Číselné charakteristiky a jejich výpočet

Jiří Neubauer

Katedra ekonometrie, FVL, UO Brno
kancelář 69a, tel. 973 442029
email:Jiri.Neubauer@unob.cz

Číselné charakteristiky

- charakteristiky polohy
- charakteristiky variability
- charakteristiky koncetrace

Číselné charakteristiky

- charakteristiky polohy
- charakteristiky variability
- charakteristiky koncetrace

Číselné charakteristiky

- charakteristiky polohy
- charakteristiky variability
- charakteristiky koncetrace

Charakteristiky polohy

Charakteristiky polohy (úrovně) měří obecnou velikost hodnot znaku v souboru a dělí se na průměry (počítané ze všech dat) a ostatní míry polohy (počítané z vybraných hodnot).

Aritmetický průměr

Definice

Aritmetický průměr je dán vztahem

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

kde $x_1, x_2 \ldots, x_n$ jsou naměřené hodnoty, n je celkový počet pozorování.

Aritmetický průměr nejčastěji užívaný druh průměru, který má uplatnění při řešení téměř všech úloh statistiky.

Aritmetický průměr

Jsou-li hodnoty statistického znaku uspořádány do tabulky rozdělení četností, určíme aritmetický průměr pomocí vztahu

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{k} n_i \cdot x_i,$$

kde $n_1, n_2, \ldots, n_k$ jsou četnosti jednotlivých variant znaku $x_1, x_2 \ldots, x_k$. Tyto četnosti udávají váhu jednotlivých variantám znaku x, proto mluvíme o **váženém aritmetickém průměru**.

Aritmetický průměr

Aritmetický průměr má tyto základní vlastnosti:

- součet jednotlivých odchylek od průměru je nulový, tj

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0,$$

- aritmetický průměr konstanty je opět roven konstantě, tj.

$$\frac{1}{n}\sum_{i=1}^{n} c = c,$$

Aritmetický průměr

Aritmetický průměr má tyto základní vlastnosti:

- součet jednotlivých odchylek od průměru je nulový, tj

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0,$$

- aritmetický průměr konstanty je opět roven konstantě, tj.

$$\frac{1}{n}\sum_{i=1}^{n} c = c,$$

- přičteme-li k jednotlivým hodnotám znaku x konstantu c, zvýší se o tuto konstantu i aritmetický průměr, tj.

$$\frac{1}{n}\sum_{i=1}^{n}(x_i + c) = c + \overline{x},$$

- násobíme-li jednotlivé hodnoty znaku x konstantou c, je touto konstantou násoben i průměr, tj.

$$\frac{1}{n}\sum_{i=1}^{n} c \cdot x_i = c \cdot \overline{x}.$$

- přičteme-li k jednotlivým hodnotám znaku x konstantu c, zvýší se o tuto konstantu i aritmetický průměr, tj.

$$\frac{1}{n}\sum_{i=1}^{n}(x_i + c) = c + \overline{x},$$

- násobíme-li jednotlivé hodnoty znaku x konstantou c, je touto konstantou násoben i průměr, tj.

$$\frac{1}{n}\sum_{i=1}^{n} c \cdot x_i = c \cdot \overline{x}.$$

Harmonický průměr

Aritmetický průměr však není jediným druhem průměru, existují i jiné, jenž se používají ve speciálních případech.

Definice

Harmonický průměr $\overline{x}_H$ je dán vztahem

$$\overline{x}_H = \frac{n}{\sum\limits_{i=1}^{n} \frac{1}{x_i}}.$$

Harmonický průměr

Harmonický průměr má specifické uplatnění v situacích, kdy má logický význam součet převrácených hodnot znaku. Bude tomu tak tehdy, kdy průměrovaná veličina má charakter části z celku, tedy průměrovat máme tzv. **poměrná čísla**. Např. průměrnou hustotu $\overline{h}$ obyvatelstva na km^2 v kraji, známe-li počet obyvatel p a hustotu h v okresech, určíme ze vztahu $\overline{h} = \frac{\sum p}{\sum r}$, kde rozloha $r = \frac{p}{h}$, nebo průměrnou rychlost $\overline{v}$ auta v km/hod., známe-li dráhu s a jí odpovídající rychlost v, určíme ze vztahu $\overline{v} = \frac{\sum s}{\sum t}$, kde čas $t = \frac{s}{t}$.

