

Analyse statistique des données forestières

Projet\_905

Membres : Ben Abdelghaffar Abir – Dumont Alban-Lamghari Mohamed

Date : 12/07/2024

**Table de Matiéres**

[**1.** **Introduction** 2](#_Toc184463879)

[**2. Méthodologie** 2](#_Toc184463880)

[2.1 Importation des Données sur RStudio 2](#_Toc184463881)

[2.2 Filtrage des Données 2](#_Toc184463882)

[**3. Analyse des Résultats** 2](#_Toc184463883)

[3.1 Question 1 : Modèle ANOVA pour les Charmes 2](#_Toc184463884)

[3.2 Question 2 : Régression Linéaire pour les Chênes 4](#_Toc184463885)

[3.3 Question 3 : Modèle ANOVA avec Agrégation par Triplets 4](#_Toc184463886)

[3.4 Question 4 : Modèle Mixte pour Étudier l'Effet de lastLog 4](#_Toc184463887)

[3.5 Question 5 (Bonus) : Modèle Linéaire Généralisé avec la Présence de Cavité Basse 4](#_Toc184463888)

[4. Discussion 4](#_Toc184463889)

[5. Conclusion 4](#_Toc184463890)

# **Introduction**

Ce projet s'inscrit dans le cadre d'une analyse statistique approfondie d'un jeu de données forestières, en utilisant des approches avancées telles que les modèles linéaires de régression, l'ANOVA, et les modèles mixtes. L'objectif principal est de comprendre les variations du diamètre des arbres, qui constitue la variable dépendante, en fonction de divers facteurs explicatifs, notamment le relevé, l'espèce d'arbre et des variables topographiques telles que l'altitude. En établissant la relation ou la corrélation entre la variable réponse et ces facteurs explicatifs, il devient possible d'estimer et de prédire le diamètre des arbres à l'échelle de la population.

# **2. Méthodologie**

# 2.1 Importation des Données sur RStudio

Les données ont été importées dans RStudio en utilisant la fonction **read.csv (),** permettant de travailler avec le fichier dataProjet\_2025.csv.

## 2.2 Filtrage des Données

Pour chaqu’une de question nous avons travaillé sur un jeu de données filtré à partir de base de donnée initiale « **dataProjet\_2025.csv** ». En effet, pour la première question, le jeu de données a été filtré selon la valeur de champ « recherche\_esp\_lb\_nom\_plantae » pour ne conserver que les observations correspondant à l’espèce «  **Carpinus betulus L., 1753** »

# **3. Analyse des Résultats**

## 3.1 Question 1 : Modèle ANOVA pour les Charmes

L’objectif de cette analyse était d’examiner l’effet des relevés sur le diamètre des arbres de l’espèce **Carpinus betulus L., 1753** (charme) en utilisant un modèle ANOVA. Nous avons vérifié la validité des hypothèses du modèle linéaire, notamment la normalité et l’homoscedasticité des résidus, et appliqué des transformations pour remédier aux violations identifiées.

* **Résultats Initiaux**

Un modèle ANOVA initial a été ajustée avec la variable « **releve »** comme facteur explicatif et le diamètre DBH comme variable dépendante. Par la suite, une vérification de l’adéquation des erreurs (ou des résidus) aux différentes hypothèses de la loi Gaussienne a été réalisé

* **Normalité des Résidus**
  + **Diagnostic** : Les tests graphiques, comprenant un Qqplot et un histogramme, ont révélé une asymétrie notable des résidus. Les points du Qqplot s’écartaient de la droite de référence, indiquant une déviation de la normalité. L’histogramme des résidus a confirmé une asymétrie positive. Les résidus n’étaient pas donc compatibles avec une distribution normale de la loi Gaussienne. Donc, une transformation logarithmique a été proposée.
* **Homoscedasticité des Résidus**
  + **Diagnostic** : Un scale-location plot a montré une tendance à une dispersion accrue des résidus pour des valeurs ajustées élevées, suggérant une violation de l’hypothèse d’homoscedasticité (hétérogénéité des variances). La solution que nous avons proposée est d’appliquer une transformation concave, telle que la transformation logarithmique, était justifiée.

**\*\*Transformation Logarithmique**

Après une transformation logarithmique de la variable dépendante DBH, le modèle ANOVA a été réajusté :

* **Normalité des Résidus**
  + **Diagnostic** : Le Q-Q plot des résidus du modèle transformé a montré une amélioration de la normalité, bien que des déviations légères aient persisté dans les queues. L’histogramme des résidus indiquait une meilleure symétrie qu’avant la transformation. Donc, on peut déduire que la transformation logarithmique a réduit l’asymétrie initiale des résidus, mais des ajustements supplémentaires pourraient être nécessaires.
* **Homoscedasticité des Résidus**
  + **Diagnostic** : Le scale-location plot du modèle transformé a révélé que la dispersion des résidus augmentait encore légèrement avec les valeurs ajustées. L’hypothèse d’homoscedasticité restait faiblement violée, nécessitant une transformation plus appropriée.

## 3.2 Question 2 : Régression Linéaire pour les Chênes

* Filtrer les données pour ne garder que les chênes.
* Appliquer un modèle de régression linéaire du diamètre en fonction de la variable « **lastLog »**
* Faire un diagnostic des résidus.
* Tester l'ajout d'un terme non-linéaire (par exemple, lastlog2).

## 3.3 Question 3 : Modèle ANOVA avec Agrégation par Triplets

* Filtrer les données pour ne garder que les chênes.
* Visualiser la structure hiérarchique des relevés.
* Appliquer un modèle ANOVA avec une nouvelle agrégation par triplet de relevés.
* Vérifier les hypothèses et tester les modèles emboîtés.

## 3.4 Question 4 : Modèle Mixte pour Étudier l'Effet de lastLog

## 3.5 Question 5 (Bonus) : Modèle Linéaire Généralisé avec la Présence de Cavité Basse

* Appliquer un modèle linéaire généralisé pour prédire la présence de cavités basses en fonction de l’altitude et du diamètre.
* Vérifier les hypothèses avec le package DHARMa.

## 4. Discussion

# 5. Conclusion