

7. Tema pri predmetu Računalniška orodja v fiziki

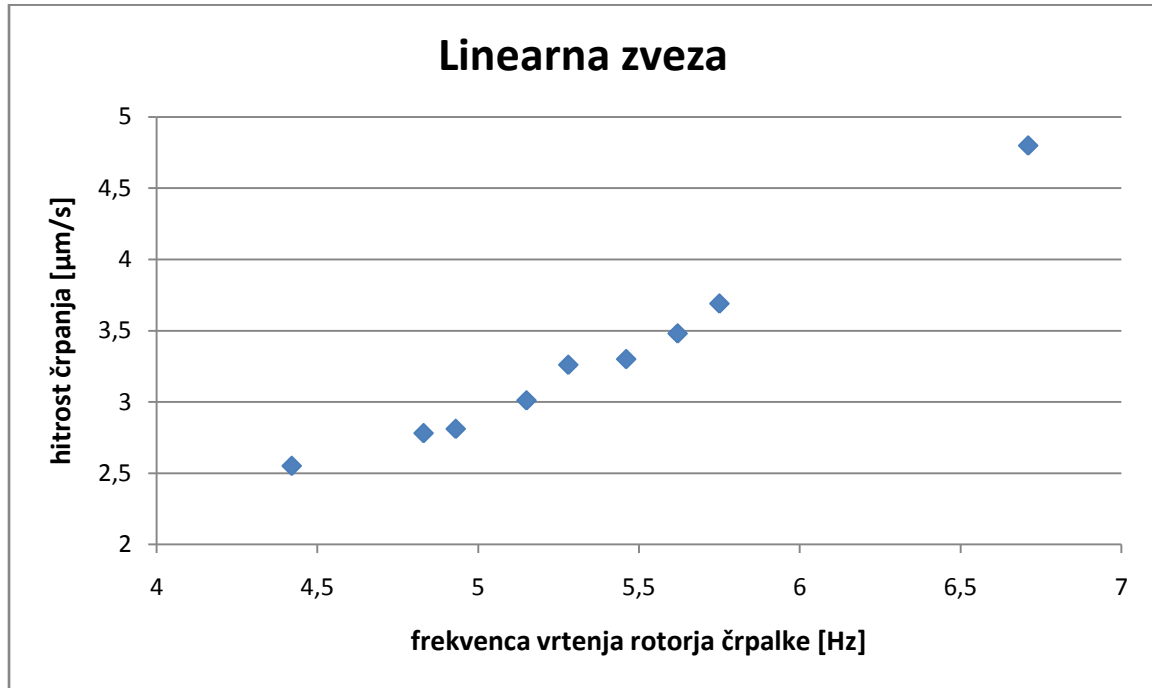
Linearna regresija

Križevci pri Ljutomeru, 4.5.2009

Jure Senčar

1. Naloga

Za meritve v datoteki »HitrostTokaOdFrekvence.txt« moramo poiskati premico ki se najbolj ujema s točkami meritev. Izračune sem opravil v Excelu, »linear fit: $mx+b$ « funkcijo pa sem uporabil pri tudi pri programu Loggerpro.



V Excelu sem sledil navodilom:

SUM(x)	48.15	SUM(C6:C14)
SUM(y)	29.68	SUM(D6:D14)
SUMPRODUCT(x^2)	261.0537	SUMPRODUCT(C6:C14;C6:C14)
SUMPRODUCT(xy)	162.2671	SUMPRODUCT(C6:C14;D6:D14)
št	9	F14
SUMPRODUCT(y^2)	101.4712	SUMPRODUCT(D6:D14;D6:D14)
Koeficient	1.008084	x $(H10*H9-H7*H6)/(H10*H8-H6*H6)$
Začetna vrednost	-2.09547	0 $(H8*H7-H9*H6)/(H10*H8-H6*H6)$
	4.904528	7 $H13+7*H*H12$

S korelacijo:

Povx	5.35	AVERAGE(C6:C14)
Povy	3.297778	AVERAGE(D6:D14)
Sigmax	0.619247	STDEVP(C6:C14)
Sigmay	0.631854	STDEVP(D6:D14)
Korelacija	0.98797	CORREL(C6:C14;D6:D14)
Naklon	1.008084	x $H20*H19/H18$
Začetna vrednost	-2.09547	0 $H17-H16*H21$
	4.961117	7 $H22+7*H21$

Ter še z napakami:

