

Examen Régression logistique

Dadi Abel

15/02/2021

1 CREATION D'UN REPERTOIRE DE TRAVAIL

```
setwd("C:/Users/Dadi_abel/Desktop/MesCours2021/Régression logistique/Contrôle de connaissance")
getwd()
```

2 CHARGER LE JEU DE DONNEE

```
hypotrophie <- read.csv("Hypotrophie.csv", sep = ";", header = T)
attach(hypotrophie)
hypotrophie$Hypotrophie.recorde <- ifelse(hypotrophie$Hypotrophie==1,"hypotrophie","pas d'hypotrophie")
hypotrophie$Tabac.recorde <- ifelse(hypotrophie$Tabac==1,"fume","ne fume pas")
require(knitr)
kable(head(hypotrophie[,c(-6,-7)]))
```

Rang_naissance	Tabac	Age_maman	Poids_maman	Hypotrophie
1	1	28	54	0
2	1	33	64	0
1	0	29	59	0
2	0	34	68	0
3	0	37	65	1
1	1	31	85	1

On dispose de données concernant l'hypotrophie néonatale et on souhaite utiliser ces données pour étudier les facteurs prédictifs de l'hypotrophie néonatale (hypotrophie néonatale = nouveau-né de poids anormalement faible). Les premières lignes du tableau :

3 EN ANALYSE UNIVARIEE, ETUDIONS LA RELATION ENTRE TABAC ET HYPOTROPHIE :

3.1 INDIQUE-S'IL EXISTE OU NON UNE RELATION STATISTIQUEMENT SIGNIFICATIVE ENTRE TABAC ET HYPOTROPHIE,

```
modele1 <- glm(Hypotrophie ~ Tabac, data = hypotrophie, family = "binomial")
summary(modele1)
```

```
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.1587      0.1370  -8.457  < 2e-16 ***
## Tabac         0.8169      0.1997   4.090  4.31e-05 ***
```

3.2 QUANTIFIONS CETTE RELATION PAR L'ODD'S RATIO ASSOCIE AU FAIT DE FUMER (AVEC SON INTERVALLE DE CONFIANCE)

```
resultat1 <- summary(modele1)
(OR_tabac <- exp(resultat1$coefficients[2,1]))
```

Odd's ratio = 2.263534

```
(borne_inf <- exp(resultat1$coefficients[2,1]-1.96*resultat1$coefficients[2,2]))
```

Borne inferieur = 1.530322

```
(borne_sup <- exp(resultat1$coefficients[2,1]+1.96*resultat1$coefficients[2,2]))
```

Borne supérieur = 3.348043

CONCLUSION : Le tabac est significativement associé au risque d'hypotrophie néonatale. Une maman fumeuse de tabac augmente significativement le risque d'hypotrophie néonatale par rapport à une maman non fumeuse ; odd's ratio associé à une maman fumeuse plutôt qu'une non fumeuse = 2.263534 (IC95% 1.530322 à 3.348043), donc l'hypotrophie néonatale est plus fréquente chez les femmes fumeuse (OR > 1).

3.3 CALCULEZ LE RISQUE RELATIF ASSOCIE AU FAIT DE FUMER (RISQUE RELATIF = PROBABILITE D'HYPOTROPHIE SI LA MAMAN FUME / PROBABILITE D'HYPOTROPHIE SI LA MAMAN NE FUME PAS)

```
kable(table(hypotrophie$Tabac.recorde, hypotrophie$Hypotrophie.recorde))
```

	Hypotrophie	Pas d'hypotrophie
Fume	81	114
Ne fume pas	70	223

```
(RR <- (81/(81+114))/(70/(70+223))) # Calcul du risque relatif
```

Risque relative = 1.738681

Vérification du résultat

```
library(Epi)
twoby2(hypotrophie$Tabac.recorde, hypotrophie$Hypotrophie.recorde)
```

```
## 2 by 2 table analysis:
```

```
## -----
```

```
## Outcome : hypotrophie
```

```
## Comparing : fume vs. ne fume pas
```

```
##
```

	hypotrophie	pas d'hypotrophie	P(hypotrophie)	95% conf. interval
fume	81	114	0.4154	0.3483 0.4858
ne fume pas	70	223	0.2389	0.1935 0.2911

```
##
```

```
##
```

