Td-TP Ch 1 : Les outils de représentation d'un échantillon

I. EXERCICES « A LA MAIN »

Les exercices ou parties avec une étoile (*) sont à traiter en priorité.

1. Eléments de Statistiques

Exercice 1 *
On considère le caractère X étudié sur 10 individus. Calculer la moyenne, la médiane, l'écart-type de X dans les cas suivants :

a)

n°	1	2	3	4	5	6	7	8	9	10
X	4	2	1	1	3	4	5	2	1	3

b)

n°	1	2	3	4	5	6	7	8	9	10
X	1	4	2	3	2	2	4	2	3	3

c)

n°	1	2	3	4	5	6	7	8	9	10
X	5	0	5	1	4	2	3	5	5	5

Exercice 2 *
Soit X un caractère étudié sur 10 individus et *a* et *b* deux réels.

n°	1	2	3	4	5	6	7	8	9	10
X	*a*	1	2	0	3	1	*b*	2	1	0

a) Calculer les valeurs de *a* et *b* telles que la moyenne du caractère X soit 2 et sa variance soit 4.
b) Faire le même calcul lorsque la moyenne et la variance sont égales à 2.

2. Calcul matriciel

Exercice 3 *
Soit les matrices :

$$\mathbf{A}=\begin{pmatrix}3&1&2\\1&3&2\\2&3&1\\3&2&1\end{pmatrix}\quad;\quad \mathbf{B}=\begin{pmatrix}1&0&1\\0&1&1\\1&1&0\end{pmatrix}\quad;\quad \mathbf{C}=\begin{pmatrix}-1&1&0\\-1&-1&0\\1&-1&1\end{pmatrix}\quad;$$

$$\mathbf{U}=\begin{pmatrix}1\\-1\\1\end{pmatrix}\quad;\quad \mathbf{V}=\begin{pmatrix}1\\2\\-2\\-1\end{pmatrix}.$$

a) La somme A+B+C a –t-elle un sens ?
b) Les produits AB, AC, BA ont-ils une sens ? Si oui, les calculer.
c) Même question pour U^TB et V^TA.
d) Même question pour V^TAU et U^TBU.

M1 Ingénierie Statistique, Dpt de Mathématiques, Université de Nantes : Analyse de données (L. Bellanger)

Exercice 4 *

Soit la matrice : $\mathbf{X}=\begin{pmatrix}1&0\\0&1\end{pmatrix}$.

a) Ecrire la matrice $\overline{\mathbf{X}}=\begin{pmatrix}\bar v_1&\bar v_2\\\bar v_1&\bar v_2\end{pmatrix}$ où $\bar v_1$ et $\bar v_2$ sont les moyennes de la première colonne et de la deuxième colonne.
b) Ecrire $\mathbf{X}-\overline{\mathbf{X}}$ (matrice centrée) et la matrice :

$$\mathbf{V}=\frac{1}{2}\left(\mathbf{X}-\overline{\mathbf{X}}\right)^T\left(\mathbf{X}-\overline{\mathbf{X}}\right).$$

c) Trouver les valeurs propres de **V** (notées λ_1,λ_2).
d) Calculer λ_1,λ_2 et trace(**V**).
e) Déterminer des vecteurs propres **U** de **V** normés (ie vérifiant $\mathbf{U}\mathbf{U}^T$=1).
Vérifier que les deux vecteurs propres normés obtenus sont orthogonaux.
f) Vérifier vos résultats sous R.

Exercice 5 *
Calculer les valeurs propres et des vecteurs propres normés des matrices :

$$a)\,\mathbf{M}=\begin{pmatrix}1&2\\4&3\end{pmatrix};b)\,\mathbf{N}=\begin{pmatrix}2&2\\3&1\end{pmatrix};c)\,\mathbf{P}=\begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}.$$

Exercice 6
Soit la matrice **X**, dans la base canonique (e_1,e_2), associée à l'application linéaire de R^2 dans R^2 telle que :

$h=\begin{pmatrix}-1\\1\end{pmatrix}$ a pour image $h'=\begin{pmatrix}1\\-2\end{pmatrix}$ et $l=\begin{pmatrix}3\\2\end{pmatrix}$ a pour image $l'=\begin{pmatrix}-1\\3\end{pmatrix}$.

