

**Analyse des Données Ecologiques Multidimensionnelles**  
- EXAMEN Mai 2010 -**NOM** .....**Prénom** .....**N° Etudiant** .....*Durée : 2h00**- Tous documents autorisés -***Répondre dans le cadre prévu pour chaque question.****Analyse en Composantes Principales (ACP) et Analyse Factorielle des Correspondances (AFC)**

Ayant suivi le module ELB cette année, vous êtes contacté par un collègue qui vous envoie le message suivant :

*Cher-e collègue,*

*Dans le cadre de mes travaux, j'analyse un jeu de données décrivant l'abondance de 25 espèces d'oiseaux mesurée sur 50 sites. Je cherche à décrire la structure de ces communautés, et j'ai réalisé deux analyses sur ce tableau : (1) une ACP centrée et (2) une AFC. Le problème est que j'obtiens des résultats très différents, et ne connaissant pas bien ces méthodes, j'avoue avoir du mal à interpréter mes résultats... Il semble que l'ACP mette surtout en évidence les espèces abondantes, alors que l'AFC fait ressortir essentiellement les espèces rares... est-ce normal ?*

*Merci d'avance pour vos conseils,  
cordialement, X.*

**1.** Que lui répondez-vous ?**2. Vrai ou Faux ?** (justifier) : L'ACP et l'AFC sont des méthodes utilisables sur tous types de données quantitatives.

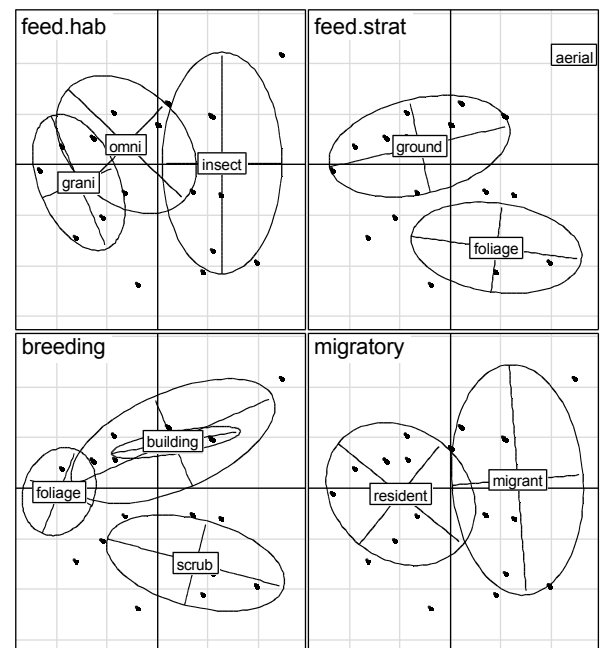
## Analyse des Correspondances Multiples (ACM)

Une étude s'intéresse à 4 caractéristiques écologiques codées en modalités pour 40 espèces d'oiseaux : le régime alimentaire (**feed.hab** : granivores, omnivores ou insectivores), le lieu d'alimentation (**feed.strat** : au sol (ground) ou dans les feuillages (foliage)), le lieu de reproduction (**breeding** : au sol, dans des constructions (building), dans les feuillages ou les buissons (scrub)), et la stratégie de migration (**migratory** : migrateur ou résident).

On réalise une ACM de ce tableau :

```
> acm1 = dudi.acm(oiseaux)
> scatter(acm1)

> acm1$cr
              RS1      RS2
feed.hab    0.7262450 0.04803740
feed.strat  0.4167404 0.73966236
breeding     0.4498672 0.71483143
migratory    0.5973046 0.01250811
```



### 3. Interprétez les résultats de cette analyse.

**4. Vrai ou Faux ?** (justifier) : En Analyse des Correspondances Multiples (ACM), la proximité entre deux modalités de deux variables différentes traduit toujours une association positive entre ces deux modalités.

## Analyse Factorielle Discriminante (AFD)

**5. Vrai ou Faux ?** (justifier) : Dans une Analyse Factorielle Discriminante, le nombre de valeurs propres est égal au nombre de groupes étudiés.

