metóda najmenších štvorcov

$x_0, \dots, x_m$ uzly, $x_i \neq x_j$ pre $i \neq j$

$f_0, \dots, f_n$

$\varphi_0, \dots, \varphi_m$ bázové funkcie

(x_i, f_i) aproximujeme funkciou $\varphi(x) = \sum_{i=0}^{m} c_i \varphi_i(x)$

tak, aby výraz

$$\sigma^2(c_0, \dots, c_m) = \sum_{j=0}^{n} \left(f_j - \varphi(x_j)\right)^2$$

bol minimálny.

príklad: Pre hodnoty

x_i	0,00	0,25	0,5	0,75	1,00
f_i	1,0000	1,2840	1,6487	2,1170	2,7183

nájdite najlepšiu aproximáciu polynómom 2. stupňa.

aproximujeme polynómom 2. stupňa

$$\Rightarrow \varphi_0 \equiv 1, \ \varphi_1 = x, \ \varphi_2 = x^2$$

$$\sigma^2(c_0, c_1, c_2) = \sum_{j=0}^{m} \left(f_j - \varphi(x_j)\right)^2 = \sum_{j=0}^{4} \left(f_j - c_0 - c_1 x_j - c_2 x_j^2\right)^2$$

$$\frac{\partial}{\partial c_0} \sigma^2 = -2 \sum_{j=0}^{\cancel{m}\,4} \left(f_j - c_0 - c_1 x_j - c_2 x_j^2\right) = 0$$

$$\frac{\partial}{\partial c_1} \sigma^2 = -2 \sum_{j=0}^{\cancel{m}\,4} \left(f_j - c_0 - c_1 x_j - c_2 x_j^2\right) x_j = 0$$

$$\frac{\partial}{\partial c_2} \sigma^2 = -2 \sum_{j=0}^{4} \left(f_j - c_0 - c_1 x_j - c_2 x_j^2\right) x_j^2 = 0$$

I. $(m+1)c_0 + \left(\sum_{j=0}^{\cancel{m}\,4} x_j\right)c_1 + \left(\sum_{j=0}^{\cancel{m}\,4} x_j^2\right)c_2 = \sum_{j=0}^{\cancel{m}\,4} f_j$

II. $\left(\sum_{j=0}^{4} x_j\right)c_0 + \left(\sum_{j=0}^{4} x_j^2\right)c_1 + \left(\sum_{j=0}^{4} x_j^3\right)c_2 = \sum_{j=0}^{4} f_j x_j$

III. $\left(\sum_{j=0}^{4} x_j^2\right)c_0 + \left(\sum_{j=0}^{4} x_j^3\right)c_1 + \left(\sum_{j=0}^{4} x_j^4\right)c_2 = \sum_{j=0}^{4} f_j x_j^2$

ak definujeme diskrétny skalárny súčin funkcií g, h predpisom $(g,h) = \sum_{j=0}^{m} g(x_j)\cdot h(x_j)$, potom má systém rovníc $\frac{\partial}{\partial c_i}\sigma^2 = 0$, $i \in \{0, \dots, m\}$ v prípade lineárnej metódy najmenších štvorcov vo všeobecnosti tvar

$$Ac = b, \text{ kde } A = \begin{pmatrix} (\varphi_0,\varphi_0) & (\varphi_0,\varphi_1) & (\varphi_0,\varphi_2) \\ (\varphi_1,\varphi_0) & (\varphi_1,\varphi_1) & (\varphi_1,\varphi_2) \\ (\varphi_2,\varphi_0) & (\varphi_2,\varphi_1) & (\varphi_2,\varphi_2) \end{pmatrix} \quad c = \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix}$$

$$b = \begin{pmatrix} (f,\varphi_0) \\ (f,\varphi_1) \\ (f,\varphi_2) \end{pmatrix} \rightarrow \text{porovnajte pre } \begin{array}{l} \varphi_0(x) \equiv 1 \\ \varphi_1(x) = x \\ \varphi_2(x) = x^2 \end{array}$$

I. $5c_0 + 2{,}5c_1 + 1{,}875c_2 = 8{,}768$

II. $2{,}5c_0 + 1{,}875c_1 + 1{,}5625c_2 = 5{,}4514$

III. $1{,}875c_0 + 1{,}5625c_1 + 1{,}3828125c_2 = 4{,}4015375$

$c_0 = 1{,}0051$ polynóm $c_2x^2 + c_1x + c_0$ najlepšie

$c_1 = 0{,}8641$ aproximuje dáta (x_i, f_i) spomedzi poly-

$c_2 = 0{,}8437$ nómov 2. stupňa v zmysle MNŠ

Príklad 1. Pomocou jednoduchých populačných modelov a metódy najmenších štvorcov študujte vývoj ľudskej populácie. Použite voľne dostupné dáta o veľkosti populácie v priebehu posledných rokov.

