

1 Exercice 1

1. calculez la moyenne et la mdiane des vitesses,

```
p = [1 2 3 4 5];  
v = [10 5 3.33 2.5 2];
```

```
mean(v)  
% 4.566  
median(v)  
% 3.33
```

2. calculez le temps de parcours. En dduire la vitesse moyenne,

```
t = sum(1./v)  
% 1.5 soit 1h et demi  
vmoy = 5/t  
% 3.33 km/h
```

3. montrez que la moyenne harmonique est le rsultat de la minimisation suivante

```
% J(c) = sum(1/v_i - 1/c)^2  
% J'(c) = 2 * 1/c^2 * sum(1/v_i - 1/c)  
% J'(c*) = 0 <=> sum(1/v_i - 1/c*) = 0  
%           <=> sum(1/v_i) - n/c* = 0  
%           <=> sum(1/v_i) = n/c*  
%           <=> c* = n/sum(1/v_i) = 1/moy(1/v_i)
```

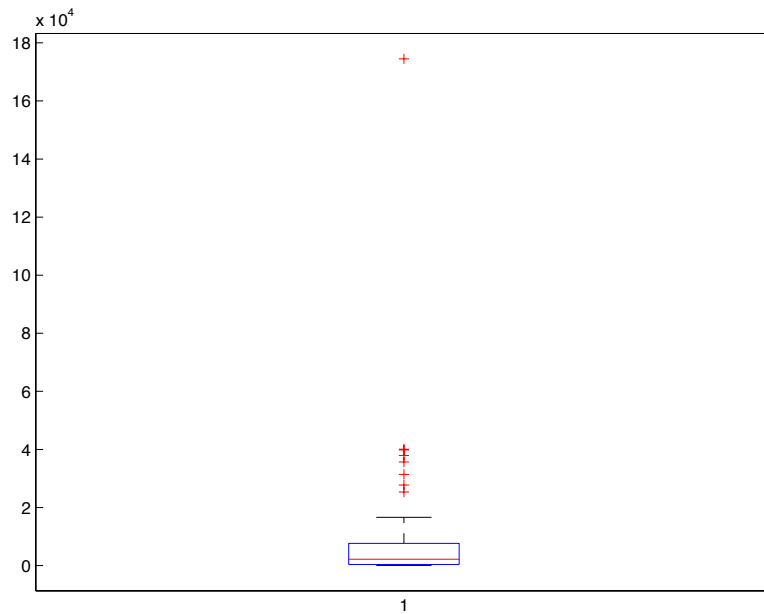
```
mh = 1/mean(1./v);  
% 3.33  
% c'est la mme que la vitesse moyenne (et diffrente de la mouyenne des  
% vitesses)  
% Ainsi la vitesse moyenne est a moyenne harmonique des vistesses
```

2 Exercice 2

```
archive.ics.uci.edu/ml/machine-learning-databases/00192/  
BreastTissue.xls
```

1. Tracez sur la feuille ci-joint la boite moustache de la vriable V5:AREA area under spectrum.

```
figure(1)  
boxplot(X(:,5))
```



