

Analyse en composantes principales

Le package FactoMineR est utiles pour les exercices suivants.

Exercice 1 ACP et effet taille

Le jeu de données considéré est constitué de 200 crabes décrits par huit variables (3 qualitatives et 5 quantitatives). Charger le jeu de données et sélectionner les variables quantitatives en utilisant le code R suivant :

```
> library(MASS)
> data(crabs)
> crabsquant<-crabs[,4:8]
```

Cette étude vise à utiliser l'ACP pour trouver une représentation des crabes qui permettent des distinguer visuellement différents groupes, liés à l'espèce et au sexe.



FIG. 1 – *l'individu en question*

1. Testez une ACP sur `crabsquant` sans traitement préalable. Que constatez vous?
2. Trouvez une solution pour améliorer la qualité de votre représentation en terme de visualisation des différents groupes.
3. Que dire de la qualité de représentation de cette nouvelle ACP? Combien d'axes retenez-vous? Pourquoi?
4. Comment interprétez-vous les axes retenus à partir du cercle des corrélations?
5. Que pouvez-vous en déduire sur la caractérisation des mâles/femelles, crabes oranges/bleus?

Exercice 2 Phylogénie des globines

On se propose d'effectuer une analyse factorielle des dissimilarités de séquence protéique de plusieurs globines issues de différentes espèces et de comparer les résultats obtenus à l'arbre phylogénétique de la Figure 2.

1. Télécharger le fichier `neighbor_globin.txt` et importer les données dans R dans un `data.frame` `d`. Elles contiennent les scores d'alignement deux à deux de diverses globines chez différentes espèces tel que décrit dans le fichier `Globines_liste.txt`.
2. Vérifier que ces scores correspondent bien à des dissimilarités. Nommer les colonnes.
3. Calculer la matrice Δ des carrés des dissimilarités.

