

MASTER 2 – Méthodologie et Statistiques en  
Recherche Biomédicales – DEVOIR  
**Rapport du devoir portant sur les données  
TPretinol**

*Nisrine BENNOR*  
*M2 MSR – Novembre 2020*  
**Groupe 2**

# TABLES DES MATIÈRES

## ENNONCE

## DESCRIPTION DES VARIABLES

- Description des variables quantitatives
  - 1- Age .....3
  - 2- BMI .....4
  - 3- Alcool .....5
  - 4- Cholestérol .....6
  - 5- Retdiet : *rétinol consommé* .....7
  - 6- Retplasma : *Concentration en rétinol plasmatique* .....7
    - Tableau final décrivant les données précédentes .....8
    - Description des variables qualitatives .....8
      - 1- Sexe .....8
      - 2- Tabac et Vitamine .....8

## CROISEMENT DES VARIABLES .....9

- 1- Variable qualitative / qualitative ..... 10
- 2- Variable quantitative / qualitative .....10
  - *Variable quantitative / qualitative (binaire= 2 modalités)* .....10
    - *Sexe* .....10
  - *Variable quantitative / qualitative (catégorielle= plus de 2 modalités)* ..... 11
    - *Tabac* .....11
    - *Vitamines*.....11
- Variables quantitative / quantitative .....12
  - *Calcul de la matrice de corrélation entre les variables quantitatives* .....12

## REGRESSION LINÉAIRE MULTIPLE.....13

Modèle de régression linéaire sans interaction .....14

Modèle de régression linéaire avec interaction .....14

## REGRESSION LOGISTIQUE .....15

Modèle de régression logistique sans interaction ..... 15

Modèle de régression logistique avec interaction ..... 16

## ÉNONCE

A partir du jeu de données TPretinol, vous pouvez rechercher les relations existant entre la concentration en rétinol plasmatique d'une part et, d'autres part, l'âge, le sexe, le BMI (indice d'obésité), le tabagisme, la consommation alimentaire de vitamines, de cholestérol, d'alcool et de rétinol.

## PRÉSENTATION DU JEU DE DONNÉES

Le jeu de données contient 14 variables. Dans notre projet nous intéresserons à 9 variables :

- Variables qualitatives :
  - Le sexe (1 = masculin / 2 = féminin)
  - La consommation de tabac (1=jamais / 2=autrefois / 3= actuellement)
  - La consommation de vitamines (1=oui, souvent/ 2 =oui, pas souvent/ 3=non)
- Variables quantitatives :
  - L'âge (en années)
  - L'indice de masse corporelle ( $BMI = \text{poids} / (\text{taille}^2)$ )
  - La consommation d'alcool (nombre de verres d'alcool consommés par semaine)
  - La consommation de cholestérol (mg par jour)
  - La consommation de rétinol (mcg par jour)
  - La consommation de rétinol plasmatique (ng/ml)

## DESCRIPTION DES VARIABLES

### • Description des variables quantitatives

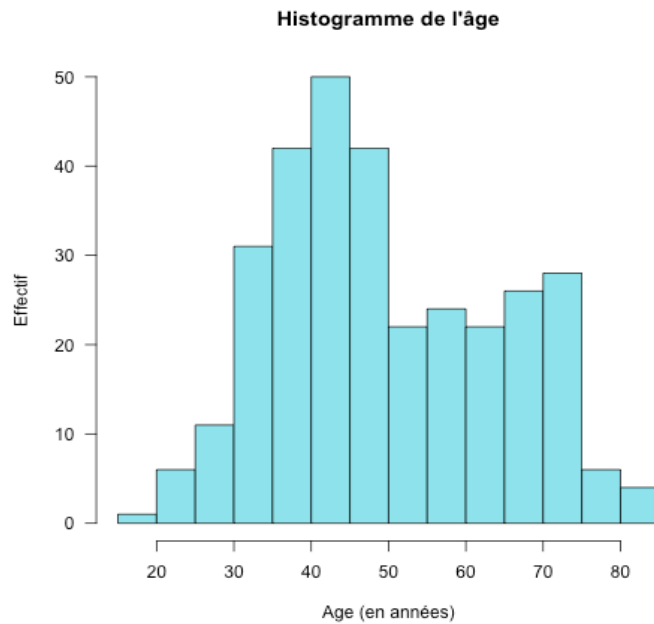
Nous sommes en présence de 14 variables, dont 3 sont qualitatives et 11 sont quantitatives. Nous sélectionnons que les 9 variables essentielles sur lesquelles nous travaillerons tout au long de ce devoir. Nous décrirons d'abord les variables quantitatives.

Pour chaque variable :

- ⇒ Calcul de l'effectif total, de la moyenne, la médiane, l'écart-type ainsi que le 1<sup>er</sup> et le 3<sup>ème</sup> quartiles. On présente aussi la valeur minimale et maximale.
- ⇒ L'utilisation de l'histogramme pour illustrer la répartition des variables quantitatives. (Cf. Code\_R en annexe)

### 1- Age :

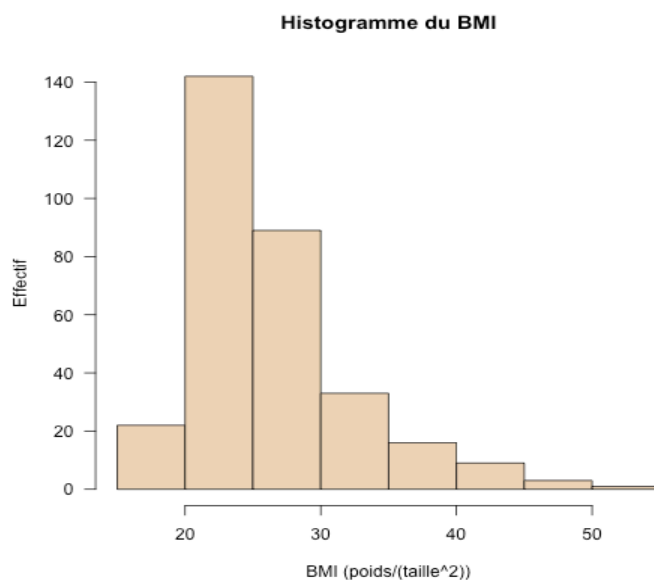
Dans notre jeu de données, l'âge s'étale de 19 à 83 ans, avec une moyenne d'environ 50 ans, et un écart-type de 14,58 ans.



Le graphique montre que l'âge semble suivre une distribution bimodale. Cela est peut-être due à une fluctuation de l'échantillon. Après, il est possible de faire une approximation par une loi normale.

## 2- BMI :

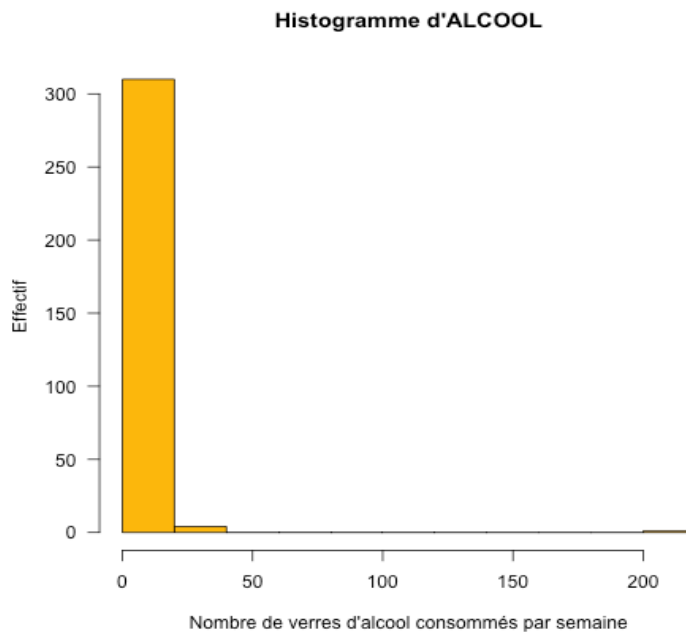
L'indice de masse corporelle, ou BMI, est compris entre *16,33* et *50,40* dans notre jeu de données. Sa moyenne est de *26,16* et son écart-type de *6,01*.



La distribution du BMI possède la forme d'une courbe en cloche. Néanmoins, elle est asymétrique, et donc vraisemblablement symétrique, le BMI ne suit pas une loi normale. Mais, il n'en est pas très loin, alors on pourra par la suite l'assimiler à une loi normale.

