



# Correction des activités

## des grains 02 et 03

Dominique Bertrand  
Data-Frame, Nantes, France

Benoit Jaillais  
INRA, Nantes, France

Sébastien Preys  
Ondalys, Montpellier, France

Eric Latrille  
INRA, Narbonne, France

V18.10



L'auteur autorise toute utilisation de l'oeuvre originale (y compris à des fins commerciales) ainsi que la création d'oeuvres dérivées, à condition qu'elles soient distribuées sous une licence identique à celle qui régit l'oeuvre originale présentée.

## Table des matières

<b>I</b>	<b>Réponses aux activités du grain 02</b>	<b>3</b>
1	Réponses à la question 1.	3
2	Réponses à la question 2	4
<b>II</b>	<b>Réponses aux activités du grain 03.</b>	<b>5</b>
3	Réponses à la question 1 : visualisation des données.	6
4	Réponses à la question 2 : réalisation et visualisation d'une ACP.	6
5	Réponses à la question 3 : interprétation des résultats d'une ACP.	8

## Première partie

# Réponses aux activités du grain 02

## 1 Réponses à la question 1.

- $\mathbf{X}(3, 2)$  indique la troisième ligne et la deuxième colonne de la matrice  $\mathbf{X}$ , et représente donc la réponse de Harald à la question « Jog : Aimez-vous faire du jogging ? ». Harald est faiblement intéressé et a donné la note 3 à cette question.
- La matrice des distances est :

	<i>Lara</i>	<i>Eric</i>	<i>Harald</i>	<i>Flora</i>
<i>Lara</i>	0.0	10.82	12.57	9.7
<i>Eric</i>	10.82	0.0	11.7	9.54
<i>Harald</i>	12.57	11.7	0.0	13.49
<i>Flora</i>	9.7	9.54	13.49	0.0

Cette matrice est symétrique, avec des valeurs nulles sur la diagonale (la distance d'une personne à elle-même est évidemment nulle). Les distances sont du même ordre de grandeur, montrant que les personnes n'ont pas beaucoup de goûts communs ! Les personnes les plus proches sont Eric et Flora. Les plus éloignées sont Harald et Flora.

- Le vecteur associé à Harald est :  $[0, 3, 10, 0, 0]$ . Sa norme est égale à :  $\sqrt{0^2 + 3^2 + 10^2 + 0^2 + 0^2} = 10.4$ .

*Note pour les matheux : lorsqu'on ne donne pas de précision, comme ici, la norme est la « 2-norme », c'est-à-dire la longueur euclidienne du vecteur.*

## 2 Réponses à la question 2

- La figure 1 montre que la moyenne et l'écart-type donnent des profils très semblables. En spectroscopie, cela indique qu'un facteur physique (indépendant de la composition chimique du produit étudié) a une très forte influence sur la mesure spectrale.

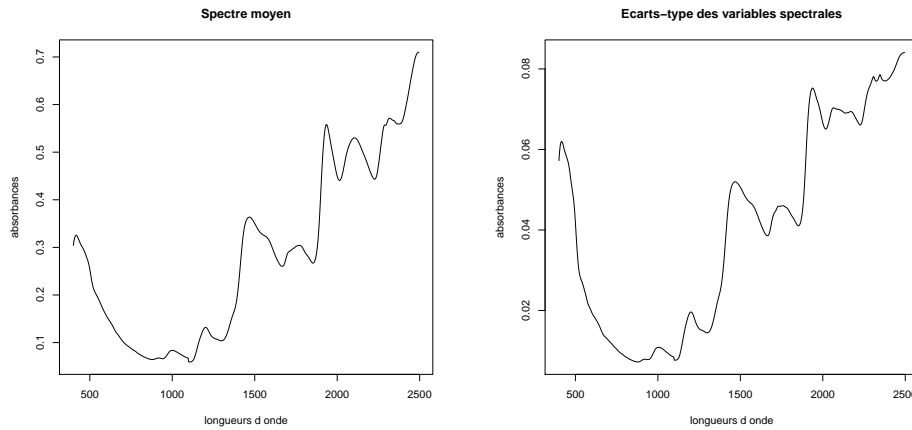


FIGURE 1 – Représentation des moyennes à gauche, des écarts-types à droite

- Sur la figure 2, seule une partie de la plage spectrale du visible (autour de 800 nanomètres) n'est pas fortement corrélée à la teneur en protéines. Toutes les autres régions spectrales présentent un coefficient de corrélation élevé, de l'ordre de 0.65.

