

Modèle linéaire généralisé

Feuille 3 : mesures d'adéquation

Robin Ryder

Mai 2019

1 Test du χ^2

On va tester le lien entre le fait de fumer et le fait de faire du sport, chez des étudiants de l'Université d'Adelaide :

```
> library(MASS)
> data(survey)
> table(survey$Exer, survey$Smoke)
```

Effectuer un test du χ^2 pour tester l'hypothèse que ces deux variables sont indépendantes. Quel est le nombre de degrés de liberté à prendre en compte dans le test ?

2 Choix de modèle à l'aide de la déviance

Charger les données `german_credit.csv`, qui contiennent pour 1000 clients d'une banque d'une part si le client est un bon emprunteur (variable à expliquer), d'autre part des covariables.

Proposer plusieurs modèles et les tester à l'aide d'un test de déviance.

3 Validation croisée

Reprendre les données Titanic.

1. Proposer un modèle de classification, et estimer son taux d'erreur par validation croisée.
2. Répéter pour d'autres modèles. Quel modèle préférez-vous ?
3. Les différentes techniques de choix de modèle rencontrées sont-elles en accord ?

4 Comparaison des modèles probit et logistique

Reprendre les données Pima (feuille 2, exercice 3). Comparer le taux de prédiction avec le modèle logistique au taux de prédiction avec le modèle probit.