

Alimentation

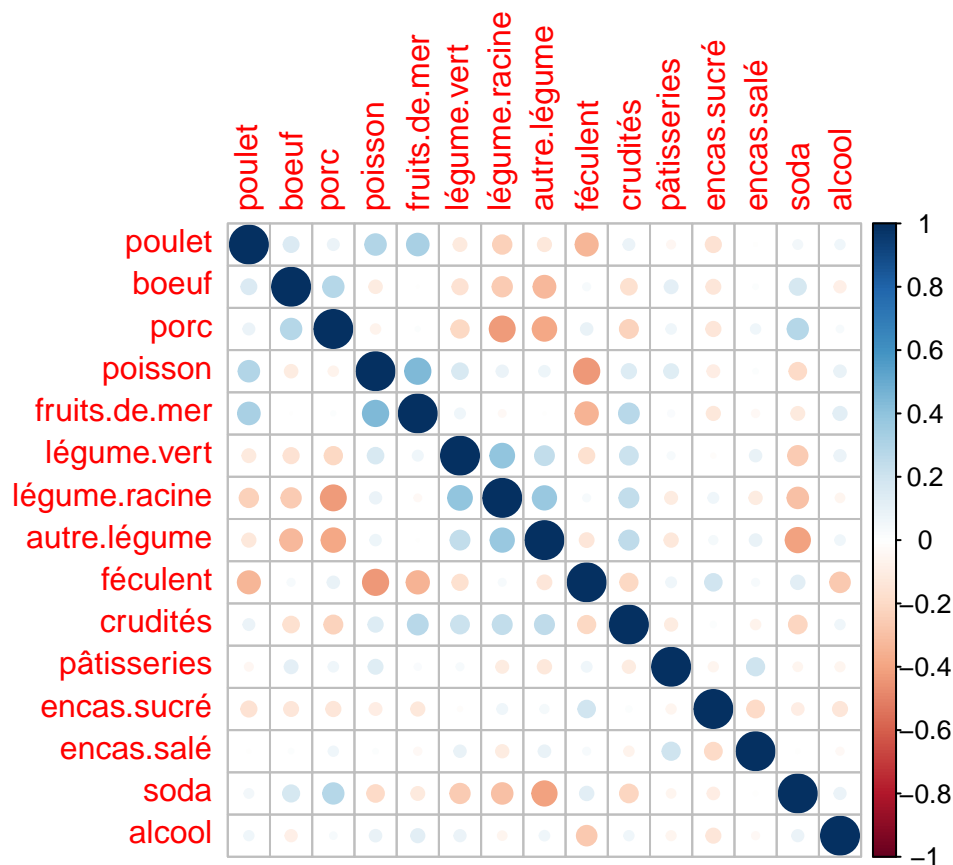
Arthur et Leonardo

12/13/2020

ACP

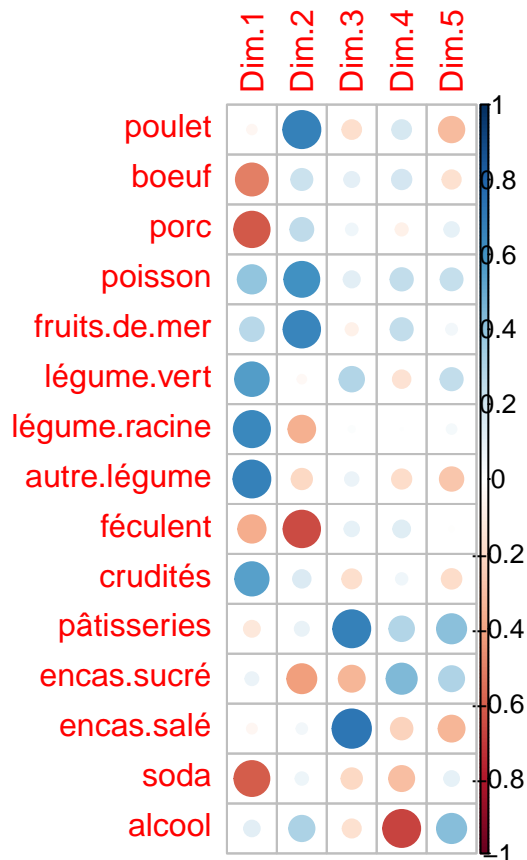
Certains aliments sont-ils corrélés ?

```
alim.cor <- cor(alim[, -c(16,17,18)])  
corrplot(alim.cor)
```



