Exercice 1 La moyenne des taux et le taux moyen 6 points

L'inflation à été très importante en Argentine. Nous allons supposer qu'il est prévue une évolution de l'inflation donnée par an dans le tableau suivant :

Année	2013	2014	2015	2016	2017
inflation (en %)	$i_1 = 250$	$i_2 = 100$	$i_3 = 200$	$i_4 = 50$	$i_5 = 400$

A une une inflation de i % on associe le taux d'inflation $t = 1 + i/100$ de sorte que le prix en fin d'année puisse être calculé par une simple multiplication.

1. calculez les taux d'inflation de 2013 à 2017, leur moyenne et leur médiane,

Année	2013	2014	2015	2016	2017
inflation (en %)	$t_1 = 3{,}5$	$i_2 = 2$	$i_3 = 4$	$i_4 = 1{,}5$	$i_5 = 3$

   La moyenne et la médaine sont 3

2. calculez le prix fin 2017 d'un article coutant 100 pesos début 2013. En déduire le taux d'inflation moyen sur la période,

$$p = 100 * 3,5 * 2 * 4 * 1,5 * 3 = 15750$$

   le taux moyen est donné par cc tel que

$$100 * c * c * c * c * c = 15750$$

   d'où $c = 2,75$

3. montrez que la moyenne géométrique

$$c = \sqrt[n]{\prod_{i=1}^{n} t_i}$$

   est le résultat de la minimisation suivante

$$\min_c \sum_{i=1}^{n} (\log t_i - \log c)^2$$

$$\frac{d \sum_{i=1}^{n} (\log t_i - \log c)^2}{dc} = \frac{-2}{c} \sum_{i=1}^{n} (\log t_i - \log c) = \frac{-2}{c} \left(\log \prod_{i=1}^{n} t_i - n \log c \right)$$

   le minimum de la fonction est atteint lorsque la dérivée s'annule soit si $c \neq 0$ on a

$$\log \prod_{i=1}^{n} t_i - n \log c = 0 \quad \Leftrightarrow \quad \prod_{i=1}^{n} t_i = c^n$$

   d'où le résultat

4. calculez la moyenne géométrique des taux d'inflation.

$$c = \sqrt[n]{3,5 * 2 * 4 * 1,5 * 3} = 2,75$$

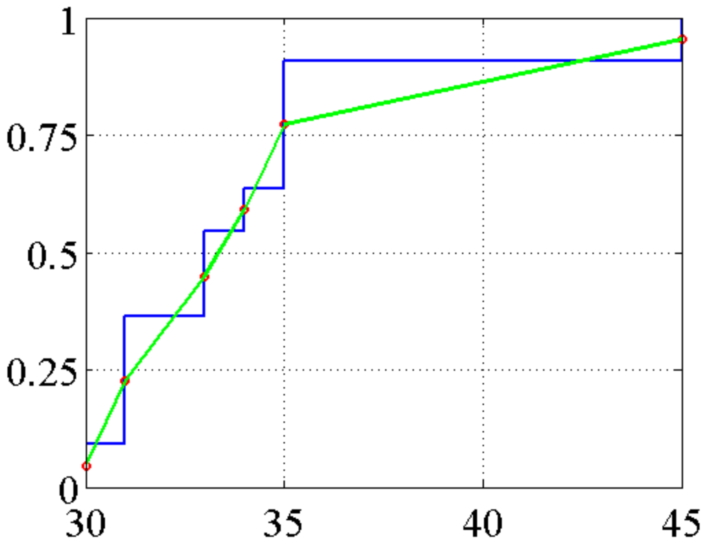
   Comparez la à la moyenne des taux et au taux moyen et commentez.
   la moyenne est inférieure au taux moyen. Attention a ne pas ce faire avoir en utilisant une moyenne qui n'a pas de signification.

