## I- Séries statistiques à une variable :

### 1. Exemples :

◆**Tableau 1** :

Dans un groupe de dix élèves, voici les notes à un devoir : 12, 4, 16, 16, 10, 7, 9, 12, 9, 12.
Cette série de note est une série statistique $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ avec $x_i$ les valeurs prises par X, à chaque valeur $x_i$ on associe un effectif $n_i$ égale au nombre d'élèves ayant obtenu la note $x_i$.

▸ La fréquence de la valeur d'une série est : $f_i = \frac{n_i}{N}$ avec N l'effectif total ( $N = \sum n_i$ ) .

| $x_i$ (Notes) | 4 | 7 | 9 | 10 | 12 | 16 |
|---|---|---|---|---|---|---|
| $n_i$ (Effectif) | 1 | 1 | 2 | 1 | 3 | 2 |
| $f_i$ (Fréquence) | 0,1 | | 0,2 | 0,1 | | 0,2 |
| effectifs cumulés croissants | 1 | | 4 | | 8 | |
| $n_i x_i$ | 4 | | 18 | 10 | | 32 |
| $n_i x_i^2$ | 16 | 49 | | 100 | 432 | |

◆**Tableau 2**

Le tableau suivant donne la durée (en heures) passée chaque jour devant la télévision pour 1000 personnes.

| Durées (en heures) | $[0;2[$ | $[2;4[$ | $[4;6[$ | $[6;8[$ | $[8;10[$ |
|---|---|---|---|---|---|
| $C_i$ (centre de classe) | 1 | | 5 | | 9 |
| $n_i$ (Effectif) | 100 | 200 | 500 | 150 | 50 |
| effectifs cumulés croissants | 100 | | 800 | | 1000 |
| $n_i C_i$ | | 600 | | 1050 | 450 |
| $n_i C_i^2$ | 100 | | 12500 | | 4050 |

◆ **Définitions** :

❖ La première série est une série quantitative discrète.
❖ La deuxième série est une série quantitative continue.

### 2. Les paramètres de position :

❖ **Mode ( ou classe modale )**

On appelle Mode ou classe Modale la valeur de X correspondant à l'effectif le plus haut.

▸ Pour la série quantitative discrète le Mode est :…………………………………………………………………..

▸ Pour la série quantitative continue la classe Modale est :……………………………………………………..

❖ **Moyenne** $(\overline{X})$ **:** La moyenne est : $\overline{X} = \frac{1}{N}\sum_{i=1}^{p} n_i x_i$

▸ Tableau 1 : La moyenne est : $\overline{X} =$ ……………………………………………………………………

▸ Tableau 2 : La moyenne est : $\overline{X} =$ ……………………………………………………………………

❖ **Médiane** ($M_e$):

On appelle Médiane d'une série statistique d'effectif total N la valeur de la variable qui correspond à l'effectif cumulé $\frac{N}{2}$ si N est pair et $\frac{N+1}{2}$ si N est impair.

- Tableau 1 : $\frac{N}{2} = 5$ alors $M_e = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$
- Tableau 2 : $\frac{N}{2} = 500$ ; $300 < 500 < 800$ alors $\bullet \langle M_e \langle \bullet$

  alors $\frac{M_e - 4}{500 - 300} = \frac{6-4}{800-300}$ D'où $M_e = \ldots\ldots\ldots\ldots\ldots\ldots$

❖ **Le premier Quartile ($Q_1$)**

Le premier Quartile $Q_1$ est la plus petite valeur de la série telle qu'au moins 25% des termes de la série ont une valeur inférieure ou égale à $Q_1$.

❖ **Le troisième Quartile ($Q_3$)**

Le troisième Quartile $Q_3$ est la plus petite valeur de la série telle qu'au moins 75% des termes de la série ont une valeur inférieure ou égale à $Q_3$.

