ZIP Regression

Zero-Inflated Poisson Regression

Ricco Rakotomalala

Université Lumière Lyon 2

Plan

1. Problématique – Modélisation – Estimation
2. Inférence statistique
3. Prédiction et résidus
4. Conclusion
5. Références

Principe et estimation des paramètres

ZERO INFLATED POISSON REGRESSION

Pourquoi la Régression ZIP

Contexte

La loi de Poisson est privilégiée dans un modèle de comptage. Mais dans certains cas elle n'est pas adaptée, notamment parce qu'il y a une surreprésentation de la valeur 0.

Pourquoi

Parce que la valeur 0 peut représenter l'absence du phénomène ou un nombre d'apparition nul sur la période étudiée c.-à-d. il y a 2 sources de la valeur 0. Ex. Nombre d'infidélité = 0 parce la personne est fidèle par nature, ou parce qu'elle n'a pas eu le temps ou l'opportunité de tromper son conjoint sur la période étudiée.

Conséquence

On note une surdispersion dans la régression de Poisson. La variance de Y/X est supérieure à sa moyenne, violant la propriété sous-jacente à l'hypothèse Poisson.

Solution

Combiner deux lois pour la modélisation : (1) loi binomiale pour la survenance ou non de Y = 0 ; (2) loi de Poisson pour le comptage des évènements, y compris possiblement la valeur 0.

Principe de la Régression ZIP

La régression ZIP procède d'une modélisation avec la combinaison de 2 lois de distribution : Binomiale et Poisson

X,Z : variables explicatives, à voir plus loin

Distribution de Y

$$P(Y = y|X,Z) = \begin{cases} \pi + (1-\pi)e^{-\mu} & \text{si } y = 0 \\ (1-\pi)\dfrac{\mu^{y} e^{-\mu}}{y!} & \text{si } y > 0 \end{cases}$$

Eléments à modéliser

- π est la probabilité que l'on ait structurellement 0, par conséquent il y a $(1 - \pi)$ de chances d'être dans la situation modélisable par Poisson
- μ est le paramètre de la loi de Poisson

Propriétés

$$\begin{cases} E[Y|X,Z] = (1-\pi)\,\mu \\ V[Y|X,Z] = (1-\pi)\,(\mu + \pi\,\mu^2) \end{cases}$$

Puisque $(0 \leq \pi \leq 1)$ et $(\mu > 0)$, on note que la variance est au moins supérieure à la moyenne dans cette configuration.

Double modélisation, combinée

Odds

Modéliser π

$$\ln \frac{\pi}{1-\pi} = \ln \lambda = a_0 + a_1 z_1 + \cdots + a_m z_m$$

Log-Odds = LOGIT

Fonction de lien LOGIT

m explicatives, Z = (1 | z_1 | … | z_m)

$$\lambda = \frac{\pi}{1-\pi} \Rightarrow \pi = \frac{\lambda}{1+\lambda}$$

$$\lambda = e^{a_0 + a_1 z_1 + \cdots}$$

Modéliser μ

$$\ln \mu = b_0 + b_1 x_1 + \cdots + b_k x_k$$

Fonction de lien LOG

k explicatives, X = (1 | x_1 | … | x_k)

$$\mu = e^{b_0 + b_1 x_1 + \cdots}$$

X, Z peuvent être identiques, avoir des variables en commun, ou être disjoints. !

Données et notations

Nombre d'infidélités d'une personne (sur 1 année)

Affairs	Constante	Gender	RatingMarriage	Occupation	YearsMarried
0	1	0	4	1	15
0	1	0	4	5	7
0	1	1	3	6	4
0	1	1	4	6	10
0	1	0	5	5	1.5
0	1	0	3	6	0.125
1	1	0	2	4	15
1	1	1	4	6	4
2	1	1	4	4	15
3	1	0	4	1	10
3	1	0	4	5	15
7	1	0	3	5	10
7	1	1	3	5	10
7	1	0	2	3	15
7	1	1	4	6	10
12	1	1	2	5	15
12	1	1	2	6	10
1	1	1	4	2	1.5
1	1	1	5	6	10
1	1	0	5	1	1.5

n = 20 observations,

m = 2 descripteurs, Z = (1, Occupation, YearsMarried)

k = 2 descripteurs, X = (1, Gender, RatingMarriage)

X_i , Z_i description de l'individu n°i.

