

La régression logistique

Ghazi Bel Mufti

ghazi.belmufti@gmail.com

ESSAI-3 / DATA MINING

Plan

Introduction

Modélisation

Estimation des paramètres

Evaluation statistique de la régression logistique

Lecture et interprétation des coefficients

Introduction

- ▶ Même si elle a été introduite plus récemment que l'analyse discriminante dans les logiciels, elle a supplanté sa rivale dans de nombreux problèmes de classement : médecine, scoring...
- ▶ Le succès de la régression logistique repose notamment sur les nombreux outils qui permettent d'interpréter de manière approfondie les résultats obtenus.
- ▶ Elle permet de traiter des variables à prédire à 2 valeurs (sans faire d'hypothèses aussi restrictives que l'analyse discriminante), à $k \geq 3$ valeurs nominales avec des variables explicatives pouvant être quantitatives ou qualitatives.
- ▶ L'analyse discriminante linéaire peut encore opposer une résistance dans le cas de variables explicatives continues multinormales, et homoscédastiques, puisqu'elle fournit une solution directe et que la régression logistique ne fournit qu'une approximation.

Notations et définitions

- ▶ On considère une variable cible $Y = 0$ ou 1 , et p variables explicatives X_j continues, binaires ou qualitatives (dont les indicatrices ramènent au cas d'une variable binaire).
- ▶ Remarquons que les variables cibles qualitatives à $k > 2$ modalités sont traitées par ce que l'on appelle la *régression logistique polytomique*, les k modalités pouvant être ordonnées (*régression logistique ordinale*) ou non (*régression logistique multinomiale* ou *nominale*).
- ▶ Dans le cas d'une variable explicative qualitative à k modalités, après avoir choisi une modalité de référence, on remplacera cette variable par les $k - 1$ variables indicatrices associées aux $k - 1$ modalités restantes. Au total, on suppose que l'on a p variables explicatives x^1, x^2, \dots, x^p dont certaines peuvent être binaires.

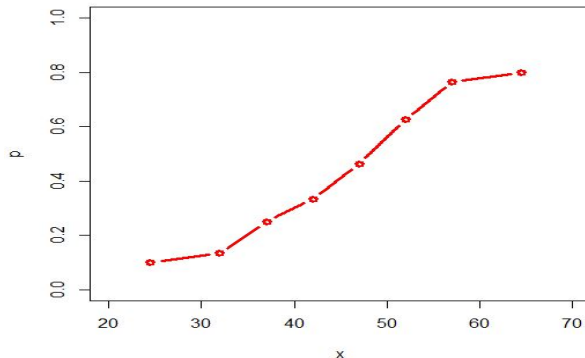
Principe

- ▶ La régression logistique consiste à modéliser $\pi(x) = E(Y|x)$, plus précisément de telle sorte que $\ln\left[\frac{\pi(x)}{1 - \pi(x)}\right]$ appelé logit de $\pi(x)$ soit une fonction linéaire des x^j .
- ▶ Notons que la plupart des résultats obtenus restent valables quand $H(\pi(x))$, la fonction H étant croissante, est une fonction linéaire des x^j .

Exemple I

- ▶ Il s'agit d'un échantillon de 100 personnes, pour lesquels la présence ($CHD = 1$) ou l'absence ($CHD = 0$) d'une maladie cardio-vasculaire a été observée. On souhaite expliquer la variable CHD par la variable explicative (AGE).
La Figure suivante qui représente les fréquences relatives de $CHD = 1$ selon les catégories d'âge suggère l'allure de cette fonction :

Exemple II



$P(Y = 1|X = x) = E(Y|X = x)$ a une forme sigmoïdale

Exemple III

```
> CHD.glm = glm(CHD~AGE, family=binomial(link="logit"))
> CHD.glm

Call:  glm(formula = CHD ~ AGE, family = binomial(link = "logit"))

Coefficients:
(Intercept)      AGE
   -5.3095      0.1109

Degrees of Freedom: 99 Total (i.e. Null);  98 Residual
Null Deviance:      136.7
Residual Deviance: 107.4      AIC: 111.4
```

