

TP noté

Exercice 1 :

Sujet : Analyser des sports et de sportifs

But : On a collecté des données sur les scores de sportifs de haut niveau pendant un Décathlon (dix épreuves). On souhaite analyser des données pour comparer ces sportifs : ceux qui ont les meilleurs scores ? les plus mauvais ? Qu'est ce qui rapproche les meilleurs ?

1)

```
# =====  
# TP Noté  
# 2 exercices  
# =====  
#  
# =====  
# Exercice 1 : decathlon  
# =====  
# indiquer Working directory  
setwd("C:/Users/davia/Downloads/ingedis/Rstudio/datas")  
# chargement des datas  
etudeSportifs<- read.table("decathlon.csv",header=TRUE,sep=";")  
etudeSportifs  
attach(etudeSportifs)|
```

2)

decathlon.csv - Bloc-notes

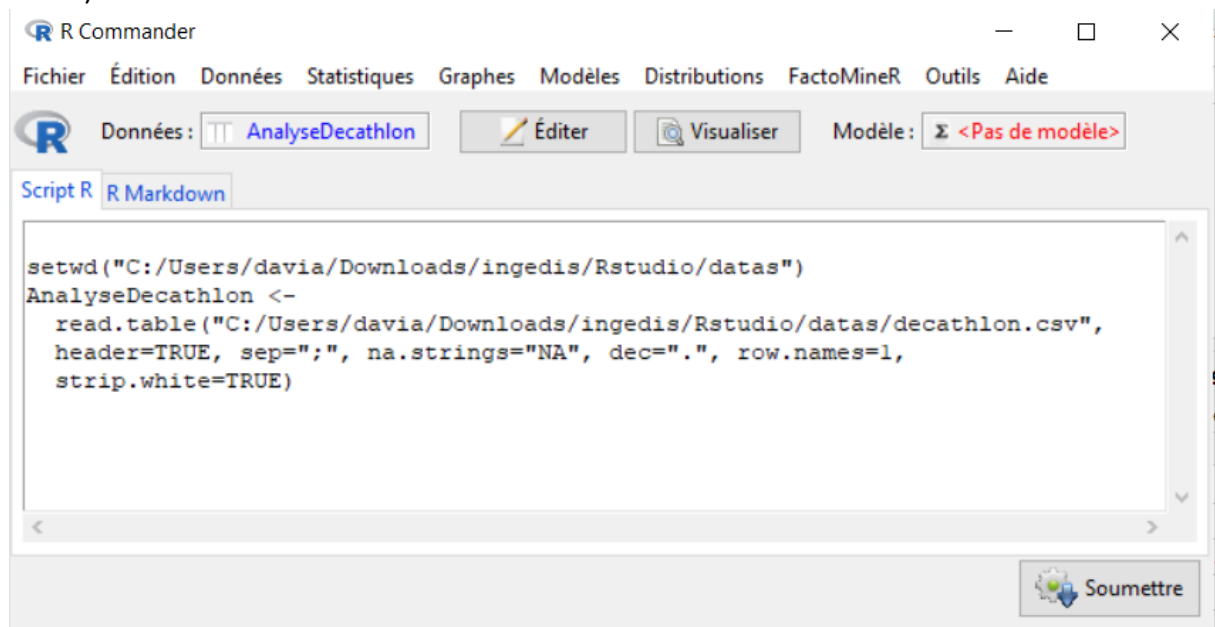
Fichier Edition Format Affichage Aide

```
;100m;Longueur;Poids;Hauteur;400m;110m H;Disque;Perche;Javelot;1500m;Classement;Points;Competition  
Sebrle;10.85;7.84;16.36;2.12;48.36;14.05;48.72;570.52;280.01;1;8893;JO  
Clay;10.44;7.96;15.23;2.06;49.19;14.13;50.11;4.9;69.71;282.2;8820;JO  
Karpov;10.5;7.81;15.93;2.09;46.81;13.97;51.65;4.6;55.54;278.11;3;8725;JO  
Macey;10.89;7.47;15.73;2.15;48.97;14.56;48.34;4.4;58.46;265.42;4;8414;JO  
Warners;10.62;7.74;14.48;1.97;47.97;14.01;43.73;4.9;55.39;278.05;5;8343;JO  
Zsivoczky;10.91;7.14;15.31;2.12;49.4;14.95;45.62;4.7;63.45;269.54;6;8287;JO  
Hernu;10.97;7.19;14.65;2.03;48.73;14.25;44.72;4.8;57.76;264.35;7;8237;JO  
Nool;10.8;7.53;14.26;1.88;48.81;14.8;42.05;5.4;61.33;276.33;8;8235;JO  
Bernard;10.69;7.48;14.8;2.12;49.13;14.17;44.75;4.4;55.27;276.31;9;8225;JO  
Schwarzl;10.98;7.49;14.01;1.94;49.76;14.25;42.43;5.1;56.32;273.56;10;8102;JO  
Pogorelov;10.95;7.31;15.1;2.06;50.79;14.21;44.6;5;53.45;287.63;11;8084;JO  
Schoenbeck;10.9;7.3;14.77;1.88;50.3;14.34;44.41;5;60.89;278.82;12;8077;JO  
Barras;11.14;6.99;14.91;1.94;49.41;14.37;44.83;4.6;64.55;267.09;13;8067;JO  
Smith;10.85;6.81;15.24;1.91;49.27;14.01;49.02;4.2;61.52;272.74;14;8023;JO  
