

ANALÝZA DAT V R

Intervaly spolehlivosti a testování hypotéz

- většina běžně používaných funkcí je implementována

STATISTICKÉ FUNKCE		
<code>mean()</code>	výběrový průměr	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
<code>var()</code>	výběrový rozptyl	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
<code>sd()</code>	výběrová směrodatná odchylka	$s = \sqrt{s^2}$
<code>min()</code>	minimum	$x_{min} = x_{(1)}$
<code>max()</code>	maximum	$x_{max} = x_{(n)}$
<code>range()</code>	minimum a maximum	x_{min}, x_{max}
<code>quantile(., p)</code>	p -tý kvantil, $p \in (0; 1)$	$x_p = (1 - \gamma)x_{(j)} + \gamma x_{(j+1)}$
<code>quantile(., 0.5), median()</code>	medián	$\tilde{x} = x_{0.5}$
<code>IQR()</code>	mezikvartilové rozpětí	$IQR = x_{0.75} - x_{0.25}$
<code>summary()</code>	minimum, dolní kvartil, medián, průměr, horní kvartil, maximum	
<code>cov(., .)</code>	výběrová kovariance	$s_{xy}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$
<code>cor(., ., method="pearson")</code>	Pearsonův korelační koeficient	$r_{xy} = \frac{s_{xy}^2}{s_x \cdot s_y}$

Poznámka: Argumenty funkcí `cov` a `cor` jsou dva vektory nebo matice o dvou sloupcích.

- není implementována funkce pro modus, populační rozptyl, směrodatnou odchylku a kovarianci $\Rightarrow$ nutno dopočítat:

- modus přes tabulku četností `table()`

- populační rozptyl, resp. populační směrodatnou odchylku přes výběrový rozptyl

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{n-1}{n} s^2$$

- populační kovarianci přes výběrovou kovarianci

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{n-1}{n} s_{xy}^2$$

- 9 možností výpočtu kvantilu

Příklad 1

Následující data jsou výšky chlapců posledních ročníků jisté střední školy:

169 176 164 155 159 159 156 164 164 195 173 171 166 192 180 174
155 166 173 156 162 139 161 158 167 164 165 157 190 155 164 179
174 177 184 168 154 172 167 166 175 175 178 174 169 160 158 161
152 177 185 166 161

- (a) Jaká je průměrná výška?
- (b) Určete minimum, maximum, medián a modus výšky.
- (c) Určete 70% kvantil a 2. decil.
- (d) Určete první a třetí kvartil.
- (e) Spočtete rozptyl, směrodatnou odchylku, rozpětí a mezikvartilové rozpětí.
- (f) Nakreslete histogram absolutních, relativních, kumulativních absolutních a kumulativních relativních třídních četností.
- (g) Znázorněte data pomocí boxplotu.

Příklad 2

V následující tabulce je uvedena délka praxe a odpovídající výška platu učitelů na jisté vesnické škole.

Spočítejte Pearsonův korelační koeficient.

Délka praxe (v letech)	16	22	3	14	3	18	12	38	5	2	15
Plat (v tis. Kč)	30	34	26	28	24	33	26	43	25	23	29