

# Université Claude Bernard - Lyon 1

---

**MADG**                      **Vendredi 6 décembre 2019**

---

P<sup>r</sup> Jean R. LOBRY  
Tous documents autorisés

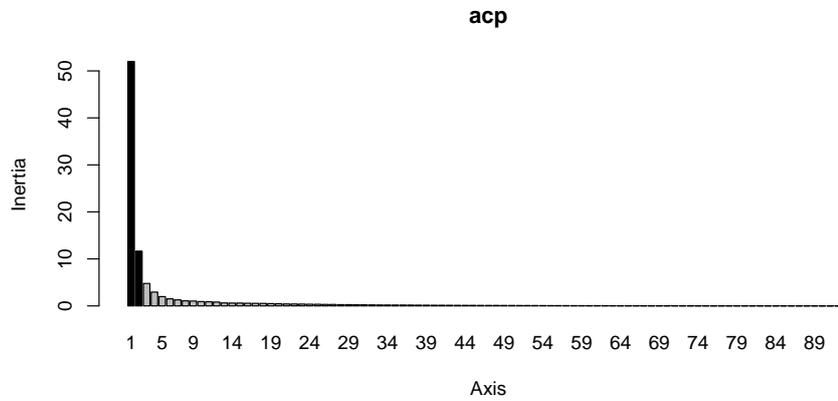
**NOM Prénom** .....

IL s'agit ici de données réelles issues de la collecte de 93 variables morphométriques sur 6068 soldats américains [2, 1]. Pensez-vous que cet échantillon soit représentatif de l'ensemble de la population américaine ? Justifiez votre réponse, si possible avec un exemple concret.

**Réponse :**

ON décide d'utiliser une analyse en composantes principales centrée-réduite pour résumer ce jeu de données.

```
load("ansur.Rda")
library(ade4)
dta <- ansur[, 2:94] # sélection des 93 variables morphométriques
acp <- dudi.pca(dta, center = TRUE, scale = TRUE, scannf = FALSE)
screeplot(acp)
```



AU vu du graphe des valeurs propres, combien de facteurs aimeriez vous retenir ?

**Réponse :**

```
summary(acp)
Class: pca dudi
Call: dudi.pca(df = dta, center = TRUE, scale = TRUE, scannf = FALSE)
Total inertia: 93

Eigenvalues:
  Ax1    Ax2    Ax3    Ax4    Ax5
52.010 11.646  4.774  2.939  1.976

Projected inertia (%):
  Ax1    Ax2    Ax3    Ax4    Ax5
55.925 12.523  5.133  3.160  2.125

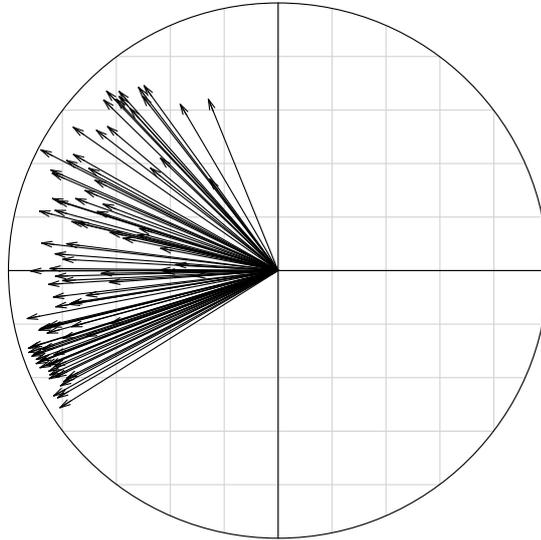
Cumulative projected inertia (%):
  Ax1  Ax1:2  Ax1:3  Ax1:4  Ax1:5
55.92  68.45  73.58  76.74  78.87

(Only 5 dimensions (out of 93) are shown)
```

QUEL pourcentage de l'inertie totale avez-vous conservé avec le nombre de facteurs retenus ?

