Metoda nejmenších čtverců

Úvod

Metodu nejmenších čtverců používáme, chceme-li naměřenými (nebo jinak získanými) body proložit křivku, např. přímku. Tedy hledáme taková reálná čísla a, b, aby graf funkce

$$f(x) = ax + b$$

nejlépe procházel mezi danými N body o souřadnicích (x_i, y_i), kde $i = 1, \cdots, N$. Požadujeme, aby proložené hodnoty $f(x_i) = ax_i + b$ byly blízké naměřeným hodnotám y_i. Tedy, aby součet druhých mocnin (čili čtverců, odtud název metody) odchylek proložených hodnot od naměřených hodnot byl co možná nejmenší. To znamená, aby funkce

$$S(a, b) = \sum_{i=1}^{N} (ax_i + b - y_i)^2$$

nabývala minima. To nastane, budou-li obě její parciální derivace nulové

$$\frac{\partial S}{\partial a} = \sum_{i=1}^{N} 2a(ax_i + b - y_i)$$

$$\frac{\partial S}{\partial b} = \sum_{i=1}^{N} 2(ax_i + b - y_i).$$

Odtud dostaneme soustavu dvou rovnic pro neznámé parametry a, b:

$$\begin{aligned} a\bar{x} + b &= \bar{y} \\ a\bar{x^2} + b\bar{x} &= \bar{xy}, \end{aligned}$$

kde pruh značí průměrné hodnoty

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$$

$$\bar{x^2} = \frac{1}{N} \sum_{i=1}^{N} x_i^2$$

$$\bar{xy} = \frac{1}{N} \sum_{i=1}^{N} x_i y_i.$$

Malý příklad

Např. pro tři body

$$(0, 0), \quad (1, 2), \quad (2, 3)$$

si sestavíme tabulku

x	y	xy	x^2
0	0	0	0
1	2	2	1
2	3	6	4
3	5	8	5

kde v posledním řádku jsou součty hodnot z příslušného sloupečku, což jsou průměrné hodnoty vynásobené počtem bodů $N = 3$. Z ní sestavíme soustavu rovnic

$$\begin{aligned} 3a + 3b &= 5 \\ 5a + 3b &= 8 \end{aligned} \tag{1}$$

a jejím řešením jsou hledané parametry

$$a = \frac{3}{2}$$

$$b = \frac{1}{6}.$$

Takže hledaná přímka má rovnici

$$y = \frac{3}{2}x + \frac{1}{6}.$$

Použití systému Maple

A jak to lze spočíst pomocí systému Maple? Můžeme použít příkaz `leastsquare`. Ten se nachází v balíku `stats` v podbalíku `fit`. Spolu s příkazem `leastsquare` zadáme ještě čtyři upřesnění:

- seznam proměnných x, y,
- rovnici funkce, jejíž graf prokládáme,
- seznam parametrů a, b,
- naměřená data jako seznam dvou seznamů čísel.

Takže zadáme tyto dva řádky:

```
> with(stats):
> fit[leastsquare[[x,y],y=a*x+b,{a,b}]]([[0,1,2],[0,2,3]]);
```

$$y = \frac{3\,x}{2} + \frac{1}{6}$$

Je povzbuzující, že jsme obdrželi stejný výsledek, jako při ručním výpočtu.

Predikce olympijských výsledků

Podívejme se na použití této metody pro reálná data. Jako příklad uvažujme olympijské výsledky ve skoku o tyči v jednotlivých letech (jak je uvádí World Almanac). Protože je čísel více, připravíme si je nejprve do dvou seznamů, které pojmenujeme `rok` a `vyska`. Údaje o výšce poté přepočítáme z palců na metry a výsledek si pojmenujeme `vyskam`. To lze zadat takto:

```
> rok:=[1896,1900,1904,1908,1912,1920,1924,
> 1928,1932,1936,1948,1952,1956,1960,1964,
> 1968,1972,1976,1980,1984,1988,1992]:
> vyska:=[130,130,137.75,146,155.5,161,155.5,
> 165.25,169.75,171.25,169.25,179,179.5,185,200.75,
> 212.75,216.5,216.5,227.5,226.25,237.25,228.25]:
> vyskam:=0.0254 * vyska:
```

Nejdříve si data prohlédneme. Pro vykreslení grafu si z nich připravíme seznam dvojic příkazem

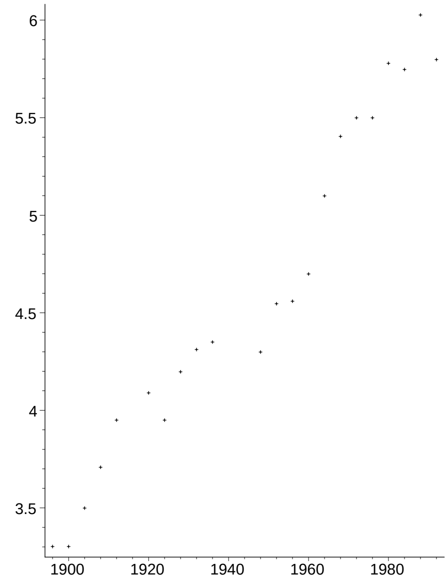
```
> body:=convert(linalg[transpose]([rok,vyskam]),listlist):
```