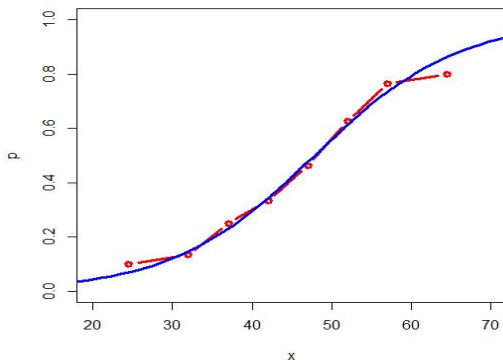
- Ainsi, si l'on suit l'expression de cette courbe, on peut écrire :

$$E(Y|X = x) = P(Y = 1|X = x) = \frac{\exp(b_0 + b_1 x)}{1 + \exp(b_0 + b_1 x)}$$

avec $b_0 = -5.309$ et $b_1 = 0.1109$.

Exemple IV

- La courbe obtenue avec les coefficients b_0 et b_1 estimés par la R.L. est illustrée par la Figure suivante.



Modélisation

La régression logistique est un cas particulier du modèle général suivant, où l'on modélise $\pi(x)$ sous la forme :

$$\pi(x) = G(\beta_0 + \beta' x), \quad (1)$$

où $G (= H^{-1})$ est une fonction connue, croissante, prenant ses valeurs dans $[0, 1]$, β_0 une constante inconnue et β un vecteur de paramètres, de dimension $p \times 1$, qui n'est pas connu et qu'il faut donc, comme β_0 , estimer.

Notons que si l'on pose $\gamma = \begin{pmatrix} \beta_0 \\ \beta \end{pmatrix}$ et $z = \begin{pmatrix} 1 \\ x \end{pmatrix}$, alors l'équation (1) se met sous la forme :

$$\pi(x) = G(\gamma' z) \quad (2)$$

Le *modèle logistique* correspond au cas où :

$$G(u) = \frac{\exp u}{1 + \exp u} \quad (3)$$

Cas d'une loi gaussienne I

- ▶ Du théorème de Bayes, on déduit :

$$\pi(x) = P(Y = 1|X = x) = \frac{P(x|y = 1)P(y = 1)}{P(x|y = 1)P(y = 1) + P(x|y = 0)P(y = 0)}$$

D'où

$$\pi(x) = \frac{P(x|y = 1)P(y = 1)/P(x|y = 0)P(y = 0)}{1 + P(x|y = 1)P(y = 1)/P(x|y = 0)P(y = 0)}$$

- ▶ Si la loi de x sachant $Y = k$ est une loi gaussienne de moyenne μ_k (pour $k = 1$ ou 2) et de matrice variance Σ connue, on a :

$$P(x|y = k) = (2\pi)^{-p/2}(\det(\Sigma))^{-p/2} \exp\{1/2(x - \mu_k)' \Sigma^{-1}(x - \mu_k)\}$$

Cas d'une loi gaussienne II

- Il en résulte que l'on a :

$$\frac{P(x|y=1)P(y=1)}{P(x|y=0)P(y=0)} = \exp\{\beta_0 + \beta'x\}$$

avec

$$\begin{aligned}\beta &= \Sigma^{-1} (\mu_1 - \mu_0) \\ \beta_0 &= \ln \frac{\pi_1}{\pi_0} - 1/2 (\|\mu_1\|_{\Sigma^{-1}}^2 - \|\mu_0\|_{\Sigma^{-1}}^2)\end{aligned}$$

où $\pi_1 = P(y=1)$ et $\pi_0 = P(y=0)$.

Estimation des paramètres I

- Rappelons que :

$$z = (x^0, x)' = ((x^0, x^1, \dots, x^p)', \text{ avec } x^0 = 1 \text{ et}$$

$$\gamma = (\beta_0, \beta)'$$

- Pour estimer les coefficients γ_j , on a recours à la méthode du maximum de vraisemblance.
- Les n observations (y_i, x_i) avec $i \in \{1, \dots, n\}$, sont des réalisations de variables indépendantes et les y_i , plus précisément d'une variable de Bernoulli.

