

**I°/ Définitions :**

Lorsqu'une étude est faite sur un ensemble à N individus présentant deux caractères X et Y quantitatifs discrets, on obtient une série statistique double  $(x_i; y_i)$

Avec  $(x_i)_{1 \leq i \leq N}$  les valeurs prises par X et  $(Y_i)_{1 \leq i \leq N}$  les valeurs prises par Y

Une série statistique est souvent présentée par un tableau appelé tableau à double entré.

**Paramètres d'une série statistiques**

F Valeur Moyenne de X (ou Moyenne Arithmétique de X) notée  $\bar{X}$  définie par :  $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$

Remarque : Si on utilise les distributions marginales on utilise la formule :  $\bar{X} = \frac{1}{n} \sum_{i=1}^n n_i x_i$  Avec  $n_i$  l'effectif de  $x_i$  ou le nombre de répétition de  $x_i$

F Variance de X noté  $V(X)$  définie par :  $V(X) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{X}^2$

Remarque : Si on utilise les distributions marginales on utilise la formule :

$V(X) = \frac{1}{n} \sum_{i=1}^n n_i x_i^2 - \bar{X}^2$  Avec  $n_i$  l'effectif de  $x_i$  ou le nombre de répétition de  $x_i$

F Ecart-type noté  $\sigma_x$  ou  $\sigma(X)$  définie par :  $\sigma_x = \sqrt{V(X)}$

Remarque : Cette formule explique bien que l'Ecart-type n'est définie que si :  $V(X) \geq 0$

**Nuage De Points**

F Soit  $(x_i; y_i)$  avec  $1 \leq i \leq n$  une série statistique double. On désigne par  $x_1, x_2, \dots, x_i, \dots, x_n$  les valeurs prises par X et  $y_1, y_2, \dots, y_i, \dots, y_n$  les valeurs prises par Y.

Dans le plan étant rapporté à un repère orthogonal. On appelle Nuage De Points associé à la série considérée l'ensemble des points  $M_i(x_i; y_i)_{1 \leq i \leq n}$ . Et on appelle Point Moyen du nuage le point de coordonnées  $\bar{X}$  et noté :  $G(\bar{X}; \bar{Y})$

Soit  $(x_i; y_i)$  avec  $1 \leq i \leq n$  une série statistique double. On appelle Covariance de X et Y notée  $\text{cov}(X, Y)$  le nombre définie par :  $\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})$

Remarque :  $\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i y_i) - \bar{X} \bar{Y}$  Est la plus commode pour les calculs "à la main"

Conséquences :

- $\text{cov}(X, Y) = \text{cov}(Y, X)$
- $\text{cov}(aX + b, Y) = a \text{cov}(Y, X); \forall a, b \in \mathbb{R}$
- $\text{cov}(X, X) = V(X)$
- $\text{cov}^2(X, Y) \leq V(X) V(Y)$

Remarque : La covariance permet une mesure de la dispersion des points du nuage par rapport au point moyen.

**II°/Nuage de points, Moyenne, Variance, Ecart-type, covariance et Coefficient de corrélation linéaire. (Réf 274)**

On considère le tableau suivant :

