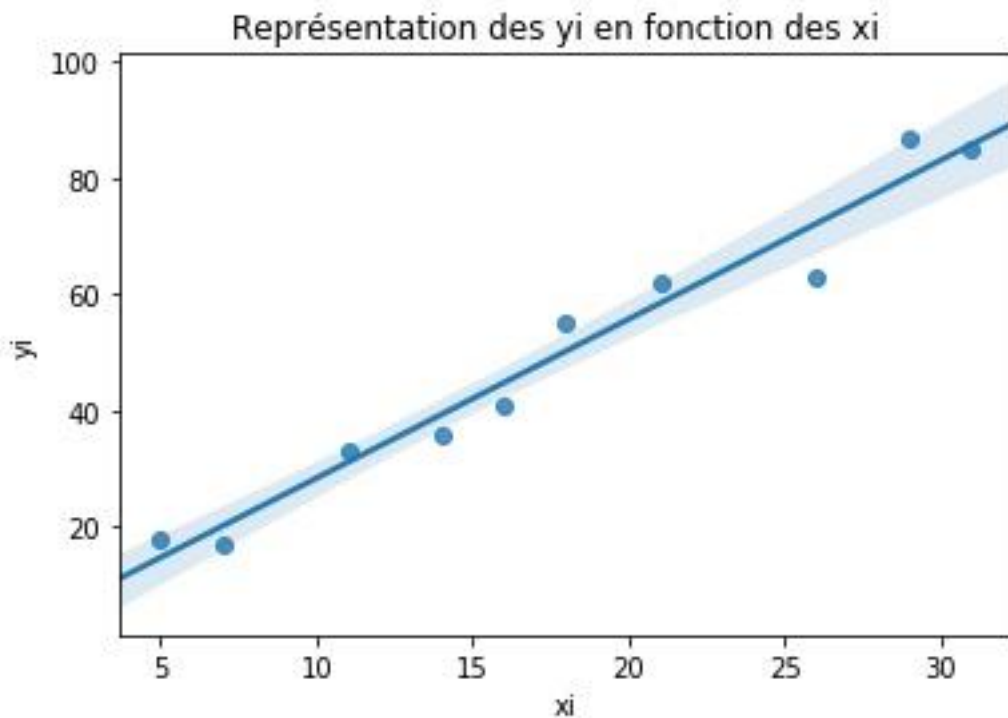


## QUESTION 2: Régression linéaire simple

1. Les données ont été enregistrées dans le logiciel Excel sous le format csv et ensuite importé dans python à l'aide la fonction `panda.read_csv ('source du fichier')`

2. Représentons  $y_i$  en fonction des  $x_i$ .

L'équation de la droite de régression est :  $y = \beta_0 + \beta_1 x$  avec  $\beta_0$  l'ordonnée à l'origine et  $\beta_1$  le coefficient directeur de la droite.



L'allure de la représentation ne nous laisse pas soupçonner une liaison linéaire entre les deux variables car la droite de régression ne passe pas par la plupart des points.

3. Déterminons les coefficients de la droite des moindres carrés pour ces observations.

L'estimateur des moindres carrés ordinaires est la solution du programme de minimisation de la somme des carrés des écarts entre les valeurs prédites et les valeurs observées par rapport aux deux paramètres  $\beta_0$  et  $\beta_1$ . Les valeurs estimées des paramètres sont :

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x} \quad \text{et} \quad \widehat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Les calculs avec python donnent :  $\widehat{\beta}_0 = 1.0213414634146432$  et

$$\widehat{\beta}_1 = 2.7347560975609753$$

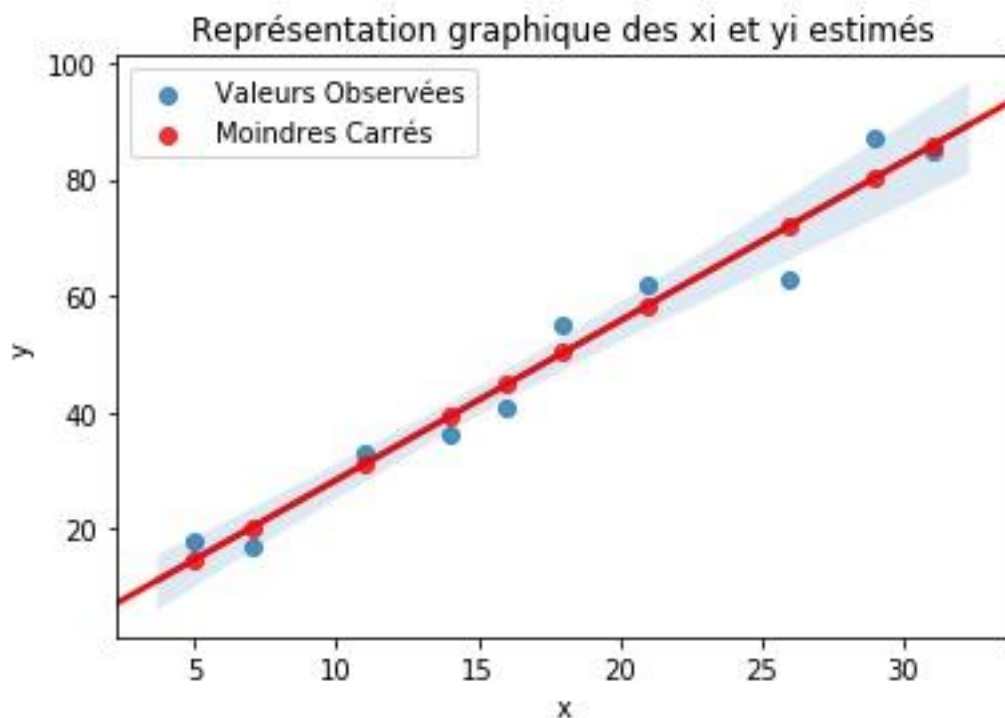
La droite des moindres carrés ordinaires est donc :

$$y = 1.0213414634146432 + 2.7347560975609753 x$$

4. Donnons les ordonnées des  $y_i$  calculés par la droite des moindres carrés correspondant aux différentes valeurs des  $x_i$ . Les valeurs prédites ont été arrondies à trois chiffres après la virgule

$x_i$	18	7	14	31	21	5	11	16	26	29
$y_i$	50.247	20.164	39.308	85.799	58.451	14.695	31.104	44.777	72.125	80.329

5. Traçons la droite des moindres carrés sur le même graphique



6. Quelle est une estimation plausible de Y à  $x_i = 21$

Pour  $x_i = 21$ ,  $Y = 1.0213414634146432 + 2.7347560975609753 * 21$  d'où  $Y = 58.4513$

7. L'écart entre la valeur observée de  $Y$  à  $x_i = 21$  et la valeur estimée avec la droite des moindres carrés est :  $\varepsilon = 62 - 58.4513 = 3.549$ . Cet écart est appelé résidu ou erreur.
8. Oui La droite des moindres carrés obtenue en 2 passe par le point  $(\bar{x}, \bar{y})$ . On ne peut pas généraliser cette conclusion à n'importe quelle droite de régression car cela concerne uniquement les droites de régressions dont les paramètres ont une liaison linéaires entre eux.