Devoir MAN Statistiques 2016

**Sarah FELDMAN**

30 octobre 2016

# Q1/Dans un premier temps décrivez vos variables.

## Nombre de variables

Le jeu de données présentation retinol fait 315 lignes et 14 colonnes ou variables.

## Type de variable

Les variables sont toutes reconnues comme quantitatives mais en regardant plus attentivement les min et max on peut voir que certaines variables ne sont pas quantitatives:

* sexe : variable binaire

sexe a un minimum de 1 et un maximum de 2 c'est donc une variables qualitative binaire. Je la recode en 0(homme)/1(femme) pour une interprétation plus facile par la suite. Je m'assure que c'est bien une valeur numérique pour pouvoir l'intégrer dans ma matrice de corrélation en question 2.

* tabac et vitamine : variables qualitatives ordinales

Les variables tabac et vitamine sont des variables qualitatives ordinales, il y a bien un ordre dans les classes mais je ne peux pas dire si l'espace entre 2 classes est le même. Pour tabac par exemple : ne pas fumer versus fumer anciennement aura-t-elle la même différence d'effet que fumer anciennement versus fumer actuellement? Je préfère donc les coder comme des variables qualitatives en attendant d'en savoir plus. Pour cela je les transforme simplement en facteur.  
Je garde comme classe de référence "souvent" pour vitaline et "jamais" pour tabac car c'est ce qui a pour moi le plus de sens médicalement parlant. Je n'indique pas qu'elles sont ordonnées car je veux les analyser d'abord comme étant non ordonnées.

* Les autres variables sont quantitatives continues: age, bmi, calories, graisses, fibres, alcool, cholestérol, betadiet,retdiet,betaplasma,retplasma

|  |  |  |
| --- | --- | --- |
| Variables quantitative continues | Variables qualitatives | Variable binaire |
| age | tabac | sexe |
| bmi | vitamine |  |
| calories |  |  |
| graisses |  |  |
| fibres |  |  |
| alcool |  |  |
| cholesterol |  |  |
| betadiet |  |  |
| retdiet |  |  |
| betaplasma |  |  |
| retplasma |  |  |

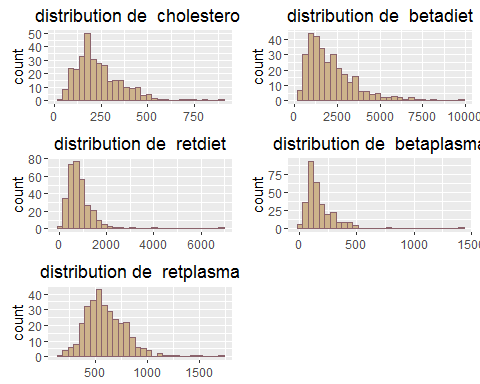
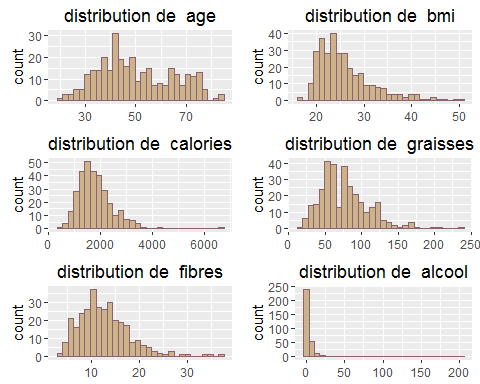
## Données abberrantes

Observer ses variables sert également à dépister les "bizarreries". Par exemple, j'ai un doute concernant le max d'alcool, je vais donc regarder plus attentivement la variable, par exemple avec une table.

Je passe de 35 à 203 verres par semaine...203/7=29 verres par jour, il y a probablement un erreur de codage.

La valeur extrême d'alcool influe beaucoup sur la symétrie de ma distribution. On peut le voir notamment en comparant la moyenne (3.28) et la médiane (0.3). Cependant je ne sais pas si je peux la retirer sans conséquence, je préfère donc la garder pour la suite du devoir.

