

Rapport du Projet SAS

par

Tehe Bommanin Yannick
yankouek@yahoo.fr

Table des matières

	I	Partie 1	3
1	Exercice 1		5
2	Exercice 2		8
3	Exercice 3		10
4	Exercice 4		11
5	Exercice 5		13
6	Mini-Conclusion et observations		16
	II	Partie2	17
7	Préparation de base de donnée		19
8	Modélisation		23
9	prediction		24
10	Mini-Conclusion et observations		26

Première partie

Partie 1

Chapitre 1

Exercice 1

Question 1)

Nous avons utilisé ici des instructions data et import (pour la dernière table) pour sélectionner les tables dans la bibliothèque partie_1 et les renommer :voir le code SAS et le fichier html de sortie nous affichons ici une des tables comme exemple d’affichage de vérification

Obs	label	start	end	fmtname	type
1	Agriculteurs sur petite exploitation	11	11	\$fcs	C
2	Agriculteurs sur moyenne exploitation	12	12	\$fcs	C
3	Agriculteurs sur grande exploitation	13	13	\$fcs	C
4	Artisans	21	21	\$fcs	C
5	Commerçants et assimilés	22	22	\$fcs	C
6	Chefs d'entreprise de 10 salariés ou plus	23	23	\$fcs	C

FIGURE 1.1: tabfcs

Question 2)

Nous avons ici utilisé une Proc contents sur toutes les tables que nous avons crée dans la bibliothèque work auparavant .

Voici une exemple de sortie des Proc Contents :

Data Set Name	WORK.LIBDEP	Observations	96
Member Type	DATA	Variables	2
Engine	V9	Indexes	0
Created	04/08/2017 18:46:19	Observation Length	25
Last Modified	04/08/2017 18:46:19	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO

FIGURE 1.2: sortie content pour LIBDEP

Question 3)

Répondons à ces trois questions posées tableau par tableau :

TABLE 1 : libdep

Variables numériques : NULL (il n'y en a pas)
Variables caractères : CODGEO et LIBGEO
Individus statistiques : les départements

TABLE 2 : liste_ze

Variables numériques : NULL (il n'y en a pas)
Variables caractères : CODGEO
Individus statistiques : les zones d'emploi

TABLE 3 : pauv12dep

Variables numériques : D112 , D912, MED12, NBMENFISC12, NBPERSOMENFISC12,
nbmenimp12, nbperspv12
variables caractères : CODGEO
Individus statistiques : les départements

TABLE 4 :pauv12zetypmen1

Variables numériques : med12
variables caractères : codgeo
Individus statistiques : les couples sans enfants dans les zones indiquées par codgeo

TABLE 5 :pauv12zetypmen2

Variables numériques : med12
Variables caractères : codgeo individus statistiques : couple avec enfant(s) dans les zones
indiquées par codgeo

TABLE 6 :pauv12zetypmen3

Variables numériques : med12
Variables caractères : codgeo
Individus statistiques : famille monoparentales dans les zones indiquées par codegeo

TABLE 8 : pauv12zetypmen4

Variables numériques : med12

variables caractères : codgeo

Individus statistiques :femmes seules dnas es zones indiqué par codegeo

TABLE 9 : pauv12zetypmen5

Variables numériques : MED12

Variables caractères : CODGEO

Individus statistiques : hommes seules dnas es zones indiqué par codegeo

TABLE 10 : pauv12zetypmen6

Variables numériques : med12

Variables caractères : codgeo

Individus statistiques :les autres types de ménages dans les zones induées par codegéo

TABLE 11 : rp12_d35

Variables numériques : IPONDI et AGE

Variables caractères : AEMM, CS1, CS3, NUMMR, SEXE, TP, TRANS, TYPFC, TYPL

individus statistiques :les individus (des personnes)

TABLE 12 : tabfcs

Variables numériques :NULL

Variables caractères :label , start, end , fmtname, type

Individus statistiques :catégries socio-professionnelles (categories sociales)

Chapitre 2

Exercice 2

Question 1)

Pour la première partie de cette question nous avons vérifié en considérant la variable `codgeo` comme une clé qu'il n'y avait pas de clés doubles.
(voir le code SAS) Et après le décompte de département faisait bien 96 d'où chaque département de France métropolitaine est présent une seule fois.

