

Analyse de seuils physiologiques adaptatifs chez les patients âgés

Étude de la fréquence cardiaque à partir de la base MIMIC-IV

1 Contexte et motivation du projet

L'utilisation de seuils physiologiques standards, notamment pour la fréquence cardiaque, est largement répandue en pratique clinique. Ces seuils sont généralement définis pour la population adulte générale, sans distinction spécifique liée à l'âge ou au contexte de soins.

Chez les patients âgés, en particulier en soins intensifs (ICU), la physiologie présente cependant des caractéristiques spécifiques : comorbidités fréquentes, traitements chroniques, diminution de la réserve physiologique et augmentation de la variabilité interindividuelle. Dans ce contexte, l'interprétation d'un signe vital à l'aide de seuils fixes peut conduire à une perte de pertinence clinique ou à une sur-détection d'alertes.

Sur le plan personnel, ce projet s'inscrit dans une volonté de développer une maîtrise pratique du langage **R** appliquée à un contexte médical réel. L'objectif était de mener un projet complet de data science médicale, depuis l'exploration des données jusqu'à une approche d'IA interprétable, tout en restant centré sur une problématique clinique concrète.

2 Données et périmètre de l'étude

Les données utilisées proviennent de la base clinique ouverte **MIMIC-IV**, qui regroupe des informations hospitalières et de soins intensifs.

Le périmètre de l'étude a été volontairement restreint afin de garantir une analyse cohérente et interprétable :

- patients âgés de 65 ans et plus,
- contexte de soins intensifs (ICU),
- signal physiologique étudié : fréquence cardiaque (Heart Rate).

Les tables principales exploitées sont : `patients`, `chartevents` et `d_items`. Aucune donnée brute identifiable n'est incluse dans ce projet.

3 Phase 1 — Prise en main de R

Cette première phase a été consacrée à l'exploration de l'environnement R, des principaux packages d'analyse de données (`tidyverse`, `data.table`, `lubridate`) et des outils de visualisation (`ggplot2`).

Elle avait pour objectif de poser les bases techniques nécessaires au reste du projet et n'a pas donné lieu à des analyses cliniques spécifiques.

4 Phase 2 — Exploration démographique

4.1 Sélection de la population

La table `patients` a été filtrée afin de ne conserver que les patients âgés de 65 ans et plus. Les variables retenues incluent l'âge, le sexe et l'identifiant patient.

4.2 Analyse descriptive

L'analyse démographique met en évidence :

- une diminution progressive des effectifs avec l'âge,
- une légère prédominance féminine,
- un regroupement artificiel des âges très élevés (> 90 ans), lié à la structure de la base MIMIC-IV.



FIGURE 1 – Répartition des patients par âge et sexe

5 Phase 3 — Nettoyage et structuration du signal

Le signal de fréquence cardiaque a été extrait de la table `chartevents` à partir de l'identifiant correspondant à *Heart Rate*. Les données ont été filtrées afin d'exclure les valeurs manquantes et les valeurs physiologiquement impossibles ($HR \leq 0$ ou $HR \geq 250$ bpm).

Les données ont ensuite été jointes aux informations démographiques, triées chronologiquement et regroupées en trois classes d'âge : 65–74 ans, 75–84 ans et 85 ans et plus.

Le jeu de données final contient 4275 mesures exploitables.

6 Phase 4 — Analyse exploratoire clinique

6.1 Distribution globale

La fréquence cardiaque présente une médiane de 83 bpm et une distribution approximativement gaussienne, avec une décroissance progressive vers les valeurs extrêmes.