**Réponse :**

```
s.corcircle(acp$co, clabel = 0)
```



QUEL est le terme technique pour décrire le phénomène visible sur le premier facteur ?

Réponse :

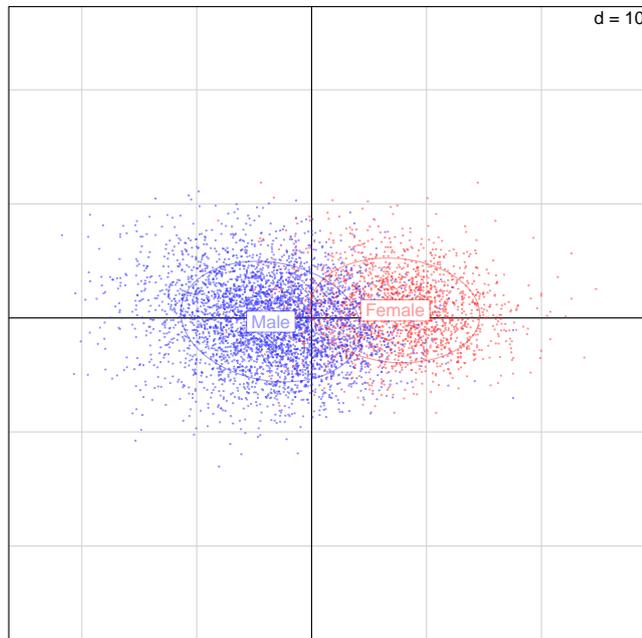
LES flèches des vecteurs ne devraient-elles pas logiquement être orientées dans l'autre sens ?

Réponse :

QUELLE interprétation biologique pourriez-vous avancer pour expliquer le premier facteur ?

Réponse :

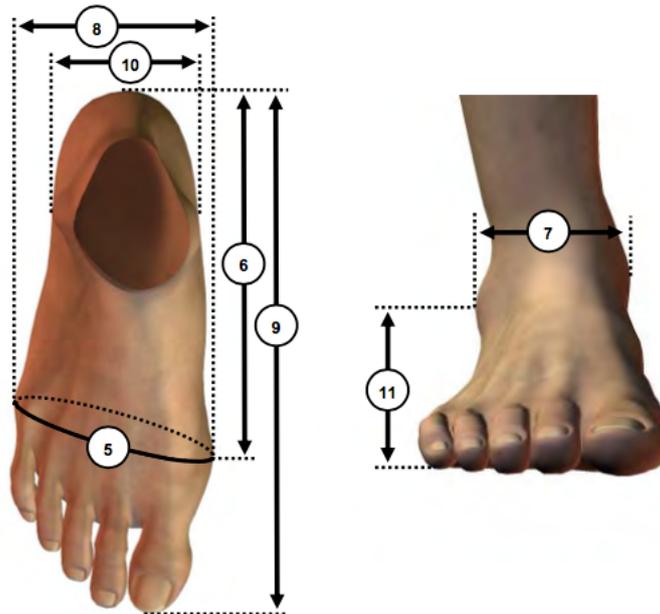
```
colM <- rgb(0.2, 0.2, 1, 0.5) ; colF <- rgb(1, 0.2, 0.2, 0.5)
mycols <- ifelse(ansur$Gender == "Male", colM, colF)
s.class(acp$li, ansur$Gender, cstar = 0, axesell = FALSE, cpoint = 0.2, col = c(colF, colM))
```



