

Problème pratique de statistique n° pps001
Discrimination et kystes de l'ovaire

D. Chessel & J.R. Lobry

1 Source des données

Les données sont issues d'une conférence (Lasne, Y., Akiki, S., Dargent, D. (1988) Dosages des marqueurs biochimiques dans les kystes de l'ovaire. Communication au 7ème Colloque International de Biologie Prospective. Pont-à-Mousson.) et de la référence [1].

2 La question posée

La question à l'origine est posée par le Pr. D. Dargent au Dr. Y. Lasne : y a-t-il un ou des marqueurs biochimiques capables de rendre compte de la fonctionnalité, de l'organicité et au sein de celle-ci de la bénignité ou de la malignité d'un kyste ovarien ? Des patientes présentant un kyste de l'ovaire constituent un échantillon de référence pour comparer diagnostic histologique et diagnostic biochimique, l'intérêt manifeste du second étant d'éviter une intervention chirurgicale qui ne s'imposerait pas.

A la mémoire de Y. Lasne, tragiquement disparu en février 2001, nous conservons une partie des données pour en faire un exercice de statistique multivariée. Les recherches en la matière ont considérablement progressé : ces données constituent un témoignage de questions posées il y a une quinzaine d'années.

3 Lecture des données

```
kyo <- read.table("http://pbil.univ-lyon1.fr/R/donnees/pps001.txt",
  header = TRUE, sep = "\t")
dim(kyo)
[1] 121 11
names(kyo)
[1] "cla"      "aceser"   "ca19ser"  "ca125ser" "oeaser"   "proser"   "acekys"
[8] "ca19kys"  "ca125kys" "oekys"    "prokys"
```

Il a 121 lignes et 11 colonnes. Les 121 lignes sont 121 patientes ayant subi une ablation d'un kyste de l'ovaire. Avant l'opération une ponction de liquide du kyste est effectuée simultanément avec une prise de sang. Dans le sérum et le liquide kystique sont dosées 5 molécules, l'antigène carcino-embryonnaire (ace en ng/ml), les antigènes carbohydrates CA 19-9 (ca19 en ui/ml) et CA 125 (ca125 ui/ml) et les hormones ovariennes l'œstradiol (oea en ng/ml) et la progestérone (pro en ng/ml). On obtient ainsi 10 variables notée ser pour sérum et kys pour kyste :

```
aceser  ca19ser  ca125ser  oeaser  proser
acekys  ca19kys  ca125kys  oeakys  prokys
```

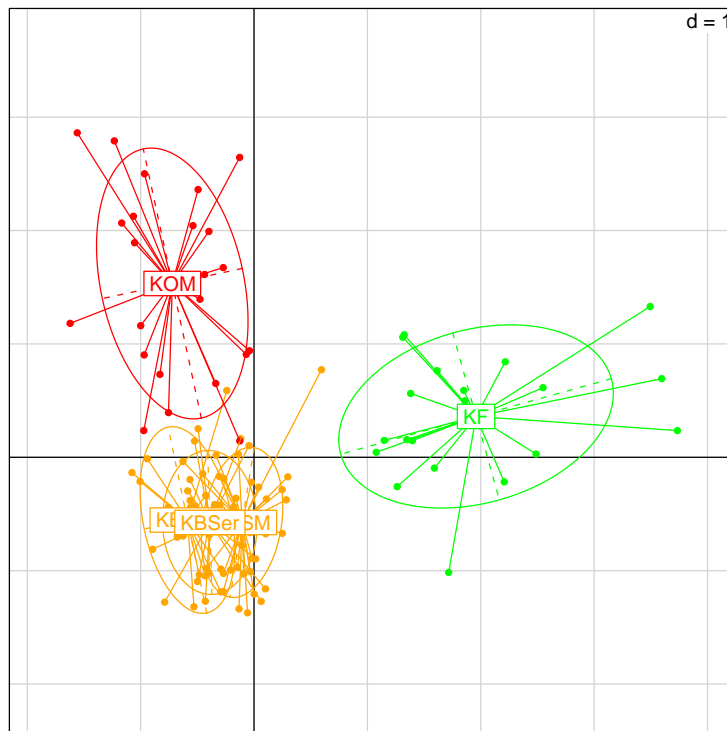
Les données ont été transformées par $x \rightarrow \log(x+1)$ et arrondies au centième le plus proche. Le tableau a en outre une première colonne qui est une variable qualitative et répartit les kystes en 5 classes :

```
table(kyo$cla)
KBEnd KBSM KBSer  KF  KOM
  12   37   28   20   24
```

KF désigne les kystes fonctionnels. Les autres sont organiques soit malins (KOM) soit bénins. Ces derniers sont endométrioïdes (KBEnd), séro-mucineux (KBSM) ou séreux (KBSer). Le but de l'analyse est de caractériser le lien entre dosages biologiques et cette classification.

4 Illustration

```
library(ade4)
acp <- dudi.pca(kyo[, 2:11], scann = FALSE, nf = 2)
dis <- discrimin(acp, kyo$cla, scann = FALSE, nf = 2)
s.class(dis$li, kyo$cla, col = c("orange", "orange", "orange", "green",
"red"))
```



Références

- [1] D. Dargent, Y. Lasne, and S. Akiki. Diagnostic de la nature des kystes de l'ovaire par le dosage dans le liquide kystique et dans le sang des stéroïdes et des "marqueurs tumoraux". déductions pratiques. *Contraception-fertilité-sexualité*, 18 :1011–1016, 1990.