Ex. $X_1 = (1, 0, 4)$, $Z_1 = (1,1,15)$

Y_i valeur de la cible pour l'individu n°i, Ex. $Y_8 = 1$

Notation matricielle : $a_0 + a_1 X_{i,1} + \ldots = X_i a$

Gender : 1, homme ; 0, femme

Satisfaction dans l'union (1: faible ; 5 : très satisfait.e)

Type de profession (1 peu gratifiante, 7 très gratifiante)

Nombre d'années de mariage

On souhaite modéliser le nombre d'infidélités (Y = Affairs) d'une personne sur l'année passée.

Fonction de vraisemblance – Estimateur du Max. de Vraisemblance

L'EMV cumule deux bonnes propriétés asymptotiques : absence de biais, normalité

La **log-vraisemblance LL** se décompose en différentes parties selon que (y = 0) ou pas. A maximiser en fonction de b et a.

$$\boxed{LL = L1 + L2 - L3}$$

Où

$$\begin{cases} L1 = \sum_{\{i:y_i=0\}} \ln[\lambda_i + e^{-\mu_i}] \\ L2 = \sum_{\{i:y_i>0\}} [y_i \ln(\mu_i) - \mu_i - \ln(y_i!)] \\ L3 = \sum_{i=1}^{n} \ln(1+\lambda_i) \end{cases}$$

Cellules variables

Optimisation avec le « solveur » Excel.

Données « Affairs »

	Affairs	Constante	Gender	RatingMarriage	Occupation	YearsMarried	Mu_i	Lambda_i	L1	L2	L3
Coefficients Logistique		-4.4574			1.1608	-0.3970					
Coefficients Poisson		3.1027	0.6240	-0.6502							
	0	1	0	4	1	15	1.6519	0.0001	-1.6514	0.0000	0.0001
	0	1	0	4	5	7	1.6519	0.2387	-0.8432	0.0000	0.2140
	0	1	1	3	6	4	5.9071	2.5072	0.9203	0.0000	1.2548
	0	1	1	4	6	10	3.0832	0.2315	-1.2825	0.0000	0.2083
	0	1	0	5	5	1.5	0.8622	2.1190	0.9327	0.0000	1.1375
	0	1	0	3	6	0.125	3.1649	11.6774	2.4613	0.0000	2.5398
	1	1	0	2	4	15	6.0637	0.0031	0.0000	-4.2614	0.0031
	1	1	1	4	6	4	3.0832	2.5072	0.0000	-1.9572	1.2548
	2	1	1	4	4	15	3.0832	0.0031	0.0000	-1.5244	0.0031
	3	1	0	4	1	10	1.6519	0.0007	0.0000	-1.9379	0.0007
	3	1	0	4	5	15	1.6519	0.0100	0.0000	-1.9379	0.0099
	7	1	0	3	5	10	3.1649	0.0725	0.0000	-3.6252	0.0700
	7	1	1	3	5	10	5.9071	0.0725	0.0000	-1.9992	0.0700
	7	1	0	2	3	15	6.0637	0.0010	0.0000	-1.9726	0.0010
	7	1	1	4	6	10	3.0832	0.2315	0.0000	-3.7266	0.2083
	12	1	1	2	5	15	11.3175	0.0100	0.0000	-2.1885	0.0099
	12	1	1	2	6	10	11.3175	0.2315	0.0000	-2.1885	0.2083
	1	1	1	4	2	1.5	3.0832	0.0651	0.0000	-1.9572	0.0631
	1	1	1	5	6	10	1.6093	0.2315	0.0000	-1.1335	0.2083
	1	1	0	5	1	1.5	0.8622	0.0204	0.0000	-1.0105	0.0202

SUM	0.5371	-31.4205	7.4853

LL(a,b)	-38.3686

Cellule cible à maximiser

$$\hat{a} = \begin{pmatrix} -4.4574 \\ 1.1608 \\ -0.3970 \end{pmatrix}$$

$$\hat{b} = \begin{pmatrix} 3.1027 \\ 0.6240 \\ -0.6502 \end{pmatrix}$$

Estimation des paramètres – Algorithme Newton-Raphson

Algorithme itératif, passage de l'étape (t) à (t+1) :

$$\begin{pmatrix} b \\ a \end{pmatrix}^{(t+1)} = \begin{pmatrix} b \\ a \end{pmatrix}^{(t)} - H^{-1} \times g$$

A l'optimum, toutes ses valeurs sont nulles.

Matrice hessienne, dérivées partielles secondes.

