



LEÇON 06 : STATISTIQUE À DEUX VARIABLES

1. SITUATION D'APPRENTISSAGE

Dans le cadre des recherches pour un exposé, des élèves d'une classe de Terminale ont été accrochés par les informations suivantes :

La prévision météorologique est une science en pleine évolution. Elle a pour objectif de prédire un ensemble de paramètres comme la pluviométrie, la pression atmosphérique, la température, etc.

Le tableau suivant donne les pluviométries et températures moyennes de septembre 2018 à août 2019 dans une ville.

	Sept 2018	Oct 2018	Nov 2018	Déc 2018	Jan 2019	Fév 2019	Mar 2019	Avr 2019	Mai 2019	Juin 2019	Juillet 2019	Août 2019
Pluviométrie (en mm)	13	23	49	49	50	64	79	48	40	10	5	6
Température (en °C)	23	17	14	10	10	11	13	15	17	23	27	28

La température moyenne d'octobre 2019 était de 32 °C.

Ils décident alors de chercher à savoir si la pluviométrie est liée à la température et dans ce cas, prévoir la pluviométrie d'octobre 2019.

MOTIVATION

Dans les classes antérieures, nous avons étudié les séries statistiques à une variable (ou à un seul caractère).

Face à la situation de la pandémie de la maladie à Coronavirus 2019 (COVID-19), des chercheurs ont établi un lien entre le nombre de décès lié à ce virus et l'âge des personnes testées positives. Cette étude a donc porté sur deux caractères quantitatifs:

- X le nombre de décès
- et Y l'âge des personnes testées positives.

D'où l'importance de notre leçon : STATISTIQUE A DEUX VARIABLES

Nous allons la traiter en quatre parties :

Plan du cours

- I. Présentation de la série statistique double
- II. L'ajustement
- III. L'estimation
- IV. Exercice de synthèse

2. RÉSUMÉ DE COURS

I. Série statistique double

1. Définition

On considère deux caractères quantitatifs X et Y sur une même population de n individus.

On note : $x_1, x_2, x_3, \dots, x_p$ les valeurs (ou les modalités) du caractère X ; $y_1, y_2, y_3, \dots, y_p$ les valeurs du caractère Y et n_{ij} l'effectif du couple (x_i, y_j) .

On appelle série statistique double de caractère (X, Y), l'ensemble des triplets (x_i, y_j, n_{ij}) .

Exemple

Une étude statistique porte sur une population de 100 ménages. Deux caractères X et Y sont étudiés :

- le caractère X est le nombre d'enfants ;
- le caractère Y est le nombre de pièces de l'appartement occupé.

On obtient le tableau suivant qui représente le série statistique de caractère (X, Y).

Y X	1	2	3	4
0	6	3	1	0
1	4	11	3	1
2	1	10	16	3
3	0	5	13	5
4	0	1	4	8
5	0	0	1	4

Sur la 1^{ère} ligne, sont inscrites les valeurs du caractère Y ; soit $y_1 = 1 ; y_2 = 2 ; y_3 = 3 ; y_4 = 4$.

La 1^{ère} colonne affiche les valeurs du caractère X qui sont : $x_1 = 0 ; x_2 = 1 ; x_3 = 2 ; x_4 = 3 ; x_5 = 4 ; x_6 = 5$.

Les nombres du tableau qui ne sont pas sur la 1^{ère} ligne et la 1^{ère} colonne, représentent les différents effectifs n_{ij} des couples (x_i, y_j) .

Ainsi considérons le nombre 6 dans ce tableau. On constate qu'il est dans la colonne de la valeur 1 du caractère Y et dans la ligne de la valeur 0 du caractère X. On dit alors il y a 6 ménages qui n'ont aucun enfant et occupent un appartement d'une pièce.

On dit que l'effectif du couple (0 ; 1) est 6.

Combien de ménages ont deux enfants et occupent un appartement de trois pièces ?

On va donc considérer la ligne ayant la valeur 2 du caractère X et la colonne ayant la valeur 3 du caractère Y. L'intersection de cette ligne et de cette colonne donne 16.

16 ménages ont donc deux enfants et occupent un appartement de trois pièces.

On dit que l'effectif du couple (2 ; 3) est 16.

Exercez-vous à la maison avec le reste des n_{ij} .

Ce tableau à double entrée ci-dessus est appelé tableau de contingence.

2. Tableau de séries marginales (*Série A1 seulement*)

Reprenons l'exemple précédent.

Il est question de trouver l'effectif de chaque modalité du caractère X et l'effectif de chaque modalité du caractère Y.

Y \ X	1	2	3	4	Total
0	6	3	1	0	10
1	4	11	3	1	19
2	1	10	16	3	30
3	0	5	13	5	23
4	0	1	4	8	13
5	0	0	1	4	5
Total	11	30	38	21	100

Considérons le caractère X

Pour trouver l'effectif de la valeur 0, on additionne tous les n_{ij} qui se trouvent sur la ligne de la valeur 0 du caractère X c'est-à-dire : $6+3+1+0=10$.

Quel est l'effectif de la valeur 3 du caractère X ?

Pour trouver l'effectif de la valeur 3 du caractère X, on additionne tous les n_{ij} qui se trouvent sur la ligne de la valeur 3 du caractère X c'est-à-dire : $0+5+13+5 = 23$.

On procède de la même manière pour trouver l'effectif des autres modalités du caractère X. Ainsi à chaque valeur du caractère X, on a son effectif dans la dernière colonne.

D'où le tableau linéaire associé à X :

x_i	0	1	2	3	4	5
n_i	10	19	30	23	13	5

La série ainsi obtenue est appelée série marginale du caractère X.

En faisant de même avec les colonnes, on obtient ~~par exemple~~ :

L'effectif de la modalité 1 du caractère Y en additionnant les n_{ij} de la colonne où se trouve cette modalité. Soit $6+4+1+0+0+0=11$.

On obtient ainsi l'effectif de chaque modalité du caractère Y dans la dernière ligne du tableau.

