1. (18 bodů) Náhodný vektor (X, Y) má diskrétní rozdělení; hodnoty sdružené pravděpodobnostní funkce $p_{X,Y}(x, y)$ jsou uvedeny v tabulce.

$y\backslash x$	-1	0	1
0	$\frac{1}{4}$	0	$\frac{1}{5}$
1	0	$\frac{3}{10}$	$\frac{1}{4}$

a) Vypočtěte hodnotu kovariance $\mathrm{cov}(X, Y)$. *(9 bodů)*

b) Napište podmínku pro nezávislost náhodných veličin X a Y. *(2 body)*

c) Rozhodněte, zda jsou náhodné veličiny X a Y závislé či nezávislé. *(2 body)*

d) Odvoďte vzorec pro rozptyl $\mathrm{D}(X + Y)$ a kdy platí, že $\mathrm{D}(X + Y) = \mathrm{D}X + \mathrm{D}Y$. *(5 bodů)*

Řešení:

Nejdříve určíme marginální pravděpodobnostní funkce p_X, p_Y. Sečtením hodnot ve sloupcích a řádcích dostaneme

$y\backslash x$	-1	0	1	p_Y
0	$\frac{1}{4}$	0	$\frac{1}{5}$	$\frac{9}{20}$
1	0	$\frac{3}{10}$	$\frac{1}{4}$	$\frac{11}{20}$
p_X	$\frac{1}{4}$	$\frac{3}{10}$	$\frac{9}{20}$	

a) Hodnotu kovariance vypočteme ze vzorce

$$\mathrm{cov}(X, Y) = \mathrm{E}XY - \mathrm{E}X \cdot \mathrm{E}Y\,.$$

Postupně dostaneme

$$\mathrm{E}X = \sum_x x\, p_X(x) = -\frac{1}{4} + \frac{9}{20} = \frac{-5+9}{20} = \frac{1}{5}\,;$$

$$\mathrm{E}Y = \sum_y y\, p_Y(y) = \frac{11}{20}\,;$$

$$\mathrm{E}XY = \sum_{(x,y)} x\, y\, p_{(X,Y)}(x, y) = \frac{1}{4}\,.$$

$$\mathrm{cov}(X, Y) = \mathrm{E}XY - \mathrm{E}X \cdot \mathrm{E}Y = \frac{1}{4} - \frac{1}{5} \cdot \frac{11}{20} = \frac{25-11}{100} = \frac{14}{100} = \mathbf{0.14}\,.$$

b) Diskrétní náhodné veličiny X a Y jsou nezávislé, právě když je

$$p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$$

pro všechna x, y.

c) Pro $x = 0$ a $y = 0$ dostaneme, že

$$0 \neq \frac{3}{10} \cdot \frac{9}{20}\,,$$

takže náhodné veličiny X a Y jsou závislé. To potvrzuje i fakt, že $\operatorname{cov}(X,Y) \neq 0$.

d) Pro výpočet rozptylu dostaneme vzorec

$$\begin{aligned}\mathrm{D}(X+Y) &= \mathrm{E}\big((X+Y)^2\big) - \big(\mathrm{E}(X+Y)\big)^2 = \\ &= \mathrm{E}(X^2+2XY+Y^2) - \big((\mathrm{E}X)^2 + 2\mathrm{E}X\mathrm{E}Y + (\mathrm{E}Y)^2\big) = \\ &= \mathrm{E}X^2 - (\mathrm{E}X)^2 + \mathrm{E}Y^2 - (\mathrm{E}Y)^2 + 2\big(\mathrm{E}XY - \mathrm{E}X\mathrm{E}Y\big) = \\ &= \mathrm{D}X + \mathrm{D}Y + 2\operatorname{cov}(X,Y)\,.\end{aligned}$$

Rovnost $\mathrm{D}(X+Y) = \mathrm{D}X + \mathrm{D}Y$ nastává, právě když jsou náhodné veličiny X a Y nekorelované, tj. $\operatorname{cov}(X,Y) = 0$. Postačující (nikoli nutnou) podmínkou je, že jsou nezávislé.

2. (18 bodů) Porovnáváme hmotnost dvou skupin stejně starých dětí (v kg); naměřená data jsou:

$$(50, 66, 64, 52, 43, 55)\,, \qquad (49, 54, 58, 46, 50, 53, 53, 48, 44, 55)\,.$$

   a) Posuďte, zda rozptyly obou souborů lze považovat za stejné. *(5 bodů)*

   b) Za předpokladu dle bodu a) otestujte, zda střední hodnoty obou souborů lze považovat za stejné. *(8 bodů)*

   c) Uveďte použité předpoklady a jejich přiměřenost. *(3 body)*

   d) Co byste navrhli změnit, kdyby se měl test provést na vyšší hladině významnosti? *(2 body)*

   V testech uvažujte hladinu významnosti 5 %.