**E**XPLIQUEZ pourquoi les hommes (**Male**) sont séparés des femmes (**Female**) sur le premier facteur.

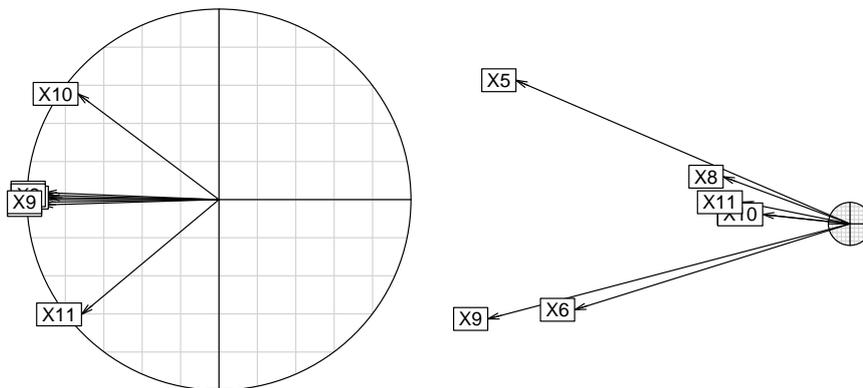
**Réponse :**

**E**N tant que fabricant de chaussures, vous êtes surtout intéressé par les mensurations relatives aux pieds. Votre objectif est de vous démarquer de la concurrence en proposant non seulement des modèles de différentes *tailles* (les pointures) mais également de différentes *formes*. Les variables disponibles sont représentées dans le schéma ci-après.



ON extrait les variables correspondant au schéma ci-avant puis on réalise une ACP centrée-réduite, à gauche, et une ACP centrée-non-réduite, à droite.

```
sel <- c("balloffootcircumference", "balloffootlength", "bimalleolarbreadth",
        "footbreadthhorizontal", "footlength", "heelbreadth",
        "lateralmalleolusheight")
isel <- which(colnames(ansur) %in% sel)
pied <- ansur[, isel]
colnames(pied) <- paste("X", 4 + (1:7), sep = "")
acpcr <- dudi.pca(pied, scannf = FALSE, scale = TRUE)
acpcnr <- dudi.pca(pied, scannf = FALSE, scale = FALSE)
par(mfrow = c(1, 2), xpd = NA)
s.corcircle(acpcr$co)
s.corcircle(acpcnr$co)
```



POUR optimiser votre production de chaussures, pensez-vous qu'il soit préférable de travailler avec des données centrées-réduites ou bien centrées-non-réduites? Justifiez votre réponse.

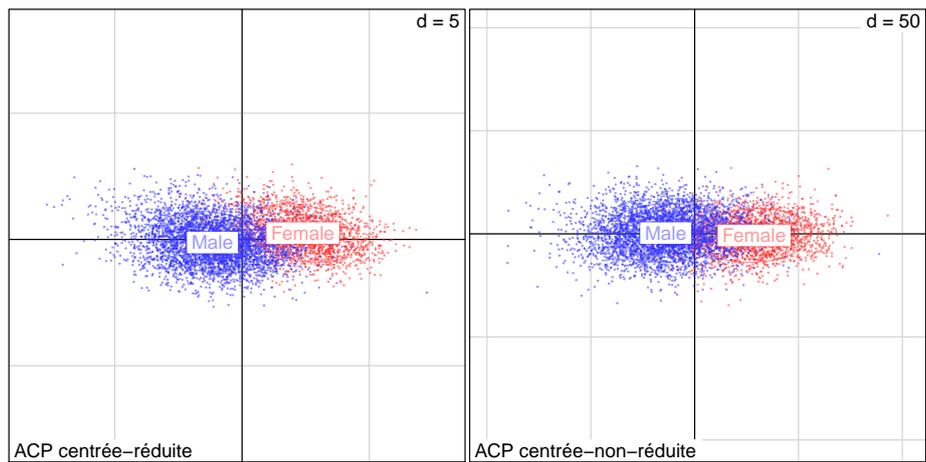
Réponse :

POUR chaque pointure vous décidez de proposer deux modèles de forme différente. Décrivez en quoi ces deux modèles diffèrent.

