5 Spojité náhodné veličiny

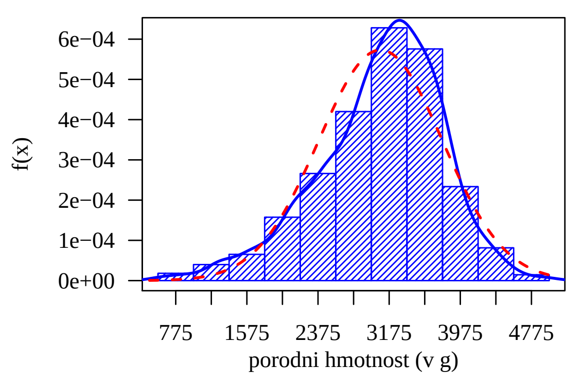
5.1 Normální rozdělení $N(\mu, \sigma^2)$

- $X_1, \ldots, X_n$ ... nezávislé náhodné veličiny
- Normální rozdělení
  - $X \sim N(\mu, \sigma^2)$
  - hustota
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}.$$
  - vlastnosti: $\mathrm{E}[X] = \mu$; $\mathrm{Var}[X] = \sigma^2$
  - dnorm(x, mu, sigma), pnorm(x, mu, sigma), qnorm(alpha, mu, sigma)
- Standardizované normální rozdělení
  - $X \sim N(0, 1)$
  - hustota
$$f(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad x \in \mathbb{R}.$$
  - vlastnosti: $\mathrm{E}[X] = 0$; $\mathrm{Var}[X] = 1$
  - dnorm(x), pnorm(x), qnorm(alpha)
- Vlastnosti normálního rozdělení
  - **Věta 1:** Nechť $X_1, \ldots, X_n$ jsou nezávislé náhodné veličiny z normálního rozdělení $N(\mu, \sigma^2)$. Potom náhodná veličina $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Příklad 5.1. Výpočet parametrů μ a σ normálního rozdělení
Mějme datový soubor 17-anova-newborns-2.txt obsahujícího údaje o porodní hmotnosti novorozenců v jedné okresní nemocnici za období jednoho roku (Alánová, 2008). Za předpokladu, že náhodná veličina X popisující porodní hmotnost novorozenců pochází z normálního rozdělení $N(\mu, \sigma^2)$ odhadněte parametr střední hodnoty μ a rozptylu σ^2. Finální rozdělení porovnejte s naměřenými údaji.

Řešení příkladu 5.1

```
  edu.M prch.N sex.C weight.C weight.K
1     2      0     m     3470        2
2     2      0     m     3240        2
3     2      0     f     2980        2
```



Interpretace výsledků: Náhodná veličina X popisující porodní hmotnost novorozenců pochází z normálního rozdělení se střední hodnotou $\mu =$.................... a rozptylem $\sigma^2 =$ ...., tj. $X \sim N($.................... ; ....$)$.
★

Příklad 5.2. Výpočet pravděpodobností na základě normálního rozdělení
Za předpokladu, že porodní hmotnost novorozenců pochází z normálního rozdělení $N(3078.027, 696^2)$, vypočítejte pravděpodobnost, že porodní hmotnost novorozence bude (a) menší než 3800 g; (b) v rozmezí 2500–4200 g; (c) větší než 4000 g, (d) rovná 2100 g.

Řešení příkladu 5.2

```
[1] 0.8502061                                                5
```

```
[1] 0.7433938                                                6
```

```
[1] 0.09263967                                               7
```

Interpretace výsledků: Pravděpodobnost, že porodní hmotnost novorozenců bude menší než 3800 g, je ....%. Pravděpodobnost, že porodní hmotnost novorozenců bude v rozmezí 2500–4200 mm, je ....%. Pravděpodobnost, že porodní hmotnost novorozenců bude větší než 4000 mm, je ....%. Pravděpodobnost, že porodní hmotnost novorozenců bude rovná 2100 g, je ....%, protože data pochází z normálního rozdělení, což je .... typ rozdělení, proto $\Pr(X = 2100) =$ ....