Řešení:

a) Test rovnosti rozptylů dvou normálních rozdělení; odhady středních hodnot, rozptylů a směrodatných odchylek:

$\bar{\boldsymbol{x}} = 55,\ s_{\boldsymbol{x}}^2 = \dfrac{380}{5} = 76,\ s_{\boldsymbol{x}} \doteq 8.72,\ m = 6,$

$\bar{\boldsymbol{y}} = 51,\ s_{\boldsymbol{y}}^2 = \dfrac{170}{9} \doteq 18.9,\ s_{\boldsymbol{y}} = 4.35,\ n = 10.$

$H_0' : s_{\boldsymbol{x}}^2 = s_{\boldsymbol{y}}^2, \qquad H_1' : s_{\boldsymbol{x}}^2 \neq s_{\boldsymbol{y}}^2.$

Podíl většího a menšího odhadu rozptylu $\dfrac{s_{\boldsymbol{x}}^2}{s_{\boldsymbol{y}}^2} = \dfrac{76}{18.9} \doteq 4.0$ porovnáme s $q_{F(5,9)}(0.975) \doteq 4.48$, hypotézu H_0' o rovnosti rozptylů nezamítáme.

b) Test rovnosti středních hodnot dvou normálních rozdělení se stejným rozptylem; společný odhad rozptylu a směrodatné odchylky:

$$s^2 = \frac{(m-1)\ s_{\boldsymbol{x}}^2 + (n-1)\ s_{\boldsymbol{y}}^2}{m+n-2} = \frac{550}{14} \doteq 39.3\,, \qquad s \doteq \sqrt{39.3} \doteq 6.27\,.$$

$H_0 : \mathrm{E}X = \mathrm{E}Y, \qquad H_1 : \mathrm{E}X \neq \mathrm{E}Y.$

$$t = \frac{\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}}{s\sqrt{\frac{1}{m} + \frac{1}{n}}} \doteq \frac{4}{6.27\sqrt{\frac{1}{6} + \frac{1}{10}}} \doteq 1.24$$

porovnáme s $q_{t(14)}(0.975) \doteq 2.14$, hypotézu H_0 o rovnosti středních hodnot nezamítáme.

c) Předpoklady: normální rozdělení (stejné uvnitř každého souboru), nezávislost, stejné rozptyly. Rozdělení není neomezené (to má minimální vliv) ani symetrické (to by mohlo výsledky ovlivnit).

d) Potřebovali bychom hlavně zvětšit rozsah obou souborů. Přesnější nebo opakované vážení by asi pomohlo jen málo.

3. (14 bodů) Znaky (A, B, C) jsou permutací stavů $(1, 2, 3)$ Markovova řetězce s maticí přechodu

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 0.1 & 0.8 & 0.1 \\ 0.5 & 0.5 & 0 \end{pmatrix}.$$

a) Odhadněte tuto permutaci z pozorované posloupnosti znaků $(B, C, C, C, A, A, B, A, C)$. *(8 bodů)*

b) Mějme nyní permutaci, která měla v části a) největší věrohodnost. Posloupnost znaků z části a) skončila ve stavu C. Najděte její nejpravděpodobnější pokračování z následujících možností (první uvedený stav C je počátečním stavem této posloupnosti):

1. (C, A, C, C, B),
2. (C, C, C, B, A),
3. (C, B, A, A, C).

(6 bodů)

Řešení:

Pravděpodobnost počátečního stavu není dána, nebudeme ji uvažovat a budeme počítat jen s pravděpodobnostmi přechodu.

a) Permutací je sice 6, ale 0 v pozici $(3, 3)$ znamená, že stav 3 se nemůže opakovat; musí mu tedy odpovídat znak B. Zbývají dvě možnosti.

1. Pokud je $A = 1$ a $C = 2$, pak pozorovaná posloupnost je $(3, 2, 2, 2, 1, 1, 3, 1, 2)$ a její věrohodnost

$$p_{32}\, p_{22}\, p_{22}\, p_{21}\, p_{11}\, p_{13}\, p_{31}\, p_{12} = 0.5 \cdot 0.8 \cdot \mathbf{0.8} \cdot 0.1 \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot 0.5 \cdot \frac{1}{3} = \frac{1}{3^3} \cdot 0.016 \doteq 0.000\,59\,.$$

2. Pokud je $A = 2$ a $C = 1$, pak pozorovaná posloupnost je $(3, 1, 1, 1, 2, 2, 3, 2, 1)$ a její věrohodnost

$$p_{31}\, p_{11}\, p_{11}\, p_{12}\, p_{22}\, p_{23}\, p_{32}\, p_{21} = 0.5 \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot 0.8 \cdot \mathbf{0.1} \cdot 0.5 \cdot 0.1 = \frac{1}{3^3} \cdot 0.002 \doteq 0.000\,074\,.$$

Věrohodnější je případ 1, $A = 1$, $C = 2$. Protože jsme potřebovali jen porovnat věrohodnosti, nemuseli jsme je vyčíslit a stačilo se podívat na činitele, v nichž se liší (vyznačeny tučně).

b) Po rozkódování dle části a) máme porovnat věrohodnosti následujících posloupností:

1. $(2, 1, 2, 2, 3)$:

$$p_{21}\, p_{12}\, p_{22}\, p_{23} = 0.1 \cdot \frac{1}{3} \cdot 0.8 \cdot 0.1 = \frac{1}{3} \cdot 0.008 \doteq 0.002\,67\,.$$

2. $(2, 2, 2, 3, 1)$:

$$p_{22}\, p_{22}\, p_{23}\, p_{31} = 0.8 \cdot 0.8 \cdot 0.1 \cdot 0.5 = 0.032\,.$$

3. $(2, 3, 1, 1, 2)$.

$$p_{23}\, p_{31}\, p_{11}\, p_{12} = 0.1 \cdot 0.5 \cdot \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9} \cdot 0.05 \doteq 0.005\,56\,.$$

Nejvěrohodnějším pokračováním je varianta 2.