

Examen (vendredi 28 avril 2023)

Exercice n°1 : Analyse de données

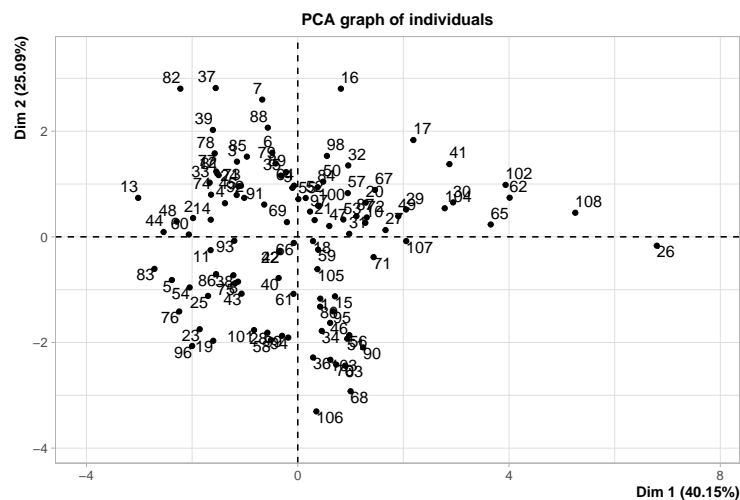
On s'intéresse à des données expérimentales mesurées chez les poules. Les données sont contenues dans le fichier `exercice1_28avril2023.txt`.

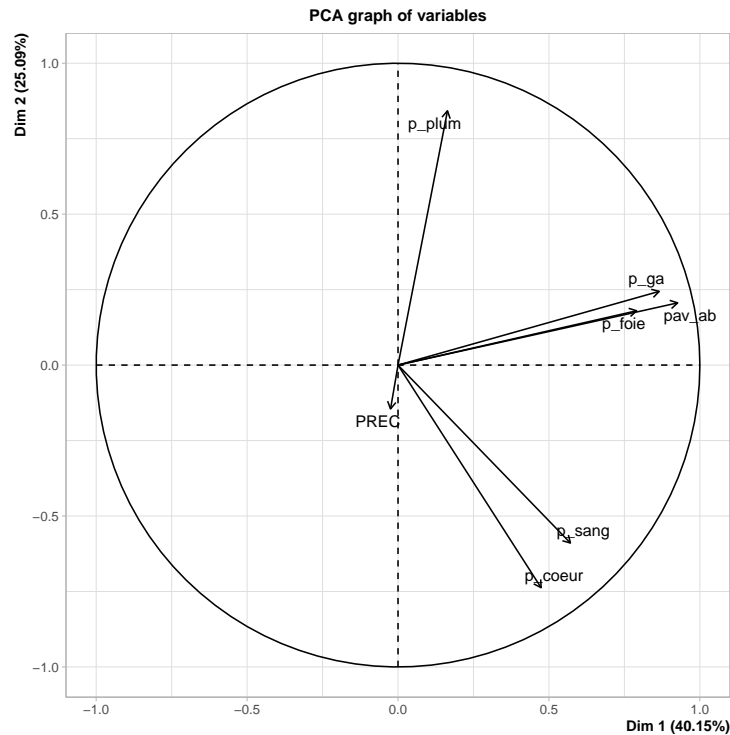
- La variable `PREC` indique la date de la première ponte.
- La variable `pav_ab` indique le poids avant abattage.
- La variable `p_ga` indique le poids du gras abdominal.
- Les variables `p_coeur`, `p_foie`, `p_sang`, `p_plum` indiquent le poids du coeur, du foie, du sang et des plumes.

```
library(FactoMineR)
load("exercice1_28avril2023.RData")
dim(data)

## [1] 108    7

PCA(data)
```





```
## **Results for the Principal Component Analysis (PCA)**
## The analysis was performed on 108 individuals, described by 7 variables
## *The results are available in the following objects:
##
##   name                description
## 1  "$eig"              "eigenvalues"
## 2  "$var"              "results for the variables"
## 3  "$var$coord"        "coord. for the variables"
## 4  "$var$cor"          "correlations variables - dimensions"
## 5  "$var$cos2"         "cos2 for the variables"
## 6  "$var$contrib"      "contributions of the variables"
## 7  "$ind"              "results for the individuals"
## 8  "$ind$coord"        "coord. for the individuals"
## 9  "$ind$cos2"         "cos2 for the individuals"
## 10 "$ind$contrib"      "contributions of the individuals"
## 11 "$call"             "summary statistics"
## 12 "$call$centre"      "mean of the variables"
## 13 "$call$ecart.type"  "standard error of the variables"
## 14 "$call$row.w"       "weights for the individuals"
## 15 "$call$col.w"       "weights for the variables"
```

- Décrire et commenter les deux graphiques obtenus sur les données.
- Sur le graphique des individus (premier graphique), que présentent les axes 1 et 2? (NB : les réponses détaillées utilisant un formalisme matriciel seront appréciées).
- En utilisant la fonction `eigen` (sans utiliser les sorties de la fonction `PCA`), retrouver les coordonnées du point d'indice 26. Ces coordonnées sont-elles cohérentes avec le graphique des individus représentés plus haut?

Exercice n°2 : Analyse de données chez *C. elegans*

On s'intéresse aux mécanismes cellulaires et moléculaires sous-jacents à l'élimination sélective des composants cellulaires par autophagie dans le cadre des processus de développement, du vieillissement et des réponses adaptatives à divers stress. *Caenorhabditis elegans* (*C. elegans* en abrégé) est un petit ver d'un millimètre environ, transparent et non parasitaire. *C. elegans* est un bon modèle pour analyser les processus autophagiques. Parce que la voie autophagique est hautement conservée chez les eucaryotes, les données obtenues chez la levure et le ver sont essentielles pour comprendre le rôle de l'autophagie chez l'homme.

Afin de comprendre l'impact de la température sur la croissance des organismes *C. elegans*, un laboratoire met au point une expérience. Ils mesurent la taille des vers soumis à température normale (20 degrés) et également la taille des vers soumis à température de 37 degrés. Les mesures sont effectuées par 4 techniciens différents (indiqués par les lettres A, B, C et D dans le jeu de données), le même jour et dans des conditions expérimentales identiques. Les longueurs de vers mesurées sont stockées dans la variable `length`.

Les données sont contenues dans le fichier `data_labo2_Examen28avril.txt`.

- (a) Les mesures effectuées sont-elles dépendantes du technicien qui a effectué ces mesures? L'impact du technicien sur les mesures est-elle significative?
- (b) Les mesures effectuées sont-elles dépendantes de la température à laquelle sont soumis les vers? L'impact de la température sur la taille des vers est-elle significative?
- (c) On s'intéresse maintenant à la taille des vers soumis à une température normale (20 degrés). Proposer une modélisation probabiliste de ces données en justifiant votre choix.
- (d) On s'intéresse maintenant à la taille des vers soumis à une température normale (37 degrés). Proposer une modélisation probabiliste de ces données en justifiant votre choix.
- (e) On cherche à distinguer plusieurs "sous-populations" d'individus parmi les vers soumis à 37 degrés. Proposer une méthode pour détecter ces sous-populations en prenant soin de bien justifier votre approche. Il convient de détailler le modèle utilisé, le choix de paramètres de ce modèle, ainsi que d'indiquer les commandes R utilisées.