I Výběrová kovariance a výběrový korelační koeficient

Pracujeme-li s dvourozměrným náhodným výběrem (X, Y), často nás zajímají výběrové analogie kovariance a korelačního koeficientu. Výběrová kovariance se značí S_{12} a udává vzájemnou provázanost změny náhodných veličin X a Y. Realizace výběrového korelačního koeficientu R_{12} nabývají hodnot z intervalu $\langle -1, 1\rangle$ a výběrový korelační koeficient udává míru lineární závislosti mezi složkami X a Y náhodného vektoru (X, Y). Pro výpočet budeme potřebovat výběrové průměry M_1 a M_2 a výběrové směrodatné odchylky S_1, S_2 jednotlivých složek X, Y náhodného vektoru (X, Y).

Nový pojem: Výběrová kovariance, výběrový korelační koeficient

Statistika

$$S_{12} = \frac{1}{n-1}\sum_{i=1}^{n}((X_i - M_1)(Y_i - M_2))$$

dvourozměrného náhodného vyběru $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ se nazývá výběrová kovariance. Výběrový korelační koeficient tohoto náhodného výběru je

$$R_{12} = \frac{S_{12}}{S_1 S_2}.$$

II Ilustrační příklad

V následující tabulce jsou data z 20 náhodně vybraných nabídek prodeje bytů v Brně. Vždy je uvedena rozloha v m^2 a cena bytu v milionech Kč.

Zajímá nás korelace mezi rozlohou bytu a jeho cenou – dá se předpokládat, že čím větší byt,

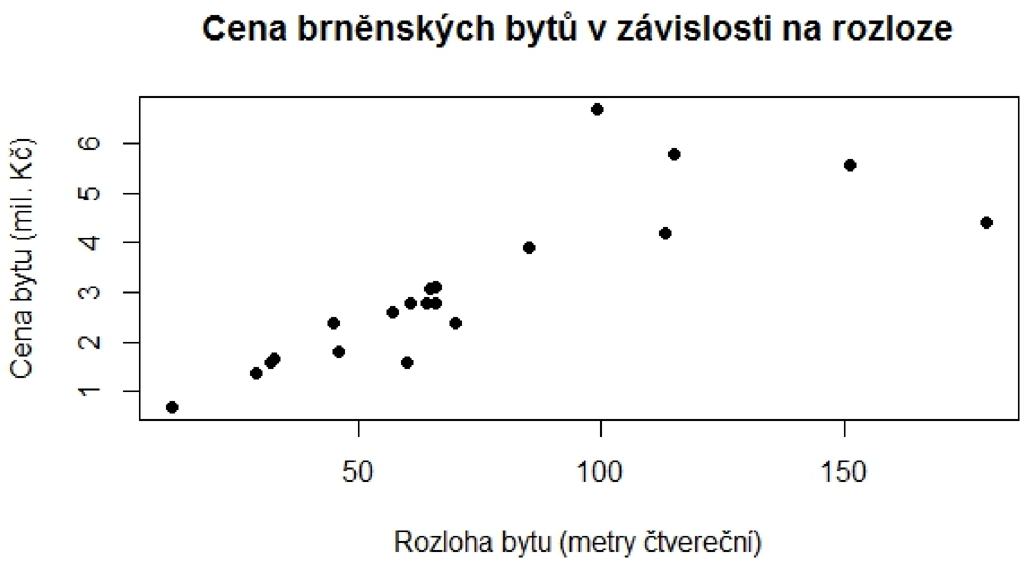
Ceny a rozlohy brněnských bytů

	rozloha m^2	cena mil. Kč		rozloha m^2	cena mil. Kč		rozloha m^2	cena mil. Kč		rozloha m^2	cena mil. Kč
1	12	0,690	6	46	1,800	11	65	3,080	16	99	6,691
2	29	1,390	7	57	2,600	12	66	2,800	17	113	4,200
3	32	1,600	8	60	1,600	13	66	3,125	18	115	5,790
4	33	1,650	9	61	2,800	14	70	2,390	19	151	5.550
5	45	2,400	10	64	2,800	15	85	3,900	20	179	4,395

tím dražší, takže očekáváme kladnou korelaci. Nicméně závislost nemusí být nutně lineární, může být např. kvadratická nebo exponenciální, a tu bychom pomocí výběrového korelačního koeficientu neodhalili. Zkuste si nejprve výběrový korelační koeficient spočítat sami, pak si výpočet zkontrolujte.

Řešení. Označíme X náhodnou veličinu, která udává rozlohu bytu a Y jeho cenu v milionech Kč. Pak:

$$m_1 = 72,4 \text{ m}^2 \qquad m_2 = 3,06255 \text{ mil. Kč}$$
$$s_1 = 41,52919 \text{ m}^2 \qquad s_2 = 1,592295 \text{ mil. Kč}$$
$$s_{12} = 53,80693 \qquad r_{12} = 0,813694$$



Jak jsme očekávali, korelační koeficient r_{12} vyšel kladný, tedy mezi rozlohou bytu a jeho cenou existuje silná pozitivní korelace.