

# Projet Pet Tracker



# Sommaire

## 1. Introduction et présentation du projet

Contexte du projet

Objectif de l'analyse statistique

## 2. Présentation de la variable quantitative étudiée

Définition d'une variable quantitative

Variable quantitative continue et discrète

Justification du choix de la variable

## 3. Analyse descriptive des données

Présentation des données observées

Calcul et interprétation de la moyenne

Calcul et interprétation de la médiane

## 4. Mesures de dispersion

Étendue

Variance

Écart-type

Interprétation dans le contexte PetTracker

## 5. Représentations graphiques des données

Histogramme avec étendue imposée (Dataset 1)

Histogramme suivant une loi normale (Dataset 2)

Comparaison des distributions

## 6. Boîte à moustaches de la distance parcourue

Quartiles (Q1, Q2, Q3)

Analyse de la dispersion et des valeurs extrêmes

## 7. Loi normale et règle des 68 % – 95 % – 99 %

Paramètres de la loi normale

Interprétation des intervalles de probabilité

## 8. Calculs de probabilités avec la table Z

Normalisation de la variable

Lecture de la table Z

Interprétation des résultats

## 9. Analyse de corrélation entre variables

Variables corrélées

Variables non corrélées

Corrélation et causalité

## 10. Régression linéaire et nuages de points

Génération des datasets sous Python

Nuages de points

Droites de régression linéaire

Interprétation des pentes

## 11. Fonctions Python utilisées

Génération des données

Régression linéaire

Visualisation graphique

## 12. Conclusion générale

Bilan des analyses réalisées

Validation de l'objectif du projet

# \* Introduction:



## Pet Tracker c'est Quoi?

PetTracker est une entreprise technologique qui propose un collier connecté intelligent destiné aux animaux de compagnie.

Grâce à une application mobile, les propriétaires peuvent suivre la position de leur animal en temps réel, recevoir des alertes en cas de fugue et analyser son activité quotidienne (distance parcourue, nombre de pas, durée de sommeil).

L'objectif de ce projet est d'analyser les données d'activité collectées par PetTracker à l'aide d'outils statistiques, afin de mieux comprendre le comportement quotidien de l'animal et d'évaluer son niveau de bien-être.

# Variable quantitative étudiée?



## c'est quoi une variable quantitaive

Une variable quantitative est une variable qui prend des valeurs numériques et permet de mesurer une quantité.

Quantitative discrète : valeurs entières issues d'un comptage.

Quantitative continue : valeurs mesurées pouvant être décimales.

## Alors c'est quoi notre variable quantitative:

**La distance parcourue par l'animal par jour, exprimée en kilomètres**, est une variable quantitative continue mesurant le niveau d'activité physique quotidien de l'animal équipé du collier PetTracker.

## Pourquoi on a choisi cette variable ?

Cette variable est le cœur des analyses de PetTracker.  
Elle permet d'évaluer la santé, le bien-être et l'état psychique de l'animal, et sert de base aux alertes et recommandations personnalisées.



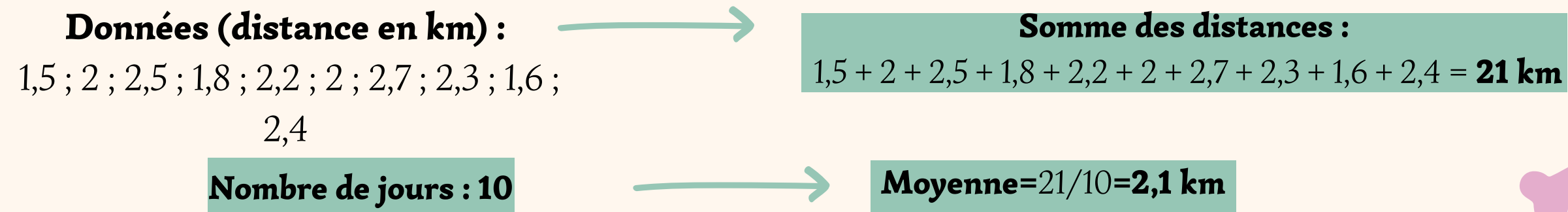
| Jour | Distance (km) |
|------|---------------|
| 1    | 1,5           |
| 2    | 2             |
| 3    | 2,5           |
| 4    | 1,8           |
| 5    | 2,2           |
| 6    | 2             |
| 7    | 2,7           |
| 8    | 2,3           |
| 9    | 1,6           |
| 10   | 2,4           |

# Calcul de la moyenne et de la médiane

- ❖ Quel est le niveau moyen d'activité de l'animal ?
- ❖ Quelle distance représente une journée "typique" pour l'animal ?

## Calculons la moyenne :

La moyenne est une mesure de tendance centrale qui représente la valeur moyenne d'un ensemble de données.  
Elle est **obtenue en additionnant toutes les valeurs observées, puis en divisant le résultat par le nombre total de valeurs.**  
**Dans le contexte de PetTracker, la moyenne permet d'estimer la distance journalière moyenne parcourue par l'animal, et donc son niveau d'activité global.**



| Jour | Distance (km) |
|------|---------------|
| 1    | 1,5           |
| 2    | 2             |
| 3    | 2,5           |
| 4    | 1,8           |
| 5    | 2,2           |
| 6    | 2             |
| 7    | 2,7           |
| 8    | 2,3           |
| 9    | 1,6           |
| 10   | 2,4           |

## Calculons la médiane:

La médiane est la valeur qui partage une série de données ordonnées en deux parties égales :  
50 % des valeurs sont en dessous et 50 % au-dessus.  
**Dans le contexte de PetTracker, la médiane représente la distance parcourue lors d'une journée typique, indépendamment des journées exceptionnellement très actives ou très calmes.**

**Étape 1 :** classer les distances par ordre croissant  
1,5 ; 1,6 ; 1,8 ; 2,0 ; **2,0 ; 2,2** ; 2,3 ; 2,4 ; 2,5 ; 2,7

**Étape 2 :** identifier la valeur médiane  
Nombre de valeurs = 10 (nombre pair) alors la médiane est la moyenne des 5<sup>e</sup> et 6<sup>e</sup> valeurs



# Calcul de la moyenne et de la médiane

**Donc:**

✿ **Quel est le niveau moyen d'activité de l'animal ?**

Le niveau moyen d'activité de l'animal est de 2,1 km par jour.

✿ **Quelle distance représente une journée “typique” pour l'animal ?**

Une journée typique pour l'animal correspond à une distance de 2,1 km.



# Calcul de l'étendue, la variance et l'écartype

- Les distances parcourues sont-elles proches ou éloignées de la moyenne ?
- L'activité quotidienne de l'animal est-elle stable ou variable ?

## l'étendue

L'étendue mesure l'écart entre la distance maximale et la distance minimale parcourues par l'animal.  
Dans le contexte de PetTracker, elle indique l'amplitude totale de variation de l'activité sur la période observée.

**Étendue**  
Valeur minimale = 1,5 km  
Valeur maximale = 2,7 km  
Etendue=2,7-1,5=1,2 km

## la variance

La variance mesure l'éloignement moyen des distances par rapport à la moyenne, en mettant ces écarts au carré.  
Dans le contexte de PetTracker, elle permet d'identifier le degré de fluctuation de l'activité physique de l'animal.

## l'écartype

L'écart-type est la racine carrée de la variance.  
Dans le contexte de PetTracker,il représente en moyenne de combien de kilomètres l'activité journalière s'écarte de la moyenne, ce qui le rend facilement interprétable pour le suivi du bien-être de l'animal.

