

Računalniška orodja v fiziki - Linearna regresija

Primož Cigler (28090039)

16. april 2010

Povzetek

Linearna zveza $y = kx$ je najpreprostejša zveza med dvema količinama. Pri merjenju je vedno dobrodošlo, da takšno linearno odvisnost dveh količin potrdimo z večjim številom meritev. Zaradi napak pri merjenju nikoli vse meritve ne bodo ležale na isti premici. Če iz meritev določamo najboljši koeficient naklona, je to najboljšo narediti tako, da je vsota kvadratov razdalj točk od premica najmanjša (princip najmanjših kvadratov).

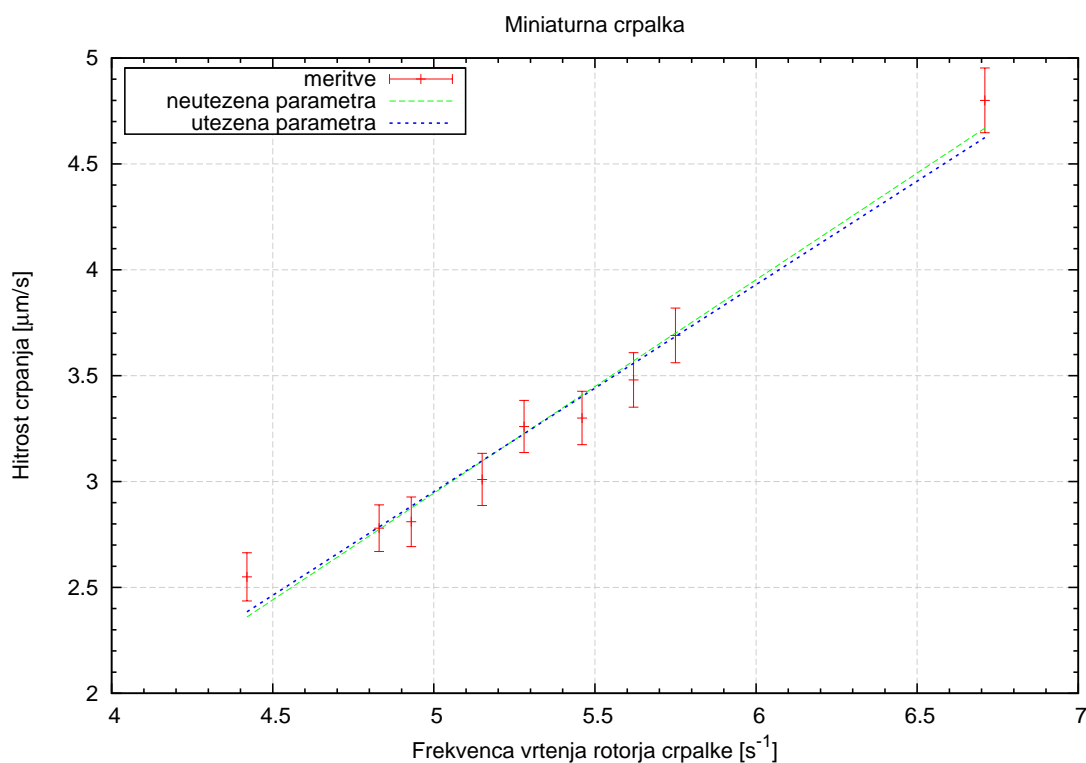
1 Linearna regresija in napake v programu Gnuplot

Program Gnuplot za linearno regresijo uporablja Marquardt-Levenbergov algoritem. Najprej definiramo modelsko funkcijo, ki opisuje nek pojav (v našem primeru linearna funkcija, lahko pa so tudi druge). Definiramo jo kot odvisnost ene spremenljivke od druge, z neznanimi parametri (recimo a in b pri linearni funkciji $f(x) = ax + b$). Nato z ukazom “fit” požene algoritem, ki poskuša po metodi najmanjših kvadratov ugotoviti najboljše vrednosti za neznane parametre.