SUM(x*ro^-2)	3102.635	SUMPRODUCT(L20:L28;N20:N28;N20:N28)
SUM(y*ro^-2)	1889.45	SUMPRODUCT(M20:M28;N20:N28;N20:N28)
SUMPRODUCT(x^2*ro^-2)	16488.52	SUMPRODUCT(L20:L28;L20:L28;N20:N28;N20:N28)
SUMPRODUCT(xy*ro^-2)	10112.41	SUMPRODUCT(L20:L28;M20:M28;N20:N28;N20:N28)
SUMPRODUCT(ro^-2)	590.731	SUMPRODUCT(N20:N28;N20:N28)
SUMPRODUCT(y^2*ro^-2)	6233.348	SUMPRODUCT(M20:M28;M20:M28;N20:N28;N20:N28)
Koeficient3	0.978093	x (Q24*Q23-Q21*Q20)/(Q24*Q22-Q20*Q20)
Začetna vrednost3	-1.93864	0 (Q22*Q21-Q23*Q20)/(Q24*Q22-Q20*Q20)
	5.061358	7 Q27+7
χ²	5.437458	Q25+Q26*Q26*Q22-2*Q26*Q23- 2*Q27*Q21+2*Q26*Q27*Q20+Q27*Q27*Q24

Excel ima tudi vgrajeno funkcijo LINEST, ki nam vrne premico, določeno po metodi najmanjših kvadratov:

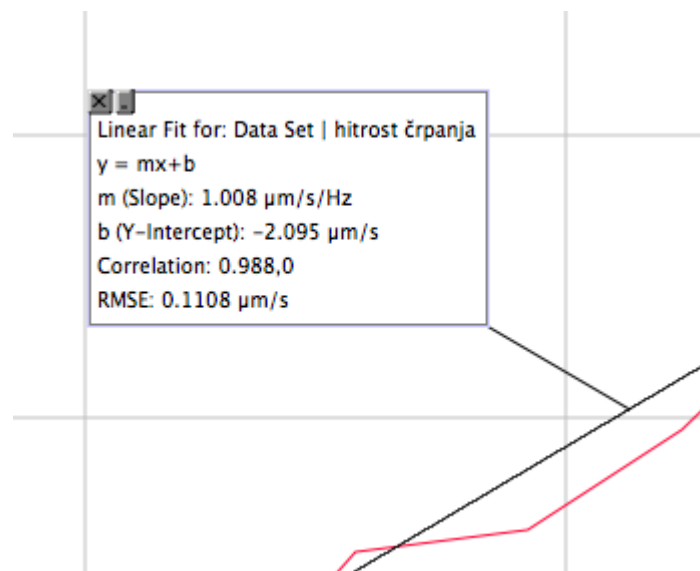
Naklon = INDEX(LINEST(B6:B14;A6:A14); 1) = 1.008084

Začetna vrednost = INDEX(LINEST(B6:B14;A6:A14); 2) = -2.09547

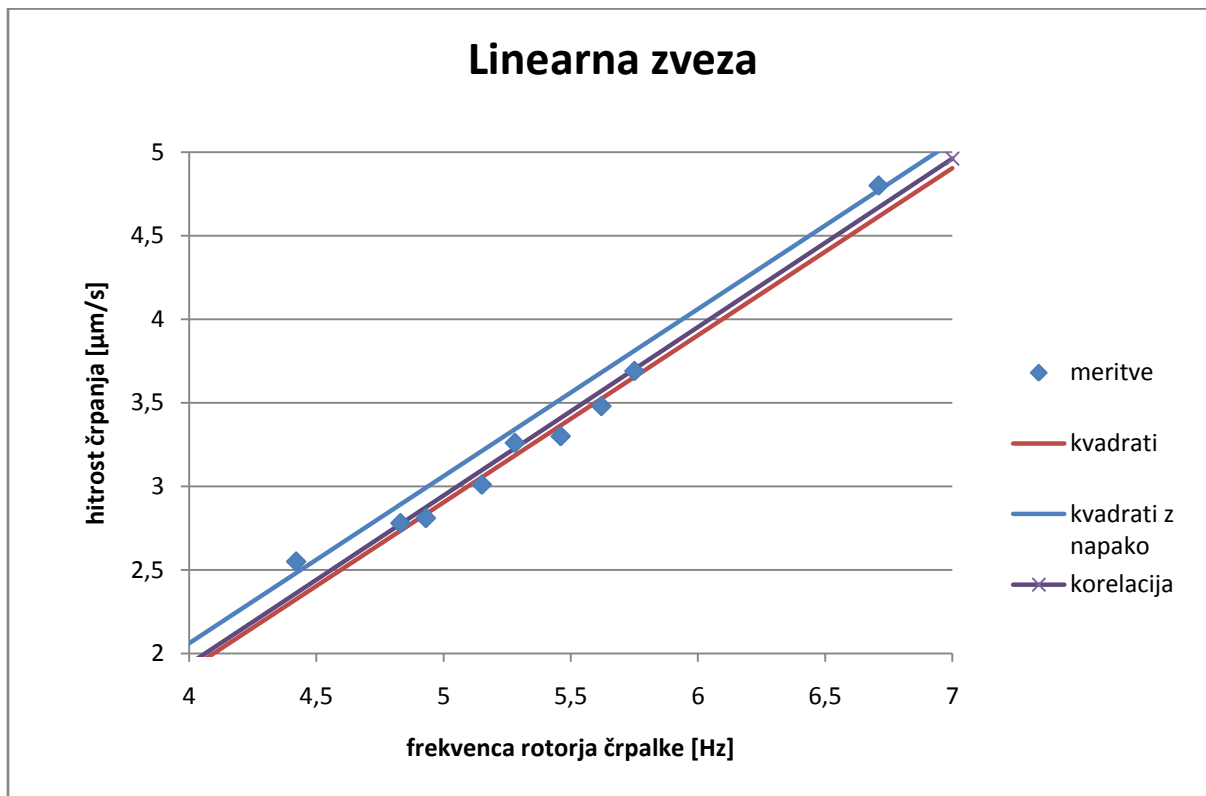
Žal pa ta funkcija ne zajema napak (ne moremo jih vnesti v formulo).