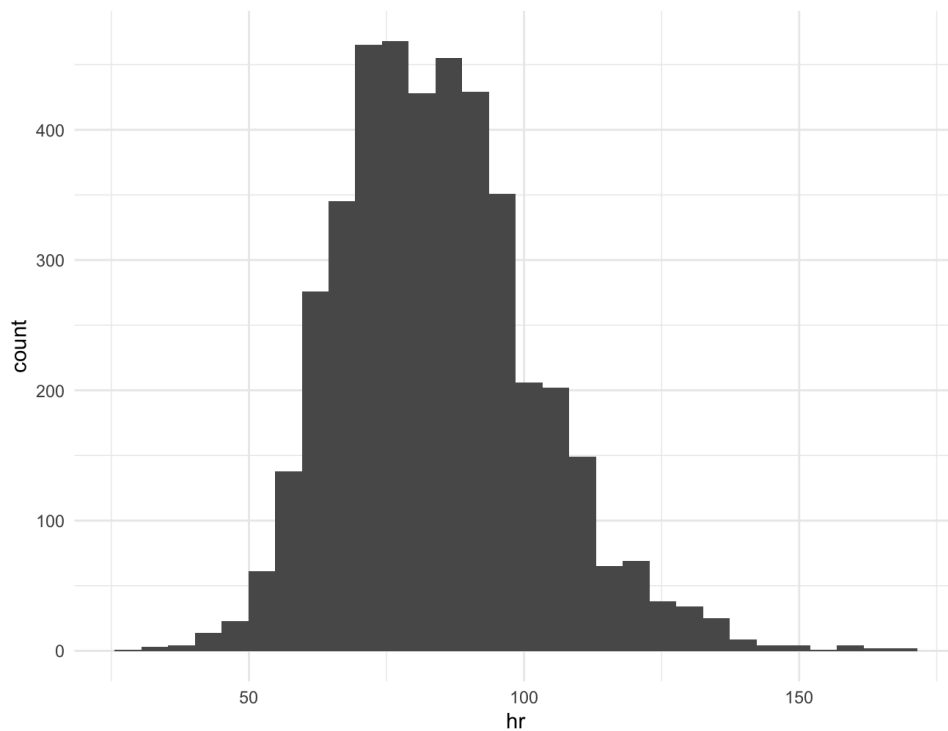


FIGURE 2 – Distribution globale de la fréquence cardiaque

6.2 Analyse par groupes d'âge

Les distributions par groupe d'âge montrent des valeurs centrales proches, un fort chevauchement entre groupes et une augmentation progressive de la variabilité avec l'âge, particulièrement chez les patients de 85 ans et plus.

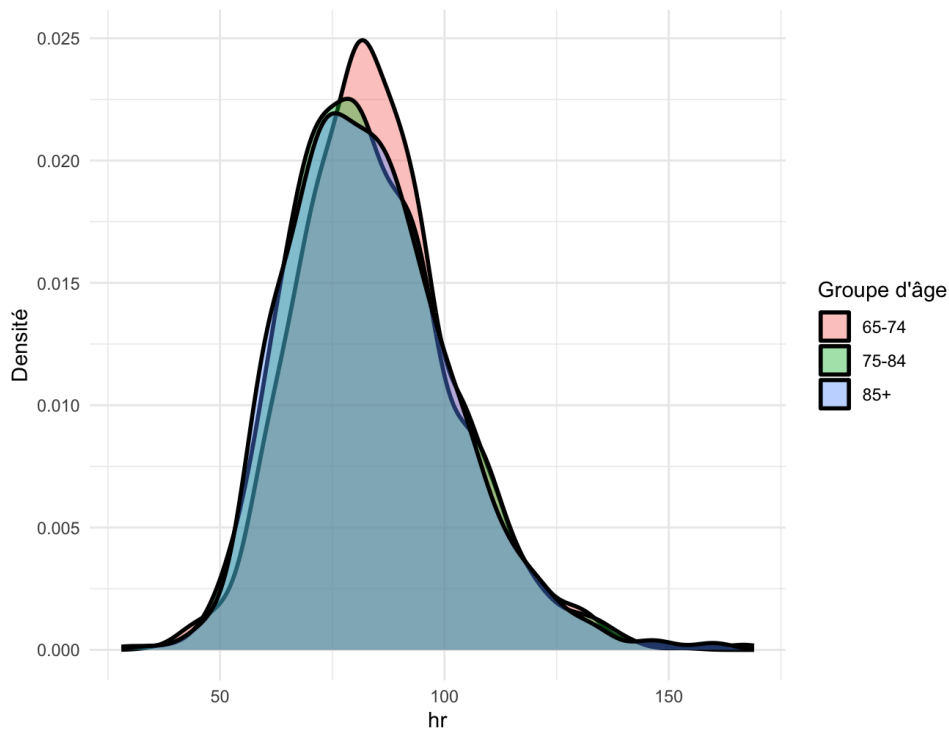


FIGURE 3 – Distribution de la fréquence cardiaque par groupe d'âge

6.3 Comparaison avec les seuils standards

La bradycardie ($HR < 60$ bpm) représente environ 5,7 % des mesures, tandis que la tachycardie ($HR > 100$ bpm) concerne près de 17,7 % des observations, suggérant que le seuil standard de tachycardie est peu discriminant dans ce contexte.

7 Phase 5 — Seuils adaptatifs et score percentile

Des seuils adaptatifs ont été définis à partir des percentiles observés (5^e et 90^e percentiles) pour chaque groupe d'âge.

Afin de synthétiser les seuils adaptatifs obtenus à partir des distributions observées, le Tableau 1 présente les seuils bas (5^e percentile) et hauts (90^e percentile) de la fréquence cardiaque pour chaque groupe d'âge.

TABLE 1 – Seuils adaptatifs de fréquence cardiaque par groupe d'âge

Groupe d'âge	Seuil bas (5%)	Seuil haut (90%)
65–74 ans	60 bpm	108 bpm
75–84 ans	58 bpm	108 bpm
85 ans et plus	58 bpm	107 bpm

Un score percentile continu a également été calculé, indiquant la position relative de chaque mesure au sein de son groupe d'âge.

