

## Chapitre 4

# Modèles linéaires généralisés (GLM)

### 4.1 La famille exponentielle linéaire

Les GLM sont basés sur la famille exponentielle linéaire. Cette famille de distributions contient les lois qui peuvent être écrites sous la forme :

$$f_Y(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\}. \quad (4.1)$$

pour des fonctions  $a$ ,  $b$  et  $c$  spécifiques. Si  $\phi$  est **connu**, ce modèle est la **famille exponentielle** ; si  $\phi$  est inconnu (doit être estimé), ce modèle est plutôt appelé **famille exponentielle de dispersion**.

- $\theta$  est le paramètre canonique
- $\phi$  est le paramètre de dispersion
- $c$  ne dépend pas de  $\theta$
- le support de  $Y$  ne dépend pas de  $\theta$  ou  $\phi$ .

Note : la fonction  $a$  est souvent l'identité, i.e.  $a(\phi) = \phi$ .

**Lemme 4.1.** Si  $l(\theta, \phi; y) = \ln f_Y(y; \theta, \phi)$ , alors on a (sous certaines conditions de régularité)

$$E_{f_Y}[\dot{l}_\theta(\theta, \phi; Y)] = 0,$$

où  $\dot{l}_\theta(\theta, \phi; Y) = \frac{\partial}{\partial \theta} l(\theta, \phi; y)$ , et

$$E_{f_Y}[\{\dot{l}_\theta(\theta, \phi; Y)\}^2] = -E_{f_Y}[\ddot{l}_{\theta\theta}(\theta, \phi; Y)],$$

où  $\ddot{l}_{\theta\theta}(\theta, \phi; Y) = \frac{\partial^2}{\partial \theta^2} l(\theta, \phi; y)$ .

Si  $Y$  fait partie de la famille exponentielle, on trouve donc que

$$E[Y] = \dot{b}(\theta) \quad \text{Var}(Y) = a(\phi)\ddot{b}(\theta).$$

On note parfois que  $\mu = \dot{b}(\theta)$  est la fonction de moyenne, alors que  $V(\mu) = \ddot{b}(\theta)$  est la fonction de variance. Cela souligne qu'il y a un lien explicite entre la moyenne et la variance de  $Y$ .

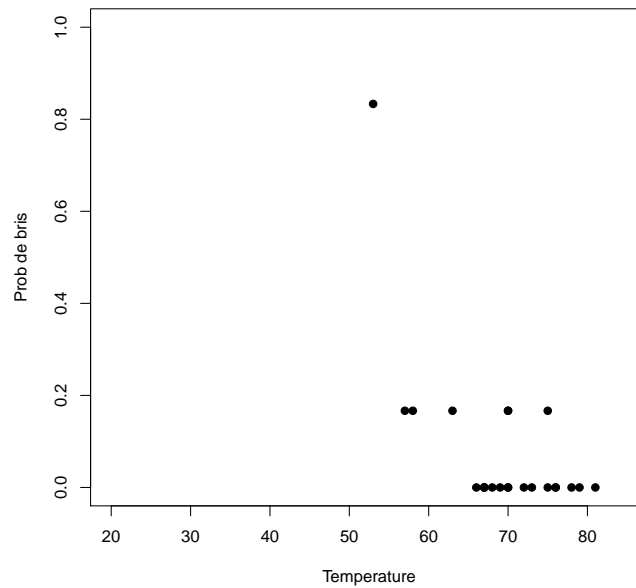
Quelques exemples de distributions faisant partie de la famille exponentielle :

- Normale
- Gamma
- Inverse-Gaussienne
- Poisson
- Binomiale( $n, \pi$ ) avec  $n$  connu
- Binomiale Négative( $r, \pi$ ) avec  $r$  connu
- Multinomiale( $n, \pi$ ) avec  $n$  connu

**Exemple 4.2.** La modélisation des risques n'est pas seulement importante pour les actuaires ! Les données suivantes sont le nombre de 'O-rings' brisés, sur six, lors de 23 lancements de la navette spatiale Challenger. La température au sol lors de ces lancements variait entre 53 et 81 degrés Fahrenheit. On s'intéresse à modéliser la probabilité de bris simultané des six 'O-rings' à différentes températures pour comprendre le risque.

Température	53	57	58	63	66	67	67	67	68	69	70	70	70	70	72	73	75	75	76	76	78	79	81
# brisés	5	1	1	1	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0

Le graphique suivant montre la proportion de bris pour les différentes températures :



On considère trois modèles différents pour ces données binomiales. Si  $Y$  est la proportion de bris et  $T$  est la température, on ajuste

- un modèle linéaire :  $E[Y_i] = \beta_0 + \beta_1 T_i$ ,
- un GLM avec lien logistique :  $E[Y_i] = \text{expit}(\beta_0 + \beta_1 T_i) = \frac{e^{\beta_0 + \beta_1 T_i}}{1 + e^{\beta_0 + \beta_1 T_i}}$ ,
- un GLM avec lien probit :  $E[Y_i] = \Phi(\beta_0 + \beta_1 T_i)$ , où  $\Phi(t)$  est la fonction de répartition d'une loi  $N(0,1)$ .

