

Bio-statistiques 2 / Contrôle Continu

Bio-statistiques 2 / M1-AMIV / 2006 / BI6003M1 CC / D. Chessel

22 mars 2006 - 9h-11 h

Tous les documents sont autorisés. Les échanges entre étudiants sont strictement interdits. Vous devez utiliser votre micro-ordinateur personnel. \mathbb{R} et `ade4` sont recommandés. L'accès à internet est inutile. Répondre directement sur la feuille en laissant les agrafes.

1 Pratique de \mathbb{R}

1.1 Que faut-il taper pour obtenir ce qui suit ?

```
[1] 1 2 3 4 5 6 7 8 9 10 11 12
```

```
[1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10" "11" "12"
```

```
[1] "A" "B" "C" "D" "E" "F" "G" "H" "I" "J" "K" "L"
```

```
[1] 1.00000000 0.50000000 0.33333333 0.25000000 0.20000000 0.16666667 0.14285714  
[8] 0.12500000 0.11111111 0.10000000 0.09090909 0.08333333
```

1.2 Que faut-il taper pour obtenir ce qui suit ?

```
[1] "a" "a" "a" "b" "b" "b" "b" "c" "c"
```

```
[1] a a a b b b b c c  
Levels: a b c
```

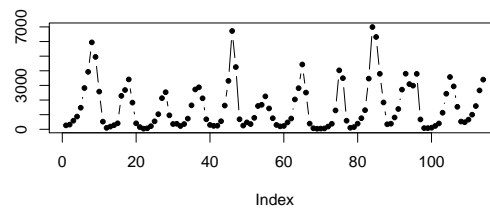
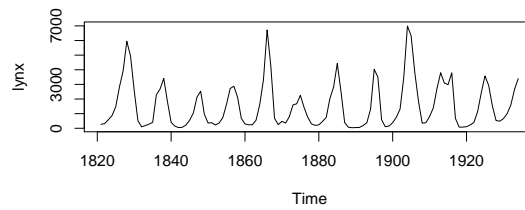
```
[1] a a a b b b b c c  
Levels: a < b < c
```

1.3 dpqr

Quelle est la valeur x telle que $P(X < x) = 0.05$ si X suit une loi du χ^2 à 22 degrés de liberté ?

1.4 Faire une figure

Après `data(lynx)`, trois lignes sont nécessaires pour refaire cette figure. Lesquelles ?



1.5 Peut-on parler de ségrégation sexuelle significative ?

Dans cette population où 8 individus sont répartis en 4 groupes de la manière suivante :

Femelles	4	1	0	0
Mâles	0	0	1	2

1.6 A quelle condition le résultat suivant est-il possible ?

```
sort(sample(c("a", "b", 12.7, "c", n0, "aaa", "bb"), 7, replace = F))
```

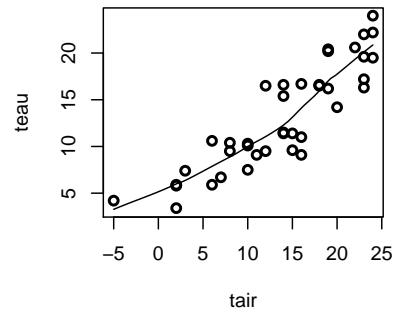
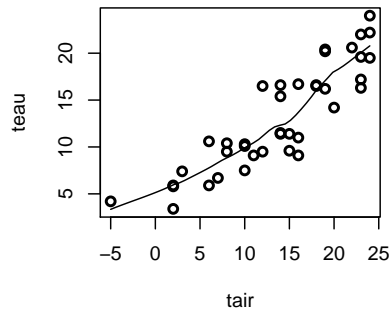
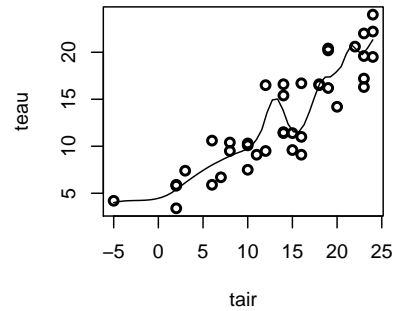
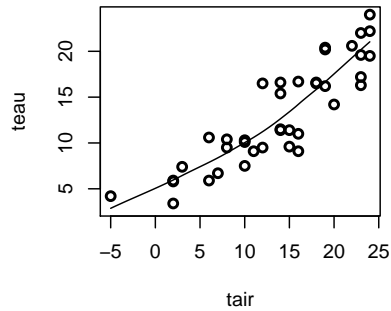
```
[1] "12.7" "a" "aaa" "b" "b" "bb" "c"
```

1.7 Lire une fiche de documentation

`teau` est la température de l'eau et `tair` est la température de l'air à 39 dates différentes.

```
library(ade4)
data(rhone)
tair <- rhone$tab[, 1]
teau <- rhone$tab[, 2]
```

On regarde le lien entre les deux variables avec la fonction `scatter.smooth`, qu'on utilise quatre fois avec les valeurs 0.25, 0.45, 0.55 et 0.75. Trouvez de quel paramètre il s'agit et inscrivez clairement dans chaque fenêtre la valeur qui est utilisée.



