

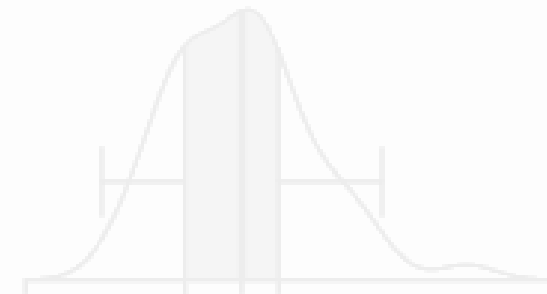
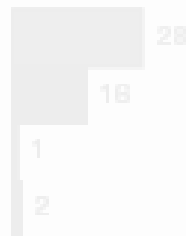
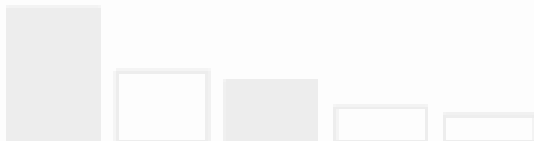
Régression logistique avec plusieurs variables explicative

Plusieurs variables explicatives

Introduction à la statistique avec R > La régression logistique



$$\text{Log} \left[\frac{p(h.r. \text{ de suicide})}{1 - p(h.r. \text{ de suicide})} \right] = a + b \times \text{duree} + c \times \text{discip} + d \times \text{abus}$$



$$\text{Log} \left[\frac{p(h.r. de suicide)}{1 - p(h.r. de suicide)} \right] = a + b \times duree + c \times discip + d \times abus$$

```
> mod2 <- glm(suicide.hr~abus+discip+duree, data=smp.l, family="binomial")
```

$$\text{Log} \left[\frac{p(h.r. \text{ de suicide})}{1 - p(h.r. \text{ de suicide})} \right] = a + b \times \text{duree} + c \times \text{discip} + d \times \text{abus}$$

```
> mod2 <- glm(suicide.hr~abus+discip+duree, data=smp.l, family="binomial")
> summary(mod2)
```

Call:

```
glm(formula = suicide.hr ~ abus + discip + duree, family = "binomial",
    data = smp.l)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.3200	-0.6655	-0.6012	-0.4997	2.0700

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.02462	0.49635	-0.050	0.960439
abus	0.62289	0.22764	2.736	0.006213 **
discip	0.52809	0.23767	2.222	0.026287 *
duree	-0.39862	0.11723	-3.400	0.000673 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 555.94 on 549 degrees of freedom
Residual deviance: 533.26 on 546 degrees of freedom
(249 observations deleted due to missingness)
AIC: 541.26

Number of Fisher Scoring iterations: 4

$$\text{Log} \left[\frac{p(h.r. \text{ de suicide})}{1 - p(h.r. \text{ de suicide})} \right] = a + b \times \text{duree} + c \times \text{discip} + d \times \text{abus}$$

```
> mod2 <- glm(suicide.hr~abus+discip+duree, data=smp.l, family="binomial")
> summary(mod2)
```

Call:

```
glm(formula = suicide.hr ~ abus + discip + duree, family = "binomial",
    data = smp.l)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.3200	-0.6655	-0.6012	-0.4997	2.0700

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.02462	0.49635	-0.050	0.960439
abus	0.62289	0.22764	2.736	0.006213 **
discip	0.52809	0.23767	2.222	0.026287 *
duree	-0.39862	0.11723	-3.400	0.000673 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 555.94 on 549 degrees of freedom
Residual deviance: 533.26 on 546 degrees of freedom
(249 observations deleted due to missingness)
AIC: 541.26

Number of Fisher Scoring iterations: 4

$$\text{Log} \left[\frac{p(h.r. \text{ de suicide})}{1 - p(h.r. \text{ de suicide})} \right] = a + b \times \text{duree} + c \times \text{discip} + d \times \text{abus}$$

```
> mod2 <- glm(suicide.hr~abus+discip+duree, data=smp.l, family="binomial")
> summary(mod2)
```

Call:

```
glm(formula = suicide.hr ~ abus + discip + duree, family = "binomial",
    data = smp.l)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.3200	-0.6655	-0.6012	-0.4997	2.0700

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.02462	0.49635	-0.050	0.960439
abus	0.62289	0.22764	2.736	0.006213 **
discip	0.52809	0.23767	2.222	0.026287 *
duree	-0.39862	0.11723	-3.400	0.000673 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 555.94 on 549 degrees of freedom
Residual deviance: 533.26 on 546 degrees of freedom
(249 observations deleted due to missingness)
AIC: 541.26

Number of Fisher Scoring iterations: 4

$$\text{Log} \left[\frac{p(h.r. \text{ de suicide})}{1 - p(h.r. \text{ de suicide})} \right] = a + b \times \text{duree} + c \times \text{discip} + d \times \text{abus}$$