Lorsque la navette Challenger s'est désintégrée le 28 janvier 1986, au cours du décollage, la température était de 31F. Est-ce que les analystes auraient pu prévenir ce désastre à l'aide des GLMs ? Les paramètres sont estimés avec **R** et les fonctions **lm** et **glm**. Cet exemple est tiré du livre de Faraway (2005).

**Code pour l'exemple**

```

library(faraway)
data(orings)

## Graphique
plot(damage/6~temp, orings, xlim=c(20,85), ylim = c(0,1),xlab="Temperature", ylab="Prob de bris",pch=19)

## Modèle linéaire
lmod <- lm(damage/6~temp, orings)
abline(lmod,lwd=2)

## GLM avec lien logistique
logitmod <- glm(cbind(damage,6-damage)~temp, family=binomial, orings)
x <- seq(20,85,1)
lines(x,ilogit(coef(logitmod)[1]+coef(logitmod)[2]*x),lty=2,lwd=2)

## GLM avec lien probit
probitmod <- glm(cbind(damage,6-damage)~temp,family=binomial(link=probit), orings)
lines(x,pnorm(coef(probitmod)[1]+coef(probitmod)[2]*x),lty=3,lwd=2)

legend(20,0.4,c("Linear","Logit","Probit"),lty=c(1:3),lwd=rep(2,3))

## résultats
summary(lmod)
summary(logitmod)
summary(probitmod)

```

**Résultats pour le modèle linéaire :**

```

> summary(lmod)

Call:
lm(formula = damage/6 ~ temp, data = orings)

Residuals:
    Min       1Q   Median       3Q      Max
-0.13786 -0.10345 -0.02369  0.06601  0.48345

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.21429    0.29993   4.049 0.000578 ***
temp        -0.01631    0.00429  -3.801 0.001043 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

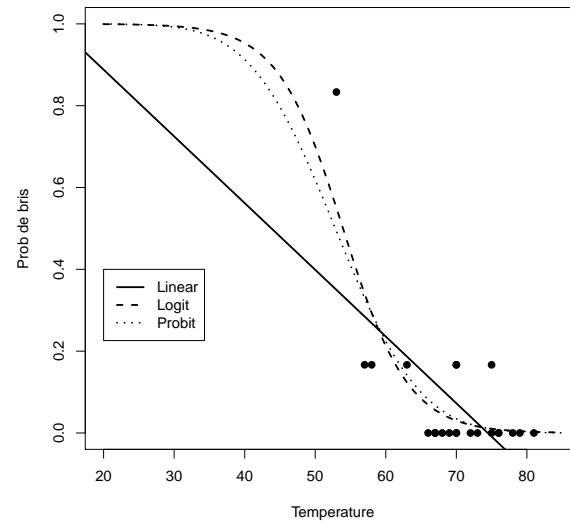


FIGURE 4.1: Ajustement des modèles linéaire, logistique et probit

Residual standard error: 0.142 on 21 degrees of freedom  
 Multiple R-squared: 0.4076, Adjusted R-squared: 0.3794  
 F-statistic: 14.45 on 1 and 21 DF, p-value: 0.001043

Résultats pour le GLM logistique :

```
> summary(logitmod)
```

Call:

```
glm(formula = cbind(damage, 6 - damage) ~ temp, family = binomial,  
     data = orings)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.9529	-0.7345	-0.4393	-0.2079	1.9565

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	11.66299	3.29626	3.538	0.000403 ***
temp	-0.21623	0.05318	-4.066	4.78e-05 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 38.898 on 22 degrees of freedom  
 Residual deviance: 16.912 on 21 degrees of freedom  
 AIC: 33.675

Number of Fisher Scoring iterations: 6

### Résultats pour le GLM probit :

```
> summary(probitmod)
```

Call:

```
glm(formula = cbind(damage, 6 - damage) ~ temp, family = binomial(link = probit),
     data = orings)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.0134	-0.7761	-0.4467	-0.1581	1.9983

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	5.59145	1.71055	3.269	0.00108 **
temp	-0.10580	0.02656	-3.984	6.79e-05 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 38.898 on 22 degrees of freedom  
 Residual deviance: 18.131 on 21 degrees of freedom  
 AIC: 34.893

Number of Fisher Scoring iterations: 6

### Comparaison de l'ajustement si T=31

