

TISD¹ – DM 2

Travail à faire en binôme. Un rapport et un script R et/ou SAS par binôme doivent être déposés sur Moodle avant le mercredi 7 décembre 23h55. Le nom des étudiants doit apparaître sur le rapport et en commentaire au début du script. Le rapport, en format .pdf, présentera les réponses aux questions, les sorties graphiques, et vos explications/commentaires. Vous êtes libres d'utiliser R et/ou SAS à votre convenance.

Exercice 1 (Criquets et température)

Certains disent que l'on peut déterminer la température ambiante en écoutant chanter les criquets. Pour vérifier ces dires on mesure, à 16 différentes températures (en °C), la fréquence de stridulation de criquets dans un champs (en Hz). Le jeu de données est disponible sur Moodle : `cricket_chirps.csv`.

1. Importer ces données et proposer un modèle statistique pour modéliser la température en fonction de la fréquence. Expliquez pourquoi ce choix vous paraît cohérent et donner les hypothèses précises que l'on fait en utilisant ce modèle.
2. Donner les estimateurs des paramètres de ce modèle et le coefficient de détermination. Est-ce que cette modélisation est pertinente ? Est-elle satisfaisante ? Justifier.
3. Prédire la température sachant que la fréquence mesurée est de 3Hz. Quelles garanties pouvez-vous fournir ? Même question avec une fréquence de 0.5Hz. Comment pourrait-on améliorer ces garanties ? Jusqu'à quel niveau de précision peut-on arriver ? Justifier.
4. Présentez vos conclusions de cette analyse dans une courte synthèse.

¹Responsable : Adrien Hardy. Laboratoire Paul Painlevé, Université des Sciences et Technologies de Lille, Bâtiment M3, Bureau 306. Email : adrien.hardy@math.univ-lille1.fr

Exercice 2 (Eaunologie)

En 2008, une expérience menée à Longwood University consistait à faire goûter aux participants quatre eaux différentes et leur demander de classer ces eaux suivant leur ordre de préférence. L'expérience s'est déroulée en double aveugle : ni les participants ni expérimentateurs ne savaient comment les eaux étaient réparties, les bouteilles n'étant labelées que par A , B , C , ou D . Le contenu réel des bouteilles est le suivant.

A : eau en bouteille (Sam's choice)	B : eau en bouteille (Aquafina)
C : eau en bouteille (Fiji)	D : eau du robinet

Le jeu de données `water_tasting.csv` contient les résultats de cette expérience. Avant les résultats, il contient aussi d'autres renseignements demandés aux participants : leur sexe (Male, Female), leur age, leur position universitaire (Freshman, SOphomore, Junior, senior SR, Other), le type d'eau usuellement consommée (Bottled, Filtered, Tap, Not Applicable), et leur marque favorite de bouteille d'eau.

1. Certaines réponses sont incomplètes : supprimer du jeu de données ces participants.
2. Donner une représentation graphique de la répartition des eaux préférées des participants. Pensez-vous qu'une eau soit préférée par rapport aux autres ? Comment tester cette hypothèse ? Faire le test et conclure. Même question pour les eaux les moins appréciées.
3. Pensez-vous que la préférence de l'eau dépende du sexe du participant ? Donner une représentation graphique puis proposer un protocole pour tester cette hypothèse. Que peut-on conclure ?

Aide : On trouve dans la littérature des conditions d'applicabilité pratique des test du χ^2 , comme la "condition de Cochran". Qu'en pensez-vous ?

4. Extraire des données toutes les autres informations que vous jugerez intéressantes.