

Examen du 16 octobre 2020
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	Alexandra Werle	alexandra.werle@gmail.com
TDmath	Gabriela Ciuperca	Gabriela.Ciuperca@univ-lyon1.fr
Ariane 03	Laure Marillet	laure_marillet@hotmail.com
Nautibus TP6	Anne Perrut	Anne.Perrut@univ-lyon1.fr
Nautibus TP13	Anne Perrut	Anne.Perrut@univ-lyon1.fr

EXERCICE 1

Les tables *Prelevement_1.csv* et *Prelevement_2.txt* sont accessibles sur la page de Mme CIUPERCA:
<http://math.univ-lyon1.fr/~gciuperca/enseign.html>

Rubrique « Fichier de données ».

- 1) Créer 2 tables SAS à partir des fichiers de données *Prelevement_1.csv* et *Prelevement_2.txt* contenant toutes les observations. La première table contient les colonnes *Echantillon*, *Age*, *Date_1* et *Resultat_1*. La seconde table contient les colonnes *Echantillon*, *Date_2* et *Resultat_2*.
- 2) Assembler les 2 tables par la variable commune *Echantillon* pour créer une unique table *Prelevement*. Cette nouvelle table contient pour chaque échantillon, le résultat de 2 prélèvements réalisés à 2 dates différentes.
- 3) Supprimer les échantillons pour lesquels il y a une incohérence dans les dates, c'est-à-dire ceux pour lesquels la date du 1^{er} prélèvement est supérieure à celle du 2^{ème} prélèvement. Combien d'échantillons présentent une incohérence ?
- 4) Tester la normalité des variables *Resultat_1* et *Resultat_2* et tracer l'histogramme ainsi que la courbe de distribution pour chacune de ces variables.
- 5) Créer une nouvelle variable *Groupe_Age* qui prend les valeurs « <=35 » si l'âge du patient est inférieur ou égal à 35, et « >35 » sinon.
Créer une autre variable *Diff_resultat* qui contient la différence entre *Resultat_2* et *Resultat_1*. On considère que si le résultat augmente de plus de 6 unités entre le 1^{er} résultat

et le 2nd résultat, l'échantillon présente un risque de maladie. Créer une nouvelle variable *Risque* qui indique si l'échantillon présente un risque ou non selon la différence entre les 2 résultats.

- 6) Standardiser la variable *Diff_Resultat* à l'aide de la PROC STDIZE. Dans une table nommée *Standard*, récupérer la variable standardisée sous le nom *StdDiff_resultat* et la variable originale *Diff_resultat* sous le nom *OrgDiff_resultat*.

EXERCICE 2 (Graphique)

On utilise la table créée à la question 5 de l'exercice 1.

- 1) Tracer les boîtes à moustaches de la variable *Diff_resultat* en fonction du groupe d'âge.
- 2) Sur le même graphique, tracer le nuage de points des variables de résultat en fonction de l'âge des échantillons. La variable *Resultat_1* doit être représentée par des symboles en forme de croix (+) bleue, et la variable *Resultat_2* par des symboles en forme de cercle orange. Faire afficher la légende.

EXERCICE 3 (Macros)

On utilise la table créée à la question 5 de l'exercice 1.

Créer un macro-programme avec 5 paramètres en entrée :

- *Table*
- *Var_group*
- *Var*
- *Moy_theo*
- *Sens*

Ce macro-programme doit comparer la moyenne d'une variable à une moyenne théorique en précisant le sens de comparaison (par exemple, $H_0 : \text{Moyenne} \leq \text{Moyenne théorique}$) selon une variable de groupe.

Avec cette macro, répondez aux questions suivantes pour chaque groupe d'âge :

- 1) La moyenne du premier résultat est-elle inférieure ou égale à 15 ?
- 2) La moyenne du deuxième résultat est-elle supérieure ou égale à 20 ?