

# Rapport projet ADD sur jeu de données Rossignol

Mohamed SOUANE  
Damian BIMBENET  
Bailin CAI  
Alexandre ZAFARI

15 décembre 2022

## 1 Introduction

Étude d'une population de cette espèce d'oiseau en Suède. Pour prédire, la réponse évolutive de la sélection dans ces traits, on présente des estimations et corrélations d'héritabilité  $h^2$

la projection des ailes, la hauteur de la bec, la largeur de la bec, la longueur de la bec, le poids. Par contre il semble avoir une sur la longueur de la queue, la longueur du tarsus, la longueur des ailes.

## 2 Analyse Descriptive

L'âge moyen des 544 sujets inclus dans la cohorte était de 32 mois ( $q_1 = 20$ ;  $q_3 = 35$ ), 35% étaient des femelles et 65% étaient des mâles. Longueur des ailes moyen mesuré à l'inclusion était de 99cm. Le poids moyen mesuré était 33,9g. La longueur de la queue moyen mesuré était de 33cm. La hauteur du bec moyen mesuré était de 5cm. La largeur du bec moyen mesuré était de 5cm.

Variable	Mâle(mean)	Femelle(mean)	P-value
Poids	6	87837	787
2	7	78	5415
3	545	778	7507
4	545	18744	7560
5	88	788	6344

TABLE 1 – Test comparaisons des caractéristiques chez les mâles et femelles

### 2.1 Visualisation des données

Décrivez et comparez à l'aide de tests statistiques les caractéristiques dans les deux groupes mâle et femelle. L'objectif ici est de décrire et de comparer les groupes mâle et femelle en fonction des caractéristiques recueillies qui sont toutes des variables quantitatives, en utilisant un test de comparaison de deux moyennes. les effectifs dans chaque groupe étant suffisant grande. Le test de wilcoxon est assez robuste au non-respect de la normalité des distributions.

Les box-plot montre qu'on ne peut pas confirmer que le sexe a une influence sur l'âge,

**Commentaires Tableau 1** Comme on pouvait s'y attendre, le poids en moyenne est significativement plus basses chez les femelles que chez les mâles ( $p\text{-value} < 0,5$ ). Il en est de même pour la longueur des ailes, la longueur des tarsus, la longueur de la queue, la longueur de la bec, la hauteur de la bec, la projection des ailes. Par contre, les deux groupes ne diffèrent pas significativement pour l'âge. La largeur des becs en moyenne est significativement plus basses chez les mâles que chez les femelles ( $p\text{-value} > 0,5$ )

## 2.2 Matrice de corrélation

Variable	Mâle(coeff)	P-value	Femelle(coeff)	P-value
Poids	6	87837	787	1
2	7	78	5415	1
3	545	778	7507	1
4	545	18744	7560	1
5	88	788	6344	1

TABLE 2 – Le test du coefficient de corrélation linéaire entre la longueur du tarsus et certaines traits morphologiques chez les mâles et chez les femelles

**Commentaires Tableau 2** D’après les résultats du Tableau 2 et les deux matrices de corrélation, chez les mâles, seules les variables longueur des ailes, longueur du bec, longueur de la queue, hauteur de la bec, largeur de la bec, poids, sont significativement liées linéairement à la PAM( $p < 0, 5$ ). Les liens sont positifs. On retrouve les mêmes résultats chez les femelles en ce qui concerne longueur des ailes, longueur du bec, largeur de la bec et poids, mais pas pour la longueur des ailes, longueur du vbec, largeur de la bec ( $p = 0, 097$ ). L’âge et la projection des ailes ne sont pas significativement liées linéairement à la longueur du tarsus dans les deux groupes.

## 3 La regression linéaire multiple

Dans cette partie on a réalisé le modèle avec toutes les variables quantitatives,

### 3.1 Le choix des variables

Pour le choix du modèle, nous vous appliquons une procédure de simplification descendante pas à pas. La procédure de la simplification descendante pas à pas est une approche visant à améliorer le modèle explicatif. On réalise un premier modèle avec toutes les variables spécifiées, puis on regarde s’il est possible d’améliorer le modèle avec toutes les variables spécifiées, puis on regarde s’il est possible d’améliorer le modèle en supprimant une des variables dont la suppression améliorera le plus le modèle. Puis on recommence le même en supprimant une des variables dont la suppression

améliorera le plus le modèle. Puis on recommence la même procédure pour voir si la suppression d’une seconde variable peut encore améliorer le modèle et ainsi de suite. Lorsque le modèle ne peut plus être amélioré par la suppression d’une variable on s’arrête.

Variable	Mâle(coeff)	P-value	Femelle(coeff)
Poids	6	87837	787
2	7	78	5415
3	545	778	7507
4	545	18744	7560
5	88	788	6344

TABLE 3 – Relation entre le poids et certains traits morphologiques. Analyse multivariée. Régression linéaire

### 3.2 Analyse des résidus

Cette étape concerne la vérification des hypothèses énoncées sur l’erreur estimée. Il s’agit du test de normalité du résidu. La validation des modèles est attachée à l’idée selon laquelle les résidus sont indépendants et identiquement distribués suivant la loi normale centrée avec une variance constante.

L’alignement des points sur la première bissectrice est presque parfaite, on en déduit l’hypothèse selon laquelle les résidus théoriques suivent une loi normale mais pas centrée réduite.

D’après les graphiques ci-dessus, 3.6% des résidus sont aberrants, ce qui est acceptable pour un jeu de données de taille 544. De plus, ces résidus ne sont pas très éloignés des limites d’intervalle. Ce ne sont pas de véritables observations aberrantes, car 96,4% des observations sont entre les deux seuils. 6% sont des points leviers et une seule distance de Cook élevée (observation 26 est à la fois levier et aberrante).

## 4 ANOVA

## 5 Conclusion