

Régression logistique : introduction



- Expliquer un risque suicidaire élevé en prison par

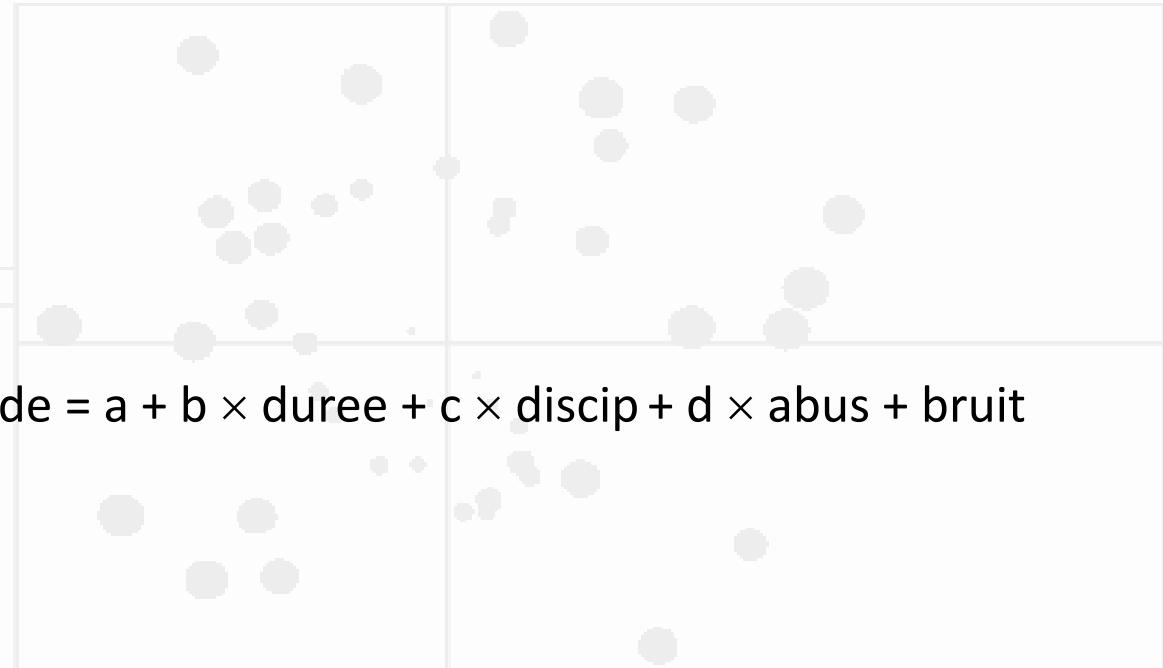
- La durée de la peine
- L'existence de mesures disciplinaire
- Des antécédents d'abus dans l'enfance

Les limites de la régression linéaire

Introduction à la statistique avec R > Régression logistique : introduction



- Haut risque de suicide = $a + b \times \text{duree} + c \times \text{discip} + d \times \text{abus} + \text{bruit}$



Les limites de la régression linéaire

Introduction à la statistique avec R > Régression logistique : introduction



- Haut risque de suicide = $a + b \times \text{duree} + c \times \text{discip} + d \times \text{abus} + \text{bruit}$
- La distribution du « bruit » est normale, donc
« $a + b \times \text{duree} + c \times \text{discip} + d \times \text{abus} + \text{bruit}$ » varie entre plus et moins l'infini, or la variable « haut risque de suicide » est binaire...