On appelle `afd1` l'Analyse Factorielle Discriminante d'un tableau X (150 individus, 3 variables X1,

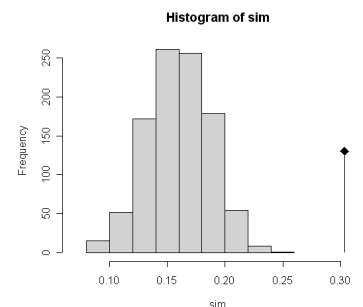
X2 et X3) et de 4 groupes (variable « groupe » en 4 modalités).

```
> afd1= lda(groupe~X1+X2+X3, CV=TRUE)
> table(groupe, afd1$class)
```

Predictions	Observations			
	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	40	5	0	0
Classe 2	9	9	11	1
Classe 3	13	26	6	5
Classe 4	0	3	10	12

**7.** Cette analyse a-t-elle une bonne valeur prédictive (justifier) ?

```
> afd2=discrimin(dudi.pca(X), groupe)
> plot(randtest.discrimin(afd2, 1000))
```



**8.** Quelles informations sont représentées sur cette figure ? Interprétez.

**9. Vrai ou Faux ? (justifier) :** Dans une AFD, lorsque le test de permutation est significatif ( $P < 0.05$ ), on peut conclure que toutes les variables (colonnes) traitées diffèrent entre les groupes.

**10. Vrai ou Faux ? (justifier) :** Dans une AFD, lorsque le test de permutation est significatif ( $P < 0.05$ ), cette analyse permettra d'affecter efficacement un nouvel individu à son groupe d'origine.

## Classification hiérarchique

Dans 12 cours d'eau de quatre bassins hydrographiques, on a récolté 306 truites (*Salmo trutta*) sur lesquelles on a mesuré :

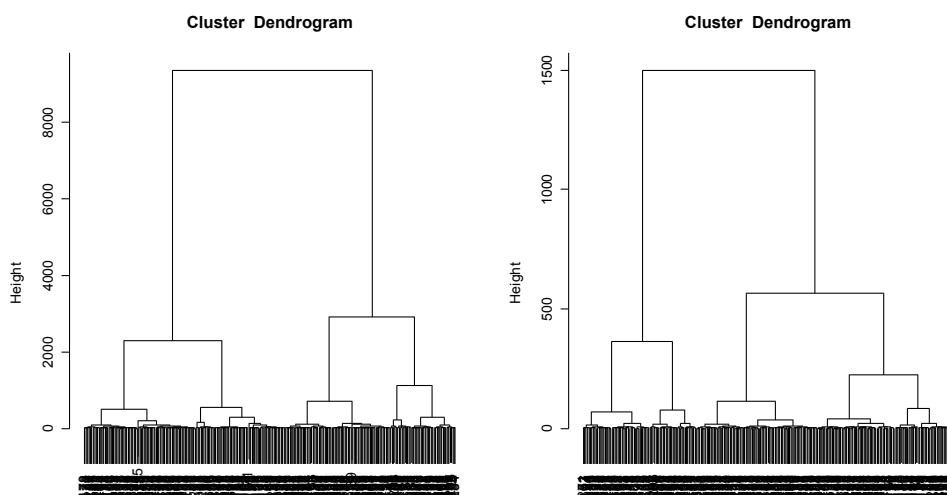
- **5 variables méristiques** (nombres de rayons sur 5 nageoires) : tableau **meris**
- **13 variables morphologiques** (patrons de coloration) : tableau **morpho**

On réalise l'analyse suivante:

```
> distmorpho=dist(morpho)
> distmeris=dist(meris)

> dendromorpho=hclust(dismorpho, method="ward")
> dendromeris=hclust(distmeris, method="ward")

> plot(dendromorpho)           > plot(dendromeris)
```



```

      distmorpho
      hclust ("ward")
> groupemorpho=cutree(dendromorpho, 4)
> groupemeris=cutree(dendromeris, 4)

      distmeris
      hclust ("ward")
> A
      groupemeris
groupemorpho  1  2  3  4
1      23 17 42 10
2      28 15 16 15
3      19 10 22  3
4      29 13 33 11

> chisq.test(A)
X-squared = 15.6231, df = 9, p-value = 0.07518
```

