Sciences des données et la décision

Jonas Couturon

Sylvestre Jeannin

Yaasine Mosafeer

– 3PI –

Promo 2026

Sommaire

Exercice 1

On se propose d'étudier, d'étudier la production laitier d'une élevage selon le modèle d'alimentation.

On trouve un modèle d'équation : $y = 0.65x + 10.7$

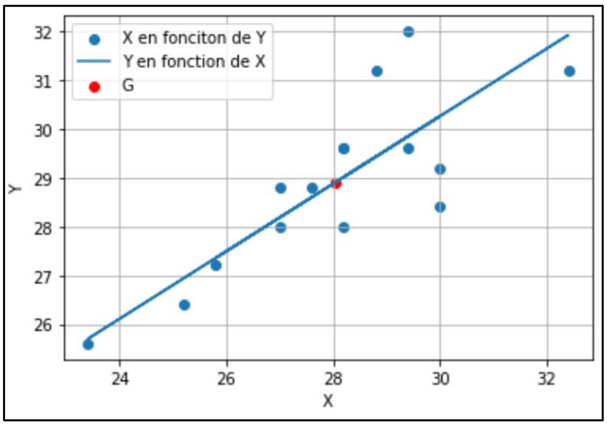


Figure 1 : Graphique du modèle trouvé en fonction de l'expérience

Selon la variable X : $m = 28.04$ et $\sigma = 2.13$ Selon la variable Y : $m = 28.91$ et $\sigma = 1.7$

On trouve : $cov(x, y) = 3.138$ et $r = 0{,}81 \approx 1$. Le modèle proposé semble suivre l'expérience.

```
Variance residuelle (M1) est  1.0288658719757016
Variance residuelle (M2) est  1.1871529292027327
Variance expliquée (M1) est  2.1720265579767544
Variance expliquée (M2) est  32.580398369651306
```

Figure 2 :Variance expliqué et résiduelle de la variable X et Y

On en déduit que le régime alimentaire du vétérinaire à un effet sur la production laitière.

Exercice 2

L'étude de la décharge d'un condensateur nous porte à supposer sa progression sous la forme :

$$V = V_0 e^{-\frac{t}{\tau}}$$

Soit sous la forme $y = ax + b$, $\ln(V) = -\frac{1}{\tau}t + \ln(V_0)$ Ainsi $\begin{cases} a = -\frac{1}{\tau} et\ \tau = -\frac{1}{a} \\ b = \ln(V_0)\ et\ V_0 = e^b \end{cases}$

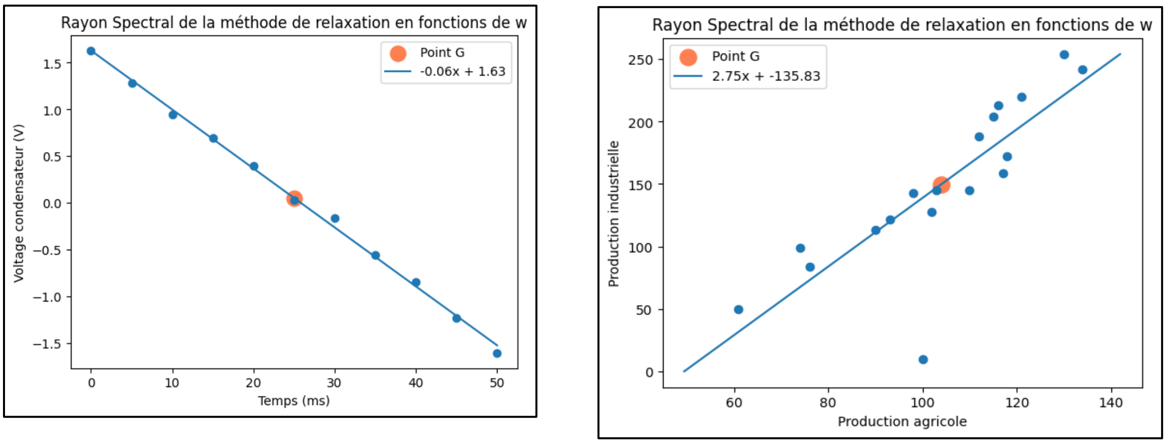
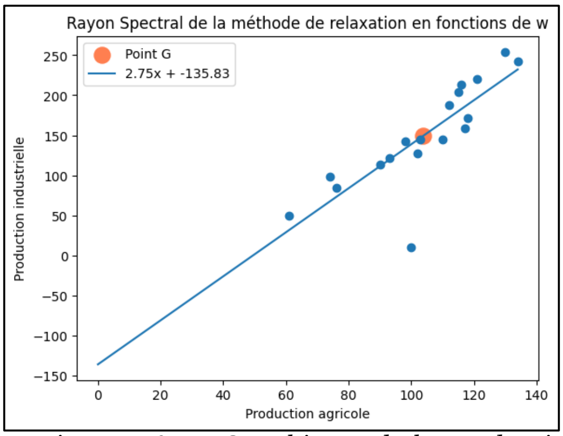


