Modèle fréquence sévérité

Marie-Pier Côté

Université Laval

4 décembre 2014

Modèle fréquence sévérité

Plusieurs base de données en actuariat sont en deux parties :

1. Fréquence : indique s'il y a eu une réclamation ou non, ou plus généralement, le nombre de réclamations.
2. Sévérité : indique le montant d'une réclamation sachant qu'elle a eu lieu.

Traditionnellement, on inclut pas de variables explicatives dans le modèle (e.g. cours de modélisation de distribution de sinistres, IARD 1).

Modèle fréquence sévérité

On a

$$S_i = \begin{cases} \sum_{j=1}^{N_i} Y_{i,j}, & \text{si } N_i > 0 \\ 0, & \text{si } N_i = 0, \end{cases}$$

où

- S_i est le montant total payé pour l'assuré i
- N_i est le nombre de réclamations pour l'assuré i
- $Y_{i,j}$ est le montant de la j^e réclamation pour l'assuré i

Modèle fréquence sévérité

On peut utiliser les techniques de modèles linéaires généralisés et de régression linéaire multiple pour estimer la distribution de N_i et $Y_{i,j}$. Par exemple,

- $N_i \in \{0, 1\}$, on peut utiliser un GLM binomial,
- $N_i \in \{0, 1, \ldots\}$, on peut utiliser un GLM Poisson ou binomiale négative,
- $Y_{i,j}$ peut être modélisé avec une régression linéaire multiple sur le logarithme naturel,
- $Y_{i,j}$ peut aussi être modélisé avec un GLM Gamma.

Exemple : dépenses médicales

- Données du *Medical Expenditure Panel Survey (MEPS)* pour 2003
  - Sondage sur la population des États-Unis
  - Informations sur l'utilisation des soins de santé, les dépenses encourues et la couverture d'assurance.
- Échantillon de 2000 individus âgés entre 18 et 65 ans.
- Source : [Frees, 2009].

Exemple : dépenses médicales

On désire modéliser conjointement :

- la fréquence N_i,
  - pour comprendre les variables exogènes qui ont un impact sur la probabilité d'être hospitalisé pour au moins une nuit,
  - $N_i \in \{0, 1\}$,
- la sévérité $Y_{i,j}$.
  - sachant que l'individu est hospitalisé, quels facteurs influencent les dépenses médicales ?
  - $Y_{i,j} \in (0, \infty)$.

Modèle pour la fréquence

On ajuste un modèle binomial avec lien logistique. Les variables explicatives sont :

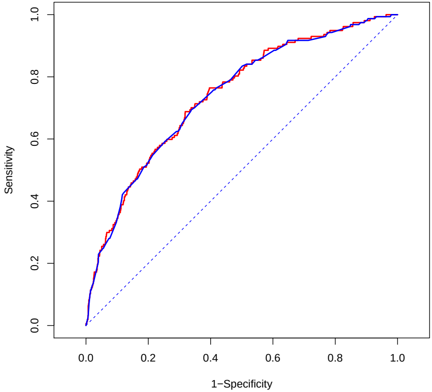
- `AGE`
- `GENDER` : 1 si femme, 0 sinon.
- `RACE: ASIAN, BLACK, NATIV, OTHER, WHITE.`
- `EDUC: COLLEGE, HIGHSCH, LHIGHSC.`
- `PHSTAT: EXCE, FAIR, GOOD, POOR, VGOO.`
- `INCOME: HINCOME, LINCOME, MINCOME, NPOOR, POOR.`
- `insure` : 1 si couvert par une assurance médicale privée pour un mois de 2013 ou plus, 0 sinon.