pour la deuxième partie une Proc SQL nous a permis de voir que le département 93 (la seine saint-Denis) est le département avec la médiane de revenu la moins élevée et que le département 75 (Paris) est celui avec la médiane de revenu la plus élevée .

voici les deux bouts de la table obtenue :

CODGEO	MED12
93	16609.44
62	17155.33
11	17230.67
23	17301.3
66	17522
2B	17613.33
31	21474.67
68	21482.67
77	21707.22
91	22615.2
74	23658
78	25143.91
92	25521.95
75	25711

FIGURE 2.1: table rangée du revenu médian par département France métropolitaine (début et fin)

Question 2)

la question précédente y avait déjà logiquement répondu mais en refaisant une proc sql se restreignant au départements d'île de France nous obtenons sans surprise que le département 93(la seine saint-Denis) est le département d'île de France avec la médiane de revenu la moins élevé et que le département 75(Paris) est celui d'île de France avec la médiane de revenu la plus élevé .

Voici la table obtenue :

CODGEO	MED12
75	25711
92	25521.95
78	25143.91
91	22615.2
77	21707.22
94	21359.13
95	20547.6
93	16609.44

FIGURE 2.2: table rangée du revenu médian en Ile de France

Question 3)

Avec un format nous avons affecté 0 aux départements en dessous du revenu médian de métropole et 1 à ceux au dessus . nous donnons ici une partie de la sortie :

Obs	CODGEO	NBMENFISC12	NBPERSMENFISC12	MED12	D112	D912	nbperspv12	nbmenimp12
1	01	243585	598328.0	1	11763.89	40783.33	61751	159819
2	02	222044	532092.0	0	9690.00	31280.00	99230	127592
3	03	157846	334816.5	0	10473.00	32199.05	49571	91302
4	04	71051	155262.5	0	10029.00	33378.33	25500	42090
5	05	62107	136164.0	0	10670.00	32884.67	19224	38293
6	06	508766	1100253.5	1	10133.00	39093.33	167336	334343
7	07	137631	314462.0	0	10435.00	32089.33	46478	81045

FIGURE 2.3: sortie apres l'ajout du format

Puis le décompte avec une proc sort en affichant les effectifs nous donne qu'il y a 68 départements en dessous de la médiane en métropole et 28 départements en dessous de celle ci.

Question4)

en utilisant la variable nbperspv12 : On la somme pour tous les départements d'île de France on obtient qu'il y a 1.765.613 personnes sous le seuil de pauvreté en ile de France

Chapitre 3

Exercice 3

Question 1) voir code SAS

Question 2)

une proc Freq nous permet d'obtenir sur la table RP12_D35 ou le format de la question 1 a été appliqué :

SEXE				
SEXE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Homme	492272.5	48.84	492272.5	48.84
Femme	515707.1	51.16	1007980	100.00

FIGURE 3.1: repartition Homme femme Ile et Vilaine

On a donc 48,84% d'homme et 51,16% de femmes en ile et vilaine .

Question 3)

en distinguant temps partiel et temps plein on obtient la table suivante :

Table of SEXE by TP					
Frequency Percent Row Pct Col Pct	SEXE(SEXE)	TP(TP)			
		1	2	Z	Total
	Homme	208719 20.71 42.40 58.97	17433.3 1.73 3.54 21.40	266120 26.40 54.06 46.48	492273 48.84
	Femme	145221 14.41 28.16 41.03	64045.7 6.35 12.42 78.60	306441 30.40 59.42 53.52	515707 51.16
	Total	353940 35.11	81479 8.08	572561 56.80	1007980 100.00

FIGURE 3.2: répartition homme femme ile et vilaine en font fonction du temps de travail(partiel, plein ou autre)

on conclut donc que : 58.97% des personnes à temps complet sont hommes contre 41,03% pour les femmes. Alors que 21,4% des personnes à temps partiel sont des hommes contre 78,6% pour les femmes.