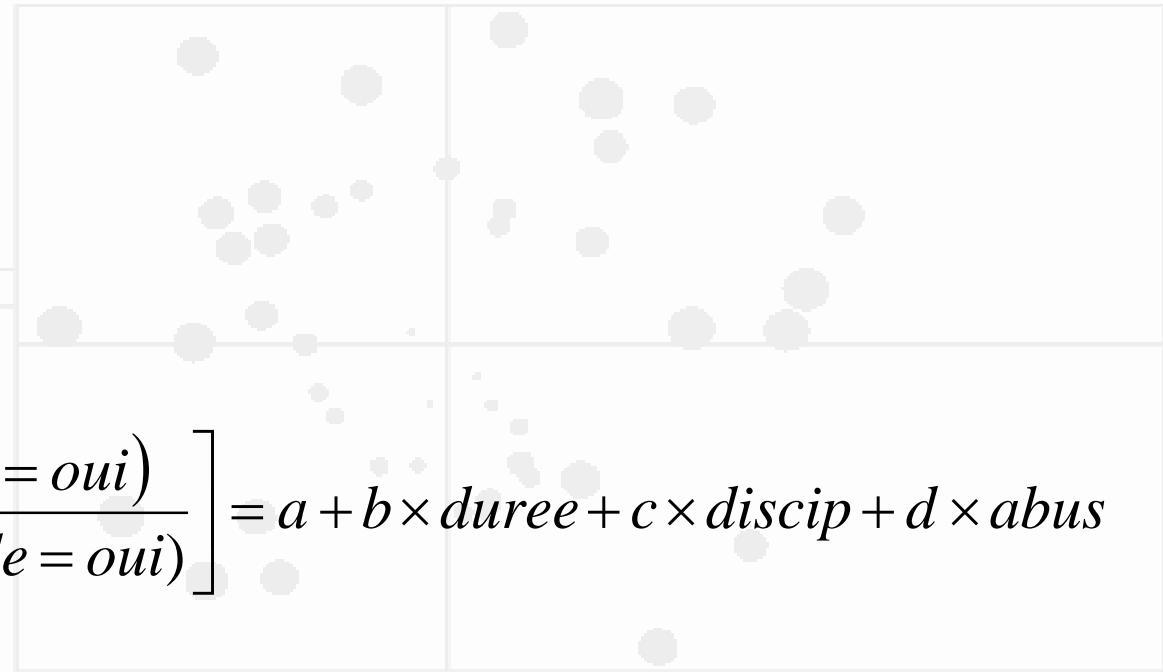
Les limites de la régression linéaire



Introduction à la statistique avec R > Régression logistique : introduction



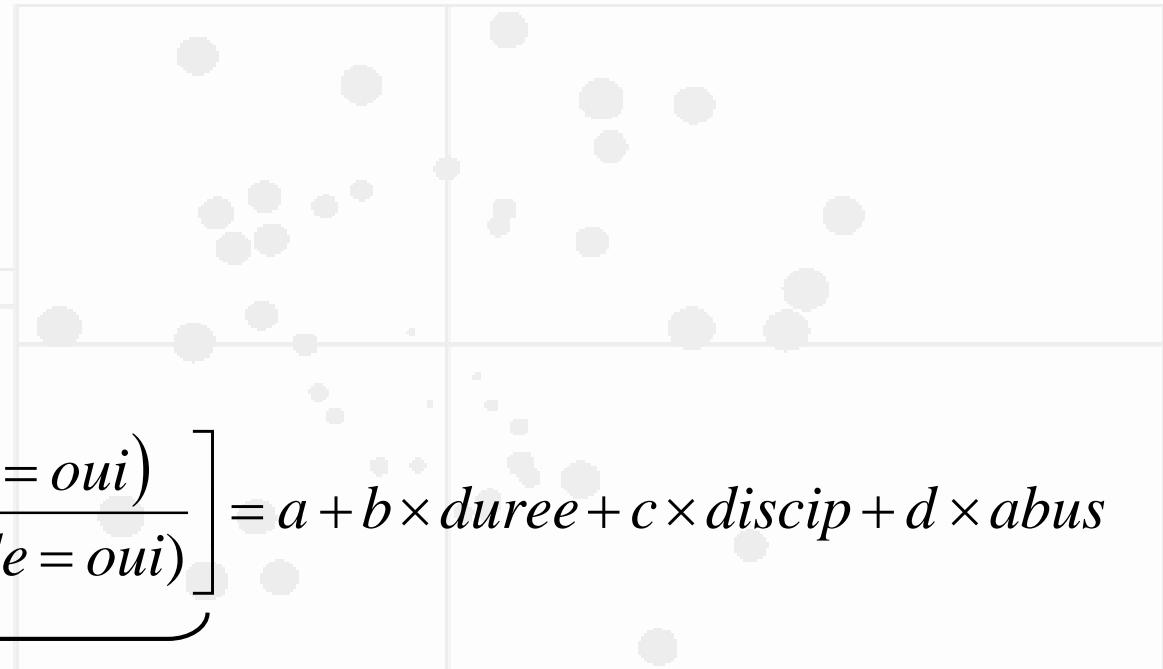
$$\text{Log} \left[\frac{\text{prob}(HR \text{ suicide} = \text{oui})}{1 - \text{prob}(HR \text{ suicide} = \text{oui})} \right] = a + b \times \text{duree} + c \times \text{discip} + d \times \text{abus}$$



