

Řešení vybraných kontrolních úkolů

Kapitola 3.

5. V akciové společnosti je průměrná mzda 13 500 Kč. Přitom 30% pracovníků s nejnižší mzdou má průměrně 9000 Kč. Na začátku roku dostal každý z těchto pracovníků přidáno 500 Kč. O kolik % vzrostla průměrná mzda v celé akciové společnosti?

Řešení:

Všech pracovníků je n . Celkový objem měsíčně vyplacených peněz je $n \cdot 13\,500$ Kč.

Označme m_2 průměrnou mzdou zbylých 70% pracovníků. Musí platit:

$$n \cdot 13\,500 = 0,3n \cdot 9\,000 + 0,7n \cdot m_2. \text{ Odtud vypočteme } m_2 = 10\,800/0,7 \text{ Kč.}$$

Po přidání je celkový objem měsíčně vyplacených peněz roven

$$(0,3n \cdot 9\,500 + 0,7n \cdot 10\,800/0,7) \text{ Kč.}$$

Tedy $n \cdot 13\,500$ je 100% a $0,3n \cdot 9\,500 + 0,7n \cdot 10\,800/0,7$ je $x\%$.

Odtud trojčlenkou vypočteme $x = 101,1\%$. Vidíme, že průměrná měsíční mzda v celé akciové společnosti vzrostla o 1,1%.

6. Při statistickém šetření pojištěnců byly získány tyto výše pojistek v Kč:

výše pojistky	390	410	430	450	470	490	510	530	550	570
abs. četnost	7	10	14	22	25	12	3	3	2	2

Určete aritmetický průměr, medián, modus, rozptyl, směrodatnou odchylku a koeficient variace výše pojistky.

Řešení:

(Ruční výpočet, bez použití software STATISTICA)

Použijeme vzorec pro vážený aritmetický průměr. Rozsah souboru $n = 7 + 10 + \dots + 2 = 100$.

$$m = \frac{1}{n} \sum_{j=1}^r n_j x_{[j]} = \frac{1}{100} (7 \cdot 390 + 10 \cdot 410 + \dots + 2 \cdot 570) = 457,4$$

Datový soubor má sudý rozsah, tedy medián je průměr dvou prostředních hodnot uspořádaného datového souboru, tj. průměr 50. a 51. uspořádané hodnoty, tedy 450.

Modus je nejčetnější varianta znaku, tj. 470.

Nejprve vypočteme rozptyl podle vzorce pro vážený rozptyl.

$$s^2 = \frac{1}{n} \sum_{j=1}^r n_j (x_{[j]} - m)^2 =$$

$$\frac{1}{100} [7 \cdot (390 - 457,4)^2 + 10 \cdot (410 - 457,4)^2 + \dots + 2 \cdot (570 - 457,4)^2] = 1493,24$$

$$\text{Směrodatná odchylka } s = \sqrt{s^2} = \sqrt{1493,24} = 38,64$$

$$\text{Koeficient variace } \frac{s}{m} = \frac{38,64}{457,4} = 0,08$$

7. V datovém souboru, z něhož byl vypočten průměr 110 a rozptyl 800, byly zjištěny 2 chyby: místo 85 má být 95 a místo 120 má být 150. Ostatních 18 údajů je správných. Opravte průměr a rozptyl.

Řešení:

Bez újmy na obecnosti lze předpokládat, že prvních 18 hodnot je správných.

Původní průměr:

$$110 = m_1 = \frac{1}{20} \sum_{i=1}^{20} x_i = \frac{1}{20} \left(\sum_{i=1}^{18} x_i + 85 + 120 \right) \Rightarrow \sum_{i=1}^{18} x_i = 20 \cdot 110 - 85 - 120 = 1995$$

Nový průměr:

$$m_2 = \frac{1}{20} \left(\sum_{i=1}^{18} x_i + 95 + 150 \right) = \frac{1}{20} (1995 + 95 + 150) = 112$$

Původní rozptyl:

$$800 = \frac{1}{20} \sum_{i=1}^{20} x_i^2 - m_1^2 = \frac{1}{20} \left(\sum_{i=1}^{18} x_i^2 + 85^2 + 120^2 \right) - 110^2 \Rightarrow \sum_{i=1}^{18} x_i^2 = 236375$$

Nový rozptyl:

$$s_2^2 = \frac{1}{20} \left(\sum_{i=1}^{18} x_i^2 + 95^2 + 150^2 \right) - m_2^2 = \frac{1}{20} (236375 + 95^2 + 150^2) - 112^2 = 851$$

8. Vážený aritmetický průměr činil 1500 a vážený rozptyl 90000. Varianty $x_{[j]}$ byly transformovány vztahem: $y_{[j]} = \frac{x_{[j]} - a}{h}$, $j = 1, \dots, r$. Po této transformaci byl vážený aritmetický průměr 5 a vážený rozptyl 9. Určete konstanty a a h .

Řešení:

Podle věty 3.18. a) je průměr lineární kombinace $Y = a + bX$ roven $m_2 = a + bm_1$. V našem případě tedy $5 = \frac{1500 - a}{h}$. Podle věty 3.18. a) je rozptyl lineární kombinace $Y = a + bX$

roven $s_2^2 = b^2 s_1^2$. V našem případě tedy $9 = \frac{90000}{h^2}$. Odtud $h = 100$. Z první rovnice dostaneme $a = 1500 - 5 \cdot 100 = 1000$.

10. Rozptyl součtů hodnot dvou znaků je 350, rozptyl rozdílů je 700. Vypočtete koeficient korelace, víte-li, že oba znaky mají stejné rozptyly.

Řešení:

Podle věty 3.18. b) je rozptyl součtu $U = X + Y$ roven $s_3^2 = s_1^2 + s_2^2 + 2s_{12}$ a analogicky rozptyl rozdílu $V = X - Y$ roven $s_4^2 = s_1^2 + s_2^2 - 2s_{12}$. Společnou hodnotu rozptylů znaků X a Y označíme s^2 . Dostáváme tedy dvě rovnice pro dvě neznámé:

$$2s^2 + 2s_{12} = 350$$

$$2s^2 - 2s_{12} = 700$$

Odtud vypočteme $s^2 = 262,5$ a $s_{12} = -87,5$. Koeficient korelace se počítá podle vzorce

$$r_{12} = \frac{s_{12}}{s_1 s_2}, \text{ v našem případě } r_{12} = \frac{s_{12}}{s^2} = -\frac{87,5}{262,5} = -\frac{1}{3}$$

Kapitola 4.

7. Je dána rovnice regresní přímky $y = 87 + 0,3(x - 25)$ a koeficient korelace $r_{12} = 0,77$. Najděte rovnici sdružené regresní přímky.

Řešení:

Původní regresní přímka má rovnici $y = m_2 + b_1(x - m_1)$, sdružená regresní přímka má rovnici $x = m_1 + \bar{b}_1(y - m_2)$, přičemž $b_1 \bar{b}_1 = r_{12}^2$, tedy $0,3 \bar{b}_1 = 0,77^2$. Odtud $\bar{b}_1 = 1,9763$. Dále $x = 25 + 1,9763(y - 87)$.

9. Jak se změní úsek a směrnice regresní přímky, když každou hodnotu závisle proměnného znaku zvětšíme o 10%?

Řešení:

Původní regresní přímka má rovnici $y = b_0 + b_1x$, nová regresní přímka má rovnici $1,1y = 1,1(b_0 + b_1x)$, tedy úsek i směrnice se zvýší o 10%.

10. Závislost mezi vnější teplotou a teplotou ve skladišti je popsána regresní přímkou $y = 8 + 0,6x$. Při jaké vnější teplotě klesne teplota ve skladišti pod bod mrazu?

Řešení:

$$0 = 8 + 0,6x, \text{ tedy; } x = -\frac{8}{0,6} = -13,3$$

Kapitola 5.

6. Z karetní hry o 32 kartách vybereme náhodně bez vracení 4 karty. Jaká je pravděpodobnost, že aspoň jedna z nich je eso?

Řešení:

Přejdeme k opačnému jevu, který znamená, že ani jedna z vybraných karet není eso. Pak

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{28 \cdot 27 \cdot 26 \cdot 25}{32 \cdot 31 \cdot 30 \cdot 29} = 0,4306$$

7. Dva hráči házejí střídavě mincí. Vyhrává ten, komu padne dřív líc. Stanovte pravděpodobnost výhry 1. hráče a pravděpodobnost výhry 2. hráče.

Řešení:

A_1 ... v 1. hodu padne líc, A_3 ... ve 3. hodu padne líc, A_5 ... v 5. hodu padne líc, atd.

A ... vyhraje první hráč, $A = A_1 \cup A_3 \cup A_5 \cup \dots$ Jevy A_1, A_3, A_5, \dots jsou neslučitelné, tedy

$$P(A) = P(A_1) + P(A_3) + P(A_5) + \dots = \frac{1}{2} + \frac{1}{2^3} + \frac{1}{2^5} + \dots = \frac{\frac{1}{2}}{1 - \frac{1}{2^2}} = \frac{2}{3}.$$

Pravděpodobnost výhry druhého hráče je tedy $P(\bar{A}) = 1 - P(A) = \frac{1}{3}$.