★

Příklad 5.3. Výpočet pravděpodobností na základě normálního rozdělení
Za předpokladu, že porodní hmotnost novorozenců pochází z normálního rozdělení $N(3078.027, 696^2)$, vypočítejte pravděpodobnost, že průměrná porodní hmotnost pěti novorozenců bude (a) menší než 3800 g; (b) v rozmezí 2500–4200 g; (c) větší než 4000 g; (d) rovná 2100 g.

Řešení příkladu 5.3

```
[1] 0.9898164                                                8
```

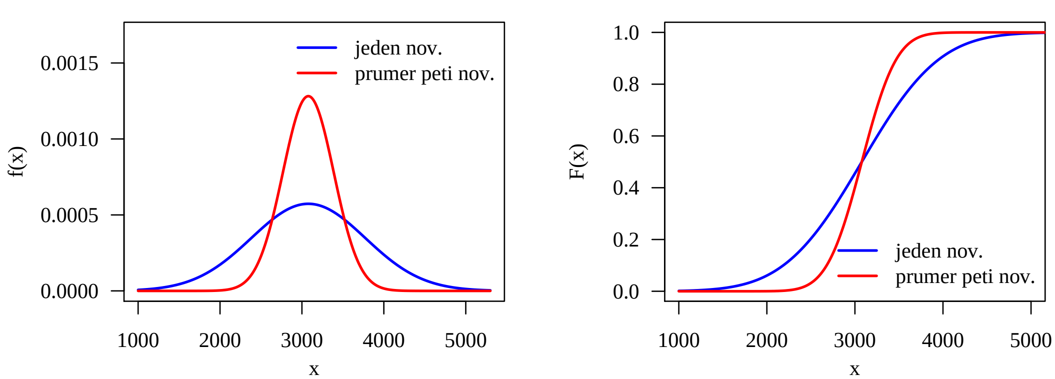
```
[1] 0.9681918                                                9
```

```
[1] 0.001527937                                              10
```

Interpretace výsledků: Pravděpodobnost, že průměrná porodní hmotnost pěti novorozenců bude menší než 3800 g je ....%. Pravděpodobnost, že průměrná porodní hmotnost pěti novorozenců bude v rozmezí 2500–4200 g je ....%. Pravděpodobnost, že průměrná porodní hmotnost pěti novorozenců bude větší než 4000 g je ....%. Pravděpodobnost, že průměrná porodní hmotnost pěti novorozenců bude rovná 2100 g je ....%, protože data pochází z normálního rozdělení, což je .... typ rozdělení, proto $\Pr(X = 2100) =$ .... ★

Příklad 5.4. Graf hustoty a distribuční funkce normálního rozdělení
Nakreslete graf hustoty a distribuční funkce náhodné veličiny $X \sim N(3078.027, 696^2)$ popisující porodní hmotnost jednoho novorozence a porovnejte je s křivkami hustoty a distribuční funkce náhodné veličiny $\bar{X} \sim N(3078.027, \frac{696^2}{5})$ popisující průměrnou porodní hmotnost pěti novorozenců.

Řešení příkladu 5.4



★

5.2 Dvourozměrné normální rozdělení $\mathbf{N}_2(\boldsymbol{\mu}, \Sigma)$

- $(X_1, Y_1)^T, \ldots, (X_n, Y_n)^T$ …dvojice nezávislých stejně rozdělených náhodných veličin
- $(X, Y)^T$ …dvourozměrný náhodný vektor
  - $(X, Y)^T \sim \mathbf{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
    - $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ …vektor středních hodnot
    - $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$ …varianční matice
  - $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)^T$, kde $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 > 0$, $\rho \in \langle -1; 1 \rangle$
  - hustota
$$f(x, y) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)}$$
  - vlastnosti $\mathrm{E}[(X, Y)^T] = \boldsymbol{\mu}$; $\mathrm{Var}[(X, Y)] = \boldsymbol{\Sigma}$
  - marginální rozdělení $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$.
- Grafická vizualizace dat
  - dvourozměrný tečkový diagram superponovaný konturovým diagramem
  - 3D-graf

Dataset: 03-paired-means-clavicle2.txt
Datový soubor 03-paired-means-clavicle2.txt obsahuje osteometrické údaje o délkách klíčních kostí na pravé a levé straně těla v párovém uspořádání. Data pochází z anglického souboru dokumentovaných skeletů (Parsons, 1916).

