Lycée Clemenceau **ECG 2**

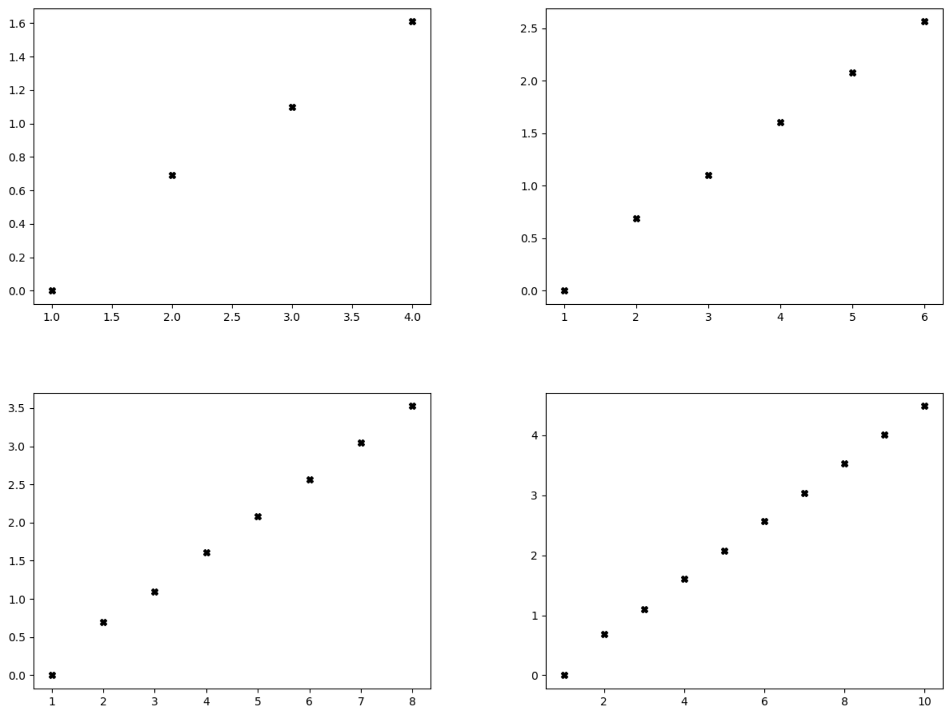
TP 4 : STATISTIQUE DOUBLE

Exercice 1. Soit $(u_n)_{n\geq 0}$ la suite de Fibonacci, c'est-à-dire la suite réelle définie par $u_0 = 0$, $u_1 = 1$ et pour tout $n \geq 2$ par $u_n = u_{n-1} + u_{n-2}$.

(1) Ecrire une fonction en Python qui, étant donné un entier $n \geq 3$, retourne le vecteur $u = (u_1, u_2, ..., u_n)$.

(2) En déduire une fonction en Python qui, étant donné un entier $n \geq 3$, construit les vecteurs $x = (1, 2, ..., n)$ et $y = (\ln(u_1), \ln(u_2), ..., \ln(u_n))$, puis calcule et affiche la somme des carrés des écarts verticaux entre les points du nuage et la droite de régression $\mathcal{D}_{y/x}$.

(3) On suppose que l'on a écrit une fonction en Python qui trace en fonction de n le nuage de points associé aux séries statistiques x et y. On a obtenu les résultats suivants pour $n = 4, 6, 8, 10$:



Que peut-on conjecturer sur le coefficient de corrélation linéaire si n est grand ?

Exercice 2. Ecrire une fonction en Python qui, à partir d'un entier $n > 0$ et de deux réels a, b tels que $a < b$, crée deux listes $x = (x_1, ..., x_n)$, $y = (y_1, ..., y_n)$ de n réels aléatoires dans $[a, b]$, calcule les nouvelles listes $u = (x_1 + y_1, ..., x_n + y_n)$ et $v = (x_1 y_1, ..., x_n y_n)$, puis détermine le coefficient de corrélation linéaire r de u, v et enfin affiche l'un des messages suivants :

— "les variables sont peu corrélées" si $|r| < 0.3$,

— "les variables sont moyennement corrélées" si $0.3 \leq |r| < 0.8$,

— "les variables sont fortement corrélées" si $0.8 \leq |r|$.

Exercice 3. Soient X et Y deux variables aléatoires discrètes. On suppose que X suit la loi $\mathcal{P}(1)$ et que, pour tout $n \in \mathbb{N}$, Y suit la loi $\mathcal{G}\left(\frac{1}{2+n}\right)$ sachant $[X = n]$. Ecrire une fonction en Python qui, étant donné un entier $m \geq 3$, réalise et affiche m simulations du couple (X, Y), puis qui trace le nuage de points et la droite de régression de X par rapport à Y, et enfin en calcule et affiche le coefficient de corrélation linéaire.

Exercice 4. (Droite de régression de Y par rapport à X) Dans cet exercice, on se propose d'établir le théorème sur l'existence et l'unicité de la droite de régression de Y par rapport à X. Pour ce faire, on considère une série statistique double $((x_i, y_i))_{1\leq i\leq n}$ telle que les x_i ne sont pas tous égaux. On munit l'espace vectoriel $\mathcal{M}_{n,1}(\mathbb{R})$ du produit scalaire canonique, et l'on pose :

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad , \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad , \quad U = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Enfin, pour tout $(a, b) \in \mathbb{R}^2$, on pose : $Q(a, b) = \|Y - aX - bU\|^2 = \sum_{k=1}^{n}(y_i - ax_i - b)^2$.

(1) (a) Justifier qu'il existe un unique couple $(a_0, b_0) \in \mathbb{R}^2$ qui minimise $Q(a, b)$.

(b) Montrer que (a_0, b_0) est solution d'un système linéaire que l'on déterminera.

(c) En déduire les expressions de a_0, b_0 en fonction des x_i, des y_i et de N.

(2) Soit n un entier quelconque ≥ 2. On considère une classe comportant n étudiants, dont les notes à deux épreuves A et B sont présentées sous la forme d'une matrice $N \in \mathcal{M}_{n,2}(\mathbb{R})$. Toutes les notes sont comprises entre 0 et 20 et les deux colonnes de N correspondent aux listes x et y des notes aux deux épreuves. Ecrire une fonction en Python qui, étant donnée la matrice N des notes, calcule le coefficient de corrélation linéaire de x et y, puis trace le nuage de points et la droite de régression $\mathcal{D}_{y/x}$.

Exercice 5. Soit $p \in]0, 1[$. On considère un individu qui se déplace le long d'un axe gradué selon le protocole suivant. Initialement (c'est-à-dire à l'instant 0), l'individu se trouve au point d'abscisse 0. S'il se trouve à l'abscisse x_n à l'instant n, alors à l'instant $n + 1$, il peut soit avancer d'un pas avec probabilité $1 - p$, soit rester à sa position avec probabilité $p/2$, soit reculer d'un pas avec probabilité $p/2$.

(1) Ecrire une fonction en Python qui, étant donnés un entier $n \in \mathbb{N}^*$ et un réel $p \in]0, 1[$, réalise une simulation des n premières positions $x_1, ..., x_n$ de l'individu suivant sa position initiale, puis calcule et affiche le coefficient de corrélation linéaire ρ des points $M_k = (k, x_k)$ pour $1 \leq k \leq n$.

(2) Que dire de ρ si p est proche de 1 ? de 0 ? Justifier.