Vecteur gradient, dérivées partielles premières.

$$g = \begin{pmatrix} \frac{\partial LL}{\partial b} \\ \frac{\partial LL}{\partial a} \end{pmatrix}$$

A l'optimum, son inverse **H**$^{-1}$ correspond à la matrice de variance covariance des coefficients $\begin{pmatrix} \hat{b} \\ \hat{a} \end{pmatrix}$

$$H = \begin{pmatrix} \frac{\partial^2 LL}{\partial b_r \partial b_s} & \frac{\partial^2 LL}{\partial b_r \partial a_s} \\ \frac{\partial^2 LL}{\partial b_r \partial a_s} & \frac{\partial^2 LL}{\partial a_r \partial a_s} \end{pmatrix}$$

$$\frac{\partial LL}{\partial b_r} = \sum_{\{i:y_i=0\}} \left[\frac{-x_{ir}\mu_i}{\lambda_i e^{\mu_i}+1}\right] + \sum_{\{i:y_i>0\}} (y_i - \mu_i)x_{ir} \qquad \text{(r = 1, 2, ..., k)}$$

$$\frac{\partial LL}{\partial a_r} = \sum_{\{i:y_i=0\}} \left[\frac{z_{ir}\lambda_i e^{\mu_i}}{\lambda_i e^{\mu_i}+1}\right] - \sum_{i=1}^{n} \left(\frac{\lambda_i}{1+\lambda_i}\right) z_{ir} \qquad \text{(r = 1, 2, ..., m)}$$

$$\frac{\partial^2 LL}{\partial b_r \partial b_s} = \sum_{\{i:y_i=0\}} \left[\frac{x_{ir}x_{is}\mu_i[(\mu_i-1)\lambda_i e^{\mu_i}-1]}{(\lambda_i e^{\mu_i}+1)^2}\right] - \sum_{\{i:y_i>0\}} \mu_i x_{ir} x_{is} \qquad \text{(r , s = 1, 2, ..., k)}$$

$$\frac{\partial^2 LL}{\partial a_r \partial a_s} = \sum_{\{i:y_i=0\}} \left[\frac{z_{ir}z_{is}\lambda_i e^{\mu_i}}{(\lambda_i e^{\mu_i}+1)^2}\right] - \sum_{i=1}^{n} \frac{z_{ir}z_{is}\lambda_i}{(\lambda_i+1)^2} \qquad \text{(r , s = 1, 2, ..., m)}$$

$$\frac{\partial^2 LL}{\partial b_r \partial a_s} = \sum_{\{i:y_i=0\}} \left[\frac{z_{ir}z_{is}\lambda_i \mu_i e^{\mu_i}}{(\lambda_i e^{\mu_i}+1)^2}\right] \qquad \text{(r = 1, 2, ..., k ; s = 1, 2, ..., m)}$$

La convergence peut être actée de différentes manières : stabilité des coefficients, stabilité de la log-vraisemblance, etc.

Un exemple sous R

```
#importation des données
library(xlsx)
D <- read.xlsx("affairs_for_R.xlsx",sheetIndex=1)
print(str(D))

## 'data.frame':    20 obs. of  9 variables:
##  $ Affairs        : num  0 0 0 0 0 0 1 1 2 3 ...
##  $ Gender         : num  0 0 1 1 0 0 0 1 1 0 ...
##  $ Age            : num  42 27 27 37 22 22 37 32 57 32 ...
##  $ YearsMarried   : num  15 7 4 10 1.5 0.125 15 4 15 10 ...
##  $ Children       : num  1 0 0 1 0 0 1 0 1 1 ...
##  $ Religiousness  : num  3 2 2 2 3 2 4 4 1 4 ...
##  $ Education      : num  12 17 18 18 16 16 14 20 17 14 ...
##  $ Occupation     : num  1 5 6 6 5 6 4 6 4 1 ...
##  $ RatingMarriage: num  4 4 3 4 5 3 2 4 4 4 ...
```

Importation des données, 9 variables dont la cible « Affairs ». n=20.

```
#library
library(pscl)

#modélisation
m1 <- zeroinfl(Affairs ~ Gender+RatingMarriage | Occupation+YearsMarried,data=D)
print(sm1 <- summary(m1))

##
## Count model coefficients (poisson with log link):
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)      3.1027     0.4214   7.362 1.81e-13 ***
## Gender           0.6240     0.2681   2.328   0.0199 *
## RatingMarriage  -0.6502     0.1325  -4.908 9.19e-07 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -4.4563     4.8644  -0.916    0.360
## Occupation     1.1605     0.8833   1.314    0.189
## YearsMarried  -0.3970     0.2419  -1.641    0.101
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 17
## Log-likelihood: -38.37 on 6 Df
```