- Pour chaque valeur, on calcule l'écart entre la distance observée et la moyenne afin de mesurer l'éloignement de chaque journée par rapport à l'activité moyenne.
- Les écarts sont élevés au carré afin d'éliminer les valeurs négatives et de mieux prendre en compte les écarts importants.
- La variance correspond à la moyenne des carrés des écarts à la moyenne. Elle permet de mesurer le niveau de dispersion de l'activité journalière.
- L'écart-type est obtenu en prenant la racine carrée de la variance, ce qui permet de revenir à l'unité initiale (le kilomètre) et de faciliter l'interprétation.

**Somme des carrés des écarts**  
0,36+0,01+0,16+0,09+0,01+0,01+0,36+0,04+0,25+0,09=1,38

**Calcul de la variance**  
Variance=1,38/10=0,138

**Calcul de l'écart-type**  
 $\sigma = \sqrt{0,138} \approx 0,37 \text{ km}$

En moyenne, la distance parcourue par l'animal s'écarte de 0,37 km autour de la moyenne de 2,1 km, ce qui indique une activité stable et régulière, signe de bon bien-être.

**l'utilité de l'écartype dans notre contexte?** Les valeurs situées au-delà d'un écart-type peuvent indiquer un changement de comportement nécessitant une attention particulière.

journées anormalement faibles : Distance <  $\mu - \sigma$   
journées anormalement élevées : Distance >  $\mu + \sigma$

| Distance (km) | Distance - 2,1 | (Distance - 2,1)² |
|---------------|----------------|-------------------|
| 1,5           | -0,6           | 0,36              |
| 2             | -0,1           | 0,01              |
| 2,5           | 0,4            | 0,16              |
| 1,8           | -0,3           | 0,09              |
| 2,2           | 0,1            | 0,01              |
| 2             | -0,1           | 0,01              |
| 2,7           | 0,6            | 0,36              |
| 2,3           | 0,2            | 0,04              |
| 1,6           | -0,5           | 0,25              |
| 2,4           | 0,3            | 0,09              |

# histogramme des valeurs de distance parcourue (en km)

Valeur minimale : 1,5 km

Valeur maximale : 2,7 km

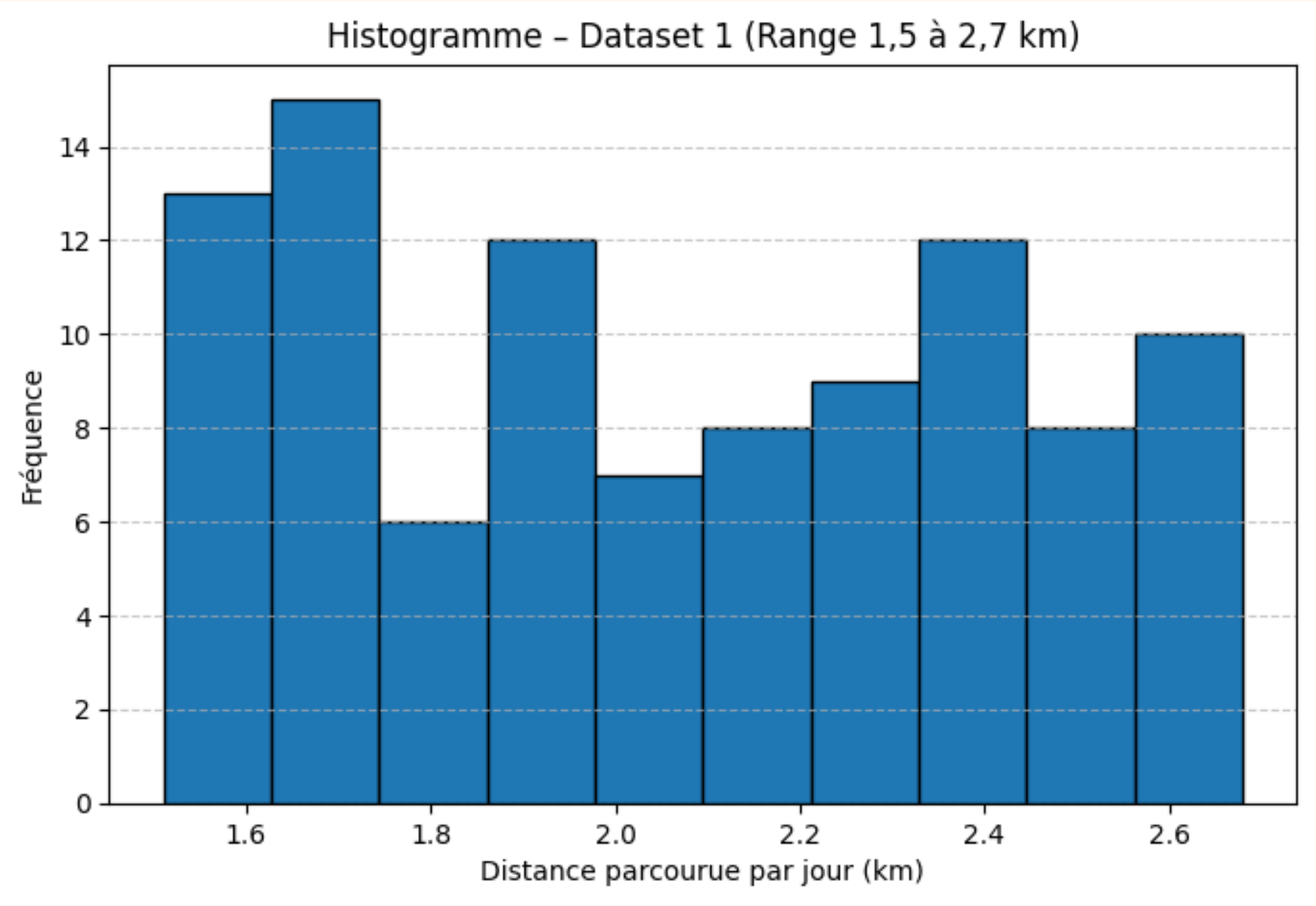
Étendue (range) :

$$2,7 - 1,5 = 1,2 \text{ km}$$

**Axe horizontal (x) :** distance parcourue par jour (en km)

**Axe vertical (y) :** fréquence (nombre de jours)

Chaque barre représente le nombre de jours pour lesquels la distance se situe dans un intervalle donné.



**Cet histogramme permet de visualiser la répartition des distances parcourues par l'animal sur 100 jours, en respectant l'étendue définie.**

# Histogramme de la distance parcourue par l'animal – Loi normale ( $\mu = 2,1$ km ; $\sigma = 0,21$ km)

Moyenne ( $\mu$ ) = 2,1 km

Écart-type ( $\sigma$ ) = 0,21 km (10 % de la moyenne)

100 valeurs

Axe horizontal (x) : distance parcourue par jour (en km)

