

1

POPISNÁ STATISTIKA

Předmět popisné statistiky

1.1. Hromadná data a náhodné veličiny. Představte si, že potřebujete zjistit podrobné a komplexní informace o určitém souboru objektů, jedinců či událostí (stromech v lese, lidech ve městě, broucích na mezi, mravencích v mraveništi, výrobcích z určité dodávky, nehodách na silnicích, povodních na řekách apod.) Za tím účelem zpravidla zjišťujeme či měříme vytypované charakteristiky jednotlivých objektů, a tak obdržíme tzv. *hromadná data*. V klasické statistické terminologii se popisované objekty nazývají *statistické jednotky*, zjišťovaným charakteristikám se říká *statistické znaky*, o vyšetřovaném souboru objektů pak mluvíme jako o *statistickém* (či datovém) *souboru*. Hromadná data tedy vznikají měřením jistých statistických znaků na jednotkách nějakého statistického souboru.

Statistickým znakem může být například tloušťka stromu, hmotnost člověka, délka krovek brouka či počet nehod v jistém úseku silnice. Tyto znaky mají proměnlivý charakter a pro různé objekty z daného statistického souboru nabývají různých hodnot. V teorii pravděpodobnosti mluvíme proto o statistických znacích jako o *náhodných veličinách*. Není tomu ovšem tak, že by náhodné veličiny nabývaly svých hodnot zcela nahodile a nepodléhaly nějakému řádu; ve skutečnosti se všechny hodnoty vyskytují s jistými pravděpodobnostmi charakterizujícími danou veličinu. Byla-li tedy hromadná data získána změřením hodnot jistého statistického znaku na jednotkách nějakého statistického souboru, lze očekávat, že více pravděpodobné hodnoty se budou v těchto datech objevovat s větší *četností* (frekvencí) než hodnoty méně pravděpodobné. Základním úkolem popisné statistiky přitom je:

- (1) určit tyto četnosti a prezentovat je ve formě přehledné tabulky či diagramu,
- (2) nahradit zpravidla veliké množství hromadných dat malým počtem *ukazatelů* vystihujících některé charakteristické vlastnosti dat; takovým ukazatelům se též říká *statistiky*.

1.2. Diskrétní a spojité náhodné veličiny. Řekneme, že náhodná veličina je *diskrétní*, nabývá-li pouze konečně nebo spočetně mnoha hodnot. *Spojité* veličiny jsou pak takové, které mohou nabývat všech hodnot z nějakého intervalu.

Příkladem diskrétní náhodné veličiny je počet šišek na stromu, počet roztočů na listu jabloně, počet nehod v roce či výsledek hodu hrací kostkou. Příkladem spojitě náhodné veličiny je pak tloušťka či výška stromu, délka krovek brouka, hmotnost

člověka či věk, kterého se tento člověk dožije apod.

Jak uvidíme dále, techniky používané při prezentaci a charakterizaci hromadných dat se poněkud liší dle toho, byla-li tato data získána změřením hodnot veličin diskretních či spojitých.

Prezentace hromadných dat

Budeme nyní ilustrovat rozličné způsoby prezentace hromadných dat na třech příkladech. V prvních dvou příkladech budeme prezentovat data, která byla získána měřením hodnot diskretní náhodné veličiny (totiž výsledky hodů hrací kostkou a počet roztočů na listech jabloně), ve třetím pak data, která byla získána měřením hodnot spojitě náhodné veličiny (tloušťky stromů).

1.3. Příklad (výsledky hodů hrací kostkou). Následující posloupnost čísel představuje výsledky série sto dvaceti hodů hrací kostkou:

5 6 4 2 3 2 4 3 6 3 3 6 1 6 4 2 6 5 6 3 3 3 1 2 3 3 6 1 2 3 2 5 6 2
 6 1 6 6 5 5 2 2 6 4 5 5 3 1 3 3 3 6 4 4 2 3 1 2 4 2 4 1 2 3 2 3 4 6
 2 1 2 1 2 5 5 3 4 3 1 5 2 4 1 4 4 4 2 4 3 1 4 2 6 4 6 5 4 5 2 2 3 5
 1 3 5 5 2 2 1 3 2 5 3 6 1 4 1 5 1 6

Jde o hromadná data, která byla získána zaznamenáním výsledků jednotlivých hodů. Výsledek hodu je diskretní náhodnou veličinou, která nabývá pouze konečně mnoha hodnot; totiž hodnot z množiny $\{1, 2, 3, 4, 5, 6\}$. Četnosti výskytu jednotlivých hodnot v sérii jsou zaznamenány v následující tabulce:

TAB. 1.1. Tabulka četností

Výsledek hodu	1	2	3	4	5	6
Četnost	17	25	24	19	17	18

Uvědomte si přitom triviální skutečnost, že součet všech četností je roven počtu dat (hodů).

Vyjádříme-li četnosti možných výsledků relativně, obdržíme tabulku *relativních četností*, tj. četností dělených počtem dat.

TAB. 1.2. Tabulka relativních četností

Výsledek hodu	1	2	3	4	5	6
Relativní četnost	0,142	0,208	0,200	0,158	0,142	0,150

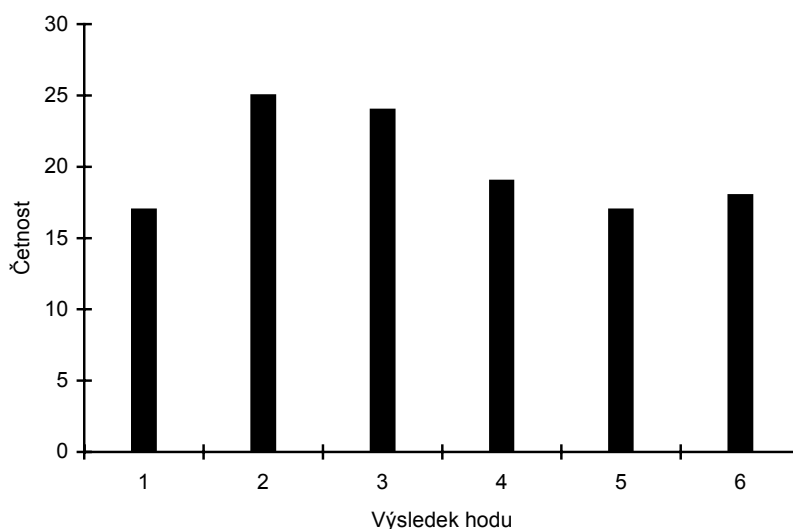
Vzhledem k tomu, že součet všech četností je roven počtu dat, je součet všech *relativních četností roven jedné*.

Někdy se relativní četnosti vyjadřují v procentech (viz následující tabulka).

TAB. 1.3. Tabulka relativních četností (%)

Výsledek hodu	1	2	3	4	5	6
Relativní četnost (%)	14,2	20,8	20,0	15,8	14,2	15,0

Seznam (relativních) četností zachycený v předchozích tabulkách se nazývá též *rozdělením (relativních) četností*. Rozdělení četností lze znázornit též graficky, například tzv. *tyčkovým diagramem* (viz obr. 1.1).



Obr. 1.1. Tyčkový diagram

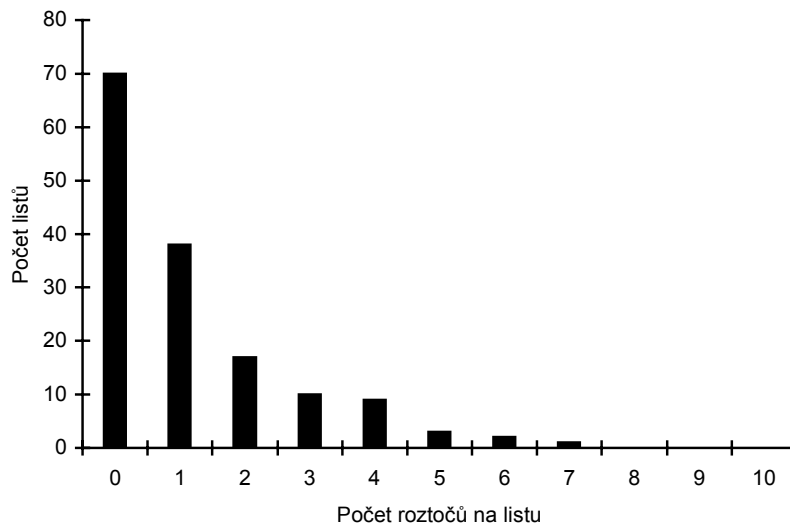
Tyčkový diagram vystihuje velmi názorně relativní rozdíly mezi četnostmi jednotlivých hodnot; přitom je irelevantní, zda jde o diagram četností či diagram četností relativních.

1.4. Příklad (počet roztočů na jabloňových listech). V následující tabulce je zaznamenáno rozdělení počtu roztočů na sto padesáti jabloňových listech.

Počet roztočů na listu	0	1	2	3	4	5	6	7	8 a více
Počet listů s daným počtem roztočů	70	38	17	10	9	3	2	1	0

Popisovanými statistickými jednotkami jsou listy jabloně, zjišťovaným statistickým znakem je počet roztočů na listu, četnost výskytu určité hodnoty tohoto znaku v datovém souboru tedy vyjadřuje počet listů s daným počtem roztočů. Počet roztočů na listu je diskrétní náhodná veličina, jejímiž hodnotami mohou být v principu

všechna nezáporná celá čísla $0, 1, 2, \dots$ (prakticky lze totiž jen těžko stanovit nějakou mez pro maximální možný počet roztočů na jednom listu). Množina hodnot této veličiny je tedy sice nekonečná, ale *spočetná*, což znamená, že lze její prvky očíslovat a seřadit do posloupnosti. Tabulku i tyčkový diagram četností lze proto vytvořit podobně jako v případě, kdy je množina hodnot zkoumané náhodné veličiny konečná s tím drobným rozdílem, že musíme sami rozhodnout, u jaké hodnoty seznam rozdělení četností ukončíme (viz obr 1.2).



Obr. 1.2. Tyčkový diagram rozdělení počtu roztočů na listech jabloně

1.5. Příklad (tloušťky stromů v porostu). Následující data jsou záznamem výčetních tloušťek jednoho sta čtyřicetiletých smrků sitka. (Tloušťky jsou měřeny v milimetrech).

104	198	76	176	82	142	106	120	132	109
111	161	167	138	124	134	139	132	172	169
136	142	146	104	117	106	163	160	154	120
182	183	206	162	163	128	129	220	100	90
133	144	128	79	115	120	91	148	144	102
134	148	123	149	190	118	140	96	122	124
108	137	170	180	114	201	214	207	208	140
189	101	89	110	156	87	120	129	126	160
134	127	141	147	92	174	94	98	150	124
192	122	118	154	141	119	242	179	93	112

Nyní se jedná o data, která byla získána změřením hodnot spojitých náhodných veličin, totiž tloušťek stromů. Hodnotami tloušťek stromů mohou být všechna reálná čísla z určitého intervalu; množina těchto hodnot je tedy nekonečná a navíc nespočetná. Budeme-li měřit tloušťky stromů velmi přesně, pak se v získaném datovém souboru bude každá hodnota vyskytovat pouze jednou. Chceme-li tedy získat

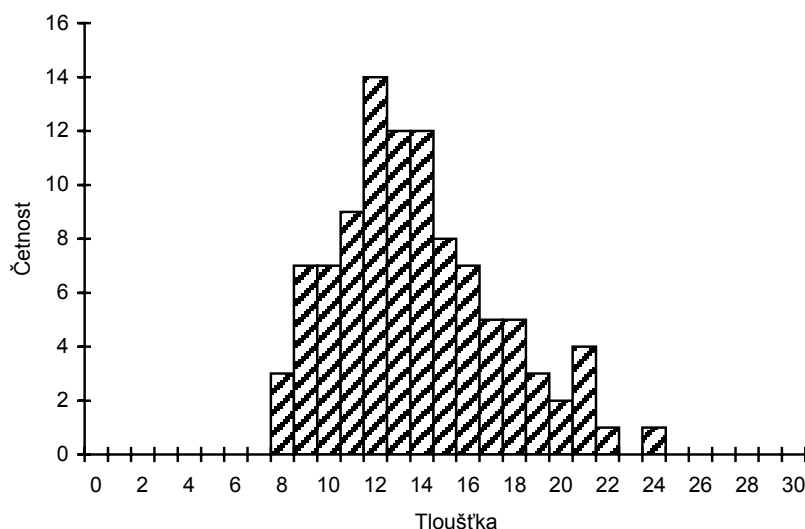
názornou představu o rozdělení četností naměřených tloušťek, je třeba namísto četností jednotlivých hodnot určit četnosti výskytu těchto hodnot v daném rozmezí (intervalu). Zvolená soustava intervalů pak představuje tzv. (*tloušťkové*) *třídy*. Sami přitom určíme, jaké budou mít jednotlivé třídy meze. Nejpřirozenější setřídění našich dat obdržíme tak, že hodnoty tloušťek vyjádříme v centimetrech a poté je zaokrouhlíme na celá čísla. Jinak řečeno, reálnou osu rozdělíme na vzájemně disjunktní třídní intervaly

$$(1) \quad (0,5; 1,5], (1,5; 2,5], (2,5; 3,5], \dots$$

a pro každý takový interval zaznamenáme četnost stromů, jejichž tloušťka se v tomto intervalu nachází. Zastoupíme-li přitom každou třídu jejím středem, obdržíme následující tabulku četností:

Tloušťka (cm)	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Četnost	3	7	7	9	14	12	12	8	7	5	5	3	2	4	1	0	1

Analogem tyčkového diagramu je nyní tzv. *histogram* (viz obr. 1.3).