```
> (coef(lmod)[1]+coef(lmod)[2]*31)
0.7086905
> ilogit(coef(logitmod)[1]+coef(logitmod)[2]*31)
0.9930342
> pnorm(coef(probitmod)[1]+coef(probitmod)[2]*31)
0.9895983
```

□

## 4.2 Fonctions de lien

On modélise  $\mu_i = E[Y_i | \mathbf{X}_i = \mathbf{x}_i]$  en posant  $\mu_i = g^{-1}(\eta_i)$  (ou  $\eta_i = g(\mu_i)$ ), où  $g$  est une fonction de lien et  $\eta_i = \mathbf{x}_i \boldsymbol{\beta}$  est un prédicteur linéaire. Différentes fonctions de lien peuvent être envisagées :

**Lien identité :**  $\eta = \mu$ .

**Lien logistique :**  $\eta = \ln\left(\frac{\mu}{1-\mu}\right)$ . L'inverse du lien logistique est la fonction **expit** :  $g^{-1}(t) = \frac{e^t}{1+e^t}$ .

**Lien probit :**  $\eta = \Phi^{-1}(\mu)$ , où  $\Phi^{-1}$  est le quantile d'une loi Normale (0,1). Dans ce cas, on trouve  $\mu = \Phi(\eta)$ .

**Lien log-log complémentaire :**  $\eta = \ln(-\ln(1 - \mu))$ , alors  $\mu = 1 - e^{-e^\eta}$ .

**Lien log :**  $\eta = \log(\mu)$ , alors  $\mu = e^\eta$ .

**Lien inverse :**  $\eta = 1/\mu$ .

Le choix de la fonction de lien est un choix de modélisation, et ce choix n'est pas toujours évident.

## 4.3 Estimation

On utilise la méthode du maximum de vraisemblance pour l'inférence statistique pour les GLMs. En général,  $\hat{\beta}_n$  tel que  $\dot{l}_\theta(\hat{\beta}_n; y_1, \dots, y_n) = 0_{p \times 1}$  doit être trouvé de façon numérique.

L'estimateur du maximum de vraisemblance a plusieurs propriétés désirables. Soit  $\boldsymbol{\beta}$  la vraie valeur du paramètre.

**Cohérence :** Lorsque  $n \rightarrow \infty$ ,  $\hat{\beta}_n \rightarrow \boldsymbol{\beta}$ .

**Normalité asymptotique :** Lorsque  $n \rightarrow \infty$ ,

$$\hat{\beta}_n \sim N\left(\boldsymbol{\beta}, \frac{\mathcal{I}(\boldsymbol{\beta})^{-1}}{n}\right),$$

où  $\mathcal{I}(\boldsymbol{\beta})$  est la matrice d'information de Fisher :

$$\mathcal{I}(\boldsymbol{\beta}) = E[\dot{l}(\boldsymbol{\beta}; Y_1, \dots, Y_n) \dot{l}(\boldsymbol{\beta}; Y_1, \dots, Y_n)^T],$$

qui est aussi égale à  $-E[\ddot{l}(\boldsymbol{\beta}; Y_1, \dots, Y_n)]$  sous certaines conditions de régularité. En pratique, on ne connaît pas la vraie valeur de  $\boldsymbol{\beta}$ , alors on utilise **l'information observée** :

$$\mathcal{I}(\hat{\beta}_n) = \begin{cases} \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}; y_i) \frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}; y_i)^T \\ - \sum_{i=1}^n \frac{\partial^2}{\partial \boldsymbol{\beta}^2} l(\boldsymbol{\beta}; y_i) \end{cases} \Big|_{\boldsymbol{\beta}=\hat{\beta}_n} \quad (4.2)$$

#### 4.4 Algorithme de Newton-Raphson et 'Fisher Scoring'

Puisqu'on doit évaluer numériquement l'estimateur du maximum de vraisemblance, on va étudier les principaux algorithmes utilisés.

On considère une approximation de Taylor de premier ordre de la fonction de score  $\frac{\partial}{\partial \beta} \ell(\beta) = \dot{\ell}(\beta; y_1, \dots, y_n)$ , autour d'un point arbitraire  $\beta^{(0)}$  :

$$\dot{\ell}(\beta) \approx \dot{\ell}(\beta^{(0)}) + \ddot{\ell}(\beta^{(0)})(\beta - \beta^{(0)}).$$

On réarrange :

$$\beta \approx \beta^{(0)} + \{\ddot{\ell}(\beta^{(0)})\}^{-1}[\dot{\ell}(\beta) - \dot{\ell}(\beta^{(0)})].$$

Si on remplace  $\beta$  par l'estimateur du MV  $\hat{\beta}$ , on a que  $\dot{\ell}(\hat{\beta}) = \mathbf{0}_p$ . Alors,

$$\hat{\beta} \approx \beta^{(0)} + \{-\ddot{\ell}(\beta^{(0)})\}^{-1}\dot{\ell}(\beta^{(0)}).$$

Cette approximation permet de développer une méthode itérative pour évaluer  $\hat{\beta}$ .

##### Algorithme de Newton-Raphson :

1. Initialiser  $\beta^{(0)}$  à une valeur quelconque  $\hat{\beta}^{(0)}$ .
