

## 7. domača naloga - Linearna regresija

Urban Škudnik

2009-05-04

### **Abstract**

V tej domači nalogi smo se ukvarjali z linearno zvezo  $y = kx$ , ki je najpreprostejša in daleč najbolj pogosta zveza med dvema fizikalnima količinama, zlasti v ozkih intervalih ko lahko aproksimiramo z  $\Delta y = k\Delta x$ . Koeficient  $k$  pri tem za majhne  $\Delta x$  limitira k vrednosti odvoda  $\frac{dy}{dx}$ .

# 1 Naloga

**Hitrost toka** Za meritve v datoteki "HitrostTokaOdFrekvence.txt" določi parametra najboljše premice. Ker so podane napake hitrosti lahko določiš tudi  $\chi^2$ .

**Reševanje** Prvo nalogo sem rešil z programom *Mathematica*. Podatke sem najprej naložil z ukazom `ReadList`, nato sem z ukazom `LinearModelFit` iz prvega in drugega stolpca izdelal linearni model, ki že po privzetih nastavitvah izpiše vrednosti  $k$  in  $n$ . Takšen linearni model še ni upošteval napak, vendar je za začetno razumevanje lažji.

```
hitmod = LinearModelFit[#[[1]], #[[2]]] & /@ hit, x, x]
```

Takšen linearni model nam je vrnil za parametra vrednosti  $k = 1.00808$  in  $n = -2.09547$ .

Vrednosti  $\chi^2$  lahko preberemo iz t.i. *tabele ANOVA (Analysis of Variance)* z ukazom `hitmod["ANOVATable"]` v stolpcu *SS (Sum of squares)* v vrstici z napakami (*Error*) in je brez upoštevanja napak  $\chi^2 = 0.08593$ .

**Obtežitev z napakami** Ker smo imeli podane tudi podatke za napake sem v ukaz `LinearModelFit` podal tudi parameter `Weights` s katerim lahko posamezne točke tudi pravilno obtežimo (vrednost vsake točke delimo z  $\epsilon_i^2$ ).

Končni ukaz je bil torej

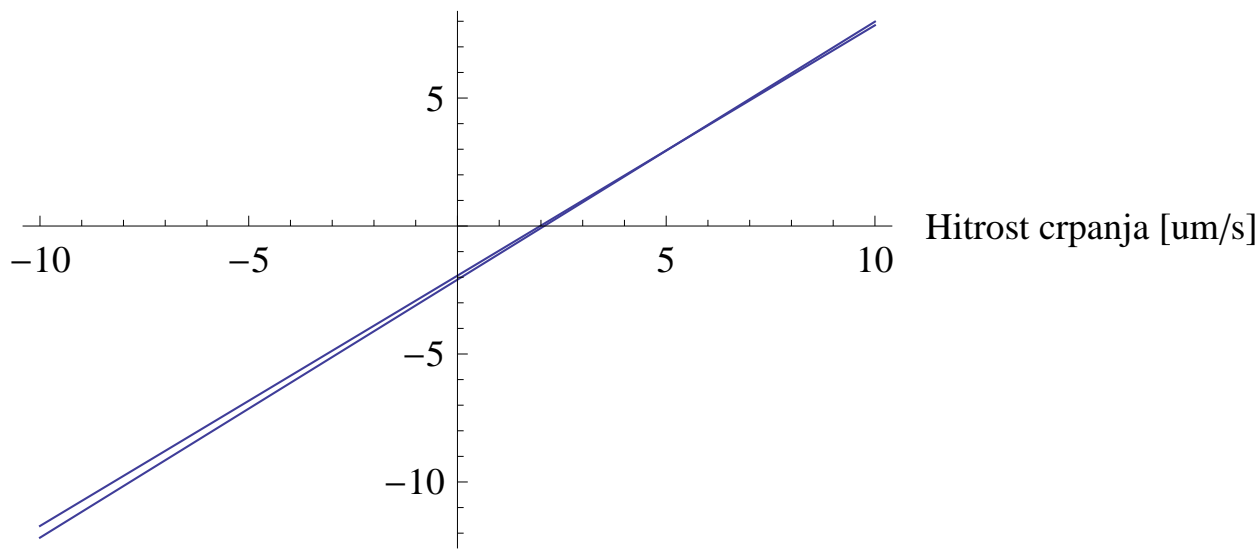
```
hitmode = LinearModelFit[#[[1]], #[[2]]] & /@ hit, x, x, Weights -> {1/#[[3]]^2 & /@ hit}[[1]]]
```

Z upoštevanjem napak nam je linearni model za parametra vrnil vrednosti  $k = 0.978093$  in  $n = -1.93864$ , iz *tabele ANOVA* pa vidimo da je vrednost  $\chi^2 = 5.43746$ .