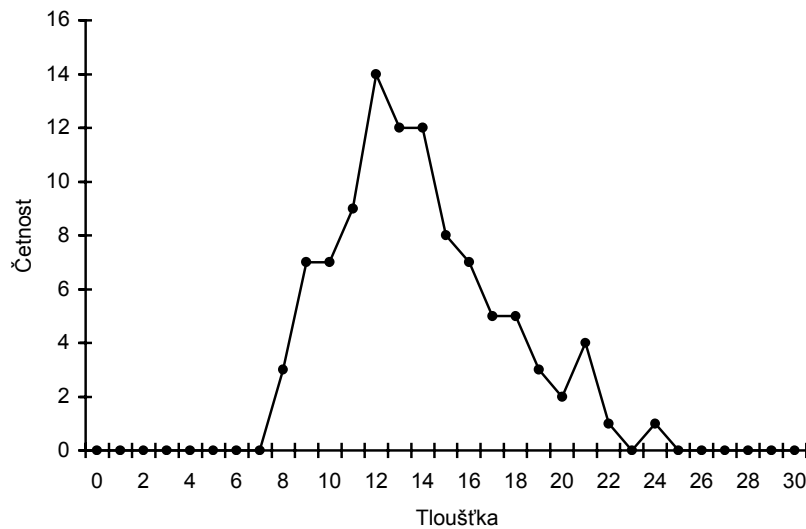
Obr. 1.3. Histogram četností

Sloupce histogramu hrají roli tyček v tyčkovém diagramu. Počátek každého sloupce je totožný s dolní mezí příslušné třídy, konec sloupce pak s její mezí horní. Sloupce tedy navazují jeden na druhý, což názorně vystihuje spojitost měřených veličin. Namísto histogramu se používá též *polygon četností* (viz obr. 1.4). Ten je velice ilustrativní prezentací „*tvaru*“ rozdělení četností. Speciálně si povšimněte, že převládají tloušťky průměrné, zatímco stromů s výrazně podprůměrnou či výrazně nadprůměrnou tloušťkou je velmi málo.

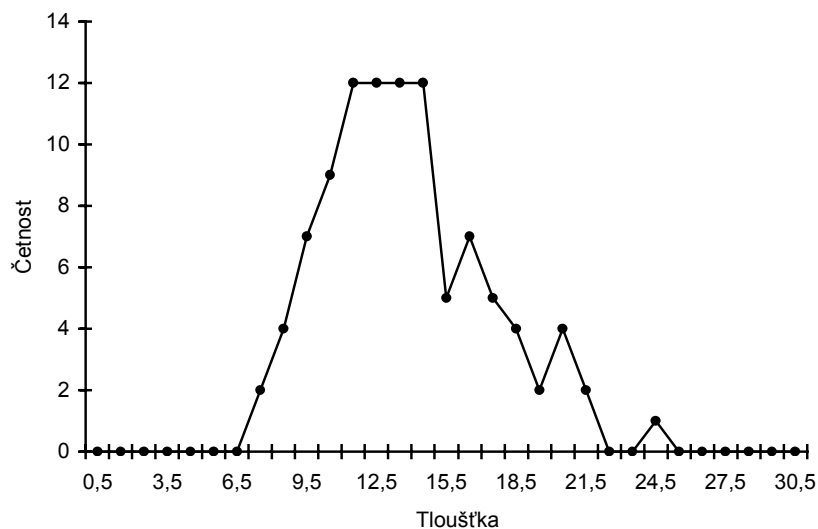
Jiná přirozená soustava třídnicích intervalů je

(2) $(0, 1], (1, 2], (2, 3], \dots$

Ve srovnání s tříděním (1) zůstala tedy zachována délka intervalů, změnil se ale „počátek třídění“. Odpovídající polygon četností je na obr. 1.5 .



Obr. 1.4. Polygon četností



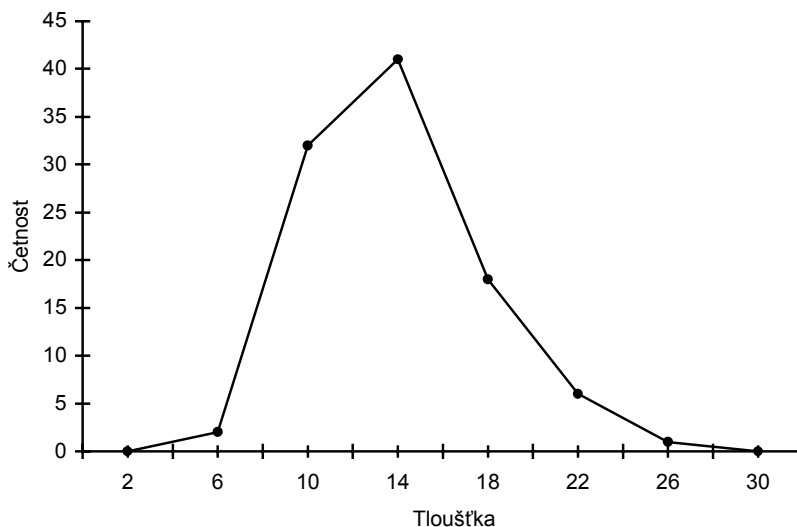
Obr. 1.5. Polygon četností

Všimněte si, že polygony na obrázcích 1.4 a 1.5 se sice co do tvaru globálně shodují, lokálně však nikoliv. Lokální kolísání četností lze přitom odstranit vytvořením

delších třídních intervalů, tj. zvýšením počtu hodnot v jednotlivých třídách. Sdružíme-li například intervaly ze soustavy (2) po čtyřech, obdržíme soustavu třídních intervalů

$$(3) \quad (0, 4], (4, 8], (8, 12], \dots$$

majících délku čtyři centimetry. Odpovídající polygon četností je na obr. 1.6 .



Obr. 1.6. Polygon četností

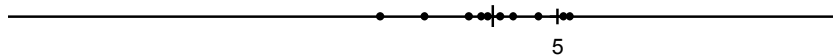
Statistické ukazatele

1.6. Motivační úloha. Při výrobě mincí je stanovena hmotnost mince pět gramů. Je podezření, že na materiálu se systematicky šetří. Cílem je toto podezření prokázat či vyvrátit.

Ukážeme, jak tato úloha přímo vybízí k zavedení některých základních statistických ukazatelů. Předně si uvědomme, že není jiné cesty jak získat informaci o hmotnostech vyráběných mincí než provést náhodnou kontrolu, při níž náhodně vybereme určitý (ne nutně příliš veliký) počet mincí a určíme jejich hmotnost. Dejme tomu, že bylo vybráno deset mincí s následujícími hmotnostmi (v gramech):

4,91 5,02 4,88 4,79 4,89 4,72 5,01 4,97 4,86 4,93

Znázorníme získané hodnoty hmotností jako body (malé kroužky) na číselné ose (viz obr. 1.7). Vidíme, že soustava těchto bodů je poměrně značně posunuta doleva vůči bodu „5“, odpovídajícímu stanovené normě. (Tato skutečnost přitom zesiluje



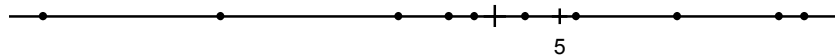
Obr. 1.7. Data jako body na číselné ose

podezření, že se mince vyrábějí systematicky lehčí.) Chceme-li velikost posunutí datové struktury vůči bodu „5“ nějak změřit, je výhodné zastoupit polohu dat na číselné ose jedním bodem. Takový bod je pak *ukazatelem (mírou) polohy* hromadných dat.

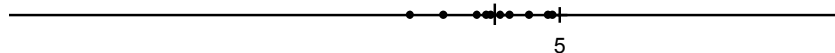
Jako velmi přirozené se jeví zastoupit polohu dat na číselné ose jejich *těžištěm*. Lze přitom snadno ukázat, že souřadnicí tohoto těžiště je *aritmetický průměr* jednotlivých dat. Jinou přirozenou mírou polohy je *medián* neboli prostřední hodnota při uspořádání dat podle velikosti. Jinak řečeno, medián je bod, pod nímž i nad nímž leží stejný počet hodnot.

V našem případě je počet dat sudý a medián proto není určen jednoznačně; ve skutečnosti je mediánem libovolné číslo ležící mezi pátou a šestou hodnotou, tj. nacházející se v intervalu $(4,89; 4,91)$. Skutečnost, že medián je menší než 5, neznamená přitom nic jiného, než že hodnotu menší než 5 má alespoň polovina dat. Aritmetický průměr je přitom roven číslu 4,898. (Na obr. 1.7 je poloha aritmetického průměru znázorněna delší svislou čarou.) Rozdíl $5 - 4,898 = 0,102$ je kvantitativním vyjádřením posunutí datové struktury z obr. 1.7 vůči bodu „5“ doleva.

Skutečnost, že průměrná hmotnost vybraných mincí je o 0,102 g menší než činí stanovená norma, nemusí ještě nutně znamenat, že se mince vyrábějí systematicky lehčí. Hodnota rozdílu mezi průměrnou a stanovenou hmotností ztrácí totiž na významu, pokud je vzorek vybraných mincí příliš malý a jestliže hmotnosti vyráběných mincí vykazují příliš velkou variabilitu. Odrazem velikosti této variability je velikost *rozptýlení* bodů reprezentujících hmotnosti vybraných mincí na číselné ose. Budou-li například hmotnosti vybraných mincí rozptýleny na ose tak silně, jak to vidíme na obr. 1.8, pak možná žádné podezření, že se mince vyrábějí systematicky lehčí, vůbec nevznikne. Naopak při malém rozptýlení, které vidíme na obr. 1.9, bude toto podezření patrně mnohem silnější než při rozptýlení na obr. 1.7. Ve všech třech uvažovaných případech je přitom průměrná hmotnost vybraných mincí stejná.



Obr. 1.8. Data s velkým rozptýlením



Obr. 1.9. Data s malým rozptýlením

Lze definovat různé *ukazatele (míry) rozptýlení* hromadných dat; zpravidla pak konstruujeme tyto ukazatele na základě odchylek jednotlivých hodnot datového souboru od nějaké centrální hodnoty.

Systematickému studiu rozličných statistických ukazatelů včetně příkladů jejich použití je věnován celý zbytek této kapitoly.

Míry polohy

1.7. Definice. Jsou-li $x_1, x_2, \dots, x_n$ reálná čísla (reprezentující hromadná data), pak jejich *aritmetický průměr* $\bar{x}$ je definován předpisem

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

1.8. Poznámka. Význam aritmetického průměru tkví v tom, že může nahradit jednotlivá data při výpočtu jejich součtu. Přesněji řečeno, nahradíme-li všechna čísla $x_1, x_2, \dots, x_n$ průměrnou hodnotou $\bar{x}$, obdržíme nový soubor čísel $\underbrace{\bar{x}, \bar{x}, \dots, \bar{x}}_{n\text{-krát}}$, který

má stejný součet jako soubor původní. Je totiž

$$x_1 + x_2 + \dots + x_n = n\bar{x} = \underbrace{\bar{x} + \bar{x} + \dots + \bar{x}}_{n\text{-krát}}.$$

1.9. Geometrický význam aritmetického průměru. Dle 1.8 je

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n \cdot \bar{x} = 0.$$

To ale znamená, že

$$\sum_{x_i > \bar{x}} (x_i - \bar{x}) = \sum_{x_i < \bar{x}} (\bar{x} - x_i).$$

Reprezentujeme-li tedy čísla $x_1, x_2, \dots, x_n$, a rovněž tak jejich průměr $\bar{x}$, jako body na číselné ose, je součet (absolutních hodnot) odchylek bodů $x_1, x_2, \dots, x_n$ od bodu $\bar{x}$ stejný pro body ležící nalevo od $\bar{x}$ jako pro body ležící napravo od $\bar{x}$. Shrnutí: *Bod $\bar{x}$ je těžištěm bodů $x_1, x_2, \dots, x_n$.*

1.10. Příklad. Uvažme data

4,91 5,02 4,88 4,79 4,89 4,72 5,01 4,97 4,86 4,93

z odstavce 1.6 (obr. 1.7). Jejich aritmetický průměr je roven 4,898, odchylky jednotlivých hodnot od průměru jsou

0,012 0,122 -0,018 -0,108 -0,008 -0,178 0,112 0,072 -0,038 0,032

(Ověřte si sami, že součet všech těchto odchylek je nulový, počítáme-li záporné odchylky i s jejich znaménkem). To ale znamená, že součet kladných odchylek

$$0,012 + 0,122 + 0,112 + 0,072 + 0,032$$

je stejný jako součet záporných odchylek

$$0,018 + 0,108 + 0,008 + 0,178 + 0,038.$$

1.11. Definice. Necht' $x_1, x_2, \dots, x_n$ jsou reálná čísla, přičemž $x_1 \leq x_2 \leq \dots \leq x_n$.

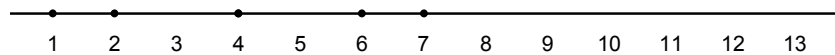
(a) Je-li $n = 2k + 1$ liché číslo, pak *medián* $\tilde{x}$ čísel $x_1, x_2, \dots, x_n$ definujeme předpisem $\tilde{x} = x_{k+1}$.

