Příklady z popisné statistiky

Příklad 1.: Hodnoty znaku X mají průměr -3 a rozptyl 4. Najděte průměr a rozptyl hodnot znaku Y = 1 – 5X.
Řešení: Podle 5.17. a) dostáváme: $m_2 = 1 - 5m_1 = 1 - 5.(-3) = 16$, $s_2^2 = 5^2.s_1^2 = 25.4 = 100$.

Příklad 2.: Hodnoty znaku X mají aritmetický průměr 20 a směrodatnou odchylku 10. Najděte aritmetický průměr hodnot znaku $Y = X^2$.
Řešení:
Průměr hodnot znaku Y je $m_2 = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} x_i^2$. Průměr součtu kvadrátů hodnot znaku X se objeví ve výpočetním vzorci pro rozptyl hodnot znaku X: $s_1^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - m_1^2$. Odtud dostaneme $m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = s_1^2 + m_1^2 = 10^2 + 20^2 = 500$

Příklad 3.: V datovém souboru zvýšíme každou hodnotu o 10%. O kolik procent se zvýší rozptyl a o kolik % se zvýší koeficient variace?
Řešení:
Transformace X na Y je dána vzorcem Y = 1,1X. Tedy podle 5.17. a) $m_2 = 1{,}1m_1$ a $s_1^2 = 1{,}1^2 s_1^2 = 1{,}21 s_1^2$. Vidíme, že rozptyl se zvýší o 21%. Podle 5.25 je koeficient variace znaku Y roven $\frac{s_2}{m_2} = \frac{1{,}1s_1}{1{,}1m_1} = \frac{s_1}{m_1}$. Koeficient variace se tedy nezmění.

Příklad 4.: Znak X nabývá variant 0 a 1, přičemž varianta 0 se vyskytuje v 80% případů. Vypočtěte průměr a směrodatnou odchylku znaku X.
Řešení:
Podle 5.18. a) je vážený průměr $m = \frac{1}{n}\sum_{j=1}^{r} n_j x_{[j]} = \sum_{j=1}^{r} p_j x_{[j]}$. V našem případě varianta $x_{[1]}$ = 0, její relativní četnost $p_1 = 0{,}8$, varianta $x_{[1]} = 1$, její relativní četnost $p_2 = 0{,}2$. Po dosazení máme: m = 0,8.0 + 0,2.1 = 0,2.
Podle 5.18. a) je vážený rozptyl $s^2 = \frac{1}{n}\sum_{j=1}^{r} n_j\left(x_{[j]} - m\right)^2 = \sum_{j=1}^{r} p_j\left(x_{[j]} - m\right)^2$. Po dosazení máme:
$s^2 = 0{,}8(0 - 0{,}2)^2 + 0{,}2(1 - 0{,}2)^2 = 0{,}8.0{,}04 + 0{,}2.0{,}64 = 0{,}16$, $s = \sqrt{0{,}16} = 0{,}4$

Příklad 5.: Vážený aritmetický průměr činil 1500 a vážený rozptyl 90000. Varianty $x_{[j]}$ byly transformovány vztahem: $y_{[j]} = \frac{x_{[j]} - a}{h}$, j = 1, ..., r. Po této transformaci byl vážený aritmetický průměr 5 a vážený rozptyl 9. Určete konstanty a a h.
Řešení:
Podle 5.17. a) je průměr lineární kombinace Y = a + bX roven $m_2 = a + bm_1$. V našem případě tedy $5 = \frac{1500 - a}{h}$. Podle 5.17. a) je rozptyl lineární kombinace Y = a + bX roven $s_2^2 = b^2 s_1^2$.
V našem případě tedy $9 = \frac{90000}{h^2}$. Odtud $h_1 = 100$, $h_2 = -100$. Z první rovnice dostaneme
$a_1 = 1500 - 5.100 = 1000$, $a_2 = 1500 + 5.100 = 2000$.

Příklad 6.: Rozptyl součtů hodnot dvou znaků je 350, rozptyl rozdílů je 700. Vypočtěte koeficient korelace, víte-li, že oba znaky mají stejné rozptyly.