2. Trouvez la mdiane de Tuckey du nuage de points donn sur la feuille ci-joint

figure(2)

```
plot(X(1:37,1),X(1:37,3),'d','LineWidth',2)
set(gcf,'Color',[1,1,1])
set(gca,'FontSize',14,'FontName','Arial','XTick',[],'YTick',[]);
hold on
```

```
x = X(1:37,1);
y = X(1:37,3);
```

```
c1 = convhull(x,y);
plot(x(c1,1),y(c1),'r');
```

```
x(c1) = [];
y(c1) = [];
```

```
c1 = convhull(x,y);
plot(x(c1,1),y(c1),'b');
```

```
x(c1) = [];
y(c1) = [];
```

```
c1 = convhull(x,y);
plot(x(c1,1),y(c1),'m');
```

```

x(c1) = [];
y(c1) = [];

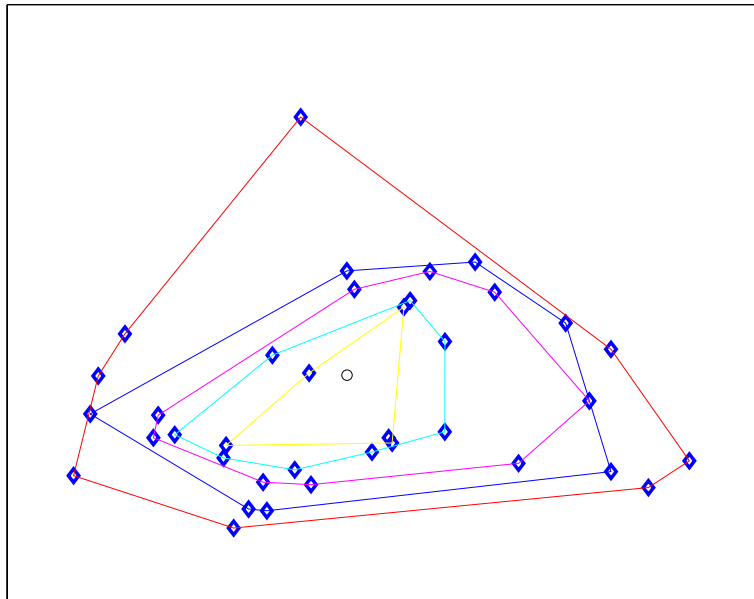
c1 = convhull(x,y);
plot(x(c1,1),y(c1),'c');

x(c1) = [];
y(c1) = [];

c1 = convhull(x,y);
plot(x(c1,1),y(c1),'y');

plot(mean(x(c1,1)),mean(y(c1)),'ok');
hold off

```



3. Quel pr traitement est-il recommand de faire avant l'ACP et pourquoi ?
il faut centrer et rdre les donnes pour que les variables soient comparables

```

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

```

4. calculez la matrice des corrlations

```

C = Xn'*Xn;
C/n

```

5. calculez ses vecteurs et valeurs propres

```
[V L] = eig(C);
vp = diag(L);
```

6. visualisez les valeurs propres et leur importance

```
100*(vp(end:-1:1))/sum(vp)
vpc = 100*cumsum(vp(end:-1:1))/sum(vp)
```

```
figure(3)
bar(vpc,.5); hold on;
plot(100*vp(end:-1:1)/vp(end),'r-o');
title('Valeurs propres')
hold off
```

Le premier plan factoriel (les deux premiers axes) représentent 70% de l'information contenue dans le tableau X

7. Calculez les composantes principales

```
V = V(:,end:-1:1);
L = L(end:-1:1,end:-1:1);
U = Xn*V;
```

8. Visualisez tous les individus sur les deux premières composantes principales

```
figure(4)
plot(U(:,1),U(:,2),'o');

xlabel('axe 1');
ylabel('axe 2');
title('Représentation du nuage avec deux axes de l''ACP')
```

9. Comment interpréter les axes en terme de variables ?

```
% L'axe 1 oppose les individus aux fortes valeurs globales sur les variables
% 1, 4, 5, 6, 7, 8 et 9
% L'axe 2 oppose les individus avec des valeurs fortes sur les variables 2
% et 3 bien représentés par cet axe
```

```
Vn = V*sqrt(L)/sqrt(n);
figure(5)
subplot(2,2,1)
title('Représentation des variables avec deux axes de l''ACP')
plot(Vn(:,1),Vn(:,2),'*');hold on
for i=1:p
    text(Vn(i,1),Vn(i,2),['v' num2str(i)]);
end
a = 0:0.01:2*pi;
plot(cos(a),sin(a),':')hold off;
```

```

xlabel('axe 1');
ylabel('axe 2');
title('Representation des variables avec deux axes de l''ACP')

subplot(2,2,2)
plot(Vn(:,1),Vn(:,3),'*');hold on
for i=1:p
    text(Vn(i,1),Vn(i,3),['v' num2str(i)]);
end
a = 0:0.01:2*pi;
plot(cos(a),sin(a),'');
xlabel('axe 1');
ylabel('axe 3');
title('Representation des variables avec deux axes de l''ACP')
hold off;
subplot(2,2,3)
plot(Vn(:,2),Vn(:,3),'*');hold on
for i=1:p
    text(Vn(i,2),Vn(i,3),['v' num2str(i)]);
end
a = 0:0.01:2*pi;
plot(cos(a),sin(a),'');
xlabel('axe 2');
ylabel('axe 3');
title('Representation des variables avec deux axes de l''ACP')
hold off;

subplot(2,2,4)
plot(Vn(:,4),Vn(:,5),'*');hold on
for i=1:p
    text(Vn(i,4),Vn(i,5),['v' num2str(i)]);
end
a = 0:0.01:2*pi;
plot(cos(a),sin(a),'');
xlabel('axe 4');
ylabel('axe 5');
title('Representation des variables avec deux axes de l''ACP')
hold off;

