Vysoká škola ekonomická
Fakulta informatiky a statistiky

Sbírka úloh pro účely bakalářských statistických předmětů

—

4ST201 Statistika a 4ST210 Statistika pro finance

Lubomír Štěpánek

Praha 2020

Sbírka úloh pro účely bakalářských statistických předmětů — 4ST201 Statistika a 4ST210 Statistika pro finance

Lubomír Štěpánek

Praha 2020

Slovo autora

Inspiration does exist, but it has to find you working.
— Pablo Picasso

Tato sbírka příkladů by měla podle autorova iniciálního záměru sloužit zejména jako zdroj typových příkladů pro opakování učiva v rámci předmětů 4ST201 Statistika a 4ST210 Statistika pro finance, které jsou dlouhodobě vyučovány a garantovány Katedrou statistiky a pravděpodobnosti Fakulty informatiky a statistiky Vysoké školy ekonomické.

Sbírka je členěna do kapitol podle logických celků odpovídajících přednáškám v rámci obou předmětů. Na začátku každé kapitoly je uvedeno několik řešených příkladů, jež jsou násedovány řadou neřešených úloh určených k vlastnímu procvičování.

Jednotlivé úlohy mají různé autory, obvykle neznámé. Část z úloh pochází z vlastní provenience tvůrce této sbírky příkladů, zbytek je převzat z různých dílčích zdrojů, které ale obvykle samy uvádějí, že i ony již úlohy převzaly. Budiž těmto anonymním původcům úloh vzdán zasloužený hold.

V Praze 25. března 2020 Lubomír Štěpánek

Obsah

1 Kombinatorika

1.1 Řešené úlohy

Úloha 1.1.1 Najděme všechna $x \in \mathbb{N}$ taková, pro která platí

(i)

$$\binom{x}{2} = 10.$$

(ii)

$$\binom{x}{3} = 455.$$

Řešení.

(i) Zřejmě podle vzorce pro kombinační číslo má platit

$$\binom{x}{2} = \frac{x!}{(x-2)! \cdot 2!} = \frac{x \cdot (x-1)}{2} = 10.$$

Odtud již zřejmě $x(x-1) = 20$, tedy $x^2 - x = 20$, neboli $x^2 - x - 20 = 0$. Rovnost je možné vyřešit jako kvadratickou rovnici pomocí diskriminantu $x_{1,2} = \frac{1 \pm \sqrt{81}}{2} = \{-4; 5\}$, nebo pomocí Vietových vztahů $x_1 + x_2 = 1 \wedge x_1 x_2 = -20$ a „uhádnout“ kořeny, zřejmě $x_{1,2} = \{-4; 5\}$. Očivině může úloze vyhovovat pouze řešení $x = 5$.

Existuje rychlý a trikový přístup, který umožní vyhnout se vyčíslení řešení kvadratické rovnice. Vyjdeme ze vztahu $x(x-1) = 20$. Uvědomíme-li si, že pro řešení nutně platí $x \in \mathbb{N}$, pak je i $(x-1)^2 < x(x-1) < x^2$, neboli $(x-1)^2 < 20 < x^2$ a $(x-1) < \sqrt{20} < x$. Hledáme tedy takové $x \in \mathbb{N}$, které je „o něco málo“ větší než $\sqrt{20} \doteq 4{,}47$, resp. takové $x - 1 \in \mathbb{N}$, které je „o něco málo“ menší než $\sqrt{20} \doteq 4{,}47$. Zřejmě je $x = 5$, resp. $x - 1 = 4$. Dosazením do $\binom{x}{2} = \binom{5}{2} = \frac{5!}{(5-2)! \cdot 2!} = 10$ ověříme správnost řešení.

(ii) Obdobně jako v předchozí části má platit

$$\binom{x}{3} = \frac{x!}{(x-3)! \cdot 3!} = \frac{x(x-1)(x-2)}{6} = 455.$$

Odtud již zřejmě $x(x-1)(x-2) = 6 \cdot 455$, tedy $x^3 - 3x^2 + 2x = 2730$, neboli $x^3 - 3x^2 + 2x - 2730 = 0$. Rovnost je možné vyřešit jako kubickou rovnici, anebo s výhodou následovně.

Pokud řešení existuje, pak pro něj platí $x \in \mathbb{N}$ a dále $(x-2)^3 < x(x-1)(x-2) < x^3$, tedy $(x-2)^3 < 2730 < x^3$, neboli $(x-2)^3 < \sqrt[3]{2730} < x^3$. Hledáme tedy takové $x \in \mathbb{N}$, které je „o něco málo“ větší než $\sqrt[3]{2730} \doteq 13{,}98$, resp. takové $x-2 \in \mathbb{N}$, které je „o něco málo“ menší než $\sqrt[3]{2730} \doteq 13{,}98$. Snadno nahlédneme, že může být buďto $x = 14 > \sqrt[3]{2730} \doteq 13{,}98$ a $x-2 = 12 < \sqrt[3]{2730} \doteq 13{,}98$, anebo $x = 15 > \sqrt[3]{2730} \doteq 13{,}98$ a $x-2 = 13 < \sqrt[3]{2730} \doteq 13{,}98$.

Dosazením obou možných hodnot do výrazu $\binom{x}{3}$ ověříme, že $\binom{14}{3} = 364$ a $\binom{15}{3} = 455$. Řešením je tedy $x = 15$. □

Úloha 1.1.2 Jirka má právě sedm různých triček, čtvery různé kalhoty a pět párů různých bot.

(i) Kolika navzájem různými způsoby může vytvořit svůj outfit?
(ii) Kolik by potřeboval triček, aby měl každý den v roce originální[1] outfit?

Řešení.

(i) Využijeme kombinatorické pravidlo součinu. Ke každému ze sedmi triček si může Jirka vzít libovolné z celkem čtver kalhot, tedy celkem počet možností, jak může kombinovat tričko a kalhoty, je $7 \cdot 4 = 28$. Ke každé možnosti, jak spároval tričko a kalhoty, si ještě postupně může vzít první, druhý, třetí, čtvrtý, nebo pátý pár bot, tedy celkem $(7 \cdot 4) \cdot 5 = 28 \cdot 5 = 140$. Tedy celkový počet možností, jak může Jirka vybrat tričko, kalhoty a pár bot do svého outfitu je 140.
(ii) Označme počet triček t, zřejmě je $t \in \mathbb{N}$. Počet možností, jak může Jirka vybrat outfit z t triček, čtver kalhot a pěti párů bot, je dle předchozí části roven $t \cdot 4 \cdot 5 = 20t$. Pokud má mít každý den v roce unikátní outfit, musí mít možnost složit svůj outfit alespoň 365 způsoby.

Hledáme tedy nejmenší takové $t \in \mathbb{N}$, aby platilo $20t \geq 365$, neboli $t \geq \frac{365}{20} = 18{,}25$, tedy $t = 19$. Jirka tedy potřebuje alespoň devatenáct triček k tomu, aby je dokázal s čtvermi kalhoty a pěti páry bot zkombinovat tak, že bude mít každý den v roce unikátní outfit. □

Úloha 1.1.3 Uvažujme sedm číslic $1, 2, 3, \ldots, 6, 7$.

(i) Kolika navzájem různými způsoby z nich můžeme sestavit trojciferné číslo, pokud se číslice mohou opakovat?
(ii) Kolika navzájem různými způsoby z nich můžeme sestavit trojciferné číslo, pokud se číslice nemohou opakovat?
(iii) Pro kolik takových trojciferných čísel bude platit, že první jejich cifra je menší než druhá a zároveň druhá jejich cifra je menší než třetí?

Řešení.

(i) Mohou-li se cifry na jednotlivých pozicích trojciferného čísla opakovat, pak zřejmě na pozici stovek může být kterákoliv ze sedmi cifer $\{1, 2, 3, \ldots, 6, 7\}$, obdobně na pozici desítek může být kterákoliv ze sedmi cifer $\{1, 2, 3, \ldots, 6, 7\}$

[1] V tom smyslu, aby měl pokaždé jiný outfit než během kteréhokoliv jiného dne v roce.

a nakonec i na pozici jednotek může být kterákoliv ze sedmi cifer $\{1, 2, 3, \ldots, 6, 7\}$. Podle kombinatorického pravidla součinu je tak celkový počet navzájem různých trojciferných čísel sestatených z cifer $\{1, 2, 3, \ldots, 6, 7\}$ s jejich možným opakováním roven

$$\begin{aligned} |\{1,2,3,\ldots,6,7\}| \times |\{1,2,3,\ldots,6,7\}| \times |\{1,2,3,\ldots,6,7\}| &= 7 \cdot 7 \cdot 7 = \\ &= 7^3 = \\ &= 343. \end{aligned}$$

(ii) Nemohou-li se cifry na jednotlivých pozicích trojciferného čísla opakovat, pak zřejmě na pozici stovek může být kterákoliv ze sedmi cifer $\{1, 2, 3, \ldots, 6, 7\}$, na pozici desítek může být kterákoliv ze sedmi cifer $\{1, 2, 3, \ldots, 6, 7\}$ kromě té na pozici stovek a nakonec na pozici jednotek může být kterákoliv ze sedmi cifer $\{1, 2, 3, \ldots, 6, 7\}$ kromě oněch dvou na pozici stovek a desítek. Tři navzájem různé cifry z množiny $\{1, 2, 3, \ldots, 6, 7\}$ můžeme vybrat celkem

$$\binom{|\{1,2,3,\ldots,6,7\}|}{3} = \binom{7}{3} = 35$$

způsoby. Protože jsou vždy číslice navzájem různé, v každém z 35 způsobů je lze ještě vzájemně uspořádat (permutovat). Celkový počet navzájem různých trojciferných čísel sestatených z cifer $\{1, 2, 3, \ldots, 6, 7\}$ bez jejich opakování[2] je roven

$$\binom{7}{3} \cdot 3! = 35 \cdot 6 = 210.$$

(iii) Stačí si uvědomit, že kdykoliv z množiny $\{1, 2, 3, \ldots, 6, 7\}$ vybereme právě tři různé číslice, lze je jednoznačně (a vždy právě jedním způsobem) uspořádat tak, aby první byla menší než druhá a druhá aby byla menší než třetí. Počet trojciferných čísel sestavených z číslic množiny $\{1, 2, 3, \ldots, 6, 7\}$ takových, že jejich první cifra je vždy menší než druhá a druhá cifra je menší než třetí, je tak roven počtu způsobů, kolika můžeme vybrat z dané množiny navzájem různé tři číslice. To je možné právě $\binom{7}{3} = 35$ způsoby. □

[2]Porovnáním řešení této a předchozí části získáme celkový počet navzájem různých trojciferných čísel sestatených z cifer $\{1, 2, 3, \ldots, 6, 7\}$ tak, že alespoň jedna z cifer je v každém takovém čísle obsažena alespoň dvakrát, ten je tedy roven $343 - 210 = 133$.

To ovšem můžeme nahlédnout i analyticky. Má-li se alespoň jedna z cifer trojciferného čísla opakovat, obsahuje takové číslo dvě různé číslice (a jednu z nich dvakrát), nebo pouze jednu číslici (a to třikrát). Dvě různé číslice můžeme z množiny $\{1, 2, 3, \ldots, 6, 7\}$ vybrat celkem $\binom{7}{2}$ způsoby, a protože jednu z nich použijeme dvakrát, tedy první, pak druhou, jde o $\binom{7}{2} \cdot 2$ způsobů. Nakonec si ještě uvědomme, že číslice, která je v čísle jen jedenkrát, může být na pozici stovek, desítek, anebo jednotek. Celkově tak můžeme trojciferné číslo s dvěma různými číslicemi, z nichž jedna je obsažena dvakrát, získat $\binom{7}{2} \cdot 2 \cdot 3 = 21 \cdot 2 \cdot 3 = 126$ způsoby. Trojciferné číslo s jednou číslicí z množiny $\{1, 2, 3, \ldots, 6, 7\}$ obsaženou třikrát můžeme očividně získat právě 7 způsoby. A zřejmě je skutečně počet trojciferných čísel setavených z množiny $\{1, 2, 3, \ldots, 6, 7\}$ tak, že alespoň jedna jeho cifra se v něm opakuje, roven $126 + 7 = 133$, což jsme odvodili dvěma různými způsoby.

1.2 Úlohy k vlastnímu procvičení

Úloha 1.2.1 Kolika způsoby je možné rozdělit zlatou, stříbrnou a bronzovou medaili mezi osm finalistů (pořadí na nemedailových pozicích nesestavujeme)? V kolika z těchto možností bude finalista ze Srí Lanky bronzový (je v osmici jen jeden)? V kolika z těchto možností bude finalista ze Srí Lanky na kterékoliv medailové pozici?

Úloha 1.2.2 Určeme počet všech nejvýše pěticiferných čísel, v nichž se žádná číslice neopakuje.

Úloha 1.2.3 Kolika způsoby se může při nástupu seřadit n táborníků do řady? V kolika způsobech stojí táborník A po levici táborníka B? V kolika stojí A po pravici B? V kolika případech stojí táborníci A a B těsně vedle sebe? V kolika případech mezi A a B stojí minimálně jeden další táborník?

Úloha 1.2.4 Regionální letecká společnost rozšířila síť svých letů na několik nových letišť (o více než jedno). Kvůli novým letům musela začít v databázi evidovat o 296 letových tras více (let z letiště A do letiště B považujeme za různý od letu z letiště B do letiště A). Mezi kolika letišti původně společnost létala a kolik nových jich přibrala?

Úloha 1.2.5 Určete počet všech přirozených šesticiferných čísel, jejichž ciferný součet je sudý.

Úloha 1.2.6 Musí mít nutně v městečku o 1500 obyvatelích dva občané zcela shodné iniciály? Jméno a příjmení začíná jedním z 32 různých znaků.

Úloha 1.2.7 Na maturitním večírku je patnáct hochů a dvanáct dívek. Kolika způsoby je možné vybrat čtyři taneční páry?

Úloha 1.2.8 Určete počet všech navzájem různých řešení rovnice

$$x_1 + x_2 + \cdots + x_k = n$$

v oboru přirozených čísel, tedy $\forall i \in \{1, 2, \ldots k\}$ je $x_i \in \mathbb{N}$, vzhledem k parametrům $k \in \mathbb{N}$ a $n \in \mathbb{N}$.

Úloha 1.2.9 Určete počet všech navzájem různých řešení nerovnice

$$x_1 + x_2 + \cdots + x_k \leq n$$

v oboru

(i) přirozených čísel, tedy $\forall i \in \{1, 2, \ldots k\}$ je $x_i \in \mathbb{N}$
(ii) celých nezáporných čísel, tedy $\forall i \in \{1, 2, \ldots k\}$ je $x_i \in \mathbb{N} \cup \{0\}$

vzhledem k parametrům $k \in \mathbb{N}$ a $n \in \mathbb{N}$.

Úloha 1.2.10 Sportovní utkání neskončilo nerozhodně a žádné z obou družstev nevstřelilo více než dvacet a méně než deset branek. Určete počet možných výsledků.

Úloha 1.2.11 Hokejové utkání skončilo prohrou domácího týmu 7 : 12 (branek). Kolik navzájem různými možnostmi se mohl zápas vyvíjet[3]? Všechny smysluplné stavy během zápasu (co do počtu branek) jsou následující

$$\begin{pmatrix} 0:0 & 0:1 & 0:2 & \cdots & 0:12 \\ 1:0 & 1:1 & 1:2 & \cdots & 1:12 \\ 2:0 & 2:1 & 2:2 & \cdots & 2:12 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 7:0 & 7:1 & 7:2 & \cdots & 7:12 \end{pmatrix}.$$

[3]Za vývoj zápasu tedy považujeme sekvenci jeho průběžných brankových skóre tak, jak se mohla skóre vyvjíjet vždy po padnutí další branky, tedy $[0:0] \to \{[0:1],[1:0]\} \to \{[0:2],[1:1],[2:0]\} \to \cdots \to [7:12]$.

2 Deskriptivní statistika

2.1 Řešené úlohy

Úloha 2.1.1 V tabulce 2.2.1 jsou uvedeny hodnoty osmiprvkového výběru $\boldsymbol{x} = (x_1, x_2, \ldots, x_8)^T$. Určete pro daný výběr jeho aritmetický průměr, medián, výběrový rozptyl, výběrovou směrodatnou odchylku, minimum, maximum, variační rozpětí a variační koeficient.

i	1	2	3	4	5	6	7	8
x_i	4	5	2	12	11	8	6	16

Tabulka 2.1.1: Hodnoty výběru $\boldsymbol{x} = (x_1, x_2, \ldots, x_8)^T$

Řešení. S použitím známých vzorců a při značení $n = 8$ hodnot pomocí symbolů $x_1, x_2, x_3, \ldots, x_8$ dostaneme postupně hodnoty pro aritmetický průměr

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{8}\sum_{i=1}^{8} x_i = 8,$$

medián

$$\tilde{x} = \frac{1}{2}\left(x_{(4)} + x_{(5)}\right) = \frac{1}{2}(6+8) = 7,$$

výběrový[4] rozptyl

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{8-1}\sum_{i=1}^{8}(x_i - \bar{x})^2 = 22,$$

[4]Kdykoliv jde o *výběr* nebo náhodně vybrané hodnoty, preferujeme k výpočtu (výběrového) rozptylu a směrodatné odchylky vzorce

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

Pokud jde naopak o celou *populaci*, preferujeme k výpočtu (populačního) rozptylu a směrodatné ochylky vzorec

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

výběrovou směrodatnou odchylku

$$s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{8-1}\sum_{i=1}^{8}(x_i - \bar{x})^2} = \sqrt{22} \doteq 4{,}69,$$

minimum $x_{\min} = x_{(1)} = 2$, maximum $x_{\max} = x_{(8)} = 16$, variační rozpětí $R = x_{\max} - x_{\min} = x_{(8)} - x_{(1)} = 16 - 2 = 14$ a variační koeficient

$$v_x = \frac{s_x}{\bar{x}} = \frac{\sqrt{22}}{8} \doteq 0{,}59.$$

□

Úloha 2.1.2 Na katederní schůzi, kde bylo původně šestnáct lidí, přišlo několik doktorandů pozdě. Jejich příchodem kleslo průměrné IQ v místnosti ze 140 na 137. Průměrné IQ ve skupince pozdě příchozích doktorandů bylo 125. Určete, kolik bylo doktorandů bylo ve skupince pozdě příchozích.

Řešení. Označme neznámý počet doktorandů symbolem k. Dále značme hodnotu IQ pro i-tého doktoranda symbolem d_i, kde $i \in \{1, 2, 3, \ldots, k\}$ a hodnotu IQ j-tého člena katederní schůze před příchodem doktorandů postupně q_j, kde $j \in \{1, 2, 3, \ldots, 16\}$. Pro aritmetický průměr IQ ve skupince doktorandů platí

$$125 = \frac{1}{k}\sum_{i=1}^{k} d_i.$$

Odsud zřejmě je $\sum_{i=1}^{k} d_i = 125k$. Dále vyjádřeme průměrnou hodnotu IQ na katederní schůzi před a po příchodu doktorandů. Před příchodem doktorandů zřejmě pro průměrné IQ v místnosti platilo

$$140 = \frac{1}{16}\sum_{j=1}^{16} q_j,$$

odsud zřejmě je $\sum_{j=1}^{16} q_j = 16 \cdot 140 = 2240$. Po příchodu doktorandů pak pro průměrné IQ v místnosti platilo

$$137 = \frac{1}{16+k}\left(\sum_{j=1}^{16} q_j + \sum_{i=1}^{k} d_i\right).$$

Dosazením za $\sum_{j=1}^{16} q_j = 2240$ a za $\sum_{i=1}^{k} d_i = 125k$ do předchozí rovnosti získáme

$$137 = \frac{1}{16+k}\left(2240 + 125k\right).$$

Odtud již dalšími úpravami dostáváme

$$\begin{aligned}137 &= \frac{1}{16+k}\left(2240+125k\right)\\137(16+k) &= 2240+125k\\2192+137k &= 2240+125k\\12k &= 48\\k &= 4.\end{aligned}$$

Ve skupince pozdě příchozích byli právě čtyři doktorandi. □

Úloha 2.1.3 V souboru deseti čísel je aritmetický průměr roven 12 a rozptyl 4. Do souboru přidáme hodnotu 23. Jak se změní průměr a rozptyl souboru?

Řešení. Značme původní hodnoty postupně $x_1, x_2, \dots, x_{10}$ a novou, přidanou hodnotu x_{11}. V původním souboru jistě pro aritmetický průměr platilo

$$\bar{x} = 12 = \frac{1}{10}\sum_{i=1}^{10} x_i,$$

odtud $\sum_{i=1}^{10} x_i = 12 \cdot 10 = 120$, a pro výběrový rozptyl (ve výpočetním tvaru) platilo

$$s_x^2 = 4 = \frac{1}{10}\sum_{i=1}^{10} x_i^2 - \left(\frac{1}{10}\sum_{i=1}^{10} x_i\right)^2 = \frac{1}{10}\sum_{i=1}^{10} x_i^2 - \left(\bar{x}\right)^2,$$

odsud je

$$\sum_{i=1}^{10} x_i^2 = 10 \cdot \left(s_x^2 + (\bar{x})^2\right) = 10 \cdot (4 + 12^2) = 1480.$$

Pro aritmetický průměr v souboru po přidání jedenácté hodnoty platí

$$\bar{x}_{\text{nový}} = \frac{1}{11}\left(\sum_{i=1}^{10} x_i + x_{11}\right),$$

odsud po dosazení $\sum_{i=1}^{10} x_i = 120$ a $x_{11} = 23$ je $\bar{x}_{\text{nový}} = \frac{1}{11}\left(\sum_{i=1}^{10} x_i + x_{11}\right) = \frac{1}{11}(120+23) = \frac{143}{11} = 13$.

Pro výběrový rozptyl v souboru po přidání jedenácté hodnoty platí

$$\begin{aligned}s_{x,\ \text{nový}}^2 &= \frac{1}{11}\left(\sum_{i=1}^{10} x_i^2 + x_{11}^2\right) - \left(\frac{1}{11}\left(\sum_{i=1}^{10} x_i + x_{11}\right)\right)^2 =\\&= \frac{1}{11}\left(\sum_{i=1}^{10} x_i^2 + x_{11}^2\right) - \left(\bar{x}_{\text{nový}}\right)^2,\end{aligned}$$

dosazením za $\sum_{i=1}^{10} x_i^2 = 1480$, za $\bar{x}_{\text{nový}} = 13$ a za $x_{11} = 23$ je

$$
\begin{aligned}
s^2_{x,\text{ nový}} &= \frac{1}{11}\left(\sum_{i=1}^{10} x_i^2 + x_{11}^2\right) - (\bar{x}_{\text{nový}})^2 = \\
&= \frac{1}{11}\left(1480 + 23^2\right) - (13)^2 = \\
&= 13\frac{7}{11}.
\end{aligned}
$$

Po přidání jedenácté hodnoty je nový aritmetický průměr souboru roven $\bar{x}_{\text{nový}} = 13$ a výběrový rozptyl je nyní roven $s^2_{x,\text{ nový}} = 13\frac{7}{11}$. □

Úloha 2.1.4 V souboru šestnácti čísel je jejich aritmetických průměr roven 10,3. Jak se změní jejich aritmetický průměr, pokud

(i) zvýšíme každé z čísel o 5,2?
(ii) zvýšíme každé z čísel třikrát?
(iii) zvýšíme polovinu čísel o 7,2 a zbytek čísel zmenšíme o 7,2?
(iv) zvýšíme polovinu čísel o 3,1 a druhou polovinu čísel zmenšíme o 1,1?

Řešení. Označme hodnoty v souboru postupně $x_1, x_2, x_3, \dots, x_{16}$. Pak dle zadání platí $\bar{x} = \frac{1}{16}\sum_{i=1}^{16} x_i$.

(i) Pokud zvýšíme každou z šestnácti hodnot o 5,2, nové hodnoty mají $\forall i \in \{1, 2, 3, \dots, 16\}$ tvar $x_i' = x_i + 5{,}2$. Pro nový aritmetický průměr $\bar{x'}$ pak platí

$$
\begin{aligned}
\bar{x'} &= \frac{1}{16}\sum_{i=1}^{16} x_i' = \\
&= \frac{1}{16}\sum_{i=1}^{16} (x_i + 5{,}2) = \\
&= \frac{1}{16}\left(\sum_{i=1}^{16} x_i + \sum_{i=1}^{16} 5{,}2\right) = \\
&= \frac{1}{16}\left(\sum_{i=1}^{16} x_i + 16 \cdot 5{,}2\right) = \\
&= \frac{1}{16}\sum_{i=1}^{16} x_i + \frac{1}{16}(16 \cdot 5{,}2) = \\
&= \bar{x} + 5{,}2.
\end{aligned}
$$

Aritmetický průměr čísel v souboru se tak po jejich zvýšení o hodnotu 5,2 zvětší rovněž o hodnotu 5,2.

(ii) Pokud zvýšíme každou z šestnácti hodnot třikrát, nové hodnoty mají $\forall i \in \{1, 2, 3, \dots, 16\}$ tvar $x_i' = 3x_i$. Pro nový aritmetický průměr $\bar{x'}$ pak platí

$$
\bar{x'} = \frac{1}{16}\sum_{i=1}^{16} x_i'
$$

$$\begin{aligned}
&= \frac{1}{16}\sum_{i=1}^{16}(3x_i) = \\
&= 3 \cdot \frac{1}{16}\sum_{i=1}^{16} x_i = \\
&= 3\bar{x}.
\end{aligned}$$

Aritmetický průměr čísel v souboru se tak po jejich zvýšení třikrát zvětší rovněž třikrát.

