

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

- X_1, \dots, X_n ... nezávislé náhodné veličiny
- Normální rozdělení

- $X \sim N(\mu, \sigma^2)$

- $\theta = (\mu, \sigma^2)^T$

- hustota

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}. \quad (7.1)$$

- vlastnosti $E[X] = \mu$; $\text{Var}[X] = \sigma^2$

- `dnorm(x, mu, sigma)`, `pnorm(x, mu, sigma)`, `rnorm(M, mu, sigma)`, `qnorm(alpha, mu, sigma)`

- Standardizované normální rozdělení

- $X \sim N(0, 1)$

- $\theta = (0, 1)^T$

- hustota

$$f(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}. \quad (7.2)$$

- vlastnosti $E[X] = 0$; $\text{Var}[X] = 1$

- `dnorm(x)`, `pnorm(x)`, `rnorm(M)`, `qnorm(alpha)`

- Vlastnosti normálního rozdělení

- Věta 1: Necht' X_1, \dots, 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: Necht' X_1, \dots, X_{n_1} jsou nezávislé náhodné veličiny pocházející z normálního rozdělení $N(\mu_1, \sigma_1^2)$ a Y_1, \dots, 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. Hustota normálního modelu

Naprogramujte v \mathbb{R} funkci `dnormal(x, mu, sigma2)` počítající hodnoty hustoty normálního rozdělení $N(\mu, \sigma^2)$ v hodnotě x . Správnost funkce otestujte na výpočtu $f(x)$, $x = -1, 0, 2$, pro $X \sim N(\mu, \sigma^2)$, kde $\mu = 0$ a $\sigma^2 = 2$. Výsledky ověřte s výsledky funkce `dnorm()`.

Řešení příkladu 7.1

	<code>f(-1)</code>	<code>f(0)</code>	<code>f(2)</code>
1	0.2196956	0.2820948	0.1037769

`f(-1) = 0.2197`; `f(0) = 0.2821`; `f(2) = 0.1038`. ★

Příklad 7.2. Základní číselné charakteristiky spojitého znaku

Načtěte datový soubor `03-paired-means-clavicle2.txt`. Nechť náhodná proměnná X popisuje délku klíční kosti z levé strany u mužů. Proměnná X je potom spojitého typu. Pro délku klíční kosti z levé strany u mužů vytvořte tabulku základních číselných charakteristik.

Řešení příkladu 7.2

V tabulce základních číselných charakteristik budou obsaženy následující charakteristiky: aritmetický průměr, směrodatná odchylka, koeficient variace, minimální hodnota, dolní kvartil, medián, horní kvartil, maximální hodnota, interkvartilové rozpětí, koeficient šikmosti a koeficient špičatosti.

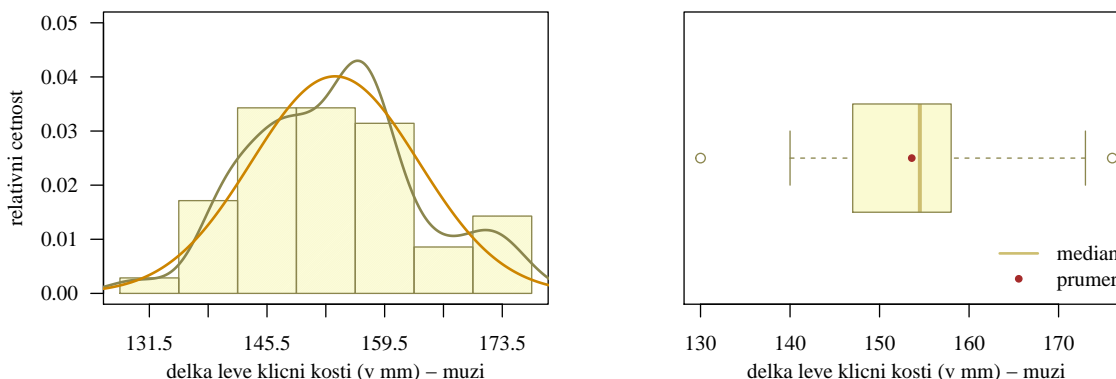
	<code>n</code>	<code>m</code>	<code>s</code>	<code>v</code>	<code>min</code>	<code>dolni.kv</code>	<code>median</code>	<code>horni.kv</code>	<code>max</code>	<code>IQR</code>	<code>sikmost</code>	<code>spicatost</code>
leva	50	153.6	9.95	0.06	130	147	154.5	158	176	11	0.21	-0.29