Les vecteurs associés aux matrices colonnes *h*, *h'*, *l* et *l'* sont notés respectivement *h, h', l* et *l'*
a) Déterminer la matrice **X** (dans la base canonique (e_1,e_2)).
b) On considère la nouvelle base (u_1,u_2) de R^2 définie par :

$$u_1=\frac{1}{2}(e_1+e_2)\text{ et }u_2=\frac{1}{2}(e_1-e_2).$$

Exprimer e_1 et e_2 en fonction de u_1 et u_2.
c) Ecrire les vecteurs *h, h', l* et *l'* dans la base (u_1,u_2).
d) Calculer la matrice **Y** associée à l'application linéaire définie par **X** dans la base (u_1,u_2).

Exercice 7
Soit dans R^2 la droite *d*, de vecteur directeur u de composantes : (1 ;-1) dans la base canonique.
a) Ecrire la matrice **P** de la projection orthogonale de R^2 sur *d*, vérifier que $\mathbf{P}^2=\mathbf{P}$.
b) Calculer les valeurs propres et les vecteurs propres de **P**.

Rappel utile* : *Projection orthogonale sur une droite vectorielle
Si *F* est une droite vectorielle engendrée par le vecteur *a*, l'ensemble des vecteurs orthogonaux à *F* est un hyperplan appelé hyperplan normal à *F* et défini par :

$$F^{\perp} = \{h \in E, (h \cdot a) = 0\}$$

Si *x* est un vecteur arbitraire de *E*, on peut toujours le décomposer de la façon suivante :

$$x = x_F + x_{\perp} \text{ avec } x_F = \frac{(a \cdot x)}{\|a\|^2} a$$

Et on constate que x_F est dans *F*, tandis que $x_{\perp} = x - x_F$ est dans l'hyperplan normal à *F*. Il est donc toujours possible d'effectuer une projection orthogonale sur une droite vectorielle.

Exercice 8 *
Soit les vecteurs *u, v, w* et leurs vecteurs colonnes dans la base canonique de R^3 :

$$\mathbf{U} = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}; \mathbf{V} = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} \text{et } \mathbf{W} = \begin{pmatrix} 0 \\ -6 \\ 6 \end{pmatrix}.$$

a) **U**, **V**, **W** sont-ils indépendants linéairement ?
b) Même question pour **U** et **V**.
c) Même question pour **V** *et* **W**, **U** et **W**.
d) Soit x, y et z 3 vecteurs linéairement indépendants 2 à 2, peut-on conclure qu'ils sont linéairement indépendants dans leur ensemble ?

Exercice 9 *
a) Calculer les normes (pour la métrique du produit scalaire classique) et les cosinus (pris 2 à 2) des vecteurs de R^4 :

$$X_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}; X_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}; X_3 = \begin{pmatrix} 0 \\ 3 \\ 2 \\ 1 \end{pmatrix}; X_4 = \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \end{pmatrix}.$$

b) Déterminer le réel λ tel que les vecteurs suivants soient normés :

$$U_1 = \begin{pmatrix} 3\lambda \\ \lambda \\ 2\lambda \end{pmatrix}; U_2 = \begin{pmatrix} \lambda \\ -\lambda \\ \lambda \\ -\lambda \end{pmatrix}; U_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \lambda \end{pmatrix}.$$

c) On se place maintenant dans R^5. Soient les vecteurs :

$$X = \begin{pmatrix} \alpha \\ 1 \\ \beta \\ 0 \\ \gamma \end{pmatrix}; A = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}; B = \begin{pmatrix} -1 \\ 1 \\ 2 \\ -2 \\ 0 \end{pmatrix}; C = \begin{pmatrix} 0 \\ -2 \\ 1 \\ 1 \\ 0 \end{pmatrix}.$$

Déterminer les réels α, β et γ pour que X soit orthogonal à chacun des vecteurs A, B et C.

M1 Ingénierie Statistique, Dpt de Mathématiques, Université de Nantes : Analyse de données (L. Bellanger)

Exercice 10 *
On considère la matrice

$$\mathbf{A} = \begin{bmatrix} 7 & 4 & 0 & 0 \\ -12 & -7 & 0 & 0 \\ 20 & 11 & -6 & -12 \\ -12 & -6 & 6 & 11 \end{bmatrix}$$

a) Calculer les valeurs propres de **A.**
b) Déterminer une base formée de vecteurs propres de A
c) **A** est-elle diagonalisable ? Inversible ?

Exercice 11
Montrer en utilisant le théorème de Cayley-Hamilton que la matrice

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 0 \\ 2 & 0 & -1 \end{bmatrix}$$

est inversible et calculer son inverse.
Exercice 12 * sous R
Soit les matrices

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \end{bmatrix} \text{ et } \mathbf{B} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

Calculer la DVS maigre de A et B (les valeurs singulières étant classées par ordre décroissant). Puis la DVS pleine.