8. Chevalier de Méré pozoroval, že při házení třemi kostkami padá součet 11 častěji než součet 12, i když podle jeho názoru (nesprávného) mají oba součty stejnou pravděpodobnost. Stanovte pravděpodobnost obou jevů.

Řešení:

jev A: padne součet 11
příznivé výsledky:

4 4 3 3 způsoby

4 5 2 3! způsoby

4 6 1 3! způsoby

5 5 1 3 způsoby

5 3 3 3 způsoby

6 3 2 3! způsoby

celkem 27 možností

$$m(\Omega) = 6^3$$

$$P(A) = 27/6^3 = 0,125$$

jev B: padne součet 12
příznivé výsledky:

4 4 4 1 možnost

4 5 3 3! způsoby

4 6 2 3! způsoby

5 5 2 3 způsoby

5 6 1 3! způsoby

6 3 3 3 způsoby

celkem 25 možností

$$P(B) = 25/6^3 = 0,1157$$

9. Student se ke zkoušce připravil na 15 otázek z 20 zadaných. Při zkoušce si vybere náhodně dvě otázky. Jaká je pravděpodobnost, že aspoň na jednu zná odpověď?

Řešení:

Přejdeme k opačnému jevu, který znamená, že student nezná odpověď ani na jednu ze dvou vybraných otázek.

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{5 \cdot 4}{20 \cdot 19} = \frac{18}{19}$$

Kapitola 6.

4. Je pravděpodobnější vyhrát se stejně silným soupeřem tři partie ze čtyř nebo pět z osmi, když nerozhodný výsledek je vyloučen a výsledky jsou nezávislé?

Řešení:

A ... výhra právě ve třech partiích ze čtyř, přičemž pravděpodobnost výhry je $\frac{1}{2}$. Podle vzorce pro binomické rozložení pravděpodobností (viz př. 6.3. b)) dostaneme:

$$P(A) = \binom{4}{3} \left(\frac{1}{2}\right)^3 \left(1 - \frac{1}{2}\right)^{4-3} = \frac{1}{4} = 0,25$$

B ... výhra právě v pěti partiích z osmi, přičemž pravděpodobnost výhry je $\frac{1}{2}$. Podle vzorce pro binomické rozložení pravděpodobností (viz př. 6.3. b)) dostaneme:

$$P(B) = \binom{8}{5} \left(\frac{1}{2}\right)^5 \left(1 - \frac{1}{2}\right)^{8-5} = 0,219$$

Je tedy pravděpodobnější vyhrát tři partie ze čtyř než pět z osmi.

5. První dělník vyrobí denně 60 výrobků, z toho 10% zmetků. Druhý dělník vyrobí denně 40 výrobků, z toho 5% zmetků. Jaká je pravděpodobnost, že náhodně vybraný výrobek z denní produkce je zmetek a pochází od prvního dělníka?

Řešení:

A ... výrobek pochází od 1. dělníka, H ... výrobek je zmetek. Přitom 1. dělník vyrobí denně 6 zmetků, druhý dva zmetky.

$$P(A \cap H) = P(H)P(A/H) = \frac{8}{100} \cdot \frac{6}{8} = 0,06$$

6. Ze šesti vajec jsou dvě prasklá. Náhodně vybereme dvě vejce. Jaká je pravděpodobnost, že budou a) obě prasklá, b) právě jedno prasklé, c) obě dobrá?

Řešení:

A_1 ... první vybrané vejce je prasklé, A_2 ... druhé vybrané vejce je prasklé.

$$\text{ad a) } P(A_1 \cap A_2) = P(A_1)P(A_2 / A_1) = \frac{2}{6} \cdot \frac{1}{5} = \frac{1}{15}$$

$$\text{ad b) } P(A_1 \cap \overline{A_2}) + P(\overline{A_1} \cap A_2) = P(A_1)P(\overline{A_2} / A_1) + P(\overline{A_1})P(A_2 / \overline{A_1}) = \frac{2}{6} \cdot \frac{4}{5} + \frac{4}{6} \cdot \frac{2}{5} = \frac{8}{15}$$

$$\text{ad c) } P(\overline{A_1} \cap \overline{A_2}) = P(\overline{A_1})P(\overline{A_2} / \overline{A_1}) = \frac{4}{6} \cdot \frac{3}{5} = \frac{6}{15}$$

10. Pojišťovací společnost rozlišuje při pojišťování tři skupiny řidičů – A, B a C. Pravděpodobnost toho, že řidič patřící do skupiny A bude mít během roku nehodu, je 0,03, zatímco u řidiče skupiny B je to 0,06 a u řidiče skupiny C 0,1. Podle dlouhodobých záznamů společnosti je 70% pojistných smluv uzavřeno s řidiči skupiny A, 20% s řidiči skupiny B a 10% s řidiči skupiny C. Jestliže došlo k nehodě řidiče pojištěného u této společnosti, jaká je pravděpodobnost, že patřil do skupiny C?

Řešení:

Označme jevy takto:

N je jev: "došlo k nehodě"

H_1 je jev: "řidič patří do skupiny A"

H_2 je jev: "řidič patří do skupiny B"

H_3 je jev: "řidič patří do skupiny C"

Pak je

$$P(H_1) = 0,7, P(H_2) = 0,2, P(H_3) = 0,1, P(N/H_1) = 0,03, P(N/H_2) = 0,06, P(N/H_3) = 0,1$$

$$P(N) = P(H_1)P(N/H_1) + P(H_2)P(N/H_2) + P(H_3)P(N/H_3) = 0,7 \cdot 0,03 + 0,2 \cdot 0,06 + 0,1 \cdot 0,1 = 0,043$$

$$P(H_3 / N) = \frac{P(H_3)P(N/H_3)}{P(N)} = \frac{0,1 \cdot 0,1}{0,043} = 0,233$$

11. U jistého druhu elektrického spotřebiče se s pravděpodobností 0,01 vyskytuje výrobní vada. U spotřebiče s touto výrobní vadou dochází v záruční lhůtě k poruše s pravděpodobností 0,5. Výrobky, které tuto vadu nemají, se v záruční lhůtě porouchají s pravděpodobností 0,01. Jaká je pravděpodobnost, že

a) u náhodně vybraného výrobku nastane v záruční lhůtě porucha,

b) výrobek, který se v záruční lhůtě porouchá, bude mít dotyčnou výrobní vadu?

Řešení:

H_1 - výrobek má dotyčnou výrobní vadu

H_2 - výrobek nemá tuto výrobní vadu

A - výrobek se v záruční době porouchá

$$\text{Pak je: } P(H_1) = 0,01, P(H_2) = 0,99, P(A/H_1) = 0,5, P(A/H_2) = 0,01$$

$$\text{ad a) } P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) = 0,01 \cdot 0,5 + 0,99 \cdot 0,01 = 0,0149$$

$$\text{ad b) } P(H_1 / A) = \frac{P(H_1)P(A/H_1)}{P(A)} = \frac{0,01 \cdot 0,5}{0,0149} = 0,3386$$

Kapitola 7.

8. Náhodná veličina udává průměrný počet ok při hodu dvěma kostkami. Nakreslete graf její pravděpodobnostní funkce.

Řešení:

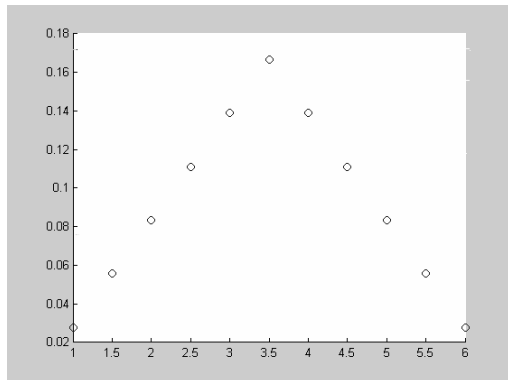
Náhodná veličina X nabývá dvanácti hodnot, a to 1, 1,5, 2, ..., 5,5, 6. Označme X_1 náhodnou veličinu, která udává počet ok při hodu první kostkou a X_2 náhodnou veličinu, která udává počet ok při hodu druhou kostkou.

$$\pi(1) = P(X = 1) = P(X_1 = 1 \wedge X_2 = 1) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

$$\pi(1,5) = P(X = 1,5) = P((X_1 = 1 \wedge X_2 = 2) \vee (X_1 = 2 \wedge X_2 = 1)) = \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{2}{36}$$

Analogicky počítáme hodnoty pravděpodobnostní funkce v dalších bodech.

Graf:



9. Diskrétní náhodný vektor (X_1, X_2) má simultánní pravděpodobnostní funkci $\pi(x_1, x_2)$ danou hodnotami: $\pi(0,0) = \pi(0,2) = \pi(1,1) = \pi(2,0) = \pi(2,2) = 0$, $\pi(0,1) = \pi(1,0) = \pi(1,2) = \pi(2,1) = 0,25$. Jsou náhodné veličiny X_1, X_2 stochasticky nezávislé?