Kočne parametre nam program poda z izračunanimi asimptotičnimi napakami, poda pa nam tudi χ^2 , ki je poimenovan WSSR (Weighted Sum of Squared Residuals). Poleg tega nam izračuna še nekaj drugih statističnih vrednosti, recimo RMS ostankov, varianco ali sigmo in korelacijsko matriko parametrov. Če programu ne podamo napak, nam vse meritve upošteva kot enako natančne.

2 Miniaturna črpalka

Pri prvi nalogi je bilo potrebno iz podatkov iz datoteke “HitrostTokaOdFrekvence.txt” določiti parametra najboljše premice. Ker so v datoteki v tretjem stolpcu podane še napake meritev, lahko posamezne točke tudi otežimo (pondiramo). S programom Gnuplot sem “pofital” linearno funkcijo $f(x) = ax + b$ preko parametrov a in b .



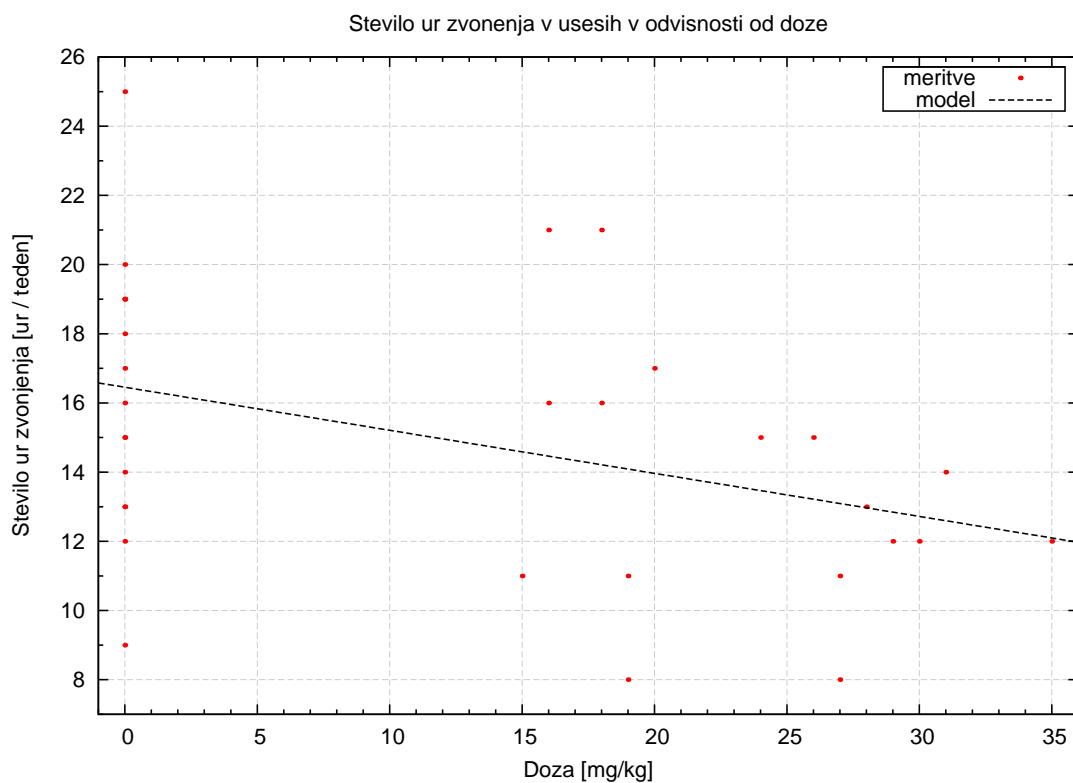
Graf 1: Linearna regresija s podatki za miniaturno črpalko

Naredil sem dve primerjavi linearne regresije. Zelena premica predstavlja “pofitana” parametra brez upoštevanih različnih napak pri posameznih meritvah, modra pa predstavlja utežena parametra z napakami, ki so podane v tretjem stolpcu.