Če obe premici narišemo na isti graf za primerjavo vidimo, da se linearna modela niti ne razlikujeta zelo.

## Primerjava linearnih modelov

Frekvenca vrtenja [/s]



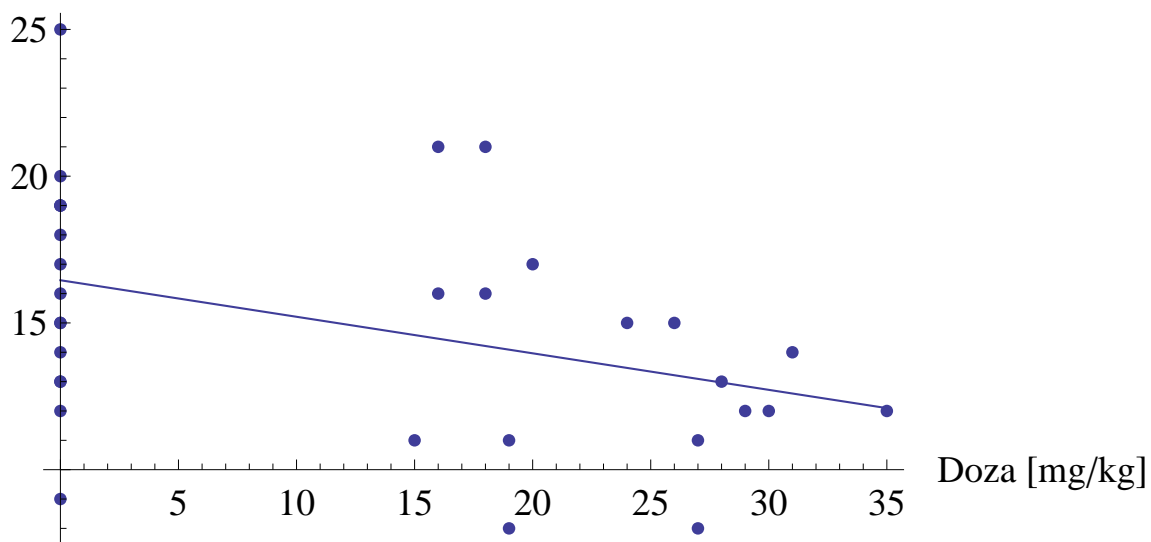
## 2 Naloga

**Tintin** Skozi oblak podatkov "Tintin.dat" potegni najboljšo premico. Uporabiš lahko kar korelacijske rezultate iz naloge 6.2.

**Reševanje** Tudi druge naloge sem se lotil z *Mathematico* z enakim postopkom kot prej, pri čemer je linearni model, ki ga ustvari *Mathematica* vrnil za parametra vrednosti  $k = -0.124488$  in  $n = 16.4546$ . Če prikažemo ta linearni model na istem grafu kot podatke precej lažje vidimo, da znatno povečanje odmerka le počasni zmanjša zvonjenje v ušesih.

### Prikaz uspesnosti zdravljenja

Trajanje zvonjenja [h]