Označme symbolom $x(t)$ veľkosť ľudskej populácie v čase t. Najjednoduchší populačný model je tzv. Malthusov model. Predpokladá, že rýchlosť rastu populácie (akéhokoľvek biologického druhu) je priamo úmerná jej aktuálnej veľkosti. To sa dá vyjadriť nasledujúcou diferenciálnou rovnicou:

$$x'(t) = \lambda x(t),$$

kde λ je miera rastu populácie. Všeobecné riešenie tejto rovnice spĺňajúce počiatočnú podmienku $x(t_0) = x_0$ má tvar

$$\begin{aligned}
x'(t) - \lambda x(t) &= 0, && / \cdot \mathrm{e}^{-\lambda t} \\
x'(t)\mathrm{e}^{-\lambda t} - \lambda x(t) \cdot \mathrm{e}^{-\lambda t} &= 0, \\
\left(x(t)\mathrm{e}^{-\lambda t}\right)' &= 0, && \Big/ \int_{t_0}^{t} \\
\int_{t_0}^{t} \left(x(s)\mathrm{e}^{-\lambda s}\right)' \, \mathrm{d}s &= 0, \\
x(t)\mathrm{e}^{-\lambda t} - x(t_0)\mathrm{e}^{-\lambda t_0} &= 0, \\
x(t)\mathrm{e}^{-\lambda t} &= x_0 \mathrm{e}^{-\lambda t_0},
\end{aligned}$$

$$x(t) = x_0 \mathrm{e}^{\lambda(t-t_0)}.$$

Aby sme presne určili priebeh a vývoj populácie, potrebujeme určiť neznáme parametre x_0 a λ (t_0 je zvolený časový okamih). Na to môžeme použiť metódu najmenších štvorcov. Ešte pred tým než začneme, mali by sme si uvedomiť, že funkcia $x(t)$ nie je lineárna v premennej λ, teda nie je vyjadrená ako lineárna kombinácia bázových funkcií tak, aby koeficienty lineárnej kombinácie boli hľadané parametre. To môžeme napraviť aplikáciou logaritmu a následnou linearizáciou problému:

$$\begin{aligned}
x(t) &= x_0 \mathrm{e}^{\lambda(t-t_0)}, \qquad / \log(\cdot) \\
\log(x(t)) &= \log(x_0 \mathrm{e}^{\lambda(t-t_0)}) = \log(x_0) + \lambda(t - t_0) \\
&= \underbrace{(\log(x_0) - \lambda t_0)}_{c_0} \cdot 1 + \underbrace{\lambda}_{c_1} \cdot t.
\end{aligned}$$

Teraz už môžeme použiť lineárnu metódu najmenších štvorcov. Bázové funkcie sú $\varphi_0(t) \equiv 1$ a $\varphi_1(t) = t$ a hľadáme najlepšiu aproximáciu funkcie $\log(x(t))$ spomedzi funkcií $\varphi(t) = c_0\varphi_0(t) + c_1\varphi_1(t)$, teda hľadáme minimum funkcie

$$\sigma^2(c_0, c_1) = \sum_{j=0}^{n} \left(\log(x_j) - c_0 - c_1 t_j\right)^2.$$

V nasledujúcej tabuľke sú uvedené veľkosti ľudskej populácie vo vybraných rokoch a ich logaritmy.

t_i	x_i	$\log(x_i)$
t_0	3 034 949 748	21,83346
t_1	3 700 437 046	22,03172
t_2	4 458 003 514	22,21797
t_3	5 327 231 061	22,39610
t_4	6 143 493 823	22,53866
t_5	6 956 823 603	22,66299
t_6	7 794 798 739	22,77672

Môžeme počítať:

$$\frac{\partial}{\partial c_0}\sigma^2(c_0,c_1) = -2\sum_{j=0}^{6}\left(\log(x_j) - c_0 - c_1 t_j\right) = 0,$$
$$\frac{\partial}{\partial c_1}\sigma^2(c_0,c_1) = -2\sum_{j=0}^{6}\left(\log(x_j) - c_0 - c_1 t_j\right) t_j = 0.$$

