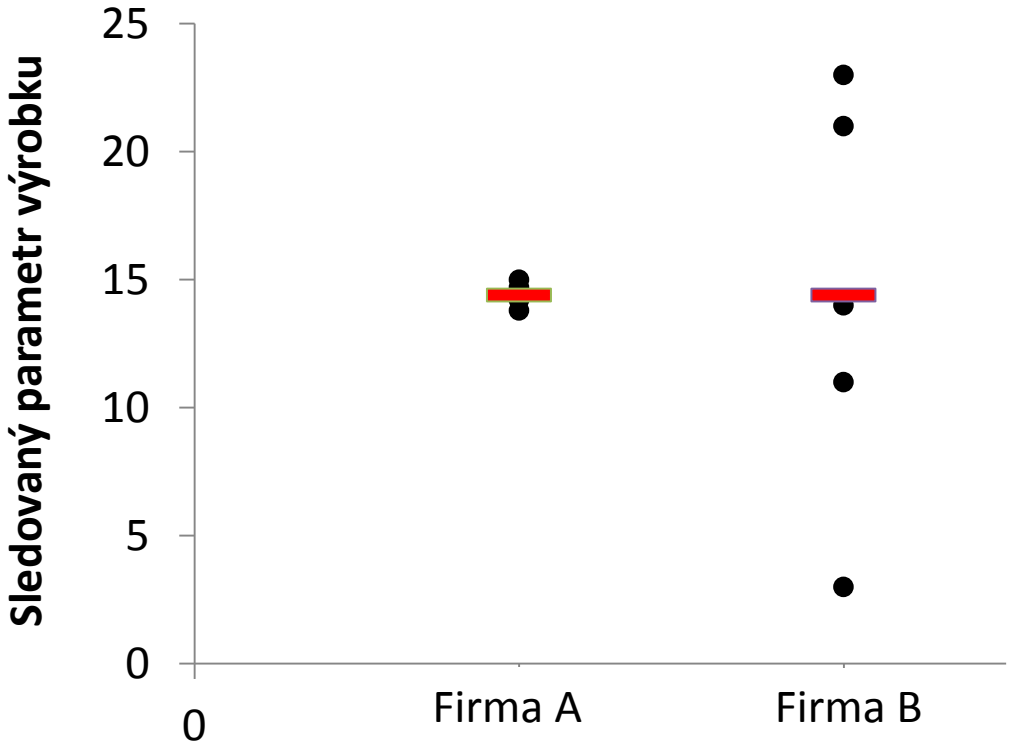
Proč potřebujeme míry variability?

Firmy A a B nabízejí dodávku jistého výrobku za srovnatelných finančních podmínek. Průměrné hodnoty sledovaného parametru výrobku udávají stejné…



Proč potřebujeme míry variability?

Firmy A a B nabízejí dodávku jistého výrobku za srovnatelných finančních podmínek. Průměrné hodnoty sledovaného parametru výrobku udávají stejné…



Výběrový rozptyl

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

Na co si dát pozor?

- Nezaměňovat s populačním rozptylem σ^2!!! $\left(\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}\right)$
- Rozměr rozptylu charakteristiky je druhou mocninou rozměru proměnné.

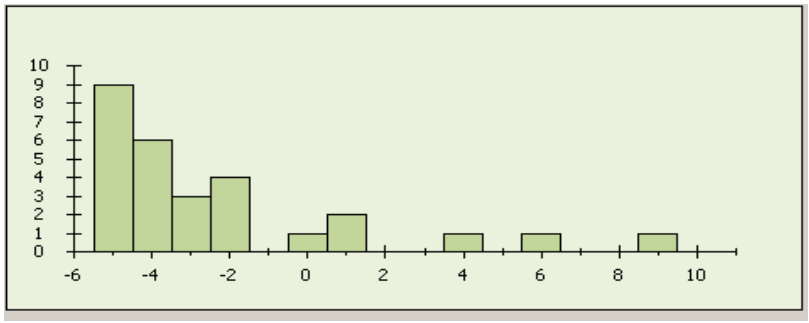
Výběrová směrodatná odchylka

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

Jakou představu o variabilitě dat nám dává sm. odchylka?

