EXPLORATORNÍ ANALÝZA DAT

7. cvičení

Teorie pravděpodobnosti x Statistika

- **Teorie pravděpodobnosti** – popisuje zákonitosti týkající se náhodných jevů, používá se k modelování náhodností a neurčitostí, které souvisí s nedostatečnou znalostí počátečních podmínek.

- **Statistika** – rozvíjí znalosti na základě dat získaných pozorováním; zabývá se metodami sběru dat a jejich zpracováním a vyhodnocováním.

Způsoby statistického šetření

Vyčerpávající šetření

prošetření všech jednotek statistického souboru (populace)

Výhody: přesnost a detailnost zjištěných informací

Nevýhody: personální, finanční a časová náročnost

Výběrové šetření

ze základního souboru (populace) o rozsahu *N* vybereme jeho část, tzv. **výběrový soubor**, zkráceně **výběr**, o rozsahu *n*. Ze zpracovaných výsledků pak usuzujeme na vlastnosti celé populace

Výhody: menší personální, finanční a časová náročnost

Nevýhody: mírou objektivnosti získaných informací je kvalita provedení výběrového šetření

Exploratorní analýza dat

= Popisná statistika (angl. Exploratory Data Analysis, EDA) – provede uspořádání proměnných do názornější formy a jejich popis několika málo hodnotami, které by obsahovaly co největší množství informací obsažených v původním souboru.

- Způsob zpracování proměnných závisí především na jejich typu.

Typy proměnných

Typy proměnných

- **Kvalitativní proměnná**
  (kategoriální, slovní)
  nemůžeme měřit, můžeme ji pouze zařadit do tříd
  - **Nominální proměnná**
    nabývá rovnocenných variant; nelze je smysluplně porovnávat ani seřadit
    (např. pohlaví, národnost, barva, ...)
  - **Ordinální proměnná**
    přechod mezi kvalitativními a kvantitativními proměnnými; jednotlivým variantám lze přiřadit pořadí a vzájemně je porovnávat nebo seřadit (např. vzdělání, velikost oděvů, známka ve škole)
- **Kvantitativní proměnná**
  (numerická, číselná)
  je měřitelná
  - **Diskrétní proměnná**
    nabývající konečného nebo spočetného množství variant.
    (např. známka z matematiky, věk v letech, výška v centimetrech, váha v kilogramech)
  - **Spojitá proměnná**
    nabývající libovolných hodnot z R nebo z nějaké podmnožiny R
    (např. výška, váha, vzdálenost, ...)

Nominální veličina

- nemá smysl porovnávání,
- nabývá v rámci souboru různých, ale rovnocenných kategorií,
- počet těchto kategorií nebývá příliš velký
- Např. barva, národnost, značka, pohlaví, kraj,...

Číselné charakteristiky

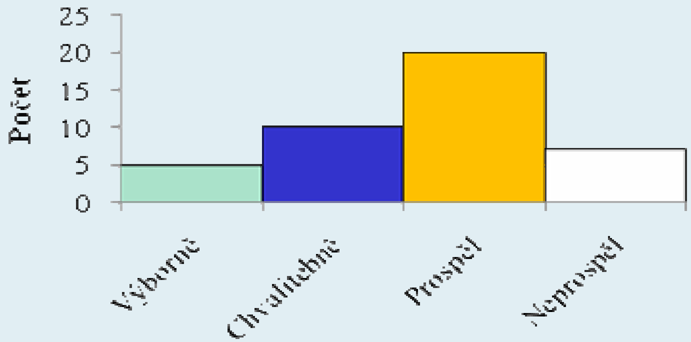
- **absolutní četnost** n_i - počet výskytu dané varianty kvalitativní proměnné.
- **relativní četnost** p_i – vyjadřuje velikost části souboru tvořené proměnnou s konkrétní variantou
  p_i = četnost/rozsah souboru = n_i/n
- **modus** – název varianty proměnné s nejvyšší četností, vyjadřuje typického reprezentanta souboru, v případě výskytu více variant s maximální četností modus neurčujeme

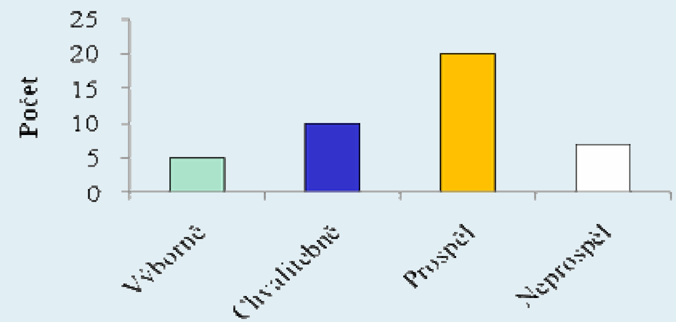
Tabulka rozdělení četnosti

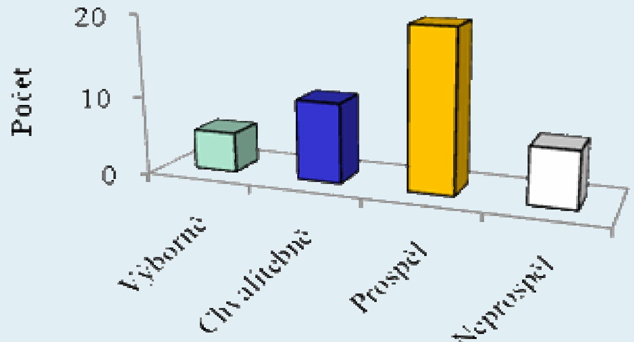
Hodnoty x_i	**Absolutní četnosti** n_i	**Relativní četnosti** p_i
x_1	n_1	$p_1 = n_1 / n$
x_2	n_2	$p_2 = n_2 / n$
...	...	...
x_k	n_k	$p_k = n_k / n$
Celkem:	$\sum_{i=1}^{k} n_i = n$	1

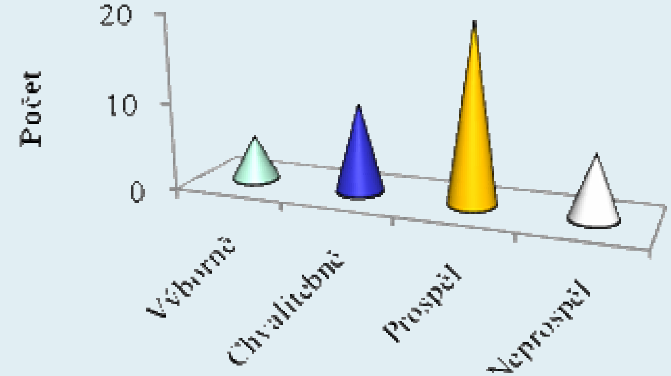
Grafické znázornění

- **Histogram** – sloupcový graf, znázorňuje **absolutní četnosti**: na jednu osu vynášíme varianty proměnné a na druhou osu jejich absolutní četnosti



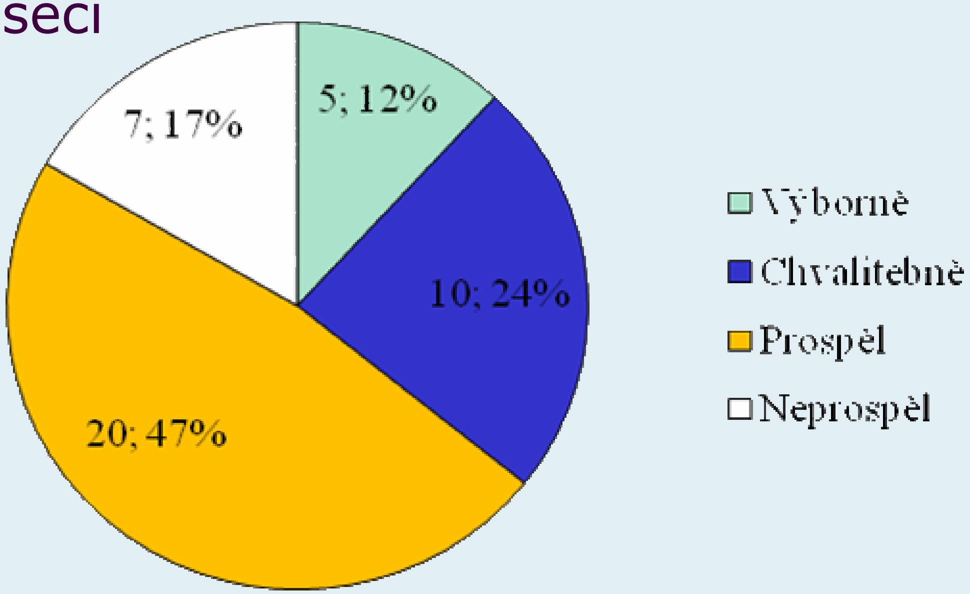






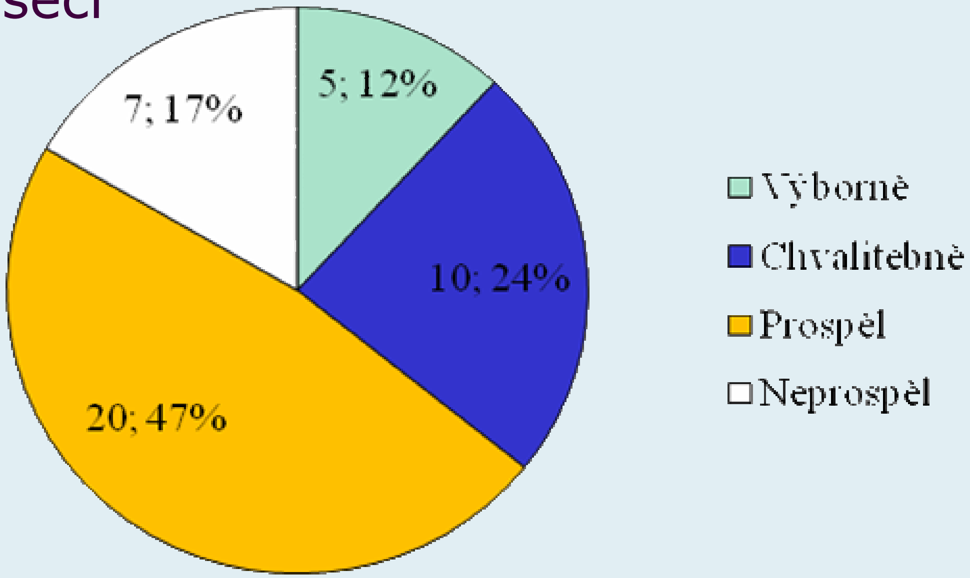
Grafické znázornění

- **Výsečový graf** – znázorňuje **relativní četnosti**: jednotlivé relativní četnosti jsou přímo úměrné obsahu příslušných kruhových výsečí



Grafické znázornění

- **Výsečový graf** – znázorňuje **relativní četnosti**: jednotlivé relativní četnosti jsou přímo úměrné obsahu příslušných kruhových výsečí



- Jednotlivé výseče musí být označeny relativními i absolutními četnostmi, neuvedení celkového počtu pozorování, by mohlo vést k matení (záměrnému nebo nechtěnému) toho, komu je graf určen.

1. Následující datový soubor představuje částečný výsledek zaznamenaný při průzkumu původu studentů studujících na VŠB-TU Ostrava a sice kraj trvalého bydliště.

   Data vyhodnoťte a graficky znázorněte.

   moravskoslezský, zlínský, olomoucký, moravskoslezský, pardubický, zlínský, zlínský, moravskoslezský, pardubický, moravskoslezský, olomoucký, moravskoslezský, pardubický, olomoucký, moravskoslezský, moravskoslezský, olomoucký, zlínský, olomoucký, moravskoslezský

Řešení:

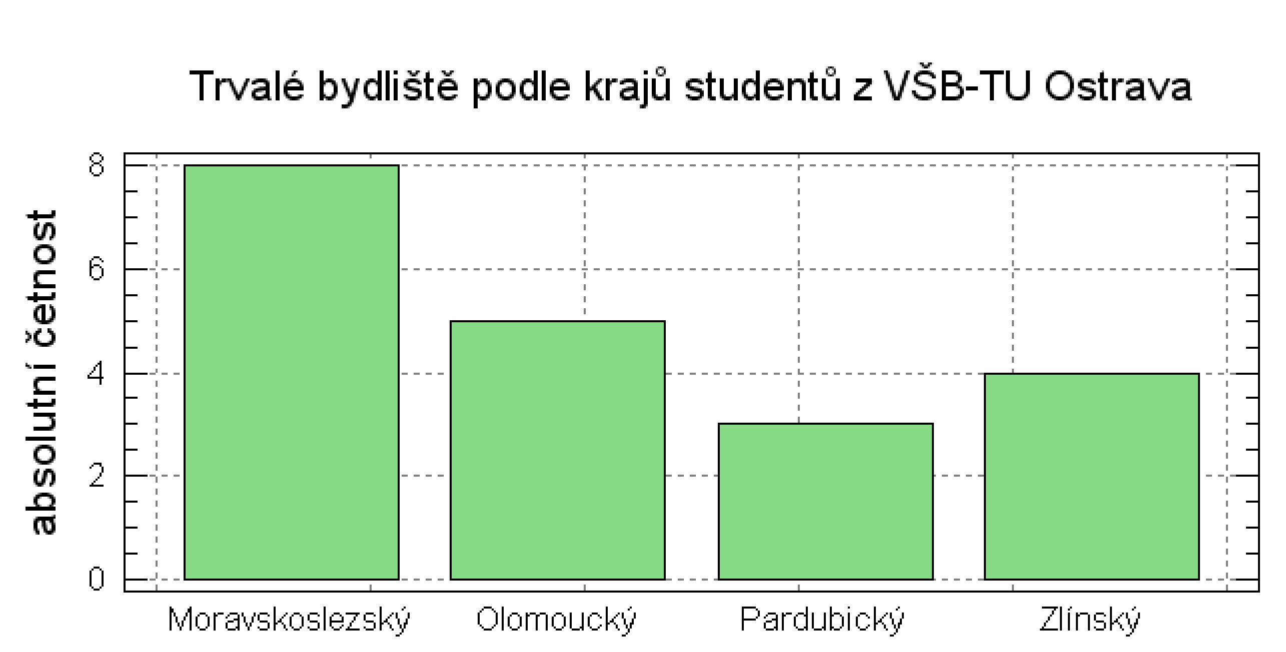
kvalitativní proměnná; kraje nemá smysl seřazovat => nominální proměnná