(b) Je-li $n = 2k$ sudé číslo, pak *medián* $\tilde{x}$ čísel $x_1, x_2, \dots, x_n$ definujeme jako libovolné číslo z intervalu $[x_k, x_{k+1}]$, zpravidla pak jako $\tilde{x} = \frac{1}{2}(x_k + x_{k+1})$.

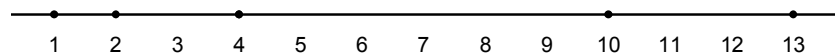
Jinak řečeno, medián čísel $x_1, x_2, \dots, x_n$ získáme tak, že tato čísla uspořádáme podle velikosti a poté vezmeme prostřední z nich, případně průměr dvou prostředních. Poznamenejme v této souvislosti, že latinské slovo „*medius*“ a anglické „*median*“ znamená střední či prostřední.

1.12. Poznámky. Jsou-li data rozložena na číselné ose *symetricky* (viz např. obrázek 1.10), pak jejich aritmetický průměr (těžiště) a medián („prostřední hodnota“) splývají.

Podstatný rozdíl mezi aritmetickým průměrem a mediánem jakožto mírami polohy hromadných dat spočívá v tom, že aritmetický průměr je v protikladu k mediánu velmi citlivý na změny hodnot. Na druhou stranu medián na některé, byť i velmi hrubé (neboli robustní) změny v datové struktuře vůbec nereaguje (srovnej obr. 1.10 s obr. 1.11). Medián proto patří mezi tzv. *robustní statistiky*.



Obr. 1.10. Symetricky rozložená data (1, 2, 4, 6, 7) na číselné ose. Aritmetický průměr i medián jsou rovny číslu 4.



Obr. 1.11. Asymetricky rozložená data (1, 2, 4, 10, 13) na číselné ose. Data vznikla z dat na obr. 1.10 posunutím jejich „pravé části“ více doprava. Aritmetický průměr se rovněž posouvá doprava a je roven 6, hodnota mediánu zůstává nezměněna (je rovna 4).

1.13. Definice. *Modus* je hodnota, která se v hromadných datech vyskytuje s největší četností. Budeme ji značit $\hat{x}$.

Má-li mít přitom pojem modu praktický smysl, musí být datová struktura dostatečně velká, zatímco počet hodnot, které se v této struktuře vyskytují, je poměrně malý. Ani pak ale nemusí být modus určen jednoznačně.

Pro ilustraci uvažme ještě jednou data z příkladu 1.3 (výsledky hodů hrací kostkou). Nejfrekventovanějším výsledkem je dvojka (padla celkem pětadvacetkrát.) Modus je tedy roven dvěma.

V případě, že data vzniknou měřením hodnot spojitě náhodné veličiny, lze jejich modus určit až po té, co je dostatečně hrubě zaokrouhlíme (setřídíme). Hodnota modu pak závisí na způsobu setřídění. Například pro data z příkladu 1.5 (tloušťky

stromů v porostu) a při seřídění znázorněném na obr. 1.3 je modem hodnota dvanáct (centimetrů). To znamená, že tloušťka většiny stromů se nachází v rozmezí 11,5–12,5 cm.

Modus je mírou polohy v tom smyslu, že jde o bod, v němž či kolem něhož jsou data nejvíce soustředěna. Latinské slovo „modus“ je vyjádřením pro (pravou) míru.

Míry rozptýlení

Naším cílem dále bude vyjádřit kvantitativně míru rozptýlení (a tedy též variability) hromadných dat. Necht' $x_1, x_2, \dots, x_n$ jsou reálná čísla (reprezentující hromadná data). Velmi jednoduchou mírou rozptýlení těchto čísel (jakožto bodů na reálné ose) je rozdíl mezi jejich maximální a minimální hodnotou nazývaný též *rozpětí*. Tato míra je ovšem příliš robustní na to, aby mohla mít nějaké příliš významné praktické použití.

Mnohem jemnější míru rozptýlení čísel $x_1, x_2, \dots, x_n$ obdržíme tak, že změříme jejich průměrnou odchylku od nějaké centrální hodnoty x . Položíme-li $x = \bar{x}$, dospějeme k následující definici:

1.14. Definice. Necht' $x_1, x_2, \dots, x_n$ jsou reálná čísla. Číslo d_a definované předpisem

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

se nazývá *průměrná odchylka* (čísel $x_1, x_2, \dots, x_n$ od jejich aritmetického průměru).

Jde o historicky nejstarší používanou míru rozptýlení hromadných dat navrženou francouzským matematikem a fyzikem Pierrem Laplacem (1749–1827). Označení d_a je odvozeno z anglického „average deviation“.

1.15. Rozptyl a směrodatná odchylka. V moderní statistice se průměrná odchylka d_a používá k vyjádření rozptýlení dat jen zřídka a nahrazuje se zpravidla *průměrnou kvadratickou odchylkou* hodnot $x_1, x_2, \dots, x_n$ od jejich aritmetického průměru, tj. výrazem

$$(4) \quad s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 .$$

Číslo s^2 je tzv. *rozptyl* čísel $x_1, x_2, \dots, x_n$, zatímco číslo s se nazývá *směrodatná odchylka*. Směrodatná odchylka je tedy odmocnina z rozptylu.

Zdůrazněme, že slovo „rozptyl“ jsme v této definici použili nikoliv v intuitivním slova smyslu, nýbrž jako odborný termín označující konkrétním způsobem

definovanou míru rozptýlení hromadných dat. V tomto významu budeme výraz rozptyl používat i v celém dalším textu. Písmeno s je v daném kontextu prvním písmenem v anglickém ekvivalentu pro směrodatnou odchylku („standard deviation“).

1.16. Vzorec pro výpočet rozptylu. Výpočet výrazu $\sum_{i=1}^n (x_i - \bar{x})^2$ lze zjednodušit takto:

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2.\end{aligned}$$

Tudíž

$$(5) \quad s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2.$$

Jinak řečeno, *rozptyl čísel $x_1, x_2, \dots, x_n$ lze spočítat tak, že od průměru druhých mocnin čísel $x_1, x_2, \dots, x_n$ odečteme druhou mocninu jejich průměru.* To bývá výhodné při ručním počítání tehdy, když čísla $x_1, x_2, \dots, x_n$ jsou celá a $\bar{x}$ nikoliv.

Obecně pak přímý výpočet rozptylu při známé hodnotě průměru $\bar{x}$ vyžaduje při výpočtu dle definice (4) řádově $3n$ operací (tj. sčítání a násobení), při výpočtu podle vzorce (5) je pak počet operací roven řádově pouze $2n$.

1.17. Příklad. Uvažme ještě jednou data

4,91 5,02 4,88 4,79 4,89 4,72 5,01 4,97 4,86 4,93

z odstavce 1.6 (obr. 1.7). Víme již, že jejich aritmetický průměr je roven 4,898, odchylky jednotlivých hodnot od průměru jsou

0,012 0,122 -0,018 -0,108 -0,008 -0,178 0,112 0,072 -0,038 0,032 .

a součet všech těchto odchylek (uvažovaných i s jejich znaménkem) je nulový. Symbolem d_a jsme označili aritmetický průměr absolutních hodnot těchto odchylek a symbolem s^2 průměr jejich druhých mocnin. Je tedy

$$d_a = \frac{0,012 + 0,122 + 0,018 + 0,108 + 0,008 + 0,178 + 0,112 + 0,072 + 0,038 + 0,032}{10}$$

a

$$s^2 = \frac{0,012^2 + 0,122^2 + 0,018^2 + 0,108^2 + 0,008^2 + 0,178^2 + 0,112^2 + 0,072^2 + 0,038^2 + 0,032^2}{10},$$

resp.

$$s^2 = \frac{4,91^2 + 5,02^2 + 4,88^2 + 4,79^2 + 4,89^2 + 4,72^2 + 5,01^2 + 4,97^2 + 4,86^2 + 4,93^2}{10} - 4,898^2,$$

použijeme-li k výpočtu rozptylu vzorce (5). Vyjde $d_a = 0,07$, $s^2 \doteq 0,0079$, $s \doteq 0,09$. Pro data z obr. 1.8 máme $d_a = 0,28$, $s \doteq 0,36$, pro data z obr. 1.9 pak dostaneme $d_a = 0,05$, $s \doteq 0,07$.

1.18. Vztah mezi směrodatnou a průměrnou odchylkou. Čísla s a d_a mají stejný fyzikální rozměr a podávají o souboru dat $x_1, x_2, \dots, x_n$ stejný typ informace (měří určitým způsobem rozptýlení čísel $x_1, x_2, \dots, x_n$ na číselné ose). Hodnoty odchylek s a d_a se ovšem liší, přičemž vždy platí, že

$$(6) \quad d_a \leq s.$$

Přirozeně definovanou průměrnou odchylku d_a nahrazujeme v matematicko-statistické teorii směrodatnou odchylkou s proto, že teorie založená na počítání s odchylkami kvadratickými namísto odchylek absolutních je mnohem jednodušší a elegantnější.

Dokažme nerovnost (6). Položme $y_i = |x_i - \bar{x}|$. Ze vzorce (5) plyne, že rozdíl

$$\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2$$

je nezáporný, a tedy

$$\sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2} \geq \frac{1}{n} \sum_{i=1}^n y_i.$$

To je ale dokazovaná nerovnost (6).

Uvažujme ještě o tom, pro jaká data se odchylky s a d_a shodují. Pokud tato situace nastane, pak též $s^2 = d_a^2$, čili

$$\frac{1}{n} \sum_{i=1}^n y_i^2 = \bar{y}^2.$$

Odtud dle vzorce (5) vyplývá, že rozptyl čísel $y_1, y_2, \dots, y_n$ je nulový, z čehož dále plyne, že $y_1 = y_2 = \dots = y_n$, a tedy $|x_1 - \bar{x}| = |x_2 - \bar{x}| = \dots = |x_n - \bar{x}|$. To však nastane právě tehdy, když buď

$$(a) \quad x_1 = x_2 = \dots = x_n$$

nebo

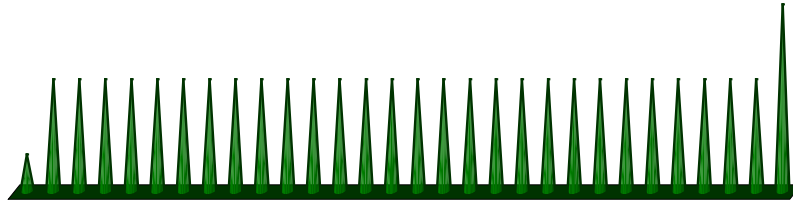
(b) n je sudé a čísla $x_1, x_2, \dots, x_n$ nabývají právě dvou hodnot; přitom každá z obou hodnot se vyskytuje ve stejném počtu.

Naopak, v obou případech (a) i (b) je $s = d_a$. Dospíváme k závěru, že *průměrná odchylka d_a a směrodatná odchylka s nabývají stejné hodnoty tehdy a jen tehdy, nastane-li některý z výše popsanych případů (a) nebo (b).*

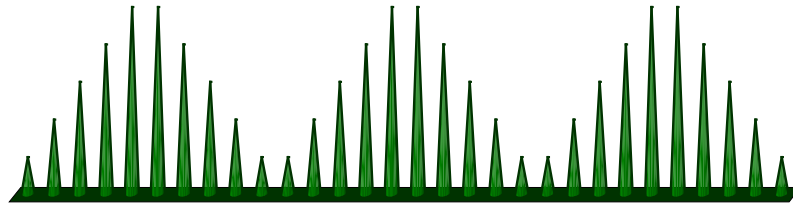
1.19. Poznámka. Všechny výše zavedené míry rozptýlení čísel $x_1, x_2, \dots, x_n$ na číselné ose, totiž rozpětí, průměrná odchylka, směrodatná odchylka a rozptyl mají následující společné vlastnosti:

- (a) jsou vždy nezáporné, přičemž mohou nabýt libovolné nezáporné hodnoty,
- (b) jsou nulové, pokud $x_1 = x_2 = \dots = x_n$,
- (c) jsou nenulové (a tedy kladné), pokud všechna čísla $x_1, x_2, \dots, x_n$ nejsou totožná.