Averyanov;10.55;7.34;14.44;1.94;49.72;14.39;39.88;4.8;54.51;271.02;15;8021;JO  
Ojanemi;10.68;7.5;14.97;1.94;49.12;15.01;40.35;4.6;59.26;275.71;16;8006;JO  
Smirnov;10.89;7.07;13.88;1.94;49.11;14.77;42.47;4.7;60.88;263.31;17;7993;JO  
Qi;11.06;7.34;13.55;1.97;49.65;14.78;45.13;4.5;60.79;272.63;18;7934;JO  
Drews;10.87;7.38;13.07;1.88;48.51;14.01;40.11;5;51.53;274.21;19;7926;JO  
Parkhomenko;11.14;6.61;15.69;2.03;51.04;14.88;41.9;4.8;65.82;277.94;20;7918;JO  
Terek;10.92;6.94;15.15;1.94;49.56;15.12;45.62;5.3;50.62;290.36;21;7893;JO  
Gomez;11.08;7.26;14.57;1.85;48.61;14.41;40.95;4.4;60.71;269.7;22;7865;JO  
Turli;11.08;6.91;13.62;2.03;51.67;14.26;39.83;4.8;59.34;290.01;23;7708;JO  
Lorenzo;11.1;7.03;13.22;1.85;49.34;15.38;40.22;4.5;58.36;263.08;24;7592;JO  
Karlivans;11.33;7.26;13.3;1.97;50.54;14.98;43.34;4.5;52.92;278.67;25;7583;JO  
Korkizoglou;10.86;7.07;14.81;1.94;51.16;14.96;46.07;4.7;53.05;317.26;7573;JO  
Uldal;11.23;6.99;13.53;1.85;50.95;15.09;43.01;4.5;60;281.7;27;7495;JO  
Casarsa;11.36;6.68;14.92;1.94;53.2;15.39;48.66;4.4;58.62;296.12;28;7404;JO  
SEBRLE;11.04;7.58;14.83;2.07;49.81;14.69;43.75;5.02;63.19;291.7;1;8217;Decastar  
CLAY;10.76;7.4;14.26;1.86;49.37;14.05;50.72;4.92;60.15;301.5;2;8122;Decastar  
KARPOV;11.02;7.3;14.77;2.04;48.37;14.09;48.95;4.92;50.31;300.2;3;8099;Decastar  
BERNARD;11.02;7.23;14.25;1.92;48.93;14.99;40.87;5.32;62.77;280.1;4;8067;Decastar  
YURKOV;11.34;7.09;15.19;2.1;50.42;15.31;46.26;4.72;63.44;276.4;5;8036;Decastar  
WARNERS;11.11;7.6;14.31;1.98;48.68;14.23;41.1;4.92;51.77;278.1;6;8030;Decastar  
ZSIVOCZKY;11.13;7.3;13.48;2.01;48.62;14.17;45.67;4.42;55.37;268.7;7;8004;Decastar  
MCMULLEN;10.83;7.31;13.76;2.13;49.91;14.38;44.41;4.42;56.37;285.1;8;7995;Decastar  
MARTINEAU;11.64;6.81;14.57;1.95;50.14;14.93;47.6;4.92;52.33;262.1;9;7802;Decastar  
HERNU;11.37;7.56;14.41;1.86;51.1;15.06;44.99;4.82;57.19;285.1;10;7733;Decastar  
BARRAS;11.33;6.97;14.09;1.95;49.48;14.48;42.1;4.72;55.4;282.11;7708;Decastar
```