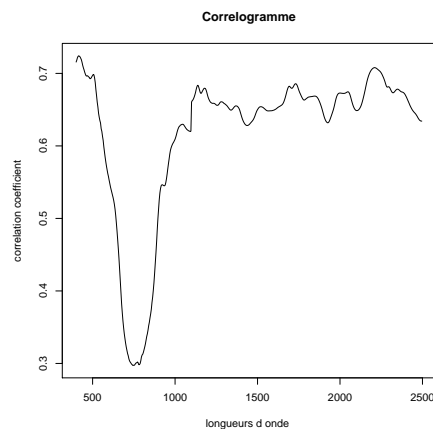


FIGURE 2 – Corrélations entre chacune des variables spectrales et la teneur en protéines

- L’histogramme de la figure 3-a présente deux pics. On observe un lien marqué entre la teneur en protéines et la nature « tendre/dur » des blés. Une explication : les blés durs sont en général plus riches en protéines que les blés tendres.

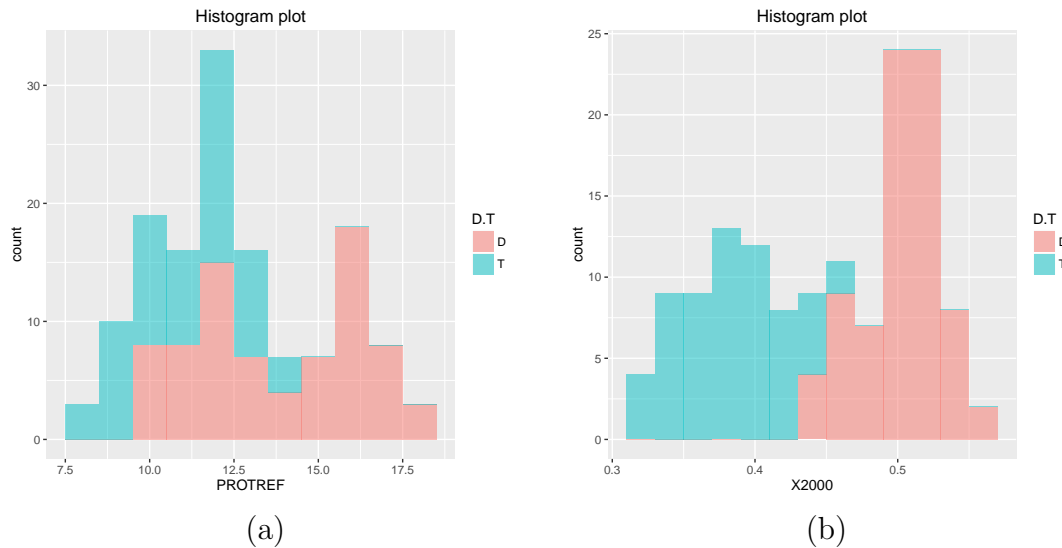


FIGURE 3 – Histogrammes des teneurs en protéines (à gauche) et de la longueur d’onde 2000nm (à droite)

- 2.4. La figure 3-b a été obtenue avec **bin width** fixé à  $0.02$ . L’histogramme des absorbances à 2000 nanomètres met en évidence un clivage presque parfait entre les blés tendres, à gauche de la figure, et les blés durs, à droite.

L’explication physique est la suivante. Les blés sont mis à l’état de farines par une étape de broyage standardisée. Par nature, les blés tendres donnent toujours une granulométrie plus fine que celle des blés durs. La spectroscopie proche infrarouge est très sensible à la granulométrie, ce qui se traduit par une hauteur générale des spectres différente entre les blés tendres et les blés durs.

## Deuxième partie

# Réponses aux activités du grain 03.

### 3 Réponses à la question 1 : visualisation des données.

Une variable de la figure 4 a une dispersion de 20 unités et d'autres variables de moins d'une unité. Les variances sont donc très différentes. Il est nécessaire de normaliser ces variables avant de faire une analyse en composantes principales.

