

Econométrie

Examen final Session 1

Note : Aucun document n'est autorisé. Tout type de calculatrice est autorisé.

Pour tous les exercices, vous effectuerez les tests à un seuil de risque de 5%.

Exercice 1 : Complémentaire santé et état de santé

Partie 1 : Montant de la cotisation annuelle pour son contrat de complémentaire santé

On souhaite analyser les déterminants des cotisations annuelles pour une complémentaire santé. En particulier, on cherche à savoir si les compagnies d'assurance ont tendance à réduire les cotisations des individus qui adoptent certains gestes pour préserver leur santé (ex : ne pas fumer, faire du sport, manger des fruits et légumes). Pour ce faire, une régression linéaire multiple, à partir d'un échantillon de 10 146 assurés, a été estimée par moindres carrés ordinaires à partir du logiciel Stata. Les résultats de cette régression sont présentés ci-dessous (les ??? correspondent à des résultats qui ont été volontairement enlevés) :

Source	SS	df	MS	Number of obs	=	10,146
Model	1.3248e+09	8	165599545	F(8, 10137)	=	??????
Residual	7.1787e+09	???????	708172.877	Prob > F	=	??????
Total	8.5035e+09	10,145	838200.574	R-squared	=	0.1551
				Adj R-squared	=	0.1551
				Root MSE	=	841.53

montant_an	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
age	12.40844	.6869441	18.06	0.000	11.06189 13.75499
sexe	153.6995	47.85919	3.21	0.001	59.88604 247.513
sexe * age	-3.246032	.8945087	-3.63	0.000	-4.999446 -1.492618
fume1	-106.3984	22.17235	-4.80	0.000	????????? ??????????
fruit	-11.03216	7.798847	-1.41	0.157	-26.31944 4.255129
sport	1.44986	4.80597	0.30	0.763	-7.970793 10.87051
entrep1	-500.7784	19.84566	-25.23	0.000	-539.6798 -461.8769
cadre	46.40204	24.04576	1.93	0.054	-.7324209 93.5365
_cons	749.1625	45.06765	16.62	0.000	660.821 837.5041

avec

- *montant_an* : le montant de la cotisation annuelle pour une complémentaire santé (en euros)
- *age* : l'âge de l'assuré
- *sexe* : une variable dichotomique qui vaut 1 si l'assuré est un homme et 0 sinon.
- *fume1* : une variable dichotomique qui vaut 1 si l'assuré est fumeur et 0 sinon
- *fruit* : le nombre de fruits et légumes mangés en moyenne par jour.
- *sport* : le nombre de jours où l'assuré pratique un sport au cours d'une semaine habituelle
- *entrep1* : une variable dichotomique qui vaut 1 si l'adhésion à la complémentaire santé était obligatoire au sein de l'entreprise et 0 sinon.
- *cadre* : une variable dichotomique qui vaut 1 si l'assuré est cadre et 0 sinon.

1. Calculer le coefficient de détermination du modèle et commenter.
2. Tester si le modèle est globalement significatif.
3. Déterminer l'intervalle de confiance (à 95%) pour le coefficient associé à la variable *fume1* et donner en une interprétation.
4. (a) Quelles sont les variables significatives dans ce modèle ?
 (b) Commenter en détails l'effet de l'âge sur les montants des cotisations et quantifier précisément ses effets.
 (c) Au vu des résultats, peut-on dire que les compagnies d'assurance ont tendance à réduire les cotisations des individus qui adoptent des comportements préventifs ? Justifier votre réponse.
 (d) Commenter littérairement l'ensemble des autres résultats.
5. Effectuer une prévision du montant des cotisations pour une employée de 35 ans, fumeuse, qui mange 3 fruits et légumes par jour, ne fait pas de sport et qui a souscrit une complémentaire santé de sa propre initiative.
6. Pour l'assurée décrite dans la question précédente, calculer l'élasticité des montants des cotisations par rapport à l'âge.
7. Quelles sont les limites de l'estimation effectuée ?

Partie 2 : Analyse des états de santé des assurés

Dans la même base de données, nous avons une variable qui vaut 1 si l'état de santé des assurés est jugé bon ou très bon et 0 sinon. En supposant que les termes d'erreur suivent une loi normale de moyenne nulle et de variance unitaire, les résultats de l'estimation du modèle approprié sont reportés ci-dessous.

		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
etasantel						
montant_an		.0000206	.0000163	1.26	0.207	-.0000114 .0000525
age		-.033978	.0013251	-25.64	0.000	-.0365751 -.0313809
sexe		-.1727614	.0939742	-1.84	0.066	-.3569474 .0114246
sexe * age		.0024515	.0016547	1.48	0.138	-.0007917 .0056947
fumel		-.089903	.037452	-2.40	0.016	-.1633076 -.0164984
fruit		-.0432489	.0132281	-3.27	0.001	-.0691755 -.0173223
sport		.0982467	.008258	11.90	0.000	.0820614 .114432
entrep1		.0926447	.0360227	2.57	0.010	.0220415 .1632478
cadre		.3628647	.0416777	8.71	0.000	.281178 .4445515
_cons		2.178564	.087648	24.86	0.000	2.006777 2.350351