Z dodatkom »Data Analysis«, ki smo ga že potrebovali pri upodabljanju histogramov, pa lahko uporabimo »correlation« oddelek, ki nam vrne izračune na oba načina. Žal pa tudi ta ne ponuja možnosti vnosa napak, kot so podane v tej nalogi.

Eden izmed programov, ki določajo »best fit«-e s pomočjo korelacije je Loggerpro.



Vse premice ter meritve sem vnesel v graf (poševno napisane vrednosti v Excelu ter meritve):



Po mojem mnenju je vizualno »ta prava« premica povprečje vseh treh :)

Kot rezultat pa bi podal meritev s pomočjo kvadratov ter napak:

Naklon: $0.9781 \frac{(\mu\text{m} / \text{s})}{(\text{Hz})}$

Začetna vrednost: $-1.9386 \frac{\mu\text{m}}{\text{s}}$

χ^2 : $5.43746 \left(\frac{\mu\text{m}}{\text{s}}\right)^2$

2. Naloga

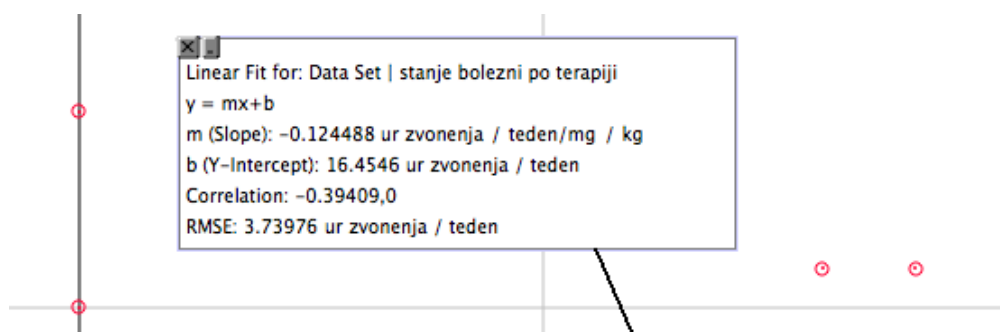
Pri tej nalogi znova preučujemo učinke zdravila za zvonjenje v ušesih. Tokrat iščemo najprimernejšo premico za podane meritve z metodo korelacije. Identično, kot pri 1. nalogi:

$$\text{Naklon} = \text{CORREL}(B2:B33;C2:C33) * \text{STDEVP}(B2:B33) / \text{STDEVP}(C2:C33) = -1.244884$$

$$\text{Točka: } x1 = \text{AVERAGE}(B2:B33) \quad y1 = \text{AVERAGE}(C2:C33)$$

$$\text{Začetna vrednost} = y1 - x1 * \text{Naklon} = 16.45457$$

Enake vrednosti nam določi tudi Loggerpro.