◆ **Remarques** :

- Le deuxième quartile $Q_2$ est la médiane $M_e$.
- Les trois quartiles partagent l'ensemble des valeurs en quatre sous ensembles de (presque) même effectif.
- On a toujours : $Q_1 \leq M_e \leq Q_3$
- Si $\frac{N}{4}$ est un entier, le premier quartile $Q_1$ est la valeur qui dans cette liste occupe le rang $\frac{N}{4}$ et le troisième quartile $Q_3$ est la valeur qui dans cette liste occupe le rang $\frac{3N}{4}$ .
- Si $\frac{N}{4}$ n'est pas un entier, le premier quartile $Q_1$ est la valeur qui dans cette liste occupe le rang immédiatement supérieur à $\frac{N}{4}$ et le troisième quartile $Q_3$ est le valeur qui dans cette liste occupe le rang immédiatement supérieur à $\frac{3N}{4}$.
- Pour Une série continue :

  $Q_1$ est la valeur correspondant à la fréquence cumulée croissante égale à 0,25 .

  $Q_3$ est la valeur correspondant à la fréquence cumulée croissante égale à 0,75 .

- Tableau 1 :
  - $\frac{N}{4} = 2,5$ alors $Q_1 = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$
  - $\frac{3}{4}N = 7,5$ alors $Q_3 = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$
- Tableau 2 :
  - $\frac{N}{4} = 250$ ; $100 < 250 < 300$ alors $\bullet < Q_1 < \bullet$

  alors $\frac{Q_1 - 2}{250 - 100} = \frac{4-2}{300-100}$ D'où $Q_1 = \ldots\ldots\ldots\ldots\ldots\ldots$

- $\frac{3}{4}N = 750$ ; $300 < 750 < 800$ <u>alors</u> $\bullet < Q_3 < \bullet$

alors $\dfrac{Q_3 - 4}{750 - 300} = \dfrac{6 - 4}{800 - 300}$ D'où $Q_3 = $ ..................

**3. <u>Les paramètres de dispersion</u>** :

❖ **Variance et écart-type** $(V(X)$ et $\sigma(X))$ **:**

- La variance est : $V(X) = \frac{1}{N}\sum_{i=1}^{p} n_i x_i^2 - \overline{X}^2$.
- L'écart-type est : $\sigma(X) = \sqrt{V(X)}$
- L'écart type mesure la dispersion des valeurs de la série autour de la moyenne.

Plus l'écart type est petit, plus les valeurs de la série sont concentrées autour de la moyenne.

▸ Tableau 1 : • $V(X) = $ ..............................................................................................................................

• $\sigma(X) = $ ..............................................................................................................................

▸ Tableau 2 : • $V(X) = $ ..............................................................................................................................

• $\sigma(X) = $ ..............................................................................................................................

❖ **Etendue**

L'étendue est la différence entre la plus grande et la petite des valeurs.

▸ Tableau 1 : L'étendue est : ..................................................................................

▸ Tableau 2 : L'étendue est : ..................................................................................

❖ **Ecart interquartile**

- L'intervalle interquartile d'une série statistique est l'intervalle $[Q_1, Q_3]$
- L'écart interquartile d'une série statistique est le nombre $Q_3 - Q_1$

▸ Tableau 1 : L'écart interquartile est : ..................................................................................

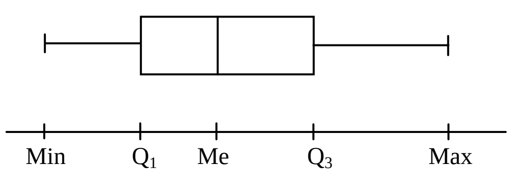
▸ Tableau 2 : L'écart interquartile est : ..................................................................................

❖ **Diagramme en boîte** ( boîtes à moustaches )

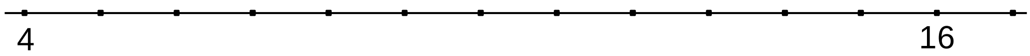
Un diagramme en boite est un rectangle délimité par le premier quartile et le troisième quartile.

Pour l'obtenir, on trace un axe horizontal ( ou vertical ) sur lequel on place les valeurs de $Q_1, Q_3$ et $M_e$

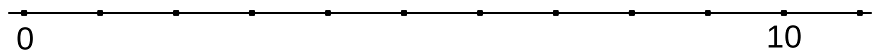
L'un des côtés du rectangle a pour longueur l'écart interquartile, l'autre est quelconque.
On complète ce diagramme en traçant deux traits horizontaux : l'un joignant $Q_1$ au minimum de la série et l'autre joignant $Q_3$ au maximum de la série.