Chapitre 4

Exercice 4

Question 1)

la proc univariate sur la variable age dans un premier temps puis en distinguant homme et femme donne entre autres les tables suivants nous ne gardons que celles qui nous intéresse :

Weighted Basic Statistical Measures			
Location		Variability	
Mean	39.46811	Std Deviation	44.03839
Median	38.00000	Variance	1939
Mode	21.00000	Range	114.00000
		Interquartile Range	37.00000

FIGURE 4.1

The UNIVARIATE Procedure Variable: age Weight: IPONDI (IPONDI) SEX=Homme			
Weighted Basic Statistical Measures			
Location		Variability	
Mean	38.08480	Std Deviation	42.66839
Median	37.00000	Variance	1821
Mode	23.00000	Range	114.00000
		Interquartile Range	36.00000

FIGURE 4.2: age médian homme ile et vilaine

The UNIVARIATE Procedure Variable: age Weight: IPONDI (IPONDI) SEX=Femme			
Weighted Basic Statistical Measures			
Location		Variability	
Mean	40.78857	Std Deviation	45.16063
Median	40.00000	Variance	2039
Mode	20.00000	Range	107.00000
		Interquartile Range	39.00000

FIGURE 4.3: age médian femme Ile et vilaine

On peut donc dire que l'âge médian pondéré est de 38 ans en ile et vilaine . L'âge médian pondéré des hommes est de 37 ans et celui des femmes est de 40 ans

Question 2)

Avec une proc univariate effectuée selon le sexe on obtient la table de décile suivante :

Obs	SEXE	quantile_10	quantile_20	quantile_30	quantile_40	quantile_50	quantile_60	quantile_70	quantile_80	quantile_90	quantile_100
1	Homme	8	15	22	30	37	44	51	60	70	114
2	Femme	9	17	24	32	40	47	55	64	76	107

FIGURE 4.4: decile de l'âge des hommes et des femmes en ile et vilaine

Question 3)

la proc univariate nous fournit les indicateurs demandés dans kes autres questions nous n'avions recupéré que ce qui nous intéressait il suffit donc de stocker tous kes résultats dans une table SAS

Chapitre 5

Exercice 5

Question 1)

pour cette question on crée une table `emploi_tze`(emploi toute zone confondue) en superposant les 6 tables auxquelles on a rajouté une nouvelle variable précisant le type de ménage.

Puis on applique une proc `sql` pour sélectionner les trois variables qui nous intéressent .

et enfin on applique une proc `sort` pour ranger la grande table obtenue en fonction de `codegeo`(des zones d'emploi) pour arriver au résultat escompté moyennant le fait de renommer la colonne `codegeo` pour être plus proche du résultat demandé. Voici le début de la table obtenue finalement :

Obs	zone_demploi	type_menage	Mediane
1	0050	1	21421.33
2	0050	2	19625.00
3	0050	3	14936.00
4	0050	4	16374.50
5	0050	5	17061.00
6	0050	6	19720.95

FIGURE 5.1: revenu médian en fonction de la zone d'emploi et du type de ménage

Question 2)

On voit bien si on veut rapprocher ce qu'on a à faire d'une opération matricielle qu'il s'agit d'une transposition . On utilise donc la proc `transpose` de SAS(voir le code SAS pour plus de détails) : voici le début de la table obtenue finalement :

zone_emploi	mediane_pour_letype_de_menage_1	mediane_pour_letype_de_menage_2	mediane_pour_letype_de_menage_3	mediane_poi
0050	21421.33	19625.00	14936.00	16374.5
0051	21387.00	18660.83	13712.61	16702.0
0052	21525.67	18695.71	13952.08	16807.0
0053	23032.00	19882.86	15457.78	17492.0
0054	21678.33	18798.00	14201.54	17059.0

FIGURE 5.2: revenu médian en fonction de la zone d'emploi et du type de ménage autre vue

Question 3)

En observant le tableau obtenu à la question précédente il semble que ce soit le type de ménage 3 (les familles mono-parentales) qui ait les revenus les plus faibles vérifions le ou démentons le .

On procède comme suit :

on crée une nouvelle variable avec une instruction data qui vaut 1 si la médiane de revenu du type de ménage 3 est le plus petit sur chaque ligne (du tableau obtenu à la question précédente) et 0 sinon .

il suffit enfin de comparer le nombre de 1 dans cette variable au nombre de champ de la table obtenue .

on conclut que le type de ménage 3 n'est pas celui avec le revenu médian le plus faible dans toutes les zones d'emploi .

Pour les autres types de ménage on voit en regardant ne serait-ce que la première ligne du tableau qu'il ne sont pas ceux avec le revenu médian le plus faible dans toutes les zones d'emploi .

On répond donc à la question qu'aucun des types de ménage n'est celui le revenu médian le plus faible dans toutes les zones d'emploi.

(la vraie réponse à la question : le mode opératoire est le chemin utilisé pour arriver à cette conclusion) voir le code SAS et les sorties pour un peu plus de détail .