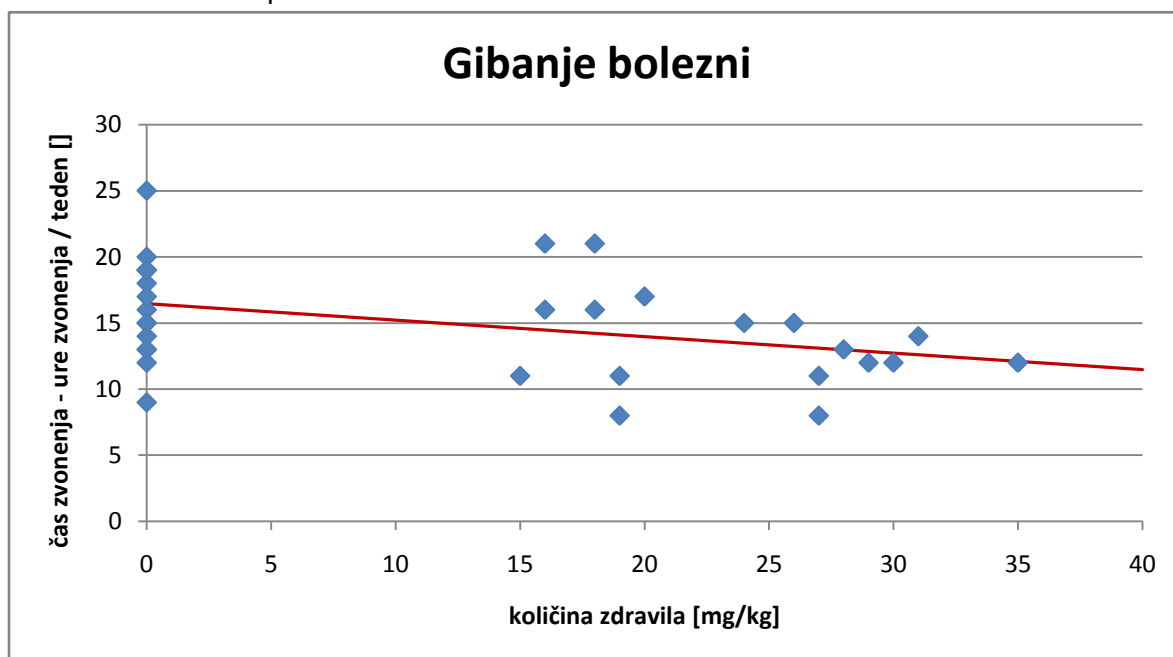


Določimo še premico z metodo najmanjših kvadratov:

$$\text{Naklon} = \text{INDEX}(\text{LINEST}(C2:C33;B2:B33);1) = -0.124488 \frac{(\text{ur zvonjenja / teden})}{(\text{mg / kg})}$$

$$\text{Začetna vrednost} = \text{INDEX}(\text{LINEST}(C2:C33;B2:B33);2) = 16.4546 \frac{\text{ur zvonjenja}}{\text{teden}}$$

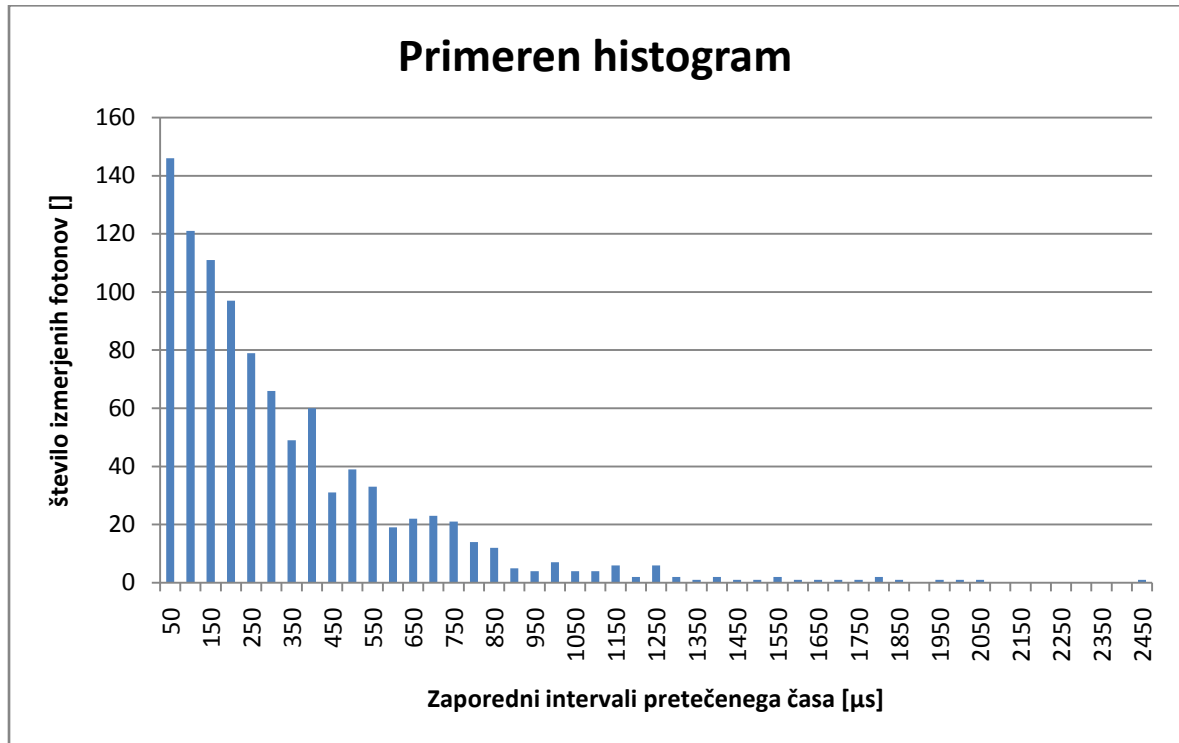
Rezultata sta tokrat pri obeh metodah enaka:



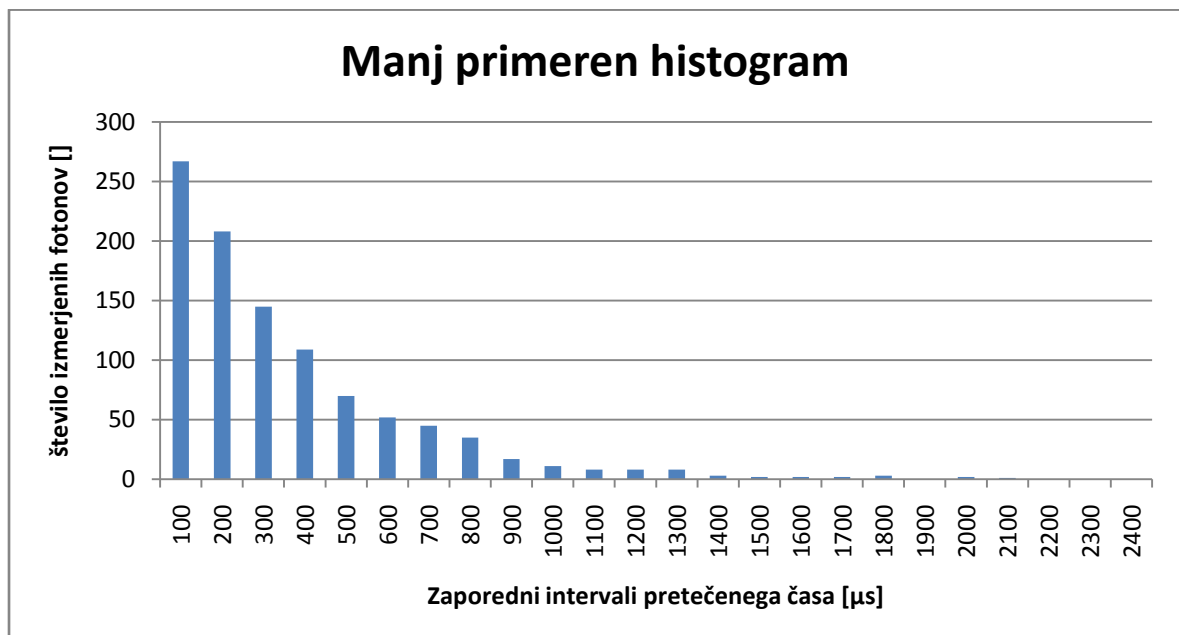
3. Naloga

Skozi histogram iz podatkov »Interval.dat« je potrebno potegniti najboljšo eksponentno funkcijo.

Odločil sem se za velikost predalčka 50 mikrosekund:



Pri tem izgubimo rep, a če izberemo večje predalčke, izgubimo precejšnje število uporabni končnik točk. Teh točk pa si želimo čim več. Poleg tega pa še opazimo, da tudi pri repu nismo dosegli zelenega učinka.



Odločil sem se, da histogram preuredim v verjetnostno gostoto ter ji določim najboljše eksponentno ujemanje.

Vrednosti y-stolpca v 1. histogramu sem delil z $50 \mu\text{s}$. Nato sem s pomočjo vgrajene funkcije v Excelu določil eksponentno ujemanje. Ta funkcija se imenuje LOGEST in deluje na enak način kot prej uporabljena funkcija LINEST, le da uporablja obliko $y = a \cdot b^x$. Nas pa zanima oblika $y = a \cdot e^{(-\lambda \cdot x)}$.

$$b = \text{INDEX}(\text{LOGEST}(D3:D33;C3:C33);1) = 0.996644$$

$$a = \text{INDEX}(\text{LOGEST}(D3:D33;C3:C33);2) = \mathbf{3.310346 \text{ (fotonov)}}$$

Nato sem še y vrednosti logaritmiral ter določil najboljšo premico:

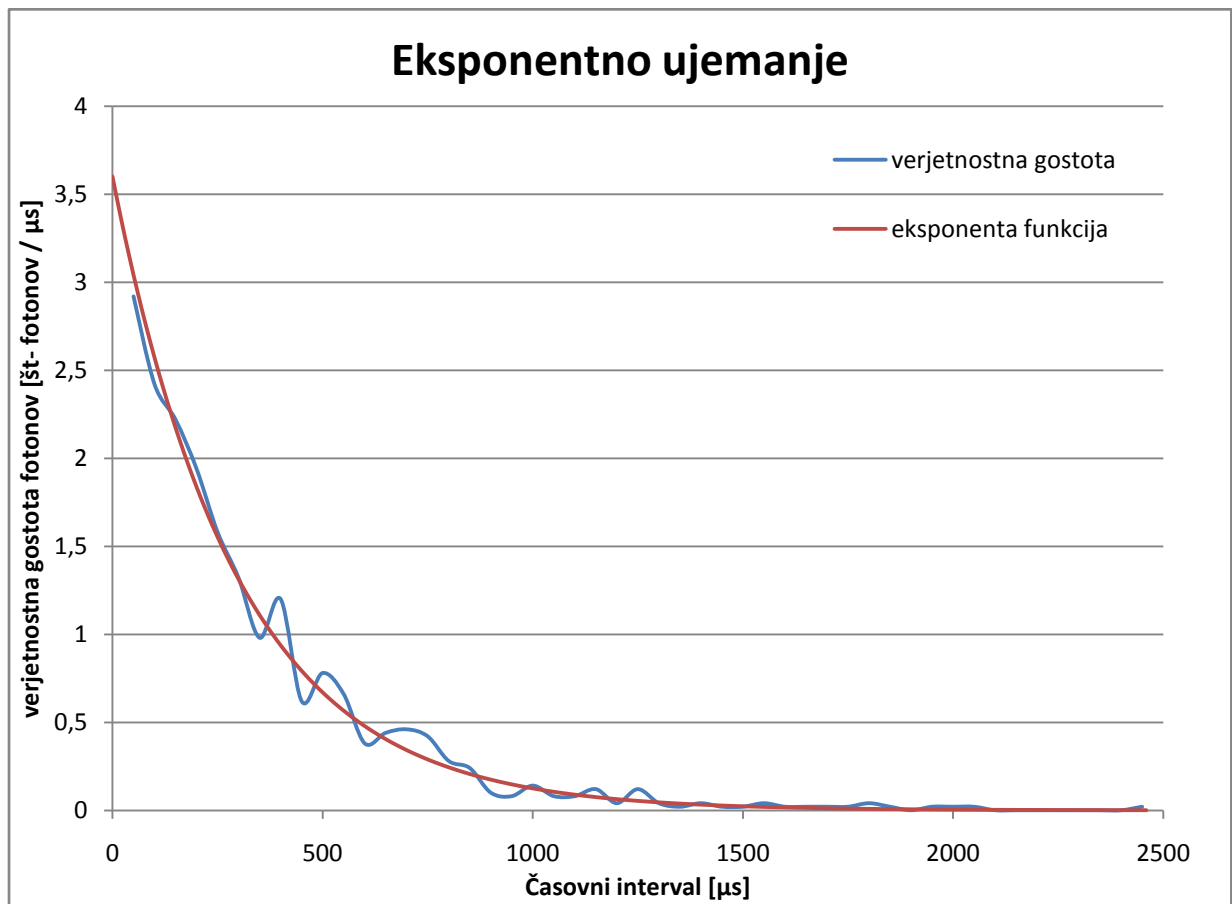
$$\lambda = -\ln(b) = -\text{INDEX}(\text{LINEST}(K3:K33;C3:C33);1) = \mathbf{0.0033619 \mu\text{s}^{-1}}$$

$$\ln(a) = \text{INDEX}(\text{LINEST}(K3:K33;C3:C33);2) = 1.19705$$

Opazimo, da se izračuni ujemajo, kar pomeni da LOGEST deluje na enakem principu kot navodila (pretvorba na linearno zvezo ter določitev s pomočjo najmanjših kvadratov).

λ se tudi še kar dobro ujema s teoretično vrednostjo histograma s najmanjšim predalčkom – to je kar povprečje vseh meritev: $1/\text{AVERAGE}(A1:A999) = 0.321456$.

Končna slika:



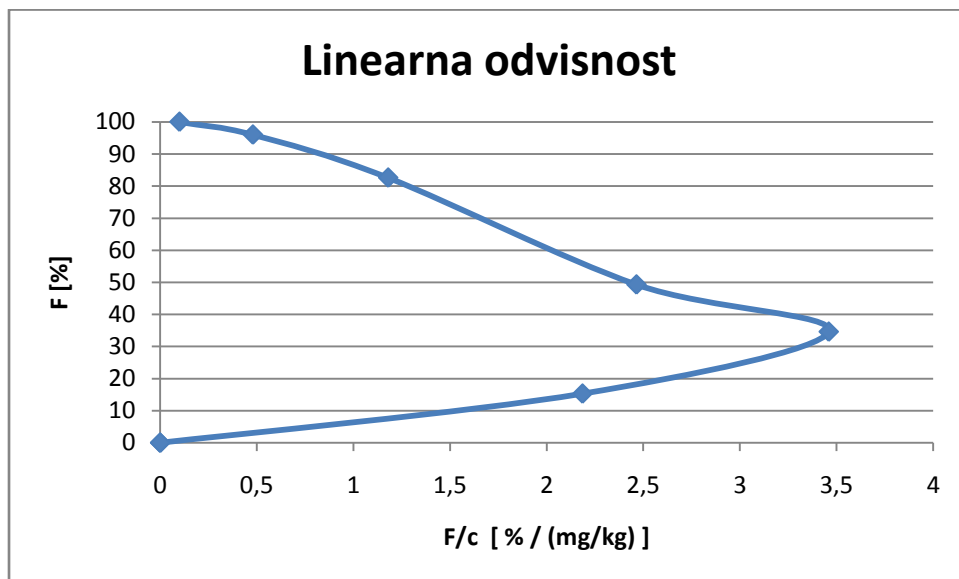
4. Naloga

Na točke meritev iz datoteke »interval.dat« je potrebno prilepiti funkcijo oblike

$$f(x) = b \cdot x / (a + x) \quad \text{v našem primeru: } F = F(c) = F_{\max} \cdot c / (a + c)$$

Odločil sem se, da bom uporabil kvocient F/c , saj mi ostale možnosti poženejo prvi dve točki v neskončnost, kar pa ni ugodno za linearno določanje. $F = -a(F/c) + F_{\max}$

Žal pa je že iz grafa $F(F/c)$ razvidno, da vse ne bo šlo gladko.