## Distribution des variables



Lecture des histogrammes :

* bmi, graisses, cholestérol, betadiet, retdiet et betaplasme on des distribution à peu près normales mais asymétriques.
* calories, fibres, ret plasma ont de distributions d'allure normales
* age a une distribution irrégulière avec 2 cloches à 40 et 75 ans.
* alcool a une distribution très asymétrique qui ne semble pas normale

|  |  |
| --- | --- |
| Variables quantitative normales | Variables quantitative non normales |
| retplasma | age |
| bmi | alcool |
| cholesterol |  |
| retdiet |  |
| graisses |  |
| betadiet |  |
| betaplasma |  |
| calories |  |
| fibres |  |

## Voici un résumé des différentes variables :

* pour les variables quantitatives :

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| variables | n | missing | moyenne | median | q1 | q3 | rmin | max | Distribution normale |
| age | 315 | 0 | 50.1 | 48.0 | 39.0 | 62.5 | 19.0 | 83.0 | non |
| bmi | 315 | 0 | 26.2 | 24.7 | 21.8 | 28.9 | 16.3 | 50.4 | oui |
| calories | 315 | 0 | 1796.7 | 1666.8 | 1338.0 | 2100.4 | 445.2 | 6662.2 | oui |
| graisses | 315 | 0 | 77.0 | 72.9 | 54.0 | 95.2 | 14.4 | 235.9 | oui |
| fibres | 315 | 0 | 12.8 | 12.1 | 9.1 | 15.6 | 3.1 | 36.8 | oui |
| alcool | 315 | 0 | 3.3 | 0.3 | 0.0 | 3.2 | 0.0 | 203.0 | non |
| cholesterol | 315 | 0 | 242.5 | 206.3 | 155.0 | 308.9 | 37.7 | 900.7 | oui |
| betadiet | 315 | 0 | 2185.6 | 1802.0 | 1116.0 | 2836.0 | 214.0 | 9642.0 | oui |
| retdiet | 315 | 0 | 832.7 | 707.0 | 480.0 | 1037.0 | 30.0 | 6901.0 | oui |
| betaplasma | 315 | 0 | 189.9 | 140.0 | 90.0 | 230.0 | 0.0 | 1415.0 | oui |
| retplasma | 315 | 0 | 602.8 | 566.0 | 466.0 | 716.0 | 179.0 | 1727.0 | oui |

* pour les variables qualitatives:

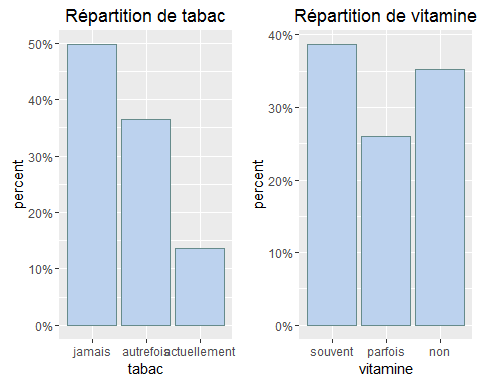
|  |  |  |
| --- | --- | --- |
| vitamine | Fréquence | Pourcentage |
| souvent | 122 | 39 |
| parfois | 82 | 26 |
| non | 111 | 35 |
| N | 315 | 100 |
| missing | 0 | 0 |

|  |  |  |
| --- | --- | --- |
| tabac | Fréquence | Pourcentage |
| jamais | 157 | 50 |
| autrefois | 115 | 37 |
| actuellement | 43 | 14 |
| N | 315 | 100 |
| missing | 0 | 0 |

|  |  |  |
| --- | --- | --- |
| sexe | Fréquence | Pourcentage |
| 0 (homme) | 42 | 13 |
| 1 (femme) | 273 | 87 |
| N | 315 | 100 |
| missing | 0 | 0 |

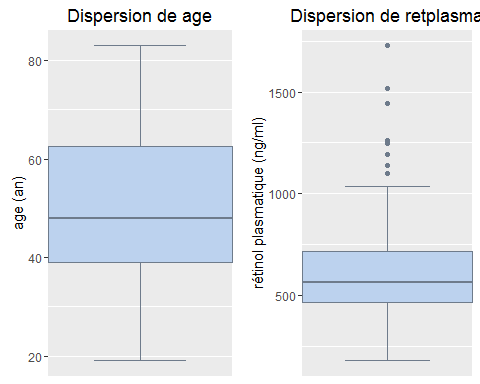
## Exemples de représentation graphique des variables :

### Diagrammes pour tabac et vitamine

Je veux représenter des pourcentages, un diagramme est pour moi plus parlant qu'un camembert.

### 

### Boîtes à moustache pour les variables quantitatives comme age ou ret plasma



# 

# Q2/Etudiez les relations existant entre toutes les paires possibles de variables.

NB : ce ne sont que les 9 variables de la régression demandées en question 3 qui sont concernées.