Estimation des paramètres II

- La vraisemblance pour une observation s'écrit :

$$\pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}$$

et pour l'ensemble des observations, la vraisemblance du modèle est donnée par :

$$L(\gamma) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \text{ où } \pi(x_i) = \frac{\exp(\sum_{j=0}^p \gamma_j z^j)}{1 + \exp(\sum_{j=0}^p \gamma_j z^j)}$$

- Donc

$$\ln(L(\gamma)) = \sum_{i=1}^n [y_i \ln(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i))]$$

L'estimation dans la pratique I

- ▶ En posant $\pi_i = \pi(x_i)$ pour alléger l'écriture, on vérifie facilement que le vecteur $\frac{\partial \ln L}{\partial \gamma}$ est donné par :

$$\frac{\partial \ln L}{\partial \gamma} = \sum_i (y_i - \pi_i) \frac{1}{\pi_i(1 - \pi_i)} \frac{\partial \pi_i}{\partial \gamma}$$

- ▶ Par ailleurs,

$$\begin{aligned} \log \frac{\pi_i}{1 - \pi_i} &= \gamma' z_i \\ \left(\frac{1}{\pi_i} + \frac{1}{1 - \pi_i} \right) \frac{\partial \pi_i}{\partial \gamma} &= z_i \\ \frac{\partial \pi_i}{\partial \gamma} &= \pi_i(1 - \pi_i) z_i \end{aligned}$$

- ▶ Ainsi, le vecteur $\frac{\partial \pi_i}{\partial \gamma}$ est donné par $\pi_i(1 - \pi_i) z_i$.

L'estimation dans la pratique II

- ▶ Sachant que $\frac{\partial \pi_i}{\partial \gamma} = \pi_i(1 - \pi_i)z_i$, on a donc

$$\frac{\partial \ln L}{\partial \gamma} = \sum_i (y_i - \pi_i)z_i$$

- ▶ Le problème n'a pas de solution analytique à l'instar des fonctions discriminantes de l'analyse discriminante que l'on obtient en inversant la matrice des covariances.
- ▶ En pratique on utilise une procédure approchée pour obtenir une solution satisfaisante de la maximisation ci-dessus.
- ▶ La procédure la plus connue est la méthode Newton-Raphson qui est une méthode itérative du gradient (Algorithme d'optimisation).
- ▶ Cette absence de solution analytique est le principal inconvénient de la RL qui la rend plus difficile à programmer...

L'estimation dans la pratique III

- En désignant par $\hat{\gamma}$ l'estimateur du maximum de vraisemblance de γ . On a alors les résultats asymptotiques suivants qui seront utiles si on veut faire des tests :

$$\begin{aligned}(\hat{\gamma} - \gamma) &\rightsquigarrow \mathcal{N}(\mathbf{0}, \mathbf{I}(\gamma)^{-1}) \\(\hat{\gamma} - \gamma)' \mathbf{I}(\gamma) (\hat{\gamma} - \gamma) &\rightsquigarrow \chi^2_{p+1} \\(\hat{\gamma} - \gamma)' \mathbf{I}(\hat{\gamma}) (\hat{\gamma} - \gamma) &\rightsquigarrow \chi^2_{p+1}\end{aligned}$$

L'estimation dans la pratique IV

- où $\mathbf{I}(\gamma)$ est la matrice d'information de Fisher :

$$\begin{aligned}\mathbf{I}(\gamma) &= \text{Var} \left(\frac{\partial \ln L}{\partial \gamma} \right) \\ &= \left(\text{Cov} \left[\frac{\partial \ln L}{\partial \gamma_j}, \frac{\partial \ln L}{\partial \gamma_{j'}} \right] \right)_{1 \leq j, j' \leq p} \\ &= \left(-\mathbb{E} \left[\frac{\partial^2 \ln L}{\partial \gamma_j \partial \gamma_{j'}} \right] \right)_{1 \leq j, j' \leq p}\end{aligned}$$

Tests de significativité des coefficients I

L'objectif des tests de significativité est de prouver le rôle d'une, de plusieurs ou de l'ensemble, des variables explicatives. Formellement, les hypothèses nulles peuvent se décliner comme suit :

1. Evaluer la contribution individuelle d'une variable

$$H_0 : \beta_j = 0$$

- Ce test de significativité est systématiquement donné par les logiciels.
- Une de ses formes (test de Wald) est systématiquement proposée.
- L'autre (test du rapport de vraisemblance) est passée sous silence par certains logiciels.
- Il arrive que ces approches ne se comportent pas de la même manière.