Question 4)

Ici la solution que je propose est une jointure interne entre les tables liste_ze et celle obtenue à la question 1 (donc déjà triée de manière appropriée pour une telle jointure) avec une proc sql. la vérification a bien sûr été de comparer le nombre de champs de cette table à ce qui devait être c'est à dire ($\text{effectif}(\text{liste_ze}) \times 6$) et le compte est effectivement bon $65 \times 6 = 390$

Voici le début de la table ainsi obtenue :

Obs	zone_demploi	type_menage	Mediane
1	0061	1	24734.00
2	0061	2	22399.23
3	0061	3	15736.20
4	0061	4	17533.00
5	0061	5	18622.50
6	0061	6	20311.14
7	0061	7	21000.00

FIGURE 5.3: revenu médian par type de ménage restreint à liste_ze

Question5)

Ici une nouvelle proc sql nous permet avec un except , d'exclure la table obtenue à la question 3 de celle de la quelle elle a été extraite par jointure interne .

Voici e début de la table ainsi obtenue :

Obs	zone_demploi	type_menage	Mediane
1	0050	1	21421.33
2	0050	2	19625.00
3	0050	3	14936.00
4	0050	4	16374.50
5	0050	5	17061.00
6	0050	6	19720.95
7	0050	7	21000.00

FIGURE 5.4: revenu médian par type de ménage excluant liste_ze

Voir le Code SAS pour plus de détail.

Chapitre 6

Mini-Conclusion et observations

De cette longue étude qui nous a appris beaucoup de chose nous pouvons en tirer des information diverses pour notre propre culture et la complexité de l'évaluation de statistiques sur le revenu médian. Nous pouvons énumérer les grandes informations suivantes :

-Le département le plus pauvre de France est la seine saint denis(le 93) et celui le plus riche est Paris l'île de France qui contient ces deux départements présente donc un contraste entre pauvreté et richesse qui en fait presque un échantillon miniature de la réalité de France métropolitaine .

-L'analyse du département d'île de France nous montre qu'il y a plus de précarité de l'emploi et de précarité tout court chez les femmes que chez les hommes.

-Et enfin on ne peut pas vraiment faire de généralité quant au type de foyer qui a le plus faible revenu médian en fonction de la zone d'emploi néanmoins les familles monoparentales sont celles qui ont dans la plupart des zones d'emploi le revenu médian le plus bas .

Deuxième partie

Partie2

Chapitre 7

Préparation de base de donnée

Voir le Code SAS pour la création de la bibliothèque et des formats. Décrivons tout de même les formats que nous avons adopté :

Sur la variable SEXE qui était à l'origine encodée comme suit ,0 pour homme et 1 pour femme nous adoptons le format qui fait passer 0 à femme et 1 à homme (sans innovation)

Sur la variable TABAC qui était encodé comme suit 0 pour non-fumeur et 1 pour fumeur nous adoptons le format qui affiche fumeur pour 1 et non-fumeur pour 0.

Enfin sur la variable apnee qui était fixée à 0 quand l'individu n'avait pas une apnée du sommeil et 1 lorsqu'il en avait une nous adoptons le format qui affiche non pour 0 et oui pour 1.

Analyse univariée des variables de la table apnee Ici nous présenterons différentes tables obtenus à partir de la proc univariate que nous commenterons.

variable AGE

Basic Statistical Measures			
Location		Variability	
Mean	52.27000	Std Deviation	11.39720
Median	52.00000	Variance	129.89606
Mode	46.00000	Range	51.00000
		Interquartile Range	19.50000

FIGURE 7.1: Moyenne et médiane variable age

Quantiles (Definition 5)	
Level	Quantile
100% Max	74.0
99%	73.0
95%	69.0
90%	67.5
75% Q3	62.5
50% Median	52.0

25% Q1	43.0
10%	38.0
5%	34.5
1%	27.5
0% Min	23.0

FIGURE 7.2

On se rend bien compte ici d'abord qu'on a affaire à une population d'adulte l'âge le plus bas est de 23 ans et une population plutôt agée avec une moyenne et une médiane proche de 52 ans la cible de l'étude doit donc être là ce ne serait donc pas étonnant de voir l'âge est un facteur explicatif en tout cas en une certaine part de l'apnée du sommeil

variable POIDS

Basic Statistical Measures			
Location		Variability	
Mean	90.4100	Std Deviation	18.73585
Median	95.0000	Variance	351.03222
Mode	108.0000	Range	78.00000
		Interquartile Range	30.00000