Datový soubor obsahuje údaje o délkách klíční kosti z levé strany 50 mužů. Naměřené hodnoty délky klíční kosti z levé strany u mužů se pohybují v rozmezí 130–176 mm. Průměrná délka klíční kosti z levé strany u mužů je 153.60 mm se směrodatnou odchylkou 9.95 mm. Směrodatná odchylka představuje 6 % aritmetického průměru. 25 % naměřených hodnot je menších nebo rovných 147.00 mm, 50 % hodnot je menších nebo rovných 154.50 mm, 75 % hodnot je menších nebo rovných 158.00 mm. Interkvartilové rozpětí naměřených hodnot je 11 mm. Hodnota koeficientu šikmosti ($b_1 = 0.21$) ukazuje na mírné kladné vychýlení dat (hodnoty vyšikmené doleva s prodlouženým pravým koncem), hodnota koeficientu špičatosti ($b_2 = -0.29$) ukazuje na mírně zploštělý charakter dat. ★

Příklad 7.3. Vizualizace dat z normálního modelu

Načtěte datový soubor `03-paired-means-clavicle2.txt`. Nechť náhodná proměnná X popisuje délku klíční kosti z levé strany u mužů. Pomocí histogramu a krabicového diagramu vhodně vizualizujte rozdělení délky klíční kosti z levé strany u mužů. Histogram superponujte (a) křivkou jádrového odhadu hustoty; (b) křivkou teoretické hustoty normálního rozdělení $N(\mu, \sigma^2)$. Hodnoty parametrů μ a σ^2 odhadněte na základě dat.

Řešení příkladu 7.3



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

Příklad 7.4. Výpočet pravděpodobností na základě normálního modelu

Za předpokladu, že náhodná veličina X udávající délku klíční kosti z levé strany u mužů pochází z normálního rozdělení $N(153.6, 9.95^2)$ vypočítejte pravděpodobnost, že délka klíční kosti z levé 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	0.0858	0.26	0.3813	0

5
6

Pravděpodobnost, že délka klíční kosti z levé strany u mužů je menší než 140 mm je 8.58 %. Pravděpodobnost, že délka klíční kosti z levé strany u mužů je větší než 160 mm je 26.00 %. Pravděpodobnost, že délka klíční kosti z levé strany u mužů je v rozmezí 150–160 mm je 38.13 %. Protože délka klíční kosti z levé strany u mužů pochází z normálního rozdělení, což je rozdělení spojitého typu, je tato délka rovná 155 mm s pravděpodobností 0 %. ★

Příklad 7.5. Výpočet pravděpodobností na základě normálního modelu

Za předpokladu, že náhodná veličina X udávající délku klíční kosti z levé strany u mužů pochází z normálního rozdělení $N(153.6, 9.95^2)$ vypočítejte pravděpodobnost, že **průměrná délka pěti** klíčních kostí z levé strany u mužů je (a) menší než 140 mm; (b) větší než 160 mm; (c) v rozmezí 150–160 mm; (d) rovná 155 mm.

