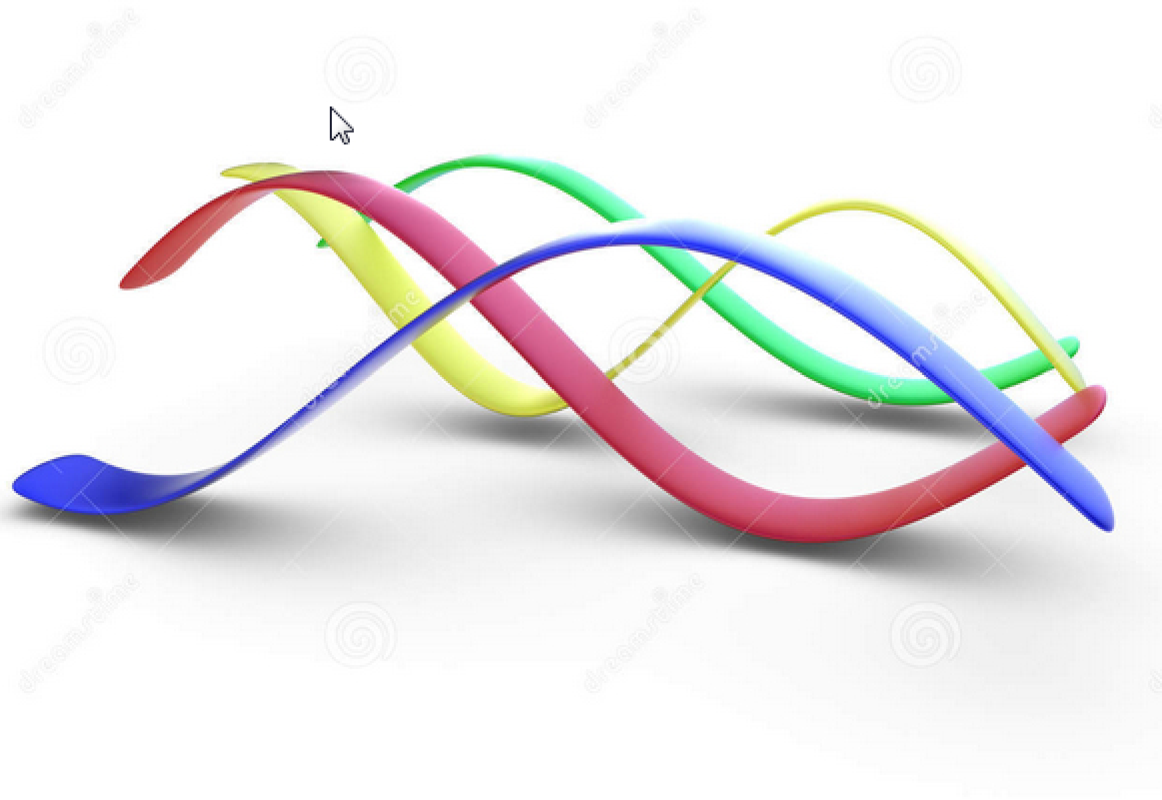
EN LETTRES CAPITALES **NOM(S) :**	**PRÉNOM(S) :**	**GROUPE :**

– Travaux pratiques de Mathématiques –

Ajustement Test



Etude 1 : A l'aide d'un tableur (Excel)

On s'intéresse à l'évolution de la population d'un village entre 1912 et 2001. Des recensements sont régulièrement effectués. Les résultats de ces recensements sont collectés dans le tableau ci-après.

Année : x_i	1912	1920	1928	1931	1940	1961	1990	2001
Nombre d'habitants : y_i	1534	2886	3467	3821	5067	7206	11350	12590

On souhaite prédire le nombre d'habitants du village en 2016.

Travail demandé

1. Tracer la courbe représentant l'évolution de la population de ce village.
2. Déterminer le nombre d'habitants en 2016, tenant compte de cette évolution.

Etude 2 : A l'aide d'un programme (Scilab)

L'objectif est d'écrire un programme permettant de réaliser une régression linéaire d'un jeu de données tout en testant si la régression a un sens ou non.

Travail demandé

1. Ecrire un programme permettant de réaliser une régression linéaire d'un jeu de données.
2. Tracer le nuage de points du jeu de données ainsi que la droite de régression linéaire.
3. Lancer votre programme pour les deux cas ci-après (Cas N°1 et Cas N°2).

Consignes à propos de ce programme

➔ Le programme doit permettre de déterminer les coefficients a et b ainsi que le coefficient de détermination R^2.