(iii) Bez újmy na obecnosti předpokládejme, že čísla $x_1, x_2, \ldots, x_8$ byla zvetšena o 7,2, zatímco čísla $x_9, x_{10}, \ldots, x_{16}$ byla zmenšena o 7,2. Tedy $\forall i \in \{1, 2, \ldots, 8\}$ mají nové hodnoty tvar $x_i' = x_i + 7{,}2$ a $\forall i \in \{9, 10, \ldots, 16\}$ mají nové hodnoty tvar $x_i' = x_i - 7{,}2$. Pro nový aritmetický průměr $\bar{x'}$ pak platí

$$\begin{aligned}
\bar{x'} &= \frac{1}{16}\sum_{i=1}^{16} x_i' = \\
&= \frac{1}{16}\left(\sum_{i=1}^{8} x_i' + \sum_{i=9}^{16} x_i'\right) = \\
&= \frac{1}{16}\left(\sum_{i=1}^{8}(x_i + 7{,}2) + \sum_{i=9}^{16}(x_i - 7{,}2)\right) = \\
&= \frac{1}{16}\left(\sum_{i=1}^{8} x_i + \sum_{i=1}^{8} 7{,}2 + \sum_{i=9}^{16} x_i - \sum_{i=9}^{16} 7{,}2\right) = \\
&= \frac{1}{16}\left(\underbrace{\sum_{i=1}^{8} x_i + \sum_{i=9}^{16} x_i}_{=\sum_{i=1}^{16} x_i} + 8 \cdot 7{,}2 - 8 \cdot 7{,}2\right) = \\
&= \frac{1}{16}\left(\sum_{i=1}^{16} x_i + 0\right) = \\
&= \frac{1}{16}\sum_{i=1}^{16} x_i = \\
&= \bar{x}.
\end{aligned}$$

Aritmetický průměr souboru čísel se po zvýšení poloviny čísel o 7,2 a zmenšení zbytku čísel o 7,2 nezměnil.

(iv) Bez újmy na obecnosti předpokládejme, že čísla $x_1, x_2, \ldots, x_8$ byla zvetšena o 3,1, zatímco čísla $x_9, x_{10}, \ldots, x_{16}$ byla zmenšena o 1,1. Tedy $\forall i \in \{1, 2, \ldots, 8\}$ mají nové hodnoty tvar $x_i' = x_i + 3{,}1$ a $\forall i \in \{9, 10, \ldots, 16\}$

mají nové hodnoty tvar $x_i' = x_i - 1{,}1$. Pro nový aritmetický průměr $\bar{x'}$ pak platí

$$
\begin{aligned}
\bar{x'} &= \frac{1}{16}\sum_{i=1}^{16} x_i' = \\
&= \frac{1}{16}\left(\sum_{i=1}^{8} x_i' + \sum_{i=9}^{16} x_i'\right) = \\
&= \frac{1}{16}\left(\sum_{i=1}^{8}(x_i + 3{,}1) + \sum_{i=9}^{16}(x_i - 1{,}1)\right) = \\
&= \frac{1}{16}\left(\sum_{i=1}^{8} x_i + \sum_{i=1}^{8} 3{,}1 + \sum_{i=9}^{16} x_i - \sum_{i=9}^{16} 1{,}1\right) = \\
&= \frac{1}{16}\left(\underbrace{\sum_{i=1}^{8} x_i + \sum_{i=9}^{16} x_i}_{=\sum_{i=1}^{16} x_i} + 8\cdot 3{,}1 - 8\cdot 1{,}1\right) = \\
&= \frac{1}{16}\left(\sum_{i=1}^{16} x_i + 16\right) = \\
&= \frac{1}{16}\sum_{i=1}^{16} x_i + 1 = \\
&= \bar{x} + 1.
\end{aligned}
$$

Aritmetický průměr souboru čísel se po zvýšení poloviny čísel o 3,1 a zmenšení zbytku čísel o 1,1 zvětšil o jednotku.

Poznámka. Z řešení prvních dvou částí snadno nahlédneme, že pokud jsou původní hodnoty souboru $n \in \mathbb{N}$ čísel zvětšeny vždy q-krát, kde $q \in \mathbb{R} \setminus \{0\}$, a současně zvětšeny vždy o konstantu r, kde $r \in \mathbb{R}$, pak se původní aritmetický průměr souboru $\bar{x}$ změní na nový aritmetický průměr $\bar{x'}$ tak, že $\bar{x'} = q\bar{x} + r$, neboť

$$
\begin{aligned}
\bar{x'} &= \frac{1}{n}\sum_{i=1}^{n} x_i' = \\
&= \frac{1}{n}\sum_{i=1}^{n}(qx_i + r) = \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} qx_i + \sum_{i=1}^{n} r\right) = \\
&= \frac{1}{n}\left(q\sum_{i=1}^{n} x_i + n\cdot r\right) =
\end{aligned}
$$

$$
\begin{aligned}
&= q \cdot \frac{1}{n} \sum_{i=1}^{n} x_i + \frac{1}{n} \cdot n \cdot r = \\
&= q \cdot \frac{1}{n} \sum_{i=1}^{n} x_i + r = \\
&= q\bar{x} + r.
\end{aligned}
$$

□

Úloha 2.1.5 Během čtyř let cena akcie meziročně postupně vzrostla o 8 %, pak o 12 %, poté o 3 % a nakonec klesla o 1 %. Pokud by se během těchto čtyřech let měnila cena akcie meziročně vždy o konstantní procento, o kolik by to bylo?

Řešení. Cena akcie je po čtyřech letech rovna $1{,}08 \cdot 1{,}12 \cdot 1{,}03 \cdot 0{,}99 \doteq 1{,}233$ násobku původní hodnoty, tedy zvýšila svou hodnotu o 23,3 procent. Předpokládejme, že se meziročně vždy zvýšila cena akcie o p procent, tedy každý ze čtyř roků zvýšila svou hodnotu $\left(1 + \frac{p}{100}\right)$-krát a za všechny čtyři roky tedy změní svou hodnotu právě $\left(1 + \frac{p}{100}\right)^4$-krát. Pak platí, že

$$\left(1 + \frac{p}{100}\right)^4 = 1{,}08 \cdot 1{,}12 \cdot 1{,}03 \cdot 0{,}99 \doteq 1{,}233,$$

odsud tedy

$$
\begin{aligned}
\left(1 + \frac{p}{100}\right) &\doteq \sqrt[4]{1{,}233} \\
\frac{p}{100} &\doteq \sqrt[4]{1{,}233} - 1 \\
p &\doteq 100 \cdot (\sqrt[4]{1{,}233} - 1) \\
p &\doteq 100 \cdot (1{,}054 - 1) \\
p &\doteq 5{,}4.
\end{aligned}
$$

Snadno nahlédneme, že meziroční změna předpokládající konstantní meziroční procento zhodnocení odpovídá geometrickému průměru nad zadanými konkrétními meziročními změnami.

Pokud by se během čtyř let měnila cena akcie meziročně vždy o konstantní procento, bylo by to vždy o 5,4 %. □

Úloha 2.1.6 Ukažme, že pokud se geometrický průměr původního množství $n \in \mathbb{N}$ kladných čísel rovnal $\bar{x}_G$, pak aritmetický průměr logaritmů původních čísel je roven $\log \bar{x}_G$.

Řešení. Pro původní geometrický průměr $\bar{x}_G$ z $n \in \mathbb{N}$ kladných čísel označených $x_1, x_2, \ldots, x_n$ platí

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

Obě strany rovnosti jsou zřejmě nezáporné, jejich zlogaritmováním dostaneme

$$\begin{aligned}
\log \bar{x}_G &= \log \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}\\
\log \bar{x}_G &= \log (x_1 \cdot x_2 \cdot \dots \cdot x_n)^{\frac{1}{n}}\\
\log \bar{x}_G &= \frac{1}{n} \cdot \log (x_1 \cdot x_2 \cdot \dots \cdot x_n)\\
\log \bar{x}_G &= \frac{1}{n} (\log x_1 + \log x_2 + \dots + \log x_n),
\end{aligned}$$

poslední řádek již značí, že logaritmus geometrického průměru kladných čísel je roven aritmetickému průměru logaritmů původních kladných čísel. □

Úloha 2.1.7 Totožná součástka se vyrábí na dvou automatech. Starší z nich vyrobí jeden kus každých 10 minut, nový každých 6 minut.

(i) Jak dlouho trvá v průměru výroba jedné součástky, pokud oba automaty pracují nezávisle na sobě?
(ii) Jak dlouho trvá v průměru výroba jedné součástky, pokud oba automaty pracují na výrobě součástek společně?
(iii) Jak dlouho trvá v průměru výroba jedné součástky, pracuje-li starší automat 5 hodin denně a nový 8 hodin denně?
(iv) Jaká je týdenní produkce součástek, pracují-li oba stroje na maximum, tj. 8 hodin denně, 7 dní v týdnu?

Řešení.

(i) Máme-li spočítat průměrnou dobu, za jakou je oběma automaty pracujícími nezávisle na sobě vyrobena jedna součástka, je třeba předpokládat, že oba automaty pracují dostatečně dlouhou dobu $t > 0$ minut. Díky tomu lze pak zanedbat např. i fakt, že automaty nezačaly pracovat oba ve stejný okamžik. Pak zřejmě platí, že podíl celkové doby práce, po kterou pracují oba automaty, a počtu všech součástek, které vyrobí první i druhý automat vždy za dobu t minut, odpovídá průměrné době výroby jedné součástky.

Celková doba práce obou automatů je $t + t = 2t$ minut. Za dobu t minut vyrobí starší automat právě $\frac{t}{10}$ součástek, za stejnou dobu t minut vyrobí mladší automat právě $\frac{t}{6}$ součástek.

Pro průměrnou dobu $\bar{\tau}$, za jakou je oběma automaty pracujícími nezávisle na sobě vyrobena jedna součástka, platí

$$\begin{aligned}
\bar{\tau} &= \frac{\text{celkové doba práce obou automatů}}{\text{celkový počet součástek, které vyrobí oba automaty za danou dobu}}\\
&= \frac{2t}{\frac{t}{10} + \frac{t}{6}} =\\
&= \frac{2t}{\frac{3t+5t}{30}} =\\
&= \frac{60t}{8t} =
\end{aligned}$$

$$\begin{aligned}\bar{\tau} &= \frac{60}{8} = \\ &= 7{,}5 \text{ [minut]}.\end{aligned}$$

Všimněme si, že tvar $\bar{\tau} = \frac{2t}{\frac{t}{10}+\frac{t}{6}} = \frac{2}{\frac{1}{10}+\frac{1}{6}}$ přesně odpovídá situaci, kdy nad hodnotami 6 minut a 10 minut počítáme *harmonický průměr*. Průměrná doba, za jakou je oběma automaty pracujícími nezávisle na sobě vyrobena jedna součástka, tedy odpovídá *harmonickému* průměru obou dob, za které každý z automatů vyrobí jednu součástku sám.

Číselně je průměrná doba, za jakou je oběma automaty pracujícími nezávisle na sobě vyrobena jedna součástka, rovna 7,5 minutám.

Současně si všimněme, že taková průměrná doba odpovídá hodnotě, která leží „někde“ mezi oběma dobami výroby součástky každým z automatů (6 a 10 minut), tedy není ani menší, nebo ani větší než obě dílčí doby.

(ii) Pokud oba automaty pracují společně, stačí si uvědomit, že starší automat, který vyrobí jednu součástku za 10 minut, vyrobí za jednu minutu právě $\frac{1}{10}$ součástky. Obdobně novější automat, který vyrobí jednu součástku za 6 minut, vyrobí za jednu minutu přesně $\frac{1}{6}$ součástky. Oba společně tedy vyrobí za jednu minutu právě $\frac{1}{10} + \frac{1}{6} = \frac{3+5}{30} = \frac{8}{30} = \frac{4}{15}$ součástky.

Jsou-li za jednu minutu vyrobeno $\frac{4}{15}$ součástky, pak celá jedna součástka je vyrobena za $\frac{1}{4/15} = \frac{1}{\frac{1}{10}+\frac{1}{6}} = \frac{15}{4} = 3{,}75$ minuty.

Uvědomme si, že v tomto zadání jde vlastně o *úlohu o společné práci*. Pokud budeme předpokládat, že stroje pracují dlouhou dobu na velkém množství součástek, lze hodnotu 3,75 minuty skutečně považovat za průměrnou dobu výroby jedné součástky, pokud oba automaty pracují na výrobě každé ze součástek společně. V takovém případě lze dokonce zanedbat i případné nepřesnosti plynoucí např. z faktu, že oba automaty nemusely začít ve zcela shodný okamžik.

V porovnání s předchozí částí tedy zde nejde (!) o výpočet harmonického průměru, neboť $3{,}75 = \frac{1}{\frac{1}{10}+\frac{1}{6}} \neq \frac{2}{\frac{1}{10}+\frac{1}{6}} = 7{,}5$.

Současně si všimněme, že taková průměrná doba odpovídá hodnotě, která neleží mezi oběma dobami výroby součástky každým z automatů (6 a 10 minut), ale je dle očekávání díky spolupráci automatů menší než obě dílčí doby, za které by součástku vyrobil každý z automatů sám.

(iii) Zřejmě opět platí, že průměrná doba $\bar{\tau}$ výroby jedné součástky je rovna podílu celkové doby práce, po kterou pracují oba automaty, a počtu všech součástek, které vyrobí první i druhý automat vždy za příslušnou dobu. Starší automat vyrobí za 5 hodin, tedy $5 \cdot 60 = 300$ minut[5] právě $\frac{300}{10} = 30$ součástek, mladší automat pak za 8 hodin, tedy $8 \cdot 60 = 480$ minut vyrobí $\frac{480}{6} = 80$ součástek. Tedy

[5]Převod na minuty není formálně nutný, ale v kontextu úlohy je velmi názorný.

$$
\begin{aligned}
\bar{\tau} &= \frac{\text{celkové doba práce obou automatů}}{\text{celkový počet součástek, které vyrobí oba automaty za danou dobu}} \\
&= \frac{300+480}{\frac{300}{10}+\frac{480}{6}} = \\
&= \frac{780}{30+80} = \\
&= \frac{780}{110} = \\
&= 7{,}09 \text{ [minuty]}.
\end{aligned}
$$

Všimněme si, že tvar $\bar{\tau} = \frac{300+480}{\frac{300}{10}+\frac{480}{6}}$ přesně odpovídá situaci, kdy nad hodnotami 6 minut a 10 minut počítáme *vážený harmonický průměr* s vahami 300 a 480 minut, zde odpovídajícími dobám práce jednotlivých automatů denně.

Opět si všimněme, že taková průměrná doba odpovídá hodnotě, která leží „někde“ mezi oběma dobami výroby součástky každým z automatů (6 a 10 minut), tedy není ani menší, nebo ani větší než obě dílčí doby.

(iv) Sedm dní v týdnu po osmi hodinách odpovídá celkem $7 \cdot 8 \cdot 60 = 3360$ minutám. Starší automat vyrobí jednu součástku za 10 minut, tedy za 3360 minut vyrobí právě $\frac{3360}{10} = 336$ součástek, novější automat vyrobí jednu součástku za 6 minut, tedy za 3360 minut vyrobí právě $\frac{3360}{6} = 560$ součástek. Celkem oba automaty tak za sedm dní v týdnu po osmi hodinách vyrobí $336 + 560 = 896$ součástek.

Ke stejnému výsledku můžeme dojít i následovně. Z první části víme, že jednu součástku vyrobí každý z automatů průměrně za 7,5 minuty. Tedy za 3360 minut první a druhý automat dohromady $\frac{3360}{7{,}5} + \frac{3360}{7{,}5} = 448+448 = 896$ součástek. □

Úloha 2.1.8 V souboru hodnot $\boldsymbol{x} = (1, 3, 2, 2, 4, 1, 4, 2, 2, 5, 1, 2)^T$ nalezněme

(i) hodnotu kvantilu $\tilde{x}_{0,25}$.
(ii) hodnotu kvantilu $\tilde{x}_{0,45}$.
(iii) hodnotu modusu souboru $\boldsymbol{x}$.

Řešení. Vyjdeme z definice p-tého kvantilu pro libovolné $0 \le p \le 1$. Nad souborem $n \in \mathbb{N}$ čísel $x_1, x_2, \dots, x_n$ se p-tý kvantil značí $\tilde{x}_p$ a je definován jako

$$
\tilde{x}_p = \begin{cases} x_{(\lfloor k \rfloor + 1)}, & \text{pro } k = np \notin \mathbb{N} \\ \frac{1}{2}\left(x_{(k)} + x_{(k+1)}\right), & \text{pro } k = np \in \mathbb{N}, \end{cases}
$$

kde $x_{(k)}$ je k-té nejmenší číslo mezi čísly $x_1, x_2, \dots, x_n$ a kde $\lfloor x \rfloor$ značí dolní celou část čísla x, tedy nejvyšší celé číslo takové, že nepřevýší x, takže např. $\lfloor 4{,}3 \rfloor = 4$, $\lfloor 1 \rfloor = 1$ a $\lfloor -2{,}8 \rfloor = -3$.

Je vhodné čísla v souboru uspořádat vzestupně. Získáme tak

$$\boldsymbol{x}_{\text{uspořádaný}} = (1, 1, 1, 2, 2, 2, 2, 2, 3, 4, 4, 5)^T,$$

kde j-té nejmenší číslo značme $x_{(j)}$ pro $j \in \{1, 2, 3, \ldots, 12\}$, takže je např. $x_{(1)} = 1$, $x_{(4)} = 2$, $x_{(6)} = 2$, $x_{(9)} = 3$, $x_{(10)} = 4$ anebo $x_{(12)} = 5$.

Hodnotu modusu souboru $\boldsymbol{x}$ pak nalezneme jednoduše jako tu, která se v souboru $\boldsymbol{x}$ vyskytuje nejčastěji. Modus $\hat{x}$ souboru $\boldsymbol{x}$ je pak taková hodnota, že

$$\hat{x} = \arg\max_{x_i \in \boldsymbol{x}} \left\{ |\{x_i : x_i \in \boldsymbol{x}\}| \right\}.$$

Praktické je pro dopočítání absolutních četností jednotlivých navzájem různých hodnot v souboru použít místo souboru $\boldsymbol{x}$ raději vzestupně uspořádaný výběr $\boldsymbol{x}_{\text{uspořádaný}}$.

(i) Podle definice p-tého kvantilu hledáme $\tilde{x}_p$ pro $p = 0{,}25$ a soubor $\boldsymbol{x}_{\text{uspořádaný}}$ o $n = 12$ číslech.

Zřejmě součin $k = np = 12 \cdot 0{,}25 = 3 \in \mathbb{N}$, tedy podle definice je

$$\begin{aligned}
\tilde{x}_{0,25} &= \frac{1}{2}\left(x_{(k)} + x_{(k+1)}\right) = \\
&= \frac{1}{2}\left(x_{(np)} + x_{(np+1)}\right) = \\
&= \frac{1}{2}\left(x_{(3)} + x_{(3+1)}\right) = \\
&= \frac{1}{2}\left(x_{(3)} + x_{(4)}\right) = \\
&= \frac{1}{2}(1 + 2) = \\
&= \frac{3}{2}.
\end{aligned}$$

(ii) Podle definice p-tého kvantilu hledáme $\tilde{x}_p$ pro $p = 0{,}45$ a soubor $\boldsymbol{x}_{\text{uspořádaný}}$ o $n = 12$ číslech.

Zřejmě součin $k = np = 12 \cdot 0{,}45 = 5{,}4 \notin \mathbb{N}$, tedy podle definice je

$$\begin{aligned}
\tilde{x}_{0,45} &= x_{(\lfloor k \rfloor + 1)} = \\
&= x_{(\lfloor np \rfloor + 1)} = \\
&= x_{(\lfloor 5,4 \rfloor + 1)} = \\
&= x_{(5+1)} = \\
&= x_{(6)} = \\
&= 2.
\end{aligned}$$

(iii) Modus v souboru $\boldsymbol{x}$ nalezneme jako nejčetnější hodnotu. S výhodou spočítejme absolutní četnosti všech navzájem různých hodnot v souboru $\boldsymbol{x}$ a vyjádřeme je v tabulce 2.1.2.

x_i	1	2	3	4	5
absolutní četnost x_i	3	5	1	2	1

Tabulka 2.1.2: Absolutní četnosti navzájem různých hodnot souboru $\boldsymbol{x}$.

Z tabulky 2.1.2 je zřejmé, že modusem souboru $\boldsymbol{x}$ je hodnota $\hat{x} = 2$, protože jeji absolutní četnost (5) je mezi všemi četnostmi nejvyšší. □

Úloha 2.1.9 Určeme, jak se změní výběrový rozptyl a výběrová směrodatná odchylka, pokud se všechny hodnoty ve výběru

(i) zmenší o pět.
(ii) zvětší dvakrát.

Řešení. Označme hodnoty v souboru postupně $x_1, x_2, x_3, \ldots, x_n$ pro neznámé $n \in \mathbb{N}$. Pak dle zadání je výběrový rozptyl $s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$ a vyběrová směrodatná odchylka $s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$, kde $\bar{x}$ je vždy aritmetický průměr z čísel $x_1, x_2, x_3, \ldots, x_n$.

Prvně připomeňme, že pokud jsou původní hodnoty souboru $n \in \mathbb{N}$ čísel zvětšeny vždy q-krát, kde $q \in \mathbb{R} \setminus \{0\}$, a současně zvětšeny vždy o konstantu r, kde $r \in \mathbb{R}$, pak se původní aritmetický průměr souboru $\bar{x}$ změní na nový aritmetický průměr $\bar{x'}$ tak, že $\bar{x'} = q\bar{x} + r$, neboť

$$
\begin{aligned}
\bar{x'} &= \frac{1}{n} \sum_{i=1}^{n} x_i' = \\
&= \frac{1}{n} \sum_{i=1}^{n} (q x_i + r) = \\
&= \frac{1}{n} \left(\sum_{i=1}^{n} q x_i + \sum_{i=1}^{n} r \right) = \\
&= \frac{1}{n} \left(q \sum_{i=1}^{n} x_i + n \cdot r \right) = \\
&= q \cdot \frac{1}{n} \sum_{i=1}^{n} x_i + \frac{1}{n} \cdot n \cdot r = \\
&= q \cdot \frac{1}{n} \sum_{i=1}^{n} x_i + r = \\
&= q\bar{x} + r.
\end{aligned}
$$

(i) Pokud zmenšíme každou z n hodnot o pět, nové hodnoty mají $\forall i \in \{1, 2, 3, \ldots, n\}$ tvar $x_i' = x_i - 5$, stejně tak podle výše uvedeného má nový aritmetický průměr souboru tvar $\bar{x'} = \bar{x} - 5$. Pro nový vyběrový rozptyl $s^2_{x,\text{ nový}}$ pak platí

$$\begin{aligned} s^2_{x,\text{ nový}} &= \frac{1}{n-1}\sum_{i=1}^{n}(x'_i - \bar{x'})^2 = \\ &= \frac{1}{n-1}\sum_{i=1}^{n}((x_i - 5) - (\bar{x} - 5))^2 = \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - 5 - \bar{x} + 5)^2 = \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \\ &= s^2_x \end{aligned}$$

a pro novou výběrovou směrodatnou odchylku platí

$$\begin{aligned} s_{x,\text{ nový}} &= \sqrt{s^2_{x,\text{ nový}}} = \\ &= \sqrt{s^2_x} = \\ &= s_x. \end{aligned}$$

Pokud tedy zmenšíme každou z n hodnot souboru vždy o pět, výběrový rozptyl ani výběrová směrodatná odchylka se nijak nezmění.

(ii) Pokud zvětšíme každou z n hodnot dvakrát, nové hodnoty mají $\forall i \in \{1, 2, 3, \ldots, n\}$ tvar $x'_i = 2x_i$, stejně tak podle výše uvedeného má nový aritmetický průměr souboru tvar $\bar{x'} = 2\bar{x}$. Pro nový vyběrový rozptyl $s^2_{x,\text{ nový}}$ pak platí

$$\begin{aligned} s^2_{x,\text{ nový}} &= \frac{1}{n-1}\sum_{i=1}^{n}(x'_i - \bar{x'})^2 = \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(2x_i - 2\bar{x})^2 = \\ &= \frac{1}{n-1}\sum_{i=1}^{n}\left(2 \cdot (x_i - \bar{x})\right)^2 = \\ &= \frac{1}{n-1}\sum_{i=1}^{n}4 \cdot (x_i - \bar{x})^2 = \\ &= 4 \cdot \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \\ &= 4 \cdot s^2_x \end{aligned}$$

a pro novou výběrovou směrodatnou odchylku platí

$$\begin{aligned} s_{x,\text{ nový}} &= \sqrt{s^2_{x,\text{ nový}}} = \\ &= \sqrt{4 \cdot s^2_x} = \\ &= 2 \cdot s_x. \end{aligned}$$

Pokud tedy zvětšíme každou z n hodnot souboru dvakrát, výběrový rozptyl vzroste čtyřikrát a výběrová směrodatná odchylka vzroste dvakrát.

Poznámka. Snadno z předchozích řešení nahlédneme, že pokud jsou původní hodnoty souboru $n \in \mathbb{N}$ čísel zvětšeny vždy q-krát, kde $q \in \mathbb{R} \setminus \{0\}$, a současně zvětšeny vždy o konstantu r, kde $r \in \mathbb{R}$, nové hodnoty mají $\forall i \in \{1, 2, 3, \ldots, n\}$ tvar $x_i' = qx_i + r$ a aritmetický průměr souboru je dle výše uvedeného roven $\bar{x}' = q\bar{x} + r$. Původní výběrový rozptyl souboru s_x^2 změní na nový výběrový rozptyl $s^2_{x,\text{ nový}}$ tak, že $s^2_{x,\text{ nový}} = q^2 s_x^2$, neboť

$$\begin{aligned}
s^2_{x,\text{ nový}} &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i' - \bar{x}')^2 = \\
&= \frac{1}{n-1}\sum_{i=1}^{n}((qx_i + r) - (q\bar{x} + r))^2 = \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(qx_i + r - q\bar{x} - r)^2 = \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(qx_i - q\bar{x})^2 = \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(q \cdot (x_i - \bar{x})\right)^2 = \\
&= \frac{1}{n-1}\sum_{i=1}^{n} q^2 \cdot (x_i - \bar{x})^2 = \\
&= q^2 \cdot \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \\
&= q^2 s_x^2
\end{aligned}$$

a původní výběrová směrodatná odchylka souboru s_x změní na novou výběrovou směrodatnou odchylku $s_{x,\text{ nový}}$ tak, že $s_{x,\text{ nový}} = qs_x^2$, neboť

$$\begin{aligned}
s_{x,\text{ nový}} &= \sqrt{s^2_{x,\text{ nový}}} = \\
&= \sqrt{q^2 \cdot s_x^2} = \\
&= qs_x.
\end{aligned}$$

□

Úloha 2.1.10 V zemi *Statlandia* žijí dva kouzelníci, jeden vždy mluví pravdu a druhý vždy lže. Mladší z nich nám řekl, že ve všech ovčích stádech *Statlandie* je čtverec průměru počtu ovcí 128,6 a průměr čtverců počtu ovcí 115,4. Starší nám řekl, že je to naopak. Který z nich určitě lhal? Mladší, nebo starší?