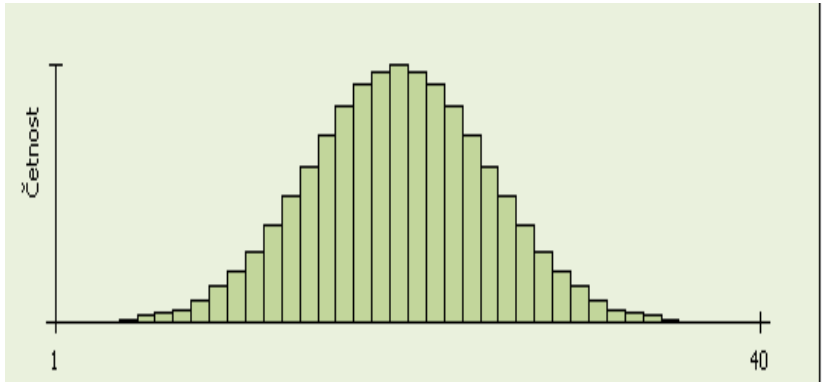
Čebyševova nerovnost: $\forall k > 0\colon\ P(\mu - k\sigma < X < \mu + k\sigma) > 1 - \frac{1}{k^2}$

k	$P(\mu - k\sigma < X < \mu + k\sigma)$
1	>0
2	>0,75
3	>0,89



Pravidlo 3 sigma

k	$P(\mu - k\sigma < X < \mu + k\sigma)$
1	0,682
2	0,954
3	0,998



Variační koeficient

(Směrodatná odchylka v procentech aritmetického průměru)

$$V_x = \frac{s}{\bar{x}} \cdot 100 \,[\%]$$

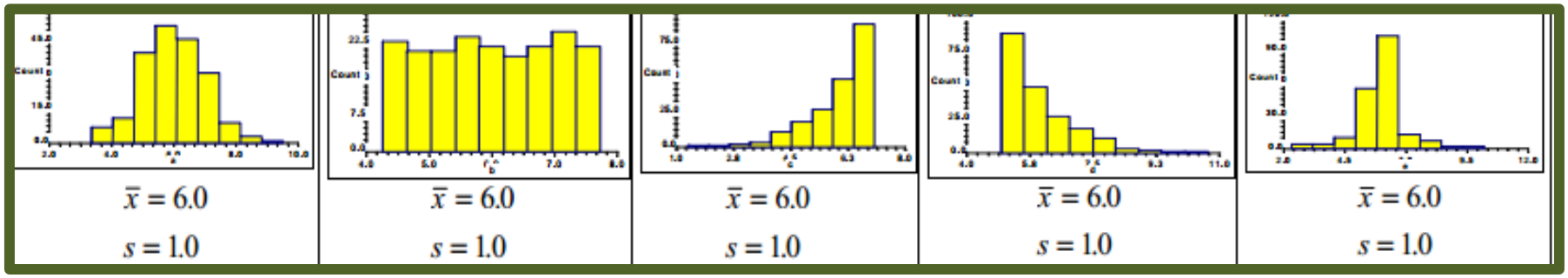
- Čím nižší var. koeficient, tím homogennější soubor.
- V_x > 50% značí silně rozptýlený soubor.

Proč potřebujeme bezrozměrnou míru variability?

Umožňuje srovnání variability proměnných, které mají různé jednotky.

Míry šikmosti a špičatosti

Jsou míry polohy a míry variability dostatečné pro posouzení rozdělení sledovaných veličin?



Zdroj: TVRDÍK, J.: Základy matematické statistiky, Ostravská univerzita, 2008

Všech pět ukázek má stejné charakteristiky polohy i variability (průměry i směrodatné odchylky jsou shodné). Přesto na první pohled vidíme, že tvary rozdělení dat jsou různé.