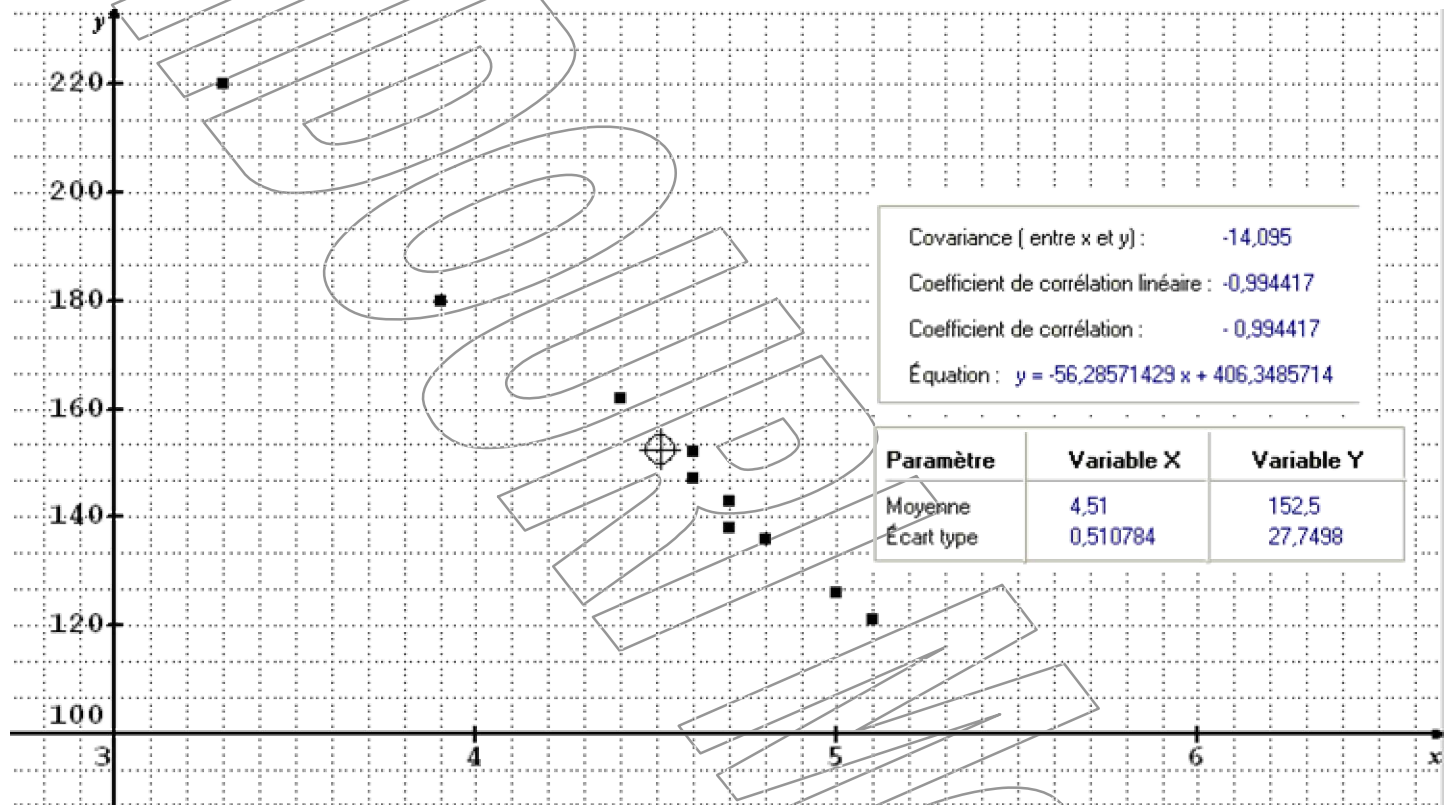
$X_i$	5,1	5	4,8	4,7	4,7	4,6	4,6	4,4	3,9	3,3
$Y_i$	121	126	136	138	143	147	152	162	180	220

1) Placer dans un repère orthogonal  $(O, \vec{i}, \vec{j})$  le nuage de points  $M_i(x_i; y_i)_{1 \leq i \leq 10}$

2) Comment semble réparti le nuage de points ? Commentez !

- 3) Calculer :  $\bar{X}$  la moyenne arithmétique de la série statistique à variable  $x_i$
- 4) Calculer :  $\bar{Y}$  la moyenne arithmétique de la série statistique à variable  $y_i$
- 5) Placer le point moyen  $G(\bar{X}, \bar{Y})$  dans le même repère
- 6) Comment semble réparti le nuage de points de par et d'autre du point  $G$  ? Commentez !
- 7) Existe-il un lien entre les variations de  $X$  et  $Y$  ?
- 8) Calculer la variance et l'écart-type de chacune des variables  $X$  et  $Y$

Correction :



**III°/Ajustements Linéaires :**

## 1) Coefficient de corrélation linéaire :

a) Définition : On appelle coefficient de corrélation linéaire du couple  $(X, Y)$ , le réel

$$r \text{ définie par : } r(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