Modélisation

$$\ln \mu = b_0 + b_1 Gender + b_2 RatingMarriage$$

$$\ln \frac{\pi}{1-\pi} = \ln \lambda = a_0 + a_1 Occupation + a_2 YearsMarried$$

```
#matrice de variance covariance des coefficients
print(round(m1$vcov,5))

##                     count_(Intercept) count_Gender count_RatingMarriage
## count_(Intercept)             0.17761     -0.04395             -0.04770
## count_Gender                 -0.04395      0.07187             -0.00140
## count_RatingMarriage         -0.04770     -0.00140              0.01755
## zero_(Intercept)             -0.01908     -0.09538              0.03738
## zero_Occupation               0.00123      0.01184             -0.00422
## zero_YearsMarried            -0.00294      0.00220              0.00091
##                     zero_(Intercept) zero_Occupation zero_YearsMarried
## count_(Intercept)           -0.01908         0.00123          -0.00294
## count_Gender                -0.09538         0.01184           0.00220
## count_RatingMarriage         0.03738        -0.00422           0.00091
## zero_(Intercept)            23.66270        -3.99549          -0.06026
## zero_Occupation             -3.99549         0.78016          -0.05735
## zero_YearsMarried           -0.06026        -0.05735           0.05853
```

$\boxed{\sqrt{0.07187} = 0.2681}$

$\boxed{\sqrt{0.05853} = 0.2419}$

Matrice de variance covariance des coefficients estimés. Nous avons les variances sur les diagonales.

Tests et intervalles de confiance

INFÉRENCE STATISTIQUE

Inférence basée sur la normalité asymptotique

Puisque : $\begin{cases} \hat{b}_r \cong \mathcal{N}(b_r, \hat{\sigma}_{\hat{b}_r}) \\ \hat{a}_r \cong \mathcal{N}(a_r, \hat{\sigma}_{\hat{a}_r}) \end{cases}$

Nous pouvons effectuer différents tests, dont les tests de significativité.

Ex. $\begin{cases} H_0 : b_{Gender} = 0 \\ H_1 : b_{Gender} \neq 0 \end{cases}$ ⇨ $z_{Gender} = \frac{0.6240}{0.2681} = 2.328 \rightarrow \mathrm{P}(> |z|) = p.value = 0.0199$

```
## Count model coefficients (poisson with log link):
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)       3.1027     0.4214   7.362 1.81e-13 ***
## Gender            0.6240     0.2681   2.328   0.0199 *
## RatingMarriage   -0.6502     0.1325  -4.908 9.19e-07 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -4.4563     4.8644  -0.916    0.360
## Occupation      1.1605     0.8833   1.314    0.189
## YearsMarried   -0.3970     0.2419  -1.641    0.101
## ---
```

Nous pouvons également calculer les intervalles de confiance.

Intervalle de confiance à 90% du coefficient de YearsMarried

$$\hat{a}_{YearsMarried} \pm u_{0.95} \times \hat{\sigma}_{\hat{a}_{YearsMarried}}$$

$$-0.3970 \pm 1.6448 \times 0.2419$$

$$[-0.79489 , 0.00089]$$

Test du rapport de vraisemblance

Permet de confronter des scénarios de modélisation

ou tester la significativité d'un bloc de coefficients.

$$\ln \mu = b_0 + b_1 Gender + b_2 RatingMarriage$$

$$\ln \frac{\pi}{1-\pi} = \ln \lambda = a_0 + a_1 Occupation + a_2 YearsMarried$$

$$\begin{cases} H_0 : a_1 = a_2 = 0 \\ H_1 : a_1 \neq 0 \; ou \; a_2 \neq 0 \end{cases}$$

Log-vraisemblance du modèle sous H0.

