

# Régression logistique sur la survie de patients dans une unité de soins intensifs

Perrine Warter

2024-11-25

L'analyse porte sur la survie de patients admis dans une unité de soins intensifs. L'objectif est est de calculer la probabilité qu'un individu appartienne à la catégorie donnée (décès ou survie).

## Statistique descriptive

### Data Frame Summary

icu

Dimensions: 255 x 13

Duplicates: 0

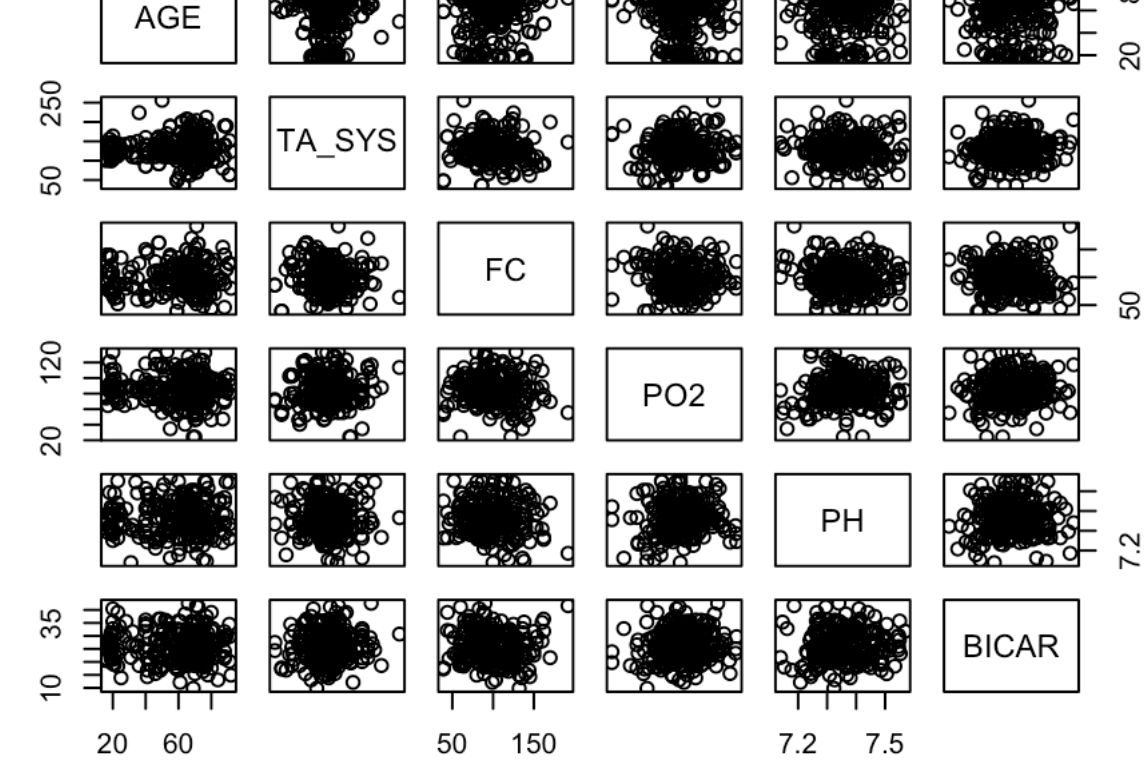
No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
1	DECEDE [integer]	Min : 0 Mean : 0.3 Max : 1	0 : 187 (73.3%) 1 : 68 (26.7%)		255 (100.0%)	0 (0.0%)
2	AGE [integer]	Mean (sd) : 58.9 (19.6) min ≤ med ≤ max: 16 ≤ 64 ≤ 92 IQR (CV) : 25 (0.3)	65 distinct values		255 (100.0%)	0 (0.0%)
3	SEXE [factor]	1. femme	162 (63.5%)		255 (100.0%)	0 (0.0%)

Generated by [summarytools](#) 1.0.1 (R version 4.3.2)

2024-11-29

Cette étude a inclus 255 patients, dont 68 (26,7%) sont décédés. L'âge moyen des patients est de 65 ans, avec une majorité de femmes (63,5%). Parmi eux, 121 n'ont pas eu recours à une intervention chirurgicale, et 108 (42,4%) présentaient une infection le jour de leur admission. Les paramètres vitaux montrent une tension artérielle systolique moyenne de 130,3 mmHg, une fréquence cardiaque moyenne de 98,4 bpm et un pH sanguin moyen de 7,4. Sur les 255 patients, 75,3% ont été admis en urgence. L'état de conscience, évalué sur une échelle ordinale de 0 à 2, indique que la majorité des patients avaient un score à 0. De plus, 68,5% des patients ont obtenu un score de 13 à 15 sur l'échelle globale d'évaluation du degré de conscience.

