

Modèle linéaire généralisé

Feuille 2 : régression logistique 2

Robin Ryder

Mai 2019

1 Prédiction : la navette Challenger

Le 28 janvier 1986, la navette spatiale Challenger s'est désintégrée peu après son décollage, tuant les 7 astronautes à bord, suite à la défaillance d'un joint torique potentiellement due au froid.

Des tests du joint torique avaient été effectués à diverses températures avant le lancement. Dalal, Fowlkes et Hoadley (1989) et Casella et Berger (2002) fournissent les données du fichier `shuttle.txt` (températures en degrés Fahrenheit ; 1=échec, 0=réussite).

1. Charger les données sous R et les explorer rapidement.
2. Obtenir une estimation ponctuelle des paramètres d'une régression logistique.
3. Le jour du lancement, la température était de 31 ° F. Quelle est la probabilité d'échec prédite par le modèle ? Cette prédiction vous paraît-elle fiable ?

2 Données binomiales : informations de la CAF

Nous disposons d'informations de la CAF au niveau de l'IRIS (équivalent à un pâté de maisons). Pour chaque IRIS de grande ville, nous avons la population, le nombre d'allocations du RSA et de l'AAH, le nombre de demandeurs d'emploi, et le nombre de quelques équipements (écoles, restaurants...)

1. Charger les données `alloc_iris.csv` et les explorer rapidement.
2. Expliquer le nombre d'allocataires du RSA par les autres covariables ; répéter pour d'autres variables observées d'intérêt.

3 Mesure de la précision : diabète chez les Pimas

Les Pimas sont un peuple amérindien de l'Arizona, chez qui le diabète est particulièrement prévalent. Nous allons chercher à expliquer le diabète chez des femmes Pimas adultes.

1. Charger les données à l'aide des commandes

```
install.packages("MASS")
require(MASS)
data(Pima.tr)
data(Pima.te)
```

Nous utiliserons les données `Pima.tr` pour entraîner notre modèle, puis le testerons sur les données `Pima.te`.

2. Explorer rapidement les données.
3. Effectuer une régression logistique sur les données `Pima.tr`, et l'affiner jusqu'à obtenir un modèle satisfaisant.
4. À l'aide de la fonction `predict()`, faire une prédiction du statut de diabétique chez les individus du jeu `Pima.te`.
5. Quel est votre taux de prédiction ?
6. Calculer le taux de prédiction pour d'autres modèles, et comparer.

4 Un cas limite : authenticité de billets de banque

Nous allons analyser le jeu de données `bank` du package `gclus`, qui donne diverses mesures prises sur des billets de banque faux ou authentiques :

```
install.packages("gclus")
require(gclus)
data(bank)
```

1. Charger les données et les explorer rapidement.
2. Implémenter une régression logistique. La variable à expliquer est `Status`.
3. Chercher sur Internet une explication du message d'erreur. Cette erreur est-elle problématique ?
4. Tracer un graphe sur lequel chaque point correspond à une observation, avec en abscisse la variable `Bottom`, en ordonnée la variable `Diagonal`, et la couleur du point déterminée par `Status`. Qu'observez-vous ?