



ENSAI

École nationale
de la statistique
et de l'analyse
de l'information

Quelles circonstances sont associées à l'expression de signes
cliniques respiratoires des porcs en croissance en élevages
alternatifs ?

Étudiants :

Antoni GUÀRDIA SANZ
Clément YVERNAULT-COLLET
Marius VINCLAIR
Xavier BRAQUAVAL

Tutrice :

Christelle FABLET

18 février 2025

Table des matières

1	Introduction	2
1.1	Introduction du sujet	2
1.2	Description de la base de données	2
2	Etude des variables respiratoires	3
2.1	Présentation et caractérisation des variables respiratoires	3
2.2	Exploration des relations entre les variables respiratoires	4
2.2.1	Relations linéaires entre les variables respiratoires	4
2.2.2	Relations logarithmiques entre les variables respiratoires	4
2.3	Analyse des regroupements : réduction de dimensions et catégorisation	5
A	Annexe : compléments de l'analyse sur les variables respiratoires	6
A.1	Test de pearson sur les corrélations des variables respiratoires	6
A.2	Test de pearson sur les corrélations du logarithme des variables respiratoires	6
A.3	Résultats régression linéaire entre les variables respiratoires et leur logarithme	6

1 Introduction

1.1 Introduction du sujet

1.2 Description de la base de données

2 Etude des variables respiratoires

Dans cette section, on se propose d'analyser les relations entre les variables caractérisant la respiration des porcs, en particulier les fréquences d'éternuements et de toux en engraissement et post-sevrage, afin d'identifier les liens existants. L'objectif est de regrouper ces variables et de simplifier l'étude en réduisant leur complexité.

2.1 Présentation et caractérisation des variables respiratoires

Dans cette partie, nous présentons et caractérisons les variables respiratoires mesurées chez les porcs. L'objectif est de mieux comprendre la répartition de ces données et d'identifier d'éventuelles particularités dans le comportement respiratoire des porcs.

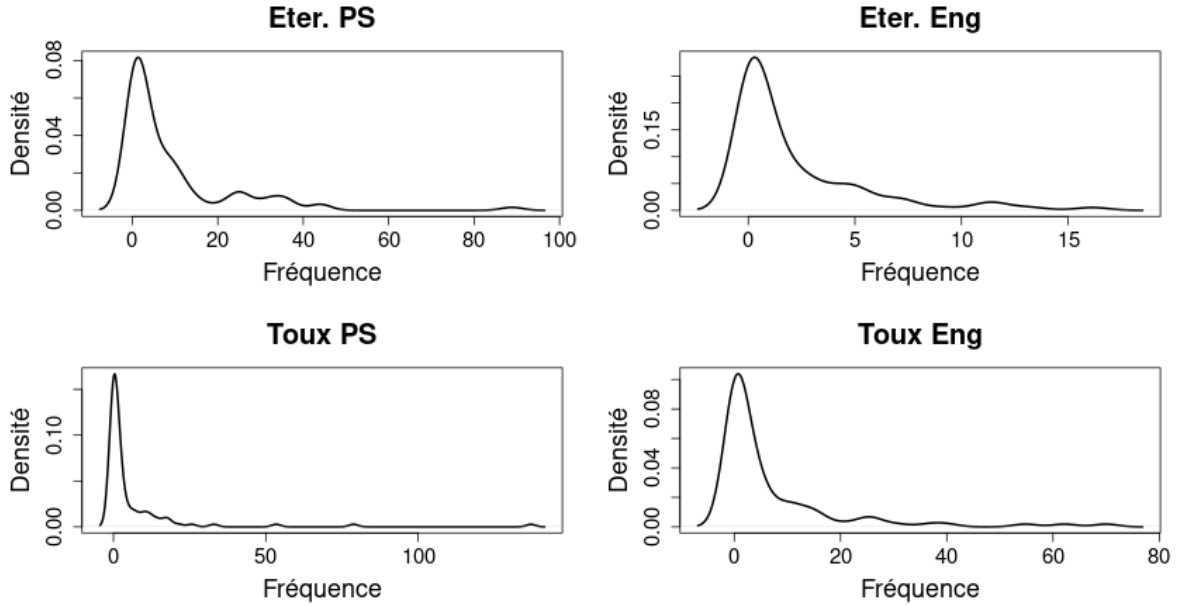


FIGURE 1 – Densité des variables respiratoires (BP¹ : nrd0²)

Notons que dans (1) les quatre variables respiratoires présentent une forte concentration de la densité autour de 0, ce qui reflète un bon état respiratoire global chez les porcs étudiés. Cependant, certains élevages montrent des exceptions où cette tendance générale n'est pas observée.