Řešení:

Podle 5.17. b) je rozptyl součtu U = X + Y roven $s_3^2 = s_1^2 + s_2^2 + 2s_{12}$ a analogicky rozptyl rozdílu V = X - Y roven $s_4^2 = s_1^2 + s_2^2 - 2s_{12}$. Společnou hodnotu rozptylů znaků X a Y označíme s^2. Dostáváme tedy dvě rovnice pro dvě neznámé:

$2s^2 + 2s_{12} = 350$

$2s^2 - 2s_{12} = 700$

Odtud vypočteme $s^2 = 262{,}5$ a $s_{12} = -87{,}5$. Koeficient korelace se počítá podle vzorce

$r_{12} = \frac{s_{12}}{s_1 s_2}$, v našem případě $r_{12} = \frac{s_{12}}{s^2} = -\frac{87{,}5}{262{,}5} = -\frac{1}{3}$

Příklad 7.: Máme k dispozici údaje od šesti obchodníků o poptávce (v kusech) po určitém výrobku loni a letos:

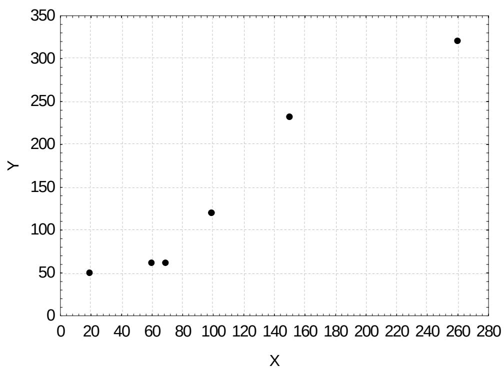
č. obchodníka	1	2	3	4	5	6
poptávka loni	20	60	70	100	150	260
poptávka letos	50	60	60	120	230	320

Pro úsporu času máte uvedeny číselné charakteristiky: $m_1 = 110$, $m_2 = 140$, $s_1^2 = 6066{,}7$, $s_2^2 = 10300$, $r_{12} = 0{,}972$.

a) Stanovte parametry regresní přímky, která vystihuje závislost letošní poptávky na loňské.
b) Byla-li loňská poptávka 110 kusů, jaký je regresní odhad letošní poptávky?

Řešení:

Nejprve nakreslíme dvourozměrný tečkový diagram.



Ze vzhledu tohoto diagramu lze soudit, že regresní přímka bude vhodným modelem závislosti letošní poptávky (znak Y) na loňské (znak X).

ad a) Podle 6.9. máme:

$$b_1 = \frac{s_{12}}{s_1^2} = \frac{r_{12} \cdot s_1 \cdot s_2}{s_1^2} = \frac{r_{12} \cdot s_2}{s_1} = 0{,}972\sqrt{\frac{10300}{6066{,}7}} = 1{,}2665$$

$$b_0 = m_2 - b_1 m_1 = 140 - 1{,}2665 \cdot 110 = 0{,}685$$

Regresní přímka má tedy rovnici y = 0,685 + 1,2665x.

ad b) $\hat{y} = 0{,}685 + 1{,}2665 \cdot 110 = 140$

Příklad 8.: Závislost mezi vnější teplotou a teplotou ve skladišti je popsána regresní přímkou y = 8+0,6x. Při jaké vnější teplotě klesne teplota ve skladišti pod bod mrazu?

Řešení:

0 = 8 + 0,6x, tedy $x = -\frac{8}{0{,}6} = -13{,}3$. Teplota ve skladišti klesne pod bod mrazu při vnější teplotě -13,3°C.

Příklad 9.: Jak se změní směrnice regresní přímky, když každou hodnotu závisle proměnného znaku zvětšíme o 10 % ?
Řešení:
Původní regresní přímka má rovnici $y = b_0 + b_1x$, nová regresní přímka má rovnici $1{,}1y = 1{,}1(b_0 + b_1x)$, tedy úsek i směrnice se zvýší o 10%.

Příklad 10.: V datovém souboru, z něhož byl vypočten průměr 110 a rozptyl 800, byly zjištěny 2 chyby: místo 85 má být 95 a místo 120 má být 150. Ostatních 18 údajů je správných. Opravte průměr a rozptyl.
Řešení:
Bez újmy na obecnosti lze předpokládat, že prvních 18 hodnot je správných.
Původní průměr:

$$110 = m_1 = \frac{1}{20}\sum_{i=1}^{20} x_i = \frac{1}{20}\left(\sum_{i=1}^{18} x_i + 85 + 120\right) \Rightarrow \sum_{i=1}^{18} x_i = 20 \cdot 110 - 85 - 120 = 1995$$

Nový průměr:

$$m_2 = \frac{1}{20}\left(\sum_{i=1}^{18} x_i + 95 + 150\right) = \frac{1}{20}(1995 + 95 + 150) = 112$$

Původní rozptyl:

$$800 = \frac{1}{20}\sum_{i=1}^{20} x_i^2 - m_1^2 = \frac{1}{20}\left(\sum_{i=1}^{18} x_i^2 + 85^2 + 120^2\right) - 110^2 \Rightarrow \sum_{i=1}^{18} x_i^2 = 236375$$

Nový rozptyl:

$$s_2^2 = \frac{1}{20}\left(\sum_{i=1}^{18} x_i^2 + 95^2 + 150^2\right) - m_2^2 = \frac{1}{20}\left(236375 + 95^2 + 150^2\right) - 112^2 = 851$$