

Prévision du temps de travail nécessaire à la fabrication d'un lot de pièces de rechange

Avant-propos

Cette étude est réalisé à partir du jeu de données mis à disposition par Christophe Chesneau, Professeur de statistiques à l'université de Caen, à l'adresse suivante: <https://chesneau.users.lmno.cnrs.fr/toluca.txt> (<https://chesneau.users.lmno.cnrs.fr/toluca.txt>)

Ce fichier contient une variable X1 représentant la taille du lot et une variable Y représentant le nombre total d'heures de travail nécessaire à sa fabrication. On a connaissance du nombre total d'heures de travail nécessaire pour un échantillon de 25 lots de pièces hétérogènes.

L'entreprise et les données sont fictives.

Merci de m'ajouter sur Linkedin pour toutes questions ou remarques: <https://fr.linkedin.com/in/romain-vimont-a296131b8> (<https://fr.linkedin.com/in/romain-vimont-a296131b8>)

Mise en situation

Toluca, une entreprise spécialisée dans la fabrication de pièce de rechange pour l'équipement de réfrigération, souhaite connaître ses besoins en main-d'œuvre pour l'année à venir. Pour cela, elle dispose des données de l'année passée rapportant le nombre d'heures total nécessaire à la fabrication d'un lot de pièces pour 25 lots de tailles variables. Pour être en bénéfice, le directeur financier estime qu'il faudra produire et vendre au minimum 100 lots de 33 pièces de rechange dans l'année. Il est demandé au Data Analyst de déterminer le nombre de contrats 35h/semaines à faire signer afin de répondre à la demande.

Analyse des données

On commence par chargé les données sur Python:

Entrée [74]:

```
import pandas as pd
dataset = pd.read_csv("C:/Users/vimon/Downloads/toluca.csv", header = 0)
```

On affiche les 5 premières lignes du jeu de données:

Entrée [83]:

```
dataset.head()
```

Out[83]:

| | X1 | Y |
|---|----|-----|
| 0 | 80 | 399 |
| 1 | 30 | 121 |
| 2 | 50 | 221 |
| 3 | 90 | 376 |
| 4 | 70 | 361 |

Par exemple, on a un lot de 80 pièces demandant 399 heures de travail au total pour être fabriquées, puis 30 pièces demandant 121h, etc...

On souhaite expliquer Y a partir de X1.

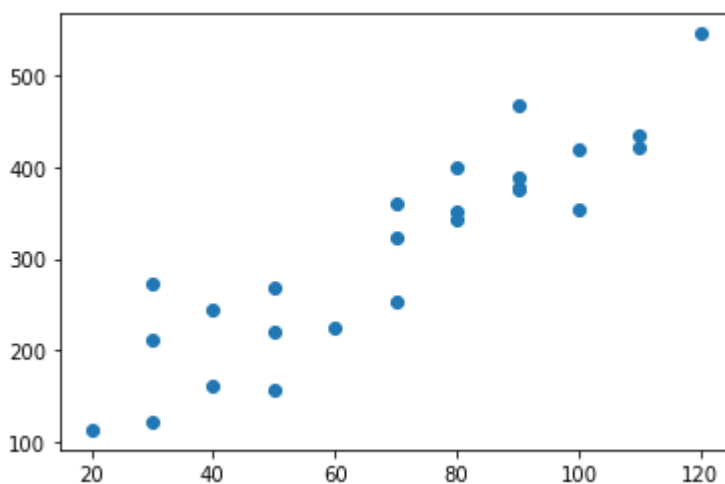
On peut tracer le nuage de points pour voir si un lien linéaire entre X1 et Y est envisageable:

Entrée [84]:

```
import matplotlib.pyplot as plt
plt.scatter(dataset["X1"],dataset["Y"])
```

Out[84]:

<matplotlib.collections.PathCollection at 0x20f1e4951f0>



Vu la forme allongée du nuage, un lien linéaire entre X1 et Y est envisageable.

Régression linéaire simple

On propose de mettre en place un modèle de régression linéaire simple de type: $Y = \beta_0 + \beta_1 X + e$ où β_0 et β_1 sont des coefficients inconnus qu'il faudra estimer avec les données, et e est une variable d'erreur qui modélise

une somme de petites erreurs aléatoires provenant de l'expérimentation d'où provient les données.

Par la méthode des moindres carrés ordinaires, on peut estimer β_0 et β_1 à l'aide des données, donnant ainsi les estimations b_0 et b_1 , respectivement. Dès lors, pour une valeur x de X donnée, une estimation de la valeur moyenne de Y , notée y , est donnée par l'équation : $y = b_0 + b_1x$

Aussi, avec x variable, la droite d'équation $y = b_0 + b_1x$ est appelée droite de régression ; c'est cette droite qui fournit un bon ajustement du nuage de points associés.