## Lien entre les variables quantitatives



L'exploration des relations entre variables numériques montre des relations faibles ou inexistantes entre la plupart des variables. En effet les point sont très dispersés, sans tendance claire. Afin de confirmer cette observation, nous réalisons le test de corrélation.

##		AGE	TA_SYS	FC	PO2	PH	BICAR
##	AGE	1.0000	0.0466	0.0514	-0.0710	0.0160	-0.0066
##	TA_SYS	0.0466	1.0000	-0.0477	0.0709	-0.0021	0.0967
##	FC	0.0514	-0.0477	1.0000	-0.0764	-0.0700	-0.0900
##	PO2	-0.0710	0.0709	-0.0764	1.0000	0.0918	0.1189
##	PH	0.0160	-0.0021	-0.0700	0.0918	1.0000	0.0428
##	BICAR	-0.0066	0.0967	-0.0900	0.1189	0.0428	1.0000

Ces résultats approuvent l'analyse ci-dessus. Aucune corrélation n'est présente entre les variables quantitatives.

## Lien entre les variables qualitatives

##		Paires	V_de_Cramer
##	1	SEXE_vs_CHIR_MED	0.04683462
##	2	SEXE_vs_INF_J0	0.02657494
##	3	SEXE_vs_URG_NURG	0.11289250
##	4	SEXE_vs_CONSC	0.08240314
##	5	SEXE_vs_GLASGOW	0.19540666
##	6	CHIR_MED_vs_INF_J0	0.20269121
##	7	CHIR_MED_vs_URG_NURG	0.52611859
##	8	CHIR_MED_vs_CONSC	0.13488734
##	9	CHIR_MED_vs_GLASGOW	0.24617605
##	10	INF_J0_vs_URG_NURG	0.17816798
##	11	INF_J0_vs_CONSC	0.13072851
##	12	INF_J0_vs_GLASGOW	0.26490557
##	13	URG_NURG_vs_CONSC	0.16728011
##	14	URG_NURG_vs_GLASGOW	0.23869285
##	15	CONSC_vs_GLASGOW	1.00000000

La fonction de Cramer permet d'évaluer l'association entre des variables qualitatives. Les résultats montrent une forte relation entre l'admission en urgences et le recours à la chirurgie. Une infection sévère ou mal contrôlée touche d'avantages les personnes dans un état critique et peut conduire à une situation nécessitant une admission urgente, enfin les admissions en urgence concernent plus souvent des patients dans un état critique ou grave. Les autres variables qualitatives montrent des associations très faibles entre elles

## Lien entre les variables qualitatives et la variable dépendante

Variable	Cramers
SEXE	0.0221079
CHIR_MED	0.2083729
INF_J0	0.1471663
URG_NURG	0.2631807
CONSC	0.4738193
GLASGOW	0.4997562

Ce test nous permet de mieux comprendre nos variables qualitatives afin de mieux les analyser. Nous observons que l'état de conscience et le score évaluant le degré de conscience sont également fortement associés à la variable décès. Nous allons donc pouvoir catégoriser ces 2 variables.

## Estimer le modèle complet et évaluer la significacité de chaque coefficient.

	Coefficient	Standard Error	Z value	P value
(Intercept)	23.205	17.202	1.349	0.1773
AGE	0.048	0.012	4.035	0.0001
SEXHomme	-0.292	0.405	-0.722	0.4704
CHIR_MEDoui	-0.449	0.461	-0.973	0.3304
INF_J0oui	-0.284	0.420	-0.677	0.4982
TA_SYS	-0.018	0.007	-2.562	0.0104
FC	-0.008	0.008	-0.934	0.3501
URG_NURGoui	2.427	0.824	2.945	0.0032
PO2	0.006	0.010	0.646	0.5184
PH	-3.758	2.319	-1.621	0.1051
BICAR	-0.012	0.031	-0.381	0.7031
CONSCModéré	21.659	1184.230	0.018	0.9854
GLASGOWModéré	1.321	1.192	1.108	0.2679
GLASGOWLéger	1.272	1.216	1.046	0.2954

