

TD 5

Exercice 1

On considère deux jeux de données (X_i, Y_i) et (X_i, Z_i) , $i = 1, \dots, n$, définis par les modèles suivants,

$$Y_i = 2X_i + 5 + \mathcal{N}(0, 4),$$

$$Y_i = X_i^2 + 5 + \mathcal{N}(0, 4).$$

Comparer par simulation les résultats de régression linéaire. Que peut on conclure sur les graphiques des résidus ? Appliquer la régression linéaire sur les données transformées en logarithme. Qu'obtient on ?

1. Générer un vecteur \mathbf{X} de 100 nombres aléatoires distribués uniformément dans $[2, 10]$.
2. Générer un vecteur \mathbf{e} de 100 nombres aléatoires selon une loi normale d'espérance 0 et de variance 4.
3. Définir les vecteur $\mathbf{Y}=2*\mathbf{X}+5+\mathbf{e}$ et $\mathbf{Z}=\mathbf{X}*\mathbf{X}+5+\mathbf{e}$.
4. Estimer les paramètres des droites de régression en utilisant `lm`. Afficher les résultats de la régression.
5. Tracer côte à côte les nuages de points (X_i, Y_i) et (X_i, Z_i) , $i = 1, \dots, 100$. Ajouter les droites de régression en rouge et les lignes de modèle réel en bleu.
6. Afficher les graphiques des résidus de deux modèles dans les fenêtres séparées.
7. Refaire les procédures précédentes, mais sur les couples $(\log X_i, \log Y_i)$ et $(\log X_i, \log Z_i)$.

Exercice 2 : Exemple « Oeufs »

Une entreprise de production d'oeufs produit 10000 oeufs par jour. Compte tenu des contrat qu'elle a passés avec ses clients, il lui faut calibrer cette production d'oeufs en fonction de leur poids. Ce que fait notre entreprise, qui dispose d'une chaîne d'emballage automatisée incorporant une unité de pesage et de tri. Mais cette chaîne est extrêmement lente. La société recherche un procédé économique de calcul, même indirect et approximatif, du poids de ses oeufs, qui permette le pesage-tirage en continu.

Un procédé pourrait être la détermination de la hauteur ou de la largeur de l'oeuf, par des moyens faisant appel à la technique des faisceaux laser. Le poids de chaque oeuf serait déduit de la hauteur ou de la largeur.

Le fichier de données "oeufs.txt" est téléchargeable à l'adresse suivante.

<http://samm.univ-paris1.fr/IMG/txt/oeufs.txt>

Enregistrez le dans votre répertoire de travail.

```
getwd()           Récupère le répertoire courant
setwd("emplacement")  Spécifie le répertoire courant
```

Le poids de l'oeuf se déduit-il approximativement de sa hauteur ou de sa largeur ? Mieux de l'un ou de l'autre ? Et comment ?

```
oeufs = read.table("emplacement/oeufs.txt", header = TRUE)
```