Řešení. Buď $\bar{x}$ průměrný počet ovcí na stádo a $\overline{x^2}$ průměrný čtverec počtu ovcí na stádo ve *Statlandii.*

Protože je $s_x^2 = \overline{x^2} - \bar{x}^2$, je vždy i $\overline{x^2} - \bar{x}^2 = s_x^2 \geq 0$, a tedy $\overline{x^2} - \bar{x}^2 \geq 0$, čili vždy $\overline{x^2} \geq \bar{x}^2$. Proto musí být průměr čtverců počtu ovcí minimálně tak velký jako čtverec průměru počtu ovcí na stádo.

Zatímco starší kouzelník tvrdí, že čtverec průměru počtu ovcí všech stád je $\bar{x}^2 = 115{,}4$ a průměr čtverců počtu ovcí ve všech stádech je $\overline{x^2} = 128{,}6$, mladší tvrdí opak, tedy že čtverec průměru počtu ovcí všech stád je $\bar{x}^2 = 128{,}6$ a průměr čtverců počtu ovcí ve všech stádech je $\overline{x^2} = 115{,}4$.

Protože je vždy $\overline{x^2} \geq \bar{x}^2$, musela by dle mladšího kouzelníka platit nerovnost $115{,}4 \geq 128{,}6$. Mladší kouzelník tedy lhal. □

Úloha 2.1.11 Obchodní řetězec odebírá určitý výrobek, jehož cena v průběhu roku sezónně kolísá, od dvou stálých dodavatelů A a B. Průměrná cena za celý rok od dodavatele A je 9 Kč, její směrodatná odchylka činí 2 Kč, výrobků od dodavatele A se nakoupilo 1000 kusů. U dodavatele B činí průměrná cena 10 Kč při směrodatné odchylce 1 Kč, nákup od dodavatele B byl 4000 kusů. Určeme

(i) variační koeficient vyjadřující variabilitu kolísání nákupní ceny během roku souhrnně za oba dva dodavatele dohromady.

(ii) zda se na celkové variabilitě nákupní ceny větší měrou podílí průběžné sezónní kolísání cen výrobku u jednotlivých dodavatelů v rámci roku, nebo zda jsou důležitější rozdíly mezi průměrnými cenami jednotlivých dodavatelů.

Řešení. Údaje ze zadání o obou dodavatelích můžeme přehledně shrnout do tabulky 2.1.3.

	dodavatel	
	A	B
i	1	2
průměrná nákupní cena [Kč], $\bar{x}_i$	9	10
počet dodaných kusů, n_i	1000	4000
směrodatná odchylka nákupní ceny [Kč], $s_{x,i}$	2	1
rozptyl nákupní ceny [Kč^2], $s_{x,i}^2$	4	1

Tabulka 2.1.3: Údaje o dodavatelích.

Nadále tak pro dodavatele A značme počet dodaných výrobků $n_1 = 1000$, průměrnou nákupní cenu $\bar{x}_1 = 9$ [Kč], smědodatnou odchylku nákupní ceny $s_{x,1} = 2$ [Kč] a rozptyl nákupní ceny $s_{x,1}^2 = 4$ [Kč^2].

Obdobně pro dodavatele B značme počet dodaných výrobků $n_2 = 4000$, průměrnou nákupní cenu $\bar{x}_2 = 10$ [Kč], smědodatnou odchylku nákupní ceny pak $s_{x,2} = 1$ [Kč] a rozptyl nákupní ceny $s_{x,2}^2 = 1$ [Kč^2].

(i) K výpočtu celkového variačního koeficientu je třeba znát celkovou směrodatnou odchylku nákupní ceny a celkový průměr nákupní ceny, neboť

$$v_x = \frac{s_x}{\bar{x}}.$$

Celkový průměr nákupní ceny je zřejmě vážený průměr nákupních cen od obou dodavatelů, A a B, kde vahami jsou počty nakoupných kusů, tedy

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \frac{1000 \cdot 9 + 4000 \cdot 10}{1000 + 4000} = 9{,}8 \text{ [Kč]}.$$

Pro výpočet celkové směrodatné odchylky nákupní ceny je vhodné použít vzorec pro rozptyl komponovaný ze sčítance pro vnitroskupinovou a meziskupinovou variabilitu, tedy

$$s_x^2 = \overline{s^2} + s_{\bar{x}}^2 = \underbrace{\frac{1}{n}\left(\sum_{i=1}^{k} n_i s_{x,i}^2\right)}_{\text{vnitroskupinová variabilita}} + \underbrace{\frac{1}{n}\left(\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2\right)}_{\text{meziskupinová variabilita}},$$

kde v našem případě je $k = 2$ pro dva dodavatele A a B a $n = n_1 + n_2$. Spočítejme nejdříve vnitroskupinovou variabilitu, také někdy značenou $\overline{s^2}$,

$$\begin{aligned}
\overline{s^2} &= \frac{1}{n}\left(\sum_{i=1}^{k} n_i s_{x,i}^2\right) = \\
&= \frac{1}{n_1 + n_1}\left(\sum_{i=1}^{2} n_i s_{x,i}^2\right) = \\
&= \frac{n_1 s_{x,1}^2 + n_2 s_{x,2}^2}{n_1 + n_1} = \\
&= \frac{1000 \cdot 4 + 4000 \cdot 1}{1000 + 4000} = \\
&= \frac{8000}{5000} = \\
&= \frac{8}{5}.
\end{aligned}$$

Nyní vyčísleme meziskupinovou variabilitu, značenou někdy $s_{\bar{x}}^2$,

$$\begin{aligned}
s_{\bar{x}}^2 &= \frac{1}{n}\left(\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2\right) = \\
&= \frac{1}{n_1 + n_2}\left(\sum_{i=1}^{2} n_i(\bar{x}_i - \bar{x})^2\right) = \\
&= \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2}{n_1 + n_2} = \\
&= \frac{1000 \cdot (9{,}0 - 9{,}8)^2 + 4000 \cdot (10{,}0 - 9{,}8)^2}{1000 + 4000} = \\
&= \frac{1000 \cdot (-0{,}8)^2 + 4000 \cdot 0{,}2^2}{5000} = \\
&= \frac{1000 \cdot 0{,}64 + 4000 \cdot 0{,}04}{5000} =
\end{aligned}$$

$$s_{\bar{x}}^2 = \frac{800}{5000} = \\ = \frac{8}{50}.$$

Celkový rozptyl nákupní ceny je tak

$$s_x^2 = \overline{s^2} + s_{\bar{x}}^2 = \frac{8}{5} + \frac{8}{50} = \frac{88}{50} = 1{,}76 \text{ [Kč}^2\text{]}.$$

Nyní při znalosti celkového průměru nákupní ceny $\bar{x} = 9{,}8$ [Kč] a celkového rozptylu nákupní ceny $s_x^2 = 1{,}76$ [Kč^2] již snadno dopočítáme variační koeficient v_x, tedy

$$v_x = \frac{s_x}{\bar{x}} = \frac{\sqrt{s_x^2}}{\bar{x}} = \frac{\sqrt{1{,}76}}{9{,}80} \doteq 0{,}135.$$

Celkový variační koeficient nákupní ceny za oba dodavatele je tedy asi 0,135. To můžeme intereptovat i tak, že celková směrodatná odchylka nákupní ceny tvoří asi 13,5 % z celkového průměru nákupní ceny.

(ii) zda se na celkové variabilitě nákupní ceny větší měrou podílí průběžné sezónní kolísání cen výrobku u jednotlivých dodavatelů v rámci roku, nebo zda jsou důležitější rozdíly mezi průměrnými cenami jednotlivých dodavatelů.

Zatímco vnitroskupinová variabilita reprezentuje kolísání nákupní ceny kolem své průměrné hodnoty nákupní ceny v rámci každého dodavatele zvlášť, odpovídá tedy sezónnímu kolísání nákupních cen kolem svých průměrů v daném roce, meziskupinová variabilita naopak reflektuje to, jak se nákupní ceny navzájem lišily mezi oběma dodavateli během daného roku.

Protože $\overline{s^2} = \frac{8}{5} \gg \frac{8}{50} = s_{\bar{x}}^2$, je vnitroskupinová variabilita podstatně větší než meziskupinová. Lze tedy předpokládat, že sezónní kolísání nákupních cen v rámci každého dodavatele podstatně více určuje celkovou variabilitu než rozdíly mezi nákupními cenami obou dodavatelů v rámci roku. □

2.2 Úlohy k vlastnímu procvičení

Úloha 2.2.1 V tabulce 2.2.1 jsou uvedeny hodnoty osmiprvkového výběru $\boldsymbol{x} = (x_1, x_2, \ldots, x_8)^T$. Určete pro daný výběr jeho aritmetický průměr, medián, výběrový rozptyl, výběrovou směrodatnou odchylku, minimum, maximum, variační rozpětí, variační koeficient, modus, medián, první kvartil a 0,65-tý kvantil.

i	1	2	3	4	5	6	7	8
x_i	11	9	8	13	6	5	3	21

Tabulka 2.2.1: Hodnoty výběru $\boldsymbol{x} = (x_1, x_2, \ldots, x_8)^T$.

Úloha 2.2.2 V tabulce 2.2.2 jsou uvedeny hodnoty osmiprvkového výběru $\boldsymbol{x} = (x_1, x_2, \ldots, x_8)^T$. Určete pro daný výběr jeho aritmetický průměr, medián, výběrový rozptyl, výběrovou směrodatnou odchylku, minimum, maximum, variační rozpětí, variační koeficient, modus, medián, třetí kvartil a 0,18-tý kvantil.

i	1	2	3	4	5	6	7	8
x_i	12	9	6	14	6	5	2	21

Tabulka 2.2.2: Hodnoty výběru $\boldsymbol{x} = (x_1, x_2, \ldots, x_8)^T$

Úloha 2.2.3 Ve třídě 8. A bylo původně 32 žáků a ve třídě 8. B téže základní školy bylo původně 20 žáků. Určitý počet žáků byl z kázeňských důvodů přesunut ze třídy 8. A do třídy 8. B. Díky tomu vzrostl průměrný inteligenční kvocient zbylých žáků v 8. A ze 127 na 128, stejně tak příchodem přesunutých žáků vzrostl průměrný inteligenční kvocient všech žáků (včetně nově přesunutých) ve třídě 8. B ze 114 na 115. Určete, kolik žáků bylo přesunuto ze třídy 8. A do třídy 8. B a jaký byl jejich průměrný inteligenční kvocient.

Úloha 2.2.4 Loď pluje z říčního přístavu k ústí řeky rychlostí $v_1 > 0$, poté se vrací stejnou trasou zpět do přístavu rychlostí $v_2 > 0$. Určete průměrnou rychlost lodi na celkové trase z říčního přístavu k ústí řeky a zpět.

Úloha 2.2.5 Investor se rozhodl pravidelně investovat do určité akcie každý měsíc fixní částku 10 000 Kč po dobu 3 měsíců. Cena jedné akcie byla první měsíc 100 Kč, druhý měsíc 125 Kč a třetí měsíc 80 Kč. Jaká byla průměrná pořizovací cena jedné akcie?

Úloha 2.2.6 Na jisté katedře je věkový profil zaměstnanců takový, že šest z nich patří věkově do dekády $\langle 20, 30)$ let, čtyři z nich do dekády $\langle 30, 40)$ let sedm z nich do dekády $\langle 40, 50)$ let, dva z nich do dekády $\langle 50, 60)$ let a jeden z nich do dekády $\langle 60, 70)$ let. Určete absolutní a relativní četnosti všech dekád na dané katedře, současně určete i jejich kumulativní protějšky.

Úloha 2.2.7 Bezpečnostní agentura má 216 zaměstnanců a skládá se ze dvou dceřiných společností. V první dceřiné společnosti je průměrná měsíční mzda 21 650 Kč a v druhé 24 800 Kč. Průměrná mzda za celý holding je 23 650 Kč. Kolik zaměstnanců pracuje ve druhé dceřiné společnosti?

Úloha 2.2.8 V hudebním tělese je rozložení věku jeho hráčů následující (v letech),

$$22,\ 82,\ 27,\ 43,\ 19,\ 47,\ 41,\ 34,\ 34,\ 42,\ 35,\ 39.$$

Určeme

(i) mediánový věk hudebníků v tělese.
(ii) modální věk hudebníků v tělese.
(iii) první a třetí kvartil věku hudebníků v tělese.

(iv) 80-tý percentil věku hudebníků v tělese.

Úloha 2.2.9 V bytovém komplexu je celkem 78 domácností, z nichž 34 nemá žádné parkovací místo v podzemních garážích, 30 domácností má jedno parkovací místo, 6 domácností má dvě parkovací místa, 5 domácností má tři parkovací místa a 3 domácnosti mají dokonce čtyři parkovací místa. Jaký průměrný počet parkovacích míst připadajících na domácnost? Sestavme tabulku absolutních a relativních četností pro počet parkovacích míst, včetně kumulativních protějšků.

Úloha 2.2.10 První dělník je schopen vyhloubit výkop za 8 hodin, druhý dělník za 6 hodin a třetí dělník pak za 10 hodin.

(i) Za kolik hodin vyhloubí jeden výkop, pokud budou pracovat společně?

(ii) Za kolik průměrně hodin je vyhlouben výkop, pokud pracují všichni tří dělníci a každý pracuje na svých výkopech?

Úloha 2.2.11 Vlivem ekonomických událostí vzrostla průměrná cena letenek v určité oblasti o 10,00 %, zatímco rozptyl ceny těchto letenek vzrostl o 46,41 %. Určeme, jak se změnil variační koeficient ceny těchto letenek.

Úloha 2.2.12 Soubor o šesti hodnotách má průměr 12 a rozptyl $4\frac{2}{3}$. Jak se změní průměr a rozptyl souboru, když do něj přibude hodnota 15?

Úloha 2.2.13 Průměrný počet bodů z testu, který psalo 30 studentů, byl původně 78 a bodový rozptyl byl 81. Jak se změní průměrný počet bodů a bodový rozptyl, pokud si test dopsali ještě další tři studenti a získali postupně 99, 92 a 83 bodů?

Úloha 2.2.14 V souboru dvaceti evidovaných hodnot je aritmetický průměr roven 110 a výběrový rozptyl 800. Poté bylo zjištěno, že dvě hodnoty byly zaevidovány chybně – místo hodnoty 85 mělo být správně 95 a místo hodnoty 120 mělo být správně 150. Jaká je správná hodnota aritmetického průměru a výběrového rozptylu?

Úloha 2.2.15 Průměrná mzda pracovního týmu složeného ze šesti finančních analytiků je 29 456 Kč. Směrodatná odchylka mezd této skupiny je 5 460 Kč. Jak se změnila průměrná mzda a směrodatná odchylka ve skupině, jestliže byl do týmu přijat další člen s nástupním platem 22 000 Kč?

Úloha 2.2.16 V tabulce 2.2.3 jsou uvedeny hodnoty představující meziroční procentuální výnosy amerického podílového fondu SLASX pro roky 1998 až 2002.

rok	1998	1999	2000	2001	2002
výnos [%]	16,2	20,3	9,3	-11,1	-17,0

Tabulka 2.2.3: Meziroční procentuální výnosy podílového fondu SLASX.

Určeme

(i) průměrnou hodnotu výnosu SLASX, použijme aritmetický a geometrický průměr.

(ii) variabilitu výnosů SLASX, použijme výběrový rozptyl a průměr.

Úloha 2.2.17 Ve firmě bylo provedeno šetření týkající se výše mezd. Zjištěné údaje jsou shrnuty v tabulce 2.2.4.

pohlaví	průměrná mzda	počet zaměstnanců	variační koeficient
ženy	25 000	20	0,35
muži	29 000	30	0,25

Tabulka 2.2.4: Údaje o mzdách zaměstnanců ve firmě.

Určeme

(i) průměrnou mzdu ve firmě.
(ii) celkovou směrodatnou odchylku mezd ve firmě.
(iii) celkový variační koeficient mezd ve firmě.

Úloha 2.2.18 Do souboru deseti čísel o průměru 20 a rozptylu 4 byly přidány dvě nové hodnoty tak, že průměr nového souboru o dvanácti číslech je 21 a rozptyl 9. Která dvě čísla byla do souboru přidána?

3 Pravděpodobnost jevů

3.1 Řešené úlohy

Úloha 3.1.1 Házíme běžnou šestistěnnou hrací kostkou. Jaká je pravděpodobnost, že

(i) při jednom hodu padne šestka?
(ii) při jednom hodu padne sudé číslo?
(iii) při dvou hodech padne dvakrát šestka?
(iv) při dvou hodech nepadne ani jednou šestka?
(v) při dvou hodech padne alespoň jednou šestka?
(vi) při dvou hodech padne právě jednou šestka?
(vii) při dvou hodech padne alespoň jednou sudé číslo?

Kolikrát musíme minimálně hodit kostkou, aby byla pravděpodobnost, že během těchto hodů padne alespoň jedna šestka, větší než 0,95?

Řešení. Označme jednotlivé jevy následujícím způsobem. Jev A_1 je, že při jednom hodu kostkou padne šestka, jev A_2 je, že při druhém hodu padne šestka, stejně tak značme $P\left(\overline{A_1}\right)$ jev, že při jednom hodu kostkou nepadne šestka a $P\left(\overline{A_2}\right)$ jev, že šestka nepadne při druhém hodu. Dále ještě značme $P(A_k)$, resp. $P\left(\overline{A_k}\right)$ jev, že pro nějaké $k \in \mathbb{N}$ padne, resp. nepadne při k-tém hodu kostkou šestka. Jev B_1 je, že při jednom hodu padne sudé číslo, jev B_2 je, že při druhém hodu padne sudé číslo.

(i) Hledáme pravděpodobnost $P(A_1)$. Zřejmě stačí vyjít z klasické definice pravděpodobnosti, protože na běžné hrací šestistěnné kostce jsou všechny možnosti, tj. $\{1, 2, 3, 4, 5, 6\}$ stejně pravděpodobné.

Můžeme tedy psát

$$P(A_1) = \frac{|\{1\}|}{|\{1, 2, 3, 4, 5, 6\}|} = \frac{1}{6}.$$

Pravděpodobnost, že při jednom hodu běžnou šestistěnnou kostkou padne šestka, je tedy $\frac{1}{6}$.

(ii) Hledáme pravděpodobnost $P(B_2)$. Opět vyjdeme z klasické definice pravděpodobnosti a pišme

$$P(B_1) = \frac{|\{2, 4, 6\}|}{|\{1, 2, 3, 4, 5, 6\}|} = \frac{3}{6} = \frac{1}{2}.$$

Pravděpodobnost, že při jednom hodu běžnou šestistěnnou kostkou padne sudé číslo, je tedy $\frac{1}{2}$.

(iii) Hledáme pravděpodobnost $P(A_1 \cap A_2)$, protože šestka musí padnout v každém ze dvou hodů, tedy v prvním i ve druhém. Protože jsou oba hody danou kostkou na sobě jistě nezávislé, můžeme psát $P(A_1 \cap A_2) = P(A_1)P(A_2)$. Pak už jen upravme

$$P(A_1 \cap A_2) = P(A_1)P(A_2) = \frac{|\{1\}|}{|\{1,2,3,4,5,6\}|} \cdot \frac{|\{1\}|}{|\{1,2,3,4,5,6\}|} = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}.$$

Pravděpodobnost, že při dvou hodech běžnou šestistěnnou kostkou padne vždy šestka, je tedy $\frac{1}{36}$.

(iv) Hledáme pravděpodobnost $P\left(\overline{A_1} \cap \overline{A_2}\right)$, protože šestka nesmí padnout ani v prvním, ani ve druhé hodu. Protože jsou oba hody danou kostkou na sobě jistě nezávislé, můžeme psát $P\left(\overline{A_1} \cap \overline{A_2}\right) = P\left(\overline{A_1}\right) \cdot P\left(\overline{A_2}\right)$. Pak už jen upravme

$$\begin{aligned} P\left(\overline{A_1} \cap \overline{A_2}\right) &= P\left(\overline{A_1}\right) \cdot P\left(\overline{A_2}\right) = \\ &= \frac{|\{1,2,3,4,5\}|}{|\{1,2,3,4,5,6\}|} \cdot \frac{|\{1,2,3,4,5\}|}{|\{1,2,3,4,5,6\}|} = \\ &= \frac{5}{6} \cdot \frac{5}{6} = \\ &= \frac{25}{36}. \end{aligned}$$

Výsledke bychom obdobně získali i úvahou, že

$$\begin{aligned} P\left(\overline{A_1} \cap \overline{A_2}\right) &= P\left(\overline{A_1}\right) \cdot P\left(\overline{A_2}\right) = \\ &= (1 - P(A_1)) \cdot (1 - P(A_2)) = \\ &= \left(1 - \frac{1}{6}\right) \cdot \left(1 - \frac{1}{6}\right) = \\ &= \frac{5}{6} \cdot \frac{5}{6} = \\ &= \frac{25}{36}. \end{aligned}$$

Pravděpodobnost, že při dvou hodech běžnou šestistěnnou kostkou nepadne ani jednou šestka, je tedy $\frac{25}{36}$.

(v) Má-li šestka během dvou hodů kostkou padnout alespoň jednou, může dojít ke třem různým scénářům: šestka padne v prvním hodu a ve druhém ne, nebo padne ve druhém hodu a v prvním ne, anebo může padnout v obou hodech. Hledáme tedy pravděpodobnost

$$\begin{aligned} P &= P\left((A_1 \cap \overline{A_2}) \cup (\overline{A_1} \cap A_2) \cup (A_1 \cap A_2)\right) = \\ &= P\left(A_1 \cap \overline{A_2} \cup \overline{A_1} \cap A_2 \cup A_1 \cap A_2\right). \end{aligned}$$

Všechny tři scénáře se navzájem vždy liší alespoň v jednom hodu, jsou tedy nezávislé; současně první a druhý hod v rámci scénáře je na sobě také očividně nezávislý (kostka nemá žádnou historickou paměť na předchozí hod).

Můžeme tedy psát

$$
\begin{aligned}
P &= P\left(A_1 \cap \overline{A_2} \cup \overline{A_1} \cap A_2 \cup A_1 \cap A_2\right) = \\
&= P(A_1)P\left(\overline{A_2}\right) + P\left(\overline{A_1}\right)P(A_2) + P(A_1)P(A_2) = \\
&= P(A_1)\left(1 - P(A_2)\right) + \left(1 - P(A_1)\right)P(A_2) + P(A_1)P(A_2) = \\
&= \frac{1}{6} \cdot \left(1 - \frac{1}{6}\right) + \left(1 - \frac{1}{6}\right) \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \\
&= \frac{1}{6} \cdot \frac{5}{6} + \frac{5}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \\
&= \frac{5}{36} + \frac{5}{36} + \frac{1}{36} = \\
&= \frac{11}{36}.
\end{aligned}
$$

Úlohu můžeme vyřešit také tak, když si uvědomíme, že k jevu „během dvou hodů kostkou padne alespoň jedna šestka“ je jevem opačným „během dvou hodů kostkou nepadne ani jednou šestka“. Podle druhého výroku tedy zjevně platí rovnost $P = P\left(\overline{\overline{A_1} \cap \overline{A_2}}\right)$. Tu můžeme dále upravit

$$
\begin{aligned}
P &= P\left(\overline{\overline{A_1} \cap \overline{A_2}}\right) = \\
&= 1 - P\left(\overline{A_1} \cap \overline{A_2}\right) = \\
&= 1 - P\left(\overline{A_1}\right) \cdot P\left(\overline{A_2}\right) = \\
&= 1 - \left(1 - P(A_1)\right) \cdot \left(1 - P(A_2)\right) = \\
&= 1 - \left(1 - \frac{1}{6}\right) \cdot \left(1 - \frac{1}{6}\right) = \\
&= 1 - \frac{5}{6} \cdot \frac{5}{6} = \\
&= 1 - \frac{25}{36} = \\
&= \frac{11}{36}.
\end{aligned}
$$

Pravděpodobnost, že při dvou hodech běžnou šestistěnnou kostkou padne alespoň jednou šestka, je tedy $\frac{11}{36}$.

(vi) Má-li šestka během dvou hodů kostkou padnout právě jednou, může dojít ke dvěma různým scénářům: šestka padne v prvním hodu a ve druhém ne, nebo padne ve druhém hodu a v prvním ne. Hledáme tedy pravděpodobnost

$$
P = P\left(\left(A_1 \cap \overline{A_2}\right) \cup \left(\overline{A_1} \cap A_2\right)\right) = P\left(A_1 \cap \overline{A_2} \cup \overline{A_1} \cap A_2\right).
$$

Oba scénáře se navzájem vždy liší alespoň v jednom hodu, jsou tedy nezávislé; současně první a druhý hod v rámci scénáře je na sobě také očividně nezávislý.

Můžeme tedy psát

$$\begin{aligned}
P &= P\left(A_1 \cap \overline{A_2} \cup \overline{A_1} \cap A_2\right) = \\
&= P(A_1)P\left(\overline{A_2}\right) + P\left(\overline{A_1}\right)P(A_2) = \\
&= P(A_1)\left(1 - P(A_2)\right) + \left(1 - P(A_1)\right)P(A_2) = \\
&= \frac{1}{6} \cdot \left(1 - \frac{1}{6}\right) + \left(1 - \frac{1}{6}\right) \cdot \frac{1}{6} = \\
&= \frac{1}{6} \cdot \frac{5}{6} + \frac{5}{6} \cdot \frac{1}{6} = \\
&= \frac{5}{36} + \frac{5}{36} = \\
&= \frac{10}{36} = \\
&= \frac{5}{18}.
\end{aligned}$$

Pravděpodobnost, že při dvou hodech běžnou šestistěnnou kostkou padne právě jednou šestka, je tedy $\frac{5}{18}$.

(vii) Má-li sudé číslo během dvou hodů kostkou padnout alespoň jednou, může dojít ke třem různým scénářům: sudé číslo padne v prvním hodu a ve druhém ne, nebo padne ve druhém hodu a v prvním ne, anebo může padnout v obou hodech. Hledáme tedy pravděpodobnost

$$\begin{aligned}
P &= P\left((B_1 \cap \overline{B_2}) \cup (\overline{B_1} \cap B_2) \cup (B_1 \cap B_2)\right) = \\
&= P\left(B_1 \cap \overline{B_2} \cup \overline{B_1} \cap B_2 \cup B_1 \cap B_2\right).
\end{aligned}$$

Všechny tři scénáře se navzájem vždy liší alespoň v jednom hodu, jsou tedy nezávislé; současně první a druhý hod v rámci scénáře je na sobě také očividně nezávislý.

