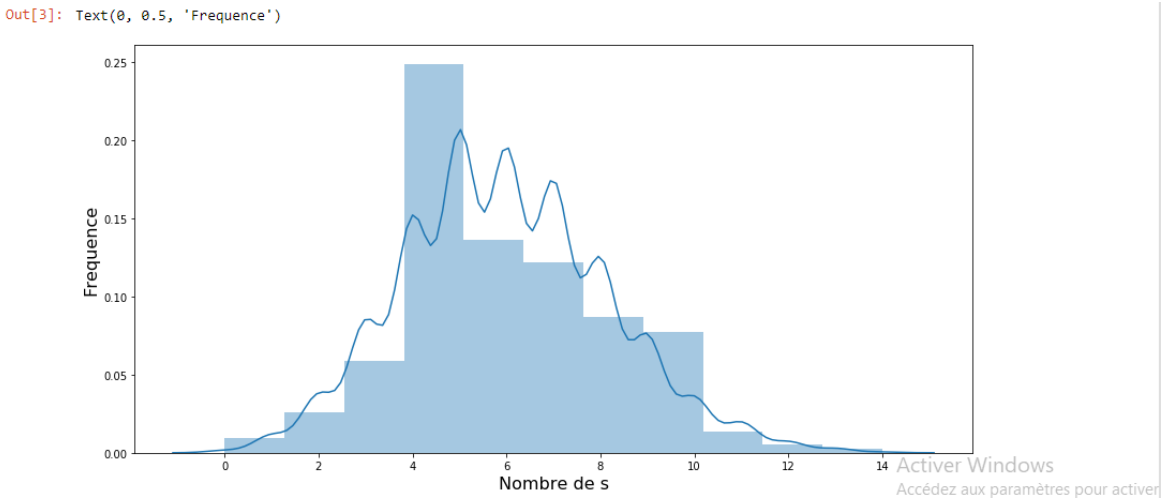
RENDU PROJET SATISTIQUES

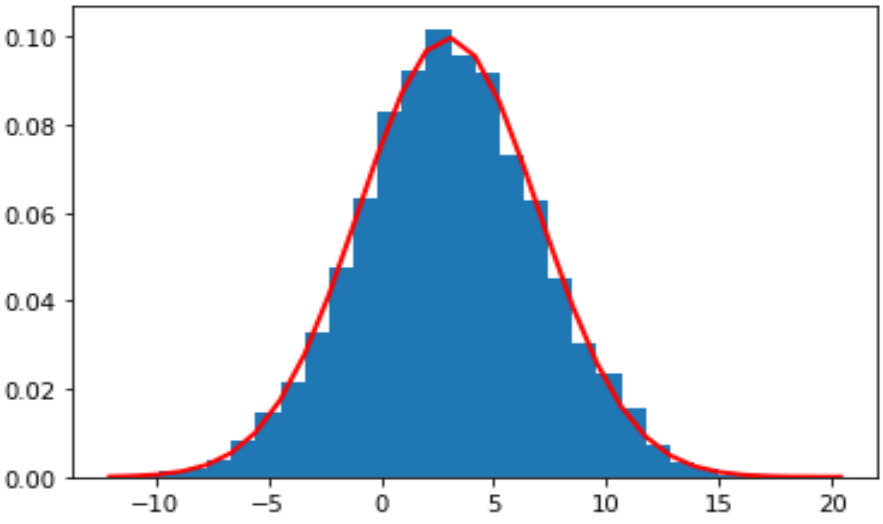
Réponses aux questions

Question 1 :