3) On peut faire la méthode ACP sur le jeu de donnée décathlon car il y a des individus en lignes et des variables quantitatives en colonnes. En effet, nous avons 41 individus et 14 variables. En effet, il s'agit de comparer des individus (des sportifs entre eux). Chaque ligne (à part la 1ere) représente un individu et elles sont décrites, par au moins 4 colonnes (variables quantitatives). Ces colonnes représentent toutes la même nature de données (épreuves). Les nombres dans chacune de ces colonnes quantitatives représentent une valeur et pas un comptage (une température et pas un nombre de températures, une note et pas un nombre de notes, ...).

```
> str(etudeSportifs)
'data.frame': 41 obs. of 14 variables:
 $ X      : chr  "Sebrle" "Clay" "Karpov" "Macey" ...
 $ X100m  : num  10.8 10.4 10.5 10.9 10.6 ...
 $ Longueur : num  7.84 7.96 7.81 7.47 7.74 7.14 7.19 7.53 7.48 7.49 ...
 $ Poids  : num  16.4 15.2 15.9 15.7 14.5 ...
 $ Hauteur : num  2.12 2.06 2.09 2.15 1.97 2.12 2.03 1.88 2.12 1.94 ...
 $ X400m  : num  48.4 49.2 46.8 49 48 ...
 $ X110m.H : num  14.1 14.1 14 14.6 14 ...
 $ Disque  : num  48.7 50.1 51.6 48.3 43.7 ...
 $ Perche  : num  5 4.9 4.6 4.4 4.9 4.7 4.8 5.4 4.4 5.1 ...
 $ Javelot : num  70.5 69.7 55.5 58.5 55.4 ...
 $ X1500m  : num  280 282 278 265 278 ...
 $ Classement : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Points  : int  8893 8820 8725 8414 8343 8287 8237 8235 8225 8102 ...
 $ Competition: chr  "JO" "JO" "JO" "JO" ...
```

4 et 5)

