Université de Pau et des Pays de l'Adour — **Semestre printemps 2023–2024**

L3 Math — Introduction à la Statistique

Feuille TD 5 : Régression linéaire et application

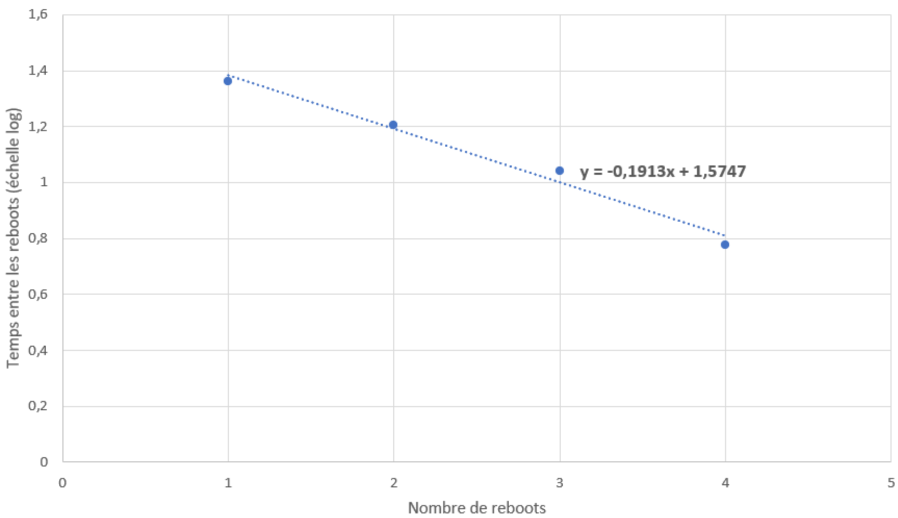
Exercice 5.1 Reboot des films *Batman*

En 2022, la quatrième série de films Batman est sortie au cinéma. Nous nous penchons ici sur le temps entre chaque remake.

1. *Batman (The movie)* de Leslie H. Martinson en 1966
2. *Batman* de Tim Burton en 1989
3. *Batman Begins* de Christopher Nolan en 2005
4. *Batman v Superman* de Zack Snyder en 2016
5. *The Batman* de Matt Reeves en 2022

Il est intéressant de noter que le logarithme (en base 10) du délai entre ces films évolue presque de manière linéaire. Dans le graphique ci-dessus, une régression linéaire a été réalisée, donnant les valeurs de

$$\hat{a} = -0.1913 \text{ et } \hat{b} = 1.5747.$$



Nous allons utiliser cette observation pour essayer de prédire la sortie de futurs films de cette série.

1. Estimer la date de sortie du cinquième reboot de Batman.

2. À quelle fréquence les reboots se produiront-ils à partir de 2033 ? Peut-on prévoir la fin du monde avec ce modèle ?

Exercice 5.2 Estimation d'une ellipse

L'équation d'une ellipse dans le plan de centre $C(a, b)$ est donnée par :

$$\frac{(x-a)^2}{\alpha^2} + \frac{(y-b)^2}{\beta^2} = 1$$

où α et β sont deux paramètres. On appelle ratio de cette ellipse la quantité $\tau = \frac{\alpha}{\beta}$. On remarque que si $\tau = 1$, alors on retrouve l'équation du cercle de centre C et de rayon α.

On dispose d'une série de n points du plan (x_i, y_i) pour $i \in \{1, ..., n\}$ non-alignés qu'on suppose appartenir à une ellipse $\mathcal{E}$ de centre $C(a, 0)$ dont le ratio $\tau = \frac{\alpha}{\beta}$ est connu. Les paramètres inconnus sont a, α et β. On cherche à déterminer par la méthode des moindres carrées des estimateurs de ces trois quantités.

1. Soit (x, y) un point appartenant à l'ellipse $\mathcal{E}$. Justifier la relation suivante :

$$(x - a)^2 + \tau^2 y^2 = \alpha^2.$$

On en déduit que les observations (x_i, y_i), $i \in \{1, \cdots, n\}$, vérifient :

$$(x_i - a)^2 + \tau^2 y_i^2 = \alpha^2 + \varepsilon_i \tag{1}$$

où ε_i est une variable aléatoire réelles suivant la loi normale $\mathcal{N}(0, \sigma^2)$.

2. L'équation (1) n'est pas linéaire par rapport aux paramètres inconnus (a, α). Montrer que l'on peut se ramener au cas linéaire de la forme $z_i = \lambda x_i + \mu + \varepsilon_i$ où (λ, μ) sont les nouveaux paramètres inconnus.

On déterminera les expressions de z_i, λ et μ en fonction de x_i, y_i, a, τ et α.

3. Déterminer $(\hat{\lambda}_{MC}, \hat{\mu}_{MC})$ les estimateurs par les moindres de carrées de (λ, μ).

4. En déduire $(\hat{a}, \hat{\alpha}, \hat{\beta})$ des estimateurs de (a, α, β).

5. **Application numérique :**

Le tableau suivant résume les observations faites pour $n = 9$:

x_i	-1.05	-0.52	0	0.49	1	1.52	1.94	2.5	3.1
y_i	0	-0.94	-1.16	1.32	1.36	-1.42	1.22	-0.95	0

À quelle estimation de (a, α, β) conduit ces observations sachant que $\tau = \frac{3}{2}$?

Indication : On pourra dans un premier temps calculer les estimations de $(\hat{\lambda}_{MC}, \hat{\mu}_{MC})$ *et ensuite* $(\hat{a}, \hat{\alpha}, \hat{\beta})$.