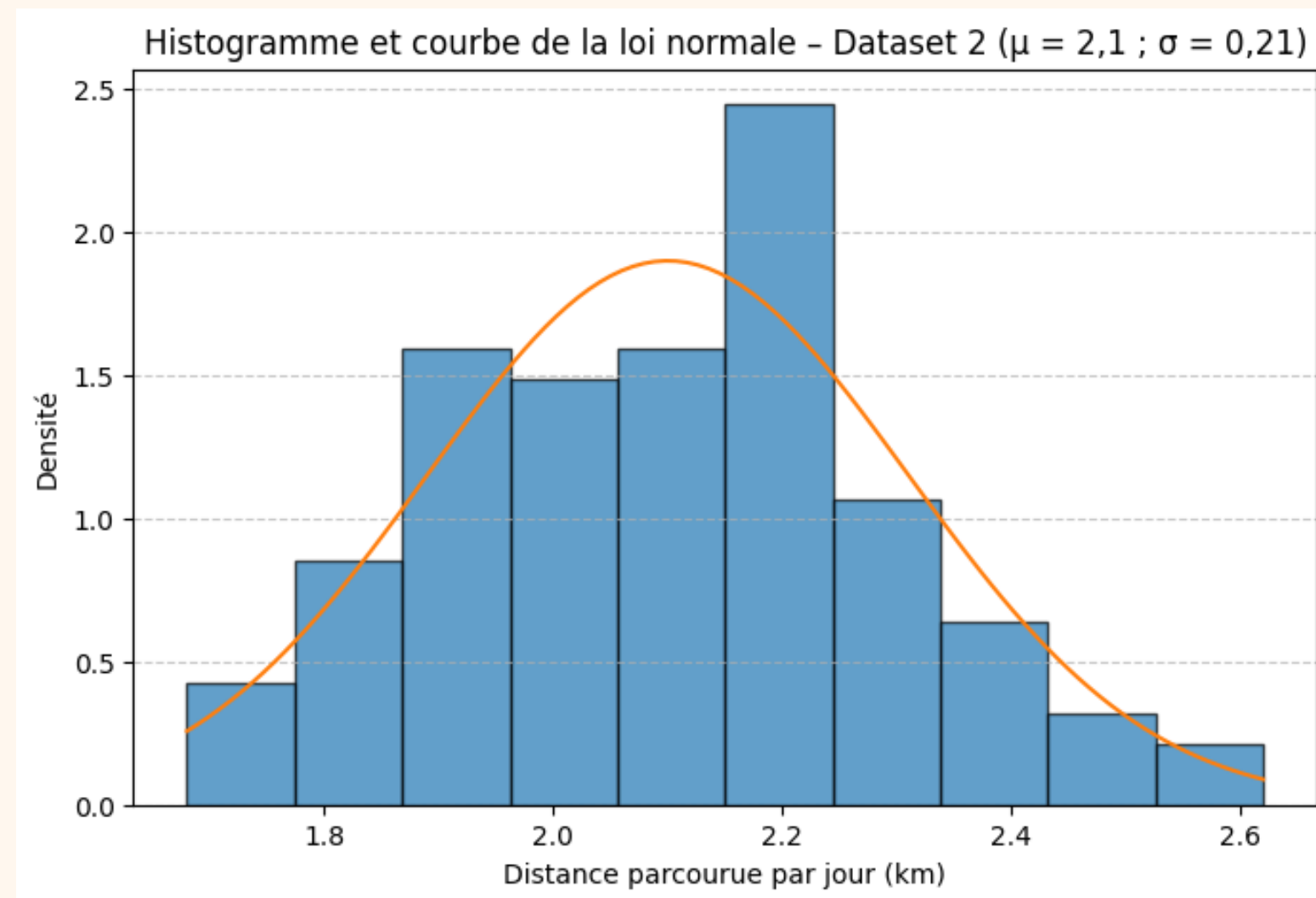
Axe vertical (y) : fréquence (nombre de jours)

Chaque barre représente le nombre de jours pour lesquels la distance se situe dans un intervalle donné.

## Loi Normale

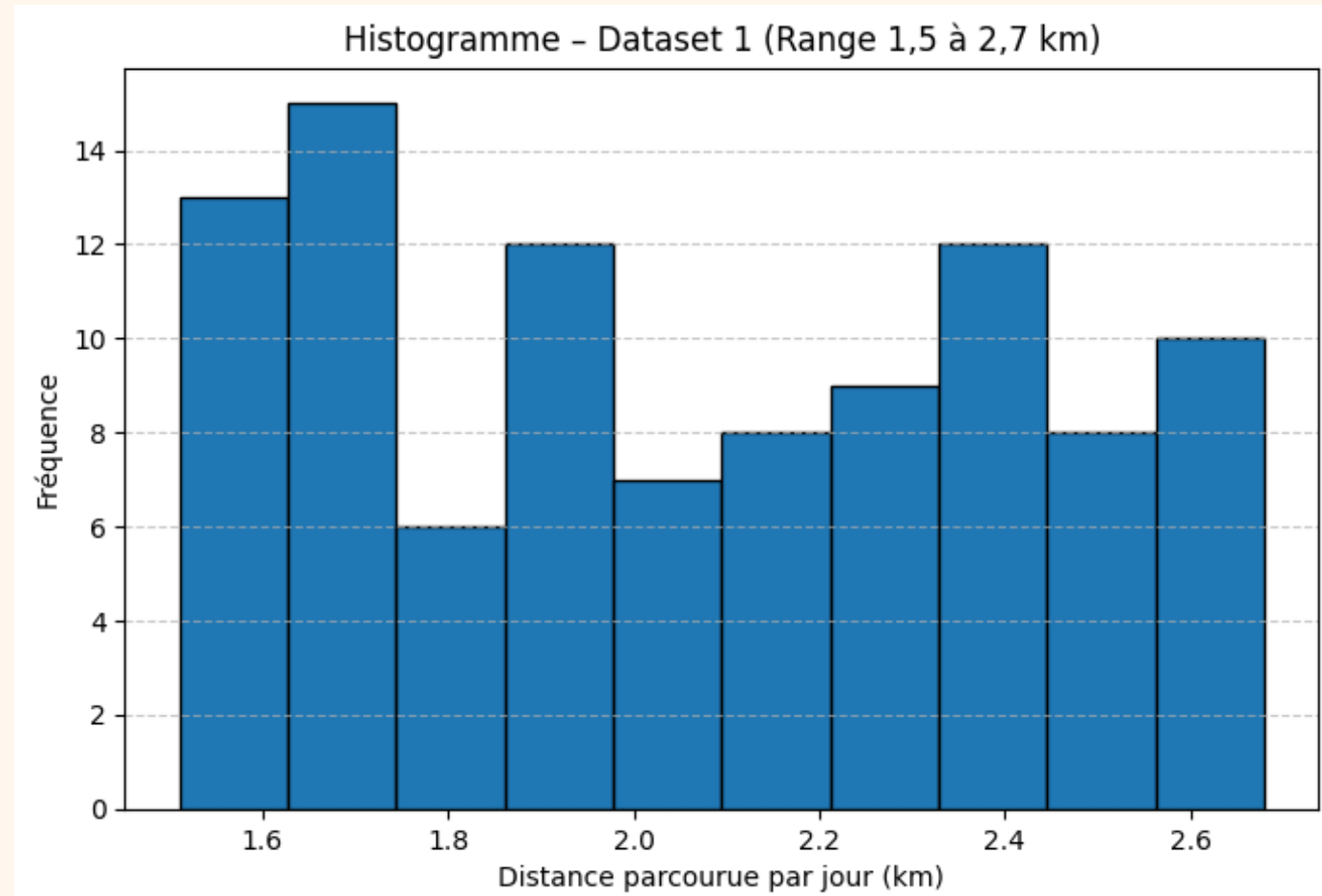
La loi normale est une distribution statistique qui décrit des données réparties de manière symétrique autour d'une moyenne.

La majorité des valeurs sont proches de la moyenne, tandis que les valeurs très petites ou très grandes sont rares.