2. Pour  $m = 1, 2, \dots$ , définir

$$\hat{\beta}^{(m)} = \hat{\beta}^{(m-1)} + \{-\ddot{\ell}(\hat{\beta}^{(m-1)})\}^{-1}\dot{\ell}(\hat{\beta}^{(m-1)}).$$

3. Continuer jusqu'à convergence de  $\{\hat{\beta}^{(m)}\}$ , soit lorsque

$$|\dot{\ell}(\hat{\beta}^{(m)})| < \varepsilon,$$

pour un seuil de tolérance  $\varepsilon$  près de 0.

Une variante de cet algorithme consiste en remplaçant  $-\ddot{\ell}(\beta)$  par son espérance :

$$\mathcal{I}(\beta) = -E[\ddot{\ell}(\beta)].$$

Il faut cependant utiliser une des expressions de l'équation 4.2 pour estimer la matrice d'Information de Fisher. Cette méthode se nomme **Fisher Scoring** et est utilisée par défaut par la fonction `glm` en R.

Note : Si le lien canonique est utilisé, l'information observée (pour l'algorithme Newton-Raphson) et l'espérance de l'information (pour le Fisher scoring) sont égales.

**Remarque 4.3.** *On peut aussi utiliser la méthode itérative des moindres carrés repondérés (Iteratively reweighted least-squares), mais nous n'aborderons pas cette méthode dans le cours. L'idée est de reformuler le problème d'optimisation en termes d'un problème de la forme bien connue  $\hat{\beta} = (X^T W X)^{-1} X^T W Y$ .*



## 4.5 Tests d'hypothèses

Les éléments sur la diagonale de l'inverse de l'information observée  $\{\hat{\mathcal{I}}(\hat{\beta})\}^{-1}$  sont les variances estimées pour les éléments de  $\hat{\beta}$ , ce qui permet de créer un test statistique. En fait, pour tester

$$H_0 : \beta_j = 0 \text{ vs } H_1 : \beta_j \neq 0,$$

où  $\beta_j$  est le  $j^e$  élément du vecteur  $\beta$ , on utilise la **statistique de Wald** :

$$\frac{\hat{\beta}_j}{\sqrt{\text{Var}(\hat{\beta}_j)}},$$

qui suit une  $N(0,1)$  sous l'hypothèse nulle. On peut donc calculer la valeur  $p$  du test et conclure si une variable exogène est nécessaire dans le modèle ou non. La statistique de Wald et la valeur  $p$  sont obtenues directement dans le **summary** d'un modèle en R.

Plus généralement, pour comparer deux hypothèses

$$H_0 : \beta \in B_0 \text{ vs } H_1 : \beta \in B_1,$$

on utilise le **test du rapport de vraisemblance** (TRV). La statistique

$$\lambda(y) = \frac{\sup_{\beta \in B_0} L_n(\beta)}{\sup_{\beta \in B_0 \cup B_1} L_n(\beta)} = \frac{L_n(\hat{\beta}^{(H_0)})}{L_n(\hat{\beta})}.$$

On remarque que, en tout temps,  $\lambda(y) \leq 1$ . Si  $\lambda(y) \leq$  valeur critique (inférieure à 1), alors on rejette l'hypothèse nulle en faveur de l'alternative.

Si  $H_0$  spécifie *entièrement* les paramètres du modèle, on peut montrer que

$$-2 \ln \lambda(Y) \approx \chi_p^2 \text{ sous } H_0,$$

où  $p$  est le nombre de paramètres.

Si  $H_0$  spécifie *partiellement* les paramètres du modèle, on doit calculer la différence entre le nombre de degrés de liberté. On suppose qu'il y a  $k_1$  paramètres non spécifiés dans l'hypothèse  $H_0$  et  $k_2$  paramètres non spécifiés dans  $H_1$ , avec  $k_2 > k_1$ . Dans ce cas, on peut montrer que

$$-2 \ln \lambda(Y) \approx \chi_{k_2 - k_1}^2 \text{ sous } H_0.$$

Le TRV est conçu pour comparer des modèles qui sont **liés**. Par exemple, il est possible de tester

$$H_0 : \eta_i = \beta_0 + \beta_1 x_i \text{ vs } H_1 : \eta_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2,$$

parce que le modèle de  $H_0$  est 'inclus' dans le modèle de  $H_1$ . Toutefois, on ne pourrait pas tester directement

$$H_0 : \eta_i = \beta_0 + \beta_1 x_{1,i} \text{ vs } H_1 : \eta_i = \beta_0 + \beta_2 x_{2,i}$$

parce que ces modèles ne sont pas liés.

