

8 Dvourozměrné normální rozdělení $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- $(X_1, Y_1)^T, \dots, (X_n, Y_n)^T \dots$ dvojice nezávislých stejně rozdělených náhodných (iid) veličin
- $(X, Y)^T \dots$ dvourozměrný náhodný vektor
- Dvourozměrné normální rozdělení

- $(X, Y)^T \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)^T$
- hustota

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right\} \\ &= \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}} \exp\left\{-\frac{1}{2}\begin{pmatrix} x-\mu_1 \\ y-\mu_2 \end{pmatrix}^T \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}^{-1} \begin{pmatrix} x-\mu_1 \\ y-\mu_2 \end{pmatrix}\right\} \\ &= \frac{1}{2\pi(\det(\boldsymbol{\Sigma}))^{1/2}} \exp\left\{-\frac{1}{2}U^T\boldsymbol{\Sigma}^{-1}U\right\}, \end{aligned}$$

kde $(x, y)^T \in \mathbb{R}$, $(\mu_1, \mu_2)^T \in \mathbb{R}$, $\sigma_i^2 > 0$, $j = 1, 2$ a korelační koeficient $\rho \in \langle -1; 1 \rangle$.

- marginální rozdělení $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$.
- `mvtnorm::dmvnorm(x, Mu, Sigma)`, `mvtnorm::rmvnorm(n, Mu, Sigma)`
- Standardizované dvourozměrné normální rozdělení

- $(X, Y)^T \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- $\boldsymbol{\theta} = (0, 0, 1, 1, \rho)^T$
- hustota

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\} \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} \begin{pmatrix} x \\ y \end{pmatrix}\right\} \end{aligned}$$

kde $(x, y)^T \in \mathbb{R}$ a korelační koeficient $\rho \in \langle -1; 1 \rangle$.

- marginální rozdělení $X \sim N(0, 1)$, $Y \sim N(0, 1)$.
- Data
 - **Dataset 7: 13-two-samples-correlations-trunk.txt**
 - Máme k dispozici soubor hodnot délky trupu (rozdíl akrominální a spinální výšky těla) a délky dolní končetiny (spinální výška těla) mladých dospělých jedinců, převážně studentů vysokých škol z Brna a Ostravy (Králík, nepublikovaná data).
 - Datový soubor obsahuje následující proměnné
 - * `sex` ... pohlaví jedince (m - muž, f - žena);
 - * `lowex.L` ... délka dolní končetiny (v mm);
 - * `tru.L` ... délka trupu (v mm).

Příklad 8.1. Hustota dvourozměrného normálního modelu

Naprogramujte v \mathbb{R} funkci `dnorm2()`, jejímiž vstupy budou hodnoty $x, y, \mu_1, \mu_2, \sigma_1, \sigma_2$ a ρ a výstupem bude hodnota hustoty dvourozměrného normálního rozdělení $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ s parametry $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ a $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ v hodnotách x a y . Správnost funkce otestujte na výpočtu $f(x, y)$ pro (a) $x = 1, y = 1, \mu_1 = 0, \mu_2 = 0, \sigma_1^2 = 1, \sigma_2^2 = 1, \rho = 0.5$; (b) $x = 2.5, y = 1.5, \mu_1 = 3, \mu_2 = 2, \sigma_1^2 = 9, \sigma_2^2 = 16, \rho = 0.75$. Výsledky ověřte s výsledky funkce `dmvnorm()` z knihovny `mvtnorm`.

Řešení příkladu 8.1

```
f(1, 1) f(2.5, 1.5)
1 0.0943539 0.01977507
```

(a) $f(1, 1) = 0.0944$; (b) $f(2, 3) = 0.0198$.

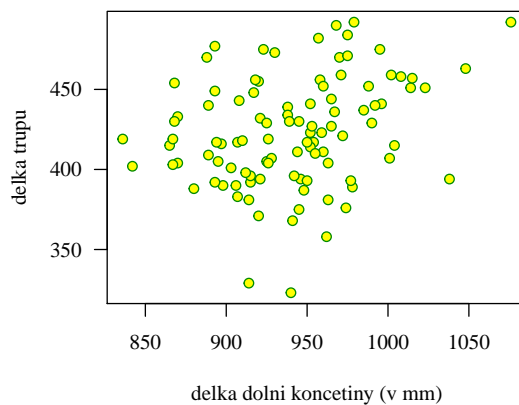
1
2
★

Příklad 8.2. Základní číselné charakteristiky dvojice spojitých znaků

Načtěte datový soubor `13-two-samples-correlations-trunk.txt`. Nechť náhodná proměnná X popisuje délku dolní končetiny (v mm) a náhodná proměnná Y popisuje délku trupu u žen. Pomocí tečkového diagramu vizualizujte vztah proměnných X a Y . Za předpokladu, že data pochází z dvourozměrného normálního rozdělení $(X, Y)^T \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ odhadněte hodnoty parametrů $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12}$ a ρ . Výsledky řádně interpretujte.

