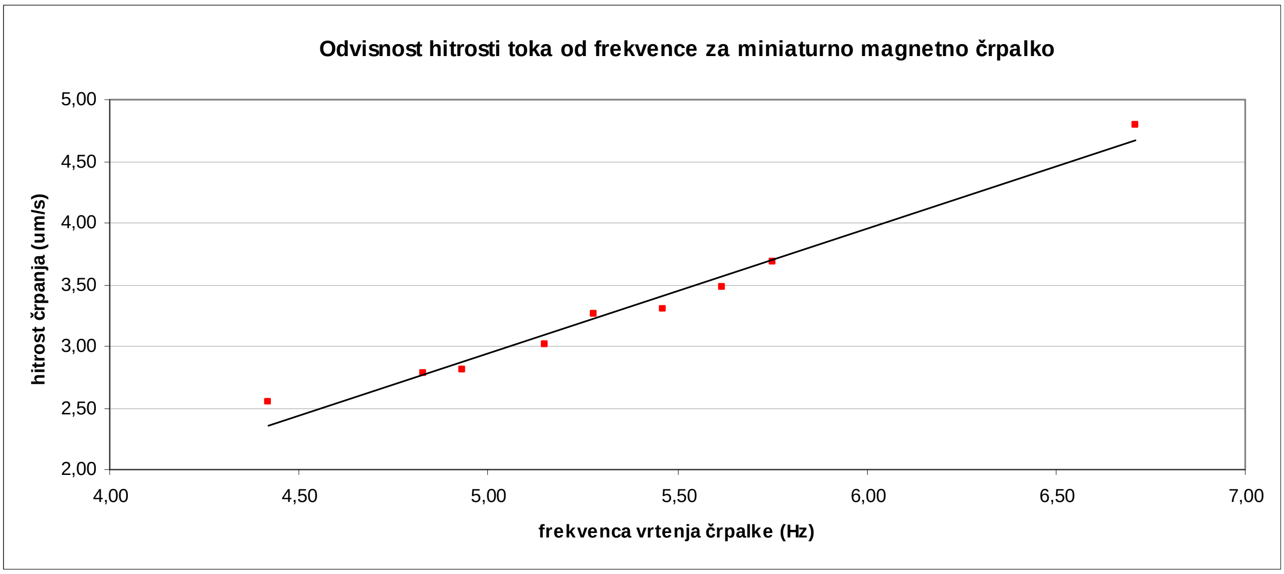
1.) V najnovejši številki Obzornika je objavljen zanimiv članek o miniaturni magnetni črpalki.[1] Avtorji napovedo linearno zvezo med frekvenco rotorja in hitrostjo toka; meritve v datoteki "HitrostTokaOdFrekvence.txt" to potrjujejo. Določi korelacijski koeficient zveze med obema količinama.



Korelacijski koeficient zveze med obema količinama sem računala v excelu in sicer na dva načina. Prvič sem v excelu obdelala podatke po dani formuli v navodilih:

`R(a,b) = (r(a,b) – apov*bpov)/(sigmaa*sigmab).`

Po tej formuli je korelacijski koeficient prišel: **0,987970185**

Obstaja pa v excelu tudi funkcija CORREL, ki za izbrana dva seta podatkov izračuna ta koeficient. V tem primeru pride enako: **0,987970185**

Funcijo lahko izberemo s seznama funkcij, obstaja pa tudi pod zavihkom Tools – Data analysis – Correlation.