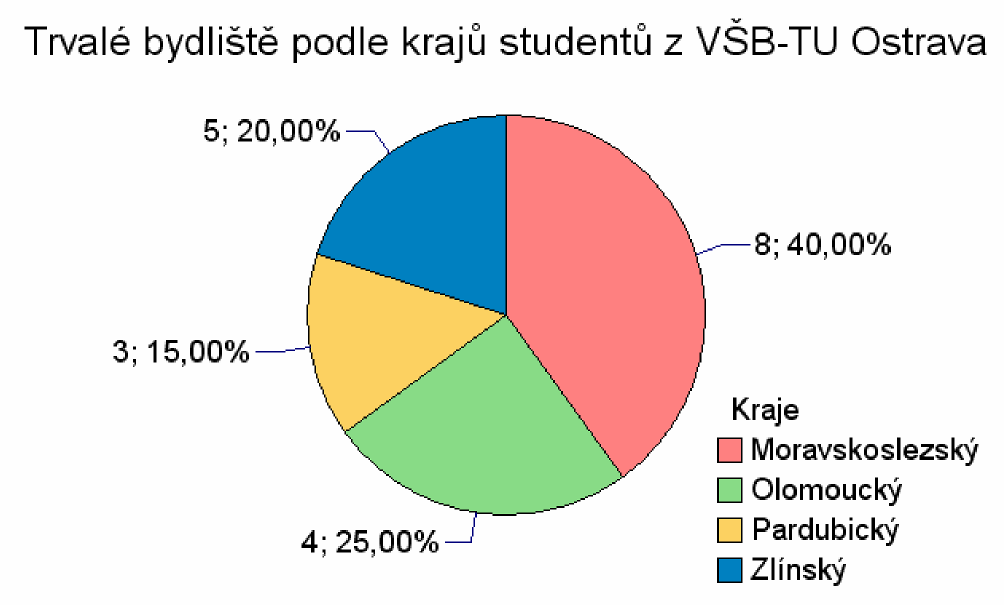
Pro popis tedy zvolíme tabulku četností, určíme modus a proměnnou znázorníme prostřednictvím histogramu a výsečového grafu.

Tabulka rozdělení četnosti původu studentů podle krajů

	Hodnoty x_i	**Absolutní četnosti** n_i		**Relativní četnosti** p_i
modus →	moravskoslezský	8	**max**	8/20 = 0,40
	zlínský	4		4/20 = 0,20
	olomoucký	5		5/20 = 0,25
	pardubický	3		3/20 = 0,15
	Celkem:	20		1

Řešení:


Trvalé bydliště podle krajů studentů z VŠB-TU Ostrava
absolutní četnost
0
2
4
6
8
Moravskoslezský
Olomoucký
Pardubický
Zlínský


Trvalé bydliště podle krajů studentů z VŠB-TU Ostrava
5; 20,00%
8; 40,00%
3; 15,00%
4; 25,00%
Kraje
Moravskoslezský
Olomoucký
Pardubický
Zlínský

Ordinální veličina

- nabývá v rámci souboru různých kategorií,
- varianty proměnné mají přirozené uspořádání, můžeme určit která je „menší“ a která „větší“
- Např. slovně vyjádřené známky ve škole (výborně, chvalitebně,…), velikosti oblečení vyjádřené písmeny (XS,S,M,L,XL,…), stupeň dosaženého vzdělání (základní, středoškolské bez maturity, středoškolské s maturitou),…

Číselné charakteristiky

- stejné charakteristiky jako pro nominální proměnnou: **absolutní četnost** n_i, **relativní četnost** p_i, **modus**
- **kumulativní četnost** $\boldsymbol{m_i}$ - počet hodnot proměnné, které nabývají varianty nižší nebo rovné *i*-té variantě.
- **relativní kumulativní četnost** $\boldsymbol{F_i}$ - vyjadřuje jakou část souboru tvoří hodnoty nabývající *i*-té a nižší varianty.

Tabulka rozdělení četnosti

Hodnoty x_i	**Absolutní četnosti** n_i	**Relativní četnosti** p_i	**Kumulativní četnosti** m_i	**Relativní kumulativní četnosti** F_i
x_1	n_1	$p_1 = n_1 / n$	n_1	p_1
x_2	n_2	$p_2 = n_2 / n$	$n_1 + n_2$	$p_1 + p_2$
…	…	…		
x_k	n_k	$p_k = n_k / n$	$n_1 + n_2 + \ldots + n_k = n$	$p_1 + p_2 + \ldots + p_k = 1$
Celkem:	$\sum_{i=1}^{k} n_i = n$	1		

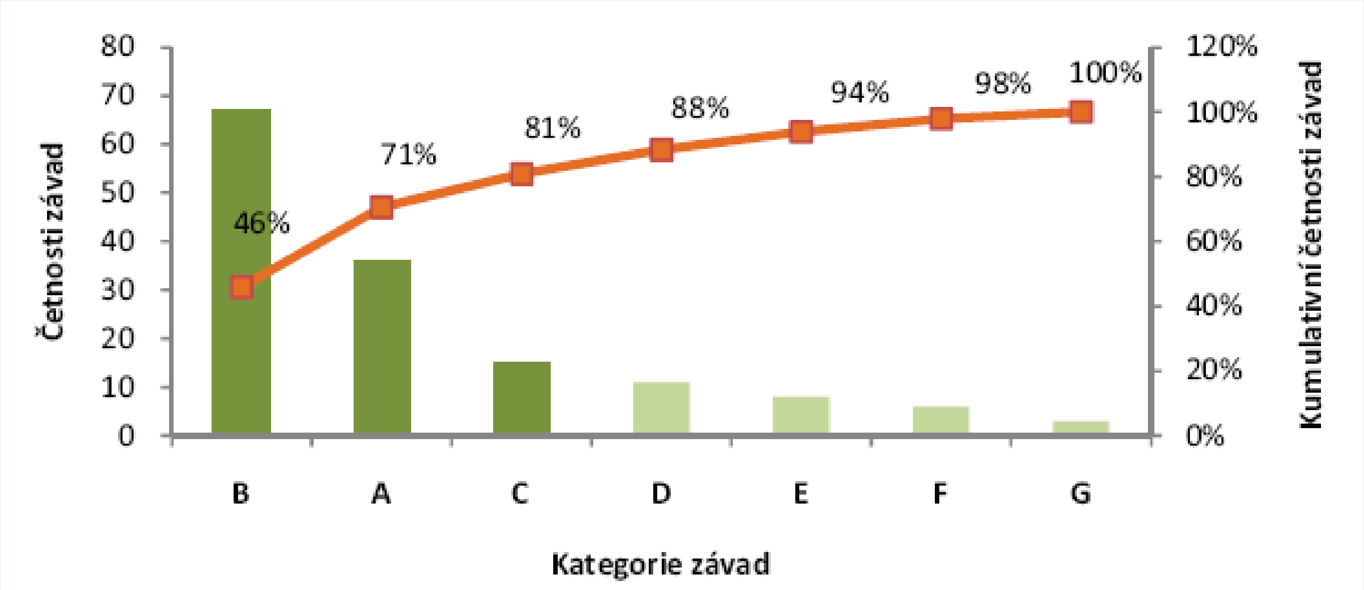
varianty proměnné jsou seřazeny od nejmenší po největší

Grafické znázornění

- **Histogram, výsečový graf**
- **Lorenzova křivka –** (polygon kumulativních četností, Galtonova ogiva, S křivka): spojnicový graf, na vodorovné ose jsou jednotlivé varianty proměnné v pořadí od „nejmenší" do „největší", na svislé ose příslušné hodnoty **kumulativních četností**.

Grafické znázornění

- **Paretův graf =** Lorenzova křivka + histogram se seřazenými sloupci od varianty s největší absolutní četností, po varantu s nejmenší absolutní četností



Paretovo pravidlo (80/20)

- V. Pareto vyvrátil základní rovnováhu mezi vynaloženým úsilím a následnou odměnou:
- **„20% všech našich činností přináší 80% zisku“** (v praxi taky 70/30, 99/1, …)
- „80% bohatství země je v rukou 20% lidí.“
- Řešený příklad:

  http://mi21.vsb.cz/flash-animace/paretova-analyza

- Využití:
  - výroba a služby
  - zajišťování kvality/jakosti (80% zmetků ve výrobě je způsobeno 20% příčin.)
  - ekonomie, management, marketing (80% zisku je vytvářeno 20% produktů.)
  - psychologie
  - sociologie (20% vašich přátel stojí za 80% vašeho zájmu)

2. Následující data představují výsledky zkoušky jedné skupiny ze statistiky roku 2011:

neprospěl, neprospěl, prospěl, výborný, prospěl, výborný, prospěl, neprospěl, výborný, prospěl, prospěl, neprospěl, prospěl, chvalitebný, neprospěl, prospěl.

a) Data vyhodnoťte a graficky znázorněte.

b) Určete kolik procent studentů dané skupiny při zkoušce uspělo.

Řešení:

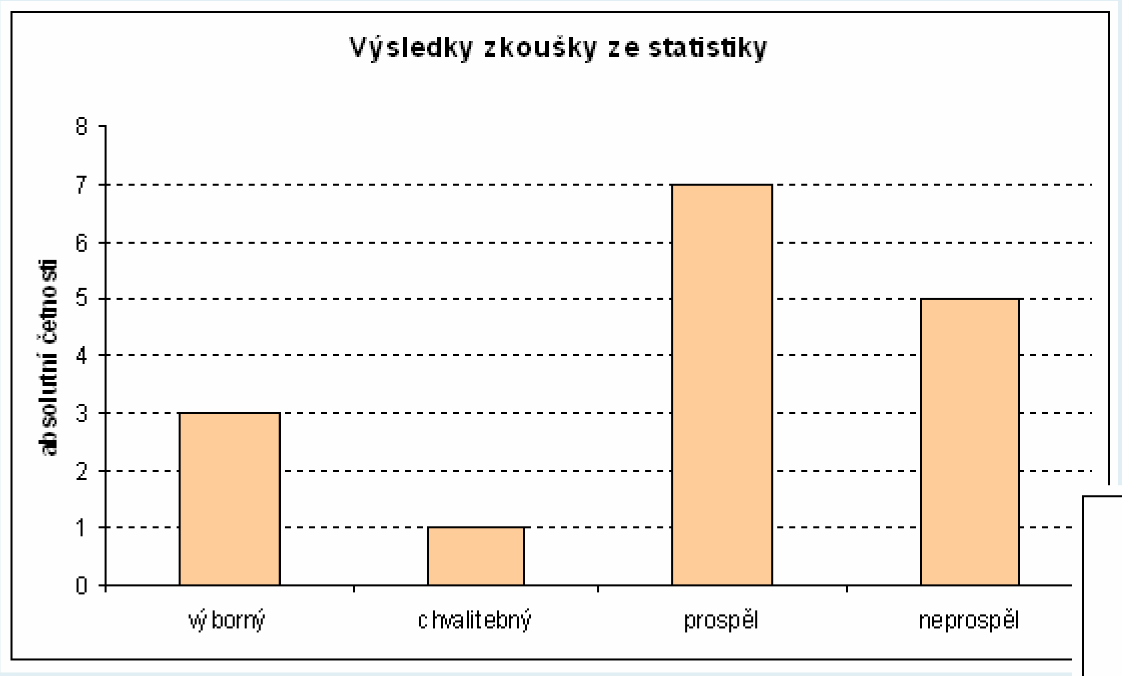
kvalitativní proměnná; známky lze seřadit =>ordinální proměnná

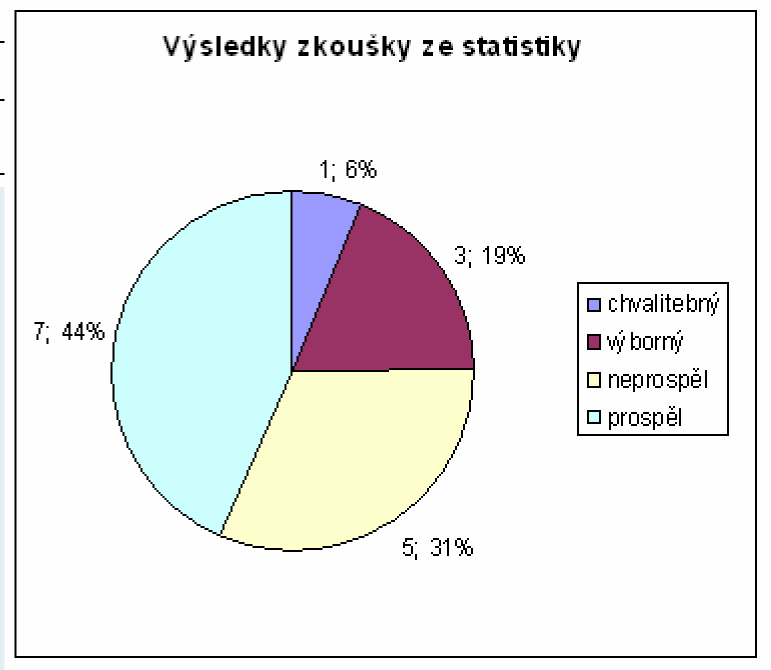
Při vyhodnocení mají smysl kumulativní charakteristiky