Řešení příkladu 8.2

```
          n   mean    sd    s12    rho
delka d.koncetiny 100 940.50 45.4712 441.3081 0.2853
delka trupu       100 423.17 34.0229 441.3081 0.2853
```



Obrázek 1: Dvourozměrný tečkový diagram pro délku dolní končetiny a délku trupu

Datový soubor obsahuje údaje o délce dolní končetiny a délce trupu 100 žen. Délka dolní končetiny se pohybuje okolo hodnoty 940.50 mm se směrodatnou odchylkou 45.47 mm. Délka trupu se pohybuje okolo hodnoty 423.17 mm se směrodatnou odchylkou 34.02 mm. Hodnota kovariance $s_{12} = 441.31$. Mezi délkou dolní končetiny a délkou trupu žen existuje nízký stupeň přímé lineární závislosti ($r_{12} = 0.2853$).

★

Příklad 8.3. Test dvourozměrné normality

Načtete datový soubor 13-two-samples-correlations-trunk.txt. Nechť náhodná proměnná X popisuje délku dolní končetiny a náhodná proměnná Y popisuje délku trupu žen. Na hladině významnosti $\alpha = 0.05$ testujte hypotézu o dvourozměrné normalitě vektoru $(X, Y)^T$. K otestování použijte Mardiův test.

Řešení příkladu 8.3

Na hladině významnosti $\alpha = 0.05$ testujeme H_0 : *Data pochází z dvourozměrného normálního rozdělení.* oproti H_1 : *Data nepochází z dvourozměrného normálního rozdělení.*

Mardiův test – sestává ze dvou testů:

(i) test šikmosti

- H_{0a} : Data nejsou kladně ani záporně vyšikmená.
- H_{1a} : Data jsou kladně nebo záporně vyšikmená.

(ii) test špičatosti

- H_{0b} : Data nejsou kladně ani záporně zešpičatělá.
- H_{1b} : Data jsou kladně nebo záporně zešpičatělá.

Poznámka: Náhodný výběr pochází z dvourozměrného normálního rozdělení, pokud nevykazuje kladné ani záporné zešikmení ani kladné nebo záporné zešpičatění.

	Test	HZ	p value	MVN
1	Henze-Zirkler	0.5045323	0.587225	YES

6
7

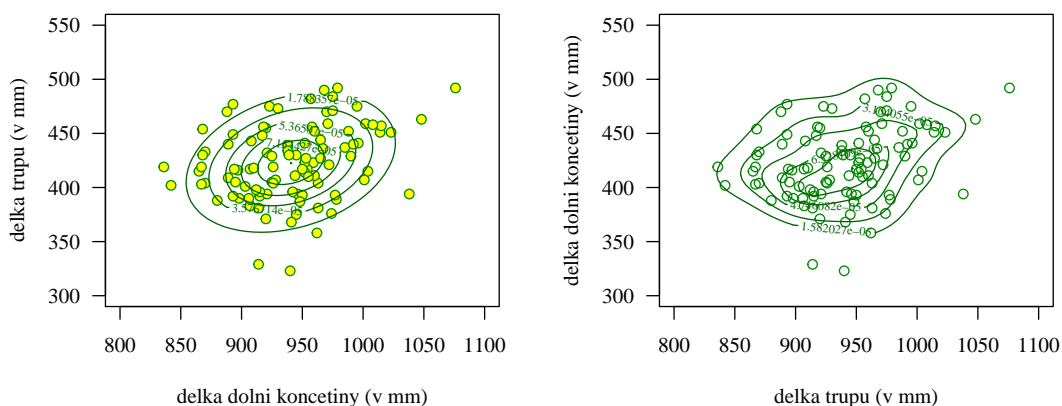
Protože p -hodnota testu šikmosti $p = 0.1769$ je větší než α , H_{0a} nezamítáme na hladině významnosti $\alpha = 0.05$. Protože p -hodnota testu špičatosti $p = 0.8360$ je větší než α , H_{0b} nezamítáme na hladině významnosti $\alpha = 0.05$. Data nejsou kladně ani záporně zešikmená ani zešpičatělá. Data pochází z dvourozměrného normálního rozdělení. ★

