Ladislav Beran; Milan Trch

Přímka, kterou hledáte (1)

Učitel matematiky, Vol. 17 (2009), No. 2, 81–85

Persistent URL: `http://dml.cz/dmlcz/150575`

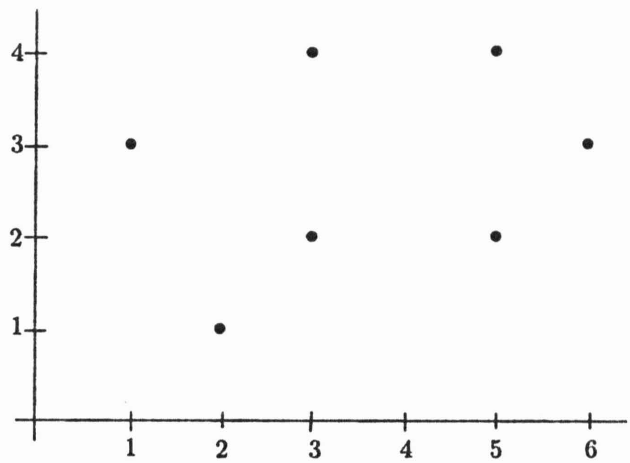
PŘÍMKA, KTEROU HLEDÁTE (1)

LADISLAV BERAN, MILAN TRCH

Článek amerických matematiků Burkeho a Hodgsona [1] oživil znovu problematiku metody lineární regrese. V tomto článku se snažíme doplnit jejich přístup z poněkud odlišného zorného úhlu a ukázat možnosti bezprostřední aplikace tohoto zobecnění.

Nejprve na jednoduchém příkladě připomeňme základní myšlenky.

Graf na obr. 1 byl získán vynesením údajů z tab. 1.



Obr. 1

1	2	3	3	5	5	6
3	1	2	4	2	4	3

Tabulka 1

Umístění bodů na obr. 1 vede zcela přirozeně na otázku, zda by rozložení těchto bodů bylo možné vystihnout „co nejlépe vedenou přímkou". Pokud mají vynášené body zachycovat například experimentálně nalezené výsledky vyšetřování jistého jevu, pak taková přímka by měla „co nejlépe lineárně popsat zkoumaný jev".

Je patrné, že uvedeným požadavkem není přímka v obecném případě určena. Obvykle se při stanovení takovéto přímky o rovnici $y = kx + q$ držíme osvědčené *metody nejmenších čtverců*. Při ní se hledá přímka mající tu vlastnost, že *součet druhých mocnin vzdáleností „ve směru osy y" bodů grafu od hledané přímky je co nejmenší.*

Podle této definice má být tedy co nejmenší součet

$$[b_1 - (ka_1 + q)]^2 + [b_2 - (ka_2 + q)]^2 + \cdots + [b_7 - (ka_7 + q)]^2.$$

Zde $a_1 = 1$, $a_2 = 2$, $a_3 = 3$, $a_4 = 3$, $a_5 = 5$, $a_6 = 5$, $a_7 = 6$, $b_1 = 3$, $b_2 = 1$, $b_3 = 2$, $b_4 = 4$, $b_5 = 2$, $b_6 = 4$, $b_7 = 3$.

Hledanou přímku nazvěme pro jednoduchost *minimální přímka* a označme ji p_{min}.

Jako ukázku vyhledávání minimální přímky uveďme možný numerický postup pro výše zadaný příklad.

Označme ještě

$$A := a_1 + a_2 + \cdots + a_7 = 25, \quad B := b_1 + b_2 + \cdots + b_7 = 19$$
$$C := a_1^2 + a_2^2 + \cdots + a_7^2 = 109, \; D := a_1b_1 + a_2b_2 + \cdots + a_7b_7 = 71.$$

Nalezené údaje lze přehledně vyčíst z tabulky 2.

a_i	b_i	a_i^2	a_ib_i
1	3	1	3
2	1	4	2
3	2	9	6
3	4	9	12
5	2	25	10
5	4	25	20
6	3	36	18
$A = 25$	$B = 19$	$C = 109$	$D = 71$

Tabulka 2

Směrnice k a úsek q na ose y hledané minimální přímky $y =$ $= kx + q$ se určí řešením soustavy dvou lineárních rovnic

$$\begin{aligned} Ak + nq &= B \\ Ck + Aq &= D, \end{aligned} \tag{1}$$

v níž n značí počet výchozích řádků tabulky, tj. zde $n = 7$.