1. Ecrire le modèle qui a été estimé et quelle a été la technique d'estimation utilisée ?
2. Calculer l'effet marginal de l'âge pour le profil d'assurée décrit à la question 5 de la partie 1 et interpréter.
3. Si on demande à Stata de calculer les effets marginaux, on obtient les résultats suivants :

		Delta-method				
		dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]
montant_an		6.90e-06	5.46e-06	1.26	0.207	-.381e-06 .0000176
age		-.0109733	.0003123	-35.14	0.000	-.0115854 -.0103612
sexe		-.0168199	.0098859	-1.70	0.089	-.0361958 .0025561
fumel		-.0301754	.0125621	-2.40	0.016	-.0547968 -.0055541
fruit		-.0145163	.0044382	-3.27	0.001	-.0232149 -.0058176
sport		.032976	.0027672	11.92	0.000	.0275524 .0383996
entrep1		.0310957	.0120911	2.57	0.010	.0073976 .0547937
cadre		.1217935	.0139852	8.71	0.000	.094383 .1492041

Note: dy/dx for factor levels is the discrete change from the base level.

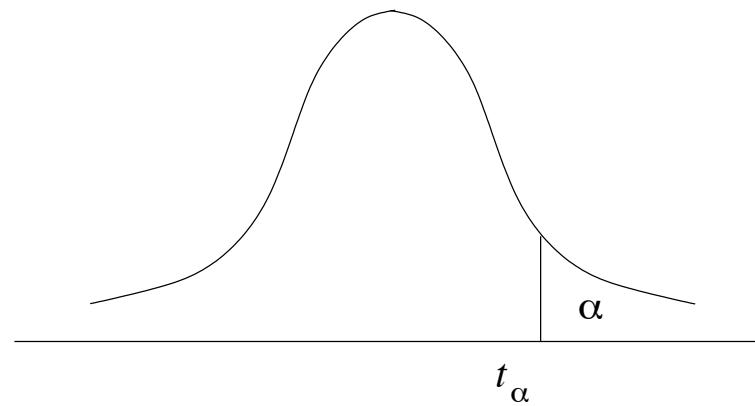
- (a) Commenter l'effet marginal donné par Stata et expliquer l'éventuelle différence obtenue par rapport à votre calcul de la question précédente.

- (b) Donner la formule qui a permis de calculer l'effet marginal de la variable $fume1$ et interpréter littérairement cet effet
- (c) Si le montant de cotisation annuelle augmente de 1%, de combien augmente la probabilité d'être en bonne santé ?
- (d) Est-ce que le fait de ne pas fumer, de faire du sport fréquemment ou de manger beaucoup de fruits et légumes améliorent l'état de santé des assurés ?

4. Au vu du tableau suivant, quelle est la validité du modèle estimé ?

Prédit	Observé		Total
	1	0	
1	6263	1902	8165
0	718	1263	1981
Total	6981	3165	10146

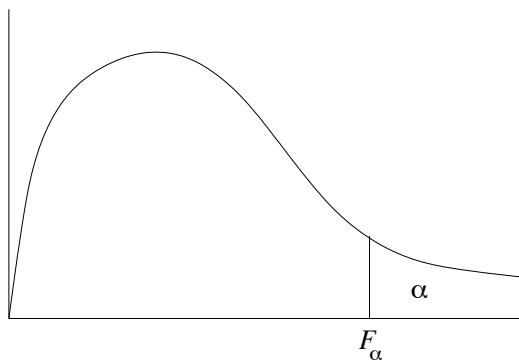
Loi de Student



d.d.l.	α	0.10	0.05	0.025	0.01	0.005
1		3.078	6.314	12.706	31.821	63.657
2		1.886	2.920	4.303	6.965	9.925
3		1.638	2.353	3.182	4.541	5.841
4		1.533	2.132	2.776	3.747	4.604
5		1.476	2.015	2.571	3.365	4.032
6		1.440	1.943	2.447	3.143	3.707
7		1.415	1.895	2.365	2.998	3.499
8		1.397	1.860	2.306	2.896	3.355
9		1.383	1.833	2.262	2.821	3.250
10		1.372	1.812	2.228	2.764	3.169
11		1.363	1.796	2.201	2.718	3.106
12		1.356	1.782	2.179	2.681	3.055
13		1.350	1.771	2.160	2.650	3.012
14		1.345	1.761	2.145	2.624	2.977
15		1.341	1.753	2.131	2.602	2.947
16		1.337	1.746	2.120	2.583	2.921
17		1.333	1.740	2.110	2.567	2.898
18		1.330	1.734	2.101	2.552	2.878
19		1.328	1.729	2.093	2.539	2.861
20		1.325	1.725	2.086	2.528	2.845
21		1.323	1.721	2.080	2.518	2.831
22		1.321	1.717	2.074	2.508	2.819
23		1.319	1.714	2.069	2.500	2.807
24		1.318	1.711	2.064	2.492	2.797
25		1.316	1.708	2.060	2.485	2.787
26		1.315	1.706	2.056	2.479	2.779
27		1.314	1.703	2.052	2.473	2.771
28		1.313	1.701	2.048	2.467	2.763
29		1.311	1.699	2.045	2.462	2.756
∞		1.282	1.645	1.960	2.326	2.576