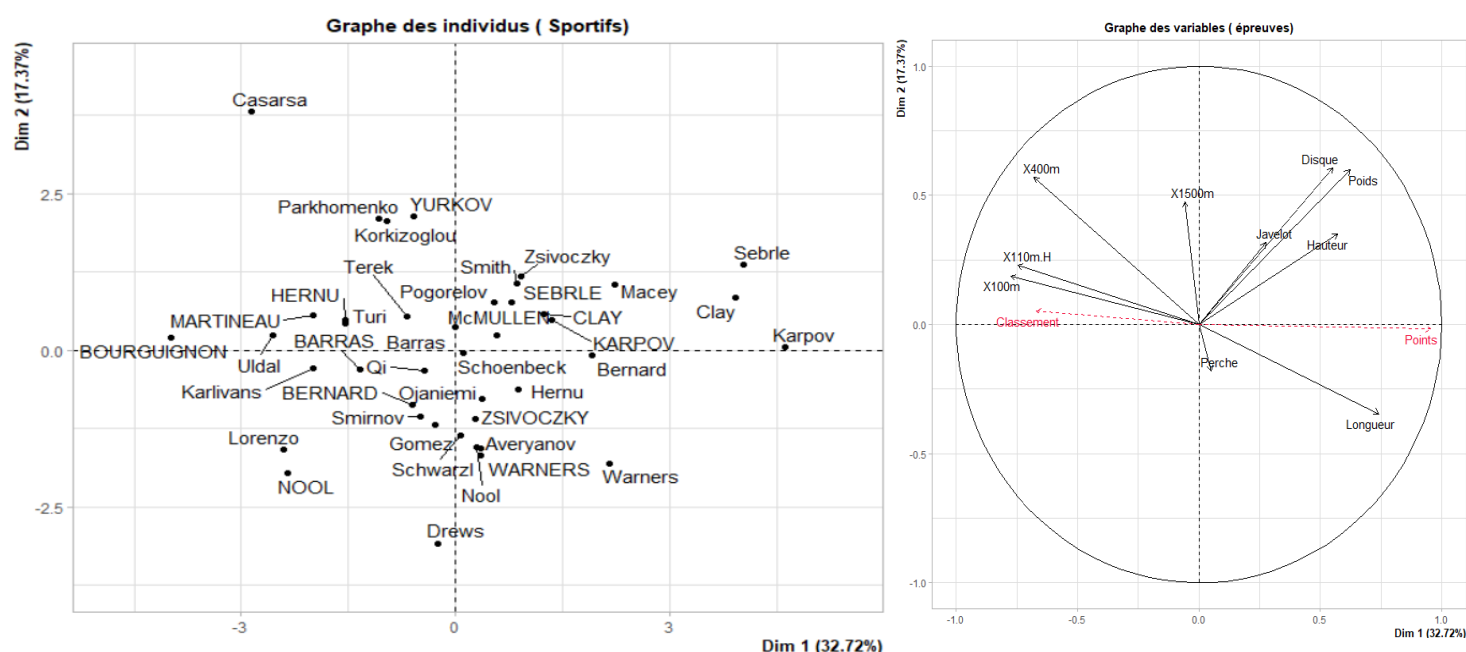


6) En Cliquant sur Visualiser. Les données se présentent correctement.

	X100m	Longueur	Poids	Hauteur	X400m	X110m.H	Disque	Perche	Javelot	X1500m	Classement	Points	Compétition
Sebrle	10.85	7.84	16.36	2.12	48.36	14.05	48.72	5.00	70.52	280.01	1	8893	JO
Clay	10.44	7.96	15.23	2.06	49.19	14.13	50.11	4.90	69.71	282.00	2	8820	JO
Karpov	10.50	7.81	15.93	2.09	46.81	13.97	51.65	4.60	55.54	278.11	3	8725	JO
Macey	10.89	7.47	15.73	2.15	48.97	14.56	48.34	4.40	58.46	265.42	4	8414	JO
Warners	10.62	7.74	14.48	1.97	47.97	14.01	43.73	4.90	55.39	278.05	5	8343	JO
Zsivoczky	10.91	7.14	15.31	2.12	49.40	14.95	45.62	4.70	63.45	269.54	6	8287	JO
Hernu	10.97	7.19	14.65	2.03	48.73	14.25	44.72	4.80	57.76	264.35	7	8237	JO
Nool	10.80	7.53	14.26	1.88	48.81	14.80	42.05	5.40	61.33	276.33	8	8235	JO
Bernard	10.69	7.48	14.80	2.12	49.13	14.17	44.75	4.40	55.27	276.31	9	8225	JO
Schwarzl	10.98	7.49	14.01	1.94	49.76	14.25	42.43	5.10	56.32	273.56	10	8102	JO

