

**Correction**

*Durée : 10 minutes. Accès au shell Python autorisé.*

**Question 1 :** A quoi correspond la fonction `__init__()` dans une classe ? (2 points)

- ☒ Elle est appelée automatiquement pour chaque nouvelle instance de la classe
- ☐ Elle est appelée automatiquement à chaque utilisation de l'instance
- ☒ Elle est utilisée pour initialiser les paramètres de l'instance

**Question 2 :** On diagnostique les résidus de 3 modèles différents de régression linéaire, en les représentant graphiquement. Dites pour chacun si le modèle semble valide, et expliquer très brièvement pourquoi. (3 points)



**Remarque :** les résidus sont définis comme étant la différence entre les points de données et leurs prédictions. Cf dans le chapitre `Regression_5_Cas_pratique.ipynb` :

```
residuals = y-regr.predict(x)
```

1 : Valide

Les résidus semblent avoir la même variance, ne dépendent pas de  $x$ , et peuvent coller avec une distribution gaussienne (un histogramme permettrait de vérifier)

2 : Pas Valide

Les résidus et leur distribution dépendent clairement de  $x$

3 : Pas Valide

Les résidus et leur distribution dépendent clairement de  $x$

**Question 3 :** Que fait le code suivant ? (1.5 points)

```
from scipy.stats import norm
norm.ppf(0.25)
```

Soit  $X$  une variable aléatoire qui suit une distribution normale de moyenne 0 et d'écart-type 1, le code précédent donne la valeur  $X_c$  telle que  $P(X \leq X_c) = 0.25$ .

Autre exemple, `norm.ppf(0.5)` retourne la valeur de la médiane.

Nous avons utilisé cette fonction dans le chapitre `Regression_4_Statistiques.ipynb`, pour calculer les intervalles de confiance.

**Question 4 :** Suite à un sondage pour connaître les intentions de votes pour les candidats A et B, on trouve pour une population de 1000 personnes, 52 % de votes pour A. Dans le cadre du théorème central limite, on approxime l'intervalle de confiance à 95% sur la distribution binomiale par la formule suivante :  $p \pm 1.96 \sqrt{\frac{p(1-p)}{n}}$ .

1. Calculer l'intervalle de confiance (1 point)

Il suffit d'appliquer la formule précédente avec  $p=0.52$  et  $n=1000$ .

On trouve  $0.52 \pm 0.03$

2. Que peut-on conclure sur la différence dans les intentions de votes entre A et B ? (1 point)

$0.52 - 0.03 = 0.49$ , donc le candidat A n'est pas sûr que les intentions de votes soient effectivement supérieures à celles pour le candidat B. La différence peut très bien s'expliquer par des fluctuations statistiques.

**Question 5 :** 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
- ☒ La distribution est représentative des revenus des ménages