

Nom, prénom :

/10

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

**Question 1 :** Rappeler la définition et les hypothèses sur les résidus pour un modèle de régression linéaire (2 points)

- les erreurs suivent une loi normale de moyenne nulle
- la variance est la même pour tous (homoscédasticité) : la variance est la même pour l'ensemble des termes d'erreurs gaussiens  $Var(\varepsilon_i) = \sigma$
- les termes d'erreurs pour les différents  $x_i$  sont indépendants les uns des autres

**Question 2 :** On diagnostique les résidus de 3 régressions linéaires, en les représentant graphiquement. Dites pour chaque cas si le modèle semble valide ou non, et expliquer brièvement pourquoi. (3 points)



1 : Les résidus semblent centrés autour d'une même valeur, avec une dispersion constante, puisque les points sont tous dans une même bande. Ces points ne violent pas à priori les hypothèses sur les résidus pour une régression linéaire, et un histogramme sur y donnerait bien une distribution de type gaussienne.

2 : Les résidus dépendent clairement de x

3 : Les résidus dépendent clairement de x, et la dispersion est de plus en plus forte pour x croissant

**Question 3 :** Quels choix ci-dessous permettent de créer une matrice 2x2 ? (2 point)

```
import numpy as np
[ ] M = np.matrix([1,2,3,4])
[ ] M = np.array([1,2;3,4])
[X] M = np.array([1,2,3,4]).reshape(2,2)
[X] M = np.array([[1,2],[3,4]])
```

**Question 4 :** Soit la loi normale  $N(20, 1)$ , cocher les propositions correctes (1,5 points)

[X] Environ 68% des valeurs générées suivant cette loi sont entre 19 et 21  
[ ] La distribution est centrée en 0  
[X] La médiane vaut 20

**Question 5 :** Que fait le code suivant (1,5 point) ?

```
import numpy as np
M = np.array([[1,2],[3,4]])
np.array([[M[i,j] for i in range(M.shape[0])] for j in
range(M.shape[1])])
```

Le code calcul la transposée de la matrice M