- Tableau 1 :




- Tableau 2 :




## II. <u>Séries Statistiques Doubles</u> :

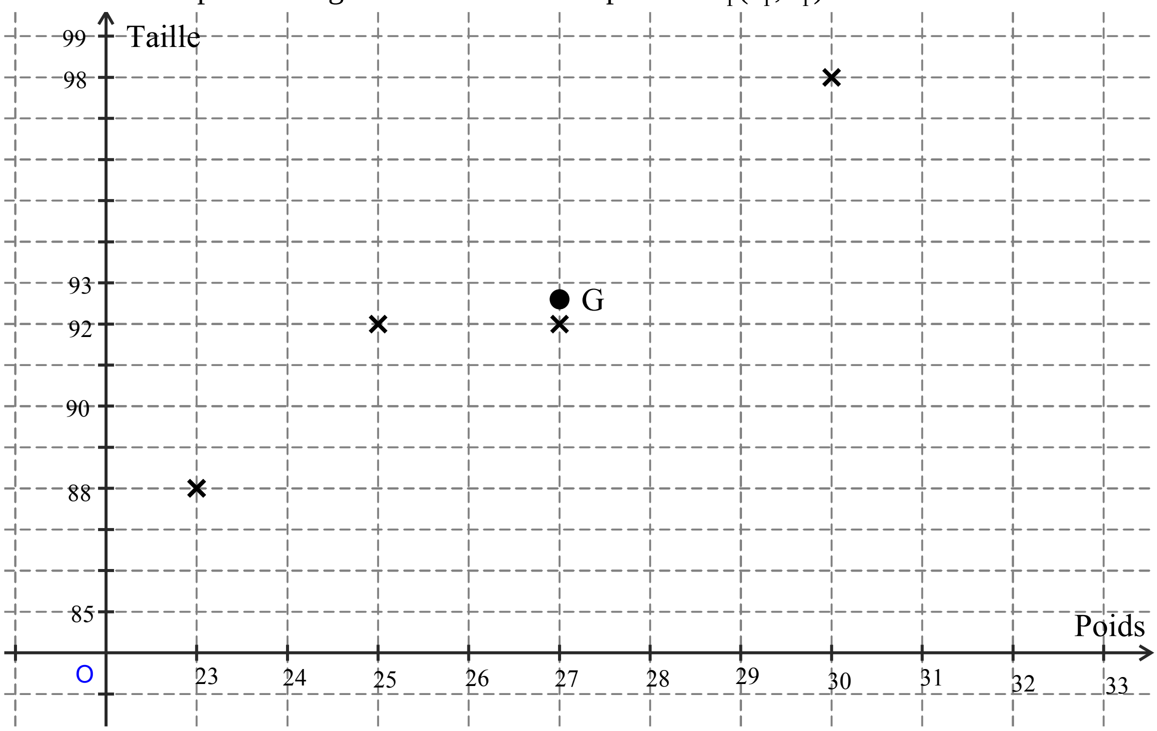
◆ **Exemple** : Le tableau suivant donne le poids en Kg et la taille en cm d'un groupe de 10 enfants :

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $P_i$ | 25 | 27 | 23 | 30 | 27 | 23 | 25 | 30 | 32 | 28 |
| $T_i$ | 90 | 92 | 85 | 99 | 93 | 88 | 92 | 98 | 99 | 90 |

- Le couple $(P_1, T_1) = (25, 90)$ veut dire que l'enfant N°1 pèse 25 Kg et mesure 90 cm.
- On a donc une population de 10 enfants sur laquelle on a observé simultanément les deux variables P et T.

◆ **Définition** : On dit qu'un couple (X,Y) de variables statistiques définies une série double si les deux variables X et Y sont observés simultanément sur une même population.

- La moyenne arithmétique des poids est : $\overline{P} =$ ........................................
- La moyenne arithmétique des Tailles est : $\overline{T} =$ ........................................
- Placer dans un repère orthogonal l'ensemble des points $M_i(P_i, T_i)$ :




◆ **Définition** :

Soit une série statistique définie par deux variables X et Y . On désigne par $x_1, x_2, ......, x_n$ les valeurs de X et par $y_1, y_2, ......, y_n$ celles de Y. Le plan étant rapporté à un repère orthogonal.