Po úprave dostaneme lineárny systém

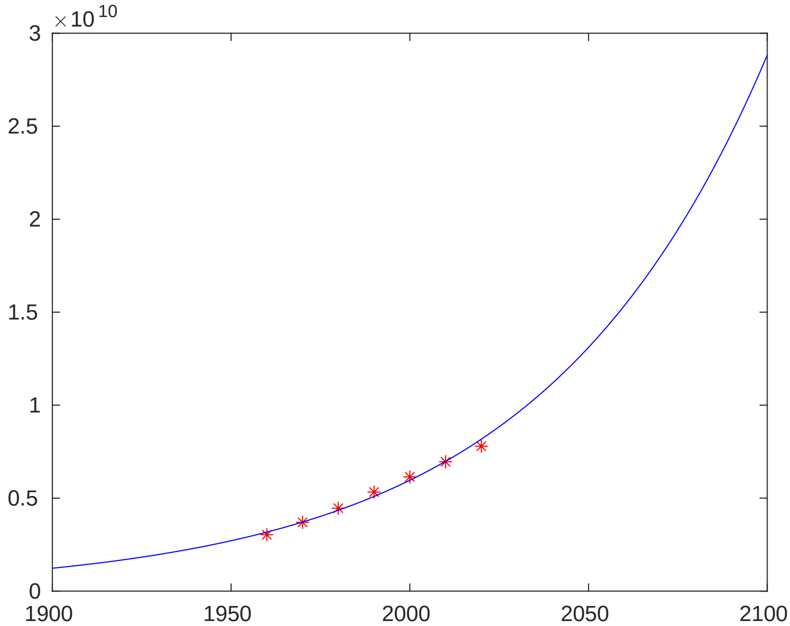
$$14c_0 + \left(2\sum_{j=0}^{6} t_j\right) c_1 = 2\sum_{j=0}^{6}\log(x_j),$$
$$\left(2\sum_{j=0}^{6} t_j\right) c_0 + \left(2\sum_{j=0}^{6} t_j^2\right) c_1 = 2\sum_{j=0}^{6}\log(x_j)\cdot t_j.$$

Dosadením hodnôt z tabuľky dostaneme systém

$$14c_0 + 27860c_1 = 312,91524,$$
$$27860c_0 + 55447000c_1 = 622789,5878.$$

Odčítaním 1990-násobku prvej rovnice od druhej eliminujeme c_0 a tým pádom môžeme postupne dopočítať c_1 a c_0 a následne aj pôvodné parametre λ a x_0:

$$5600c_1 = 88,2524 \quad\Rightarrow\quad c_1 = \lambda = 0,0157593$$
$$\Rightarrow\quad x_0 = \mathrm{e}^{c_0+\lambda t_0} \approx 3\,173\,801\,400.$$



Pre parameter x_0 sme mohli použiť tabuľkovú hodnotu a pomocou najmenších štvorcov dopočítať len jeden parameter λ. Avšak je zrejmé, že hodnoty v tabuľke nie sú presné rovnako, ako je to vo všeobecnosti so všetkými meranými veličinami. To, že aj koeficient x_0 považujeme za neurčitý, nám umožní lepšie vyhladiť chyby v dátach.

Funkcia $x(t) = x_0\mathrm{e}^{\lambda}(t - t_0)$ by teda mala približne vyjadrovať veľkosť ľudskej populácie v ľubovoľnom čase t. Je zrejmé, že tento model má mnohé nedostatky, z ktorých najevidentnejší je dôsledkom skutočnosti $\lim_{t\to\infty} x(t) = \infty$. To znamená, že tento model predpovedá neobmedzený rast populácie, čo je nereálne. Je to spôsobené tým, že v Malthusovom modeli sa neuvažuje o kapacite prostredia, v ktorom populácia žije. Ako poslednú poznámku k tomuto modelu uveďme, že je použiteľný v prípade, kedy je veľkosť populácie dostatočne malá v porovnaní s kapacitou prostredia.

O niečo lepším modelom je tzv. Verhulstova logistická rovnica, ktorej tvar je nasledujúci:

$$x'(t) = rx(t)\left(1 - \frac{x(t)}{K}\right).$$

Ak $x(t)$ je podstatne menšie ako K, potom je výraz v zátvorke blízky 1 a rovnica je blízka Malthusovmu modelu, teda opäť môžeme očakávať exponenciálny rast. Avšak pre $x(t)$ blízke K je výraz v zátvorke blízky nule. To znamená, že rast (zmena) populácie bude veľmi pomalá. Preto môžeme parameter K interpretovať ako kapacitu prostredia. Pozrime sa na riešenie tejto rovnice.

$$\frac{x'(t)}{x(t)\left(1 - \frac{x(t)}{K}\right)} = r, \qquad \Big/ \int_{t_0}^{t}$$

