

# 7. Linearna regresija

Marko Pirc

5. maj 2009

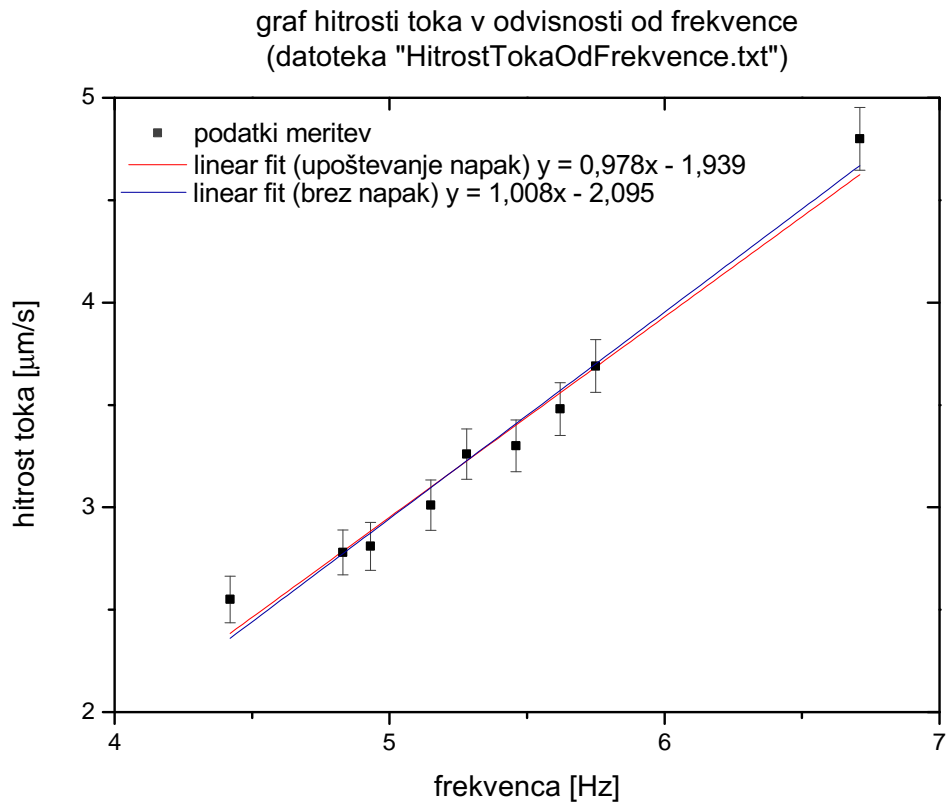
## Naloga 1

### Navodila

Za meritve<sup>1</sup> v datoteki "HitrostTokaOdFrekvence.txt" (naloga 6.1) določi parametra najboljše premice. Ker so podane napake hitrosti, lahko določiš tudi  $\chi^2$ .

### Rezultati

Rezultate sem izračunal s programom Origin, kateri ima že vgrajeno funkcijo v meniju Analysis → Fitting → Fit Linear in Residual sum of square, kateri vrne vrednost  $\chi^2$ .



<sup>1</sup>Blaz Kavčič, Dušan Babič in Igor Poberaj, Mikrofluidično vezje z mikročrpalko, OMF 56 (2009) 1.

Za primerjavo sem naredil fitanje in izračune za oba primera - z in brez upoštevanja napak. Na zgornjem grafu se vidi rahlo odstopanje med premico, fitano z upoštevanjem napak in tisto brez upoštevanja. Rezultat sem preveril tudi s programom **Mathematica** in funkcijo, opisano na spletni strani ROVF, katera mi je dala enake rezultate:

-enačba linearne zveze in vrednost  $\chi^2$  z upoštevanjem napak:

$$y = 0,978 \cdot x - 1,939; \quad \chi^2 = 0,086.$$

**(Opomba:** Če računamo  $\chi^2$  po formuli v navodilih dobimo vrednost 5,44 ker tam ni napisano, da je potrebno vsak člen vsote deliti s kvadratom napake!)