## Matrice de corrélation

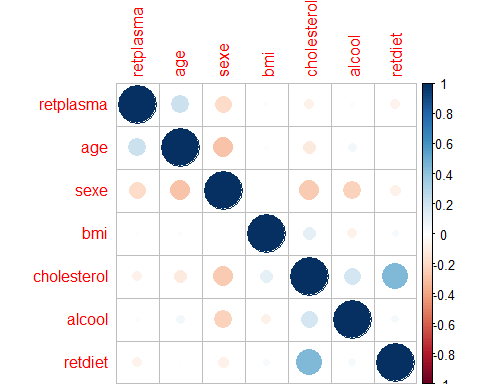
Je peux faire une matrice de corrélation. Je n'inclue ni vitamine ni tabac dans cette matrice car je les ai considérées comme des variables qualitatives à plusieurs classes. Je garde sexe qui est binaire.

Aucune condition n'est nécessaire pour faire des coefficients de corrélation. C'est pour les tester que nous avons besoin de vérifier les confidions de validité.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | retplasma | age | sexe | bmi | cholesterol | alcool | retdiet |
| retplasma | 1.000 | 0.212 | -0.184 | 0.013 | -0.070 | 0.017 | -0.063 |
| age | 0.212 | 1.000 | -0.280 | -0.017 | -0.114 | 0.052 | -0.010 |
| sexe | -0.184 | -0.280 | 1.000 | -0.007 | -0.255 | -0.228 | -0.074 |
| bmi | 0.013 | -0.017 | -0.007 | 1.000 | 0.110 | -0.073 | 0.032 |
| cholesterol | -0.070 | -0.114 | -0.255 | 0.110 | 1.000 | 0.182 | 0.443 |
| alcool | 0.017 | 0.052 | -0.228 | -0.073 | 0.182 | 1.000 | 0.045 |
| retdiet | -0.063 | -0.010 | -0.074 | 0.032 | 0.443 | 0.045 | 1.000 |

Il faut ensuite interprêter la matrice. Pour cela je peux faire des schémas:

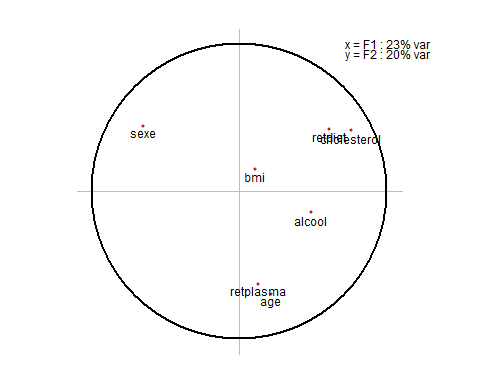
**Visualisation de la matrice de corrélation**



Dans cette visualisation de la matrice de corrélation, on voit surtout la corrélation entre cholestérol et retdiet car les cercles ont une densité élevée et on voit que la corrélation est positive car le cercle est de couleur bleu.

Je trouve que la représentation graphique de l'analyse en composantes principales (ACP) s'interprête plus facilement :

**Analyse en composantes principales**



* Lecture de l'ACP :
  + bmi est très proches du centre du cercle donc non interprétable.
  + les paires retdiet-cholestérol, retplasma-age sont fortement associés (corrélation positive), et ces deux groupes de variables sont indépendants l'un de l'autre car forme un angle droit avec le centre.
  + sexe est également indépendant de retdiet et cholestérol.
  + alcool-sexe est négativement corrélée.
* Je peux aussi sélectionner dans la matrice les valeurs absolues supérieures ou égales à certain niveau de corrélation :

|  |  |  |
| --- | --- | --- |
| variable 1 | variable 2 | coefficient de corrélation |
| Age | retplasma | 0.212 |
| Sexe | age | -0.280 |
| Cholesterol | sexe | -0.255 |
| alcool | sexe | -0.228 |
| retdiet | cholesterol | 0.443 |

4 couples ont un coefficient de corrélation entre 0.2 et 0.4 (en nombre absolu).  
age-retplasma, sexe-age, cholesterol-sexe, alcool-sexe.

Il n'y a qu'un couple avec un coefficient de corrélation supérieur ou égal à 0.4 :  
retdiet-cholestérol

## Tests de corrélation

### Pertinence

* Est-ce pertinent de faire de tests de corrélation pour chaque variable?  
  On peut se poser la question, en effet je n'avais aucune hypothèse de départ quant à ces corrélations et multiplier le nombre de tests augmente le risque alpha. Cependant pour les besoins du devoir je le fais quand même, mais je ne testerai que les corrélations supérieur à 0.2. En effet je ne sais pas quel sens je donnerai à une corrélation significative de 0.01 par exemple...

### Conditions de validité

Avant de faire une test de corrélation, il faut tester les conditions de validité :

* Une des 2 variables du couple testé doit suivre une loi normale.