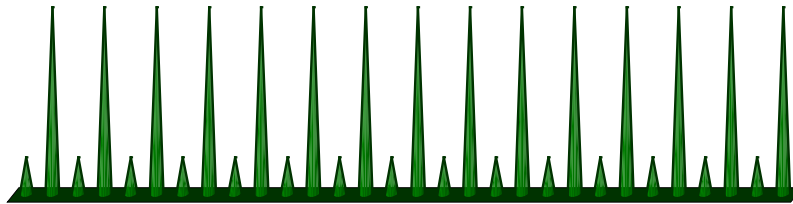
1.20. Ilustrace. Na následujících třech obrázcích jsou schematicky znázorněny výšky tří stejně početných skupin stromů. Přestože výšky mají ve všech třech souborech totéž rozpětí, intuitivně vzato je rozptýlení výšek stromů na obr. 1.12 menší než na obr. 1.13 a u stromů na obr. 1.13 je zase menší než u stromů na obr. 1.14. Tento pocit je přitom velmi dobře kvantifikován hodnotou jak směrodatné, tak průměrné odchylky.



Obr. 1.12. Rozptýlení výšek stromů ($\bar{x} = 3$, rozpětí je 4, $d_a = 2/15 \doteq 0,13$, $s \doteq 0,52$)

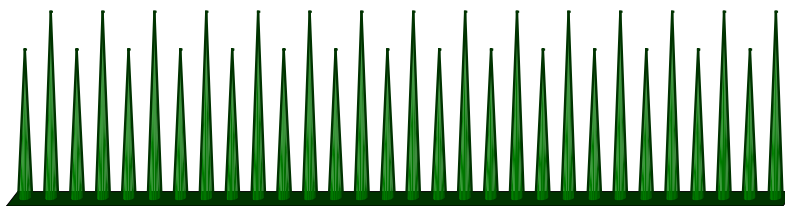


Obr. 1.13. Rozptýlení výšek stromů ($\bar{x} = 3$, rozpětí je 4, $d_a = 1,20$, $s = \sqrt{2} \doteq 1,41$)

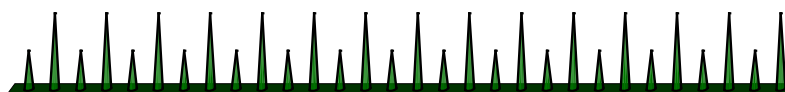


Obr. 1.14. Rozptýlení výšek stromů ($\bar{x} = 3$, rozpětí je 4, $d_a = s = 2$)

1.21. Variační koeficient. Při porovnávání variability několika datových souborů je někdy žádoucí vyjádřit míru rozptýlení hromadných dat relativně vzhledem k jejich průměrné hodnotě. Například rozpětí, průměrná odchylka i směrodatná odchylka výšek stromů znázorněných na obrázcích 1.15 a 1.16 jsou stejné. „Relativně“ však výšky stromů na obr. 1.15 vykazují mnohem menší rozptýlení než výšky stromů na obr. 1.16. Statistickým ukazatelem, který tento rozdíl v rozptýlení hromadných dat dobře vystihne, je kupříkladu poměr $s/\bar{x}$, nazývaný *variační koeficient*. Hodnota tohoto koeficientu se přitom často vyjadřuje v procentech.



Obr. 1.15. Variabilita výšek stromů ($\bar{x} = 9$, rozpětí je 1, $d_a = s = 1$, $s/\bar{x} = 1/9 \doteq 11,1\%$)



Obr. 1.16. Variabilita výšek stromů ($\bar{x} = 3$, rozpětí je 1, $d_a = s = 1$, $s/\bar{x} = 1/3 \doteq 33,3\%$)

Příklady

1.22. Určíme míry polohy a rozptýlení pro výsledky hodů hrací kostkou z příkladu 1.3. Uspořádáme-li data podle velikosti (vzestupně), dostaneme následující posloupnost:

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 2 2 2 2 2 2 2 3
 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6

Jde o řadu sto dvaceti čísel, „prostředními hodnotami“ jsou tedy šedesátá a šedesátá první. Ty jsou obě rovny třem, a tedy i *medián je roven třem*. To znamená, že malé hodnoty 1, 2 a 3 převládají nad velkými hodnotami 4, 5 a 6. Přitom *modus* (nejčastěji se vyskytující hodnota) je *roven dvěma*. Uvědomte si, že tyto skutečnosti lze velmi rychle zjistit též nahlédnutím do tabulky 1.1 či 1.2.

Počítáme-li *aritmetický průměr* ze seřazených dat v tabulce 1.1, je potřeba zahrnout všechny hodnoty tolikrát, kolik činí četnost jejich výskytu v datech. Dostaneme

$$\bar{x} = \frac{17 \cdot 1 + 25 \cdot 2 + 24 \cdot 3 + 19 \cdot 4 + 17 \cdot 5 + 18 \cdot 6}{120} = 3,4.$$

Je důležité si povšimnout, že výpočet lze provést též na základě znalosti relativních četností z tabulky 1.2, aniž bychom znali počet měření. Lze totiž psát

$$\begin{aligned}\bar{x} &= \frac{17 \cdot 1 + 25 \cdot 2 + 24 \cdot 3 + 19 \cdot 4 + 17 \cdot 5 + 18 \cdot 6}{120} \\ &= \frac{17}{120} \cdot 1 + \frac{25}{120} \cdot 2 + \frac{24}{120} \cdot 3 + \frac{19}{120} \cdot 4 + \frac{17}{120} \cdot 5 + \frac{18}{120} \cdot 6 \\ &\doteq 0,142 \cdot 1 + 0,208 \cdot 2 + 0,200 \cdot 3 + 0,158 \cdot 4 + 0,142 \cdot 5 + 0,150 \cdot 6.\end{aligned}$$

Četnosti resp. relativní četnosti hrají tedy při výpočtu aritmetického průměru roli *vah* jednotlivých hodnot.

Podobně pro *průměrnou odchylku* dostaneme

$$d_a = \frac{17 \cdot |1 - 3,4| + 25 \cdot |2 - 3,4| + 24 \cdot |3 - 3,4| + 19 \cdot |4 - 3,4| + 17 \cdot |5 - 3,4| + 18 \cdot |6 - 3,4|}{120} \doteq 1,42.$$

Pro *rozptyl* pak máme

$$\begin{aligned}s^2 &= \frac{17 \cdot (1 - 3,4)^2 + 25 \cdot (2 - 3,4)^2 + 24 \cdot (3 - 3,4)^2 + 19 \cdot (4 - 3,4)^2 + 17 \cdot (5 - 3,4)^2 + 18 \cdot (6 - 3,4)^2}{120} \\ &= \frac{17 \cdot 1^2 + 25 \cdot 2^2 + 24 \cdot 3^2 + 19 \cdot 4^2 + 17 \cdot 5^2 + 18 \cdot 6^2}{120} - 3,4^2 = 2,69.\end{aligned}$$

Konečně pro *směrodatnou odchylku* obdržíme $s \doteq 1,64$.

(Přitom stejně jako výpočet měr polohy lze i výpočet odchylek d_a a s provést pouze na základě znalosti relativních četností.)

Shrnuto: $\bar{x} = 3,4$; $\hat{x} = 2$; $\tilde{x} = 3$; $d_a \doteq 1,42$; $s^2 = 2,69$; $s \doteq 1,64$.

1.23. Lze ukázat, že pro tloušťky stromů v porostu z příkladu 1.5 je:

$$\bar{x} = 138,5; \quad \tilde{x} = 134; \quad d_a = 27,61; \quad s^2 = 1203,15; \quad s \doteq 34,69.$$

Vyjádříme-li přitom naměřené hodnoty tlouštěk v centimetrech (bez zaokrouhlení), dostaneme:

$$\bar{x} = 13,85; \quad \tilde{x} = 13,4; \quad d_a \doteq 2,76; \quad s^2 \doteq 12,03; \quad s \doteq 3,47.$$

Zmenší-li se totiž všechna data desetkrát, zmenší se desetkrát i všechny charakteristiky s výjimkou rozptylu, který se v takovém případě zmenší stokrát. (Zdůvodněte to!)

1.24. Sheppardova korekce a interpolace mediánu. Kumulativní četnosti. V praxi se občas stává, že nejsou k dispozici originální data, nýbrž pouze data setříděná, přitom však veličiny, jejichž změřením byla data získána, jsou spojitě. (To vede samozřejmě k jisté ztrátě informace.) Výpočet statistických ukazatelů na základě takových setříděných dat pak provádíme tak, že původní naměřené hodnoty nahradíme středy odpovídajících tříd. Ilustrujme tento postup na datech z příkladu 1.5 (tloušťky stromů v porostu) setříděných po čtyřech centimetrech (viz obr. 1.6). Tabulka četností odpovídající danému setřídění je následující:

Tloušťka (cm)	4 – 8	8 – 12	12 – 16	16 – 20	20 – 24	24 – 28
Četnost	2	32	41	18	6	1

Použijeme-li data z této tabulky k výpočtu mediánu, aritmetického průměru, rozptylu a směrodatné

odchylky tloušťek, dostaneme

$$\tilde{x} = 14,$$

$$\bar{x} = \frac{2 \cdot 6 + 32 \cdot 10 + 41 \cdot 14 + 18 \cdot 18 + 6 \cdot 22 + 1 \cdot 26}{100} = 13,88,$$

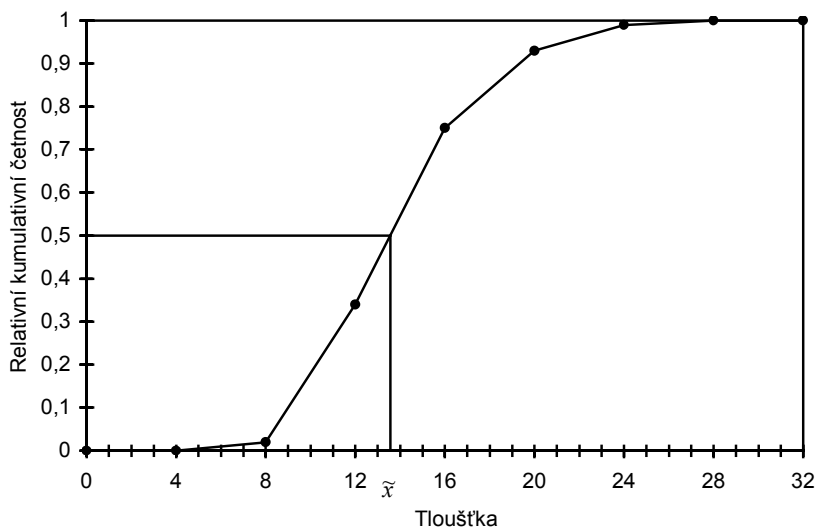
$$s^2 = \frac{2 \cdot 6^2 + 32 \cdot 10^2 + 41 \cdot 14^2 + 18 \cdot 18^2 + 6 \cdot 22^2 + 1 \cdot 26^2}{100} - 13,88^2 = 14,5456,$$

$$s \doteq 3,81.$$

Výsledky se přirozeně liší od těch které byly vypočítány z původních nesetříděných dat (viz 1.23). Lze přitom ukázat, že chyba, ke které došlo při výpočtu aritmetického průměru, je zcela náhodné povahy. Při výpočtu rozptylu dochází ovšem k jeho systematickému nadhodnocení. Velikost tohoto nadhodnocení je v případě stejně širokých tříd rovna řádově $h^2/12$, kde h je šířka třídy. Oprava spočívající v odečtení čísla $h^2/12$ od rozptylu vypočteného ze setříděných dat, se nazývá *Sheppardova korekce*. V našem případě je opravená hodnota rozptylu rovna $14,5456 - 16/12 \doteq 13,21$. Odpovídající hodnota směrodatné odchylky je pak asi 3,63.

TAB. 1.4. Tabulka kumulativních četností

Tloušťka (cm)	≤ 8	≤ 12	≤ 16	≤ 20	≤ 24	≤ 28
Kumulativní četnost	2	34	75	93	99	100
Relativní kumulativní četnost	0,02	0,34	0,75	0,93	0,99	1,00



Obr. 1.17. Polygon relativních kumulativních četností a grafická interpolace mediánu

Co se týče mediánu, snadno nahlédneme, že se nachází v třídě 12 – 16 cm. Tuto hodnotu lze dále zpřesnit pomocí *lineární interpolace*, při níž předpokládáme, že uvnitř jednotlivých tříd jsou hodnoty dat rozloženy zhruba rovnoměrně. Při realizaci interpolace je výhodné nahradit četnosti tzv. *kumulativními četnostmi* či ještě lépe relativními kumulativními četnostmi (viz tabulka 1.4 a obr. 1.17). Užitím tabulky relativních kumulativních četností obdržíme pro lineární interpolaci mediánu následující vztahy:

$$\frac{\tilde{x} - 12}{16 - 12} = \frac{50 - 34}{75 - 34} = \frac{0,50 - 0,34}{0,75 - 0,34}.$$

Odtud pak

$$\tilde{x} = 12 + 4 \cdot \frac{50 - 34}{41} = 12 + 4 \cdot \frac{0,50 - 0,34}{0,41} \doteq 12 + 1,56 = 13,56.$$

Na obr. 1.17 je interpolace mediánu znázorněna graficky.