Můžeme tedy psát

$$\begin{aligned}
P &= P\left(B_1 \cap \overline{B_2} \cup \overline{B_1} \cap B_2 \cup B_1 \cap B_2\right) = \\
&= P(B_1)P\left(\overline{B_2}\right) + P\left(\overline{B_1}\right)P(B_2) + P(B_1)P(B_2) = \\
&= P(B_1)\left(1 - P(B_2)\right) + \left(1 - P(B_1)\right)P(B_2) + P(B_1)P(B_2) = \\
&= \frac{1}{2} \cdot \left(1 - \frac{1}{2}\right) + \left(1 - \frac{1}{2}\right) \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \\
&= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \\
&= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \\
&= \frac{3}{4}.
\end{aligned}$$

Úlohu můžeme vyřešit také tak, když si uvědomíme, že k jevu „během dvou hodů kostkou padne alespoň jedna sudé číslo“ je jevem opačným „během dvou hodů kostkou nepadne ani jednou sudé číslo“. Podle druhého výroku tedy zjevně platí rovnost $P = P\left(\overline{\overline{B_1} \cap \overline{B_2}}\right)$. Tu můžeme dále upravit

$$\begin{aligned}
P = P\left(\overline{\overline{B_1} \cap \overline{B_2}}\right) = \\
&= 1 - P\left(\overline{B_1} \cap \overline{B_2}\right) = \\
&= 1 - P\left(\overline{B_1}\right) \cdot P\left(\overline{B_2}\right) = \\
&= 1 - (1 - P(B_1)) \cdot (1 - P(B_2)) = \\
&= 1 - \left(1 - \frac{1}{1}\right) \cdot \left(1 - \frac{1}{2}\right) = \\
&= 1 - \frac{1}{2} \cdot \frac{1}{2} = \\
&= 1 - \frac{1}{4} = \\
&= \frac{3}{4}.
\end{aligned}$$

Pravděpodobnost, že při dvou hodech běžnou šestistěnnou kostkou padne alespoň jednou sudé číslo, je tedy $\frac{3}{4}$.

Předpokládeje, že běžnou šestistěnnou kostkou musíme hodit právě k-krát k tomu, aby byla pravděpodobnost, že během těchto hodů padne alespoň jedna šestka, větší než 0,95, kde $k \in \mathbb{N}$ je neznámá. Hledáme tedy takové $k \in \mathbb{N}$, aby platilo

$$P = P(\text{během } k \text{ hodů padne alespoň jedna šestka}) > 0{,}95.$$

Snadno nahlédneme, že k jevu „během k hodů kostkou padne alespoň jednou šestka“ je jevem opačným „během k hodů kostkou nepadne ani jednou šestka“. Toho využijme a vyjádřeme složený jev z druhého výroku. Pokud nemá během k hodů padnout ani jednou šestka, můžeme tento jev podle výše uvedené notace zapsat jakosti $\overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_k}$. Jev k němu opačný, tedy $\overline{\overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_k}}$, pak odpovídá jevu „během k hodů kostkou padne alespoň jednou šestka“. Můžeme tedy psát

$$\begin{aligned}
P = P(\text{během } k \text{ hodů padne alespoň jedna šestka}) = \\
&= P\left(\overline{\overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_k}}\right) = \\
&= 1 - P\left(\overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_k}\right) > \\
&> 0{,}95.
\end{aligned}$$

Jevy $\overline{A_1}, \overline{A_2}, \ldots, \overline{A_k}$, tedy že šestka nepadne v prvním, ve druhém, …, v k-tém hodu, jsou jistě na sobě nezávislé (neboť kostka nemá historickou paměť). Můžeme tedy psát

$$\begin{aligned}
P &= P(\text{během } k \text{ hodů padne alespoň jedna šestka}) = \\
&= P\left(\overline{\overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_k}}\right) = \\
&= 1 - P\left(\overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_k}\right) = \\
&= 1 - P\left(\overline{A_1}\right) P\left(\overline{A_2}\right) \cdots P\left(\overline{A_k}\right) = \\
&= 1 - (1 - P(A_1))\,(1 - P(A_2)) \cdots (1 - P(A_k)) = \\
&= 1 - \underbrace{\left(1 - \frac{1}{6}\right) \cdot \left(1 - \frac{1}{6}\right) \cdot \cdots \cdot \left(1 - \frac{1}{6}\right)}_{k \text{ členů}} = \\
&= 1 - \underbrace{\frac{5}{6} \cdot \frac{5}{6} \cdot \cdots \cdot \frac{5}{6}}_{k \text{ členů}} = \\
&= 1 - \left(\frac{5}{6}\right)^k > \\
&> 0{,}95.
\end{aligned}$$

Poslední nerovnost upravme

$$\begin{aligned}
1 - \left(\frac{5}{6}\right)^k &> 0{,}95 \\
\left(\frac{5}{6}\right)^k &< 1 - 0{,}95 \\
\left(\frac{5}{6}\right)^k &< 0{,}05 \\
\log \left(\frac{5}{6}\right)^k &< \log 0{,}05 \\
k \cdot \log \left(\frac{5}{6}\right) &< \log 0{,}05,
\end{aligned}$$

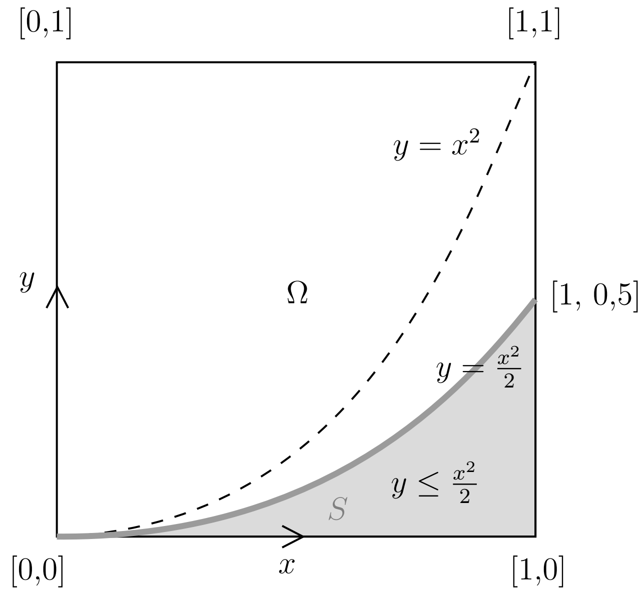
protože je $\left(\frac{5}{6}\right) < 0$, po vydělení obou stran nerovnosti tímto výrazem se změní znaménko nerovnosti,

$$\begin{aligned}
k \cdot \log \left(\frac{5}{6}\right) &< \log 0{,}05 \\
k &> \frac{\log 0{,}05}{\log \left(\frac{5}{6}\right)} \\
k &\gtrsim 16{,}4 \\
k &\geq 17.
\end{aligned}$$

Odtud již vidíme, že musíme hodit běžnou šestistěnnou kostkou alespoň 17-krát, aby byla pravděpodobnost, že během těchto hodů padne alespoň jedna šestka, větší než 0,95. $\square$

Úloha 3.1.2 Z intervalu $\langle 0,1\rangle$ náhodně vybereme dvě čísla x a y. Jaká je pravděpodobnost jevu, že $2y \leq x^2$?

Řešení. Předpokládejme, že čísla $x \in \langle 0,1\rangle$ a $y \in \langle 0,1\rangle$ jsou souřadnice bodu v rovině Oxy. Pak geometrickou množinou všech takových bodů v rovině, jejichž souřadnice $[x,y]$ splňují $x \in \langle 0,1\rangle \wedge y \in \langle 0,1\rangle$, je čtverec s vrcholy v bodech $[0,0],[1,0],[0,1],[1,1]$ v soustavě souřadnic Oxy, viz čtverec Ω na obrázku 3.1.1.



Obrázek 3.1.1: Čtverec s vrcholy v bodech $[0,0],[1,0],[0,1],[1,1]$ v soustavě souřadnic Oxy.

Hledáme pravděpodobnost, že $2y \leq x^2$, tedy hodnotu výrazu $P\left(2y \leq x^2\right)$ neboli $P\left(2y \leq x^2\right) = P\left(y \leq \frac{x^2}{2}\right)$. Snadno nahlédneme, že body, pro jejichž souřadnice $[x,y]$ platí $y = x^2$, leží na parabole s rovnicí $y \leq x^2$ v soustavě souřadnic Oxy (na obrázku 3.1.1 přerušovaná čára). Naopak body, pro jejichž souřadnice $[x,y]$ platí $y = \frac{x^2}{2}$, leží na parabole s rovnicí $y = \frac{x^2}{2}$ v soustavě souřadnic Oxy (na obrázku 3.1.1 šedá tučná čára). Souřadnice všech bodů, které leží v jednotkovém čtverci s vrcholy $[0,0],[1,0],[0,1],[1,1]$ a současně je jejich svislá, y-ová souřadnice pod parabolou danou rovnicí $y = \frac{x^2}{2}$, splňují nerovnost $y \leq \frac{x^2}{2}$ (na obrázku 3.1.1 šedá plocha S).

Zajímá nás, s jakou pravděpodobností leží náhodně vybraný bod ve čtverci Ω s vrcholy $[0,0],[1,0],[0,1],[1,1]$ současně v šedé ploše S. Protože jsou čísla x a y vybírána náhodně, je i bod v rámci čtverce Ω volen zcela náhodně a pravděpodobnost, že padne do šedé oblasti S, je rovna podílu plochy šedé oblasti S a celkové plochy čtverce s vrcholy $[0,0],[1,0],[0,1],[1,1]$, tedy $P\left(2y \leq x^2\right) = P\left(y \leq \frac{x^2}{2}\right) = \frac{|S|}{|\Omega|}$. Celková plocha jednotkového čtverce s vrcholy $[0,0],[1,0],[0,1],[1,1]$ je zjevně $|\Omega| = 1 \cdot 1 = 1$. Plocha šedé oblasti S je rovna určitému integrálu funkce $y = \frac{x^2}{2}$ v mezích $\langle 0,1\rangle$, tedy

$$|S| = \int_0^1 \frac{x^2}{2}\mathrm{d}x = \left[\frac{x^3}{6}\right]_0^1 = \frac{1^3}{6} - \frac{0^3}{6} = \frac{1}{6}.$$

Nakonec je tedy

$$P\left(2y \leq x^2\right) = P\left(y \leq \frac{x^2}{2}\right) = \frac{|S|}{|\Omega|} = \frac{\frac{1}{6}}{1} = \frac{1}{6}.$$

Vybíráme-li čísla x a y vždy z intervalu $\langle 0, 1\rangle$ zcela náhodně, pak splňují nerovnost $2y \leq x^2$ s pravděpodobností $\frac{1}{6}$. □

3.2 Úlohy k vlastnímu procvičení

Úloha 3.2.1 Směs sazenic je tvořena ze 80 % sazenicemi první jakosti a z 20 % sazenicemi druhé jakosti. Sazenice první jakosti vyklíčí s pravděpodobností 0,9, sazenice druhé jakosti vyklíčí s pravděpodobností 0,6. S jakou pravděpodobností vyklíčí náhodně vybraná sazenice ze směsi?

Úloha 3.2.2 Z intervalu $\langle 0, 1\rangle$ náhodně vybereme dvě čísla x a y. Jaká je pravděpodobnost jevu, že $\frac{1}{2} \leq x + y \leq \frac{3}{2}$?

Úloha 3.2.3 V nákladu 150 pytlů ořechů z Turecka je 5 pytlů se zkaženými ořechy. Podobně v nákladu 250 pytlů z Afghánistánu je 5 pytlů se zkaženými ořechy.

(i) S jakou pravděpodobností obsahuje náhodně vybraný pytel zkažené ořechy, pokud oba náklady zkompletujeme do jednoho?
(ii) S jakou pravděpodobností obsahuje náhodně vybraný pytel zkažené ořechy, pokud nejdříve vybereme náhodně náklad a pak náhodně jeden pytel?

Úloha 3.2.4 Je možné, aby dva jevy byly neslučitelné a současně i nezávislé? Zkoumejme.

Úloha 3.2.5 Házíme čtyřmi hracími kostkami. Vypočítejme pravděpodobnost následujících jevů.

(i) Na všech kostkách padnou různá čísla.
(ii) Na dvou kostkách padnou stejná čísla a na dalších dvou jiná různá čísla.
(iii) Na třech kostkách padnou stejná čísla a na další jiné číslo.
(iv) Padnou dvě (různé) dvojice stejných čísel.
(v) Na všech kostkách stejná čísla.
(vi) Padnou alespoň tři stejná čísla.

Úloha 3.2.6 Je dána úsečka AB. Náhodně na ní zvolíme dva body X a Y. Jaká je pravděpodobnost, že úsečka XY obsahuje střed úsečky AB?

Úloha 3.2.7 Ve třídě je dohromady lichý počet žáků, chlapců a dívek, zároveň předpokládáme, že jich není méně než 10 ani více než 45. Pravděpodobnost, že ze dvou náhodně vybraných žáků je jedna dívka a jeden chlapec, je stejná jako pravděpodobnost, že tomu tak není. Určete počet žáků ve třídě.

Úloha 3.2.8 Adam s Bětkou jedou ve stejném vlaku, v kterém je právě $k \in \mathbb{N}$ kupé vždy o šesti místech. S jakou pravděpodobností budou sedět oba ve stejném kupé, pokud si místa k sezení vybírají zcela náhodně ve chvíli, kdy jsou ještě všechna místa volná? Obecnou pravděpodobnost nakonec ověřme pro $k = 1$.

Úloha 3.2.9 S jakou pravděpodobností má kvadratická rovnice

$$x^2 + \sqrt{10 - a^2 - b^2} \cdot x + (a^2 + b^2) = 0$$

pro přípustné hodnoty parametrů $a, b \in \mathbb{R}$ oba kořeny reálné?

Úloha 3.2.10 Házíme jedenkrát šipkou na kruhový jednotkový[6] terč, který s jistotou zasáhneme; každý bod terče má stejnou pravděpodobnost zásahu. Pravděpodobnost, že se trefíme do vzdálenosti větší než p od středu terče, je p. Určete p.

Úloha 3.2.11 Házíme pětkrát spravedlivou mincí. Jaká je pravděpodobnost, že

(i) padne právě dvakrát hlava?
(ii) padne nejvýše dvakrát hlava?
(iii) padne alespoň dvakrát hlava?

Úloha 3.2.12 Házíme takovou nespravedlivou (biasovanou) šestistěnnou hrací kostkou, že pravděpodobnost padnutí každé stěny je přímo úměrná počtu ok na této stěně. Určeme pravděpodobnost, že padne sudý počet ok.

Úloha 3.2.13 Jaká je pravděpodobnost toho, že ve skupině n osob mají alespoň dva lidé narozeniny ve stejný den? Předpokládejme, že rok má 365 dní.

Úloha 3.2.14 Správce má v kapse celkem n klíčů od různých dveří budovy, mezi nimi i jeden klíč k hlavním dveřím. Před hlavními dveřmi postupně tahá klíče z kapsy jeden po druhém a zkouší jimi dveře odemknout. Vyzkoušené klíče zpět do kapsy nevrací. Jaká je pravděpodobnost, že dveře otevře právě až na k-tý pokus? Jaká je pravděpodobnost, že dveře odemkne nejpozději na l-tý pokus? A jaká je pravděpodobnost, že dveře neodemkne dříve než na m-tý pokus? Diskutujte řešení i vzhledem k přirozeným parametrům n, k, l, m. Změní se nějak pravděpodobnost, pokud budou klíče navlečeny na kroužku a v tomto pořadí je bude správce zkoušet?

Úloha 3.2.15 Hokejové utkání skončilo prohrou domácího týmu 7 : 12 (branek). S jakou pravděpodobností byl někdy během zápasu stav branek 5 : 3? Všechny smysluplné možnosti stavu zápasu (co do počtu branek), tj.

$$\begin{pmatrix} 0:0 & 0:1 & 0:2 & \cdots & 0:12 \\ 1:0 & 1:1 & 1:2 & \cdots & 1:12 \\ 2:0 & 2:1 & 2:2 & \cdots & 2:12 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 7:0 & 7:1 & 7:2 & \cdots & 7:12 \end{pmatrix}$$

považujeme za stejně pravděpodobné.

[6] Poloměr terče je roven přesně 1,0.

4 Diskrétní náhodná veličina a diskrétní rozdělení pravděpodobnosti

4.1 Řešené úlohy

Úloha 4.1.1 Mějme spravedlivou šestistěnnou kostku s jedním až šesti oky na první až šesté stěně. Uvažujme náhodnou veličinu X, která vrací počet ok, jež padnou na horní stěně kostky při jednom jejím hodu.

(i) Najděme pravděpodobnostní funkci náhodné veličiny X a načrtněme obrázek takové pravděpodobnostní funkce.
(ii) Najděme disktribuční funkci náhodné veličiny X a načrtněme obrázek takové distribuční funkce.
(iii) Spočítejme střední hodnotu a rozptyl náhodné veličiny X.

Řešení. Postupně vyřešme obě části.

(i) Nalézt pravděpodobnostní funkci (diskrétní) náhodné veličiny znamená najít ke každé reálné hodnotě příslušnou pravděpodobnost, s jakou může daná náhodná veličina téty hodnoty nabýt. Obvykle se u diskrétních náhodných veličin zabýváme pouze takovými reálými hodnotami, kterých může veličina skutečně nabýt. V našem případě jsou to počty ok na horní stěně kostky při jejím hodu, tedy $\{1, 2, 3, \ldots, 6\}$. Protože jde o spravedlivou šestistěnnou kostku, zřejmě může každý počet ok padnout se stejnou pravděpodobností. Tedy

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

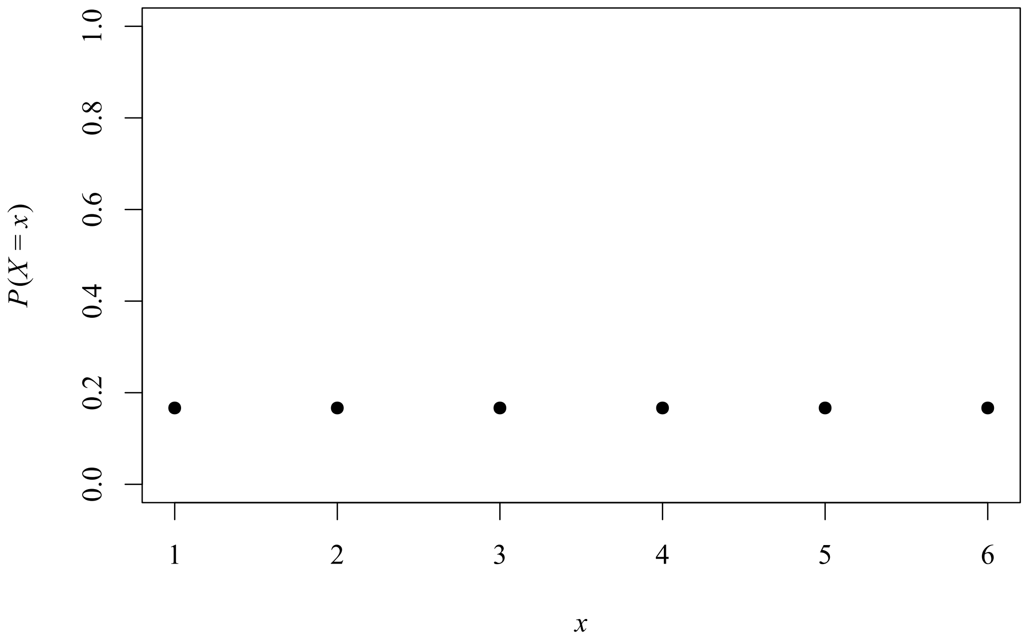
Neboli

$$\begin{aligned} &P(X = 1) = P(X = 2) = P(X = 3) = \\ &= P(X = 4) = P(X = 5) = P(X = 6) = \frac{1}{6}, \end{aligned}$$

což lze přepsat v podobě pravděpodobnostní funkce $P_X(x)$ též jako

$$P_X(x) = P(X = x) = \begin{cases} \frac{1}{6}, & x \in \{1, 2, 3, \dots, 6\} \\ 0, & \text{jinak.} \end{cases}$$

Díky výše uvedenému tvaru pravděpodobnostní funkce $P_X(x)$ již snadno načtrtneme i obrázek pravděpodobnostní funkce, viz obrázek 4.1.1.



Obrázek 4.1.1: Pravděpodobnostní funkce $P(X = x)$ vracející počet ok při jednom hodu spravedlivou šestistěnnou kostkou.

(ii) Nalézt distribuční funkci (diskrétní) náhodné veličiny X znamená nalézt takovou funkci, která pro libovolné reálné x vrací pravděpodobnost, že náhodná veličina X nebude větší než X, tedy $P(X \leq x)$. Známe-li již pravděpodobnostní funkci $P_X(x) = P(X = x)$ diskrétní náhodné veličiny X, můžeme obvykle sestavit distribuční funkci $F_X(x) = P(X \leq x)$ pomocí vztahu

$$F_X(x) = P(X \leq x) = \sum_{\forall \chi \in X : \chi \leq x} P(X = \chi).$$

To nám umožní sestavit následující tabulku.

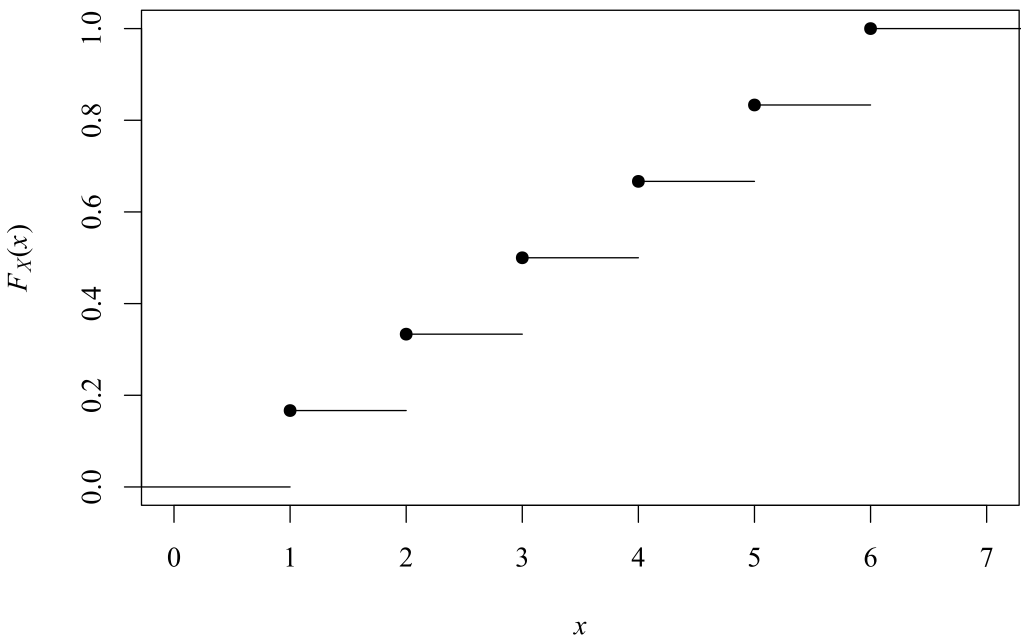
x	$F_X(x)$	číselně
$(-\infty, 1)$	$P(X < 1) = 0$	0
$\langle 1, 2)$	$P(X < 1) + P(X = 1) = 0 + \frac{1}{6}$	$\frac{1}{6}$
$\langle 2, 3)$	$P(X < 2) + P(X = 2) = \frac{1}{6} + \frac{1}{6}$	$\frac{1}{3}$
$\langle 3, 4)$	$P(X < 3) + P(X = 3) = \frac{1}{3} + \frac{1}{6}$	$\frac{1}{2}$
$\langle 4, 5)$	$P(X < 4) + P(X = 4) = \frac{1}{2} + \frac{1}{6}$	$\frac{2}{3}$
$\langle 5, 6)$	$P(X < 5) + P(X = 5) = \frac{2}{3} + \frac{1}{6}$	$\frac{5}{6}$
$\langle 6, +\infty)$	$P(X < 6) + P(X \geq 6) = \frac{5}{6} + \frac{1}{6}$	1

Tedy např. je-li $x \in (-\infty, 1)$, pak zřejmě na kostce nemůže padnout žádný počet ok nepřevyšující $x < 1$, tedy $P(X < 1) = P(X = 0) = 0$. Pro $x \in \langle 1, 2)$ ale již může padnout jedno oko, protože $1 \leq x \in \langle 1, 2)$, tedy $P(X < 2) = P(X = 1) = F_X(1) = \frac{1}{6}$. Naopak je-li $x \in \langle 6, +\infty)$, pak na kostce může padnout jedno až šest ok a vždy bude platit, že $6 \leq x \in \langle 6, +\infty)$, tedy $P(X \geq 6) = P(X = 1) + P(X = 2) + \cdots + P(X = 6) = F_X(6) = 1$.

Tabulku lze přepsat v podobě distribuční funkce $F_X(x) = P(X \leq x)$ též jako

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & x < 1 \\ \frac{\lfloor x \rfloor}{6}, & x \in \langle 1, 6) \\ 1, & x \geq 6. \end{cases}$$

Díky výše uvedenému tvaru distribuční funkce $F_X(x) = P(X \leq x)$ již snadno načtrtneme i obrázek distriburční funkce, viz obrázek 4.1.2.



Obrázek 4.1.2: Distribuční funkce $F_X(x) = P(X \leq x)$ vracející pravděpodobnost, že počet ok při jednom hodu spravedlivou šestistěnnou kostkou nepřevýší x.

