

Pragersko, 4. 5. 2009

Računalniška orodja v fiziki

7. naloga

Linearna regresija

Mitja Predikaka

1. naloga

Navodilo: Za meritve[1] v datoteki "HitrostTokaOdFrekvence.txt" (naloga 6.1) določi parametra najboljše premice. Ker so podane napake hitrosti, lahko določiš tudi χ^2 .

Podatke, podane v datoteki »HitrostTokaOdFrekvence.txt« sem obdelal v programu Excel na sledeč način:

	1	x	y	xy	x2	y2
	76,9467528	340,1046	196,2142	867,2669	1503,263	500,3463
	82,6446281	399,1736	229,7521	1109,702	1928,008	638,7107
	73,0513551	360,1432	205,2743	1012,002	1775,506	576,8208
	66,098222	340,4058	198,9556	1024,622	1753,09	598,8565
	66,098222	348,9986	215,4802	1137,735	1842,713	702,4655
	62,9881582	343,9153	207,8609	1134,921	1877,778	685,941
	60,0925425	337,7201	209,122	1175,266	1897,987	727,7447
	60,0925425	345,5321	221,7415	1275,014	1986,81	818,2261
	42,7186125	286,6419	205,0493	1375,881	1923,367	984,2368
sum	590,731036	3102,635	1889,45	10112,41	16488,52	6233,348
k:	0,97809318					
n:	-1,9386423					
hi2:	5,43745758					

V posameznih stolpcih so izračuni narejeni po sledeči formuli: $z = z[i]/\text{sqr}(\varepsilon[i])$; pri čemer je $z[i]$ po vrsti 1, $x[i]$, $y[i]$, $x[i]*x[i]$ in $x[i]*y[i]$.

V vrstici sum so vsote posameznih stolpcev.

Podatke za k, n in χ^2 (hi2) sem dobil po sledečih formulah:

- $k = (\text{sum}1 * \text{sum}xy - \text{sum}x * \text{sum}y) / (\text{sum}1 * \text{sum}x2 - \text{sum}x * \text{sum}x)$;
- $n = (\text{sum}x2 * \text{sum}y - \text{sum}x * \text{sum}xy) / (\text{sum}1 * \text{sum}x2 - \text{sum}x * \text{sum}x)$;
- $hi2 = \text{sum}y2 + k * k * \text{sum}x2 - 2 * k * \text{sum}xy - 2 * n * \text{sum}y + 2 * k * n * \text{sum}x + n * n * \text{sum}i$;

Rezultati izračunani po formulah z upoštevanimi napakami so torej:

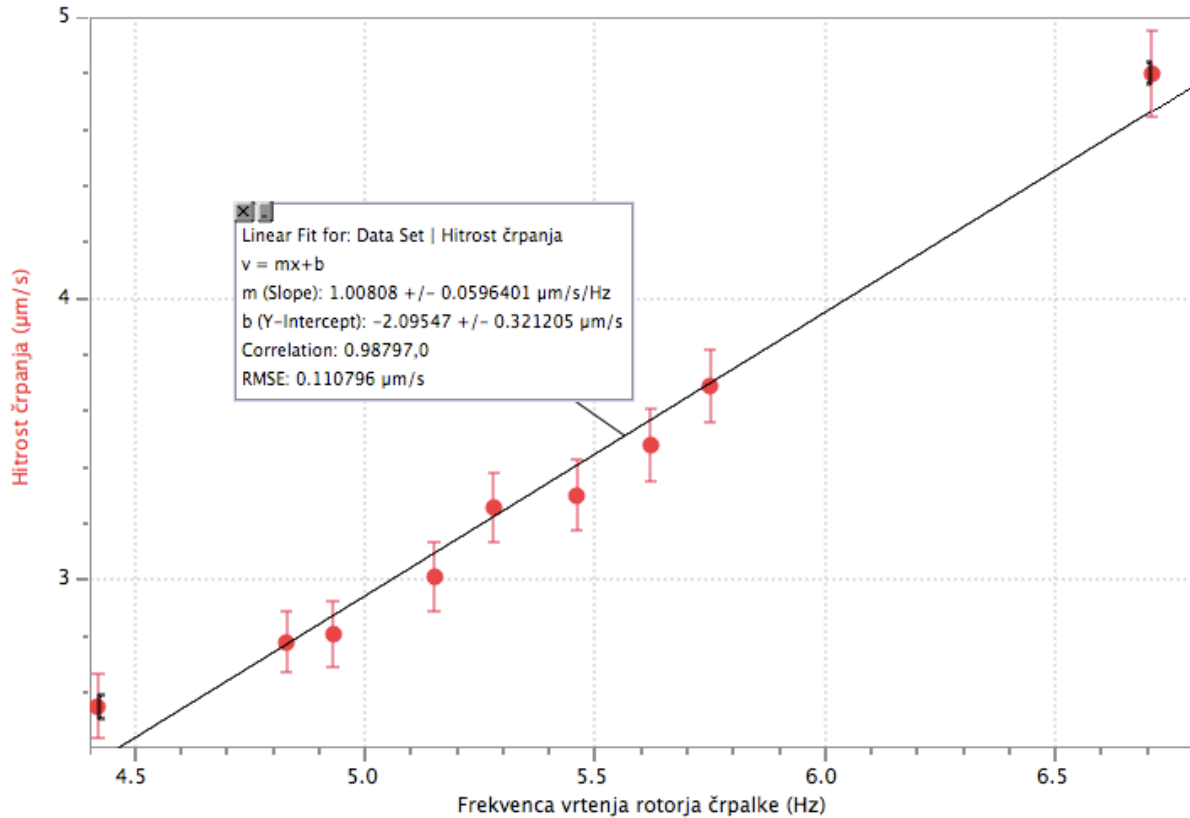
- $k = 0,97809 \mu\text{m}$
- $n = -1,93864 \mu\text{m/s}$
- $\chi^2 = 5,43745 (\mu\text{m/s})^2$
- $m \pm v2m = 9 \pm 4,243$