Figure 1 : Graphique de régression linéaire avec son point G et sa fonction

La décharge du condensateur suit donc la loi : $V = 5{,}10 * e^{-\frac{t}{16{,}66}}$

Exercice 3

On cherche à étudier l'existence d'un lien corrélatif qu'il pourrait y avoir entre la production agricole et la production industrielle entre 1944 et 1962.



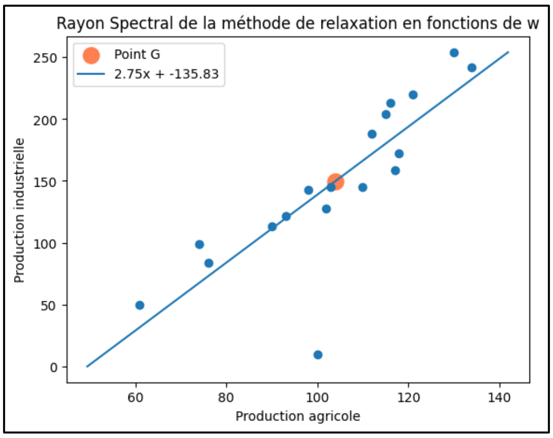


Figure 2 & 3 : Graphiques de la production agricole en fonction de la production industrielle (Y rectifié à gauche, X rectifié à droite)

On a : $cov(x, y) = 992{,}05$ et $r = 0{,}823 \approx 1$, mais reste bien en dessous de 0,95. L'expérience s'ajuste donc au modèle trouvé.

On ajuste successivement les valeurs de y puis de x :

```
variance :  4020.138888888887
variance résiduelle :  2724.660340761498
variance expliqué :  1295.4785481273886
```

```
variance :  361.20987654321107
variance résiduelle :  532.9522546418665
variance expliqué :  -171.74237809865537
```

Figure 6 : Variance, variance expliqué et résiduelle (pour Y à gauche, pour X à droite)

On cherche tout de même les formules de la variance expliquée et résiduelle pour X :

$$\frac{V_e}{V(X)} = \frac{V\left(\frac{1}{a}x - \frac{b}{a}\right)}{V(X)} = \left[\frac{V(Y)}{cov(X,Y)}\right]^2 \frac{V(X)}{V(Y)} = \frac{V(X)V(Y)}{[cov(X,Y)]^2} = \frac{1}{r^2}$$

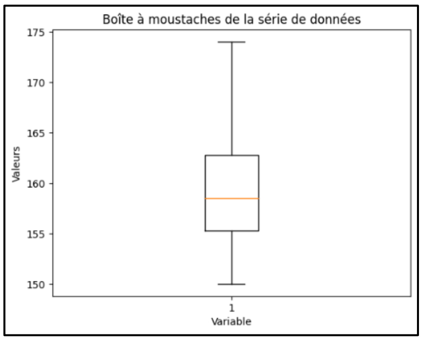
$$Vr = V(X) - V_e = V(X) - \frac{V(X)}{r^2} = V(X) * \left(1 - \frac{1}{r^2}\right)$$