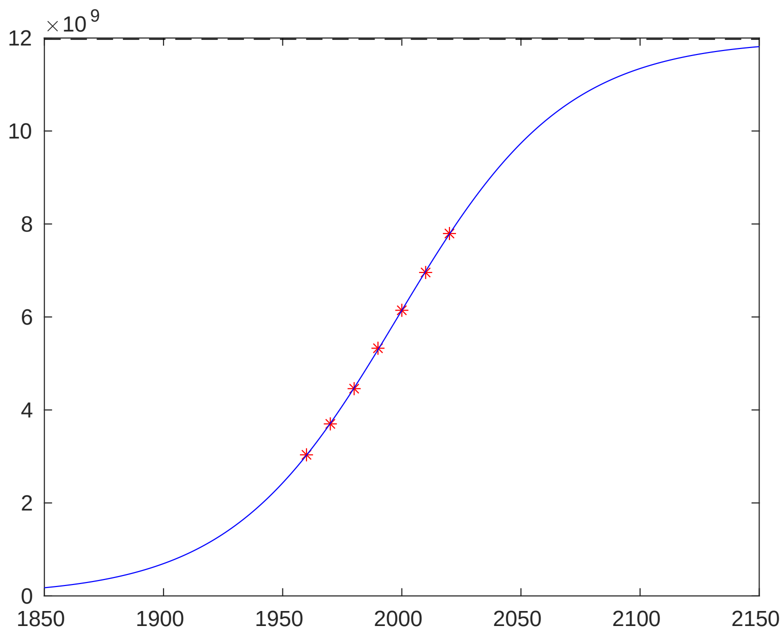
$$\left|\begin{array}{c} u = x(s) \\ \mathrm{d}u = x'(s)\,\mathrm{d}s \\ t \rightsquigarrow x(t) \\ t_0 \rightsquigarrow x_0 \end{array}\right| \qquad \int_{t_0}^{t} \frac{x'(s)}{x(s)\left(1 - \frac{x(s)}{K}\right)}\,\mathrm{d}s = \int_{t_0}^{t} r\,\mathrm{d}s = r(t - t_0),$$

$$\begin{aligned}
\int_{t_0}^{t} \frac{1}{u\left(1 - \frac{u}{K}\right)}\,\mathrm{d}u &= r(t - t_0),\\
\int_{t_0}^{t} \frac{1}{u} + \frac{1}{K - u}\,\mathrm{d}u &= r(t - t_0),\\
\left[\log\left(\frac{u}{K - u}\right)\right]_{x_0}^{x(t)} = \left[\log(u) - \log(K - u)\right]_{x_0}^{x(t)} &= r(t - t_0),\\
\log\left(\frac{\frac{x(t)}{K - x(t)}}{\frac{x_0}{K - x_0}}\right) = \log\left(\frac{x(t)}{K - x(t)}\right) - \log\left(\frac{x_0}{K - x_0}\right) &= r(t - t_0),\\
\frac{x(t)}{K - x(t)} &= \frac{x_0}{K - x_0}\mathrm{e}^{r(t - t_0)},\\
x(t)\left(1 + \frac{x_0}{K - x_0}\mathrm{e}^{r(t - t_0)}\right) &= \frac{Kx_0}{K - x_0}\mathrm{e}^{r(t - t_0)},\\
x(t) &= \frac{\frac{Kx_0}{K - x_0}\mathrm{e}^{r(t - t_0)}}{1 + \frac{x_0}{K - x_0}\mathrm{e}^{r(t - t_0)}},
\end{aligned}$$

$$x(t) = \frac{Kx_0}{x_0 + (K - x_0)\mathrm{e}^{-r(t-t_0)}}.$$

V tomto prípade už nie je zrejmé, ako vyjadriť funkciu $x(t)$ v tvare lineárnej kombinácie bázových funkcií tak, aby jej koeficienty boli funkciami voľných parametrov K, x_0 a r. Nič nám však nebráni v tom, aby sme použili matlab a nelineárnu metódu najmenších štvorcov.

```
function x = logisticka(t,K,r,x0)

x = (K*x0)./(x0+(K-x0)*exp(-r*(t-1960)));

end

rok=[1960; 1970; 1980; 1990; 2000; 2010; 2020];
pop=[3034949748; 3700437046; 4458003514; 5327231061; 6143493823;
   6956823603; 7794798739];

ft=fittype('logisticka(t,K,r,x0)','independent','t','dependent','
   x');
op=fitoptions('Method','NonlinearLeastSquares','Lower',[0 0 0],'
   StartPoint',[10000000000 0.02 3000000000],'TolX',1e-15,'
   MaxFunEvals',10000,'MaxIter',10000,'TolFun',1e-15);
log=fit(rok,pop,ft,op);

xx=linspace(1850,2150,1000);
yy=feval(log,xx);
K=coeffvalues(log);
plot(rok,pop,'r*',xx,yy,'b-',[1850 2150],[K(1) K(1)],'k--')
```



Podľa tohto modelu je pravdepodobná kapacita našej planéty pribižne 12 miliárd ľudí.

✈