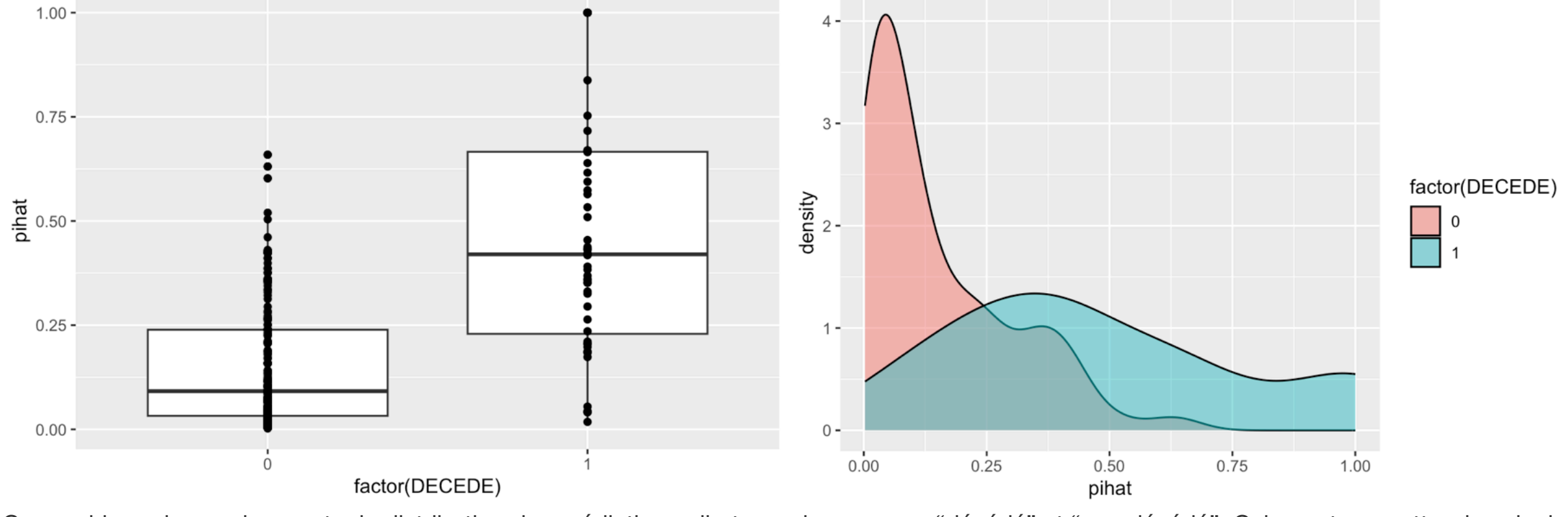
Pour l'analyse statistique, un seuil de significativité de 0,05 % (p < 0,0005) a été retenu pour déterminer si les relations observées étaient statistiquement significatives." Les p-valeurs de summary() testent l'importance de chaque variable individuellement dans le modèle. L'âge, l'admission en urgences et la tension artérielle systolique ont des effets significatifs sur le décès. Ce qui signifie que pour chaque année supplémentaire, la probabilité de décès augmente légèrement, de plus être admis en urgences augmente fortement la probabilité de décès. Enfin, une tension artérielle systolique plus élevée (coefficient négatif) pourrait réduire légèrement cette probabilité.

En ayant ensuite observés les p-valeurs de anova() qui testent la signification globale de l'ajout des variables dans le modèle comparé à un modèle sans variables. À la différence de notre test précédent, nous observons que l'état de conscience à un effet significatif sur la variable décès. Cela signifie que l'état de conscience pourrait avoir un effet indirect sur le décès. L'hypothèse est donc que l'état de conscience à un effet sur une variable qui a un impact direct sur la variable décès.

## Prédictions de la variable décès.

##	[1]	0.01181981	0.37649464	0.07213069	0.18979242	0.35099616	0.41092058
----	-----	------------	------------	------------	------------	------------	------------

Ce tableau représente la probabilité estimée de l'événement d'intérêt c'est à dire la probabilité de décès. Par exemple pour le patient 1, la probabilité de décès pour cet individu avec les caractéristiques spécifiées est d'environ 1.18%.