- L'ensemble des points $M_i(x_i, y_i)$ ; $i \in \{1, 2, ....., n\}$ est appelé Nuage De Points.
- Le point $G(\overline{x}, \overline{y})$ est appelé point moyen du nuage.

◆ **Distributions marginales** :

Soit le tableau statistique suivant : X : note en mathématiques ; Y : nombre de frères et sœurs. $N = 100$

| Y \ X | 0 | 1 | 2 | 3 | 4 | 5 | 6 | Totaux |
|---|---|---|---|---|---|---|---|---|
| [0,4[ | 1 | 0 | 1 | 1 | 0 | 1 | 1 | |
| [4,8[ | 2 | 2 | 4 | 3 | 3 | 4 | 2 | 20 |
| [8,12[ | 5 | 5 | 10 | 7 | 6 | 4 | 3 | |
| [12,16[ | 2 | 3 | 5 | 5 | 4 | 4 | 2 | 25 |
| [16,20[ | 1 | 1 | 2 | 3 | 2 | 1 | 0 | |
| Totaux | 11 | | 22 | | | | 8 | 100 |

▸ Les totaux inscrits en marge de chaque tableau à double entrée définissent deux distributions marginales, l'une associée à la première variable statistique et l'autre associée à la deuxième variable statistique.

| X : Note en Maths | [0,4[ | [4,8[ | [8,12[ | [12,16[ | [16,20[ | Total |
|---|---|---|---|---|---|---|
| Effectif | | 20 | | | 10 | |

Distribution marginale de X

| Y : nombre de frères et sœurs | 0 | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|---|
| Effectif | | | 22 | | | | 8 | |

Distribution marginale de Y

• Calcul de **la moyenne (** $\overline{X}$ **) ; la variance** $(V(X))$ **et l'écart-type** $(\sigma(X))$

▸ $\overline{X} = \dfrac{\sum_{i=1}^{p} x_i n_i}{N} = \dfrac{(2\times5)+(6\times20)+(10\times40)+(14\times25)+(18\times10)}{100} = \ldots\ldots\ldots$

▸ $V(X) = \dfrac{\sum_{i=1}^{p} x_i^{2}\cdot n_i}{N} - \overline{X}^2 = \dfrac{(2^2\times5)+(6^2\times20)+(10^2\times40)+(14^2\times25)+(18^2\times10)}{100} - \overline{X}^2 = \ldots\ldots$

▸ $\sigma(X) = \sqrt{V(X)} \approx \ldots\ldots\ldots$

▸ $\overline{Y} = \dfrac{\sum_{i=1}^{q} y_i n_i}{N} = \dfrac{(0\times11)+(1\times11)+(2\times22)+(3\times19)+(4\times15)+(5\times14)+(6\times8)}{100} = \ldots\ldots\ldots$

▸ $V(Y) = \dfrac{\sum_{i=1}^{q} y_i^{2}\cdot n_i}{N} - \overline{Y}^2 = \dfrac{(0^2\times11)+(1^2\times11)+(2^2\times22)+\ldots\ldots\ldots+(6^2\times8)}{100} - \overline{Y}^2 = \ldots\ldots$