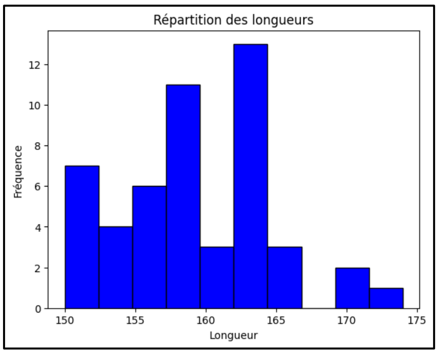
Exercice 4

1) La population étudier est une liste de tailles de la rectrice centrale de 50 male gelinotte huppées
2) La variable statistique est la taille des plumes de chaque mal.
3) C'est une variable discrète, elles sont toute numériques et il n'y a pas une infinité de valeur dans cette liste, il n'y a que 50 valeurs par une de plus
4) /
5) min = 150 max= 174 etendue= 24 moyenne= 159.08 mediane= 158.5 quartille Q1 Q2 et Q3= 155.25 158.5 162.75 iqr= 7.5
6) varaince 30.4736 ecart_type 5.520289847462722 moyenne 159.08

Si on calcul moyenne(+ou-) 2*ecartype ==> (140 à 170) on trouve presque tous les valeurs du tableau (sauf 3) ce qui est plutôt cohérant.

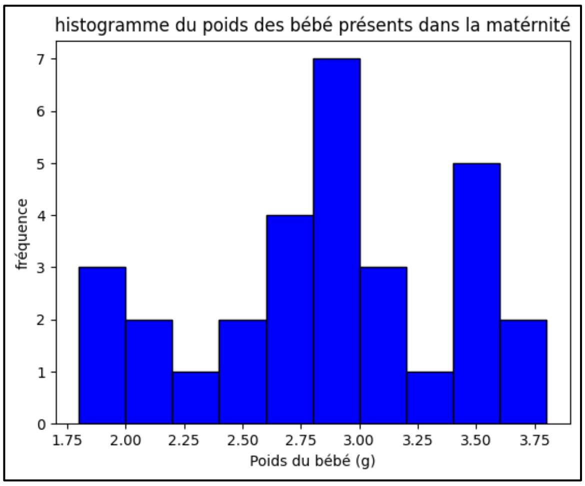
7) et 8) On retrouve bien les mêmes valeurs supposer plus haut sur le tableau si dessous de meme pour le second table les valeurs sont cohérant





Exercice 5

Dans cet exercice, nous étudions la pesé de bébé dans une maternité.



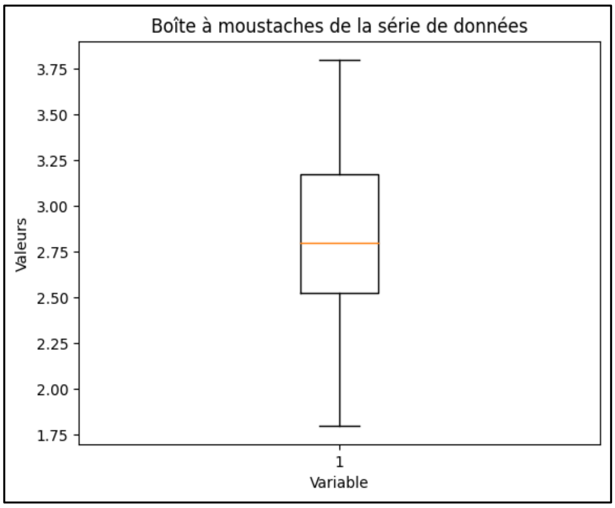


Figure 1& 2 : histogramme (à gauche) et boîte à moustache (à droite) du poids des bébés

Cette sérié a un espérance de $e = 2{,}76$ et un écart-type de $\sigma = 0.53$.

En séparant les donnés en deux partie égales, on trouve une moyenne de $m_1 = 2{,}93g$ pour les 15 premier bébés et une moyenne de $m_2 = 2{,}68g$ pour les 15 bébés suivants. En moyenne un bébé pèse 2,8 g dans la maternité.

Exercice 6

Dans cette expérience, on souhaite étudier le passage d'un portique dans un aéroport. On utilise une loi binominale, permettant de suivre le succès de l'expérience X, répété, « la personne qui passe le portique sonne ». Donc $X \hookrightarrow b(500\,; 0{,}031175)$.

3) E[X] = 15,5875 : environ 16 personnes passent le portique en le faisant sonner.

4) La probabilité qu'au moins une personne fasse sonner le portique est d'environ 1,000.