1.8 Pourquoi utilise-t-on la fonction `set.seed` dans cet exercice et les suivantes ?

```
set.seed(22032006)
runif(8)
```

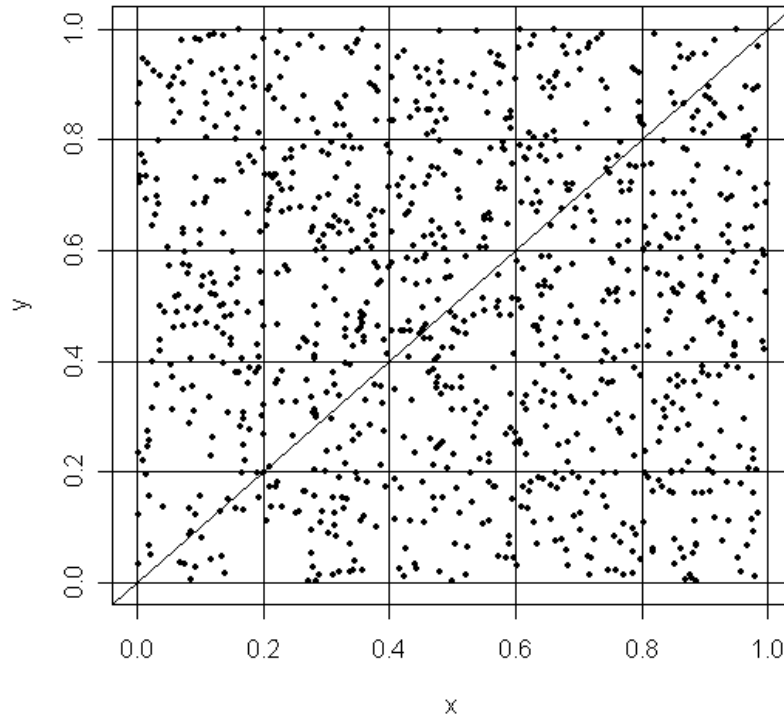
```
[1] 0.7820305 0.7360664 0.1498964 0.8437935 0.9094233 0.8828020 0.1316912 0.4611798
```

```
set.seed(22032006)
runif(8)
```

```
[1] 0.7820305 0.7360664 0.1498964 0.8437935 0.9094233 0.8828020 0.1316912 0.4611798
```

```
set.seed(22032006)
x <- runif(1000)
y <- runif(1000)
plot(x, y, pch = 20)
abline(c(0, 1))
abline(v = seq(0, 1, by = 0.2), h = seq(0, 1, by = 0.2))
```

1.9 Combien y-a-t-il de points en dessous de la droite $y = x$?



1.10 Compter les points

Indiquer directement sur la figure combien il y a de points dans chacun des carrés de la grille.