	95% conf. interval
Relative Risk: 1.7387	1.3358 2.2631
Sample Odds Ratio: 2.2635	1.5303 3.3480
Conditional MLE Odds Ratio: 2.2596	1.5009 3.4123
Probability difference: 0.1765	0.0918 0.2600

```
##
```

```
## Exact P-value: 0.0001
```

```
## Asymptotic P-value: 0.0000
```

```
## -----
```

CONCLUSION : Les mamans fumeurs de tabac ont 1.74 fois plus de chance de donner naissance à un enfant dont le poids est anormalement faible qu'une maman qui ne fume pas de tabac.

4 EN ANALYSE UNIVARIEE, ETUDIEZ LA RELATION ENTRE AGE DE LA MAMAN ET HYPOTROPHIE :

4.1 INDIQUE-S'IL EXISTE OU NON UNE RELATION STATISTIQUEMENT SIGNIFICATIVE ENTRE AGE DE LA MAMAN ET HYPOTROPHIE,

```
modele2 <- glm(Hypotrophie ~ Age_maman, data = hypotrophie, family = "binomial")
summary(modele2)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.50501	0.46089	-3.265	0.00109 **
Age_maman	0.02639	0.01682	1.569	0.11664

4.2 QUANTIFIONS CETTE RELATION PAR L'ODD'S RATIO ASSOCIE AU FAIT D'UN AGE AUGMENTE DE 5 ANS (AVEC SON INTERVALLE DE CONFIANCE)

```
resultat2 <- summary(modele2)
(OR_Age.maman.5ans <- exp(resultat2$coefficients[2,1])^5)
```

Odd's ratio = 1.141053

```
(borne_inf <- exp(resultat2$coefficients[2,1]-1.96*resultat2$coefficients[2,2])^5)
```

Borne inferieur = 0.9676563

```
(borne_sup <- exp(resultat2$coefficients[2,1]+1.96*resultat2$coefficients[2,2])^5)
```

Borne supérieur = 1.345521

CONCLUSION : L'âge de la maman n'est pas significativement associé au risque d'hypotrophie néonatale (pvalue > 0.05). L'odd's ratio associé à l'augmentation de l'âge de la maman de 5 ans vaut 1.141053 (IC95% 0.9676563 à 1.345521), donc l'hypotrophie néonatale est plus fréquente chez les femmes fumeuse (OR > 1).

4.3 CALCULEZ LE RISQUE RELATIF ASSOCIE AU FAIT, POUR LA MAMAN, D'AVOIR 25 ANS PLUTOT QUE 20 ANS (RISQUE RELATIF = PROBABILITE D'HYPOTROPHIE SI LA MAMAN A 25 ANS / PROBABILITE D'HYPOTROPHIE SI LA MAMAN A 20 ANS)

```
don <- subset(hypotrophie, (Age_maman==25 | Age_maman==20))
kable(table(don$Age_maman, don$Hypotrophie.recorde))
```

	Hypotrophie	Pas d'hypotrophie
20	6	17
25	8	23

```
(RR <- (8/(23+8))/(6/(17+6)))
```

Risque Relatif = 0.9892473

CONCLUSION : Les mamans âgées de 25 ans ont 1 fois plus de chance de donner naissance à un enfant dont le poids est anormalement faible qu'une maman âgée de 20 ans.

4.4 FAITES LE CALCUL DU RISQUE RELATIF CETTE FOIS POUR UNE MAMAN AYANT 35 ANS PAR RAPPORT A UNE MAMAN AYANT 30 ANS (RISQUE RELATIF = PROBABILITE D'HYPOTROPHIE SI LA MAMAN A 35 ANS / PROBABILITE D'HYPOTROPHIE SI LA MAMAN A 30 ANS).