Ob upoštevanju napake, se korelacijski koeficient nekoliko spremeni in sicer, če prištejemo napake vrednostim hitrosti se korelacijski koeficient zmanjša, če pa napako odštejemo se poveča:

brez napake	**0,987970185**
prišteta napaka	0,987948677
odšteta napaka	0,987975012

---

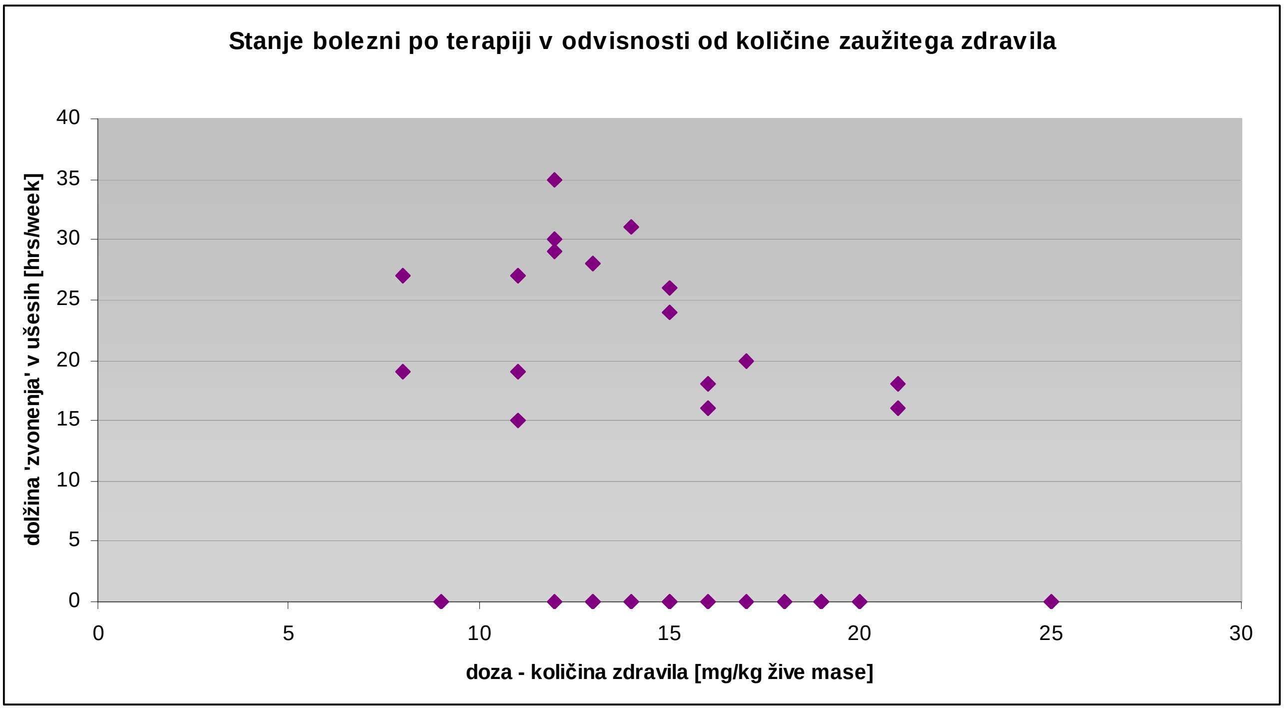
*Za ugotavljanje kako močno sta povezana dana seta podatkov sem uporabila tudi Spermanov rank. Prva formula je poenostavljena verzija druge za namen lažjega programiranja. Rezultati naj bi bili enaki.

Spearman Rank Correlation Coefficient l

$$r' \equiv 1 - 6 \sum \frac{d^2}{N(N^2-1)},$$: **0,999572**

$$r \equiv \frac{\sum x\, y}{\sqrt{\sum x^2 \sum y^2}}$$: **0,996997685**

2.) Ameriška uprava za zdravila (FDA – Food and Drug Administration) je preskusila čudežno zdravilo mirabilitin za zvonjenje v ušesih (tintinabulus). V datoteki "Tintin.dat" so podani rezultati dvojno slepega preskusa. Določi korelacijski koeficient med dozo (v mg/kg žive mase) in stanjem bolezni po terapiji (ur zvonjenja na teden).