Geometrický průměr

Definice

Geometrický průměr $\overline{x}_G$ je dán vztahem

$$\overline{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdots x_n}.$$

Geometrický průměr je např. využíván při jednoduché analýze časové řady pro určení tzv. průměrného tempa růstu nebo průměrného tempa poklesu. Např. pro tři meziroční indexy výroby 1,05; 1,06 a 1,02 je průměrné tempo růstu výroby rovno $\overline{x}_G = \sqrt[3]{1{,}05 \cdot 1{,}06 \cdot 1{,}02} \doteq 1{,}043$, což znamená, že průměrně za rok činil nárůst výroby 4,3 %.

Příklad

Určete aritmetický, harmonický, geometrický a kvadratický průměr z hodnot 1, 2, 5, 6, 7, 8, 8, 9.

- Aritmetický průměr

$$\overline{x} = \frac{1+2+5+6+7+8+8+9}{8} = 5{,}75.$$

- Harmonický průměr

$$\overline{x}_H = \frac{8}{\frac{1}{1}+\frac{1}{2}+\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}+\frac{1}{8}+\frac{1}{9}} \doteq 3{,}375.$$

- Geometrický průměr

$$\overline{x}_G = \sqrt[8]{1\cdot 2\cdot 5\cdot 6\cdot 7\cdot 8\cdot 8\cdot 9} \doteq 4{,}709.$$

Všimněte si, že pro naše průměry platí $\overline{x}_H \leq \overline{x}_G \leq \overline{x}$, tento vztah mezi průměry platí obecně.

Průměry

Pro výpočet aritmetického průměru v programu EXCEL existuje funkce PRŮMĚR, pro určení harmonického a geometrického průměru funkce HARMEAN a GEOMEAN. Harmonický průměr hodnot z předchozího příkladu by se např. určil příkazem HARMEAN(1;2;5;6;7;8;8;9) nebo zadáním daných hodnot do polí A1 až A8 a příkazem HARMEAN(A1:A8).

Kvantil

Definice

Kvantil x_p je hodnota znaku, pro kterou platí, že $100p\,\%$ jednotek uspořádaného souboru má hodnotu menší nebo rovnu x_p a $100(1-p)\,\%$ jednotek má hodnotu větší nebo rovnu x_p.

Takto definovaný kvantil není určen jednoznačně. Na jednoduchém příkladu ukážeme, jak počítají kvantily některé softwarové produkty.

Kvantil

Mějme následující datový soubor 2 5 7 10 12 13 18 21.
Možné výpočty kvantilů

- Uspořádejme data vzestupně od nejmenší hodnoty k největší. Určíme pořadový index i_p kvantilu x_p, který musí vyhovovat nerovnosti

$$np < i_p < np + 1.$$

Kvantil x_p je potom roven hodnotě znaku na pozici i_p, tedy $x_p = x_{(i_p)}$. Jsou-li hodnoty np, $np + 1$ celočíselné, určíme kvantil jako aritmetický průměr hodnot $x_{(np)}$ a $x_{(np+1)}$, tj. $x_p = \frac{x_{(np)}+x_{(np+1)}}{2}$. Tímto způsobem určuje kvantily např. statistický software STATISTICA.

Kvantil

Mějme následující datový soubor 2 5 7 10 12 13 18 21.
Možné výpočty kvantilů

- Uspořádejme data vzestupně od nejmenší hodnoty k největší. Určíme pořadový index i_p kvantilu x_p, který musí vyhovovat nerovnosti

$$np < i_p < np + 1.$$

Kvantil x_p je potom roven hodnotě znaku na pozici i_p, tedy $x_p = x_{(i_p)}$. Jsou-li hodnoty $np, np + 1$ celočíselné, určíme kvantil jako aritmetický průměr hodnot $x_{(np)}$ a $x_{(np+1)}$, tj. $x_p = \frac{x_{(np)} + x_{(np+1)}}{2}$. Tímto způsobem určuje kvantily např. statistický software STATISTICA.