Vrednost χ^2 je tudi znotraj intervala $m \pm v2m$, kar pomeni da bi premica z zgornjimi podatki dobro opisala funkcijsko odvisnost izmerjenih točk.

Naloga se lahko lotimo tudi preko korelacijskega koeficienta po formuli $k = R\sigma_y / \sigma_x$ in upoštevamo da gre sledeča premica skozi točko $(x_{\text{pov}}, y_{\text{pov}})$. Vendar pri tem zanemarimo napake, ki so podane za vsako izmerjeno vrednost hitrosti črpanja.

Za potrditev zgornjega odstavka sem to metodo preizkusil tudi v programu Logger Pro.

Graf 1: Hitrost črpanja v odvisnosti od frekvence vrtenja rotorja črpalke



Rdeče točke na grafu predstavljajo posamezno meritev. Rdeča okolica točke predstavlja napako. Črna črta predstavlja linearni fit glede na meritve, brez upoštevanja napak.

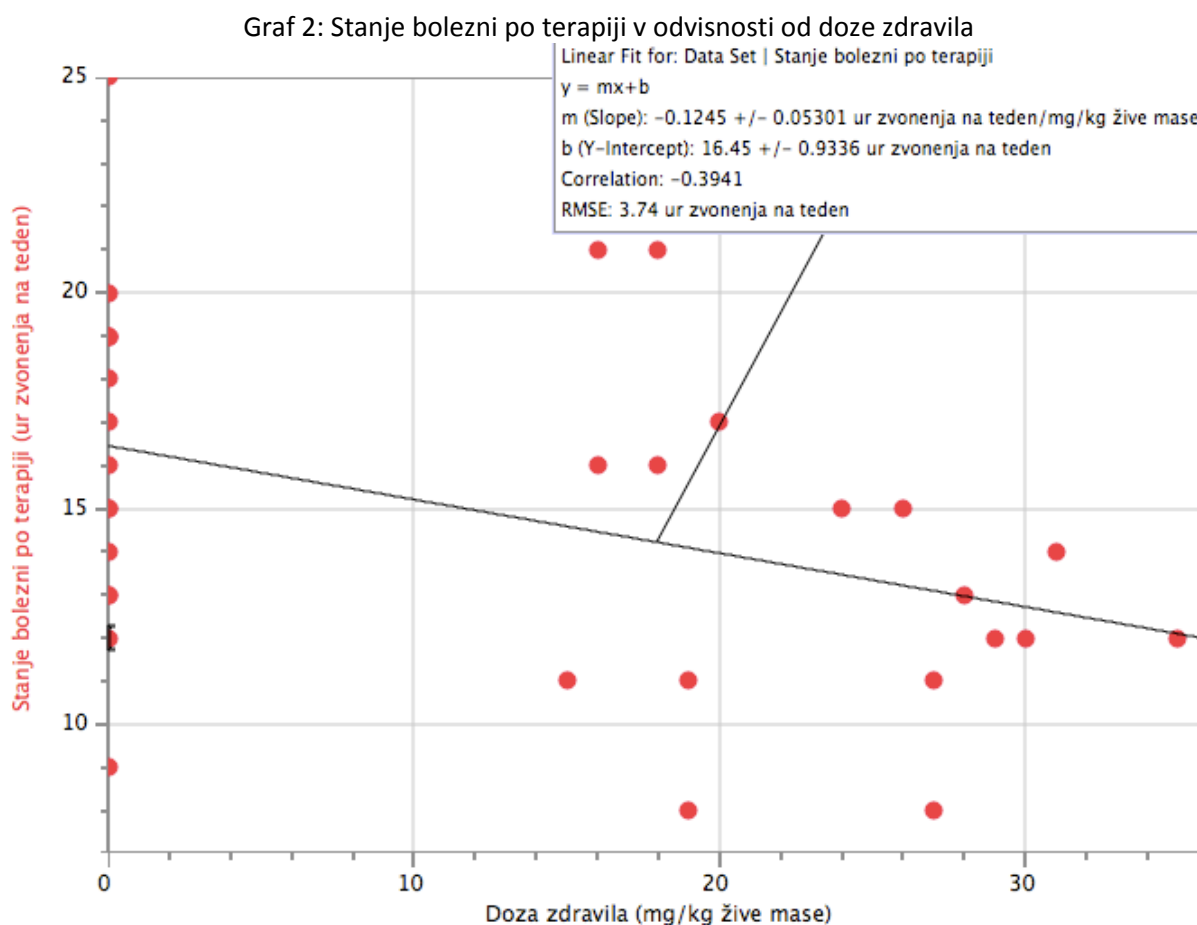
Kljub temu da sem mu da sem med podatke nanesel tudi napako posamezne meritve hitrosti črpanja, jih program ni upošteval pri izračunu smernega koeficienta in začetne vrednosti.

2. naloga