### 3- Alcool :

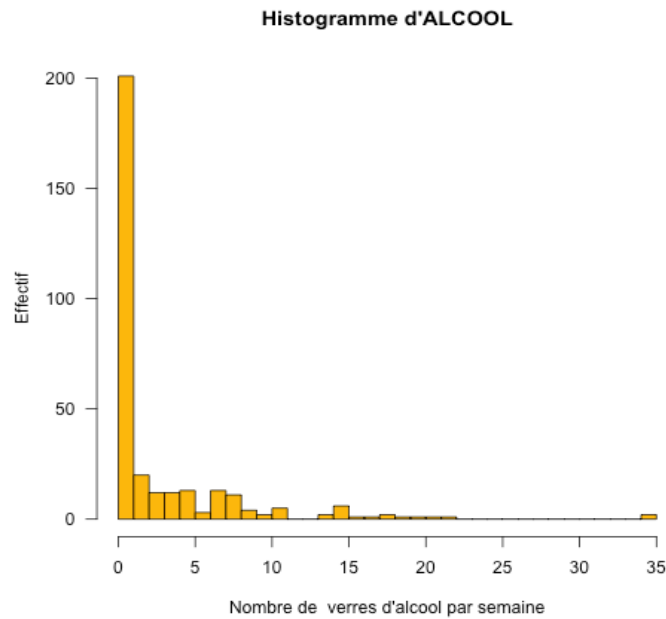
La variable alcool représente le nombre de verres d'alcool consommé par chaque individu en une semaine dont le minimum est 0 et le maximum est 203. Le nombre moyen de verres d'alcool consommés chaque semaine est plus au moins 3 verres, avec un écart-type de 12,32.



La présentation graphique obtenue est un peu particulière. L'inspection visuelle du graphique montre toute de suite la valeur extrême, de 203 valeurs par semaine. Est-ce qu'il s'agit d'une **donnée aberrante** ? Il est possible que ça soit vrai pour plusieurs raisons ; puisque cela est l'équivalent de 50.57 grammes d'alcools par litre dans le sang (selon « **la réglementation de l'alcool au volant /Sécurité routière en France** » chaque verre fait monter le taux d'alcool de 0.20g à 0.25g en moyenne quelle que soit la boisson alcoolisée consommée), soit à peu près 4 fois plus élevé que le record de 13.74g/l de sang en 2013, sachant qu'un taux à 3g/litre peut être mortel.

On propose de retirer la valeur maximale pour plus de précision puisqu'il s'agit que d'une seule valeur, et on réaffiche le graphique. La différence en la moyenne et la médiane est très grande. Il semble que cette valeur aberrante entraîne une élévation de la moyenne, nous considérons donc comme mieux la médiane qui est plus représentative, caractéristique et parlante.

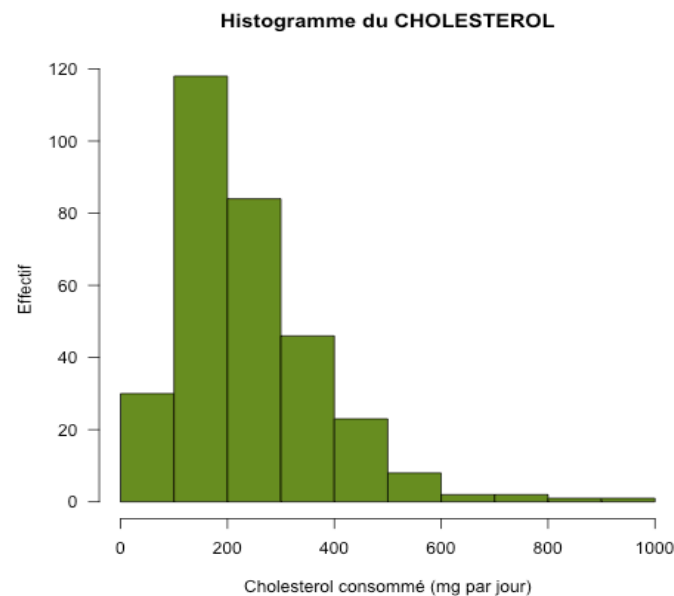
L'interprétation de ces résultats nécessite une grande prudence.



Effectivement, la plupart des individus ne consomment pas beaucoup d'alcool. La distribution de la variable alcool est très asymétrique. Et donc la distribution affichée ne semble pas suivre une loi normale.

#### 4- Cholestérol :

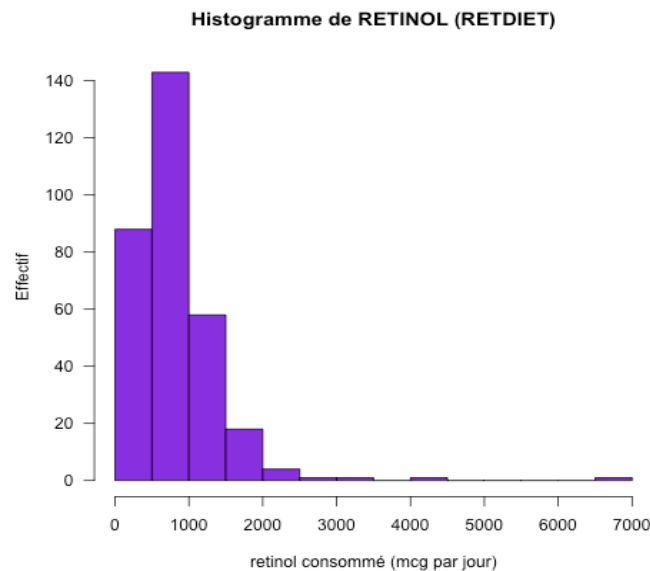
Les sujets étudiés consomment entre 37,7 et 900,7 mg de cholestérol par jour. La moyenne est de 242,46 et l'écart-type est de 132.



On remarque que le cholestérol possède la forme d'une courbe en cloche. Néanmoins, elle est un peu asymétrique. Mais notre présentation graphique n'est pas assez à l'extrême. Donc, on pourra par la suite l'approximer par une loi normale.

### 5- Retdiet : *rétinol consommé*

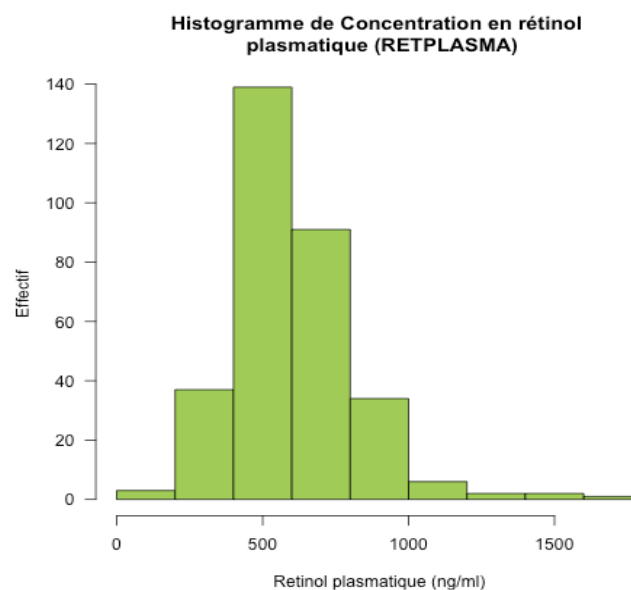
Le taux de rétinol consommé quotidiennement varie entre 30 et 6901(mcg) par jour, avec une moyenne de 832,71 et un écart-type de 589,3.



On remarque que le rétinol consommé possède la forme d'une courbe en cloche. Néanmoins, elle est un peu asymétrique. Donc, on pourra par la suite l'approximer par une loi normale.

### 6- Retplasma : *Concentration en rétinol plasmatique*

La concentration en rétinol plasmatique est entre 179 et 1727 ng/ml, avec une moyenne de 602,8 et un écart-type de 208,9.



Visiblement, Cette variable semble suivre une loi normale.