**11.** Interprétez les résultats de cette analyse.

## Analyse Canonique des Correspondances (ACC)

A l'aide de la librairie vegan de R, on réalise une ACC entre un tableau floristique tabflo (contenant les abondances de différentes espèces de plantes sur n sites) et un tableau tabenv décrivant les conditions environnementales sur ces mêmes sites. On s'intéresse plus précisément à l'effet du Calcium (Ca) et du pH sur la composition floristique des sites. Deux analyses sont réalisées :

### • Analyse n°1

```
> cca1=cca(tabflo ~ Ca, tabenv)
> anova(cca1)
```

	Df	Chisq	F	N.Perm	Pr(>F)
Model	1	0.1572	1.7959	3599	0.06028
Residual	22	1.9260			

### • Analyse n°2

```
> cca2=cca(tabflo ~ Ca + Condition(pH), tabenv)
> anova(cca2)
```

	Df	Chisq	F	N.Perm	Pr(>F)
Model	1	0.1827	2.1864	399	0.025
Residual	21	1.7547			

## 12. Interprétez les résultats de ces deux analyses.

## Modèles Additifs Généralisés (GAM)

Pour compléter l'analyse précédente, on s'intéresse à l'effet du Calcium sur le nombre d'espèce (richesse spécifique rs). Deux modèles sont réalisés :

<pre>&gt; mod1 = gam(rs ~ Ca)</pre>						<pre>&gt; mod2 = gam(rs ~ s(Ca,2))</pre>							
<pre>&gt; anova(mod1)</pre>						<pre>&gt; anova(mod2)</pre>							
	Df	Npar	Df	Npar	F	Pr(F)		Df	Npar	Df	Npar	F	Pr(F)
(Intercept)	1						(Intercept)	1					
Ca	1		1	0.8125	0.2564		s(Ca, 2)	1		1	3.3291	0.03282	

## 13. Interprétez les résultats.

### Modèles linéaires

On réalise une expérience agronomique où la teneur en azote du sol est manipulée grâce des apports d'engrais. Suite à cet apport, on mesure la croissance en biomasse de plants de maïs. Cette expérience est réalisée dans des parcelles, chaque parcelle recevant une dose d'engrais pouvant aller, sur une échelle arbitraire, de un à cinq. Pour chaque dose d'engrais, cinq parcelles ont été étudiées.

Les résultats de l'expérience ont été étudiés par le modèle linéaire. L'analyse graphique des résidus est représentée dans la Figure 1.

**14.** A l'aide des graphiques de la Fig.1, concluez quant la qualité du modèle ajusté aux données.

La personne ayant réalisé l'étude avait en fait oublié de vous prévenir que deux variétés différentes de maïs avaient été étudiées dans cette expérience. Pour chaque dose d'engrais les deux variétés ont donc été mesurées dans chaque parcelle. Vous réalisez maintenant une nouvelle analyse prenant en compte à la fois l'apport azoté, le facteur variété et l'interaction entre ces deux facteurs pour expliquer la croissance des plants de maïs.

**15.** Quel nom donne-t-on à ce type d'analyse ?

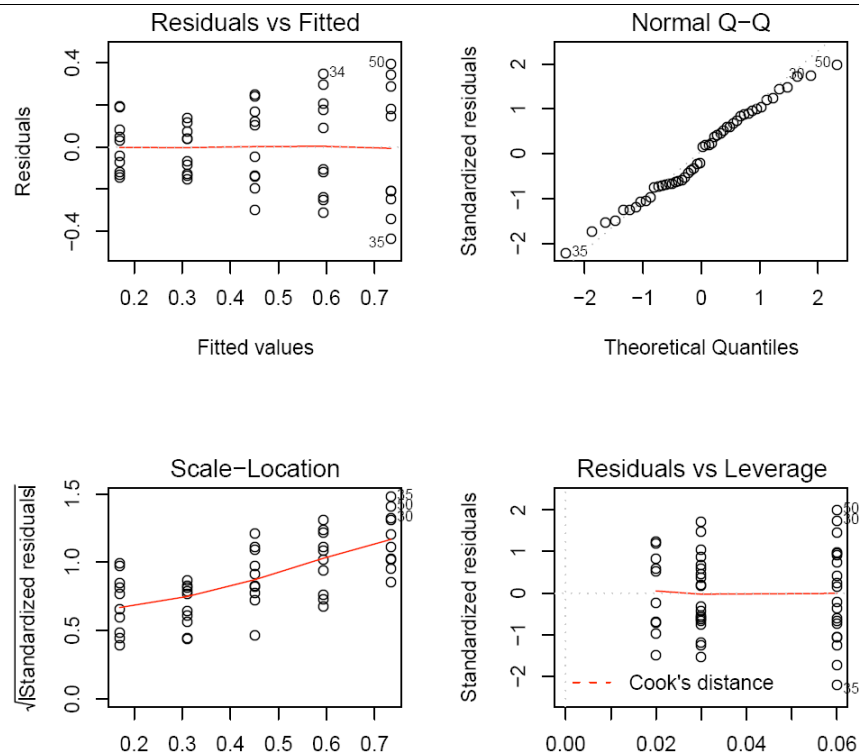
Les résidus de ce nouveau modèle sont présents dans la Figure 2.