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

Reševanja te naloge sem se lotil preko korelacijskega koeficienta, saj v podatkih nimamo podanih vrednosti napak posameznih meritev, tako da metoda $add(sumz, z[i])$; kjer je $z[i]$ po vrsti $1, x[i], y[i], x[i]*x[i]$ in $x[i]*y[i]$ vrne enake rezultate kot pri prvi nalogi navedena metoda z korelacijskim koeficientom.

Prav tako dobimo enake rezultate če nalogo rešimo z »Linear fit« funkcijo programa Logger pro, saj tudi ta izračuna korelacijski koeficient in nato naklonski koeficient in začetno vrednost funkcije, ter tudi preveri, da je RMSE (Root mean square error) minimalen.



Zapišemo:

- $k = -0,1245 \pm 0,0530$ (ur zvonjenja na teden)/(mg/kg žive mase)
- $n = 16,45 \pm 0,93$ ur zvonjenja na teden
- $\chi^2 = 419,57$ (ur zvonjenja na teden)²
- $m \pm \sqrt{2m} = 32 \pm 8$

Opazimo da zgornja $kx + n$ funkcija le slabo opiše zgornjo meritev.

Opomba: χ^2 se ne da izračunati preko korelacijskega koeficienta, zato sem uporabil enak postopek kot pri prvi nalogi.