Tests de significativité des coefficients II

2. Evaluer la contribution d'un bloc de "q" variables. Sans restreindre la généralité du propos (les coefficients à tester ne sont pas forcément consécutifs dans la régression)

$$H_0 : \beta_j = \beta_{j+1} = \dots = \beta_{j+q} = 0$$

On ne peut pas le transformer en une succession de tests individuels. En effet, les coefficients ne sont pas indépendants (en tous les cas, ils ont une covariance non-nulle). Il faut bien tester la nullité simultanée des q coefficients.

3. Evaluer l'apport de l'ensemble des variables explicatives.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

Il s'agit d'une évaluation globale de la régression. En effet, si l'hypothèse nulle est compatible avec les données, cela signifierait qu'aucun des descripteurs ne contribue à l'explication de la variable dépendante.

Deux approches pour les tests I

[1.] Le principe du rapport de vraisemblance.

[2.] Test de Wald.

Principe du rapport de vraisemblance I

- ▶ L'approche est générique, elle est en cohérence avec la démarche d'estimation des paramètres.
- ▶ Elle est puissante c.-à-d. elle détecte mieux l'hypothèse alternative lorsqu'elle est vraie.
- ▶ L'inconvénient est qu'elle est plus gourmande en ressources machines : chaque hypothèse à évaluer donne lieu à une nouvelle estimation des paramètres, donc à un processus d'optimisation.

Principe du rapport de vraisemblance II

$$\begin{aligned} LR &= -2 \log\left(\frac{L(\text{modèle réduit})}{L(\text{modèle complet})}\right) \\ &= -2[\log(L(\text{modèle réduit})) - \log(L(\text{modèle complet}))] \end{aligned}$$

- *modèle réduit* : modèle à j variables
- *modèle complet* : modèle à $j + l$ variables
- LR suit asymptotiquement une loi du χ^2 à (l) degrés de liberté.
- La déviance d'un modèle $D(\text{modèle})$ correspond à la valeur de LR lorsque le modèle complet est le modèle saturé (i.e modèle de vraisemblance maximale).

Tests de validité du modèle (1)

1. En particulier, l' H_0 que le modèle s'ajuste bien aux données peut se reformuler de la manière suivante :

"les j paramètres du modèle (en incluant la constante) sont suffisants pour l'ajustement"

ou encore

"les paramètres supplémentaires pour arriver au nombre de paramètres du modèle saturé sont tous nul" :

$$H_0 : \beta_{j+1} = \beta_{j+2} = \dots = \beta_n = 0$$

2. Sous cette hypothèse nulle, LR suit une loi du χ^2 à $n - j$ degrés de liberté
3. **Si H_0 est acceptée, le modèle est bien ajusté aux données.**

Tests de validité du modèle (2)

1. $H_0 : \beta_1 = \beta_2 = \dots = \beta_j = 0$
2. Sous l'hypothèse nulle de nullité de tous les coefficients $\beta_1, \beta_2, \dots, \beta_j$, LR suit une loi du χ^2 à j ddl
3. **Si H_0 est rejetée, le modèle est bien ajusté aux données.**
 - On retrouve cette exigence sous une forme équivalente dans deux autres critères classiques :
 - **Critère d'Akaïké** $AIC = -2 \log[L(j)] + 2(j + 1)$
 - Critère de Schwartz $SC = -2 \log[L(j)] + (j + 1) \log(n)$ souvent noté BIC (Bayesian Information Criterion).

Evaluation d'une variable

Cas d'une variable continue :

- $H_0 : \beta_j = 0$
- Sous l'hypothèse nulle de nullité du coefficient d'une variable LR suit une loi du χ^2 à 1 ddl

Cas d'une variable catégorielle à $q + 1$ modalités :

- $H_0 : \beta_j = \dots = \beta_{j+q} = 0$
- Pour évaluer le rôle de la variable catégorielle prise dans son ensemble, quelle que soit la modalité considérée, nous devons tester simultanément les coefficients associés aux variables indicatrices.

Test de Wald I

- ▶ S'appuyer sur la normalité asymptotique des estimateurs (du maximum de vraisemblance). On parle de test de Wald.
- ▶ Le principal avantage est que les informations que l'on souhaite exploiter sont toutes disponibles à l'issue de l'estimation du modèle complet, incluant l'ensemble des variables.
- ▶ L'obtention des résultats est donc immédiate.
- ▶ L'inconvénient est que le test de Wald est conservateur. Il a tendance à favoriser l'hypothèse nulle.