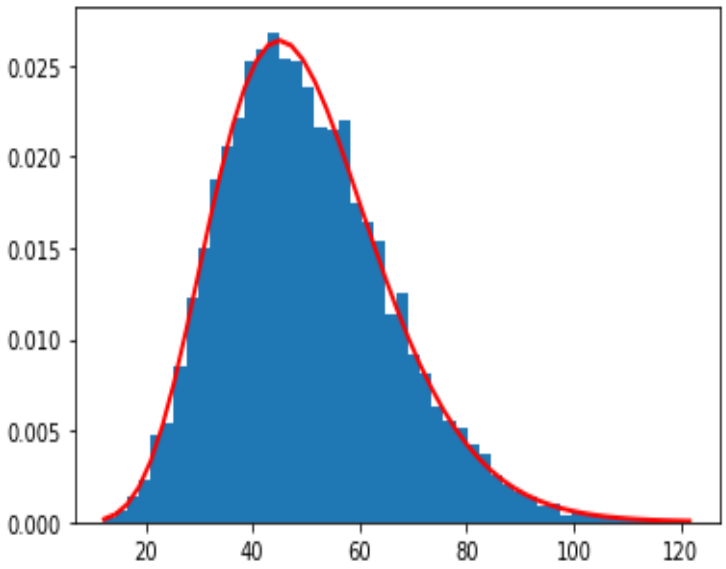
1-



2-



3-

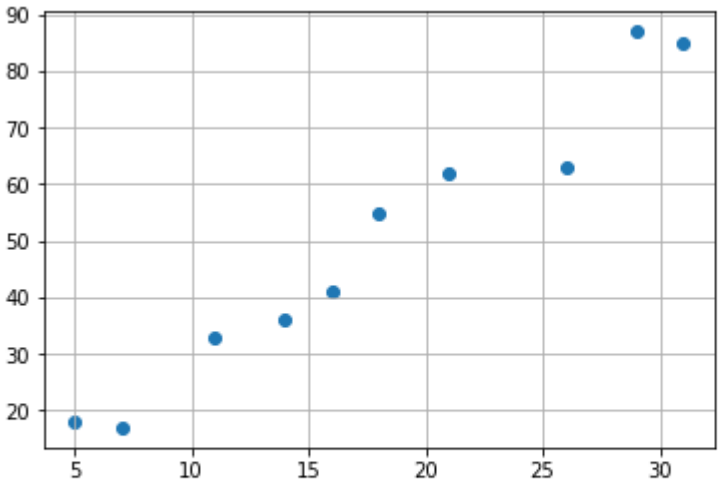


Question 2 :

1-

```
Entrée [2]: X = np.array([18, 7, 14, 31, 21, 5, 11, 16, 26, 29])
            Y = np.array([55, 17, 36, 85, 62, 18, 33, 41, 63, 87])

Entrée [3]: X

    Out[3]: array([18,  7, 14, 31, 21,  5, 11, 16, 26, 29])

Entrée [4]: Y

    Out[4]: array([55, 17, 36, 85, 62, 18, 33, 41, 63, 87])
```

2-



""" Vu l'allure de cette courbe on peut dire qu'il existe une relation entre X et Y """

3-

Entrée [7]:
```
# LES coefficients de la doite sont :
print("La pente est :" + str(slope))
print("la constante est : " +str(intercept))
```

```
La pente est :2.7347560975609757
la constante est : 1.021341463414636
```

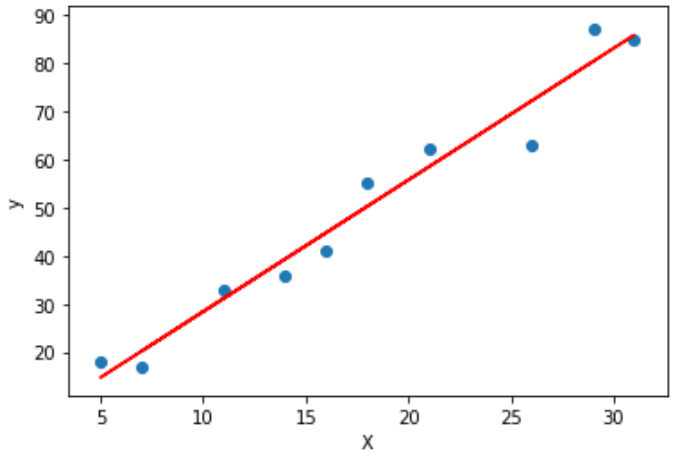
4-

Entrée [9]:
```
for i in range(len(X)):
    print(predict(X[i]))
```

```
50.2469512195122
20.164634146341466
39.3079268292683
85.79878048780489
58.451219512195124
14.695121951219514
31.103658536585368
44.77743902439025
72.125
80.32926829268294
```

5-

```
plt.scatter(X,Y)
plt.xlabel("X")
plt.ylabel("y")
plt.show()
```



6-

```
Entrée [11]: print(predict(21))
```

```
58.451219512195124
```

7-

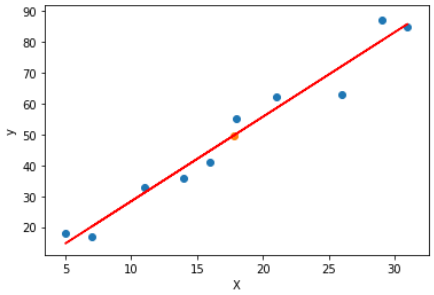
```
Entrée [12]: ecart = Y[4] - predict(21)
             ecart
```

```
Out[12]: 3.5487804878048763
```