Exercice 2 Affections respiratoires 8 points

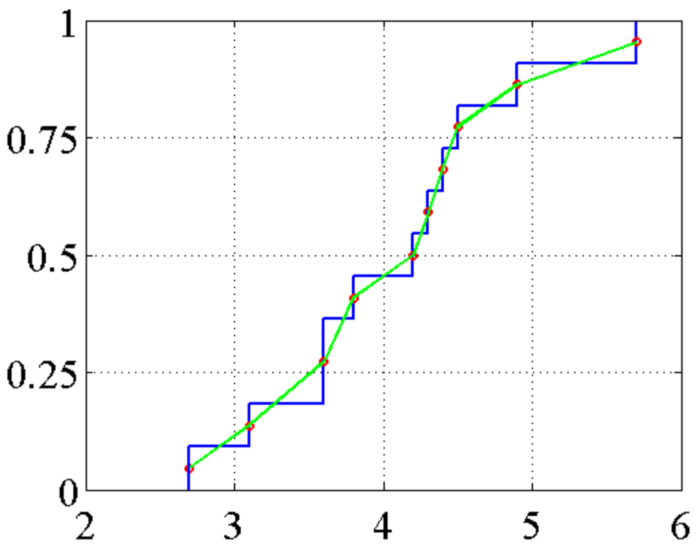
L'une des mesures qui sont faites lors de l'investigation des affections respiratoires est celle du volume expiratoire moyen par seconde, appelé Vems. Sur 11 sujets tirés au sort parmi la population saine d'âge compris entre 30 et 45 ans, on a mesuré la taille, T en mètres et le Vems V en litres par seconde, et obtenu les résultats suivants :

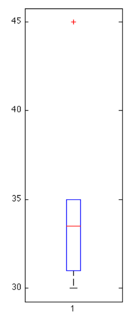
Sujet	1	2	3	4	5	6	7	8	9	10	11	$\sum T_i$ =18,94
Age	30	31	31	31	33	33	34	35	35	35	45	$\sum V_i$ =44,8
T	1,85	1,72	1,51	1,62	1,60	1,80	1,75	1,68	1,81	1,72	1,88	$\sum T_i^2$ =32,74
V	4,5	3,6	2,7	3,1	3,6	4,4	4,3	3,8	4,2	4,9	5,7	$\sum V_i^2$ =189,5
												$\sum T_i V_i = 77,95$

1. Tracer la fonction de répartition empirique des ages,



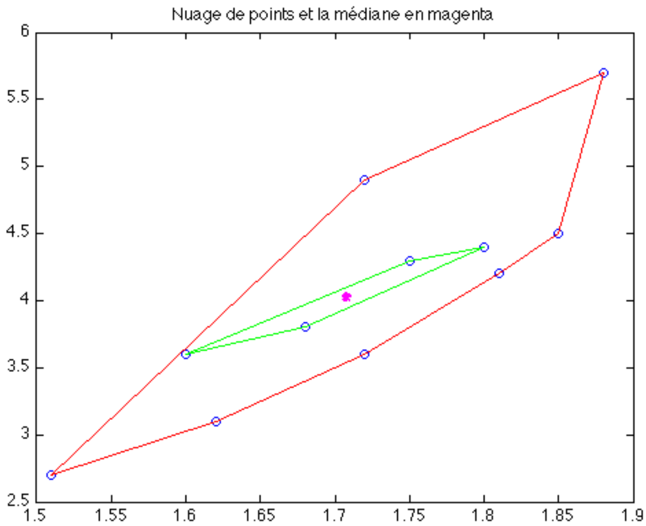
2. Tracer la fonction de répartition empirique du Vems,





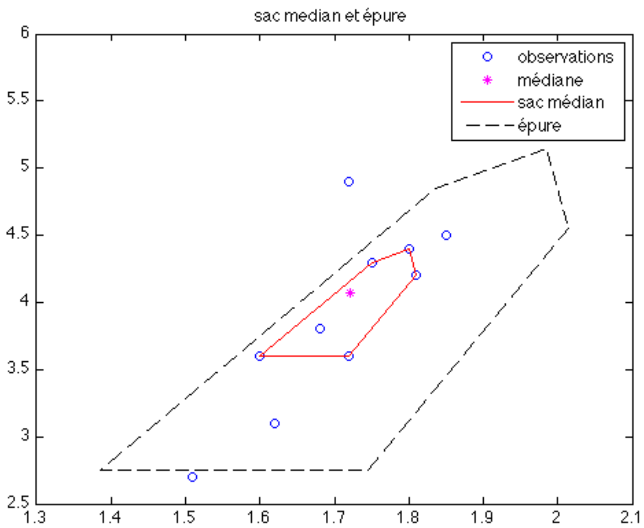
3. Dessinez la boite à moustache des ages,

4. Quelle est la médiane du couple de variables T et V ?



$M = (1.7075,\ 4.0250)$

5. Calculez la correlation entre les variables T et V 0.8570
6. Y-a t'il à votre avis des points aberrants dans cet échantillon pour le couple de variables T et V ?



les deux points hors épure sont suspect : les points 10 (1.72,4.9) et 11 (1.88,5.7)

Exercice 3 — Anna-Lise en CP — 6 points

À la suite de l'énoncé, vous trouverez la copie d'une session Matlab. On y trouve des lignes de code qui commencent par le chargement d'un tableau de données V comportant $n = 24$ points de mesures différents de $p = 5$ variables concernant la qualité de l'eau (dans l'ordre) :

1. conductivité électrique (C),
2. pluviométrie (Pℓ),
3. concentration en sulfate (SO_4)
4. pH
5. concentration en aluminium (Aℓ)

Rappel : la fonction matlab `[V,lam] = eig(M)` calcule les vecteurs propres (V) et les valeurs propres associées (sur la diagonale de `lam`) de la matrice M.