## 4.6 Adéquation du modèle

### 4.6.1 $X^2$ de Pearson

La statistique du  $X^2$  de Pearson est une mesure de l'ajustement ou de l'adéquation du modèle. Si  $n$  est grand, on a

$$X^2 = \sum_{i=1}^n \left( \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}} \right)^2 \approx \chi_{n-p'}^2,$$

où  $V(\hat{\mu}_i) = \ddot{b}(\hat{\theta}_i)$  est la relation entre la moyenne et la variance de  $Y_i$ . Par défaut, l'estimateur de  $\phi$  en  $\mathbf{R}$  est

$$\hat{\phi} = \frac{X^2}{n - p'}.$$

### 4.6.2 Déviance

Le modèle le plus simple que l'on peut imaginer pour  $E[Y_i|X_i = x_i] = \mu_i$  est appelé le **modèle nul** :

$$\hat{\mu}_i = \hat{\mu}_0, i = 1, \dots, n.$$

Il n'y a donc pas de dépendance entre  $y_i$  et  $x_i$ . D'un autre côté, on a aussi le modèle le plus complexe, i.e. le **modèle complet**, où chaque prévision est exactement égale à la donnée :

$$\hat{\mu}_i = y_i, i = 1, \dots, n.$$

Pour tous les autres modèles, on a

$$\hat{\mu}_i = g^{-1}(\mathbf{x}_i \hat{\beta}), i = 1, \dots, n,$$

où  $g^{-1}$  est l'inverse de la fonction de lien.

Soit les vecteurs  $n \times 1$  de paramètres canoniques suivants :

- $\bar{\theta}$  développé sous le modèle nul,
- $\tilde{\theta}$  développé sous le modèle complet,
- $\hat{\theta}$  développé sous le modèle avec une fonction de lien.

Pour la famille exponentielle de dispersion (avec pondérations, pour inclure le cas binomial), on a

$$\begin{aligned} 2(\ell(\tilde{\theta}) - \ell(\hat{\theta})) &= 2 \sum_{i=1}^n \frac{w_i}{\phi} \left( y_i \tilde{\theta} - b(\tilde{\theta}) - [y_i \hat{\theta} - b(\hat{\theta})] \right) \\ &= 2 \sum_{i=1}^n \frac{w_i}{\phi} \left( y_i (\tilde{\theta} - \hat{\theta}) - [b(\tilde{\theta}) - b(\hat{\theta})] \right) \\ &= \frac{D(y; \hat{\mu})}{\phi}. \end{aligned}$$

$D(y; \hat{\mu})$  est appelée la **déviante** du modèle qui définit  $\hat{\mu}$ . On appelle  $D^*(y; \hat{\mu}) = \frac{D(y; \hat{\mu})}{\phi}$  la déviante réduite (*scaled deviance*).

La déviante mesure la différence de l'ajustement entre le modèle complet et le modèle choisi. Selon la théorie du TRV, on a

$$2(l_n(\tilde{\theta}) - l_n(\hat{\theta})) \sim \chi_{n-p}^2,$$

si le modèle représenté par  $\hat{\theta}$  (avec  $p$  paramètres) est adéquat. L'espérance de la déviante réduite  $\frac{D(y; \hat{\mu})}{\phi}$  est approximativement égale à  $n - p$  si le modèle est adéquat.

### 4.6.3 Critère AIC

Le critère AIC peut aussi être utilisé pour comparer des modèles linéaires généralisés liés.

## 4.7 Comparaison de modèles

On suppose que  $\phi$  est connu et que l'on ajuste deux modèles liés. On rappelle que deux modèles  $m_A$  (avec  $p_A$  paramètres) et  $m_B$  (avec  $p_B$  paramètres,  $p_B > p_A$ ) sont liés si  $m_A$  peut être obtenu en imposant des contraintes sur les paramètres de  $m_B$ . Alors, si  $m_A$  est une simplification adéquate du modèle  $m_B$ ,

$$\frac{D(y; \hat{\mu}_A) - D(y; \hat{\mu}_B)}{\phi} = 2(l_n(\hat{\theta}_B) - l_n(\hat{\theta}_A)) \sim \chi_{p_B - p_A}^2.$$

**Exemple 4.4.** On a les données suivantes pour des contrats d'assurance automobile pendant l'année 2012 :

Sexe	Âge	Nombre de réclamations
H	17	2
F	23	0
H	45	0
F	65	0
F	31	1
H	45	0
F	28	1
H	54	0
F	48	0
H	37	1

Ces données sont représentées dans le graphique 4.2. On modélise le nombre de réclamations avec une loi de Poisson et un lien logarithmique. La variable explicative **Âge** est continue, alors que la variable explicative **sexe** est discrète (le niveau de base est H). On s'intéresse aussi à l'effet de l'interaction entre l'âge et le sexe.

**Code pour l'exemple**

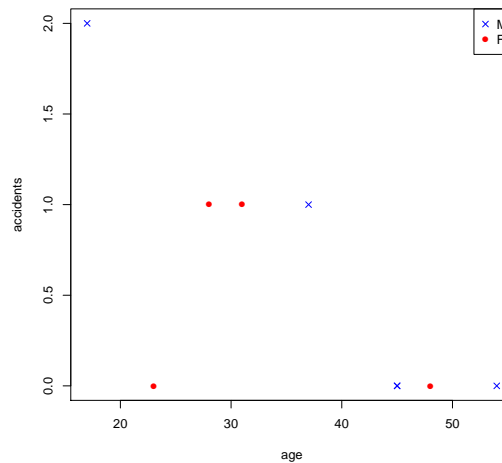


FIGURE 4.2: Nuage de points des données de l'exemple 4.4