2 ACP et clémentines

2.1 Introduction

```
library(ade4)
data(clementines)
dim(clementines)
```

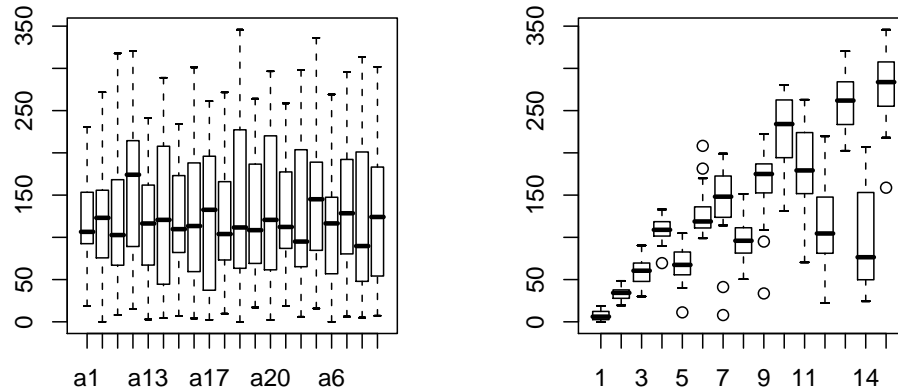
[1] 15 20

L'objet `clementines` est un data frame avec 15 lignes et 20 colonnes. Les 20 colonnes sont des arbres (clémentiniers) et les 15 lignes sont des années successives. Les valeurs du tableau donnent pour chaque année (ligne) et chaque arbre (colonne) la production de fruits totale. Il s'agit d'un extrait à vocation pédagogique proposé par D. Tisne-Agostini à partir d'un vaste essai porte-greffe de la Station de Recherches Agronomiques de San Giuliano (<http://www.corse.inra.fr/sra/sra.htm>).

Appeler `anlig` une copie de `clementines` et `anco1` une copie de son transposé. Dans `anlig` on a 15 lignes-années et 20 colonnes-arbres, dans `anco1` on a 20 lignes-arbres et 15 colonnes-années.

Appeler `tot` le vecteur qui déroule les 20 chroniques individuelles les unes sous les autres (`as.numeric(as.matrix(anlig))`), `anfac` le facteur associé à l'année (`factor(rep(1:15,20))`) et `arbfac` le facteur associé à l'arbre (`factor(rep(names(clementines),rep(15,20)))`).

Comment est obtenue la figure ci-dessous et que signifie-t-elle ?

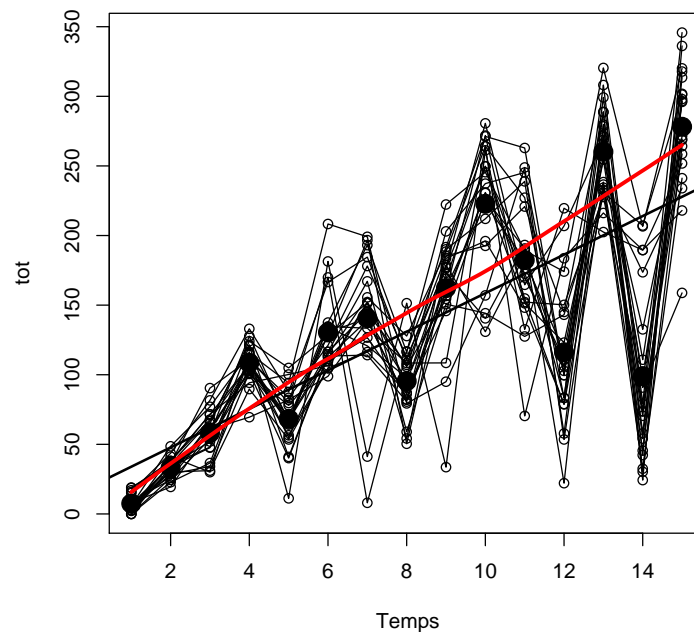


2.2 Tests

Que peut-on faire pour prouver que la variabilité de la production entre arbres est négligeable par rapport à la variabilité de la production entre années ?

2.3 Les années

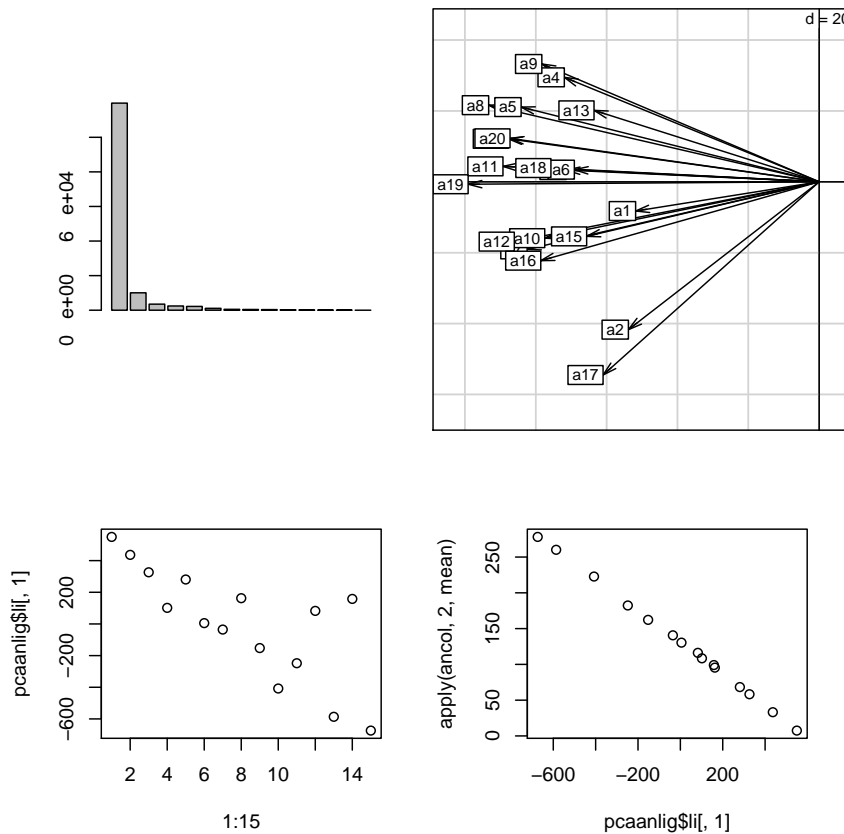
```
plot(as.numeric(anfac), tot, xlab = "Temps")
invisible(lapply(anlig, function(x) lines(1:15, x)))
points(1:15, tapply(tot, anfac, mean), cex = 3, pch = 20, type = "b")
abline(lm(tot ~ as.numeric(anfac)), lwd = 2)
lines(loess.smooth(as.numeric(anfac), tot), col = "red", lwd = 3)
anova(lm(tot ~ as.numeric(anfac)), lm(tot ~ anfac))
```