Tabulka rozdělení četnosti výsledků zkoušky ze statistiky				
Hodnoty x_i	**Absolutní četnosti** n_i	**Relativní četnosti** p_i	**Kumulativní četnosti** m_i	**Relativní kumulativní četnosti** F_i
výborný	3	3/16	3	3/16
chvalitebný	1	1/16	4	4/16
prospěl	7	7/16	11	11/16
neprospěl	5	5/16	16	16/16=1
Celkem:	16	1		

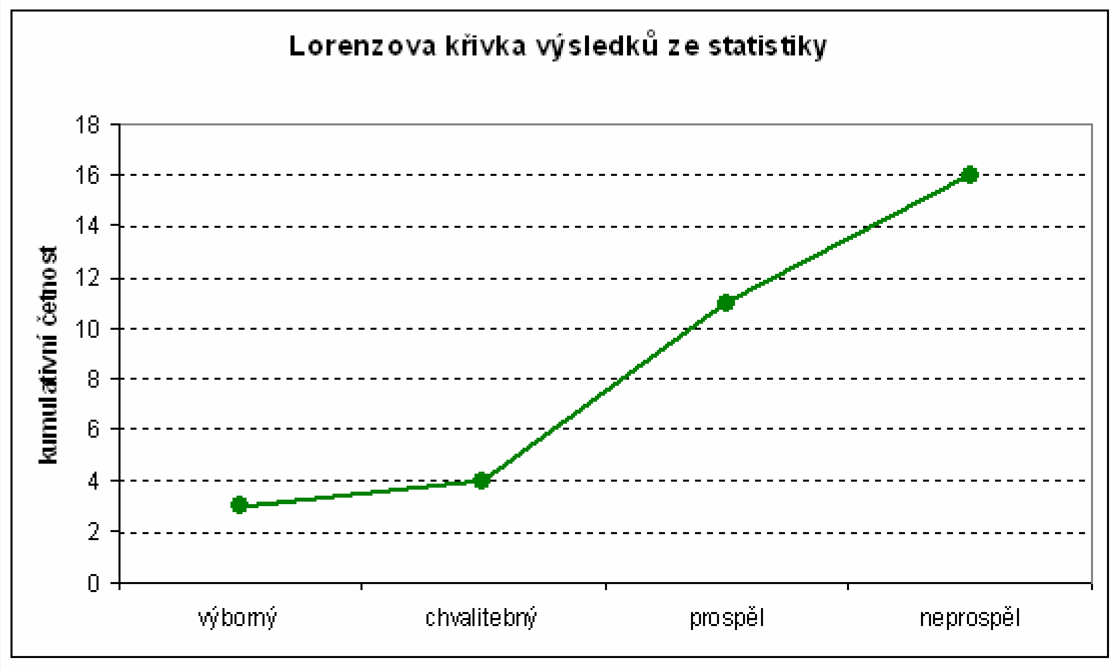
modus = prospěl

Řešení:

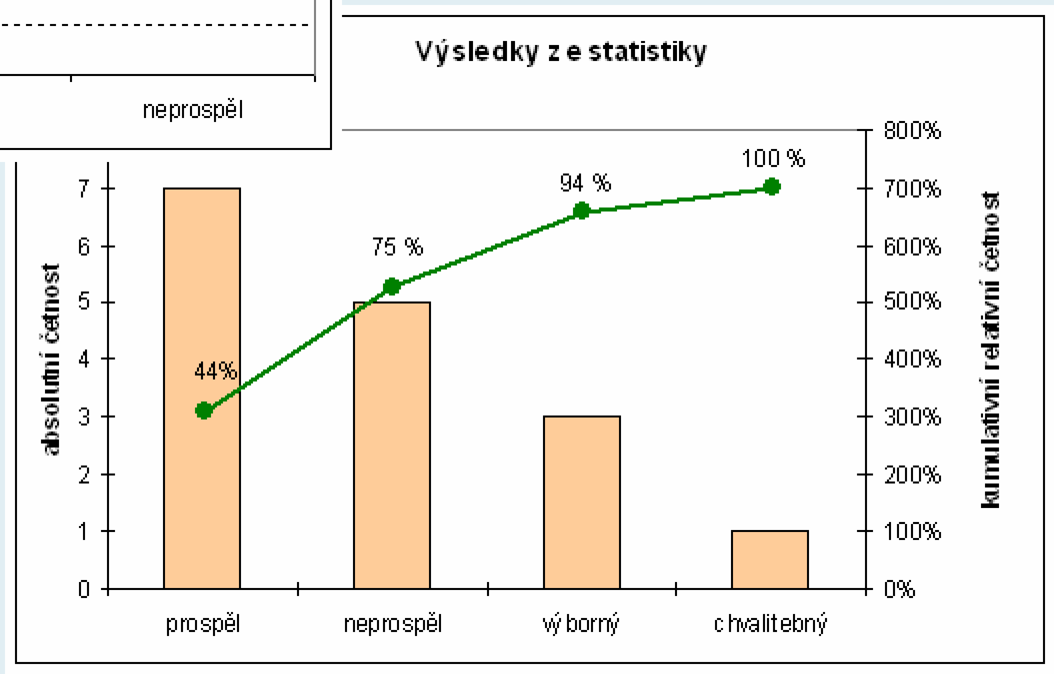

Výsledky zkoušky ze statistiky
absolutní četnosti
0
1
2
3
4
5
6
7
8
výborný
chvalitebný
prospěl
neprospěl


Výsledky zkoušky ze statistiky
1; 6%
3; 19%
7; 44%
5; 31%
chvalitebný
výborný
neprospěl
prospěl

Řešení:



Lorenzova křivka výsledků ze statistiky
kumulativní četnost
18
16
14
12
10
8
6
4
2
0
výborný
chvalitebný
prospěl
neprospěl



Výsledky ze statistiky
absolutní četnost
kumulativní relativní četnost
7
6
5
4
3
2
1
0
800%
700%
600%
500%
400%
300%
200%
100%
0%
44%
75 %
94 %
100 %
prospěl
neprospěl
výborný
chvalitebný

Řešení:

b) Kolik procent studentů dané skupiny u zkoušky uspělo ?
= relativní kumulativní četnost pro variantu prospěl =

$$11/16 \cdot 100\ \% = \underline{68{,}75\ \%}$$

U zkoušky ze statistiky uspělo 68,75 % studentů.

Kvantitativní (numerická) proměnná

- Pro popis používáme všechny charakteristiky ordinální proměnné +
- **Míry polohy** – určují typické rozložení hodnot proměnné (jejich rozmístění proměnných na číselné ose)
- **Míry variability** – určují rozptyl hodnot kolem své typické polohy

Míry polohy

- Aritmetický průměr:

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

x_i – hodnoty proměnné

n – rozsah výběru

- Vlastnosti:
  - součet odchylek od průměru je roven nule $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$
  - jestliže přičteme ke každé hodnotě dat. souboru konstantu, průměr se o tuto konstantu změní

$$\forall (a \in \Re): \left(\bar{x} = \frac{\sum_{i=!}^{n} x_i}{n} \Rightarrow \frac{\sum_{i=!}^{n} (a + x_i)}{n} = a + \bar{x} \right)$$

  - jestliže vynásobíme každou hodnotu dat. souboru konstantou, průměr se změní s násobkem této konstanty.

$$\forall (b \in \Re): \left(\bar{x} = \frac{\sum_{i=!}^{n} x_i}{n} \Rightarrow \frac{\sum_{i=!}^{n} (b x_i)}{n} = b\bar{x} \right)$$

Míry polohy

- Vážený průměr:

hodnoty proměnné

$$\overline{x} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \cdots + x_k \cdot n_k +}{n_1 + n_2 + \cdots + n_k} = \frac{\sum_{i=1}^{k} x_i \cdot n_i}{\sum_{i=1}^{k} n_i}$$

jednotlivé váhy

- Geometrický průměr:
    - Používáme pracujeme-li s kladnou proměnnou představující relativní změny (růstové indexy, cenové indexy...)

$$\overline{x_G} = \sqrt[n]{x_1 \cdot x_2 \cdot \cdots x_n}$$

- Vážený geometrický průměr:

$$\overline{x_G} = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \cdots x_k^{n_k}} \quad \text{, kde} \quad \sum_{i=1}^{k} n_i = n$$

Míry polohy

- Harmonický průměr:
  - používáme, jestliže proměnná má charakter části z celku (úlohy o společné práci...), proměnné vyjadřující čas na jednotku výkonu (poměrná čísla)

$$\overline{x_H} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

- Vážený harmonický průměr:

$$\overline{x_H} = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{n} \frac{n_i}{x_i}}$$

3. Vypočítejte průměrnou výšku žáků 9. tříd, výšky a počty žáků s danou výškou jsou uvedeny v následující tabulce:

x_i	160	165	170	175	180	185	190
n_i	9	20	36	82	35	14	4

Řešení:

Dosazením do vzorce pro aritmetický průměr:

$$\bar{x} = \frac{160 \cdot 9 + 165 \cdot 20 + 170 \cdot 36 + 175 \cdot 82 + 180 \cdot 35}{200} =$$

$$+ \frac{185 \cdot 14 + 190 \cdot 4}{200} = \underline{\underline{174{,}3}}$$

Průměrná výška žáků 9. tříd je 174,3 cm.

4. Ve škole jsou čtyři 6. třídy, označené A, B, C, D; počty žáků a průměrné známky z matematiky jsou uvedeny v tabulce. Určete průměrnou známku z matematiky ve všech 6. třídách dohromady.

Třída	A	B	C	D
Průměrná známka	2,21	1,82	2,33	2,11
Počet žáků	28	24	32	30

Řešení:

Počty žáků ve třídách určují váhy příslušné danému průměru.

$$\bar{x} = \frac{2{,}21 \cdot 28 + 1{,}82 \cdot 24 + 2{,}33 \cdot 32 + 2{,}11 \cdot 30}{28 + 24 + 32 + 30} = \underline{\underline{2{,}14}}$$

Průměrná známka z matematiky v 6. třídách je 2,14.

5. V rozvojové zemi došlo během posledních 12 měsíců k prudkému vzrůstu inflace. V tabulce je pro každý měsíc uveden index spotřebitelských cen v procentech proti minulému měsíci:

I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
102	104	104	104	110	120	130	140	150	160	170	180

a) Jaké bylo průměrné měsíční tempo růstu cen?

b) Kolikrát vyšší byly ceny v prosinci ve srovnání s prosincem minulého roku?

Řešení:

a) průměrné tempo růstu za jedno období vyjadřuje geometrický průměr:

$$\overline{x_G} = \sqrt[12]{102 \cdot 104 \cdot 104 \cdot 104 \cdot 110 \cdot 120 \cdot 130 \cdot 140}$$
$$\cdot \sqrt[12]{150 \cdot 160 \cdot 170 \cdot 180} = \underline{\underline{128{,}5}}$$

Průměrné měsíční tempo růstu cen je 128,5 %.

Řešení:

b)

měsíc	koeficient růstu	relativní přírůstek [%]	
XII_p			
I	102	$I / XII_p = 1{,}02$	$I = 1{,}02 \cdot XXI$
II	104	$II / I = 1{,}04$	$II = 1{,}04 \cdot I$
…	…	…	…
XI	170	$XI / X = 1{,}7$	$XI = 1{,}7 \cdot X$
XII	180	$XII / XI = 1{,}8$	$XII = 1{,}8 \cdot XI$

prosinec předešlého roku

$$XII_p \cdot k = XII$$

prosinec letošního roku

$$k = ?$$

$$k \cdot XII_p = XII = 1{,}8 \cdot XI = 1{,}8 \cdot 1{,}7 \cdot X = 1{,}8 \cdot 1{,}7 \cdot \dots \cdot 1{,}02 \cdot XII_p =$$

$$= 20{,}24 \cdot XII_p$$

$$k \cdot XII_p = 20{,}24 \cdot XII_p \quad \Rightarrow \quad k = 20{,}24$$

Řešení:

b)

měsíc	koeficient růstu	relativní přírůstek [%]	
XII_p			
I	102	$I / XII_p = 1{,}02$	$I = 1{,}02 \cdot XXI$
II	104	$II / I = 1{,}04$	$II = 1{,}04 \cdot I$
…	…	…	…
XI	170	$XI / X = 1{,}7$	$XI = 1{,}7 \cdot X$
XII	180	$XII / XI = 1{,}8$	$XII = 1{,}8 \cdot XI$

prosinec předešlého roku

$$XII_p \cdot k = XII$$

prosinec letošního roku

$$k = ?$$

$$k \cdot XII_p = XII = 1{,}8 \cdot XI = 1{,}8 \cdot 1{,}7 \cdot X = 1{,}8 \cdot 1{,}7 \cdot \cdots \cdot 1{,}02 \cdot XII_p = $$

$$= 20{,}24 \cdot XII_p$$

$$k \cdot XII_p = 20{,}24 \cdot XII_p \quad \Rightarrow \quad k = 20{,}24$$

V prosinci letošního roku byly ceny oproti roku minulému vyšší více než dvacetkrát.