D'où le tableau linéaire associé à Y :

y_i	1	2	3	4
n_i	11	30	38	21

La série ainsi obtenue est appelée série marginale du caractère Y.

• Dressons le tableau des fréquences marginales du caractère X

On rappelle que la fréquence est l'effectif de la modalité sur l'effectif total

On obtient le tableau suivant :

x_i	0	1	2	3	4	5
f_i	$\frac{10}{100}$	$\frac{19}{100}$	$\frac{30}{100}$	$\frac{23}{100}$	$\frac{13}{100}$	$\frac{5}{100}$

De la même manière, définit le tableau des fréquences marginales du caractère Y.

y_i	1	2	3	4
f_i	$\frac{11}{100}$	$\frac{30}{100}$	$\frac{38}{100}$	$\frac{21}{100}$

3. Nuage de points

Définition

On considère deux caractères quantitatifs X et Y sur une même population de n individus.

On note $x_1, x_2, x_3, \dots, x_p$ les valeurs du caractère X,

$y_1, y_2, y_3, \dots, y_p$ les valeurs du caractère Y,

On appelle nuage de points associé à la série statistique double de caractère (X,Y) les points de couple de coordonnées $(x_i; y_j)$ d'effectifs non nuls.

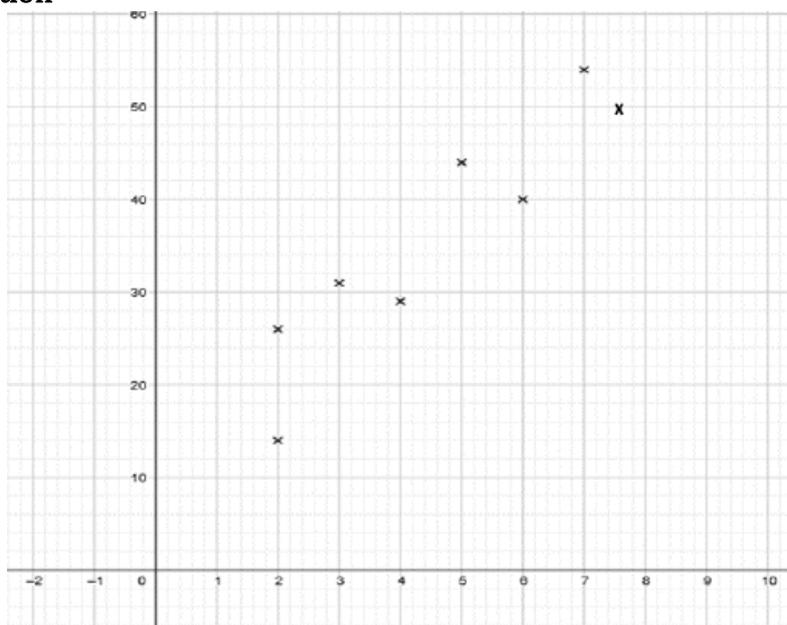
Exercice

Le tableau suivant donne le nombre d'exploitations agricoles d'une région selon leur superficie en hectares.

Superficie X	2	2	3	4	5	6	7	7,6
Nombre d'exploitations Y	14	26	31	29	44	40	54	50

Représente le nuage de points associé à cette série.

Solution



Remarque

Dans la suite, les séries doubles considérées seront comme la série de l'exemple précédent ; c'est-à-dire l'effectif n_{ij} de chaque couple (x_i, y_j) vaut 1.

4. Point moyen

Définition

On appelle point moyen d'un nuage de n points M_i de coordonnées $(x_i; y_i)$ le point G de coordonnées $(x_G; y_G)$ telles que :

$$x_G = \bar{X} = \frac{x_1+x_2+\dots+x_n}{n} ; y_G = \bar{Y} = \frac{y_1+y_2+\dots+y_n}{n}.$$

Exercice

Détermine les coordonnées du point moyen du nuage de points de la série statistique suivante :

Superficie X	2	2	3	4	5	6	7	7,6
Nombre d'exploitations Y	14	26	31	29	44	40	54	50

Solution

C'est le point de coordonnées $(\bar{X}; \bar{Y})$.

$$\text{On a : } \bar{X} = \frac{2+2+3+4+5+6+7+7,6}{8} = \frac{36,6}{8} = 4,575$$

$$\text{et } \bar{Y} = \frac{14+26+31+29+44+40+54+50}{8} = \frac{288}{8} = 36$$

Donc G(4,575 ; 36)

Exercice de maison

On considère la série statistique double suivante :

x_i	0	1	2	3	4	5	6	7	8
y_i	160	110	100	72	36	29	20	10	3

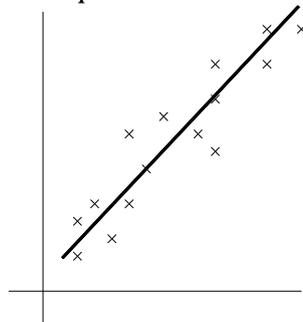
- 1) Représente le nuage de points de cette série statistique.
- 2) Détermine les coordonnées du point moyen de cette série statistique.

II. Ajustement

Soit un nuage de points associé à une série statistique double représenté dans un repère orthogonal.

Faire un ajustement de ce nuage de points, c'est trouver une courbe qui passe le plus près « possible » du maximum de points de ce nuage.

Lorsque cette courbe est une droite, on dit que l'ajustement est affine ou linéaire.



Exemple d'ajustement par une droite.

1. Ajustement linéaire par la méthode de Mayer

a) Droite d'ajustement

Pour déterminer la droite d'ajustement linéaire d'un nuage de points, on peut procéder comme suit :

- On range la série statistique double (X; Y) suivant les valeurs croissantes de x_i .
- On partage la série statistique en deux séries statistiques de même effectif : $(x_1; y_1), (x_2; y_2), \dots, (x_p; y_p)$ et $(x_{p+1}; y_{p+1}), (x_{p+2}; y_{p+2}), \dots, (x_n; y_n)$.
- On détermine le point moyen G_1 du premier sous-nuage et le point moyen G_2 du deuxième sous-nuage.
- La droite (G_1G_2) est la droite d'ajustement par la méthode de Mayer.