➔ Le programme doit permettre de déterminer si le jeu de données est linéaire. Si c'est le cas les coefficients a et b sont calculés sinon le programme renvoie un message.

CAS N°1

On donne les valeurs d'un indice boursier au premier de chaque mois entre les mois de janvier et de juillet d'une année donnée.

Mois	janvier	février	mars	avril	mai	juin	juillet
Rang : x_i	1	2	3	4	5	6	7
Indice : y_i	7100	6900	6800	6600	6500	6350	6400

CAS N°2

Une entreprise de Génie civil achète une machine-outil pour la réalisation de pièces d'assemblages destinées à la construction de charpentes métalliques. Cette machine coûte 10 000 euros à l'achat. Le tableau d'amortissement ci-dessous donne la valeur y en euros de cette machine après x années d'utilisation :

Année : x_i	1	2	3	4	5	6	7	8	9	10
Valeur : y_i	8500	7230	6140	5220	4440	3770	3200	2720	2300	1970

Remarques :

1. Pour saisir une liste de valeurs, il suffit d'utiliser la typographie suivante :
`x = [1,5,7,10,14]` permet de créer un vecteur ligne comportant 5 colonnes.
`x(1,4)` renvoie 10
`x(1,5)` renvoie 14

2. Créer des fonctions à l'aide de la commande « function ». Cela permet de simplifier l'écriture du programme principal.

3. Afficher 4 décimales pour les valeurs numériques calculées.

Droite de régression linéaire au sens des moindres carrés

Il s'agit de déterminer les réels a et b tels que l'écart quadratique moyen $EQ(a;b)$ soit minimal afin d'obtenir le coefficient directeur et l'ordonnée à l'origine de la droite, notée $D_{Y/X}$, la plus proche possible du nuage de points. Une telle droite est appelée droite de régression linéaire de Y en X au sens des moindres carrés.

• Covariance

La covariance de X et de Y est le nombre, noté $\operatorname{cov}(X,Y)$, défini par la formule suivante :

$$\operatorname{cov}(X,Y)=\frac{1}{N}\left(\sum_{i=1}^{N}\left(x_i-\overline{X}\right)\left(y_i-\overline{Y}\right)\right)=\frac{1}{N}\left(\sum_{i=1}^{N}x_i.y_i\right)-\overline{X}\times\overline{Y}$$

avec : $\overline{X}=\frac{1}{N}\sum_{i=1}^{N}x_i=\frac{x_1+x_2+\ldots+x_N}{N}$ et $\overline{Y}=\frac{1}{N}\sum_{i=1}^{N}y_i=\frac{y_1+y_2+\ldots+y_N}{N}$

• Écart quadratique moyen minimal

Les valeurs de a et b telles que l'écart quadratique moyen $EQ(a;b)$ soit minimal ont les expressions suivantes :

$$a=\frac{\operatorname{cov}(X,Y)}{Var(X)}$$

avec $\operatorname{cov}(X,Y)=\frac{1}{N}\left(\sum_{i=1}^{N}x_i.y_i\right)-\overline{X}\times\overline{Y}$ et $Var(X)=\left(\frac{1}{N}\sum_{i=1}^{N}{x_i}^2\right)-\overline{X}^2$

Le point moyen appartient à la droite $D_{Y/X}$, on en déduit la valeur de b :

$$b=\overline{Y}-a\overline{X}$$

• Remarque

La droite de régression linéaire de Y en X au sens des moindres carrés passe toujours par le point moyen du nuage $G\left(\overline{X};\overline{Y}\right)$.

Coefficient de détermination – Coefficient de corrélation

On note $EQ_{\min}$ la valeur minimale de l'écart quadratique moyen $EQ(a;b)$ pour les valeurs suivantes de a et de b :

$$a = \frac{\text{cov}(X,Y)}{Var(X)} \quad \text{et} \quad b = \overline{Y} - a\overline{X}$$

Pour ces valeurs de a et de b, la valeur minimale $EQ_{\min}$ de l'écart quadratique moyen a l'expression suivante :

$$EQ_{\min} = Var(Y) \times \left(1 - r^2\right)$$

avec $r^2 = \dfrac{\text{cov}^2(X,Y)}{Var(X) \times Var(Y)}$ et $Var(Y) = \left(\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} y_i^{\,2}\right) - \overline{Y}^2$

• **Coefficient de détermination**

Le coefficient r^2 est appelé coefficient de détermination. On a les inégalités suivantes :

$$0 \le r^2 \le 1 \quad \text{et} \quad 0 \le EQ_{\min} \le Var(Y)$$

• **Remarques**

1. Si la valeur de r^2 est proche de 1 alors la valeur de $EQ_{\min}$ est proche de 0. Dans ce cas, les points du nuage sont tous très proches de la droite $D_{Y/X}$.