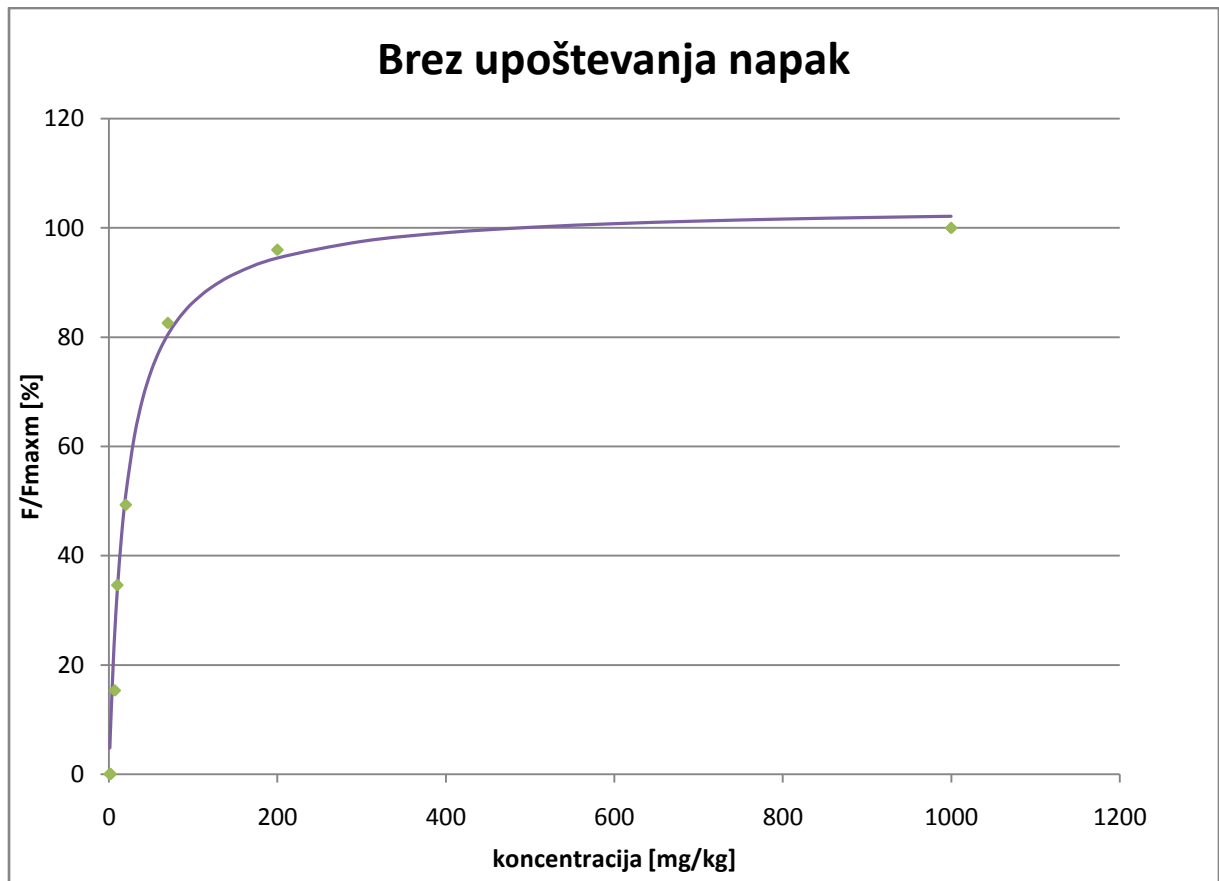


Odločil sem se, da točk z ničlami ne bom upošteval, saj jih je tudi pri drugih metodah skoraj nemogoče, poleg tega pa so gotovo obremenjene z zelo veliko relativno napako. Znova sem uporabil LINEST funkcijo ter določil parametra a ter F_{\max} .

$$a = \text{Naklon} = \text{INDEX}(\text{LINEST}(F5:F9;G5:G9);1) = 20.6352 \frac{\text{mg}}{\text{kg}}$$

$$F_{\max} = \text{Začetna vrednot} = -\text{INDEX}(\text{LINEST}(F5:F9;G5:G9);2) = 104.2163 \%$$

Ker je v podatkih podano razmerje F/F_{\max} velja omeniti, da je zgoraj izračunan F_{\max} najvišja (verjetno možna) vrednost sile. F_{\max} pa najvišja vrednost izmerjene sike. Naš dosedanji F je v bistvu razmerje F/F_{\max} :

