

Université Claude Bernard - Lyon 1

M1 - EEME

Année 2005-2006

Analyse et Modélisation des Systèmes Biologiques Session 1

Mardi 6 juin 2006

Durée : 1 heure

A.B. Dufour

Tous les documents sont autorisés.

1 Lacs de l'Etat du Wisconsin

Des limnologues ont déterminé le pH de l'eau (pHW) et le pourcentage de matière organique des sédiments (matorgW) de 21 lacs de l'état de Wisconsin.

- 1) On donne la représentation graphique du pourcentage de matière organique en fonction du pH ainsi que les coefficients de corrélation et de détermination. Commenter.

```
summary(pHW)
```

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
5.600  6.400   7.300   7.162  7.800   8.400
```

```
summary(matorgW)
```

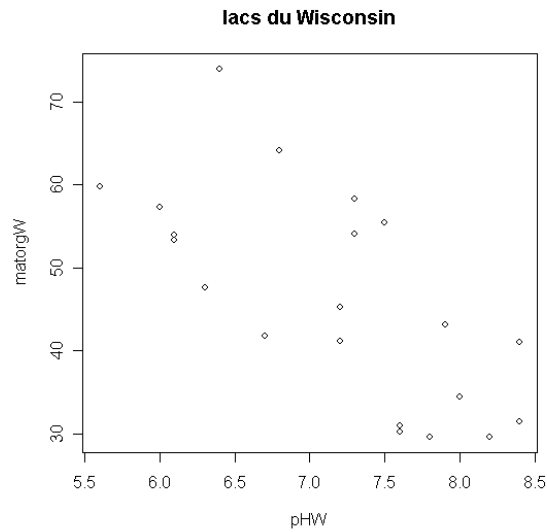
```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
29.57  34.50   45.26   46.52  55.44   74.00
```

```
plot(pHW, matorgW, main = "lacs du Wisconsin", pch = 21)
cor(pHW, matorgW)
```

```
[1] -0.6958636
```

```
cor(pHW, matorgW)^2
```

```
[1] 0.4842262
```



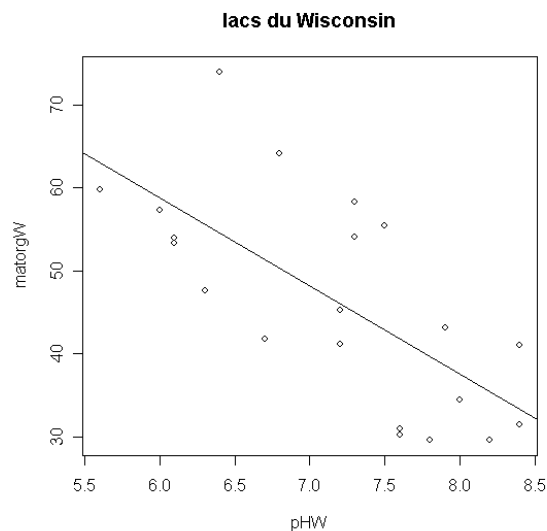
- 2) Ecrire le modèle linéaire associé.
- 3) Le pH est-il une variable aléatoire ou une variable contrôlée ?
- 4) L'équation de la droite est :

```
plot(pHW, matorgW, main = "lacs du Wisconsin", pch = 21)
lm(matorgW ~ pHW)
```

```
Call:
lm(formula = matorgW ~ pHW)
```

```
Coefficients:
(Intercept)      pHW
    122.71      -10.64
```

```
abline(lm(matorgW ~ pHW))
```



Connait-on un point solution de la droite ? Préciser.

- 5) On donne la table de décomposition de la variation associée au modèle.

```
options(show.signif.stars = FALSE)
anova(lm(matorgW ~ pHW))
```

Analysis of Variance Table

Response: matorgW

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
pHW	1	1590.03	1590.03	17.838	0.00046
Residuals	19	1693.63	89.14		

- Pourquoi les degrés de liberté de la variation résiduelle sont-ils égaux à 19?
 - Interpréter cette analyse.
- 6) En quoi le test de Shapiro permet-il de donner une conclusion finale à l'étude?

