

Uvod

Linearna zveza $y = kx$ je najpreprostejša in najpogostejša zveza med dvema fizikalnima količinama, zlasti še, ker lahko tudi druge funkcijske odvisnosti v ozkem intervalu aproksimiramo z linearno zvezo: $\Delta y = k\Delta x$. Vemo, da sorazmernostni koeficient k za majhne Δx limitira k odvodu dy/dx .

Problem najboljše premice skozi dane merske točke je mogoče definirati na mnogo načinov, ki pa se vsi prevedejo na eno od dveh matematičnih oblik. Če smo za meritve že določili korelacijski koeficient (prim. 6.1-2), zadošča vedeti, da gre najboljša premica vedno skozi težišče oblaka točk (x_{pov}, y_{pov}) in da ima naklon $R\sigma_y / \sigma_x$. Sicer pa lahko koeficienta k in n najboljše premice $y = kx + n$ določimo tudi z elementarnim računom, kjer uporabljamo test hi-kvadrat.

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

Za izračun parametrov k in n , ter vrednosti χ^2 sem uporabila dane formule:

$$k := (\text{sum}1 * \text{sum}xy - \text{sum}x * \text{sum}y) / (\text{sum}1 * \text{sum}x2 - \text{sum}x * \text{sum}x);$$

$$n := (\text{sum}x2 * \text{sum}y - \text{sum}x * \text{sum}xy) / (\text{sum}1 * \text{sum}x2 - \text{sum}x * \text{sum}x)$$

in $S = \sum (y[i] - k*x[i] - n)^2 = \min$, ki jo lahko predelamo v obliko:

$$\chi^2 := \text{sum}y^2 + k*k*\text{sum}x^2 - 2*k*\text{sum}xy - 2*n*\text{sum}y + 2*k*n*\text{sum}x + n*n*\text{sum}1$$

Napake sem vključila tako, da sem vse sumande delila s kvadratom podane napake pri posameznih podatkih $(\epsilon_i^2) - x[i] / \epsilon_i^2, x*y/\epsilon_i^2$, itd.

V Excelu obstaja tudi opcija risanja linearne 'Trendline', katere rezultate sem primerjala s temi izračunanimi po prej omenjeni poti.

Primerjava rezultatov (m=9)

- izračun po danih formulah:

$$k = 1,003905$$

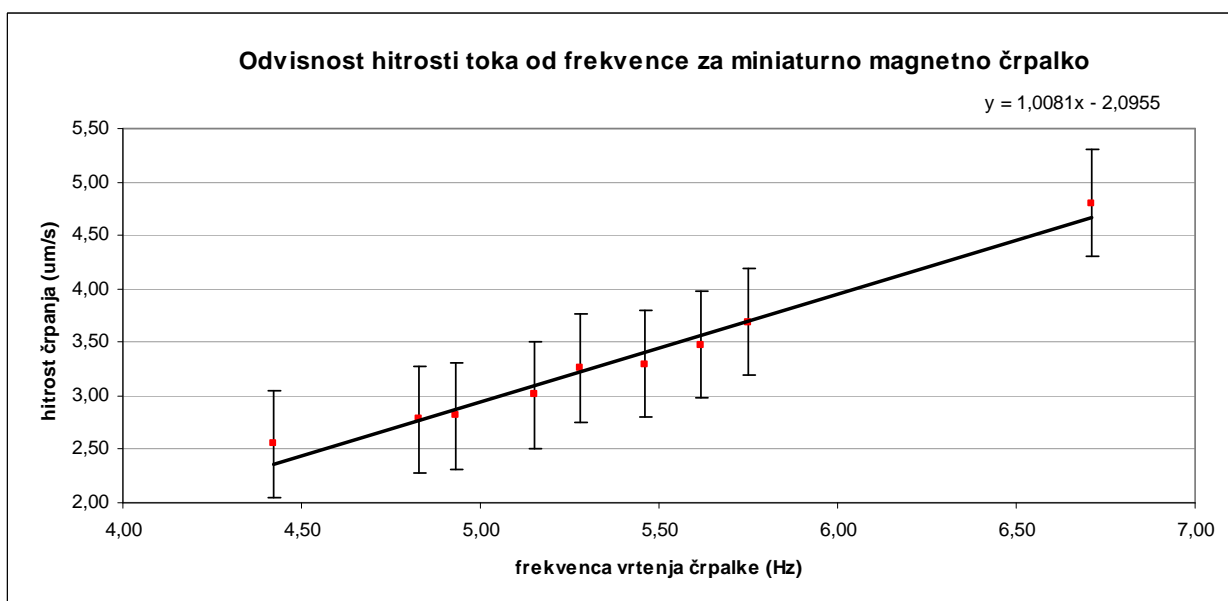
$$n = -2,07089$$

- Excel:

$$k = 1,0081$$

$$n = -2,0955$$

$\chi^2 = 5,437$, kar je v sklopu pričakovane vrednosti za dobro ujemanje - $m \pm \sqrt{2m} = 9 \pm 4,24$ (kar se sklada z veliko vrednostjo korelacijskega koeficienta izračunanega pri nalogi 6.1)



