

TP 3 Régression logistique

Les données ICU (Intensive Care Unit) du fichier *icu.csv* sont constituées d'un échantillon de 255 patients admis dans un service hospitalier de soins intensifs pour adultes (ICU). L'objectif de l'étude est de mettre en place un modèle de régression logistique pour prédire la probabilité de survie de ces patients et les facteurs de risque associés à la mortalité dans un service de soins intensifs.

Intitulé des variables :

DECEDE : statut vital (0 = vivant, 1 = mort)
AGE : age en années
SEXE : (0 = homme, 1 = femme)
SERV : service lors de l'admission en USI (0 = Medical, 1 = Chirurgical)
INF_J0 : Infection probable à l'admission en USI (0 = Non, 1 = Oui)
TAS : tension artérielle systolique à l'admission en USI (en mm Hg)
FC : fréquence cardiaque à l'admission en USI (battements/min)
TYP_AD : type d'admission (0 = normal, 1 = en urgence)
PO2 : PO2 de la gazométrie artérielle initiale ($0 = > 60$, $1 = \leq 60$)
PH : PH de la gazométrie artérielle initiale ($0 = \geq 7.25$, $1 = < 7.25$)
BICAR : bicarbonate de la gazométrie artérielle initiale ($0 = \geq 18$, $1 = < 18$)
CONSC : niveau de conscience à l'admission (0 = ni coma ni stupeur, 1 = stupeur profonde, 2 = coma)
GLASGOW : score.

1. Faire les analyses descriptives univariées des variables SEXE et CONSC. Quelles seront les catégories de références pour ces variables lorsqu'elles seront incluses dans un modèle de régression ? Le codage dans R est-il optimal ? Le modifier s'il ne l'est pas.
2. Quelle est la probabilité de décès dans notre population ? Appliquer un modèle associant la mortalité à l'état de conscience. Que remarquez vous ? et que proposez vous ?
3. Quelle variable pourrait être variable de confusion pour l'association entre la mortalité et le service d'admission ?
Tester le modèle associant la mortalité au service d'admission, et celui la rattachant au service d'admission et à sa potentielle variable de confusion. Qu'en déduisez vous ?
4. A partir du premier modèle, quel est la probabilité de survie pour un individu ayant été admis dans un service médical ? dans un service chirurgical ?
5. Afin de savoir quels facteurs jouent sur cet évènement, réaliser une sélection de variables en utilisant toutes les variables de la base de données. Le modèle obtenu sera nommé modSel.
Quelle est l'augmentation du risque de décéder par rapport au risque de survivre lié à l'âge pour chaque année supplémentaire ? Pour 10 années supplémentaire ?
6. Tracer la courbe ROC associée à ce modèle et trouver son optimum. Quel seuil conseilleriez-vous d'utiliser pour prédire au mieux la survie individuelle ? Combien de décès seraient alors prédits ?

7. Effectuer un test de Hosmer et Lemeshow pour ce modèle. Vous observerez les valeurs observées et prédites données en sortie de cette fonction.
8. Repérez les individus pour lesquels les résidus de déviance sont en valeur absolue supérieurs à 2. Ces individus sont-ils tous décédés ?
Tester la robustesse du modèle en excluant ces individus. A l'aide d'une nouvelle courbe ROC, vous regarderez comment ont évolué la sensibilité et spécificité des prédictions.
9. On se demande si l'âge a un effet quadratique sur la chance de survie. Le modèle incluant ce terme quadratique de l'âge est-il meilleur que le précédent ? Vous testerez cela de deux manières différentes.
Quelle sensibilité et spécificité des prédictions obtenez vous ?