Řešení:

Sestavíme kontingenční tabulku, v níž budou hodnoty simultánní pravděpodobnostní funkce a obou marginálních pravděpodobnostních funkcí.

x_1	x_2			$\pi_1(x_1)$
	0	1	2	
0	0	0,25	0	0,25
1	0,25	0	0,25	0,5
2	0	0,25	0	0,25
$\pi_2(x_2)$	0,25	0,5	0,25	1

Ověříme splnění multiplikativního vztahu $\forall (x_1, x_2) \in \mathbb{R}^2: \pi(x_1, x_2) = \pi_1(x_1) \pi_2(x_2)$. Již pro $x_1 = 0, x_2 = 0$ vztah splněn není, protože $\pi(0,0) = 0$, avšak $\pi_1(0) = 0,25$ a $\pi_2(0) = 0,25$. Veličiny X_1, X_2 tedy nejsou stochasticky nezávislé.

10. Necht' spojité vektor (X_1, X_2) má simultánní hustotu pravděpodobnosti

$$\varphi(x_1, x_2) = \begin{cases} 24x_1^2x_2(1-x_1) & \text{pro } 0 \leq x_1 < 1, 0 \leq x_2 < 1 \\ 0 & \text{jinak} \end{cases}. \text{ Doka\u017ee, \u017ee n\u00e1hodn\u00e9 veli\u010diny } X_1, X_2$$

jsou stochasticky nezávislé.

Řešení:

Vypočítáme obě marginální hustoty a ověříme platnost multiplikativního vztahu $\forall (x_1, \dots, x_n) \in \mathbb{R}^n: \varphi(x_1, \dots, x_n) = \varphi_1(x_1) \dots \varphi_n(x_n)$ s případnou výjimkou na množině bodů neovlivňujících integraci.

$$\varphi_1(x_1) = \int_0^1 24x_1^2x_2(1-x_1) dx_2 = 24x_1^2(1-x_1) \left[\frac{x_2}{2} \right]_0^1 = 12x_1^2(1-x_1) \text{ pro } 0 \leq x_1 < 1,$$

$$\varphi_1(x_1) = 0 \text{ jinak.}$$

$$\varphi_2(x_2) = \int_0^1 24x_1^2x_2(1-x_1) dx_1 = 24x_2 \left[\frac{x_1^3}{3} - \frac{x_1^4}{4} \right]_0^1 = 2x_2 \text{ pro } 0 \leq x_2 < 1,$$

$$\varphi_2(x_2) = 0 \text{ jinak.}$$

Vidíme, že multiplikativní vztah je splněn, tudíž veličiny X_1, X_2 jsou stochasticky nezávislé.

Kapitola 8.

2. Pojišťovna zjistila, že 12% pojistných událostí je způsobeno vloupáním. Jaká je pravděpodobnost, že mezi 30 náhodně vybranými pojistnými událostmi bude způsobeno vloupáním nejvýše 6?

Řešení:

$$X - \text{počet vloupání, } X \sim \text{Bi}(30; 0,12), P(X \leq 6) = \sum_{x=0}^6 \binom{30}{x} 0,12^x (1-0,12)^{30-x} = 0,939$$

3. Doba (v hodinách), která uplyne mezi dvěma naléhavými příjmy v jisté nemocnici, se řídí rozložením $\text{Ex}(0,5)$. Jaká je pravděpodobnost, že uplyne více než 5 hodin bez naléhavého příjmu?

Řešení:

X – doba, která uplyne mezi dvěma naléhavými příjmy, $X \sim \text{Ex}(0,5)$,

$$P(X > 5) = \int_5^{\infty} 0,5e^{-0,5x} dx = \left[-e^{-0,5x} \right]_5^{\infty} = e^{-2,5} = 0,0821$$

4. Jaká je pravděpodobnost, že náhodná veličina $X \sim N(20, 16)$ nabude hodnotu menší než 12 nebo větší než 28?

Řešení:

$$\begin{aligned} P(X < 12 \vee X > 28) &= P(X \notin \langle 12, 28 \rangle) = 1 - P(12 \leq X \leq 28) = 1 - P\left(\frac{12-20}{4} \leq \frac{X-20}{4} \leq \frac{28-20}{4}\right) = \\ &= 1 - P(-2 \leq U \leq 2) = 1 - [\Phi(2) - \Phi(-2)] = 1 - \Phi(2) + 1 - \Phi(2) = 2[1 - \Phi(2)] = 2(1 - 0,97725) = 0,0455 \end{aligned}$$

$$5. \text{ Necht' } X \sim \text{Rs}(a, b), \text{ přičemž } \Phi(x) = \begin{cases} 0 & \text{pro } x \leq a \\ \frac{x+20}{55} & \text{pro } a < x < b \\ 1 & \text{pro } x \geq b \end{cases}. \text{ Určete } a, b.$$

Řešení:

X je spojitá náhodná veličina, tedy její distribuční funkce je spojitá. Proto platí

$$\frac{a+20}{55} = 0, \frac{b+20}{55} = 1. \text{ Odtud dostaneme } a = -20, b = 35.$$

6. Necht' X_1, X_2 jsou stochasticky nezávislé náhodné veličiny, $X_i \sim N(0, 1), i = 1, 2$. Jaké rozložení má transformovaná náhodná veličina $X = \frac{X_1^2}{X_2^2}$?

Řešení:

Náhodná veličina $X_i^2 \sim \chi^2(1), i = 1, 2$ podle 8.6. e). Náhodná veličina $\frac{X_1^2}{X_2^2}$ má rozložení $F(1,1)$ podle 8.6. g).

Kapitola 9.

2. Necht' $X \sim N(-1, 4)$. Najděte $K_{0,025}(X)$.

Řešení:

$X \sim N(-1, 4) \Rightarrow U = \frac{X+1}{2} \sim N(0,1)$, tedy $X = 2U - 1$. Podle 9.5. a) dostáváme

$$K_{0,025}(X) = 2 \cdot u_{0,025} - 1 = 2 \cdot (-u_{0,975}) - 1 = -2 \cdot 1,95996 - 1 = -4,9199$$

3. Necht' X_1, X_2 jsou stochasticky nezávislé náhodné veličiny, $X_1 \sim N(2, 4), X_2 \sim N(-1, 9)$. Vypočtete 99% kvantil transformované náhodné veličiny $Y = 2X_1 - 3X_2 + 5$.

Řešení:

Protože náhodná veličina Y je lineární kombinací dvou náhodných veličin s normálním rozložením, bude mít podle 9.23. také normální rozložení se střední hodnotou

$$E(Y) = 2E(X_1) - 3E(X_2) + 5 = 2 \cdot 2 - 3 \cdot (-1) + 5 = 12$$

a rozptylem

$$D(Y) = 4D(X_1) + 9D(X_2) = 4 \cdot 4 + 9 \cdot 9 = 97.$$

$Y \sim N(12, 97) \Rightarrow U = \frac{Y-12}{\sqrt{97}} \sim N(0,1)$, tedy $Y = \sqrt{97} U + 12$. Podle 9.5. a) dostáváme

$$K_{0,99}(Y) = \sqrt{97} \cdot u_{0,99} + 12 = 34,9119$$

4. V zásilce 15 výrobků je 5 nekvalitních. Náhodná veličina X udává počet nekvalitních výrobků mezi čtyřmi náhodně vybranými výrobky. Vypočtete její střední hodnotu a rozptyl, jestliže výběr byl proveden a) s vracením, b) bez vracení. (Návod: v bodě (a) má X binomické rozložení, v bodě (b) hypergeometrické.)

Řešení:

ad a) Provádíme $n = 4$ nezávislé opakované pokusy, pravděpodobnost úspěchu v každém

z nich je $\vartheta = \frac{5}{15} = \frac{1}{3}$, tedy $X \sim \text{Bi}\left(4, \frac{1}{3}\right)$.

Podle 9.13. c) dostáváme $E(X) = n\vartheta = \frac{4}{3}$, $D(X) = n\vartheta(1 - \vartheta) = \frac{4}{3} \cdot \frac{2}{3} = \frac{8}{9}$.

ad b) Celkový počet výrobků je $N = 15$, mezi nimi je $M = 5$ nekvalitních, náhodně vybíráme $n = 4$ výrobky. Tedy $X \sim \text{Hg}(15, 5, 4)$.

Podle 9.13. e) dostáváme $E(X) = \frac{M}{N}n = \frac{4}{3}$, $D(X) = \frac{Mn}{N} \cdot \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1} = \frac{44}{63}$

5. Sledovaná železniční trasa vykazuje velké nerovnosti, takže zatížení jednotlivé vozové nápravy náhodně kolísá, teoreticky spojitým způsobem. Prakticky jsou známy jen částečné informace, takže uvažujeme o diskrétní náhodné veličině X (náhodné zatížení v tunách) s pravděpodobnostní funkcí $\pi(x) = 0,15$ pro $x=6$, $\pi(x) = 0,65$ pro $x=30$, $\pi(x) = 0,2$ pro $x=70$, $\pi(x) = 0$ jinak. Při kalkulaci nákladů se ekonom zajímá o střední opotřebením náprav dané vzorcem $Y = 1,15 X^2$. Vypočítejte střední hodnotu opotřebením.