Que pensez-vous de l'évolution dans le temps de la production de fruit ?

2.4 ACP de anlig

```
pcaanlig <- dudi.pca(anlig, scan = F, nf = 2, scale = F)
par(mfrow = c(2, 2))
barplot(pcaanlig$eig)
s.arrow(pcaanlig$co)
plot(1:15, pcaanlig$li[, 1], pch = 21)
plot(pcaanlig$li[, 1], apply(ancol, 2, mean), pch = 21)
```



On a fait l'ACP du tableau arbre en colonnes. Faites l'interprétation de cette analyse et rédigez une légende pour cette figure.

2.5 Production et coordonnée

Que vaut la corrélation entre la production totale d'une année et la première coordonnée des lignes de l'ACP de anlig ?

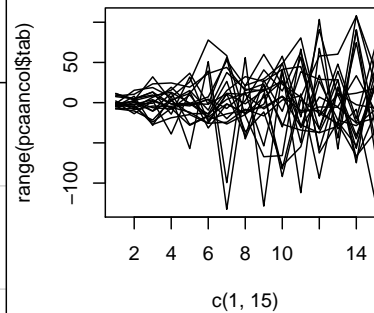
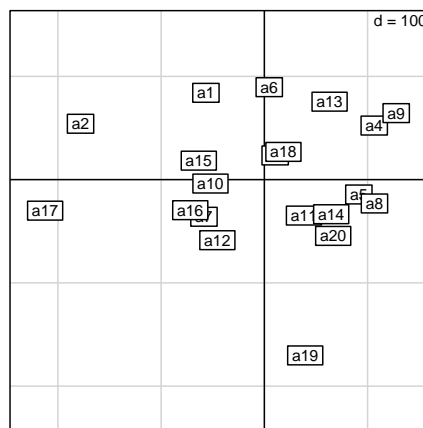
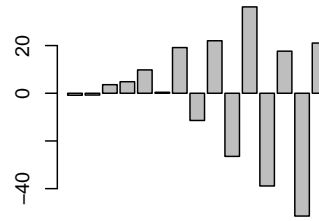
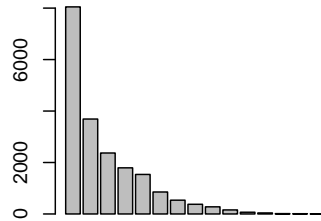
2.6 Commentaire

Que pensez-vous de l'assertion : *l'analyse qui précède ne nous a rien appris ?*

2.7 ACP de ancol

```
pcaancol <- dudi.pca(ancol, scan = F, nf = 2, scale = F)
par(mfrow = c(2, 2))
barplot(pcaancol$eig)
barplot(pcaancol$co[, 1])
s.label(pcaancol$li)
plot(c(1, 15), range(pcaancol$tab), type = "n")
apply(pcaancol$tab, 1, function(x) lines(1:15, x))
```

On a fait l'ACP du tableau arbre en lignes. Faites l'interprétation de cette analyse et rédiger une légende pour cette figure.



2.8 Il faut ou il ne faut pas

Faut-il faire une analyse plutôt que l'autre ou les deux ou aucune ? A-t-on appris quelque chose qu'on ne voyait pas ? Où, quand, quoi ? Peut-on lui donner un nom ?