Kvantil

- Podle MATLABu
  Spočteme se číslo

$$\bar{i}_p = \frac{np + np + 1}{2} = \frac{2np + 1}{2}$$

určující polohu kvantilu. Hodnota kvantilu se určí lineární interpolací

$$x_p = x_{([\bar{i}_p])} + (x_{([\bar{i}_p]+1)} - x_{([\bar{i}_p])})(\bar{i}_p - [\bar{i}_p]),$$

kde $[\cdot]$ značí celou část čísla. Je-li $\bar{i}_p < 1$ položíme $x_p = x_{(1)}$, je-li $\bar{i}_p > n$ položíme $x_p = x_{(n)}$.

Kvantil

- Podle EXCELu
  Hodnotám uspořádaného souboru se přiřadí postupně hodnoty $0, \frac{1}{n-1}, \frac{2}{n-1}, \dots, \frac{n-2}{n-1}, 1$. Pokud je hodnota P rovna násobku $\frac{1}{n-1}$, je kvantil x_p roven hodnotě znaku odpovídající danému násobku. Jestliže P není násobkem $\frac{1}{n-1}$, určí se hodnota kvantilu lineární interpolací.

Kvantil

x_p	0,10	0,25	0,50	0,75	0,90
Sylabus	2	6	11	15,5	21
MATLAB	2,9	6	11	15,5	20,1
EXCEL	4,1	6,5	11	14,25	18,9

Příklad

Určete medián, dolní kvartil a horní decil z hodnot 1, 2, 5, 6, 7, 8, 8, 9.

Nejprve určíme medián, tedy prostřední hodnotu uspořádaného souboru. Rozsah souboru je $n = 8$, neexistuje tedy jedna prostřední hodnota, ale hodnoty dvě (6 a 7). Hodnotu mediánu učíme jako aritmetický průměr těchto hodnot

$$\tilde{x} = x_{0,50} = \frac{6+7}{2} = 6{,}5.$$

Tento výsledek budeme interpretovat takto: 50 % uspořádaných hodnot v souboru je menší nebo rovno 6,5, tedy nepřekročí hodnotu 6,5.

Příklad

Nyní určíme dolní kvartil $x_{0,25}$. Vyjdeme ze vztahu

$$np < i_p < np + 1$$

a dostáváme $8 \cdot 0{,}25 < i_p < 8 \cdot 0{,}25 + 1 \Leftrightarrow 2 < i_p < 3$. V případě, že žádné přirozené číslo nesplňuje danou nerovnici (i_p je pořadový index, tedy přirozené číslo), určíme hledaný kvartil jako aritmetický průměr hodnot, které jsou na pořadí np a $np + 1$, v našem případě průměr druhé a třetí hodnoty v uspořádaném souboru

$$x_{0,25} = \frac{x_{(2)} + x_{(3)}}{2} = \frac{2+5}{2} = 3{,}5.$$

Analogicky určíme horní decil $x_{0,90}$,
$8 \cdot 0{,}90 < i_p < 8 \cdot 0{,}90 + 1 \Leftrightarrow 7{,}2 < i_p < 8{,}2$, odkud $i_p = 8$ a

$$x_{0,90} = x_{(8)} = 9.$$

Řekneme, že 25 % uspořádaných hodnot v souboru je menší nejvýše rovno 3,5. Analogicky 90 % hodnot nepřekročí 9.

Příklad

Pro určení mediánu je v EXCELu k dispozici funkce MEDIAN, libovolný kvantil lze spočítat pomocí funkce PERCENTIL, PERCENTIL.INC. Dolní kvartil z příkladu by se potom určil příkazem PERCENTIL.INC(A1:A8;0,25) = 4,25. Dané hodnoty jsou zapsány v polích A1 až A8. Všimněte si, že hodnota určená v EXCELu je odlišná od hodnoty spočítané v příkladu.

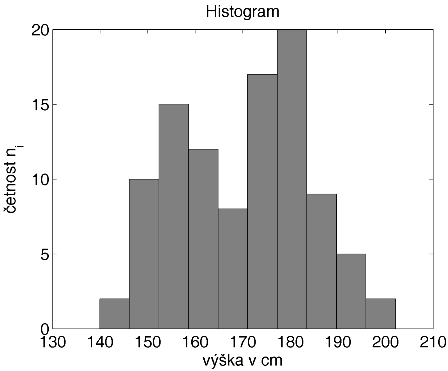
Modus

Definice

Modus $\hat{x}$ je hodnota znaku s největší četností.