Řešení:

$$E(Y) = E(1,15X^2) = 1,15E(X^2) = 1,15 \sum_{x=-\infty}^{\infty} x^2 \pi(x) = 1,15(6^2 \cdot 0,15 + 30^2 \cdot 0,65 + 70^2 \cdot 0,2) = 1805,96$$

6. Počet různých druhů zboží, které zákazník nakoupí při jedné návštěvě obchodu, je náhodná veličina X . Dlouhodobým sledováním bylo zjištěno, že X nabývá hodnot 0, 1, 2, 3, 4 s pravděpodobnostmi 0,25, 0,55, 0,11, 0,07 a 0,02.

- Najděte distribuční funkci náhodné veličiny X a nakreslete její graf.
- Vypočítejte střední hodnotu náhodné veličiny X .
- Vypočítejte rozptyl náhodné veličiny X .

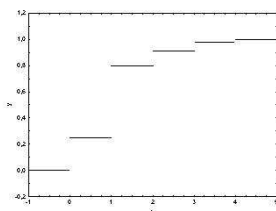
Řešení:

ad a)

$$x \in (-\infty, 0): \Phi(x) = 0, x \in \langle 0, 1): \Phi(x) = 0,25, x \in \langle 1, 2): \Phi(x) = 0,8, x \in \langle 2, 3): \Phi(x) = 0,91,$$

$$x \in \langle 3, 4): \Phi(x) = 0,98, x \in \langle 4, \infty): \Phi(x) = 1$$

Graf:



ad b)

$$E(X) = \sum_{x=-\infty}^{\infty} x\pi(x) = 0 \cdot 0,25 + 1 \cdot 0,55 + 2 \cdot 0,1 + 3 \cdot 0,07 + 4 \cdot 0,02 = 1,06$$

ad c)

$$\begin{aligned} D(X) &= E(X^2) - [E(X)]^2 = \\ &= \sum_{x=-\infty}^{\infty} x^2 \pi(x) - [E(X)]^2 = 0^2 \cdot 0,25 + 1^2 \cdot 0,55 + 2^2 \cdot 0,1 + 3^2 \cdot 0,07 + 4^2 \cdot 0,02 - 1,06^2 = 0,8164 \end{aligned}$$

7. Střelec střelí 3× nezávisle na sobě do terče. Při každém výstřelu se trefí s pravděpodobností 3/4. Za zásah získá 2 body, jinak ztratí 2 body. Vypočtěte střední hodnotu a rozptýl počtu získaných bodů.

Řešení:

X – počet získaných bodů, $X \in \{-6, -2, 2, 6\}$

X = -6, když střelec 3x nezasáhne. Pak $\pi(-6) = \left(\frac{1}{4}\right)^3 = \frac{1}{64}$.

X = -2, když střelec 2x nezasáhne, 1x zasáhne. Pak $\pi(-2) = \binom{3}{1} \frac{3}{4} \left(\frac{1}{4}\right)^2 = \frac{9}{64}$.

X = 2, když střelec 1x nezasáhne, 2x zasáhne. Pak $\pi(2) = \binom{3}{2} \left(\frac{3}{4}\right)^2 \frac{1}{4} = \frac{27}{64}$.

X = 6, když střelec 3x zasáhne. Pak $\pi(6) = \left(\frac{3}{4}\right)^3 = \frac{27}{64}$.

$$E(X) = \sum_{x=-\infty}^{\infty} x\pi(x) = -6 \cdot \frac{1}{64} - 2 \cdot \frac{9}{64} + 2 \cdot \frac{27}{64} + 6 \cdot \frac{27}{64} = \frac{192}{64} = 3$$

$$\begin{aligned} D(X) &= E(X^2) - [E(X)]^2 = \\ &= \sum_{x=-\infty}^{\infty} x^2 \pi(x) - [E(X)]^2 = (-6)^2 \cdot \frac{1}{64} + (-2)^2 \cdot \frac{9}{64} + 2^2 \cdot \frac{27}{64} + 6^2 \cdot \frac{27}{64} - 3^2 = 18 - 9 = 9 \end{aligned}$$

8. Uvažme rodinu se třemi dětmi. Předpokládáme, že pravděpodobnost narození chlapce i dívky je stejná. Náhodná veličina X udává počet dívek v této rodině (má binomické rozložení), transformovaná náhodná veličina $Y = -100X^2 + 300X + 500$ udává roční náklady (v dolarech) na ošacení dětí. Vypočtěte střední hodnotu náhodné veličiny Y.

Řešení:

$X \sim \text{Bi}(3, \frac{1}{2})$, tedy podle 9.13. c) $E(X) = \frac{3}{2}$, $D(X) = \frac{3}{4} = E(X^2) - [E(X)]^2$, tedy $E(X^2) = 3$.

Podle 9.17 $E(Y) = -100 \cdot E(X^2) + 300 \cdot E(X) + 500 = 650$

9. Náhodná veličina X udává příjem manžela (v tisících dolarů) a náhodná veličina Y udává příjem manželky (v tisících dolarů). Je známa simultánní pravděpodobnostní funkce $\pi(x,y)$ diskrétního náhodného vektoru (X,Y): $\pi(10,10) = 0,2$, $\pi(10,20) = 0,04$, $\pi(10,30) = 0,01$, $\pi(10,40) = 0$, $\pi(20,10) = 0,1$, $\pi(20,20) = 0,36$, $\pi(20,30) = 0,09$, $\pi(20,40) = 0$, $\pi(30,10) = 0$, $\pi(30,20) = 0,05$, $\pi(30,30) = 0,1$, $\pi(30,40) = 0$, $\pi(40,10) = 0$, $\pi(40,20) = 0$, $\pi(40,30) = 0$, $\pi(40,40) = 0,05$, $\pi(x,y) = 0$ jinak.

- a) Vypočtete korelační koeficient náhodných veličin X, Y .
- b) Vypočtete střední hodnotu a směrodatnou odchylku náhodné veličiny $Z = 0,1X + 0,2Y$, která vyjadřuje příspěvek obou manželů na důchod. (Náhodná veličina Z vyjadřuje, že příspěvek na důchod činí 10% manželova platu a 20% manželčina platu.)

Řešení:

Hodnoty simultánní pravděpodobnostní funkce i obou marginálních pravděpodobnostních funkcí zapíšeme do kontingenční tabulky:

x	y				$\pi_1(x)$
	10	20	30	40	
10	0,2	0,04	0,01	0	0,25
20	0,1	0,36	0,09	0	0,55
30	0	0,05	0,1	0	0,15
40	0	0	0	0,05	0,05
$\pi_2(y)$	0,3	0,45	0,2	0,05	1

ad a) Pro výpočet koeficientu korelace musíme znát kovarianci náhodných veličin X, Y a směrodatné odchylky těchto veličin. Nejprve vypočteme střední hodnoty a rozptyly:

$$E(X) = \sum_{x=-\infty}^{\infty} x\pi_1(x) = 10 \cdot 0,25 + 20 \cdot 0,55 + 30 \cdot 0,15 + 40 \cdot 0,05 = 20$$

$$E(Y) = \sum_{y=-\infty}^{\infty} y\pi_2(y) = 10 \cdot 0,3 + 20 \cdot 0,45 + 30 \cdot 0,2 + 40 \cdot 0,05 = 20$$

$$D(X) = E(X^2) - [E(X)]^2 = \sum_{x=-\infty}^{\infty} x^2\pi_1(x) - [E(X)]^2 =$$

$$= 10^2 \cdot 0,25 + 20^2 \cdot 0,55 + 30^2 \cdot 0,15 + 40^2 \cdot 0,05 - 20^2 = 460 - 400 = 60$$

$$D(Y) = E(Y^2) - [E(Y)]^2 = \sum_{y=-\infty}^{\infty} y^2\pi_2(y) - [E(Y)]^2 =$$

$$= 10^2 \cdot 0,3 + 20^2 \cdot 0,45 + 30^2 \cdot 0,2 + 40^2 \cdot 0,05 - 20^2 = 470 - 400 = 70$$

Nyní vypočteme kovarianci:

$$C(X, Y) = E(XY) - E(X)E(Y) = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} xy\pi(x, y) - E(X)E(Y) =$$

$$= 10 \cdot 10 \cdot 0,2 + 10 \cdot 20 \cdot 0,04 + 10 \cdot 30 \cdot 0,01 + 20 \cdot 10 \cdot 0,1 + 20 \cdot 20 \cdot 0,36 + 20 \cdot 30 \cdot 0,09 + 30 \cdot 20 \cdot 0,05 +$$

$$+ 30 \cdot 30 \cdot 0,1 + 40 \cdot 40 \cdot 0,05 - 20 \cdot 20 = 449 - 400 = 49$$

Dosadíme do vzorce pro výpočet koeficientu korelace:

$$R(X, Y) = \frac{C(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = \frac{49}{\sqrt{60}\sqrt{70}} = 0,76$$

ad b) $E(Z) = 0,1E(X) + 0,2E(Y) = 0,1 \cdot 20 + 0,2 \cdot 20 = 6$

$$D(Z) = 0,1^2 D(X) + 0,2^2 D(Y) + 2 \cdot 0,1 \cdot 0,2 \cdot C(X, Y) = 0,01 \cdot 60 + 0,04 \cdot 70 + 0,04 \cdot 49 = 5,36$$

$$\sqrt{D(Z)} = \sqrt{5,36} = 2,32$$

10. Náhodné veličiny X_1, X_2 mají kovarianci 12. Vypočtete kovarianci náhodných veličin $Y_1 = -8 + 11X_1, Y_2 = 6 - 4X_2$.