Řešení příkladu 7.5

	p1	p2	p3	p4
1	0.0011	0.0751	0.7157	0

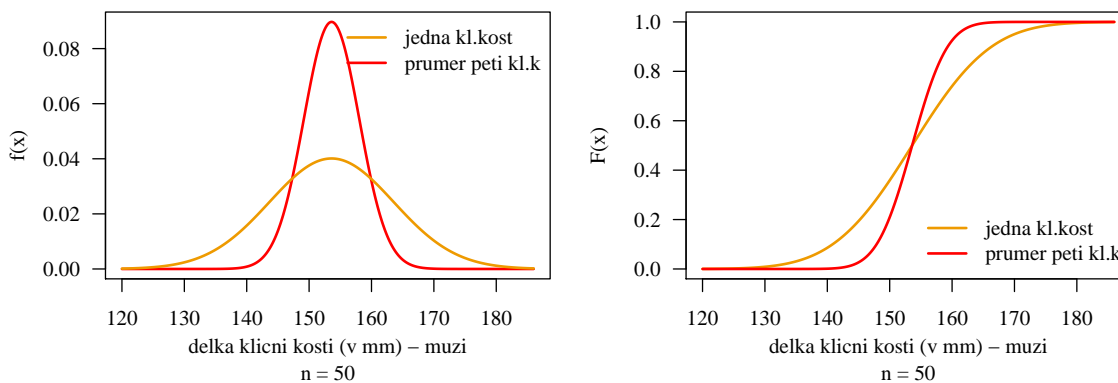
7
8

Pravděpodobnost, že průměrná délka pěti klíčních kostí z levé strany u mužů je menší než 140 mm je 0.11 %. Pravděpodobnost, že průměrná délka pěti klíčních kostí z levé strany u mužů je větší než 160 mm je 7.51 %. Pravděpodobnost, že průměrná délka pěti klíčních kostí z levé strany u mužů je v rozmezí 150–160 mm je 71.57 %. Protože průměrná délka pěti klíčních kostí z levé strany u mužů pochází z normálního rozdělení, což je rozdělení spojitého typu, je tato délka rovná 155 mm s pravděpodobností 0.00 %. ★

Příklad 7.6. Graf funkce hustoty a distribuční funkce normálního modelu

V příkladech 7.2 a 7.3 jsme odhadli hodnoty parametrů μ a σ^2 normálního rozdělení znaku $X =$ délka klíční kosti z levé strany u mužů jako $\hat{\mu}_X = 153.60$ a $\hat{\sigma}_X^2 = 9.95^2$. Nakreslete graf hustoty a distribuční funkce tohoto normálního rozdělení.

Řešení příkladu 7.6



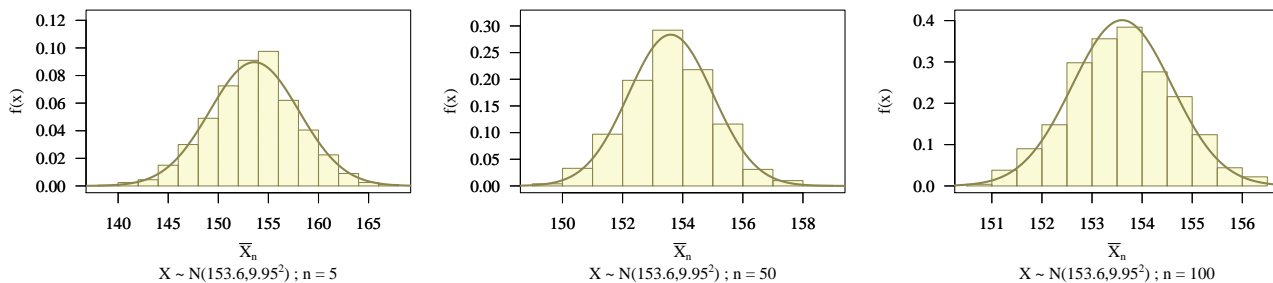
Obrázek 2: Funkce hustoty (vlevo) a distribuční funkce (vpravo) normálního modelu

★

Příklad 7.7. Simulační studie: Věta 1: Rozdělení výběrového průměru \bar{X}_n

Na základě simulační studie ověřte, že pokud $X \sim N(\mu, \sigma^2)$, potom $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$. Zvolte $\mu = 153.60$, $\sigma^2 = 9.95^2$, (a) $n = 5$, (b) $n = 50$, (c) $n = 100$. Vygenerujte M pseudonáhodných výběrů X_1, \dots, X_n , $M = 1000$. Pro každý výběr vypočítejte realizaci aritmetického průměru $\bar{x}_{n,m}$, $m = 1, \dots, M$. Následně vygenerujte histogram pro hodnoty $\bar{x}_{n,m}$ a superponujte jej teoretickou křivkou hustoty pro \bar{X}_n . Pro všechny tři případy (a), (b) i (c) vypočítejte $\Pr(\bar{X}_n > 152)$ na základě empirického a teoretického rozdělení \bar{X}_n . Pravděpodobnosti porovnejte.

Řešení příkladu 7.7



Obrázek 3: Histogramy průměrů 1000 náhodných výběrů o rozsahu (a) $n = 5$ (vlevo); (b) $n = 50$ (uprostřed); (c) $n = 100$ (vpravo)

	n = 5	n = 50	n = 100
teoreticka	0.6405	0.8723	0.9461
exaktni	0.6520	0.8660	0.9340

9
10
11