Je considère comme normale une variable dont l'histogramme montre une distribution en cloche. Lorsque la distribution ne semble pas normale, il faut interpréter les tests avec prudence. Je ne préfère pas faire des tests non paramétriques type test de corrélation de spearman, ni sur certaines variables pour garder une cohérence, ni sur toutes les variables pour ne pas m'empêcher de faire des tests paramétriques par la suite.

### 

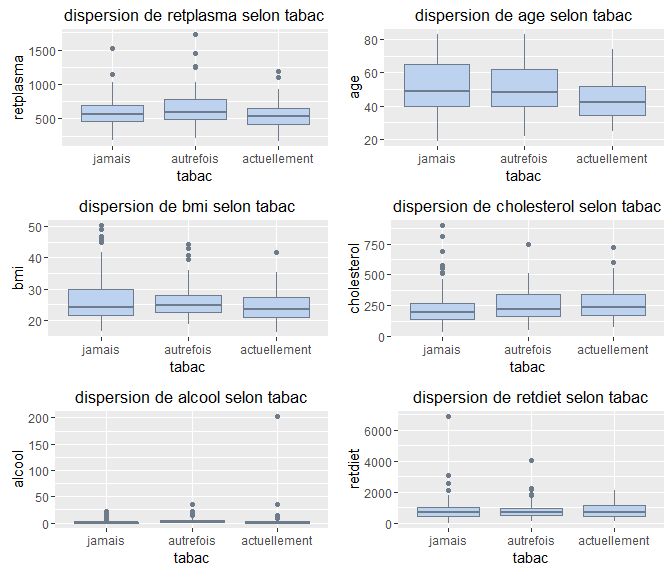
### Résultats

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| var1 | var2 | r | testvalid | test |
| age | retplasma | 0.212 | TRUE | 0.00015 |
| sexe | age | -0.280 | FALSE | 0.00000 |
| cholesterol | sexe | -0.255 | TRUE | 0.00000 |
| alcool | sexe | -0.228 | FALSE | 0.00005 |
| retdiet | cholesterol | 0.443 | TRUE | 0.00000 |

Toutes les corrélations supérieures ou égales à 0.2 sont significatives car p<= 0.05. Il faut cependant prêter attention au fait que 2 couples ont des conditions de validité probablement non remplies : sexe-age et sexe-alcool car age et alcool n'ont pas une allure normale (et sexe est binaire donc ne peut pas être normale).

## étude des liens entre une variable qualitative à plusieurs classe et des variables quantitatives:

### Graphiquement: je regarde la dispersion des variables quantitatives dans les différentes classes grâce à des boîtes à moustache.



### 

### Comparaison de moyenne entre plusieurs groupes (plus de 2) : ANOVA

#### Conditions de validité :

* Variance du même ordre de grandeur dans tous les sous groupes :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| tabac | retplasma | age | bmi | cholesterol | alcool | retdiet |
| jamais | 188 | 15 | 7 | 134 | 3 | 659 |
| autrefois | 231 | 14 | 5 | 122 | 6 | 530 |
| actuellement | 207 | 14 | 5 | 146 | 31 | 468 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| vitamine | retplasma | age | bmi | cholesterol | alcool | retdiet |
| souvent | 224 | 15 | 6 | 151 | 4 | 557 |
| parfois | 192 | 14 | 6 | 100 | 5 | 463 |
| non | 205 | 14 | 6 | 131 | 20 | 702 |

C'est le cas partout sauf pour la variable alcool. Je préfère donc ne pas faire de test avec alcool.

* Distribution normale :  
  Je fais l'approximation que toutes les variables suivent une loi normale, d'autant plus que l'anova est un test qui résiste bien à des ditribution qui s'éloignent un peu de la normale.

#### Je fais donc des ANOVA entre ma variable qualitative (vitamine ou tabac) et mes variables quantitatives (alcool exclue)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| variable\_qualitative | retplasma | age | bmi | cholesterol | retdiet |
| tabac | 0.024 | 0.024 | 0.128 | 0.109 | 0.899 |
| vitamine | 0.782 | 0.013 | 0.310 | 0.826 | 0.938 |

interprétation :

* tabac est significativement associée à retplasma et age.
* vitamine est significativement associée à age.

### Etude du lien entre variables qualitatives

Je fais un test du chi 2 pour les couples :

* tabac-vitamine
* tabac-sexe
* sexe-vitamine

#### Condition de validité : Effectifs théoriques supérieurs à 5

Tous les effectifs théoriques sont supérieurs à 5 pour les 3 tableaux de contingence, je peux faire un chi2.

