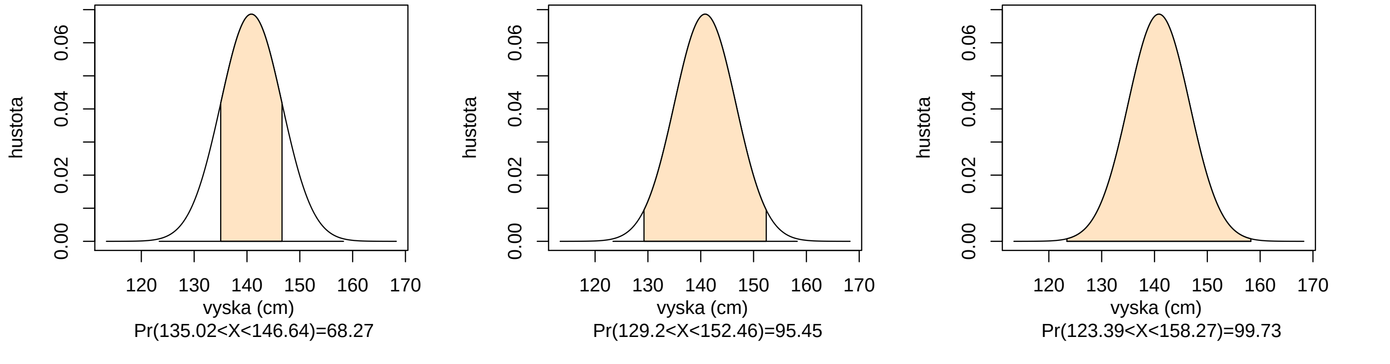
Zadání příkladů - cvičení č.1 - 15-9-23

Příklad č.1 (porovnání dvou typů modelů) *(přednáška)*
Model rozdělení pravděpodobnosti je modelem náhodné proměnné X, např. (1) model rozdělení pravděpodobnosti náhodné proměnné X šířka dolní čelisti, nebo (2) model rozdělení pravděpodobnosti náhodné proměnné X hrubost kožních řas u dospělých zdravých žen. *Statistický model* je modelem náhodné proměnné $Y|X$ (Y kauzálně závisí na X), např. (1) model závislosti náhodné proměnné Y šířka dolní čelisti na proměnné X pohlaví, nebo (2) model závislosti náhodné proměnné Y hrubost kožních řas u dospělých zdravých žen na proměnné X BMI. Všimněme si, že náhodné proměnné označujeme X anebo Y podle toho, jaký model je charakterizuje.

Příklad č.2 (jednoduchý náhodný výběr)
V jednoduchém náhodném výběru o rozsahu n z populace s konečným rozsahem N má každý prvek stejnou pravděpodobnost vybrání. Pokud vybíráme bez vracení (opakování), mluvíme o *jednoduchém náhodném výběru bez vracení* (Dalgaard, 2008). Pokud vybíráme s vracením, mluvíme o *jednoduchém náhodném výběru s vracením.* Mějme množinu $\mathcal{M}$ s $N = 10$ prvky a chceme z ní vybrat $n = 3$ prvky (a) bez vracení, (b) s vracením. Kolik máme možností? Jak vypadá jedna takováto možnost, pokud $\mathcal{M} = \{1, 2, \ldots, 10\}$? Zopakujte to samé pro $N = 100$, $n = 30$ a množinu $\mathcal{M} = \{1, 2, \ldots, 100\}$.

Příklad č.3 (jednoduchý náhodný výběr)
Mějme skupinu lidí označených identifikačními čísly (ID) od 1 do 30. Vyberte (a) náhodně 5 lidí z 30-ti bez návratu, (b) náhodně 5 lidí ze 30-ti s návratem a nakonec (c) náhodně 5 lidí ze 30-ti bez návratu, přičemž lidé s ID od 28-mi do 30-ti mají pravděpodobnost vybrání $4\times$ vyšší než lidé s ID od 1 do 27.

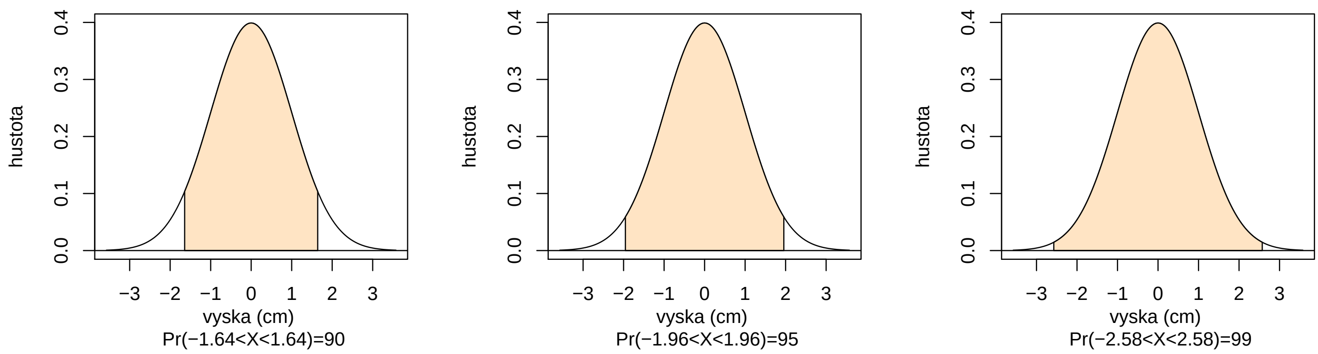
Příklad č.4 (normální rozdělení)
Mějme náhodnou proměnnou X (může to být např. výška postavy desetiletých dívek) a předpokládejme, že tato náhodná proměnná má normální rozdělení s parametry μ (střední hodnota) a σ^2 (rozptyl), což zapisujeme jako $X \sim N(\mu, \sigma^2)$, $\mu = 140.83$, $\sigma^2 = 33.79$. Normální rozdělení představuje model rozdělení pravděpodobnosti pro tuto náhodnou proměnnou. Vypočítejte pravděpodobnost $\Pr(a \leq X \leq b) = \Pr(X < b) - \Pr X < a) = F_X(b) - F_X(a)$, kde $a = \mu - k\sigma$, $b = \mu + k\sigma$, $k = 1, 2, 3$. Nakreslete hustotu rozdělení pravděpodobnosti, vybarvěte oblast mezi body a a b a popište osy x a y tak, jako je uvedeno na obrázku 1.