- **Tableau final décrivant les données précédentes**

<i>Variables</i>	<i>Effectifs d'observations</i>	<i>Moyenne</i>	<i>Médiane</i>	<i>Écart- type</i>	<i>Q1</i>	<i>Q3</i>	<i>Min</i>	<i>Max</i>
<i>Age</i>	315	50.15	48	14.58	39	62.5	19	83
<i>BMI</i>	315	26.16	24.47	6.01	21.80	28.85	16.33	50.40
<i>Alcool</i>	315	3.28	0.3	12.32	0	3.2	0	203
<i>Cholesterol</i>	315	242.46	206.3	132	155	308.85	37.7	900.7
<i>Retdiet</i>	315	832.71	707	589.3	480	1037	30	6901
<i>Retplasma</i>	315	602.8	566	208.9	466	716	179	1727

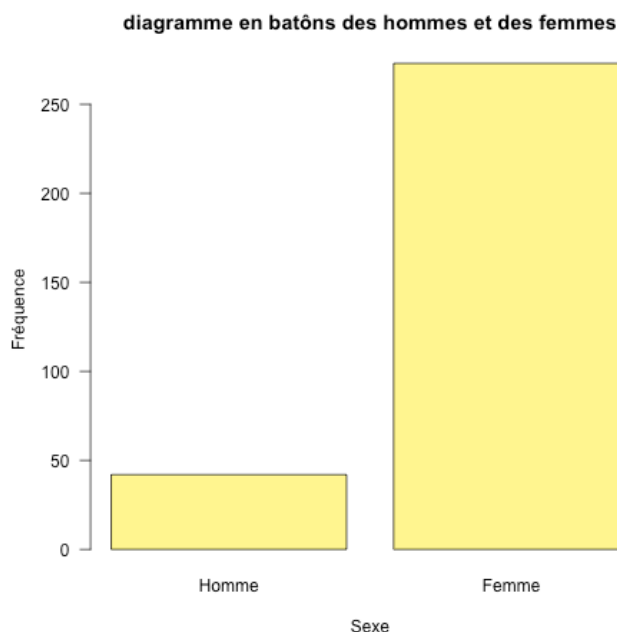
- **Description des variables qualitatives**

On a procédé pour la méthode suivante :

- ⇒ Calcul de l'effectif de chaque variable en fonction leur modalité.
- ⇒ Calcul de pourcentage pour chaque catégorie.
- ⇒ L'utilisation du diagramme « barplot ou diagramme en bâtons » pour illustrer la répartition des variables selon chaque modalité. (Cf. *Code\_R en annexe*)

### 1- Sexe :

Sur les 315 individus, 42 sont des hommes (13.33%) et 273 sont des femmes (86.67%).

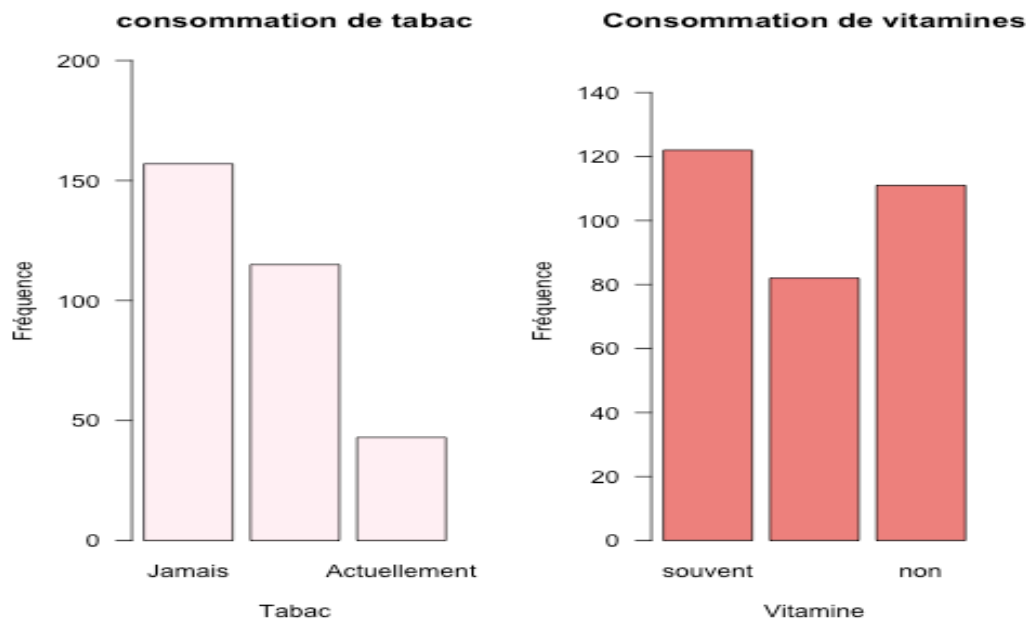


<i>Sexe</i>	<i>Hommes</i>	<i>Femmes</i>	<i>NA</i>
<i>Effectif</i>	42	273	0
<i>Pourcentage</i>	13.33%	86.67%	0%

### 2- Tabac et Vitamines :

Ces deux variables concernent la consommation de tabac et de vitamines. Elles sont qualitatives à trois modalités.





Tabac	Jamais	Autrefois	Actuellement
Effectif	157	115	43
Pourcentage	50%	36.51%	13.65%

Vitamines	Oui, souvent	Oui, pas souvent	Non
Effectif	157	115	43
Pourcentage	50%	36.51%	13.65%

On observe que presque 50% des sujets étudiés dans notre jeu de données n'ont jamais fumé, contre 13.65% de fumeurs actuels et 36.51% d'anciens fumeurs.

La consommation de vitamines est fréquente chez 38.57% des sujets, et pas souvent chez 26%, et 35.23% n'en consomment jamais.

## CROISEMENT DES VARIABLES

Nous allons examiner et étudier les relations existantes entre tous les couples de variables.

On va donc calculer le petit « p », qui représente la probabilité que le hasard puisse expliquer à lui seul une différence au moins aussi importante que celle qui a été observée. Et donc formellement de considérer que si cette probabilité petit « p » était plus petite que 5%, alors on pourrait dire que la différence est statistiquement significative (présence de significativité entre les 2 variables).

Nous privilégions l'utilisation de la théorie de Fisher, où « p » sera un moyen pour conclure et expliquer les résultats obtenus dans le rapport.

### 1- Variable qualitative / qualitative :

Les variables sont : Sexe – Tabac- Vitamines.

Pour tester la liaison entre deux variables qualitatives, on utilise le test du Chi-2, qui a pour condition de validité que les effectifs théoriques ne soient soit pas trop petits (Quelques dizaines). Pour notre jeu de données, l'effectif minimal est 42.

Tous les croisements possibles que nous allons réaliser remplissent les conditions de validité, le test du Chi-2 peut donc être appliqué.

***Voici le résumé des résultats du test Chi-2.***

<i>Var QL/QL</i>	<i>Sexe</i>	<i>Tabac</i>
<i>Tabac</i>	0.0282*	-----
<i>Vitamines</i>	0.003944*	0.06491

À un seuil de 5%, les tests mettent en évidence une liaison significative entre les variables Sexe et Tabac ( $p = 0,02$ ) et entre les variables Sexe et Vitamines ( $p = 0,004$ ). En revanche, il ne semble pas y avoir de relation entre les variables Tabac et Vitamines ( $p = 0,06$ ).

## **2- Variable quantitative / qualitative :**

### ○ Variable quantitative / qualitative (binaire= 2 modalités)

#### **• Sexe**

Le test de liaison entre une variable quantitative et une variable qualitative est un test de comparaison des moyennes de la variable quantitative en fonction des modalités de la variable qualitative.

Lorsque la variable qualitative a 2 modalités et que les effectifs dans chaque groupe sont supérieurs à 30, et l'ensemble des variables quantitatives peuvent être considérées comme suivant une loi normale, on utilise un **test de Student**. Reste aussi la vérification de l'égalité des écart-types (variances) entre les groupes à comparer (lorsque les variances sont très différentes, on utilise un **test de Welch**).

L'utilisation des boîtes à moustaches (Boxplot de ggplot2) pour illustrer la répartition des variables quantitatives par rapport aux variables qualitatives. (Cf. *Code\_R en annexe*).

C'est le cas ici de la variable Sexe (variable à 2 modalités masculin=1 et féminin=2), que nous allons croiser avec toutes les variables quantitatives.

### ***Résumé des résultats obtenus grâce aux tests:***

<i>Sexe</i>	<i>Age</i> <i>Moyenne</i> <i>(sd)</i>	<i>BMI</i> <i>Moyenne</i> <i>(sd)</i>	<i>Alcool</i> <i>Moyenne</i> <i>(sd)</i>	<i>Cholesterol</i> <i>Moyenne</i> <i>(sd)</i>	<i>Retdiet</i> <i>Moyenne</i> <i>(sd)</i>	<i>Retplasma</i> <i>Moyenne</i> <i>(sd)</i>
<i>Hommes</i>	60.54 (13.46)	26.27 (4.17)	14.42 (31.46)	328.12 (145.43)	944.14 (642.26)	700.73 (307.80)
<i>Femmes</i>	48.54 (14.09)	26.13 (6.25)	2.18 (4.11)	229.28 (124.96)	815.57 (580.07)	587.72 (185.43)
<i>p-value</i>	4.238e-07*	0.895	0.097	4.58e-06*	0.188	0.025*

À un seuil de 5%, les tests effectués montrent une liaison significative entre la variable Sexe et : Age ( $p < 0,001$ ), Cholestérol ( $p < 0,001$ ) et Concentration de rétinol plasmatique (*Retplasma*) ( $p = 0,025$ ).

On peut dire alors que la concentration de rétinol plasmatique chez les hommes est plus élevée que chez les femmes, ainsi que les hommes consomment beaucoup plus de cholestérol par jour que les femmes.