II. ENVIRONNEMENT R ET 1ERS PAS EN ANALYSE FACTORIELLE

0. Introduction à l'environnement R

Voir document `AD-TdTP0-Stat-IntroGaleR-PPT.pdf` (p 1- 75) et `AS-02-Outils-LB&RT.pdf`.

1. Calcul matriciel

Voir document `AD-TdTP0-Stat-IntroGaleR-PPT.pdf` (p 76 à la fin) et ce qui suit !

Matrices

```
> x1=c(5,2,4,8,9,10,3,2,5,6,1,9) # x1 est un vecteur
> x2=c(1,2,8,5,6,4,9,12,1,16,7,8) # x2 est un vecteur
> m1=matrix(data=x1, nrow=3, ncol=4)
> m2=matrix(data=x2, nrow=3, ncol=4)
> m3=matrix(data=x2, nrow=4, ncol=3)
```

$$m_1=\begin{bmatrix}5&8&3&6\\2&9&2&1\\4&10&5&9\end{bmatrix},\ m_2=\begin{bmatrix}1&5&9&16\\2&9&12&7\\8&4&1&8\end{bmatrix}\text{ et } m_2=\begin{bmatrix}1&6&1\\2&4&16\\8&9&7\\5&12&8\end{bmatrix}$$

Pour vérifier le format de `m1`

```
> is.matrix(m1)
```

Autres manières de créer des matrices :

```
> cbind( 1,3:1,1:3)
> rbind(-1,c(1:2,0),0:5)
```

Dimension d'une matrice

```
> dim(m1)
```

Sélection d'une valeur, d'une colonne ou d'une ligne dans une matrice

```
> m1[2,3]
> m1[ ,3]
> m1[2, ]
```

Supprimer une ligne ou une colonne

```
> m1[-2, ]
> m1[, -3]
```

Assembler des matrices

```
> rbind(m1,m2)
> cbind(m1,m2)
```

Addition, soustraction

```
> m1+m2
> m1-m2
```

Multiplication par un scalaire (k)

```
> k * m1
```

Multiplication de deux matrices : « `%*%` » => produit matriciel

```
> m1 %*% m2
> m1 %*% m3
```

Matrice transposée

```
> t(m1)
```

Matrice diagonale

Créer une matrice diagonale (n,n) : `diag(x= , nrow=, ncol= )`

```
> v=c(10, 20, 30) ; diag(v)
```

Extraire la diagonale d'une matrice : `Créer la matrice carrée` $\mathbf{A}=\begin{bmatrix}2&3\\1&-1\end{bmatrix}$

```
> diag(A)
```

Déterminant

```
> D=det(A)
```


Inversion de matrice

Déterminer manuellement la matrice B inverse de A

Solution dans R :

```
> B=solve(A) ; B
```

On vérifie que AB=I (matrice identité)

```
> A%*%B
```

Système d'équations linéaires

`Créer la matrice` $\mathbf{Y}=\begin{bmatrix}9\\2\end{bmatrix}$. On cherche $\mathbf{X}=\begin{bmatrix}x_1\\x_2\end{bmatrix}$ qui vérifie **Y=AX**

Ecrire cette relation sous la forme d'un système d'équations linéaires

Déterminer manuellement **X** :

1) en résolvant le système d'équation
2) en utilisant le calcul matriciel

Solution dans R :

```
> solve(A,Y)
```

Valeurs propres et vecteurs propres

`Créer la matrice` $\mathbf{M}=\begin{bmatrix}1&3\\4&2\end{bmatrix}$

Calculer manuellement les valeurs propres et vecteurs propres de **M**

Solution dans R :

```
> eigen(M)
> eigen(cbind( 1,3:1,1:3))
> eigen(cbind(-1,2:1)) # valeurs complexes
```

Diagonalisation d'une matrice carrée

$\mathbf{M}=\begin{bmatrix}1&3\\4&2\end{bmatrix}$. On cherche la matrice **P** telle que $\mathbf{P}^{-1}\,\mathbf{M}\,\mathbf{P}=\Lambda$

```
> P=eigen(M)$vectors
> T=solve(P) # T=P-1
```

On vérifie que $P^{-1}\,M\,P=\Lambda$:

```
> T %*% M %*% P
```

⇨ **Exercices** : retrouver les résultats des exercices 3, 4, 5, 10 et 11

Remarques : utiliser la fonction `fractions(MASS)`

Décomposition en valeurs singulières (attention aux notations du cours qui vont être différentes !!!! u ↔ v)

```
> hilbert <- function(n) { i <- 1:n; 1 / outer(i - 1, i, "+") }
> X <- hilbert(9)[,1:6]
> (s <- svd(X))
> D <- diag(s$d) # matrix of singular values of x
> s$u %*% D %*% t(s$v) #  X = U D V'
> t(s$u) %*% X %*% s$v #  D = U' X V
```

Il existe ausssi la fonction `svd.triplet`FactoMineR) `"Compute the singular-value decomposition of a rectangular matrix with weights for rows and columns".`

⇨ **Exercice** : retrouver les résultats de l'exercice 12.