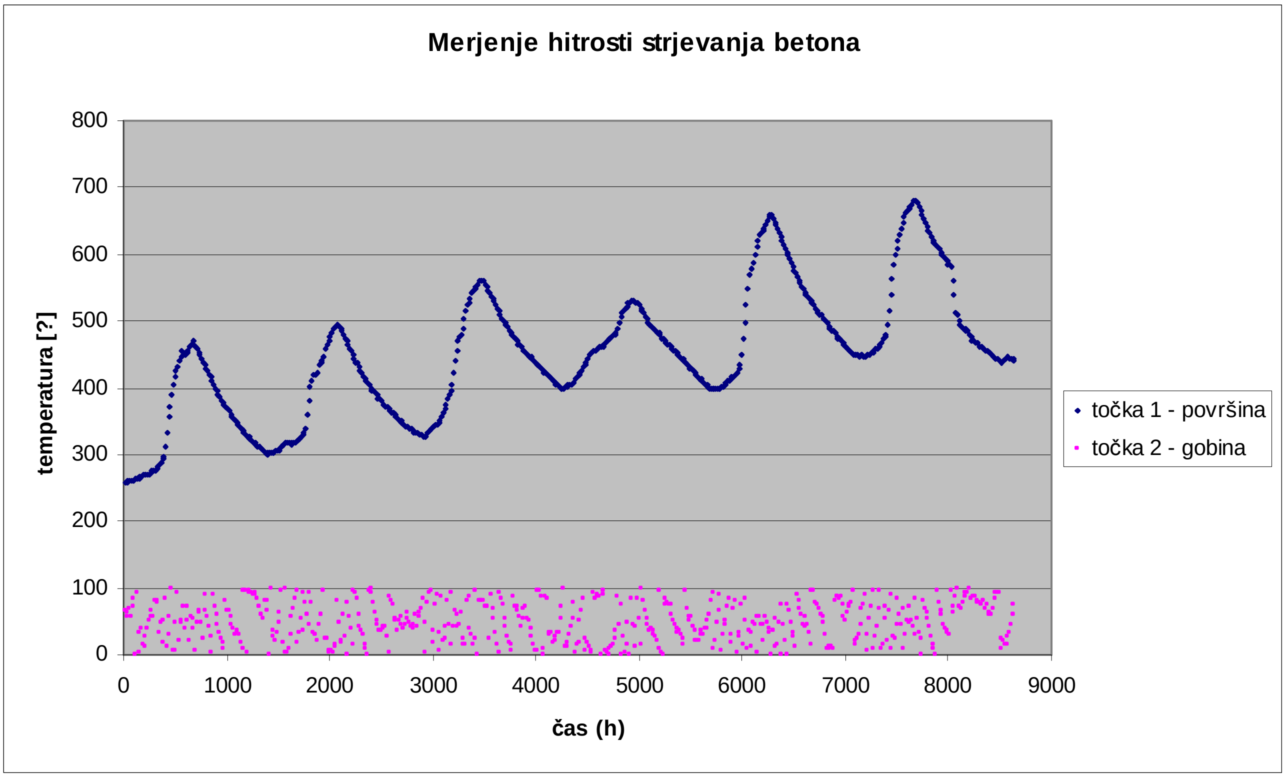
Izračuni po istem postopku kot v prejšnji nalogi:

Korelacijski koeficient (po dani formuli): **-0,394090046**

Korelacijski koeficient (s funkcijo CORRREL): **-0,394090046**

Rezultati kažejo šibko povezavo med količino zaužitega zdravila in količino 'zvonjenja' v ušesih. Vseeno, glede na to da je za oddobritev zdravila potrebna vrednost korelacijskega koeficienta vsaj 0,3 absolutno, bi bilo to zdravilo praviloma odobreno.

3.) Pred leti smo v okviru mednarodnega projekta v našem znanem gradbenem podjetju merili hitrost strjevanja betona. Ulili so nekaj metrov velik betonski blok, v katerega je bila vdelana cela vrsta termočlenov za sprotno merjenje temperature. Datoteka "Beton.dat" podaja izmerke v razdobju šestih dni v dveh merilnih točkah. Prva je blizu površine, druga globoko v notranjosti. (Prvi stolpec je zaporedna številka meritve – časovni interval med njimi lahko oceniš iz očitnih dnevnih nihanj temperature.) Določi efektivno zakasnitev med obema signaloma iz njune korelacijske funkcije.