- Variable quantitative / qualitative (catégorielle= plus de 2 modalités)

Quand la variable qualitative a plus de 2 modalités, on a alors recours à une analyse de variance (ou ANOVA) qui permet de faire une comparaison de plusieurs moyennes.

- *Tabac*

**Résumé des résultats obtenus grâce aux tests:**

<i>Tabac</i>	<i>Age</i> Moyenne (sd)	<i>BMI</i> Moyenne (sd)	<i>Alcool</i> Moyenne (sd)	<i>Cholesterol</i> Moyenne (sd)	<i>Retdiet</i> Moyenne (sd)	<i>Retplasma</i> Moyenne (sd)
<i>Jamais</i>	51.22 (15.02)	26.72 (6.86)	1.67 (3.36)	228.39 (134.22)	838.85 (659.12)	583.30 (187.64)
<i>Autrefois</i>	50.77 (13.97)	25.92 (4.99)	3.84 (5.81)	250.42 (121.69)	838.75 (529.63)	644.24 (231.16)
<i>Actuellement</i>	44.53 (13.50)	24.69 (4.91)	7.63 (31.16)	272.53 (145.91)	794.13 (468.38)	563.06 (206.57)
<i>p-value</i>	0.023*	0.127	0.015*	0.108	0.899	0.023*

À un seuil de 5%, les tests ANOVA que nous avons réalisé montrent une liaison significative entre la variable Tabac et les variables : Age ( $p = 0,02$ ), Alcool ( $p = 0,015$ ) et concentration de rétinol plasmatique ( $p = 0,02$ ).

On peut conclure que les fumeurs sont plus jeunes que les non-fumeurs et les ex-fumeurs, et que les fumeurs consomment beaucoup plus d'alcool que les non-fumeur, ainsi que la concentration de rétinol plasmatique chez les ex-fumeurs est plus importantes que chez fumeurs.

- *Vitamines*

**Résumé des résultats obtenus grâce aux tests:**

<i>Vitamines</i>	<i>Age</i> Moyenne (sd)	<i>BMI</i> Moyenne (sd)	<i>Alcool</i> Moyenne (sd)	<i>Cholesterol</i> Moyenne (sd)	<i>Retdiet</i> Moyenne (sd)	<i>Retplasma</i> Moyenne (sd)
<i>Oui , souvent</i>	51.71 (15.29)	25.50 (5.81)	1.84 (3.75)	236.7 (151.1)	846.27 (556.52)	846.27 (556.52)
<i>Oui , pas souvent</i>	46.07 (13.92)	26.62 (6.46)	2.62 (5.1)	245.44 (99.62)	816.47 (462.65)	816.47 (462.65)
<i>Non</i>	51.43 (13.76)	26.52 (5.87)	5.33 (19.8)	246.6 (131.33)	829.81 (701.82)	829.81 (701.82)
<i>p-value</i>	0.0126*	0.309	0.083	0.826	0.937	0.782

À un seuil de 5%, une liaison significative entre la variable Vitamine et la variable Age ( $p = 0,012$ ). Il y a une différence de consommation des vitamines en fonction d'âge.

### 3- Variabes quantitative / quantitative :

Le test de liaison entre deux variables quantitatives s'effectue par le biais d'un test de nullité du coefficient de corrélation entre ces deux variables. Ce test est valable quand au moins une des deux distributions suit une loi normale. On a vu lors de la description des variables que la normalité est fortement contestable pour les variables BMI, Cholestérol et Retdiet (Rétinol consommé), et totalement exclue pour la variable Alcool. Nous avons cependant testé la nullité du coefficient de corrélation pour tous les couples de variables quantitatives. L'analyse portera sur p-value et la corrélation.

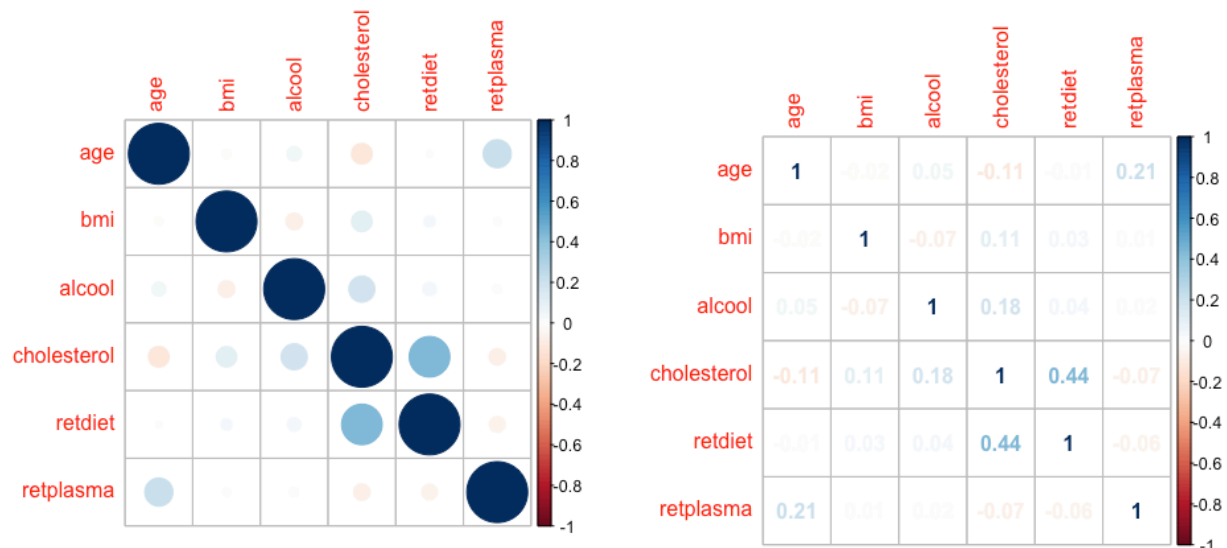
L'utilisation des matrices pour illustrer la corrélation entre les variables des variables quantitatives. La présentation de la distribution des variables a aussi été faite par diagramme « nuages des points ». (Cf. Code\_R en annexe)

<i><b>Var QT/QT</b></i>	<i><b>p-value</b></i>
Age / BMI	0.757
Age / Alcool	0.361
Age / Cholesterol	<b>0.043*</b>
Age / Retdiet	0.865
Age / Retplasma	<b>&lt; 0.001*</b>
BMI / Alcool	0.198
BMI / Cholesterol	0.05
BMI / Retdiet	0.57
BMI / Retplasma	0.816
Alcool / Cholesterol	<b>0.001*</b>
Alcool / Retdiet	0.426
Alcool / Retplasma	0.761
Cholesterol / Retdiet	<b>&lt; 0.001*</b>
Cholesterol / Retplasma	0.214
Retdiet / Retplasma	0.266

#### ○ *Calcul de la matrice de corrélation entre les variables quantitatives.*

À un seuil de 5%, ces tests ont permis de mettre en évidence une relation entre les variables : Cholestérol et Age ( $p = 0,04$ ), Cholestérol et Alcool ( $p = 0,001$ ), Cholestérol et Rétinol consommé ( $p < 0,001$ ), et Concentration de rétinol plasmatique et Age ( $p < 0,001$ ).

La relation entre la concentration du rétinol et le cholestérol est moyennement forte( $r=0.44$ ).

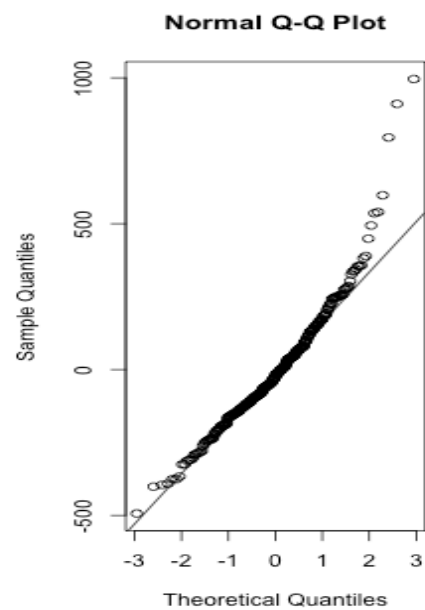
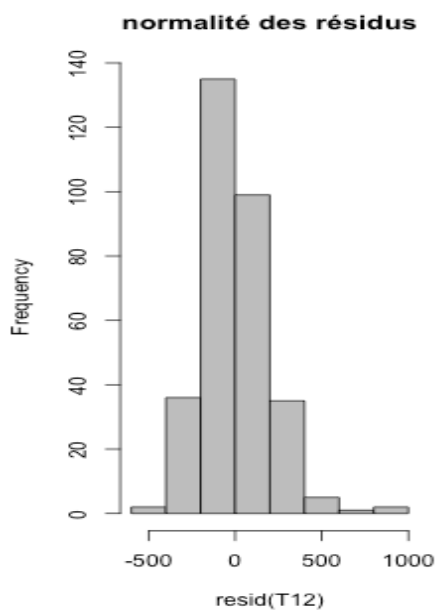


## REGRESSION LINÉAIRE MULTIPLE

On utilisera la variable « concentration en rétinol plasmatique(retplasma) » comme variable à expliquer et on réalisera des régressions linéaires multiples .