Vecteur M-normé

```
> M<-matrix(c(3,1,-1,1,1,0,-1,0,2),ncol=3) ;M
> u<-matrix(c(2,1,-1),ncol=1);u
> NormeMu<-(t(u)%*%M%*%u)^0.5 ; NormeMu
> NormeMu1<-u%*%diag(1/NormeMu)
> t( NormeMu1)%*% M%*%NormeMu1
     [,1]
[1,]    1
```

Matrice de Vecteur M-orthonormés

```
> M<-matrix(c(3,1,-1,1,1,0,-1,0,2),ncol=3) ;M
     [,1] [,2] [,3]
[1,]    3    1   -1
[2,]    1    1    0
[3,]   -1    0    2
> u<-matrix(c(2,1,-1,1,0,2),ncol=2);u # vecteurs M-orhtog.
     [,1] [,2]
[1,]    2    1
[2,]    1    0
[3,]   -1    2
> NormeMu<-(t(u)%*%M%*%u)^0.5 ; NormeMu
         [,1]     [,2]
[1,] 4.795832 0.000000
[2,] 0.000000 2.645751
> NormeMu1<-u%*%diag(1/diag(NormeMu)) # vecteurs M-orthonormaux
> NormeMu1
           [,1]      [,2]
[1,]  0.4170288 0.3779645
[2,]  0.2085144 0.0000000
[3,] -0.2085144 0.7559289
> round( t( NormeMu1)%*% M%*%NormeMu1,3)
     [,1] [,2]
[1,]    1    0
[2,]    0    1
```

2. Examen et transformation du tableau de données sous R

Tableau de données : n individus x p variables

Exemple : Relevé des bénéfices de 6 magasins pour 3 produits lait (lait, eau, huile)

$$\begin{pmatrix} 16 & 2 & 0 \\ 8 & 12 & 10 \\ 12 & 16 & 14 \\ 20 & 8 & 14 \\ 16 & 4 & 10 \\ 0 & 6 & 12 \end{pmatrix}$$

Création possible du fichier

```
> ben<-array(rep(0,18),dim=c(6,3)) ; > ben<-edit(ben)
> ben<-as.data.frame(ben)
> colnames(ben)<-c("lait","eau","huile")
```

Examen du fichier

```
> apply(ben,2,mean) ; apply(ben,2,var)
> summary(ben)
> plot(ben) ; boxplot(ben)
> par(mfrow=c(3,1))
> apply(ben,2,hist,nclass=3)
```

Création d'un tableau centré réduit et de la matrice des corrélations

```
Rben<-scale(ben)*sqrt(dim(ben)[[1]]/(dim(ben)[1]-1)) ; Rben
# creation d'une fonction qui centre et réduit
cr=function(t)
{
return(scale(t)*sqrt(dim(t)[1]/(dim(t)[1]-1)))
}
cr(ben); dim(ben)
```



```
scale(cr(ben))
apply(Rben,2,mean) ; apply(Rben^2,2,sum) ; apply(Rben^2,2,mean)
Corben<-cor(ben) ; Corben
t(Rben)%*%Rben/dim(Rben)[[1]]
```

3. Centrage et réduction d'un nuage de points

Construction d'un tableau de données de 30 individus et de 2 variables suivant une loi Normale à 2 dimensions.

```
> library(MASS)
> XY<-mvrnorm(30,mu=c(0,0),Sig=matrix(c(1,0.75,0.75,1),2,2))
> plot(XY) ; plot(XY,asp=1) ; abline(v=0); abline(h=0)
> XY
> XY<-mvrnorm(30,mu=c(5,2),Sig=matrix(c(4,1.5,1.5,1),2,2))
```

→ Représenter le point moyen du nuage.

→ Construire le nuage centré `cXY`, puis réduire `RXY` et les représenter dans le même graphique avec une couleur différente (`option col= ''red''` voir aussi `colors()`).

→ Reprendre avec `XY<-mvrnorm(30,mu=c(3,0),Sig=matrix(c(1.0,0.75,0.75,5),2,2))`.