χ^2 v prvem primeru je 0.08593, v drugem pa 5.43746.

3 Zvonenje v ušesih

Skozi oblak podatkov, ki so zbrani v datoteki "Tintin.dat" sem potegnil najboljšo premico.

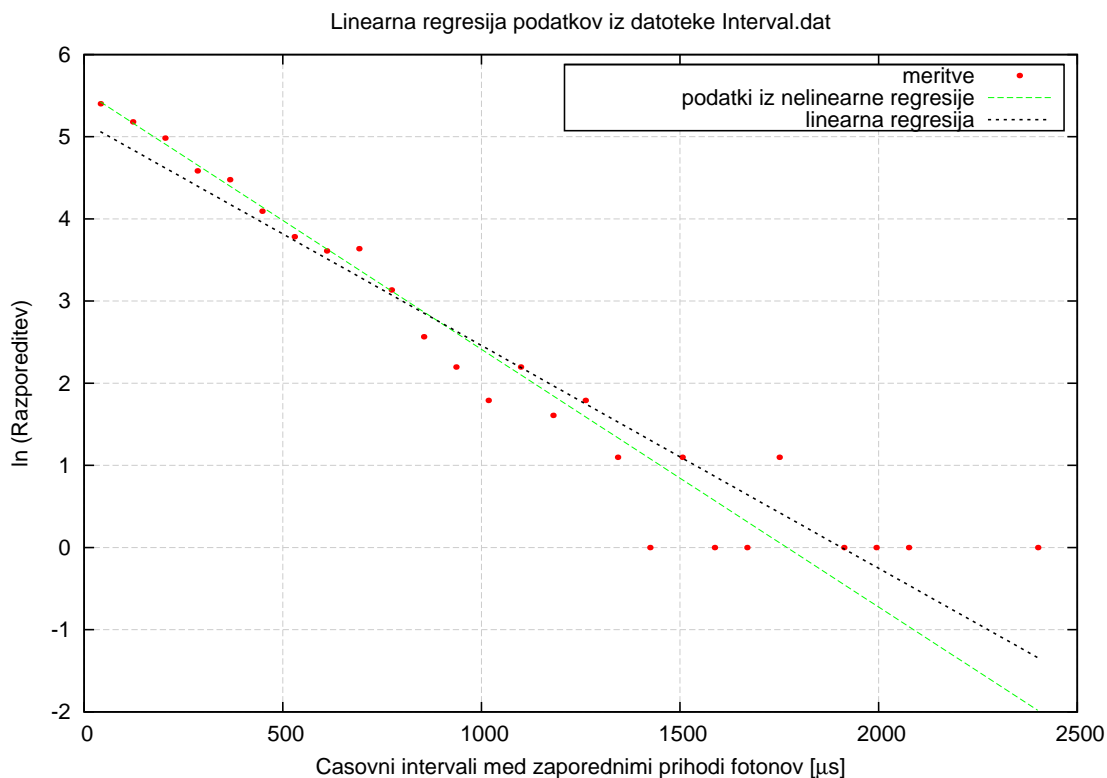


Graf 2: Linearna regresija podatkov za zdravilo tintinabulus

Vidimo, da ima premica negativen parametra a , vendar pada bolj počasi. Vrednost parametra a je približno -0.125 , χ^2 pa 419.575, ker so podatki zelo razpršeni.