#### Test du Chi2

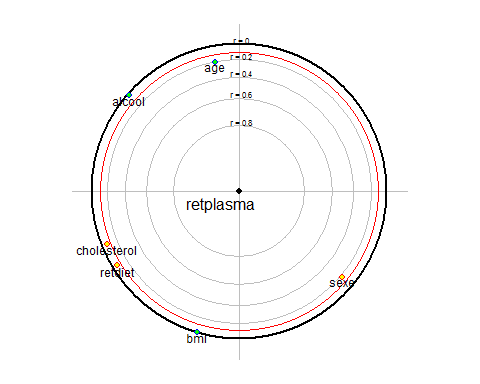
Le seuil de significativité est p<=0.05 :

* tabac et vitamine ne sont pas significativement liées : 0.065
* tabac et sexe sont significativement liées : p value = 0.028
* sexe et vitamine sont significativement liées : p value = 0.004

# Q3/Effectuez ensuite une régression linéaire où la variable à expliquer sera la concentration en rétinol plasmatique, les autres variables étant explicatives. Recherchez des interactions entre les variables explicatives.

## Analyse en composante principale focalisée

Je peux commencer par regarder les interaction enter retplasma et les variables explicatives avec une acp focalisée. Je retire cependant vitamine et tabac car fpca se base sur une matrice de corrélation, il faut donc faire attention à l'interprêtation du schéma car les interactions seront peutêtre modifiées en rajoutant vitamine et tabac:



2 variables semblent significativement liée à retplasma : age et sexe.

* age-retplasma est un couple qui ressortait dans la matrice de corrélation avec un r=0.212 significativement différent de 0.
* sexe-retplasma ne ressortait pas car inférieur à 0.2, comme on peut le voir sur l'acp focalisée.

## régression linéaire multiple

La variable à expliquer retplasma étant une variable Quantitative, je peux faire une régression linéaire. Et dans la mesure où j'introduis plusieurs variables explicatives, ce sera une régression linéaire multiple. Vitamine et tabac feront parti du modèle, je me suis bien assurée auparavent de les coder en facteur afin qu'elles soit recodées automatiquement en (nclasses-1) variables binaires.

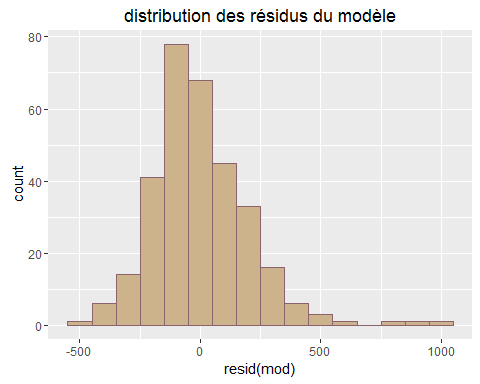
mod <- lm(retplasma ~ age + sexe + bmi + tabac + vitamine + cholesterol + alcool + retdiet,rt)

### Vérification des conditions de validité:

Il y a 3 conditions de validité aux modèles :

* la normalité de bruit
* la variance du bruit ne doit dépendre ni de la variable à expliquer ni de la variable explicative
* le bruit doit être du vrai bruit

En pratique on ne teste que la première des conditions :



La distribution des résidus a une allure normale, les conditions de validité de mon modèle sont vérifiées.

### Interprêtation des coefficients beta du modèle

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | estimateur | CI95 | pval | signif |
| (Intercept) | 578.50 | [400.7 - 756.3] | 0.000 | \* |
| age | 2.26 | [0.59 - 3.94] | 0.008 | \* |
| sexe | -98.18 | [-172.88 - -23.48] | 0.010 | \* |
| bmi | 1.12 | [-2.7 - 4.94] | 0.565 |  |
| tabacautrefois | 58.07 | [8.23 - 107.92] | 0.023 | \* |
| tabacactuellement | 1.12 | [-70.88 - 73.13] | 0.976 |  |
| vitamineparfois | -0.90 | [-58.77 - 56.97] | 0.976 |  |
| vitaminenon | -31.25 | [-85.03 - 22.53] | 0.254 |  |
| cholesterol | -0.13 | [-0.34 - 0.07] | 0.204 |  |
| alcool | -0.07 | [-1.99 - 1.85] | 0.940 |  |
| retdiet | -0.01 | [-0.06 - 0.03] | 0.533 |  |

Légende : \* p value <= 0.05 soit coefficient beta significativement différent de 0.

Interprétation :

* L'age et le sexe sont significativement associés à la concentration plasmatique en rétinol (p value<=0.05 et un intervalle de confiance à 95% ne contenant pas 0). Quand l'âge du sujet augmente d'un an, la concentration augmente de 2.26 unités. J'avais recodé les femmes en 1 et homme en 0, donc les femmes ont une concentration plasmatique en rétinol plus faible que les hommes d'environ 100 unités.
* Le coefficient de tabac "autrefois" est significativement différent de "jamais". ceux qui fumaient autrefois ont donc une concentration plasmatique en rétinol significativement plus élevée de 58 unités par rapport à ceux qui n'ont jamais fumé. Par contre on ne retrouve pas cette augmentation chez ceux qui fument actuellement.
* Les autres estimateurs ne sont pas significatifs.