Test de Wald II

- ▶ Soit l'hypothèse (H_0) :

$$H_0 : D\gamma = d$$

où D est une matrice de dimension $(s, p + 1)$, γ le vecteur de dimension $p + 1$ et d le vecteur de dimension s .

- ▶ Par exemple, si l'on veut tester $H_0 : \beta = (\beta_1, \dots, \beta_p) = 0$, on prendra

$$D = (0, Id_p) \text{ et } d = 0 \in \mathbb{R}^p$$

où D est de dimension $(p, p + 1)$.

- ▶ Si l'on veut tester $H_0 : \beta_1 = 0$, on prendra

$$D = (0, 1, 0, \dots, 0) \text{ et } d = 0 \text{ (i.e. } s = 1)$$

dans ce cas, $D\gamma = \beta_1 = d = 0$.

Test de Wald III

- ▶ Comme $\hat{\gamma}$ suit asymptotiquement une loi normal de matrice variance $\mathbf{I}(\gamma)^{-1}$, en posant

$$T = (D\hat{\gamma} - d)'[D\mathbf{I}(\gamma)^{-1}D']^{-1}(D\hat{\gamma} - d),$$

on a $T \rightsquigarrow \chi_s^2$ sous H_0 .

- ▶ Donc on rejette H_0 si $T > \chi_{1-\alpha}^2(s)$ où $\chi_{1-\alpha}^2(s)$ est le fractile d'ordre $1 - \alpha$ de la loi du χ^2 à s d.d.l.
- ▶ Cas du test du modèle : $H_0 : \beta = (\beta_1, \dots, \beta_p) = 0$

$D\mathbf{I}(\gamma)^{-1}D' = \mathbf{I}^{11}(\hat{\gamma})$ si l'on écrit $\mathbf{I}(\gamma)^{-1}$ sous la forme :

$$\mathbf{I}(\gamma)^{-1} = \begin{matrix} & \begin{matrix} 1 & \xleftarrow{p} \end{matrix} \\ \begin{matrix} \downarrow \\ p \end{matrix} & \begin{pmatrix} \mathbf{I}^{00}(\hat{\gamma}) & \mathbf{I}^{01}(\hat{\gamma}) \\ \mathbf{I}^{10}(\hat{\gamma}) & \mathbf{I}^{11}(\hat{\gamma}) \end{pmatrix} \end{matrix}$$

Test de Wald IV

- Cas du test du rôle significatif d'une variable : $H_0 : \beta_j = 0$

$$\frac{|D\hat{\gamma} - d|}{\sqrt{D\mathbf{I}(\gamma)^{-1}D'}} = \frac{|\hat{\beta}_j|}{\sqrt{[\mathbf{I}(\hat{\gamma})]_{jj}^{-1}}} > u_{1-\alpha/2}$$

Test de Wald sur une variable

- Ainsi, dans le cas où l'on cherche à tester le rôle significatif d'une variable :

$$H_0 : \beta_j = 0$$

- La statistique de WALD répond à ce test (de nullité d'un des coefficients), elle s'écrit

$$W = \frac{\hat{\beta}^2}{\hat{V}(\hat{\beta})} \rightsquigarrow \chi^2 \text{ à } 1 \text{ ddl}$$

- Le logiciel R, propose la statistique Z à la place de W , avec

$$Z = \frac{\hat{\beta}}{\hat{\sigma}(\hat{\beta})} \rightsquigarrow \mathcal{N}(0, 1)$$

Rapport de vraisemblance ou Wald ? I

Le test du rapport de vraisemblance :

- ▶ Il est plus puissant. Il détecte mieux l'hypothèse alternative lorsque cela est justifié.
- ▶ Il est, en revanche, plus gourmand en ressources car il impose de recalculer le modèle sous la contrainte de l'hypothèse nulle. Encore une fois, le problème ne se pose véritablement que lorsque nous avons à traiter une grande base de données.

Rapport de vraisemblance ou Wald ? II

Le test de Wald :

- ▶ Il est moins puissant, plus conservateur. Il favorise l'hypothèse nulle H_0 .
- ▶ Lorsque la valeur du coefficient est élevé, l'estimation de l'écart type gonfle exagérément. De nouveau H_0 est favorisé lors des tests individuels, cela nous emmène à supprimer à tort des variables importantes du modèle.
- ▶ Il repose sur des propriétés asymptotiques de l'estimateur. Il est par conséquent peu précis lorsque nous traitons de petits effectifs.