-enačba linearne zveze brez upoštevanja napak:  $y = 1,008 \cdot x - 2,095$

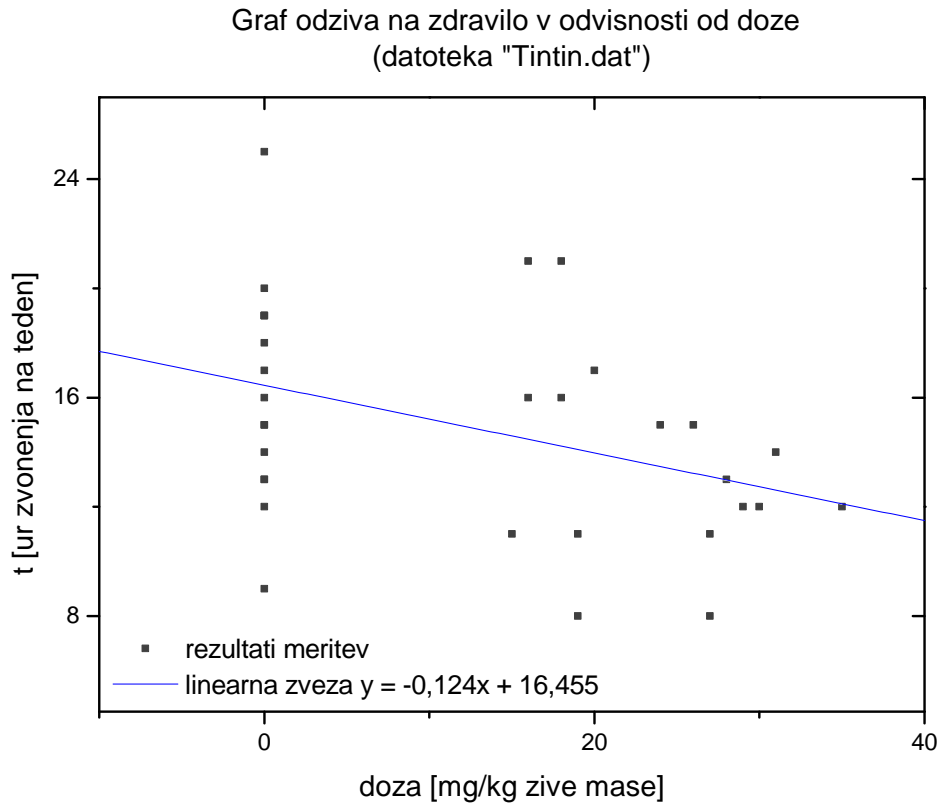
Pri tem je prvi parameter linearne zveze naklonski koeficient  $k$  in drugi parameter začetna vrednost  $n$ .  
Opomba: Vse vrednosti so zaokrožene na tri decimalna mesta.

## Naloga 2

### Navodila

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

### Rezultati



Na zgornji graf nisem narisal primerjalne premice pofitane s programom, saj se popolnoma ujema z izračunano in se je na grafu sploh ne vidi. Za izračun rezultatov pri tej nalogi sem uporabil korelacijske rezultate iz naloge 6.2:  $R = -0,394$ ,  $\bar{x} = 12,438$ ,  $\bar{y} = 14,906$ ,  $\sigma_x = 12,672$ ,  $\sigma_y = 4,003$ .

Iz tega hitro sledi:

$$k = R \cdot \frac{\sigma_y}{\sigma_x} = -0,124$$

Če vemo, da premica s tem naklonskim koeficientom poteka skozi točko s koordinatami  $(\bar{x}, \bar{y})$  lahko preprosto izračunamo začetno vrednost po enačbi:

$$n = \bar{y} - k \cdot \bar{x} = 16,455.$$

Iskana enačba premice je:

$$y = -0,124 \cdot x + 16,455.$$

## Naloga 3

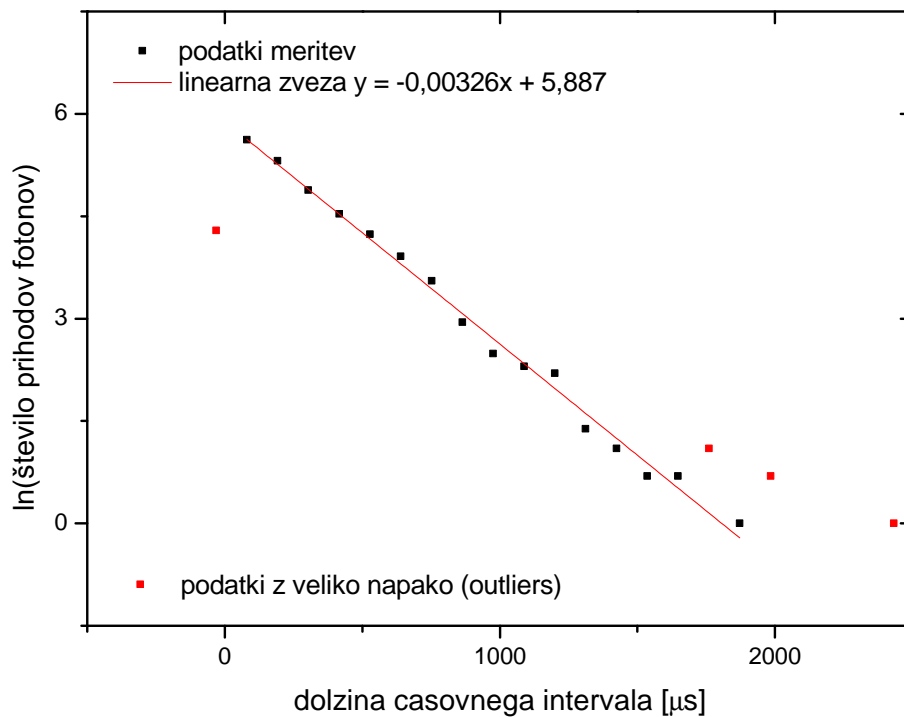
### Navodila

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.

### Rezultati

Najprej sem pripravil histogram iz podatkov datoteke "Interval.dat", v katerem sem podatke opredelil v 25 predalčkov (kasneje za preizkus tudi za 5) (program `Origin` in vgrajena funkcija `Frequency Counts`). Nato sem vrednosti na  $y$  osi logaritmiral (orodje `Excel`), drugače označil (vrednosti v nadaljnjih izračunih niso upoštewane) vse "outliere", ter dobljeni graf pofital z linearno premico:

histogram prihodov fotonov v detektor v odvisnosti od casovnega intervala med njimi (datoteka "Interval.dat")



Iz česar sledi:

$$k = -\lambda = -0,00326$$

in

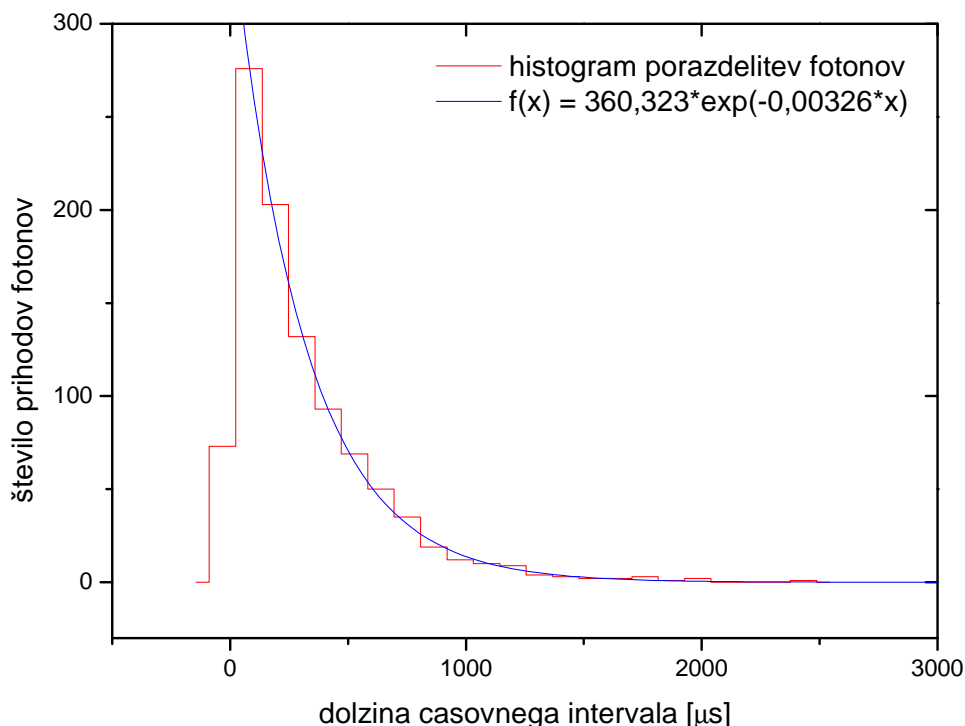
$$n = 5,887 = \ln(A) \longrightarrow A = 360,323.$$