Graf nezamaknjenih funkcij

Korelacijski koeficient med setoma podatkov za primer, ko druga krivulja ni zamaknjena po x osi je: -0,06666, kar pomeni zelo šibko povezavo.

Ko drugo krivuljo (rezultati meritev v globini) zamaknemo po x osi, se korelacijski koeficient postopoma povečuje in dosežće maksimalno vrednost pri zamiku za 9 merilnih točk/kvadratkov, kar časovno ustreza približno 2h in 12min.

Takrat je korelacijski koeficient enak **-0,0906961**

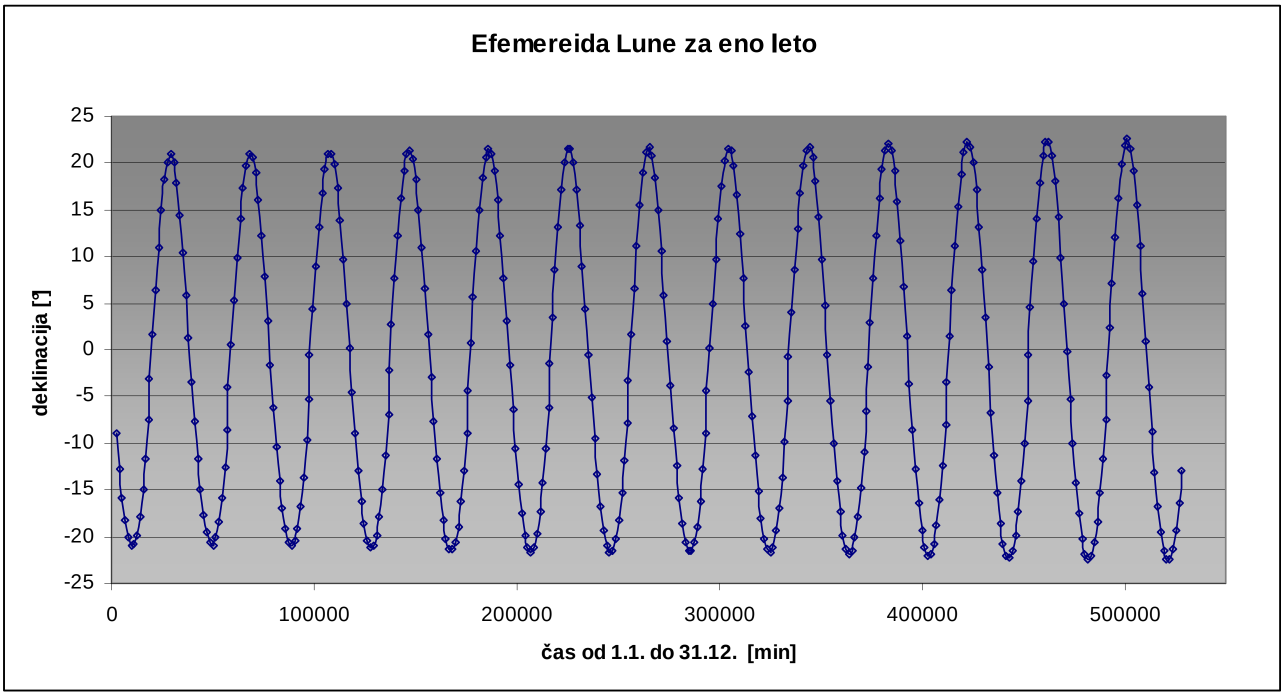
Korelacijska funkcija pa bi se glasila:

K(y) = (a(x), b(x+9)),

oz. izraženo s časom v urah:

K(y) = (a(x), b(x+2.19))

4.)V datoteki "Luna.efe" je dana efemerida Lune za eno od preteklih let. Stolpci so: dan začenši s 1.1., nato rektascenzija (nebesna dolžina) v urah in minutah, in nazadnje deklinacija (nebesna širina) v stopinjah, ob 0^h svetovnega časa tega dne. Iz avtokorelacijske funkcije deklinacije čim bolj natančno določi Lunino periodo tira. (Lahko si pomagaš z odvajanjem.)