Primerjava

### 3 Naloga

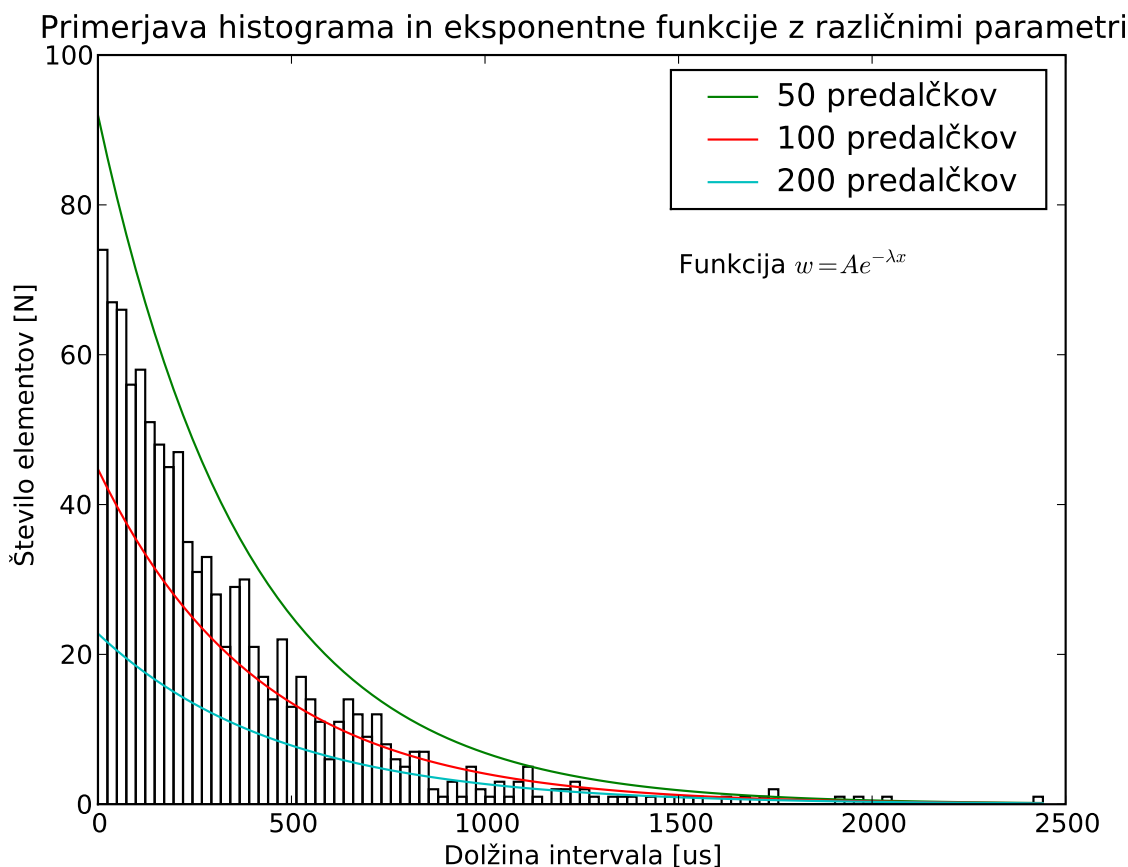
**Interval** Skozi točke v histogramu podatkov "Interval.dat" poskusi potegniti najboljšo eksponentno funkcijo  $w = Ae^{-\lambda x}$ , ki jo moramo najprej predelati v linearno zvezo. Z logaritmiranjem dobimo  $\ln(w) = \ln(A) - \lambda x$ . V grafu  $y = \ln(w)$  od  $x$  sta koeficienta premice  $k = -\lambda$  in  $n = \ln(A)$ . Po teoriji verjetnosti mora biti koeficient  $\lambda$  enak recipročni povprečni vrednosti histograma.

**Reševanje** Za reševanje te naloge nisem več uporabljal *Mathematice* pač pa sem posegel po bolj naprednem orodju (*Matplotlib*), pri čemer sem postopal po v nalogi predpisanem postopku, ki po logaritmicizaciji eksponentno krivuljo  $w = Ae^{-\lambda x}$  preoblikuje v izraz  $\ln w = -\lambda x + \ln A$ , kar je običajna linearna zveza s koeficientom  $-\lambda$  in ostankom  $\ln A$ . Tako lahko po logaritmiranju števila časovnih intervalov v določenem predalčku v histogramu iz dobljenega niza podatkov z orodjem *scipy* z ukazom `polyfit` dobimo koeficiente členov polinoma oziroma v našem primeru, ko imamo opravka z polinomom prve stopnje, vrednosti  $k$  in  $n$ .

Po malo eksperimentiranja sem se odločil, da bom izdelal in primerjal vrednosti za histogram z 50, 100 in 200 predalčki.

/	50	100	200
k	-0.00259840	-0.00239389	-0.00213357
n	4.52174573	3.80062196	3.12579433
A	91.99605	44.72899	22.77798

Za te različne vrednosti sedaj narišemo graf, ki ga primerjamo s samim histogramom (s 100 predalčki).



## Koda

```
inter = csv.reader(open("Interval.dat"), delimiter="\n")
data = list()
for row in inter:
    data.append(float(row[0]))

p = 100
hist = plot.hist(data, bins=p, fill=False)
histx = hist[1] # vrednost intervala
histy = hist[0] # stevilo elementov v predalcku

lnx, lny = list(), list()
for s in range(p):
    if histx[s] == 0 or histy[s] == 0:
        pass
    else:
        lnx.append(histx[s])
        lny.append(math.log(histy[s]))

k, n = polyfit(lnx, lny, 1)
A = math.exp(n)

print "predalckov %s: k = %s, n = %s, A = %s" % (p, k, n, math.exp(n))
plot.plot(histx, [A*math.exp(k*x) for x in histx], label="$w = Ae^{-\lambda x}$")

plot.xlabel(u"Dolžina intervala [us]")
plot.ylabel(u"Število elementov [N]")
plot.title(u"Primerjava histograma in eksponentne funkcije z različnimi parametri")
plot.legend()
plot.show()
```

