

Regresná priamka a iné

Počty 19 najkradnutejších značiek r. 2007 na Slovensku a počty oficiálne predaných vozidiel týchto značiek v kategóriách M1 a N1:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Meno	Skoda	VW	Lada	Peugeot	Audi	Seat	BMW	Fiat	Renault	Citroen	Ford	Opel	MB	Toyota	Mitsubishi	Nissan	Mazda	Chevrolet	Honda
Krad. 2007	2087	552	163	149	130	121	104	104	88	60	59	52	36	31	16	14	14	13	9
M1 2007	13017, 3474, 103, 4141, 803, 1745, 544, 1975, 3153, 3541, 3390, 2831, 1260, 2522, 377, 1231, 665, 977, 1223																		
N1 2007	4946, 2326, 11, 1920, 696, 349, 789, 1298, 1368, 2287, 721, 734, 1607, 408, 317, 259, 70, 103, 341																		

>> Kr=[2087 552 163 149 130 121 104 104 88 60 59 52 36 31 16 14 14 13 9]

>> M=[13017, 3474, 103, 4141, 803, 1745, 544, 1975, 3153, 3541, 3390, 2831, 1260, 2522, 377, 1231, 665, 977, 1223]

>> N=[4946, 2326, 11, 1920, 696, 349, 789, 1298, 1368, 2287, 721, 734, 1607, 408, 317, 259, 70, 103, 341]

Sme si vedomí, že sa kradnú aj iné ako nové autá. Ak predsa porovnáваме počty predaných a ukradnutých vozov, porovnáваме tým aktuálny záujem trhu o značku a pýtame sa na súvis (koreláciu) medzi záujmom legálnej a kriminálnej zložky trhu.

Ktorá je najkradnutejšia značka vzhľadom na počet predaných vozidiel? Inými slovami, pri ktorej značke je záujem kriminálneho trhu v pomere k legálnemu najsilnejší? Pri práci v Matlabe nebudeme používať značky áut, ale ich poradové číslo v rebríčku absolútnych počtov kradnutých kusov.

>> P1=Kr./M; ma1=max(P1), find(P1==ma1)

ma1 = 1.582524271844660, ans = 3

Vítazom je značka Lada, kde počet kradnutých vozov o skoro 60 percent prevyšuje počet predaných. Ľudové ponaučenie – ak si do firmy kúpite dve Lady, počítajte s tým, že tri z nich vám ukradnú... Bol to extrémny príklad, pozrime sa preto radšej na ďalšie pozície:

>> for i=1:19, por(i)=find(P1s(i) ==P1); end, por, P1(por)

por = 19 16 14 18 10 11 12 17 9 13 4 15 8 6 2 1 5 7 3

ans =

```
7.358953393295176e-003  1.137286758732738e-002  1.229183187946075e-002  1.330603889457523e-002
1.694436599830556e-002  1.740412979351032e-002  1.836806782055811e-002  2.105263157894737e-002
2.790992705359975e-002  2.857142857142857e-002  3.598164694518232e-002  4.244031830238727e-002
5.265822784810127e-002  6.934097421203439e-002  1.588946459412781e-001  1.603288007989552e-001
1.618929016189290e-001  1.911764705882353e-001  1.582524271844660e+000
```

Druhým najkradnutejším je BMW (č. 7), kde záujem kriminálneho trhu dosahuje 19% záujmu oficiálneho trhu. Tretím je Audi (16%), štvrtou Škoda (16.03%), piatym VW (15.88%), šiestym Seat (7%) a siedmym Fiat (5.3%). Všimnime si, že pozície 3.-6. kontinuálne obsadili štyri značky jedného koncernu...

Urobte podobné porovnanie s použitím hodnôt vo vektore N a M+N.

1. Aký je celkový súvis medzi absolútnymi počtami kradnutosti a predávanosti značiek? Je rozdiel medzi M1 a N1?

```
>> corrcoef(Kr, M)
ans =
    1.000000000000000e+000    9.076502494305891e-001
    9.076502494305891e-001    1.000000000000000e+000
```