6. Stezkou, která vede na vrchol hory, vystupuje turista rychlostí 2,5 km·hod^{-1}, sestupuje rychlostí 5 km·hod^{-1}. Jakou průměrnou rychlostí jde, jestliže:

   a) se vrací stejnou cestou?

   b) se vrací cestou, která je 2x delší?

Řešení:

a) vzdálenost nahoru i dolů je stejná

	nahoru	dolů	celkem
dráha [km]	x	x	2x
rychlost [km·hod^{-1}]	2,5	5	?

proměnná má charakter části z celku => harmonický průměr, části jsou stejně velké

$$\overline{x}_H = \frac{2}{\frac{1}{2,5}+\frac{1}{5}} = \underline{\underline{3,3}}$$

Průměrná rychlost turisty je 3,3 km·hod^{-1}.

Řešení:

b) vzdálenost dolů je dvakrát delší než nahoru

	nahoru	dolů	celkem
dráha [km]	x	2x	3x
rychlost [km·hod^{-1}]	2,5	5	?

proměnná má charakter části z celku => harmonický průměr, části nejsou stejně velké => vážený harmonický průměr

$$\bar{x}_H = \frac{3}{\frac{1}{2,5} + \frac{2}{5}} = \underline{\underline{3,75}}$$

Průměrná rychlost turisty je 3,75 km·hod^{-1}.

Míry polohy

- Kvantily: varianty proměnné rozdělující výběrový soubor

  100*p* %-ní kvantil x_p

  odděluje 100*p*% menších hodnot od zbytku souboru (100*p*% hodnot datového souboru je menších než toto číslo.)
- Nejvýznamnější kvantily:
  - kvartily
    - dolní kvartil - $\tilde{x}_{0,25}$, medián - $\tilde{x}_{0,5}$, horní kvartil - $\tilde{x}_{0,75}$
  - decily - $\tilde{x}_{0,1}; \tilde{x}_{0,2}; \cdots; \tilde{x}_{0,9}$
  - percentily - $\tilde{x}_{0,01}; \tilde{x}_{0,02}; \cdots; \tilde{x}_{0,99}$

Míry polohy – určování kvantilů

1. Výběrový soubor uspořádáme podle velikosti
2. Jednotlivým hodnotám proměnné přiřadíme pořadí, a to tak, že nejmenší hodnota bude mít pořadí 1 a nejvyšší hodnota pořadí *n* (rozsah souboru)
3. 100*p*%- ní kvantil je roven hodnotě proměnné s pořadím z_p, kde: $z_p = np + 0{,}5$
4. Není-li z_p celé číslo, pak daný kvantil určíme jako průměr prvků s pořadím $\lfloor z_p \rfloor$ a $\lfloor z_p \rfloor + 1$.

Míry polohy

- Interkvartilové rozpětí (IQR): $\tilde{x}_{0,75} - \tilde{x}_{0,25}$
  využití při identifikaci odlehlých pozorování
- MAD – medián absolutních odchylek od mediánu
  využití při identifikaci odlehlých pozorování
- Short – nejkratší interval, ve kterém leží alespoň 50 % hodnot proměnné
- Modus – střed shortu

Odlehlá pozorování

- Takové hodnoty proměnné. které se výrazně liší od ostatních hodnot v souboru, ovlivňují např. vypovídací hodnotu průměru

- Co s tím?
  - Odlehlost pozorování je způsobena:
    - hrubými chybami, překlepy, prokazatelným selháním lidí či techniky ...
    - důsledky poruch, chybného měření, technologických chyb ...

  jestliže známe příčinu odlehlosti a předpokládáme, že již nenastane, můžeme tato pozorování vyloučit z dalšího zpracování (vymazat).

  - V ostatních případech je nutno zvážit, zda se vyloučením odlehlých pozorování nepřipravíme o důležité informace o jevech vyskytujících se s nízkou četností.

Identifikace odlehlých pozorování

- Metoda vnitřních hradeb: 1,5 násobek IQR

$$\left[\left(x_i < x_{0,25} - 1{,}5 \cdot IQR\right) \vee \left(x_i > x_{0,75} + 1{,}5 \cdot IQR\right)\right] \Rightarrow$$

x_i je odlehlým pozorováním

- Metoda vnějších hradeb: 3 násobek IQR – identifikuje extrémní pozorování

$$\left[\left(x_i < x_{0,25} - 3 \cdot IQR\right) \vee \left(x_i > x_{0,75} + 3 \cdot IQR\right)\right] \Rightarrow$$

x_i je odlehlým pozorováním

Identifikace odlehlých pozorování

- Z-souřadnice

$$z - sou\check{r}._i = \frac{x_i - \bar{x}}{s}$$

jinak přepsané pravidlo 6σ

$$\left(\left|z - sou\check{r}._i\right| > 3\right) \Rightarrow x_i \text{ je odlehlým pozorováním}$$

- Mediánová souřadnice

$$medi\acute{a}nov\acute{a}\ sou\check{r}._i = \frac{x_i - x_{0,5}}{1{,}483.MAD}$$

$$\left(\left|medi\acute{a}nov\acute{a}\ sou\check{r}._i\right| > 3\right) \Rightarrow x_i \text{ je odlehlým pozorováním}$$

Míry variability

- Výběrový rozptyl:

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

rozměr výběrového rozptylu charakteristiky je kvadrátem rozměru charakteristiky

- Vlastnosti výběrového rozptylu:
  - výběrový rozptyl konstanty je roven nule, tzn. jestliže jsou všechny hodnoty proměnné stejné, soubor má nulovou rozptýlenost,
  - jestliže přičteme ke všem hodnotám proměnné stejnou konstantu, rozptyl se nezmění,
  - jestliže vynásobíme všechny hodnoty proměnné konstantou, výběrový rozptyl se změní se zvětší s kvadrátem této konstanty.

Míry variability

- Výběrová směrodatná odchylka:

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

  rozměr výběrové směrodatné odchylky je stejný jako rozměr charakteristiky
- Výběrový rozptyl ani výběrová směrodatná odchylka neumožňují srovnávat rozptýlenost proměnných s jinými jednotkami => zavádí se variační koeficient

- Variační koeficient: $V_x = \frac{s}{\overline{x}} \cdot 100\ [\%]$

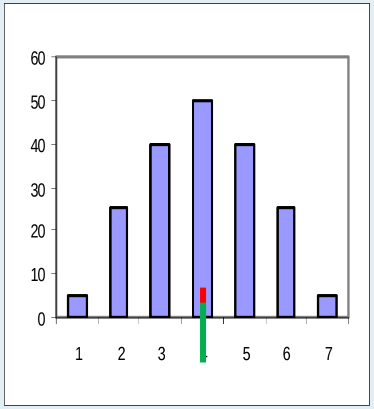
  bezrozměrná veličina, čím nižší var. koeficient, tím homogennější soubor, V_x>50% značí silně rozptýlený soubor.

Míry šikmosti a špičatosti

- Výběrová šikmost:

$$a = \frac{n}{(n-1)(n-2)} \cdot \frac{\sum_{i=1}^{n} (x_i - \overline{x})^3}{s^3}$$

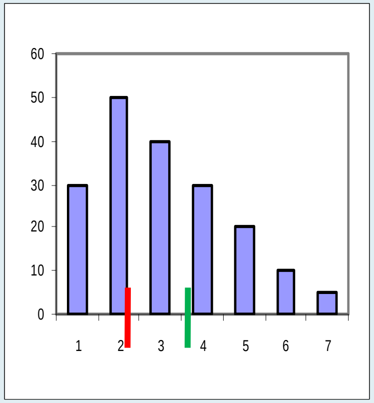
Symetrická data
$a = 0$



Průměr = medián

Polovina dat.souboru je menší než průměr

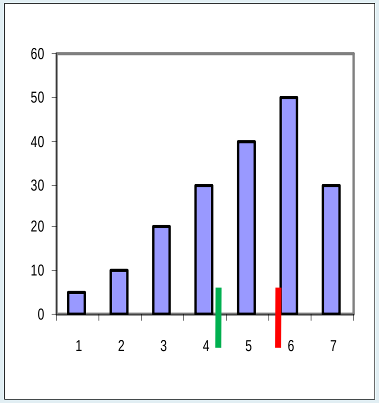
Pozitivně zešikmená data
$a > 0$



Medián < průměr

Nadpoloviční většina dat.souboru je menší než průměr

Negativně zešikmená data
$a < 0$



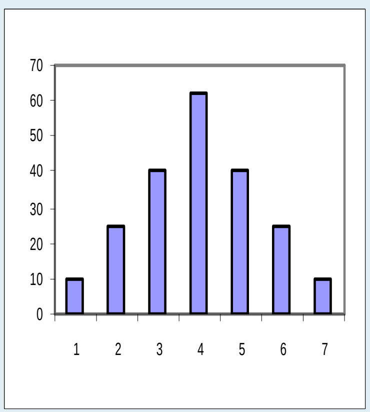
Průměr < medián

Nadpoloviční většina dat.souboru je větší než průměr

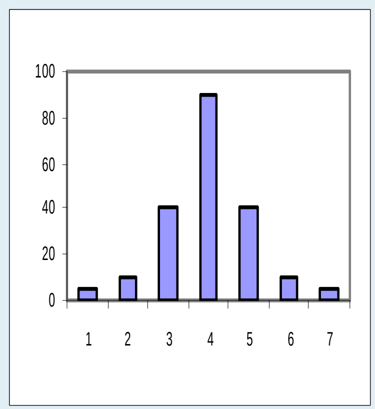
Míry šikmosti a špičatosti

- Výběrová špičatost:

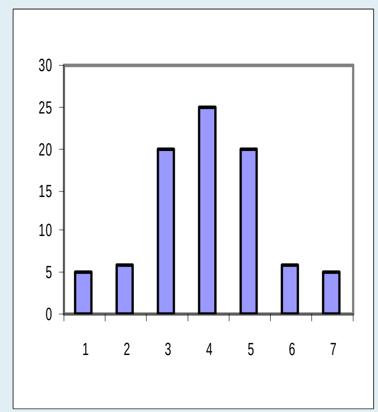
$$b = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \cdot \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4} - 3\frac{(n-1)^2}{(n-2)(n-3)}$$