NB :

- La droite (G_1G_2) passe par le point moyen G du nuage de points.
- Si le nombre n de points du nuage de points est impair, alors on partage le nuage de points en deux sous-nuages d'effectif $\frac{n+1}{2}$ et $\frac{n+1}{2} - 1$.

Exercice

Partage la série statistique ci-dessous en deux séries et détermine le point moyen de chacune d'elles.

Superficie X	2	2	3	4	5	6	7	7,6
Nombre d'exploitations Y	14	26	31	29	44	40	54	50

Solution

Les valeurs du caractère X sont rangées dans l'ordre croissant.

L'effectif total de la série est 8.

On va partager la série en deux séries d'effectif 4 chacune.

Série 1

X	2	2	3	4
Y	14	26	31	29

Point moyen G_1

$$G_1(\bar{X}_1; \bar{Y}_1) \text{ avec } \bar{X}_1 = \frac{2+2+3+4}{4} = 2,75 \text{ et } \bar{Y}_1 = \frac{14+26+31+29}{4} = 25$$

Donc : $G_1(2,75; 25)$

Série 2

x_i	5	6	7	7,6
y_i	44	40	54	50

Point moyen G_2

$$G_2(\bar{X}_2; \bar{Y}_2), \text{ avec } \bar{X}_2 = \frac{5+6+7+7,6}{4} = 6,4 \text{ et } \bar{Y}_2 = \frac{44+40+54+50}{4} = 47$$

Donc : $G_2(6,4; 47)$

b) Equation

Soit $G_1(\bar{X}_1; \bar{Y}_1)$ et $G_2(\bar{X}_2; \bar{Y}_2)$ les points moyens des sous-nuages.

On détermine une équation de la droite (G_1G_2) à l'aide d'un vecteur directeur ou du coefficient directeur.

Exercice

On considère la série statistique précédente.

Détermine une équation de la droite d'ajustement linéaire du nuage de points par la méthode de Mayer.

Trace cette droite.

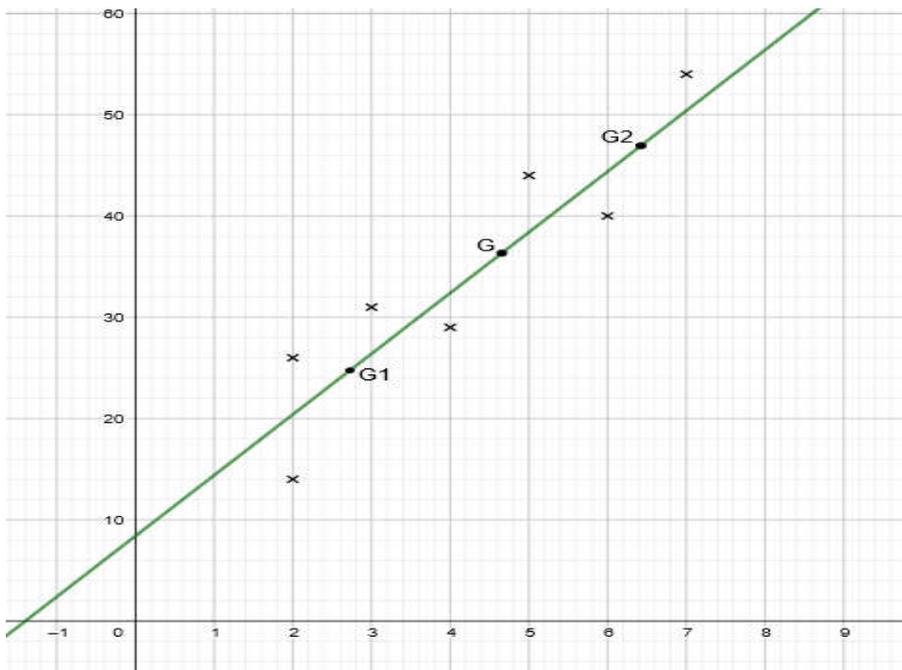
Solution

C'est la droite (G_1G_2) avec $G_1(2,75 ; 25)$ et $G_2(6,4 ; 47)$.

Elle a pour équation $y = ax + b$ où $a = \frac{47-25}{6,4-2,75} \approx 6$

et $b = 25 - 6 \times 2,75 = 8,4$

Donc $(G_1G_2) : y = 6x + 8,4$



2. Ajustement linéaire par la méthode des moindres carrés (Série A1 seulement)

a) Covariance

Définition

On appelle covariance de la série statistique double de caractère $(X ; Y)$, le nombre réel noté $COV(X ; Y)$ tel que :

$$COV(X, Y) = \frac{1}{n} \sum n_{ij} (x_i - \bar{X})(y_j - \bar{Y}) \text{ ou } COV(X, Y) = \frac{\sum n_{ij} x_i y_j}{n} - \bar{X} \bar{Y}.$$

Exercice

Calcule la covariance de la série statistique suivante.

Superficie X	2	2	3	4	5	6	7	7,6
Nombre d'exploitations Y	14	26	31	29	44	40	54	50

Solution

La covariance $COV(X, Y)$ de cette série statistique est $\frac{\sum x_i y_j}{n} - \bar{X} \bar{Y}$
 $COV(X, Y) = \frac{2 \times 14 + 2 \times 26 + 3 \times 31 + 4 \times 29 + 5 \times 44 + 6 \times 40 + 7 \times 54 + 7,6 \times 50}{8} - 4,575 \times 36$
 $COV(X, Y) = \frac{1503}{8} - 164,7$
 Donc : $COV(X, Y) = 23,675$

b) Coefficient de corrélation linéaire

Définition

Soit $V(X)$ la variance de la série statistique de caractère X, $V(Y)$ la variance de la série statistique de caractère Y et $COV(X ; Y)$ la covariance de la série statistique (X ; Y).
 On appelle coefficient de corrélation linéaire de la série statistique double (X ; Y), le nombre réel noté r tel que : $r = \frac{COV(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$.