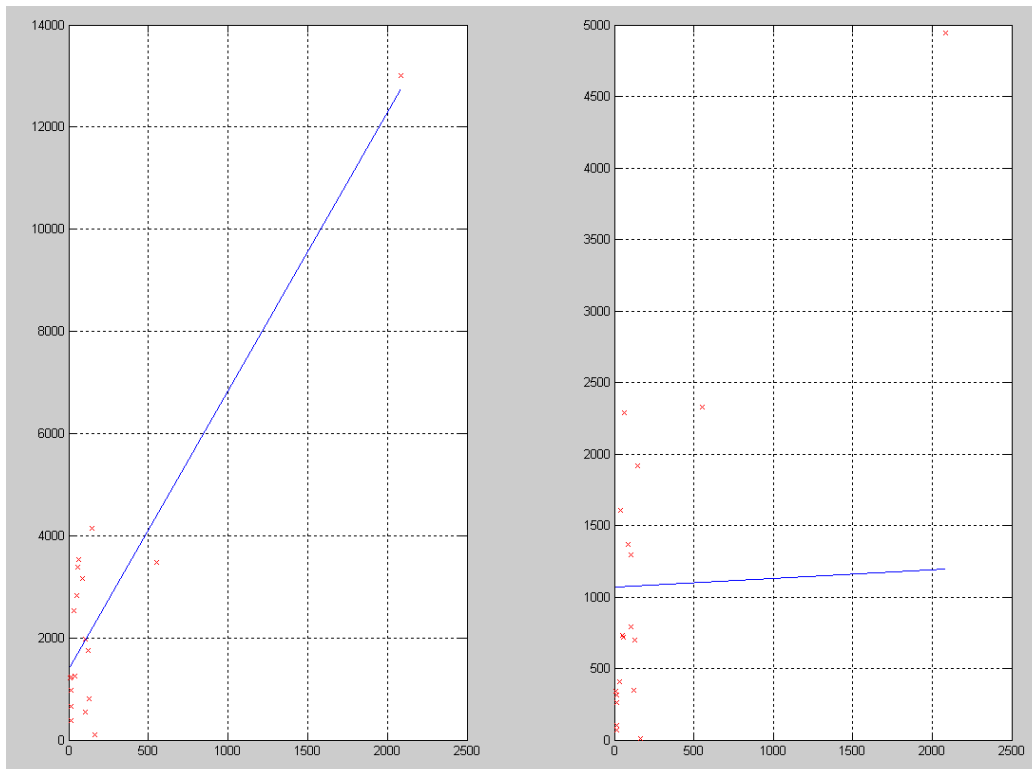
```
>> corrcoef(Kr, N)
ans =
    1.000000000000000e+000    8.454416264396026e-001
    8.454416264396026e-001    1.000000000000000e+000
```

Je tu určitá lineárna závislosť, a to zvlášť pri kategórii M1. Nakreslíme regresné priamky:

```
>> A=[Kr',M',N'];
>> m=mean(A)
m = 2.001052631578947e+002    2.472210526315790e+003    1.081578947368421e+003
```

```
>> c=cov(A)
c = 1.0e+006 *
    0.2234    1.2174    0.4753
    1.2174    8.0509    3.0423
    0.4753    3.0423    1.4142
```

```
>> k=c(1,2)/c(1,1); subplot(1,2,1) % rozdelí obrázok na dva grafy (1 riadok, 2 stĺpce) a ide kresliť do prvého
>> plot(A(:,1),A(:,2),'xr'), hold on, grid on, xm=[min(Kr), max(Kr)]; plot(xm, m(2)+k*(xm-m(1)))
>> k=c(1,3)/c(2,2); subplot(1,2,2) % potvrdí delenie obrázku na dva grafy a ide kresliť do druhého
>> plot(A(:,1),A(:,3),'xr'), hold on, grid on, xm=[min(Kr), max(Kr)]; plot(xm, m(3)+k*(xm-m(1)))
```



Na obrázku vidíme v prípade M1 regresnú priamku položenú medzi zistené hodnoty – korelačný koeficient 0.9 už znamená slabšiu väzbu a vidno to aj na obrázku. Koeficient 0.84 je ešte menší a druhý obrázok s priamkou už nie veľmi presvedčí. Tu štatistika vidí viac než ľudské oko.

2. Aký je súvis medzi relatívnou kradnutosťou a predávanosťou značiek?

```
>> corrcoeff(Kr./M, M)
ans =
    1.000000000000000e+000  -1.488929024302320e-001
   -1.488929024302320e-001  1.000000000000000e+000
```

```
>> corrcoeff(Kr./N, N)
ans =
    1.000000000000000e+000  -2.035367726551459e-001
   -2.035367726551459e-001  1.000000000000000e+000
```

Odpovedzte sami na základe získaných hodnôt. Má význam kresliť regresnú priamku?

Úloha 1: Nájdite korelačný koeficient pre štyri dvojice vektorov: M+N a Kr, M-N a Kr, M a N+Kr, M a N. Poprípade si zvolte vy sami také štyri dvojice, ktoré považujete za zaujímavé pre porovnanie.

Nakreslite príslušné regresné priamky do štyroch okien v jednom obrázku – využite príkaz subplot(2,2,i) pre i=1 až 4, ktorý treba uviesť vždy pred príkazom plot.

Úloha 2: Realizujte riešenie úlohy 1 cyklom (tj. nie na štyrikrát), poprípade ju formulujte ako m-file.