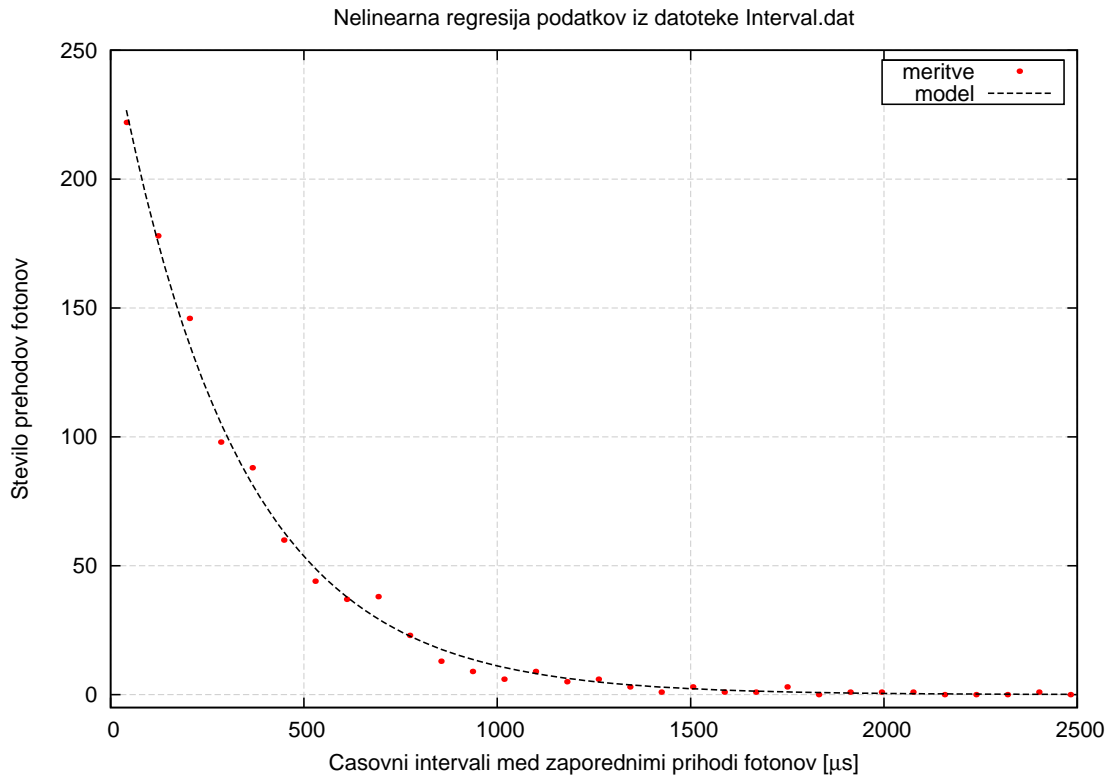
4 Interval.dat

Pri tretji nalogi je bilo potrebno najti najboljšo premico za predalčene vrednosti iz datoteke "Interval.dat". Ker pri nalogi 3.1 ni bilo potrebno predalčiti vrednosti, pač pa le izračunati σ in povprečje, sem sedaj podatke spustil skozi program, ki ga že imam napisanega v PHP-ju za predalčenje. Izbral sem 30 enako velikih predalčkov. Za določitev parametrov sem moral logaritmirati modelsko eksponentno funkcijo, da sem dobil linearno zvezo.



Graf 3: Linearna regresija logaritmiranih predalčenih podatkov iz datoteke "Interval.dat"

Pri tej nalogi sem spet naredil primerjavo, in sicer sem funkcijo še nelinearno pofital z vhodnimi podatki, kar predstavlja zelena premica, ki sem ji določena parametra iz nelinearne regresije ročno vnesel in narisal na isti graf kot za fit iz linearne regresije.