Après les tests de déviance, le modèle retenu est :

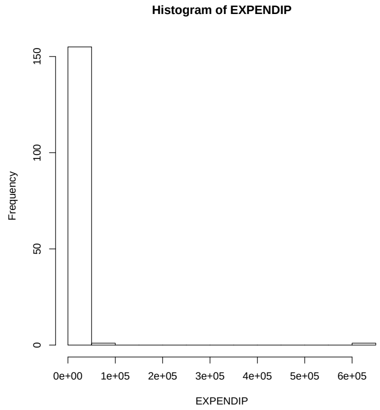
```
glm(formula = Freq ~ GENDER + PHSTAT + INCOME + insure, family = binomial)
Deviance Residuals:
    Min        1Q    Median        3Q       Max
-1.2358   -0.4339   -0.3333   -0.2487    2.9706
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -4.8264     0.3954 -12.207  < 2e-16 ***
GENDER          0.7190     0.1904   3.776 0.000159 ***
PHSTATFAIR      0.3128     0.3404   0.919 0.358120
PHSTATGOOD      0.4264     0.2562   1.664 0.096025 .
PHSTATPOOR      1.9953     0.3310   6.029 1.65e-09 ***
PHSTATVGOO      0.1757     0.2644   0.664 0.506374
INCOMELINCOME   0.4758     0.2827   1.683 0.092386 .
INCOMEMINCOME   0.2853     0.2432   1.173 0.240758
INCOMENPOOR     0.6602     0.3852   1.714 0.086538 .
INCOMEPOOR      0.8827     0.2644   3.339 0.000840 ***
insure          1.3655     0.2991   4.565 4.99e-06 ***
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 1100.36  on 1999  degrees of freedom
Residual deviance:  991.72  on 1989  degrees of freedom
AIC: 1013.7
```

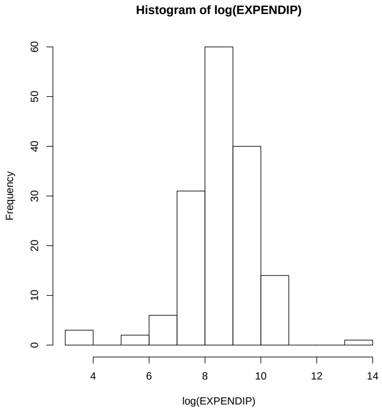
Modèle pour la fréquence - comparaison des courbes ROC

Il semble que nous ne perdons pas de pouvoir de prévision en laissant tomber les variables `AGE`, `RACE` et `EDUC`.



Modèle pour la sévérité





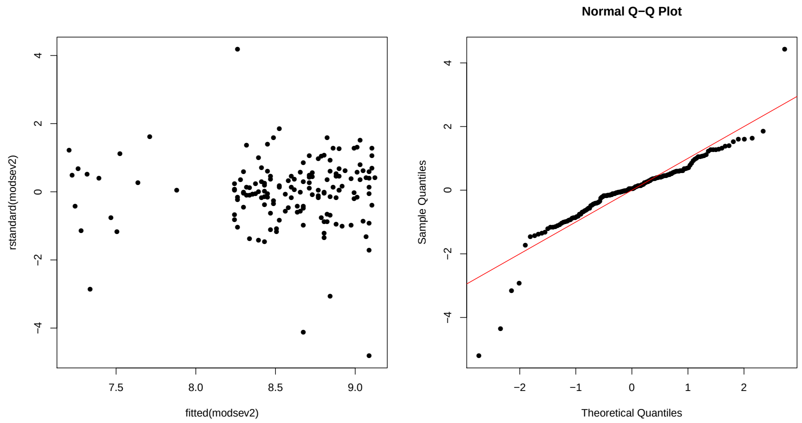
Modèle pour la sévérité

Le modèle sélectionné avec la méthode pas-à-pas est le suivant (AIC=513.895).