```
> ## Données: nombre de réclamations pendant l'année de contrat
> accidents <- c(2,0,0,0,1,0,1,0,0,1)
> age <- c(17,23,45,65,31,45,28,54,48,37)
>
> ## variable discrète, on utilise la fonction "factor"
> sex <- factor(c("H","F","H","F","F","H","F","H","F","H"),c("H","F"))
>
> ## glm ajuste le modèle poisson avec le lien canonique log.
> mod <- glm(accidents~age*sex,family=poisson)
>
> ## Résultats
> summary(mod)
```

```
Call:
glm(formula = accidents ~ age * sex, family = poisson)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2311  -0.5879  -0.3673   0.3639   0.8684
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.32703     1.32384   1.758   0.0788 .
```

```

age          -0.09031    0.05034   -1.794    0.0728 .
sexF         -1.19293    2.75662   -0.433    0.6652
age:sexF      0.02894    0.09315    0.311    0.7560
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for poisson family taken to be 1)

```

Null deviance: 9.7041 on 9 degrees of freedom
Residual deviance: 4.3399 on 6 degrees of freedom
AIC: 20.954

```

Number of Fisher Scoring iterations: 5

```

> ## Analyse de la déviance
> anova(mod)
Analysis of Deviance Table

```

Model: poisson, link: log

Response: accidents

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev
NULL			9	9.7041
age	1	5.0926	8	4.6114
sex	1	0.1839	7	4.4276
age:sex	1	0.0877	6	4.3399

```

> qchisq(0.95,1)
[1] 3.841459

```

```

>
> ## On peut aussi procéder comme suit.
> mod0 <- glm(accidents~1,family=poisson)
> mod1 <- glm(accidents~sex,family=poisson)
> mod2 <- glm(accidents~age,family=poisson)
> anova(mod0,mod1,mod2,mod)
Analysis of Deviance Table

```

```

Model 1: accidents ~ 1
Model 2: accidents ~ sex
Model 3: accidents ~ age
Model 4: accidents ~ age * sex
      Resid. Df Resid. Dev Df Deviance

```

```

1      9      9.7041
2      8      9.5027  1    0.2014
3      8      4.6114  0    4.8913
4      6      4.3399  2    0.2716
>
> summary(mod2)

Call:
glm(formula = accidents ~ age, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4692  -0.5487  -0.3104   0.2865   0.9574

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.07084    1.25930   1.644   0.1001
age          -0.08672    0.04528  -1.915   0.0555 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 9.7041  on 9  degrees of freedom
Residual deviance: 4.6114  on 8  degrees of freedom
AIC: 17.225

Number of Fisher Scoring iterations: 5

```

On trouve donc que le modèle contenant seulement la variable explicative  $\hat{\text{Age}}$  est une simplification adéquate des autres modèles. Toutefois, on ne peut pas dire que ce modèle soit bien ajusté !  $\square$

## 4.8 Résidus

Il y a trois types de résidus qui peuvent être utilisés pour vérifier l'ajuster d'un GLM :

**Résidus de Pearson :**

$$r_{Pi} = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$$

Ces résidus ont une moyenne de zéro et une variance constante si l'ajustement est adéquat. Toutefois, la distribution de ces résidus peut être très asymétrique.