```
shapiro.test(lm(matorgW ~ pHW)$residuals)
```

```
Shapiro-Wilk normality test
```

```
data: lm(matorgW ~ pHW)$residuals
W = 0.9112, p-value = 0.05788
```

2 Lacs du Wisconsin et de la côte Atlantique du Canada

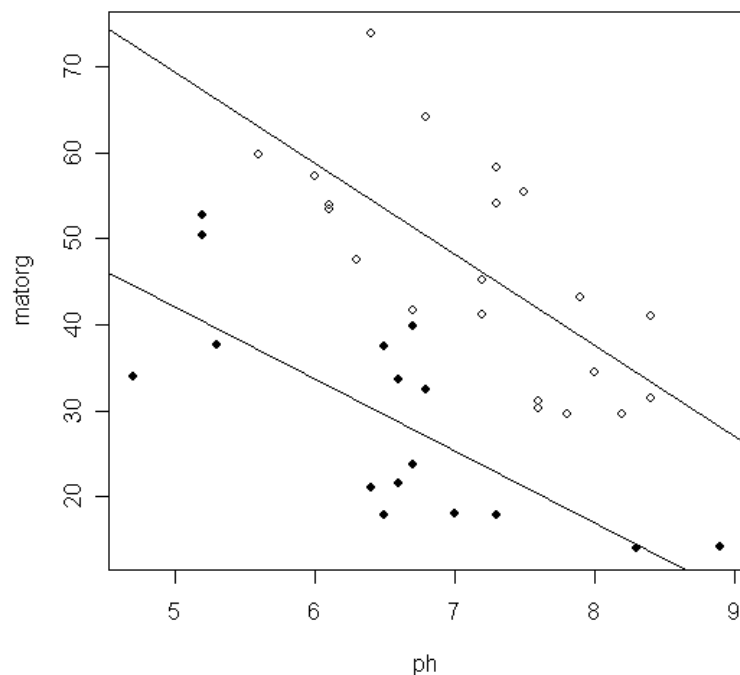
Les mêmes informations ont été recueillies sur des lacs de la côte atlantique du Canada.

```
summary(reference)
```

```
CotAtlantique  Wisconsin
           16           21
```

- 1) On donne la représentation graphique du pourcentage de matière organique (`matorg`) en fonction du pH (`ph`) des deux régions (`reference`) ainsi que les modèles linéaires¹ associés aux lacs du Wisconsin (`lmW`, points blancs) et aux lacs de la côte Atlantique (`lmCA`, points noirs). Commenter.

```
plot(ph, matorg, pch = c(19, 21)[unclass(reference)])
abline(lmW)
abline(lmCA)
```



¹P. Jolicoeur (1997) *Introduction à la biométrie* Edition Décarie

2) Modèle 1 : $lm1 \leftarrow lm(matorg \sim ph + reference)$

- a) Sur quel sous-espace est projetée la variable "pourcentage de matière organique" ?
- b) Commenter le résultat obtenu.

```
lm1 = lm(matorg ~ ph + reference)
anova(lm1)
```

Analysis of Variance Table

```
Response: matorg
      Df Sum Sq Mean Sq F value    Pr(>F)
ph      1 1171.4   1171.4   14.482 0.0005633
reference 1 4413.0   4413.0   54.557 1.456e-08
Residuals 34 2750.2     80.9
```

3) Modèle 2 : $lm2 \leftarrow lm(matorg \sim ph * reference)$

- a) Qu'apporte ce second modèle par rapport au premier ?
- b) Commenter le résultat obtenu.

```
lm2 = lm(matorg ~ ph * reference)
anova(lm2)
```

Analysis of Variance Table

```
Response: matorg
      Df Sum Sq Mean Sq F value    Pr(>F)
ph      1 1171.4   1171.4   14.2582 0.0006324
reference 1 4413.0   4413.0   53.7147 2.056e-08
ph:reference 1 39.0     39.0   0.4747 0.4956287
Residuals 33 2711.2     82.2
```

- 4) Expliquer en quoi les deux modèles sont emboîtés.
- 5) On compare les résidus du modèle 1 avec ceux du modèle 2.

Res	Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	34	2750.2				
2	33	2711.2	xx	xxx	xxxx	0.4956

- 1) Remplacer les xxx du tableau ci-dessus par leur valeur correspondante.
- 2) Lequel des deux modèles conserver ?
- 3) La représentation graphique donnait-elle déjà une idée du modèle le plus approprié aux données ?
- 4) Que faudrait-il faire pour achever complètement cette étude ?