Exercice

Calcule le coefficient de corrélation linéaire de la série statistique suivante.

Superficie X	2	2	3	4	5	6	7	7,6
Nombre d'exploitations Y	14	26	31	29	44	40	54	50

Solution

Le coefficient de corrélation linéaire r de cette série statistique est : $r = \frac{COV(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$

On a :

- $V(X) = \frac{\sum x_i^2}{n} - (\bar{X})^2 = \frac{2^2+2^2+3^2+4^2+5^2+6^2+7^2+(7,6)^2}{8} - 4,575^2$
 $V(X) = \frac{200,76}{8} - 4,575^2 \approx 4,16$
- $V(Y) = \frac{\sum y_i^2}{n} - (\bar{Y})^2 = \frac{14^2+26^2+31^2+29^2+44^2+40^2+54^2+50^2}{8} - 36^2$
 $V(Y) = \frac{11626}{8} - 36^2 = 157,25$

Donc : $r = \frac{23,675}{\sqrt{4,16 \times 157,25}} \approx 0,92$

Remarques

- Le coefficient de corrélation linéaire permet de voir la dépendance linéaire des deux caractères X et Y.
- Le coefficient de corrélation linéaire r est un nombre réel de même signe que $COV(X,Y)$ et on a : $-1 \leq r \leq 1$.
- Si $|r|$ est proche de 1, c'est-à-dire en pratique : $0,87 \leq r < 1$ ou $-1 < r \leq -0,87$, alors on dit qu'il y a une bonne corrélation linéaire ou une forte corrélation linéaire entre les deux caractères X et Y.

Exercice

Interprète le coefficient de corrélation linéaire ci-dessus.

Solution

On a : $r = 0,92$.

Comme $0,87 \leq r < 1$, il y a une forte corrélation entre la superficie et le nombre d'exploitations agricoles de cette région.

c) Droites de régressions

Propriété

Soit $V(X)$ la variance de la série statistique de caractère X , $V(Y)$ la variance de la série statistique de caractère Y et $\text{COV}(X, Y)$ la covariance de X et Y .

i. Droite de régression de Y en X .

En supposant qu'il y ait une forte corrélation entre les caractères X et Y alors, la droite (D) d'équation : $y = ax + b$ où $a = \frac{\text{Cov}(X,Y)}{V(X)}$ et $b = \bar{Y} - a\bar{X}$ est appelée la droite de régression de Y en X par la méthode des moindres carrés.

ii. Droite de régression de X en Y .

La droite (D') d'équation : $x = a'y + b'$ avec : $a' = \frac{\text{Cov}(X,Y)}{V(Y)}$ et $b' = \bar{X} - a'\bar{Y}$ est appelée la droite de régression de X en Y par la méthode des moindres carrés.

Exercice

On considère la série statistique précédente.

On sait que : $0,87 \leq r < 1$.

1. Détermine une équation de la droite d'ajustement linéaire de Y en X par la méthode des moindres carrés. (On donnera les arrondis d'ordre 2 de a et b .)
2. Détermine une équation de la droite d'ajustement linéaire de X en Y par la méthode des moindres carrés. (On donnera les arrondis d'ordre 2 de a' et b' .)

Solution

1. C'est la droite (D) d'équation : $y = ax + b$ où $a = \frac{\text{Cov}(X,Y)}{V(X)}$ et $b = \bar{Y} - a\bar{X}$

$$a = \frac{\text{Cov}(X,Y)}{V(X)} = \frac{23,675}{4,16} = 5,69 \quad \text{et} \quad b = \bar{Y} - a\bar{X} = 36 - 5,69 \times 4,575 = 9,97$$

Donc (D) : $y = 5,69x + 9,97$

2. C'est la droite (D') d'équation : $x = a'y + b'$ avec $a' = \frac{\text{Cov}(X,Y)}{V(Y)}$ et $b' = \bar{X} - a'\bar{Y}$

$$a' = \frac{\text{Cov}(X,Y)}{V(Y)} = \frac{23,675}{157,25} = 0,15 \quad \text{et} \quad b' = \bar{X} - a'\bar{Y} = 4,575 - 0,15 \times 36 = -0,825$$

Donc : (D') : $x = 0,15y - 0,83$

Remarques

- Les droites (D) et (D') passent par le point moyen G du nuage de points.
- Si r est le coefficient de corrélation linéaire on a :
 - $aa' = r^2$ et $|r| = \sqrt{aa'}$
 - Si $a > 0$ et $a' > 0$, alors $r = \sqrt{aa'}$.
 - Si $a < 0$ et $a' < 0$, alors $r = -\sqrt{aa'}$.
 - Si $r^2 = 1$, alors $a = \frac{1}{a'}$ et les deux droites sont confondues.

III. Estimation

- La droite d'ajustement tracée du nuage de points permet graphiquement une estimation de y connaissant x (resp. x connaissant y).
- L'équation de la droite d'ajustement permet de calculer une estimation de y connaissant x (resp. x connaissant y).

Exercice

On considère la série statistique précédente.

En considérant que la tendance se poursuit ainsi, détermine le nombre d'exploitations agricoles pour une superficie de 9 ha.

Solution

Une superficie de 9 ha correspond à $x = 9$.

En utilisant l'équation de la droite de Mayer, on a : $y = 6 \times 9 + 8,4 = 62,4$

Donc pour une superficie de 9 ha, le nombre d'exploitations est estimé à 63.

Série A1 seulement

En utilisant l'équation de la droite par la méthode des moindres carrés, on a :

$$y = 5,69x + 9,97$$

$$y = 5,69 \times 9 + 9,97 = 61,8$$

Donc pour une superficie de 9 ha, le nombre d'exploitations agricoles est estimé à 62.