Notre **modèle de la régression linéaire multiple** sera de la forme :

- $\text{Retplasma} = a + b \cdot \text{Age} + c \cdot \text{Sexe} + d \cdot \text{Tabac} + e \cdot \text{bmi} + f \cdot \text{Vitamine} + g \cdot \text{Alcool} + h \cdot \text{Cholestérol} + i \cdot \text{Retdiet} + \varepsilon(\text{bruit})$
- Les conditions de validité sont respectées : La variable retplasma suit approximativement une loi normale, et les résidus suivent également une loi normale. (Utilisation de l'histogramme pour illustrer la normalité des résidus (Cf. Code\_R en annexe))



On voit sur l'histogramme que la distribution des résidus à l'air suivre une loi normale, et les points sont très proches de la droite de normalité sur le Q-Q Plot. On peut donc conclure à la validité du modèle de régression linéaire.

**Modèle de régression linéaire sans interaction :** (*extraits des résultats obtenus depuis R*)

	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>Pr(&gt; t )</i>
<i>(Intercept)</i>	578.50193	90.35333	6.403	5.78e-10 ***
<i>age</i>	2.26446	0.85257	2.656	0.00832 **
<i>Sexe_Femme</i>	-98.17895	37.96318	-2.586	0.01017 *
<i>Tabac_Autrefois</i>	58.07162	25.33008	2.293	0.02255 *

Le tableau contient les variables significatives.

À un seuil de 5%, on peut conclure que quand l'âge augmente d'un an, la concentration de rétinol plasmatique augmente également de 2.26ng/ml, et que les ex-fumeurs ont un taux élevés de concentration de rétinol plasmatique par rapport au fumeurs (autour de 58.07ng/ml), ainsi que le sexe féminin a une concentration de rétinol plasmatique inférieure à celle eu chez les hommes d'environ 98.17ng/ml.

**Modèle de régression linéaire sans interaction avec effet global – sans modalité de référence :** (*extraits des résultats obtenus depuis R*)

	<i>Df</i>	<i>Sum of Sq</i>	<i>RSS</i>	<i>AIC</i>	<i>F Value</i>	<i>Pr(&gt;F)</i>
<i>age</i>	1	288245	12709499	3360.7	7.0546	0.00832 **
<i>Sexe</i>	1	273277	12694531	3360.3	6.6882	0.01017 *
<i>Tabac</i>	2	239256	12660510	3357.4	2.9278	0.055025.

À un seuil de 5%, il n'y a pas d'effet, et donc il n'y pas de différence moyenne de la concentration de rétinol plasmatique en fonction de la consommation du tabac.

**Modèle de régression linéaire avec interaction :** (*extraits des résultats obtenus depuis R*)

Afin de savoir quelles sont les interactions à inclure dans le modèle, on commence par inclure dans le modèle qu'une interaction. Si l'effet de cette interaction est significatif dans ce modèle, on décide de la garder dans le modèle final. Et on répète alors la même l'opération pour toutes les interactions possibles entre les paires des variables de notre jeu de données.

<i>Interactions</i>	<i>p-value</i>
<i>Sexe</i>	BMI 0.0117367
<i>Alcool</i>	Age 0.0037838
	Sexe 0.0202679
	Tabac 0.0176757
	BMI 0.0016718
	Vitamine 0.0002020
	Cholesterol 0.0005727
	Retdiet 0.0117564

À un seuil de 5%, on affirme qu'il y a des interactions significatives entre les variables suivantes :

- Sexe / BMI
- Alcool / Age
- Alcool / Sexe
- Alcool / Tabac
- Alcool / BMI
- Alcool / Vitamine
- Alcool / Cholesterol
- Alcool /Retdiet

On peut dire que l'alcool présente plusieurs interactions avec d'autres variables dans notre jeu de données.

## REGRESSION LOGISTIQUE

La régression logistique permet de prédire une variable binaire. On transforme la variable Concentration en rétinol plasmatique (Retplasma) en variable binaire en attribuant la valeur 0 aux individus ayant une concentration en rétinol plasmatique inférieure à la valeur médiane (566ng/ml). Et pour les individus ayant une concentration supérieure ou égale à la valeur médiane reçoivent eux la valeur 1. (Cf. annexe Code\_R)

Notre **modèle de la régression logistique** sera de la forme :

- $\text{Log}[\text{prob}(\text{Retplasma\_Recode}=1)/1 - \text{prob}(\text{Retplasma\_Recode}=1)] = a + b \cdot \text{Age} + c \cdot \text{Sexe} + d \cdot \text{Tabac} + e \cdot \text{bmi} + f \cdot \text{Vitamine} + g \cdot \text{Alcool} + h \cdot \text{Cholestérol} + i \cdot \text{Retdiet}$
- Les conditions de validité sont respectées : il y a plus de 10 événements par variable explicative.
- On va calculé d'exponentiel du coefficient quand p\_value est inférieur à 0.05 .

### **Modèle de régression logistique sans interaction :** (extraits des résultats obtenus depuis R)

	<i>Estimate</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&lt; z )</i>
<i>Age</i>	2.821e-02	8.831e-03	3.194	0.0014 **

À un seuil de 5%, on affirme que l'âge est statistiquement lié à la variable concentration de rétinol plasmatique.

### **Modèle de régression logistique sans interaction avec effet global – sans modalité de référence :** (extraits des résultats obtenus depuis R)

	<i>Df</i>	<i>Deviance</i>	<i>AIC</i>	<i>LRT</i>	<i>Pr(Chi)</i>
<i>Age</i>	<i>1</i>	<i>430.71</i>	<i>450.71</i>	<i>10.6042</i>	<i>0.001128 **</i>

À un seuil de 5%, on affirme que l'âge est statistiquement lié à la variable concentration de rétinol plasmatique.

**Modèle de régression logistique avec interaction :** (*extraits des résultats obtenus depuis R*)

	<i>Interactions</i>	<i>p-value</i>
<i>Alcool</i>	<i>Age</i>	<i>0.0006719 ***</i>
	<i>Sexe</i>	<i>0.0048272 **</i>
	<i>Tabac</i>	<i>0.0187805 *</i>
	<i>Vitamine</i>	<i>0.0012028 **</i>
	<i>Cholesterol</i>	<i>0.0073170 **</i>

À un seuil de 5%, on affirme qu'il y a une interaction significative entre les variables suivantes :

- Alcool / Age
- Alcool / Sexe
- Alcool / Tabac
- Alcool / Vitamine
- Alcool / Cholesterol



## ANNEXE

### Script R

*### iii- Importation des données:*

```
TPretinol<- read.csv2("presentationTPretinol.csv")
```

*#### - déclaration des variables qualitatives*

```
TPretinol$sexe <- as.factor(TPretinol$sexe)
```

```
TPretinol$tabac <- as.factor(TPretinol$tabac)
```

```
TPretinol$vitamine <- as.factor(TPretinol$vitamine)
```

*##### DESCRIPTION DES VARIABLES QUANTITATIVES*

```
library(prettyR)
```

```
describe(TPretinol,num.desc=c("mean","median","var","sd","valid.n","min","max"))
```

*##Quantile age*

```
quantile(TPretinol$age,na.rm = TRUE)
```

*##Quantile bmi*

```
quantile(TPretinol$bmi,na.rm = TRUE)
```

*##Quantile cholesterol*

```
quantile(TPretinol$cholesterol,na.rm = TRUE)
```

*##Quantile alcool*

```
quantile(TPretinol$alcool,na.rm = TRUE)
```

*##Quantile retdiet*

```
quantile(TPretinol$retdiet,na.rm = TRUE)
```

*##Quantile retplasma*

```
quantile(TPretinol$retplasma,na.rm = TRUE)
```

*###Histogramme*

```
hist(TPretinol$age,main="Histogramme de l'âge", xlab="Age (en années)",ylab="Effectif", col="cadetblue2", las=1)
```

```
hist(TPretinol$bmi,main="Histogramme du BMI",xlab="BMI (poids/(taille^2))",ylab="Effectif",col="bisque2", las=1)
```

```
hist(TPretinol$alcool,main="Histogramme d'ALCOOL",xlab="Nombre de verres d'alcool consommés par semaine",ylab="Effectif",col="darkgoldenrod1", las=1)
```

```
table(TPretinol$alcool == 203)
```

```
hist(TPretinol$alcool[TPretinol$alcool<203],nclass=30,main="",xlab="Nombre de verres d'alcool par semaine",ylab="Effectif",col="darkgoldenrod1", las=1)
```

```

hist(TPretinol$cholesterol,main="Histogramme du
CHOLESTEROL",xlab="Cholesterol consommé (mg par
jour)",ylab="Effectif",col="darkgoldenrod1", las=1)
hist(TPretinol$retdiet,main="Histogramme de RETINOL (RETDIET)",xlab="retinol
consommé (mcg par jour)",ylab="Effectif",col="blueviolet", las=1)
hist(TPretinol$retplasma,main="Histogramme de Concentration en rétinol
plasmatique (RETPLASMA)",xlab="Retinol plasmatique
(ng/ml)",ylab="Effectif",col="darkolivegreen3", las=1)