Statistique de test :

$$LR = 2 \times [LL(M) - LL(M0)]$$

Sous H0, suit une loi du χ² à (q = 2) degrés de liberté

```
#modèle sous H0 - Proba πi estimée par une constante
m0 <- zeroinfl(Affairs ~ Gender+RatingMarriage | 1,data=D)
print(summary(m0))

##
## Count model coefficients (poisson with log link):
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)        3.0955     0.4188   7.392 1.44e-13 ***
## Gender             0.6106     0.2798   2.182   0.0291 *
## RatingMarriage    -0.6427     0.1338  -4.802 1.57e-06 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.2879     0.6674   -1.93   0.0536 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 9
## Log-likelihood: -41.73 on 4 Df

#stat. de test de rapport de vraisemblance
LR <- 2*(m1$loglik - m0$loglik)
print(LR)

## [1] 6.71756

#p-value
print(pchisq(LR,df=2,lower.tail=FALSE))

## [1] 0.03477765
```

LL(M)=-38.37 (page 6)

LL(M0) = -41.73

LR = 6.71756

p-value = 0.034

Au risque 5%, les données contredisent H0. L'estimation de la probabilité π_i avec une constante a dégradé significativement le modèle.

$$\hat{\lambda}_i = e^{(\hat{a}_0)} = e^{-1.2879} = 0.2758$$

$$\hat{\pi}_i = \frac{\hat{\lambda}_i}{1+\hat{\lambda}_i} = 0.2162, \forall i$$

Pertinence du scénario « zero-inflated »

Attention, la configuration ($a_0 = a_1 = \ldots = a_k = 0$) n'équivaut pas à une régression de Poisson « classique », parce que dans ce cas, $\hat{\pi}_i = 0.5\,, \forall i$

On passe par le **Test de Vuong** (1989) pour tester la légitimité de la régression ZIP par rapport à la régression de Poisson (modèles non-imbriqués).

Probabilité d'obtenir le bon compte y_i pour l'individu n°i avec le modèle MZ

Stratégie : Créer les deux modèles : MZ pour ZIP Régression, MP pour régression de Poisson. Calculer les $\hat{P}_{MZ}(Y = y_i) = \hat{p}_{MZ}(i)$ et $\hat{P}_{MP}(Y = y_i) = \hat{p}_{MP}(i)$ fournies par les modèles en prédiction (en resubstitution). Les confronter avec $v_i = \hat{p}_{MZ}(i) - \hat{p}_{MP}(i)$

```
#modèle ZIP à évaluer
MZ <- zeroinfl(Affairs ~ Gender+RatingMarriage | Occupation+YearsMarried,data = D)

#modèle de Poisson équivalent sans la partie zero-inflation
MP <- glm(Affairs ~ Gender+RatingMarriage, data = D, family = "poisson")

#confrontation
vuong(MZ,MP)

## Vuong Non-Nested Hypothesis Test-Statistic:
## (test-statistic is asymptotically distributed N(0,1) under the
##  null that the models are indistinguishible)
## -------------------------------------------------------------
##              Vuong z-statistic             H_A  p-value
## Raw                  1.3800296 model1 > model2 0.083789
```

$$V = \sqrt{n}\frac{\bar{v}}{\sigma_v} \cong \mathcal{N}(0,1)$$

H0 : les deux modèles sont équivalents

V > 0, MZ a tendance à être meilleur que MP

V < 0, c'est l'inverse

MZ est légèrement meilleur, mais ne se démarque pas significativement pour un test à 5%

PRÉDICTION ET RÉSIDUS

Prédiction pour un individu

Calcul de la prédiction pour l'individu n°2, et de la probabilité d'attribution du bon compte $\mathbf{y_2=0}$

Coefficients Logistique		-4.457			1.161	-0.397
Coefficients Poisson		3.103	0.624	-0.650		
n°	Affairs	Constante	Gender	RatingMarriage	Occupation	YearsMarried
1	0	1	0	4	1	15
2	0	1	0	4	5	7

$$\hat{\lambda}_2 = \exp(-4.457 + 1.161 \times 5 - 0.397 \times 7) = 0.239 \Rightarrow \hat{\pi}_2 = \frac{\hat{\lambda}_2}{1+\hat{\lambda}_2} = 0.193$$

(page 6)

$$\hat{\mu}_2 = \exp(3.103 + 0.624 \times 0 - 0.650 \times 4) = 1.652$$

(page 6)

$$\hat{y}_2 = (1-\hat{\pi}_2)\ \hat{\mu}_2 = (1-0.193) \times 1.652 = 1.334$$

⇦ Prédiction de l'espérance de Y pour l'individu n°2 (page 5)

$$\hat{p}_{MZ}(2) = \hat{P}_{MZ}(Y=0) = \hat{\pi}_2 + (1-\hat{\pi}_2)\ e^{-\hat{\mu}_2} = 0.347$$

⇦ Probabilité d'obtenir la « bonne » valeur de Y (Y = 0) pour l'individu n°2 (page 5)

Remarque : $$\sum_{y=0,1,2,\cdots,\max(y)} \hat{P}_{MZ}(Y=y) \times y = \hat{y}_2 = 1.334$$

⇦ Moyenne pondérée de l'affectation aux valeurs de Y = estimation de l'espérance de Y pour l'individu n°2

Résidus

Les résidus permettent de diagnostiquer le modèle : identification des points atypiques, recherche des régularités. Bon modèle = résidus dispersés aléatoirement.