""" Cet écart est appelé résidu """

8-

Entrée [13]:
```
fitLine = predict(X)
plt.plot(X, fitLine, c='r')
plt.scatter(X,Y)
plt.scatter(np.mean(X),np.mean(Y))
plt.xlabel("X")
plt.ylabel("y")
plt.show()
```



""" Oui on voit bien que la droite passe bien par le point (xbar,ybar) en rouge et cette observation peut être généraliser pour n'importe quelle droite de regression """

Question 3 :

1-

```
 0   age          797 non-null    float64
 1   prof         793 non-null    object
 2   duree        576 non-null    float64
 3   discip       793 non-null    float64
 4   n.enfant     773 non-null    float64
 5   n.fratrie    799 non-null    int64
 6   ecole        794 non-null    float64
 7   separation   788 non-null    float64
 8   juge.enfant  794 non-null    float64
 9   place        792 non-null    float64
 10  abus         792 non-null    float64
 11  grav.cons    795 non-null    float64
 12  dep.cons     799 non-null    int64
 13  ago.cons     799 non-null    int64
 14  ptsd.cons    799 non-null    int64
 15  alc.cons     799 non-null    int64
 16  subst.cons   799 non-null    int64
 17  scz.cons     799 non-null    int64
 18  char         703 non-null    float64
 19  rs           696 non-null    float64
 20  ed           692 non-null    float64
 21  dr           688 non-null    float64
 22  suicide.s    758 non-null    float64
 23  suicide.hr   760 non-null    float64
 24  suicide.past 785 non-null    float64
 25  dur.interv   749 non-null    float64
dtypes: float64(18), int64(7), object(1)
```

2-

```
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   age            797 non-null    float64
 1   prof           793 non-null    category
 2   duree          576 non-null    category
 3   discip         793 non-null    category
 4   n.enfant       773 non-null    float64
 5   n.fratrie      799 non-null    int64
 6   ecole          794 non-null    category
 7   separation     788 non-null    category
 8   juge.enfant    794 non-null    category
 9   place          792 non-null    category
 10  abus           792 non-null    category
 11  grav.cons      795 non-null    category
 12  dep.cons       799 non-null    category
 13  ago.cons       799 non-null    category
 14  ptsd.cons      799 non-null    category
 15  alc.cons       799 non-null    category
 16  subst.cons     799 non-null    category
 17  scz.cons       799 non-null    category
 18  char           703 non-null    category
 19  rs             696 non-null    category
 20  ed             692 non-null    category
 21  dr             688 non-null    category
 22  suicide.s      758 non-null    category
 23  suicide.hr     760 non-null    category
 24  suicide.past   785 non-null    category
 25  dur.interv     749 non-null    float64
dtypes: category(22), float64(3), int64(1)
```

3-

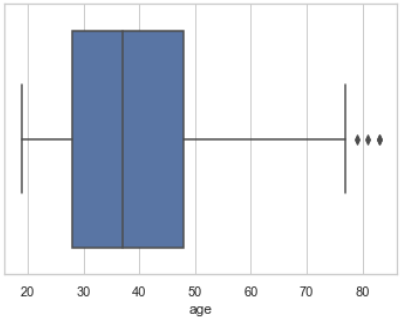
```
Entrée [6]: data.describe()
```

Out[6]:

	age	n.enfant	n.fratrie	dur.interv
count	797.000000	773.000000	799.000000	749.000000
mean	38.899624	1.755498	4.286608	61.891856
std	13.280978	1.834044	3.441485	19.669605
min	19.000000	0.000000	0.000000	0.000000
25%	28.000000	0.000000	2.000000	48.000000
50%	37.000000	1.000000	3.000000	60.000000
75%	48.000000	3.000000	6.000000	75.000000
max	83.000000	13.000000	21.000000	120.000000

4-

```
Entrée [7]: sns.set(style="whitegrid")
ax = sns.boxplot(x=data["age"])
```



""""" On constate que la moyenne d'age est 39 ans, et que 75 % des détenus ont 48 ans. On note aussi l'apparition de valeurs aberrantes ce qui veut dire que la variable age cotient des valeurs nulles(NaN) """

5-

```
Entrée [8]: data[(data.prof== "agriculteur") & (data["n.enfant"] > 2)]
```