b=0



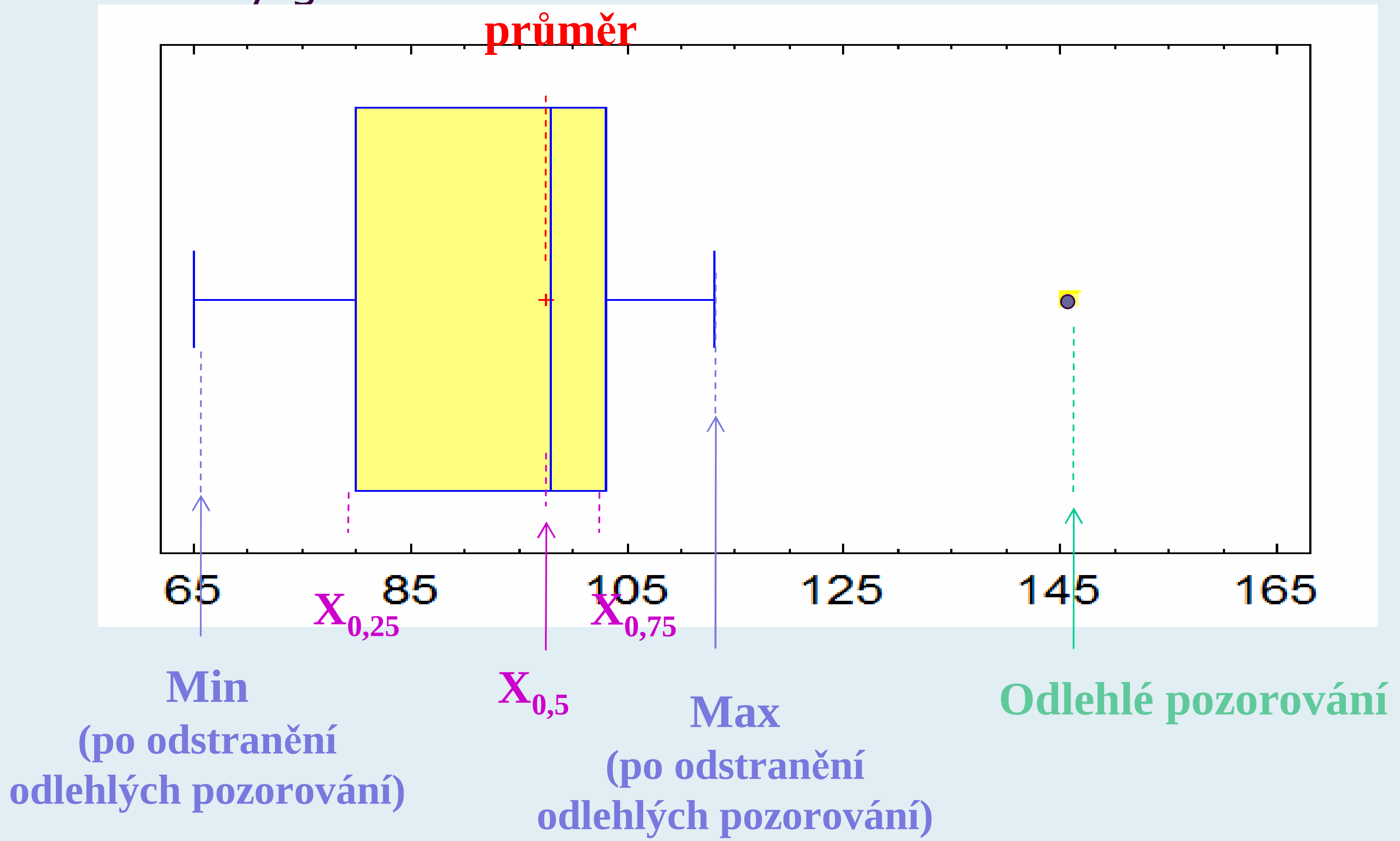
b>0



b<0

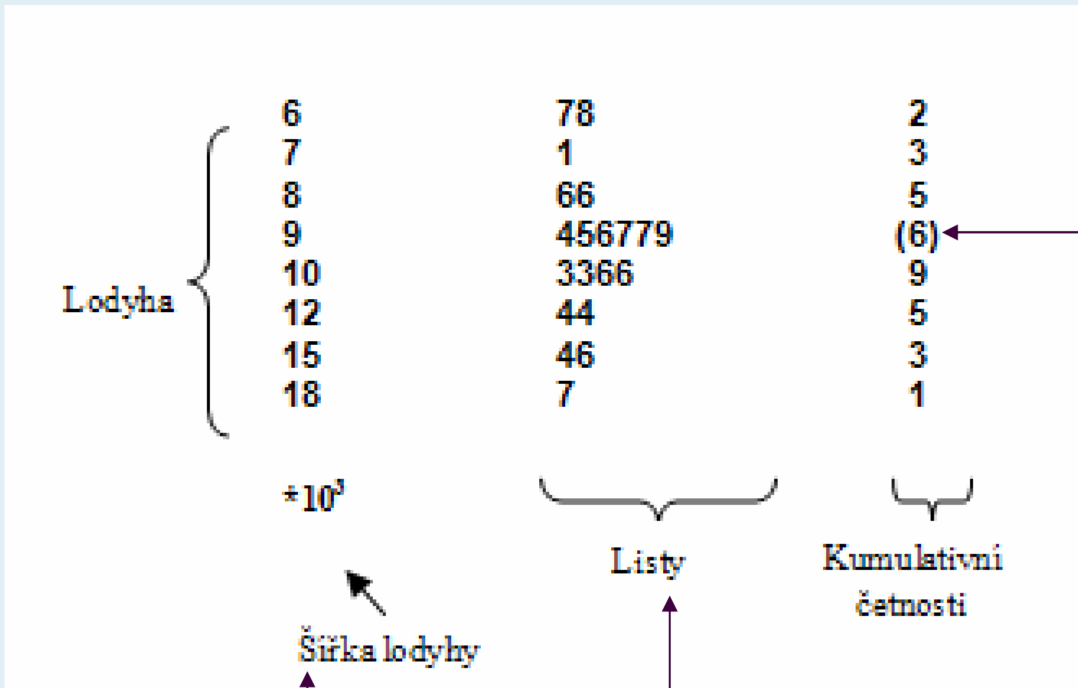
Grafické znázornění

- Krabicový graf



Grafické znázornění

- Číslicový histogram



v řádku, kde je medián se uvádí absolutní četnost v závorce

řád o jednu vyšší než

zvolený, důležitý řád

Test

1. Test ze Statistiky píše velké množství studentů. Představte si, že každý z nich odpoví správně přesně na polovinu otázek. V tomto případě bude směrodatná odchylka počtu správných odpovědí

a) rovna průměru,
b) rovna mediánu,
c) rovna nule,
d) směrodatnou odchylku nelze určit bez dalších informací,
e) dvojnásobku módu.

1. Test ze Statistiky píše velké množství studentů. Představte si, že každý z nich odpoví správně přesně na polovinu otázek. V tomto případě bude směrodatná odchylka počtu správných odpovědí

a) rovna průměru,
b) rovna mediánu,
c) rovna nule,
d) směrodatnou odchylku nelze určit bez dalších informací,
e) dvojnásobku módu.

PROČ?

Vycházíme z vlastnosti výběrového rozptylu: rozptyl konstantního souboru je roven nule (jestliže jsou všechny hodnoty proměnné stejné, má soubor nulovou rozptýlenost), směrodatná odchylka, definovaná jako kladná odmocnina z výběrového rozptylu, je potom odmocnina z nuly => nula.

2. Největší kumulativní absolutní četnost v množině čísel se rovná

a) součtu všech absolutních četností,
b) 1,
c) dvojnásobku průměru,
d) dvojnásobku mediánu,
e) dvojnásobku módu.

2. Největší kumulativní absolutní četnost v množině čísel se rovná

a) součtu všech absolutních četností,
b) 1,
c) dvojnásobku průměru,
d) dvojnásobku mediánu,
e) dvojnásobku módu.

PROČ?

Kumulativní četnost m_i je definována jako počet hodnot proměnné, které nabývají varianty nižší, nebo rovné i-té variantě. Jsou-li jednotlivé varianty uspořádány podle „velikosti“ („$x_1 < x_2 < \ldots < x_k$“), platí: $m_i = \sum_{j=1}^{i} n_j$, odtud kumulativní četnost k-té („nejvyšší “) varianty je rovna rozsahu proměnné $m_k = n$.

3. Několik studentů píše test ze Statistiky s 10-ti otázkami. Nejhorší výsledek jsou 3 správné odpovědi, nejlepší výsledek je 10 správných odpovědí. Jakou hodnotu má medián?

a) 7 (=10-3)
b) 6,5 (=(3+10)/2)
c) Medián nelze určit, pokud neznáme konkrétní výsledky jednotlivých žáků.

3. Několik studentů píše test ze Statistiky s 10-ti otázkami. Nejhorší výsledek jsou 3 správné odpovědi, nejlepší výsledek je 10 správných odpovědí. Jakou hodnotu má medián?

a) 7 (=10-3)

b) 6,5 (=(3+10)/2)

c) Medián nelze určit, pokud neznáme konkrétní výsledky jednotlivých žáků.

PROČ?

Medián $x_{0,5}$ = 50%-ní kvantil (rozděluje datový soubor tak, že polovina (50%) hodnot je menších než medián a polovina (50%) je větších (nebo rovných)).
K určení kvantilů potřebujeme uspořádat výběrový soubor podle velikosti, což není možné, pokud neznáme jednotlivé hodnoty.

4. Představte si, že jste absolvovali normovaný test (např. SCIO test) a že Vám sdělili, že patříte do 91. percentilu. To znamená, že

a) 90 žáků, kteří se podrobili stejnému testu, dosáhlo lepších výsledků než vy.
b) 90 žáků, kteří se podrobili stejnému testu, dosáhlo horších výsledků než vy.
c) 90% žáků, kteří se podrobili stejnému testu, dosáhlo lepších výsledků než vy.
d) 90% žáků, kteří se podrobili stejnému testu, dosáhlo horších výsledků než vy.

4. Představte si, že jste absolvovali normovaný test (např. SCIO test) a že Vám sdělili, že patříte do 91. percentilu. To znamená, že

~~a)~~ 90 žáků, kteří se podrobili stejnému testu, dosáhlo lepších výsledků než vy.

~~b)~~ 90 žáků, kteří se podrobili stejnému testu, dosáhlo horších výsledků než vy.

~~c)~~ 90% žáků, kteří se podrobili stejnému testu, dosáhlo lepších výsledků než vy.

d) 90% žáků, kteří se podrobili stejnému testu, dosáhlo horších výsledků než vy.

PROČ?

Zejména v souvislosti s hodnocením normovaných testů (SCIO testy, biometrické normy, …) se často setkáváme s vyjádřením „Patříte do *p*. percentilu“, přičemž *p* je celé číslo mezi 1 a 100. Je tím myšleno, že nejméně (p-1)% a zároveň méně než p% účastníků testu dosáhlo nižšího hodnocení než vy. (Tzn. „Patříte do 91. percentilu“ znamená, že nejméně 90% (a nejvýše 91%) účastníku testu dosáhlo nižšího výsledku než vy.)

5. Průměrná mzda je 60% kvantil mzdy. Lze tedy říci, že

a) medián mzdy je vyšší než průměrná mzda,
b) medián mzdy je nižší než průměrná mzda,
c) medián mzdy je stejný jako průměrná mzda,
d) o vztahu mezi mediánem mzdy a průměrnou mzdou nelze rozhodnout.

5. Průměrná mzda je 60% kvantil mzdy. Lze tedy říci, že

a) medián mzdy je vyšší než průměrná mzda,
b) medián mzdy je nižší než průměrná mzda,
c) medián mzdy je stejný jako průměrná mzda,
d) o vztahu mezi mediánem mzdy a průměrnou mzdou nelze rozhodnout.

PROČ?

Medián $x_{0,5}$ = 50%-ní kvantil - 50% hodnot je menších než medián

Průměrná mzda $x_{0,6}$ = 60%-ní kvantil - 60% hodnot je menších než medián

6. Průměrná mzda je 60% kvantil mzdy. Lze tedy říci, že

a) mzdy mají kladnou šikmost,
b) mzdy mají zápornou šikmost,
c) mzdy mají kladnou špičatost,
d) mzdy mají zápornou špičatost,
e) vztah mezi průměrem a 60% kvantilem nevypovídá nic o šikmosti ani o špičatosti dat.

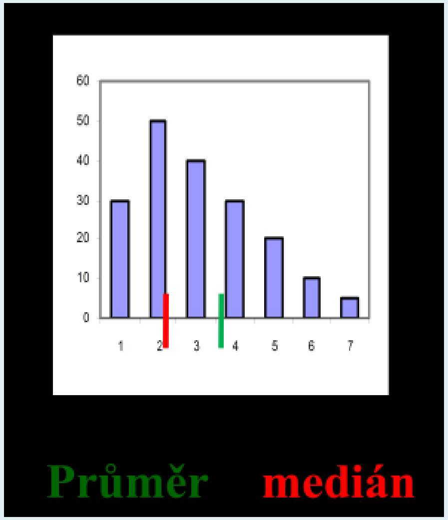
6. Průměrná mzda je 60% kvantil mzdy. Lze tedy říci, že

a) mzdy mají kladnou šikmost,
~~b)~~ mzdy mají zápornou šikmost,
~~c)~~ mzdy mají kladnou špičatost,
~~d)~~ mzdy mají zápornou špičatost,
~~e)~~ vztah mezi průměrem a 60% kvantilem nevypovídá nic o šikmosti ani o špičatosti dat.

PROČ?

$$\bar{x} = x_{0,6} > x_{0,5}$$

$a > 0$ ... u proměnné převažují hodnoty menší než průměr,



7. Lékař Petře sdělil, že patří do 3. percentilu ohledně BMI (Body mass index – poměr váhy (kg) ke kvadrátu výšky (m)). Petra má pravděpodobně

a) podváhu,
b) normální váhu,
c) nadváhu,
d) Bez dalších informací nelze usuzovat na Petřinu váhu.

7. Lékař Petře sdělil, že patří do 3. percentilu ohledně BMI (Body mass index – poměr váhy (kg) ke kvadrátu výšky (m)). Petra má pravděpodobně

a) podváhu,
b) normální váhu,
c) nadváhu,
d) Bez dalších informací nelze usuzovat na Petřinu váhu.

PROČ?

3. percentil – $x_{0,03}$ – rozděluje datový soubor tak, že 3% hodnot je menších než tento percentil a zbytek, tj. 97% větších (nebo rovných), tzn. že pouze 3% lidí mají nižší BMI než Petra.

8. Představte si, že jste absolvovali normovaný test (např. SCIO test). Měl(a) jste lepší výsledek než 85 studentů ze 100. To znamená, že

a) patříte do 99. decilu,
b) patříte do 95. decilu,
c) patříte do 10. decilu,
d) patříte do 9. decilu,
e) patříte do 2. kvartilu.

8. Představte si, že jste absolvovali normovaný test (např. SCIO test). Měl(a) jste lepší výsledek než 85 studentů ze 100. To znamená, že

a) patříte do 99. decilu,
b) patříte do 95. decilu,
c) patříte do 10. decilu,
d) patříte do 9. decilu,
e) patříte do 2. kvartilu.

PROČ?

Decily dělí výběrový soubor na 10 přibližně stejně četných částí. Výsledek lepší než 85 studentů ze 100 => 85% studentů má horší výsledek => 9. decil (zahrnuje 80% - 90%)

9. Pro srovnání variability váhy a výšky je možné použít

a) průměr,
b) rozptyl,
c) směrodatnou odchylku,
d) variační koeficient,
e) šikmost.

9. Pro srovnání variability váhy a výšky je možné použít

a) průměr,
b) rozptyl,
c) směrodatnou odchylku,
d) variační koeficient,
e) šikmost.

PROČ?

Nevýhodou výběrového rozptylu i výběrové směrodatné odchylky je skutečnost, že neumožňují porovnávat variabilitu proměnných vyjádřených v různých jednotkách (nejsou to bezrozměrné veličiny). Variační koeficient V_X vyjadřuje relativní míru variability proměnné x (je bezrozměrný). Průměr a šikmost necharakterizují variabilitu dat(variabilitu charakterizují: rozptyl, směrodatná odchylka, variační koeficient, short, interkvartilové rozpětí).

