

Travail Pratique de SDD4142-Statistiques

TEAM 6 : MATHIAM FAYE

BAYE THILOR SENE

FATOU DIARRA

Question :2 Régression linéaire simple (4 pts) Nous donnons les couples d'observations suivants :

xi 18 7 14 31 21 5 11 16 26 29

yi 55 17 36 85 62 18 33 41 63 87

1. La première étape est d'obtenir les données. Enregistrer-les dans un format adapté pour une lecture par la suite avec Python.

xi=[18, 7, 14, 31, 21, 5, 11, 16, 26, 29]

yi=[55, 17, 36, 85, 62, 18, 33, 41, 63, 87]

X = np.array(xi)

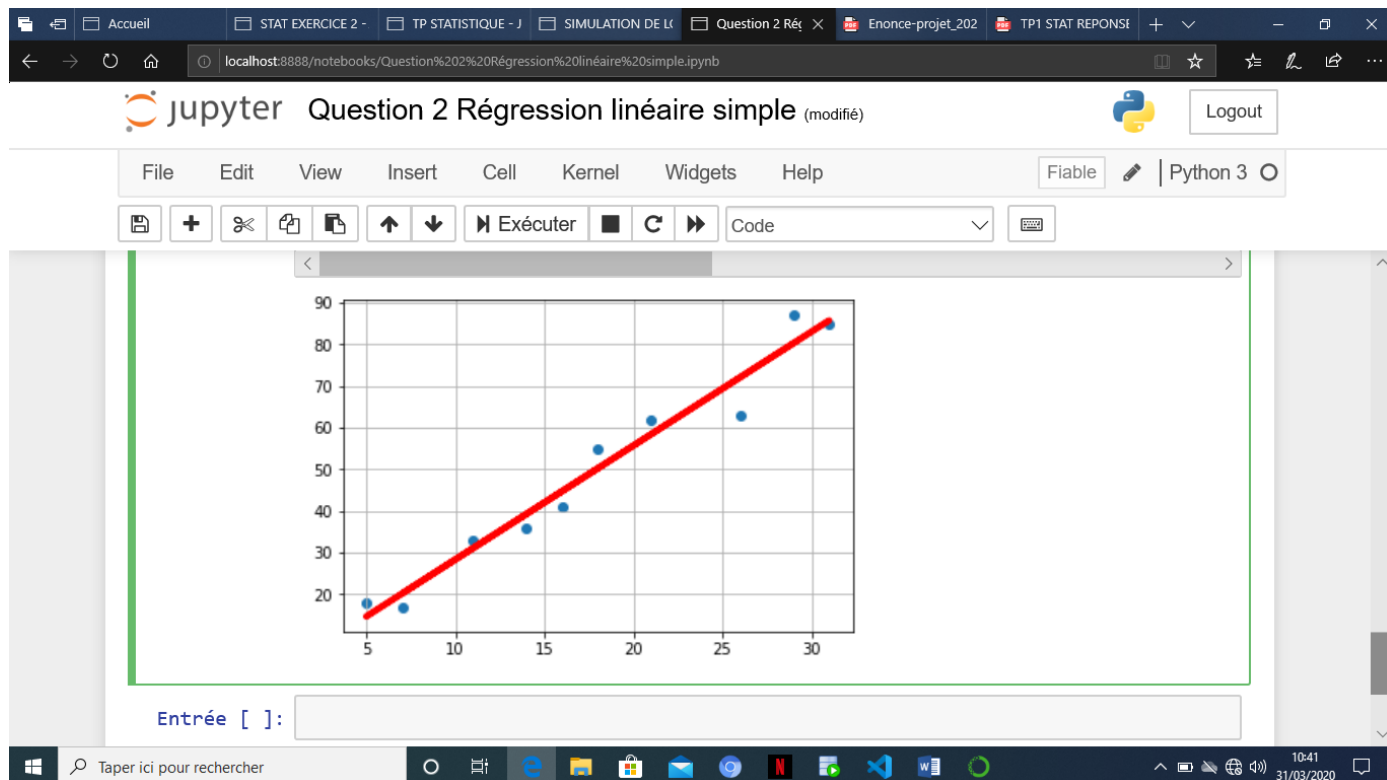
Y = np.array(yi)

2. Représentez les y_i en fonction des x_i . A la vue de cette représentation, pouvons-nous soupçonner une liaison linéaire entre ces deux variables?

A la vue de cette représentation, nous pouvons soupçonner une liaison linéaire entre ces deux variables.

De plus le calcul du coefficient de corrélation entre ces deux variables nous donne `print(r)` #

`r=0.9808024912490334` => une forte corrélation entre X et Y



3. Déterminer pour ces observations la droite des moindres carrés, c'est-à-dire donner les coefficients de la droite des moindres carrés.

#les coefficients a et b de la droite de regression par la methode des moindres carrées es

`a=covariance/v1`

`b=m2-a*m1`

`print(m1)` #17.8

`print(m2)` #49.7

`print(v1)` #72.16

`print(v2)` #561.01

```

print(ecartype1)  #8.49470423263812
print(ecartype2)  #23.68564966387876
print(covariance) #197.33999999999992
print(r)          # 0.9808024912490334 => une forte
correllation entre X et Y
print(a)          #2.734756097560975
print(b)          #1.0213414634146503
#la droite de regression des moindres carrés
est:Y=(2.734756097560975)*X+1.0213414634146503

```

4. Donner les ordonnées des y_i calculés par la droite des moindres carrés correspondant aux différentes valeurs des x_i .

```

yci=[50.2469512195122,20.164634146341474,
      39.3079268292683, 85.79878048780486,
      58.451219512195124, 4.695121951219525,
      31.103658536585375, 44.77743902439025,
      72.125, 80.32926829268291]

```

5.Tracer ensuite la droite sur le même graphique.

Voir figure

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

```

print(Y21) # A x=21 , Y sera estimé à 58.451219512195124

```

7. Quel est l'écart entre la valeur observée de Y à $x_i = 21$ et la valeur estimée avec la droite des moindres carrés?

```

print(ec) #L'ecart est de:3.5487804878048763.

```

Comment appelons-nous cet écart?

cet écart c'est ce qu'on appelle RESIDUS.

8. Est-ce que la droite des moindres carrés obtenue en 2. passe par le point (\bar{x}, \bar{y}) ?

La droite de régression de Y en fonction de X
 $(Y=2.734756097560975)*X + 1.0213414634146503$)

introduit l'hypothèse que les valeurs de Y dépendent de celles de X,

c'est-à-dire postulent que la connaissance des valeurs de X permet d'estimer les valeurs de Y. Il s'agit donc d'un modèle de prévision et l'objectif est de minimiser l'erreur de prévision c'est-à-dire la distance entre les valeurs Y_i observées et les valeurs Y_{ci} estimées par la relation.

Les résidus seront donc la distance à la droite par rapport à l'axe des y.

Pouvons-nous généraliser cette conclusion à n'importe laquelle droite de régression?

Oui la droite des moindres carrés obtenue en 2 passe par le point moyen $G(m_1, m_2)$ avec m_1 =moyenne de x_i et m_2 =moyenne de y_i . Ceci peut être généralisé car le coefficient b de la droite de régression des moindres carrés est obtenu en vérifiant la relation $b=m_2-a*m_1$. Cela permet de déduire la relation $a=(Y-m_2)/(X-m_1)$ toute droite de régression vérifie cette relation donc cette conclusion peut être généralisée à n'importe laquelle droite de régression.