Out[8]:

	age	prof	duree	discip	n.enfant	n.fratrie	ecole	separation	juge.enfant	place	...	subst.cons	scz.cons	char	rs	ed	dr	suicide.s	suicide
14	64.0	agriculteur	NaN	0.0	3.0	2	1.0	0.0	0.0	0.0	...	0	0	1.0	1.0	1.0	3.0	0.0	(
311	42.0	agriculteur	4.0	0.0	3.0	6	1.0	0.0	0.0	0.0	...	0	0	2.0	1.0	3.0	2.0	3.0	'
390	36.0	agriculteur	4.0	1.0	3.0	4	3.0	1.0	1.0	1.0	...	1	0	1.0	NaN	3.0	1.0	0.0	(
441	79.0	agriculteur	5.0	0.0	5.0	6	2.0	0.0	0.0	0.0	...	0	0	1.0	2.0	1.0	1.0	0.0	(

4 rows × 26 columns

6-

```
Entrée [9]: data["prof"].describe()

Out[9]: count          793
        unique           8
        top       ouvrier
        freq           227
        Name: prof, dtype: object
```

8-

```
Out[12]: prof
         agriculteur           48.833333
         artisan               45.111111
         autre                 34.935484
         cadre                 50.083333
         employe               38.711111
         ouvrier               37.396476
         prof.intermediaire    43.258621
         sans emploi           35.896396
         Name: age, dtype: float64
```

9-

```
Entrée [13]: data[data.prof.isnull()]
```

Out[13]:

	age	prof	duree	discip	n.enfant	n.fratrie	ecole	separation	juge.enfant	place	...	subst.cons	scz.cons	char	rs	ed	dr	suicide.s	suicide.hr
1	49.0	NaN	NaN	0.0	7.0	3	2.0	1.0	0.0	0.0	...	0	0	1.0	2.0	2.0	1.0	0.0	0.0
11	NaN	NaN	NaN	NaN	NaN	1	NaN	NaN	NaN	NaN	...	1	0	NaN	NaN	NaN	NaN	0.0	0.0
18	NaN	NaN	NaN	NaN	NaN	1	NaN	NaN	NaN	NaN	...	0	0	NaN	NaN	NaN	NaN	0.0	0.0
336	47.0	NaN	5.0	0.0	5.0	5	3.0	0.0	0.0	0.0	...	0	0	1.0	1.0	3.0	2.0	0.0	0.0
342	28.0	NaN	4.0	1.0	3.0	3	2.0	0.0	0.0	0.0	...	1	0	1.0	2.0	2.0	3.0	0.0	0.0
724	48.0	NaN	NaN	NaN	NaN	1	NaN	NaN	NaN	NaN	...	0	0	NaN	NaN	NaN	NaN	1.0	0.0

6 rows × 26 columns

10-

```
Entrée [14]: data.isna().sum()
```

```
Out[14]: age              2
         prof             6
         duree          223
         discip           6
         n.enfant        26
         n.fratrie        0
         ecole            5
         separation      11
         juge.enfant      5
         place            7
         abus             7
         grav.cons        4
         dep.cons         0
         ago.cons         0
         ptsd.cons        0
         alc.cons         0
         subst.cons       0
         scz.cons         0
         char            96
         rs             103
         ed             107
         dr             111
         suicide.s       41
         suicide.hr      39
```

11-

```
Entrée [15]: data
```

Out[15]:

	age	prof	duree	discip	n.enfant	n.fratrie	ecole	separation	juge.enfant	place	...	subst.cons	scz.cons	char	rs	ed	dr	suicide.s	su
0	31.0	autre	4.0	0.0	2.0	4	1.0	0.0	0.0	0.0	...	0	0	1.0	2.0	1.0	1.0	0.0	
1	49.0	NaN	NaN	0.0	7.0	3	2.0	1.0	0.0	0.0	...	0	0	1.0	2.0	2.0	1.0	0.0	
2	50.0	prof.intermediaire	5.0	0.0	2.0	2	2.0	0.0	0.0	0.0	...	0	0	1.0	2.0	3.0	2.0	0.0	
3	47.0	ouvrier	NaN	0.0	0.0	6	1.0	1.0	0.0	1.0	...	0	0	1.0	2.0	2.0	2.0	1.0	
4	23.0	sans emploi	4.0	1.0	1.0	6	1.0	1.0	NaN	1.0	...	0	0	1.0	2.0	2.0	2.0	0.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
794	28.0	sans emploi	5.0	0.0	1.0	4	1.0	NaN	1.0	0.0	...	0	0	NaN	3.0	1.0	3.0	1.0	
795	44.0	ouvrier	4.0	1.0	1.0	12	2.0	0.0	0.0	0.0	...	0	0	1.0	1.0	1.0	1.0	0.0	
796	31.0	cadre	4.0	0.0	3.0	6	4.0	1.0	1.0	1.0	...	1	0	2.0	1.0	1.0	1.0	1.0	
797	38.0	employe	5.0	0.0	0.0	1	3.0	0.0	0.0	0.0	...	0	0	1.0	1.0	1.0	1.0	0.0	
798	71.0	autre	4.0	0.0	2.0	4	1.0	1.0	0.0	0.0	...	0	0	1.0	1.0	1.0	1.0	0.0	