**Cet histogramme illustre la distribution normale des distances parcourues par l'animal, centrée autour de 2,1 km, avec une dispersion modérée.**

# Comparaison entre les histogrammes – Dataset 1 vs Dataset 2



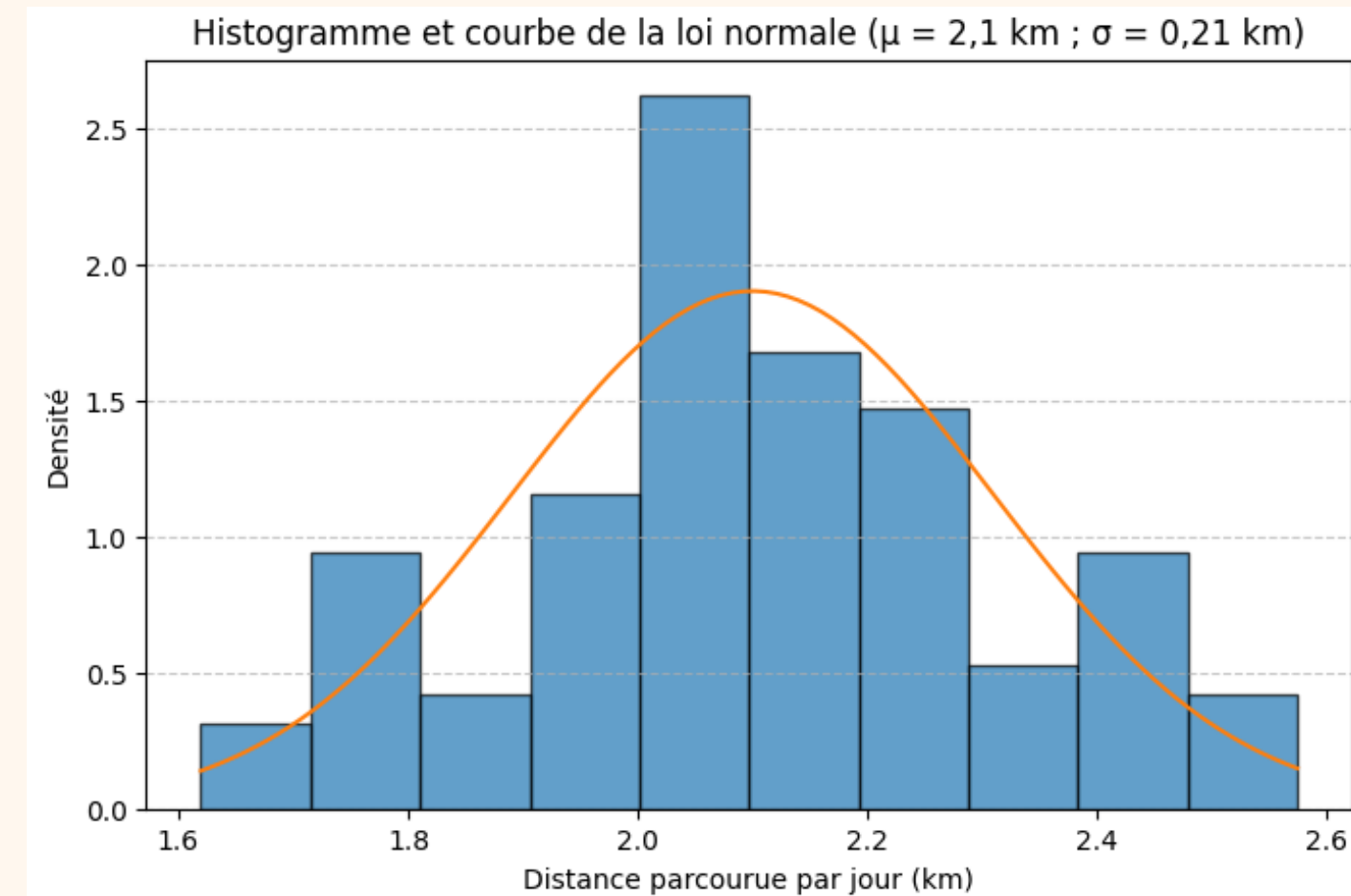
Dataset 1 (avec étendue imposée)

Les valeurs sont réparties sur un intervalle fixé [1,5;2,7]  
[1,5 ; 2,7][1,5;2,7]

L'histogramme ne suit pas une forme particulière  
Il montre uniquement la dispersion des valeurs dans la  
plage définie

Aucune hypothèse statistique n'est faite sur la  
distribution

Il sert surtout à visualiser l'amplitude et la variation  
globale des distances.



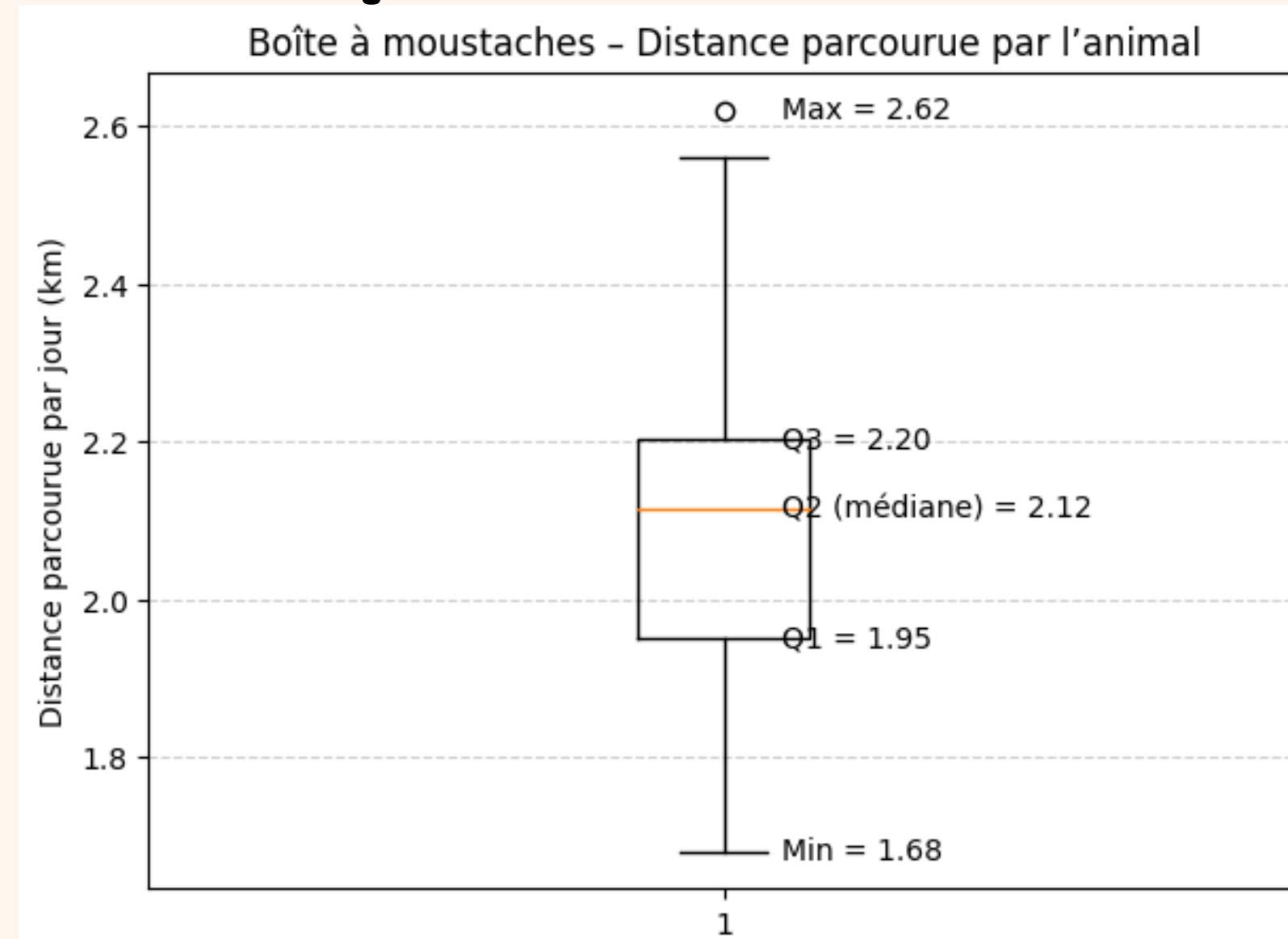
Dataset 2 (loi normale)

Les valeurs sont concentrées autour de la moyenne  
(2,1 km)

L'histogramme présente une forme en cloche  
Les distances proches de la moyenne sont les plus  
fréquentes

Les valeurs extrêmes sont plus rares  
Il représente un comportement plus réaliste et stable  
de l'activité quotidienne de l'animal.

