

Examen du 18 octobre 2019
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 enregistré sur la clé USB fournie. Ecrivez en commentaire où commence chaque exercice et la réponse à chaque question.

EXERCICE 1 (Etape data et statistiques descriptives)

1. Importer sous SAS le fichier *data_sommeil.csv* disponible à l'adresse suivante : http://math.univ-lyon1.fr/~gciuperca/data_sommeil.csv.
Ce fichier contient des informations sur 100 patients : l'âge, le poids (en kg), la taille (en cm), le sexe (0=homme, 1=femme) et des informations sur leur consommation d'alcool (en nombre de verres par semaine) et des écrans (en nombre d'heures par semaine), si les patients font de l'insomnie (0=non, 1=oui) et s'ils fument (0=non, 1=oui).
2. Corriger la valeur 294 de la colonne TAILLE de l'individu ID_042 par 194.
3. Ajouter une colonne correspondant à l'IMC de chaque individu : $\text{poids} / \text{taille}^2$, avec le poids en kg et la taille en m. Créer ensuite une colonne IMC_INTERPRET qui prend la modalité « Surpoids » si l'IMC est supérieur ou égal à 25, et « Normal » sinon.
4. La colonne IDEN est parfois présentée avec des caractères en minuscule, parfois en majuscule. Créer une nouvelle variable IDEN2 pour laquelle l'ensemble des identifiants sont en majuscules.
5. Réaliser un histogramme de la variable AGE, ainsi que le test de normalité de cette variable.
6. Calculer la moyenne et l'écart-type de la variable ECRAN et conserver ces deux informations dans une table nommée STAT_ECRAN1 contenant deux colonnes : MOY1 (pour la moyenne) et STD1 (pour l'écart-type).
7. Faites de même sans considérer la valeur aberrante 16 et conserver alors les deux statistiques (moyenne et écart-type) dans une table STAT_ECRAN2 avec comme nom de colonnes MOY2 (pour la moyenne) et STD2 (pour l'écart-type)
8. Calculer le rapport $\text{STD1}/\text{STD2}$ en regroupant les deux tables STAT_ECRAN1 et STAT_ECRAN2 dans une même table.

EXERCICE 2 (procédures)

1. Représenter sous forme de nuage de points la variable ECRAN en fonction de la variable AGE avec une couleur différente en fonction de la variable INSOMNIE.
2. Représenter les boxplots de la variable AGE en fonction de la variable INSOMNIE et relier les moyennes des 2 boxplots par un trait.
3. Réaliser un test de Student pour comparer les moyennes de la variable AGE en fonction de la variable INSOMNIE.
4. Calculer la proportion de patients souffrant d'insomnie dans l'étude ainsi que l'intervalle de confiance binomiale de cette proportion.
5. Réaliser un test du Chi2 d'indépendance entre les variables INSOMNIE et TABAC.

EXERCICE 3 (Macro)

Créer une macro nommée *graph* qui prend en entrée 4 variables :

- *Table* : une table de données SAS
- *Var_x* : une variable numérique présente dans la table *Table*
- *Var_y* : une variable numérique présente dans la table *Table*
- *Var_group* : une variable de groupe présente dans la table *Table*

Cette macro doit réaliser les nuages de points de la variable *Var_y* en fonction de la variable *Var_x* avec une couleur différente pour chacun des groupes de la variable *Var_group* de la table *Table*.
Tester cette macro sur la table utilisée au premier exercice, *data_sommeil*.

Exemples :

```
%graph(Table=data_sommeil, var_x=ECRAN, var_y=AGE, var_group=INSOMNIE);
```

```
%graph(Table=data_sommeil, var_x=TABAC, var_y =AGE, var_group=SEXE);
```