V případě spojitého statistického znaku pojem nejčetnější hodnota obvykle nedává smysl, neboť četnosti jednotlivých hodnot znaku jsou buď jedničky, nebo velice malá čísla. (Budeme-li vážit rohlíky na dostatečně přesné váze, hodnoty zjištěné hmotnosti se nebudou zpravidla vůbec opakovat.) Taková data se obvykle zpracovávají pomocí intervalového rozdělení četností a zobrazí pomocí histogramu. Ten interval, který má největší četnost, nazveme **modálním intervalem**.

Modus



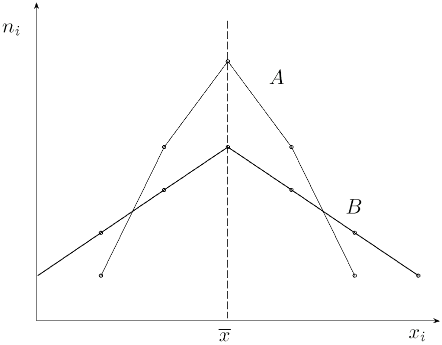
Obrázek: Dvoumodální rozdělení četností

Modus se v EXCELu určí pomocí funkce MODE, MODE.SNGL.

Charakteristiky variability

Průměry, kvantily a modus, tedy charakteristiky o jež byly zmíněny v předchozím odstavci, v sobě shrnují informaci pouze o jedné vlastnosti rozdělení četností, o poloze. Při zpracování dat je možné se setkat s případem, kdy rozdělení četností budou mít shodnou polohu, ale přesto se od sebe budou lišit.
Existuje řada měr variability, zmíníme pouze ty nejdůležitější.

Charakteristiky variability



Obrázek: Rozdělení lišící se variabilitou

Variační rozpětí

Definice

Variační rozpětí R je definováno jako rozdíl největší a nejmenší hodnoty znaku

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

Je to nejjednodušší, ale i nejhrubší míra variability. Udává šířku intervalu, v němž se nacházejí všechny hodnoty znaku.

Kvantilová rozpětí

Kvantilová rozpětí jsou dalšími jednoduchými měrami variability

Definice

- **kvartilové rozpětí**

$$R_Q = x_{0,75} - x_{0,25}$$

- **decilové rozpětí**

$$R_D = x_{0,90} - x_{0,10}$$

- **percentilové rozpětí**

$$R_C = x_{0,99} - x_{0,01}$$

Kvantilová rozpětí

Kvantilová rozpětí jsou dalšími jednoduchými měrami variability

Definice

- **kvartilové rozpětí**

$$R_Q = x_{0,75} - x_{0,25}$$

- **decilové rozpětí**

$$R_D = x_{0,90} - x_{0,10}$$

- **percentilové rozpětí**

$$R_C = x_{0,99} - x_{0,01}$$

Kvantilová rozpětí

Kvantilová rozpětí jsou dalšími jednoduchými měrami variability

Definice

- **kvartilové rozpětí**

$$R_Q = x_{0,75} - x_{0,25}$$

- **decilové rozpětí**

$$R_D = x_{0,90} - x_{0,10}$$

- **percentilové rozpětí**

$$R_C = x_{0,99} - x_{0,01}$$

Kvantilová rozpětí

Kvartilové rozpětí udává šířku intervalu, ve kterém leží 50 % prostředních hodnot uspořádaného souboru. Analogicky decilové resp. percentilové rozpětí určuje šířku intervalu, ve kterém leží 80 % resp. 98 % prostředních hodnot uspořádeného souboru.

Příklad

V dřívějším příkladu jsme určovali kvantily z dat 2, 5, 7, 10, 12, 13, 18 a 21. Vyjdeme z hodnot vypočítaných první metodou (podle programu STATISTICA: $x_{0,10} = 2$, $x_{0,25} = 6$, $x_{0,50} = 11$, $x_{0,75} = 15{,}5$, $x_{0,90} = 21$) a určíme variační, kvartilové a decilové rozpětí.