## b) Propriétés :

- ∅ Si  $r(X, Y) = 0$  alors X et Y sont non corrélés (X et Y ne sont pas liées par une relation affine autrement un tel ajustement affine n'est pas possible)
- ∅ Si  $r(X, Y) \neq 0$  alors X et Y sont corrélés et un tel ajustement est possible. (Les ajustements affines ou linéaires ainsi que les non linéaires seront possibles sous des conditions sur  $r$ )
- ∅  $-1 \leq r(X, Y) \leq 1$
- ∅ Si  $|r(X, Y)| = 1$  alors il y a une dépendance totale entre X et Y. L'une est une fonction affine de l'autre.
- ∅ Si  $|r(X, Y)| < 0,75$  alors la corrélation linéaire entre X et Y est faible (On s'intéresse souvent d'autres ajustement non affines)
- ∅ Si  $|r(X, Y)| \geq 0,75$  alors la corrélation linéaire entre X et Y est forte
- ∅ Si  $|r(X, Y)| \in ]0,95; 1]$  alors la corrélation linéaire entre X et Y est très forte (Les ajustements non linéaires non aucune importances)

## c) Théorème (admis)

Soient X et Y deux variables d'une série statistiques quantitatives

Lorsque le coefficient de corrélation linéaire vérifie:  $0,70 \leq |r(X, Y)| \leq 1$ , ou lorsque Le Nuage De Points à Une Forme Allongée, alors il est possible d'approcher la liaison entre X et Y par deux fonctions affines représenter graphiquement par deux droites passant par le point moyen  $G(\bar{X}, \bar{Y})$

## 2) Méthode de Mayer (Réf 280)

## Activité :

Le tableau suivant donne le PNB (en euros, par habitants) ainsi que le nombre d'hôpitaux (pour 1 million d'habitants) dans quelques pays européens.

Pays	P1	P2	P3	P4	P5	P6	P7	P8
$x_i$ = PNB en euros par habitant	5100	7800	11200	15800	20100	22500	26200	28900
$y_i$ = Nombre d'hôpitaux par million d'habitants	620	1080	1550	2100	3000	3250	3800	4200

1) Représenter le nuage de points associé à la série statistique  $(X, Y)$ .

Unités graphiques :

- En abscisses : 7 cm pour 10000 euros.
- En ordonnées : 1,5 cm pour 400 hôpitaux.
- On prendra pour origine le point (5000 ; 500).

2) Déterminer les coordonnées du point moyen G de ce nuage de points. Placer G sur le graphique.

3) Déterminer le coefficient de corrélation linéaire entre X et Y. Un ajustement affine est-il justifié ?

## 4) Un premier ajustement affine : la Droite de Mayer

Dans cette question, on considère deux sous-nuages : celui constitué des points correspondants aux pays P1, P2, P3 et P4 et celui constitué des points correspondants aux pays P5, P6, P7 et P8.