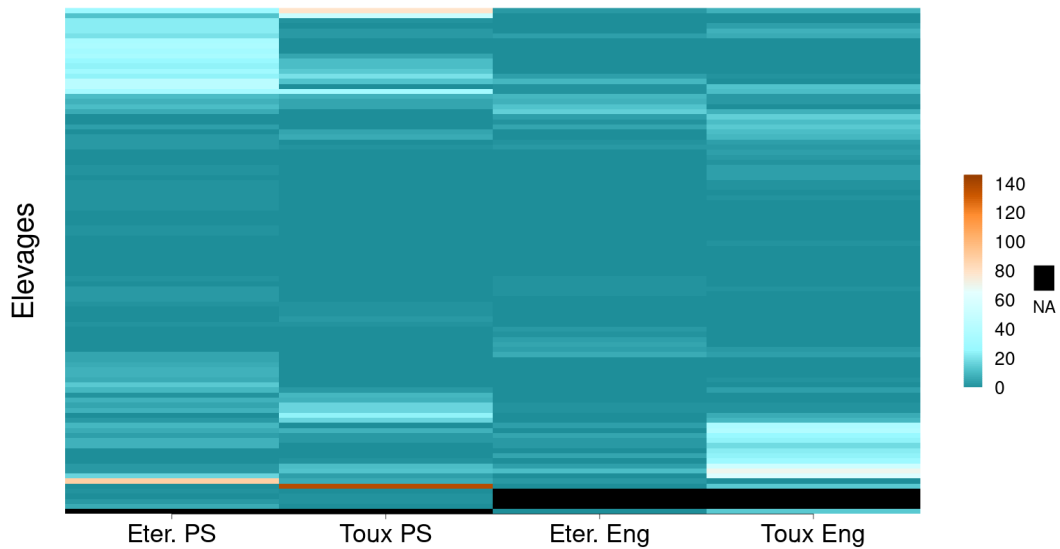


FIGURE 2 – Visualisation des variables respiratoires par élevages

1. Bande passante.
2. La méthode `nrd0` fait référence à la largeur de bande de référence normale.

Dans (2), nous constatons une fois de plus que la majorité des individus présentent des valeurs faibles pour l'ensemble des variables respiratoires. Nous relevons également la présence de cinq individus avec des valeurs manquantes. Il apparaît que ces valeurs manquantes concernent soit les variables issues de l'engraissement, soit celles du post-sevrage. Cette observation est cohérente avec le protocole de l'étude, qui prévoyait la mesure simultanée des éternuements et des toux. Des légères corrélations entre la toux et les éternuements sont également observées au sein d'un même lot.

2.2 Exploration des relations entre les variables respiratoires

L'objectif de cette section est d'analyser la force des liens entre les variables respiratoires, afin de déterminer si une régression linéaire pourrait être pertinente pour réduire le nombre de variables à considérer dans les étapes suivantes. De plus, cette exploration permet d'évaluer la pertinence de l'Analyse en Composantes Principales (ACP) pour la réduction de dimensions, afin de juger de son adéquation avec nos données.

2.2.1 Relations linéaires entre les variables respiratoires

Examinons les corrélations ainsi que la répartition des élevages selon les variables respiratoires.