Résidu brut $$r_i = y_i - \hat{y}_i$$

Résidu de Pearson

(normalisé par l'écart-type de y)

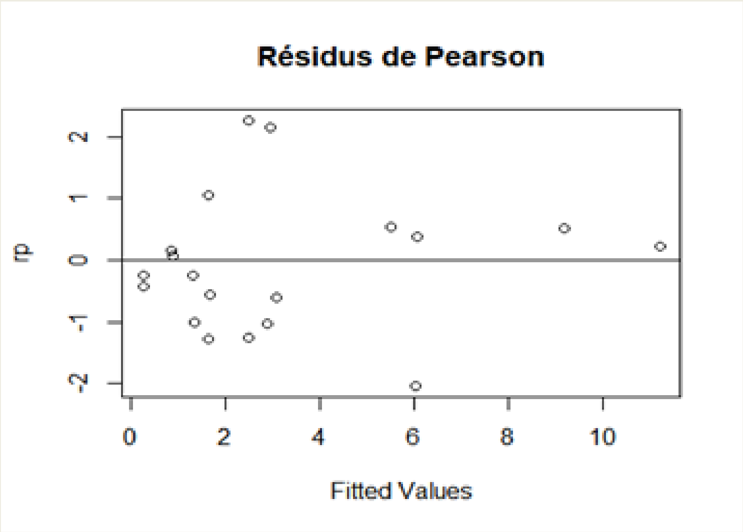
$$rp_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{\mu}_i(1-\hat{\pi}_i)(1+\hat{\mu}_i\hat{\pi}_i)}}$$

```
#résidus bruts
r <- MZ$residuals
print(r)

##          1          2          3          4          5          6
## -1.6517707 -1.3335920 -1.6844774 -2.5035108 -0.2764425 -0.2497220
##          7          8          9         10         11         12
## -5.0447962  0.1207892 -1.0736012  1.3492248  1.3643732  4.0491533
##         13         14         15         16         17         18
##  1.4924705  0.9422539  4.4964892  0.7943063  2.8104591 -1.8945478
##         19         20
## -0.3067043  0.1550322

#résidus de Pearson (résidus bruts normalisés)
rp <- residuals(MZ)

#graphique y^ vs. residus de Pearson
plot(MZ$fitted.values,rp,main="Résidus de Pearson",xlab="Fitted Values")
abline(h=0)
```



CONCLUSION

Régression ZIP - Conclusion

- Adaptée dès lors que la variable cible Y représente un comptage avec une surreprésentation de la valeur (Y = 0), conduisant à une surdispersion dans la régression de Poisson « classique ».

- Nous disposons d'un modèle à 2 étages : Binomial + Poisson

- L'arbitrage entre les variables à introduire dans le modèle Binomial et/ou celui de Poisson repose pour beaucoup sur la connaissance du domaine

- Procédure de sélection automatique de variables très compliquée à mettre en œuvre

- Le comptage peut s'effectuer sur une durée qui peut être différente d'un individu à l'autre (ex. nombre d'infidélité sur une durée variable selon les individus). On introduit alors le temps comme une variable offset() dans le modèle c.-à-d. une variable à laquelle on attribue d'emblée le coefficient 1. D'autres variables que la durée (temps) peuvent jour le rôle d'offset (ex. Y = comptage nombre de sinistres pour un foyer, offset = nombre d'objets assurés).

RÉFÉRENCES

Références

- Documentation NCSS Statistical Software, « Zero-Inflated Poisson Regression », Chapter 329.
- D. Gilles, « Notes on the Zero-Inflated Poisson Regression Model », University of Victoria, March 2010.
- UCLA Institute for Digital Research & Education, « Zero-Inflated Poisson Regresssio – R Data Analysis Examples ».
- S. Jackman et al., « Package ''pscl'' : Political Science Computational Laboratory », version 1.5.2 ; Reference manual ; Vignettes : « Regression Models for Count Data in R ».