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

Izračuna najboljše linearne premice skozi oblak podatkov sem se lotila kar na vse tri načine za primerjavo. In sicer:

- s pomočjo korelacijskega koeficienta (R): $k = R\sigma_y / \sigma_x$ in $n = y_{pov} - k * x_{pov}$;
- z uporabo Excelove funkcije za najboljši fit podatkom in
- po danih formulah prek vsot

Najprej sem zajela vse podatke in po prvi metodi dobila sledeče rezultate:

Za seta podatkov:

R = -0,39409

	x	y
pov	12,4375	14,9063
sigma	12,4723	3,9398

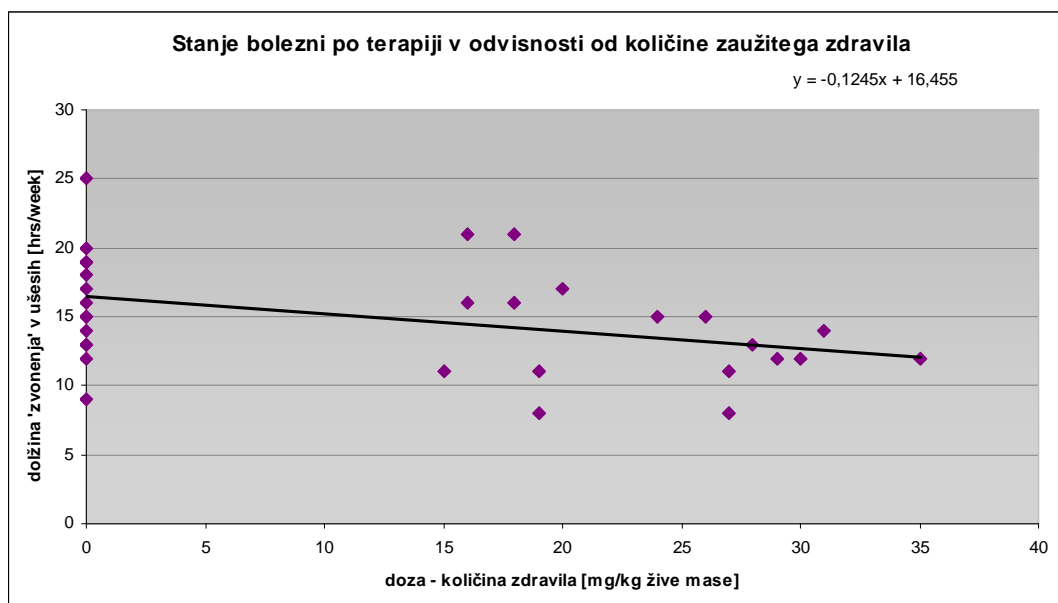
k = -0,12448

n = 16,4544

Excel da za najboljši fit sledeče rezultate:

k = -0,1245

n = 16,455



Če vzamemo za obdelavo le podatke, kjer so pacienti zaužili zdravilo pa so rezultati taki:

Za seta podatkov:

R = -0,380963033

	x	y
pov	23,41176	13,70588
sigma	5,990766	3,658426

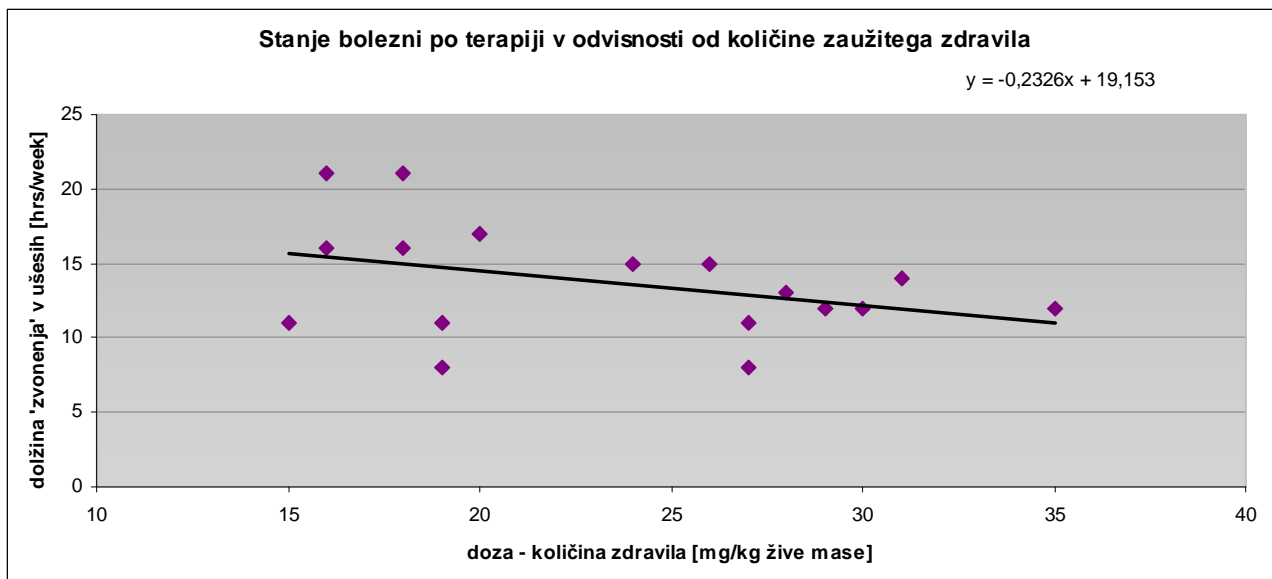
k = -0,2326x

n = 19,15

Excel da za najboljši fit sledeče rezultate:

k = -0,2326

n = 19,153



Po danih formulah prek vsot pa dobim prav tako zelo podobne rezultate:

k = -0,2326
 n = 19,153

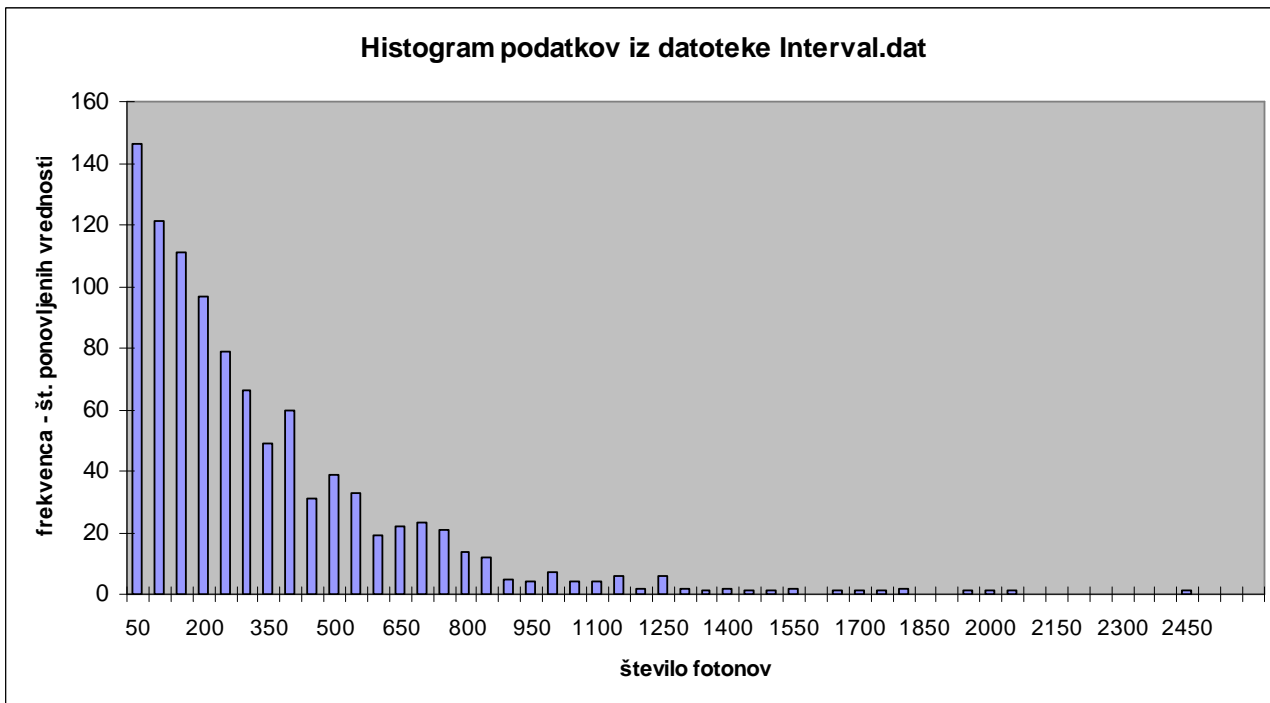
za χ^2 v tem primeru dobim 194,507, kar pa je skoraj 11-krat večje od vrednosti $m \pm \sqrt{2m} = 17 \pm 5,83$ in to se tudi sklada z majhnim korelacijskim koeficientom, ki kaže na blago zvezo med spremenljivkama.