Variační rozpětí $R = x_{\max} - x_{\min} = 21 - 2 = 19$, všechny hodnoty se nacházejí v intervalu šířky 19. Kvartilové rozpětí má hodnotu $R_Q = x_{0,75} - x_{0,25} = 15{,}5 - 6 = 9{,}5$. Znamená to, že 50 % prostředních hodnot se nachází v intervalu šířky 9,5. Decilové rozpětí je rovno $R_D = x_{0,90} - x_{0,10} = 21 - 2 = 19$.

Kvantilové odchylky

Definice

Kvantilové odchylky

- **kvartilová odchylka**

$$Q = R_Q/2$$

- **decilová odchylka**

$$D = R_D/8$$

- **percentilová odchylka**

$$C = R_C/98$$

Kvantilové odchylky

Definice

Kvantilové odchylky

- **kvartilová odchylka**

$$Q = R_Q/2$$

- **decilová odchylka**

$$D = R_D/8$$

- **percentilová odchylka**

$$C = R_C/98$$

Kvantilové odchylky

Definice

Kvantilové odchylky

- **kvartilová odchylka**

$$Q = R_Q/2$$

- **decilová odchylka**

$$D = R_D/8$$

- **percentilová odchylka**

$$C = R_C/98$$

Kvantilové odchylky

Hodnota kvartilové odchylky udává průměrnou vzdálenost mezi dvěma kvartily, analogicky decilová resp. percentilová odchylka určuje průměrnou vzdálenost mezi sousedními decily, resp. percentily.

Příklad

Určete kvartilovou a decilovou odchylku z hodnot 2, 5, 7, 10, 12, 13, 18 a 21. Využijte dřívějších výsledků.

Kvartilová odchylka $Q = R_Q/2 = 9{,}5/2 = 4{,}75$. Decilová odchylka má hodnotu $D = R_D/8 = 19/8 = 2{,}375$. To znamená, že průměrná délka dvou (osmi) prostředních kvartilových (decilových) intervalů je 4,75 (2,375).

Průměrná odchylka

Definice

Průměrná odchylka je definována jako aritmetický průměr absolutních odchylek jednotlivých hodnot od aritmetického průměru

$$\bar{d}_{\overline{x}} = \frac{1}{n} \sum_{i=1}^{n} |x_i - \overline{x}|.$$

Příklad

Určete průměrnou odchylku z hodnot 1, 2, 5, 6, 7, 8, 8 a 9.

Hodnota aritmetického průměru je $\overline{x} = 5{,}75$. Dosazením do definičního vzorce dostáváme

$$\overline{d}_{\overline{x}} = \frac{|1-5{,}75| + |2-5{,}75| + |5-5{,}75| + |6-5{,}75|}{8} + \\ + \frac{|7-5{,}75| + |8-5{,}75| + |8-5{,}75| + |9-5{,}75|}{8} = 2{,}3125.$$

Průměrnou odchylku získáme v EXCELu pomocí funkce PRŮMODCHYLKA, pro dané hodnoty příkazem PRŮMODCHYLKA(1;2;5;6;7;8;8;9) nebo PRŮMODCHYLKA(A1:A8), pokud jsou data zadána v polích A1 až A8.

Rozptyl

Definice

Rozptyl s_n^2 je definován jako aritmetický průměr čtverců odchylek jednotlivých hodnot znaku od aritmetického průměru

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

Patří k nejpoužívanějším mírám variability.

Rozptyl

Pro ruční výpočty rozptylu je možné odvodit jednodušší vzorec

$$
\begin{aligned}
s_n^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2 = \frac{1}{n}\left(\sum_{i=1}^{n}x_i^2 - 2\overline{x}\sum_{i=1}^{n}x_i + \sum_{i=1}^{n}\overline{x}^2\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n}x_i^2 - 2n\overline{x}^2 - n\overline{x}^2\right) = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \overline{x}^2 = \overline{x^2} - \overline{x}^2.
\end{aligned}
$$

Rozptyl

Rozptyl má tyto základní vlastnosti:

- rozptyl konstanty je roven nule, tj.