10. Zvýšíme-li každému zaměstnanci ve firmě plat o 100,- Kč, průměrný plat ve firmě se zvýší

a) o 100,- Kč,
b) o 1000,- Kč,
c) průměrný plat se nezmění.

10. Zvýšíme-li každému zaměstnanci ve firmě plat o 100,- Kč, průměrný plat ve firmě se zvýší

a) o 100,- Kč,
~~b)~~ o 1000,- Kč,
~~c)~~ průměrný plat se nezmění.

PROČ?

Z vlastnosti průměru:

$$\forall\left(a \in \Re\right):\left(\bar{x} = \frac{\sum_{i=!}^{n} x_i}{n} \Rightarrow \frac{\sum_{i=!}^{n} \left(a + x_i\right)}{n} = a + \bar{x}\right)$$

Přičteme-li ke každé hodnotě datového souboru konstantu, průměr se o tuto konstantu změní.

11. Zvýšíme-li každému zaměstnanci ve firmě plat dvojnásobně, průměrný plat ve firmě se zvýší

a) dvojnásobně,
b) čtyřnásobně,
c) průměrný plat se nezmění.

11. Zvýšíme-li každému zaměstnanci ve firmě plat dvojnásobně, průměrný plat ve firmě se zvýší

a) dvojnásobně,
~~b)~~ čtyřnásobně,
~~c)~~ průměrný plat se nezmění.

PROČ?

Z vlastnosti průměru:

$$\forall\left(b \in \Re\right):\left(\bar{x}=\frac{\sum_{i=!}^{n} x_i}{n} \Rightarrow \frac{\sum_{i=!}^{n}\left(b x_i\right)}{n}=b\bar{x}\right)$$

Vynásobíme-li každou hodnotu datového souboru konstantou, průměr se změní také s násobkem této konstanty.

12. Zvýšíme-li každému zaměstnanci ve firmě plat o 20%, průměrný plat ve firmě se zvýší

a) o 20%,
b) o 400%,
c) o 40%,
d) o 44%,
e) průměrný plat se nezmění.

12. Zvýšíme-li každému zaměstnanci ve firmě plat o 20%, průměrný plat ve firmě se zvýší

a) o 20%,
~~b)~~ o 400%,
~~c)~~ o 40%,
~~d)~~ o 44%,
~~e)~~ průměrný plat se nezmění.

PROČ?

Jestliže zvyšujeme každý plat ve firmě o 20% => každý plat násobíme 1,2 krát.

Z vlastnosti průměru $\overline{x} = \frac{\sum_{i=!}^{n} x_i}{n} \Rightarrow \frac{\sum_{i=!}^{n} (1,2 \cdot x_i)}{n} = 1,2 \cdot \overline{x}$, průměrný plat ve firmě stoupne 1,2 krát, tzn. o 20%.

13. Zvýšíme-li každému zaměstnanci ve firmě plat o 100,- Kč, rozptyl platů ve firmě se zvýší

a) o 100,- Kč,
b) o 1000,- Kč,
c) rozptyl platů se nezmění.

13. Zvýšíme-li každému zaměstnanci ve firmě plat o 100,- Kč, rozptyl platů ve firmě se zvýší

a) o 100,- Kč,
b) o 1000,- Kč,
c) rozptyl platů se nezmění.

PROČ?

Z vlastnosti výběrového rozptylu:

Přičteme-li ke všem hodnotám proměnné konstantu, výběrový rozptyl se nezmění.

14. Zvýšíme-li každému zaměstnanci ve firmě plat dvojnásobně, rozptyl platů ve firmě se zvýší

a) dvojnásobně,
b) čtyřnásobně,
c) rozptyl platů se nezmění.

14. Zvýšíme-li každému zaměstnanci ve firmě plat dvojnásobně, rozptyl platů ve firmě se zvýší

~~a)~~ dvojnásobně,
b) čtyřnásobně,
~~c)~~ rozptyl platů se nezmění.

PROČ?

Z vlastnosti výběrového rozptylu:

Vynásobíme-li všechny hodnoty proměnné konstantou, výběrový rozptyl se zvětší kvadrátem této konstanty (b^2 krát).

15. Zvýšíme-li každému zaměstnanci ve firmě plat o 20%, rozptyl platů ve firmě se zvýší

a) o 20%,
b) o 400%,
c) o 40%,
d) o 44%,
e) rozptyl platů se nezmění.

15. Zvýšíme-li každému zaměstnanci ve firmě plat o 20%, rozptyl platů ve firmě se zvýší

a) o 20%,
b) o 400%,
c) o 40%,
d) o 44%,
e) rozptyl platů se nezmění.

PROČ?

Jestliže zvyšujeme každý plat ve firmě o 20% => každý plat násobíme 1,2 krát.
Z vlastnosti výběrového rozptylu o násobení konstantou víme, že pokud vynásobíme všechny hodnoty proměnné 1,2 krát, výběrový rozptyl se zvětší kvadrátem této konstanty = 1,44 krát => průměrný plat ve firmě stoupne 1,44 krát, tzn. o 44%.

16. Největší kumulativní relativní četnost se rovná

a) dvojnásobku průměru,
b) dvojnásobku mediánu,
c) dvojnásobku módu,
d) součtu všech jednotlivých hodnot absolutních četností,
e) 1.

16. Největší kumulativní relativní četnost se rovná

a) ~~dvojnásobku průměru,~~
b) ~~dvojnásobku mediánu,~~
c) ~~dvojnásobku módu,~~
d) ~~součtu všech jednotlivých hodnot absolutních četností,~~
e) 1.

PROČ?

Kumulativní četnost *k*-té („nejvyšší “) varianty je rovna rozsahu proměnné $m_k = n$.

Kumulativní relativní četnost F_i vyjadřuje jakou část souboru tvoří hodnoty nabývající *i*-té a nižší varianty, což není nic jiného než relativní vyjádření kumulativní četnosti.

$$F_k = \frac{m_k}{n} = \frac{n}{n} = 1$$

17. Určete, která tvrzení jsou pravdivá.

a) Geometrický průměr je definován pro proměnné, které nabývají pouze kladných hodnot.
b) Jedna čtvrtina hodnot je větší než 25% kvantil, zatímco tři čtvrtiny hodnot jsou menší.
c) Mají-li dvě proměnné stejný průměr a stejný rozptyl, mají stejný variační koeficient.
d) Mzdy v ČR mají kladnou šikmost. (V ČR mají zhruba 2/3 lidí podprůměrný plat.)
e) Nejčetnější hodnota v souboru se nazývá medián.
f) Rozptyl má vždy kladnou hodnotu.

17. Určete, která tvrzení jsou pravdivá.

a) Geometrický průměr je definován pro proměnné, které nabývají pouze kladných hodnot.
b) Jedna čtvrtina hodnot je větší než 25% kvantil, zatímco tři čtvrtiny hodnot jsou menší.
c) Mají-li dvě proměnné stejný průměr a stejný rozptyl, mají stejný variační koeficient.
d) Mzdy v ČR mají kladnou šikmost. (V ČR mají zhruba 2/3 lidí podprůměrný plat.)
e) Nejčetnější hodnota v souboru se nazývá medián.
f) Rozptyl má vždy kladnou hodnotu.

17. Určete, která tvrzení jsou pravdivá.

a) Geometrický průměr je definován pro proměnné, které nabývají pouze kladných hodnot.
b) Jedna čtvrtina hodnot je větší než 25% kvantil, zatímco tři čtvrtiny hodnot jsou menší.
c) Mají-li dvě proměnné stejný průměr a stejný rozptyl, mají stejný variační koeficient.
d) Mzdy v ČR mají kladnou šikmost. (V ČR mají zhruba 2/3 lidí podprůměrný plat.)
e) Nejčetnější hodnota v souboru se nazývá medián.
f) Rozptyl má vždy kladnou hodnotu.

17. Určete, která tvrzení jsou pravdivá.

a) Geometrický průměr je definován pro proměnné, které nabývají pouze kladných hodnot.

b) Jedna čtvrtina hodnot je větší než 25% kvantil, zatímco tři čtvrtiny hodnot jsou menší.

c) Mají-li dvě proměnné stejný průměr a stejný rozptyl, mají stejný variační koeficient.

d) Mzdy v ČR mají kladnou šikmost. (V ČR mají zhruba 2/3 lidí podprůměrný plat.)

e) Nejčetnější hodnota v souboru se nazývá medián.

f) Rozptyl má vždy kladnou hodnotu.

17. Určete, která tvrzení jsou pravdivá.

a) Geometrický průměr je definován pro proměnné, které nabývají pouze kladných hodnot.
b) Jedna čtvrtina hodnot je větší než 25% kvantil, zatímco tři čtvrtiny hodnot jsou menší.
c) Mají-li dvě proměnné stejný průměr a stejný rozptyl, mají stejný variační koeficient.
d) Mzdy v ČR mají kladnou šikmost. (V ČR mají zhruba 2/3 lidí podprůměrný plat.)
e) Nejčetnější hodnota v souboru se nazývá medián.
f) Rozptyl má vždy kladnou hodnotu.

17. Určete, která tvrzení jsou pravdivá.

a) Geometrický průměr je definován pro proměnné, které nabývají pouze kladných hodnot.

b) Jedna čtvrtina hodnot je větší než 25% kvantil, zatímco tři čtvrtiny hodnot jsou menší.

c) Mají-li dvě proměnné stejný průměr a stejný rozptyl, mají stejný variační koeficient.

d) Mzdy v ČR mají kladnou šikmost. (V ČR mají zhruba 2/3 lidí podprůměrný plat.)

e) Nejčetnější hodnota v souboru se nazývá medián.

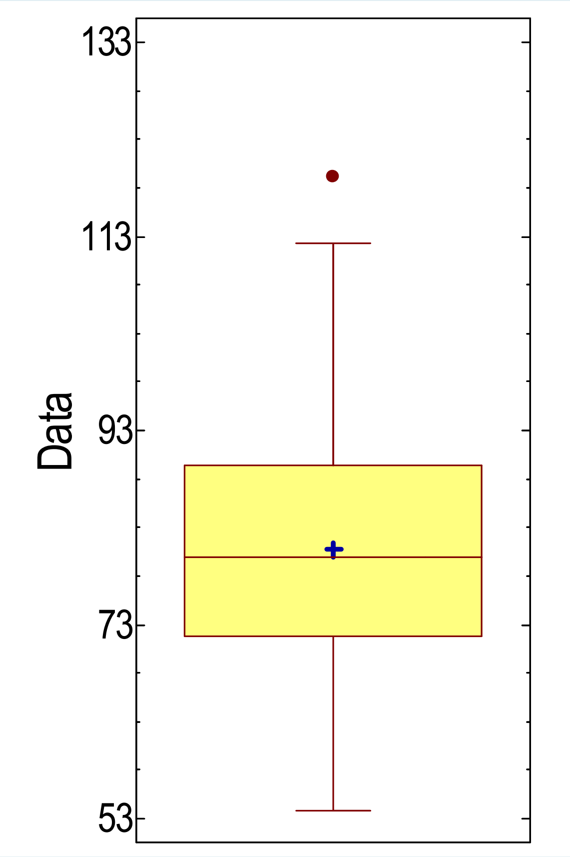
f) Rozptyl má vždy kladnou hodnotu.

17. Určete, která tvrzení jsou pravdivá.

a) Geometrický průměr je definován pro proměnné, které nabývají pouze kladných hodnot.
b) Jedna čtvrtina hodnot je větší než 25% kvantil, zatímco tři čtvrtiny hodnot jsou menší.
c) Mají-li dvě proměnné stejný průměr a stejný rozptyl, mají stejný variační koeficient.
d) Mzdy v ČR mají kladnou šikmost. (V ČR mají zhruba 2/3 lidí podprůměrný plat.)
e) Nejčetnější hodnota v souboru se nazývá medián.
f) Rozptyl má vždy kladnou hodnotu.