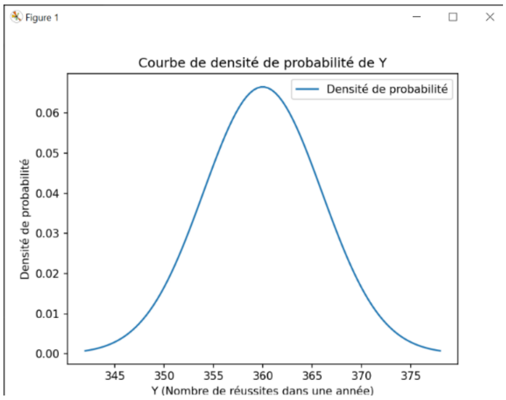
5) La probabilité qu'au maximum 3 personnes fassent sonner le portique est d'environ 0,0001.

6) Impossible de calculer P(X>250) facilement à l'aide de la loi binomiale.

8) et 9) A l'aide l'approximation normale, les probabilités que plus de 50 personnes fassent sonner le portique ou qu'exactement 70 personnes fassent sonner le portique sont nulles.

Exercice 7

1) La courbe de la densité de probabilité de Y



La ligne des Y varie d'espérance - 3*sigma à espérance + 3*sigma

La fonction $f(y) = \frac{1}{\sigma \times \sqrt{2\pi}} \times \exp\left(-\frac{(y-esp)^2}{2\times\sigma^2}\right)$

2) norme.cdf donne la proba d'un événement <ou= on cherche l'inverse. Le résultat est cohérent, car il on 90% de chance de réussir 400 opérations donc réussir 345 serait plus probable encore.

3) Rater plus de 28 opérations dans l'année soit réussir moins de 400-38 opération dans l'année

4) 1% de chance de rater ==> 99% de réussir on cherche le nombre de réussite en utilisant la probaliter de cette événement

```
2) probabilité de réussir au moin 345 0.9937903346742238
3) probabilité de rater plus de 28 opération (réusir moin de 400-28 opération) 0.9772498680518208
4) nb de opération rater dépassent 1% = 55.0
```

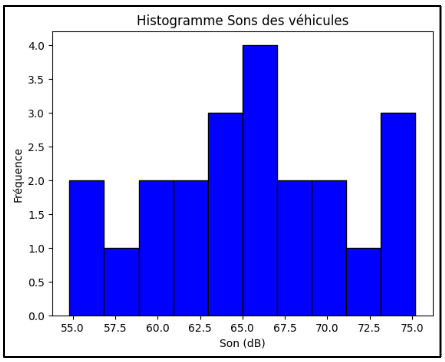
Exercice 8

Nous utilisons les formules suivantes :

$$\sigma_e = cov(x,x) = \frac{1}{n}\sum_{i=1}^{n=22}(x_i - \overline{x})^2 = 32{,}7$$

$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 65{,}22$ et $m_e = \frac{(\sum_{i=1}^{n=22} x_i)}{n} = 65{,}22$

Nous pouvons assimiler le bruit par une loi normale car celui-ci suit une loi normale de paramètres :
$\sigma^2 = 4253{,}65$ et $m = 65{,}22$
La loi de T suit une loi normale centrée réduite.



Exercice 9

On étudie la vente d'œuf sur une marché avec un échantillon $n = 100$

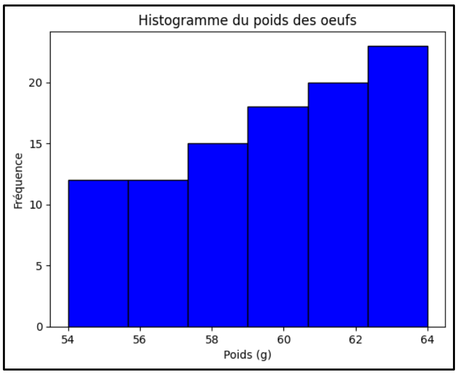


Figure 1 : Histogramme du poids de oeufs

Le vendeur souhaite assimiler son modèle selon une loi normale : $T \hookrightarrow N(0,1)$

On obtient donc : $m_e = 59{,}82$ (moyenne observé), $\sigma_e = 3{,}36$ (variance observé), $\sigma = 3{,}34$, $\beta = 0{,}95$