Řešení:

$$C(Y_1, Y_2) = C(-6 + 11X_1, 6 - 4X_2) = 11 \cdot (-4) \cdot C(X_1, X_2) = 11 \cdot (-4) \cdot 12 = -528$$

11. Náhodná veličina X udává výšku v metrech a náhodná veličina Y udává hmotnost v gramech. Jak se změní kovariance a koeficient korelace, jestliže výšku vyjádříme v cm a hmotnost v kg?

Řešení:

X ... výška v metrech, $100X$... výška v cm

Y ... hmotnost v gramech, $Y/1000$... hmotnost v kg

$$C\left(100X, \frac{Y}{1000}\right) = \frac{100}{1000} C(X, Y) = 0,1C(X, Y)$$

$$R\left(100X, \frac{Y}{1000}\right) = \frac{C\left(100X, \frac{Y}{1000}\right)}{\sqrt{D(100X)} \cdot \sqrt{D\left(\frac{Y}{1000}\right)}} = \frac{0,1C(X, Y)}{\sqrt{100^2 D(X)} \cdot \sqrt{\frac{1}{1000^2} D(Y)}} = \frac{C(X, Y)}{\sqrt{D(X)} \cdot \sqrt{D(Y)}} = R(X, Y)$$

Kovariance se 10x zmenší, koeficient korelace se nezmění.

12. Náhodná veličina X má střední hodnotu μ a směrodatnou odchylku σ . Kolik procent realizací této náhodné veličiny se bude nacházet v intervalu $(\mu - 2\sigma, \mu + 2\sigma)$?

Řešení:

K řešení této úlohy použijeme Čebyševovu nerovnost: $\forall t > 0 : P(|X - \mu| > t\sigma) \leq \frac{1}{t^2}$ neboli

$\forall t > 0 : P(|X - \mu| \leq t\sigma) > 1 - \frac{1}{t^2}$. V našem případě $t = 2$, tedy $1 - \frac{1}{t^2} = \frac{3}{4}$. Znamená to, že v intervalu $(\mu - 2\sigma, \mu + 2\sigma)$ se nachází aspoň 75% realizací náhodné veličiny X .

13. Použijte Čebyševovu nerovnost k odhadu pravděpodobnosti, že při 600 hodech kostkou padne šestka aspoň $75 \times$ a nejvýše $125 \times$.

Řešení:

X – počet šestek při 600 hodech kostkou, $X \sim \text{Bi}\left(600, \frac{1}{6}\right)$, $E(X) = 100 = \mu$,

$D(X) = \frac{500}{6} = \sigma^2$. Čebyševova nerovnost: $\forall t > 0 : P(|X - \mu| > t\sigma) \leq \frac{1}{t^2}$ neboli

$\forall t > 0 : P(|X - \mu| \leq t\sigma) > 1 - \frac{1}{t^2}$. V našem případě: $P(|X - 100| \leq 25) > 1 - \frac{1}{t^2}$, tudíž $25 = t\sigma$.

Odtud $t = \frac{25}{\sigma} = \frac{25}{\sqrt{\frac{500}{6}}} = 2,7386$ a $1 - \frac{1}{t^2} = 1 - \frac{1}{2,7386^2} = 0,8667$. Vidíme tedy, že

s pravděpodobností aspoň 0,8667 padne šestka při 600 hodech kostkou aspoň $75 \times$ a nejvýše $125 \times$.

Kapitola 11.

5. Předpokládáme, že velký ročník na vysoké škole má výsledky ze statistiky normálně rozloženy kolem střední hodnoty 72 bodů se směrodatnou odchylkou 9 bodů. Vypočtete pravděpodobnost, že

- náhodně vybraný student bude mít výsledek nad 80 bodů
- průměr výsledků náhodně vybraných 10 studentů bude nad 80 bodů.

Řešení:

ad a) X ... výsledek ze statistiky, $X \sim N(72, 9^2)$

$$\begin{aligned} P(X > 80) &= 1 - P(X \leq 80) = 1 - P\left(\frac{X - 72}{9} \leq \frac{80 - 72}{9}\right) = 1 - P(U \leq 0,89) = 1 - \Phi(0,89) = \\ &= 1 - 0,81327 = 0,18673 \end{aligned}$$

ad b) X_1, \dots, X_{10} ... výsledky ze statistiky 10 náhodně vybraných studentů, tvoří náhodný výběr z rozložení $N(72, 9^2)$.

$$\begin{aligned} P(M > 80) &= 1 - P(M \leq 80) = 1 - P\left(\frac{M - 72}{\frac{9^2}{\sqrt{10}}} \leq \frac{80 - 72}{\frac{9^2}{\sqrt{10}}}\right) = 1 - P\left(U \leq \sqrt{10} \cdot \frac{8}{9}\right) = 1 - \Phi(2,81) = \\ &= 1 - 0,99752 = 0,00248 \end{aligned}$$

6. Necht' X_1, \dots, X_{20} je náhodný výběr z $N(\mu, \sigma^2)$. Najděte čísla k_1, k_2 tak, aby platilo

$$P\left(\frac{S^2}{\sigma^2} < k_1\right) = 0,05 \text{ a } P\left(\frac{S^2}{\sigma^2} > k_2\right) = 0,05.$$

Řešení:

Využijeme toho, že statistika $K = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

$$P\left(\frac{S^2}{\sigma^2} < k_1\right) = P\left(\frac{(20-1)S^2}{\sigma^2} < (20-1)k_1\right) = P(K < 19k_1) = 0,05. \text{ Znamená to, že číslo } 19k_1 \text{ je}$$

kvantil $\chi^2_{0,05}(19)$. Ve statistických tabulkách najdeme $\chi^2_{0,05}(19) = 10,117$, tedy

$$k_1 = \frac{10,117}{19} = 0,532.$$

$$P\left(\frac{S^2}{\sigma^2} > k_2\right) = P\left(\frac{(20-1)S^2}{\sigma^2} > (20-1)k_2\right) = 1 - P(K \leq 19k_2) = 0,05, \text{ tedy } P(K \leq 19k_2) = 0,95.$$

Znamená to, že číslo $19k_2$ je kvantil $\chi^2_{0,95}(19)$. Ve statistických tabulkách najdeme $\chi^2_{0,95}(19) =$

$$30,144, \text{ tedy } k_2 = \frac{30,144}{19} = 1,587.$$

Kapitola 12.

3. Necht' X_1, \dots, X_n je náhodný výběr z rozložení $Rs(0,b)$, kde $b > 0$ je neznámý parametr.

Jsou definovány statistiky $T_1 = X_1 + \frac{1}{2}X_2 + \frac{1}{3}X_3 + \frac{1}{6}X_4$ a

$T_2 = \frac{1}{2}(X_1 + X_2 + X_3 + X_4)$. Ukažte, že T_1, T_2 jsou nestranné odhady parametru b a určete, který odhad je lepší.

Řešení:

Uvědomíme si, že střední hodnota náhodné veličiny s rozložením $Rs(0,b)$ je $\frac{b}{2}$ (viz 9.13. h).

Nyní vypočteme střední hodnoty obou odhadů T_1, T_2 .

$$\begin{aligned} E(T_1) &= E\left(X_1 + \frac{1}{2}X_2 + \frac{1}{3}X_3 + \frac{1}{6}X_4\right) = E(X_1) + \frac{1}{2}E(X_2) + \frac{1}{3}E(X_3) + \frac{1}{6}E(X_4) = \\ &= \frac{b}{2}\left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{6}\right) = b \end{aligned}$$

Statistika T_1 je tedy nestranný odhad parametru b .

$$\begin{aligned} E(T_2) &= E\left(\frac{1}{2}(X_1 + X_2 + X_3 + X_4)\right) = \frac{1}{2}(E(X_1) + E(X_2) + E(X_3) + E(X_4)) = \\ &= \frac{1}{2}\left(\frac{b}{2} + \frac{b}{2} + \frac{b}{2} + \frac{b}{2}\right) = b \end{aligned}$$

Statistika T_2 je tedy nestranný odhad parametru b .