```
don <- subset(hypotrophie, (Age_maman==35 | Age_maman==30))
kable(table(don$Age_maman, don$Hypotrophie.recorde))
```

	hypotrophie	pas d'hypotrophie
30	13	14
35	6	5

```
(RR <- (6/(5+6))/(13/(13+14)))
```

Risque Relatif = 1.132867

CONCLUSION : Les mamans âgées de 35 ans ont 1.13 fois plus de chance de donner naissance à un enfant dont le poids est anormalement faible qu'une maman âgée de 30 ans.

5 DEFINISSONS LE MEILLEUR MODELE EN NE GARDANT QUE LES VARIABLES QUI NOUS SEMBLANT PERTINENTES PUIS VALIDEZ CE MODELE PAR UN TEST DE STUKEL.

5.1 DEFINISSONS LE MEILLEUR MODELE

```
modele3 <- glm(Hypotrophie ~ Rang_naissance + Tabac + Age_maman + Poids_maman, family = "binomial")
summary(modele3)
```

```
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.070633   0.594151  -1.802   0.0716 .
## Rang_naissance  0.102624   0.150226   0.683   0.4945
## Tabac          0.801432   0.202625   3.955 7.65e-05 ***
## Age_maman      0.036867   0.022171   1.663   0.0963 .
## Poids_maman   -0.019577   0.007827  -2.501   0.0124 *
```

```
step(modele3)
```

```
## Start: AIC=587.08
## Hypotrophie ~ Rang_naissance + Tabac + Age_maman + Poids_maman
##
##              Df Deviance    AIC
## - Rang_naissance  1    577.54 585.54
## <none>              577.08 587.08
## - Age_maman        1    579.86 587.86
## - Poids_maman       1    583.77 591.77
## - Tabac             1    592.85 600.85
##
## Step: AIC=585.54
## Hypotrophie ~ Tabac + Age_maman + Poids_maman
##
##              Df Deviance    AIC
## <none>          577.54 585.54
## - Age_maman     1    583.61 589.61
```

```
## - Poids_maman 1 583.94 589.94
## - Tabac 1 593.71 599.71

##
## Call: glm(formula = Hypotrophie ~ Tabac + Age_maman + Poids_maman,
## family = "binomial")
##
## Coefficients:
## (Intercept) Tabac Age_maman Poids_maman
## -1.14090 0.80950 0.04524 -0.01898
##
## Degrees of Freedom: 487 Total (i.e. Null); 484 Residual
## Null Deviance: 603.8
## Residual Deviance: 577.5 AIC: 585.5
```

Étape 1 : On enlève Rang_naissance Le modèle sans Rang_naissance est celui qui a le meilleur AIC (AIC plus petit).

Étape 2 : Le modèle complet est celui qui a le meilleur AIC. Il n'y a plus rien à enlever. On s'arrête là.

```
meilleur.modele <- glm(formula = Hypotrophie ~ Tabac + Age_maman + Poids_maman, family =
"binomial")
summary(meilleur.modele)
```

```
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.140901 0.584237 -1.953 0.0508 .
## Tabac 0.809496 0.202227 4.003 6.26e-05 ***
## Age_maman 0.045237 0.018488 2.447 0.0144 *
## Poids_maman -0.018979 0.007746 -2.450 0.0143 *
```

5.2 VALIDATION DU MODELE PAR UN TEST DE STUKEL

ETUDE DE LA RELATION ENTRE TABAC, AGE_MAMAN, POIDS_MAMAN ET HYPOTROPHIE - TEST DE STUKEL

```
logit <- meilleur.modele$linear.predictors # on récupère Les Logit
```

on crée Les variables supplémentaires

```
lin2.plus <- c()
```

```
lin2.moins <- c()
```

```
for (i in 1:length(logit)) {# on démarre une boucle qui balaye chaque Logit
```

```
  if(logit[i]>0){ # on teste si la i ème valeur de Logit est positive
    lin2.plus[i] <- 0.5*logit[i]^2
    lin2.moins[i] <- 0
  }
  else{
    lin2.plus[i] <- 0
    lin2.moins[i] <- -0.5*logit[i]^2
  }
}
```