# Boîte à moustaches représentant la distance parcourue par l'animal



**Cette boîte à moustaches représente la distribution des distances parcourues par l'animal par jour (en km).**

**La valeur minimale observée est de 1,68 km**, ce qui correspond aux journées les moins actives.

**Le premier quartile (Q1) est égal à 1,95 km** : 25 % des journées présentent une distance inférieure à cette valeur.

**La médiane (Q2) est de 2,12 km**, ce qui représente une journée d'activité typique pour l'animal.

**Le troisième quartile (Q3) est égal à 2,20 km** : 75 % des journées présentent une distance inférieure à cette valeur.

**La valeur maximale observée est de 2,62 km**, correspondant aux journées les plus actives.

**L'intervalle interquartile (Q3 – Q1)** est relativement réduit, ce qui indique une variabilité modérée de l'activité quotidienne.

La distribution est globalement centrée autour de la médiane, sans présence de valeurs aberrantes marquées.

# la courbe de la loi normale

## Paramètres PetTracker

Moyenne ( $\mu$ ) = 2,1 km

Écart-type ( $\sigma$ ) = 0,21 km

## Intervalles 68/95/99

≈ 68% des distances sont entre :  $\mu - \sigma = 2,1 - 0,21 = 1,89$  et  $\mu + \sigma = 2,1 + 0,21 = 2,31$  [1,89 ; 2,31] km

≈ 95% des distances sont entre :  $\mu \pm 2\sigma = 2,1 \pm 0,42$  [1,68 ; 2,52] km

≈ 99% des distances sont entre :  $\mu \pm 3\sigma = 2,1 \pm 0,63$  [1,47 ; 2,73] km

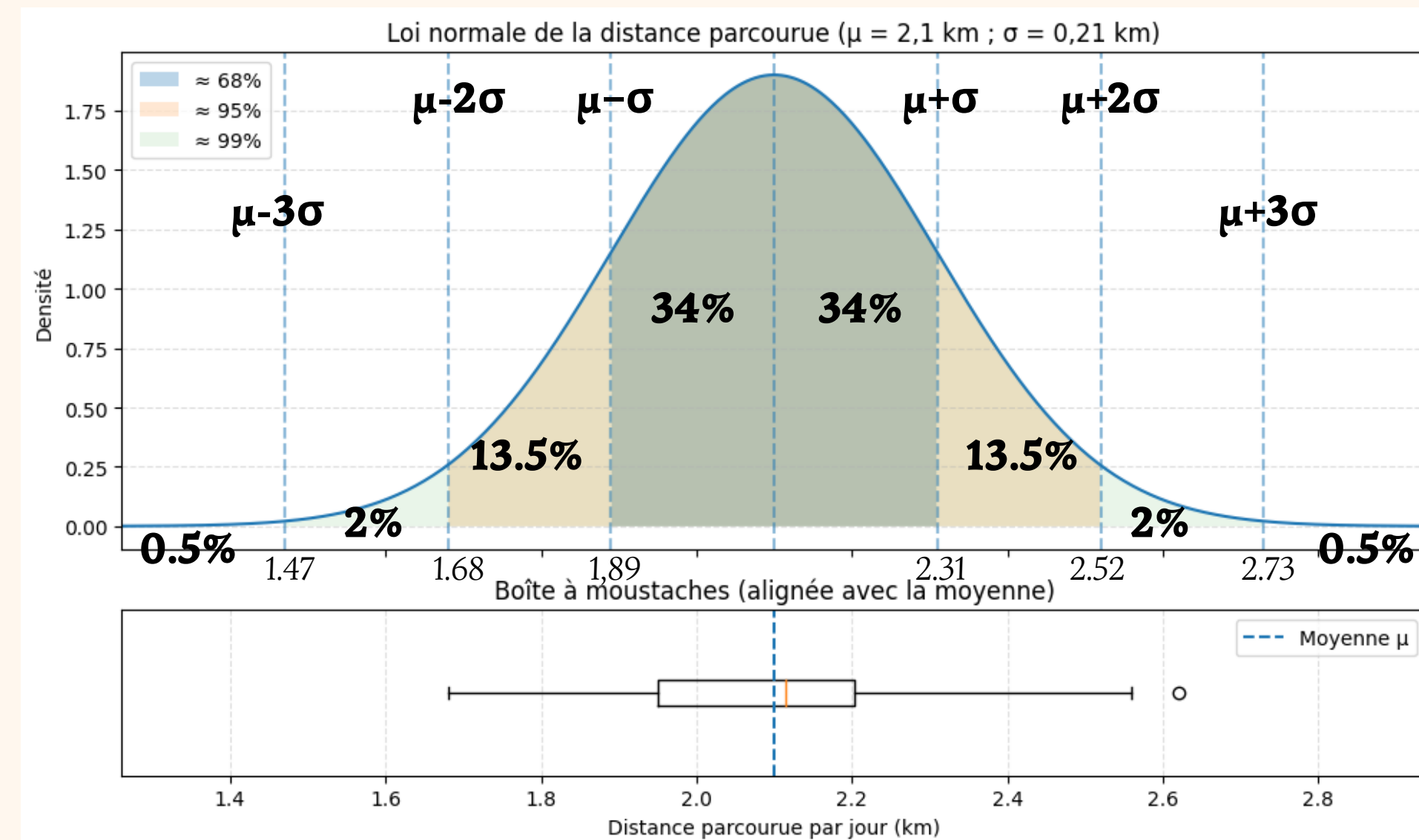
« Il y a environ 68% de chance que la distance parcourue par jour soit comprise entre 1,89 km et 2,31 km. »

« Il y a environ 95% de chance que la distance parcourue par jour soit comprise entre 1,68 km et 2,52 km. »

« Il y a environ 99% de chance que la distance parcourue par jour soit comprise entre 1,47 km et 2,73 km. »

« La probabilité que la distance parcourue soit supérieure à la moyenne (2,1 km) est de 50% (symétrie de la loi normale). »

« La probabilité que la distance parcourue dépasse 2,31 km ( $\mu + 1\sigma$ ) est d'environ 16%. »



# Calcul de probabilité avec la table Z

La variable suivant une loi normale, on la normalise afin d'utiliser la table de la loi normale et déterminer la probabilité recherchée.

On recherche la probabilité que la distance parcourue par jour soit inférieure à 2,31 km.

On sait que la variable suit une loi normale de :

$$\begin{aligned}\text{moyenne } (\mu) &= 2,1 \text{ km} \\ \text{écart-type } (\sigma) &= 0,21 \text{ km}\end{aligned}$$

On transforme la variable **X** en une variable **Z** suivant une loi normale centrée réduite :

$$Z = \frac{X - \mu}{\sigma}$$

Pour  $X=2,31$  :

$$Z = \frac{2.31 - 2.1}{0.21} = \frac{0.21}{0.21} = 1$$

On recherche dans la table de la loi normale :  **$P(Z < 1)$**

D'après la table Z :

$$P(Z < 1) = \mathbf{0,8413}$$

La probabilité que la distance parcourue par l'animal par jour soit inférieure à 2,31 km est donc d'environ 84,13 %.