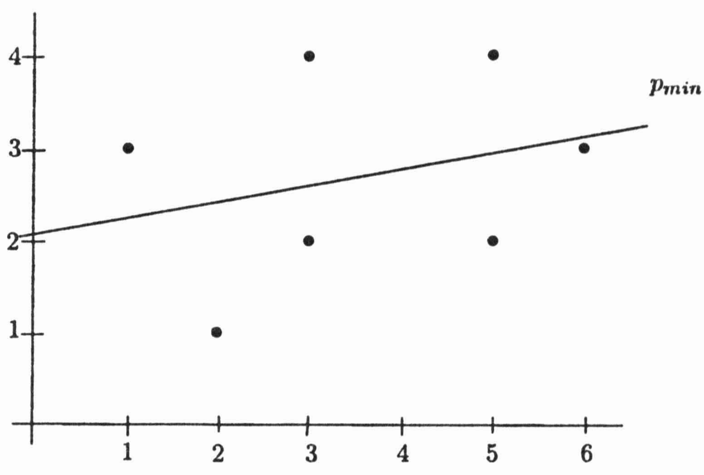
Soustava

$$\begin{aligned} 25k + 7q &= 19 \\ 109k + 25q &= 71 \end{aligned} \tag{1'}$$

má řešení

$$k = \frac{11}{69} \doteq 0{,}16; q = \frac{148}{69} \doteq 2{,}14,$$

což s dostatečnou přesností dává $y = 0{,}16x + 2{,}14$ (srovn. obr. 2).



Obr. 2

Pro další účely si přichystáme dva pomocné výsledky.

Věta 1. *Nechť $a_1, a_2, \cdots, a_n$ jsou libovolná reálná čísla a $n \geq 1$. Potom*

$$\Delta_n := n(a_1^2 + a_2^2 + \cdots + a_n^2) - (a_1 + a_2 + \cdots + a_n)^2 \geq 0.$$

Je-li $n \geq 2$, pak $\Delta_n = 0$ právě když $a_1 = a_2 = \cdots = a_n$.

Důkaz: 1. Nejprve ukážeme, že pro každé $n \geq 2$ platí $\Delta_n \geq \Delta_{n-1}$.
Vskutku, $\Delta_n \geq \Delta_{n-1}$ právě když

$$n(a_1^2 + a_2^2 + \cdots + a_n^2) - (n-1)(a_1^2 + a_2^2 + \cdots + a_{n-1}^2) \geq$$
$$\geq (a_1 + a_2 + \cdots + a_n)^2 - (a_1 + a_2 + \cdots + a_{n-1})^2.$$

To je ekvivalentní s tím, že

$$a_1^2 + a_2^2 + \cdots + a_{n-1}^2 + na_n^2 \geq a_n(2a_1 + 2a_2 + \cdots 2a_{n-1} + a_n)$$

a tedy také s tím, že

$$a_1^2 + a_2^2 + \cdots + a_{n-1}^2 + (n-1)a_n^2 \geq 2a_1a_n + 2a_2a_n + \cdots + 2a_{n-1}a_n.$$

Tento vztah je ale důsledkem toho, že

$$a_1^2 + a_n^2 \geq 2a_1a_n \quad \& \quad a_2^2 + a_n^2 \geq 2a_2a_n \quad \& \quad \cdots$$
$$\cdots \quad \& \quad a_{n-1}^2 + a_n^2 \geq 2a_{n-1}a_n.$$

2. Je-li $n \geq 2$, je podle prvního bodu důkazu

$$\Delta_n \geq \Delta_2 = 2(a_1^2 + a_2^2) - (a_1 + a_2)^2 = (a_1 - a_2)^2 \geq 0.$$