Exercices de maison

Exercice 1

On considère la série statistique suivante :

x_i	0	1	2	3	4	5	6	7	8
y_i	160	110	100	72	36	29	20	10	3

- 1) Détermine une équation de la droite d'ajustement linéaire du nuage de points par la méthode de Mayer.
- 2) Trace cette droite.

Exercice 2 (Série A1 seulement)

On considère la série statistique suivante :

x_i	0	1	2	3	4	5	6	7	8
y_i	160	110	100	72	36	29	20	10	3

- 1) Détermine la covariance de la série statistique.
- 2) Détermine le coefficient de corrélation linéaire de cette série.
- 3) Interprète ce coefficient de corrélation linéaire.
- 4) Détermine une équation de la droite d'ajustement linéaire de Y en X du nuage de points de la série par la méthode des moindres carrés.
- 5) Détermine une équation de la droite d'ajustement linéaire de X en Y.

Exercice de synthèse

Le tableau suivant donne le chiffre d'affaires X (en millions de francs) réalisé au cours des 6 derniers mois par un site de vente en ligne en fonction du nombre de commandes Y reçues.

Nombre de commandes (x_i)	6 400	8 350	9 125	9 600	10 050	12 000
Chiffre d'affaires mensuel (y_i)	250	320	335	350	370	400

Partie A

- 1) Représente le nuage de points associé à cette série statistique (X,Y).
- 2) Détermine les coordonnées du point moyen de ce nuage.
- 3) a) Détermine une équation de la droite d'ajustement linéaire du nuage de points par la méthode de Mayer.
 b) Trace cette droite.
- 4) En considérant que la tendance se poursuit ainsi, détermine le chiffre d'affaires pour 15 000 commandes reçues.

Partie B (Série A1 seulement)

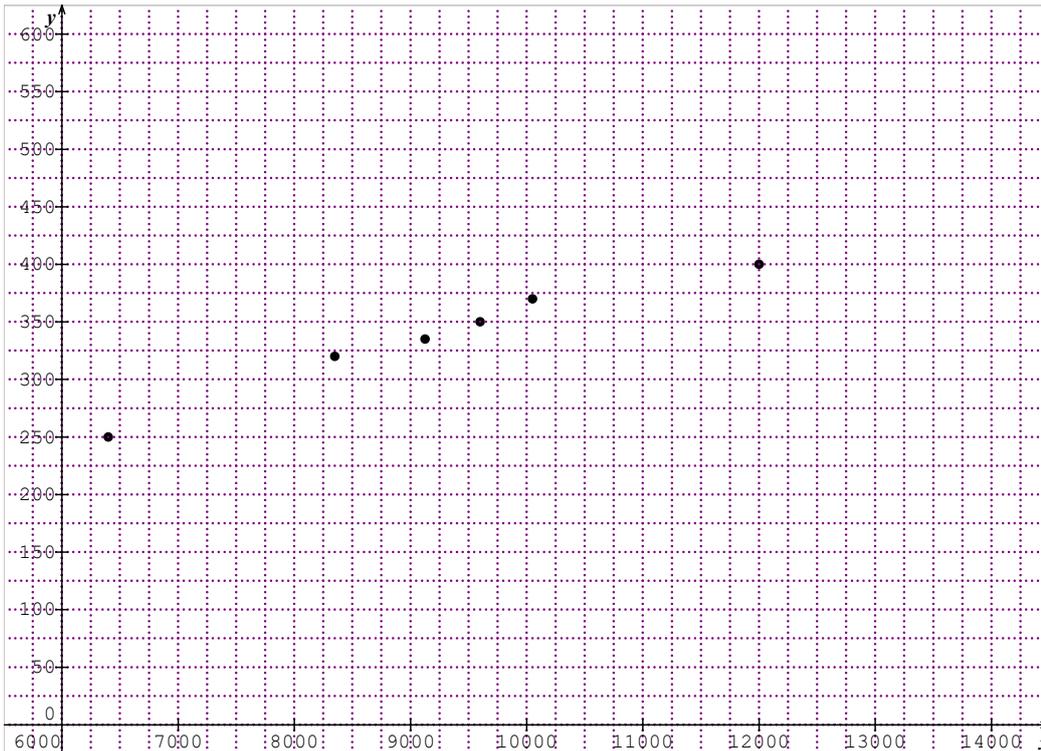
- 1) Calcule la covariance de la série statistique (X,Y).
- 2) Calcule le coefficient de corrélation linéaire de la série statistique (X,Y).
- 3) Interprète le coefficient de corrélation linéaire.
- 4) Détermine une équation de la droite d'ajustement linéaire de Y en X du nuage de points par la méthode des moindres carrés.
- 5) Détermine une équation de la droite d'ajustement linéaire de X en Y du nuage de points par la méthode des moindres carrés.
- 6) En considérant que la tendance se poursuit ainsi, détermine le chiffre d'affaires pour 15 000 commandes reçues.

Corrigé

Partie A

- 1) Nuage de points associé à la série

Nombre de commandes (x_i)	6 400	8 350	9 125	9 600	10 050	12 000
Chiffre d'affaires mensuel (y_i)	250	320	335	350	370	400



2) Point moyen G

Soit $G(\bar{X}; \bar{Y})$ le point moyen . On a :

$$\bar{X} = \frac{6400 + 8350 + 9125 + 9600 + 10050 + 12000}{6} = 9254,17$$

$$\bar{Y} = \frac{250 + 320 + 335 + 350 + 370 + 400}{6} = 337,5$$

Donc : $G(9254,17 ; 337,5)$

3) a) Une équation de la droite d'ajustement linéaire du nuage de points par la methode de mayer

Série 1

x_i	6400	8350	9125
y_i	250	320	335

Soit $G_1(\bar{X}_1; \bar{Y}_1)$ le point moyen de la série 1. On a :

$$\bar{X}_1 = \frac{6400 + 8350 + 9125}{3} = 7958,33$$

$$\bar{Y}_1 = \frac{250 + 320 + 335}{3} = 301,66$$

Ainsi : $G_1 (7958,33 ; 301,66)$.