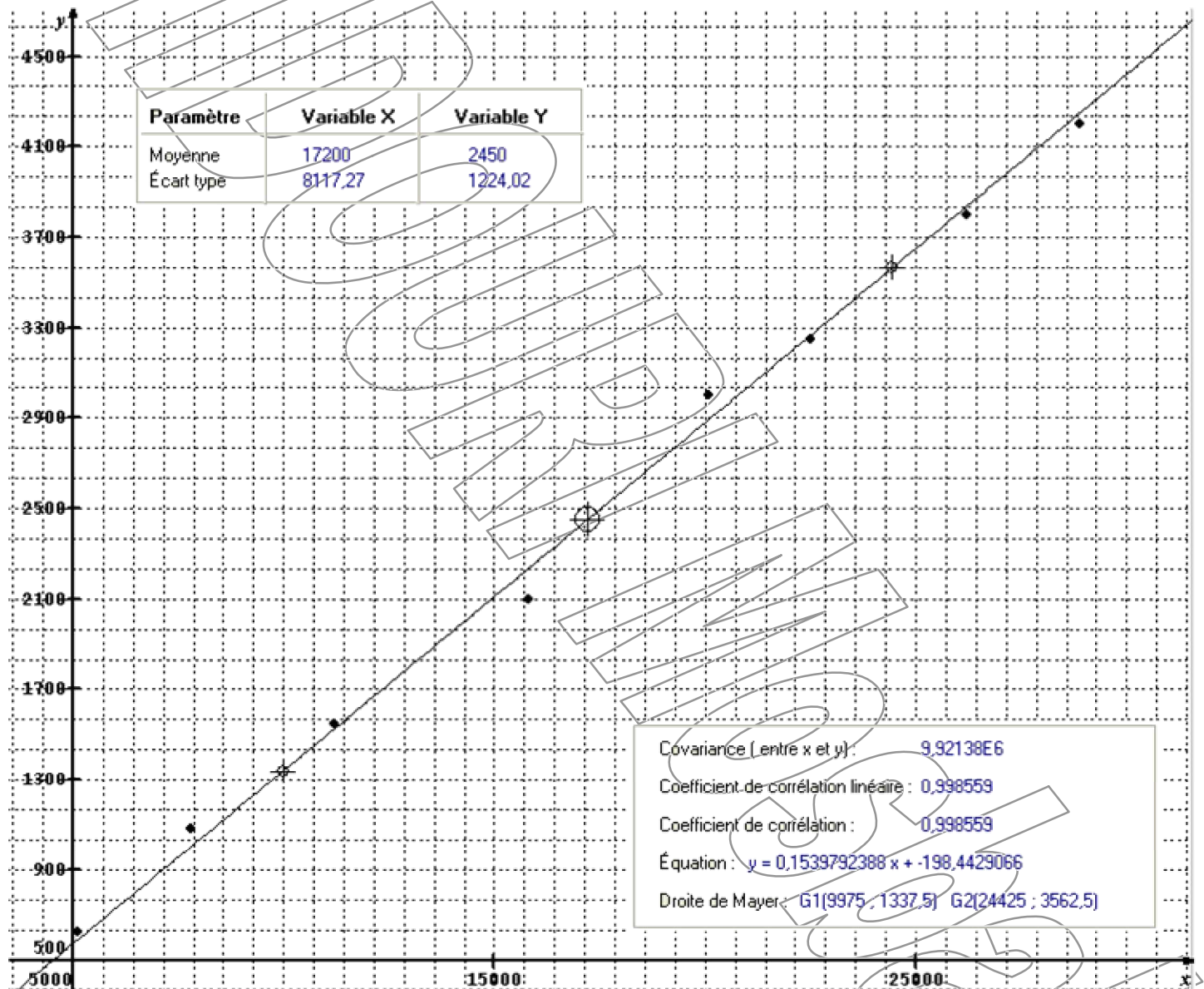
Ø Calculer les coordonnées des points moyens  $G_1$  et  $G_2$  des deux sous-nuages. Placer les points  $G_1$  et  $G_2$  sur le graphique.

Ø Démontrer qu'une équation de la droite  $(G_1G_2)$  sous la forme  $y = mx + p$  est :  
 $Y = 0,1539X - 198,4429$  (On détaillera les calculs).

(On arrondira  $m$  et  $p$  à  $10^{-4}$  près)

La droite  $(G_1G_2)$  s'appelle la "droite de Mayer". Représenter cette droite sur le graphique.

Correction :





**Commentaires**

a) Le principe d'ajustement affine par la méthode de Mayer consiste à partager le nuage associé à une série  $(x_i, y_i)$  en deux nuages dont le nombre de points diffère d'au plus un. On désigne par  $G_1$  et  $G_2$  les points moyens respectifs du première et du deuxième nuage. La droite  $(G_1G_2)$  est appelée droite de Mayer

5) Un deuxième ajustement affine : la Droite de Régression : Méthode des moindres carrés.

a) Déterminer une équation de la droite de régression de  $y$  en  $x$  par la méthode des moindres carrés. On notera  $D$  cette droite. Représenter  $D$  sur la graphique.

b) Laquelle des deux droites  $(G_1G_2)$  et  $D$  réalise-t-elle le meilleur ajustement affine ?