(iii) Pro náhodnou veličinu X, kterou je počet ok, jež padnou při jednom hodu spravedlivou šestistěnnou kostkou, platí, že počet možných výsledků hodu kostkou jakožto náhodného pokusu je $k = 6$ různých počtů ok, tedy $X \in \{1, 2, 3, \ldots, 6\}$. Pro střední hodnotu $\mathbb{E}(X)$ náhodné veličiny X tak je

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{i=1}^{k} x_i P(X = x_i) = \\
&= \sum_{i=1}^{k} x_i P(x_i) = \\
&= \sum_{i=1}^{6} i \cdot \frac{1}{6} = \\
&= \frac{1}{6}(1+2+3+4+5+6) \\
&= \frac{7}{2}.
\end{aligned}
$$

Pro rozptyl $\operatorname{var}(X)$ náhodné veličiny X platí

$$
\begin{aligned}
\operatorname{var}(X) &= \sum_{i=1}^{k} (x_i - \mathbb{E}(X))^2 P(X = x_i) = \\
&= \sum_{i=1}^{k} (x_i - \mathbb{E}(X))^2 P(x_i) = \\
&= \sum_{i=1}^{k} x_i^2 P(x_i) - \left(\sum_{i=1}^{k} x_i P(x_i) \right)^2 = \\
&= \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) - \left(\frac{1}{6}(1+2+3+4+5+6) \right)^2 = \\
&= \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) - \left(\frac{7}{2} \right)^2 = \\
&= \frac{105}{36}.
\end{aligned}
$$

□

Úloha 4.1.2 V urně je deset bílých a třicet černých koulí. S jakou pravděpodobností budou mezi šesti náhodně vytaženými koulemi právě čtyři černé, pokud

(i) každou kouli po vytažení vždy vrátíme zpět do urny?
(ii) koule po vytažení do urny zpět nevracíme?

Řešení. Postupně vyřešme obě části.

(i) Pokud po každém tahu vrátíme kouli zpět do urny, je pravděpodobnost, že bude vytažena černá koule, konstantní během všech šesti tahů, a sice je rovna podílu počtu černých a všech koulí v urně, tedy $p_{\text{černá}} = \frac{30}{30+10} = \frac{3}{4}$. Obdobně pravděpodobnost vytažení bílé koule se díky vracení vytažených koulí vždy zpět do urny nemění a je během všech šesti tahů rovna podílu počtu bílých a všech koulí, tedy $p_{\text{bílá}} = \frac{10}{10+30} = \frac{1}{4}$.

Označme symbolem X náhodnou veličinu vracející počet vytažení černých koulí během šesti tahů z urny s vracením. Zřejmě, protože jednotlivé tahy koulí z urny jsou díky jejich zpětnému vracení do urny na sobě nezávislé, sleduje náhodná veličina X binomické rozdělení s počtem pokusů $n = 6$ a pravděpodobností úspěchu rovnou $p_{\text{černá}} = \frac{3}{4}$, formálně tedy $X \sim \text{binom}(n, p) \sim \text{binom}(n, p_{\text{černá}}) \sim \text{binom}\left(6, \frac{3}{4}\right)$.

Hledáme hodnotu výrazu $P(X = 4)$. Podle definice pravděpodobnostní funkce binomického rozdělení dostáváme pravděpodobnost pro vytažení právě k černých koulí rovnou

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

kde v našem případě je $k = 4$. Postupně tedy dostáváme

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\begin{aligned}
P(X = k) &= \binom{n}{k} p_{\text{černá}}^k (1 - p_{\text{černá}})^{n-k} \\
P(X = 4) &= \binom{6}{4} \left(\frac{3}{4}\right)^4 \left(1 - \left(\frac{3}{4}\right)\right)^{6-4} \\
P(X = 4) &= \binom{6}{4} \left(\frac{3}{4}\right)^4 \left(\frac{1}{4}\right)^2 \\
P(X = 4) &\doteq 15 \cdot 0{,}316 \cdot 0{,}0625 \\
P(X = 4) &\doteq 0{,}297.
\end{aligned}$$

Pravděpodobnost, že mezi šesti nezávislými tahy koulí z dané urny budou právě čtyři černé, uvažujeme-li vždy tažené koule po každém tahu zpět do urny, je tak rovna přibližně 0,297.

(ii) Pokud koule po jejich vytažení z urny nevracíme zpět, jednotlivé tahy již nejsou nezávislé a stejně tak ani pravděpodobnost vytažení černé koule není po každém z tahů shodná.

Označme symbolem Y náhodnou veličinu vracející počet vytažení černých koulí během šesti tahů z urny, ale bez zpětného vracení koulí zpět vždy po jejich vytažení. Zřejmě jednotlivé tahy nejsou nezávislé a náhodná veličina Y sleduje hypergeometrické rozdělení s parametrem $N = 40$ odpovídajícím počtu všech koulí v urně, parametrem $M = 30$ odpovídajícím počtu černých koulí v urně a parametrem $n = 6$ odpovídajícímu počtu tahů z urny bez vracení. Formálně tedy $Y \sim \text{hypergeom}(N, M, n) \sim \text{hypergeom}(40, 30, 6)$.

Hledám hodnotu výrazu $P(Y = 4)$. Podle definice pravděpodobnostní funkce hypergeometrického rozdělení dostáváme pravděpodobnost pro vytažení právě k černých koulí rovnou

$$P(Y = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},$$

kde v našem případě je $k = 4$. Postupně tedy dostáváme

$$\begin{aligned}
P(Y = k) &= \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \\
P(Y = 4) &= \frac{\binom{30}{4}\binom{10}{2}}{\binom{40}{6}} \\
P(X = 4) &= \frac{27405 \cdot 45}{3838380} \\
P(X = 4) &\doteq 0{,}321.
\end{aligned}$$

Pravděpodobnost, že mezi šesti tahy koulí z dané urny budou právě čtyři černé, neuvažujeme-li vrácení tažené koule po každém tahu zpět do urny, je tak rovna přibližně 0,321. $\square$

Úloha 4.1.3 V malé nemocnici zaznamenají vždy od pondělí do neděle průměrně 35 porodů.

(i) Kolik porodů můžeme v dané nemocnici očekávat následující den?
(ii) Kolik porodů můžeme v dané nemocnici očekávat během následujících tří dní?
(iii) Jaký je očekávaný rozptyl počtu porodů v dané nemocnici během následujících tří dní?
(iv) S jakou pravděpodobností dojde v dané nemocnici během následujících tří dní právě k 12 porodům?
(v) S jakou pravděpodobností dojde v dané nemocnici během následujícího dne k nejvýše dvěma porodům?
(vi) S jakou pravděpodobností dojde v dané nemocnici během následujícího dne k alespoň jednomu porodu?
(vii) S jakou pravděpodobností dojde v dané nemocnici během následujícího dne právě ke čtyřem, pěti nebo šesti porodům?

Řešení. Označme symbolem X náhodnou veličinu udávající počet porodů za jeden den v dané malé nemocnici. Předpokládejme, že počty porodů mezi jednotlivými dny jsou navzájem nezávislé, pak náhodná veličina X sleduje Poissonovo rozdělení o parametru $\lambda = \frac{35}{7} = 5$ [porodů/den]. Formálně tedy pišme $X \sim \text{poiss}(\lambda) \sim \text{poiss}(5)$. Ještě dodejme, že střední hodnota náhodné veličiny X a její rozptyl jsou pak oba rovny parametru λ, tedy $\mathbb{E}(X) = \text{var}(X) = \lambda = 5$.

(i) Předpokládejme, že počet porodů „následující“ den je zcela nezávislý na všech ostatních dnech. Pak počet porodů, kolik můžeme v dané malé nemocnici očekávat během následujícího dne, je roven střední hodnotě náhodné veličiny X, tedy $\mathbb{E}(X) = \lambda = 5$ [porodů].

(ii) Předpokládejme, že počty porodů během „následujících“ tří dnů jsou zcela nezávislé na všech ostatních dnech a rovněž nezávislé navzájem. Pak označme symbolem Y náhodnou veličinu udávající počet porodů během následujících tří dnů. Zřejmě jde o součet počtu porodů následující, druhý následující a třetí následující den, tedy $Y = X + X + X = 3X$. Protože víme, že součet k veličin sledujících Poissonovo rozdělení vždy o parametru λ rovněž sleduje Poissonovo rozdělení, avšak o parametru $k\lambda$, platí, že $Y = 3X \sim \text{poiss}(3\lambda) \sim \text{poiss}(3 \cdot 5) \sim \text{poiss}(15)$. Stejně tak platí, že $\mathbb{E}(Y) = \text{var}(Y) = 3\lambda = 3 \cdot 5 = 15$. Proto můžeme během následujících tří dní očekávat v dané nemocnici právě $\mathbb{E}(Y) = 15$ porodů.

(iii) Počet očekávaných porodů v dané nemocnici během následujících tří dní je popsán veličinou Y. Proto očekávaný rozptyl počtu porodů během následujících tří dní je roven $\text{var}(Y) = 15$ [porodů].

(iv) Hledáme hodnotu pravděpodobnosti $P(Y = 12)$. Ze vztahu pro pravděpodobnostní funkci Poissonova rozdělení dostáváme

$$P(Y = k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

kde v našem případě je $k = 12$. Postupně tedy dostáváme

$$\begin{aligned} P(Y = k) &= \frac{\lambda^k e^{-\lambda}}{k!} \\ P(Y = 12) &= \frac{15^{12} e^{-15}}{12!} \\ P(X = 12) &\doteq 0{,}083. \end{aligned}$$

Během následujících tří dní tedy v dané nemocnici dojde ke dvanácti porodům s pravděpodobností přibližně 0,083.

(v) Hledáme hodnotu pravděpodobnosti $P(X \leq 2)$, což lze vyjádřit vzhledem k nezávislosti jednotlivých počtů porodů také jako $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$. Ze vztahu pro pravděpodobnostní funkci Poissonova rozdělení víme, že $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ a dostáváme

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ P(Y \leq 2) &= \frac{5^0 e^{-5}}{0!} + \frac{5^1 e^{-5}}{1!} + \frac{5^2 e^{-5}}{2!} \\ P(X \leq 2) &\doteq 0{,}0067 + 0{,}0337 + 0{,}0842 \\ P(X \leq 2) &\doteq 0{,}125. \end{aligned}$$

Během následujícího dne tedy v dané nemocnici dojde nejvýše ke dvěma porodům s pravděpodobností přibližně 0,125.

(vi) Hledáme hodnotu pravděpodobnosti $P(X \geq 1)$, což lze vyjádřit vzhledem k nezávislosti jednotlivých počtů porodů také jako $P(X \geq 1) = 1 - P(X = 0)$. Ze vztahu pro pravděpodobnostní funkci Poissonova rozdělení víme, že $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ a dostáváme

$$\begin{aligned}
P(X \geq 1) &= 1 - P(X = 0)\\
P(Y \geq 1) &= 1 - \frac{5^0 e^{-5}}{0!}\\
P(Y \geq 1) &= 1 - 0{,}0067\\
P(X \geq 1) &\doteq 0{,}993.
\end{aligned}$$

Během následujícího dne tedy v dané nemocnici dojde alespoň k jednomu porodu s pravděpodobností přibližně 0,993.

(vii) Hledáme hodnotu pravděpodobnosti $P(X \in \{4,\ 5,\ 6\})$, což lze vyjádřit vzhledem k nezávislosti jednotlivých počtů porodů také jako $P(X \in \{4,\ 5,\ 6\}) = P(X = 4) + P(X = 5) + P(X = 6)$. Ze vztahu pro pravděpodobnostní funkci Poissonova rozdělení víme, že $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ a dostáváme

$$\begin{aligned}
P(X \in \{4,\ 5,\ 6\}) &= P(X = 4) + P(X = 5) + P(X = 6)\\
P(X \in \{4,\ 5,\ 6\}) &= \frac{5^4 e^{-5}}{4!} + \frac{5^5 e^{-5}}{5!} + \frac{5^6 e^{-5}}{6!}\\
P(X \in \{4,\ 5,\ 6\}) &\doteq 0{,}1755 + 0{,}1755 + 0{,}1462\\
P(X \in \{4,\ 5,\ 6\}) &\doteq 0{,}497.
\end{aligned}$$

Během následujícího dne tedy v dané nemocnici dojde k právě čtyřem, pěti nebo šesti porodům s pravděpodobností přibližně 0,497. □

4.2 Úlohy k vlastnímu procvičení

Úloha 4.2.1 V krabici je $b \in \mathbb{N}$ bílých koulí a $c \in \mathbb{N}$ černých koulí. Z krabice vytáhneme najednou tři koule.

(i) S jakou pravděpodobností jsou všechny tři bílé?
(ii) S jakou pravděpodobností jsou všechny dvě bílé a jedna černá?

Úloha 4.2.2 V urně je sedm bílých a tři černé koule. Náhodně vybereme tři koule.

(i) S jakou pravděpodobností jsou všechny bílé, pokud jsme je vybírali tak, že byly vždy vraceny zpět do urny po každém tahu?
(ii) S jakou pravděpodobností jsou všechny bílé, pokud jsme je vybírali tak, že nebyly po jednotlivých tazích vraceny zpět do urny?

Úloha 4.2.3 Určitá léčba je úspěšná průměrně u 80 % pacientů. S jakou pravděpodobností bude v kohortě deseti pacientů léčba neúspěšná nejvýše u jednoho z pacientů?

Úloha 4.2.4 Házíme desetkrát nespravedlivou (biasovanou) mincí, na které padne hlava s pravděpodobností 0,6.

(i) S jakou pravděpodobností padne ve všech deseti hodech desetkrát hlava?
(ii) S jakou pravděpodobností padne mezi deseti hody právě čtyřikrát hlava?
(iii) S jakou pravděpodobností padne mezi deseti hody alespoň osmkrát hlava?
(iv) Jaký počet padnutí hlavy je mezi deseti hody nejpravděpodobnější?

Úloha 4.2.5 V zásilce třiceti žárovek jsou tři vadné.

(i) Vybereme-li ze zásilky náhodně pět žárovek, s jakou pravděpodobností mezi nimi nebude žádná vadná?
(ii) Vybereme-li ze zásilky náhodně pět žárovek, s jakou pravděpodobností mezi nimi bude právě jedna vadná?
(iii) Kolik žárovek musíme náhodně ze zásilky vyzkoušet, aby pravděpodobnost, že objevíme alespoň jednu vadnou, byla větší než 0,7?

Úloha 4.2.6 Student odpoví na otázku v testu bodovanou jedním bodem správně s pravděpodobnostní $p = 0{,}7$. Jaká je očekávaná střední hodnota a rozptyl počtu bodů, které student za takovou otázku získá?

Úloha 4.2.7 Házíme desetkrát klasickou šestistěnnou kostkou.

(i) S jakou pravděpodobností padne právě čtyřikrát šestka?
(ii) S jakou pravděpodobností padne šestka nejvýše dvakrát?
(iii) S jakou pravděpodobností padne šestka alespoň třikrát?
(iv) A s jakou pravděpodobností padne šestka právě čtyřikrát, pětkrát nebo šestkrát?

Úloha 4.2.8 Ve vybraném periodiku se objevují průměrně čtyři překlepy na každých jeho deset stránek. S jakou pravděpodobností bude na náhodně vybrané stránce periodika

(i) žádný překlep?
(ii) jeden překlep?
(iii) dva překlepy?
(iv) více než dva překlepy?

Úloha 4.2.9 V sérii po dvě stě kusech je deset zmetků. Při přejímce náhodně vybereme pět kusů a podrobíme je destrukční zkoušce. Pokud není mezi pěti vybranými ani jeden zmetek, sérii přijmeme. S jakou pravděpodobností to nastane?

Úloha 4.2.10 Bětka napsala na tabuli náhodně pět čísel z množiny $\{1, 2, 3, \dots, 20\}$. Jaká je pravděpodobnost, že

(i) právě dvě čísla z pěti napsaných na tabuli jsou větší než 12, pokud Bětka napsala čísla na tabuli tak, že se mohla opakovat[7]?
(ii) právě dvě čísla z pěti napsaných na tabuli jsou větší než 12, pokud Bětka napsala čísla na tabuli tak, že se opakovat nemohla[8]?

[7]Takže např. pětice čísel 2, 3, 3, 7, 15 se na tabuli mohla objevit.
[8]Takže např. pětice čísel 2, 3, 3, 7, 15 se na tabuli objevit nemohla.

(iii) největší z čísel napsaných na tabuli je alespoň 16, pokud Bětka napsala čísla na tabuli tak, že se mohla opakovat?
(iv) nejmenší z čísel napsaných na tabuli je nejvýše 3, pokud Bětka napsala čísla na tabuli tak, že se opakovat nemohla?
(v) pět čísel, která Bětka napsala na tabuli a postupně je označila x, y, z, u, v, jsou řešením rovnice
$$x + y + z + u + v = 20,$$
pokud se čísla v takové pětici mohla opakovat[9]?

[9] Takže např. pětice čísel $x = 2$, $y = 3$, $z = 3$, $u = 5$, $v = 7$ jsou řešením rovnice.

5 Spojitá náhodná veličina a spojitá rozdělení pravděpodobnosti

5.1 Řešené úlohy

Úloha 5.1.1 Lze považovat funkci $F_X(x) = \sin x$ za distribuční funkci v intervalu

(i) $0 < x < \pi$?
(ii) $0 < x < \frac{\pi}{2}$?

Řešení. Obecně pro $\forall x \in \mathbb{R}$ je $F_X(x) = P(X \leq x)$, tedy nutně

$$0 \leq F_X(x) \leq 1.$$

Zároveň nahlédněme, že je-li $x_1 < x_2$, pak

$$\begin{aligned} F_X(x_2) &= P(X \leq x_2) = \\ &= P(X \leq x_1 \vee X \in \langle x_1, x_2 \rangle) = \\ &= P(X \leq x_1) + P(X \in \langle x_1, x_2 \rangle) = \\ &= F(x_1) + P(X \in \langle x_1, x_2 \rangle). \end{aligned}$$

To přepišme jako $F_X(x_2) - F_X(x_2) = P(X \in \langle x_1, x_2 \rangle) \geq 0$, tedy nutně $F_X(x_2) - F_X(x_2) \geq 0$ pro $\forall x_1, x_2 \in \mathbb{R} : x_1 < x_2$. Distribuční funkce libovolné náhodné veličiny X je tedy neklesající a svými hodnotami nikdy mimo interval $\langle 0, 1 \rangle$.

(i) V intervalu $(0, \pi)$ není $F_X(x) = \sin x$ distribuční funkcí, neboť na podintervalu $\left\langle 0, \frac{\pi}{2} \right)$ je klesající.
(ii) V intervalu $\left(0, \frac{\pi}{2}\right)$ může být $F_X(x) = \sin x$ distribuční funkcí, neboť je neklesající a není svými hodnotami mimo interval $\langle 0, 1 \rangle$. Současně snadno nahlédneme, že pravděpodobnostní hustotou je pak

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x} = \frac{\mathrm{d}\sin x}{\mathrm{d}x} = \cos x.$$

□

Úloha 5.1.2 Spojitá náhodná veličina X je definována v mezích od 0 do 4 s hustotou pravděpodobnosti $f_X(x) = \frac{1}{2} - ax$. Určeme

(i) konstantu a.
(ii) distribuční funkci náhodné veličiny X.
(iii) pravděpodobnost, že náhodná veličina X nabude hodnoty od 1 do 2.

Řešení. Zřejmě je pro náhodnou veličinu X dle zadání $P(0 \leq X \leq 4) = 1$. Proto můžeme psát $P(0 \leq X \leq 4) = F_X(4) - F_X(0) = 1$ (†). Obecně platí $F_X(x) = \int_{-\infty}^{x} f_X(\tau)\mathrm{d}\tau$.

(i) Dosaďme předchozí vztah do rovnosti (†). Získáme

$$\begin{aligned} F_X(4) - F_X(0) &= \int_{-\infty}^{4} f_X(\tau)\mathrm{d}\tau - \int_{-\infty}^{0} f_X(\tau)\mathrm{d}\tau \\ &= \int_{0}^{4} f_X(\tau)\mathrm{d}\tau \\ &= \int_{0}^{4} \left(\frac{1}{2} - a\tau\right) \mathrm{d}\tau \\ &= \left[\frac{\tau}{2} - \frac{a\tau^2}{2}\right]_0^4 \\ &= \left(\frac{4}{2} - \frac{a \cdot 4^2}{2}\right) - \left(\frac{0}{2} - \frac{a \cdot 0^2}{2}\right) \\ &= 2 - 8a \\ &\overset{(\dagger)}{=} 1. \end{aligned}$$

Je tedy $2 - 8a = 1$. Teď již snadno nahlédneme, že $a = \frac{1}{8}$.

(ii) Už víme, že je $F_X(x) = \int_{-\infty}^{x} f_X(\tau)\mathrm{d}\tau$. Dosaďme do předchozího vztahu za $f_X(\tau) = \left(\frac{1}{2} - a\tau\right)$ pro $a = \frac{1}{8}$. Dostaneme

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(\tau)\mathrm{d}\tau \\ &= \int_{-\infty}^{x} \left(\frac{1}{2} - a\tau\right) \mathrm{d}\tau \Big|_{\tau \geq 0 \ \wedge \ a = \frac{1}{8}} \\ &= \int_{0}^{x} \left(\frac{1}{2} - \frac{\tau}{8}\right) \mathrm{d}\tau \\ &= \left[\frac{\tau}{2} - \frac{\tau^2}{16}\right]_0^x \\ &= \left(\frac{x}{2} - \frac{x^2}{16}\right) - \left(\frac{0}{2} - \frac{0^2}{16}\right) \\ &= \frac{8x - x^2}{16}, \qquad 0 \leq x \leq 4. \end{aligned}$$

(iii) Je $P(1 \leq X \leq 2) = F_X(2) - F_X(1)$, po dosazení z (ii) dostaneme

$$\begin{aligned} P(1 \leq X \leq 2) &= F_X(2) - F_X(1) = \\ &= \frac{8 \cdot 2 - 2^2}{16} - \frac{8 \cdot 1 - 1^2}{16} = \\ &= \frac{3}{4} - \frac{7}{16} = \\ &= \frac{5}{16}. \end{aligned}$$

□

Úloha 5.1.3 Distribuční funkce spojité náhodné veličiny X má tvar

$$F_X(x) = 1 - \frac{1}{x^3}, \qquad x > 1.$$

Určeme střední hodnotu náhodné veličiny X a její medián.

Řešení. Obecně platí $f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}$ a také $\mathbb{E}(X) = \int_{\forall x \in X} x f_X(x)\mathrm{d}x$. Nejdříve najděme pravděpodobnostní hustotu $f_X(x)$, zřejmě je

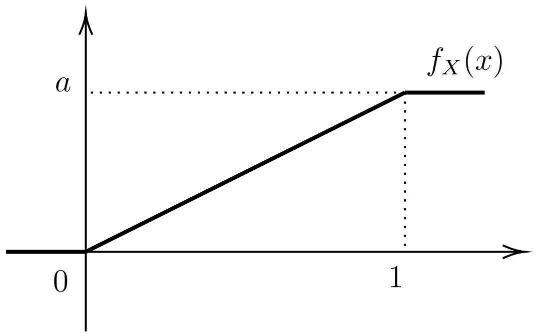
$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x} = \frac{\mathrm{d}\left(1 - \frac{1}{x^3}\right)}{\mathrm{d}x} = \frac{3}{x^4}, \qquad x > 1.$$

Nyní dopočítejme střední hodnotu. Je

$$\begin{aligned} \mathbb{E}(X) &= \int_{\forall x \in X} x f_X(x)\mathrm{d}x = \\ &= \int_1^\infty x \cdot \frac{3}{x^4} \cdot \mathrm{d}x = \\ &= \int_1^\infty \frac{3}{x^3} \cdot \mathrm{d}x = \\ &= \left[-\frac{3}{2x^2}\right]_1^\infty = \\ &= 0 - (-\frac{3}{2}) = \\ &= \frac{3}{2}. \end{aligned}$$

Pro medián $\tilde{x}$ platí, že pravděpodobnost toho, že náhodná veličina X nabude hodnoty nanejvýš rovné mediánu, je $\frac{1}{2}$. Je tedy $P(X \leq \tilde{x}) = F_X(\tilde{x}) = \frac{1}{2}$. Získáváme rovnost $1 - \frac{1}{\tilde{x}^3} = \frac{1}{2}$, jejímž řešením je medián $\tilde{x} = \sqrt[3]{2}$. □

Úloha 5.1.4 Hustota pravděpodobnosti $f_X(x)$ náhodné veličiny X je naznačena na obrázku 5.1.1.



Obrázek 5.1.1: Hustota pravděpodobnosti $f_X(x)$ náhodné veličiny X.

Určeme

(i) konstantu a a rovnici hustoty pravděpodobnosti $f_X(x)$ náhodné veličiny X.
(ii) distribuční funkci náhodné veličiny X.
(iii) pravděpodobnost, že náhodná veličina X nabude hodnoty z intervalu $\left(\frac{1}{3}, \frac{2}{3}\right)$.

Řešení. Z obrázku 5.1.1 nahlédneme, že je

$$f_X(x) = \begin{cases} ax, & 0 \le x \le 1 \\ 0, & \text{jinak.} \end{cases}$$

(i) Víme, že musí být $P(X = \{\forall x : x \in \mathcal{D}_{f_X}\}) = 1 = \int_{\forall x \in \mathcal{D}_{f_X}} f_X(x)\mathrm{d}x$. Z toho vyjděme při vyčíslení konstanty a; je tedy

$$\begin{aligned} P(X = \langle 0, 1\rangle) = 1 &= \\ &= \int_0^1 f_X(x)\mathrm{d}x = \\ &= \int_0^1 ax\mathrm{d}x = \left[\frac{ax^2}{2}\right]_0^1 = \\ &= \frac{a \cdot 1^2}{2} - \frac{a \cdot 0^2}{2} = \\ &= \frac{1}{2}a. \end{aligned}$$

Takže $1 = \frac{1}{2}a$, tedy $a = 2$. Teď už jen přepišme vztah pro pravděpodobnostní hustotu $f_X(x)$

$$f_X(x) = \begin{cases} 2x, & 0 \le x \le 1 \\ 0, & \text{jinak.} \end{cases}$$

(ii) Obecně je $F_X(x) = \int_{-\infty}^{x} f_X(\tau)\mathrm{d}\tau\Big|_{\tau \in \mathcal{D}_{f_X}}$. Dosaďme za $f_X(\tau) = 2x$ a řešme

$$F_X(x) = \int_{-\infty}^{x} f_X(\tau)\mathrm{d}\tau\Big|_{\tau \in \mathcal{D}_{f_X}} =$$

$$\begin{aligned} F_X(x) &= \int_0^x 2\tau \mathrm{d}\tau = \\ &= \left[\tau^2\right]_0^x = \\ &= x^2 - 0^2 = \\ &= x^2, \end{aligned}$$

kde $0 \leq x \leq 1$. Pro distribuční funkci tedy platí

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \leq x \leq 1 \\ 1, & x > 1. \end{cases}$$

(iii) Je $P\left(\frac{1}{3} \leq X \leq \frac{2}{3}\right) = F_X\left(\frac{2}{3}\right) - F_X\left(\frac{1}{3}\right)$, po dosazení z (ii) dostaneme

$$P\left(\frac{1}{3} \leq X \leq \frac{2}{3}\right) = F_X\left(\frac{2}{3}\right) - F_X\left(\frac{1}{3}\right) = \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2 = \frac{1}{3}.$$

□

Úloha 5.1.5 Čemu musí být rovny konstanty c, aby nálsedující funkce byly hustotami pravděpodobnosti náhodné veličiny X?