```
> mod2 <- glm(suicide.hr~abus+discip+duree, data=smp.l, family="binomial")
> summary(mod2)
```

Call:

```
glm(formula = suicide.hr ~ abus + discip + duree, family = "binomial",
    data = smp.l)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.3200	-0.6655	-0.6012	-0.4997	2.0700

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.02462	0.49635	-0.050	0.960439
abus	0.62289	0.22764	2.736	0.006213 **
discip	0.52809	0.23767	2.222	0.026287 *
duree	-0.39862	0.11723	-3.400	0.000673 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 555.94 on 549 degrees of freedom
 Residual deviance: 533.26 on 546 degrees of freedom
 (249 observations deleted due to missingness)
 AIC: 541.26

Number of Fisher Scoring iterations: 4

Codage

abus : « 1 » oui, « 0 » non

discip : « 1 » oui, « 0 » non

duree : 1 à 5 (gradation)

```
> mod2 <- glm(suicide.hr~abus+discip+duree, data=smp.l, family="binomial")
> summary(mod2)
```

Call:

```
glm(formula = suicide.hr ~ abus + discip + duree, family = "binomial",
    data = smp.l)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.3200	-0.6655	-0.6012	-0.4997	2.0700

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.02462	0.49635	-0.050	0.960439
abus	0.62289	0.22764	2.736	0.006213 **
discip	0.52809	0.23767	2.222	0.026287 *
duree	-0.39862	0.11723	-3.400	0.000673 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 555.94 on 549 degrees of freedom
Residual deviance: 533.26 on 546 degrees of freedom
(249 observations deleted due to missingness)
AIC: 541.26

Number of Fisher Scoring iterations: 4

```
> exp(coefficients(mod2))
```

(Intercept)	abus	discip	duree
0.9756803	1.8643147	1.6956873	0.6712485

Codage

abus : « 1 » oui, « 0 » non

discip : « 1 » oui, « 0 » non

duree : 1 à 5 (gradation)

- Ce qui change peu ou pas avec le chapitre sur la régression linéaire
 - Variables catégorielles à plus de 2 classes (prof),

- Ce qui change peu ou pas avec le chapitre sur la régression linéaire
 - Variables catégorielles à plus de 2 classes (prof),
 - `drop1(mod1, .~., test="Chisq")`,

- Ce qui change peu ou pas avec le chapitre sur la régression linéaire
 - Variables catégorielles à plus de 2 classes (prof),
 - `drop1(mod1, .~., test="Chisq")`,
 - Interaction (`smp.l$duree*smp.l$discip`)

- Conditions de validité de la régression logistique

- Conditions de validité de la régression logistique
« Au moins 5-10 évènements par variable explicative »

- Conditions de validité de la régression logistique

« Au moins 5-10 évènements par variable explicative »

– Fichier « smp.l » : 799 détenus

– Variable à expliquer : « forme sévère de schizophrénie » : 54 détenus

- Conditions de validité de la régression logistique
 - « Au moins 5-10 évènements par variable explicative »
 - Fichier « smp.l » : 799 détenus
 - Variable à expliquer : « forme sévère de schizophrénie » : 54 détenus
 - Variables explicatives : age, trauma, prof

- Conditions de validité de la régression logistique
 - « Au moins 5-10 évènements par variable explicative »
 - Fichier « smp.l » : 799 détenus
 - Variable à expliquer : « forme sévère de schizophrénie » : 54 détenus
 - Variables explicatives : age (1), trauma (1), prof (7)

- Conditions de validité de la régression logistique
 - « Au moins 5-10 évènements par variable explicative »
 - Fichier « smp.l » : 799 détenus
 - Variable à expliquer : « forme sévère de schizophrénie » : 54 détenus
 - Variables explicatives : age, trauma, prof
 - $(1+1+7) \times 10 = 90 > 54 \rightarrow$ pas bon

- Conditions de validité de la régression logistique

« Au moins 5-10 évènements par variable explicative »

- Fichier « smp.l » : 799 détenus
- Variable à expliquer : « forme sévère de schizophrénie » : 54 détenus
- Variables explicatives : age, trauma, prof
- $(1+1+7) \times 10 = 90 > 54 \rightarrow$ pas bon
- $(1+1+7) \times 5 = 45 < 54 \rightarrow$ OK mais franchement limite

Conclusion

Introduction à la statistique avec R > La régression logistique



```
mod2 <- glm(suicide.hr~abus+discip+duree, data=smp.l,  
            family="binomial")  
summary(mod2)  
exp(coefficients(mod2))
```