Iskana eksponentna funkcija je torej:

$$w = 360,323 \cdot e^{-0,00326 \cdot x}.$$

Naš histogram z izračunano najboljšo eksponentno funkcijo izgleda:

histogram prihodov fotonov v detektor v odvisnosti od casovnega intervala med njimi (datoteka "Interval.dat")



**Komentar:** Pravilneje bi bilo, če bi izbral število predalčkov tako, da nebi bila vrednost nobenega 0. To je pri  $n=11$ . Če bi histogram po predalčenju tudi normiral, bi bila izračunana vrednost  $\lambda$  enaka, drugačna bi bil le koeficient  $A$ , kateri funkcijo reztegne ali skrči v smeri  $y$  osi.

**Dodatek:**

Navodilo naloge nas sprašuje tudi o primerjavi med koeficientom  $\lambda$  in povprečno vrednostjo  $\bar{y}$  histograma. emo, da se slednjo izračuna kot (delal sem z Excelom in Originom):

$$\bar{y} = \frac{1}{\sum ("BinCenter" \times "RelativeFrequency")}$$

V mojem primeru (25 predalčkov) tako dobljena vrednost  $\lambda$  znasa 0,00321 kar se dobro ujema s prvotno izračunano vrednostjo.

Pogledal sem ujemanje tudi za druge  $n$ -je. Za  $n=5$  je enačba iskane funkcije:

$$w = 1225,4 \cdot e^{-0,00262 \cdot x}$$

in znaša  $\lambda$  izračunana iz histograma 0,00344 kar je že nekoliko večje od prave vrednosti.