```
d.plus <- data.frame(hypotrophie,lin2.plus,lin2.moins) # on ajoute Les variables créées
```

à la table de base

```
r.plus <- glm(Hypotrophie ~ Tabac + Age_maman + Poids_maman + lin2.plus + lin2.moins, data=d.plus, family = "binomial")
```

`summary(r.plus)` # on refait la régression avec les variables supplémentaires

```
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.24671    0.77478  -1.609   0.1076
## Tabac        0.93775    0.52591   1.783   0.0746 .
## Age_maman    0.04721    0.03675   1.285   0.1989
## Poids_maman -0.02165    0.01575  -1.374   0.1695
## lin2.plus    37.91864   33.02264   1.148   0.2509
## lin2.moins   -0.32330    0.78558  -0.412   0.6807
```

```
a.plus <- anova(r.plus, test="Chisq") # on fait l'analyse de déviance
```

```
x = a.plus[5,2] + a.plus[6,2] # on récupère la déviance supprimée
```

```
p.value = 1 - pchisq(x, 2) # on calcule la p.value (distribution du Chi2 avec 2ddl)
```

p. value = 0.22

5.3 AVEC CE MODELE, DETERMINEZ LA MATRICE DE VRAISEMBLANCE

```
proba_hypotrophie <- predict(meilleur.modele, hypotrophie, type = "response")
summary(proba_hypotrophie)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.09108 0.22956 0.28207 0.30943 0.38794 0.59535
```

```
p_class <- ifelse(proba_hypotrophie > 0.5, "Hypotrophie_prévue", "Pas_d'hypotrophie_prévue.")
kable(table(p_class, hypotrophie$Hypotrophie.recorde))
```

	hypotrophie	pas d'hypotrophie
Hypotrophie_prévue	16	13
Pas_d'hypotrophie_prévue.	135	324

5.3.1 CALCULONS LE TAUX D'ERREUR

5.3.2 LE TAUX D'ERREUR = MAL CLASSES / EFFECTIF TOTAL

```
mal.classés <- 13 + 135
effectif.total <- (16+13+135+324)
(taux.erreur <- (mal.classés/effectif.total)*100)
```

Taux d'erreur = 30.32787

5.3.3 CALCULEZ LA SENSIBILITE

5.3.4 LA SENSIBILITE = PROPORTION D'EVENEMENTS PREDITS PARMI LES REELS

```
(sensibilite <- (16/(16 + 135))*100)
```

Sensibilité = 10.59603

La probabilité de prévoir l'hypotrophie quand le poids du nouveau-né est anormalement faible vaut 10.6%

5.3.5 CALCULEZ LA SPECIFICITE

5.3.6 LA SPECIFICITE = PROPORTION DE NON-EVENEMENTS PREDITS PARMIS LES REELS

(specificite <- (324/(13 + 324))*100)

Spécificité = 96.14243

La probabilité de ne pas prévoir l'hypotrophie quand il n'y a pas d'hypotrophie néonatale vaut 96.14%

5.3.7 CALCULEZ LES VALEURS PREDICTIVES POSITIVES.

5.3.8 LA VALEUR PREDICTIVE POSITIVE = PROPORTION D'EVENEMENTS REELS PARMIS LES PREDITS

(val.pred.positive <- (16/(16 + 13))*100)

Valeur prédictive positive = 55.17241

La probabilité d'hypotrophie néonatale quand le test est positif (hypotrophie prévue) est de 55.17%

5.3.9 CALCULEZ LES VALEURS PREDICTIVES NEGATIVES.

5.3.10 LA VALEUR PREDICTIVE NEGATIVE = PROPORTION DE NON-EVENEMENTS REELS PARMIS LES PREDITS

(val.pred.negative <- (324/(135 + 324))*100)

Valeur prédictive négative = 70.58824

La probabilité de ne pas être malade (pas d'hypotrophie) quand le test est négatif (pas d'hypotrophie prévue) est de 70.59%