```

#### #### DESCRIPTION DES VARIABLES QUALITATIVES

##### ## sexe

```

SEXE<-table(factor(TPretinol$sexe, levels = c(1,2), labels =
c("Homme","Femme")))
SEXE
prop.table(SEXE) * 100
barplot(SEXE, main = "diagramme en bâtons des hommes et des
femmes",xlab="Sexe",ylab="Fréquence", las=1, col = "khaki1")

```

```

par(mfrow=c(1,2))

```

##### ## Tabac

```

TABAC<-table(factor(TPretinol$tabac, levels = c(1,2,3), labels
=c("Jamais","Autrefois","Actuellement")))
TABAC
prop.table(TABAC) * 100

```

##### ##Vitamine

```

VITAMINE<-table(factor(TPretinol$vitamine, levels = c(1,2,3), labels
=c("souvent","pas souvent","non")))
VITAMINE
prop.table(VITAMINE) * 100

```

```

barplot(TABAC, main = "consommation de
tabac",xlab="Tabac",ylab="Fréquence", ylim = c(0,200), las=1, col =
"lavenderblush1")

```

```

barplot(VITAMINE, main = "Consommation de
vitamines",xlab="Vitamine",ylab="Fréquence", las=1,ylim = c(0,150), col =
"lightcoral")

```

#### #### CROISEMENTS DES VARIABLES:

##### ##### 1- Variables qualitative / qualitative

##### ##Sexe/Tabac

```
table(TPretinol$sexe,TPretinol$tabac,deparse.level = 2)
chisq.test(TPretinol$sexe,TPretinol$tabac,correct=F)
```

### *##Sexe/Vitamine*

```
table(TPretinol$sexe,TPretinol$vitamine, deparse.level = 2)
chisq.test(TPretinol$sexe,TPretinol$vitamine,correct=FALSE)
```

### *##Vitamine/Tabac*

```
table(TPretinol$vitamine,TPretinol$tabac,deparse.level = 2)
chisq.test(TPretinol$vitamine,TPretinol$tabac,correct=FALSE)
```

### *##V.QL : Sexe croisée avec les Autres V.QT*

#### *##Age/Sexe*

```
by(TPretinol$age,TPretinol$sexe,mean,na.rm=TRUE)
by(TPretinol$age,TPretinol$sexe,sd,na.rm=TRUE) #égalité var
t.test(TPretinol$age~TPretinol$sexe, var.equal=TRUE) # différence significative
```

#### *##Boxplot*

```
library(gplots)
boxplot(age~sexe,TPretinol)
```

#### *##OU*

```
library(ggplot2)
ggplot(TPretinol, aes(x=sexe, y=age, fill=sexe)) +
  geom_boxplot()+theme_grey()+
  ggtitle("L'age en fonction du sexe")
```

### *## Alcool/Sexe*

```
by(TPretinol$alcool,TPretinol$sexe,mean,na.rm=TRUE)
by(TPretinol$alcool,TPretinol$sexe,sd,na.rm=TRUE) #pas d'égalité var = Test
WELCH
t.test(TPretinol$alcool~TPretinol$sexe, var.equal=FALSE)# différence non
significative
ggplot(TPretinol, aes(x=sexe, y=alcool, fill=sexe)) +
  geom_boxplot()+scale_fill_manual(values=c("#999999",
"#E69F00"))+ylim(0,40)+theme_grey()+ ggtitle("Consommation quotidienne
d'alcool en fonction du sexe")
```

### *## Cholesterol/Sexe*

```
by(TPretinol$cholesterol,TPretinol$sexe,mean,na.rm=TRUE)
by(TPretinol$cholesterol,TPretinol$sexe,sd,na.rm=TRUE) #égalité var
```

```
t.test(TPretinol$cholesterol~TPretinol$sexe, var.equal=TRUE)# différence
significative
ggplot(TPretinol, aes(x=sexe, y=cholesterol, fill=sexe)) +
  geom_boxplot()+scale_fill_manual(values=c("#994499",
"#E69F22"))+theme_grey()+ ggtitle("Taux de cholesterol en fonction du sexe")
```

### *## BMI/Sexe*

```
by(TPretinol$bmi,TPretinol$sexe,mean,na.rm=TRUE)
by(TPretinol$bmi,TPretinol$sexe,sd,na.rm=TRUE) #égalité var
t.test(TPretinol$bmi~TPretinol$sexe, var.equal=TRUE)# différence non
significative
ggplot(TPretinol, aes(x=sexe, y=bmi, fill=sexe)) +
  geom_boxplot()+scale_fill_manual(values=c("#999974",
"#E69F88"))+theme_grey()+ ggtitle("BMI en fonction du sexe")
```

### *## RETDIET/Sexe*

```
by(TPretinol$retdiet,TPretinol$sexe,mean,na.rm=TRUE)
by(TPretinol$retdiet,TPretinol$sexe,sd,na.rm=TRUE) #égalité var
t.test(TPretinol$retdiet~TPretinol$sexe, var.equal=TRUE)# différence non
significative
ggplot(TPretinol, aes(x=sexe, y=retdiet, fill=sexe)) +
  geom_boxplot()+scale_fill_manual(values=c("#999823",
"#E69D88"))+theme_grey()+ ggtitle("Retinol consommé(mcg par jour) en
fonction du sexe")
```

### *## Retplasma/Sexe*

```
by(TPretinol$retplasma,TPretinol$sexe,mean,na.rm=TRUE)
by(TPretinol$retplasma,TPretinol$sexe,sd,na.rm=TRUE) #pas d'égalité var =
test WELCH
t.test(TPretinol$retplasma~TPretinol$sexe, var.equal=FALSE)# différence
significative
ggplot(TPretinol, aes(x=sexe, y=retplasma, fill=sexe)) +
  geom_boxplot()+scale_fill_manual(values=c("#998774",
"#E27F48"))+theme_grey()+ ggtitle("Retinol plasmatique(ng/ml) en fonction du
sexe")
```

### *### Tabac*

#### *#Tabac/Age*

```
by(TPretinol$age,TPretinol$tabac,mean,na.rm=TRUE)
by(TPretinol$age,TPretinol$tabac,sd,na.rm=TRUE)
T1<- lm(age~tabac,data=TPretinol)
```

```
summary(T1)
drop1(T1,~, test="F")#significative
boxplot(TPretinol$age~TPretinol$tabac,varwidth =TRUE, notch = TRUE, outline
= TRUE,
        col=(c("yellow4","yellow3","yellow1")),
        main="Consommation du tabac en fonction de l'âge",
xlab="Tabac",ylab="Age")
legend("bottomright", inset=.02, title="Consommation du tabac",
      c("Jamais","Autrefois","actuellement"), fill=c("yellow4","yellow3","yellow1"),
horiz=TRUE, cex=0.8)
```

### *#Tabac/BMI*

```
by(TPretinol$bmi,TPretinol$tabac,mean,na.rm=TRUE)
by(TPretinol$bmi,TPretinol$tabac,sd,na.rm=TRUE)
T2<- lm(bmi~tabac,data=TPretinol)
summary(T2)
drop1(T2,~, test="F")# non significative
boxplot(TPretinol$bmi~TPretinol$tabac,varwidth =TRUE, notch = TRUE, outline
= TRUE,
        col=(c("yellow4","yellow3","yellow1")),
        main="Consommation du tabac en fonction du bmi",
xlab="Tabac",ylab="BMI")
legend("bottomright", inset=.02, title="Consommation du tabac",
      c("Jamais","Autrefois","actuellement"), fill=c("yellow4","yellow3","yellow1"),
horiz=TRUE, cex=0.8)
```