Obrázek 1: Míry normálního rozdělení; křivka hustoty s vybarveným obsahem pod touto křivkou mezi příslušnými kvantily na ose x; obsah je rovný pravděpodobnosti výskytu subjektů s danou výškou v rozpětí těchto kvantilů.

Dostaneme pravidlo $68.27 - 95.45 - 99.73$ (tzv. *míry normálního rozdělení*.

Příklad č.5 (normální rozdělení)
Mějme $X \sim N(\mu, \sigma^2)$, kde $\mu = 150$, $\sigma^2 = 6.25$. Vypočítejte $a = \mu - x_{1-\alpha/2}\sigma$ a $b = \mu + x_{1-\alpha/2}\sigma$ tak, aby $\Pr(a \leq X \leq b) = 1 - \alpha$, byla rovná 0.9, 0.95, 0.99. Číslo $x_{1-\alpha/2}$ je kvantil normovaného normálního rozdělení, t.j. $\Pr(Z = \frac{X-\mu}{\sigma} < x_{1-\alpha}$, $Z \sim N(0,1)$. Nakreslete hustotu rozdělení pravděpodobnosti, vybarvěte oblast mezi body a a b a popište osy x a y tak, jako je uvedeno na obrázku 2.



Obrázek 2: Upravené míry normálního rozdělení; křivka hustoty s vybarveným obsahem pod touto křivkou mezi příslušnými kvantily na ose x; obsah je rovný pravděpodobnosti výskytu subjektů s danou normovanou výškou v rozpětí těchto kvantilů.

Dostaneme pravidlo $90 - 95 - 99$ (tzv.*upravené míry normálního rozdělení*. Použili jsme nerovnost $\Pr(u_{\alpha/2<Z<u_{1-\alpha/2}} = \Phi(x_{1-\alpha/2}) - \Phi(x_{\alpha/2}) = 1 - \alpha$, kde Φ je distribuční funkce normálního normovaného rozdělení a všeobecně $(\alpha \in (0, 1/2)$; v příkladě $\alpha = 0.1, 0.05$ a 0.01.

Příklad č.6 (normální rozdělení)
Předpokládejme model normálního rozdělení $N(130, 13^2)$ pro systolický krevní tlak. Jaká část populace (v %) bude mít hodnoty vyšší než 160 mm Hg?

Příklad č.7 (binomické rozdělení)
Předpokládejme, že počet lidí upřednostňujících léčbu A před léčbou B se řídí modelem binomického rozdělení s parametry N (rozsah náhodného výběru) a p (pravděpodobnost výskytu), ozn. $Bin(N, p)$, kde $N = 20$, $p = 0.5$, t.j. lidé preferují oba dva typy léčby stejnou měrou. (a) Jaká je pravděpodobnost, že 16 a více pacientů upřednostní léčbu A před léčbou B? (b) Jaká je pravděpodobnost, že 16 a více a zároveň 4 a méně pacientů upřednostní léčbu A před léčbou B?

Příklad č.8 (binomické rozdělení)
Předpokládejme, že $\Pr(vir) = 0.533 = p_1$ je pravděpodobnost výskytu dermatoglyfického vzoru vír na palci pravé ruky mužů české populace a $\Pr(ostatni) = 0.467 = p_2$ je pravděpodobnost výskytu ostatních vzorů na palci pravé ruky mužů české populace, přičemž X je počet vírů a Y je počet

ostatních vzorů, kde $X \sim Bin(N, p_1)$ a $Y \sim Bin(N, p_2)$. Vypočítejte (1) $\Pr(X \leq 120)$, když $N = 300$ a (2) $\Pr(Y \leq 120)$, když $N = 300$.

Příklad č.9 (parametry) *(přednáška)*
Příklady parametrů θ - střední hodnota μ, rozptyl σ^2, korelační koeficient ρ, pravděpodobnost p výskytu nějaké události, rozdíl dvou středních hodnot $\mu_1 - \mu_2$, podíl dvou rozptylů σ_1^2/σ_2^2, rozdíl dvou korelačních koeficientů $\rho_1 - \rho_2$, rozdíl dvou pravděpodobností $p_1 - p_2$ apod.

Příklad č.10 (binomické rozdělení) *(přednáška)*
Pokud $X \sim Bin(N, \theta)$, $\theta = p \in \langle 0; 1 \rangle$, potom $\mathcal{Y}_\theta$ je stejný pro všechny θ a koinciduje s výběrovým prostorem $\mathcal{Y} = \{0, 1, \ldots, N\}$.

Příklad č.11 (počet členů v mnohorozměrném LRM) *(z přednášky)*
Mějme mnohorozměrný lineární regresní model $\mathcal{L}$ o 20-ti proměnných, ve kterém jsou obsaženy všechny možné interakce těchto proměnných (dvojné, trojné,...). Kolik členů (jednoduché regresory + všechny interakce) má takový model?