Graf 4: Nelinearna regresija predalčenih podatkov iz datoteke “Interval.dat”

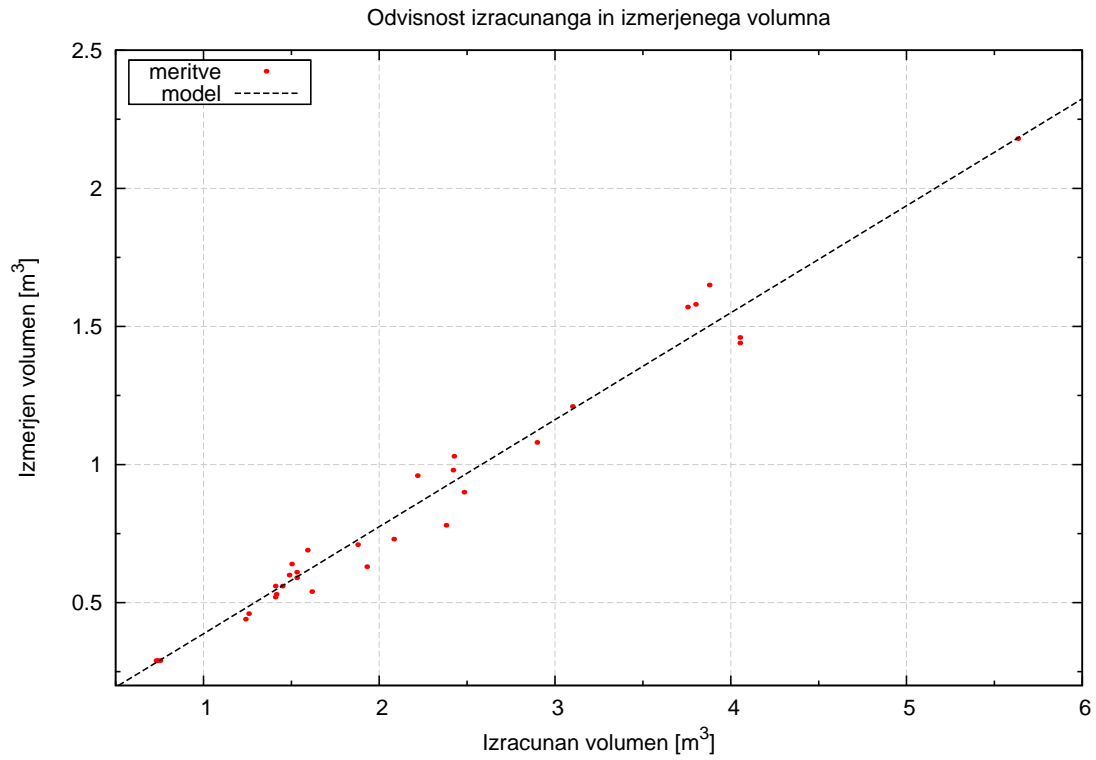
V teoriji bi se morali premici prekrivati, ne vem zakaj se v praksi ne.

Povprečje vseh meritev je 311, recipročna vrednost parametra λ pa je 369 za linearno regresijo in 319 za nelinearno regresijo. Red velikosti je vsekakor pravi, odstopanje pa se verjetno pojavi zaradi predalčenja.

5 Oblika dreves

V četrti nalogi je bilo potrebno potrebno iz podatkov v datoteki “drevesa.txt” določiti koeficient stožčastosti oz. oblike dreves. Iz meritev dolžine in premera pri tleh je možno izračunati kakšen volumen bi imeli valjasti hlodi, ker pa se proti vrhu ožajo, je bilo potrebno določiti ta koeficient. Vemo, da je prostornina stožca $\frac{1}{3}$ prostornine valja. Torej pričakujemo koeficient med $\frac{1}{3}$ in 1.

Koeficient k sem ponovno določil z Gnuplotom, in sicer sem na eno os nanese izračunane vrednosti, na drugo pa izmerjene in pofital linearno funkcijo, ki gre čez ta graf. Naklonski koeficient te premice je istočasno tudi k , ki ga iščemo.



Graf 5: Določanje koeficienta za obliko dreves med izmerjenimi in izračunanimi vrednostmi.

Za koeficient k sem dobil vrednost približno 0.39, kar pomeni, da so drevesa precej stožčaste oblike.