

Consignes générales.

- Vous disposez d'une heure pour traiter le sujet qui vous a été fourni.
- Vous avez droit à vos notes de cours, le seul logiciel utile pour ce TP est Python.
- Vous créez sur votre ordinateur un dossier nommé de la façon suivante : numéro du sujet suivi du nom de famille (par exemple 8dupont), dans lequel seront sauvegardés tous les documents (programmes Python, fichiers textes) nécessaires à votre évaluation, qui sera effectuée uniquement sur cette base (pas de copie papier). Le surveillant de salle s'occupera de récupérer les fichiers sur une clé USB, laissez simplement les ordinateurs allumés à la fin de l'heure.
- Le surveillant de salle ne peut intervenir qu'en cas de problème technique sur votre ordinateur.
- Bon courage !

Sujet n°3 : Droite de régression linéaire.

Le principe de la régression linéaire est de trouver, à partir de deux listes de données $l = [l[0], \dots, l[n-1]]$ et $m = [m[0], \dots, m[n-1]]$, l'équation de la droite qui s'approche le plus possible de tous les points du plan de coordonnées $(l[i], m[i])$. Naturellement, une telle approximation sera d'autant plus possible qu'il y a une forte corrélation entre les deux listes de valeurs.

1. Écrire une fonction en Python calculant la moyenne \bar{l} des valeurs d'une liste l donnée en argument.
2. Construire de même une fonction **var** donnant la variance d'une liste l (on rappelle que celle-ci est donnée par la formule $v(l) = \frac{1}{n} \sum_{i=0}^{n-1} (l[i] - \bar{l})^2$, où n est le nombre d'éléments de la liste).
3. Écrire une fonction **covar** prenant comme arguments deux listes l et m de même longueur, et calculant la covariance des deux listes, définie par $\text{covar}(l, m) = \frac{1}{n} \sum_{i=0}^{n-1} (l[i] - \bar{l})(m[i] - \bar{m})$.
4. En déduire une fonction prenant comme arguments deux listes l et m , et ressortant le coefficient directeur a et l'ordonnée à l'origine b de leur droite de régression linéaire, donnés par les formules $a = \frac{\text{covar}(l, m)}{\text{var}(l)}$, et $b = \bar{m} - a\bar{l}$.
5. À l'aide du module matplotlib, tracer la droite de régression linéaire, ainsi que les points de coordonnées $(l[i], m[i])$, pour les listes de données suivantes (l correspond au temps de révision, m à la note obtenue), issue d'un sondage sur une classe de PTSI à l'issue d'un devoir de maths :

temps de révision	1	2	2.5	3	4	4.5	5	6
note obtenue	5	10	12	15	10	15	13	17

6. Estimer la note que peut espérer un élève ayant révisé pendant 7 heures.