

Projet SAS de l'UE Statistique avec SAS

Date de rendu : 5 Juin 2024

Les exercices sont indépendants.

L'énoncé comporte 2 pages.

La notation tiendra particulièrement compte de la rédaction.

Les hypothèses nulles et alternatives ainsi que les conditions d'application des tests utilisés devront être présentées.

Exercice 1 :

Nous disposons d'un échantillon composé de garçons et de filles d'un district écossais. Cet échantillon est composé d'observations indépendantes. Les données sont résumées dans le tableau suivant :

	BLOND	ROUX	CHATAIN	BRUN	NOIR DE JAIS
GARCON	592	119	849	504	36
FILLE	544	97	677	451	14

Procéder à l'analyse statistique adéquate pour pouvoir déterminer si la couleur des cheveux est indépendante du sexe.

Exercice 2 :

Une entreprise pharmaceutique expérimente une nouvelle molécule censée faire baisser une mesure physiologique pour des patients porteurs d'une pathologie. Pour cela, 54 patients sans lien de parenté ont été recrutés. Chaque patient s'est vu attribuer soit la nouvelle molécule (traitement A) soit un placebo (traitement B). Les groupes ont été construits de telle manière à ce que les distributions de l'âge et du sexe soient similaires dans les deux groupes. Les données sont contenues dans le fichier molecule.csv.

Quelles sont les moyennes et écart-types dans chacun des groupes?

Est-ce que la molécule testée est efficace?

Exercice 3 :

La pollution de l'air constitue actuellement une des préoccupations majeures de santé publique. De nombreuses études épidémiologiques ont permis de mettre en évidence l'influence sur la santé de certains composés chimiques comme l'ozone (O_3). Le jeu de données *ozone.xls* comporte 112 observations indépendantes relevées durant l'été 2001.

- La variable *max03* est le maximum journalier de la concentration en ozone (en $\mu g/m^3$).
- Les variables *T9*, *T12*, *T15* correspondent aux températures relevées respectivement à 9h, 12h et 15h.
- Les variables *Ne9*, *Ne12*, *Ne15* correspondent aux nébusolistés relevées respectivement à 9h, 12h et 15h.
- Les variables *Vx9*, *Vx12*, *Vx15* correspondent à la composante Est-Ouest du vent relevée respectivement à 9h, 12h et 15h.

Le but de l'exercice est d'étudier l'éventuelle relation linéaire entre la variable *max03* et l'ensemble des autres variables. Pour cela, à l'aide de SAS, après avoir présenté les statistiques descriptives des variables, il faut appliquer un protocole de sélection du modèle. Le critère de sélection utilisé devra être présenté. Une fois le modèle optimal trouvé, il est demandé d'interpréter les résultats.