2.9 prcomp et princomp

Peut-on faire ces deux analyses avec la fonction `prcomp` ? Peut-on faire ces deux analyses avec la fonction `princomp` ? Un commentaire ?

2.10 ACP centrée et ACP normée

Quelle est la corrélation entre la première coordonnée de l'ACP centrée de `anlig` et la première coordonnée de l'ACP normée de `anlig` ? C'est normal ? Il en est toujours ainsi ou est-ce un caractère particulier de la situation ?

3 ACP et géométrie euclidienne

Ici on démontre ce qu'on avance ou on passe.

- 3.1 Vrai ou faux ? La trace d'une matrice de corrélation entre p variables vaut 1.
- 3.2 Vrai ou faux ? Le produit des éléments de la diagonale d'une matrice de corrélation entre p variables vaut 1.
- 3.3 Vrai ou faux ? Une matrice de corrélation entre p variables est toujours diagonalisable.
- 3.4 Vrai ou faux ? Une matrice de corrélation entre p variables est toujours inversible.
- 3.5 Vrai ou faux ? Une matrice de corrélation entre p variables n'a jamais de valeurs propres négatives.
- 3.6 Vrai ou faux ? Une matrice de corrélation entre p variables est toujours la matrice d'un produit scalaire de \mathbb{R}^p .
- 3.7 Vrai ou faux ? Cette matrice peut être la matrice de corrélation entre 3 variables

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix}$$

3.8

Soit a strictement compris entre 0 et 1. Quelle est la plus grande valeur propre de la matrice \mathbf{B} ?

$$\mathbf{B} = \begin{bmatrix} 1 & a & a & a \\ a & 1 & a & a \\ a & a & 1 & a \\ a & a & a & 1 \end{bmatrix}$$

Indiquer ici clairement votre numéro d'anonymat

1.1

1.2

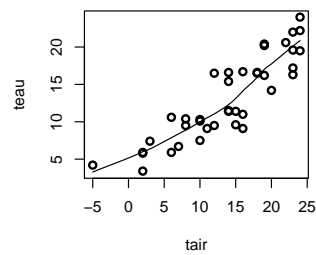
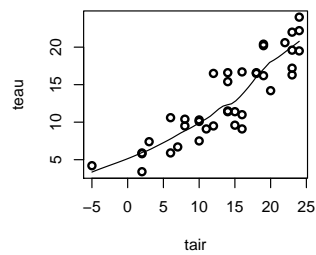
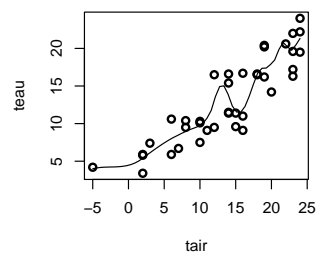
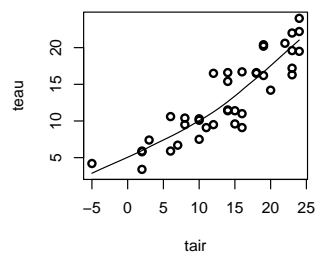
1.3 Indiquez comment vous la trouvez.

1.4

1.5

1.6

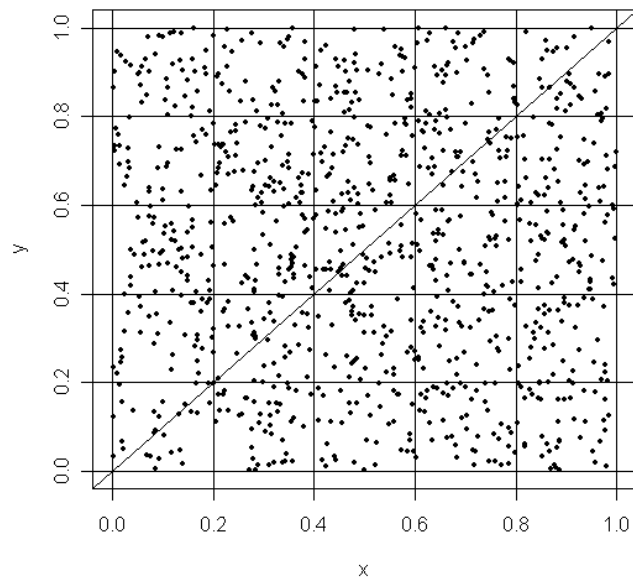
1.7



1.8

1.9

1.10



2.1

2.2

2.3

2.4

2.5

2.6

2.7

2.8

2.9

2.10

3.1

3.2

3.3

3.4

3.5

3.6

3.7

3.8

Place pour un commentaire personnel :