4. Axe factoriel pour un tableau à 2 variables

Construction du tableau de données de 100 individus et de 2 variables

```
> library(MASS)
> XY<-mvrnorm(100,mu=c(0,0),Sig=matrix(c(1,0.75,0.75,1),2,2))
> plot(XY) ; plot(XY,asp=1)
```

→ Représenter le point moyen du nuage.

```
> XY<-as.data.frame(XY)
> points(mean(XY[,1]),mean(XY[,2]),col="red")
> colnames(XY)<-c("X","Y")
```

Régression de Y en X et son principe

```
> abline(lm(XY$Y~XY$X))
> segments(XY$X,XY$Y,XY$X,predict(lm(XY$Y~XY$X)))
> attach(XY) ; X ; Y
```

Régression de X en Y et son principe

→ Représenter le nuage et le point moyen

→ Construire la figure correspondante.

Axe principal du nuage et son principe

Dans les régressions précédentes les variables n'ont pas un rôle symétrique. Pour construire un axe principal, où les deux variables ont un rôle symétrique, on minimise la somme des écarts entre les points observés et l'axe principal. On procèdera ainsi en ACP. Les axes principaux sont alors les vecteurs propres de la matrice de covariance.

```
> XY<-mvrnorm(100,mu=c(0,0),Sig=matrix(c(1,0.75,0.75,2),2,2))
```

→ Représenter le nuage et le point moyen

→ Représenter le principe de l'ajustement

```
> covXY<-var(XY) ; vp<-eigen(covXY)$vectors
> p<-vp[2,1]/vp[1,1] ; p
> q<-vp[2,2]/vp[1,2] ; q
> abline(c(mean(XY$Y)-p*mean(XY$X),p))
> abline(c(mean(XY$Y)-q*mean(XY$X),q))
```

→ Représenter les 2 vecteurs propres u_1 et u_2 de `covXY` avec pour origine le point moyen du nuage et avec légende (`arrows` et `text`)

```
> u1<-vp[,1]; u2<-vp[,2]
> arrows(mean(XY)[1],mean(XY)[2],mean(XY)[1]+u1[1],mean(XY)[2]+u1[2],length=0.05,angle=30,code=2,col="red")
> arrows(mean(XY)[1],mean(XY)[2],mean(XY)[1]+u2[1],mean(XY)[2]+u2[2],length=0.05,angle=30,code=2,col="blue")
> legend("topright",c("u1","u2"),text.col=c("red","blue"),lty=c(1,1),col=c("red","blue"),merge=T,bg='aliceblue',cex=0.7)
> segments(XY[,1],XY[,2],as.matrix(XY)%*%vp[,1]*vp[1,1],as.matrix(XY)%*%vp[,1]*vp[2,1])
> plot(as.matrix(XY)%*%vp[,1],as.matrix(XY)%*%vp[,2])
> points(mean(XY)[1],mean(XY)[2],col="green")
```

5. Exercices sous R

(Algèbre)

1. Soit la matrice **A** contenant les coordonnées de 8 individus en fonction de 2 variables **X** et **Y** :

$$\mathbf{A}^{T} = \begin{pmatrix} 2 & 5 & 8 & 3 & -2 & -5 & -8 & -3 \\ -4 & 0 & 4 & 4 & 4 & 0 & -4 & -4 \end{pmatrix}$$

a. Centrer le tableau et tracer le nuage de points.
b. Déterminer la matrice des covariances et en déduire le 1er axe factoriel du nuage. Le tracer.

2. Soit la forme quadratique q définie dans R^3 par $q(x,y,z)=3x^2+y^2+2z^2+2xy-2xz$

a. Déterminer la matrice **M** de q et construire **M** sous R.

b. Déterminer une base orthonormale de R^3 pour le produit scalaire classique et une base M-orthonormale constituée de vecteurs propres de **M**.

c. Soit $<,>_M$ le produit scalaire associé à q

i. Calculer $\|u\|_{\mathbf{M}}$, $\|v\|_{\mathbf{M}}$, $<u,v>_{\mathbf{M}}$ avec u'=(1,2,-1) et v'=(0,1,1).

ii. Calculer les affixes des projetés orthogonaux de A=(2,1,0) et B=(5,-2,3) sur les axes (O,u) et les axes (O,v).

iii. Calculer les affixes des projetés M-orthogonaux de A=(2,1,0) et B=(5,-2,3) sur les axes (O,u) et les axes (O,v).

Données du travail sur le projet à rendre lors du dernier TP semaine 43

Etude des données **`Projet M1 AD 1920.csv`**

Voir sujet projet **`PROJET M1 AD-1920.pdf`**.