### Je peux regarder si les variables tabac et vitamine sont globalement associées à retplasma:

**p values obtenues avec la fonction drop1 pour les variables tabac et vitamine.**

|  |  |
| --- | --- |
|  | p value |
| tabac | 0.055 |
| vitamine | 0.455 |

Ni tabac ni vitamine ne sont globablement associées à retplasma.

## Recherche d'interactions entre les variables explicatives:

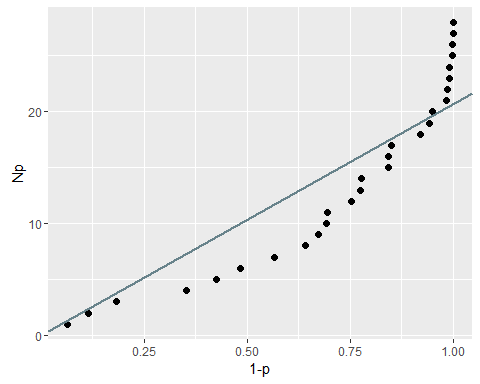
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Df | Sum.of.Sq | RSS | AIC | Pr..Chi. | signif |
| vitamine:alcool | 2 | 680454.0770 | 11740799 | 3341.692 | 0.0001 | \* |
| cholesterol:alcool | 1 | 477673.0449 | 11943580 | 3345.087 | 0.0004 | \* |
| bmi:alcool | 1 | 399093.0404 | 12022160 | 3347.152 | 0.0013 | \* |
| age:alcool | 1 | 339565.7626 | 12081688 | 3348.708 | 0.0031 | \* |
| sexe:bmi | 1 | 258033.5031 | 12163220 | 3350.827 | 0.0101 | \* |
| alcool:retdiet | 1 | 257913.8681 | 12163340 | 3350.830 | 0.0101 | \* |
| tabac:alcool | 2 | 327568.6014 | 12093685 | 3351.021 | 0.0149 | \* |
| sexe:alcool | 1 | 219309.5973 | 12201944 | 3351.828 | 0.0178 | \* |
| sexe:tabac | 2 | 234156.1788 | 12187097 | 3353.444 | 0.0499 | \* |
| sexe:vitamine | 2 | 221396.5505 | 12199857 | 3353.774 | 0.0589 |  |
| sexe:cholesterol | 1 | 120836.9181 | 12300417 | 3354.360 | 0.0793 |  |
| sexe:retdiet | 1 | 81694.8710 | 12339559 | 3355.361 | 0.1494 |  |
| cholesterol:retdiet | 1 | 78852.8292 | 12342401 | 3355.433 | 0.1567 |  |
| bmi:cholesterol | 1 | 78661.8340 | 12342592 | 3355.438 | 0.1572 |  |
| age:bmi | 1 | 58314.1212 | 12362939 | 3355.957 | 0.2234 |  |
| bmi:vitamine | 2 | 116638.7908 | 12304615 | 3356.467 | 0.2263 |  |
| age:sexe | 1 | 52682.3075 | 12368571 | 3356.100 | 0.2472 |  |
| bmi:tabac | 2 | 93217.8520 | 12328036 | 3357.066 | 0.3053 |  |
| age:tabac | 2 | 92611.7314 | 12328642 | 3357.082 | 0.3077 |  |
| vitamine:retdiet | 2 | 87846.3254 | 12333407 | 3357.204 | 0.3270 |  |
| age:retdiet | 1 | 33043.2127 | 12388210 | 3356.600 | 0.3597 |  |
| vitamine:cholesterol | 2 | 65621.2250 | 12355632 | 3357.771 | 0.4342 |  |
| tabac:vitamine | 4 | 127730.3757 | 12293523 | 3360.183 | 0.5159 |  |
| tabac:cholesterol | 2 | 43656.3939 | 12377597 | 3358.330 | 0.5743 |  |
| tabac:retdiet | 2 | 34190.9266 | 12387063 | 3358.571 | 0.6478 |  |
| age:cholesterol | 1 | 2104.6167 | 12419149 | 3357.386 | 0.8173 |  |
| age:vitamine | 2 | 9610.5602 | 12411643 | 3359.195 | 0.8852 |  |
| bmi:retdiet | 1 | 256.0591 | 12420997 | 3357.433 | 0.9358 |  |

Légende : \* p value <= 0.05 soit interaction significative

Interprétation :

J'ai 9 interactions significatives (p value<=0.05) : vitamine-alcool, cholestérol-alcool, bmi-alcool, age-alcool,sexe-bmi,alcool-retdiet, tabac-alcool,sexe-alcool,sexe-tabac.