▸ $\sigma(Y) = \sqrt{V(Y)} \approx \ldots\ldots\ldots$

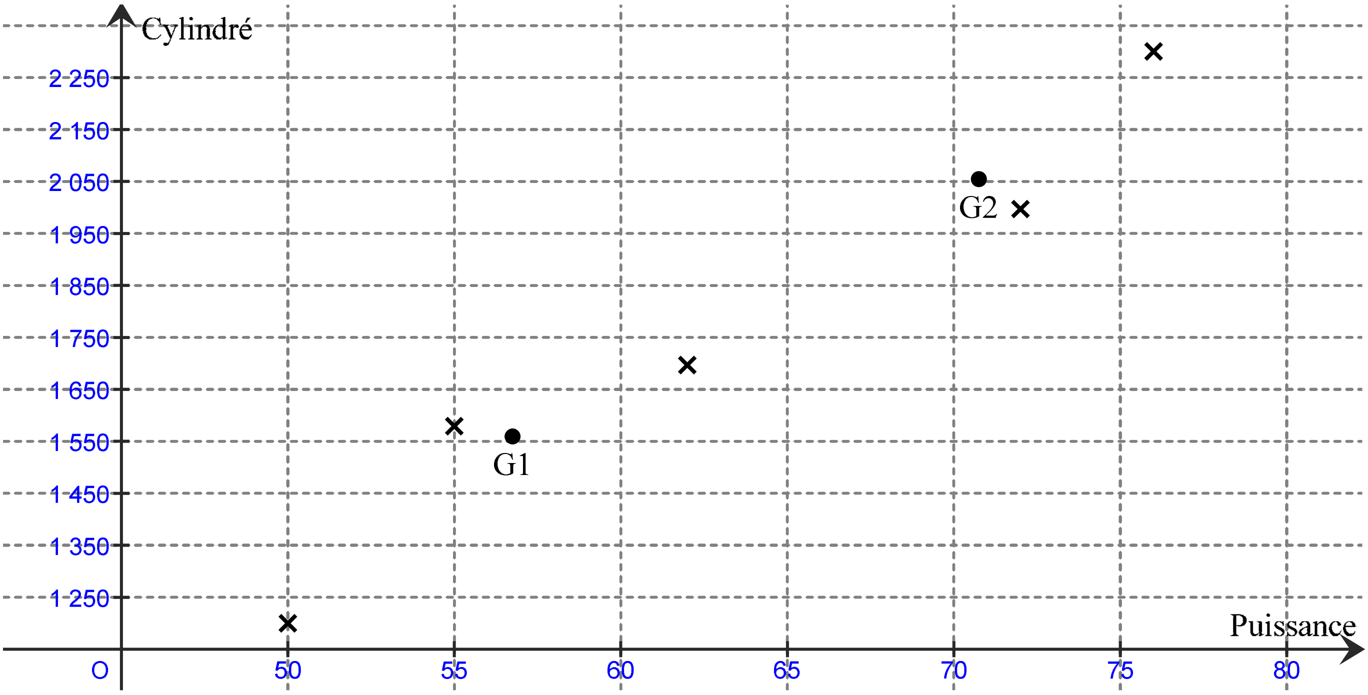
## III. <u>Ajustement affine d'une série statistique double</u> :

Lorsque le nuage des points, représentant graphiquement une série statistique à deux caractères X et Y, a une forme allongée, on peut approcher la relation entre les deux variables X et Y par une relation affine définie par : $Y = aX + b$ ou $X = a'Y + b'$.

◆ **Exemple** :
Le tableau ci-dessous indique la puissance X en chevaux DIN (Deutsche Industrie Normen) et la cylindrée Y en $cm^3$ de 8 voitures à moteur diesel.

| Voiture | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| Puissance X | 50 | 55 | 60 | 62 | 65 | 70 | 72 | 76 |
| Cylindrée Y | 1200 | 1579 | 1761 | 1697 | 1935 | 1986 | 1997 | 2300 |

**1.** Compléter le nuage de points associé à ce tableau.




**2.** Calculer la puissance moyenne et la cylindrée moyenne de ces huit voitures.

La puissance moyenne est $\overline{X} = \frac{1}{8}(50 + 55 + \ldots\ldots\ldots\ldots + 76) = \ldots\ldots\ldots\ldots\ldots\ldots$

La cylindrée moyenne est $\overline{Y} = \frac{1}{8}(1200 + 1579 + \ldots\ldots\ldots\ldots + 2300) = \ldots\ldots\ldots\ldots\ldots\ldots$

**3.** On appelle $G_1$ et $G_2$ les points moyens des sous-nuages constitués d'une part par les voitures A, B, C et D, d'autre part, par les voitures E, F, G et H.