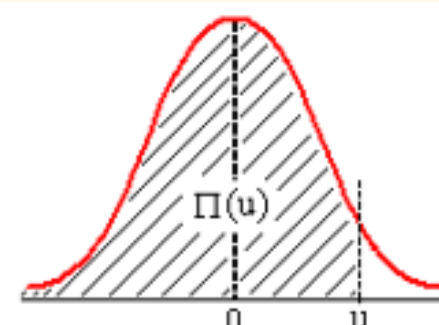
La variable Z est une variable dite centrée réduite, qui suit une loi normale de :

moyenne 0  
écart-type 1

La transformation en Z permet de mesurer l'écart entre une valeur X et la moyenne, exprimé en nombre d'écarts-types.

**Cette transformation ne modifie pas la probabilité, elle permet uniquement d'utiliser la table de la loi normale standard.**

Table de Loi Normale  
 $P(x < u)$



|     | 0,00   | 0,01   | 0,02   | 0,03   | 0,04   | 0,05   | 0,06   | 0,07   | 0,08   | 0,09   |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0,0 | 0,5000 | 0,5040 | 0,5080 | 0,5120 | 0,5160 | 0,5199 | 0,5239 | 0,5279 | 0,5319 | 0,5359 |
| 0,1 | 0,5398 | 0,5438 | 0,5478 | 0,5517 | 0,5557 | 0,5596 | 0,5636 | 0,5675 | 0,5714 | 0,5753 |
| 0,2 | 0,5793 | 0,5832 | 0,5871 | 0,5910 | 0,5948 | 0,5987 | 0,6026 | 0,6064 | 0,6103 | 0,6141 |
| 0,3 | 0,6179 | 0,6217 | 0,6255 | 0,6293 | 0,6331 | 0,6368 | 0,6406 | 0,6443 | 0,6480 | 0,6517 |
| 0,4 | 0,6554 | 0,6591 | 0,6628 | 0,6664 | 0,6700 | 0,6736 | 0,6772 | 0,6808 | 0,6844 | 0,6879 |
| 0,5 | 0,6915 | 0,6950 | 0,6985 | 0,7019 | 0,7054 | 0,7088 | 0,7123 | 0,7157 | 0,7190 | 0,7224 |
| 0,6 | 0,7257 | 0,7291 | 0,7324 | 0,7357 | 0,7389 | 0,7422 | 0,7454 | 0,7486 | 0,7517 | 0,7549 |
| 0,7 | 0,7580 | 0,7611 | 0,7642 | 0,7673 | 0,7704 | 0,7734 | 0,7764 | 0,7794 | 0,7823 | 0,7852 |
| 0,8 | 0,7881 | 0,7910 | 0,7939 | 0,7967 | 0,7995 | 0,8023 | 0,8051 | 0,8078 | 0,8106 | 0,8133 |
| 0,9 | 0,8159 | 0,8186 | 0,8212 | 0,8238 | 0,8254 | 0,8289 | 0,8315 | 0,8340 | 0,8365 | 0,8389 |
| 1,0 | 0,8413 | 0,8438 | 0,8461 | 0,8485 | 0,8508 | 0,8531 | 0,8554 | 0,8577 | 0,8599 | 0,8621 |
| 1,1 | 0,8645 | 0,8665 | 0,8686 | 0,8708 | 0,8729 | 0,8749 | 0,8770 | 0,8790 | 0,8810 | 0,8830 |
| 1,2 | 0,8849 | 0,8869 | 0,8888 | 0,8907 | 0,8925 | 0,8944 | 0,8962 | 0,8980 | 0,8997 | 0,9015 |
| 1,3 | 0,9032 | 0,9049 | 0,9066 | 0,9082 | 0,9099 | 0,9115 | 0,9131 | 0,9147 | 0,9162 | 0,9177 |
| 1,4 | 0,9192 | 0,9207 | 0,9222 | 0,9236 | 0,9251 | 0,9265 | 0,9279 | 0,9292 | 0,9306 | 0,9319 |
| 1,5 | 0,9332 | 0,9345 | 0,9357 | 0,9370 | 0,9382 | 0,9394 | 0,9406 | 0,9418 | 0,9429 | 0,9441 |
| 1,6 | 0,9452 | 0,9463 | 0,9474 | 0,9484 | 0,9495 | 0,9505 | 0,9515 | 0,9525 | 0,9535 | 0,9545 |
| 1,7 | 0,9554 | 0,9564 | 0,9573 | 0,9582 | 0,9591 | 0,9599 | 0,9608 | 0,9616 | 0,9625 | 0,9633 |
| 1,8 | 0,9641 | 0,9649 | 0,9656 | 0,9664 | 0,9671 | 0,9678 | 0,9686 | 0,9693 | 0,9699 | 0,9706 |
| 1,9 | 0,9713 | 0,9719 | 0,9726 | 0,9732 | 0,9738 | 0,9744 | 0,9750 | 0,9756 | 0,9761 | 0,9767 |
| 2,0 | 0,9772 | 0,9778 | 0,9783 | 0,9788 | 0,9793 | 0,9798 | 0,9803 | 0,9808 | 0,9812 | 0,9817 |
| 2,1 | 0,9821 | 0,9826 | 0,9830 | 0,9834 | 0,9838 | 0,9842 | 0,9846 | 0,9850 | 0,9854 | 0,9857 |
| 2,2 | 0,9861 | 0,9864 | 0,9868 | 0,9871 | 0,9875 | 0,9878 | 0,9881 | 0,9884 | 0,9887 | 0,9890 |
| 2,3 | 0,9893 | 0,9896 | 0,9898 | 0,9901 | 0,9904 | 0,9906 | 0,9909 | 0,9911 | 0,9913 | 0,9916 |
| 2,4 | 0,9918 | 0,9920 | 0,9922 | 0,9925 | 0,9927 | 0,9929 | 0,9931 | 0,9932 | 0,9934 | 0,9936 |
| 2,5 | 0,9938 | 0,9940 | 0,9941 | 0,9943 | 0,9945 | 0,9946 | 0,9948 | 0,9949 | 0,9951 | 0,9952 |
| 2,6 | 0,9953 | 0,9955 | 0,9956 | 0,9957 | 0,9959 | 0,9960 | 0,9961 | 0,9962 | 0,9963 | 0,9964 |
| 2,7 | 0,9965 | 0,9966 | 0,9967 | 0,9968 | 0,9969 | 0,9970 | 0,9971 | 0,9972 | 0,9973 | 0,9974 |
| 2,8 | 0,9974 | 0,9975 | 0,9976 | 0,9977 | 0,9977 | 0,9978 | 0,9979 | 0,9979 | 0,9980 | 0,9981 |
| 2,9 | 0,9981 | 0,9982 | 0,9982 | 0,9983 | 0,9984 | 0,9984 | 0,9985 | 0,9985 | 0,9986 | 0,9986 |
| 3,0 | 0,9987 | 0,9987 | 0,9987 | 0,9988 | 0,9988 | 0,9989 | 0,9989 | 0,9989 | 0,9990 | 0,9990 |
| 3,1 | 0,9990 | 0,9991 | 0,9991 | 0,9991 | 0,9992 | 0,9992 | 0,9992 | 0,9992 | 0,9993 | 0,9993 |
| 3,2 | 0,9993 | 0,9993 | 0,9994 | 0,9994 | 0,9994 | 0,9994 | 0,9994 | 0,9995 | 0,9995 | 0,9995 |
| 3,3 | 0,9995 | 0,9995 | 0,9995 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9997 |
| 3,4 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9998 |