Les limites de la régression linéaire



Introduction à la statistique avec R > Régression logistique : introduction



Varie entre + et - l'infini

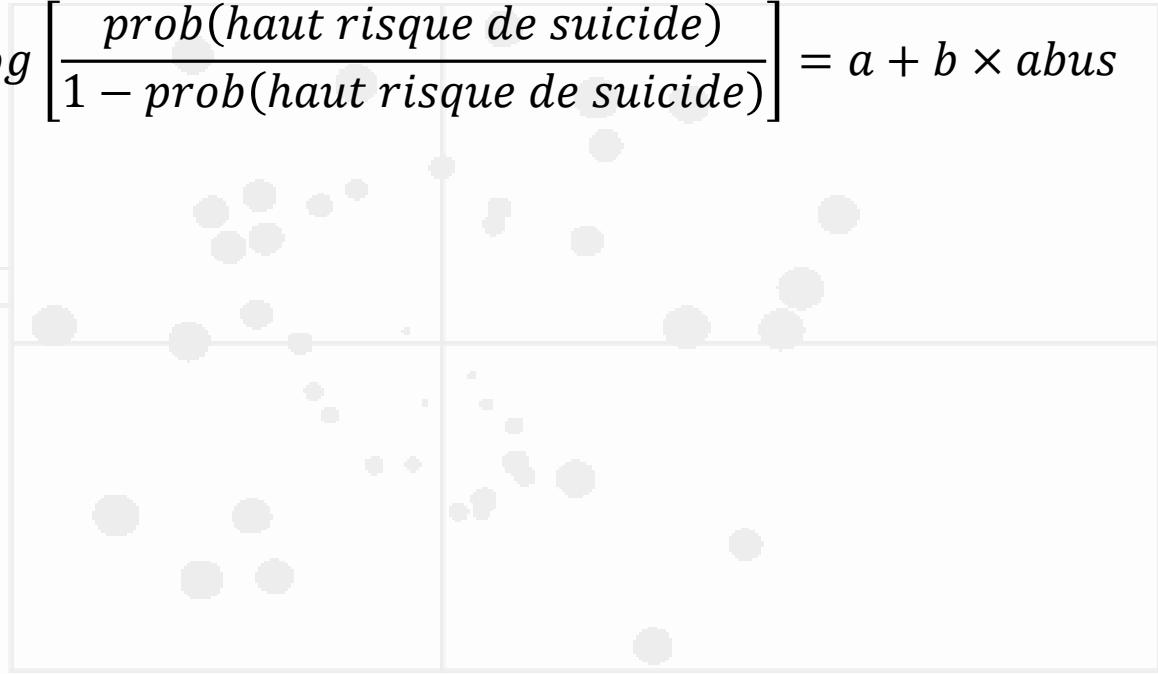


Une seule variable explicative



Introduction à la statistique avec R > Régression logistique : introduction

$$\text{Log} \left[\frac{\text{prob(haut risque de suicide)}}{1 - \text{prob(haut risque de suicide)}} \right] = a + b \times \text{abus}$$



Une seule variable explicative

Introduction à la statistique avec R > Régression logistique : introduction

$$\text{Log} \left[\frac{\text{prob(haut risque de suicide)}}{1 - \text{prob(haut risque de suicide)}} \right] = a + b \times \text{abus}$$

```
> mod1 <- glm(suicide.hr~abus, data=smp.1, family="binomial")
```



