Výpočet kvantilov

Počítanie kvantilov je pri menšom počte dát závislé na tom, ako presne kvantil definujeme. Pre nás budú zaujímavé dva prístupy. Jeden vychádza z PČLEDF a hovorí sa o ňom na inom mieste. Druhý je založený na nasledujúcej definícii:

*Nech n je počet dát vo vektore x. Číslo q_p pre $p \in [0,1]$ nazveme p-kvantilom, ak platí, že aspoň p*n dát je menších alebo rovných ako q_p a aspoň (1-p)*n dát je väčších alebo rovných ako q_p.*

Prakticky to znamená, že ak p*n nie je celé číslo, napríklad 7.25, tak definíciu splníme, ak aspoň 8 údajov bude menších alebo rovných ako q_p a aspoň (n-7) údajov bude väčších alebo rovných ako q_p. Najväčší z údajov menších alebo rovných ako q_p bude preto zhodný s najmenším z údajov väčších alebo rovných ako q_p a práve tento údaj bude hľadaným kvantilom.
Ak je p*n celé číslo, potom definícii vyhovuje akékoľvek číslo medzi n*p-tou a (n*p+1)-ou hodnotou. Aby sme sa vyhli nejednoznačnosti, v tomto prípade za kvantil považujeme aritmetický priemer týchto hodnôt.

Úloha: Napíšte m-funkciu, ktorá bude počítať rôzne kvantily v súbore dát. Inšpirovať sa môžete prednáškou alebo nižšie uvedeným riešením. Nedajte sa tým však obmedzovať a zostavte si vlastnú verziu.

Príklad. Výpočet podľa danej definície nám umožní funkcia kvantil, pri ktorej pre jednoduchosť žiadame, aby vstupom p boli čísla z otvoreného intervalu (0,1).

→ Skúste do m-funkcie kvantil doplniť testovanie vhodnosti parametra p.

Súbor **kvantil.m**:

```
function q=kvantil(p,v)
x=sort(v);
n=length(x); s=n*p; ta=ceil(s); tb=ceil(n*(1-p));
if ta+tb==n,
   q=0.5*(x(ta)+x(ta+1));
else q=x(ta);
end
```

Príklady:

1.
Voľte si vektory dát rôznych dĺžok a počítajte rôzne kvantily s pomocou funkcie kvantil.m.
Ako možno počítať kvantily v zmysle našej definície aj jednoduchším spôsobom?
(napr. úvahou nad výsledkom súčinu p*n).
Pokiaľ nás zaujíma iba výsledok a nepotrebujeme ho používať v kontexte dlhšieho algoritmu bez vstupu človeka, použitie funkcie kvantil.m nie je potrebné.

2.
Navrhnite vektor dát *a* tak, aby súčasne platili nasledujúce podmienky:

a) length(a) < 50
b) $q_{0.31} = 27$
c) $q_{0.32} = 29$
d) $q_{0.65} = \bar{a}$

Skusmo zistíme, že napr. pri počte 16 sa oba kvantily budú pri zaokrúhľovaní nadol či nahor líšiť:

```
[0.31, 0.32]*16

        ans =      4.9600    5.1200
```

To znamená, že pri 16 prvkoch vektora musí byť $q_{0.31} = 27$ piaty a $q_{0.32} = 29$ šiesty prvok zľava.
Zostavme taký vektor:

```
x=[10 11 15 17 27 29 31 33 34 35 36 37 39 41 45 47 ]

mean(x)
        ans =    30.4375
```

0,65-kvantil je 11-ty prvok zľava (lebo 0.65*16 = 10.4), teda číslo 36. Je vhodné s touto hodnotou nehýbať a skúsiť radšej zvýšiť priemer, a to napr. zmenou posledného prvku vo vektore. Po pár pokusoch vychádza:

```
x=[10 11 15 17 27 29 31 33 34 35 36 37 39 41 45 136 ];  mean(x)

ans =    36
```

3.

Navrhnite čo najkratší vektor dát *a* tak, aby súčasne platili podmienky:

a) $q_{0.61} = 22$
b) $q_{0.62} = 24$
c) $q_{0.25}+q_{0.75} = 1.5*\bar{a}$

Skúsime, či nájdeme vhodný vektor s dĺžkou do 20:

```
a=(1:20)'; b=a*[0.61 0.62];[a,b]

ans =     1.0000    0.6100    0.6200
          2.0000    1.2200    1.2400
          3.0000    1.8300    1.8600
          4.0000    2.4400    2.4800
          5.0000    3.0500    3.1000
          6.0000    3.6600    3.7200
          7.0000    4.2700    4.3400
          8.0000    4.8800    4.9600
          9.0000    5.4900    5.5800
         10.0000    6.1000    6.2000
         11.0000    6.7100    6.8200
         12.0000    7.3200    7.4400
         13.0000    7.9300    8.0600
         14.0000    8.5400    8.6800
         15.0000    9.1500    9.3000
         16.0000    9.7600    9.9200
         17.0000   10.3700   10.5400
         18.0000   10.9800   11.1600
         19.0000   11.5900   11.7800
         20.0000   12.2000   12.4000
```

Pre 13 prvkov to pôjde – v zmysle našej definície kvantilu budú 0,61 a 0,62 kvantily rôzne a budú zodpovedať pozíciám 8 a 9:

x = [? ? ? ? ? ? ? 22 24 ? ? ? ?]

Ďalej upresníme, že

```
[0.25 0.75]*13
     ans =    3.2500    9.7500
```

teda kvartily zodpovedajú pozíciám 4 a 10. Aby sme ostali pri celých číslach (pohodlie nadovšetko), zvolíme kvartily tak, aby ich súčet bol 1,5-násobkom celého čísla.

x = [? ? ? 5 ? ? ? 22 24 25 ? ? ?]

Naša voľba znamená, že $q_{0.25}+q_{0.75} = 5+25 = 30$. Aby sme splnili bod c, priemer hodnôt vo vektore musí byť 20, inými slovami súčet hodnôt vo vektore musí byť 13*20 = 260. Doplníme vhodné čísla na zvyšné pozície, skusmo to pôjde azda najrýchlejšie:

```
x = [5 5 5 5 20 20 20 22 24 25 30 30 49];
mean(x)
     ans =    20
```