

**Examen du 8 Décembre 2014**  
**Durée 3 heures: documents autorisés**

---

*NB: Chaque étudiant enregistre son programme dans un fichier sauvé sous le nom: "nom\_prénom.sas"  
Ne pas oublier d'écrire en tête de programme, en commentaire, le nom et le prénom. A la fin de  
l'examen le fichier sera enregistré sur la clé USB fournie  
Ecrivez, en commentaire, où commence chaque exercice et la réponse à chaque question.*

---

**EXERCICE 1**

Le fichier de données, appelé *Mercury.txt*, se trouve sur ma page web

<http://math.univ-lyon1.fr/~gciuperca/enseign.html>

La description des données sur la dernière page du présent sujet. Les variables sont dans l'ordre donnée dans cette description.

(Le fichier de données provient de l'adresse internet " <http://lib.stat.cmu.edu/DASL/Datafiles/MercuryinBass.html>")

1) Enregistrez le fichier sur votre ordinateur, en gardant le même nom pour le fichier. Créez un tableau SAS, appelé *Mercury*, à partir du fichier *Mercury.txt*, contenant toutes les observations et toutes les variables.

2) Affichez le tableau SAS *Mercury*. Quel est le nombre d'observations du fichier?

3) La variable " *age\_data*" est un indicateur sur la disponibilité des données sur l'âge des poissons: 1 si on dispose de l'âge et 0 si on ne dispose pas de l'âge. C'est pourquoi, on va partager le tableau *Mercury* en deux tableaux. Ces deux nouveaux tableaux ne doivent plus contenir la variable " *age\_data*".

- Créez le tableau SAS, qu'on va appeler *Mercury\_age*, qui contient toutes les observations du tableau *Mercury* pour lesquelles on connaît l'âge du poisson.
- Créez le tableau SAS, qu'on va appeler *no\_age*, qui contient toutes les observations du tableau *Mercury* pour lesquelles on ne connaît pas l'âge du poisson.

Spécifiez pour chacun des deux tableaux le nombre d'observations et de variables.

4) Réalisez une analyse descriptive univariée sur les variables *Alkalinity*, *pH*, *Calcium*, *Chlorophyll*, *Avg\_Mercury*, *No.samples*, *min*, *max*, *3\_yr\_Standard* du tableau de données *Mercury\_age*. Pour le même tableau de données, testez si les variables *Alkalinity*, *pH* sont corrélées avec les variables *Avg\_Mercury*, *3\_yr\_Standard*.

Il y a-t-il des données manquantes dans le tableau de données *Mercury\_age*?

5) Pour le tableau de données *Mercury\_age*, testez si les variables *Avg\_Mercury* et *3\_yr\_Standard* sont de loi Normale.

Tracez l'histogramme des variables *Avg\_Mercury* et *3\_yr\_Standard* et aussi la densité de la loi Normale correspondante.

6) En utilisant le tableau de données *Mercury\_age*, créez deux nouvelles variables:

- *log\_alkalility* qui sera le logarithme de la variable *Avg\_Mercury*.
- *log\_yrstd* qui sera le logarithme de la variable *3\_yr\_Standard*.

7) Est-ce que les variables *log\_alkalility* et *log\_yrstd* sont de loi Normale? Si oui, tracez les histogrammes correspondantes.

8) Quelles sont les moyennes des variables *log\_alkalility* et *log\_yrstd*?

9) Est-ce que les variables *log\_alkalility* et *log\_yrstd* ont la même moyenne (faire un test d'hypothèse)?

### EXERCICE 2 (à utiliser PROC IML)

On utilise le tableau *Mercury* créé à l'Exercice 1, question 1).

- 1) Lire, dans une matrice, qu'on va noter  $X$ , toutes les observations des variables *Alkalinity*, *pH*, *Avg\_Mercury*, *3\_yr\_Standard* du tableau SAS *Mercury*.
- 2) Créez une nouvelle variable  $V$  qui sera la différence des variables *Avg\_Mercury* et *3\_yr\_Standard*. Affichez la variable  $V$ .
- 3) On crée maintenant la variable  $Z$  qui aura les valeurs:

$$\begin{aligned} & -1 \text{ si } V < 0; \\ & 0 \text{ si } V = 0; \\ & 1 \text{ si } V > 0; \end{aligned}$$

Affichez les deux vecteurs  $V$  et  $Z$  et vérifiez que vous avez fait la bonne transformation.

- 4) Créez une nouvelle matrice, notée  $Y$ , en concaténant la colonne de la matrice  $X$  et les vecteurs  $V$  et  $Z$ . Vérifiez (par affichage) qu'à la question précédente vous avez fait la bonne transformation.
- 5) Créez un tableau SAS appelé *nouveau* qui va contenir toutes les variables de la matrice  $Y$ . Quittez PROC IML.

### EXERCICE 3

A partir du tableau de données créé à l'Exercice 1, question 1) et du tableau *nouveau* créé à l'Exercice 2, question 5), créez un nouveau tableau, appelé *final*. Ce tableau doit contenir seulement les variables *ID*, *Lake*, *Alkalinity*, *pH*, *Calcium*, *Cholophyll*, *Avg\_Mercury*, *No.samples*, *min*, *max*, *3\_yr\_Standard*, *age\_data*,  $V$ ,  $Z$ . Affichez le nouveau tableau.

### EXERCICE 4 (à utiliser PROC GPLOT)

On utilise le tableau *final* créé à l'Exercice 3.

Tracer le graphique de la variable *pH* en abscisse, fonction de la variable *Alkalinity* en ordonnée. Les points de représentation sont "\*" et la courbe, obtenue par interpolation, sera de couleur bleu. Donner un titre à ce graphique et pour les axes, spécifier les noms des variables (le nom de l'*Alkalinity* sera en noir et le nom du *pH* sera en vert).

### EXERCICE 5 (à utiliser les macros SAS)

1) Créez une macro-variable *tableau* associée au nom du tableau *final*, créé à l'Exercice 3. Créez également les macro-variables:

- $V1$  correspondant à la variable *pH*.
- $V2$  correspondant à la variable *Alkalinity*.
- $V3$  correspondant à la variable *Calcium*.
- $V4$  correspondant à la variable *Avg\_Mercury*.

2) Ecrivez un programme macro (paramétré) avec trois paramètres en entrée. Ce macro-programme doit réaliser un modèle de régression linéaire d'une variable fonction de deux autres variables. La procédure SAS incluse dans le macro-programme doit utiliser comme tableau en entrée la macro-variable correspondant du tableau *tableau*.

Appliquez pour les macro-variables  $V4$  fonction de  $V1$  et  $V2$ .

Appliquez pour les macro-variables  $V4$  fonction de  $V1$  et  $V3$ .

**Remarque** La procédure SAS qui fait une régression multiple est PROC REG, avec une description à l'adresse:

[http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug-reg\\_sect006.htm](http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug-reg_sect006.htm)

Vous trouvez aussi une description de la procédure dans la documentation écrite par N. COQUE (que je vous ai distribuée), page 38.

*Idée: pour ce que l'on vous demande à l'Exercice 5, il faut spécifier seulement les instructions PROC REG et MODEL*

---