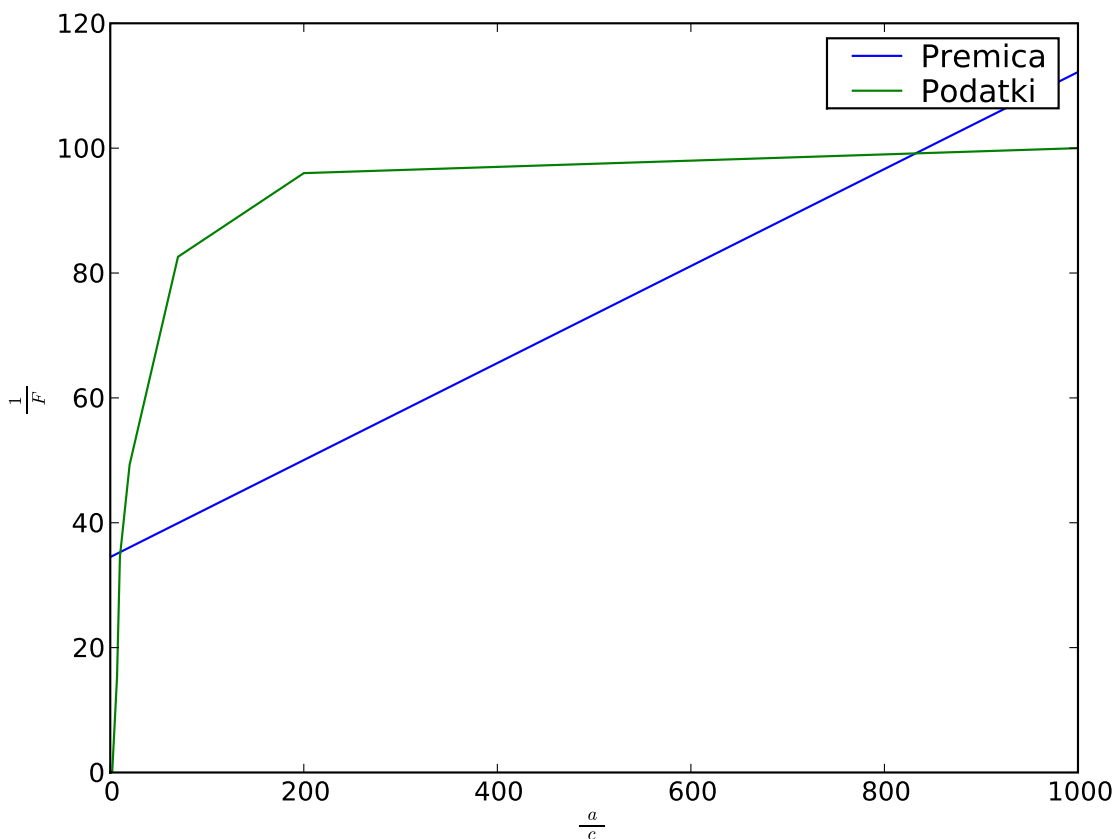
## 4 Naloga

**Adrenalin** Teorija kemijske kinetike napove za sigmoidno krivuljo iz podatkov "Adrenalin.dat" (naloga 1.1) naslednjo odvisnost  $\frac{F}{F_{max}} = \frac{c}{a+c}$ , kjer pomeni  $a$  koncentracijo s polovičnim maksimalnim učinkom. Določi koeficienta  $F_{max}$  in  $a$ . Pretvori v linearno zvezo – ena pot je uvedba recipročnih spremenljivk  $\frac{1}{F}$  in  $\frac{1}{c}$ , druga pa je uvedba spremenljivke  $\frac{c}{F}$ .

**Reševanje** Kljub temu da naloga zahteva izračun  $F_{max}$ , tega ne moremo storiti saj imamo podatke za silo podane v odstotkih in ne v absolutnih enotah.

Za izračun le-tega sem najprej pretvoril v enačbo z recipročnimi spremenljivkami:  $\frac{1}{F} = \frac{a}{cF_{max}} + \frac{1}{F_{max}}$ , nakar sem predvideval, da lahko  $F_{max}$  zanemarim saj je to relativni del in ker je to maksimum (torej 1) v nadaljnjih računih naj nebi igral vloge. Tako sem nadaljeval z enačbo  $y = \frac{1}{F} = \frac{a}{F} + 1 = kx + n$ .

Iz te enačbe sem na enak način kot pri prejšnji nalogi z orodjem *scipy* z ukazom `polyfit` izračunal  $k = 0.077690$  in  $n = 34.503163$ . Za primerjavo sem narisal graf z podatki in premico s tako pridobljenima vrednostima parametrov.



### Primerjava

**Reševanje** Po tem poročilu sem iz enačbe  $\frac{a}{c} + 1 = kc + n$  izrazil  $a$  in dobil za le-tega vrednost  $a = 36.9437470585$ .