Abychom posoudili, který z odhadů je lepší, vypočítáme jejich rozptyly. Přitom si uvědomíme, že rozptyl náhodné veličiny s rozložením $Rs(0,b)$ je $\frac{b^2}{12}$ (viz 9.13. h)

$$\begin{aligned} D(T_1) &= D\left(X_1 + \frac{1}{2}X_2 + \frac{1}{3}X_3 + \frac{1}{6}X_4\right) = D(X_1) + \frac{1}{4}D(X_2) + \frac{1}{9}D(X_3) + \frac{1}{36}D(X_4) = \\ &= \frac{b^2}{12}\left(1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{36}\right) = \frac{25b^2}{216} = 0,116b^2 \end{aligned}$$

$$\begin{aligned} D(T_2) &= D\left(\frac{1}{2}(X_1 + X_2 + X_3 + X_4)\right) = \frac{1}{4}(D(X_1) + D(X_2) + D(X_3) + D(X_4)) = \\ &= \frac{1}{4}\left(\frac{b^2}{12} + \frac{b^2}{12} + \frac{b^2}{12} + \frac{b^2}{12}\right) = \frac{b^2}{12} = 0,083b^2 \end{aligned}$$

Menší rozptyl má statistika T_2 , odhad T_2 je tedy lepší než T_1 .

7. Hloubka moře se měří přístrojem, jehož systematická chyba je nulová a náhodné chyby měření mají normální rozložení se směrodatnou odchylkou $\sigma = 1$ m. Kolik měření je nutno provést, aby se hloubka moře stanovila s chybou nejvýše $\pm 0,25$ m při riziku 0,05?

Řešení:

$$\text{Podle 12.9. (a) dostáváme: } 0,5 \geq h - d = m + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} - m + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} = 2 \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} \Rightarrow$$

$$n \geq \frac{4\sigma^2 u_{1-\alpha/2}^2}{0,16^2} = \frac{4 \cdot 1 \cdot u_{0,975}^2}{0,5^2} = \frac{4 \cdot 1,96^2}{0,5^2} = 61,47 \Rightarrow n \geq 62$$

8. U jistého měřicího zařízení má být posouzena jeho přesnost. Proto na něm byla nezávisle změřena délka téhož výrobku. Výsledky měření v cm byly: 15,15; 15,20; 15,04; 15,14; 15,22. Předpokládáme, že tyto výsledky jsou číselné realizace náhodného výběru rozsahu 5 z rozložení $N(\mu, \sigma^2)$. Sestrojte 95% empirický interval spolehlivosti pro rozptyl σ^2 .

Řešení:

Podle 12.9. (c) máme: $(d, h) = \left(\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}, \frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)} \right)$. Nejprve spočítáme $m = 15,15$,

$s^2 = 0,0049$. Ve statistických tabulkách najdeme $\chi^2_{1-\alpha/2}(n-1) = \chi^2_{0,975}(4) = 11,143$,

$\chi^2_{\alpha/2}(n-1) = \chi^2_{0,025}(4) = 0,484$. Po dosazení do vzorců pro dolní a horní mez dostaneme

$$d = \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)} = \frac{4 \cdot 0,0049}{11,143} = 0,0018, \quad h = \frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)} = \frac{4 \cdot 0,0049}{0,484} = 0,0405$$

Tedy $0,0018 \text{ cm}^2 < \sigma^2 < 0,0405 \text{ cm}^2$ s pravděpodobností aspoň 0,95.

9. Sponzor televizních pořadů pro děti chce vědět, kolik času stráví děti sledováním televize, protože na těchto informacích závisí typy a počty programů. Náhodným výběrem 100 dětí se zjistilo, že sledování televize věnují týdně průměrně 27,5 h se směrodatnou odchylkou 8 h. Za předpokladu, že počet hodin strávený za týden sledováním televize se řídí normálním rozložením, sestrojte 95% empirický interval spolehlivosti pro střední hodnotu počtu hodin strávených týdně sledováním televize.

Řešení:

X_1, \dots, X_{100} je náhodný výběr z rozložení $N(\mu, \sigma^2)$, kde parametry μ, σ^2 neznáme. Bylo spočteno, že $m = 27,5$, $s = 8$. Podle 12.9. (b) máme

$(d, h) = \left(m - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1), m + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) \right)$. V tabulkách najdeme $t_{1-\alpha/2}(n-1) = t_{0,975}(99) =$

1,96. Po dosazení do vzorců pro dolní a horní mez dostaneme

$$d = m - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) = 27,5 - \frac{8}{\sqrt{100}} 1,96 = 25,93,$$

$$h = m + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) = 27,5 + \frac{8}{\sqrt{100}} 1,96 = 29,07$$

Tedy $25,93 \text{ h} < \mu < 29,07 \text{ h}$ s pravděpodobností aspoň 0,95.

9. Na jisté velké americké univerzitě bylo v r. 1969 náhodně vybráno 5 profesorů a nezávisle na tom 5 profesorek a byl zjištěn jejich roční příjem (v tisících dolarů). Muži: 16, 19, 12, 11, 22, ženy: 9, 12, 8, 10, 16. Předpokládáme, že uvedené údaje tvoří realizace dvou nezávislých náhodných výběrů z rozložení $N(\mu_1, \sigma_1^2)$ a $N(\mu_2, \sigma_2^2)$.

a) Sestrojte 95% empirický interval spolehlivosti pro podíl rozptylů příjmů mužů a žen.

b) Pokud bude uvedený interval spolehlivosti obsahovat 1, sestrojte 95% empirický interval spolehlivosti pro rozdíl středních hodnot příjmů mužů a žen. V opačném případě sestrojte aspoň přibližný interval spolehlivosti.

Řešení:

$$\text{ad a) Podle 12.13. (d, h) = \left(\frac{s_1^2 / s_2^2}{F_{1-\alpha/2}(n_1 - 1, n_2 - 1)}, \frac{s_1^2 / s_2^2}{F_{\alpha/2}(n_1 - 1, n_2 - 1)} \right)$$

Nejprve vypočítáme $m_1 = 16$, $m_2 = 11$, $s_1^2 = 21,5$, $s_2^2 = 10$. Dále ve statistických tabulkách najdeme $F_{1-\alpha/2}(n_1 - 1, n_2 - 1) = F_{0,975}(4, 4) = 9,6045$,

$$F_{\alpha/2}(n_1 - 1, n_2 - 1) = F_{0,025}(4, 4) = \frac{1}{F_{0,975}(4,4)} = \frac{1}{9,6045} = 0,1041.$$

Dosadíme do vzorců pro dolní a horní mez:

$$d = \frac{s_1^2 / s_2^2}{F_{1-\alpha/2}(n_1 - 1, n_2 - 1)} = \frac{21,5/10}{9,6045} = 0,22, h = \frac{s_1^2 / s_2^2}{F_{\alpha/2}(n_1 - 1, n_2 - 1)} = \frac{21,5/10}{0,1041} = 20,65$$

Vidíme tedy, že $0,22 < \frac{\sigma_1^2}{\sigma_2^2} < 20,65$ s pravděpodobností aspoň 0,95.

ad b) Podle 12.13.(b) máme

$$(d, h) = (m_1 - m_2 - s^* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{1-\alpha/2}(n_1+n_2-2), m_1 - m_2 + s^* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{1-\alpha/2}(n_1+n_2-2))$$

Vypočteme tedy $s^{*2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{4 \cdot 21,5 + 4 \cdot 10}{8} = 15,75$. V tabulkách najdeme

$t_{0,975}(8) = 2,306$. Dosadíme do vzorců pro dolní a horní mez:

$$d = m_1 - m_2 - s^* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{1-\alpha/2}(n_1+n_2-2) = 16 - 11 - \sqrt{15,75} \cdot \sqrt{\frac{1}{5} + \frac{1}{5}} \cdot 2,306 = -0,79,$$

$$h = m_1 - m_2 + s^* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{1-\alpha/2}(n_1+n_2-2) = 16 - 11 + \sqrt{15,75} \cdot \sqrt{\frac{1}{5} + \frac{1}{5}} \cdot 2,306 = 10,79.$$

Zjistili jsme, že $-0,79$ tisíc dolarů $< \mu_1 - \mu_2 < 10,79$ tisíc dolarů s pravděpodobností aspoň 0,95.

11. Pět mužů se rozhodlo, že budou hubnout. Zjistili svou hmotnost před zahájením diety a po ukončení diety.

Číslo osoby	1	2	3	4	5
Hmotnost před dietou	84	77,5	91,5	84,5	97,5
Hmotnost po dietě	78,5	73,5	88,5	80	97

Za předpokladu, že uvedené údaje jsou číselné realizace náhodného výběru rozsahu 5

z dvourozměrného normálního rozložení $N_2 \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right)$, sestrojte 95% empirický

interval spolehlivosti pro střední hodnotu úbytku hmotnosti.