2. Si la valeur de r^2 est proche de 0 alors la valeur de $EQ_{\min}$ est élevée comparativement à la variance de Y. Dans ce cas, on considère que la dispersion des valeurs y_i (dispersion mesurée par la variance de Y) n'est pas directement liée à la dispersion des valeurs x_i. Il semble alors peu probable que Y dépende linéairement de X.

3. En général, on conclue en faveur d'une dépendance linéaire de Y par rapport à X lorsque le coefficient de détermination r^2 est supérieur ou égal à 0.80. Dans le cas contraire, on rejette l'hypothèse d'une dépendance linéaire.

• **Coefficient de corrélation**

Le coefficient r est appelé coefficient de corrélation de Y et de X :

$$r = \frac{\text{cov}(X,Y)}{s(X) \times s(Y)} \quad \text{et} \quad -1 \le r \le 1$$

avec $s(X) = \sqrt{Var(X)}$ et $s(Y) = \sqrt{Var(Y)}$

La valeur de r est du signe du coefficient directeur a de la droite de régression linéaire $D_{Y/X}$.

Les fonctions et structures

➔ Structure répétitive

Pour k **de** 1 **jusqu'à** n **Faire**
{Traitement 1}
FinPour

➔ Structure alternative

Si {condition} **Alors**
{Traitement 1}
Sinon
{Traitement 2}
FinSi

➔ rand()

La fonction rand() permet de générer un nombre aléatoire strictement compris entre 0 et 1. La loi sélectionnée par défaut est la loi uniforme.

➔ modulo(a , m)

Renvoie le reste de la division euclidienne de a par m.
Exemple :
x = modulo(23 , 4) ; x contient la valeur 3 car 23 = 4*5 + 3
y = modulo(5 , 2) ; y contient la valeur 1
z = modulo(8 , 2) ; z contient la valeur 0

➔ t = zeros(1 , 100)

La fonction zeros(n , p) définit une matrice de n lignes et de p colonnes dont tous les termes sont nuls.
La fonction zeros(1 , 100) définit un vecteur ligne de 100 colonnes dont tous les termes sont nuls.
t(1 , 2) = 6 place la valeur 6 dans la deuxième colonne du vecteur ligne t.

➔ disp(t)

disp(t) : Affiche les éléments d'un vecteur ligne (ou d'une matrice ou d'une variable).

➔ Affichage de plusieurs variables

```
printf ("Encadrement : %f%s%f\n",a," < xsol < ",b);
```

L'affichage est le suivant : Encadrement : 3.412 < xsol < 3.413

La chaîne de caractères "%s%f%\n" est appelée chaîne de formatage :
%s : affichage d'une chaîne de caractères (string).
%i : affichage d'un nombre entier (integer).
%f : affichage d'un nombre réel (float).
\n : Effectue un retour à la ligne après l'affichage.
%0.8f : Force l'affichage du réel avec 8 décimales.

➔ function

L'instruction « function » permet de définir une fonction utilisateur.

```
function z = f(x)
z = x^2 + x + 2
endfunction
```

f(1) renvoie 4

➔ int

La fonction int() renvoie la troncature à l'unité d'un nombre.
int(3.5) renvoie 3

➔ return
L'instruction return permet de sortir d'une fonction et de retourner à l'endroit de l'appel.

➔ break
L'instruction break permet de sortir d'une boucle for (interruption d'une boucle).

➔ length(a)
length(a) revoie le nombre d'élément d'un vecteur ou d'une matrice.

➔ sum
Soit a = [5,8]
a(1,1) a pour valeur 5 et (1,2) a pour valeur 8.

s = sum(a) ; s est un reel et s = 13.
n = a.^2 ; n est un vecteur et n = [25,64]
p = sum(n) ; p est un réel et p = 89.

Exemple d'affichage de la console :

Entrer les valeurs des abscisses : x = [x1,x2,...,xn] ; x = [0,1,2,3,4,5,6]
Entrer les valeurs des ordonnées : y = [y1,y2,...,yn] ; y = [8.44,8.67,8.82,8.86,9.09,9.31,9.43]
Point moyen : G(3.0000 ; 8.9457)
Equation de la droite :
y = 0.1614 x + 8.4614