Povšimněte si, že úsečky polygonu kumulativních četností leží „nad“ jednotlivými třídami; strmost těchto úseček je přitom úměrná třídním četnostem. Třída, nad kterou leží nejstrmější úsečka lokalizuje tedy modus.

Koeficient disperze

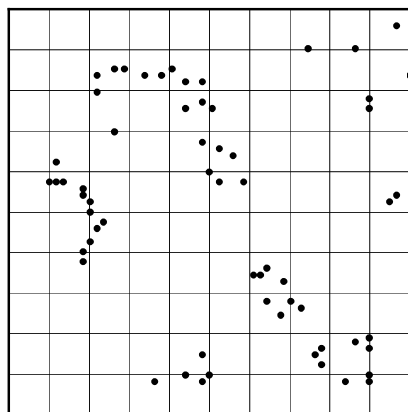
1.25. Definice a vlastnosti. *Koeficient disperze* je definován jako poměr rozptylu a aritmetického průměru a bývá používán v disciplíně nazývané *prostorová statistika* jako *míra agregovanosti* či *regularity* prostorových bodových struktur. Vysvětlíme nejprve příslušné pojmy.

Prostorem rozumíme libovolný Eukleidovský prostor nebo jeho část; může jít tedy o prostor třírozměrný, ale též dvojrozměrný (rovinu) nebo jednorozměrný (přímku). *Prostorová bodová struktura* (stručněji *bodová struktura*) je definována jako náhodné rozmístění bodů v prostoru. *Body* přitom obvykle reprezentují pozice určitých hmotných objektů (*jedinců*) či místa výskytu jistých náhodných událostí.

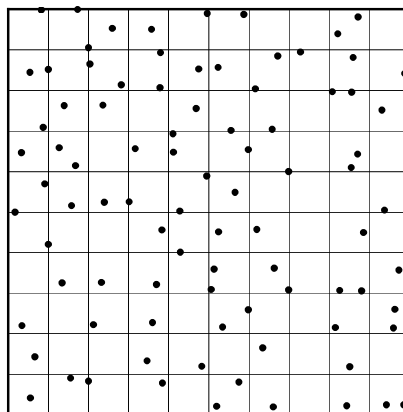
Jako příklad bodové struktury lze uvést rozmístění hvězd v galaxii, bakterií či molekul nějaké látky v ovzduší, květin na louce, stromů v lese či podél potoka, lidí v parku nebo vlaštovek na drátě. Typickým příkladem jednorozměrné bodové struktury je posloupnost okamžiků výskytu nějakých náhodných událostí; prostorem je v takovém případě časová osa. Zkoumanými událostmi mohou být třeba nehody na dálnici, poruchy jistého stroje, pracovní úrazy, příchody hovorů na telefonní ústřednu apod. Ve všech výše uvedených příkladech se může jedinec či událost vyskytnout v libovolném místě prostoru. Někdy je však tento výskyt omezen pouze na určitá oddělená místa v prostoru. S takovou situací se setkáme například při studiu rozmístění housenek či brouků na rostlinách, roztočů na listech apod.

Na obrázcích 1.18 a 1.19 je znázorněno rozmístění stromů na čtvercovém stanovišti; na každém z těchto obrázků vidíme tedy dvojrozměrnou bodovou strukturu. Povaha obou struktur je ovšem značně odlišná. Zatímco sekvoje na prvním obrázku vytvářejí dobře patrné shluky, rozmístění smrků na druhém obrázku je víceméně pravidelné, tj. s řádově srovnatelnými rozestupy mezi jedinci. V prvním případě mluvíme o *agregované* (nahloučené), kdežto v druhém o *regulární* (pravidelné) struktu-

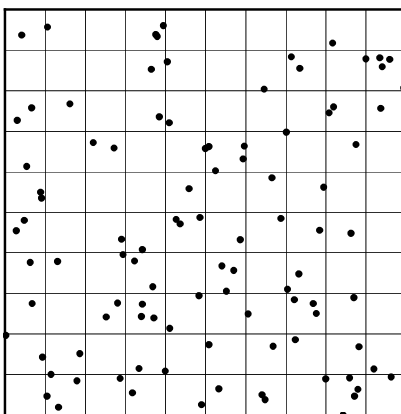
ře. Bodová struktura se nám ovšem nemusí jevit ani jako agregovaná, ani jako regulární; jako příklad může sloužit struktura na obr. 1.20. Poznamenejme přitom, že výskyt takových struktur je v přírodě poměrně vzácný.



Obr. 1.18. Prostorové rozmístění sekvojí (koeficient disperze je roven 1,9)



Obr. 1.19. Prostorové rozmístění smrků (koeficient disperze je roven 0,5)



Obr. 1.20. Výsledek zcela náhodného rozmístění bodů ve čtverci (koeficient disperze je roven 1,1)

Pokryjme nyní každé ze stanovišť na obrázcích 1.18 – 1.20 pravidelnou čtvercovou sítí (dejme tomu o rozměrech 10×10 čtverců) a spočítejme aritmetický průměr, rozptyl a koeficient disperze počtu jedinců v jednotlivých čtvercích (viz tabulky 1.5 – 1.7). Všimněte si, že v případě agregované struktury na obr. 1.18 je na rozdíl od struktur na obrázcích 1.19 a 1.20 relativně velké množství čtverců prázdných, zatímco poměrně málo jich obsahuje právě jednoho jedince. Tato skutečnost má zřejmě za následek poměrně vysokou hodnotu rozptylu počtu jedinců ve čtvercích, a tedy též vysokou hodnotu koeficientu disperze. Naopak ze všech tří struktur vykazuje nejmenší hodnotu rozptylu počtu jedinců ve čtvercích, a rovněž tak nejmenší hodnotu koeficientu disperze regulární struktura na obr. 1.19.

TAB. 1.5. Prostorové rozmístění sekvojí

Počet jedinců ve čtverci	0	1	2	3	4	5
Počet čtverců s daným počtem jedinců	68	14	9	7	1	1

$$\bar{x} = 0,62, s^2 \doteq 1,16, s^2/\bar{x} \doteq 1,9$$

TAB. 1.6. Prostorové rozmístění smrků

Počet jedinců ve čtverci	0	1	2	3	4	5
Počet čtverců s daným počtem jedinců	28	54	18	0	0	0

$$\bar{x} = 0,9, s^2 = 0,45, s^2/\bar{x} = 0,5$$

TAB. 1.7. Výsledek zcela náhodného rozmístění bodů ve čtverci

Počet bodů ve čtverci	0	1	2	3	4	5
Počet čtverců s daným počtem bodů	35	43	13	6	2	1

$$\bar{x} = 1, s^2 = 1,06, s^2/\bar{x} \doteq 1,1$$

Obecně lze ukázat, že při vhodné volbě velikosti čtverců je hodnota koeficientu disperze ve strukturách, které se jeví jako agregované, výrazně větší než jedna; čím více se přitom struktura zdá být agregovaná, tím větší je hodnota koeficientu disperze. Naopak v regulárních strukturách je koeficient disperze výrazně menší než jedna; čím více je přitom struktura regulární, tím je hodnota koeficientu disperze menší. Konečně ve strukturách, které se nejeví ani agregované, ani regulární se koeficient disperze neodlišuje příliš od jedničky.

1.26. Příklad (rozmístění roztočů na jabloňových listech). Vyšetřujeme prostorové rozmístění roztočů na listech jabloně na základě dat z příkladu 1.4. Prostorovými jednotkami, v nichž zaznamenáváme počet vyskytujících se jedinců (roztočů), nechť jsou jednotlivé jabloňové listy. Mají-li přitom uvedená data poskytnout smysluplnou informaci o prostorovém rozmístění roztočů, je třeba předpokládat, že všechny listy jsou (alespoň přibližně) stejně veliké. Aritmetický průměr a rozptyl počtu roztočů na listech je $\bar{x} \doteq 1,15$ a $s^2 \doteq 2,26$; koeficient disperze $s^2/\bar{x} \doteq 2,0$. Jde proto o značně agregovanou strukturu.

Průměry

Při řešení řady praktických úloh je třeba vypočítat průměrnou hodnotu čísel $x_1, x_2, \dots, x_n$, přičemž výsledkem nemusí být průměr aritmetický. Podáme dále definice některých často se vyskytujících typů průměrů a příklady jejich použití.

1.27. Kvadratický průměr. Necht' $x_1, x_2, \dots, x_n$ jsou kladná čísla. *Kvadratický průměr* $\bar{x}_K$ těchto čísel definujeme předpisem

$$\bar{x}_K = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}.$$

Je tedy

$$x_1^2 + x_2^2 + \dots + x_n^2 = n\bar{x}_K^2 = \underbrace{\bar{x}_K^2 + \bar{x}_K^2 + \dots + \bar{x}_K^2}_{n\text{-krát}},$$

což znamená, že kvadratickým průměrem můžeme nahradit jednotlivé hodnoty $x_1, x_2, \dots, x_n$ při výpočtu součtu jejich druhých mocnin.

1.28. Poznámka. Necht' $x_1, x_2, \dots, x_n$ jsou kladná čísla a s^2 je jejich rozptyl. Vzor (5) lze pak přepsat ve tvaru

$$s^2 = \bar{x}_K^2 - \bar{x}^2.$$

Je tedy

$$\bar{x}_K = \sqrt{\bar{x}^2 + s^2}.$$

Odtud plyne následující tvrzení:

1.29. Tvrzení. Necht' $x_1, x_2, \dots, x_n$ jsou kladná čísla. Pak $\bar{x}_K \geq \bar{x}$, přitom rovnost nastává právě tehdy, když $x_1 = x_2 = \dots = x_n$.

Uvědomte si též, že průměrná odchylka d_a je aritmetickým průměrem z odchylek čísel $x_1, x_2, \dots, x_n$ od jejich aritmetického průměru, zatímco směrodatná odchylka s je kvadratickým průměrem těchto odchylek. Nerovnost (6) mezi průměrnou odchylkou a odchylkou směrodatnou lze tedy považovat za speciální případ tvrzení 1.29.

1.30. Příklad (dendrometrický). V dendrometrii se s pojmem kvadratického průměru setkáváme při výpočtu průměrné (kruhové) výčetní základny. Představme si porost čítající n stromů s výčetními tloušťkami $d_1, d_2, \dots, d_n$. (*Výčetní tloušťkou* stromu rozumíme tloušťku změřenou v tzv. *prsni výšce*, tj. ve výšce 1,3 metru nad zemí.) Pro každý strom uvažme řez kmene rovinou vedenou v prsní výšce kolmo ke kmeni. Předpokládejme, že tento řez má pro všechny stromy kruhový tvar. Řez i -tým stromem je tedy kruh s průměrem d_i a obsahem $\frac{1}{4}\pi d_i^2$. Tento kruh se nazývá *kruhová výčetní základna* či stručněji *výčetní základna* (anglicky „*basal area*“). Obsah celkové výčetní základny, tj. hodnota součtu

$$\sum_{i=1}^n \frac{1}{4} \pi d_i^2$$

je veličinou, jejíž znalost je důležitá při odhadu objemu dřeva v porostu. Při výpočtu obsahu celkové výčetní základny můžeme ovšem výčetní základny jednotlivých stromů zastoupit *průměrnou kruhovou výčetní základnou*, tj. kruhem o obsahu

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{4} \pi d_i^2 .$$

Výčetní tloušťku stromu s průměrnou kruhovou výčetní základnou označme d . Zřejmě je

$$\frac{1}{4} \pi d^2 = \frac{1}{n} \sum_{i=1}^n \frac{1}{4} \pi d_i^2 ,$$

neboli

$$(7) \quad n \cdot \frac{1}{4} \pi d^2 = \sum_{i=1}^n \frac{1}{4} \pi d_i^2 .$$

Rovnost (7) lze interpretovat tak, že *obsah celkové výčetní základny stromů s výčetními tloušťkami $d_1, d_2, \dots, d_n$ je stejný jako obsah celkové výčetní základny n stejně tlustých stromů s výčetní tloušťkou d* .

Tloušťka d není ovšem aritmetickým průměrem tlouštěk $d_1, d_2, \dots, d_n$. Z rovnosti (7) totiž postupně dostaneme

$$nd^2 = \sum_{i=1}^n d_i^2 , \quad d^2 = \frac{1}{n} \sum_{i=1}^n d_i^2 , \quad d = \sqrt{\frac{1}{n} \sum_{i=1}^n d_i^2} = \bar{d}_K .$$

Jinak řečeno, *výčetní tloušťka stromu s průměrnou kruhovou výčetní základnou je kvadratickým průměrem výčetních tlouštěk jednotlivých stromů v porostu*. Označíme-li tedy po řadě $\bar{d}$ a s_d^2 aritmetický průměr a rozptyl čísel $d_1, d_2, \dots, d_n$, pak

$$d = \sqrt{\bar{d}^2 + s_d^2} ,$$

a proto vždy $d \geq \bar{d}$.