```

ans =

Columns 1 through 7

0.9906	-0.3899	0.0282	0.8119	0.5548	0.6063	0.8159
-0.3899	0.9906	0.5042	-0.0890	0.0828	0.2277	-0.0499

0.0282	0.5042	0.9906	0.1060	0.2041	0.3527	0.3673
0.8119	-0.0890	0.1060	0.9906	0.7242	0.6422	0.7461
0.5548	0.0828	0.2041	0.7242	0.9906	0.8223	0.7283
0.6063	0.2277	0.3527	0.6422	0.8223	0.9906	0.8051
0.8159	-0.0499	0.3673	0.7461	0.7283	0.8051	0.9906
0.7263	-0.0763	0.0115	0.9650	0.6694	0.5356	0.5946
0.9794	-0.3425	0.1014	0.7667	0.5687	0.6730	0.8537

Columns 8 through 9

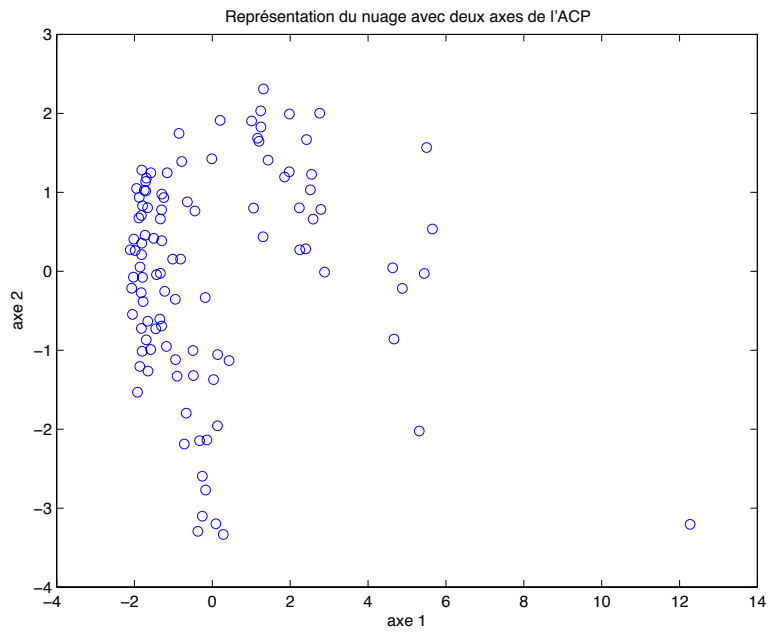
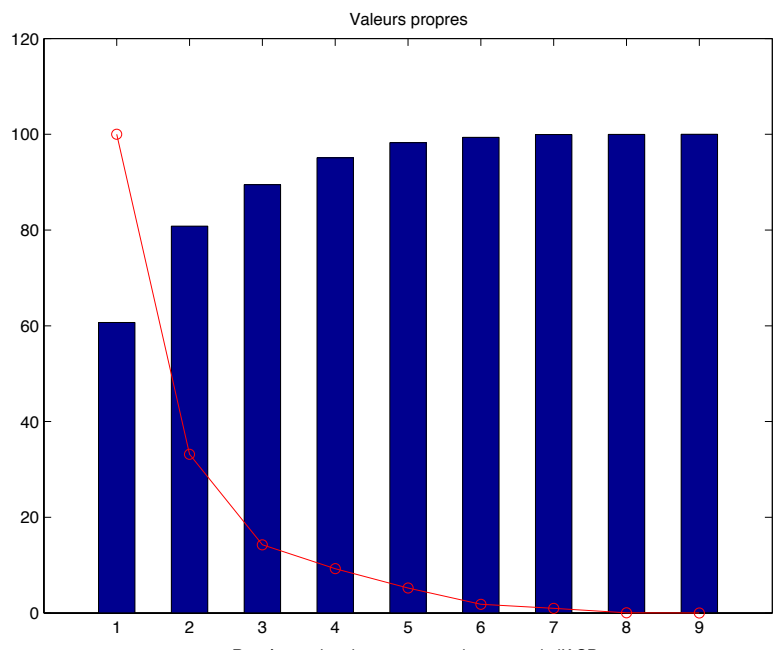
0.7263	0.9794
-0.0763	-0.3425
0.0115	0.1014
0.9650	0.7667
0.6694	0.5687
0.5356	0.6730
0.5946	0.8537
0.9906	0.6597
0.6597	0.9906