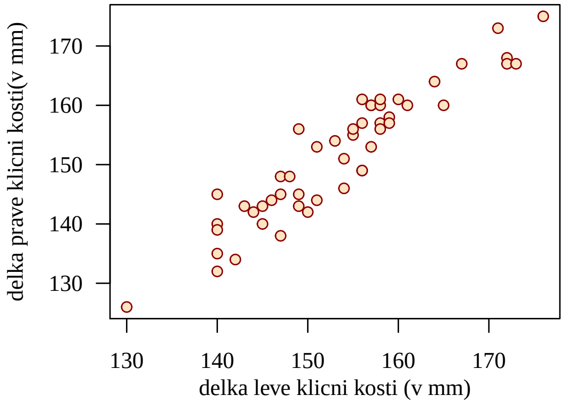
Popis proměnných v datasetu:

- id …ID jedince;
- sex …pohlaví jedince (m - muž, f - žena);
- length.L …délka levé klíční kosti (v mm);
- length.R …délka pravé klíční kosti (v mm).

Příklad 5.5. Výpočet parametrů $\boldsymbol{\mu}$ a $\boldsymbol{\Sigma}$ dvourozměrného normálního rozdělení
Načtěte datový soubor 03-paired-means-clavicle2.txt. Nechť náhodná veličina X popisuje délku levé klíční kosti a náhodná veličina Y popisuje délku pravé klíční kosti u mužů. Pomocí tečkového diagramu vizualizujte vztah délky levé a pravé klíční kosti. Za předpokladu, že data pochází z dvourozměrného normálního rozdělení $(X, Y)^T \sim \mathbf{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ odhadněte hodnoty parametrů μ_1, μ_2, σ_1^2, σ_2^2 a ρ a stanovte tvar vektoru středních hodnot a varianční matice.

Řešení příkladu 5.5

```
  id sex length.R length.L
1 66   m      126      130
2 69   m      158      159
3 71   m      153      151
4 72   m      145      147
```




```
                mean      sd    rho
leva strana  153.60  9.9468 0.9371
prava strana 151.74 10.9969 0.9371
```

Interpretace výsledků: Náhodný vektor $(X,Y)^T$ popisující délku klíční kosti z levé a pravé strany u mužů pochází z dvourozměrného normálního rozdělení s vektorem středních hodnot $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$, kde $\mu_1 =$ .... mm a $\mu_2 =$ .... mm a s varianční maticí $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, kde $\sigma_1 =$ .... mm, $\sigma_2 =$............................. mm a $\rho =$ .... Délka klíční kosti z levé strany u mužů pochází marginálně z normálního rozdělení se střední hodnotou $\mu_1 =$....................... mm a směrodatnou odchylkou $\sigma_1 =$ .... mm. Délka klíční kosti z pravé strany u mužů pochází marginálně z normálního rozdělení se střední hodnotou $\mu_2 =$ .... mm a směrodatnou odchylkou $\sigma_2 =$ .... mm. ★

Příklad 5.6. Parametrické a neparametrické odhady dat z $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Načtěte datový soubor 03-paired-means-clavicle2.txt. Za předpokladu, že náhodný vektor $(X,Y)^T$ popisující délku klíční kosti z levé a pravé strany u mužů pochází z dvourozměrného normálního rozdělení, tj. $(X,Y)^T \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ s odhadem středních hodnot $\widehat{\mu}_1 = 153.6$, $\widehat{\mu}_2 = 151.74$, rozptylů $\widehat{\sigma}_1^2 = 9.95^2$ a $\widehat{\sigma}_2^2 = 11^2$ a odhadem korelačního koeficientu $\widehat{\rho} = 0.9371$.