On constate une corrélation entre les variables poisson/fruits.de.mer, poulet/fruit.de.mer, legume.racine/legume.vert et entre le soda/porc. On constate cependant une corrélation négative entre les legumes.racine/porc, legume.vert/porc, féculents/poisson et le soda avec tous les légumes.

```
alim.pca <- PCA(alim, quali.sup = c(16,17,18), scale.unit=TRUE, ncp=5, graph = FALSE)  
corrplot(alim.pca$var$cor)
```



On peut constater que ce sont majoritairement les végétaux qui ont contribué à la formation du premier axe. Pour la création du deuxième axe, il s'agit plutôt des aliments animaux (poisson, fruit de mer, poulet).

Pour résumer on pourra globalement quantifier la masse de l'alimentation végétarienne sur le premier axe et la masse de l'alimentation carnivore sur le deuxième axe.

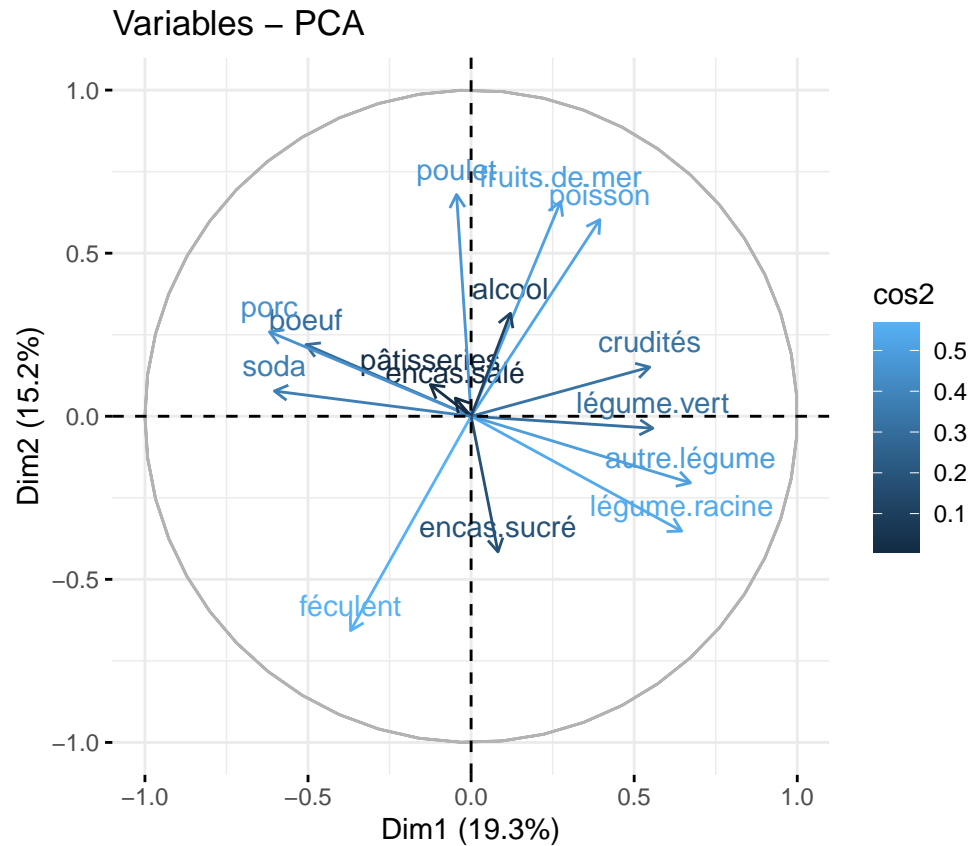
Peut-on résumer des variables fortement corrélées par des variables synthétiques ?

