



**RAPPORT DE « TP noté »**

présenté par

**FOUNTIR Mohamed**

Matière :

**Méthodes Mathématiques : Approche aléatoire**

chargé de TP : **METELKINA Asya**

---

# Table des matières

0.1	Exercice 1 . . . . .	2
0.2	Exercice 2 . . . . .	2
0.2.1	Calcule de $\sigma^2(XX)$ . . . . .	2
0.2.2	Différence entre la méthode de Monte-Carlo et Bootstrap . . . . .	2
0.2.3	Comparaison de la proportion des IC et $\alpha$ . . . . .	2
0.3	Exercice 3 . . . . .	3
0.3.1	Observation de la distribution du nuage de points . . . . .	3
0.3.2	Prédiction de Y pour X = 28 . . . . .	3
0.3.3	La loi de distribution des résidus : Loi normal N(0,1) ? . . . . .	3
0.3.4	Comparaison entre t(A) et les sorties de la fonction lm . . . . .	4

## 0.1 Exercice 1

Sur la partie graphique de cette exercice, j'ai utilisé un animation graphique toute les 2 seconde visualisant les 100 intervalles de confiance de 95% ainsi que la droite  $y = m$  et en variant les valeurs de  $n$ ,  $\alpha$ ,  $\sigma^2$

**NB :** Sur le terminal, on trouve la proportion de ces intervalles en temps réel

En observant la variation des intervalles de confiance en modifiant les paramètres  $n$ ,  $\alpha$ , et  $\sigma^2$ , nous pouvons tirer plusieurs conclusions :

**Variation de  $n$  :** Plus  $n$  est grand, plus les intervalles de confiance devraient se resserrer autour de la vraie valeur de  $m$ . Cela signifie que nous avons une plus grande confiance dans notre estimation de la moyenne  $m$  lorsque nous avons plus de données.

**Variation de  $\alpha$  :** Plus le niveau de confiance est grand  $\alpha$ , plu les intervalles de confiance devraient s'élargir. Cela signifie que nous sommes plus prudents dans notre estimation de  $m$  et nous permettons une marge d'erreur plus grande pour notre intervalle de confiance.

**Variation de  $\sigma^2$  :** Plus  $\sigma^2$  de la distribution est grande, plus les intervalles de confiance devraient également s'élargir. Cela reflète une plus grande incertitude dans notre estimation de la moyenne  $m$  lorsque la variabilité des données est plus importante.

En ce qui concerne la comparaison de la proportion d'intervalles de confiance qui ne contiennent pas la vraie valeur de  $m$  avec  $\alpha$ , si cette proportion est plus grande que  $\alpha$ , cela suggère que nos intervalles de confiance sont trop larges et que nous sommes trop conservateurs dans notre estimation de  $m$ . Si cette proportion est inférieure à  $\alpha$ , cela suggère que nos intervalles de confiance sont trop étroits et que nous prenons trop de risques en sous-estimant la variabilité de  $m$ . En ajustant  $\alpha$  en fonction de la proportion observée, nous pouvons trouver un équilibre entre précision et prudence dans nos estimations.