3. naloga

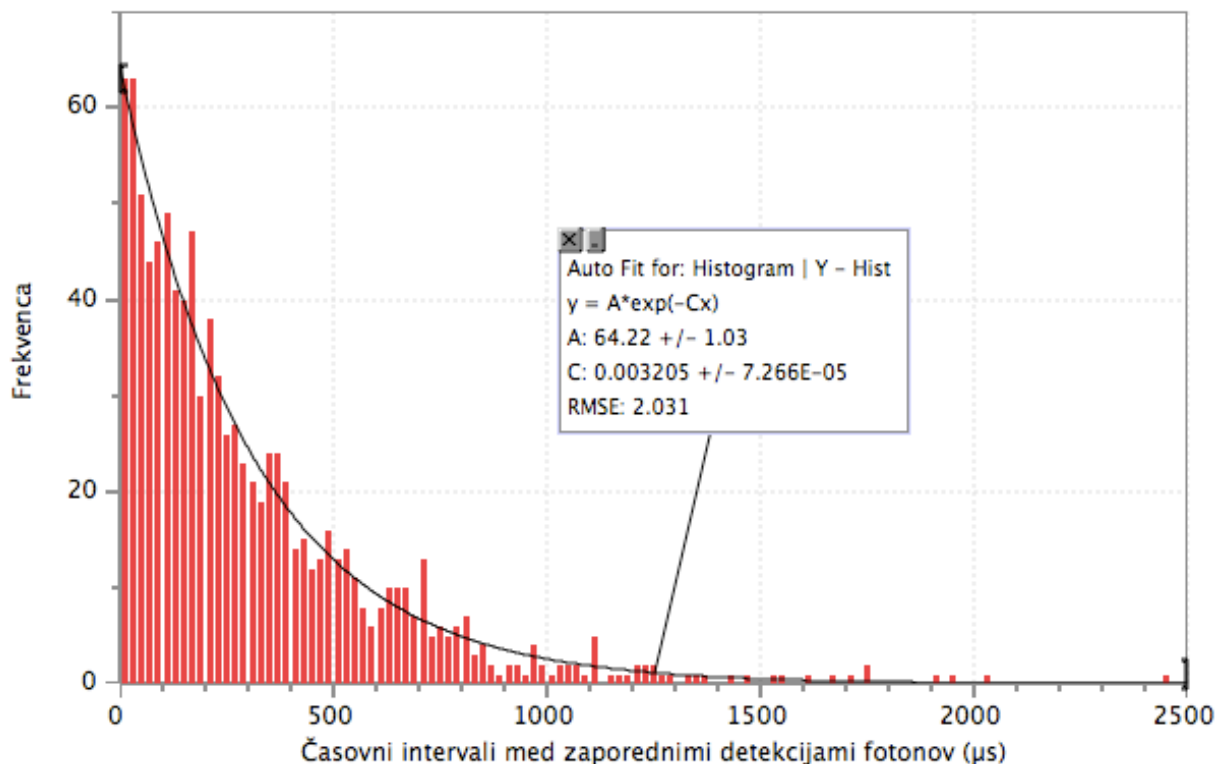
Navodilo: Skozi točke v histogramu podatkov "Interval.dat" poskusi potegniti najboljšo eksponentno funkcijo $w = Ae^{-\lambda x}$, ki jo moramo najprej predelati v linearno zvezo. Z logaritmiranjem dobimo $\ln(w) = \ln(A) - \lambda x$. V grafu $y = \ln(w)$ od x sta koeficienta premice $k = -\lambda$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

Datoteka "Interval.dat" podaja časovne intervale (v mikrosekundah) med zaporednimi prihodi posameznih fotonov v detektor.

Recipročna povprečna vrednost histograma je 0,003215

Odločil sem se za primerjavo dveh različnih načinov predalčenja. Prvi način bo predalčil fotone, katerih zamak med zaporednimi prihodi je manjši od 20 μs , drugi pa fotone z zamakom 40 μs . Pri prvem predalčenju bo tako 123 predalčkov, pri drugem pa 62.

Histogram 1: Histogram z predalčenjem po 20 μs in najboljšim fitom $y = A \cdot \exp(-Cx)$