Les variables peuvent être très corrélées à axe (variable synthétique) sans être bien représentées, donc non. Cependant si leur qualité de représentation est très bonne pour les variables dans ce cas oui.

```
alim.pca$var$cos2
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
## poulet	0.001984174	0.462363074	0.0299300919	3.051103e-02	1.012592e-01
## boeuf	0.256557466	0.048189897	0.0124925694	3.435169e-02	2.700304e-02
## porc	0.382306926	0.066961026	0.0044211114	5.355027e-03	1.100448e-02
## poisson	0.155794714	0.363447948	0.0165144890	6.035843e-02	5.651962e-02
## fruits.de.mer	0.075343775	0.433036026	0.0053521189	5.882072e-02	3.462705e-03
## légume.vert	0.310525221	0.001345215	0.0879500735	2.300418e-02	5.894960e-02
## légume.racine	0.416619678	0.123996588	0.0003155472	3.353905e-06	2.097627e-03
## autre.légume	0.452124890	0.041622301	0.0080227057	3.341261e-02	7.322615e-02
## féculent	0.136591981	0.431916591	0.0108543916	1.959333e-02	9.751237e-05
## crudités	0.299905822	0.022689195	0.0318386161	4.160690e-03	3.566892e-02
## pâtisseries	0.015616327	0.009400564	0.4570010759	8.918050e-02	1.759377e-01
## encas.sucré	0.006788678	0.172806191	0.1106419170	2.020123e-01	9.565745e-02
## encas.salé	0.002318975	0.003100544	0.5286144452	5.264695e-02	1.101746e-01

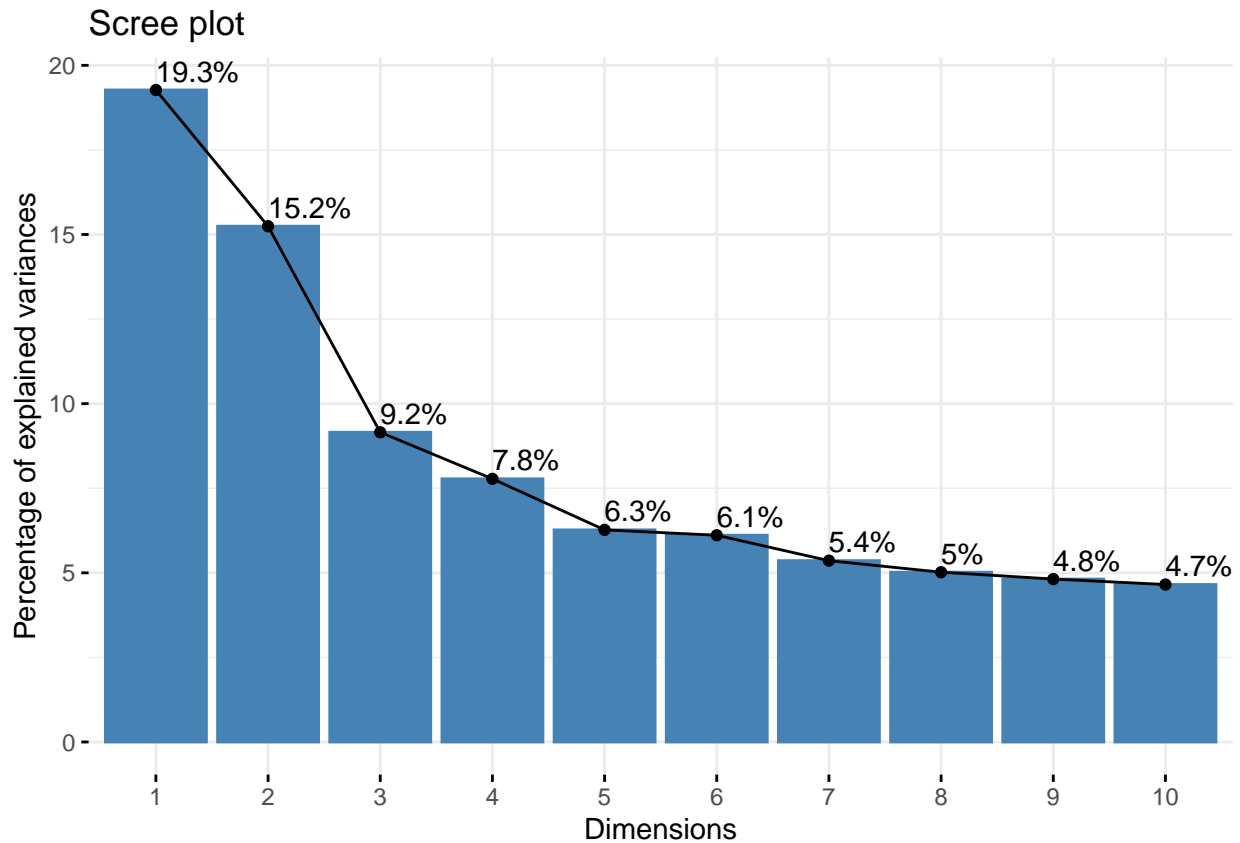
```
## soda      0.363571186 0.005985935 0.0429901032 9.558189e-02 1.182336e-02
## alcool    0.014572699 0.099704547 0.0260290204 4.575430e-01 1.774436e-01
fviz_pca_var(alim.pca, col.var = "cos2")
```



On constate que la qualité de représentation des aliments à base végétale sont un peu inférieure à 0.5. Cependant elles ont environ la même qualité de représentation sur le premier axe. Cela ne permet pas d'être sûr qu'elles sont totalement représentative de la réalité mais ça réduit l'erreur d'interprétation. De même pour le deuxième axe et l'alimentation carnivore.

Le nombre d'axes factoriels retenus est-il pertinent ?

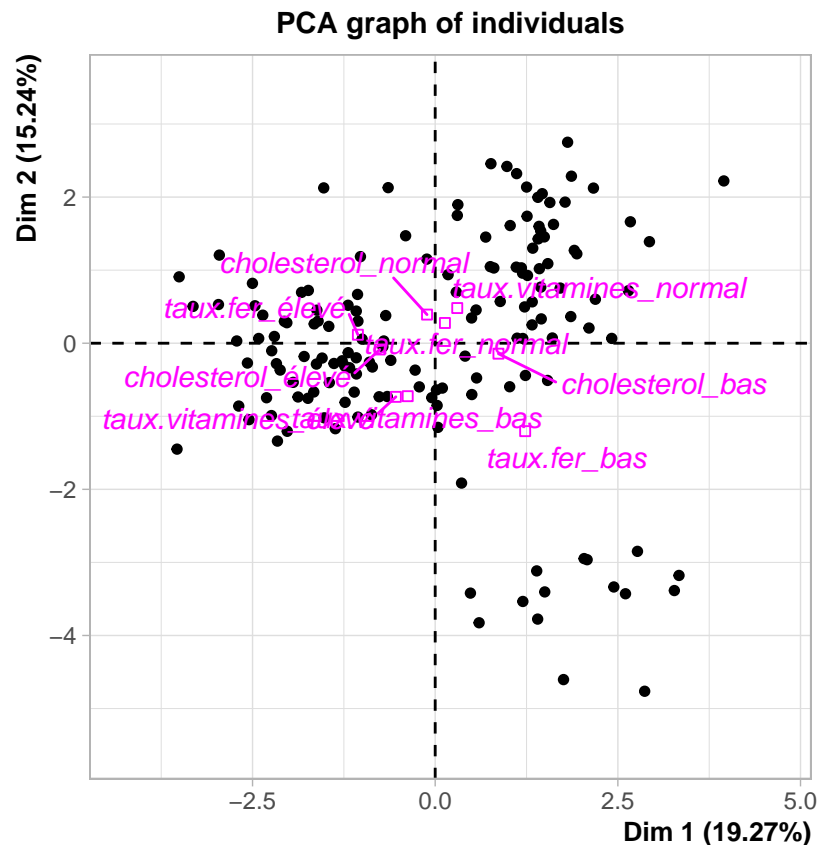
```
fviz_eig(alim.pca, addlabels = TRUE, choice = "variance" )
```



On utilise le critère du coude et on retient donc 2 axes ici. Cela semble pertinent car on peut distinguer deux types d'alimentation différentes: Végétarienne et Carnivore.

Y a-t-il des ressemblances ou des oppositions entre les individus ?