Une seule variable explicative



Introduction à la statistique avec R > Régression logistique : introduction

$$\text{Log} \left[\frac{\text{prob(haut risque de suicide)}}{1 - \text{prob(haut risque de suicide)}} \right] = a + b \times \text{abus}$$

```
> mod1 <- glm(suicide.hr~abus, data=smp.1, family="binomial")
> summary(mod1)

Call:
glm(formula = suicide.hr ~ abus, family = "binomial", data = smp.1)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.8446	-0.6020	-0.6020	-0.6020	1.8959

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.6161	0.1154	-14.003	< 2e-16 ***
abus	0.7688	0.1897	4.052	5.07e-05 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 760.21 on 752 degrees of freedom
Residual deviance: 744.26 on 751 degrees of freedom
(46 observations deleted due to missingness)
AIC: 748.26

Number of Fisher Scoring iterations: 4



Une seule variable explicative



Introduction à la statistique avec R > Régression logistique : introduction

$$\text{Log} \left[\frac{\text{prob(haut risque de suicide)}}{1 - \text{prob(haut risque de suicide)}} \right] = a + b \times \text{abus}$$

```
> mod1 <- glm(suicide.hr~abus, data=smp.1, family="binomial")
> summary(mod1)

Call:
glm(formula = suicide.hr ~ abus, family = "binomial", data = smp.1)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.8446	-0.6020	-0.6020	-0.6020	1.8959

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.6161	0.1154	-14.003	< 2e-16 ***
abus	0.7688	0.1897	4.052	5.07e-05 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 760.21 on 752 degrees of freedom
Residual deviance: 744.26 on 751 degrees of freedom
(46 observations deleted due to missingness)
AIC: 748.26

Number of Fisher Scoring iterations: 4



Une seule variable explicative



Introduction à la statistique avec R > Régression logistique : introduction

$$\text{Log} \left[\frac{\text{prob(haut risque de suicide)}}{1 - \text{prob(haut risque de suicide)}} \right] = a + b \times \text{abus}$$

```
> mod1 <- glm(suicide.hr~abus, data=smp.1, family="binomial")
> summary(mod1)

Call:
glm(formula = suicide.hr ~ abus, family = "binomial", data = smp.1)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.8446	-0.6020	-0.6020	-0.6020	1.8959

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.6161	0.1154	-14.003	< 2e-16 ***
abus	0.7688	0.1897	4.052	5.07e-05 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 760.21 on 752 degrees of freedom
Residual deviance: 744.26 on 751 degrees of freedom
(46 observations deleted due to missingness)
AIC: 748.26

Number of Fisher Scoring iterations: 4



Une seule variable explicative



Introduction à la statistique avec R > Régression logistique : introduction

$$\text{Log} \left[\frac{\text{prob(haut risque de suicide)}}{1 - \text{prob(haut risque de suicide)}} \right] = a + b \times \text{abus}$$

```
> mod1 <- glm(suicide.hr~abus, data=smp.1, family="binomial")
> summary(mod1)

Call:
glm(formula = suicide.hr ~ abus, family = "binomial", data = smp.1)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.8446	-0.6020	-0.6020	-0.6020	1.8959

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.6161	0.1154	-14.003	< 2e-16 ***
abus	0.7688	0.1897	4.052	5.07e-05 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 760.21 on 752 degrees of freedom
Residual deviance: 744.26 on 751 degrees of freedom
(46 observations deleted due to missingness)
AIC: 748.26

Number of Fisher Scoring iterations: 4

```
> exp(0.7688)
[1] 2.157176
```

Pr. Bruno Falissard



Une seule variable explicative



Introduction à la statistique avec R > Régression logistique : introduction

```
> exp(0.7688)
[1] 2.157176
> library(Epi)
> twoby2(1-smp.1$suicide.hr, 1-smp.1$abus)
2 by 2 table analysis:
```

Outcome : 0

Comparing : 0 vs. 1

	0	1	P(0)	95% conf. interval	
0	63	90	0.4118	0.3366	0.4913
1	147	453	0.2450	0.2122	0.2810

95% conf. interval

Relative Risk: 1.6807 1.3276 2.1276

Sample Odds Ratio: 2.1571 1.4873 3.1287

Conditional MLE Odds Ratio: 2.1547 1.4577 3.1764

Probability difference: 0.1668 0.0837 0.2525

Exact P-value: 1e-04

Asymptotic P-value: 1e-04



Conclusion



Introduction à la statistique avec R > Régression logistique : introduction



```
mod1 <- glm(suicide.hr~abus, data=smp.l, family="binomial")
summary(mod1)
exp(0,7688)
library(Epi)
twoby2(1-smp.l$suicide.hr, 1-smp.l$abus)
```

