



département
Mathématiques

Master MAS, parcours M2 SMSD, Université Claude Bernard Lyon 1
Remise à niveau en Statistique - Logiciel SAS,
Année 2023-2024

Examen du 17 octobre 2023
Durée 2 heures: documents autorisés

*NB: Chaque étudiant enregistre son programme dans un fichier sauvé sous le nom: "nom-prénom.sas"
Ecrire en en-tête du programme en commentaire le nom et le prénom. A la fin de l'examen, le fichier sera envoyé par e-mail à l'enseignant surveillant de votre salle (ne quittez pas la salle tant que l'enseignant ne vous a pas confirmé la réception de votre mail avec le programme SAS).
Ecrivez en commentaire où commence chaque exercice et la réponse à chaque question.*

Salle	Enseignant	Adresse mail de l'enseignant
Mathesis	Gilles COHEN	gilles.cohen@univ-lyon1.fr
TDmath	Gabriela CIUPERCA	Gabriela.Ciuperca@univ-lyon1.fr

EXERCICE 1

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

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

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

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 tableau SAS *Mercury* et donc du fichier *Mercury.txt*?

3) La variable "*age_data*" est un indicateur sur la disponibilité des données sur l'âge des poissons. Les valeurs pour cette variable sont: 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)?

10) A partir du tableau *Mercury_age* créé à la question 3), créez un nouveau tableau qu'on va appeler *n3_Mercury_age* où une nouvelle variable, appelée *conc*, qui prend trois valeurs possibles sera obtenue à partir de la variable *calcium*:

$$conc = \begin{cases} faible & \text{si } calcium < 25 \\ moyen & \text{si } calcium \geq 25 \text{ et } calcium < 75 \\ eleve & \text{si } calcium \geq 75 \end{cases}$$

11) Réalisez une analyse descriptive pour la variable *conc*. Testez si les trois proportions de la variable *conc* sont égales.

EXERCICE 2

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

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 3

1) Créez un macro-programme nommé *calcul* qui prend en entrée trois paramètres :

- *Tab* : un tableau de données SAS
- *var* : une variable numérique présente dans le tableau *Tab*
- *seuil* : une valeur numérique.

Dans ce macro-programme, il faut d'abord créer un nouveau tableau qui contiendra seulement les observations pour lesquelles $var < seuil$. Ensuite, il faut trouver le nombre d'observations qu'on va noter *nligne*, du nouveau tableau.

Et enfin, il faut afficher une phrase dans laquelle vous spécifiez que "le nombre d'observations du tableau *Tab* qui vérifient la condition $var < seuil$ est égal à *nligne*".

Tester ce macro-programme sur:

- *Tab=Mercury, var=pH, seuil=5;*
- *Tab=Mercury, var=calcium, seuil=25.*

Description données fichier Mercury.txt

1. *ID*: ID number
2. *Lake*: Name of the lake
3. *Alkalinity*: Alkalinity (mg/L as Calcium Carbonate)
4. *pH*: pH
5. *Calcium*: Calcium (mg/l)
6. *Chlorophyll*: Chlorophyll (mg/l)
7. *Avg_Mercury*: Average mercury concentration (parts per million) in the muscle tissue of the fish sampled from that lake
8. *No.samples*: How many fish were sampled from the lake
9. *min*: Minimum mercury concentration amongst the sampled fish
10. *max*: Maximum mercury concentration amongst the sampled fish
11. *3_yr_Standard_mercury* : Regression estimate of the mercury concentration in a 3 year old fish from the lake (or = Avg Mercury when age data was not available)
12. *age_data*: Indicator of the availability of age data on fish sampled