799 rows × 26 columns

```
Entrée [16]: data= data.dropna()

Entrée [17]: data
```

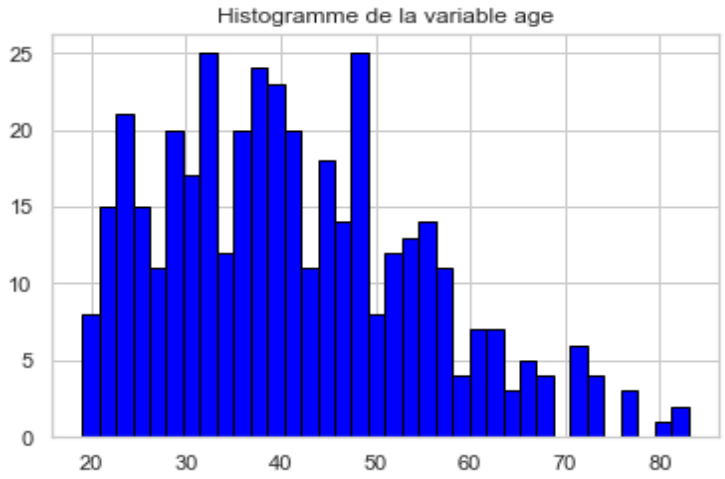
Out[17]:

	age	prof	duree	discip	n.enfant	n.fratrie	ecole	separation	juge.enfant	place	...	subst.cons	scz.cons	char	rs	ed	dr	suicide.s	su
7	52.0	artisan	5.0	0.0	2.0	9	2.0	0.0	0.0	0.0	...	0	0	1.0	2.0	2.0	2.0	0.0	
8	42.0	ouvrier	4.0	1.0	1.0	12	1.0	1.0	1.0	0.0	...	1	0	4.0	3.0	3.0	1.0	4.0	
12	21.0	employe	4.0	0.0	0.0	3	2.0	1.0	1.0	1.0	...	1	0	1.0	2.0	2.0	2.0	0.0	
13	40.0	artisan	4.0	0.0	3.0	5	1.0	0.0	1.0	0.0	...	0	0	1.0	2.0	2.0	2.0	0.0	
16	60.0	prof.intermediaire	5.0	0.0	2.0	4	2.0	1.0	0.0	1.0	...	0	0	1.0	1.0	2.0	3.0	0.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
793	27.0	prof.intermediaire	4.0	0.0	2.0	0	2.0	1.0	0.0	1.0	...	0	0	1.0	3.0	1.0	1.0	0.0	
795	44.0	ouvrier	4.0	1.0	1.0	12	2.0	0.0	0.0	0.0	...	0	0	1.0	1.0	1.0	1.0	0.0	
796	31.0	cadre	4.0	0.0	3.0	6	4.0	1.0	1.0	1.0	...	1	0	2.0	1.0	1.0	1.0	1.0	
797	38.0	employe	5.0	0.0	0.0	1	3.0	0.0	0.0	0.0	...	0	0	1.0	1.0	1.0	1.0	0.0	
798	71.0	autre	4.0	0.0	2.0	4	1.0	1.0	0.0	0.0	...	0	0	1.0	1.0	1.0	1.0	0.0	

403 rows × 26 columns

12-

```
Entrée [18]: plt.hist(data['age'], color = 'blue', edgecolor = 'black',
                       bins = int(180/5))
             plt.title("Histogramme de la variable age")

Out[18]: Text(0.5, 1.0, 'Histogramme de la variable age')
```



13-

```
Entrée [19]: from scipy import stats
a = [19,28,37,48,83]
b = (0.1, 0.2, 0.3, 0.2, 0.2)
age_classe = stats.rv_discrete(name='age_classe', values=(a, b))
```

14-

Entrée [21]:

```
fig, ax = plt.subplots(1, 1)
ax.plot(a, age_classe.pmf(a), 'ro', ms=12, mec='r')
ax.vlines(a, 0, age_classe.pmf(a), colors='r', lw=4)
plt.show()
```