FIGURE 7.3: Moyenne et médiane variable poids

Quantiles (Definition 5)	
Level	Quantile
100% Max	120.0
99%	119.5
95%	116.5
90%	112.0
75% Q3	107.0
50% Median	95.0

25% Q1	77.0
10%	61.5
5%	57.0
1%	45.5
0% Min	42.0

FIGURE 7.4: quantiles de la variable poids

la population ciblée par cette étude a plutôt un poids élevé globalement avec même des personnes en sur-poids une moyenne à 90 kilogrammes et une médiane à 95 nous indiquant même qu'il y a plus de personnes de poids élevée dans la population étudiée que de personne de faible poids .

variable TAILLE :

Basic Statistical Measures			
Location		Variability	
Mean	181.1000	Std Deviation	13.36625
Median	186.0000	Variance	178.65657
Mode	194.0000	Range	50.00000
		Interquartile Range	28.00000

FIGURE 7.5: Moyenne et médiane variable taille

Quantiles (Definition 5)	
Level	Quantile
100% Max	208.0
99%	203.0
95%	197.0
90%	195.5
75% Q3	194.0
50% Median	186.0
25% Q1	166.0
10%	163.5
5%	160.0
1%	158.0
0% Min	158.0

FIGURE 7.6: quantiles de la variable taille

Au niveau des tailles là encore on peut dire qu'on a affaire à des personnes plutôt grandes globalement avec une extrême à 2 mètres 08.

SEXE :

Moments			
N	100	Sum Weights	100
Mean	0.25	Sum Observations	25
Std Deviation	0.43519414	Variance	0.18939394
Skewness	1.17235972	Kurtosis	-0.6387545
Uncorrected SS	25	Corrected SS	18.75
Coeff Variation	174.077656	Std Error Mean	0.04351941

FIGURE 7.7: indicateurs variable SEXE

ici on peut dire qu'on a 25% de femmes dans la population d'individu étudiée

variable ALCOOL :

Moments			
N	100	Sum Weights	100
Mean	2.95	Sum Observations	295
Std Deviation	3.36462584	Variance	11.3207071
Skewness	1.12058999	Kurtosis	0.86020903
Uncorrected SS	1991	Corrected SS	1120.75
Coeff Variation	114.055113	Std Error Mean	0.33646258

FIGURE 7.8: indicateurs variable Alcool

Quantiles (Definition 5)	
Level	Quantile
100% Max	15.0
99%	13.5
95%	10.0
90%	8.0
75% Q3	4.5
50% Median	2.0
25% Q1	0.0

FIGURE 7.9: indicateurs variable Alcool

Au niveau du nombre de verre bus par jour il est assez hétérogène tout type de buveur (et de non-buveur) est représenté dans cet échantillon. Avec une moyenne autour de 3 verres et une médiane à 2 verres : comme dans la vie de tous les jours dans cette population 50% des personnes boivent peu de verres par jour.

on peut dire que vu que l'écart est quand même élevé entre les personnes qui boivent le plus et celles qui consomment peu si l'alcool est un facteur on aura peu de mal à ne pas le voir dans nos modèles .

variable TABAC :

Moments			
N	100	Sum Weights	100
Mean	0.64	Sum Observations	64
Std Deviation	0.48241815	Variance	0.23272727
Skewness	-0.5922544	Kurtosis	-1.6833119
Uncorrected SS	64	Corrected SS	23.04
Coeff Variation	75.3778361	Std Error Mean	0.04824182

FIGURE 7.10: indicateurs variable Tabac

on a 64% de fumeur dans cette étude.

VARIABLE APNEE :

Moments			
N	100	Sum Weights	100
Mean	0.35	Sum Observations	35
Std Deviation	0.47937249	Variance	0.22979798
Skewness	0.63858994	Kurtosis	-1.6251159
Uncorrected SS	35	Corrected SS	22.75
Coeff Variation	136.963567	Std Error Mean	0.04793725

FIGURE 7.11: indicateurs variable Apnee

Enfin la variable à prédire on peut 35% des personne de la population étudiée souffrent d'apnée du sommeil dans cet échantillon .

On a donc qu'environ 4 personne sur 10 souffrent d'apnée du sommeil.

