7 Normální rozdělení $N(\mu, \sigma^2)$ - Procvičovací příklady

- $X_1, \ldots, X_n$ … nezávislé náhodné veličiny
- Normální rozdělení
  - $X \sim N(\mu, \sigma^2)$
  - $\boldsymbol{\theta} = (\mu, \sigma^2)^T$
  - hustota
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}. \tag{7.1}$$
  - vlastnosti $\mathrm{E}[X] = \mu$; $\mathrm{Var}[X] = \sigma^2$
  - dnorm(x, mu, sigma), pnorm(x, mu, sigma), rnorm(M, mu, sigma), qnorm(alpha, mu, sigma)
- Standardizované normální rozdělení
  - $X \sim N(0, 1)$
  - $\boldsymbol{\theta} = (0, 1)^T$
  - hustota
$$f(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad x \in \mathbb{R}. \tag{7.2}$$
  - vlastnosti $\mathrm{E}[X] = 0$; $\mathrm{Var}[X] = 1$
  - dnorm(x), pnorm(x), rnorm(M), 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 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.
  - Věta 2: Nechť $X_1, \ldots, X_{n_1}$ jsou nezávislé náhodné veličiny pocházející z normálního rozdělení $N(\mu_1, \sigma_1^2)$ a $Y_1, \ldots, Y_{n_2}$ jsou nezávislé náhodné veličiny pocházející z normálního rozdělení $N(\mu_2, \sigma_2^2)$. Potom rozdíl náhodných veličin $\bar{X}_{n_1} - \bar{Y}_{n_2} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$.
- Data:
  - **Dataset 6:** 03-paired-means-clavicle2.txt
  - Datový soubor obsahuje osteometrické údaje o délkách klíčních kostí (*clavicula*). Data pochází z anglického souboru dokumentovaných skeletů (Parsons, 1916). V souboru se nachází délky klíčních kostí na pravé a levé straně těla v párovém uspořádání. Jednotlivé kosti bez druhostranné kosti nebyly do souboru zařazeny.
  - Přehled proměnných v datasetu:
    - id … ID jedince;
    - sex … pohlaví jedince (m - muž, f - žena);
    - length.L … délka klíční kosti z levé strany (v mm);
    - length.R … délka klíční kosti z pravé strany (v mm).

Příklad 7.1. Základní číselné charakteristiky spojitého znaku
Načtěte datový soubor 03-paired-means-clavicle2.txt. Nechť náhodná proměnná Y popisuje délku klíční kosti z pravé strany. Proměnná Y je potom spojitého typu. Pro délku klíční kosti z pravé strany vytvořte tabulku základních číselných charakteristik.

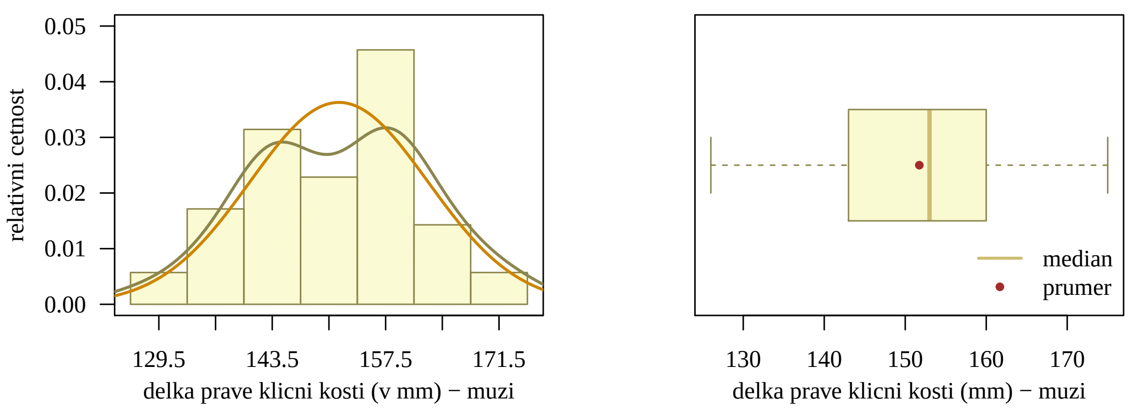
Řešení příkladu 7.1
V tabulce základních číselných charakteristik budou obsaženy následující charkteristiky: aritmetický průměr, směrodatná odchylka, minimální hodnota, dolní kvartil, medián, horní kvartil, maximální hodnota, mezikvartilové rozpětí, koeficient šikmosti a koeficient špičatosti.

```
           m  s min dolni.kv median horni.kv max IQR sikmost spicatost
prava 151.74 11 126      143    153      160 175  17   -0.06     -0.65
```