(i) $f_X(x) = c \sin x$, kde $0 < x < \pi$.
(ii) $f_X(x) = cx^2$, kde $0 < x < 2$.

Řešení. Víme, že musí být $P(X = \{\forall x : x \in \mathcal{D}_{f_X}\}) = 1 = \int_{\forall x \in \mathcal{D}_{f_X}} f_X(x)\mathrm{d}x$. Z toho vyjděme při vyčíslování konstant c.

(i) Je

$$\begin{aligned} P(X = \{\forall x : x \in \mathcal{D}_{f_X}\}) &= 1 \\ &= \int_{\forall x \in \mathcal{D}_{f_X}} f_X(x)\mathrm{d}x = \\ &= \int_0^\pi c \cdot \sin x \cdot \mathrm{d}x = \\ &= [-c \cdot \cos x]_0^\pi = \\ &= c - (-c) = \\ &= 2c. \end{aligned}$$

Takže $1 = 2c$, tudíž $c = \frac{1}{2}$.

(ii) Je

$$
\begin{aligned}
P(X = \{\forall x : x \in \mathcal{D}_{f_X}\}) = 1 &= \\
&= \int_{\forall x \in \mathcal{D}_{f_X}} f_X(x)\mathrm{d}x = \\
&= \int_0^2 cx^2 \mathrm{d}x = \\
&= \left[\frac{cx^3}{3}\right]_0^2 = \\
&= \frac{8}{3}c - 0 = \\
&= \frac{8}{3}c.
\end{aligned}
$$

Takže $1 = \frac{8}{3}c$, tudíž $c = \frac{3}{8}$. □

Úloha 5.1.6 Náhodná veličina X je definována v intervalu od -5 do 2 a její distribuční funkce $F_X(x)$ je na tomto intervalu

$$F_X(x) = \frac{x+2}{7}.$$

Určeme

(i) hustotu pravděpodobnosti $f_X(x)$.
(ii) $P(-2 < X < 2)$.
(iii) $P(X = 2)$.
(iv) $P(-6 < X < 1)$.

Řešení.

(i) Platí $f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}$, tedy $f_X(x) = \frac{\mathrm{d}\left(\frac{x+5}{7}\right)}{\mathrm{d}x} = \frac{1}{7}$ pro všechna $-5 \leq x \leq 2$. Pišme tedy pro pravděpodobnostní hustotu $f_X(x)$

$$f_X(x) = \begin{cases} \frac{1}{7}, & -5 \leq x \leq 2 \\ 0, & \text{jinak.} \end{cases}$$

□

(ii) Je $P(-2 < X < 2) = F_X(2) - F_X(-2)$, po dosazení ze zadání dostaneme

$$
\begin{aligned}
P(-2 < X < 2) = F_X(2) - F_X(-2) &= \\
&= \frac{2+5}{7} - \frac{-2+5}{7} = \\
&= 1 - \frac{3}{7} = \\
&= \frac{4}{7}.
\end{aligned}
$$

(iii) Je $P(X = 2) = P(2 \leq X \leq 2) = F_X(2) - F_X(2)$, po dosazení ze zadání dostaneme

$$\begin{aligned} P(2 \leq X \leq 2) &= F_X(2) - F_X(2) = \\ &= \frac{2+5}{7} - \frac{2+5}{7} = \\ &= 1 - 1 = \\ &= 0. \end{aligned}$$

To jsme mohli nahlédnout i apriorně, neboť pravděpodobnost, že spojitá veličina nabude právě jedné bodové hodnoty na spojitém intervalu hodnot, je již z geometrické definice pravděpodobnosti nulová.

(iv) Je $P(-6 < X < 1) = P(-6 < X < -5) + P(-5 \leq X < 1) = 0 + (F_X(1) - F_X(-5))$, po dosazení ze zadání dostaneme

$$\begin{aligned} P(-6 < X < 1) &= P(-6 < X < -5) + P(-5 \leq X < 1) = \\ &= 0 + (F_X(1) - F_X(-5)) = \\ &= \frac{1+5}{7} - \frac{-5+5}{7} = \\ &= \frac{6}{7} - 0 = \\ &= \frac{6}{7}. \end{aligned}$$

□

Úloha 5.1.7 Vypočítejme střední hodnotu, rozptyl a modus náhodné veličiny X, jejíž hustota pravděpodobnosti $f_X(x)$ je

$$f_X(x) = 12x^2(1-x), \qquad \text{kde } 0 < x < 1.$$

Řešení. Ze zadání víme, že pro pravděpodobnostní hustotu $f_X(x)$ platí

$$f_X(x) = \begin{cases} 12x^2(1-x), & 0 \leq x \leq 1 \\ 0, & \text{jinak.} \end{cases}$$

Víme, že obecně platí $\mathbb{E}(X) = \int_{\forall x \in X} x f_X(x) \mathrm{d}x$. Dosaďme do předchozího vztahu pravděpodobnostní hustotu $f_X(x)$ ze zadání, získáváme

$$\begin{aligned} \mathbb{E}(X) &= \int_{\forall x \in X} x f_X(x) \mathrm{d}x = \\ &= \int_0^1 x \cdot 12x^2(1-x) \cdot \mathrm{d}x = \end{aligned}$$

$$
\begin{aligned}
\mathbb{E}(X) &= \int_0^1 (12x^3 - 12x^4)\mathrm{d}x = \\
&= \left[3x^4 - \frac{12}{5}x^5\right]_0^1 = \\
&= (3 \cdot 1^4 - \frac{12}{5} \cdot 1^5) - 0 = \\
&= \frac{3}{5}.
\end{aligned}
$$

Je tedy $\mathbb{E}(X) = \frac{3}{5}$.

Pro výpočet rozptylu použijeme tzv. výpočtový tvar, tedy $\mathrm{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$. Zřejmě musíme nejdříve spočítat $\mathbb{E}(X^2)$. To provedeme podle obecného vztahu pro $\mathbb{E}(X^2)$, kdy je $\mathbb{E}(X^2) = \int_{\forall x \in \mathcal{D}_{f_X}} x^2 f_X(x)\mathrm{d}x$.

Dosaďme tedy

$$
\begin{aligned}
\mathbb{E}(X^2) &= \int_0^1 x^2 \cdot 12x^2(1-x) \cdot \mathrm{d}x = \\
&= \int_0^1 (12x^4 - 12x^5)\mathrm{d}x = \\
&= \left[\frac{12}{5}x^5 - 2x^6\right]_0^1 = \\
&= \left(\frac{12}{5} \cdot 1^5 - 2 \cdot 1^6\right) - 0 = \\
&= \frac{2}{5}.
\end{aligned}
$$

Teď již vyčísleme rozptyl

$$
\begin{aligned}
\mathrm{var}(X) &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \\
&= \frac{2}{5} - \left(\frac{3}{5}\right)^2 = \\
&= \frac{2}{5} - \frac{9}{25} = \\
&= \frac{1}{25}.
\end{aligned}
$$

Je tedy $\mathrm{var}(X) = \frac{1}{25}$.

Modus x^* je nejvíce pravděpodobnou hodnotou náhodné veličiny X. Musí tedy být

$$\begin{aligned}
x^* &= \arg\max_{0\leq x\leq 1}\left\{\lim_{\Delta\to 0} P\left(x-\frac{\Delta}{2}\leq X\leq x+\frac{\Delta}{2}\right)\right\} = \\
&= \arg\max_{0\leq x\leq 1}\left\{\lim_{\Delta\to 0}\left(F_X\left(x+\frac{\Delta}{2}\right)-F_X\left(x-\frac{\Delta}{2}\right)\right)\right\} = \\
&= \arg\max_{0\leq x\leq 1}\left\{\lim_{\Delta\to 0}\left(F_X\left(x+\Delta\right)-F_X\left(x\right)\right)\right\} = \\
&= \arg\max_{0\leq x\leq 1}\left\{\lim_{\Delta\to 0}\left(\frac{F_X(x+\Delta)-F_X(x)}{\Delta}\right)\right\} = \\
&= \arg\max_{0\leq x\leq 1}\left\{\frac{\mathrm{d}F_X(x)}{\mathrm{d}x}\right\} = \\
&= \arg\max_{0\leq x\leq 1}\left\{f_X(x)\right\}
\end{aligned}$$

Využijme spojitosti $f_X(x)$ na intervalu $0 \leq x \leq 1$, najděme první derivaci $f_X(x)$ podle x a položme ji rovnou nule a hledejme extrém, pak ověřme, že jde o maximum.

Tedy $\frac{\mathrm{d}f_X(x)}{\mathrm{d}x} = \frac{\mathrm{d}(12x^2(1-x))}{\mathrm{d}x} = \frac{\mathrm{d}(12x^2-12x^3)}{\mathrm{d}x} = 24x - 36x^2 \equiv 0$, odtud je $x^* \equiv x \in \left\{0, \frac{2}{3}\right\}$. Pro druhou derivaci platí $\frac{\mathrm{d}^2 f_X(x)}{\mathrm{d}x^2} = \frac{\mathrm{d}(24x-36x^2)}{\mathrm{d}x} = 24-72x$, tedy pouze pro $x = \frac{2}{3}$ je $24 - 72x < 0$ a jde v tomto bodě skutečně o maximum. Modem x^* náhodné veličiny X je tedy hodnota $x^* = \frac{2}{3}$. □

Úloha 5.1.8 Náhodná veličina X sleduje normální rozdělení $\mathcal{N}(20, 4^2)$. Jaká je pravděpodobnost, že nabude hodnoty

(i) menší než 16?
(ii) větší než 20?
(iii) v rozmezí mezi 12 a 28?
(iv) menší než 12 nebo větší než 28?

Řešení. Podle zadání je tedy zřejmě $X \sim \mathcal{N}(20, 4^2)$, tedy i $\mathbb{E}(X) = 20$ a $\mathrm{var}(X) = 4^2$, takže $\sqrt{\mathrm{var}(X)} = \sqrt{4^2} = 4$. Zřejmě tedy platí i $U \equiv \frac{X-\mathbb{E}(X)}{\sqrt{\mathrm{var}(X)}} = \frac{X-20}{4} \sim \mathcal{N}(0, 1^2)$, kde U je náhodná veličina sledující standardní normální rozdělení. Pak je tedy i $P(U \leq u) = F_U(u) = \Phi(u)$.

(i) Hledáme hodnotu výrazu $P(X < 16)$. Postupně upravujme nerovnost uvnitř argumentu výrazu $P(X < 16)$ i s využitím faktu, že $\sqrt{\mathrm{var}(X)} = 4 > 0$ a znaménko nerovnosti se nemění, tedy

$$\begin{aligned}
P(X < 16) &= P\left(X - \mathbb{E}(X) < 16 - \mathbb{E}(X)\right) = \\
&= P\left(\frac{X-\mathbb{E}(X)}{\sqrt{\mathrm{var}(X)}} < \frac{16-\mathbb{E}(X)}{\sqrt{\mathrm{var}(X)}}\right) =
\end{aligned}$$

$$
\begin{aligned}
P(X < 16) = P\left(\underbrace{\frac{X-20}{4}}_{=U} < \frac{16-20}{4}\right) &= \\
= P\left(U < \frac{-4}{4}\right) &= \\
= P\left(U < -1\right) &= \\
= F_U(-1) &= \\
= \Phi(-1) &= \\
= 1 - \Phi(1) &\doteq \\
\doteq 1 - 0{,}841 &\doteq \\
\doteq 0{,}159.
\end{aligned}
$$

Náhodná veličina $X \sim \mathcal{N}(20, 4^2)$ nabude hodnoty menší než 16 s pravděpodobností asi 0,159.

(ii) Hledáme hodnotu výrazu $P(X > 20)$, kterou můžeme také přepsat jako $P(X > 20) = 1 - P(X \leq 20)$. Postupně upravujme nerovnost uvnitř argumentu výrazu $P(X > 20)$ i s využitím faktu, že $\sqrt{\mathrm{var}(X)} = 4 > 0$ a znaménko nerovnosti se nemění, tedy

$$
\begin{aligned}
P(X > 20) &= 1 - P(X \leq 20) = \\
&= 1 - P\left(X - \mathbb{E}(X) \leq 20 - \mathbb{E}(X)\right) = \\
&= 1 - P\left(\frac{X - \mathbb{E}(X)}{\sqrt{\mathrm{var}(X)}} \leq \frac{20 - \mathbb{E}(X)}{\sqrt{\mathrm{var}(X)}}\right) = \\
&= 1 - P\left(\frac{X-20}{4} \leq \frac{20-20}{4}\right) = \\
&= 1 - P\left(U \leq \frac{0}{4}\right) = \\
&= 1 - P\left(U \leq 0\right) = \\
&= 1 - F_U(0) = \\
&= 1 - \Phi(0) = \\
&= 1 - 0{,}500 = \\
&= 0{,}500.
\end{aligned}
$$

Náhodná veličina $X \sim \mathcal{N}(20, 4^2)$ nabude hodnoty větší než 20 s pravděpodobností 0,5.

(iii) Hledáme hodnotu výrazu $P(12 < X < 28)$, kterou můžeme také přepsat jako $P(12 < X < 28) = P(X < 28) - P(X < 12)$. Postupně upravujme nerovnost uvnitř argumentu výrazu $P(12 < X < 28)$ i s využitím faktu, že $\sqrt{\mathrm{var}(X)} = 4 > 0$ a znaménko nerovnosti se nemění, tedy

$$
\begin{aligned}
P(12 < X < 28) &= P(X < 28) - P(X < 12) = \\
&= P\left(X - \mathbb{E}(X) < 28 - \mathbb{E}(X)\right) - \\
&\qquad - P\left(X - \mathbb{E}(X) < 12 - \mathbb{E}(X)\right) = \\
&= P\left(\frac{X - \mathbb{E}(X)}{\sqrt{\operatorname{var}(X)}} < \frac{28 - \mathbb{E}(X)}{\sqrt{\operatorname{var}(X)}}\right) - \\
&\qquad - P\left(\frac{X - \mathbb{E}(X)}{\sqrt{\operatorname{var}(X)}} < \frac{12 - \mathbb{E}(X)}{\sqrt{\operatorname{var}(X)}}\right) = \\
&= P\left(\frac{X - 20}{4} < \frac{28 - 20}{4}\right) - P\left(\frac{X - 20}{4} < \frac{12 - 20}{4}\right) = \\
&= P\left(U < \frac{8}{4}\right) - P\left(U < \frac{-8}{4}\right) = \\
&= P\left(U < 2\right) - P\left(U < -2\right) = \\
&= F_U(2) - F_U(-2) = \\
&= \Phi(2) - \Phi(-2) = \\
&= \Phi(2) - (1 - \Phi(2)) = \\
&= 2\Phi(2) - 1 \doteq \\
&\doteq 2 \cdot 0{,}977 - 1 \doteq \\
&\doteq 0{,}954.
\end{aligned}
$$

Náhodná veličina $X \sim \mathcal{N}(20, 4^2)$ nabude hodnoty v rozmezí mezi 12 a 28 s pravděpodobností asi 0,954.

(iv) Hledáme hodnotu výrazu $P(X < 12 \ \vee X > 28)$, což lze též přepsat jako $1 - P(12 < X < 28)$. Protože však z předchozího zadání víme, že $P(12 < X < 28) \doteq 0{,}954$, můžeme ihned psát, že

$$
\begin{aligned}
P(X < 12 \ \vee X > 28) &= 1 - P(12 < X < 28) \doteq \\
&\doteq 1 - 0{,}954 \doteq \\
&\doteq 0{,}046.
\end{aligned}
$$

Náhodná veličina $X \sim \mathcal{N}(20, 4^2)$ nabude hodnoty menší než 12 nebo větší než 28 s pravděpodobností asi 0,046. □

Úloha 5.1.9 Hmotnost bochníku chleba z jedné konkrétní pekárny sleduje normální rozdělení o střední hodnotě 0,8 kg a směrodatné odchylce 0,1 kg. Určete pravděpodobnost, že náhodně zakoupený bochník chleba z dané pekárny

(i) bude mít hmotnost menší než 0,6 kg.
(ii) bude mít hmotnost větší než 0,9 kg.
(iii) bude mít hmotnost v rozmezí mezi 0,5 kg až 1,1 kg.

Řešení. Označme symbolem X hmotnost náhodně zakoupeného bochníku z dané pekárny. Podle zadání je tedy zřejmě $X \sim \mathcal{N}(0{,}8,\ 0{,}1^2)$. Takže je i $\mathbb{E}(X) = 0{,}8$ a $\mathrm{var}(X) = 0{,}1^2$, tedy $\sqrt{\mathrm{var}(X)} = \sqrt{0{,}1^2} = 0{,}1$. Zřejmě tedy platí i $U \equiv \frac{X-\mathbb{E}(X)}{\sqrt{\mathrm{var}(X)}} = \frac{X-0{,}8}{0{,}1} \sim \mathcal{N}(0, 1^2)$, kde U je náhodná veličina sledující standardní normální rozdělení. Pak tedy je rovněž $P(U \leq u) = F_U(u) = \Phi(u)$.

(i) Hledáme hodnotu výrazu $P(X < 0{,}6)$. Postupně upravujme

$$
\begin{aligned}
P(X < 0{,}6) &= P\left(X - \mathbb{E}(X) < 0{,}6 - \mathbb{E}(X)\right) = \\
&= P\left(\frac{X - \mathbb{E}(X)}{\sqrt{\mathrm{var}(X)}} < \frac{0{,}6 - \mathbb{E}(X)}{\sqrt{\mathrm{var}(X)}}\right) = \\
&= P\left(\frac{X - 0{,}8}{0{,}1} < \frac{0{,}6 - 0{,}8}{0{,}1}\right) = \\
&= P\left(U < \frac{-0{,}2}{0{,}1}\right) = \\
&= P\left(U < -2\right) = \\
&= F_U(-2) = \\
&= \Phi(-2) = \\
&= 1 - \Phi(2) \doteq \\
&\doteq 1 - 0{,}977 \doteq \\
&\doteq 0{,}023.
\end{aligned}
$$

Hmotnost náhodného bochníku chleba[10] z dané pekárny bude menší než 0,6 kg s pravděpodobnostní přibližně 0,023.

(ii) Hledáme hodnotu výrazu $P(X > 0{,}9)$, kterou můžeme také přepsat jako $P(X > 0{,}9) = 1 - P(X \leq 0{,}9)$. Postupně upravujme

$$
\begin{aligned}
P(X > 0{,}9) &= 1 - P(X \leq 0{,}9) = \\
&= 1 - P\left(X - \mathbb{E}(X) \leq 0{,}9 - \mathbb{E}(X)\right) = \\
&= 1 - P\left(\frac{X - \mathbb{E}(X)}{\sqrt{\mathrm{var}(X)}} \leq \frac{0{,}9 - \mathbb{E}(X)}{\sqrt{\mathrm{var}(X)}}\right) = \\
&= 1 - P\left(\frac{X - 0{,}8}{0{,}1} \leq \frac{0{,}9 - 0{,}8}{0{,}1}\right) = \\
&= 1 - P\left(U \leq \frac{0{,}1}{0{,}1}\right) =
\end{aligned}
$$

[10] Intuitivně, hmotnost bochníku chleba se pohybuje „někde“ kolem 0,8 kg, proto situace, kdy je jeho hmotnost menší než 0,6 kg, je spíše nepravděpodobná a očekáváme tedy výslednou malou pravděpodobnost (0,023).

$$\begin{aligned}P(X > 0{,}9) &= 1 - P\left(U \leq 1\right) = \\ &= 1 - F_U(1) = \\ &= 1 - \Phi(1) \doteq \\ &\doteq 1 - 0{,}841 \doteq \\ &\doteq 0{,}159.\end{aligned}$$

Hmotnost náhodného bochníku chleba[11] z dané pekárny bude větší než 0,9 kg s pravděpodobností přibližně 0,159.

(iii) Hledáme hodnotu výrazu $P(0{,}5 < X < 1{,}1)$, kterou můžeme také přepsat jako $P(0{,}5 < X < 1{,}1) = P(X < 1{,}1) - P(X < 0{,}5)$. Postupně upravujme

$$\begin{aligned}P(0{,}5 < X < 1{,}1) &= P(X < 1{,}1) - P(X < 0{,}5) = \\ &= P\left(X - \mathbb{E}(X) < 1{,}1 - \mathbb{E}(X)\right) - \\ &\qquad - P\left(X - \mathbb{E}(X) < 0{,}5 - \mathbb{E}(X)\right) = \\ &= P\left(\frac{X - \mathbb{E}(X)}{\sqrt{\operatorname{var}(X)}} < \frac{1{,}1 - \mathbb{E}(X)}{\sqrt{\operatorname{var}(X)}}\right) - \\ &\qquad - P\left(\frac{X - \mathbb{E}(X)}{\sqrt{\operatorname{var}(X)}} < \frac{0{,}5 - \mathbb{E}(X)}{\sqrt{\operatorname{var}(X)}}\right) = \\ &= P\left(\frac{X - 0{,}8}{0{,}1} < \frac{1{,}1 - 0{,}8}{0{,}1}\right) - P\left(\frac{X - 0{,}8}{0{,}1} < \frac{0{,}5 - 0{,}8}{0{,}1}\right) = \\ &= P\left(U < \frac{0{,}3}{0{,}1}\right) - P\left(U < \frac{-0{,}3}{0{,}1}\right) = \\ &= P\left(U < 3\right) - P\left(U < -3\right) = \\ &= F_U(3) - F_U(-3) = \\ &= \Phi(3) - \Phi(-3) = \\ &= \Phi(3) - (1 - \Phi(3)) = \\ &= 2\Phi(3) - 1 \doteq \\ &\doteq 2 \cdot 0{,}998 - 1 \doteq \\ &\doteq 0{,}996.\end{aligned}$$

Hmotnost náhodného bochníku chleba[12] z dané pekárny bude v rozmezí 0,5 kg až 1,1 kg s pravděpodobností přibližně 0,996. □

[11] Opět intuitivně, hmotnost bochníku chleba se pohybuje „někde“ kolem 0,8 kg, proto situace, kdy je jeho hmotnost větší než 0,9 kg, není tolik pravděpodobná a očekáváme tedy výsledně spíše malou pravděpodobnost (0,159).

[12] Opět, hmotnost bochníku chleba se pohybuje „někde“ kolem 0,8 kg, proto situace, kdy je jeho hmotnost v rozmezí mezi 0,5 kg až 1,1 kg, je velmi pravděpodobná a očekáváme tedy výsledně velkou pravděpodobnost (0,996).

5.2 Úlohy k vlastnímu procvičení

Úloha 5.2.1 Značme distribuční funkci standardního normálního rozdělení jako $\Phi(\bullet)$. Pomocí úvahy, tabulek nebo vhodného softwaru

(i) najděte $\Phi(0)$.
(ii) najděte $x \in \mathbb{R}$ takové, aby $\Phi(x) = 0{,}5$.
(iii) najděte $x \in \mathbb{R}$ takové, aby $\Phi(x) = \Phi(-x)$.
(iv) najděte $\Phi(1{,}96)$.
(v) najděte $\Phi(-1{,}96)$.
(vi) najděte $x \in \mathbb{R}$ takové, aby $\Phi(x) = 0{,}025$.
(vii) najděte $x \in \mathbb{R}$ takové, aby $\Phi(x) = 1$.
(viii) najděte $x \in \mathbb{R}$ takové, aby $\Phi(x) \leq 2$.
(ix) najděte $x \in \mathbb{R}$ takové, aby $\Phi(x) = 0$.
(x) najděte $x \in \mathbb{R}$ takové, aby $\Phi(x) = +\infty$.
(xi) najděte $\Phi(+\infty)$.
(xii) najděte $\Phi(-\infty)$.
(xiii) najděte $x \in \mathbb{R}$ takové, aby $\Phi(x) - \Phi(-x) = 0{,}5$.
(xiv) najděte $x \in \mathbb{R}$ takové, aby $P(|U| \leq x) = 0{,}95$.

Úloha 5.2.2 Pravděpodobnost narození chlapce je 0,515. Jaká je pravděpodobnost, že mezi 10 000 narozenými novorozenci bude

(i) více děvčat než chlapců?
(ii) relativní četnost chlapců bude v mezích od 0,515 do 0,517?

Úloha 5.2.3 Zaměstnanec jezdí do zaměstnání pravidelně tam i zpět metrem. Doba čekání na příjezd soupravy metra se pohybuje rovnoměrně mezi 0 až 3 minutami. Jaká je pravděpodobnost, že celková doba čekání zaměstnance během 23 pracovních dní[13] bude kratší než 80 minut?

Úloha 5.2.4 Hmotnost jednoho mraženého kuřete sleduje normální rozdělení o střední hodnotě 1,20 kg a rozptylu 0,04 kg^2. Adam si jedno takové mražené kuře koupil. Jaká je pravděpodobnost, že

(i) zakoupené kuře má hmotnost menší než 0,80 kg?
(ii) zakoupené kuře má hmotnost větší než 1,40 kg?
(iii) ve čtveřici mražených kuřat, která Adam koupil během příštího nákupu, je nejvýše jedno kuře o hmotnosti menší než 0,80 kg?

[13] Zaměstnanec jede vždy dvakrát denně – do zaměstnání a zpět.

6 Odhady parametrů

6.1 Řešené úlohy

Úloha 6.1.1 V zásilce 100 kusů součástek byl vždy změřen dlouhý rozměr součástky. Průměrná hodnota dlouhého rozměru součástky byla 156 cm, směrodatná odchylka dlouhého rozměru součástky byla 2 cm. Určeme

(i) 95 % oboustranný interval spolehlivosti pro průměr dlouhého rozměru součástky.
(ii) 95 % levostranný interval spolehlivosti pro průměr dlouhého rozměru součástky.
(iii) 95 % pravostranný interval spolehlivosti pro průměr dlouhého rozměru součástky.