**Résidus d'Anscombe :** On transforme  $Y \rightarrow A(Y)$  afin que la distribution de  $A(Y)$  soit la moins asymé-

trique possible (approximativement normale). La transformation

$$A(t) = \int_{-\infty}^t \frac{ds}{V^{1/3}(s)}$$

est appropriée pour la famille exponentielle (de dispersion), voir McCullagh et Nelder (1989) pour les détails. On a donc

$$r_{Ai} = \frac{A(y_i) - A(\hat{\mu}_i)}{\dot{A}(\hat{\mu}_i)\sqrt{V(\hat{\mu}_i)}}.$$

**Résidus de déviance :**

$$r_{Di} = \text{signe}(y_i - \hat{\mu}_i)\sqrt{d_i},$$

où

$$\text{signe}(t) = \begin{cases} 1 & t > 0 \\ -1 & t < 0 \end{cases},$$

et

$$d_i = 2w_i(y_i(\tilde{\theta}_i - \hat{\theta}_i) - [b(\tilde{\theta}_i) - b(\hat{\theta}_i)])$$

est la contribution de la donnée  $i$  dans la déviance :  $D(y; \hat{\theta}) = \frac{1}{\phi} \sum_{i=1}^n d_i$ .

**Exemple 4.5.** Si  $Y_i \sim \text{Poisson}(\mu_i)$ , alors  $V(\mu_i) = \mu_i$ . On trouve :

**Résidus de Pearson**

$$r_{Pi} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}}$$

**Résidus d'Anscombe**

$$A(t) = \int_0^t \frac{d\mu}{\mu^{1/3}} = \frac{3}{2}t^{2/3},$$

et

$$r_{Ai} = \frac{3(y_i^{2/3} - \hat{\mu}_i^{2/3})}{2\hat{\mu}_i^{1/6}}$$

**Résidus de déviance**

$$r_{Di} = \text{signe}(y_i - \hat{\mu}_i)\sqrt{2 \left[ y_i \ln \left( \frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right]}$$

**Code pour graphique de résidus :** On considère un modèle de Poisson avec lien logarithmique :

$$\mu_i = e^{1+0.02x_i}.$$

On compare les résidus pour le modèle nul avec ceux du modèle incluant la variable explicative  $x$ .

```
x <- round(runif(200,0,100),0)
y <- rpois(200,exp(1+0.02*x))

m0 <- glm(y~1,family=poisson)
m1 <- glm(y~x,family=poisson)

pearson0 <- residuals(m0,type="pearson")
dev0 <- residuals(m0,type="deviance")
hatmu0 <- fitted(m0)
anscombe0 <- 1.5*(y^(2/3)-hatmu0^(2/3))/hatmu0^(1/6)

pearson1 <- residuals(m1,type="pearson")
dev1 <- residuals(m1,type="deviance")
hatmu1 <- fitted(m1)
anscombe1 <- 1.5*(y^(2/3)-hatmu1^(2/3))/hatmu1^(1/6)
```