Řešení:

Vypočteme realizace rozdílového náhodného výběru:

$$z_1 = 84 - 78,5 = 5,5, z_2 = 77,5 - 73,5 = 4, z_3 = 91,5 - 88,5 = 3, z_4 = 84,5 - 80 = 4,5, z_5 = 97,5 - 97 = 0,5. \text{ Odtud vypočteme } m = 3,5, s^2 = 3,625, s = 1,904.$$

Podle 12.9. (b) máme $(d, h) = (m - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1), m + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1))$. V tabulkách najdeme $t_{1-\alpha/2}(n-1) = t_{0,975}(4) = 2,7764$. Po dosazení do vzorců pro dolní a horní mez dostaneme

$$d = m - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) = 3,5 - \frac{1,904}{\sqrt{5}} 2,7764 = 1,14,$$

$$h = m + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) = 3,5 + \frac{1,904}{\sqrt{5}} 2,7764 = 5,84$$

Tedy $1,14 \text{ kg} < \mu_1 - \mu_2 < 5,84 \text{ kg}$ s pravděpodobností aspoň 0,95.

Kapitola 13.

7. Podle údajů na obalu čokolády by její čistá hmotnost měla být 125 g. Výrobce dostal několik stížností od kupujících, ve kterých tvrdili, že hmotnost čokolád je nižší než deklarovaných 125 g. Z tohoto důvodu oddělení kontroly náhodně vybralo 50 čokolád a zjistilo, že jejich průměrná hmotnost je 122 g a směrodatná odchylka 8,6 g. Za předpokladu, že hmotnost čokolád se řídí normálním rozložením, můžeme na hladině významnosti 0,01 považovat stížnosti kupujících za oprávněné?

Řešení:

Na hladině významnosti 0,01 testujeme nulovou hypotézu $H_0: \mu = 125 \text{ g}$ proti levostranné alternativě $H_1: \mu \leq 125 \text{ g}$. Je to úloha na jednovýběrový t-test.

a) Testování pomocí kritického oboru: Podle 13.9. (b) nulovou hypotézu zamítáme na hladině významnosti α , když realizace testové statistiky $t_0 = \frac{m - c}{\frac{s}{\sqrt{n}}} \leq -t_{1-\alpha}(n-1)$. V našem případě:

$$\frac{m - c}{\frac{s}{\sqrt{n}}} = \frac{122 - 125}{\frac{8,6}{\sqrt{50}}} = -2,4667, \quad t_{1-\alpha}(n-1) = t_{0,99}(49) = 2,3263. \text{ Protože } -2,4667 \leq -2,3263,$$

nulovou hypotézu zamítáme na hladině významnosti 0,01 a stížnosti kupujících považujeme za oprávněné.

b) Testování pomocí intervalu spolehlivosti: Podle 13.5. (b) sestojíme 99% empirický pravostranný interval spolehlivosti pro střední hodnotu μ při neznámém rozptylu σ^2 . Podle

$$12.9. (b) \text{ máme } h = m + \frac{s}{\sqrt{n}} t_{1-\alpha}(n-1) = 122 + \frac{8,6}{\sqrt{50}} 2,3263 = 124,83.$$

Protože $125 \notin (-\infty, 124,83)$, nulovou hypotézu zamítáme na hladině významnosti 0,01.

c) Testování pomocí p-hodnoty: Podle 13.5. (c) dostáváme $p = P(T_0 \leq t_0) = P(T_0 \leq -2,4667) = \Phi(-2,4667)$, kde $\Phi(x)$ je distribuční funkce Studentova rozložení se 49 stupni volnosti.

Pomocí statistického software zjistíme, že $p = 0,0086$. Protože $p \leq 0,01$, nulovou hypotézu zamítáme na hladině významnosti 0,01.

8. V restauraci "U bílého koníčka" měřili ve 20 případech čas obsluhy zákazníka. Výsledky v minutách: 6,8, 11,4, 7,6, 10,6, 9,8, 5,12, 13,10, 9,8, 7,11, 10,5. V restauraci "Zlatý lev" bylo dané pozorování uskutečněno v 15 případech s těmito výsledky: 9, 11, 10, 7, 6, 4, 8, 13, 5, 15, 8, 5, 6, 8, 7. Na hladině významnosti 0,05 testujte hypotézu, že střední hodnoty doby obsluhy jsou v obou restauracích stejné.

Řešení:

Na hladině významnosti 0,05 testujeme nulovou hypotézu $H_0: \mu_1 - \mu_2 = 0$ proti oboustranné alternativě $H_1: \mu_1 - \mu_2 \neq 0$. Je to úloha na dvouvýběrový t-test. Před provedením tohoto testu je však nutné pomocí F-testu ověřit platnost hypotézy o shodě rozptylů. Na hladině

významnosti 0,05 tedy testujeme $H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$ proti $H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$. Podle 13.9. (e) nulovou

hypotézu zamítáme na hladině významnosti α , jestliže $\frac{s_1^2}{s_2^2} \leq F_{\alpha/2}(n_1 - 1, n_2 - 1)$ nebo

$\frac{s_1^2}{s_2^2} \geq F_{1-\alpha/2}(n_1 - 1, n_2 - 1)$. Vypočteme $m_1 = 8,25$, $m_2 = 8,13$, $s_1^2 = 6,307$, $s_2^2 = 9,41$.

V našem případě $\frac{s_1^2}{s_2^2} = \frac{6,307}{9,41} = 0,6702$. V tabulkách najdeme

$$F_{\alpha/2}(n_1 - 1, n_2 - 1) = F_{0,025}(19,14) = \frac{1}{F_{0,975}(14,19)} = \frac{1}{2,6469} = 0,3778,$$

$$F_{1-\alpha/2}(n_1 - 1, n_2 - 1) = F_{0,975}(19,14) = 2,8607.$$

Protože 0,6702 nepatří do kritického oboru $\langle 0; 0,3778 \rangle \cup \langle 2,8607; \infty \rangle$, hypotézu o shodě rozptylů nezamítáme na hladině významnosti 0,05.

Nyní se vrátíme k dvouvýběrovému t-testu.

a) Testování pomocí kritického oboru: Podle 13.9. (c) nulovou hypotézu zamítáme na hladině významnosti α , když absolutní hodnota realizace testové statistiky

$$|t_0| = \frac{|m_1 - m_2 - c|}{s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geq t_{1-\alpha}(n_1 + n_2 - 2).$$

$$s_*^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{19 \cdot 6,307 + 14 \cdot 9,41}{33} = 7,623. \text{ V tabulkách najdeme } t_{0,975}(33) =$$

1,96. Dosadíme do vzorce pro výpočet absolutní hodnoty realizace testové statistiky

$$\frac{|m_1 - m_2 - c|}{s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{|8,25 - 8,13|}{\sqrt{7,623} \cdot \sqrt{\frac{1}{20} + \frac{1}{15}}} = 0,124. \text{ Protože } 0,124 < 1,96, \text{ nulovou hypotézu}$$

nezamítáme na hladině významnosti 0,05.

b) Testování pomocí intervalu spolehlivosti: Podle 12.13. (b) máme

$$(d, h) = (m_1 - m_2 - s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{1-\alpha/2}(n_1+n_2-2), m_1 - m_2 + s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{1-\alpha/2}(n_1+n_2-2))$$

$$d = 8,25 - 8,13 - \sqrt{7,623} \cdot \sqrt{\frac{1}{20} + \frac{1}{15}} 1,96 = -1,73,$$

$$h = 8,25 - 8,13 + \sqrt{7,623} \cdot \sqrt{\frac{1}{20} + \frac{1}{15}} 1,96 = 1,97.$$

Protože $0 \in (-1,73; 1,97)$, nulovou hypotézu nezamítáme na hladině významnosti 0,05.

c) Testování pomocí p-hodnoty: Podle 13.5 (c) dostáváme $p = 2 \min\{P(T_0 \leq t_0), P(T_0 \geq t_0)\} = 2 \min\{P(T_0 \leq 0,124), P(T_0 \geq 0,124)\} = 2 \min\{\Phi(0,124), 1 - \Phi(0,124)\}$, kde $\Phi(x)$ je distribuční funkce Studentova rozložení s počtem stupňů volnosti 33. Pomocí statistického

software získáme $\Phi(0,124) = 0,549$, tedy $p = 2 \cdot (1 - 0,549) = 0,902$. Protože $0,902 > 0,05$, nulovou hypotézu nezamítáme na hladině významnosti 0,05.

9. Na 10 automobilech stejného typu se testovaly dva druhy benzínu lišící se oktanovým číslem. U každého automobilu se při průměrné rychlosti 90 km/h měřil dojezd (tj. dráha, kterou ujede na dané množství benzínu) při použití každého z obou druhů benzínu. Výsledky:

č. auta	1	2	3	4	5	6	7	8	9	10
benzín A	17,5	20,0	18,9	17,9	16,4	18,9	17,2	17,5	18,5	18,2
benzín B	17,8	20,8	19,5	18,3	16,6	19,5	17,5	17,9	19,1	18,6

Za předpokladu, že dojezd se řídí normálním rozložením, testujte na hladině významnosti 0,05 hypotézu, že rozdíl středních hodnot dojezdu při dvou druzích benzínu se neliší.

