

# TP1 ADM

MARIAC Damien, Matteo Scaia

October 9, 2024



## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Partie 1</b>	<b>3</b>
2.1	inertie et barycentre . . . . .	3
2.2	Autour des types forestiers . . . . .	5
<b>3</b>	<b>Conclusion</b>	<b>6</b>

# 1 Introduction

On dispose d'un jeu de données présentant une étude de 27 espèces d'arbres dans 1000 parcelles d'une forêt. Il s'agit d'étudier la variabilité des densités de peuplement d'espèces arborées dans différentes parcelles de la forêt du bassin du Congo. Nous disposons dans notre jeu de données 30 variables quantitatives dont : 27 variables de comptage des espèces, la surface de la parcelle, 2 variables forestières et une géologique. Et une variable qualitative "code".

## 2 Partie 1

### 2.1 inertie et barycentre

Nous cherchons à calculer la densité de peuplement de chaque espèce par unité de surface. Nous calculons alors pour chaque parcelle :

$$(d_j^i)_{1 \leq i \leq 27} = \frac{x_j^i}{s_j}$$

Table 1: Extrait de densité

Code	Gen1	Gen5	Gen10
1	0	0	2.200
2	0.6	0.133	1.333
3	0.514	0.057	3.6
4	0	0.439	0.244
5	0.095	0	0.476

Nous utiliserons des densités plutôt que des comptages car cela permet de normaliser les données par rapport à la taille de la parcelle, ce qui rend les comparaisons entre les parcelles équitables.

Nous devons centrer et réduire les variables quantitatives dans le but de mieux comparer celles qui décrivent les différentes densités. Nous allons alors utiliser :

$$(x_j^i)_{1 \leq i \leq 27} = \frac{x_j^i - \bar{x}_j}{\sigma_j}$$

Avec  $\bar{x}_j$  la moyenne pour la  $j$ ème variable et  $\sigma_j$  l'écart-type de la variable quantitative  $j$ .

**Par conséquent on a :**

Barycentre à l'origine : (preuve théorique)

Supposons que nous avons un ensemble de données  $X$  composé de  $n$  observations et  $p$  variables. Après le centrage et la réduction, la matrice transformée  $X'$  est définie par :

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

où  $\bar{x}_j$  est la moyenne et  $s_j$  l'écart-type de la  $j$ -ème variable.

Le barycentre de  $X$  est donné par la moyenne de chaque colonne de  $X$ .  
Calculons cette moyenne pour n'importe quelle  $j$ :

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n \tilde{x}_{ij} = \frac{1}{n} \sum_{i=1}^n \frac{x_{ij} - \bar{x}_j}{s_j} = \frac{1}{s_j} \left( \frac{1}{n} \sum_{i=1}^n x_{ij} \right) - \frac{\bar{x}_j}{s_j} = 0$$

Ainsi, le barycentre de chaque variable dans  $X'$  est zéro.

Inertie totale égale à 27 : (preuve théorique)

Considérons la même matrice de données  $X$ . Après centrage et réduction, chaque élément de la matrice transformée  $X'$  est défini par:

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

En reprenant les mêmes notations que avant.

L'inertie de l'ensemble des points  $X'$  par rapport à leur barycentre  $\mathbf{y}_M$  est définie par:

$$I_{Y,W} = \sum_{i=1}^n w_i \|\mathbf{x}_i - \mathbf{y}\|^2$$

Pour les données centrées-réduites, chaque  $\mathbf{x}'_i$  est déjà centré, donc le barycentre  $\mathbf{y} = \mathbf{0}$ . Par conséquent, la formule de l'inertie se simplifie à:

$$I_{Y,W} = \sum_{i=1}^n w_i \|\mathbf{x}'_i\|^2$$

Si tous les poids  $w_i$  sont égaux (par exemple,  $w_i = \frac{1}{n}$  ce qui est notre cas), alors l'inertie devient:

$$I_{Y,W} = \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}'_i\|^2$$

Comme chaque  $\mathbf{x}'_i$  est une observation centrée-réduite et la variance de chaque variable est 1, nous avons (pour  $i$  fixé):

$$\|\mathbf{x}'_i\|^2 = \sum_{j=1}^p (x'_{ij})^2 = p$$

Ainsi, l'inertie totale est:

$$I_{Y,W} = \frac{1}{n} \sum_{i=1}^n p = p$$

ce qui montre que l'inertie totale du nuage des données centrées-réduites est égale au nombre de variables  $p$  (qui dans notre cas vaut 27).

## 2.2 Autour des types forestiers

Pour calculer le poids de chaque type forestiers, nous avons divisé le nombre d'occurrences de chaque type forestier par 1000 (qui correspond au nombre de parcelle). Ensuite nous avons stocké dans une matrice le barycentre relatif à chaque typer forestier. On obtient alors le tableau de valeur :

▲	V1	V2	V3	V4	V5
1	-0.24262294	0.23082910	0.03611401	-0.08889818	0.06922725
2	-0.05236911	-0.28419720	-0.34740544	-0.34547466	-0.05978232
3	-0.49949475	0.34566947	0.52721272	-0.34547466	-0.12237513
4	-0.57903936	2.26944388	-0.38335630	-0.12426722	1.68021305
5	-0.02450022	-0.15852951	-0.16319450	-0.24582392	-0.22309466
6	-0.16040638	-0.26337031	-0.17338984	-0.14638021	0.51625842
7	0.36338233	-0.09895475	0.21019653	0.40343818	-0.16569062

figure 2: matrice des barycentres pour chaque type forestier

### 3 Conclusion