

Nom, prénom :

/10

*Durée : 15 minutes. Accès à python et cours sur internet.*

**Question 1 :** Rappeler la définition de  $R^2$  et son interprétation dans le cadre d'une régression linéaire simple (2 points)

Le coefficient de détermination compare la variance des données par rapport à la moyenne à la somme du carré des erreurs par rapport à la régression. Il est donné par la formule

suivante : 
$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y})^2}{\sum_i (y_i - \bar{y})^2},$$

Pour un jeu de données où le bruit domine et il n'y a pas de tendance linéaire qui ressort, il tend vers 0. Dans le cas contraire, il tend vers 1. Il n'est pas robuste et sensible aux valeurs extrêmes.

**Question 2 :** Cocher les affirmations justes (2 points)

- ☐ Un score  $R^2$  proche de 1 est suffisant pour dire qu'il existe une relation linéaire forte entre 2 variables  $x$  et  $y$
- ☐ La fonction `plt.points` de matplotlib permet de représenter un couple  $x, y$  graphiquement sous forme de nuages de points
- ☒ Un tuple sous python est une structure de données équivalente à la liste, mais qui est immuable (ne peut pas être changé)
- ☒ Une distribution symétrique de type gaussienne est généralement caractérisée par une valeur centrale et une mesure de son étendu (ou dispersion)

**Question 3 :** Soit la distribution suivante :



Quelles sont les affirmations vraies ? (2 points)

- ☐ La moyenne est à gauche de la médiane
- ☐ La moyenne et la médiane sont confondues
- ☒ La médiane est inférieure à la moyenne
- ☐ La distribution est symétrique

**Question 4 :** Soit un vecteur suivant une distribution log-normale :  
(4 points)

```
import numpy as np
X = np.random.lognormal(0,1,size=100)
```

1. Calculez la moyenne et la médiane de X en utilisant `np.mean()` et `np.median()`  
(1 point)

```
np.mean(X), np.median(X)
```

2. Que constatez-vous ? que pouvez-vous en déduire sur cette distribution ? (1.5 point)

La médiane est inférieure à la moyenne, la distribution est asymétrique vers la gauche

3. Ecrivez le code qui affiche l'histogramme de la distribution des points (1.5 point)

```
import matplotlib.pyplot as plt
plt.hist(X)
```