

Examen du 15 octobre 2021
Durée : 1 heure 30 – Documents autorisés

NB : Chaque étudiant enregistre son programme dans un fichier nommé « 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	Laure Marillet	lauremarillet@gmail.com
TDmath	Gabriela Ciuperca	Gabriela.Ciuperca@univ-lyon1.fr
Quai 43	Antoine Denoyer	denoyer.antoine@gmail.com

EXERCICE 1

Le fichier de données « NO2.txt » se trouve sur la page de Mme CIUPERCA :

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

rubrique « Fichier de données ».

Les données de ce fichier proviennent du site <http://lib.stat.cmu.edu/datasets/> et elles représentent : des données concernant la pollution de NO2 fonction des variables météorologiques. Dans le fichier NO2.txt il y a huit variables, dans l'ordre : la valeur horaire du logarithme de la concentration de NO2 (particules), le logarithme du nombre de voitures par heure, la température à 2 mètres au-dessus du sol, la vitesse du vent, la différence de température entre 25 et 2 mètres au-dessus du sol, la direction du vent, l'heure du jour et le numéro du jour à partir du 1er octobre 2001.

- 1) Créer la table SAS qu'on va appeler NO2 à partir du fichier de données NO2.txt contenant toutes les observations. Les variables de cette table vont s'appeler logNO2, logV, temp2m, vitesse, diffTEMP, direction, heure, nbjour.
- 2) Trier la table NO2 en ordre croissant par rapport aux valeurs de la variable nbjour. Afficher la nouvelle table NO2.

Calculer la moyenne par jour des six variables logNO2, logV, temp2m, vitesse, diffTEMP, direction. Pour la procédure utilisée, donner l'option pour que les résultats de statistique descriptive ne s'affichent pas. Conserver ces moyennes dans une table nommée sortie1, avec les noms des colonnes correspondantes : mlogNO2, mlogV, mtemp2m, mvitesse, mdiffTEMP, mdirection.

- 3) A partir de la table *sortie1* créer une nouvelle table *sortie2* dans laquelle il faut créer la variable binaire *vit* qui prend la valeur « *positif* » si $mtemp2m \geq 0$ et la valeur « *negatif* » si $mtemp2m < 0$. Supprimer les variables *_TYPE_* et *_FREQ_* de la table *sortie2*.
Afficher la table *sortie2*.
- 4) A partir de la table *sortie2* créer deux tables : une qui contient les observations pour lesquelles la variable *vit* prend la valeur *positif* et une autre avec la valeur *negatif*. Il y a combien d'observations dans chacune des deux tables ? (la réponse est à mettre en commentaire)
- 5) Pour la table *sortie2*, réaliser de la statistique descriptive sur la variable *vit*. Réalisez un test d'hypothèse de Chi2 d'égalité des deux probabilités : $P[vit=positif]=P[vit=negatif]$.
- 6) Pour la table *sortie2*, tracer l'histogramme ainsi que la densité de la loi Normale s'ajustant au mieux aux données pour la variable *mlogNO2*. Tester sa normalité.
- 7) Pour la table *NO2*, en utilisant la procédure MEANS pour la variable *logNO2*, sauvegarder dans la table *results* la valeur de la médiane journalière (calculée par rapport à la variable *nbjour*). La variable sauvegardée sera nommée *medNO2*. Dans la table *results* il faut garder seulement les variables *medNO2* et *nbjour*. Afficher la table *results*.
- 8) Assembler les trois tables *NO2*, *sortie2* et *results* dans une seule, nommée *final*, par rapport à la variable *nbjour*. Créer une nouvelle variable, nommée *dif* comme la différence entre les valeurs *medNO2* et *logNO2*.

EXERCICE 2 (Graphique)

On utilise la table *final* créée à la question 8) de l'exercice 1.

- 1) Représenter les boxplots de la variable *mlogNO2* en fonction de la variable *vit* et relier les moyennes des deux boxplots par un trait.
- 2) Tracer sur le même graphique le nuage de points des variables *medNO2* et *logNO2* en fonction du numéro du jour. La variable *medNO2* doit être représentée par des symboles en forme de croix (+) bleue, et la variable *logNO2* par des symboles en forme de cercle rouge. Faire afficher la légende.

EXERCICE 3 (Macros)

Créer un macro-programme nommé *modele* qui prend en entrée trois variables :

- *Table* : une table de données SAS
- *y* : une variable numérique présente dans la table *Table*
- *z* : une variable numérique présente dans la table *Table*

Ce macro-programme doit d'abord calculer la médiane de la variable *y* pour chaque valeur de la variable *z*. Cette médiane de *y* sera sauvegardée dans un tableau. Ensuite, pour cette nouvelle table, tracer le nuage de points de la médiane fonction de la variable *z*. Les valeurs de la médiane doivent être représentées par des symboles en forme de croix (+) bleue. Mettre un titre pour spécifier que les valeurs de la médiane de la variable *y* sont représentées fonction de *z*.

Tester cette macro sur la table *NO2* et les variables (dans l'ordre) *logNO2*, *nbjour*.