## Naloga 4

### Navodila

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

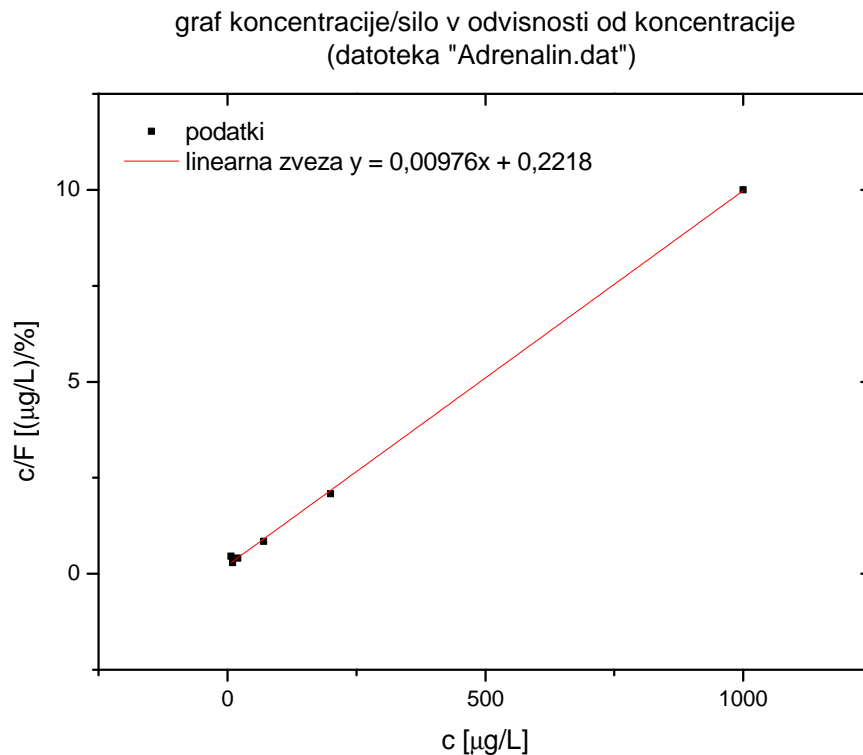
### Rezultati

Pri tej nalogi nisem bil najbolj siguren v pravilnost reševanja, saj nam datoteka z meritvami podaja le delež sile v %, ne pa same sile v  $N$ . Nalogo sem potem rešil na način kot nam je pri predavanju povedal prof. Kodre. Vse podatke sem si pripravil v programu Excel, narisal ter fital pa z vgrajeno funkcijo v programu Origin. Opomba: Vrednosti  $F_{max}$  in  $a$  sem izračunal na obe načina, ki sta opisana v navodilih.

(I) uvedba spremenljivke  $c/F$ :

$$\frac{F}{F_{max}} = \frac{c}{a + c} \quad \text{preoblikujemo v} \quad \frac{c}{F} = \frac{1}{F_{max}} \cdot c + \frac{a}{F_{max}}.$$

Tako si pripravimo tudi podatke, ter narisemo graf in določimo premico z najboljšim linearnim ujemanjem:



Velja:

$$k = \frac{1}{F_{max}} \quad \longrightarrow \quad F_{max} = 102,46 \text{ N} \quad \text{in} \quad n = \frac{a}{F_{max}} \quad \longrightarrow \quad a = 22,73 \frac{\mu}{L}$$

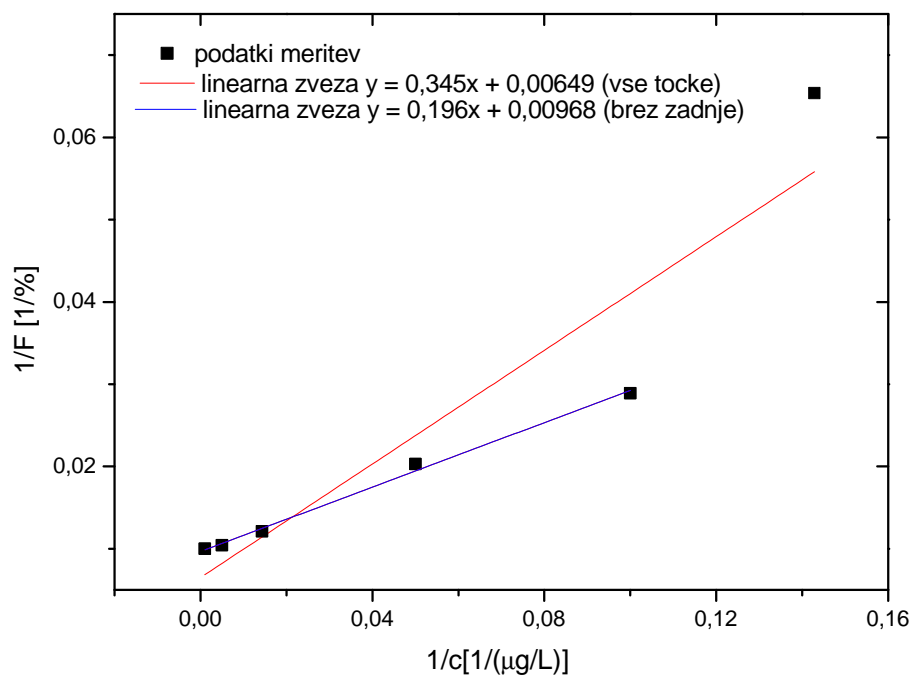
## (II) uvedba spremenljivk $1/F$ in $1/c$ :