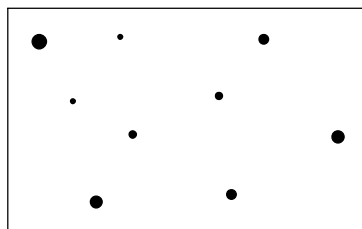
Pro konkrétní ilustraci uvažme stanoviště s devíti stromy, jejichž prostorové rozmístění včetně výčetních kruhových základů je znázorněno na obr. 1.21. Numerické hodnoty výčetních tlouštěk jednotlivých stromů (v centimetrech) nechť jsou přitom následující:

20 20 30 30 40 40 50 50 60

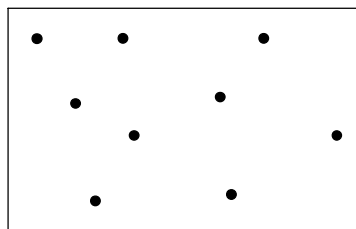
Na obr. 1.22 jsou pak znázorněny výčetní kruhové základny stejně tlustých stromů zaujímajících na daném stanovišti tutéž polohu jako stromy na obr. 1.21. Obsah celkové výčetní základny je přitom na obou obrázcích stejný. Jinak řečeno, výčetní základny stromů na obr. 1.22 jsou aritmetickým průměrem výčetních základů stromů na obr. 1.21. Tloušťka stromů na obr. 1.22 je tedy kvadratickým průměrem tlouštěk stromů na obr. 1.21. Hodnota této tloušťky je

$$\sqrt{\frac{20^2 + 20^2 + 30^2 + 30^2 + 40^2 + 40^2 + 50^2 + 50^2 + 60^2}{9}} = 40$$

centimetrů



Obr. 1.21



Obr. 1.22

1.31. Geometrický průměr. Necht' $x_1, x_2, \dots, x_n$ jsou kladná čísla. *Geometrický průměr* $\bar{x}_G$ těchto čísel definujeme předpisem

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

Je tedy

$$(8) \quad x_1 \cdot x_2 \cdot \dots \cdot x_n = \bar{x}_G^n = \underbrace{\bar{x}_G \cdot \bar{x}_G \cdot \dots \cdot \bar{x}_G}_{n\text{-krát}},$$

což znamená, že geometrickým průměrem můžeme nahradit jednotlivé hodnoty $x_1, x_2, \dots, x_n$ při výpočtu jejich součinu. Z rovnosti (8) plyne, že

$$\ln x_1 + \ln x_2 + \dots + \ln x_n = \underbrace{\ln \bar{x}_G + \ln \bar{x}_G + \dots + \ln \bar{x}_G}_{n\text{-krát}},$$

čili

$$(9) \quad \ln \bar{x}_G = \frac{\ln x_1 + \ln x_2 + \dots + \ln x_n}{n}.$$

To znamená, že *logaritmus geometrického průměru čísel $x_1, x_2, \dots, x_n$ je roven aritmetickému průměru logaritmů těchto čísel*. Rovnost (9) bývá používána při numerickém výpočtu geometrického průměru a platí zřejmě pro logaritmus o libovolném základu.

1.32. Úloha (bankovní). Určete celkovou naspořenou částku z vkladu 60 000 Kč po pěti letech, jestliže vklad měl roční úročení a úroková míra činila v prvním roce 4%, ve druhém 8%, ve třetím 6% a ve čtvrtém a pátém roce 12%. Určete též průměrnou úrokovou míru během celého pětiletého období.

Řešení. Naspořená částka na konci pětiletého období činila

$$60000 \cdot 1,04 \cdot 1,08 \cdot 1,06 \cdot 1,12 \cdot 1,12 \doteq 89608,72 \text{ Kč}.$$

Označme p průměrnou úrokovou míru (v %) a položme $r = 1 + \frac{p}{100}$. Význam čísla p je takový, že v případě pevné roční úrokové míry $p\%$ by celková naspořená částka na konci pětiletého období byla stejná jako při výše popsané pohyblivé úrokové míře. Je tedy

$$60000 \cdot r \cdot r \cdot r \cdot r \cdot r = 60000 \cdot 1,04 \cdot 1,08 \cdot 1,06 \cdot 1,12 \cdot 1,12.$$

To ale znamená, že

$$r \cdot r \cdot r \cdot r \cdot r = 1,04 \cdot 1,08 \cdot 1,06 \cdot 1,12 \cdot 1,12,$$

čili

$$r = \sqrt[5]{1,04 \cdot 1,08 \cdot 1,06 \cdot 1,12 \cdot 1,12}.$$

Číslo r je tedy geometrickým průměrem čísel 1,04; 1,08; 1,06; 1,12; 1,12. Vyjde $r \approx 1,0835$. Průměrná úroková míra činila tedy asi 8,35%. Aritmetický průměr procentuálních úrokových měr, tj. čísel 4, 8, 6, 12, 12, je přitom 8,4, tedy větší než je správně vypočtených 8,35.

Poznamenejme, že podobným způsobem by se počítala též průměrná fertilita, mortalita či růstová intenzita v dané populaci.

1.33. Příklad (dendrometrický). V příkladu 1.30 byl zaveden pojem kruhové výčetní základny stromu (kmene). Chceme-li být více realističtí, můžeme předpokládat, že tato základna není kruhová, nýbrž že má tvar elipsy. Dejme tomu, že umíme odhadnout osy této elipsy, tj. dva navzájem kolmé směry, ve kterých má kmen nejmenší a největší výčetní tloušťku. Změřme tyto tloušťky a označme jejich velikosti d_1 a d_2 . Obsah elipsy s průměry d_1 a d_2 je jak známo roven

$$\frac{1}{4} \pi d_1 d_2.$$

Trváme-li ovšem na tom, že obsah výčetní základny chceme počítat jako obsah kruhu, je třeba jeho průměr d zvolit tak, aby tento kruh a elipsa s průměry d_1 a d_2 měly stejný obsah. To vede k rovnici

$$\frac{1}{4} \pi d^2 = \frac{1}{4} \pi d_1 d_2,$$

z níž dále plyne, že

$$d^2 = d_1 d_2 \quad \text{a} \quad d = \sqrt{d_1 d_2}.$$

Číslo d je tedy geometrickým průměrem čísel d_1 a d_2 .

Závěr: Provádíme-li měření tloušťek stromů ve dvou navzájem kolmých směrech, tato měření jsou prováděna za účelem výpočtu obsahu výčetní základny a obě naměřené tloušťky nahrazujeme z úsporných důvodů jedinou (průměrnou) hodnotou, je třeba použít průměr geometrický (a nikoli aritmetický!)

1.34. Harmonický průměr. Necht' $x_1, x_2, \dots, x_n$ jsou kladná čísla. Harmonický průměr $\bar{x}_H$ těchto čísel definujeme předpisem

$$\bar{x}_H = \frac{1}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} .$$

Je tedy

$$\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} = \frac{n}{\bar{x}_H} = \underbrace{\frac{1}{\bar{x}_H} + \frac{1}{\bar{x}_H} + \dots + \frac{1}{\bar{x}_H}}_{n\text{-krát}} ,$$

což znamená, že harmonickým průměrem můžeme nahradit jednotlivé hodnoty $x_1, x_2, \dots, x_n$ při výpočtu součtu jejich převrácených hodnot. Poznamenejme ještě, že lze psát

$$\bar{x}_H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

a že harmonický průměr dvou kladných čísel x, y je

$$\frac{1}{\frac{\frac{1}{x} + \frac{1}{y}}{2}} = \frac{2}{\frac{1}{x} + \frac{1}{y}} = \frac{2xy}{x+y} .$$

1.35. Úloha (dopravní). Předpokládejme, že automobil jede nejprve do kopce rychlostí čtyřicet km/hod a poté stejnou trasou zpátky rychlostí osmdesát km/hod. Jaká je průměrná rychlost automobilu během této projížďky?

Řešení. Průměrnou rychlostí rozumíme takovou rychlost v (km/hod), že jízda, při níž bychom celou trasu projeli tam i zpět touto rychlostí, by trvala stejně dlouho jako jízda čtyřicetikilometrovou rychlostí do kopce následovaná jízdou osmdesátakilometrovou rychlostí z kopce. Necht' s je délka trasy (v jednom směru) v kilometrech. Porovnáním časů při rovnoměrném a nerovnoměrném způsobu jízdy obdržíme rovnici

$$\frac{s}{v} + \frac{s}{v} = \frac{s}{40} + \frac{s}{80} .$$

Odtud vyplývá, že

$$\frac{1}{v} + \frac{1}{v} = \frac{1}{40} + \frac{1}{80}, \quad \frac{1}{v} = \frac{\frac{1}{40} + \frac{1}{80}}{2}$$

a

$$v = \frac{1}{\frac{\frac{1}{40} + \frac{1}{80}}{2}} .$$

Rychlost v je tudíž harmonickým průměrem čísel 40 a 80. Vyjde

$$v = \frac{2 \cdot 40 \cdot 80}{40 + 80} = 53,3 \text{ km/hod}.$$

Vypočítaná průměrná rychlost je menší než aritmetický průměr čísel 40 a 80, což je v souladu se skutečností, že menší rychlostí (do kopce) se jelo déle.

1.36. Poznámka. Necht' $x_1, x_2, \dots, x_n$ jsou kladná čísla, přičemž $x_{(1)}$ je nejmenší a $x_{(n)}$ největší z nich. Lze ukázat, že platí následující nerovnosti:

$$(10) \quad x_{(1)} \leq \bar{x}_H \leq \bar{x}_G \leq \bar{x} \leq \bar{x}_K \leq x_{(n)}.$$

Jestliže je přitom $x_1 = x_2 = \dots = x_n$, pak všechny nerovnosti v (10) přecházejí v rovnosti. Naopak nejsou-li všechna čísla $x_1, x_2, \dots, x_n$ stejná, pak jsou všechny nerovnosti v (10) ostré.

1.37. Průměr stupně α . Všechny výše definované typy průměrů lze považovat za speciální případy tzv. průměru stupně α . Konkrétně necht' $\alpha \neq 0$ je dané reálné číslo a $x_1, x_2, \dots, x_n$ jsou kladná čísla (reprezentující hromadná data). *Průměr stupně α* z čísel $x_1, x_2, \dots, x_n$ definujeme předpisem

$$(11) \quad \bar{x}_\alpha = \left(\frac{1}{n} \sum_{i=1}^n x_i^\alpha \right)^{1/\alpha}.$$

Okamžitě vidíme, že aritmetický průměr je průměrem stupně jedna, kvadratický průměr je průměrem stupně dva a harmonický průměr je průměrem stupně -1 .

Ze vztahu (11) plyne, že

$$\bar{x}_\alpha^\alpha = \frac{1}{n} \sum_{i=1}^n x_i^\alpha \quad \text{a} \quad n\bar{x}_\alpha^\alpha = \sum_{i=1}^n x_i^\alpha.$$

Poslední vztah lze psát jako

$$x_1^\alpha + \dots + x_n^\alpha = \underbrace{\bar{x}_\alpha^\alpha + \dots + \bar{x}_\alpha^\alpha}_{n\text{-krát}},$$

což znamená, že průměrem stupně α můžeme nahradit jednotlivé hodnoty $x_1, x_2, \dots, x_n$ při výpočtu součtu jejich α -tých mocnin.

1.38. Příklad. Uvažme soubor n borůvek sesbíraných na dané lokalitě. Předpokládejme, že borůvky mají kulový tvar a že známe poloměry $r_1, r_2, \dots, r_n$ jednotlivých borůvek. Chceme určit poloměr borůvky s průměrným objemem. Označme tento poloměr r . Zřejmě platí:

$$\frac{4}{3}\pi r^3 = \frac{1}{n} \sum_{i=1}^n \frac{4}{3}\pi r_i^3.$$

Odtud pak plyne, že

$$r^3 = \frac{1}{n} \sum_{i=1}^n r_i^3, \quad \text{resp. } nr^3 = \sum_{i=1}^n r_i^3,$$

čili

$$r = \sqrt[3]{\frac{1}{n} \sum_{i=1}^n r_i^3} = \left(\frac{1}{n} \sum_{i=1}^n r_i^3 \right)^{1/3}.$$

Jinak řečeno, *poloměr r průměrně objemné borůvky je roven průměru třetího stupně z poloměrů $r_1, r_2, \dots, r_n$ jednotlivých borůvek*. Tento průměr zastupuje čísla $r_1, r_2, \dots, r_n$ při sčítání jejich třetích mocnin.