Ce score est borné entre 0 et 100, comparable entre groupes et directement interprétable cliniquement.

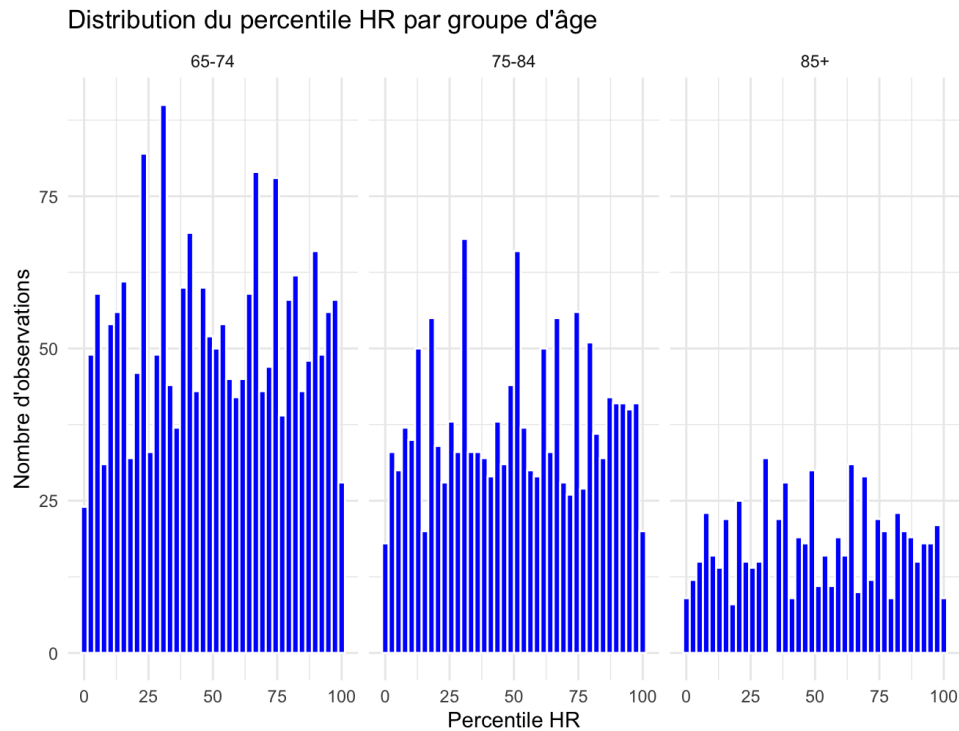


FIGURE 4 – Distribution du score percentile par groupe d'âge

8 Phase 6 — Modélisation IA interprétable

Une observation est définie comme atypique si son percentile est inférieur ou égal à 5 ou supérieur ou égal à 95, ce qui représente environ 10 % des mesures.

Une régression logistique simple a été utilisée. La relation non linéaire entre percentile et atypicité a été capturée via une variable de *distance aux extrêmes*, permettant une séparation quasi parfaite entre observations typiques et atypiques, sans complexification algorithmique inutile.

```
glm(formula = atypique ~ distance_extreme, family = binomial, data = df)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	886.3	5142.2	0.172	0.863
distance_extreme	-178.1	1033.8	-0.172	0.863

Null deviance: 2768.5 on 4274 degrees of freedom

Residual deviance: 0.000158 on 4273 degrees of freedom

AIC: 4.0

Le modèle logistique basé sur la distance aux extrêmes du percentile présente une déviance résiduelle quasi nulle, indiquant une séparation presque parfaite entre les observations typiques et atypiques. Cette performance élevée est attendue, car la variable cible atypique est directement définie à partir du score percentile dont dérive la variable *distance_extreme*.

9 Phase 7 — Discussion et validation qualitative

Les seuils adaptatifs sont stables, cliniquement plausibles et mieux alignés avec les distributions observées que les seuils standards. Les principales limites incluent le contexte ICU, l'absence de variables cliniques détaillées et la présence de mesures répétées par patient.

L'approche proposée doit être considérée comme un outil d'aide à la décision, complémentaire au jugement clinique.

10 Conclusion générale

Ce projet montre que les seuils physiologiques fixes peuvent être inadaptés chez les patients âgés en soins intensifs. Une approche basée sur les distributions observées permet de proposer des seuils plus cohérents et un score continu interprétable.

Au-delà des résultats cliniques, ce travail met en évidence l'intérêt d'une IA simple et explicable, où la valeur ajoutée provient davantage de la représentation du signal que de la complexité du modèle.

Cette méthodologie est généralisable à d'autres signaux vitaux et constitue une base solide pour des travaux futurs en data science médicale.