Régression logistique : introduction

Pr. Bruno Falissard
UNIVERSITÉ PARIS SUD

- 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

- Haut risque de suicide = a + b × duree + c × discip + d × abus + bruit

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

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

$$\underbrace{Log\left[\frac{prob(HR\,suicide = oui)}{1 - prob(HR\,suicide = oui)}\right]}_{\text{Varie entre + et - l'infini}} = a + b \times duree + c \times discip + d \times abus$$

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

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

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

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

```
> 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
```

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

```
> 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
```

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

```
> 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
```

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

```
> 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
```

```
> exp(0.7688)
[1] 2.157176
> library(Epi)
> twoby2(1-smp.l$suicide.hr, 1-smp.l$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
------------------------------------------------------
```

```
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)
```