Série 2

x_i	9600	10050	12000
y_i	350	370	400

Soit $G_2(\bar{X}_2; \bar{Y}_2)$ le point moyen de la série 2. On a :

$$\bar{X}_2 = \frac{9600 + 10050 + 12000}{3} = 10550$$

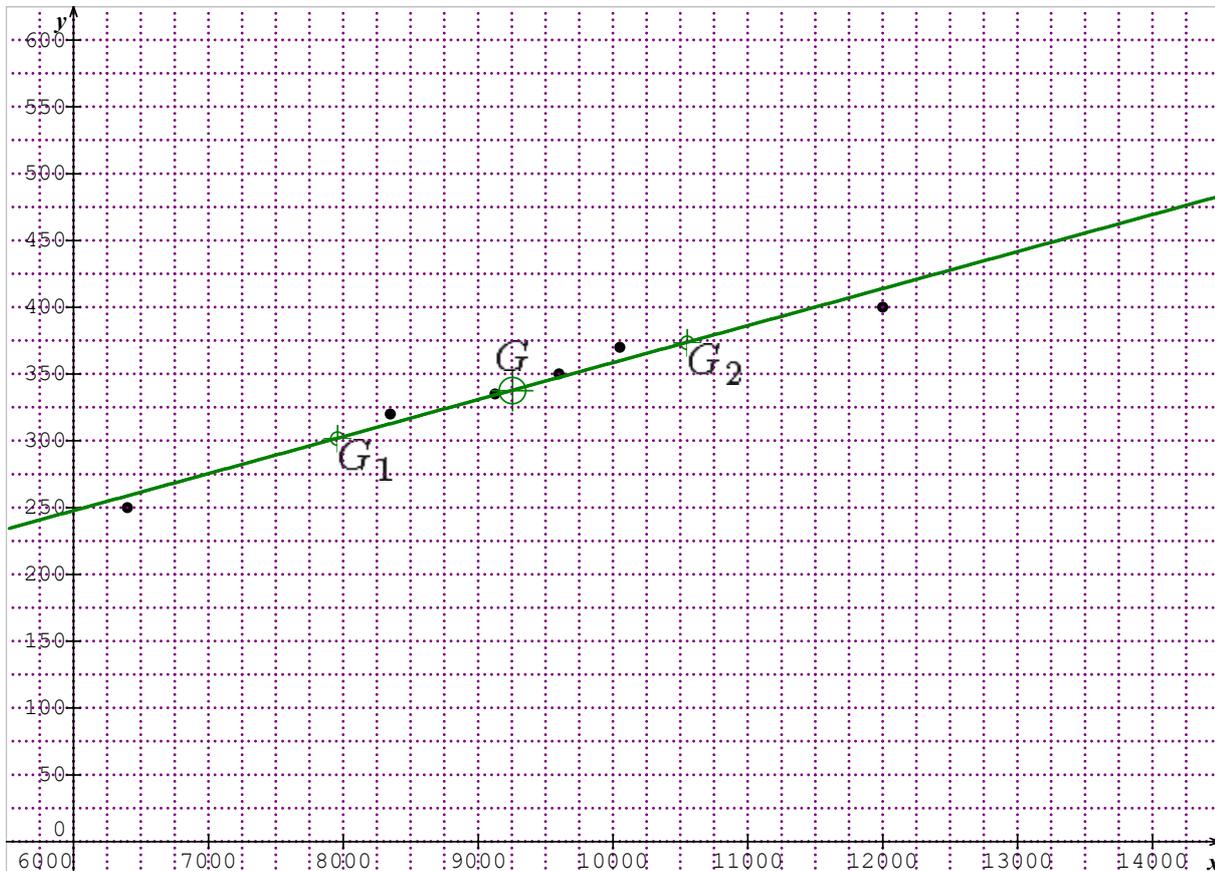
$$\bar{Y}_2 = \frac{350 + 370 + 400}{3} = 373,33$$

Ainsi : $G_2(10550 ; 373,33)$

$(G_1 G_2): y = ax + b$ avec $a = \frac{\bar{Y}_1 - \bar{Y}_2}{\bar{X}_1 - \bar{X}_2} = \frac{301,66 - 373,33}{7958,33 - 10550} = 0,028$ et $b = \bar{Y}_2 - a \bar{X}_2 = 78$

C'est-à-dire $(G_1 G_2): y = 0,028x + 78$

b) Tracé de la droite d'ajustement linéaire



4) Déterminons le chiffre d'affaires pour 15 000 commandes reçues.

Une commande de 15 000 correspond à $x = 15\ 000$.

L'équation de la droite de Mayer est : $y = 0,028x + 78$

On remplace x par 15 000 dans cette équation, on obtient $y = 498$.

Le chiffre d'affaires est d'environ 498 000 000 francs.

Partie B (Série A1 seulement)

1) Covariance de la série $(X ; Y)$

Pour calculer les moyennes, les variances et la covariance, on peut se servir du tableau suivant :

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1er mois	6400	250	40960000	62500	1600000
2e mois	8350	320	69722500	102400	2672000
3e mois	9125	335	83265625	112225	3056875
4e mois	9600	350	92160000	122500	3360000
5e mois	10050	370	101002500	136900	3718500
6e mois	12000	400	144000000	160000	4800000
Total	55525	2025	531110625	696525	19207375
Moyenne	9254,167	337,5	88518438	116088	3201229

La covariance

$$\text{COV}(X, Y) = \frac{\sum x_i y_j}{n} - \bar{X} \bar{Y} = \frac{19207375}{6} - 9254,17 \times 337,5 = 77\,947,9$$

2) Coefficient de corrélation r

$$V(X) = \frac{\sum x_i^2}{n} - \bar{X}^2 = \frac{531\,110\,625}{6} - (9254,167)^2 = 2\,878\,824,824$$

$$V(Y) = \frac{\sum y_i^2}{n} - \bar{Y}^2 = \frac{696\,525}{6} - (337,5)^2 = 2\,181,254$$

$$r = \frac{\text{COV}(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{77\,947,9}{\sqrt{2\,878\,824,824} \times \sqrt{2181,254}} = 0,98$$

3) Interprétation du coefficient de corrélation r

On remarque que : $0,87 \leq r < 1$, ainsi, on peut conclure qu'il y a une forte corrélation entre le nombre de commandes reçues et le chiffre d'affaires réalisé au cours des 6 derniers mois .