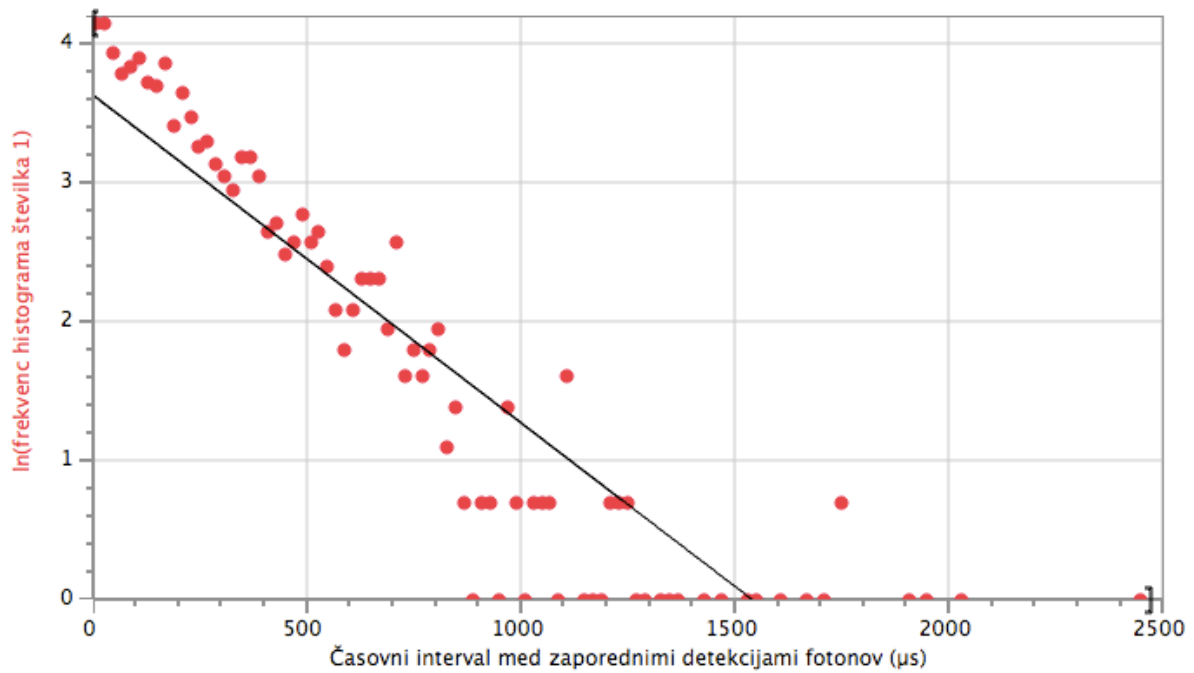


Rezlutati:

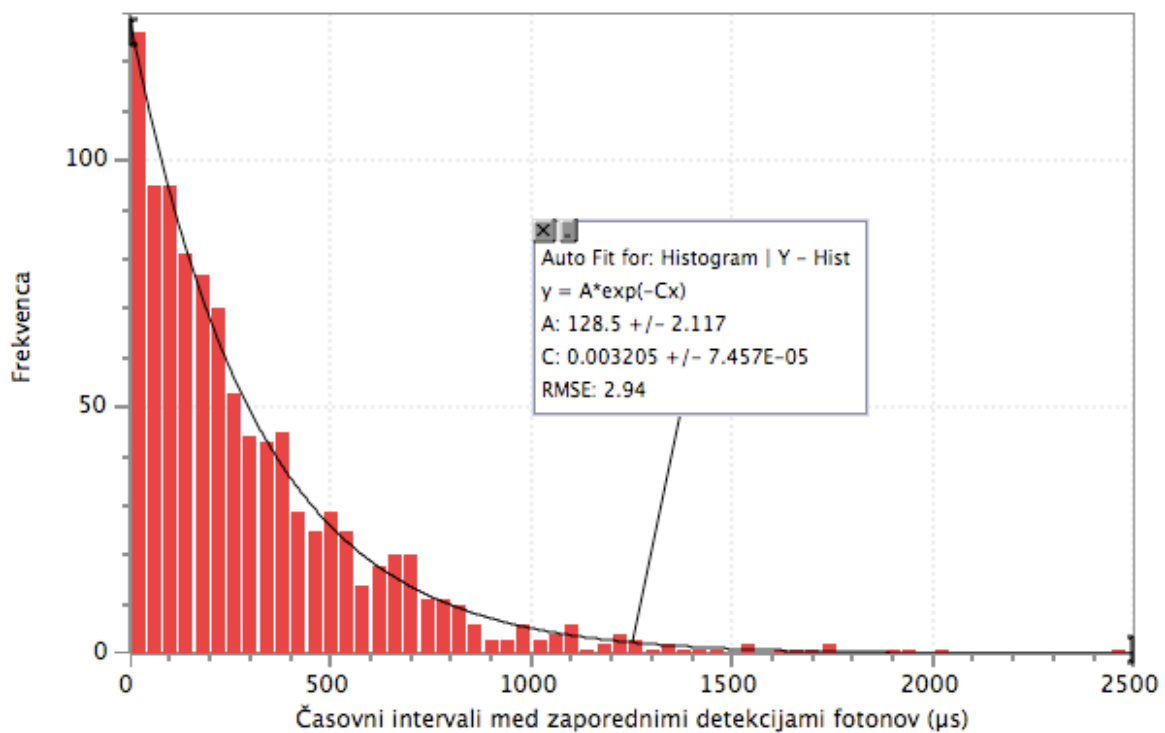
- $A = 64,22 \pm 1,03 \rightarrow n = \ln(A) = 4,162 \pm 0,029$
- $C = 0.003205 \pm 7E-5 \rightarrow k = -C = -0.003205 \pm 7E-5$