Ce graphique de gauche montre la distribution des prédictions pihat pour les groupes "décédé" et "non décédé". Cela peut permettre de voir si les prédictions sont plus élevées pour ceux qui sont décédés et plus faibles pour ceux qui ne le sont pas. Celui de droite donne une idée plus précise de la répartition des prédictions dans les deux groupes. Donc, ces 2 graphiques montrent que le modèle fait une bonne distinction entre les deux groupes et qu'il est performant.

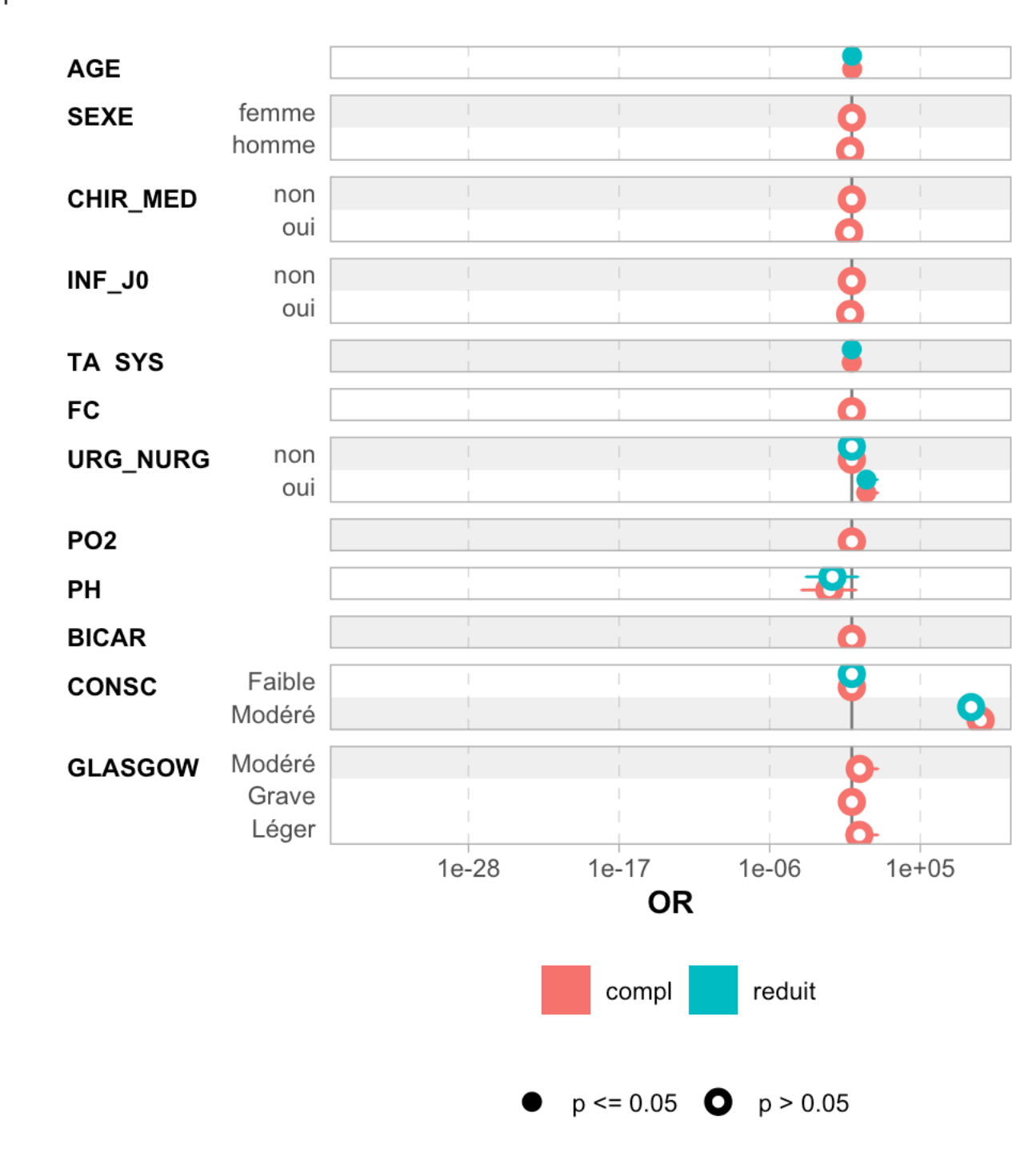
##		non	oui
##	0	0.8524590	0.1475410
##	1	0.3461538	0.6538462