$$\frac{1}{n}\sum_{i=1}^{n}(c-c)^2 = 0,$$

- přičteme-li k jednotlivým hodnotám znaku x konstantu c, hodnota rozptylu se nezmění, tj.

$$\frac{1}{n}\sum_{i=1}^{n}[(x_i+c)-(\overline{x}+c)]^2 = s_n^2,$$

- násobíme-li jednotlivé hodnoty znaku x konstantou c, je rozptyl násoben čtvercem této konstanty, tj.

$$\frac{1}{n}\sum_{i=1}^{n}(c\cdot x_i - c\cdot\overline{x})^2 = c^2\cdot s_n^2.$$

Rozptyl

Rozptyl má tyto základní vlastnosti:

- rozptyl konstanty je roven nule, tj.

$$\frac{1}{n}\sum_{i=1}^{n}(c-c)^2 = 0,$$

- přičteme-li k jednotlivým hodnotám znaku x konstantu c, hodnota rozptylu se nezmění, tj.

$$\frac{1}{n}\sum_{i=1}^{n}[(x_i+c)-(\overline{x}+c)]^2 = s_n^2,$$

- násobíme-li jednotlivé hodnoty znaku x konstantou c, je rozptyl násoben čtvercem této konstanty, tj.

$$\frac{1}{n}\sum_{i=1}^{n}(c\cdot x_i - c\cdot\overline{x})^2 = c^2\cdot s_n^2.$$

Rozptyl

Rozptyl má tyto základní vlastnosti:

- rozptyl konstanty je roven nule, tj.

$$\frac{1}{n}\sum_{i=1}^{n}(c-c)^2=0,$$

- přičteme-li k jednotlivým hodnotám znaku x konstantu c, hodnota rozptylu se nezmění, tj.

$$\frac{1}{n}\sum_{i=1}^{n}[(x_i+c)-(\overline{x}+c)]^2=s_n^2,$$

- násobíme-li jednotlivé hodnoty znaku x konstantou c, je rozptyl násoben čtvercem této konstanty, tj.

$$\frac{1}{n}\sum_{i=1}^{n}(c\cdot x_i-c\cdot\overline{x})^2=c^2\cdot s_n^2.$$

Směrodatná odchylka

Definice

Odmocnina z rozptylu se nazývá **směrodatná odchylka**

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

Směrodatná odchylka je, na rozdíl od rozptylu, vyjádřena ve stejných jednotkách jako sledovaný znak. Tvoří-li např. statistický soubor výsledky ve skoku vysokém vyjádřené v centimetrech, má i směrodatná odchylka jednotku *cm*, rozptyl je potom vyjádřen v jednotkách cm^2.

Výběrový rozptyl a směrodatná odchylka

Definice

Výběrový rozptyl s^2 je definovaný vztahem

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

odmocnina z výběrového rozptylu se nazývá **výběrová směrodatná odchylka**

$$s = \sqrt{s^2}.$$

Používá se v induktivní statistice. Jak plyne z definic rozptylu a výběrového rozptylu, platí mezi nimi vztah

$$s_n^2 = \frac{n-1}{n}s^2.$$

Příklad

Určete rozptyl, směrodatnou odchylku, výběrový rozptyl a výběrovou směrodatnou odchylku z hodnot 1, 2, 5, 6, 7, 8, 8 a 9.

Určili hodnotu aritmetického průměru $\overline{x} = 5{,}75$. Nejprve spočítáme hodnotu rozptylu z definičního vzorce

$$s_n^2 = \frac{(1-5{,}75)^2 + (2-5{,}75)^2 + (5-5{,}75)^2 + (6-5{,}75)^2}{8} + \\ + \frac{(7-5{,}75)^2 + (8-5{,}75)^2 + (8-5{,}75)^2 + (9-5{,}75)^2}{8} = 7{,}4375.$$

Příklad

Rozptyl je možné také určit pomocí vztahu $s_n^2 = \overline{x^2} - \overline{x}^2$. Určíme tedy hodnotu