Lecture et interprétation des coefficients

- ▶ Dans certains domaines, l'explication est bien plus importante que la prédiction.
- ▶ On souhaite comprendre les phénomènes de causalité, mettre à jour les relations de cause à effet.
- ▶ La régression logistique propose des outils qui permettent d'interpréter les résultats sous forme de chances (*odds*) et de rapports de chances (*odds ratio*).

Exemple 1

- Le tableau suivant croise la variable dépendante cœur (avoir une maladie cardiaque ou pas +/-) avec la variable explicative angine (groupe "exposé" vs. groupe "témoin" 1/0). deux variables sont nominales.

Nombre de cœur	angine ▼		
cœur2 ▼	1	0	Total
1	3	3	6
0	2	12	14
Total	5	15	20

Y / X	1	0	
1	a	b	a+b
0	c	d	c+d
	a+c	b+d	n

RR	3
----	---

Odds(+ 1)	1.5
-----------	-----

Odds(+ 0)	0.25
-----------	------

OR(+)	6
-------	---

Odds

- ▶ L'odds ou rapport de chances est défini comme un rapport de probabilités dans un groupe. Par exemple, dans le groupe exposé, il s'écrit :

$$odds(1) = \frac{P(+/1)}{P(-/1)} = \frac{a/(a+c)}{c/(a+c)} = 1.5$$

- ▶ Dans le groupe des personnes ayant une angine de poitrine, on a 1.5 fois plus de chances d'avoir une maladie cardiaque que de ne pas en avoir.
- ▶ Nous pouvons de la même manière définir l'odds dans le groupe témoin $odds(0)$:

$$odds(0) = \frac{P(+/0)}{P(-/0)} = \frac{b/(b+d)}{d/(b+d)} = 0.25$$

Odds ratio

- L'odds ratio est égal au rapport entre l'odds du groupe exposé (avoir une angine) et l'odds du groupe témoin

$$OR = \frac{odds(1)}{odds(0)} = 6$$

- L'OR indique que dans le groupe exposé, on a 6 fois plus de chances d'avoir une maladie cardiaque que dans le groupe témoin.

Log odds ratio I

- Le logarithme de l'odds-ratio. Développons son expression, nous verrons ainsi le rapport avec la régression logistique.

$$\begin{aligned} \ln(OR) &= \ln\left(\frac{\text{odds}(1)}{\text{odds}(0)}\right) \\ &= \ln(\text{odds}(1)) - \ln(\text{odds}(0)) \\ &= \ln\frac{P(Y = +/1)}{P(Y = -/1)} - \ln\frac{P(Y = +/0)}{P(Y = -/0)} \\ &= \ln\frac{P(Y = +/1)}{1 - P(Y = +/1)} - \ln\frac{P(Y = +/0)}{1 - P(Y = +/0)} \\ &= \text{Logit}(1) - \text{Logit}(0) \end{aligned}$$

Variable explicative binaire I

- ▶ Nous essayons de prédire la variable coeur en fonction de taux max.
- ▶ En considérant un cas de RL simple, plus précisément, un logit de la forme

$$\ln\left[\frac{\pi(x)}{1 - \pi(x)}\right] = a_0 + a_1 \times X$$

$$X = 1 \rightarrow \text{logit}(1) = a_0 + a_1 \times 1 = a_0 + a_1$$

$$X = 0 \rightarrow \text{logit}(0) = a_0 + a_1 \times 0 = a_0$$

$$\text{On a alors : } \ln(OR) = \text{Logit}(1) - \text{Logit}(0) = a_1$$

Ainsi :

$$OR = e^{a_1}$$

Variable explicative binaire II

Attributes in the equation

Attribute	Coef.	Std-dev	Wald	Signif
constant	-1.386294	.	.	.
angine	1.791759	1.1181	2.5682	0.1090

Odds ratios and 95% confidence intervals

Attribute	Coef.	Low	High
angine	6.0000	0.6706	53.6844

Variable explicative quantitative I

- Pour comprendre l'interprétation des coefficients dans le cas d'une variable explicative quantitative, voyons l'évolution du Logit lorsqu'on fait varier X d'une unité.

$$\text{logit}(X + 1) = a_0 + a_1 (X + 1)$$

$$\text{logit}(X) = a_0 + a_1 X$$

$$\text{D'où } \text{Logit}(X + 1) - \text{Logit}(X) = a_1$$

- Dans ce cas, la quantité e^{a_1} s'interprète comme l'odds ratio consécutif à l'augmentation d'une unité de la variable explicative.
- Si l'on augmente de b unités la variable explicative, l'odds-ratio devient alors e^{ba_1} .