Řešení:

Na hladině významnosti 0,05 testujeme nulovou hypotézu $H_0: \mu_1 - \mu_2 = 0$ proti oboustranné alternativě $H_1: \mu_1 - \mu_2 \neq 0$. Je to úloha na párový t-test. Přejdeme k rozdílovému náhodnému výběru, jehož realizace jsou: -0,3 -0,8 -0,6 -0,4 -0,2 -0,6 -0,3 -0,4 -0,6 -0,4. Vypočteme $m = -0,46$, $s = 0,1838$.

a) Testování pomocí kritického oboru:

Podle 13.9. (b) nulovou hypotézu zamítáme na hladině významnosti α , když absolutní

hodnota realizace testové statistiky $|t_0| = \frac{|m - c|}{\frac{s}{\sqrt{n}}} \geq t_{1-\alpha/2}(n-1)$

Po dosazení dostaneme: $|t_0| = \frac{|-0,46 - 0|}{\frac{0,1838}{\sqrt{10}}} = 7,91$.

V tabulkách najdeme $t_{1-\alpha/2}(n-1) = t_{0,975}(9) = 2,2622$. Protože absolutní hodnota realizace testové statistiky je větší nebo rovna 2,2622, nulovou hypotézu zamítáme na hladině významnosti 0,05. S rizikem omylu nanejvýš 5% jsme tedy prokázali že rozdíl středních hodnot dojezdu při dvou druzích benzínu se liší.

b) Testování pomocí intervalu spolehlivosti:

Podle 13.5. (b) sestojíme 95% empirický interval spolehlivosti pro střední hodnotu $\mu = \mu_1 - \mu_2$ při neznámém rozptylu σ^2 . Podle 12.9. (b) máme

$$d = m - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) = -0,46 - \frac{0,1838}{\sqrt{10}} 2,2622 = -0,59,$$

$$h = m + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) = -0,46 + \frac{0,1838}{\sqrt{10}} 2,2622 = -0,33.$$

Protože $0 \notin (-0,59, 0,33)$, nulovou hypotézu zamítáme na hladině významnosti 0,01.

c) Testování pomocí p-hodnoty:

Podle 13.5 (c) dostáváme

$$p = 2 \min\{P(T_0 \leq t_0), P(T_0 \geq t_0)\} = 2 \min\{P(T_0 \leq -7,91), P(T_0 \geq -7,91)\} =$$

$2 \min\{\Phi(-7,91), 1 - \Phi(-7,91)\}$, kde $\Phi(x)$ je distribuční funkce Studentova rozložení s počtem stupňů volnosti 9. Pomocí statistického software získáme $\Phi(-7,91) = 0,00001$, tedy $p = 0,00002$. Protože $0,00002 \leq 0,05$, nulovou hypotézu zamítáme na hladině významnosti 0,05.

10. Pevnost vlákna bavlněné příze lze pokládat za náhodnou veličinu s rozložením $N(\mu, \sigma^2)$. Je-li $\sigma^2 > 0,36 \text{ kg}^2$, vznikají potíže při tkaní. Při zkoušce 11 náhodně vybraných vláken byly zjištěny hodnoty jejich pevnosti a vypočten empirický rozptyl $s^2 = 0,92 \text{ kg}^2$. Na hladině významnosti 0,05 je třeba zjistit, zda je příze vyhovující.

Řešení:

Na hladině významnosti 0,05 testujeme nulovou hypotézu $H_0: \sigma^2 \leq 0,36 \text{ kg}^2$ proti pravostranné alternativě $H_1: \sigma^2 > 0,36 \text{ kg}^2$. Jde o test o rozptylu.

a) Testování pomocí kritického oboru:

Podle 13.9. (f) nulovou hypotézu zamítáme ve prospěch pravostranné alternativy, když

realizace testové statistiky $\frac{(n-1)s^2}{c} \geq \chi^2_{1-\alpha}(n-1)$. V našem případě

$$\frac{(n-1)s^2}{c} = \frac{10 \cdot 0,92}{0,36} = 25,556. \text{ V tabulkách najdeme kvantil } \chi^2_{0,95}(10) = 18,307. \text{ Protože}$$

$25,556 \geq 18,307$, nulovou hypotézu zamítáme na hladině významnosti 0,05.

b) Testování pomocí intervalu spolehlivosti:

Podle 13.5. (b) sestojíme 95% empirický levostranný interval spolehlivosti pro rozptyl σ^2 při neznámé střední hodnotě μ . Podle 12.9. (c) máme

$$d = \frac{(n-1)s^2}{\chi^2_{1-\alpha}(n-1)} = \frac{10 \cdot 0,92}{\chi^2_{0,95}(10)} = \frac{10 \cdot 0,92}{1,8125} = 5,076. \text{ Protože } 0,36 \notin (5,076; \infty), \text{ nulovou hypotézu}$$

zamítáme na hladině významnosti 0,05.

c) Testování pomocí p-hodnoty:

Podle 13.5. (c) máme $p = P(T_0 \geq t_0) = P(T_0 \geq 25,556) = 1 - P(T_0 \leq 25,556) = 1 - \Phi(25,556)$, kde $\Phi(x)$ je distribuční funkce Pearsonova rozložení $\chi^2(10)$. Pomocí statistického software najdeme $\Phi(25,556) = 0,9956$, tedy $p = 1 - 0,9956 = 0,0044$. Protože $0,0044 \leq 0,05$, nulovou hypotézu zamítáme na hladině významnosti 0,05.

11. Normálně rozložené náhodné veličiny představují výsledek měření tíže konstanty dvěma různými metodami a jejich neznámé směrodatné odchylky σ_1, σ_2 charakterizují nespolehlivost těchto metod způsobenou náhodnými chybami. Při realizaci dvou nezávislých náhodných výběrů rozsahu $n_1 = 25, n_2 = 31$ jsme získali empirické směrodatné odchylky $s_1 = 0,523, s_2 = 0,363$. Je možno na hladině významnosti 0,05 považovat obě metody za stejně spolehlivé?

Řešení:

Na hladině významnosti 0,05 testujeme nulovou hypotézu $H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$ proti oboustranné

alternativě $H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$.

a) Testování pomocí kritického oboru:

Podle 13.9. (e) nulovou hypotézu zamítáme na hladině významnosti α , jestliže realizace

testové statistiky $\frac{s_1^2}{s_2^2} \leq F_{\alpha/2}(n_1 - 1, n_2 - 1)$ nebo $\frac{s_1^2}{s_2^2} \geq F_{1-\alpha/2}(n_1 - 1, n_2 - 1)$. V našem případě:

$$\frac{s_1^2}{s_2^2} = \frac{0,523^2}{0,363^2} = 2,076. \text{ V tabulkách najdeme}$$

$$F_{\alpha/2}(n_1 - 1, n_2 - 1) = F_{0,025}(24,30) = \frac{1}{F_{0,975}(30,24)} = \frac{1}{2,209} = 0,4527,$$

$F_{1-\alpha/2}(n_1 - 1, n_2 - 1) = F_{0,975}(24,30) = 2,1359$. Protože realizace testového kritéria 2,076 nepatří do kritického oboru $\langle 0; 0,4527 \rangle \cup \langle 2,1359; \infty \rangle$, nulovou hypotézu nezamítáme na hladině významnosti 0,05. S rizikem omylu nanejvýš 5% tedy obě metody můžeme považovat za stejně spolehlivé.

b) Testování pomocí intervalu spolehlivosti:

Podle 12.13. (d) máme $(d, h) = \left(\frac{s_1^2 / s_2^2}{F_{1-\alpha/2}(n_1 - 1, n_2 - 1)}, \frac{s_1^2 / s_2^2}{F_{\alpha/2}(n_1 - 1, n_2 - 1)} \right)$, tedy

$$d = \frac{0,523^2 / 0,363^2}{2,1359} = 0,9719, \quad h = \frac{0,523^2 / 0,363^2}{0,4527} = 4,5854$$

Protože $1 \in (0,9719; 4,5854)$, nulovou hypotézu nezamítáme na hladině významnosti 0,05.

c) Testování pomocí p-hodnoty:

Podle 13.5 (c) dostáváme

$p = 2 \min\{P(T_0 \leq t_0), P(T_0 \geq t_0)\} = 2 \min\{P(T_0 \leq 2,076), P(T_0 \geq 2,076)\} = 2 \min\{\Phi(2,076), 1 - \Phi(2,076)\}$, kde $\Phi(x)$ je distribuční funkce Fischerova - Snedecorova rozložení s počtem stupňů volnosti čitatele 24 a jmenovatele 30. Pomocí statistického software získáme $\Phi(2,076) = 0,9705$, tedy $p = 2 \cdot (1 - 0,9705) = 0,059$. Protože $0,059 > 0,05$, nulovou hypotézu nezamítáme na hladině významnosti 0,05.