3. Je-li $n \geq 2$ a $\Delta_n = 0$, pak podle druhého bodu důkazu je $a_1 = a_2$. Protože Δ_n nezávisí na pořadí čísel $a_1, a_2, \cdots, a_n$, je rovněž $a_2 = a_3 = \cdots = a_n$. □

Věta 2. *Nechť γ, δ, Γ, Δ a ε jsou reálná čísla, přičemž γ a Γ jsou kladná. Pak*

(i) *funkce $f\colon \mathbb{R} \to \mathbb{R}, f\colon x \mapsto \gamma x^2 - 2\delta x$ nabývá své nejmenší hodnoty právě když $x = \frac{\delta}{\gamma} =: x_0$;*

(ii) *funkce $g\colon \mathbb{R}^2 \to \mathbb{R}^2$,*

$$g : (x, y) \mapsto \gamma(x + \varepsilon y)^2 - 2\delta \cdot (x + \varepsilon y) + \Gamma y^2 - 2\Delta \cdot y$$

nabývá své nejmenší hodnoty právě když $x = x_1$ a $y = y_1$, kde x_1 a y_1 jsou čísla jednoznačně určená vztahy

$$x_1 + \varepsilon y_1 = \frac{\delta}{\gamma} \quad \& \quad y_1 = \frac{\Delta}{\Gamma}.$$

Důkaz: (i) Můžeme buď vycházet z dobře známého průběhu grafu kvadratické funkce $f(x)$ nebo uvážit, že

$$\frac{f(x)}{\gamma} = x^2 - \frac{2\delta}{\gamma}x = \left(x - \frac{\delta}{\gamma}\right)^2 - \frac{\delta^2}{\gamma^2} \geq -\frac{\delta^2}{\gamma^2}.$$

Je tedy $-\frac{\delta^2}{\gamma}$ nejmenší hodnotou funkce $f(x)$. Z právě formulovaného vztahu je patrné, že $f(x)$ nabývá své nejmenší hodnoty právě když $x = x_0$.

(ii) Podle (i) je pro každé $y \in \mathbb{R}$

$$y \neq y_1 = \frac{\Delta}{\Gamma} \Rightarrow \Gamma y^2 - 2\Delta \cdot y > \Gamma y_1^2 - 2\Delta \cdot y_1. \tag{2}$$

(Místo nerovnosti „>“ zde nastane rovnost právě když $y = y_1$.)

Rovněž podle (i) je pro každé $x, y \in \mathbb{R}$

$$x + \varepsilon y \neq x_1 + \varepsilon y_1 \Rightarrow \gamma(x + \varepsilon y)^2 - 2\delta(x + \varepsilon y) >$$
$$> \gamma(x_1 + \varepsilon y_1)^2 - 2\delta(x_1 + \varepsilon y_1). \tag{3}$$

(Místo nerovnosti „>“ zde nastane rovnost právě když $x + \varepsilon y = x_1 + \varepsilon y_1$.)

Je-li $y \neq y_1$, plyne z (2) s přihlédnutím ke (3), že $g(x, y) > g(x_1, y_1)$. Je-li $y = y_1$ a $x \neq x_1$, je $x + \varepsilon y \neq x_1 + \varepsilon y_1$; ze (3) plyne pak s přihlédnutím ke (2), že opět $g(x, y) > g(x_1, y_1)$.

Nejmenší hodnotu $g(x_1, y_1)$ nabývá proto funkce $g(x, y)$ právě v bodě (x_1, y_1). □

Literatura

[1] Burke, M., Hodgson T., *Using Technology to Optimize and Generalize: The Least-Squares Line,* Mathematics Teacher **101** (2007), 102 - 107.

Doc. RNDr. Ladislav Beran, DrSc.
Doc. RNDr. Milan Trch, CSc., Ph.D.
Katedra matematiky České zemědělské univerzity
Kamýcká 129, 165 21 Praha 6