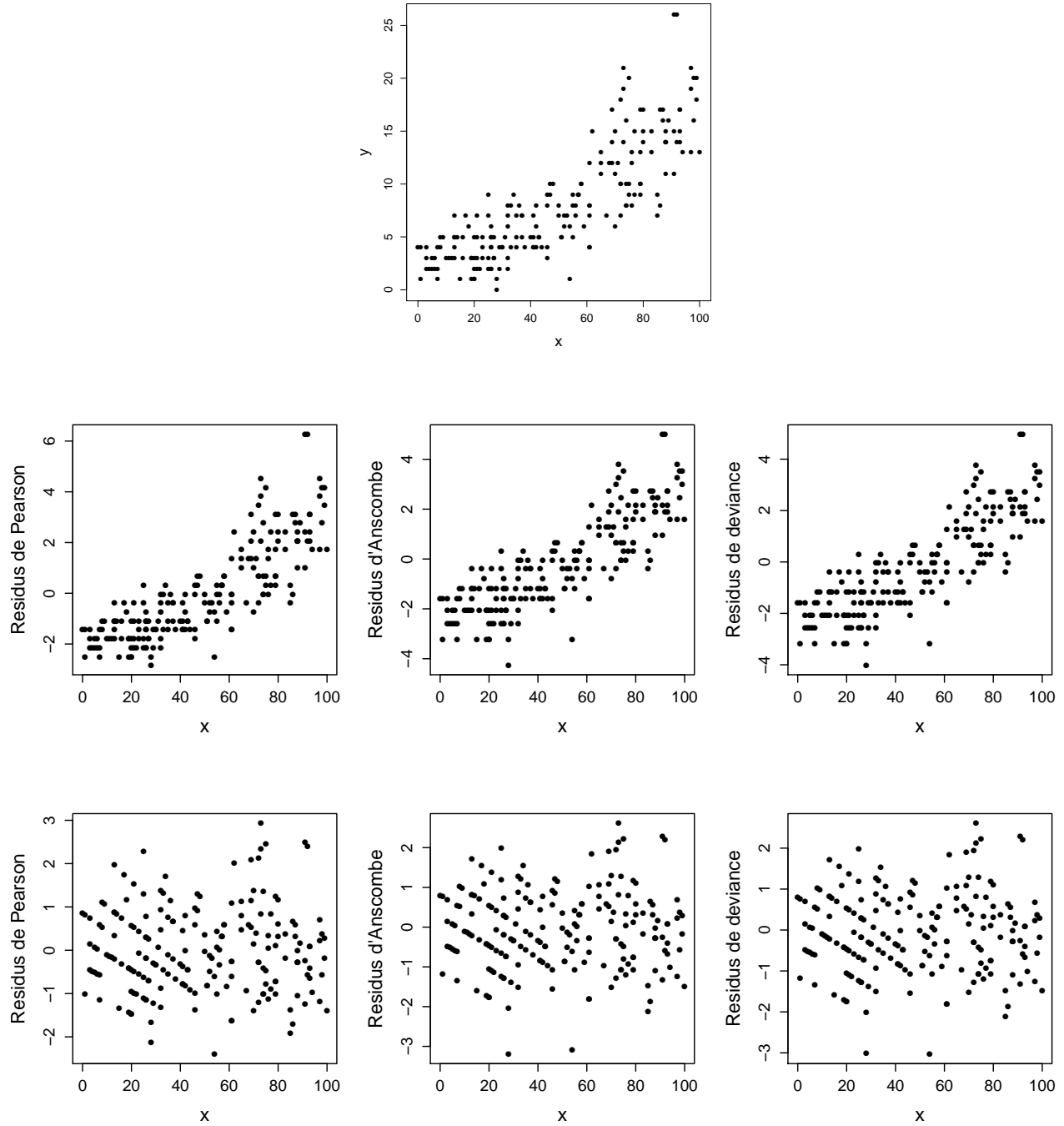


FIGURE 4.3: Données Poisson en fonction de la variable explicative  $x$  (graphique du haut); Résidus de l'ajustement du modèle nul (première ligne) et du modèle avec une variable explicative (deuxième ligne)

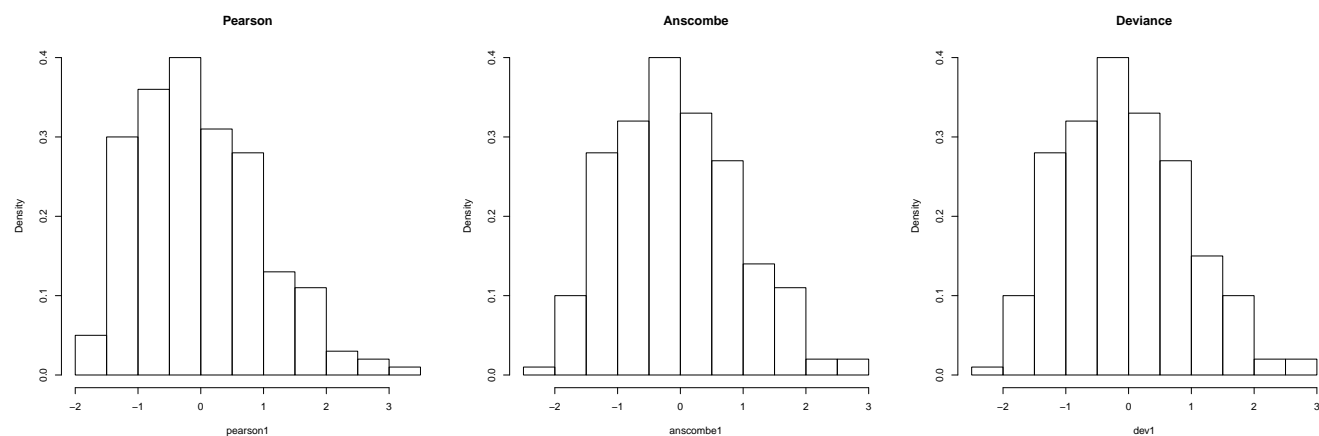


FIGURE 4.4: Histogrammes des résidus du modèle avec une variable explicative

# Bibliographie

AGRESTI, A. (2013). *Categorical Data Analysis. Third Edition*. Wiley Series in Probability and Statistics. Wiley, New Jersey, USA.

FARAWAY, J. J. (2005). *Extending the linear model with R : generalized linear, mixed effects and nonparametric regression models*. CRC press., [http ://www.maths.bath.ac.uk/~jjf23/ELM/](http://www.maths.bath.ac.uk/~jjf23/ELM/).

MCCULLAGH, P. et NELDER, J. A. (1989). *Generalized linear models (Monographs on statistics and applied probability 37)*. Chapman Hall, London.