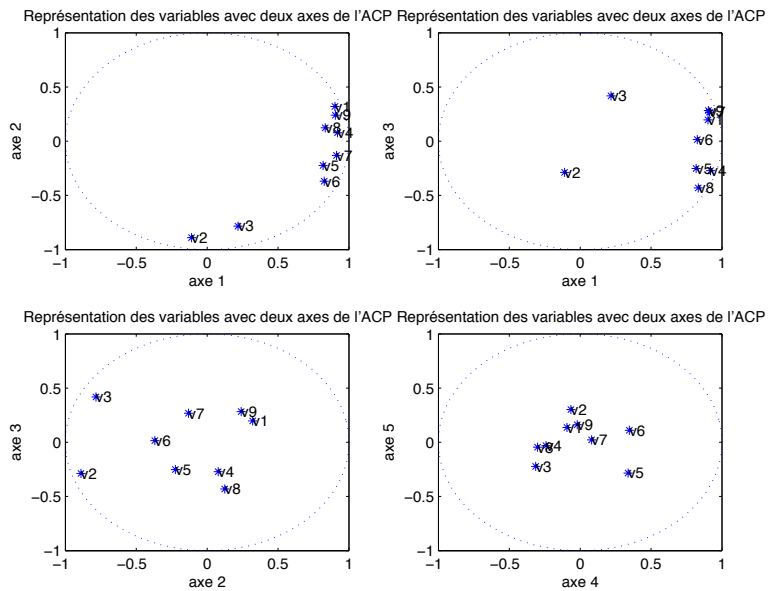
ans =

60.7015
20.1226
8.6524
5.6217
3.1635
1.1042
0.5900
0.0342
0.0099

vpc =

60.7015
80.8241
89.4765
95.0982
98.2616
99.3658
99.9559
99.9901
100.0000





10. Il se trouve que les 21 premiers individus appartiennent à un même groupe (en fait ils ont la même maladie). En utilisant l'ACP, comment décrivez-vous ce groupe ?

Ce groupe se projette dans la même zone du premier plan factoriel. Ces individus sont donc homogènes dans ce plan. Ce groupe se caractérise par une faible valeur sur l'axe 2 donc forte variables 2 et 3 et une valeur entre -1 et 1 sur l'axe 1. Il est donc possible, à partir de ce premier plan factoriel de prédire avec une bonne confiance si un individu appartient ou non à ce groupe (donc si il a la maladie en question)

figure(4)

```
% le premier groupe
for i=1:21
    text(U(i,1),U(i,2),num2str(i));
end

% le dernier groupe aussi est homogène
for i=86:n
    text(U(i,1),U(i,2),num2str(i));
end
```