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1^2+2^2+5^2+6^2+7^2+8^2+8^2+9^2}{8} = 40{,}5,$$

odtud potom dostáváme

$$s_n^2 = \overline{x^2} - \overline{x}^2 = 40{,}5 - 5{,}75^2 = 7{,}4375.$$

Směrodatná odchylka je

$$s_n = \sqrt{s_n^2} = \sqrt{7{,}4375} \doteq 2{,}72718.$$

Příklad

Výběrový rozptyl můžeme samozřejmě určit z definice, jednodušší bude ale využít vztahu

$$s^2 = \frac{n}{n-1} s_n^2 = \frac{8}{7} \cdot 7{,}4375 = 8{,}5.$$

Výběrová směrodatná odchylka má potom hodnotu

$$s = \sqrt{s^2} = \sqrt{8{,}5} \doteq 2{,}91548.$$

Pomocí EXCELu můžeme vypočítat hodnotu rozptylu pomocí funkce VAR, VAR.P, směrodatnou odchylku pomocí funkce SMODCH, SMODCH.P, výběrový rozptyl příkazem VAR.VÝBĚR, VAR.S a výběrovou směrodatnou odchylku příkazem SMODCH.VÝBĚR, SMODCH.VÝBĚR.S. Syntaxe zadávání je podobná jako např. u aritmetického průměru.

Variační koeficient

Definice

Nejznámější mírou relativní variability je **variační koeficient**

$$\nu = \frac{s_n}{|\overline{x}|},$$

který je definován jako poměr směrodatné odchylky a absolutní hodnoty aritmetického průměru.

Variační koeficient je bezrozměrné číslo, lze jej vyjádřit i v procentech. Využít ho můžeme v případě, když budeme chtít porovnávat variabilitu ve dvou nebo více statistických souborech, jejichž hodnoty budou vyjádřeny v jiných jednotkách.

Obecný a centrální moment

Definice

r-tý obecný moment je definován vztahem

$$m'_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r,$$

r-tý centrální moment je definován vztahem

$$m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^r.$$

Koeficient šikmosti

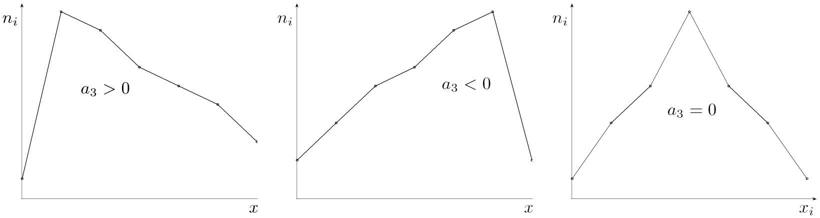
Definice

Koeficient šikmosti je dán vztahem

$$a_3 = \frac{m_3}{m_2^{3/2}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^3}{ns_n^3} = \frac{m_3}{s_n^3}$$

Je-li $a_3 = 0$, je stupeň hustoty malých a velkých hodnot stejný, což představuje souměrné rozdělení četností. Je-li $a_3 > 0$, je stupeň hustoty malých hodnot ve srovnání s hustotou velkých hodnot větší a rozdělení četností je proto zešikmené doleva. Analogicky je-li $a_3 < 0$, je rozdělení četností zešikmené doprava.

Koeficient šikmosti



Obrázek: Rozdělení lišící se šikmostí

Koeficient špičatosti

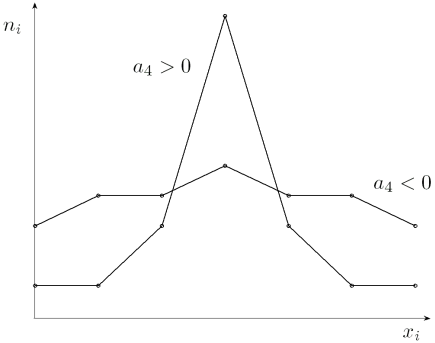
Definice

Koeficient špičatosti je dán vztahem

$$a_4 = \frac{m_4}{m_2^2} - 3 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^4}{n s_n^4} - 3$$

Je-li $a_4 > 0$, je stupeň koncentrace prostředních hodnot ve srovnání s koncentrací všech hodnot větší a rozdělení četností se potom projeví špičatým tvarem. Analogicky je-li $a_4 < 0$, má rozdělení četností plochý tvar.

Koeficient špičatosti



Obrázek: Rozdělení lišící se špičatostí