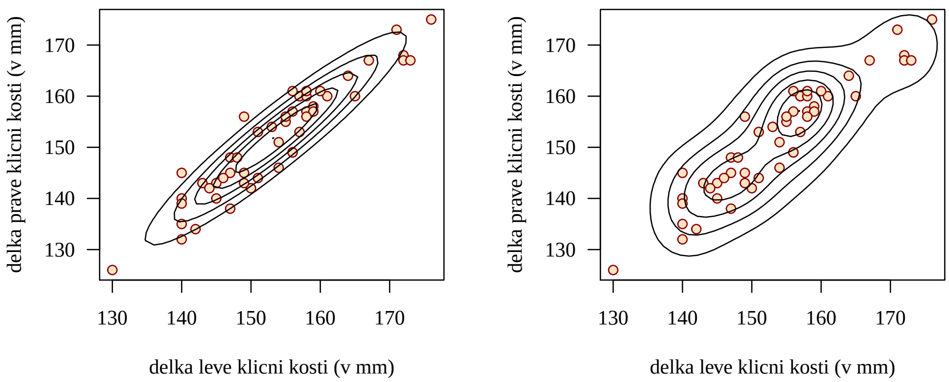
a. sestrojte tečkový diagram délky klíční kosti z levé a pravé strany superponovaný teoretickými konturami teoretického dvourozměrného normálního rozdělení;

b. sestrojte tečkový diagram délky klíční kosti z levé a pravé strany superponovaný konturami jádrového odhadu hustoty;

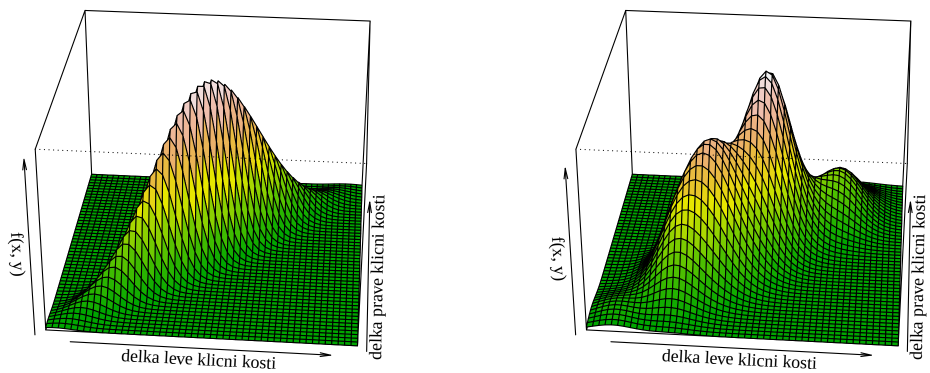
c. sestrojte 3D-diagram hustoty teoretického dvourozměrného normálního rozdělení délky klíční kosti z levé a pravé strany;

d. sestrojte 3D-diagram jádrového odhadu hustoty délky klíční kosti z levé a pravé strany.

Řešení příkladu 5.6



Obrázek 1: Dvourozměrný tečkový diagram délky klíční kosti z levé a pravé strany superponovaný konturami (a) teoretické hustoty (vlevo); (b) jádrového odhadu hustoty dvourozměrného normálního rozdělení (vpravo)



Obrázek 2: 3D diagram (a) teoretické hustoty; (b) jádrového odhadu hustoty (vpravo) dvourozměrného normálního rozdělení

Interpetace výsledků: Na základě grafické vizualizace předpokládáme, že data pochází / nepochází z dvourozměrného normálního rozdělení. ★

Poznámka: Hodnocení normality na základě grafické vizualizace je pouze subjektivním hodnocením. V sedmém cvičení budeme normalitu, případně dvourozměrnou normalitu, náhodného výběru posuzovat objektivně, a to na základě testů normality.