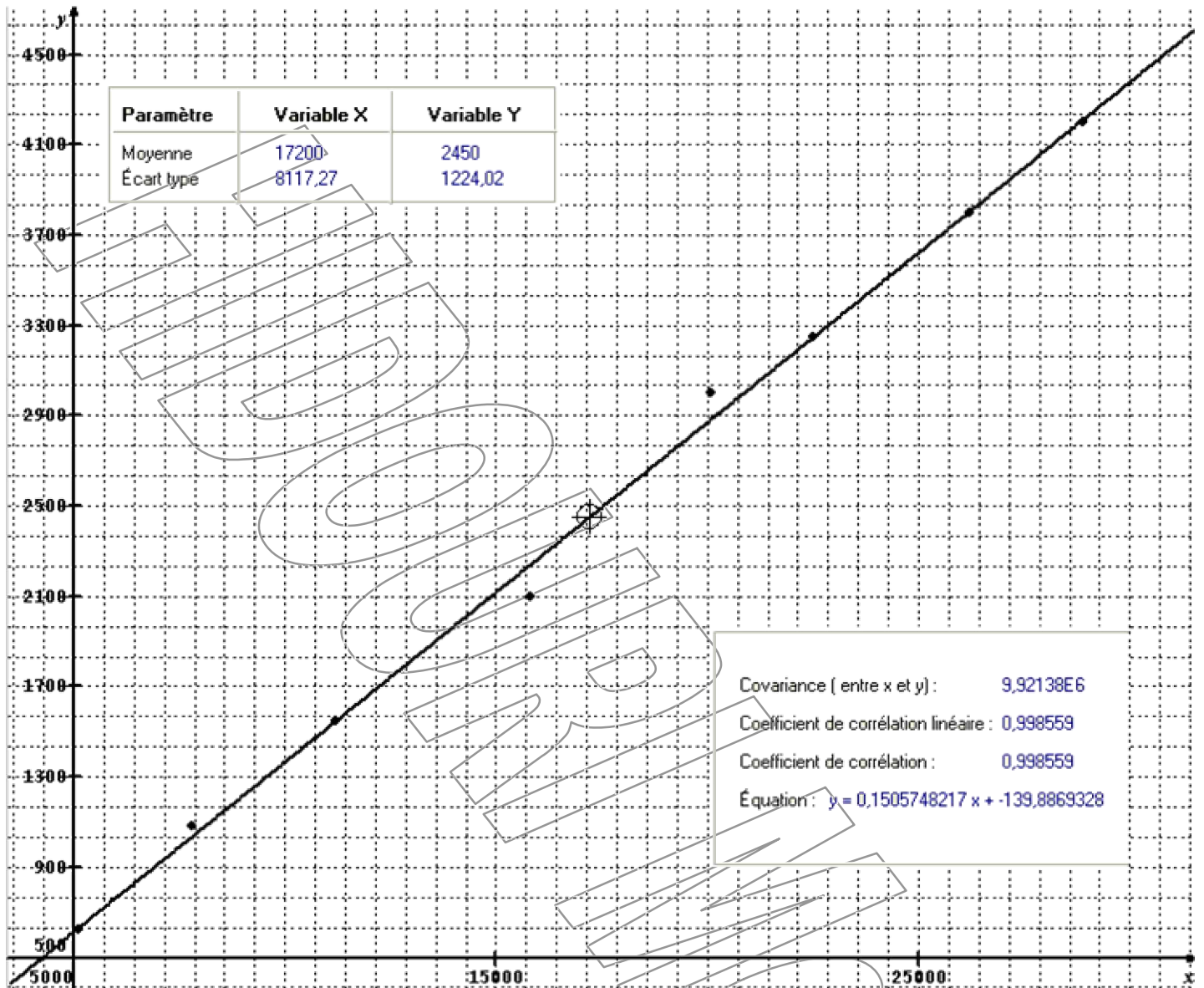
6) Estimations. À l'aide de l'équation de la droite  $(D)$  (ou à défaut celle de  $(G_1G_2)$ ), et en détaillant les calculs,

Répondre aux deux questions suivantes :

a) Un pays a un PNB de 23400 € par habitant. Quelle estimation peut-on faire du nombre d'hôpitaux (par million d'habitants) dans ce pays ? (On arrondira à l'unité près)