Pro „lepší“ představu o datech používáme pro jejich popis i míry šikmosti a špičatosti, ale to již nepatří do učiva SŠ…

Přesnost číselných charakteristik

Směrodatnou odchylku jakožto míru nejistoty měření zaokrouhlujeme **nahoru** na jednu, maximálně dvě platné cifry a míry polohy (průměr, kvantily…) zaokrouhlujeme tak, aby nejnižší zapsaný řád odpovídal nejnižšímu zapsanému řádu směrodatné odchylky.

Chybný zápis číselných charakteristik

	Délka [m]	Váha [kg]	Teplota [^{0}C]
Průměr	2,26	127,6	14 567
Medián	2,675	117,8	13 700
Směrodatná odchylka	0,78	23,7	1 200 (před zaokrouhlením 1235)
Proč je zápis chybný?			

Chybný zápis číselných charakteristik

	Délka [m]	**Váha [kg]**	**Teplota [^{0}C]**
Průměr	2,26	127,6	14 567
Medián	2,675	117,8	13 700
Směrodatná odchylka	0,78	23,3	1 200 (před zaokrouhlením 1235)
Proč je zápis chybný?	*Různý počet des. míst.*		

Chybný zápis číselných charakteristik

	Délka [m]	**Váha [kg]**	**Teplota [^{0}C]**
Průměr	2,26	127,6	14 567
Medián	2,675	117,8	13 700
Směrodatná odchylka	0,78	23,3	1 200 (před zaokrouhlením 1235)
Proč je zápis chybný?	*Různý počet des. míst.*	*3 platné cifry u směrodatné odchylky.*	

Chybný zápis číselných charakteristik

	Délka [m]	Váha [kg]	Teplota [^{0}C]
Průměr	2,26	127,6	14 567
Medián	2,675	117,8	13 700
Směrodatná odchylka	0,78	23,3	1 200 (před zaokrouhlením 1235)
Proč je zápis chybný?	*Různý počet des. míst.*	*3 platné cifry u směrodatné odchylky.*	*Nejnižší zapsaný řád průměru (jednotky) neodpovídá nejnižšímu zapsanému řádu směrodatné odchylky (stovky)+ směr. odch. není zaokrouhlena nahoru.*

Oprava

	Délka [m]	Váha [kg]	Teplota [^{0}C]
Průměr	2,26	127,6	14 567
Medián	2,68	117,8	13 700
Směrodatná odchylka	0,78	23,3	1 200 (před zaokrouhlením 1235)
Proč je zápis chybný?		*3 platné cifry u směrodatné odchylky.*	*Nejnižší zapsaný řád průměru (jednotky) neodpovídá nejnižšímu zapsanému řádu směrodatné odchylky (stovky)+ směr. odch. není zaokrouhlena nahoru.*

Oprava

	Délka [m]	Váha [kg]	Teplota [^{0}C]
Průměr	2,26	128	14 567
Medián	2,675	118	13 700
Směrodatná odchylka	0,78	24	1 200 (před zaokrouhlením 1235)
Proč je zápis chybný?			*Nejnižší zapsaný řád průměru (jednotky) neodpovídá nejnižšímu zapsanému řádu směrodatné odchylky (stovky)+ směr. odch. není zaokrouhlena nahoru.*

Správný zápis číselných charakteristik

	Délka [m]	Váha [kg]	Teplota [^{0}C]
Průměr	2,26	127,6	14 600
Medián	2,675	117,8	13 700
Směrodatná odchylka	0,78	23,7	1 300

Zajímavé odkazy k tématu Exploratorní statistika

- Slovníček pojmů z exploratorní statistiky aneb co by se Vám mohlo hodit při práci se statistickým softwarem v angličtině
- Interstat – sylabus popisné statistiky (nedokončeno)
- Jak nevytvářet grafy (anglicky) The Evil Tutor‘s Guide
- Real Time Statistics Project
- Projekt Gapminder
- Circular Statistics Toolbox (Matlab)

A to už je opravdu konec!

Děkuji za pozornost