4) Une équation de la droite d'ajustement linéaire de Y en X par la méthode des moindres carrés

Puisqu'il y a une forte corrélation entre les commandes reçues et le chiffre d'affaires alors,

(D) a pour équation : $y = ax + b$ où $a = \frac{\text{Cov}(X,Y)}{V(X)}$ et $b = \bar{Y} - a\bar{X}$.

$$a = \frac{77947,9}{2878824,824} = 0,027 \text{ et } b = 337,5 - 0,027 \times 9254,17 = 87,63$$

Soit (D) : $y = 0,027x + 87,63$

5) Droite de régression de X en Y.

La droite (D') a pour équation : $x = a'y + b'$ avec : $a' = \frac{\text{Cov}(X,Y)}{V(Y)}$ et $b' = \bar{X} - a'\bar{Y}$.

$$a' = \frac{77947,9}{2181,254} = 35,735 \text{ et } b' = 9254,17 - 35,735 \times 337,5 = -2806,39$$

Soit (D') : $x = 35,735y - 2806,39$

6) Estimation du chiffre d'affaires pour 15 000 commandes reçues

Ici, $x = 15\,000$ et en utilisant l'équation de la droite de régression de Y en X, on a :

$$y = 0,027 \times 15000 + 87,63$$

$$y = 492,63$$

Pour 15 000 commandes reçues, le site de vente en ligne fera un chiffre d'affaires d'environ 492 630 000 francs.

3. EXERCICES

3-1. Exercices de fixation

Exercice 1

On considère la série statistique suivante :

x_i	1	4	7	8	10
y_i	2	7	8	10	13

Représente le nuage de points de cette série statistique.

Exercice 2

On considère la série statistique suivante :

x_i	0	1	2	3	4	5	6	7
y_i	160	110	100	72	36	29	20	10

Représente le nuage de points de cette série statistique

Exercice 3

Détermine le point moyen de la série de l'exercice 1.

Exercice 4

Détermine le point moyen de la série de l'exercice 2.

Exercice 5

Détermine une équation de la droite d'ajustement linéaire du nuage de points de l'exercice 1 par la méthode de Mayer.

Exercice 6

Détermine une équation de la droite d'ajustement linéaire du nuage de points de l'exercice 2 par la méthode de Mayer.

Exercice 7 (Série A1 seulement)

Détermine la covariance de la série de l'exercice 2.

Exercice 8 (Série A1 seulement)

Détermine le coefficient de corrélation linéaire de la série de l'exercice 2.

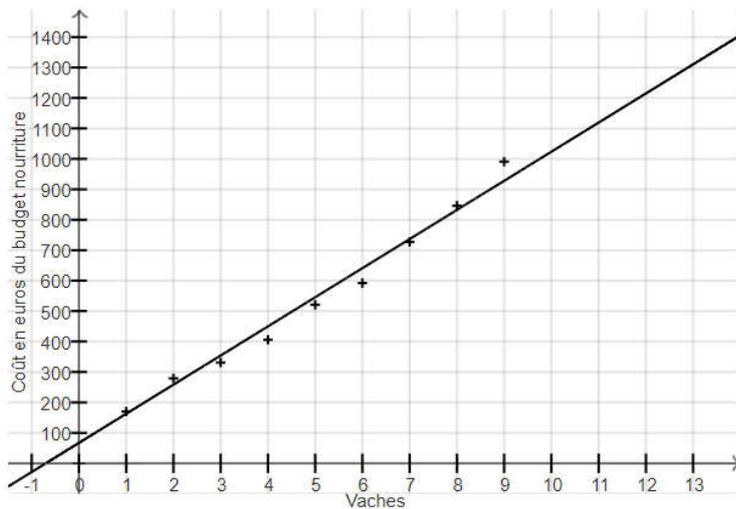
Exercice 9 (Série A1 seulement)

Détermine une équation de la droite d'ajustement linéaire de Y en X du nuage de points de la série de l'exercice 2 par la méthode des moindres carrés.

Exercice 10

Un agriculteur a estimé son budget annuel alloué, en euros, à la nourriture de ses bovins en fonction de la taille de son troupeau. Il sait que son troupeau va encore grandir d'ici 2 ans. Il a estimé qu'alors, son troupeau comportera 10 individus.

En extrapolant, détermine, à 100 euros près, le budget nourriture de l'agriculteur dans deux ans.



3-2. Exercices de renforcement

Exercice 11

La tension artérielle est une donnée médicale correspondant à la pression du sang dans les artères. On la mesure chez les patients car une tension anormale peut-être le symptôme de pathologies cardiovasculaires comme l'hypertension artérielle.

La tension artérielle d'une personne comporte deux mesures :

- la Tension Artérielle Systolique (notée TAS)
- la Tension Artérielle Diastolique (notée TAD).

Le tableau suivant regroupe les mesures de la tension artérielle pour un groupe de personnes saines :

Age	26	39	40	50	53	56
TAS (en mmHg)	128	126	118	136	142	145
TAD (en mmHg)	80	83	92	91	87	93

On s'intéresse à l'évolution de la TAS en fonction de l'âge.

Pour cela on symbolise les données du tableau à l'aide de points de coordonnées $(x;y)$ où x est l'âge de la personne et y sa TAS.

Détermine les coordonnées du point moyen des 3 points dont l'âge est le plus petit.

Détermine les coordonnées du point moyen des 3 autres points.

Exercice 12

Dans le tableau ci-dessous, on donne la taille moyenne (en cm) des nouveaux nés en fonction du nombre de l'âge gestationnel (en semaines).