```
plot.PCA(alim.pca, axes=c(1, 2), choix="ind", habillage="none", col.ind="black", col.ind.sup="blue", col
```



On peut dresser la topologie suivante:

À vue d'oeil on peut séparer trois groupes:

- 1 - Un groupe mangeant peu de légumes (à gauche) et une quantité de viande normale. Ces individus ont un taux de fer et de cholestérol élevé.
- 2 - Un groupe aurait une alimentation variée (Au centre, en haut à droite). Ces individus sont dans la moyennes par rapport aux variables qualitatives.
- 3 - Un groupe mangeant très peu de viande et beaucoup de légumes (en bas à droite). Ces individus ont un taux de fer et de cholestérol bas.

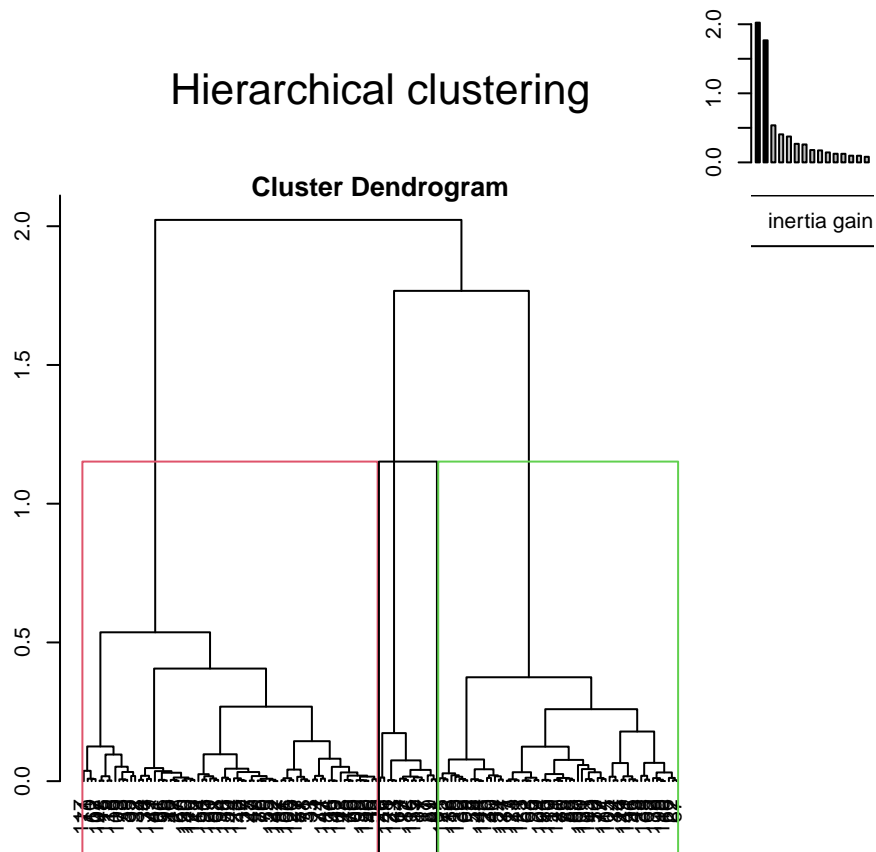
Pour conclure ne mangez ni trop gras, ni trop salée, ni trop sucrée. Et l'abus d'alcool est dangereux pour la santé.

CAH