### Je trace la représentation graphique des p values des 28 interactions possibles :



J'ai bien 9 points au dessus de la ligne qui sont mes 9 interactions significatives.

# Q4/Transformez la variable "rétinol plasmatique" en une variable binaire (en la coupant en deux au niveau de la médiane). Refaites les calculs précédents en ayant recours cette fois à une régression logistique.

Je transforme retplasma en variable binaire : les concentrations strictement inférieures à la médiane valent 0, celles supérieures ou égales valent 1.

## régression logistique : construction du modèle

La variable à expliquer est binaire, il faut donc faire une régression logistique

mod.bin <- glm(retplasma.bin ~ age + sexe + bmi + tabac + vitamine + cholesterol + alcool + retdiet, data = rt, family = "binomial")

### Vérification des conditions de validité

Pour que le modèle soit valide, il faut au moins 5 à 10 évènements par variable explicative. J'ai 6 variables explicatives comptant comme une variable, et 2 variables comptant comme n(classe) - 1 variables soit 2 pour vitamines et 2 pour tabac (sexe est binaire et ne compte donc que pour une variable). J'ai donc un total de 10 variables. 10*5=50 10*10=100

retplasma.bin : 157 sujets n'ont pas la variable, 158 l'ont. Donc 158 sujets ont des variables explicatives. 100 est inférieur à 158 donc le modèle est valable.

### Interprêtation des coefficients beta du modèle

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | estimate | CI95 | pval | signif |
| (Intercept) | -1.3105 | [-3.1405 - 0.4892] | 0.156 |  |
| age | 0.0282 | [0.0111 - 0.0458] | 0.001 | \* |
| sexe | -0.2852 | [-1.0668 - 0.4757] | 0.466 |  |
| bmi | 0.0101 | [-0.0286 - 0.049] | 0.608 |  |
| tabacautrefois | 0.2276 | [-0.2764 - 0.735] | 0.377 |  |
| tabacactuellement | 0.0514 | [-0.6833 - 0.7785] | 0.890 |  |
| vitamineparfois | -0.0360 | [-0.6211 - 0.5502] | 0.904 |  |
| vitaminenon | -0.2873 | [-0.8367 - 0.2573] | 0.302 |  |
| cholesterol | -0.0003 | [-0.0024 - 0.0018] | 0.773 |  |
| alcool | 0.0004 | [-0.0209 - 0.0252] | 0.967 |  |
| retdiet | 0.0000 | [-5e-04 - 4e-04] | 0.919 |  |

Seul le coefficient beta de age est significativement différent de 0 (c'est également le seul à avoir un intervalle de confiance à 95% ne contenant pas 0).

Je ne peux pas interpréter les coefficients tels quels, j'interprète uniquement l'exponentiel de ces coefficients qui sont égales aux Odds ratio.

**Tableau des odds ratio**

|  |  |  |  |
| --- | --- | --- | --- |
|  | OR | CI95 | signif |
| (Intercept) | 0.270 | [0.043 - 1.631] |  |
| age | 1.029 | [1.011 - 1.047] | \* |
| sexe | 0.752 | [0.344 - 1.609] |  |
| bmi | 1.010 | [0.972 - 1.05] |  |
| tabacautrefois | 1.256 | [0.759 - 2.085] |  |
| tabacactuellement | 1.053 | [0.505 - 2.178] |  |
| vitamineparfois | 0.965 | [0.537 - 1.734] |  |
| vitaminenon | 0.750 | [0.433 - 1.293] |  |
| cholesterol | 1.000 | [0.998 - 1.002] |  |
| alcool | 1.000 | [0.979 - 1.026] |  |
| retdiet | 1.000 | [1 - 1] |  |

légende : \* pvalue <= 0.05 soit OR significativement différent de 1 NB: pour retdiet l'intervalle de confiance est [1-1] à cause de l'arrondi mais contient 1 en réalité.

retplasma.bin n'est pas rare(car j'ai pris la médiane pour le construire...) donc les OR ne peuvent pas être interprétés comme des risques relatifs. On voit cependant que le coefficient de l'âge a beau être significatif, le rapport des odds est très proche de 1 donc l'âge est significativement différent mais de très très peu...

### Je peux regarder si les variables tabac et vitamine sont globalement associées à retplasma.bin:

**p values obtenues avec la fonction drop1 pour les variables tabac et vitamine.**

|  |  |
| --- | --- |
|  | p value |
| tabac | 0.667 |
| vitamine | 0.548 |

Ni tabac ni vitamine ne sont globablement associées à retplasma.

