Samir Chergui

STA110 : Devoir N°2

A l’attention de Mr Jaupi.

Etude des 3 jeux de données

Exercice 3 e1

Dans le cadre de cette exercice, nous allons essayer de déterminer le risque de décès d’un automobiliste compte tenu de trois facteurs : l’âge, la vitesse et l’accélération.

Le jeu de donnée décrit 58 individus selon leur âge, la vitesse et l’accélération du véhicule.

Le jeu de donnée comporte également une variable catégorielle : Fatalité.

Cette variable possède deux réponses possibles 1 et 0 représentant respectivement la mort et la survie de l’individu.

Pour résoudre cet exercice, nous allons utiliser la régression logistique car les caractéristiques de ce jeu de donnée sont propices (variables explicatives déterminant une variable de Bernoulli) et on cherchera à établir si il éxiste un lien entre les facteurs et la fatalité (fatalité étant une variable dummy).

En utilisant Statgraphics, nous avons les valeurs suivantes :

Nombre d'observations: 58

**Modèle estimé de régression (Maximum de vraisemblance)**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | *Erreur* | *Rapports des chances* |
| *Paramètre* | *Estimation* | *type* | *estimées* |
| CONSTANTE | -15,0536 | 5,29655 |  |
| Vitesse | 0,0161775 | 0,0144943 | 1,01631 |
| Acceleration | 0,146279 | 0,11218 | 1,15752 |
| Age | 0,170905 | 0,0432274 | 1,18638 |

**Analyse de l'écart**

|  |  |  |  |
| --- | --- | --- | --- |
| *Source* | *Ecart* | *Ddl* | *Proba.* |
| Modèle | 34,6964 | 3 | 0,0000 |
| Résidu | 43,9759 | 54 | 0,8331 |
| Total (corr.) | 78,6723 | 57 |  |

Le modèle ici convient car la p-value est égale à 0,0000 ; le résidu (p-value égale à 83,31%) montre qu’il n’y a pas d’écart à l’ajustement

Le pourcentage d'écart expliqué par le modèle est de 44,1024%

Pourcentage ajusté = 33,9336

**Tests sur les rapports de vraisemblance**

|  |  |  |  |
| --- | --- | --- | --- |
| *Facteur* | *Khi-carré* | *Ddl* | *Proba.* |
| Vitesse | 1,3556 | 1 | 0,2443 |
| Acceleration | 1,83353 | 1 | 0,1757 |
| Age | 31,2312 | 1 | 0,0000 |

L'équation du modèle ajusté est :

Fatalite = exp(eta)/(1+exp(eta))

où

eta = -15,0536 + 0,0161775\*Vitesse + 0,146279\*Acceleration + 0,170905\*Age

Le modèle est donc exp(-15,0536 + 0,0161775\*Vitesse + 0,146279\*Acceleration + 0,170905\*Age)/1+exp(-15,0536 + 0,0161775\*Vitesse + 0,146279\*Acceleration + 0,170905\*Age)

Essayons d’améliorer le pourcentage ajusté en utilisant la procédure backward : On commence par un modèle complet c’est-à-dire un modèle qui prends en compte l’ensemble des facteurs puis un facteur est retiré si il y a une augmentation du R carré ajusté.

Etape 0:

3 facteur(s) dans le modèle. 54 ddl pour l'erreur.

Pourcentage d'écart expliqué = 44,10% Pourcentage ajusté = 33,93%

Etape 1:

Suppression du facteur Vitesse avec P-en-sortie = 0,244299

2 facteur(s) dans le modèle. 55 ddl pour l'erreur.

Pourcentage d'écart expliqué = 42,38% Pourcentage ajusté = 34,75%

L'équation du modèle ajusté est :

Fatalite = exp(eta)/(1+exp(eta))

où

eta = -16,9845 + 0,233906\*Acceleration + 0,162501\*Age.