Příklad 8.4. Vizualizace dat z dvourozměrného normálního modelu

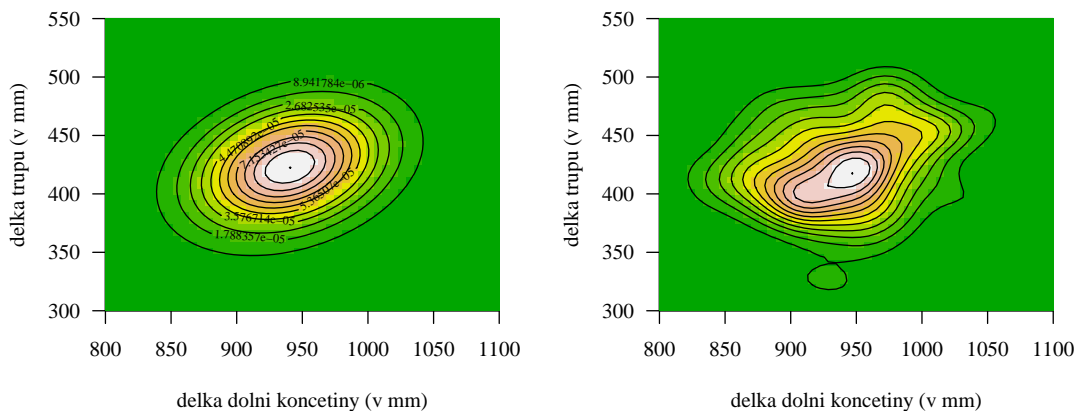
Načtete datový soubor 13-two-samples-correlations-trunk.txt. Nechť náhodná proměnná X popisuje délku dolní končetiny a náhodná proměnná Y popisuje délku trupu žen. Na základě řešení příkladů 8.2 a 8.3 předpokládáme, že data pochází z dvourozměrného normálního rozdělení $(X, Y)^T \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ s odhadem středních hodnot $\hat{\mu}_1 = 940.50$, $\hat{\mu}_2 = 423.17$, s odhadem rozptylů $\hat{\sigma}_1^2 = 45.47^2$ a $\hat{\sigma}_2^2 = 34.02^2$ a s odhadem korelačního koeficientu $\hat{\rho} = 0.2853$.

- Sestrojte tečkový diagram pro délku dolní končetiny a délku trupu. Tečkový diagram superponujte (i) konturami hustoty dvourozměrného normálního rozdělení (funkce `dnorm2()` + `contour()`); (ii) jádrovým odhadem dvourozměrné hustoty (funkce `kde2d()` z knihovny `MASS` + funkce `contour()`).
- Sestrojte (i) vrstevnicový diagram hustoty dvourozměrného normálního rozdělení délky dolní končetiny a délky trupu superponovaný svými konturami (`dnorm2()` + `image()` + `contour()`); (ii) vrstevnicový diagram jádrového odhadu hustoty délky dolní končetiny a délky trupu superponovaný svými konturami (`MASS::kde2d()` + `image()` + `contour()`).
- Sestrojte (i) 3D-diagram hustoty dvourozměrného normálního rozdělení délky dolní končetiny a délky trupu (`dnorm2()` + `persp()`); (ii) 3D-diagram jádrového odhadu hustoty dvourozměrného normálního rozdělení délky dolní končetiny a délky trupu (`MASS::kde2d()` + `persp()`).

