

Lineárna väzba dvoch dátových súborov

Je dané

```
>> x=rand(1,50)*10;  
>> a=0.5; y=x+a*rand(1,50);
```

Úloha: Vypočítajte priemer a rozptyl súborov x, y (s delením $1/n$ aj $1/(n-1)$).

Počítajme kovarianciu priamo podľa vzorca:

```
>> n=length(x); xx=mean(x); yy=mean(y); sxy=(x-xx)*(y-yy)/n
```

```
sxy = 8.4606
```

S korekciou:

```
>> yy=mean(y); sxy=(x-xx)*(y-yy)/(n-1)
```

```
sxy = 8.6332
```

To isté nám MatLab odpovie aj pri kratšie formulovanej otázke:

```
>> cov(x,y)
```

```
ans =
```

```
8.6799 8.6332  
8.6332 8.6047
```

Ak si neželáme korekciu, pýtame sa s parametrom 1:

```
>> cov(x,y,1)
```

```
ans =
```

```
8.5063 8.4606  
8.4606 8.4326
```

Odpoveďou je v oboch prípadoch matica – Matlab teda odpovedá širšie. Kde v matici je hľadaná kovariancia? Čo znamenajú ostatné čísla v matici?

Korelačný koeficient priamo zo vzorca:

```
>> cc=(x-xx)*(y-yy)/(((x-xx)*(x-xx))*((y-yy)*(y-yy)))^0.5  
cc = 0.9990
```

A teraz „exkluzívne od Matlabu“:

```
>> ccm=corrcoef(x,y)  
ccm =  
    1.0000    0.9990  
    0.9990    1.0000
```

Kde v matici je ukrytá odpoveď na otázku?

Nakreslíme regresnú priamku. Vysvetlite (sami sebe a susedovi) detailne každý z príkazov:

```
>> cm=cov(x,y); k=cm(1,2)/cm(1,1)  
k = 0.9946  
>> yr=yy+k*(x-xx);  
>> xt=[min(x), max(x)]; yt=yy+k*(xt-xx);  
>> plot(x,y,'x'), hold on, plot(xt,yt), plot(xx,yy,'dr')
```

Ak ste sa nepomýlili, obrázok vidíte v okne matlabu....

Ešte jedna drobnosť:

```
>> dy=y-yr; sse=dy*dy'  
sse = 0.8803
```

Čo sme to práve vypočítali??

Úlohy:

- Meňte koeficient a v predošlom príklade – zväčšujte ho (smelo) a sledujte, ako sa bude meniť kovariancia a korelačný koeficient. To isté pre $a = -1, -2, \dots$ (smelo znižujte).
- Zvoľte si rôzne iné dvojice vektorov x, y (nápadito!) a kreslite regresnú priamku.
- Napište celý postup kreslenia regresnej priamky ako m-funkciu.