Definici průměru stupně α nelze bezprostředně použít pro případ $\alpha = 0$. Pak totiž na pravé straně rovnosti (11) stojí neurčitý výraz typu 1^∞ . Je ovšem přirozené definovat *průměr stupně nula* jako limitní hodnotu výrazu (11) pro $\alpha \rightarrow 0$, tj. předpisem

$$\bar{x}_0 = \lim_{\alpha \rightarrow 0} \bar{x}_\alpha.$$

Ukážeme nyní, že pro libovolný soubor kladných čísel $x_1, x_2, \dots, x_n$ tato limita existuje a určíme její hodnotu.

1.39. Tvzení (o průměru stupně nula). *Nechť $x_1, x_2, \dots, x_n$ jsou pevně daná kladná čísla. Pak*

$$\lim_{\alpha \rightarrow 0} \bar{x}_\alpha = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}.$$

(Za průměr stupně nula je tedy přirozené považovat průměr geometrický.)

Důkaz. Dle definice obecné mocniny je

$$\bar{x}_\alpha = \left(\frac{1}{n} \sum_{i=1}^n x_i^\alpha \right)^{1/\alpha} = e^{(1/\alpha) \ln \left(\frac{1}{n} \sum_{i=1}^n x_i^\alpha \right)}.$$

Užitím l'Hospitalova pravidla dostaneme

$$\lim_{\alpha \rightarrow 0} \frac{\ln \left(\frac{1}{n} \sum_{i=1}^n x_i^\alpha \right)}{\alpha} = \lim_{\alpha \rightarrow 0} \frac{\sum_{i=1}^n x_i^\alpha \ln x_i}{\sum_{i=1}^n x_i^\alpha} = \frac{\sum_{i=1}^n \ln x_i}{n} = \ln \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}.$$

Tudíž dle věty o limitě složené funkce

$$\lim_{\alpha \rightarrow 0} \bar{x}_\alpha = e^{\ln \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} ,$$

což bylo dokázat. $\square$

1.40. Průměry stupně $\pm\infty$. Necht' $x_1, x_2, \dots, x_n$ jsou kladná čísla, přičemž $x_{(1)}$ je nejmenší a $x_{(n)}$ největší z nich. Pro libovolné reálné číslo α zřejmě platí, že

$$(12) \quad x_{(1)} \leq \bar{x}_\alpha \leq x_{(n)} .$$

Vhodnou volbou čísla α se lze přitom k mezím $x_{(1)}$ a $x_{(n)}$ libovolně přiblížit. Platí totiž:

$$(13) \quad \lim_{\alpha \rightarrow -\infty} \bar{x}_\alpha = x_{(1)} \quad \text{a} \quad \lim_{\alpha \rightarrow \infty} \bar{x}_\alpha = x_{(n)} .$$

Je tedy přirozené považovat číslo $x_{(1)}$ za průměr stupně $-\infty$ a číslo $x_{(n)}$ za průměr stupně ∞ .

Nerovnosti (10) a (12) jsou speciálním případem věty (viz 1.41), která říká, že *pro pevně daný soubor čísel $x_1, x_2, \dots, x_n$ roste hodnota průměru $\bar{x}_\alpha$ s rostoucí hodnotou stupně α . Mění-li se přitom stupeň α spojitě, mění se i hodnota průměru $\bar{x}_\alpha$ spojitě; průměr $\bar{x}_\alpha$ nabude tedy pro vhodnou hodnotu stupně α libovolné hodnoty z intervalu $[x_{(1)}, x_{(n)}]$.*

Dodatky

1.41. Věta (o nerovnostech mezi průměry). *Necht' $x_1, x_2, \dots, x_n$ je pevně daný soubor kladných reálných čísel, přičemž tato čísla nejsou všechna stejná. Označme $\bar{x}_\alpha$ průměr stupně α z čísel $x_1, x_2, \dots, x_n$; přitom klademe*

$$\bar{x}_0 = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} , \quad \bar{x}_{-\infty} = x_{(1)} \quad \text{a} \quad \bar{x}_\infty = x_{(n)} ,$$

kde $x_{(1)}$ je minimum a $x_{(n)}$ maximum čísel $x_1, x_2, \dots, x_n$. Tímto způsobem je na rozšířené reálné ose $[-\infty, \infty]$ definována reálná funkce

$$\alpha \rightarrow \bar{x}_\alpha$$

s hodnotami v intervalu $[x_{(1)}, x_{(n)}]$. Tato funkce je spojitá a rostoucí.

Důkaz^{*}. Jelikož obecná mocnina je spojitá funkce a součet i složení spojitých funk-

cí je spojitá funkce, je přiřazení $\alpha \rightarrow \bar{x}_\alpha$ v intervalu $(-\infty, 0)$, a rovněž tak v intervalu $(0, \infty)$, spojitou funkcí. Jelikož však v bodech $0, -\infty, \infty$ je průměr $\bar{x}_\alpha$ dodefinován limitou, je toto přiřazení spojitou funkcí v celé rozšířené reálné ose.

Ukážeme nyní, že přiřazení $\alpha \rightarrow \bar{x}_\alpha$ je funkcí rostoucí, tj. že platí:

$$-\infty \leq \alpha < \beta \leq +\infty \Rightarrow \bar{x}_\alpha < \bar{x}_\beta.$$

(I) Nejprve ukážeme, že pro $\alpha > 1$ je $\bar{x}_\alpha > \bar{x}_1$. Budeme přitom dokazovat zesílení tohoto tvrzení pro průměry vážené. Necht' tedy $w_1, w_2, \dots, w_n$ jsou kladná čísla (váhy) taková, že $\sum_{i=1}^n w_i = 1$. Chceme ukázat, že

$$(14) \quad \left(\sum_{i=1}^n w_i x_i^\alpha \right)^{1/\alpha} > \sum_{i=1}^n w_i x_i.$$

(Volbou $w_1 = w_2 = \dots = w_n = \frac{1}{n}$ odtud obdržíme nerovnost $\bar{x}_\alpha > \bar{x}_1$.) Vzhledem k tomu, že číslo α je kladné, je nerovnost (14) ekvivalentní s nerovností

$$(15) \quad \sum_{i=1}^n w_i x_i^\alpha > \left(\sum_{i=1}^n w_i x_i \right)^\alpha.$$

Důkaz nerovnosti (15) provedeme indukcí dle n .

(a) Necht' $n = 2$. Máme ukázat, že pro libovolná dvě kladná čísla γ, δ taková, že $\gamma + \delta = 1$ a pro libovolná dvě vzájemně různá kladná čísla x_1, x_2 je

$$\gamma x_1^\alpha + \delta x_2^\alpha > (\gamma x_1 + \delta x_2)^\alpha.$$

Za tím účelem zkoumejme funkci $f(x) = x^\alpha$ v proměnné x , kde $x \in (0, \infty)$. Jelikož $\alpha > 1$, je tato funkce v celém svém definičním oboru konvexní, a tedy pro libovolná dvě různá kladná čísla x_1, x_2 leží všechny vnitřní body úsečky s krajními body $[x_1, f(x_1)]$ a $[x_2, f(x_2)]$ „nad“ grafem funkce $f(x)$. To ale znamená, že

$$\gamma f(x_1) + \delta f(x_2) > f(\gamma x_1 + \delta x_2),$$

což bylo dokázat.

(b) Předpokládejme nyní, že pro nějaké přirozené číslo n je již nerovnost (15) dokázána. Necht' $x_1, \dots, x_n, x_{n+1}$ jsou kladná čísla, přičemž alespoň dvě z nich jsou vzájemně různá. Dále necht' $w_1, \dots, w_n, w_{n+1}$ jsou kladná čísla taková, že $\sum_{i=1}^{n+1} w_i = 1$.

Lze psát

$$w_1 x_1^\alpha + \dots + w_n x_n^\alpha + w_{n+1} x_{n+1}^\alpha = (1 - w_{n+1}) \left(\frac{w_1}{1 - w_{n+1}} \cdot x_1^\alpha + \dots + \frac{w_n}{1 - w_{n+1}} \cdot x_n^\alpha \right) + w_{n+1} x_{n+1}^\alpha ,$$

přitom dle indukčního předpokladu

$$\frac{w_1}{1 - w_{n+1}} \cdot x_1^\alpha + \dots + \frac{w_n}{1 - w_{n+1}} \cdot x_n^\alpha \geq \left(\frac{w_1}{1 - w_{n+1}} \cdot x_1 + \dots + \frac{w_n}{1 - w_{n+1}} \cdot x_n \right)^\alpha .$$

Vezmeme-li tedy v úvahu fakt, že pro $n = 2$ je již nerovnost (15) dokázána, dostaneme

$$\begin{aligned} & w_1 x_1^\alpha + \dots + w_n x_n^\alpha + w_{n+1} x_{n+1}^\alpha \\ & \geq (1 - w_{n+1}) \cdot \left(\frac{w_1}{1 - w_{n+1}} \cdot x_1 + \dots + \frac{w_n}{1 - w_{n+1}} \cdot x_n \right)^\alpha + w_{n+1} x_{n+1}^\alpha \\ & \geq \left[(1 - w_{n+1}) \cdot \left(\frac{w_1}{1 - w_{n+1}} \cdot x_1 + \dots + \frac{w_n}{1 - w_{n+1}} \cdot x_n \right) + w_{n+1} x_{n+1} \right]^\alpha \\ & = (w_1 x_1 + \dots + w_n x_n + w_{n+1} x_{n+1})^\alpha , \end{aligned}$$

přitom alespoň jedna z předchozích nerovností musí být ostrá (rozmyslete si proč). Tím je proveden indukční krok, a tedy i důkaz nerovnosti (15).

(II) Ukážeme, že $\bar{x}_\alpha < \bar{x}_\beta$, pokud $0 < \alpha < \beta < \infty$. To je ale téměř bezprostřední důsledek nerovností mezi průměry dokázaných v části (I). Je totiž $\beta/\alpha > 1$, a tedy dle (I)

$$\left(\frac{x_1^\beta + \dots + x_n^\beta}{n} \right)^{\alpha/\beta} = \left(\frac{[x_1^\alpha]^{\beta/\alpha} + \dots + [x_n^\alpha]^{\beta/\alpha}}{n} \right)^{\alpha/\beta} > \frac{x_1^\alpha + \dots + x_n^\alpha}{n} .$$

Odtud pak

$$\left(\frac{x_1^\beta + \dots + x_n^\beta}{n} \right)^{1/\beta} > \left(\frac{x_1^\alpha + \dots + x_n^\alpha}{n} \right)^{1/\alpha} ,$$

což bylo dokázat.

Přechodem k limitám pro $\alpha \rightarrow 0$ a $\alpha \rightarrow \infty$ dále dostaneme, že $\bar{x}_\alpha < \bar{x}_\beta$ pro $0 \leq \alpha < \beta \leq \infty$.

(III) Ukážeme, že $\bar{x}_\alpha < \bar{x}_\beta$, pokud $-\infty < \alpha < \beta < 0$. To ale ihned vyplývá z nerovností mezi průměry dokázaných v části (II). Je totiž $0 < -\beta < -\alpha < \infty$, a tedy dle (II)

$$\begin{aligned} \left(\frac{x_1^\beta + \dots + x_n^\beta}{n} \right)^{-1/\beta} &= \left(\frac{[x_1^{-1}]^{-\beta} + \dots + [x_n^{-1}]^{-\beta}}{n} \right)^{-1/\beta} \\ &< \left(\frac{[x_1^{-1}]^{-\alpha} + \dots + [x_n^{-1}]^{-\alpha}}{n} \right)^{-1/\alpha} = \left(\frac{x_1^\alpha + \dots + x_n^\alpha}{n} \right)^{-1/\alpha}. \end{aligned}$$

Přechodem k převráceným hodnotám dostaneme, že $\bar{x}_\alpha < \bar{x}_\beta$. Přechodem k limitám pro $\alpha \rightarrow 0$ a $\alpha \rightarrow \infty$ dále dostaneme, že $\bar{x}_\alpha < \bar{x}_\beta$ pro $-\infty \leq \alpha < \beta \leq 0$.

(IV) Spojením nerovností dokázaných v (II) a (III) obdržíme dokazovanou větu. $\square$

V principu by bylo možné definovat průměrnou odchylku i směrodatnou odchylku i od jiné centrální hodnoty než od aritmetického průměru. V jistém smyslu nejlepší volba této centrální hodnoty je taková, pro niž příslušná odchylka nabývá minimální hodnoty. Vzniká otázka, zda aritmetický průměr má tuto vlastnost. Odpověď dávají následující dvě tvrzení.

1.42. Tvrzení. *Nechť $x_1, x_2, \dots, x_n$ je pevně daný soubor čísel. Pak funkce*

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

nabývá v bodě $x = \bar{x}$ svého minima.

Důkaz. Snadno nahlédneme, že

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

což znamená, že $f(x)$ je kvadratická funkce a jejím grafem je parabola. Vrchol V této paraboly určíme doplněním na úplný čtverec. Konkrétně

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

kde s^2 je rozptyl čísel $x_1, x_2, \dots, x_n$. Je tedy $V = (\bar{x}, s^2)$, což bylo dokázat.