# Calcul de probabilité avec la table Z

**On recherche la probabilité que la distance parcourue par jour soit supérieure à 2,31 km.**

On sait que la variable suit une loi normale de :

moyenne ( $\mu$ ) = 2,1 km

écart-type ( $\sigma$ ) = 0,21 km

D'après la table Z :

$P(Z < 1) = \mathbf{0,8413}$

$$P(Z > 1) = 1 - 0,8413 = 0,1587$$

**La probabilité que la distance parcourue par l'animal par jour soit supérieure à 2,31 km est donc d'environ 15,87 %.**

# Calcul de probabilité avec la table Z

**On recherche la probabilité que la distance parcourue par jour soit comprise entre 1,89 km et 2,31 km.**

On sait que la variable suit une loi normale de :

$$\begin{aligned}\text{moyenne } (\mu) &= 2,1 \text{ km} \\ \text{écart-type } (\sigma) &= 0,21 \text{ km}\end{aligned}$$

**On transforme la variable X en une variable Z suivant une loi normale centrée réduite :**

$$Z = \frac{X - \mu}{\sigma}$$

Pour  $X=1.89$

$$Z_1 = \frac{1.89 - 2.1}{0.21} = \frac{-0.21}{0.21} = -1$$

Pour  $X=2,31$  :

$$Z_2 = \frac{2.31 - 2.1}{0.21} = \frac{0.21}{0.21} = 1$$

**On recherche dans la table de la loi normale :**

D'après la table Z :

$$P(-1 < Z < 1) = 0,8413 - 0,1587 = 0,6826$$

**La probabilité que la distance parcourue par l'animal par jour soit comprise entre 1,89 km et 2,31 km est donc d'environ 68,26 %.**

# Variables quantitatives du projet PetTracker



## Variables corrélées

- Distance parcourue par jour (km)
- Calories brûlées par jour

**Logique :** plus l'animal marche, plus il brûle de calories

**Corrélation positive**



## Variables non corrélées

- Distance parcourue par jour (km)
- Durée de sommeil par jour (heures)

**Logique :** La durée de sommeil ne dépend pas directement de la distance parcourue

**Pas de corrélation**

**but**

Générer un dataset pour ces variables- représenter les 2 nuages de points correspondants. - calculer la droite de régression linéaire.

# code python qui génère le nuage de points de distance parcourue vs calories brûlées

```
: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

: np.random.seed(42) #Fixe le hasard pour avoir toujours les mêmes résultats (pas obligatoire)

: # DATASET 1 : VARIABLES CORRÉLÉES
# Distance parcourue vs calories brûlées
#np.random.uniform(a, b, n) Génère n nombres au hasard entre a et b
distance = np.random.uniform(1.5, 3.0, 100) # km
calories = 50 * distance + np.random.normal(0, 10, 100)

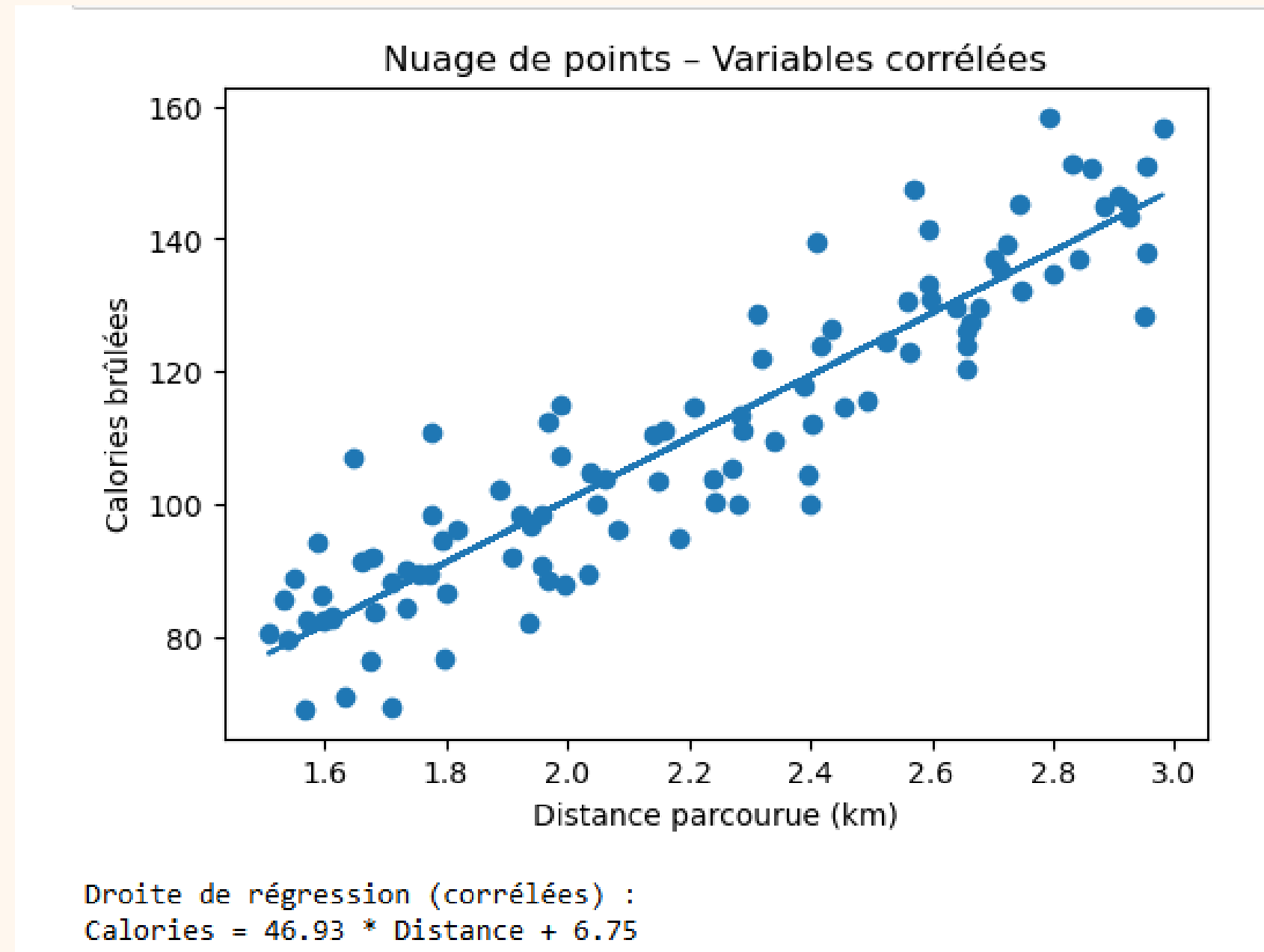
: df_corr = pd.DataFrame({
    "Distance_km": distance,
    "Calories_brulees": calories
})

: # Droite de régression linéaire
#np.polyfit(x, y, 1) Calcule la droite qui passe le mieux entre les points 1 signifie 1 seule droite
coef_corr = np.polyfit(distance, calories, 1)
#np.poly1d(...) Transforme le résultat en une fonction de droite
reg_corr = np.poly1d(coef_corr)

: # Nuage de points corrélié
plt.figure(figsize=(6,4))
plt.scatter(distance, calories)
plt.plot(distance, reg_corr(distance))
plt.xlabel("Distance parcourue (km)")
plt.ylabel("Calories brûlées")
plt.title("Nuage de points - Variables corrélées")
plt.show()

print("Droite de régression (corrélées) :")
print(f"Calories = {coef_corr[0]:.2f} * Distance + {coef_corr[1]:.2f}")
```

# le nuage de points de distance parcourue vs calories brulées



X : Distance parcourue (km)

Y : Calories brûlées

1 point = 1 journée

Ce nuage de points montre une corrélation positive entre la distance parcourue et les calories brûlées. La droite de régression est croissante.