**16.** Que pensez-vous de la validité de votre modèle ?

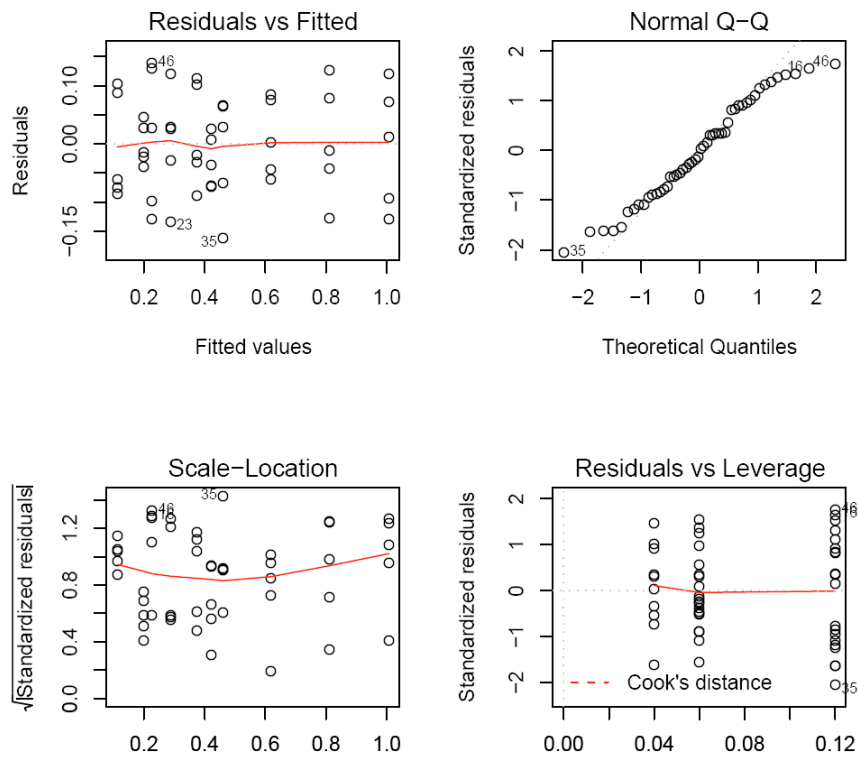
En discutant avec la personne ayant réalisé l'étude, vous vous rendez compte qu'en fait les données que vous avez analysées sont des moyennes réalisées sur vingt placettes à l'intérieur de chaque parcelle et pour chaque variété. La taille totale du jeu de données est donc de  $5 \times 2 \times 5 \times 20$  et non-pas de  $5 \times 2 \times 5$  comme dans le jeu de données que vous avez analysé.

**17.** Pourquoi peut-on dire qu'analyser les moyennes par placette au lieu des données brutes est une mauvaise idée ?

**18.** En quoi, malgré tout, cette façon de procéder permet-elle de régler un problème statistique sérieux ? Connaissez-vous d'autres façon de procéder pour régler le même problème ?



**Fig. 1 :** Analyse des résidus du premier modèle.



**Fig. 2** : Analyse des résidus du deuxième modèle.