

Spracovanie dátového súboru

Naším cieľom je zistiť, aké hodnoty sú prítomné vo vektore dát v , usporiadať ich podľa veľkosti, zistiť ich početnosti a kumulatívne početnosti. Je viacero spôsobov, ako sa to dá urobiť, preto môžete uvažovať aj o iných postupoch s využitím iných príkazov. Pre názornosť budeme postup ilustrovať na príklade a nechávame všetky výstupy vypisovať. Ak budete tento postup (po jeho pochopení) používať pri väčšom súbore dát, nezabudnite doplniť bodkočiarky tam, kde treba.

Vstupné dáta:

```
>> v=round(rand(1,23)*40-20)
v = 18 -11 4 -1 16 10 -2 -19 13 -2 5 12 17 10 -13 -4 17 17 -4 16 -18 -6 13
```

Usporiadanie:

```
>> x=sort(v), n=length(x)
x = -19 -18 -13 -11 -6 -4 -4 -2 -2 -1 4 5 10 10 12 13 13 16 16 17 17 17 18
n = 23
```

Podľa toho, ako sme vyrábali vektor x , je zrejmé, že jeho zložky nadobúdajú hodnoty medzi -20 a 20 a to iba celočíselné. Vektor prípustných hodnôt je teda

```
>> phx= -20:20;
```

Početnosť jednotlivých hodnôt zistíme pomocou príkazu `hist` alebo `histc`. (Pomocou príkazov `help hist` alebo `help histc` zistíte, čo presne uvedené príkazy robia a prečo je výsledok rovnaký.)

```
>> pc=hist(x,phx)
{ >> pc=histc(x,phx) }
```

```
pc = 0 1 0 0 1 1 2 0 0 0 0 1 0 0 0 0 0 1 0 1 0 0 1 0 0 1 1 0 0 0 0 0 2 1 1 0 1 2 3 2 0
```

Vyberieme iba tie pozície, kde je početnosť nenulová:

```
>> nenul=find(pc)
```

```
nenul = 2 5 6 7 12 18 20 23 26 27 33 34 35 37 38 39 40
```

Nenulovej početnosti sa tešia tieto čísla:

```
>> pnn=phx(nenul)
```

Nakoniec výsledok spracovania vypíšeme – čísla s nenulovou početnosťou, ich početnosť a kumulatívnu početnosť:

```
>> pcn=pc(nenul); pcc=cumsum(pcn);  
>> [pnn', pcn', pcc']
```

```
ans =
```

```
-19  1  1  
-16  1  2  
-15  1  3  
-14  2  5  
-9   1  6  
-3   1  7  
-1   1  8  
 2   1  9  
 5   1 10  
 6   1 11  
12   2 13  
13   1 14  
14   1 15  
16   1 16  
17   2 18  
18   3 21  
19   2 23
```

Stručnejší postup

Majme vektor celočíselných dát **v**:

```
>> v=round(rand(1,25)*20);
```

Miesto teoretickej špekulácie si necháme od ML zistiť, v akom rozmedzí sa pohybujú hodnoty zložiek vektora **v** a aké sú početnosti jednotlivých hodnôt:

```
>> [f y]=hist(v, min(v):max(x));
```

Výpis:

```
>> [y',f']
```

Už sme si zvykli, že čísla s nulovou početnosťou by nemali zbytočne zavádzať vo výhlade, preto ich odstránime:

```
>> nn=find(f); yn=y(nn); fn=f(nn); cf=cumsum(fn);
```

Výpis:

```
>> [yn',fn',cf']
```

Úloha:

- a) Urobte analýzu početností a relatívnych početností pre súbor

>> $v = \text{round}(\text{rand}(1,999) * K - L)$

kde K / L je ciferný súčet písmen¹ Vášho priezviska / mena (bez mäkčeňov a dĺžňov).

- b) Nájdite prehľadnejšie riešenie úlohy použitím vhodného **triedenia** (početnosti sa budú zisťovať pre triedy).

¹ A=1, B=2, C=3, ... ; napr. „Laco“ dáva hodnotu $L=12+1+3+15=31$.