Samotný graf pak vytvoříme takto:

```
> with(plots):
> obr1:=listplot(body, style=POINT):
```

A takto si jej prohlédneme

```
> display(obr1);
```



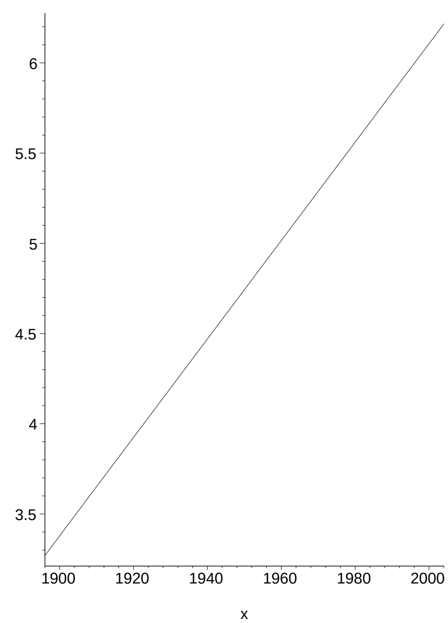
Nyní příkazem `leastsquare` proložíme přímku

```
> primka:=fit[leastsquare[[x,y],y=a*x+b,{a,b}]]([rok,vyskam]);
```

$$primka := y = 0.02727079203\,x - 48.43532721$$

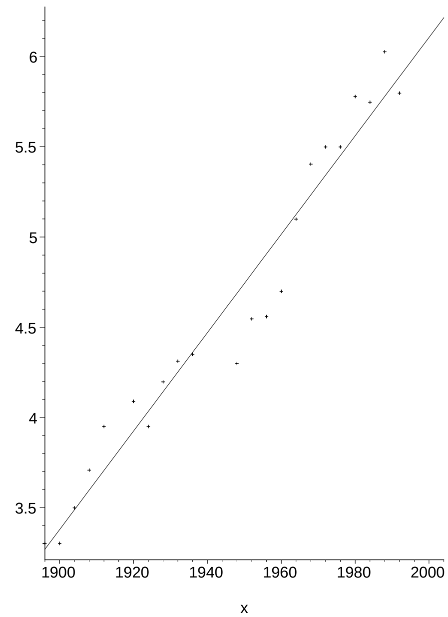
Tu si lze také prohlédnout příkazem

```
> obr2:=plot(rhs(primka),x=1896..2004):
> display(obr2);
```



Zajímavější je ale obrázek, ve kterém jsou naměřená data současně s proloženou přímkou. Ten vytvoříme takto:

```
> display({obr1,obr2});
```



Máme-li takto vytvořen model (tak se říká matematickému popisu experimentálních dat), můžeme se pokusit činit předpovědi. Budou-li se shodovat se skutečností, bude to známka, že náš model je užitečný. V našem případě se můžeme pokusit předpovědět rekord ve skoku o tyči na letních olympijských hrách v roce 2004 v Athénách:

```
> eval(rhs(primka),x=2004);
```

$$6.21534002$$

Podle našeho modelu by to mělo být přibližně 6,22 m.

Je samozřejmé, že od tohoto modelu nelze očekávat zázraky. Např. růst atletických výkonů nebude lineární s časem. Na druhou stranu u řady přírodních i technických procesů lze nalézt často překvapivě dobré modely právě tímto postupem.

Příklad: Proložte metodou nejmenších čtverců přímku body, jejichž souřadnice jsou uloženy na disku v souboru `uschovna`.

Řešení:

Soubor `uschovna` zatím nemáme, proto si jej vytvoříme. Nejdříve si připravíme x-ové s y-ové souřadnice bodů, např.

```
> xd:=[1,2,8,9];
```

$$xd := [1, 2, 8, 9]$$

```
> yd:=[5,4,3,0];
```

$$yd := [5, 4, 3, 0]$$

Tato data uložíme do souboru `uschovna` příkazem `save`

```
> save xd,yd,uschovna;
```

Nyní jsou data uložena na disku počítače. Příkazem `restart` uvedeme systém Maple do stavu, v jakém je po spuštění, tedy do stavu, kdy data nejsou uložena v operační paměti systému Maple. Tím napodobíme situaci, kdy data vznikla dříve, nezávisle na naší současné práci.

```
> restart;
```

Přesvědčíme se, že proměnná `xd` neobsahuje žádná data

```
> xd;
```

$$xd$$

Nyní příkazem `read` načteme data uložená v textové podobě v souboru na disku

```
> read uschovna;
```

$$xd := [1, 2, 8, 9]$$

$$yd := [5, 4, 3, 0]$$

Snadno se přesvědčíme, že proměnná `xd` (a podobně i proměnná `yd`) nyní obsahuje správná data

```
> xd;
```

$$[1, 2, 8, 9]$$

Nyní načteme balík `stats`

```
> with(stats):
```

a příkazem `leastsquare` proložíme přímku

```
> fit[leastsquare[[x,y], y=a*x+b, {a,b}]] ([xd,yd]);
```

$$y = -\frac{23\,x}{50} + \frac{53}{10}$$