Réponse :

ON examine la position des individus sur le premier plan factoriel en portant comme information supplémentaire le sexe des individus.

```
par(mfrow = c(1, 2))
s.class(acpqr$li, ansur$Gender, cstar = 0, axesell = FALSE, cpoint = 0.2,
        col = c(colF, colM), sub = "ACP centrée-réduite")
s.class(acpcnr$li, ansur$Gender, cstar = 0, axesell = FALSE, cpoint = 0.2,
        col = c(colF, colM), sub = "ACP centrée-non-réduite")
```



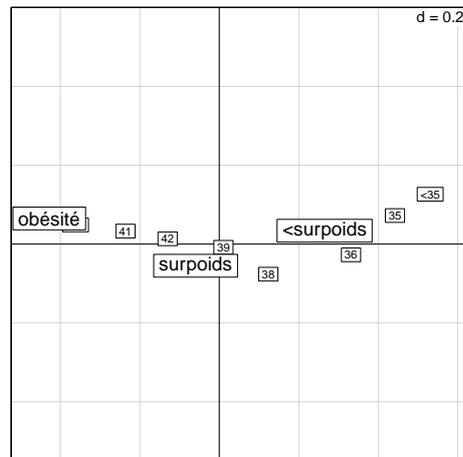
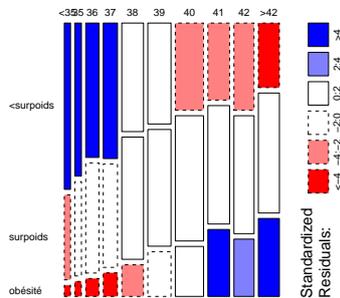
PENSEZ-VOUS qu'il serait intéressant de proposer un modèle de chaussure pour les hommes et un modèle pour les femmes ou bien qu'un modèle unisexe est suffisant ?

Réponse :

ON s'intéresse aux éventuelles correspondances entre les tailles de chaussures et l'indice de masse corporelle. La taille de chaussure est la longueur du pied exprimée en fractions de  $\frac{2}{3}$  de centimètres, on a choisi de regrouper les classes extrêmes (<35 et >42). L'indice de masse corporelle est la masse (kg) divisée la taille (m) au carré. Les bornes de l'OMS permettent de définir après regroupement les trois classes suivantes : <surpoids, surpoids et obésité.

```
ansur$pointure <- as.factor(round((ansur$footlength/10)*3/2))
levels(ansur$pointure) <- c(rep("<35", 4), 35:42, rep(">42", 6))
masse <- ansur$weightkg/10
taille <- ansur$Heightin*2.54/100
ansur$imc <- cut(masse/taille^2, breaks = c(0, 25, 30, 100))
levels(ansur$imc) <- c("<surpoids", "surpoids", "obésité")
tcont <- with(ansur, as.data.frame.matrix(table(list(imc, pointure))))
tcont.coa <- dudi.coa(t(tcont), scannf = FALSE)
par(mfrow = c(1, 2))
mosaicplot(t(tcont), shade = TRUE, main = "", las = 1)
scatter(tcont.coa, posieig = "none")
```

NULL



DONNEZ une interprétation au premier facteur de l'analyse factorielle des correspondances.

Réponse :

## Références

- [1] GORDON, C., BLACKWELL, C., BRADTMILLER, B., PARHAM, J., BARRIENTOS, P., PAQUETTE, S., CORNER, B., CARSON, J., VENEZIA, J., ROCKWELL, B., MUCHER, M., AND KRISTENSEN, S. 2012 anthropometric survey of U.S. army personnel : methods and summary statistics. Tech. rep., U.S. Army Natick Soldier Research, Development and Engineering Center Natick, Massachusetts 01760-2642, U.S.A., 2012.
- [2] HOTZMAN, J., GORDON, C., BRADTMILLER, B., CORNER, B., MUCHER, M., KRISTENSEN, S. AND PAQUETTE, S., AND BLACKWELL, C. Measurer's handbook : US army and marine corps anthropometric surveys, 2010-2011. Tech. rep., U.S. Army Natick Soldier Research, Development and Engineering Center Natick, Massachusetts 01760-2642, U.S.A., 2011.