Âge gestationnel (semaines)	30	31	32	33	34	35	36	37
Taille (cm)	47,5	48,5	49	49,7	50	50,5	50,8	51,2

Âge gestationnel (semaines)	38	39	40	41	42	43	44	45
Taille (cm)	51,5	51,8	52,2	52,5	52,8	53	53,5	53,7

1. Représente le nuage de points dans un repère orthogonal en prenant comme unités :
 En abscisses : 1 cm pour 1 semaine (commencer la graduation à 20 semaines)
 En ordonnées : 2 cm par unité (commencer la graduation à 45 cm)
2. On se propose de tracer la droite d'ajustement de ce nuage de points.
 - a) Calcule les coordonnées des points moyens G_1 et G_2
 - b) Trace la droite d'ajustement passant par les points G_1 et G_2 .
3. Détermine une équation de cette droite d'ajustement.

Exercice 13

Le tableau suivant donne le chiffre d'affaires (en millions de francs) réalisé au cours des 6 derniers mois par un site de vente en ligne en fonction du nombre de commandes reçues.

Nombre de commandes (x_i)	6 400	8 350	9 125	9 600	10 050	12 000
Chiffre d'affaires mensuel (y_i)	250	320	335	350	370	400

- 1) Représente le nuage de points associé à cette série.
- 2) Détermine les coordonnées du point moyen de ce nuage.
- 3) Détermine une équation de la droite d'ajustement linéaire du nuage de points par la méthode de Mayer.
Trace cette droite.
- 4) Calcule la covariance de la série statistique. (*Série A1 seulement*)
- 5) Calcule le coefficient de corrélation linéaire de la série statistique. (*Série A1 seulement*)
Interprète le coefficient de corrélation linéaire. (*Série A1 seulement*)
- 6) Détermine une équation de la droite d'ajustement linéaire de Y en X du nuage de points par la méthode des moindres carrés. (*Série A1 seulement*)
- 7) Détermine une équation de la droite d'ajustement linéaire de X en Y du nuage de points par la méthode des moindres carrés. (*Série A1 seulement*)
- 8) En considérant que la tendance se poursuit ainsi, détermine le chiffre d'affaires pour une commande de 15 000.

3-3. Exercices d'approfondissement

Exercice 14

Un chef d'entreprise reçoit de la part de ses collaborateurs la demande d'obtenir des véhicules de fonction plus confortables et plus puissants. Il sollicite alors son comptable afin que celui-ci examine la demande et sa faisabilité.

Le comptable utilise le tableau ci-dessous, donnant le prix de revient kilométrique (PRK) des véhicules d'une puissance fiscale de 4 à 8 CV et en fait une projection sur les véhicules plus puissants.

Puissance fiscale des véhicules (CV)	4	5	6	7	8
Prix de revient kilométrique (€)	0,424	0,471	0,492	0,513	0,555

- 1) Représente cette série statistique par un nuage.
- 2) Calcule les coordonnées du point moyen G.
- 3) On admet que la droite d'ajustement de cette série a pour équation : $y = 0,03x + 0,311$
 - a. Justifie que le point G appartient à cette droite.
 - b. Trace cette droite dans le repère précédent.
- 4) En utilisant la droite d'ajustement, détermine le prix de revient d'une voiture de 10 CV ? . Laisser apparents les traits nécessaires à la lecture.
- 5). Le comptable fixe le prix de revient kilométrique maximum à 0,650 €. Calcule la puissance maximale du véhicule qui correspond à cette exigence.

Exercice 15 (Série A1 seulement)

La consommation d'une voiture, z , est donnée en fonction de sa vitesse, x , par le tableau suivant :

x (en km/h)	80	90	100	110	120
z (en litres/ 100 km)	4	5	6,5	8	10

- 1) Examine la proportionnalité des deux grandeurs variables consommation et vitesse. Justifie ta réponse.
- 2) Complète le tableau ci-dessus, après l'avoir reproduit, par une ligne : $y = \ln z$ dont on donnera les valeurs approchées à 6 décimales (les meilleures possibles).
- 3) Dans un repère d'origine 0 ($x_0 = 70$; $y_0 = 1,30$ en prenant comme unités 1 cm pour 10 km/h en abscisse et 1 cm pour 0,10 en ordonnée, représente le nuage de points (x, y) .
- 4) Détermine une équation de la droite d'ajustement des points de coordonnées (x, y) du nuage, par la méthode des moindres carrés. Donne cette équation sous la forme $y = Bx + A$, avec les valeurs approchées de B et A (les meilleures possibles) à 3 décimales.
- 5) Estime y pour une vitesse de 140 km/h.
Estime la consommation aux 100 km pour cette vitesse de 140 km/h, à 0,5 L près comme dans le tableau initialement donné.

4. SITUATION COMPLEXE

Exercice 16

Dans le cadre des recherches pour un exposé, des élèves d'une classe de Terminale ont été accrochés par les informations suivantes :

La prévision météorologique est une science en pleine évolution. Elle a pour objectif de prédire un ensemble de paramètres comme la pluviométrie, la pression, la température, etc.

Le tableau suivant donne les pluviométries et températures moyennes de septembre 2018 à août 2019 d'une ville.

	Sept 18	Oct 18	Nov 18	Déc 18	Jan 19	Fév 19	Mars 19	Avril 19	Mai 19	Juin 19	Juillet 19	Août 19
Pluviométrie (en mm)	13	23	49	49	50	64	79	48	40	10	5	6
Température (en °C)	23	17	14	10	10	11	13	15	17	23	27	28

La température moyenne d'octobre 2019 était de 32 °C.

Ils décident alors de chercher à savoir si la pluviométrie est liée à la température et dans ce cas, prévoir la pluviométrie d'octobre 2019.

À l'aide des outils mathématiques au programme, justifie que la pluviométrie est liée à la température et détermine une estimation de la pluviométrie d'octobre 2019.