(Přirozeně bylo též možno vypočítat derivaci funkce $f(x)$ a ptát se, kdy je tato derivace nulová.) $\square$

1.43. Tvrzení. *Nechť $x_1, x_2, \dots, x_n$ je pevně daný soubor čísel. Pak funkce*

$$f(x) = \frac{1}{n} \sum_{i=1}^n |x_i - x|$$

nabývá v bodě $x = \tilde{x}$ svého minima.

Důkaz. Přenecháváme jej čtenáři jako cvičení. $\square$

Vidíme tedy, že je správné, když směrodatná odchylka se definuje jako odchylka od aritmetického průměru, na druhou stranu průměrná odchylka by měla být definována spíše jako odchylka od mediánu.

1.44. Tvzení. *Necht' R je rozpětí a s směrodatná odchylka souboru čísel $x_1, x_2, \dots, x_n$. Pak $s \leq R/2$.*

Důkaz. Označme $x_{(1)}$ nejmenší a $x_{(n)}$ největší z čísel $x_1, x_2, \dots, x_n$ a položme $x = (x_{(1)} + x_{(n)})/2$. Bod x je středem úsečky $[x_{(1)}, x_{(n)}]$, a proto $|x_i - x| \leq R/2$ pro libovolné z čísel x_i . Použijeme-li navíc tvrzení 1.42, dostaneme, že

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \leq \frac{1}{n} \sum_{i=1}^n (x_i - x)^2 \leq \frac{R^2}{4}.$$

Odtud již bezprostředně plyne dokazovaná nerovnost. $\square$

1.45. Samuelsonova nerovnost. *Necht' s je směrodatná odchylka souboru čísel $x_1, x_2, \dots, x_n$. Pak*

$$\max_i |x_i - \bar{x}| \leq s\sqrt{n-1}.$$

Důkaz. Viz []. $\square$

Vzhledem k tomu, že je zřejmě $R \leq 2 \cdot \max_i |x_i - \bar{x}|$, obdržíme spojením tvrzení 1.44 a Samuelsonovy nerovnosti následující vztahy mezi směrodatnou odchylkou a rozpětím:

$$(16) \quad 2s \leq R \leq 2s\sqrt{n-1}, \quad \text{resp.} \quad R/2\sqrt{n-1} \leq s \leq R/2.$$

Cvičení

1. Při kontrole jakosti bylo náhodně vybráno devět výrobků; jejich hmotnosti (v gramech) jsou přitom následující:

43,0 51,2 49,7 48,1 53,8 49,8 53,0 47,0 49,1

Určete aritmetický průměr, rozptyl, směrodatnou odchylku a průměrnou odchylku zaznamenaných hmotností.

Výsledek: $\bar{x} \doteq 49,41$; $s^2 \doteq 9,37$; $s \doteq 3,06$; $d_a \doteq 2,32$

2. Ve dvanáctičlenné studijní skupině bylo při zápočtovém testu dosaženo následujících bodových výsledků (maximální možný počet bodů je roven deseti):

3 5 7 10 10 10 10 8 10 0 8 3

Vypočítejte modus, medián a aritmetický průměr zaznamenaných výsledků.

Výsledek: $\hat{x} = 10$, $\tilde{x} = 8$, $\bar{x} = 7$

3. Uveďte příklad pěti vzájemně různých kladných čísel vyhovujících současně následujícím dvěma podmínkám:

- (1) aritmetický průměr čísel je menší než jejich medián,
- (2) součet všech čísel je roven jedné.

- 4*. Dokažte tvrzení 1.43.

5. Určete medián a aritmetický průměr všech lichých přirozených čísel menších než jeden tisíc.

6. Datový soubor sestává z deseti čísel, přičemž platí:

- (1) součet všech čísel v souboru je roven dvaceti,
- (2) součet jejich druhých mocnin je dvě stě.

Vypočítejte směrodatnou odchylku.

Výsledek: $s = 4$

7. Jak se změní modus, medián, aritmetický průměr, rozpětí, průměrná odchylka, rozptyl, směrodatná odchylka a variační koeficient souboru čísel $x_1, x_2, \dots, x_n$, jestliže:

- a) všechna tato čísla vynásobíme dvěma,
- b) u všech čísel změníme znaménko,
- c) všechna čísla zvětšíme o deset jednotek?

8. Jak se změní průměr stupně α souboru kladných čísel $x_1, x_2, \dots, x_n$, jestliže všechna tato čísla vynásobíme kladnou konstantou c ?

9. Vypočítejte aritmetický, harmonický a geometrický průměr, průměr druhého stupně a rozptyl následujícího souboru čísel:

$$\underbrace{1, 1, \dots, 1}_{30\text{-krát}}, \underbrace{2, 2, \dots, 2}_{60\text{-krát}}, \underbrace{3, 3, \dots, 3}_{90\text{-krát}}$$

Proveďte zkoušku správnosti seřazením vypočtených průměrů podle velikosti. Zdůvodněte, proč zcela stejný výsledek obdržíme pro soubor čísel 1, 2, 2, 3, 3, 3.

Výsledek: $\bar{x} = 2, \bar{3}$; $\bar{x}_g \doteq 2,18$; $\bar{x}_h = 2$; $\bar{x}_2 \doteq 2,45$; $s^2 = \bar{x}_2^2 - \bar{x}^2 = 5/9$

10. Dokažte elementárním způsobem, že aritmetický průměr dvou kladných čísel x, y je vždy alespoň tak velký jako jejich průměr geometrický. Přechodem k převráceným hodnotám odtud odvoďte nerovnost mezi průměrem geometrickým a harmonickým.
11. a) V rovině jsou narysovány čtverec a obdélník o stejném obvodu. Dokažte, že obsah čtverce je větší než obsah obdélníka.
 b*) V rovině jsou narysovány kružnice a elipsa stejné délky. Dokažte, že plocha ohraničená kružnicí má větší obsah než plocha ohraničená elipsou.
 c*) Dokažte, že ze všech trojúhelníků o stejném obvodu má největší obsah trojúhelník rovnostranný.
 d*) Dokažte, že ze všech kvádrů o stejném objemu má nejmenší povrch krychle.

Návod: Použijte větu o nerovnostech mezi průměry.

- 12*. Na první pohled by se mohlo zdát, že výrazně citlivější míru rozptýlení čísel $x_1, x_2, \dots, x_n$ na reálné ose než jakou je jejich rozptyl obdržíme tak, že spočítáme průměrnou hodnotu druhých mocnin vzdáleností $(x_i - x_j)^2$ všech dvojic čísel $\{x_i, x_j\}$ ze souboru $x_1, x_2, \dots, x_n$. Ve skutečnosti je však v takovéto charakteristice obsažena stejná informace jako v námi definovaném rozptylu. Platí totiž, že

$$\sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2 = 2n \sum_{i=1}^n (x_i - \bar{x})^2.$$

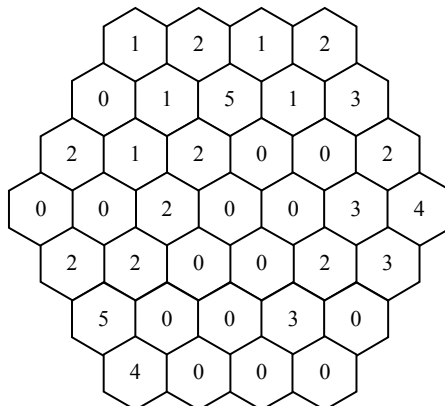
Dokažte to.

13. **Prostorové rozmístění stromů v porostu.** Šestihektarový borový porost byl rozdělen na šest set stejně velikých, vzájemně se nepřekrývajících částí („čtverců“). Počty stromů v jednotlivých čtvercích jsou zaznamenány v následující tabulce:

Počet stromů ve čtverci	0	1	2	3	4	5	6
Počet čtverců s daným počtem stromů	42	130	208	151	60	8	1

- a) Znázorněte rozdělení počtu stromů ve čtvercích tyčkovým diagramem.
 b) Vypočítejte koeficient disperze a interpretujte získaný výsledek.

Výsledek: $\bar{x} \doteq 2,14$; $s^2 \doteq 1,27$; $s^2/\bar{x} \doteq 0,6$; stromy jsou na daném stanovišti rozmístěny velmi pravidelně.

14. Prostorové rozmístění velkých stínek (*Philoscia muscorum*).


Na obrázku je zaznamenán výsledek analýzy prostorového rozmístění stínek ve spadaném listí a humusu v části bukového háje poblíž Oxfordu. Studovaná plocha byla pokryta pravidelnou šestiúhelníkovou sítí s šířkou šestiúhelníku jedna stopa (0,30 m) a poté byl spočítán počet stínek připadajících na jeden šestiúhelník. (Data jsou převzata z článku ‘Mean crowding’ od M. Lloyda otištěného v roce 1967 v časopisu *Journal of Animal Ecology*.) Vypočítejte koeficient disperze a interpretujte získaný výsledek.

Výsledek: $\bar{x} = 53/37 \doteq 1,43$; $s^2 \doteq 2,25$; $s^2/\bar{x} \doteq 1,6$; prostorová struktura je poměrně značně agregovaná.

15. Pracovní úrazy. V následující tabulce je zaznamenán počet pracovních úrazů v určitém úseku hlubinného dolu připadajících na jednu směnu:

Počet úrazů během směny	0	1	2	3	4	5	6
Počet směn s daným počtem úrazů	161	40	11	7	1	1	1

Prezentujte získaná data pomocí tyčkového diagramu. Dále vypočítejte koeficient disperze počtu úrazů připadajících na jednu směnu a interpretujte získaný výsledek.

Výsledek: $\bar{x} \doteq 0,44$; $s^2 \doteq 0,81$; $s^2/\bar{x} \doteq 1,8$; vysoká hodnota koeficientu disperze prozrazuje, že úrazy nejsou patrně zcela náhodnými událostmi.

16. Počet bliznových laloků makovice.

Počet laloků	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Počet makovic	3	11	38	106	152	238	305	315	302	234	128	50	19	3	1

Vypočítejte modus, medián, aritmetický průměr a směrodatnou odchylku počtu bliznových laloků a prezentujte data pomocí tyčkového diagramu.

Výsledek: $\hat{x} = 13$, $\tilde{x} = 13$; $\bar{x} \doteq 12,76$; $s \doteq 2,24$

17. Počet lístků na listech jasanu.

Počet lístků	3	5	7	9	11	13	15
Počet listů s daným počtem lístků	8	142	876	2674	2947	753	59

Vypočítejte modus, medián, aritmetický průměr a směrodatnou odchylku počtu lístků na listu.

Výsledek: $\hat{x} = 11$, $\tilde{x} = 11$; $\bar{x} \doteq 9,92$; $s \doteq 1,85$

18. Výčetní tloušťky jedlí. Následující seznam zachycuje výčetní tloušťky sto až sto deset let starých jedlí rostoucích na daném stanovišti. Hodnoty tlouštěk jsou uvedeny v milimetrech.

427	365	367	340	425	481	322	333	426	285
292	374	368	318	472	551	447	273	312	322
312	422	496	437	372	328	377	342	271	268
417	345	523	314	404	362	362	254	263	259
462	470	533	335	226	314	284	283	456	502
332	380	297	421	178	451	413	521	445	309
296	568	374	519	424	339	459	261	531	196
219	321	314	464	275	367	412	233	296	362
395	269	449	312	275	438	292	300	501	390
380	407	344	482	406	311	288	345	250	445
359	467	404	374	437	558	366	365	305	310
404	453	361	316	263	414	438	463	461	345
307	349	521	524	379	263	549	385	247	514
436	288	344	224	442	387	234	212	388	304
363	404	496	412	417	333	392	270	365	288
279	342	319	264	307	353	377	362	272	391
332	411	282	521	223	314	277	560	604	257
242	417	278	276	589	276	237	393	402	312
291	220	210	176	415	210	339	144	153	170
326	354	544	311	362	223	335	360	168	346

a) Zaznamenaná data seřadíte a výsledek tohoto seřazení prezentujte graficky. Volte přitom různou šířku a počátek tloušťkových tříd.

b) Určete základní statistické ukazatele.

(Použijte vhodný tabulkový kalkulátor či soubor statistických programů.)

19. V následující tabulce je uvedena hmotnost novorozenců z chudých čínských rodin v Singapuru v letech 1950–1951. Hmotnosti jsou měřeny v uncích ($1 \text{ unce} = 28,3495 \cdot 10^{-3} \text{ kg}$), data jsou přitom pro přehlednost sdružena do tříd po osmi uncích. Hmotnosti v tabulce odpovídají středům příslušných tříd.

Hmotnost	59,5	67,5	75,5	83,5	91,5	99,5	107,5	115,5	123,5	131,5	139,5	147,5	155,5	163,5	171,5
Četnost	2	6	39	385	888	1729	2240	2007	1233	641	201	74	14	5	1

Určete modus, medián, aritmetický průměr, rozptyl a směrodatnou odchylku hmotností. Korigujte chyby způsobené seřazením užitím lineární interpolace při výpočtu mediánu a Sheppardovy korekce při výpočtu rozptylu.