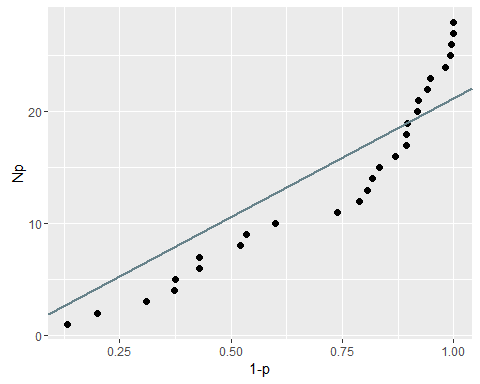
## Recherche d'interactions entre les variables explicatives:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Df | Deviance | AIC | LRT | Pr..Chi. | signif |
| age:alcool | 1 | 408.5431 | 432.5431 | 11.5653979 | 0.0007 | \* |
| vitamine:alcool | 2 | 406.6623 | 432.6623 | 13.4461750 | 0.0012 | \* |
| sexe:alcool | 1 | 412.1654 | 436.1654 | 7.9430772 | 0.0048 | \* |
| cholesterol:alcool | 1 | 412.9150 | 436.9150 | 7.1934532 | 0.0073 | \* |
| tabac:alcool | 2 | 412.1586 | 438.1586 | 7.9498765 | 0.0188 | \* |
| vitamine:retdiet | 2 | 414.2083 | 440.2083 | 5.9001197 | 0.0523 |  |
| sexe:cholesterol | 1 | 416.5274 | 440.5274 | 3.5810981 | 0.0584 |  |
| alcool:retdiet | 1 | 417.0398 | 441.0398 | 3.0686293 | 0.0798 |  |
| cholesterol:retdiet | 1 | 417.0518 | 441.0518 | 3.0566627 | 0.0804 |  |
| bmi:alcool | 1 | 417.4791 | 441.4791 | 2.6293068 | 0.1049 |  |
| sexe:vitamine | 2 | 415.6189 | 441.6189 | 4.4895862 | 0.1059 |  |
| age:sexe | 1 | 417.5073 | 441.5073 | 2.6011927 | 0.1068 |  |
| tabac:vitamine | 4 | 413.0035 | 443.0035 | 7.1049567 | 0.1304 |  |
| sexe:retdiet | 1 | 418.1929 | 442.1929 | 1.9155134 | 0.1664 |  |
| bmi:vitamine | 2 | 416.7095 | 442.7095 | 3.3989931 | 0.1828 |  |
| sexe:bmi | 1 | 418.4180 | 442.4180 | 1.6904666 | 0.1935 |  |
| age:bmi | 1 | 418.5490 | 442.5490 | 1.5594369 | 0.2117 |  |
| tabac:retdiet | 2 | 417.4232 | 443.4232 | 2.6852380 | 0.2612 |  |
| age:vitamine | 2 | 418.2739 | 444.2739 | 1.8345122 | 0.3996 |  |
| age:cholesterol | 1 | 419.5769 | 443.5769 | 0.5315997 | 0.4659 |  |
| bmi:tabac | 2 | 418.6334 | 444.6334 | 1.4750222 | 0.4783 |  |
| vitamine:cholesterol | 2 | 418.9844 | 444.9844 | 1.1240558 | 0.5701 |  |
| bmi:retdiet | 1 | 419.7874 | 443.7874 | 0.3210643 | 0.5710 |  |
| age:tabac | 2 | 419.1693 | 445.1693 | 0.9391251 | 0.6253 |  |
| age:retdiet | 1 | 419.8734 | 443.8734 | 0.2350310 | 0.6278 |  |
| sexe:tabac | 2 | 419.3692 | 445.3692 | 0.7392477 | 0.6910 |  |
| tabac:cholesterol | 2 | 419.6653 | 445.6653 | 0.4431853 | 0.8012 |  |
| bmi:cholesterol | 1 | 420.0803 | 444.0803 | 0.0281585 | 0.8667 |  |

Légende : \* p value <= 0.05 soit coefficient beta significativement différent de 0.

Interprétation :  
J'ai 5 interactions significatives (p value<=0.05) : age-alcool, vitamine-alcool, sexe-alcool, cholesterol-alcool, tabac-alcool.

### Je trace la représentation graphique des pvalue des 28 interactions possibles



J'ai 9 points au dessus de la ligne mais 5 points qui se détachent du groupe, correspondant à mes 5 interactions significatives. L'interprétation n'est pas aisée.