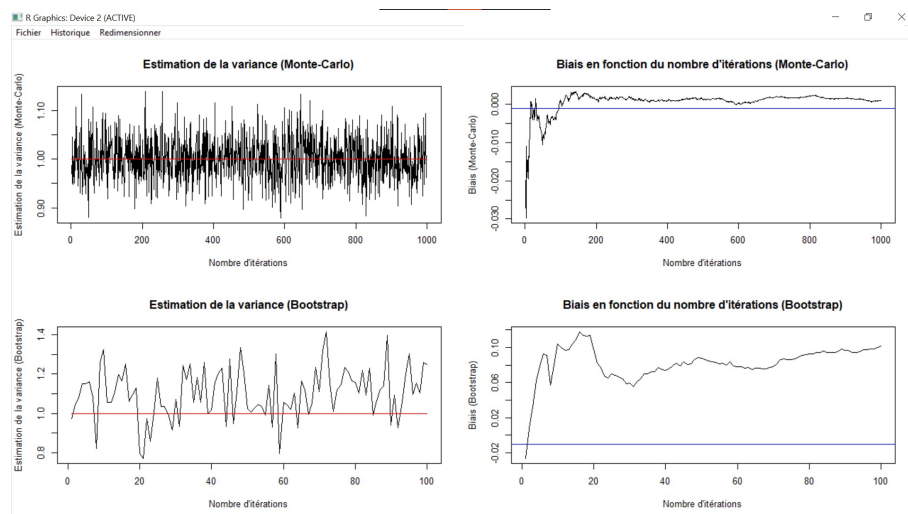
## 0.2 Exercice 2

### 0.2.1 Calcule de $\sigma^2(XX)$

Après l'estimation de 100 echantillons de loi normal  $N(0,1)$  on obtient :  $\sigma^2(XX) = 0.9396047$

### 0.2.2 Différence entre la méthode de Monte-Carlo et Bootstrap

### 0.2.3 Comparaison de la proportion des IC et $\alpha$



Les deux méthodes, Monte-Carlo et Bootstrap, partagent des objectifs similaires en matière d'estimation de la variance et d'analyse de la distribution de l'estimateur de la variance. Cependant, elles présentent des différences notables en termes de principes et d'applications :

**Principes :**

Monte-Carlo : La méthode Monte-Carlo repose sur la simulation d'un grand nombre d'échantillons aléatoires à partir d'une distribution connue ou approchée. Les estimations de la variance sont calculées sur chaque échantillon simulé, et la moyenne de ces estimations est utilisée comme estimateur final de la variance de la population.

Bootstrap : La méthode Bootstrap s'appuie sur le rééchantillonnage de l'échantillon original pour générer de nouveaux échantillons artificiels. Les estimations de la variance sont calculées sur chaque échantillon rééchantillonné, et la distribution de ces estimations est utilisée pour inférer la distribution de l'estimateur de la variance.

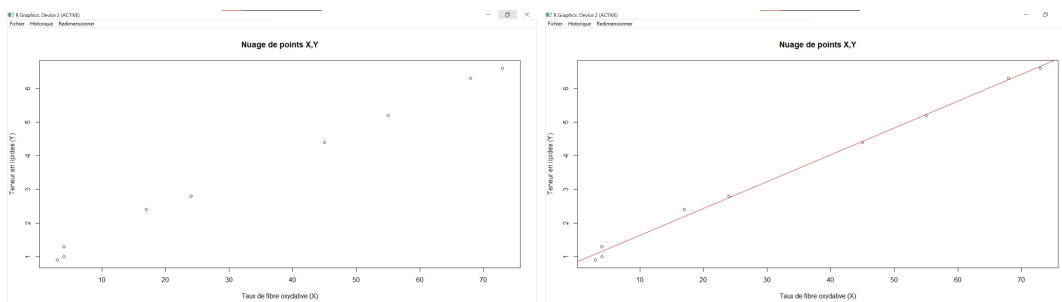
#### Performance dans ce cas d'étude :

Le biais de l'estimation de la variance de Monte-Carlo converge plus rapidement vers  $-1/N = 0.001$  qui est la valeur approchée de notre estimateur biaisé, surtout pour un grand nombre d'itération  $N = 1000$ . Le biais de l'estimation de la variance de Bootstrap reste loin de converger vers sa valeurs approché  $-1/N = 0.01$  surtout avec un nombre d'itérations très faible  $N = 100$

**Conclusion :** Dans ce cas d'étude la méthode Monte-Carlo semble performante pour estimer la variance.

## 0.3 Exercice 3

### 0.3.1 Observation de la distribution du nuage de points



Après la visualisation du nuage de points comme le montre la figure, il est possible de décrire certaines caractéristiques pour avoir une idée générale sur la nature de la relation entre le taux de fibre oxydative (X) et la teneur en lipides (Y) dans la chair des 9 lapins :

**forme :** Le nuage de points n'est ni parfaitement linéaire ni parfaitement aléatoire. Il présente une certaine tendance à s'organiser en forme de courbe.