### *#Tabac/Alcool*

```
by(TPretinol$alcool,TPretinol$tabac,mean,na.rm=TRUE)
by(TPretinol$alcool,TPretinol$tabac,sd,na.rm=TRUE)
T3<- lm(alcool~tabac,data=TPretinol)
summary(T3)
drop1(T3,~, test="F")#significative
boxplot(TPretinol$alcool~TPretinol$tabac,varwidth =TRUE, notch = TRUE,
outline = TRUE,
        col=(c("yellow4","yellow3","yellow1")),
        main="Consommation du tabac par rapport à la consommation de l'alcool
par semaine ", xlab="Tabac",ylab="Alcool")
legend("topleft", inset=.02, title="Consommation du tabac",
      c("Jamais","Autrefois","actuellement"), fill=c("yellow4","yellow3","yellow1"),
horiz=TRUE, cex=0.8)
```

### *#Tabac/Cholesterol*

```
by(TPretinol$cholesterol,TPretinol$tabac,mean,na.rm=TRUE)
by(TPretinol$cholesterol,TPretinol$tabac,sd,na.rm=TRUE)
T4<- lm(cholesterol~tabac,data=TPretinol)
summary(T4)
drop1(T4,~, test="F")# non significative
boxplot(TPretinol$cholesterol~TPretinol$tabac,varwidth =TRUE, notch = TRUE,
outline = TRUE,
      col=(c("yellow4","yellow3","yellow1")),
      main="Consommation du tabac par rapport au cholestérol (mg par jour)",
xlab="Tabac",ylab="Cholestérol (mg par jour)")
legend("topright", inset=.02, title="Consommation du tabac",
      c("Jamais","Autrefois","actuellement"), fill=c("yellow4","yellow3","yellow1"),
horiz=TRUE, cex=0.8)
```

### *#Tabac/RETDIET*

```
by(TPretinol$retdiet,TPretinol$tabac,mean,na.rm=TRUE)
by(TPretinol$retdiet,TPretinol$tabac,sd,na.rm=TRUE)
T5<- lm(retdiet~tabac,data=TPretinol)
summary(T5)
drop1(T5,~, test="F")#non significative
boxplot(TPretinol$retdiet~TPretinol$tabac,varwidth =TRUE, notch = TRUE,
outline = TRUE,
      col=(c("yellow4","yellow3","yellow1")),
      main="Consommation du tabac en fonction du rétinol",
xlab="Tabac",ylab="Rétinol consommé (mcg par jour)")
legend("topright", inset=.02, title="Consommation du tabac",
      c("Jamais","Autrefois","actuellement"), fill=c("yellow4","yellow3","yellow1"),
horiz=TRUE, cex=0.8)
```

### *#Tabac/RETPLASMA*

```
by(TPretinol$retplasma,TPretinol$tabac,mean,na.rm=TRUE)
by(TPretinol$retplasma,TPretinol$tabac,sd,na.rm=TRUE)
T5<- lm(retplasma~tabac,data=TPretinol)
summary(T5)
drop1(T5,~, test="F")#significative
boxplot(TPretinol$retplasma~TPretinol$tabac,varwidth =TRUE, notch = TRUE,
outline = TRUE,
      col=(c("yellow4","yellow3","yellow1")),
      main="Consommation du tabac en fonction du rétinol plasmatique (ng/ml) ",
xlab="Tabac",ylab="Rétinol plasmatique (ng/ml) ")
legend("bottomright", inset=.02, title="Consommation du tabac",
```

```
c("Jamais","Autrefois","actuellement"), fill=c("yellow4","yellow3","yellow1"),
horiz=TRUE, cex=0.8)
```

### ### Vitamine

#### #Vitamine/Age

```
by(TPretinol$age,TPretinol$vitamine,mean,na.rm=TRUE)
by(TPretinol$age,TPretinol$vitamine,sd,na.rm=TRUE)
T6<- lm(age~vitamine,data=TPretinol)
summary(T6)
drop1(T6,~, test="F")#significative
boxplot(TPretinol$age~TPretinol$vitamine,varwidth =TRUE, notch = TRUE,
outline = TRUE,
        col=(c("steelblue4","steelblue1","lightsteelblue1")),
        main="Vitamine en fonction de l'âge", xlab="vitamine",ylab="Age")
legend("bottomright", inset=.02, title="Vitamine",
        c("souvent","pas souvent","non"),
        fill=c("steelblue4","steelblue1","lightsteelblue1"), horiz=TRUE, cex=0.8)
```

#### #Vitamine/BMI

```
by(TPretinol$bmi,TPretinol$vitamine,mean,na.rm=TRUE)
by(TPretinol$bmi,TPretinol$vitamine,sd,na.rm=TRUE)
T7<- lm(bmi~vitamine,data=TPretinol)
summary(T7)
drop1(T7,~, test="F")#non significative
boxplot(TPretinol$bmi~TPretinol$vitamine,varwidth =TRUE, notch = TRUE,
outline = TRUE,
        col=(c("steelblue4","steelblue1","lightsteelblue1")),
        main="Vitamine en fonction du bmi", xlab="Vitamine",ylab="BMI")
legend("bottomright", inset=.02, title="Vitamine",
        c("souvent","pas souvent","non"),
        fill=c("steelblue4","steelblue1","lightsteelblue1"), horiz=TRUE, cex=0.8)
```

#### #Vitamine/Alcool

```
by(TPretinol$alcool,TPretinol$vitamine,mean,na.rm=TRUE)
by(TPretinol$alcool,TPretinol$vitamine,sd,na.rm=TRUE)
T8<- lm(alcool~vitamine,data=TPretinol)
summary(T8)
drop1(T8,~, test="F")# non significative
boxplot(TPretinol$alcool~TPretinol$vitamine,varwidth =TRUE, notch = TRUE,
outline = TRUE,
        col=(c("steelblue4","steelblue1","lightsteelblue1")),
```

```

    main="Vitamine par rapport à la consommation de l'alcool par semaine ",
    xlab="Vitamine",ylab="Alcool")
legend("topleft", inset=.02, title="Vitamine",
      c("souvent","pas souvent","non"),
    fill=c("steelblue4","steelblue1","lightsteelblue1"), horiz=TRUE, cex=0.8)

```

### *#Vitamine/Cholesterol*

```

by(TPretinol$cholesterol,TPretinol$vitamine,mean,na.rm=TRUE)
by(TPretinol$cholesterol,TPretinol$vitamine,sd,na.rm=TRUE)
T9<- lm(cholesterol~vitamine,data=TPretinol)
summary(T9)
drop1(T9,~., test="F")#non significative
boxplot(TPretinol$cholesterol~TPretinol$vitamine,varwidth =TRUE, notch =
TRUE, outline = TRUE,
      col=(c("steelblue4","steelblue1","lightsteelblue1")),
      main="Vitamine par rapport au cholestérol (mg par jour)",
    xlab="Vitamine",ylab="Cholesterol (mg par jour)")
legend("topright", inset=.02, title="Vitamine",
      c("souvent","pas souvent","non"),
    fill=c("steelblue4","steelblue1","lightsteelblue1"), horiz=TRUE, cex=0.8)

```

### *#Vitamine/RETDIET*

```

by(TPretinol$retdiet,TPretinol$vitamine,mean,na.rm=TRUE)
by(TPretinol$retdiet,TPretinol$vitamine,sd,na.rm=TRUE)
T10<- lm(retdiet~vitamine,data=TPretinol)
summary(T10)
drop1(T10,~., test="F")#non significative
boxplot(TPretinol$retdiet~TPretinol$vitamine,varwidth =TRUE, notch = TRUE,
outline = TRUE,
      col=(c("steelblue4","steelblue1","lightsteelblue1")),
      main="Vitamine en fonction du rétinol", xlab="Vitamine",ylab="Retinol
consommé (mcg par jour)")
legend("topright", inset=.02, title="Vitamine",
      c("souvent","pas souvent","non"),
    fill=c("steelblue4","steelblue1","lightsteelblue1"), horiz=TRUE, cex=0.8)

```

### *#Vitamine/RETPLASMA*

```

by(TPretinol$retdiet,TPretinol$vitamine,mean,na.rm=TRUE)
by(TPretinol$retdiet,TPretinol$vitamine,sd,na.rm=TRUE)
T11<- lm(retplasma~vitamine,data=TPretinol)
summary(T11)
drop1(T11,~., test="F")#non significative

```



```

boxplot(TPretinol$retplasma~TPretinol$vitamine,varwidth =TRUE, notch =
TRUE, outline = TRUE,
      col=(c("steelblue4","steelblue1","lightsteelblue1")),
      main="Vitamine en fonction du retinol plasmatique (ng/ml) ",
xlab="Vitamine",ylab="Retinol plasmatique (ng/ml) ")
legend("bottomright", inset=.02, title="Vitamine",
      c("souvent","pas souvent","non"),
fill=c("steelblue4","steelblue1","lightsteelblue1"), horiz=TRUE, cex=0.8)