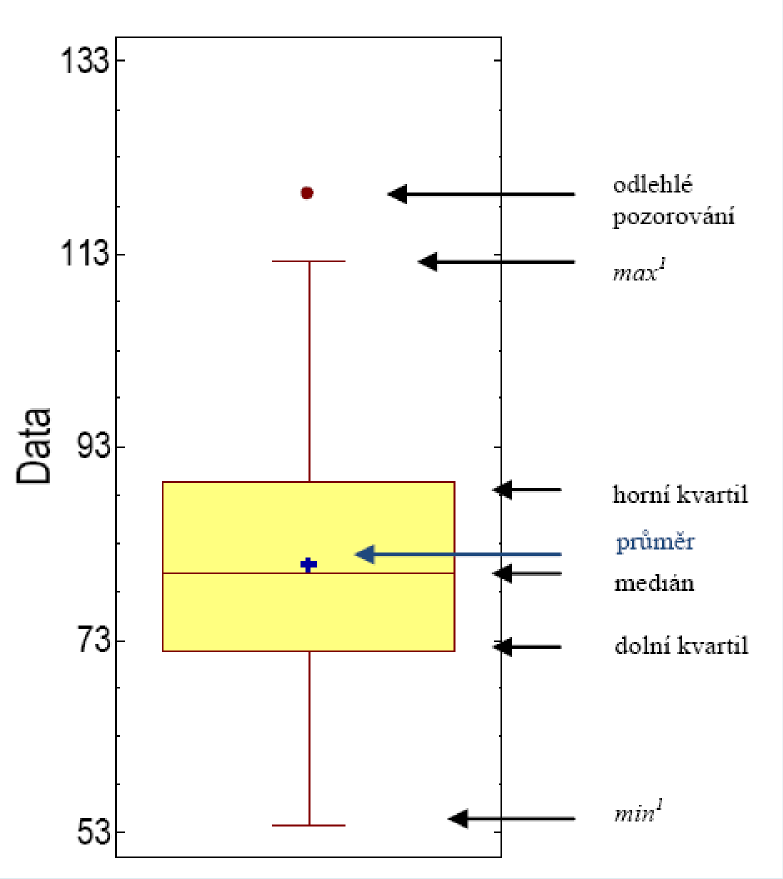
18. Modrý křížek v uvedeném grafu označuje

a) medián,
b) průměr,
c) modus,
d) interkvartilové rozpětí



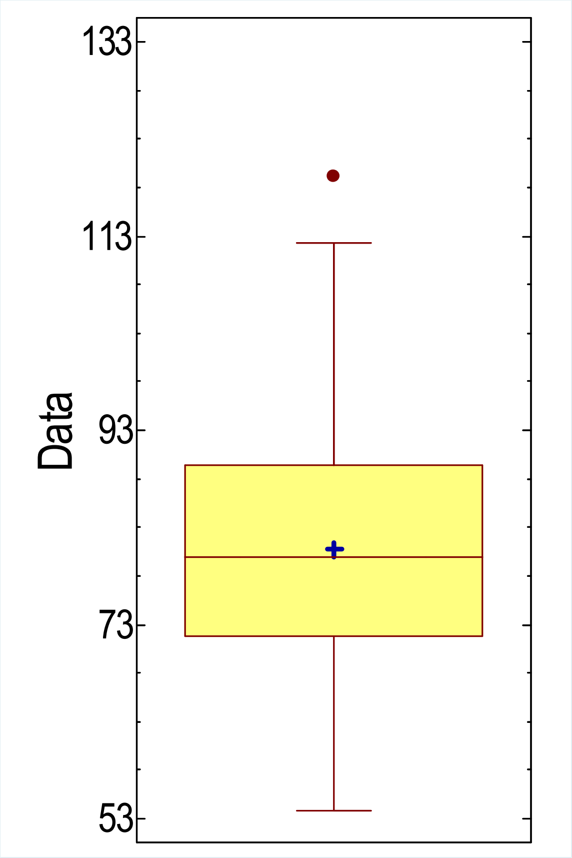
18. Modrý křížek v uvedeném grafu označuje

~~a)~~ medián,
b) průměr,
~~c)~~ modus,
~~d)~~ interkvartilové rozpětí.



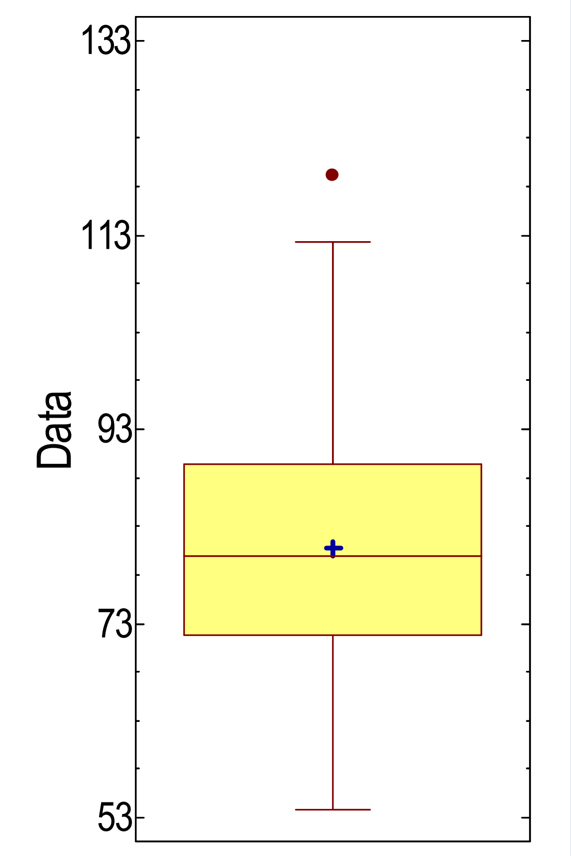
19. Určete která tvrzení jsou pravdivá. Proměnná znázorněna na obrázku

a) neobsahuje odlehlá pozorování,
b) má kladnou šikmost,
c) je kladná,
d) má více než polovinu hodnot větších než 83.



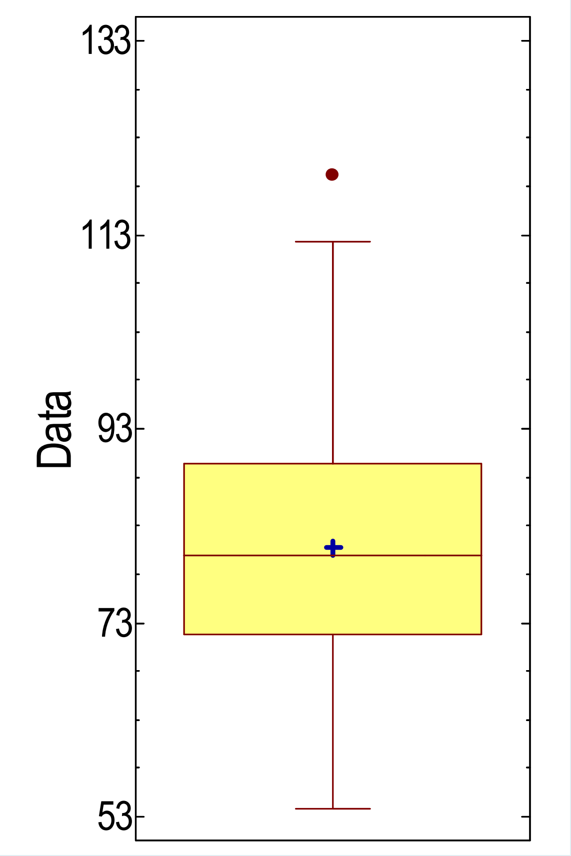
19. Určete která tvrzení jsou pravdivá. Proměnná znázorněna na obrázku

a) neobsahuje odlehlá pozorování,
b) má kladnou šikmost,
c) je kladná,
d) má více než polovinu hodnot větších než 83.



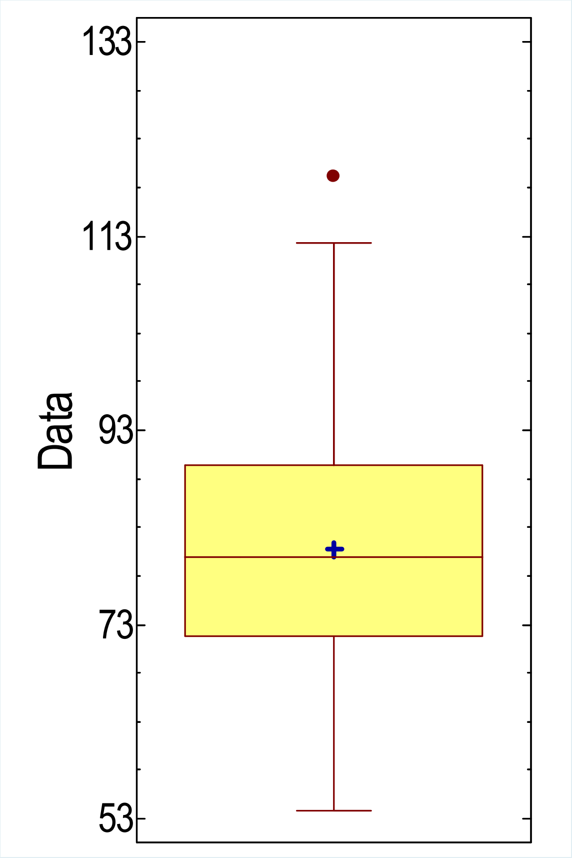
19. Určete která tvrzení jsou pravdivá. Proměnná znázorněna na obrázku

a) neobsahuje odlehlá pozorování,
b) má kladnou šikmost,
c) je kladná,
d) má více než polovinu hodnot větších než 83.



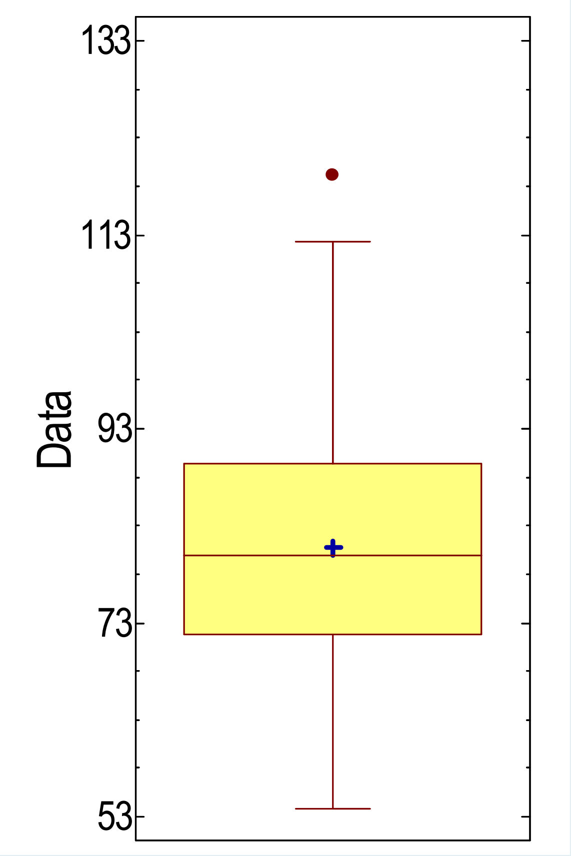
19. Určete která tvrzení jsou pravdivá. Proměnná znázorněna na obrázku

a) neobsahuje odlehlá pozorování,
b) má kladnou šikmost,
c) je kladná,
d) má více než polovinu hodnot větších než 83.



19. Určete která tvrzení jsou pravdivá. Proměnná znázorněna na obrázku

a) neobsahuje odlehlá pozorování,
b) má kladnou šikmost,
c) je kladná,
d) má více než polovinu hodnot větších než 83.



20. Na atletických závodech mládeže žáci soutěžili ve 4 kategoriích. Určete, který výrok je nepravdivý.

a) Na obrázku je znázorněn histogram a nejméně soutěžících bylo ve skoku do dálky.
b) Celkem ve čtyřech kategoriích soutěžilo 80 žáků.
c) Modus = hod koulí.
d) Modus = 30.

20. Na atletických závodech mládeže žáci soutěžili ve 4 kategoriích. Určete, který výrok je nepravdivý.

~~a)~~ Na obrázku je znázorněn histogram a nejméně soutěžících bylo ve skoku do dálky.

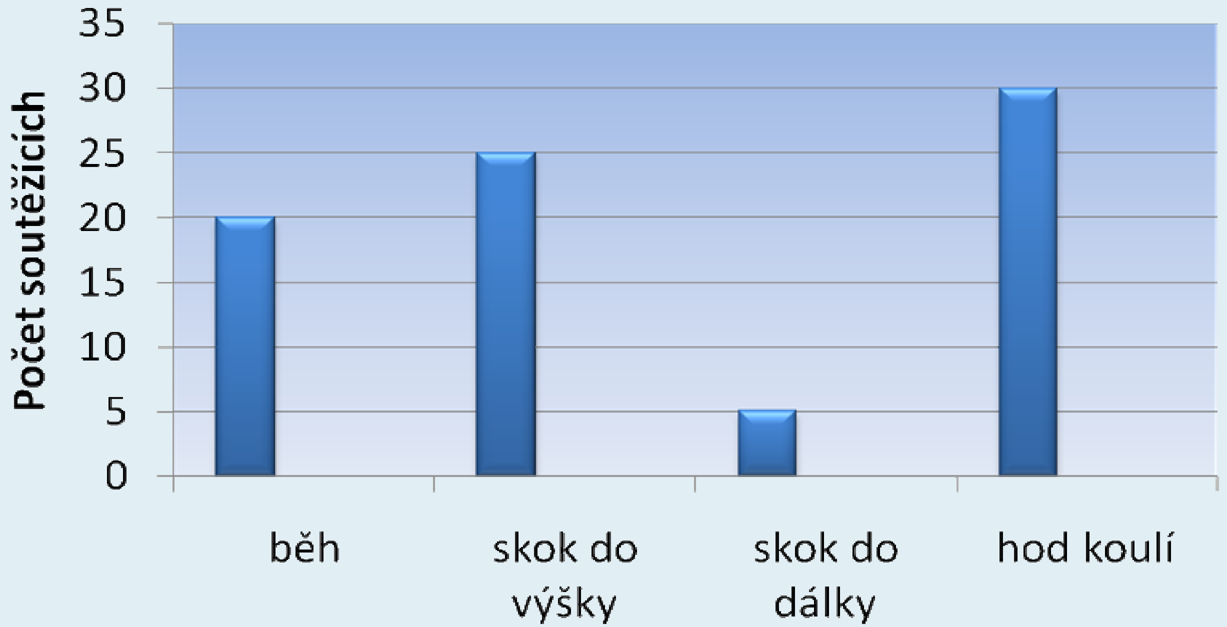
b) Celkem ve čtyřech kategoriích soutěžilo 80 žáků.

c) Modus = hod koulí.

d) Modus = 30.

20. Na atletických závodech mládeže žáci soutěžili ve 4 kategoriích. Určete, který výrok je nepravdivý.

~~a)~~ Na obrázku je znázorněn histogram a nejméně soutěžících bylo ve skoku do dálky.
~~b)~~ Celkem ve čtyřech kategoriích soutěžilo 80 žáků.
c) Modus = hod koulí.
d) Modus = 30.