NB: Corrélation  $\neq$  causalité (même si ici Distance  $\rightarrow$  Calories est logique)

$$\text{Calories} = 46.93 \times \text{Distance} + 6.75$$

+1 km  $\approx$  +46.93 calories en moyenne.

Si la pente a est clairement  $\neq 0$  et que les points suivent une tendance  $\rightarrow$  corrélation

# code python qui génère le nuage de points de distance parcourue vs durée de sommeil

```
# DATASET 2 : VARIABLES NON CORRÉLÉES  
# Distance parcourue vs durée de sommeil
```

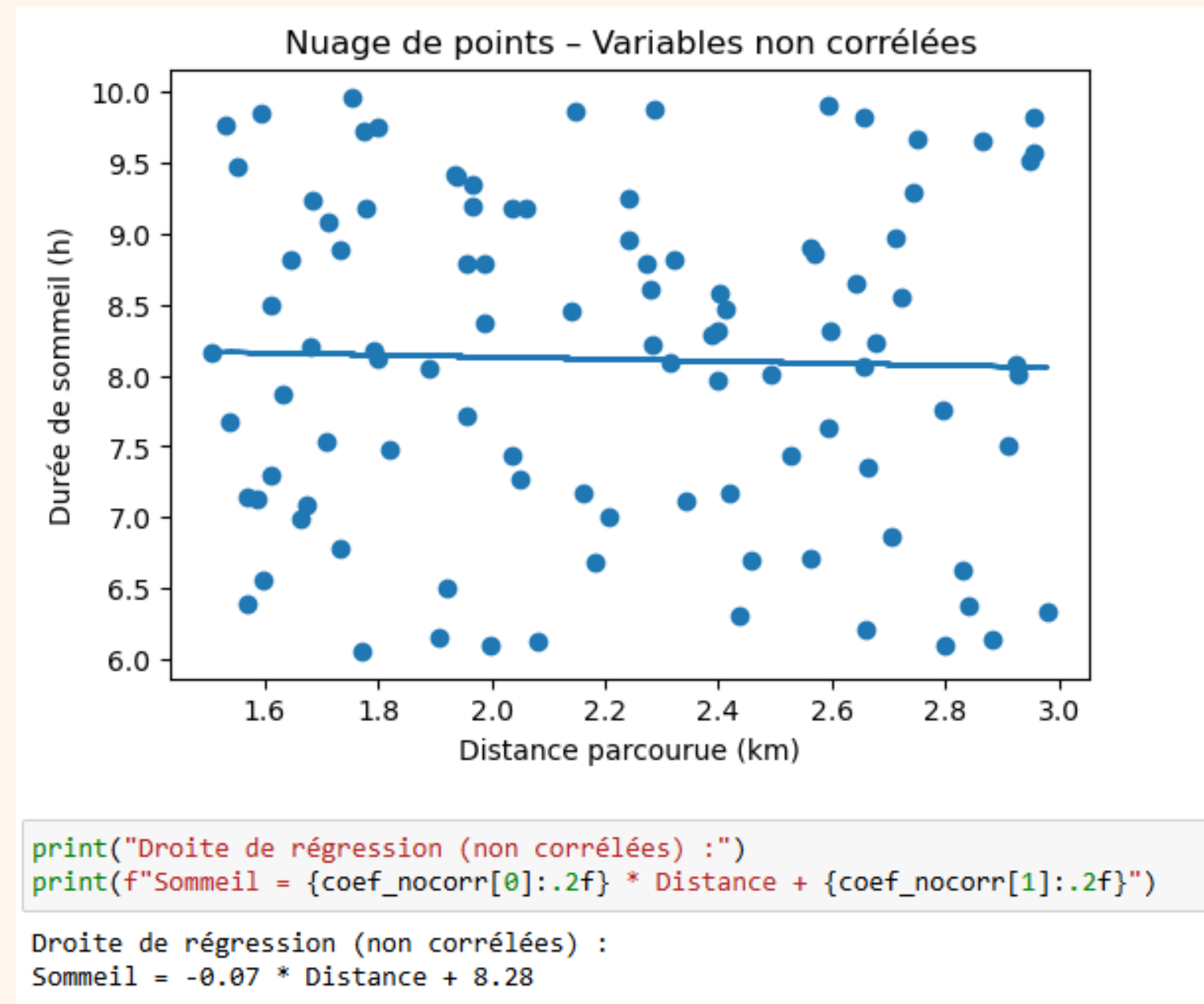
```
sleep = np.random.uniform(6, 10, 100) # heures de sommeil indépendantes
```

```
df_nocorr = pd.DataFrame({  
    "Distance_km": distance,  
    "Duree_sommeil_h": sleep  
})
```

```
# Droite de régression Linéaire  
coef_nocorr = np.polyfit(distance, sleep, 1)  
reg_nocorr = np.poly1d(coef_nocorr)
```

```
# Nuage de points non corrélé  
plt.figure(figsize=(6,4))  
plt.scatter(distance, sleep)  
plt.plot(distance, reg_nocorr(distance))  
plt.xlabel("Distance parcourue (km)")  
plt.ylabel("Durée de sommeil (h)")  
plt.title("Nuage de points - Variables non corrélées")  
plt.show()
```

# le nuage de points de distance parcourue vs durée de sommeil



X : Distance parcourue (km)

Y : Durée de sommeil (h)

Points très dispersés → pas de tendance  
Droite presque horizontale → absence de  
corrélacion

Ce nuage de points ne montre pas  
de relation entre la distance  
parcourue et la durée de sommeil.  
La droite de régression est presque  
horizontale.

$$\text{Sommeil} = -0.07 \times \text{Distance} + 8.28$$

pente proche de 0  $\Rightarrow$  distance n'explique pas le sommeil.

Si la pente est proche de 0 et les points sont dispersés → pas de corrélation



# Fonctions Python utilisées

## 1) Générer les données

```
np.random.seed(42)
```

Fixe le hasard → mêmes résultats à chaque exécution (utile pour rendre un travail stable)

```
np.random.uniform(a, b, n)
```

Crée n valeurs aléatoires entre a et b (répartition uniforme)

```
np.random.normal(m, s, n)
```

Crée n valeurs autour d'une moyenne m avec une dispersion s (bruit réaliste)

## 2) Construire la droite de régression

```
np.polyfit(x, y, 1)
```

Calcule la meilleure droite  $\hat{Y} = aX + b$

→ renvoie a (pente) et b (intercept)

```
np.poly1d([a, b])
```

Transforme (a,b) en une fonction utilisable :  $\hat{Y}(x)$

## 3) Visualiser

```
plt.scatter(x, y) : nuage de points
```

```
plt.plot(x,  $\hat{Y}(x)$ ) : trace la droite de régression
```

```
plt.xlabel / plt.ylabel / plt.title : rend le graphique lisible
```



# Conclusion

L'objectif du projet a été atteint, puisque l'analyse statistique réalisée (statistiques descriptives, dispersion, probabilités, corrélation et régression linéaire) a permis de :

- caractériser le niveau d'activité moyen de l'animal,
- identifier une journée d'activité typique,
- mettre en évidence des relations entre certaines variables,
- et illustrer l'apport de l'analyse de données pour le suivi du bien-être animal.