- 7) a. Les lignes sont des individus = ici des sportifs, (41)
- b. Les colonnes sont des variables (les caractères) : des épreuves (10)
- c. Variables numériques qui vont servir à l'analyse (variables actives) : 10 épreuves
- d. Variables numériques supplémentaires (ou variables illustratives) : Classement, points
- e. Variables qualitatives supplémentaires (ou facteurs illustratifs) : Compétition
- f. Le contenu d'une cellule de ces colonnes : Le résultat moyen à une épreuve

8)



9) Analyse des résultats

Analyser les individus

- 1) Sommes des dimensions : $32.72 + 17.37 = 50.09$ Or $50.09 \% < 70\%$ donc on ne peut pas faire confiance en cet ACP.
- 2) Interprétation de la dimension 1 : à droite les meilleurs résultats aux épreuves et à gauche les plus mauvais. Les sportifs les mieux représentés sur la Dim 1 sont : Karpov, Serbrle, Bourguignon et Casarsa

Interprétation individus très proches : « Bourguignon » et « Uldal » sont très proches : ils ont un peu près en moyenne les mêmes résultats aux épreuves.

Idem pour « Karlivans » et « Carras »

Par contre Karpov et Bourguignon ont des comportements très différents. Ces deux sportifs sont complètement opposés sur la Dim 1. En effet, l'un a des meilleurs résultats que l'autre.

- 3) Interprétation de la dimension 2 : plus on monte, plus il y a des différences entre les meilleurs résultats aux épreuves et les plus mauvaises. Terek et Hermu s'oppose à la variabilité des résultats aux épreuves. En effet, quand l'une augmente, l'autre diminue.

Analyser les variables

- 1) On interprète que les vecteurs proches du cercle. Les variables les mieux représentées et celles qui sont corrélées à Dim1 sont : Disque, Poids, X400m et Longueur.

Vecteur colinéaire : Disques et Poids

Longueur et X400m ont une corrélation négative

Interprétation des individus à partir du graphe des variables

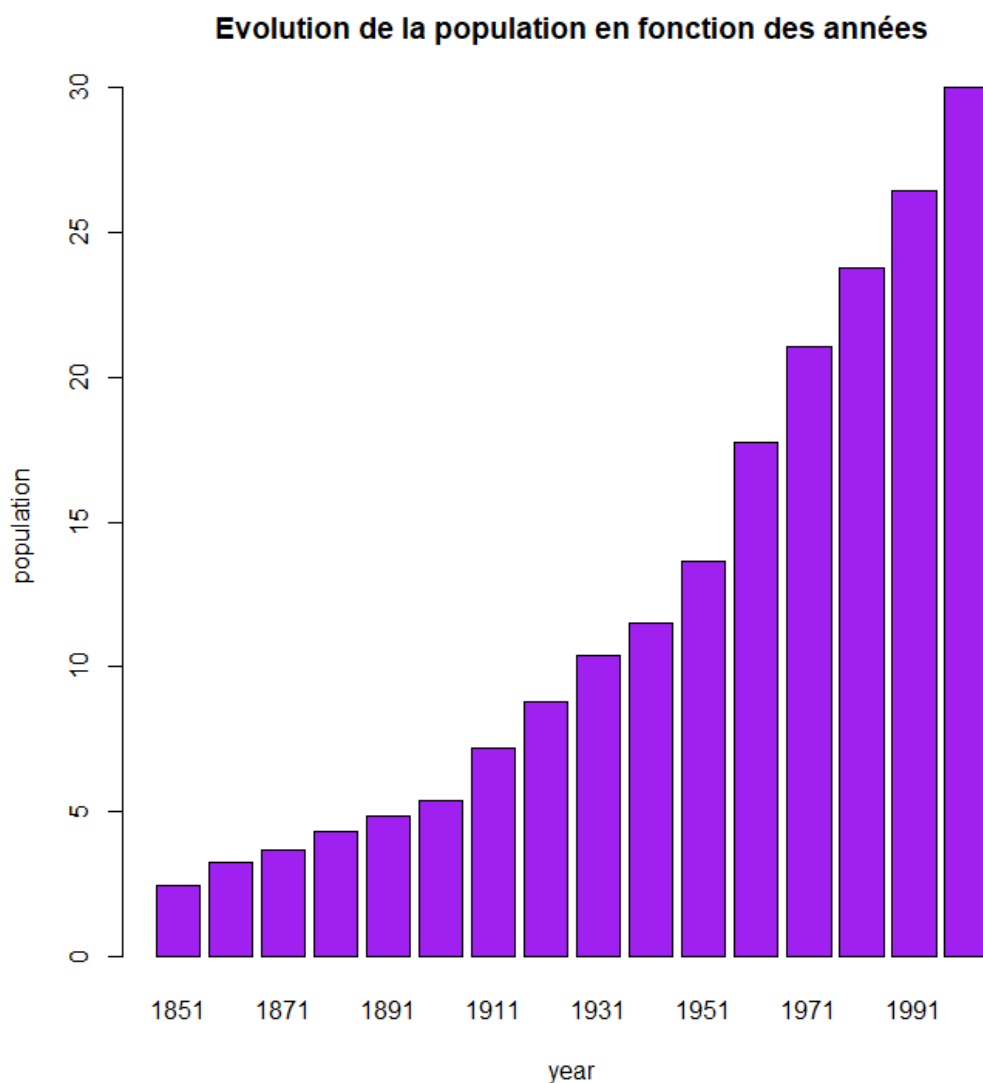
Si nous prenons Poids qui est une des meilleures variables corrélées à Dim 1 : Les Sportifs Karpov , Clay et Serbrle ont les meilleurs résultats à l'épreuve du Poids tandis Le sportif Bourguignon à les résultats les plus mauvais à cette épreuve.