Opazimo da koeficient C zares ustreza recipročni vrednosti histograma.

Graf 3: Logaritmiran histogram 1 z dodanim fitom iz podatkov dobljenih iz enačbe $y = A \cdot \exp(-Cx)$



Histogram 2: Histogram z predalčenjem po 40 μs in najboljšim fitom $y = A \cdot \exp(-Cx)$



Rezultati:

- $A = 128,5 \pm 2,1 \rightarrow n = \ln(A) = 4,855 \pm 0,746$
- $C = 0.003205 \pm 7E-5 \rightarrow k = -C = -0.003205 \pm 7E-5$

Opazimo da tudi pri tem predalčenju koeficient C ustreza recipročni vrednosti histograma.

Kljub temu da se smerna koeficienta k razlikujeta le pri višjih decimalkah ugotovimo iz zapisa RMSE, da je prvi histogram primernejši, saj to pomeni manjšo napako.

4. naloga

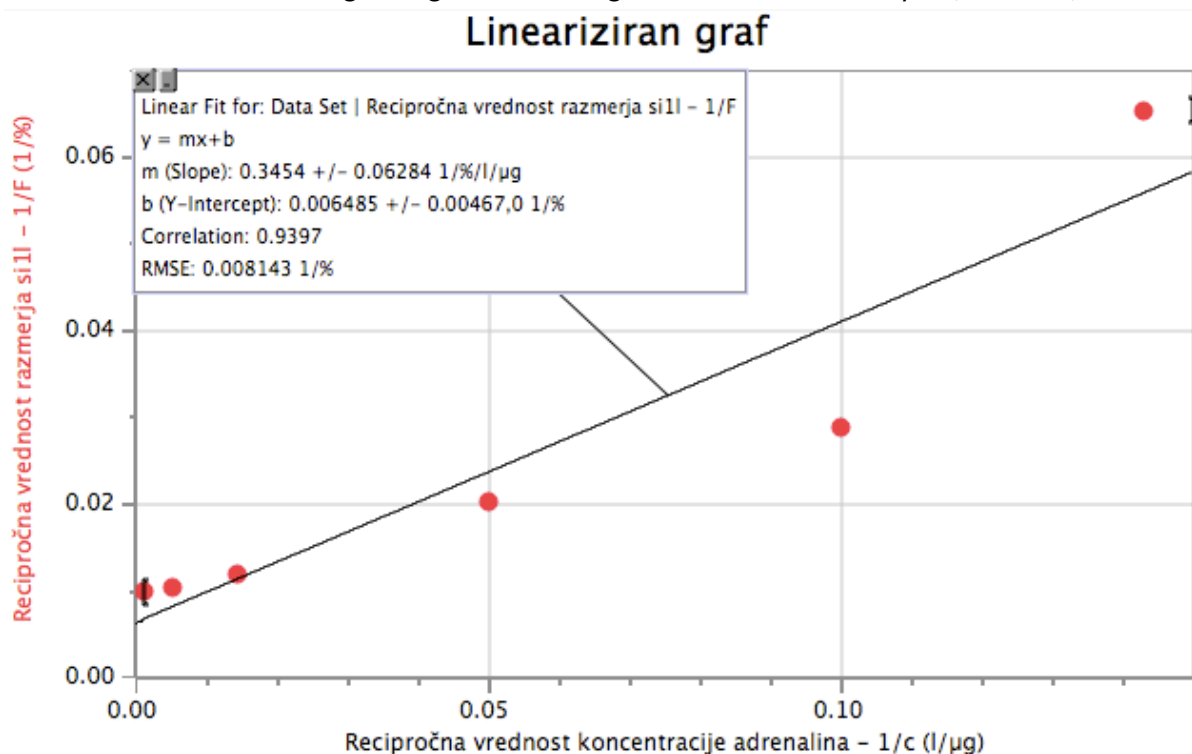
Navodilo: Teorija kemijske kinetike napove za sigmoidno krivuljo iz podatkov "Adrenalin.dat" (naloga 1.1) naslednjo odvisnost $F / F_{\max} = c / (a + c)$, kjer pomeni a koncentracijo s polovičnim maksimalnim učinkom. Določi koeficienta F_{\max} in a . Pretvori v linearno zvezo – ena pot je uvedba recipročnih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