```
alim.hcpc <- HCPC(alim.pca,
  nb.clust = 3,
  graph=FALSE)
```

Le nombre de clusters choisi est-il pertinent? D'après quel critère ?

```
plot(alim.hcpc, choice = "tree")
```



Le gain d'inertie le plus important a lieu pour $k = 3$. On garde 3 clusters.

Le premier plan factoriel suffit-il à résumer leurs différences ?

Non pour la même raison.

```
alim.hcpc$desc.var$quanti
```

```
## $`1`
##               v.test Mean in category Overall mean sd in category Overall sd
## soda          6.469034      14.799028      10.540530      8.438890      7.696891
## porc          6.403434       7.561667       5.391060      4.155063      3.963388
## boeuf         4.835161       7.105694       5.317682      4.843905      4.323723
## féculent      4.636643      18.805694      15.677550      6.169057      7.888266
## fruits.de.mer -4.092475       2.034167       3.044503      1.338251      2.886545
## poisson      -5.494623       2.564583       4.164768      2.113760      3.405106
## crudités     -5.527815       3.191528       5.515430      2.738332      4.915444
## autre.légume  -5.895673       4.840833       8.112185      3.466885      6.487723
## légume.vert  -6.391072       8.074722      11.792517      5.545635      6.801583
## légume.racine -7.020865      11.797500      16.683576      6.054055      8.137068
##               p.value
## soda          9.863147e-11
## porc          1.519205e-10
## boeuf         1.330379e-06
## féculent      3.541134e-06
## fruits.de.mer 4.267929e-05
## poisson       3.915457e-08
## crudités      3.242432e-08
```