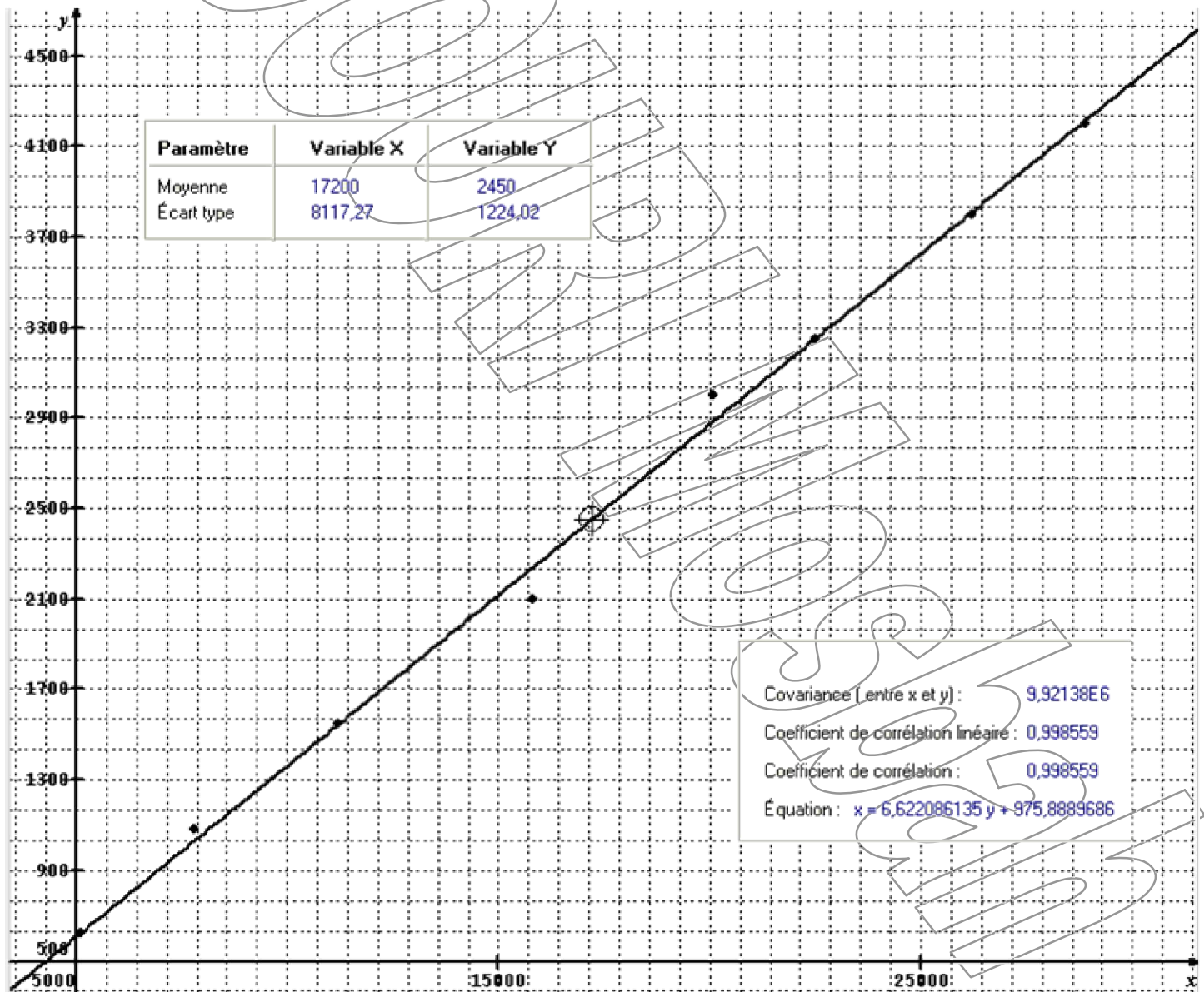
b) Un pays a 3500 hôpitaux par million d'habitants. À combien peut-on estimer son PNB (en €, par habitants) ? (On arrondira à l'euro près)

Correction :



Remarque:

On peut faire une régression de X en Y



## Commentaires

**Théorème et définition :**

∅ Lors d'un ajustement affine de Y en X par la méthode des moindres carrés la droite obtenue passe le point moyen du nuage et a pour équation :  $y = (ax - \bar{X}) + \bar{Y}$  où  $a = \frac{\text{cov}(X, Y)}{V(X)}$

∅ Cette droite s'appelle la droite de régression de Y en X

∅ La droite de régression de X en y obtenue par la méthode des moindres carrés, lors d'un ajustement affine, passe par le point moyen du nuage et à pour équation :  $x = (ay - \bar{Y}) + \bar{X}$  où

$$a' = \frac{\text{cov}(X, Y)}{V(Y)}$$

## Remarque :

Les deux droites de régressions de Y en X et de X en Y passent par le point moyen  $G(\bar{X}, \bar{Y})$