Řešení. Víme, že pro $n = 100$ součástek byla jejich průměrná délka $\bar{x} = 156$ cm a směrodatná odchylka $s_x = 2$ cm. Protože $n \gg 30$, lze ve vzorcích pro $100(1-p)$ % i $100(1-p/2)$ % intervalové odhady použít kvantily standardního normálního rozdělení, kde $p = 0{,}05$. Pak je tedy

(i) protože je $P\left(\bar{x} - u_{1-p/2} \cdot \sqrt{\frac{s_x^2}{n}} \le \mu \le \bar{x} + u_{1-p/2} \cdot \sqrt{\frac{s_x^2}{n}}\right) = 1-p$, dostáváme pro průměrnou délku dlouhého rozměru součástky konkrétní meze

$$\begin{aligned}\mu \in & \left\langle \bar{x} - u_{1-p/2} \cdot \sqrt{\frac{s_x^2}{n}}; \bar{x} + u_{1-p/2} \cdot \sqrt{\frac{s_x^2}{n}} \right\rangle = \\ &= \left\langle 156 - 1{,}96 \cdot \sqrt{\frac{2^2}{100}}; 156 + 1{,}96 \cdot \sqrt{\frac{2^2}{100}} \right\rangle = \\ &= \langle 155{,}60; 156{,}39\rangle \ \text{[cm]}.\end{aligned}$$

(ii) protože je $P\left(\bar{x} - u_{1-p} \cdot \sqrt{\frac{s_x^2}{n}} \le \mu\right) = 1-p$, dostáváme pro průměrnou délku dlouhého rozměru součástky konkrétní meze levostranného intervalu

$$\begin{aligned}\mu \in & \left\langle \bar{x} - u_{1-p} \cdot \sqrt{\frac{s_x^2}{n}}; \infty \right) = \\ &= \left\langle 156 - 1{,}64 \cdot \sqrt{\frac{2^2}{100}}; \infty \right) = \\ &= \langle 155{,}67; \infty) \ \text{[cm]}.\end{aligned}$$

(iii) protože je $P\left(\mu \le \bar{x} + u_{1-p}\cdot\sqrt{\frac{s_x^2}{n}}\right) = 1-p$, dostáváme pro průměrnou délku dlouhého rozměru součástky konkrétní meze pravostranného intervalu

$$\begin{aligned}\mu \in \left(-\infty; \bar{x} + u_{1-p}\cdot\sqrt{\frac{s_x^2}{n}}\right\rangle &= \\ = \left(-\infty; 156 + 1{,}64\cdot\sqrt{\frac{2^2}{100}}\right\rangle &= \\ = (-\infty; 156{,}33\rangle \quad [\mathrm{cm}].\end{aligned}$$

□

Úloha 6.1.2 V zásilce je 80 kusů zboží. Náhodně vybereme osm kusů a postupně je zvážíme. Získané hmotnosti (v kilogramech) jsou

$$8 \quad 8 \quad 9 \quad 7 \quad 6 \quad 8 \quad 9 \quad 9.$$

Určeme bodový odhad střední hodnoty hmotnosti jednoho kusu zboží a odhad rozptylu hodnoty hmotnosti jednoho kusu zboží. Měly-li dva z vybraných osmi kusů zboží vadu, jaký je bodový odhad populační četnosti zboží s vadou?

Řešení. Zřejmě je nejlepším bodovým odhadem střední hodnoty hmotnosti μ jednoho kusu zboží aritmetický výběrový průměr $\bar{x}$, tedy $\hat{\mu} = \bar{x}$, číselně pak $\hat{\mu} = \bar{x} = \frac{1}{8}\sum_{i=1}^{8} x_i = \frac{1}{8}(8+8+9+7+6+8+9+9) = 8$ kg.

Obdobně nejlepším bodovým odhadem rozptylu σ^2 hodnoty hmotnosti jednoho kusu zboží je výběrový rozptyl s_x^2, tedy $\hat{\sigma}^2 = s_x^2$, číselně pak $\hat{\sigma}^2 = s_x^2 = \sum_{i=1}^{8} \frac{(x_i-\bar{x})^2}{8-1} \doteq 1{,}14$ kg^2.

Jsou-li dva z osmi vybraných kusů zboží s vadou, je nejlepším bodovým odhadem populační četnosti π zboží s vadou výběrová četnost zboží s vadou, tedy $\hat{\pi} = \frac{2}{8} = \frac{1}{4} = 0{,}25$. □

Úloha 6.1.3 Změřili jsme tělesnou výšku šesti žen a získali jsme následující hodnoty (v centimetrech)

$$163 \quad 175 \quad 177 \quad 165 \quad 171 \quad 174.$$

Předpokládáme, že tělesná výška ženy sleduje normální rozdělení s (populačním) rozptylem 49 cm^2. Určeme

(i) bodový odhad populační střední hodnoty tělesné výšky ženy;

(ii) oboustranný 95% interval spolehlivosti pro populační střední hodnotu tělesné výšky ženy;

(iii) nutný počet žen, jejichž tělesnou výšku bychom museli změřit, abychom snížili přípustnou chybu odhadu[14] populační střední hodnoty tělesné výšky ženy na 1 cm.

Řešení.

(i) Nejlepším bodovým odhadem střední hodnoty tělesné výšky μ ženy je aritmetický výběrový průměr $\bar{x}$, tedy $\hat{\mu} = \bar{x}$, číselně pak $\hat{\mu} = \bar{x} = \frac{1}{6}\sum_{i=1}^{8} x_i = \frac{1}{6}(163 + 175 + 177 + 165 + 171 + 174) \doteq 170{,}83$ cm.

(ii) Protože je $\bar{x} \doteq 170{,}83$ cm, $s_x = \sqrt{49} = 7{,}00$ cm, $n = 6$ a $p = 0{,}05$, s použitím $P\left(\bar{x} - u_{1-p/2} \cdot \sqrt{\frac{s_x^2}{n}} \leq \mu \leq \bar{x} + u_{1-p/2} \cdot \sqrt{\frac{s_x^2}{n}}\right) = 1-p$ dostáváme pro populační střední hodnotu tělesné výšky ženy konkrétní meze

$$
\begin{aligned}
\mu \in & \left\langle \bar{x} - u_{1-p/2} \cdot \sqrt{\frac{s_x^2}{n}}; \bar{x} + u_{1-p/2} \cdot \sqrt{\frac{s_x^2}{n}} \right\rangle \doteq \\
\doteq & \left\langle 170{,}83 - 1{,}96 \cdot \sqrt{\frac{7^2}{6}}; 170{,}83 + 1{,}96 \cdot \sqrt{\frac{7^2}{6}} \right\rangle \doteq \\
\doteq & \langle 165{,}23; 176{,}43 \rangle
\end{aligned}
$$

(iii) Víme-li, že přípustná chyba (tedy polovina délky intervalu spolehlivosti pro střední hodnotu) odhadu populační střední hodnoty tělesné výšky ženy má být maximálně Δ, platí

$$u_{1-\alpha/2} \cdot \sqrt{\frac{s_x^2}{n}} \leq \Delta,$$

tedy

$$n \geq \left\lceil \frac{u_{1-\alpha/2}^2 \cdot s_x^2}{\Delta^2} \right\rceil.$$

Číselně pak pro $\Delta = 1{,}00$ cm, $s_x = 7{,}00$ cm a $p = 0{,}05$ dostáváme

$$n \geq \left\lceil \frac{1{,}96 \cdot 7{,}00^2}{1{,}00^2} \right\rceil \doteq \lceil 188{,}2 \rceil = 189.$$

Minimální nutný počet žen, které je třeba změřit tak, že předpokládáme stále stejnou hodnotu variability jejich tělesné výšky a chceme snížit přípustnou chybu odhadu populační střední hodnoty jejich tělesné výšky na 1 cm, je roven 189. □

Úloha 6.1.4 Byla sledována účinnost léku na snížení krevního tlaku. Ke snížení krevního tlaku došlo u 140 z 225 pacientů. Sestavme 95% interval spolehlivosti pro průměrnou účinnost léku.

[14] Jde o poloviční délku intervalu spolehlivosti pro střední hodnotu.

Řešení. Protože je $n = 225$, $p = \frac{140}{225} \doteq 0{,}622$ a $q = 0{,}05$, s použitím

$$P\left(p - u_{1-q/2} \cdot \sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p + u_{1-q/2} \cdot \sqrt{\frac{p(1-p)}{n}}\right) = 1 - q$$

dostáváme pro průměrnou účinnost léku π její 95% interval spolehlivosti

$$\begin{aligned}
\pi \in & \left\langle p - u_{1-q/2} \cdot \sqrt{\frac{p(1-p)}{n}};\ p + u_{1-q/2} \cdot \sqrt{\frac{p(1-p)}{n}}\right\rangle \doteq \\
\doteq & \left\langle 0{,}622 - 1{,}960 \cdot \sqrt{\frac{0{,}622 \cdot (1 - 0{,}622)}{225}};\right. \\
& \qquad\qquad \left. 0{,}622 + 1{,}960 \cdot \sqrt{\frac{0{,}622 \cdot (1 - 0{,}622)}{225}}\right\rangle \doteq \\
\doteq & \langle 0{,}56; 0{,}69\rangle.
\end{aligned}$$

□

6.2 Úlohy k vlastnímu procvičení

Úloha 6.2.1 U 100 náhodně vybraných výrobků činila průměrná spotřeba materiálu 150 gramů a výběrový rozptyl spotřeby byl 16 čtverečních gramů. Sestrojme 95% interval spolehlivosti pro očekávanou spotřebu na jeden výrobek.

Úloha 6.2.2 Ze 42 náhodně vybraných účastníků sportovního odpoledne bylo 16 dívek a 26 chlapců. Odhadněme podíl dívek mezi účastníky.

7 Testování hypotéz

7.1 Řešené úlohy

Úloha 7.1.1 Ředitel tvrdí, že průměrná mzda v jeho firmě je 25000 korun. Chceme ověřit jeho tvrzení, neboť se domníváme, že údaj o průměrné mzdě nadhodnotil. Máme údaje o 49 mzdách, ze který jsme spočítali průměr 23929 korun a výběrovou směrodatnou odchylku 5603 korun. Předpokládáme, že mzdy mají normální rozdělení. Hladinu významnosti volme 0,05.

Řešení. Označme jako μ střední hodnotu celkové průměrné mzdy v dané firmě. Dále víme, že $n = 49$, $\bar{x} = 23929$ Kč, $s_x = 5603$ Kč a $\alpha = 0{,}05$. Dle tvrzení ředitele je $\mu = 25000$ Kč, máme však podezření, že tento údaj je nadhodnocený, tedy že reálně je $\mu < 25000$ Kč. Testujme tedy nulovou hypotézu $H_0 : \mu = 25000$ proti alternativní hypotéze $H_1 : \mu < 25000$ pomocí testového kritéria $T = \frac{\bar{x}-\mu}{s_x}\sqrt{n}$.

Za platnosti nulové hypotézy je tedy $\mu = 25000$ a dále

$$T = \frac{\bar{x} - \mu}{s_x}\sqrt{n} = \frac{23929 - 25000}{5603}\sqrt{49} \doteq -1{,}338.$$

Protože je alternativní hypotéza jednostranná, je kritický obor pro zamítnutí nulové hypotézy $W_\alpha = \{t : t \leq -t_{1-\alpha}(n-1)\}$. Pro $n = 49$ a $\alpha = 0{,}05$ je

$$t_{1-\alpha}(n-1) = t_{0,95}(48) \doteq 1{,}677,$$

takže

$$W_\alpha = W_{0,05} = \{t : t \leq -t_{1-\alpha}(n-1)\} = \{t : t \leq -1{,}677\}.$$

Protože $T \doteq -1{,}338 \notin W_{0,05} = \{t : t \leq -1{,}677\}$, nezamítáme[15] nulovou hypotézu o tom, že průměrná mzda ve firmě je rovna 25000 Kč. □

Úloha 7.1.2 V malé nemocnici se za poslední měsíc narodilo 90 dívek a 60 chlapců. Na hladinách významnosti (i) 0,05 a (ii) 0,01 chceme otestovat hypotézu, zda se dívky a chlapci rodí se stejnou pravděpodobností.

Řešení. Označme symbolem P náhodnou veličinu vracející počet dívek, které se v dané nemocnici narodí za každý měsíc. Pro předchozí měsíc bylo zřejmě $P = \frac{90}{90+50} = \frac{3}{5} = 0{,}6$. Máme-li testovat, zda se dívky a chlapci rodí se stejnou pravděpodobností, testujeme tak vlastně nulovou hypotézu $H_0 : P = \pi_0 = 0{,}5$ o tom, že pravděpodobnost narození dívky je $P = 0{,}5$ a pravděpodobnost narození

[15] Vzhledem k $n > 30$ by bylo možné rovněž použít kvantily standardního normálního rozdělení, tj. namísto použitého kvantilu Studentova t-rozdělení $t_{1-\alpha}(n-1) = t_{0,95}(48) \doteq 1{,}677$ bylo možné počítat i s kvantilem standarního normálního rozdělení $u_{1-\alpha} = u_{0,95} \doteq 1{,}645$.

chlapce je $1 - P$, tedy konkrétně $1 - \pi_0 = 1 - 0{,}5 = 0{,}5$, a to proti alternativní hypotéze $H_1 : P \neq 0{,}5$.

Značme $n = 90 + 60 = 150$ celkový počet narozenýh dětí. Za platnosti nulové hypotézy sleduje náhodná veličina $U = \frac{P-\pi_0}{\sqrt{\pi_0(1-\pi_0)}}\sqrt{n}$, kterou použijeme jako testové kritérium, standardní normální rozložení. Zřejmě je

$$U = \frac{P - \pi_0}{\sqrt{\pi_0(1 - \pi_0)}}\sqrt{n} = \frac{0{,}6 - 0{,}5}{\sqrt{0{,}5 \cdot (1 - 0{,}5)}}\sqrt{150} \doteq 2{,}449.$$

(i) Pro $\alpha = 0{,}05$ je

$$W_\alpha = W_{0,05} = \{u : |u| \geq u_{1-\alpha/2}\} = \{u : |u| \geq 1{,}960\}.$$

Protože $U \doteq 2{,}449 \in W_{0,05} = \{u : |u| \geq 1{,}960\}$, zamítáme na hladině významnosti 0,05 nulovou hypotézu o tom, že dívky a chlapci se rodí se stejnou pravděpodobností.

(ii) Pro $\alpha = 0{,}01$ je

$$W_\alpha = W_{0,01} = \{u : |u| \geq u_{1-\alpha/2}\} = \{u : |u| \geq 2{,}576\}.$$

Protože $U \doteq 2{,}449 \notin W_{0,01} = \{u : |u| \geq 2{,}576\}$, nezamítáme na hladině významnosti 0,01 nulovou hypotézu o tom, že dívky a chlapci se rodí se stejnou pravděpodobností.

Snadno nahlédneme, že pro nižší hladinu významnosti je obecně obtížnější zamítnout nulovou hypotézu. U nižší hladiny významnosti totiž tolerujeme nižší pravděpodobnost nastání chyby prvního typu, tj. že bychom nesprávně zamítli nulovou hypotézu, když by ta reálně platila. Proto, aby byla tato pravděpodobnost nastání chyby prvního typu snížena, je obecně obtížnější zamítnout nulovou hypotézu – když totiž není nulová hypotéza zamítnuta (resp. je to obtížnější a možné jen s daty více odporujícími nulové hypotéze), nemůje obecně dojít k chybě prvního typu, ať už jde o nulovou hypotézu odpovídající, či neodpovídající realitě. To je dobře ilustrováno v tomto příkladu. □

Úloha 7.1.3 Inspekce v pivovaru chtěla ověřit, zda výrobní linka plní půllitrové lahve dostatečně přesně. Vybrala proto náhodných dvacet jedna lahví naplněných výrobní linkou a zjistila jednotlivé naplněné objemy; zjištěné hodnoty (v mililitrech) jsou v tabulce níže.

i	x_i	i	x_i	i	x_i
1	495,2	8	501,5	15	500,2
2	496,8	9	501,8	16	501,1
3	502,1	10	499,1	17	499,9
4	498,5	11	500,9	18	500,2
5	501,0	12	502,2	19	501,1
6	503,0	13	501,7	20	500,8
7	500,7	14	500,4	21	499,3

Mohla inspekce na hladině významnosti 0,05 učinit závěr, že průměrný objem piva plněný výrobní linkou je 500 mililitrů?

Řešení. Označme průměrný objem piva na láhev plněný výrovní linkou symbolem μ, dále výberový průměrný objem z dvaceti jedna pozorování symbolem $\bar{x}$ a výběrovou směrodatnou odchylku s. Ze zadání rovněž víme, že bylo zajištěno $n = 21$ měření objemů v lahvích. Inspekce testovala nulohou hypotézu $H_0 : \mu = 500$ [ml] proti alternativní hypotéze $H_1 : \mu \neq 500$ [ml] pomocí testovacího kritéria $T = \frac{\bar{x}-\mu}{s}\sqrt{n}$.

Z tabulky hodnot dopočítáme $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1+x_2+\cdots+x_{21}}{21} \doteq 500{,}357$ [ml] a $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_{21}-\bar{x})^2}{21-1}} \doteq 1{,}822$ [ml]. Pro testovací kritérium dostáváme

$$T = \frac{\bar{x}-\mu}{s}\sqrt{n} \doteq \frac{500{,}357 - 500{,}000}{1{,}822}\sqrt{21} \doteq 0{,}898.$$

Protože je $n = 21 < 30$, budeme při sestavování kritického oboru volit kvantily Studentova t-rozdělení (nikoliv kvantily standardního normálního rozdělení), a protože je alternativní hypotéza oboustranná, je kritický obor pro zamítnutí nulové hypotézy $W_\alpha = \{t : |t| \geq t_{1-\alpha/2}(n-1)\}$. Pro $n = 21$ a $\alpha = 0{,}05$ je

$$t_{1-\alpha/2}(n-1) = t_{1-0{,}05/2}(21-1) = t_{0{,}975}(20) \doteq 2{,}086,$$

takže

$$W_\alpha = W_{0{,}05} = \{t : |t| \geq t_{1-\alpha/2}(n-1)\} = \{t : |t| \geq 2{,}086\}.$$

Protože $T \doteq 0{,}898 \notin W_{0{,}05} = \{t : |t| \geq 2{,}086\}$, nezamítáme na hladině významnosti 0,05 nulovou hypotézu o tom, že průměrný objem piva na láhev plněný výrobní linkou je roven 500 mililitrů.

Řešení v MS Excel®. Příklad je přehledově vyřešen v záložce

`objemy_pivnich_lahvi`

přiloženého excelového souboru

`_sbirka_uloh_pro_ucely_bakalarskych_statistickych_predmetu_.xlsx`.

□

Úloha 7.1.4 Bylo vybráno 73 polí stejné kvality. Na 38 z nich se zkoušel nový způsob hnojení, zbývajících 35 bylo ošetřeno původním způsobem. Průměrné výnosy pšenice při novém způsobu hnojení byly 5,4 tun na hektar a výběrový rozptyl byl 0,27. Průměrné výnosy pšenice při původním způsobu hnojení byly 4,7 tun na hektar a výběrový rozptyl byl 0,24. Chceme zjistit, zda nový způsob hnojení má vliv na výnosy pšenice. Hladinu významnosti volme 0,05.

Řešení. Označme průmerný výnos pšenice při původním hnojení pomocí μ_1 a novém hnojení μ_2. Ze zadání víme, že průměrné výnosy byly pro původní typ hnojení $\bar{x}_1 = 4{,}7$ a nový typ hnojení $\bar{x}_2 = 5{,}4$, příslušné rozptyly byly $s_{x,1}^2 = 0{,}24$ a $s_{x,2}^2 = 0{,}27$, a to pro $n_1 = 35$ a $n_2 = 38$. Testujme nyní nulovou hypotézu

$H_0 : \mu_1 = \mu_1$ proti alternativní hypotéze $H_0 : \mu_1 \neq \mu_1$ pomocí testového kritéria $U = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$. Protože je obecně $\hat{\mu}_1 = \bar{x}_1$ a $\hat{\mu}_2 = \bar{x}_2$, zřejmě je

$$U = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_{x,1}^2}{n_1} + \frac{s_{x,2}^2}{n_2}}} = \frac{4{,}7 - 5{,}4}{\sqrt{\frac{0{,}24}{35} + \frac{0{,}27}{38}}} \doteq -5{,}924.$$

Pro $\alpha = 0{,}05$ je $W_\alpha = W_{0,05} = \{u : |u| \geq u_{1-\alpha/2}\} = \{u : |u| \geq 1{,}960\}$. Protože $|U| \doteq 5{,}924 \in W_{0,05} = \{u : |u| \geq 1{,}960\}$, zamítáme na hladině významnosti 0,05 nulovou hypotézu o tom, že by průměrné hektarové výnosy pšenice mohly být u obou typů hnojení srovnatelné. □

Úloha 7.1.5 Každou ze tří variant testu (A, B a C) napsali vždy čtyři náhodní studenti. Je třeba rozhodnout, zda jsou na základě výsledků v tabulce 7.1.1 varianty testu obdobně náročné, předpokládáme-li, že apriorní znalosti všech dvanácti studentů jsou srovnatelné.

		výsledek testu			
		1. student	2. student	3. student	4. student
	A	91	81	74	57
varianta	B	83	72	63	47
	C	71	69	58	40

Tabulka 7.1.1: Tabulka s výsledky z jednotlivých variant testů.

Řešení. Závisle proměnnou je zde skóre z daného testu (spojitá proměnná), které závisí na variantě testu (tedy jedna kategorická proměnná). To lze zkoumat jednorozměrnou analýzou rozptylu. Intuitivně, čím více se bude lišit variabilita skóre mezi variantami testu oproti variabilitě skóre uvnitř jednotlivých variant testu, tím spíše se mezi sebou varianty testu svou náročností navzájem liší.

Variabilitu skóre mezi variantami testu můžeme popsat meziskupinovým součtem čtverců $S_{y.m}$, zatímco variabilitu skóre uvnitř variant testu můžeme popsat vnitroskupinovým součtem čtverců $S_{y.v}$.

Označme všech dvanáct hodnot z tabulky postupně tak, jak naznačuje tabulka 7.1.2.

		výsledek testu			
		1. student	2. student	3. student	4. student
	A	$y_{1,1} = 91$	$y_{1,2} = 81$	$y_{1,3} = 74$	$y_{1,4} = 57$
varianta	B	$y_{2,1} = 83$	$y_{2,2} = 72$	$y_{2,3} = 63$	$y_{2,4} = 47$
	C	$y_{3,1} = 71$	$y_{3,2} = 69$	$y_{3,3} = 58$	$y_{3,4} = 40$

Tabulka 7.1.2: Tabulka s výsledky z jednotlivých variant testů.

Víme-li ze zadání, že $n = 12$, počet variant je $k = 3$ a $n_1 = n_2 = n_3 = 4$, můžeme spočítat celkový průměr skóre testu $\bar{y}$, tedy

$$\bar{y} = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} y_{j,i} = \frac{1}{12}(y_{1,1} + y_{1,2} + \cdots + y_{3,4}) \doteq 67{,}17.$$

Dále můžeme spočítat průměrná skóre pro každou variantu (varianta A ať je první, varianta B druhá, varianta C třetí), tedy

$$\bar{y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_{1,i} = \frac{1}{4}(y_{1,1} + y_{1,2} + \cdots + y_{1,4}) \doteq 75{,}75,$$
$$\bar{y}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} y_{2,i} = \frac{1}{4}(y_{2,1} + y_{2,2} + \cdots + y_{2,4}) \doteq 66{,}25,$$
$$\bar{y}_3 = \frac{1}{n_3}\sum_{i=1}^{n_3} y_{3,i} = \frac{1}{4}(y_{3,1} + y_{3,2} + \cdots + y_{3,4}) \doteq 59{,}50.$$

Nyní již můžeme spočítat meziskupinový součet čtverců $S_{y.m}$, tedy

$$\begin{aligned} S_{y.m} &= \sum_{j=1}^{k}(\bar{y}_j - \bar{y})^2 n_j \doteq \\ &\doteq \sum_{j=1}^{3}(\bar{y}_j - 67{,}17)^2 \cdot 4 \doteq \\ &\doteq 4 \cdot (75{,}75 - 67{,}17)^2 + 4 \cdot (66{,}25 - 67{,}17)^2 + 4 \cdot (59{,}50 - 67{,}17)^2 \doteq \\ &\doteq 533{,}17. \end{aligned}$$

Podobně můžeme spočítat vnitroskupinový součet čtverců $S_{y.v}$, tedy

$$\begin{aligned} S_{y.v} &= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{j,i} - \bar{y}_j)^2 \doteq \\ &\doteq \sum_{j=1}^{3}\sum_{i=1}^{4}(y_{j,i} - \bar{y}_j)^2 \doteq \\ &\doteq \sum_{i=1}^{4}(y_{1,i} - 75{,}75)^2 + \sum_{i=1}^{4}(y_{2,i} - 66{,}25)^2 + \sum_{i=1}^{4}(y_{3,i} - 59{,}50)^2 \doteq \\ &\doteq 1914{,}50. \end{aligned}$$

Nyní již můžeme dokončit analýzu rozptylu. Označíme-li průměrné skóre v první, druhé a třetí variantě testu postupně μ_1, μ_2 a μ_3, pak testujeme nulovou hypotézu $H_0 : \mu_1 = \mu_2 = \mu_3$ pomocí testové statistiky $F = \frac{S_{y.m}/(k-1)}{S_{y.v}/(n-k)}$, která za platnosti nulové hypotézy sleduje Fisherovo-Snedecorovo rozdělení F o stupních volnosti $k - 1$ a $n - k$.

Po dosazení dostáváme

$$F = \frac{S_{y.m} \ / \ (k-1)}{S_{y.v} \ / \ (n-k)} \doteq \frac{533{,}17 \ / \ (3-1)}{1914{,}50 \ / \ (12-3)} \doteq \frac{533{,}17 \ / \ 2}{1914{,}50 \ / \ 9} \doteq 1{,}253.$$

Kritický obor pro zamítnutí nulové hypotézy je $W_\alpha = \{F : F \geq F_{1-\alpha}(k - 1, n - k)\}$. Pro $n = 12$, $k = 3$ a $\alpha = 0{,}05$ je

$$F_{1-\alpha}(k-1, n-k) = F_{1-0,05}(3-1, 12-3) = F_{0,95}(2, 9) \doteq 4{,}257,$$

takže

$$W_\alpha = W_{0,05} = \{F : F \geq F_{1-\alpha}(k-1, n-k)\} \doteq \{F : F \geq 4{,}257\}.$$

Protože $F \doteq 1{,}253 \notin W_{0,05} \doteq \{F : F \geq 4{,}257\}$, nezamítáme na hladině významnosti 0,05 nulovou hypotézu o tom, že průměrná skóre se mezi variantami testu statisticky významně neliší.