k := (sum1*sumxy - sumx*sumy) / (sum1*sumx2 - sumx*sumx);	-0,232645584
n := (sumx2*sumy - sumx*sumxy) / (sum1*sumx2 - sumx*sumx)	19,15252603
$\chi^2 = \text{sumy}^2 + k^2 * \text{sumx}^2 - 2 * k * \text{sumxy} - 2 * n * \text{sumy} + 2 * k * n * \text{sumx} + n^2 * \text{sumi}$	194,5074238

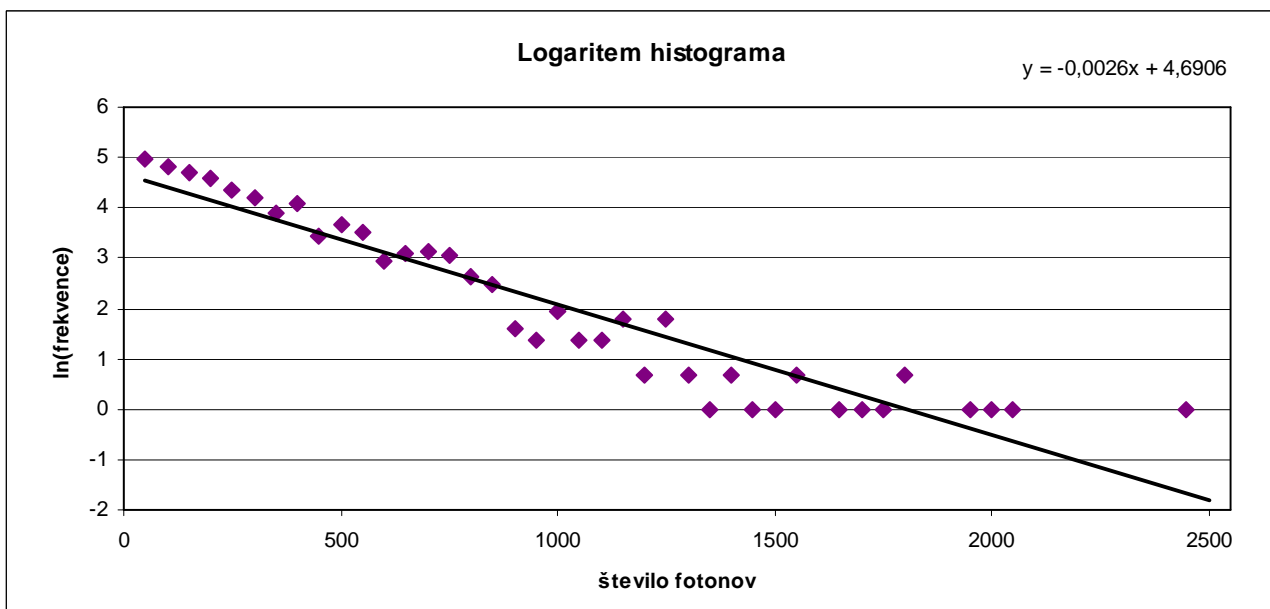
Iz dobrega ujemanja rezultatov prvih dveh nalogah lahko torej sklepam, da Excel uporablja načelo najmanjših kvadratov ali hi-kvadrat test za prilagajanje grafov.

3.)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.

Iz danih podatkov sem v Excelu narisala naslednji histogram.



