M5120 – cvičení

Číselné charakteristiky náhodného výběru a grafická zobrazení

Ondřej Pokora (pokora@math.muni.cz)

Ústav matematiky a statistiky, Přírodovědecká fakulta, Masarykova univerzita, Brno

(podzim 2015)

Výběrové momenty, histogram, ECDF

Výběrový průměr: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Výběrový rozptyl: $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$

Výběrová (empirická) distribuční funkce (ECDF):

$$\widehat{F}(x) = F_n(x) = \frac{1}{n}\left|\{i \in 1, \ldots, n : X_i \leq x\}\right|$$

Konstrukce histogramu: Rozpětí náhodného výběru rozdělíme na vhodný počet k (doporučeno $k \approx \sqrt{n}$) intervalů šířky r. V každém z nich pak počítáme relativní četnost $f_i = \frac{n_i}{n}$. Histogram se skládá z k obdélníků o základnách délky r určených dělením rozpětí na intervaly a výškách $\frac{f_i}{r}$.

Výběrové kvantily

α-kvantil x_a je číslo, rozdělující **uspořádaný náhodný výběr** $X_{(1)} \leq \cdots \leq X_{(n)}$ na dolní úsek, obsahující aspoň podíl α všech dat a na horní úsek obsahující aspoň podíl $1 - \alpha$ všech dat. Algoritmus výpočtu **výběrového α-kvantilu**:

$$n\,\alpha = \begin{cases} \text{celé číslo } c & \Rightarrow x_\alpha = \frac{X_{(c)} + X_{(c+1)}}{2} \\ \text{necelé číslo} & \Rightarrow x_\alpha = X_{(c)}, \text{ kde } c = \lceil n\,\alpha \rceil \end{cases}$$

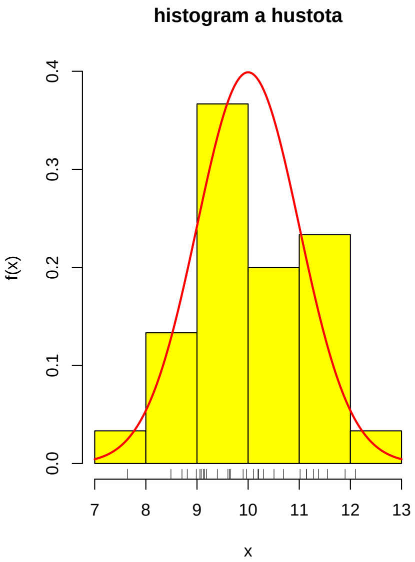
Pro speciálně zvolená α užíváme názvů: $x_{0,50}$ – medián, $x_{0,25}$ – dolní kvartil, $x_{0,75}$ – horní kvartil, $x_{0,1}, \ldots, x_{0,9}$ – decily, $x_{0,01}, \ldots, x_{0,99}$ – percentily.

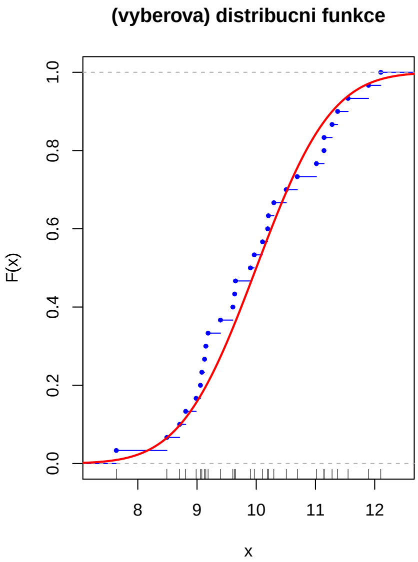
Charakteristikou variability je **kvartilová odchylka**: $q = x_{0,75} - x_{0,25}$ (**interquartile range, IQR**).

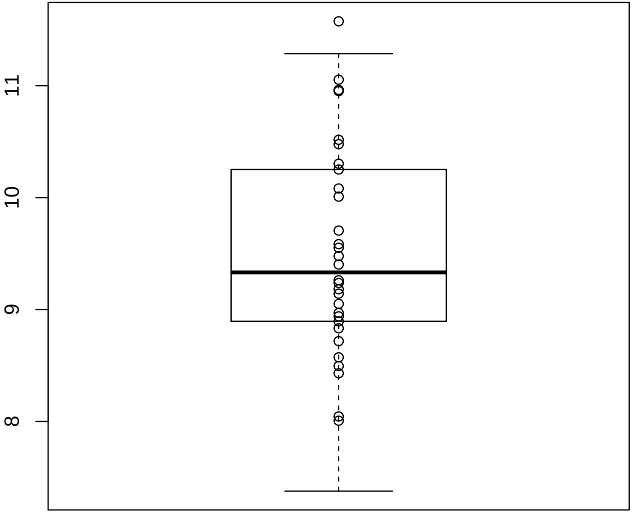
Krabicový graf (**boxplot**) je grafické znázornění číselných charakteristik

$$x_{0,50} - 1,5\,q \leq x_{0,25} \leq x_{0,50} \leq x_{0,75} \leq x_{0,50} + 1,5\,q.$$

Pozor, různé softwary (včetně *R*) vykreslují boxploty různými způsoby!


histogram a hustota
f(x)
0.0
0.1
0.2
0.3
0.4
7
8
9
10
11
12
13
x


(vyberova) distribucni funkce
F(x)
0.0
0.2
0.4
0.6
0.8
1.0
8
9
10
11
12
x



8
9
10
11