Pri tem načinu naš začetni izraz dobi obliko:

$$\frac{1}{F} = \frac{a}{F_{max}} \cdot c + \frac{1}{F_{max}}$$

Najprej sem si v Excelu ustrezno pripravil podatke, jih narisal in pofital v Originu. Za primerjavo sem naredil dva fita, enega za vse točke in enega brez točk, za katere sem ocenil da preveč odstopajo od linearnega modela (v tem primeru je to zadnja točka):

graf recipročnih vrednosti sile v odvisnosti od koncentracije raztopine  
(datoteka "Adrenalin.dat")



kjer velja:  $k = \frac{a}{F_{max}}$  in  $n = \frac{1}{F_{max}}$

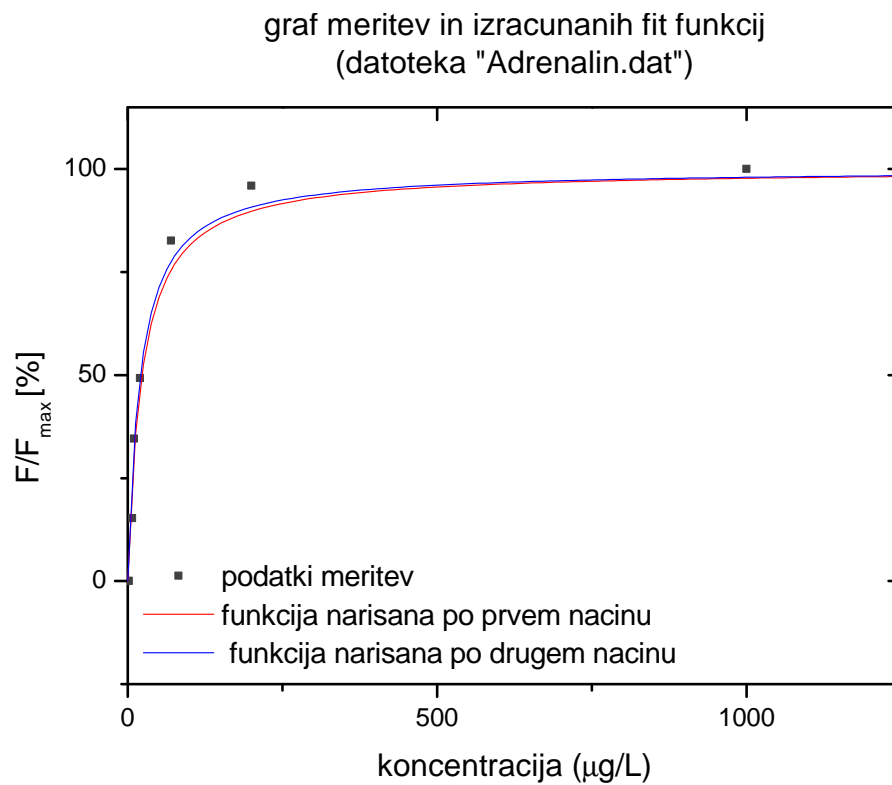
Ko izračunamo  $F_{max}$  in  $a$  za oba primera fitanja, dobimo:

-uporaba vseh točk:  $F_{max} = 154,08 N$  in  $a = 53,16 \frac{\mu g}{L}$

-brez "outliersev" (pravilnejši rezultat):  $F_{max} = 103,31 N$  in  $a = 20,25 \frac{\mu g}{L}$

**Komentar:** Vidi se, da je pravi način tisti, pri katerem sem odstranil zadnjo točko (ki je izjemno odstopala). Tako dobljeni rezultati se zelo dobro ujema z rezultati dobljenimi s prvim načinom.

Graf naše funkcije izračunane po obeh načinih v primerjavi z originalnimi podatki izgleda nekako takole:



Funkcija v parih točkah nekoliko odstopa (verjetno ker nisem upošteval nekaj točk), še boljše ujemanje pa se vidi če ga narisemo v logaritemski skali.