```

*##Variable quantitative / quantitative*

*## Calcul de matrice de corrélation*

```

png("cor.png")
par(mfrow=c(1,3))
library(corrplot)
TPretinol_cut<- TPretinol[, c(1,4,9,10,12,14)]
head(TPretinol_cut)
corrplot(cor(TPretinol_cut,use="complete.obs"),method = "circle")
##OU
corrplot(cor(TPretinol_cut,use="complete.obs"),method = "number")
dev.off()
##OU
corrplot(cor(TPretinol_cut,use="complete.obs"),method = "color")

```

*### Autre manière*

```

library(Hmisc)
rcorr(as.matrix(TPretinol_cut))
names(TPretinol)
Retcor <- cor(TPretinol_cut)
Retcor
library(corrplot)
corrplot(Retcor, type="full", order="original", tl.col="blue", tl.srt=45)

```

*#### Var\_QT: Age*

*##Age/BMI*

```

cor.test(TPretinol$age,TPretinol$bmi)## non significative
plot(TPretinol$age,TPretinol$bmi,ylab='BMI',xlab="Age (en années)",
main="BMI en fonction de l'age", col="blue")

```

*##Age/Alcool*

```

cor.test(TPretinol$age,TPretinol$alcool)## non significative

```

```
plot(TPretinol$age,TPretinol$alcohol,ylab='Cons.Alcool par semaine',xlab='Age (en années)', main="Consommation d'alcool par semaine en fonction de l'age", col="blue")
```

#### *##Age/Cholesterol*

```
cor.test(TPretinol$age,TPretinol$cholesterol)## significative  
plot(TPretinol$age,TPretinol$cholesterol,ylab='Cholesterol (mg par jour)',xlab="Age (en années)", main="Cholesterol en fonction de l'age", col="blue")
```

#### *##Age/Retdiet*

```
cor.test(TPretinol$age,TPretinol$retdiet)## non significative  
plot(TPretinol$age,TPretinol$retdiet,ylab='retinol consommé (mg par jour)',xlab="Age (en années)", main="Retinol consommé en fonction de l'age",col="blue")
```

#### *##Age/Retplasma*

```
cor.test(TPretinol$age,TPretinol$retplasma)## significative  
plot(TPretinol$age,TPretinol$retplasma,ylab='Retinol plasmatique (ng/ml)',xlab="Age (en années)", main="Retinol plasmatique en fonction de l'age", col="blue")
```

#### *#### Var\_QT: BMI*

##### *##BMI/Alcool*

```
cor.test(TPretinol$bmi,TPretinol$alcohol)## non significative  
plot(TPretinol$bmi,TPretinol$alcohol,ylab='Cons.Alcool par semaine',xlab="BMI", main="Consommation d'alcool par semaine en fonction du BMI", col="green")
```

##### *##BMI/Cholesterol*

```
cor.test(TPretinol$bmi,TPretinol$cholesterol)## non significative  
plot(TPretinol$bmi,TPretinol$cholesterol,ylab='Cholesterol (mg par jour)',xlab="BMI", main="Cholesterol en fonction du BMI", col="green")
```

##### *##BMI/Retdiet*

```
cor.test(TPretinol$bmi,TPretinol$retdiet)## non significative  
plot(TPretinol$bmi,TPretinol$retdiet,ylab='retinol consommé (mg par jour)',xlab="BMI", main="Retinol consommé en fonction du BMI",col="green")
```

### *##BMI/Retplasma*

```
cor.test(TPretinol$bmi,TPretinol$retplasma)## non significative  
plot(TPretinol$bmi,TPretinol$retplasma,ylab='Retinol plasmatique  
(ng/ml)',xlab="BMI", main="Retinol plasmatique en fonction du BMI",  
col="green")
```

### *#### Var\_QT: Alcool*

#### *##Alcool/Cholesterol*

```
cor.test(TPretinol$alcool,TPretinol$cholesterol)## significative  
plot(TPretinol$alcool,TPretinol$cholesterol,ylab='Cholesterol (mg par  
jour)',xlab="Alcool", main="Cholesterol en fonction de Alcool", col="red")
```

### *##Alcool/Retdiet*

```
cor.test(TPretinol$alcool,TPretinol$retdiet)## non significative  
plot(TPretinol$alcool,TPretinol$retdiet,ylab='retinol consommé (mg par  
jour)',xlab="Alcool", main="Retinol consommé en fonction de  
Alcool",col="red")
```

### *##Alcool/Retplasma*

```
cor.test(TPretinol$alcool,TPretinol$retplasma)## non significative  
plot(TPretinol$alcool,TPretinol$retplasma,ylab='Retinol plasmatique  
(ng/ml)',xlab="Alcool", main="Retinol plasmatique en fonction de Alcool",  
col="red")
```

### *#### Var\_QT: Cholesterol*

#### *##Cholesterol/Retdiet*

```
cor.test(TPretinol$cholesterol,TPretinol$retdiet)## significative  
plot(TPretinol$cholesterol,TPretinol$retdiet,ylab='retinol consommé (mg par  
jour)',xlab="Cholesterol", main="Retinol consommé en fonction du  
Cholesterol",col="deeppink4")
```

### *##Cholesterol/Retplasma*

```
cor.test(TPretinol$cholesterol,TPretinol$retplasma)## non significative  
plot(TPretinol$cholesterol,TPretinol$retplasma,ylab='Retinol plasmatique  
(ng/ml)',xlab="Cholesterol", main="Retinol plasmatique en fonction du  
Cholesterol", col="deeppink4")
```

```
### Var_QT: Retdiet
```

```
##Retdiet/Retplasma
```

```
cor.test(TPretinol$retdiet,TPretinol$retplasma)## non significative  
plot(TPretinol$retdiet,TPretinol$retplasma,ylab='Retinol plasmatique  
(ng/ml)',xlab="Retinol consommé", main="Retinol plasmatique en fonction du  
retinol consommé", col="gray20")
```

```
### 4- Régression linéaire multiple
```

```
#### Régression linéaire multiple sans interaction
```

```
str(TPretinol)  
T12 <-  
lm(retplasma~age+sexe+tabac+bmi+vitamine+alcool+cholesterol+retdiet,data=  
TPretinol)  
summary(T12)  
drop1(T12,~,test="F")  
### vérification de la normalité de bruit  
hist(resid(T12), main="normalité du bruit et du terme résiduel", col=  
"lightblue1")
```

```
#### Régression linéaire multiple avec interaction
```

```
## Au lieu de calculer les 28 interactions possibles entre les paires du jeu de  
données en adoptant la forme suivante:
```

```
T13<-lm(retplasma~age*sexe+bmi+alcool+cholesterol+retdiet+tabac+vitamine,  
data=TPretinol)  
summary(T13)  
drop1(T13,~,test="F")
```

```
## on pourra le remplacer par une formule plus courte et plus efficace à l'aide  
de la fonction "add"
```

```
T12_all <-  
lm(retplasma~age+sexe+tabac+bmi+vitamine+alcool+cholesterol+retdiet,data=  
TPretinol)  
add1(T12_all,~.^2,test="F")
```

```
### test de la normalité du bruit et du terme résiduel.
```

```
png("res.png")  
par(mfrow=c(1,2))  
hist(resid(T12), col = "grey", main="normalité des résidus")  
qqnorm(resid(T12))  
qqline(resid(T12))  
dev.off()
```

### ### 5- Régression logistique

#### ### recoder la variable

```
median(TPretinol$retplasma) ## 566
TPretinol$retplasma_Recode <- ifelse(TPretinol$retplasma < 566,0,1)
class(TPretinol$retplasma_Recode)
TPretinol$retplasma_Recode<-as.factor(TPretinol$retplasma_Recode)
str(TPretinol$retplasma_Recode)
table(TPretinol$retplasma_Recode)
```

#### #### Régression logistique sans interaction

```
T14<-
glm(retplasma_Recode~age+sexe+tabac+vitamine+bmi+alcool+cholesterol+ret
diet, data=TPretinol, family="binomial")
summary(T14)
drop1(T14,~,test="Chisq")
exp(coefficients(T14))
```

#### #### Régression logistique avec interaction

*## Au lieu de calculer les 28 interactions possibles entre les paires du jeu de données en adoptant la forme suivante:*

```
T15<-
glm(retplasma2~age*vitamine+sexe+tabac+bmi+alcool+cholesterol+ret
diet, data=TPretinol, family="binomial")
summary(TP15)
drop1(TP15,~,test="Chisq")
```

*## on pourra le remplacer par une formule plus coute et plus efficace à l'aide de la fonction "add"*

```
T14_all<-
glm(retplasma_Recode~age+sexe+tabac+vitamine+bmi+alcool+cholesterol+ret
diet, data=TPretinol, family="binomial")
summary(T14_all)
add1(T14_all,~.^2,test="C")
```