Variable explicative quantitative II

- En réalisant la régression logistique, nous obtenons les coefficients estimés donnés dans le tableau suivant :

Attributes in the equation

Attribute	Coef.	Std-dev	Wald	Signif
constant	8.478165	.	.	.
taux_max	-0.062653	0.0360	3.0351	0.0815

Odds ratios and 95% confidence intervals

Attribute	Coef.	Low	High
taux_max	0.9393	0.8753	1.0079

Variable explicative qualitative nominale I

- ▶ La variable dépendante est toujours la présence/absence d'une maladie cardiaque (coeur). La variable explicative est "chest pain" (type de douleur dans la poitrine) avec 4 modalités : "type. angina" (code 1), "atyp. angina" (2), "asympt." (3) et "non anginal" (4).
- ▶ Le choix de la modalité de référence est crucial pour l'interprétation. Il ne peut pas être dissocié de l'analyse qualitative des résultats que l'on veut mener par la suite.
- ▶ Dans notre exemple, admettons qu'il s'agisse de la dernière (non anginal - code 4). Nous aurons à calculer $L - 1 = 4 - 1 = 3$ odds-ratio.

Variable explicative qualitative nominale II

Nombre de cœur	chest_pain ▼				
cœur ▼	typ_angina	atyp_angina	asympt	non_anginal	Total
presence	4	6	75	7	92
absence	2	59	27	29	117
Total	6	65	102	36	209

Odds(+/-)	2.000	0.102	2.778	0.241
OR(x/ non_anginal)	8.286	0.421	11.508	

Variable explicative qualitative nominale III

- ▶ L'odds de la catégorie 1 est obtenue avec $odds(1) = 4/2 = 2.0$: les personnes présentant une douleur de type "typ. angina" ont 2.0 fois plus de chances d'avoir une maladie cardiaque.
- ▶ De même pour les autres catégories, nous pouvons calculer : $odds(2) = 0.102$; $odds(3) = 2.778$; $odds(4) = 0.241$.
- ▶ La 4^{ème} catégorie représentant la situation de référence, nous calculons les 3 odds-ratio en l'opposant aux autres c.-à-d

$$OR(1/4) = \frac{odds(1)}{odds(4)} = 8.286$$

- ▶ Nous le lisons de la manière suivante "les personnes qui ont une douleur dans la poitrine de type typ. angina ont 8.286 fois plus de chances de développer une maladie cardiaque que ceux qui présentent une douleur de type non anginal".
- ▶ De même, nous pouvons produire $OR(2/4) = 0.421$, $OR(3/4) = 11.508$.

Régression logistique sur R : Déviance I

- ▶ Null Deviance = $2(LL(\text{Saturated Model}) - LL(\text{Null Model}))$ on $df = df \text{ Sat} - df \text{ Null}$
- ▶ Residual Deviance = $2(LL(\text{Saturated Model}) - LL(\text{Proposed Model}))$ $df = df \text{ Sat} - df \text{ Proposed}$
- ▶ The Saturated Model is a model that assumes each data point has its own parameters (which means you have n parameters to estimate.)
- ▶ The Null Model assumes the exact "opposite", in that it assumes one parameter for all of the data points, which means you only estimate 1 parameter.
- ▶ The Proposed Model assumes you can explain your data points with p parameters + an intercept term, so you have $p+1$ parameters.

Régression logistique sur R : Déviance II

- ▶ If your Null Deviance is really small, it means that the Null Model explains the data pretty well. Likewise with your Residual Deviance.
- ▶ What does really small mean ? If your model is "good" then your Deviance is approx Chi 2 with $(df_{sat} - df_{model})$ degrees of freedom.
- ▶ If you want to compare your Null model with your Proposed model, then you can look at : $(Null\ Deviance - Residual\ Deviance)$ approx Chi 2 with $df_{Proposed} - df_{Null} = (n - (p + 1)) - (n - 1) = p$