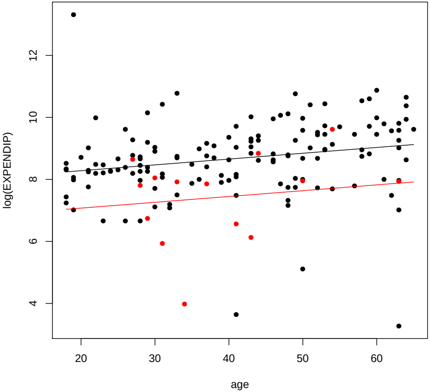
```
Call:
lm(formula = I(log(EXPENDIP)) ~ AGE + insure, data = DatSev)
Residuals:
    Min      1Q  Median      3Q     Max
-5.8170 -0.5088  0.0602  0.5991  5.0552
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.69978    0.42736  15.677  < 2e-16 ***
AGE          0.01874    0.00708   2.647 0.008974 **
insure       1.20661    0.34304   3.517 0.000573 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.224 on 154 degrees of freedom
Multiple R-squared:  0.1164,Adjusted R-squared:  0.1049
F-statistic: 10.14 on 2 and 154 DF,  p-value: 7.281e-05
```

Modèle pour la sévérité



Modèle pour la sévérité



Modèle Gamma pour la sévérité

On peut aussi considérer un modèle Gamma pour la sévérité. La densité de la loi Gamma fait partie de la famille exponentielle de dispersion. Avec $\theta = -1/\mu = \beta/\alpha$, et $a(\phi) = 1/\alpha$, on a

$$f_Y(y; \theta, \phi) = \exp\left\{ \frac{y\theta + \ln(-\theta)}{\phi} + \alpha \ln \alpha + (\alpha - 1) \ln y - \ln \Gamma(\alpha) \right\}.$$

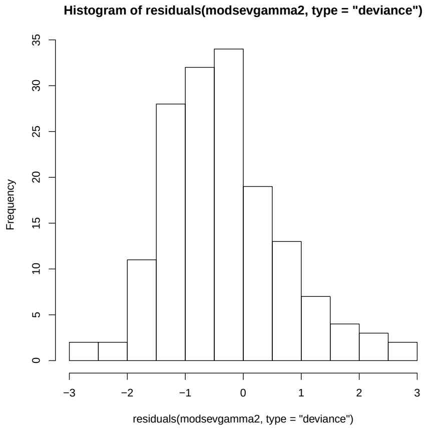
Il faut estimer le paramètre de dispersion ϕ, ce qui est fait avec le chi-carré de Pearson en R.

On utilise le lien canonique, qui es le lien inverse $\eta = 1/\mu$.

Modèle Gamma pour la sévérité

```
glm(formula = EXPENDIP ~ EDUC + PHSTAT + INCOME + insure, family = Gamma,
    data = DatSev)
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.449e-04  1.024e-04   3.367 0.000973 ***
EDUCHIGHSCH   -1.645e-05  2.671e-05  -0.616 0.538999
EDUCLHIGHSC   -6.947e-05  2.443e-05  -2.844 0.005103 **
PHSTATFAIR     5.078e-06  5.083e-05   0.100 0.920558
PHSTATGOOD    -5.931e-05  3.357e-05  -1.767 0.079400 .
PHSTATPOOR    -4.708e-05  3.472e-05  -1.356 0.177210
PHSTATVGOO    -2.547e-05  3.718e-05  -0.685 0.494439
INCOMELINCOME  1.304e-05  2.651e-05   0.492 0.623578
INCOMEMINCOME -2.302e-05  2.263e-05  -1.017 0.310755
INCOMENPOOR    4.952e-05  5.308e-05   0.933 0.352410
INCOMEPOOR     6.599e-05  3.243e-05   2.035 0.043675 *
insure        -1.901e-04  9.488e-05  -2.004 0.046921 *
(Dispersion parameter for Gamma family taken to be 1.418724)
    Null deviance: 281.14  on 156  degrees of freedom
Residual deviance: 177.58  on 145  degrees of freedom
AIC: 3209
```

Modèle pour la sévérité



Bibliographie

FREES, E. W. (2009).
Regression modeling with actuarial and financial applications.
Cambridge University Press.