Enačbo $F / F_{\max} = c / (a + c)$ lahko preoblikujemo na sledeča dva načina:

- (1) $1/F = (1/c) \cdot (a/F_{\max}) + 1/F_{\max} \rightarrow k = a/F_{\max} \quad n = 1/F_{\max}$
- (2) $c/F = c \cdot (1/F_{\max}) + a/F_{\max} \rightarrow k = 1/F_{\max} \quad n = a/F_{\max}$

Način številka ena (1):

Graf 4: lineariziran graf odgovora mišičnega vlakna na adrenalin s $y = 1/F$ in $x = 1/c$



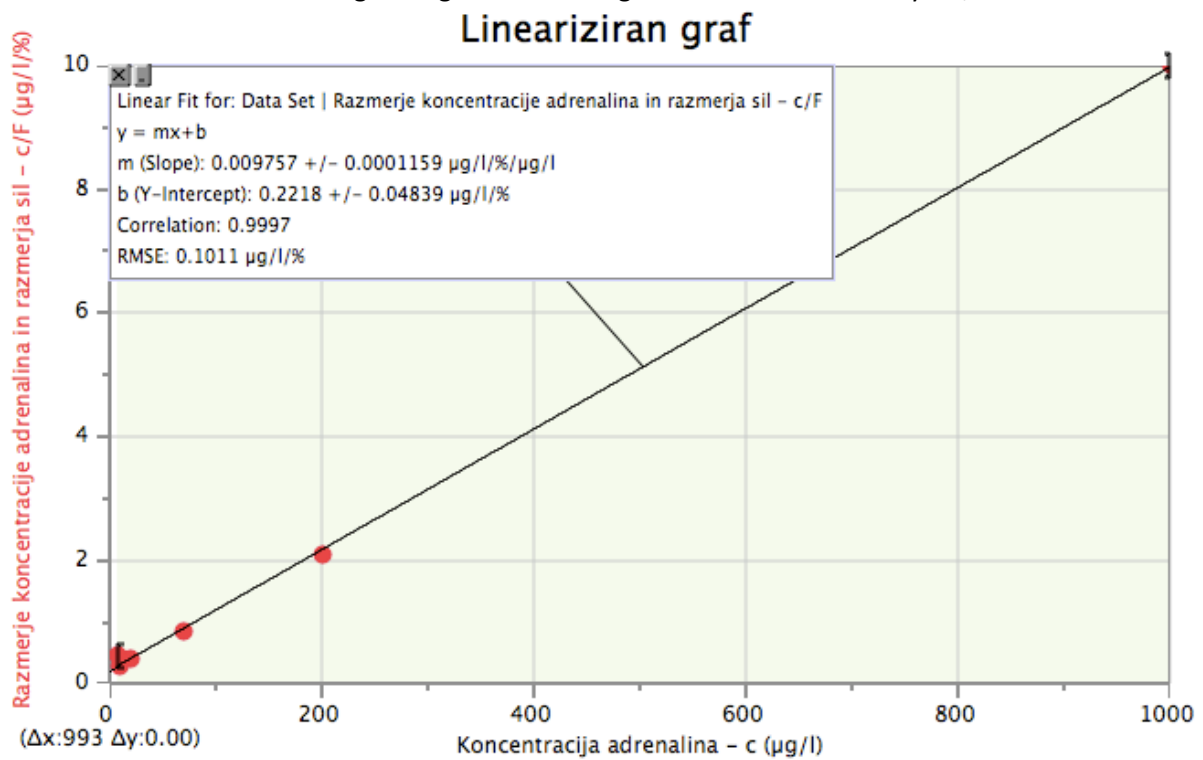
Rezultati

- $k = (0,3454 \pm 0,0628) \mu\text{g}/(\% \cdot \text{l}) \rightarrow a = (53.258 \pm 9.688) \mu\text{g}/\text{l}$
- $n = 0,00648 \%^{-1} \pm 0.00467 \%^{-1} \rightarrow F_{\max} = 154,320\% \pm 111,215\%$