On met en œuvre le modèle de régression linéaire simple en faisant :

Entrée [86]:

```
import scipy.stats
import numpy as np
import pylab
import seaborn as sns
import statistics
import statsmodels.api as smi
X = smi.add_constant(dataset["X1"])
model = smi.OLS(dataset["Y"], X)
results = model.fit()
print(results.summary())
```

OLS Regression Results

```
=====
==
Dep. Variable:          Y    R-squared:          0.8
22
Model:                OLS    Adj. R-squared:      0.8
14
Method:             Least Squares    F-statistic:      10
5.9
Date:                Sat, 24 Sep 2022    Prob (F-statistic):  4.45e-
10
Time:                22:37:08    Log-Likelihood:    -131.
64
No. Observations:      25    AIC:              26
7.3
Df Residuals:          23    BIC:              26
9.7
Df Model:              1
Covariance Type:      nonrobust
=====
==
               coef    std err          t      P>|t|      [0.025    0.97
5]
-----
--
const          62.3659    26.177      2.382    0.026     8.214    116.5
18
X1              3.5702     0.347    10.290    0.000     2.852     4.2
88
=====
==
Omnibus:           0.608    Durbin-Watson:      1.4
32
Prob(Omnibus):     0.738    Jarque-Bera (JB):    0.6
84
Skew:              0.298    Prob(JB):            0.7
10
Kurtosis:          2.450    Cond. No.            20
2.
=====
==
```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

On a $R\text{-squared} = 0.822$, qui est proche de 1, ce qui signifie que le modèle est très bon en terme prédictif. La p-valeur associée à β_1 est $P > |t| 0.000$, ce qui signifie que p-valeur < 0.001 , et que X_1 à une influence hautement significative sur Y . On a également les valeurs de b_0 et b_1 , avec $b_0 = 62.3659$ et $b_1 = 3.5702$, d'où l'équation de la droite de régression : $y = 62.3659 + 3.5702x$

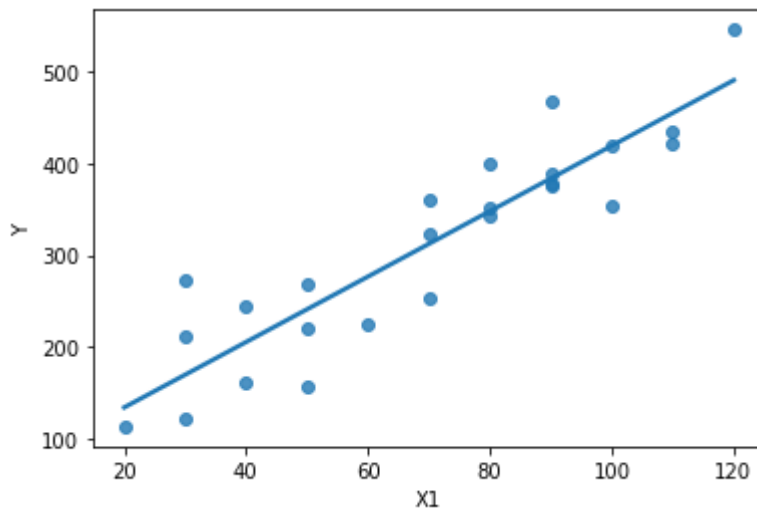
On peut visualiser cette droite en faisant :

Entrée [88]:

```
sns.regplot(x = dataset["X1"], y = dataset["Y"], ci = None)
```

Out[88]:

```
<AxesSubplot:xlabel='X1', ylabel='Y'>
```



Résolution du problème à l'aide du modèle

On obtient le nombre d'heures total nécessaire à la fabrication d'un lot de 33 pièces en faisant:

Entrée [91]:

```
results.predict((1,33))
```

Out[91]:

```
array([180.18252525])
```

Soit environ 180 heures.

Pour l'année, les besoins seront donc de $100 \times 180.18252525 = 18\,018$ heures de travail.

Pour obtenir le nombre de contrats 35h/semaines à signer pour une année de 52 semaines, on fait:

Entrée [96]:

18018 / (52*35)

Out[96]:

9.9

Pour combler les besoins de l'année à venir, les ressources humaines devront recruter 10 personnes.

Limites

Ce type d'étude peut comporter certaines limites:

1- Il pourrait être intéressant de baser la prédiction sur d'autres modèles et de comparer la qualité de chaque modèle. De cette manière, on s'assure d'avoir l'estimation la plus précise possible de nombre total d'heure de travail.

2- Le modèle ne prévoit pas une augmentation du prix des matières premières qui pourrait intervenir au cours de l'année ou une augmentation ponctuelle de la demande. On pourrait dans ce cas recourir à des contrats court type CDD ou intérim afin de combler les besoins.

Romain Vimont Septembre 2022