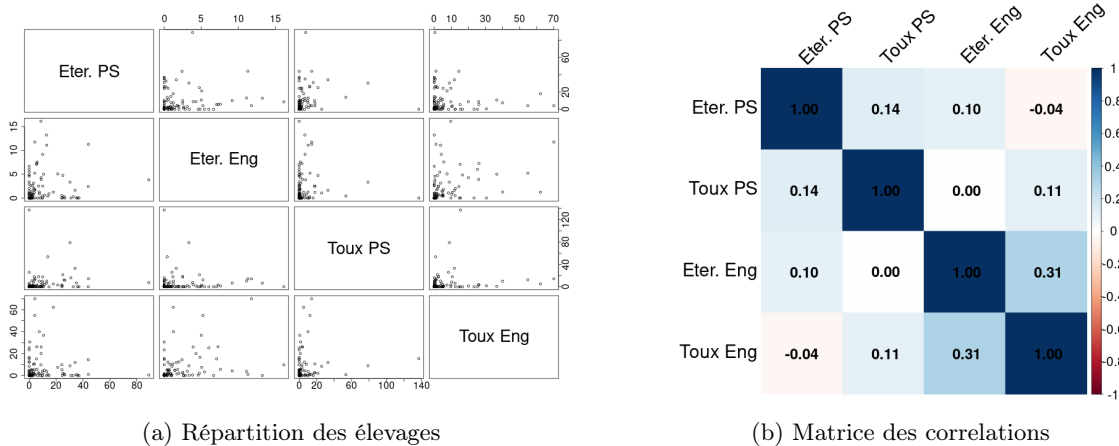


FIGURE 3 – Illustration étude linéaire

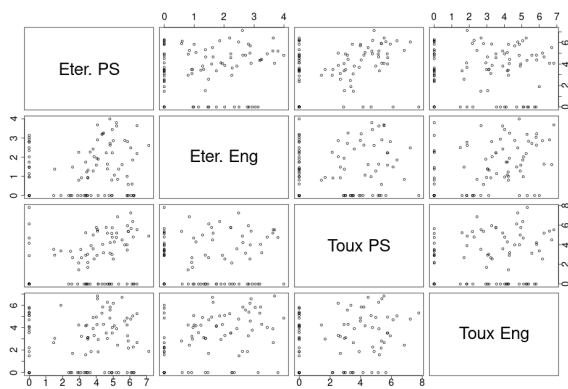
Dans la figure (3), aucune relation linéaire forte n'apparaît entre les différentes variables. Ce constat est particulièrement évident dans la figure (3a), où les nuages de points ne présentent pas la structure triangulaire typique d'une dispersion linéaire. Cependant, la matrice de corrélation (3b) met en évidence des faibles corrélations entre la plupart des variables, à l'exception notable des éternuements en engraissement et des toux en post-sevrage. On observe en effet une corrélation marquée entre les toux et les éternuements, qui appartiennent au même groupe (soit engraissement, soit post-sevrage), cette corrélation étant particulièrement prononcée chez les porcs en engraissement. Le test de Pearson confirme ces observations (Cf. A.1) : la seule corrélation statistiquement significative concerne les variables toux et éternuements en engraissement.

Les analyses graphiques et le test de Pearson révèlent une absence de relations linéaires globales entre les variables, à l'exception de la corrélation entre toux et éternuements en engraissement. Par conséquent, l'ACP, qui repose sur des hypothèses de linéarité, ne semble pas être la méthode d'analyse la mieux adaptée à ce jeu de données.

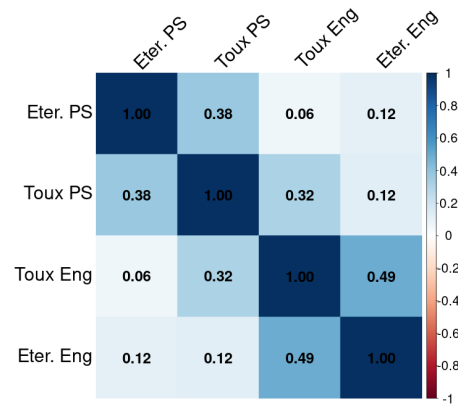
2.2.2 Relations logarithmiques entre les variables respiratoires

Nous appliquons la transformation $x \mapsto \ln(x \cdot \hat{\sigma}_x + 1)$ ³ dans le but d'étudier une éventuelle relation logarithmique entre les variables. Examinons maintenant les corrélations ainsi que la répartition des élevages selon les nouvelles données.

3. Nous utilisons la transformation $x \mapsto \ln(x \cdot \hat{\sigma}_x + 1)$ plutôt que $x \mapsto \ln(x) \cdot \hat{\sigma}_x$ afin de tenir compte de la présence de zéros dans nos données, tout en conservant une fonction bien définie. Le facteur $\text{sd}(x)$ (écart-type) est ajouté pour normaliser les données, en atténuant l'effet des différences de dispersion entre les variables, ce qui permet de stabiliser la variance et de rendre les données plus comparables. Cette approche est donc appropriée, car elle garantit des valeurs positives et améliore la gestion de la variabilité dans les données.



(a) Répartition des élevages



(b) Matrice des corrélations

FIGURE 4 – Illustration étude logarithmique

Observons que cette fois ci, il semble y avoir une corrélation plus évidente entre les données TODO

2.3 Analyse des regroupements : réduction de dimensions et catégorisation

A Annexe : compléments de l'analyse sur les variables respiratoires

A.1 Test de pearson sur les corrélations des variables respiratoires

Variable 1	Variable 2	Coefficient de corrélation	p-valeur	Significatif
Eter. PS	Eter. Eng	0.104	0.317	Non
Eter. PS	Toux PS	0.135	0.191	Non
Eter. PS	Toux Eng	-0.041	0.696	Non
Eter. Eng	Toux PS	0.003	0.979	Non
Eter. Eng	Toux Eng	0.305	0.003	Oui
Toux PS	Toux Eng	0.114	0.273	Non

TABLE 1 – Résultats des tests de corrélation de Pearson entre les variables respiratoires.

Constatons que la seule corrélation significative est celle des variables de toux et éternouements en engraissement.

A.2 Test de pearson sur les corrélations du logarithme des variables respiratoires

Variable 1	Variable 2	Coefficient de corrélation	p-valeur	Significatif
Eter. PS	Eter. Eng	0.119	0.251	Non
Eter. PS	Toux PS	0.377	0.000	Oui
Eter. PS	Toux Eng	0.061	0.557	Non
Eter. Eng	Toux PS	0.122	0.240	Non
Eter. Eng	Toux Eng	0.491	0.000	Oui
Toux PS	Toux Eng	0.315	0.002	Oui

TABLE 2 – Résultats des tests de corrélation de Pearson entre les variables respiratoires.

TODO

A.3 Résultats régression linéaire entre les variables respiratoires et leur logarithme

TODO GLS with log transform