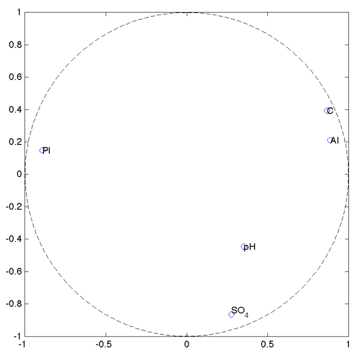
1. quelle(s) est/sont, selon vous, l'(es) étape(s) indispensable(s) à appliquer avant d'effectuer une ACP ?

   Centrer et réduire les données et éliminer les points aberrants
2. quelle relation permet de représenter les données par rapport à une composante principale ?

   $X * U$
3. caractérisez (en pourcentage) la qualité de la représentation du nuage de points par rapport à la première composante de l'ACP ?

   $\frac{\lambda_1}{\sum_j \lambda_j} = \frac{62.60}{1.43+3.54+12.61+34.79+62.6} = 0,54$ soit 54% de l'information est portée par le premier axe
4. quelle sont les moyennes et variances des variables de X ?

   0 et 1
5. représentez les variables dans le plan des deux premières composantes principales. Quelle interprétation (relations entre les variables et les composantes principales) donnez vous à cette représentation ?

   L'axe 1 oppose les variables C et Al d'un coté à la variable Pl. C'est une variable qui traduit une certaine forme de « clarté » de l'eau L'axe 2 appose lui les fort pH et SO_4 aux faibles valeurs. C'est une variable liée au pH en général
6. écrire en une ligne de code une instruction permettant de visualiser les données par rapport aux deux premiers axes de l'ACP

   ```
   plot(X*U(:,5),X*U(:,4),'o')
   ```

```
load V

V =
    1.3836   -1.0930    0.3262    5.4349    2.1660
    1.7532   -1.3385    0.5524    7.2205    2.7734
    1.1362   -1.1019    0.5590    7.3748    2.6056
    1.4546   -0.7685    0.0804    6.3739    2.6261
    1.8885   -1.1024    0.0441    7.3420    2.5703
    1.1729   -0.6155   -0.1404    5.0500    1.6798
    1.9062   -1.4222    0.5256    5.4026    2.7753
    1.4481   -1.2618    0.6298    7.3568    2.0557
    1.4978   -1.1238    0.3180    5.8705    2.6134
    1.3020   -0.7546    0.2977    6.6751    2.3673
    0.5239   -0.4024    0.2212    5.7350    1.5513
    1.4333   -0.9875    0.6963    7.6980    2.3885
    1.5134   -1.3074    0.4735    4.4032    2.4529
    0.6534   -0.4714    0.2450    6.2207    1.5849
    1.2102   -0.7971    0.2916    7.3181    1.6703
    1.4984   -1.1181    0.1538    6.1666    2.2949
    1.6517   -0.8363   -0.1578    6.0425    2.2022
    1.3006   -1.0394    0.4082    4.9735    2.0400
    1.0875   -0.6797    0.0983    4.0810    1.6893
    0.6563   -0.7660    0.7251    4.8285    1.3046
    1.7318   -0.9731    0.0936    5.3171    2.6138
    1.2850   -0.9363    0.3896    5.0632    2.1440
    1.3300   -0.8850   -0.0242    6.2067    2.0917
    1.1681   -1.0494    0.4480    8.1255    1.9529

[n,p] = size(V);
X = (V - ones(n,1)*mean(V))./(ones(n,1)*std(V));

[U,d]=eig(X'*X)
U = 0.6838    0.3193   -0.0233    0.3640    0.5454
    0.5876   -0.4810   -0.3032    0.1366   -0.5593
    0.3891   -0.1915    0.3759   -0.8001    0.1744
   -0.0440    0.1218   -0.8739   -0.4127    0.2219
   -0.1838   -0.7843   -0.0511    0.1957    0.5569

diag(d) =  1.43    3.54    12.61    34.79    62.60
```