Si nous prenons Longueur qui est une des meilleures variables corrélées à Dim 1 : Les Sportifs Karpov et Serbrle ont les meilleurs résultats à l'épreuve de longueur tandis Le sportif Bourguignon à les résultats les plus mauvais à cette épreuve.

Exercice 2 :

```
# =====
# Exercice 2 : nombre d'habitant région Canada
# =====
etudeCanPop<- read.table("CanPop.txt",header=TRUE,sep=";")
etudeCanPop
attach(etudeCanPop)
#
# Q1 : Evolution population en fonction des années
par(mfrow=c(1,1))
barplot(population~year,col="purple",main="Evolution de la population en fonction des années")
# La population ne cesse de croître en fonction des années
# Entre les années 1851 et 1881, la population a augmenté de 2.000 environ
# Entre 1881 et 1911, la population a augmenté de 3.000 environ
# Entre 1911 et 1941, la population a augmenté de 4.300 environ
# Entre 1941 et 1971, la population a augmenté de 9.500 environ
# entre 1971 et 2001, la population a augmenté de 8.960 environ

# En effet, elle augmente énormément entre les années 1941 et 1971.
```



```
# Q2 : Evolution population en fonction des années
# On peut supposer qu'il y a un lien entre la population et l'année, en effet comme
# le graphique précédent nous montre et comme les chiffres également nous montrent
# il semble avoir une relation linéaire entre l'année et la population.
# Avec les années qui passent, la population ne cesse de croître et de plus en plus.
```

```
# Q3 : rechercher un modèle linéaire pour la relation
# entre population et année :
# - détermination de l'équation de la droite
# - traçage de la droite d'équation
```

```
b1=sum((year-mean(year))*(population-mean(population)))/sum((year-mean(year))^2)
b1
b0=mean(population)-b1*mean(year)
b0
```

```
# l'équation du modèle est donc
#  $y = b_0 + b_1 \cdot x$  en théorie
# ici ça devient :  $population = -337.098 + 0.181 \cdot year$ 
```

```
plot(population~year,col="purple",main="Evolution de la population en fonction des années")
abline(lm(year,b0+b1*year))
# on a la meilleure droite possible comme modèle
# Elle passe ou s'approche de beaucoup de points
# on peut s'en servir comme modèle de prédiction
```