Logaritmirane vrednosti y sem nato narisala v odvisnosti od števila fotonov in dobila približno linearno odvisnost:

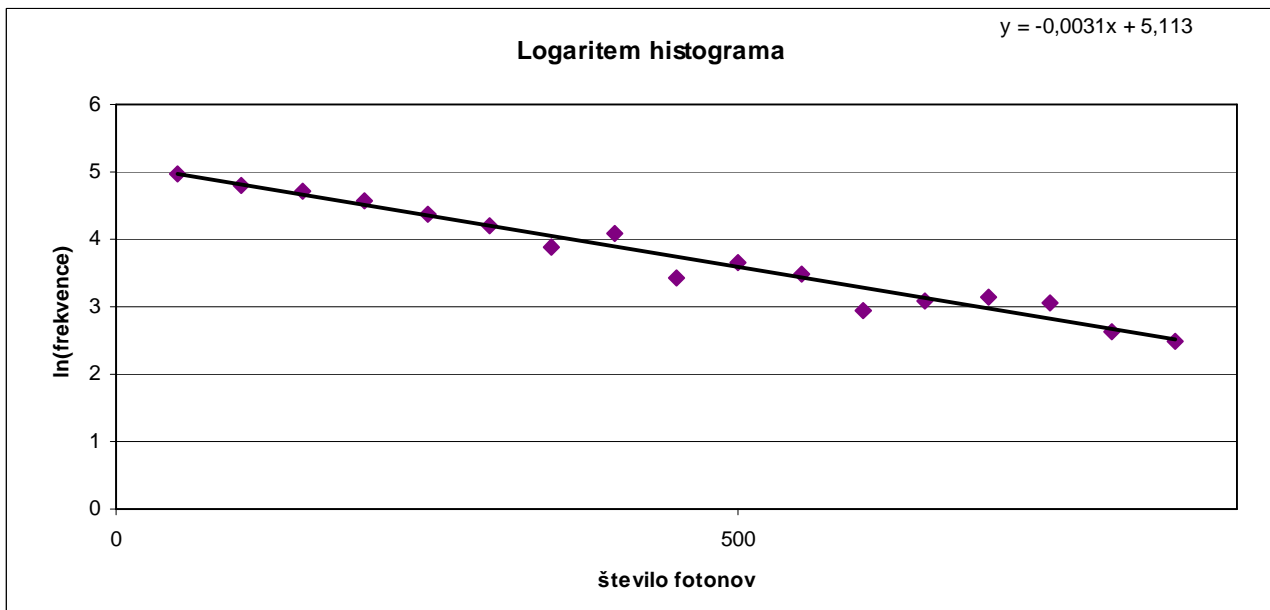


V prvem primeru sem izračunala najboljšo prilagojeno linearno premico za vse podatke. Parametri so zelo podobno tistim dobljenim s funkcijo Trendline v Excelu:

	iz formul	v Excelu
$k = -\lambda$	-0,00221	-0,0026
$n = \ln(A)$	4,828006	4,6906

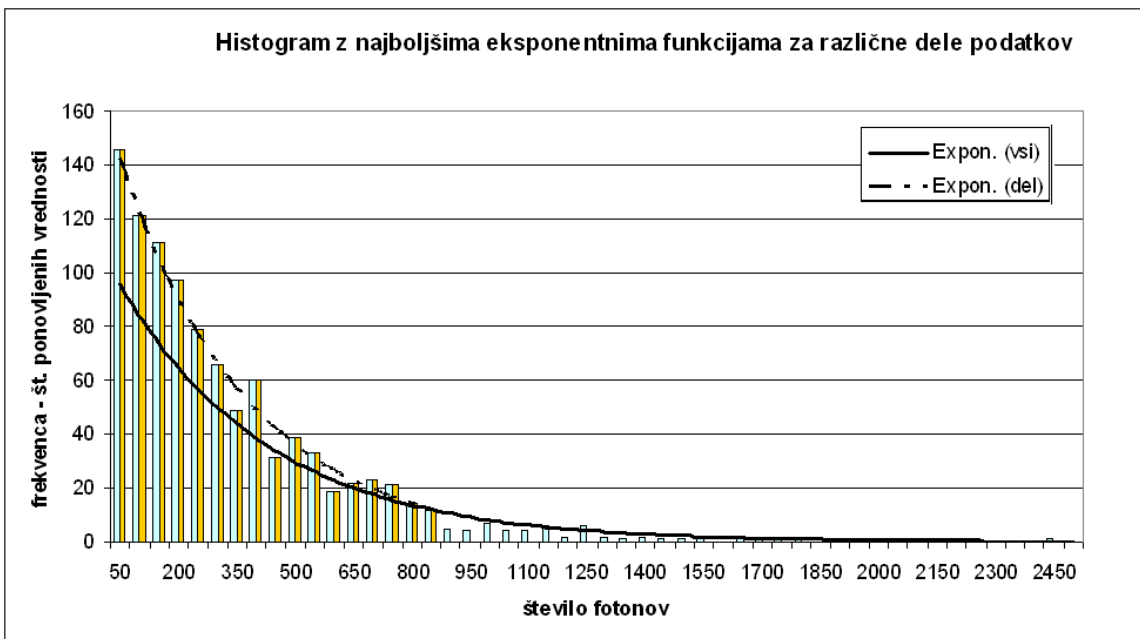
Iz formul:

Nato pa še le za del podatkov (pri prejšnjem grafu namreč vidimo, da so točke od vrednosti 1000 naprej močno raztresene v primerjavi s tistimi pred $x=1000$, slednje so veliko bolj linearno urejene):



	iz formul	v Excelu
$k = -\lambda$	-0,003056	-0,0031
$n = \ln(A)$	5,1130144	5,113

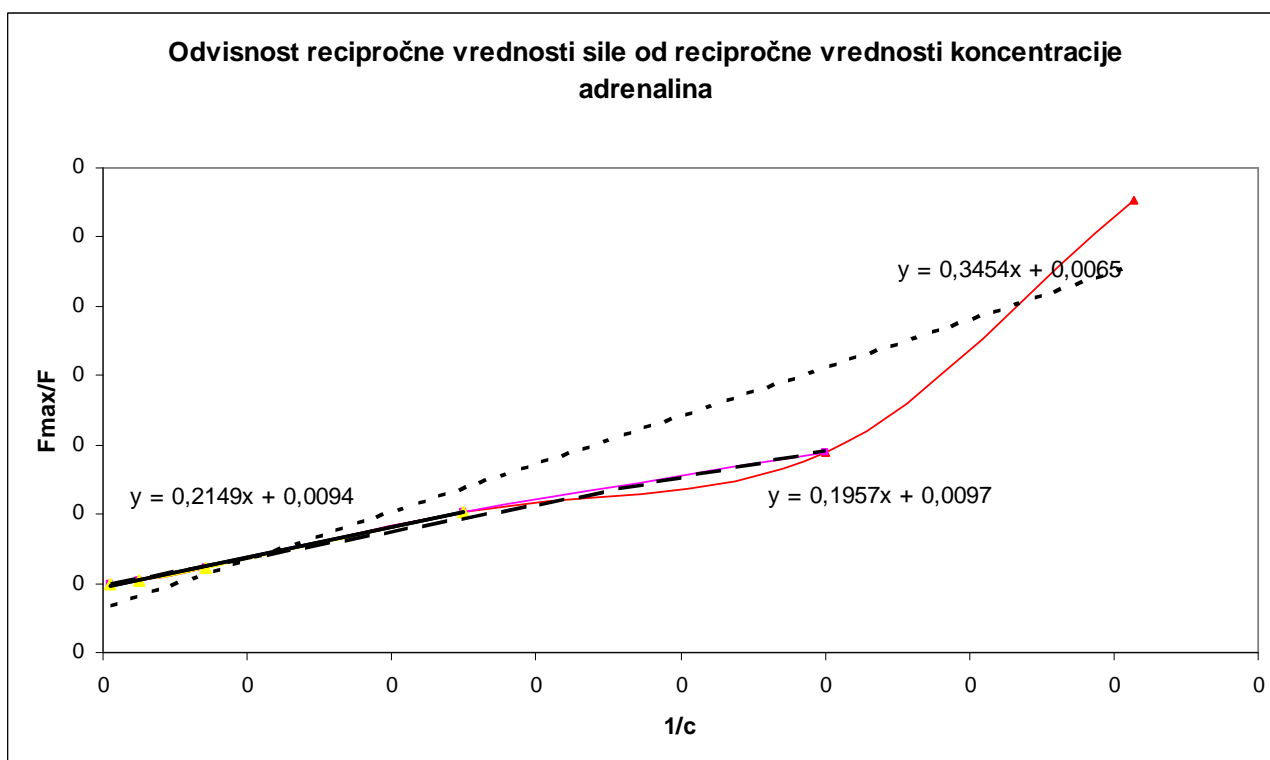
Recipročna povprečna vrednost histograma je zelo podobna vrednosti λ , tako namreč Excel računa najboljše prilegajočo eksponentno krivuljo.



4.) 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 .