Ces valeurs représentent le nombre de patients qui ont une probabilité de décès > 35% d'après notre modèle de régression logistique. Nous observons qu'il y a 34 vrais positifs et 156 vrais négatifs. Ce qui signifie que pour 85.2% des patients non décédés et pour 65.4% des patients décédés le modèle a correctement prédit la probabilité de décès.

## Selection automatique des variables du modèle

La sélection des variables est : DECEDE ~ AGE + TA\_SYS + URG\_NURG + PH + CONSC

En utilisant le critère d'information d'Akaike (AIC) cela simplifie le modèle en supprimant les variables non significative pour obtenir un modèle plus parcimonieux. Ainsi, le modèle ajusté contient l'âge, la tension artérielle systolique, l'admission en urgences, le score évaluant l'état du patient et l'état de conscience.



Les Odds ratios montre les mêmes résultats que pour l'ajustement du modèle, l'âge, la tension artérielle systolique, l'admission en urgences ont un effet significatif sur la même dépendante décès. L'âge et l'admission en urgence augmente les chances de décès. De plus, une tension artérielle systolique entraine une diminution des chances de décès. Le score évaluant l'état du patient n'a que très peu d'impact sur la variable décès. Nous pouvons voir un OR beaucoup trop élevée pour l'état de conscience avec un intervalle de confiance énorme indiquant un manque de précision. De plus le graphique compare le modèle complet et le modèle réduit et montre que seules les variables âge, tension artérielle systolique et l'admission en urgences ont un effet significatif sur la variable décès.

## Prédictions du modèle ajusté

##	[1]	"L'AUC du modèle complet est : 0.841635140815469 et celui du modèle réduit : 0.841635140815469"
----	-----	---

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	-0.135497	-0.024077	0.002647	0.000000	0.031926	0.057241

Les deux modèles ont des AUC égales, indiquant qu'ils sont comparables en termes de performance globale. Si l'un des modèle inclut des variables ou interactions supplémentaires, cela n'améliore pas significativement la capacité prédictive. Donc nous pouvons garder le modèle 2 (modèle simplifié).

## Validation croisée : entraînement du modèle

##		pihat	pihat2	pihattest
##	1	0.011819813	0.04537231	9.1923336e-09
##	2	0.072130694	0.06679117	3.813078e-02
##	3	0.048470552	0.05745120	5.687352e-02
##	4	0.023945428	0.04907355	7.570491e-02
##	5	0.006354070	0.04379226	3.706299e-02
##	6	0.008413741	0.04438141	4.393760e-02

En disant nos données en deux groupes (groupe d'entraînement et groupe de test) nous pouvons entraîner notre modèle afin qu'il puisse prédire la variable cible (décès) en fonction des autres variables de l'ensemble d'entraînement. Une fois le modèle entraîné, nous pouvons voir sa capacité à généraliser et à prédire. Ici, nous voyons que nous obtenons des valeurs de prédictions très proches entre les prédictions du modèlecalculer auparavant et les nouvelles prédictions (pihattest) réalisé par la validation croisée.

Enfin, ce modèle n'a que 12 valeurs dont les résidus sont supérieurs à 1,9 parmi les 255 patients. Nous pouvons en conclure que ce modèle est précis et prédits bien les données.

## Conclusion

Cette étude a permis d'examiner les facteurs de risque associés au décès des patients admis en unité de soins intensifs. En utilisant des techniques d'analyse statistique, telles que la régression logistique, nous avons pu identifier les variables les plus influentes dans la prédiction du risque de décès. L'analyse des variables a révélé que des facteurs cliniques tels que l'âge, la tension artérielle systolique et l'admission en urgence jouent un rôle crucial dans le pronostic des patients en soins intensifs. Les patients âgés, nécessitant une admission en urgence, présentent un risque significativement plus élevé de décéder. Cependant, une tension artérielle systolique élevée diminue fortement ce risque.