**a.** Calculer les coordonnées de $G_1$ et de $G_2$.

$G_1\left(\frac{50+55+60+62}{4} \; ; \; \frac{1200+1579+1761+1697}{4}\right)$ alors $G_1(\ldots\ldots\ldots\ldots \; ; \; \ldots\ldots\ldots\ldots)$

$G_2\left(\frac{65+70+72+76}{4} \; ; \; \frac{1935+1986+1997+2300}{4}\right)$ alors $G_2(\ldots\ldots\ldots\ldots \; ; \; \ldots\ldots\ldots\ldots)$

**b.** Déterminer l'équation réduite de la droite $(G_1G_2)$. $(Y = aX + b)$. La tracer sur le graphique.

$a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \approx 35,38$

$b = y_{G_1} - a \cdot x_{G_1} = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \approx -448,57$ donc $(G_1G_2) : y = \ldots\ldots\ldots\ldots$

**◆ Utilisation de la calculatrice :** ( Exemple pour une calculatrice Sharp EL-531WH )

**❖ Série à une variable :**

Le service comptable d'une banque donne les données statistiques suivantes pour 100 clients :

| Somme déposée en dinars : $x_i$ | 300 | 500 | 1000 | 2000 | 5000 |
|---|---|---|---|---|---|
| Effectifs : $n_i$ | 22 | 28 | 25 | 20 | 5 |

- Choisir le mode statistique simple : $\boxed{\text{MODE}}$ $\boxed{1}$ $\boxed{0}$
- Entrer les données en tapant
  $\boxed{300}$ $\boxed{\text{STO}}$ $\boxed{22}$ $\boxed{\text{M+}}$
  $\boxed{500}$ $\boxed{\text{STO}}$ $\boxed{28}$ $\boxed{\text{M+}}$
  $\boxed{1000}$ $\boxed{\text{STO}}$ $\boxed{25}$ $\boxed{\text{M+}}$
  $\boxed{2000}$ $\boxed{\text{STO}}$ $\boxed{20}$ $\boxed{\text{M+}}$
  $\boxed{5000}$ $\boxed{\text{STO}}$ $\boxed{5}$ $\boxed{\text{M+}}$
- $\overline{X}$ : $\boxed{\text{RCL}}$ $\boxed{4}$ la calculatrice affiche $\overline{X} =$ ............................
- $\sigma(X)$ : $\boxed{\text{RCL}}$ $\boxed{6}$ la calculatrice affiche $\sigma(X) =$ ............................
- $V(X)$ : $\boxed{\text{RCL}}$ $\boxed{6}$ $\boxed{x^2}$ $\boxed{=}$ la calculatrice affiche $V(X) =$ ............................

**❖ Série à deux variables :**

On a relevé dans le tableau ci-dessous, l'âge et la tension artérielle maximale de 5 femmes

| X : Age | 36 | 42 | 48 | 54 | 60 |
|---|---|---|---|---|---|
| Y : Tension artérielle | 11.8 | 14 | 12.6 | 15 | 15.1 |

- Choisir le mode statistique double : $\boxed{\text{MODE}}$ $\boxed{1}$ $\boxed{1}$
- Entrer les données en tapant
  $\boxed{36}$ $\boxed{\text{STO}}$ $\boxed{11.8}$ $\boxed{\text{M+}}$
  $\boxed{42}$ $\boxed{\text{STO}}$ $\boxed{14}$ $\boxed{\text{M+}}$
  $\boxed{48}$ $\boxed{\text{STO}}$ $\boxed{12.6}$ $\boxed{\text{M+}}$
  $\boxed{54}$ $\boxed{\text{STO}}$ $\boxed{15}$ $\boxed{\text{M+}}$
  $\boxed{60}$ $\boxed{\text{STO}}$ $\boxed{15.1}$ $\boxed{\text{M+}}$
- $\overline{X}$ : $\boxed{\text{RCL}}$ $\boxed{4}$ la calculatrice affiche $\overline{X} =$ ............................
- $\sigma(X)$ : $\boxed{\text{RCL}}$ $\boxed{6}$ la calculatrice affiche $\sigma(X) =$ ............................
- $V(X)$ : $\boxed{\text{RCL}}$ $\boxed{6}$ $\boxed{x^2}$ $\boxed{=}$ la calculatrice affiche $V(X) =$ ............................
- $\overline{Y}$ : $\boxed{\text{RCL}}$ $\boxed{7}$ la calculatrice affiche $\overline{Y} =$ ................
- $\sigma(Y)$ : $\boxed{\text{RCL}}$ $\boxed{9}$ la calculatrice affiche $\sigma(X) =$ ............................
- $V(Y)$ : $\boxed{\text{RCL}}$ $\boxed{9}$ $\boxed{x^2}$ $\boxed{=}$ la calculatrice affiche $V(Y) =$ ............................