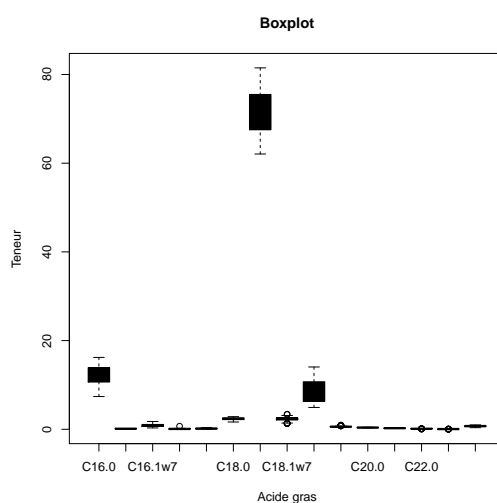


FIGURE 4 – Boite à moustaches, ou graphe de dispersion des valeurs pour chacune des 15 variables analysées.

### 4 Réponses à la question 2 : réalisation et visualisation d'une ACP.

- D'après l'éboullis des valeurs propres, figure 5, il faut au moins 3 composantes principales pour expliquer plus de 70% de la variance. La 4ème et la 5ème composante principale représentent 8% et 7% de la variance, respectivement. Les suivantes représentent moins de 4% de la variance et n'ont pas à être interprétées. Les trois premières composantes principales sont donc représentatives mais il n'est pas interdit d'aller regarder jusqu'à la 5<sup>e</sup> composante principale.

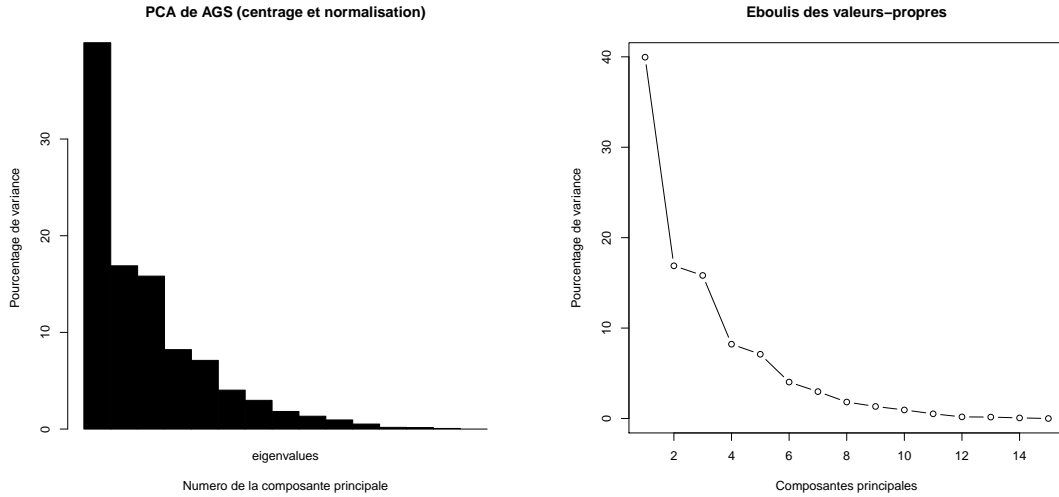


FIGURE 5 – Eboulis des valeurs-propres, soit le pourcentage de variance expliquée par chaque composante principale; représentation sous forme de graphique en barres (à gauche) ou de courbe (à droite).

— Les représentations des observations sur les axes 1 et 2 ou 1 et 3 de l’ACP, figure 6, montrent qu’il n’y a pas véritablement de point atypique ou outlier.

Les huiles sont différenciées selon leurs origines. Ainsi, NY (Nyons) est visiblement différente de toutes les autres origines. NI (Nice), NM (Nîmes) et HP (Haute Provence) sont aussi assez bien différenciées. Par contre, les origines AP (Aix en Provence) et VB (Vallée des Baux de Provence) restent très imbriquées entre elles.