Maksimalni korelacijski koeficienti se pojavljajo pri zamikih za 14 dni. Ta maksimalna vrednost na določenem intervalu se povečuje. Takšno povečevanje, širjenje, vidimo tudi zgoraj na grafu (maksimalne vrednosti so vedno večje in minimalne vrednosti vedno manjše s časom).

Prva maksimalna vrednost korelacijskega koeficienta se pojavi pri zamiku za 13 dni : **0,058923** naslednja pa potem čez 27 dni: **0,065025** itd.

Perioda tira Lune: cca. 39300min = 27,3 dni

*** Spearman Rank Correlation Coefficient**

A nonparametric (distribution-free) rank statistic proposed by Spearman in 1904 as a measure of the strength of the associations between two variables (Lehmann and D'Abrera 1998). The Spearman rank correlation coefficient can be used to give an R-estimate, and is a measure of monotone association that is used when the distribution of the data make Pearson's correlation coefficient undesirable or misleading.

The Spearman rank correlation coefficient is defined by

$r' \equiv 1 - 6 \sum \frac{d^2}{N(N^2-1)},$ where d is the difference in statistical rank of corresponding variables, and is an approximation to the exact correlation coefficient

$$r \equiv \frac{\sum x\, y}{\sqrt{\sum x^2 \sum y^2}}$$

computed from the original data. Because it uses ranks, the Spearman rank correlation coefficient is much easier to compute.

Source: [http://mathworld.wolfram.com/SpearmanRankCorrelationCoefficient.html]

V fiziki naletimo na pojem skalarnega produkta v dveh poglavitnih povezavah. Prva je skalarni produkt vektorjev $\vec{a} \cdot \vec{b}$, ki se s komponentami vektorjev zapiše kot

```
skalab := 0;
for i := 1 to 3 do skalab := skalab + a[i]*b[i]
```

Postopek se razširi na končnodimenzionalne vektorje iz linearne algebre preprosto tako, da poženemo zanko od 1 do `n`, kjer je `n` število dimenzij.

Druga povezava, v kateri naletimo na skalarni produkt, so funkcijski prostori. V Hilbertovem prostoru, ki je prostor funkcij *f*(*x*), $p \leq x \leq q$, za katere obstaja integral

$$\int_p^q f^2(x)\mathrm{d}x,$$

je skalarni produkt vektorjev/funkcij *a*(*x*) in *b*(*x*) definiran kot

$$(a, b) = \int_p^q a(x)b(x)\mathrm{d}x.$$

Naj si predstavljamo stolpca `a[i]` in `b[i]` v tabeli podatkov kot `n`-razsežna vektorja ali kot funkciji, ki sta tabelirani v enakomernih razmikih, obakrat pridemo do enakega recepta za skalarni produkt, namreč tistega v prvi vrstici. Kvečjemu bi v primeru funkcij dobljeni rezultat `skalab` pomnožili še s `h`, z velikostjo koraka v tabeli. Za neenakomerno tabelo bi skalarni produkt uvedli po vzorcu, ki smo ga uporabili pri integriranju (5. tema).

Skalarni produkt meri projekcijo enega vektorja na drugega. V harmonski analizi razstavimo signal *a*(*x*) v sinusne sestavine, tako da poiščemo skalarni produkt signala s sinusi in kosinusi različnih valovnih dolžin:

$$C(k) = (a(x), \cos(kx)),$$
$$S(k) = (a(x), \sin(kx)).$$

Funkcije *C*(*k*) in *S*(*k*) in še bolje, njuna pitagorejska vsota $\sqrt{C(k)^2 + S(k)^2}$, imajo ekstrem pri tistih periodah 2π / *k*, ki so tudi periode signala.

