
TP noté B du mercredi 4 décembre 2024 de 9h45 à 11h15

Un aide mémoire est fourni avec le sujet

LES AUTRES DOCUMENTS, CALCULATRICES ET TÉLÉPHONES NE SONT PAS AUTORISÉS

Nom, prénom et numéro d'étudiant·e :

.....

Il faut charger sur Tomuss, dans la colonne "RenduTP9h45", une feuille de calcul python (format .py ou .ipynb) et rendre le sujet avec les réponses aux questions. On pourra utiliser le programme Spyder ou <https://jupyter.univ-lyon1.fr/> en "Mode Examen".

On aura besoin des paquets suivants :

```
import numpy as np
import scipy.stats as st
import matplotlib.pyplot as plt
```

```
import pandas as pan
import scipy.stats.mstats as ms
```

2 points seront accordés à la présentation du code (Commentez le!)

2 points seront accordés à la présentation des graphiques (Mettez des titres et légendes!)

Exercice 1. Statistiques

Charger le jeu de données suivant :

```
df=pan.read_csv("http://tinyurl.com/y39an7ef/Data34698.csv", sep="\t")
for nom in df.keys():
    globals()[nom] = df[nom]
```

Ce jeu de données contient des informations issues de relevés de la pollution industrielle à Feyzin entre 2016 et 2017. Ils sont mesurés en microgrammes par mètre cube d'air. Il contient les relevés de 4 polluants :

- un composé organique volatil cancérigène : le benzène.
- d'autres composés organiques volatils : l'éthane, le propane et le propène.

On donnera les réponses numériques arrondies avec TROIS décimales.

- Quel est le type de variable statistique de chacune des variables (qualitative, quantitative discrète ou continue) ?

.....

- Quel est le nombre de jours de 2017 pour lesquels l'échantillon contient une observation ?

.....

- Créez une nouvelle variable statistique **Seuil**, qui, pour chaque jour de mesure,

- vaut 2 si la mesure d'**Éthane** dépasse la limite de pollution de $5\mu\text{g}/\text{m}^3$,
- vaut 1, si cette limite n'est pas dépassée par la mesure d'**Éthane** mais que l'objectif de qualité de $2\mu\text{g}/\text{m}^3$ est dépassé,

— vaut 0 si cet objectif n'est pas dépassé.

Donnez la table de contingence des fréquences de cette variable **Seuil**.

.....
.....

4. Tracez un diagramme en tuyau d'orgue de **Seuil**.

5. Trouvez la moyenne empirique, variance empirique, variance empirique non-biaisée et le quartile à 25 % de la mesure du **Éthane**.

.....
.....

6. Soit X l'échantillon des mesures de la pollution journalière à l'**Éthane** à Feyzin et x son logarithme népérien $x = \ln(X)$. On suppose que x peut être modélisé par une loi normale de moyenne m et écart-type σ inconnus. Trouver (avec python) un intervalle de confiance (bilatéral) au niveau de confiance 0.90 pour la moyenne m .

.....

7. Soit Y l'échantillon des mesures de la pollution journalière au Propane et y son logarithme népérien $y = \ln(Y)$. Trouver et donner la droite de régression $y = ax + b$ de y en fonction de x (défini au 6). Calculer la corrélation empirique de x et y . Est-ce qu'une approximation par la droite de régression est raisonnable ?

.....

8. Tracer le nuage de points de (x, y) en bleu et la droite de régression en rouge sur le même graphique.

Exercice 2. Simulations

1. Construire un vecteur $X = (X_1, \dots, X_N)$ simulant un échantillon de $N = 500$ variables de loi exponentielle $\mathcal{E}(2)$. Vérifier la moyenne empirique de X .
2. Construire un échantillon donnant $U = (U_1, \dots, U_N) = (\exp(-2X_1), \dots, \exp(-2X_N))$. Calculer le minimum et la maximum de U .
3. Tracer un histogramme de l'échantillon U .
4. En comparant l'histogramme de U avec la densité d'une loi uniforme bien choisie (on tracera la courbe de la densité en rouge par dessus l'histogramme), faites une hypothèse sur la loi de U en précisant l'intervalle de support de la loi uniforme.

.....
.....

On testera pour des N plus grands. Quelles valeurs de la taille de l'échantillon N vous ont permis de conclure ?

.....
.....

5. On pose

$$Y_n = \frac{X_1 + \dots + X_n}{n},$$

Tracer le vecteur $Y = (Y_1, \dots, Y_n, \dots, Y_N)$ en fonction de n .

6. Quelle est la limite attendue pour Y_n . Quel théorème du cours utilise-t-on ?

.....
.....

7. Tracer, sur le même graphique que Y , une droite constante pour la limite de Y_n pour illustrer ce théorème.