```

## autre.légume 3.731584e-09
## légume.vert 1.647266e-10
## légume.racine 2.204991e-12
##
## $`2`
##          v.test Mean in category Overall mean sd in category Overall sd
## poisson      8.623615      6.998413      4.164768      2.754539      3.405106
## fruits.de.mer 6.902304      4.967143      3.044503      3.276724      2.886545
## crudités      4.703558      7.746508      5.515430      5.654544      4.915444
## poulet        4.634517      5.863810      4.455563      2.846355      3.148827
## légume.vert    4.512969     14.754603     11.792517      5.881191      6.801583
## autre.légume    3.244321     10.143333      8.112185      6.747194      6.487723
## légume.racine    2.762224     18.852540     16.683576      5.440758      8.137068
## alcool         2.416499     10.381111      9.072848      5.334963      5.610262
## boeuf          -2.051056      4.461905      5.317682      2.546712      4.323723
## porc           -3.008749      4.240317      5.391060      2.038467      3.963388
## soda           -4.227196      7.400794     10.540530      4.074790      7.696891
## féculent       -7.931888      9.639683     15.677550      5.316151      7.888266
##          p.value
## poisson      6.487312e-18
## fruits.de.mer 5.116594e-12
## crudités      2.556668e-06
## poulet        3.577726e-06
## légume.vert    6.392637e-06
## autre.légume    1.177309e-03
## légume.racine    5.740911e-03
## alcool         1.567056e-02
## boeuf          4.026152e-02
## porc           2.623256e-03
## soda           2.366217e-05
## féculent       2.158385e-15
##
## $`3`
##          v.test Mean in category Overall mean sd in category Overall sd
## légume.racine    6.967834     30.130625     16.683576      6.05034655      8.137068
## féculent         5.183401     25.375000     15.677550      4.80515478      7.888266
## autre.légume      4.369563     14.835625      8.112185      7.36767345      6.487723
## encas.sucré       3.781013      8.625625      4.888278      7.71002186      4.167677
## légume.vert       3.140997     16.859375     11.792517      6.29193876      6.801583
## alcool           -2.090029      6.291875      9.072848      4.13175541      5.610262
## soda             -3.725342      3.740000     10.540530      2.56580153      7.696891
## fruits.de.mer    -4.416972      0.020625      3.044503      0.07988028      2.886545
## boeuf            -4.560322      0.641250      5.317682      2.48355057      4.323723
## poisson          -4.899311      0.208125      4.164768      0.80606466      3.405106
## porc            -5.570941      0.154375      5.391060      0.59789180      3.963388
## poulet          -5.901686      0.048125      4.455563      0.18638732      3.148827
##          p.value
## légume.racine    3.218587e-12
## féculent         2.178764e-07
## autre.légume      1.244956e-05
## encas.sucré       1.561918e-04
## légume.vert       1.683737e-03
## alcool           3.661518e-02
## soda             1.950507e-04

```

```
## fruits.de.mer 1.000931e-05
## boeuf         5.107533e-06
## poisson       9.617313e-07
## porc          2.533672e-08
## poulet        3.598055e-09
```

Quels sont les différents types de comportements alimentaires ?

On observe 3 types de comportement alimentaire:

1 - Un premier à base de viande et de féculent (le fameux steak / coquille). Apparemment ils aiment bien les sodas, l'hygiène de vie semble pas top.

2 - On a ici une alimentation qui paraît saine (d'après mes connaissances) avec comme seule viande, du poulet et du poisson et en accompagnement des légumes. (Apparemment ils compensent avec l'alcool, ce sont les seuls à avoir un "test" positif :)

3 - Ici on a les végétariens.

Qu'est-ce qui les caractérise ?

- 1 - Le soda
- 2 - Le poisson
- 3 - Les légumes racines

– À noter que les p-values sont très basses et qu'il y a donc peu de chance que l'on fasse un faux positif.

Ce qui pourrait être amusant est d'ajouter au jeu de données la longévité de chaque individu (à moins qu'ils soient encore en vie), et de regarder si il y a une corrélation.

Adisc

Quelles sont les variables permettant de discriminer les différents groupes ?