Après exclusion du facteur vitesse, nous avons les résultats suivants :

**Tableau des prévisions inverses pour Acceleration**

Vitesse=122,793

Age=39,7759

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | *LC inf. à 95,0%* | *LC sup. à 95,0%* |
| *Pourcentage* | *Acceleration* |  |  |
| 0,1 | 15,4512 | -67,9411 | 29,0608 |
| 0,5 | 22,3491 | -43,1148 | 33,1284 |
| 1,0 | 25,334 | -32,3841 | 34,9009 |
| 2,0 | 28,3407 | -21,5872 | 36,6987 |
| 3,0 | 30,118 | -15,2136 | 37,7699 |
| 4,0 | 31,3922 | -10,6494 | 38,5432 |
| 5,0 | 32,391 | -7,07583 | 39,1532 |
| 6,0 | 33,2157 | -4,12811 | 39,66 |
| 7,0 | 33,9205 | -1,61177 | 40,0957 |
| 8,0 | 34,5376 | 0,589269 | 40,4796 |
| 9,0 | 35,0878 | 2,54988 | 40,824 |
| 10,0 | 35,5855 | 4,32119 | 41,1374 |
| 15,0 | 37,5633 | 11,3374 | 42,4059 |
| 20,0 | 39,0524 | 16,5849 | 43,3958 |
| 25,0 | 40,2823 | 20,8835 | 44,2491 |
| 30,0 | 41,3568 | 24,5986 | 45,0346 |
| 35,0 | 42,3326 | 27,9242 | 45,7966 |
| 40,0 | 43,2457 | 30,973 | 46,5726 |
| 45,0 | 44,1212 | 33,8102 | 47,4028 |
| 50,0 | 44,9791 | 36,4665 | 48,3402 |
| 55,0 | 45,8371 | 38,9393 | 49,461 |
| 60,0 | 46,7126 | 41,1934 | 50,8744 |
| 65,0 | 47,6257 | 43,177 | 52,7155 |
| 70,0 | 48,6015 | 44,8677 | 55,1125 |
| 75,0 | 49,676 | 46,313 | 58,1678 |
| 80,0 | 50,9059 | 47,6182 | 62,0144 |
| 85,0 | 52,395 | 48,9191 | 66,951 |
| 90,0 | 54,3728 | 50,409 | 73,7458 |
| 91,0 | 54,8705 | 50,7593 | 75,4802 |
| 92,0 | 55,4207 | 51,1387 | 77,4058 |
| 93,0 | 56,0378 | 51,556 | 79,5734 |
| 94,0 | 56,7426 | 52,0239 | 82,0576 |
| 95,0 | 57,5673 | 52,5619 | 84,9742 |
| 96,0 | 58,5661 | 53,2024 | 88,5173 |
| 97,0 | 59,8403 | 54,0061 | 93,0509 |
| 98,0 | 61,6176 | 55,1089 | 99,393 |
| 99,0 | 64,6243 | 56,9421 | 110,155 |
| 99,5 | 67,6092 | 58,7367 | 120,863 |
| 99,9 | 74,5071 | 62,8324 | 145,661 |

Pour une probabilité de 50% de mort ou survie d’un individu, nous avons une accélération égale à 47,98 (avec une vitesse de 122,79 et un âge de 39,78.

Exercice 3 e2

On cherche à tester l’effet d’une substance toxique sur un type d’insecte donné. Un échantillon de 3000 insectes a été choisi au hasard, ensuite divisé en cinq groupes de tailles 600 chacun. Les individus de chaque groupe ont été exposé a une dose déterminée du pesticide. Un jour plus tard les insectes ont été observés. La mort par exposition a été noté par 1 et la survie par 0.

Notre objectif sera de chercher à estimer la dose médiane :

Exercice 4 e2

Le fichier colle10.xls contient les données de 10 lots de colle qui ont été sélectionnés au hasard. Dans chaque lot, 4 déterminations de l’humidité ont été effectués.

L’objectif ici sera de déterminer quel lot possède le taux d’humidité le plus faible.