**Direction :** La tendance générale de la courbe est positive, c'est-à-dire que les valeurs de Y augmentent en moyenne avec les valeurs de X.

**Dispersion :** Les points ne sont pas tous regroupés au même endroit, mais sont plutôt dispersés autour de la courbe.

On peut conclure que le nuage de points suggère l'existence d'une liaison positive curvilinéaire entre les variables X et Y. Nous allons chercher par la suite la nature de la dépendance entre ces deux variables quantitative à l'aide de notre modèle linéaire représenté par la droite de régression rouge.

### 0.3.2 Prédiction de Y pour X = 28

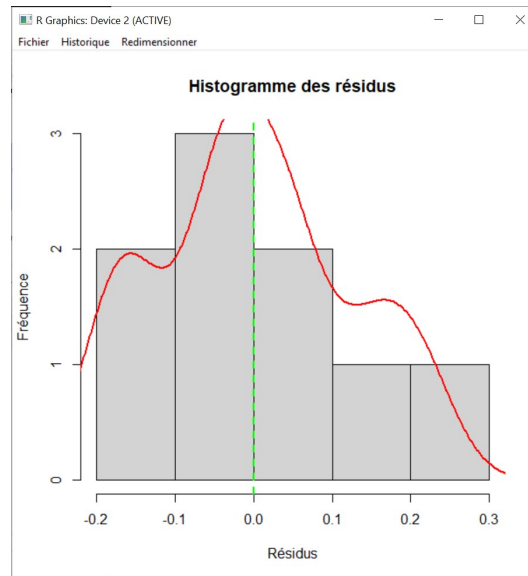
On utilisant la fonction `predict()` de R on trouve la prediction de Y quand  $X = 28$  :

La valeur prédite de Y pour  $X = 28$  est 3.06995

### 0.3.3 La loi de distribution des résidus : Loi normal $N(0,1)$ ?

Pour déterminer si les résidus suivent une loi normal  $N(0,1)$ , il existe deux façons :

**Interpretation graphique :**



On sait que la courbe de densité de probabilité d'une loi normale est une courbe en cloche symétrique par rapport à la moyenne. La plupart des valeurs se concentrent autour de la moyenne, et la probabilité de trouver des valeurs éloignées de la moyenne diminue à mesure que l'on s'éloigne de cette valeur.

Sur l'image, la distribution est asymétrique, avec une queue plus longue d'un côté que de l'autre. Cela signifie que les valeurs ne sont pas réparties symétriquement autour de la moyenne. De plus, la courbe n'est pas en forme de cloche, mais plutôt aplatie. Cela signifie que les valeurs sont plus dispersées que dans une loi normale.

#### Test de Shapiro-Wilk :

Ce test statistique permet de vérifier si une distribution suit une loi  $N(0,1)$ . Il renvoie un p-value. Si le p-value est supérieur au seuil de signification (généralement 0,05), on ne rejette pas l'hypothèse nulle selon laquelle la distribution suit une loi  $N(0,1)$ .

Voici les sorties de ce test :

data : residues W = 0.96242, p-value = 0.8233

On observe que  $p\text{-value} > 0.05$

**Conclusion :** cette analyse renforce l'idée que la distribution des résidus suit approximativement une distribution normale.

#### 0.3.4 Comparaison entre $t(A)$ et les sorties de la fonction `lm`

On a trouvé que  $b = (0.8364753, 0.07976697)$  par la méthode de transposé de la matrice  $A$ , ce qui presque identique aux sorties de la fonction `lm` ci-dessous :

(Intercept) = 0.83647527, X = 0.07976697

Cela indique que notre calcul de  $b$  est correct et que notre modèle de régression linéaire est bien ajusté aux données. C'est une validation supplémentaire de la cohérence entre la méthode analytique de calcul des coefficients de régression et la méthode basée sur la fonction `lm`.