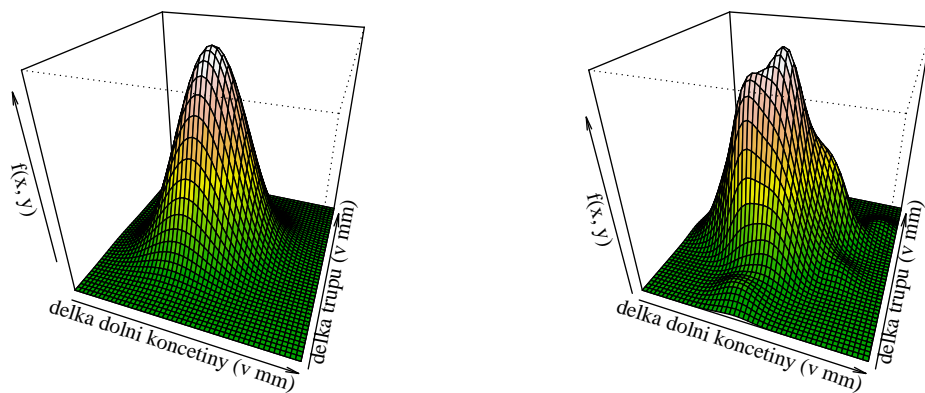
Řešení příkladu 8.4



Obrázek 2: Dvourozměrný tečkový diagram pro délku dolní končetiny a délku trupu žen superponovaný (a) konturami hustoty dvourozměrného normálního rozdělení (vlevo); (b) jádrovým odhadem hustoty dvourozměrného normálního rozdělení (vpravo)



Obrázek 3: (a) Hustota dvourozměrného normálního rozdělení pro délku dolní končetiny a délku trupu žen superponovaná konturami (vlevo); (b) jádrový odhad hustoty dvourozměrného normálního rozdělení superponovaný konturami (vpravo)



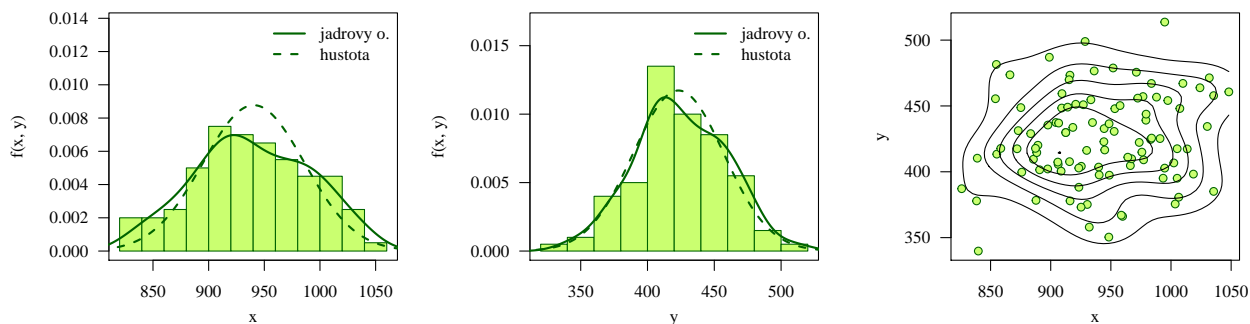
Obrázek 4: (a) 3D-diagram hustoty normálního rozdělení pro délku dolní končetiny a délku trupu (vlevo); (b) 3D-diagram jádrového odhadu hustoty dvourozměrného normálního rozdělení (vpravo)

★

Příklad 8.5. Simulace dat z dvourozměrného normálního rozdělení

Nasimulujte data $(X, Y)^T$ z dvourozměrného normálního rozdělení s parametry $\mu_1 = 940.50$, $\mu_2 = 423.17$, $\sigma_1^2 = 45.47^2$, $\sigma_2^2 = 34.02^2$ a $\rho = 0.2853$; $n = 100$. Simulaci pseudonáhodných čísel z $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ proveďte použitím funkce `rmvnorm()` z knihovny `mvtnorm`. Pro nasimulovaná data vykrestete (a) histogram náhodné veličiny X a histogram náhodné veličiny Y , přičemž každý histogram superponujte křivkou marginálního normálního rozdělení $N(\mu_i, \sigma_i^2)$, $i = 1, 2$ a křivkou jádrového odhadu hustoty; (b) tečkový diagram náhodných veličin X a Y , přičemž graf superponujte konturami jádrového odhadu dvourozměrné hustoty.

Řešení příkladu 8.5



Poznámka: Simulaci dat z dvourozměrného normálního rozdělení lze dále provést

1. pomocí funkce `rmvnorm()` z knihovny `MASS`
2. použitím funkce `rnorm()` a následujícího algoritmu: Nechť $U_1 \sim N(0, 1)$ a $U_2 \sim N(0, 1)$; potom $(X, Y)^T \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, kde $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ je vektor středních hodnot a σ_1^2 a σ_2^2 a ρ jsou parametry kovarianční matice $\boldsymbol{\Sigma}$, přičemž síla lineárního vztahu X a Y je daná velikostí a znaménkem ρ ; $X = \sigma_1 U_1 + \mu_1$ a $Y = \sigma_2(\rho U_1 + \sqrt{1 - \rho^2} U_2) + \mu_2$.

V případě zájmu si můžete simulaci těmito dvěma způsoby také vyzkoušet.

★