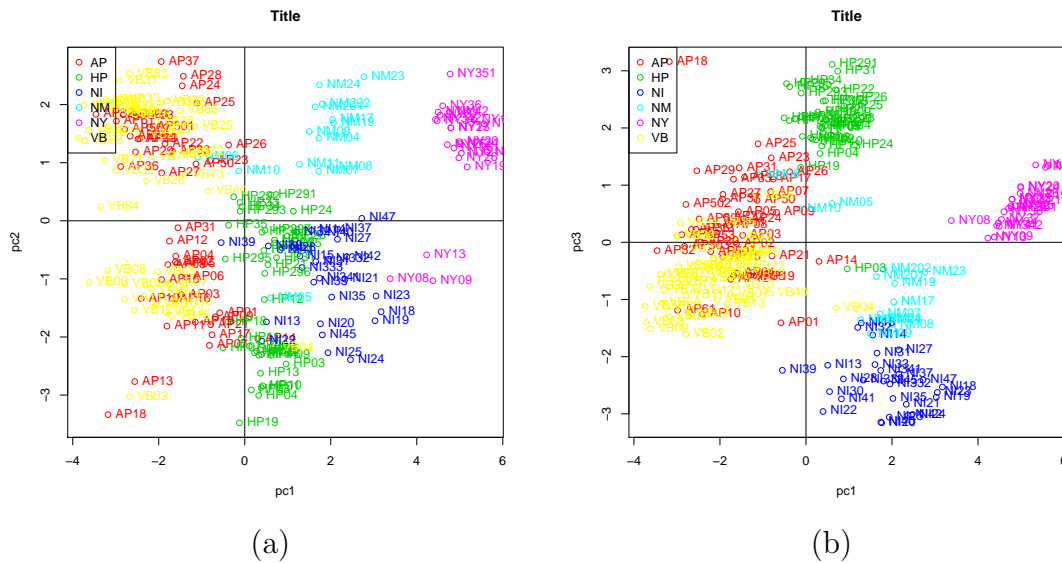


FIGURE 6 – Représentation des observations sur les axes 1 et 2 (a) ou 1 et 3 (b) de l’ACP; chacune des 6 origines d’huile est représentée par une couleur différente.

## 5 Réponses à la question 3 : interprétation des résultats d'une ACP.

- Le cercle de corrélation, figure 7, montre sur le plan 1-2 que C18 :1w9 forme un groupe à part entière. Il s'oppose au groupe représenté par les variables C16 :0, C16 :1w7 et C18 :1w7. Les acides gras C24 :0, C22 :0 et C18 :2w6 forment un autre groupe ; et C18 :0 et C20 :0 forment aussi un autre groupe.

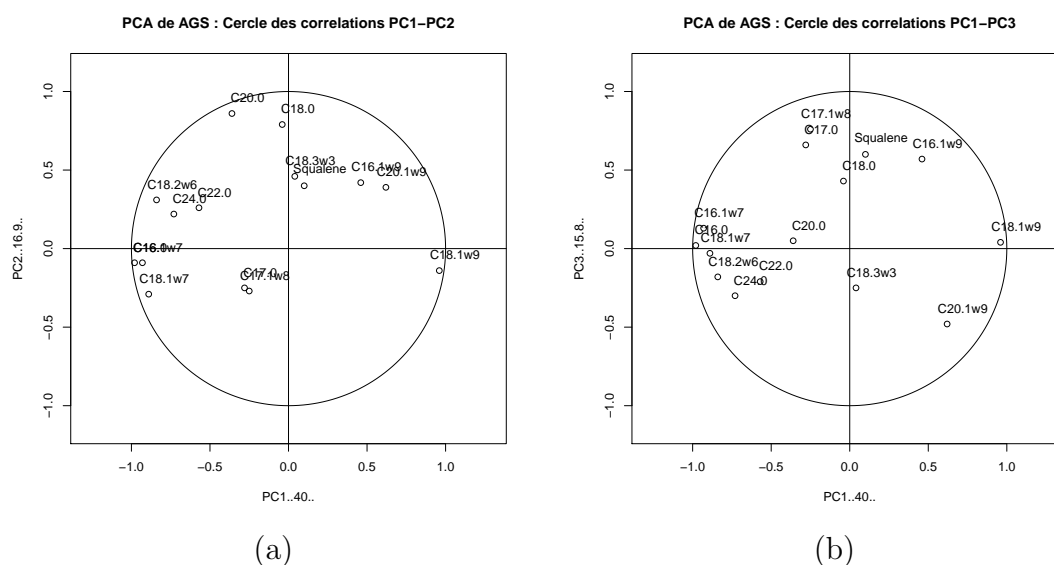


FIGURE 7 – Représentation du cercle de corrélations des variables d'origine avec les axes 1 et 2 (a) ou 1 et 3 (b) de l'ACP.

- Une interprétation des axes factoriels est la suivante :
  - l'axe PC1 oppose les acides linoléique C18 :2w6 et oléique C18 :1w9 ;
  - l'axe PC2 représente les acides gras C18 :0 et C20 :0 ;
  - l'axe PC3 représente les acides gras en C17.