Les deux coefficients a et  $a'$  sont de même signe et le coefficient de corrélation r vérifie

$$r^2 = aa'$$

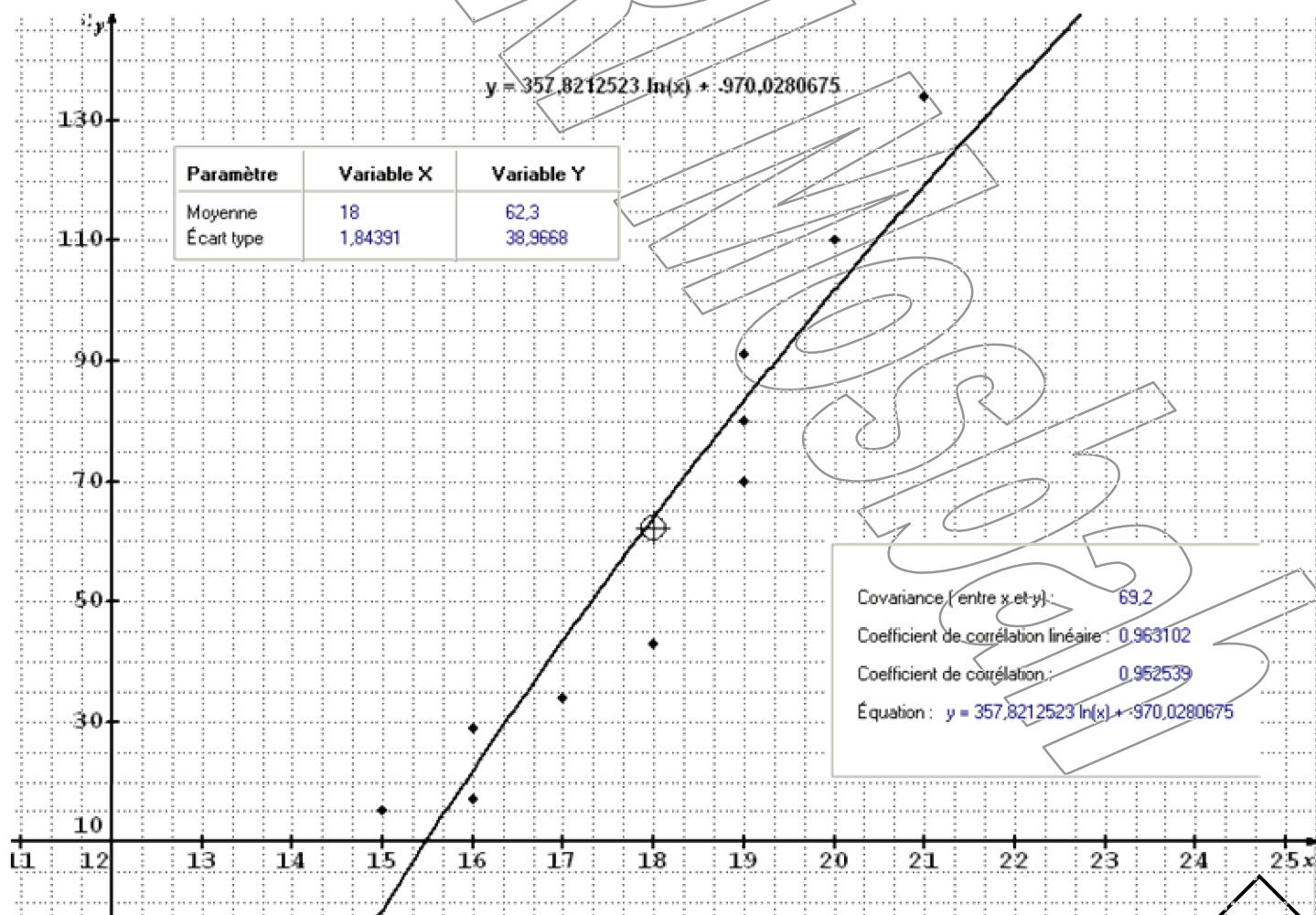
## Exemple d'ajustement non linéaire

## F Ajustement logarithmique : (Réf 296) (EX 13)

Soit le tableau suivant :

$X_i$	18	20	19	16	19	16	19	21	15	17
$Y_i$	43	110	70	17	91	29	80	134	15	34

## Correction





- F Ajustement hyperbolique : (Réf 297) (EX 14)
- F Ajustement exponentiel : (Réf 286) (EX 1)
- F Ajustement parabolique : (Réf 286) (EX 1)