Pri tej nalogi sem na osi vnesla recipročne vrednosti podanih spremenljivk in tako dobila linearno zvezo. Seveda bi bilo, kot je že omenjeno v uvodu, v primeru, da bi imeli na voljo več parov (x,y), to razmerje veliko bolj podobno linearnemu. Nam namreč od že tako malo, 8 točk, ostane le še 6 točk za odbelavo.

	Vsi podatki	5 zadnjih točk	4 zadnje točke
k	0,0345397	0,1957	0,2149
n	0,006467	0,0097	0,0094



Najboljše premice za različno število podatkov (točk)

Vrednost a dobimo iz formule:

$$1/F = F_{max} * c / (c+a)$$

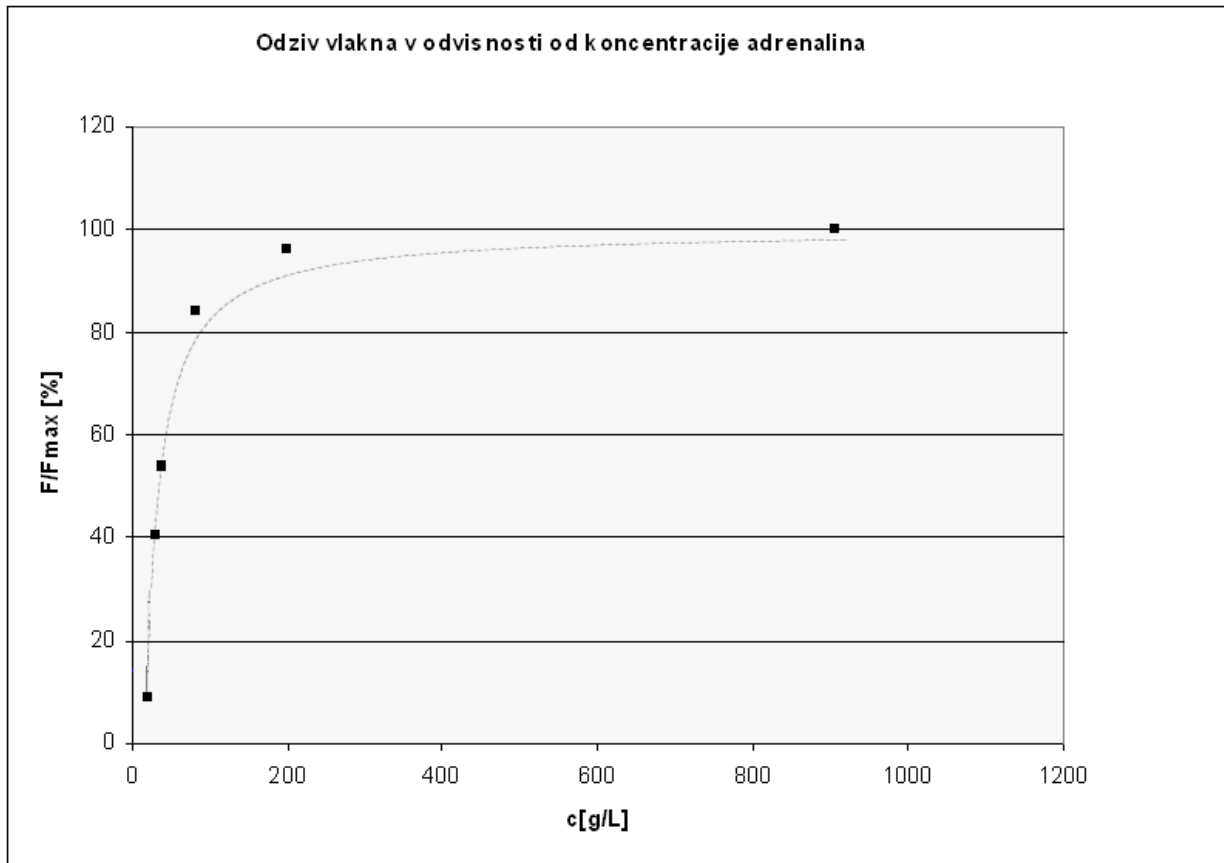
če enačbo zapišemo z recipročnimi vrednostmi, dobimo zapis za linearno funkcijo:

$$1/(F/F_{max}) = 1 + a * 1/c$$

Iz tega je razvidno, da je a enak koeficientu najbolj prilegajoče se linearne funkcije podatkom, iz česar sledi, da je $a = 0,2149$ (tu sem izbrala enačbo funkcije, ki se najbolj prilega)