Skalarni produkt dveh stolpcev podatkov (*a*,*b*) meri tudi korelacijo teh dveh količin, kadar ju razumemo kot statistični (naključni) spremenljivki. Če vpeljemo povprečno vrednost produkta

```
r(a,b) = skalab/n
```

in jo še primerno premaknemo in normiramo

```
R(a,b) = (r(a,b) – apov*bpov)/(sigmaa*sigmab)
```

smo dobili brezdimenzijski korelacijski koeficient, ki lahko zavzame vrednosti med -1 in 1. Pozitivne vrednosti pomenijo, da sta spremenljivki korelirani, se pravi, da ena količina v povprečju raste, ko raste druga. Negativne vrednosti `R` pomenijo antikorelacijo, ko vrednosti prve količine v povprečju padajo, ko druge naraščajo. Pri tem so vrednosti na obeh konceh intervala rezervirane za jasne linearne odvisnosti količin; tiste v sredi, recimo med -0.3 in 0.3, pa za blago, komaj razločno povezavo med gibanjem obeh spremenljivk.

Pomembna karakteristika povezave signalov *a* in *b* je njuna korelacijska funkcija

$$K(y) = (a(x), b(x+y)),$$

torej skalarni produkt pri zakasnitvi *y*. Iz nje zelo jasno razberemo morebitne fazne premike signalov: zgodi se, da je skalarni produkt signalov majhen, da pa zrase, ko enega od signalov premaknemo po osi *x*. Izračunamo jo lahko za vrednosti *y* od -*n* do *n*. Seveda postaja z večanjem premika *y* interval, po katerem lahko računamo skalarni produkt, vse krajši, zato je definicijo smiselno popraviti tako, da računamo povprečno vrednost produkta – se pravi, da pri vsakem *y* delimo rezultat s številom točk, na katerih se signala sploh še prekrivata.

Tako funkcijo je zanimivo izračunati celo za en sam signal, takrat jo imenujemo avtokorelacijska funkcija:

$Q(y) = (a(x), a(x + y))$.

Pri periodičnih signalih lahko iz ekstremov *Q* zelo natančno odčitamo periodo signala, pri neperiodičnih pa zvemo, kako daleč se signal "spominja" prejšnjega poteka.

Naloge:

1. V najnovejši številki Obzornika je objavljen zanimiv članek o miniaturni magnetni črpalki.[1] Avtorji napovedo linearno zvezo med frekvenco rotorja in hitrostjo toka; meritve v datoteki "HitrostTokaOdFrekvence.txt" to potrjujejo. Določi korelacijski koeficient zveze med obema količinama.
2. Ameriška uprava za zdravila (FDA – Food and Drug Administration) je preskusila čudežno zdravilo mirabilitin za zvonjenje v ušesih (tintinabulus). V datoteki "Tintin.dat" so podani rezultati dvojno slepega preskusa. Določi korelacijski koeficient med dozo (v mg/kg žive mase) in stanjem bolezni po terapiji (ur zvonjenja na teden).
3. Pred leti smo v okviru mednarodnega projekta v našem znanem gradbenem podjetju merili hitrost strjevanja betona. Ulili so nekaj metrov velik betonski blok, v katerega je bila vdelana cela vrsta termočlenov za sprotno merjenje temperature. Datoteka "Beton.dat" podaja izmerke v razdobju šestih dni v dveh merilnih točkah. Prva je blizu površine, druga globoko v notranjosti. (Prvi stolpec je zaporedna številka meritve – časovni interval med njimi lahko oceniš iz očitnih dnevnih nihanj temperature.) Določi efektivno zakasnitev med obema signaloma iz njune korelacijske funkcije.
4. V datoteki "Luna.efe" je dana efemerida Lune za eno od preteklih let. Stolpci so: dan začenši s 1.1., nato rektascenzija (nebesna dolžina) v urah in minutah, in nazadnje deklinacija (nebesna širina) v stopinjah, ob 0^h svetovnega časa tega dne. Iz avtokorelacijske funkcije deklinacije čim bolj natančno določi Lunino periodo tira. (Lahko si pomagaš z odvajanjem.)