Naměřené hodnoty délky klíční kosti z pravé strany u mužů se pohybují v rozmezí 126–175 mm. Průměrná délka klíční kosti z pravé strany u mužů je 151.74 mm se směrodatnou odchylkou 11.00 mm. 25 % naměřených hodnot je menších nebo rovných 143.00 mm, 50 % hodnot je menších nebo rovných 153.00 mm, 75 % hodnot je menších nebo rovných 160.00 mm. Interkvantilové rozpětí naměřených hodnot je 17 mm. Hodnota koeficientu šikmosti ($b_1 = -0.06$) ukazuje na nevyšikmený charakter dat, hodnota koeficientu špičatosti ($b_2 = -0.65$) ukazuje na zploštělý charakter dat. ★

Příklad 7.2. Vizualizace dat z normálního modelu
Načtěte datový soubor 03-paired-means-clavicle2.txt. Nechť náhodná proměnná Y popisuje délku klíční kosti z pravé strany. Pomocí histogramu a krabicového diagramu vhodně vizualizujte rozdělení délky klíční kosti z pravé strany **pro muže**. Histogram superponujte (a) křivkou jádrového odhadu hustoty; (b) teoretickou křivkou hustoty normálního rozdělení $N(\mu, \sigma^2)$. Hodnoty parametrů μ a σ^2 odhadněte na základě dat.

Řešení příkladu 7.2



Obrázek 1: Vizualizace délky pravé klíční kosti u mužů pomocí histogramu (vlevo) a krabicového diagramu (vpravo)

★

Příklad 7.3. Výpočet pravděpodobností na základě normálního modelu
Za předpokladu, že náhodná veličina Y udávající délku klíční kosti z pravé strany pochází z normálního rozdělení $N(151.74, 11^2)$ vypočítejte pravděpodobnost, že délka délka klíční kosti z pravé strany je (a) menší než 140 mm; (b) větší než 160 mm; (c) v rozmezí 150–160 mm; (d) rovná 155 mm.

Řešení příkladu 7.3

```
      p1     p2     p3 p4
1 0.1429 0.2263 0.3366  0
```

Pravděpodobnost, že délka klíční kosti z pravé strany je menší než 140 mm je 14.29 %. Pravděpodobnost, že délka klíční z pravé strany kosti je větší než 160 mm je 22.63 %. Pravděpodobnost, že délka klíční kosti z pravé strany je v rozmezí 150–160 mm je 33.66 %. Protože délka klíční kosti z pravé strany pochází z normálního rozdělení, což je rozdělení spojitého typu, je tato délka rovná 155 mm s pravděpodobností 0 %.

★

Příklad 7.4. Výpočet pravděpodobností na základě normálního modelu
Za předpokladu, že náhodná veličina X udávající délku klíční kosti z pravé strany pochází z normálního rozdělení $N(151.74, 11^2)$ vypočítejte pravděpodobnost, že **průměrná délka sedmi** klíčních kostí z pravé strany je (a) menší než 140 mm; (b) větší než 160 mm; (c) v rozmezí 150–160 mm; (d) rovná 155 mm.

Řešení příkladu 7.4

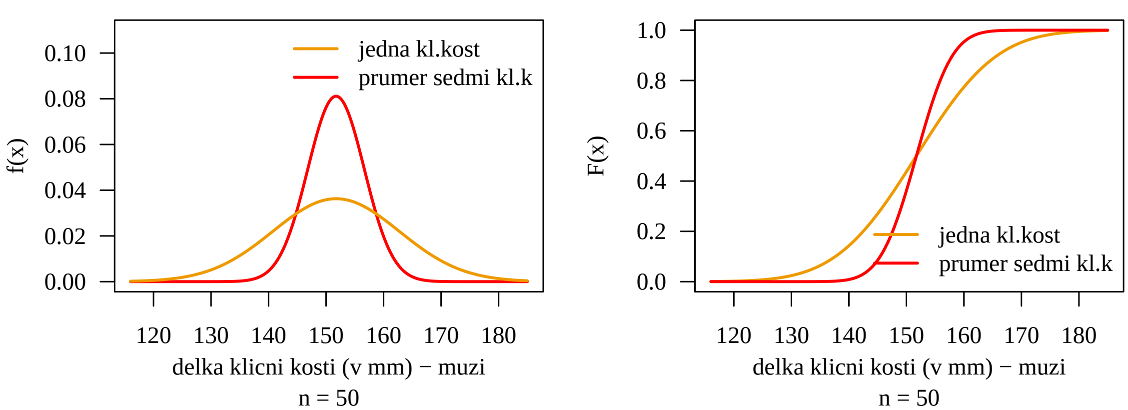
```
      p1     p2     p3 p4
1 0.0024 0.0234 0.6388  0
```

Pravděpodobnost, že průměrná délka sedmi klíčních kostí z pravé strany je menší než 140 mm je 0.24 %. Pravděpodobnost, že průměrná délka sedmi klíčních kostí z pravé strany je větší než 160 mm je 2.34 %. Pravděpodobnost, že průměrná délka sedmi klíčních kostí z pravé strany je v rozmezí 150–160 mm je 63.88 %. Protože průměrná délka sedmi klíčních kostí z pravé strany pochází z normálního rozdělení, což je rozdělení spojitého typu, je tato průměrná délka rovná 155 mm s pravděpodobností 0 %.