Řešení v MS Excel®. Příklad je přehledově vyřešen v záložce

`vysledky_vs_varianty_testu`

přiloženého excelového souboru

`_sbirka_uloh_pro_ucely_bakalarskych_statistickych_predmetu_.xlsx`.

□

Úloha 7.1.6 Z průzkumu trhu byla sestavena kontingenční tabulka preference budoucího bydliště dle pohlaví. Výsledky jsou uvedeny níže.

		preference budoucího bydliště	
		město	venkov
pohlaví	muž	71	91
	žena	82	56

Závisí preference budoucího bydliště na pohlaví? Zkoumejme na hladině významnosti 0,05.

Řešení. Označme symbolem $n_{i,\bullet}$ součet i-tého řádku výše uvedené (kontingenční) tabulky a symbolem $n_{\bullet,j}$ součet j-tého sloupce výše uvedené (kontingenční) tabulky, kde $i \in \{1, 2\}$ a $j \in \{1, 2\}$. Celkový součet výše uvedené (kontingenční) tabulky značme n.

Pomocí χ^2 testu nezávislosti pro čtyřpolní tabulku 2×2 testujme nulovou hypotézu H_0 o tom, že řádková a sloupcová kvalitativní proměnná jsou nezávislé, tj. že preference budoucího bydliště nezávisí na pohlaví. Testovou statistikou je

$$\mathcal{X}^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\left(n_{i,j} - \frac{n_{i,\bullet} n_{\bullet,j}}{n}\right)^2}{\frac{n_{i,\bullet} n_{\bullet,j}}{n}}$$

a za platnosti nulové hypotézy sleduje[16] statistika $\mathcal{X}^2$ na hladině významnosti α rozdělení χ^2 o jednom stupni volnosti, tedy $\mathcal{X}^2 \sim \chi^2_{1-\alpha}(1)$.

[16] Členy $\frac{n_{i,\bullet} n_{\bullet,j}}{n}$ pro $i \in \{1, 2\}$ a $j \in \{1, 2\}$ by měly být rovny vždy alespoň 5, jinak je nutné buňky kontingenční tabulky slučovat (to lze doporučit pouze u větších tabulek než 2×2; u testu nezávislosti nad čtyřpolní tabulkou 2×2 to nedává příliš smysl), anebo nahradit χ^2 test nezávislosti Fisherovým přesným testem.

Z tabulky snadno nahlédneme, že

$$n_{1,\bullet} = 71 + 91 = 162, \qquad n_{2,\bullet} = 82 + 56 = 138,$$
$$n_{\bullet,1} = 71 + 82 = 153, \qquad n_{\bullet,2} = 91 + 56 = 147.$$

Nyní již spočítejme testovou statistiku, tedy

$$\begin{aligned}
\mathcal{X}^2 &= \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{\left(n_{i,j} - \frac{n_{i,\bullet}n_{\bullet,j}}{n}\right)^2}{\frac{n_{i,\bullet}n_{\bullet,j}}{n}} = \\
&= \frac{\left(n_{1,1} - \frac{n_{1,\bullet}n_{\bullet,1}}{n}\right)^2}{\frac{n_{1,\bullet}n_{\bullet,1}}{n}} + \frac{\left(n_{1,2} - \frac{n_{1,\bullet}n_{\bullet,2}}{n}\right)^2}{\frac{n_{1,\bullet}n_{\bullet,2}}{n}} + \\
&\qquad + \frac{\left(n_{2,1} - \frac{n_{2,\bullet}n_{\bullet,1}}{n}\right)^2}{\frac{n_{2,\bullet}n_{\bullet,1}}{n}} + \frac{\left(n_{2,2} - \frac{n_{2,\bullet}n_{\bullet,2}}{n}\right)^2}{\frac{n_{2,\bullet}n_{\bullet,2}}{n}} = \\
&= \frac{\left(71 - \frac{162\cdot 153}{300}\right)^2}{\frac{162\cdot 153}{300}} + \frac{\left(91 - \frac{162\cdot 147}{300}\right)^2}{\frac{162\cdot 147}{300}} + \\
&\qquad + \frac{\left(82 - \frac{138\cdot 153}{300}\right)^2}{\frac{138\cdot 153}{300}} + \frac{\left(56 - \frac{138\cdot 147}{300}\right)^2}{\frac{138\cdot 147}{300}} \doteq \\
&\doteq 1{,}634 + 1{,}701 + 1{,}919 + 1{,}997 \doteq \\
&\doteq 7{,}251.
\end{aligned}$$

Kritický obor pro zamítnutí nulové hypotézy je $W_\alpha = \{\mathcal{X}^2 : \mathcal{X}^2 \geq \chi^2_{1-\alpha}(1)\}$. Pro $\alpha = 0{,}05$ je

$$\chi^2_{1-\alpha}(1) = \chi^2_{1-0{,}05}(1) = \chi^2_{0{,}95}(1) \doteq 3{,}841,$$

takže

$$W_\alpha = W_{0{,}05} = \{\mathcal{X}^2 : \mathcal{X}^2 \geq \chi^2_{1-\alpha}(1)\} \doteq \{\mathcal{X}^2 : \mathcal{X}^2 \geq 3{,}841\}.$$

Protože $\mathcal{X}^2 \doteq 7{,}251 \in W_{0{,}05} \doteq \{\mathcal{X}^2 : \mathcal{X}^2 \geq 3{,}841\}$, zamítáme na hladině významnosti 0,05 nulovou hypotézu o tom, že preference budoucího bydliště jsou nezávislé na pohlaví. Závěr tedy naznačuje, že preference budoucího bydliště *jsou závislé* na pohlaví.

Řešení v MS Excel®. Příklad je přehledově vyřešen v záložce

`preferovane_bydliste_vs_pohlavi`

přiloženého excelového souboru

`_sbirka_uloh_pro_ucely_bakalarskych_statistickych_predmetu_.xlsx`.

□

Úloha 7.1.7 Při 600 hodech hrací kostkou byly zjištěny následující četnosti jednotlivých šesti stran

$$85, \quad 99, \quad 91, \quad 108, \quad 119, \quad 98.$$

Lze na hladině významnosti 0,05 považovat takovou kostku za spravedlivou?

Řešení. Označme symbolem $\pi_{0,j}$ očekávanou relativní četnost padnutí j-té stěny, kde $j \in \{1, 2, \ldots, 6\}$, a symbolem $n = 600$ celkový počet hodů kostkou. Předpokládejme, že kostka je spravedlivá, a podle toho sestavme nulovou hypotézu H_0. Za platnosti nulové hypotézy jsou zjevně relativní četnosti padnutí každé ze stěn shodné a rovné jedné šestině, tedy

$$H_0 : \pi_{0,1} = \pi_{0,2} = \pi_{0,3} = \pi_{0,4} = \pi_{0,5} = \pi_{0,6} = \frac{1}{6}.$$

Pomocí χ^2 testu dobré shody lze porovnat očekávané absolutní četnosti padnutí jednotlivých stěn kostky $n\pi_{0,j}$, kde $j \in \{1, 2, \ldots, 6\}$, s pozorovanými absolutními četnostmi padnutí jednotlivých stěn kostky, n_j, kde opět $j \in \{1, 2, \ldots, 6\}$.

Testovou statistikou je

$$\mathcal{X}^2 = \sum_{j=1}^{k} \frac{(n_j - n\pi_{0,j})^2}{n\pi_{0,j}},$$

která za platnosti nulové hypotézy sleduje χ^2 rozdělení o počtu stupňů volnosti rovném počtu jednotlivých pozorvaných tříd zmenšených o jedna, tedy $6 - 1 = 5$, takže $\mathcal{X}^2 \sim \chi^2_{1-\alpha}(5)$.

Pro očekávané absolutní četnosti padnutí první až šesté stěny můžeme za platnosti nulové hypotézy psát

$$n\pi_{0,1} = n\pi_{0,2} = n\pi_{0,3} = n\pi_{0,4} = n\pi_{0,5} = n\pi_{0,6} = n\frac{1}{6} = 600 \cdot \frac{1}{6} = 100.$$

Pozorované absolutní četnosti padnutí první až šesté stěny můžeme známe ze zadání,

$$\begin{aligned} n_1 &= 85, \qquad & n_2 &= 99, \qquad & n_3 &= 91, \\ n_4 &= 108, \qquad & n_5 &= 119, \qquad & n_6 &= 98. \end{aligned}$$

Nyní již můžeme spočítat testovou statistiku

$$\begin{aligned}
\mathcal{X}^2 &= \sum_{j=1}^{k} \frac{(n_j - n\pi_{0,j})^2}{n\pi_{0,j}} = \\
&= \sum_{j=1}^{6} \frac{\left(n_j - 600 \cdot \frac{1}{6}\right)^2}{600 \cdot \frac{1}{6}} = \\
&= \frac{(85-100)^2}{100} + \frac{(99-100)^2}{100} + \frac{(91-100)^2}{100} + \\
&\qquad + \frac{(108-100)^2}{100} + \frac{(119-100)^2}{100} + \frac{(98-100)^2}{100} = \\
&= 2{,}25 + 0{,}01 + 0{,}81 + 0{,}64 + 3{,}61 + 0{,}04 \doteq \\
&\doteq 7{,}36.
\end{aligned}$$

Kritický obor pro zamítnutí nulové hypotézy je $W_\alpha = \{\mathcal{X}^2 : \mathcal{X}^2 \geq \chi^2_{1-\alpha}(5)\}$. Pro $\alpha = 0{,}05$ je

$$\chi^2_{1-\alpha}(5) = \chi^2_{1-0,05}(5) = \chi^2_{0,95}(5) \doteq 11{,}07,$$

takže

$$W_\alpha = W_{0,05} = \{\mathcal{X}^2 : \mathcal{X}^2 \geq \chi^2_{1-\alpha}(5)\} \doteq \{\mathcal{X}^2 : \mathcal{X}^2 \geq 11{,}07\}.$$

Protože $\mathcal{X}^2 \doteq 7{,}36 \notin W_{0,05} \doteq \{\mathcal{X}^2 : \mathcal{X}^2 \geq 11{,}07\}$, nezamítáme na hladině významnosti 0,05 nulovou hypotézu o tom, že relativní četnosti padnutí jednotlivých stěn kostky jsou shodné. Závěr tedy naznačuje, že kostka *se jeví* jako spravedlivá.

Řešení v MS Excel®. Příklad je přehledově vyřešen v záložce

`600_hodu_kostkou`

přiloženého excelového souboru

`_sbirka_uloh_pro_ucely_bakalarskych_statistickych_predmetu_.xlsx`.

□

Úloha 7.1.8 V následující tabulce jsou uvedeny hodnoty nákladů na údržbu deseti domů (v dolarech) a tržní cena těchto domů (v tisících dolarů).

i	náklady na údržbu [\$]	cena domu [tisíce \$]
1	835	136
2	63	24
3	240	52
4	1005	143
5	184	42
6	213	43
7	313	67
8	658	106
9	195	61
10	545	99

(i) Modelujme závislost nákladů na údržbu na ceně tržní domu regresní přímkou.
(ii) Zhodnoťme kvalitu modelu pomocí koeficientu determinace.
(iii) Odhadněme střední hodnotu nákladů u domů za 80 tisíc dolarů.
(iv) Spočítejme korelační koeficient mezi tržní cenou domu a náklady na jeho údržbu.

Řešení. Dvojic pozorování tržní cena domu a náklady na jeho údržbu je zjevně $n = 10$. Označme hodnoty nákladů na údržbu (v dolarech) postupně y_i pro $\forall i \in \{1, 2, 3, \dots, 10\}$ a hodnoty tržní ceny (v tisících dolarů) postupně x_i pro $\forall i \in \{1, 2, 3, \dots, 10\}$.

(i) Sestavujeme regresní přímkovou závislost ve tvaru

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

kde β_0 je absolutní člen (průsečík regresní přímky se svislou osou), β_1 je směrnice regresní přímky a ε_i je i-té reziduum, tj. svislá vzdálenost mezi regresní přímkou a skutečnou hodnotou i-tého pozorování závisle proměnné, $i \in \{1, 2, 3, \dots, 10\}$. Samotná regresní přímka využívá odhadu absolutního členu $\hat{\beta}_0$ a odhadu směrnice $\hat{\beta}_1$ a má tvar $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, kde $i \in \{1, 2, 3, \dots, 10\}$.

Regresní přímka pro dané x_i protíná střední hodnotu y_i, neboť $\mathbb{E}(y_i) = \mathbb{E}(\beta_0 + \beta_1 x_i + \varepsilon_i) = \mathbb{E}(\beta_0) + \mathbb{E}(\beta_1 x_i) + \mathbb{E}(\varepsilon_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i + 0 = \hat{\beta}_0 + \hat{\beta}_1 x_i$, neboť střední hodnotu reziduí předpokládáme nulovou, $\mathbb{E}(\varepsilon_i) = 0$ pro $\forall i \in \{1, 2, 3, \dots, 10\}$.

Snadno též nahlédneme, že bod $[\overline{x}, \overline{y}]$ leží na regresní přímce, protože pokud $\forall i \in \{1, 2, 3, \dots, 10\}$ je $y_i = \beta_0 + \beta_1 x_i$, pak je i $\sum_{i=1}^n y_i = \sum_{i=1}^n (\beta_0 + \beta_1 x_i) = \beta_0 + \beta_1 \sum_{i=1}^n x_i$ a též $\frac{1}{n} \sum_{i=1}^n y_i = \beta_0 + \beta_1 \frac{1}{n} \sum_{i=1}^n x_i$, neboli $\overline{y} = \beta_0 + \beta_1 \overline{x}$.

Koeficienty β_0 a β_1 regresní přímky lze odhadnout pomocí následujících vzorců (které lze snadno odvodit)

$$\begin{aligned} \hat{\beta}_1 &= \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2}, \\ \hat{\beta}_0 &= \overline{y} - \hat{\beta}_1 \overline{x}, \end{aligned}$$

kde symbol typu $\overline{z}$ značí průměrnou hodnotu veličiny $z \in \{x, y, x^2, xy\}$.

Dopočítejme nyní potřebné hodnoty,

$$\begin{aligned} \overline{x} &= \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{10} \sum_{i=1}^{10} x_i = \frac{1}{10}(136 + 24 + \dots + 99) \doteq \\ &\doteq 77{,}3 \text{ [tisíc \$]}, \\ \overline{y} &= \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{10} \sum_{i=1}^{10} y_i = \frac{1}{10}(835 + 63 + \dots + 545) \doteq \\ &\doteq 425{,}1 \text{ [\$]}. \\ \overline{x^2} &= \frac{1}{n} \sum_{i=1}^n x_i^2 = \frac{1}{10} \sum_{i=1}^{10} x_i^2 = \frac{1}{10}(136^2 + 24^2 + \dots + 99^2) \doteq \\ &\doteq 7508{,}5 \text{ [tisíc}^2 \text{ \$}^2\text{]}, \\ \overline{xy} &= \frac{1}{n} \sum_{i=1}^n x_i y_i = \frac{1}{10} \sum_{i=1}^{10} x_i y_i = \frac{1}{10}(136 \cdot 835 + 24 \cdot 63 + \dots + 99 \cdot 545) \doteq \\ &\doteq 44472{,}3 \text{ [tisíc \$}^2\text{]}. \end{aligned}$$

Nyní již můžeme odhadnout koeficienty β_0 a β_1 regresní přímky,

$$\hat{\beta}_1 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2} \doteq \frac{44472{,}3 - 77{,}3 \cdot 425{,}1}{7508{,}5 - 77{,}3^2} \doteq 7{,}57,$$
$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} \doteq 425{,}1 - 7{,}57 \cdot 77{,}3 \doteq -160{,}35.$$

Regresní přímka má tedy tvar $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = -160{,}35 + 7{,}57 x_i$, kde $i \in \{1, 2, 3, \dots, 10\}$. Hodnota koeficientu $\hat{\beta}_0 = -160{,}35$ značí hodnotu nákladů na údržbu u domu s tržní cenou 0 tisíc dolarů, neboť $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \,|_{x_i=0} = \hat{\beta}_0 + \hat{\beta}_1 \cdot 0 = \hat{\beta}_0 \doteq -160{,}35$. Taková tržní cena (0 tisíc dolarů) ovšem není smysluplná, stejně jako záporné náklady na údržbu. Hodnota koeficientu $\hat{\beta}_1 = 7{,}57$ značí, o kolik dolarů průměrně vzrostou náklady na údržbu domu, pokud jeho tržní cena vzroste o jeden tisíc dolarů, neboť $\hat{y}_i' - \hat{y}_i = (\hat{\beta}_0 + \hat{\beta}_1 x_i') - (\hat{\beta}_0 + \hat{\beta}_1 x_i) = (\hat{\beta}_0 + \hat{\beta}_1 (x_i + 1)) - (\hat{\beta}_0 + \hat{\beta}_1 x_i) = \hat{\beta}_1 \doteq 7{,}57$.

(ii) Index determinace I^2 je podílem $I^2 = \frac{S_T}{S_T + S_R}$, kde S_T je regresní součet čtverců a S_R je reziduální součet čtverců tak, že

$$S_T = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2,$$
$$S_R = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \hat{\varepsilon}_i^2.$$

Protože $I^2 = \frac{S_T}{S_T + S_R}$, snadno nahlédneme, že $0 \leq I^2 \leq 1$. Čím vyšší je hodnota I^2, tím větší podíl variability závisle proměnné v přímkové regresi lze vysvětlit pomocí vysvětlující proměnné.

Dopočítejme hodnoty regresního součtu čtverců S_T a reziduálního součtu čtverců S_R,

$$\begin{aligned}
S_T &= \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2 = \sum_{i=1}^{n} (\hat{\beta}_0 + \hat{\beta}_1 x_i - \overline{y})^2 \doteq \\
&\doteq \sum_{i=1}^{10} (-160{,}35 + 7{,}57 x_i - 425{,}1)^2 \doteq \\
&\doteq (-160{,}35 + 7{,}57 \cdot 136 - 425{,}1)^2 + (-160{,}35 + 7{,}57 \cdot 24 - 425{,}1)^2 + \dots + \\
&\qquad + (-160{,}35 + 7{,}57 \cdot 99 - 425{,}1)^2 \doteq \\
&\doteq 879463{,}15, \\
S_R &= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \sum_{i=1}^{n} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 \doteq \\
&\doteq \sum_{i=1}^{10} (y_i - (-160{,}35 + 7{,}57 x_i))^2 \doteq \\
&\doteq (835 - (-160{,}35 + 7{,}57 \cdot 136))^2 + (63 - (-160{,}35 + 7{,}57 \cdot 24))^2 + \dots + \\
&\qquad + (545 - (-160{,}35 + 7{,}57 \cdot 99))^2 \doteq \\
&\doteq 27463{,}75.
\end{aligned}$$

Odtud již

$$I^2 = \frac{S_T}{S_T + S_R} \doteq \frac{879463{,}15}{879463{,}15 + 27463{,}75} \doteq 0{,}9697,$$

což je hodnota velmi blízká jedné. Regresní přímkový model tak lze považovat za velmi dobře vystihující daný vztah mezi náklady na údržbu domu a jeho tržní cenou. Variabilitu závisle proměnné v naší přímkové regresi lze vysvětlit pomocí vysvětlující proměnné z asi 96,97 %.

(iii) Víme-li, že střední hodnota reziduí je nulová, $\mathbb{E}(\varepsilon_i) = 0$ pro $\forall i \in \{1, 2, 3, \dots, 10\}$, můžeme spočítat střední hodnotu nákladů na údržbu domu s tržní cenou 80 tisíc dolarů jako

$$\begin{aligned}
\mathbb{E}(y_i \mid x_i = 80\ [\text{tisíc } \$]) &= \mathbb{E}(\beta_0 + \beta_1 x_i + \varepsilon_i \mid x_i = 80\ [\text{tisíc } \$]) = \\
&= \mathbb{E}(\beta_0) + \mathbb{E}(\beta_1 x_i \mid x_i = 80) + \mathbb{E}(\varepsilon_i) = \\
&= \hat{\beta}_0 + \hat{\beta}_1 (x_i \,|_{x_i = 80}) + 0 = \\
&= \hat{\beta}_0 + \hat{\beta}_1 (x_i \,|_{x_i = 80}) \doteq \\
&\doteq -160{,}35 + 7{,}57 \cdot 80 \doteq \\
&\doteq 445{,}25\ [\$].
\end{aligned}$$

Očekávané náklady na údržbu domu s tržní cenou 80 tisíc dolarů jsou přibližně 445,25 dolaru.

(iv) Korelační koeficient (Pearsonův) mezi tržní cenou domu a náklady na jeho údržbu spočítáme pomocí následujícího vzorce

$$r_{xy} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{\left(\overline{x^2} - \overline{x}^2\right)\left(\overline{y^2} - \overline{y}^2\right)}}$$

při již zavedeném značení. Dopočítejme ještě průměrný čtverec nákladů na údržbu domu $\overline{y^2}$,

$$\begin{aligned}
\overline{y^2} &= \frac{1}{n}\sum_{i=1}^{n} y_i^2 = \frac{1}{10}\sum_{i=1}^{10} y_i^2 = \frac{1}{10}(835^2 + 63^2 + \dots + 545^2) \doteq \\
&\doteq 271402{,}7\ [\$^2].
\end{aligned}$$

Nyní již dopočítejme korelační koeficient

$$\begin{aligned}
r_{xy} &= \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{\left(\overline{x^2} - \overline{x}^2\right)\left(\overline{y^2} - \overline{y}^2\right)}} \doteq \\
&\doteq \frac{44472{,}3 - 77{,}3 \cdot 425{,}1}{\sqrt{(7508{,}5 - 77{,}3^2) \cdot (271402{,}7 - 425{,}1^2)}} \doteq \\
&\doteq 0{,}9847.
\end{aligned}$$

Všimněme si, že korelační koeficient můžeme dopočítat (pouze) v případě přímkové lineární regrese též jako odmocnicnu z indexu determinace, tedy

$$\begin{aligned} r_{xy} &= \sqrt{I^2} = \sqrt{\frac{S_T}{S_T + S_R}} \doteq \\ &\doteq \sqrt{\frac{879463{,}15}{879463{,}15 + 27463{,}75}} \doteq \sqrt{0{,}9697} \doteq \\ &\doteq 0{,}9847. \end{aligned}$$

Uvedený vztah $I^2 = r_{xy}^2$ lze naopak u přímkové (!) lineární regrese využít i pro výpočet indexu determinace I^2 ze známého korelačního koeficientu r_{xy}.

Řešení v MS Excel®. Příklad je přehledově vyřešen v záložce

`udrzba_domu_vs_jeho_cena`

přiloženého excelového souboru

`_sbirka_uloh_pro_ucely_bakalarskych_statistickych_predmetu_.xlsx`.

□

7.2 Úlohy k vlastnímu procvičení

Úloha 7.2.1 Spotřeba téhož auta byla testována u jedenácti řidičů s výsledky (vždy v litrech na 100 kilometrů)

8,8, 8,9, 9,0, 8,7, 9,3, 9,0, 8,7, 8,8, 9,4, 8,6, 8,9.

Je pravdivá výrobcem udávaná spotřeba 8,8 l/100 km? Testujme na hladině významnosti 0,05.

Úloha 7.2.2 Je padnutí 22 líců při 40 hodech mincí důkazem její nevyváženosti na hladině významnosti 0,05? Od jakého rozsahu výběru je 55% líců na hladině významnosti 0,05 již významný výsledek?

Úloha 7.2.3 Ze 100 hodů jednou mincí padla hlava 60-krát.

(i) Je pravděpodobnost padnutí hlavy a orla shodná na hladině významnosti 0,05? Testujme pomocí testu o parametru pro alternativního rozdělení.

(ii) Nyní pomocí následující kontingenční tabulky

	strana mince	
	hlava	orel
počet padnutí	60	40

a vhodného χ^2 testu otestujme zcela shodnou hypotézu, opět na hladině významnosti 0,05. Docházíme ke stejnému závěru? Je možné legitimně použít oba předchozí přístupy? Krátce vysvětleme.

(iii) Kolikrát by během 100 hodů musela padnout hlava, abychom zamítli hypotézu o shodné pravděpodobnosti padnutí hlavy a orla na hladině významnosti 0,01?

Úloha 7.2.4 V následující tabulce jsou uvedeny hmotnosti těl různých zvířecích druhů (v kilogramech) a současně i hmotnosti jejich mozků (v gramech).

pořadí druhu, i	1	2	3	4	5	6	7	8	9	10
hmotnost těla druhu [kg], x_i	4	5	7	12	15	16	20	25	28	35
hmotnost mozku druhu [g], y_i	40	45	42	50	55	60	75	79	80	88

(i) Modelujme závislost hmotnoti mozku druhů na hmotnosti těla druhů pomocí regresní přímky založené na minimalizaci součtu čtverců reziduí. Odhadněme parametry regresní přímky a zapišme tuto přímku ve tvaru

$$\hat{\boldsymbol{y}} = \hat{\beta}_0 + \hat{\beta}_1 \boldsymbol{x}.$$

(ii) Zhodnoťme kvalitu regresního modelu pomocí indexu determinace.
(iii) Interpretujme věcně hodnotu odhadu regresního koeficientu $\hat{\beta}_0$ a odhadu koeficientu $\hat{\beta}_1$.
(iv) Odhadněme střední hodnotu hmotnosti mozku zvířecího druhu, jehož hmotnost těla je 22 kg.

Reference

Následující prameny byly inspirací pro některé úlohy zařazené v této sbírce. Lze je doporučit jako zdroje nepřeberného množství dalších příkladů.

[1] Diana Bílková, Adam Čabla, Tomáš Karel, Elena Makhalova, Ivana Malá, Lukáš Sobíšek a Zdeněk Šulc. *Sbírka řešených příkladů ze statistiky*. 2015. URL: `https://statsbirka.vse.cz/` (cit. 15. 03. 2020).

[2] Petr Hebák. *Počet pravděpodobnosti v příkladech*. Praha: Informatorium, 2010. ISBN: 9788073330774.

[3] Petr Otipka. *Pravděpodobnost a statistika*. Ostrava: Vysoká škola báňská - Technická univerzita, 2006. ISBN: 80-248-1194-4.

ŠTĚPÁNEK, Lubomír. *Sbírka příkladů pro účely bakalářských statistických předmětů — 4ST201 Statistika a 4ST210 Statistika pro finance.* Praha, 2020. Neoficiální učební pomůcka. Katedra statistika a pravděpodobnosti. Fakulta informatiky a statistiky. Vysoká škola ekonomická.