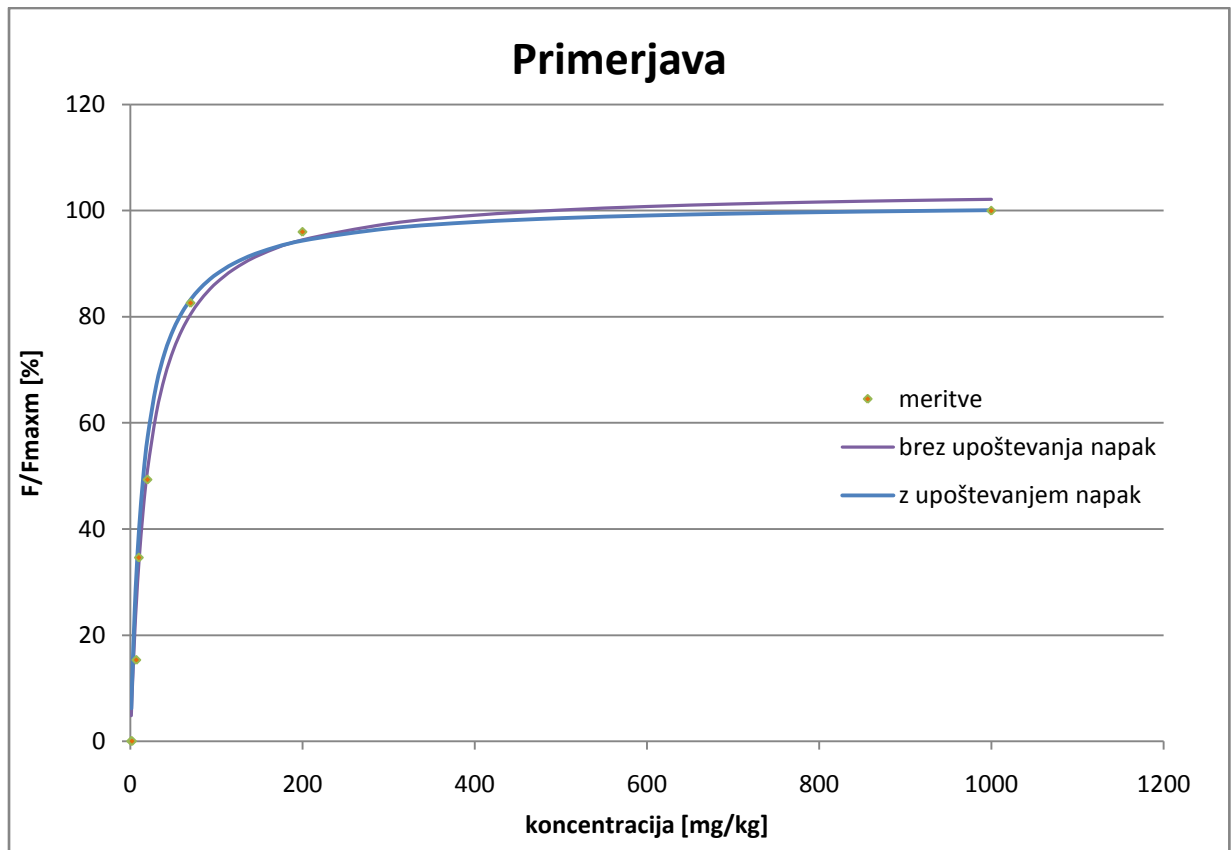


Dodamo lahko še predpostavko, da so vse meritve sile obremenjene z enako absolutno napako. Tako je v našem primeru napaka (F/c) enaka $Const/c$. Velikost $Const$ v našem primeru ne vpliva na rezultat, saj se po naravi računanja koeficienta in začetne vrednosti $Const$ uniči. Vplivala bi le na χ^2 . S to predpostavko lahko v nalogi zavzamemo vse točke:

x	y	1/σ		
0	0	1	SUM($x \cdot ro^{-2}$)	126421.1
0	0	2	SUM($y \cdot ro^{-2}$)	104268669.7
2.185714	15.3	7	SUMPRODUCT($x^2 \cdot ro^{-2}$)	29900.5
3.46	34.6	10	SUMPRODUCT($xy \cdot ro^{-2}$)	12383013.23
2.465	49.3	20	SUMPRODUCT(ro^{-2})	1045449
1.18	82.6	70	SUMPRODUCT($y^2 \cdot ro^{-2}$)	10403174906
0.48	96	200	k	-15.44470466
0.1	100	1000	n	101.6034319

$$a = -k = 15.4447 \frac{mg}{kg}$$

$$F_{max} = n = 101.6034 \% \text{ (od } F_{maxm} \text{)}$$



Po moji presoji je metoda z napakami boljša (vizualno), saj je dotika vseh točk, razen tiste, od katere sta obe metodi enako oddaljeni.