★

Příklad 7.5. Graf funkce hustoty a distribuční funkce normálního modelu
V příkladech 7.1 a 7.2 jsme odhadli hodnoty parametrů μ a σ^2 normálního rozdělení znaku X = *délka klíční kosti z pravé strany u mužů* jako $\widehat{\mu}_Y = 151.74$ a $\widehat{\sigma}^2_Y = 11^2$. Nakreslete graf hustoty a distribuční funkce tohoto normálního rozdělení.

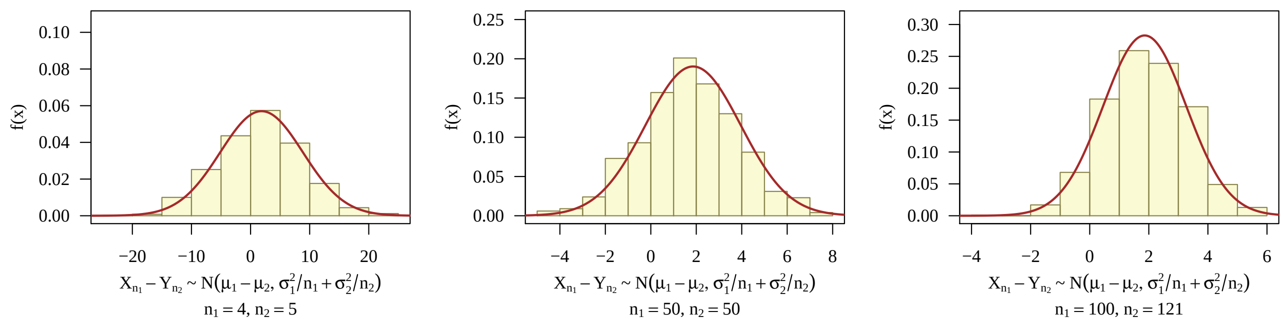
Řešení příkladu 7.5



Obrázek 2: Funkce hustoty (vlevo) a distribuční funkce (vpravo) normálního modelu

★

Příklad 7.6. Simulační studie: Věta 2: Rozdělení rozdílu výběrových průměrů $\bar{X}_{n_1} - \bar{Y}_{n_2}$
Na základě simulační studie ověřte že pokud náhodná veličina $X \sim N(\mu_1, \sigma_1^2)$ a náhodná veličina $Y \sim N(\mu_2, \sigma_2^2)$, potom $\bar{X}_{n_1} - \bar{Y}_{n_2} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$. Generujte pseudonáhodná čísla $X \sim N(\mu_1, \sigma_1^2)$ a $Y \sim N(\mu_2, \sigma_2^2)$, kde $\mu_1 = 153.60$, $\sigma_1^2 = 9.95^2$, $\mu_2 = 151.74$, $\sigma_2 = 11^2$ při (a) $n_1 = 4$, $n_2 = 5$; (b) $n_1 = 50$, $n_2 = 50$; (c) $n_1 = 100$, $n_2 = 121$. Pro všechny tři případy (a), (b) i (c) vypočítejte $\Pr(\bar{X}_{n_1} - \bar{Y}_{n_2} < 2)$ na základě empirického a teoretického rozdělení $\bar{X}_{n_1} - \bar{Y}_{n_2}$. Pravděpodobnosti porovnejte.



Obrázek 3: Histogramy rozdílů průměrů 1000 náhodných výběrů o rozsazích (a) $n_1 = 4, n_2 = 5$ (vlevo); (b) $n_1 = 50, n_2 = 50$ (uprostřed); (c) $n_1 = 100, n_2 = 121$ (vpravo)

Tabulka 1: Tabulka teoretických a exaktních pravděpodobností $\Pr(X_{n_1} - Y_{n_2} < 2\,\mathrm{mm})$

	teoretická	exaktní
$n_1 = 4$, $n_2 = 5$	0.5080	0.4970
$n_1 = 45, n_2 = 50$	0.5266	0.5630
$n_1 = 100, n_2 = 121$	0.5395	0.5280

★