Le groupement de communes du pays de Guéret-St Vancry comprend 17 communes dont la population était la suivante en janvier 1999 :

982 ; 519 ; 348 ; 15 105 ; 336 ; 162 ; 532 ; 133 ; 2 250 ; 709 ; 446 ; 1 879 ; 2 059 ; 297 ; 326 ; 198 ; 124 .

1/ Calculer la moyenne et l'écart-type de cette série.

Moyenne : $\overline{X} = \dfrac{x_1 + x_2 + \ldots + x_{17}}{17}$,

$$\overline{X} = \frac{982+519+348+15\,105+336+162+532+133+2\,250+709+446+1\,879+2\,059+297+326+198+124}{17}$$

$\overline{X}$ = 1 553,24 habitants par commune (on arrondira à 1553).

La Moyenne est un *indicateur de Valeur Centrale*.

L'écart-Type $\sigma(X)$ (sigma) est égal à la racine carrée de la Variance $V(X)$.

$$V(X) = \frac{(x_1 - \overline{X})^2 + (x_2 - \overline{X})^2 + (x_3 - \overline{X})^2 + \ldots + (x_{17} - \overline{X})^2}{17},$$

$$V(X) = \frac{(982 - 1553)^2 + (519 - 1553)^2 + (384 - 1553)^2 + \ldots.. + (124 - 1553)^2}{17} = 202\,725\,774 .$$

$\sigma(X) = \sqrt{V(X)}$ = 14 238 habitants, après arrondi.

L'Ecart-Type est un *indicateur de dispersion* autour de la Moyenne.
Plus il est grand, plus les données sont dispersées autour de la moyenne, ce qui signifie qu'on a de grands écarts entre les petites et les grandes communes.

Explication de son mode de calcul :

On mesure les écarts entre les populations de chaque commune x_i et la Moyenne d'entre elles, $\overline{X}$, soit la moyenne des écarts $x_i - \overline{X}$.

Certains écart sont *négatifs* (communes plus petites que la moyenne) et d'autre *positifs* (communes plus grande que la moyenne), si bien qu'en les additionnant ces écarts *algébriques* se compensent, jusqu'à donner une somme nulle.

Il faut rendre équivalents les écarts négatifs et positifs, en les rendant tous positifs par leur mise au carré.

On obtient la Variance $V(X) = \dfrac{(x_1 - \overline{X})^2 + (x_2 - \overline{X})^2 + (x_3 - \overline{X})^2 + \ldots + (x_{17} - \overline{X})^2}{17}$.

La Variance, par ses mises au carré, donne une importance excessive aux grands écarts, qui mis au carré deviennent encore plus grands, par rapport aux petits écarts.
Il faut compenser cet excès en reprenant la racine carrée du résultat, soit : $\sigma(X) = \sqrt{V(X)}$.

2/ Déterminer la médiane et l'intervalle interquartile de la série.

On classe les valeurs x_i dans l'ordre croissant :

124 ; 133 ; 162 ; 198 ; 297 ; 326 ; 336 ; 348 ; 446 ; 519 ; 532 ; 709 ; 982 ; 1 879 ; 2 059 ; 2 250 ; 15 105 .

Il y a 17 valeurs, d'où : $\frac{17}{2} = 17 \times \frac{50}{100} = 8{,}5$. Me est la 9ème valeur, puisque 17 est impair.

La Médiane est la 9ème valeur, 446 habitants. (50% des communes ont moins de 446 habitants, et 50% en ont plus).

Le 1er Quartile Q_1 est en position $17 \times \frac{25}{100} = 4{,}25$. Q_1 est la 5ème valeur, puisque 17 est impair.

$Q_1 = 297$ habitants (25% des communes ont moins de 297 habitants).

Le 3ème Quartile Q_3 est en position $17 \times \frac{75}{100} = 12{,}75$. Q_3 est la 13ème valeur, puisque 17 est impair.

$Q_3 = 982$ habitants (75% des communes ont moins de 982 habitants).

Intervalle interquartile : $Q_3 - Q_1 = 982 - 297 = 685$ hab. (50% entre 297 hab. et 982 hab.).

3/ Expliquer pourquoi on observe une différence importante entre moyenne et médiane.

$\overline{X}$ = 1 553 hab. et Me = 446 hab. . La différence est considérable.

La raison est que la *Médiane ne tient pas compte des valeurs* réelles des populations des communes, mais *seulement de leur classement* .
On cherche la commune placée au milieu du classement, soit 446 habitants.
Par contre, la *Moyenne tient compte des valeurs de chaque population*, donc est sensiblement relevée par les grosses communes, en particulier celle avec 15 105 habitants, ce qui explique $\overline{X}$ = 1 553 hab.