Iz tega podatka, lahko naredimo najboljše prilegajočo se funkcijo osnovnemu grafu:

$$F_{max}/F = c/(c+a)$$



Linearna zveza $y = kx$ je najpreprostejša in najpogostejša zveza med dvema fizikalnima količinama, zlasti še, ker lahko tudi druge funkcijske odvisnosti v ozkem intervalu aproksimiramo z linearno zvezo: $\Delta y = k\Delta x$. Vemo, da sorazmernostni koeficient k za majhne Δx limitira k odvodu dy / dx . Kadar moramo določiti koeficient v linearni zvezi dveh količin x in y , je s stališča merske tehnike koristno izmeriti več kot en par (x, y) . Če določimo vrednosti y_i , ki ustrezajo celi vrsti izbranih x_i , $i = 1, 2, \dots, m$, lahko z njimi določimo k z večjo natančnostjo. Obenem preverimo, ali linearna zveza zares velja, pa tudi, ali naša merska naprava dobro deluje v širšem intervalu spremenljivk. Zlasti se lahko zgodi, da ima merilec te ali druge količine premaknjeno ničlo. Tedaj se naša linearna zveza ne pokaže kot premica skozi izhodišče koordinatnega sistema, pač pa je nekoliko premaknjena. Zato velja, da iz vrste meritev x_i, y_i nikoli ne računamo koeficienta kot povprečje kvocientov y_i / x_i , pač pa vedno kot naklon premice, ki jo potegnemo skozi izmerjene točke.

Problem najboljše premice skozi dane merske točke je mogoče definirati na mnogo načinov, ki pa se vsi prevedejo na eno od dveh matematičnih oblik. Če smo za meritve že določili korelacijski koeficient (prim. 6.1-2), zadošča vedeti, da gre najboljša premica vedno skozi težišče oblaka točk (x_{pov}, y_{pov}) in da ima naklon $R\sigma_y / \sigma_x$.

Sicer pa lahko koeficienta k in n najboljše premice $y = kx + n$ določimo tudi z elementarnim računom. Zahtevamo, naj bo vsota kvadratov razdalj točk od premice najmanjša (načelo najmanjših kvadratov – least squares principle):

$$S = \sum (y_i - kx_i - n)^2 = \min$$

Sistem dveh enačb za k in n rešimo z naslednjim programom:

```
sum1 := 0;
sumx := 0;
sumy := 0;
sumx2 := 0;
sumxy := 0;
for i:=1 to m do begin
  add(sum1, 1);
  add(sumx, x[i]);
  add(sumy, y[i]);
  add(sumx2, x[i]*x[i]);
  add(sumxy, x[i]*y[i])
end;
k := (sum1*sumxy - sumx*sumy)/(sum1*sumx2 - sumx*sumx);
n := (sumx2*sumy - sumx*sumxy)/(sum1*sumx2 - sumx*sumx);
```

V zaresni merilni tehniki se spodobi, da so meritve opremljene s statistično napako, torej

$y_i = \bar{y}_i \pm \epsilon_i$. Kadar se ϵ_i od izmerka do izmerka znatno spreminja, je teža posameznih točk različna. Statistično neoporečen rezultat dobimo, če v gornjem programu vse sumande utežimo z ϵ_i^{-2} :

```
add(sumz, z[i]/sqr(epsilon[i]));
```

kjer je $z[i]$ po vrsti 1, $x[i]$, $y[i]$, $x[i]*x[i]$ in $x[i]*y[i]$. To je druga, popolnejša oblika za najboljšo premico. V njej vsoto S , ki smo jo minimizirali, običajno označimo s χ^2 in jo lahko uporabljamo tudi za diagnozo kvalitete ujemanja. Za dobro ujemanje je njena pričakovana vrednost $m \pm \sqrt{2m}$. Govorimo o testu hi-kvadrat.

Mnoga računska orodja imajo že vgrajene ukaze za najboljše premice. Velja preveriti, katero od obeh oblik uporabljajo.

Naloge

1. Za meritve^[1] v datoteki "HitrostTokaOdFrekvence.txt" (naloga 6.1) določi parametra najboljše premice. Ker so podane napake hitrosti, lahko določiš tudi χ^2 .
2. Skozi oblak podatkov "Tintin.dat" potegni najboljšo premico. Uporabiš lahko kar korelacijske rezultate iz naloge 6.2.
3. 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.

4. 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 .

-0,394090046

pov	12,4375	14,90625
sigma	12,4723131	3,939855446