Chapitre 8

Modélisation

Nous préconisons ici l'utilisation d'une régression logistique pour prédire la variable apnée en fonction de toutes les autres :

nous utilisons un proc logistic(voir CODE SAS) avec un choix pour les variables explicatives et nous remarquons que certaines ne sont pas très significatives dans le modèle avec toutes variables :

Nous effectuons alors une sélection automatisée(voir code SAS et sorties SAS) sur laquelle nous pouvons faire les commentaires suivants :

la sélection de variables stepwise que nous avons choisi fait rentrer pas à pas les variables dans le modèle elles entrent si la valeur de la probabilité dans le test de suppression (la valeur de f-statistique) est en dessus 0.3 et reste si cette valeur est en dessous a 0.3 , on a pas été très sévère mais en regardant le résultat uniquement 3 variables sont sélectionnées. une sélection backward et forward donne le même triplet de variable sélectionné .(Voir CODE SAS) :

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
AGE	1.067	1.020	1.117
ALCOOL	1.300	1.110	1.522
TABAC	0.348	0.123	0.985

FIGURE 8.1: ods ratios du modèle retenu

En analysant les indicateurs disponibles pour ce modèle on peut dire que l'alcool reste la variable(la plus significative) qui explique le mieux la présence d'apnée du sommeil.

Chapitre 9

prediction

Nous allons utiliser les tableaux " model fit statistics" et "testing global hypotesis null" pour comparer les différents modèles.

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
AGE	0.937	0.895	0.980
ALCOOL	0.769	0.657	0.901
TABAC	2.873	1.015	8.131

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	18.8297	3	0.0003
Score	16.8749	3	0.0007
Wald	14.2367	3	0.0026

FIGURE 9.1: statistiques pour le modèle retenu

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	131.489	123.418
SC	134.094	141.654
-2 Log L	129.489	109.418

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	20.0716	6	0.0027
Score	17.7452	6	0.0069
Wald	14.8584	6	0.0214

FIGURE 9.2: statistiques pour le modèle complet(avec toutes le variables explicatives

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	131.489	132.401
SC	134.094	137.611
-2 Log L	129.489	128.401

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1.0884	1	0.2968
Score	1.0989	1	0.2945
Wald	1.0924	1	0.2959

FIGURE 9.3: statistiques pour le modèle avec la variable Tabac

Au niveau du test de nullité les différentes probabilités des différents test sont plus élevés pour le modèle complet il comporte donc sans surprise plus de variable moins significatives . Mais au niveau des indicateurs en rapport avec la vraisemblance du modèle complet a logiquement un AIC plus élevé(l'AIC est un critère de selection de variable c'est logique) que le modèle retenu.

Mais les deux modèles ont une vraisemblance proche il n'en demeure pas moins que globalement et à beaucoup d'égards le modèle sélectionné c'est à dire avec comme variables explicatives le tabac , l'alcool et l'age reste le meilleur.

le modèle avec que le tabac comme variable est particulièrement mauvais par rapport au modèle retenu :

D'une le test de nullité des coefficients renvoie des probabilités 10 fois supérieurs pour lui par rapport à celui du modèle retenu.

De deux par rapport aux critères en rapport avec la vraisemblance c'est à dire l'AIC et $-2 \cdot \log$ -vraisemblance ils sont beaucoup plus hauts que ceux du modèle retenu .

Chapitre 10

Mini-Conclusion et observations

Cette étude sur l'apnée du sommeil que nous n'avons pu approfondir comme nous le voulions faute de temps nous a mené à des observations importantes :

Au vu des modèles que nous avons sélectionné nous pouvons dire que les facteurs qui explique le mieux cette pathologie sont dans l'ordre d'importance de la plupart des modèle le Tabac l'alcool et l'age.

cette étude aurait grand intérêt de santé publique car elle nous montre bien sur quoi doit porter la prévention pour prévenir cette pathologie qui a un stade avancée peut annoncer de problème de santé plus grave .

Et d'autre part à l'intérieur même de ces probables cause l'alcool reste l'une des causes la plus corrélée à l'apnée du sommeil nous pouvons faire l'hypothèse qu'elle joue un grand rôle dans son apparition et sensibiliser encore une fois.

Merci d'avoir lu ce rapport et bonne continuation dans vos activités.
Tehe Bommanin Yannick joignable aux adresses électroniques suivantes :

yankouek@yahoo.fr
bommanin.tehe@etu.upmc.fr