20. Na atletických závodech mládeže žáci soutěžili ve 4 kategoriích. Určete, který výrok je nepravdivý.

- a) Na obrázku je znázorněn histogram a nejméně soutěžících bylo ve skoku do dálky.
- b) Celkem ve čtyřech kategoriích soutěžilo 80 žáků.
- c) Modus = hod koulí.
- d) Modus = 30.

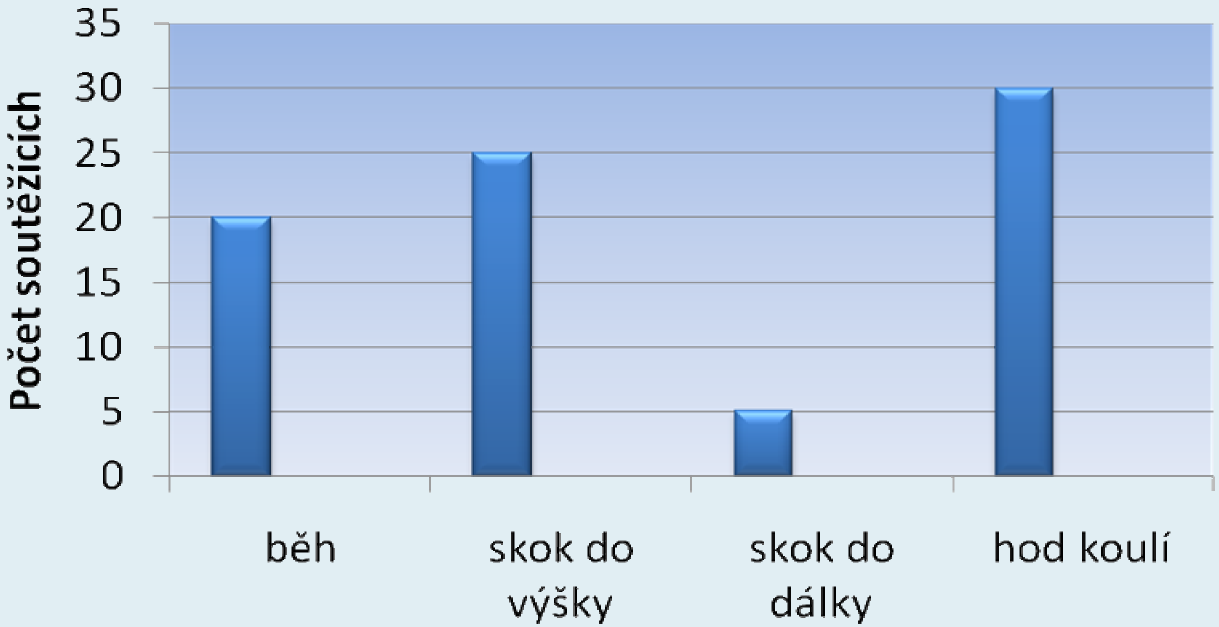
20. Na atletických závodech mládeže žáci soutěžili ve 4 kategoriích. Určete, který výrok je nepravdivý.

a) Na obrázku je znázorněn histogram a nejméně soutěžících bylo ve skoku do dálky.

b) Celkem ve čtyřech kategoriích soutěžilo 80 žáků.

c) Modus = hod koulí.

d) Modus = 30.



21. Číslicový histogram reprezentuje množství peněz, které studenti jedné třídy vybrali na humanitární účely. Které z následujících výroků jsou určitě nepravdivé?

a) 10 studentů věnovalo méně než 120 Kč.
b) Medián vybrané částky činí 120 Kč.
c) Na humanitarní účely přispělo v této třídě 23 studentů.
d) Přispívající studenti věnovali na humanitární účely částky od 1,- Kč do 35,- Kč.

0	11555889	8
1	112344555	(9)
2	005	6
3	025	3

Multiply by 10^2

21. Číslicový histogram reprezentuje množství peněz, které studenti jedné třídy vybrali na humanitární účely. Které z následujících výroků jsou určitě nepravdivé?

~~a)~~ 10 studentů věnovalo méně než 120 Kč.

b) Medián vybrané částky činí 120 Kč.

c) Na humanitarní účely přispělo v této třídě 23 studentů.

d) Přispívající studenti věnovali na humanitární účely částky od 1,- Kč do 35,- Kč.

0	11555889	8
1	112344555	(9)
2	005	6
3	025	3

Multiply by 10^2

21. Číslicový histogram reprezentuje množství peněz, které studenti jedné třídy vybrali na humanitární účely. Které z následujících výroků jsou určitě nepravdivé?

a) 10 studentů věnovalo méně než 120 Kč.
b) Medián vybrané částky činí 120 Kč.
c) Na humanitarní účely přispělo v této třídě 23 studentů.
d) Přispívající studenti věnovali na humanitární účely částky od 1,- Kč do 35,- Kč.

0	11555889	8
1	112344555	(9)
2	005	6
3	025	3

Multiply by 10^2

21. Číslicový histogram reprezentuje množství peněz, které studenti jedné třídy vybrali na humanitární účely. Které z následujících výroků jsou určitě nepravdivé?

a) 10 studentů věnovalo méně než 120 Kč.

b) Medián vybrané částky činí 120 Kč.

c) Na humanitarní účely přispělo v této třídě 23 studentů.

d) Přispívající studenti věnovali na humanitární účely částky od 1,- Kč do 35,- Kč.

0	11555889	8
1	112344555	(9)
2	005	6
3	025	3

Multiply by 10^2

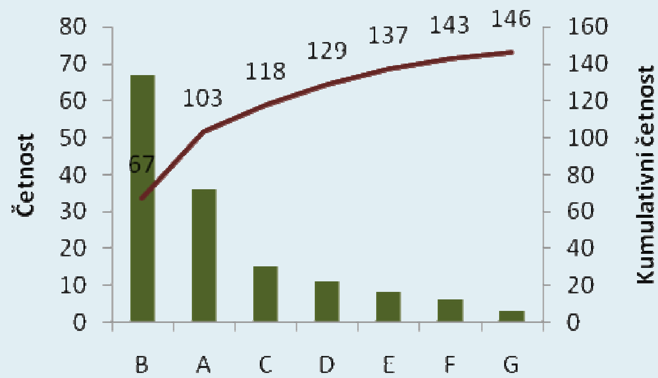
21. Číslicový histogram reprezentuje množství peněz, které studenti jedné třídy vybrali na humanitární účely. Které z následujících výroků jsou určitě nepravdivé?

a) 10 studentů věnovalo méně než 120 Kč.
b) Medián vybrané částky činí 120 Kč.
c) Na humanitarní účely přispělo v této třídě 23 studentů.
d) Přispívající studenti věnovali na humanitární účely částky od 1,- Kč do 35,- Kč.

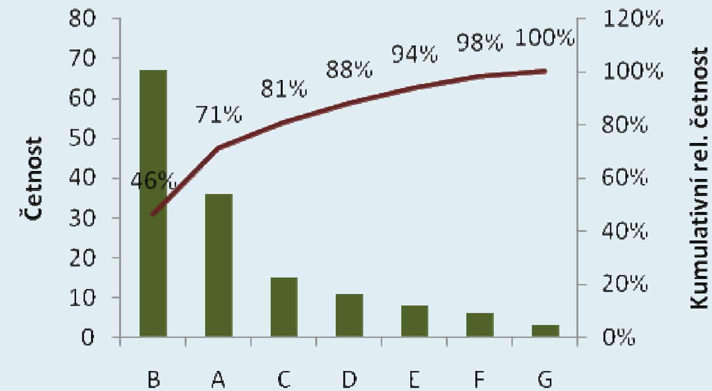
0	11555889	8
1	112344555	(9)
2	005	6
3	025	3

Multiply by 10^2

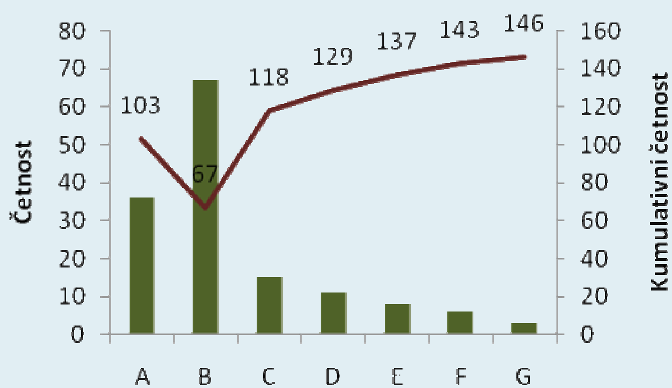
22. Určete na kterém obrázku je zobrazen Paretův graf.



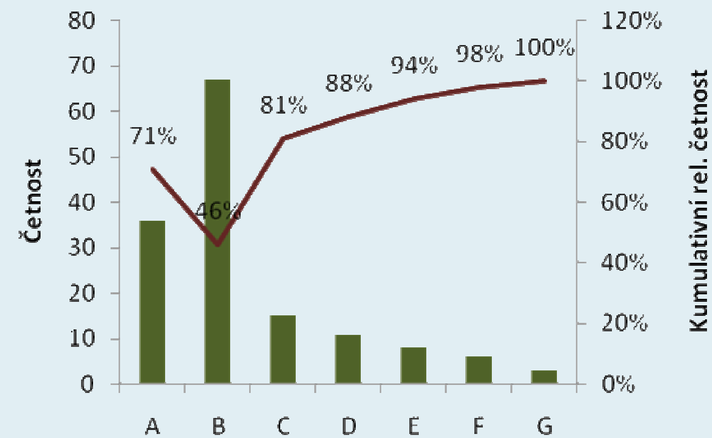
a)



b)

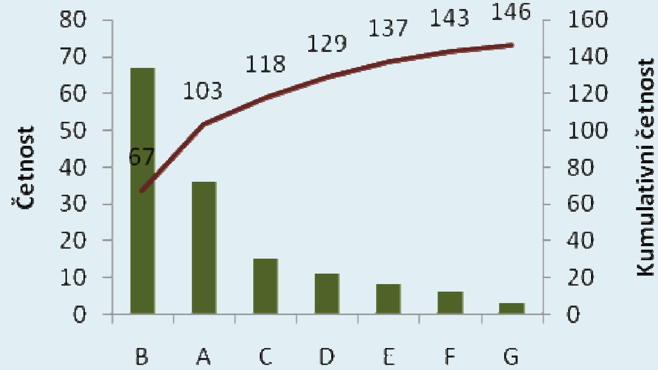


c)

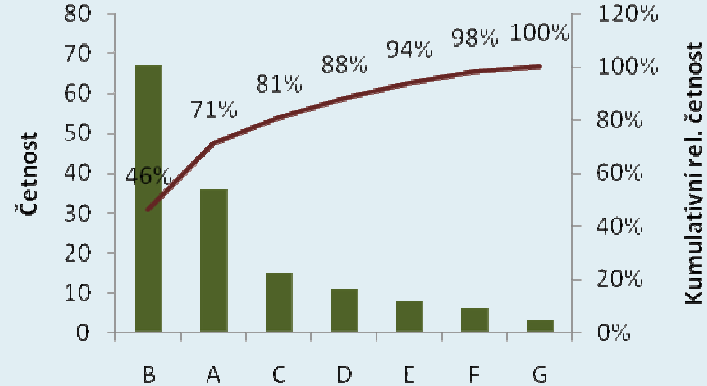


d)

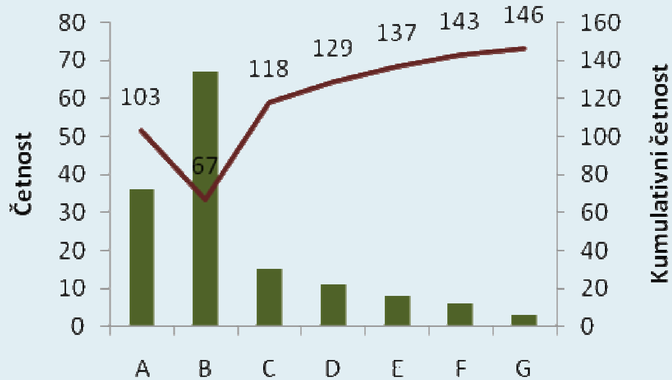
22. Určete na kterém obrázku je zobrazen Paretův graf.



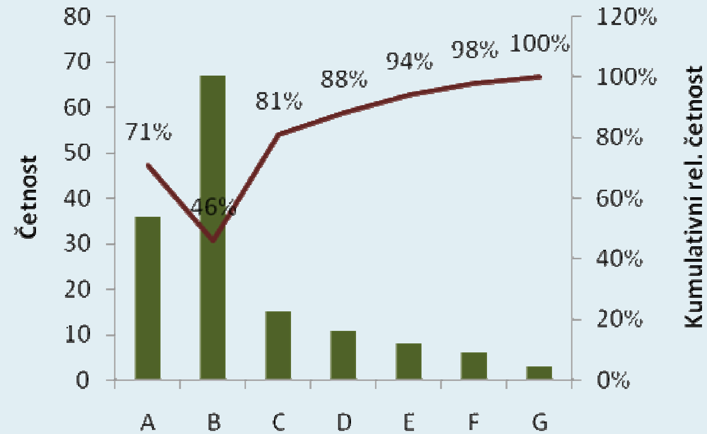
a)



b)



c)



d)