```
alim.des.fer <- desDA(alim[,1:15], alim$taux.fer)
as.data.frame(alim.des.fer$power)
```

##	cor_ratio	wilks_lamb	F_statistic	p_values
## poulet	0.038734919	0.9612651	2.98188716	5.375196e-02
## boeuf	0.088393243	0.9116068	7.17535271	1.061103e-03
## porc	0.123104012	0.8768960	10.38857176	6.000095e-05
## poisson	0.038716929	0.9612831	2.98044646	5.382645e-02
## fruits.de.mer	0.035816145	0.9641839	2.74884790	6.727121e-02
## légume.vert	0.128144377	0.8718556	10.87643834	3.916502e-05
## légume.racine	0.115638856	0.8843611	9.67622262	1.123555e-04
## autre.légume	0.133174508	0.8668255	11.36897069	2.552372e-05
## féculent	0.029563205	0.9704368	2.25432217	1.085368e-01
## crudités	0.021453226	0.9785468	1.62234319	2.009267e-01
## pâtisseries	0.009938513	0.9900615	0.74283261	4.775297e-01
## encas.sucré	0.048657393	0.9513426	3.78480589	2.494216e-02
## encas.salé	0.001293509	0.9987065	0.09584361	9.086624e-01
## soda	0.064188308	0.9358117	5.07573785	7.378134e-03
## alcool	0.034710283	0.9652897	2.66092229	7.322651e-02

Le porc et les différents légumes discriminent au mieux les groupes pour le taux de fer.


```
alim.des.vit <- desDA(alim[,1:15], alim$taux.vitamines)
as.data.frame(alim.des.vit$power)
```

```
##               cor_ratio wilks_lamb F_statistic      p_values
## poulet          0.0541244140  0.9458756  4.23439055 1.628235e-02
## boeuf           0.0087837576  0.9912162  0.65575809 5.205499e-01
## porc           0.0001620947  0.9998379  0.01199696 9.880757e-01
## poisson        0.1805011332  0.8194989 16.29908764 4.004778e-07
## fruits.de.mer  0.0941944643  0.9058055  7.69523931 6.615883e-04
## légume.vert    0.0099645315  0.9900355  0.74479688 4.766019e-01
## légume.racine  0.0027612322  0.9972388  0.20489695 8.149620e-01
## autre.légume   0.0135280212  0.9864720  1.01480183 3.649823e-01
## féculent       0.1336113513  0.8663886 11.41201470 2.458916e-05
## crudités       0.0315264770  0.9684735  2.40890354 9.343157e-02
## pâtisseries    0.0062730595  0.9937269  0.46713678 6.277153e-01
## encas.sucré    0.0059372143  0.9940628  0.44197798 6.436094e-01
## encas.salé     0.0238819446  0.9761181  1.81050221 1.671759e-01
## soda          0.0116184029  0.9883816  0.86986830 4.211377e-01
## alcool         0.0208378531  0.9791621  1.57481693 2.104949e-01
```

Le poisson et les féculents discriminent au mieux les groupes pour le taux de vitamine.

```
alim.des.chol <- desDA(alim[,1:15], alim$cholesterol)
as.data.frame(alim.des.chol$power)
```

```
##               cor_ratio wilks_lamb F_statistic      p_values
## poulet          0.011989548  0.9880105  0.89799310 0.4095942707
## boeuf           0.064359542  0.9356405  5.09020971 0.0072788951
## porc           0.080229250  0.9197707  6.45483077 0.0020525321
## poisson        0.052487950  0.9475121  4.09927058 0.0185042252
## fruits.de.mer  0.002370731  0.9976293  0.17585099 0.8389179470
## légume.vert    0.116373645  0.8836264  9.74580455 0.0001056528
## légume.racine  0.071061893  0.9289381  5.66085083 0.0042758420
## autre.légume   0.076682951  0.9233170  6.14581781 0.0027287469
## féculent       0.025968019  0.9740320  1.97286481 0.1426980445
## crudités       0.076134034  0.9238660  6.09819904 0.0028514362
## pâtisseries    0.005326789  0.9946732  0.39629339 0.6735210554
## encas.sucré    0.013475183  0.9865248  1.01078405 0.3664317473
## encas.salé     0.018805545  0.9811945  1.41828192 0.2454009812
## soda          0.076607064  0.9233929  6.13923118 0.0027453932
## alcool         0.001068320  0.9989317  0.07914022 0.9239494322
```

Les légumes verts discriminent au mieux les groupes pour le taux de cholesterol

Quelles sont celles qui décrivent au mieux les groupes ?

Pour le fer:

```
rownames(alim.des.fer$discor)[abs(alim.des.fer$discor)==max(abs(alim.des.fer$discor))]
```

```
## [1] "autre.légume"
```

Pour les vitamines:

```
rownames(alim.des.vit$discor)[abs(alim.des.vit$discor)==max(abs(alim.des.vit$discor))]
```

```
## [1] "poisson"
```

Pour le cholesterol:

```
rownames(alim.des.chol$discor)[abs(alim.des.chol$discor)==max(abs(alim.des.chol$discor))]  
## [1] "légume.vert"
```