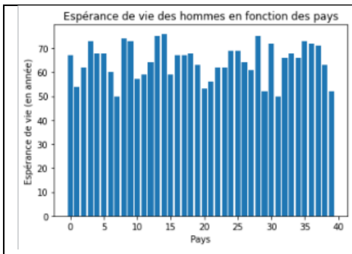
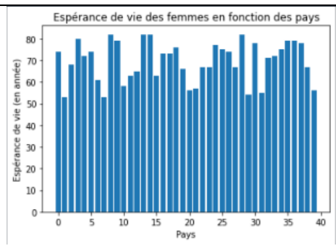
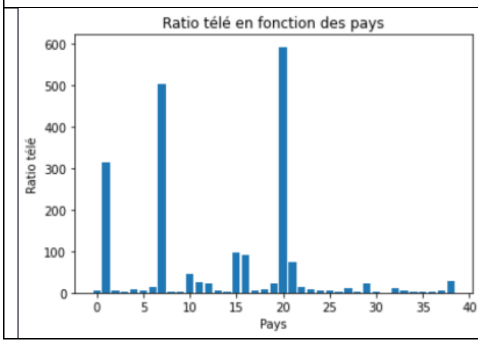
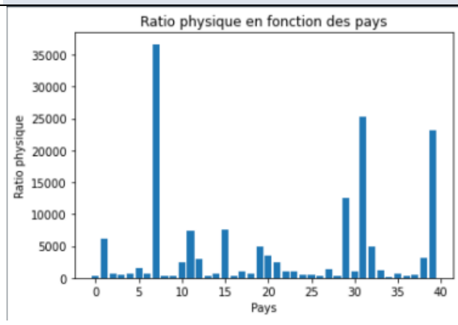
$$\beta = p(-u < T < u) = 0{,}95 \text{ et } p(T < u) = \beta + \frac{1-\beta}{2} = \frac{1}{2} + \frac{\beta}{2} = 0{,}975 \text{ soit } u = 1{,}96$$

Donc $I_c = [59{,}82 - 0{,}65\,;59{,}82 + 0{,}65] = [59{,}16\,; 60{,}47]$

On remarque que cet intervalle est très rapproché par rapport à la moyenne et situé dans l'intervalle $]59, 61]$. On peut donc considérer que le vendeur à bien choisi d'interpréter son expérience selon une loi normale.

Exercice 10

Dans exercice nous sommes invité à traiter un certain nombre de donné sur des statistiques par pays de la manière la plus utile possible.

<table>
<tr><td>
</td><td>
</td></tr>
<tr><td>nobs=40,
minmax=(50, 76),
mean=64.5,
variance=55.07692307692307, skewness=-0.3924245613203462, kurtosis=-0.8061832235781243</td><td>nobs=40,
minmax=(53, 82),
mean=69.575,
variance=83.9429487179487, skewness=-0.4071244791924928, kurtosis=-1.042109300337584</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

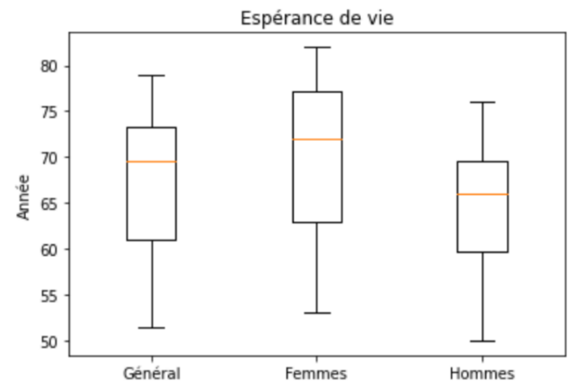
Jonas Couturon Sylvestre Jeannin Yaasine Mosafeer – 3PI -Promo 2026

nobs=40 minmax=(nan, nan), mean=nan variance=nan, skewness=nan urtosis=nan	nobs=40, minmax=(226, 36660), mean=3997.65, variance=58769573.720512815, skewness=2.9508177669079743, kurtosis=8.316645336032934

Les données atypiques du ratio télé sont : Pays n°32 et n°40

Les données atypiques du ratio physique : Aucune valeur

Voici les boites à moustache des espérances de vie des homme, femme, et en des deux



On remarque tout de suite que les femmes vivent plus longtemps que les hommes