Takoj opazimo ogromno napako meritev in posledično izračunanih vrednosti za a in F_{\max} . K tako veliki napaki prispevata prvi dve meritvi, katerih recipročna vrednost sile ne obstaja, tako da pri tako majhnem številu meritev, ki niso lepo linearne, ne moremo upati, da je to rezultat, katerega smo iskali

Način številka dva (2):

Graf 5: lineariziran graf odgovora mišičnega vlakna na adrenalin s $y = c/F$ in $x = c$

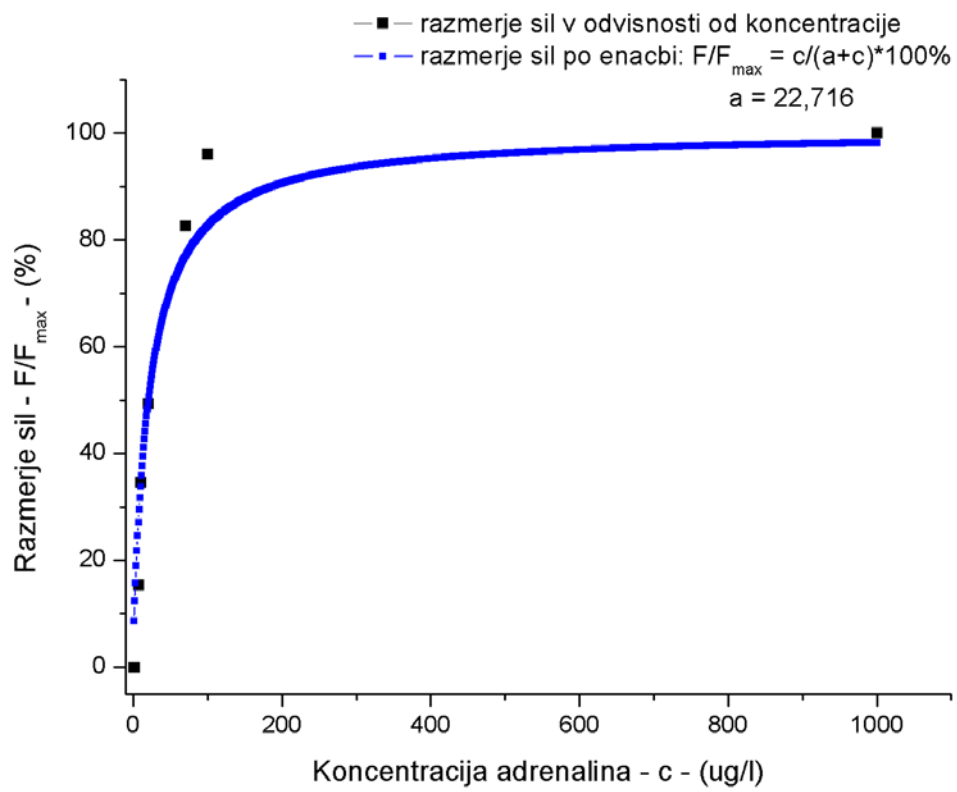


Rezultati

- $k = (0,009757 \pm 0,000116) \%^{-1} \rightarrow F_{\max} = 102,491\% \pm 1,217\%$
- $n = (0,2218 \pm 0.04839) \mu\text{g/l}/\% \rightarrow a = (22,716 \pm 4.963) \mu\text{g/l}$

Po primerjavi obeh načinov, bi izbral način številka dva, saj vrne lepše rešitve z veliko manjšimi napakami. Prav tako tudi ne bi bilo logično da bi z meritvijo konačili tako hitro, če bi se mišica lahko raztegnila še za 64%, kot nakazuje prvi način.

Graf 6: graf odgovora mišičnega vlakna na adrenalin



OPOMBA: Za modro krivuljo sem sem izračunal 1000 točk po formuli $F/F_{\max} = c/(a+c) \cdot 100\%$, pri čemer sem za a uporabil 22,716, kot nam vrne način številka dva (2).