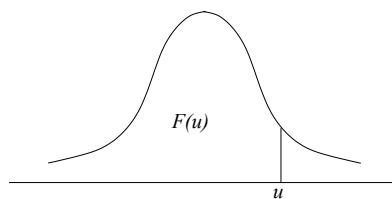
Loi de Fisher : $\alpha = 0.05$

© Théo Jalabert

	Degrés de liberté du numérateur : ν_1								
	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Loi normale



u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Table pour les grandes valeurs de u :

u	3.0	3.1	3.2	3.3	3.4	3.5
$F(u)$	0.998650	0.999032	0.999313	0.999517	0.999663	0.999767

u	3.6	3.7	3.8	3.9	4.0	4.5
$F(u)$	0.999841	0.999892	0.999928	0.999952	0.999968	0.999997

Exercice 1:

$$1) R^2 = \frac{SCE}{SCT} = \frac{1.3248 \times 10^9}{8.5035 \times 10^9} = 0,158$$

2) Test de significativité globale du modèle :

$$H_0: \log a_i = 0$$

$$\text{Stat de test: } F = \frac{R^2}{1-R^2} \cdot \frac{T-k}{k-1} \sim F(k-1, T-k)$$

$$= 233,84 > \underbrace{F_{\alpha=0,05}}_{(8, \infty)}(18, \infty)$$

D'où on rejette H_0 .

Le modèle est globalement significatif.

3) L'intervalle de confiance pour le coeff associé à la variable jume1 est :

$$\hat{\beta}_1 = \hat{\beta}_1 \pm t_{T-k}^{\alpha/2} \times \hat{s}_{\hat{\beta}_1}$$

$$\Leftrightarrow \hat{\beta}_1 = -106.3984 \pm \underbrace{t_{10,37}^{0,025}}_{1,96} \times 22.17235$$

D'où

$$\begin{aligned} IC_{jume1} &= [-106.3984 - 1.96 \times 22.17235; -106.3984 + 1.96 \times 22.17235] \\ &= [-149,85621; -62.94059]. \end{aligned}$$

Interprétation: Il y a 95% de chance que la valeur du coeff associé à la variable jume1 soit compris dans cet intervalle (pour $\alpha=0,05$).

$$4) a) t_{T-k}^{\alpha/2} = 1,96$$

$$\cdot \hat{\beta}_1 \cdot \dots \cdot 1,96$$

On considère qu'une variable est significative si $|t_i| > t_{\alpha/2}$

© Théo Jalabert

où t_i est le coeff associé à cette variable

* age: $\left| \frac{12.40844}{0.6863441} \right| = 18.06 > 1.96$

Donc variable significative

* Sexe: $\left| \frac{153.6985}{57.85319} \right| = 3,21 > 1.96$

Donc variable significative

* sexe*age: $\left| \frac{-3.246032}{0.8945087} \right| = 3,63 > 1.96$

Donc variable significative

* femme1: $4,80 > 1.96$ Donc variable significative

* fruit: $1,41 < 1.96$ Donc variable non significative

* Sport: $0,31 < 1.96$ Donc variable non significative

* entrep1: $25,23 > 1.96$ Donc variable significative

* cadre: $1,93 < 1.96$ Donc variable non significative

* -cons: $16,62 > 1.96$ Donc variable significative

Au lieu d'être tenté regarde les données colonne "t"

Les va significatives sont: age

Sexe

Sexe*age

femme1

entrep

-cons

b) Plus l'individu est âgé plus le montant de la cotisation annuelle pour une complémentaire santé augmente.

Plus précisément, pour une unité d'âge supplémentaire le montant de la cotisation annuelle augmente de $\sim 12,41 \text{ €}$

© Théo Jalabert

Jalabert

Cotisat° annuelle augmente de $\sim 12,41 \text{ €}$

c) $\hat{\beta}_{\text{fem}} = -106,3984 \Rightarrow$ Si l'individu femme alors Montant_an \rightarrow de 106€ chelou?

d) Si l'individu est un homme + 153,7€

Cadre + 46,4€

5) $\text{Montant_an} = 12,40844 \times \underbrace{35}_{\text{age}} + 153,6995 \times \underbrace{0}_{\text{femme}} + (-3,246032) \times 0$

$$+ (-106,3984) \times \underbrace{1}_{\text{fumage}} + (-11,03216) \times \underbrace{3}_{\text{fruits et légumes}} + 1,64986 \times \underbrace{0}_{\text{pas de sport}}$$

$$+ (-500,7784) \times \underbrace{0}_{\text{adhésion initiative}} + 46,4044 \times \underbrace{0}_{\text{non cadre}} + 749,1625$$

$$= 1043,96 \text{ €}$$

6) $e = \frac{\Delta \text{montant}_\text{an}}{\text{Montant}_\text{an}} \times \frac{\text{âge}}{\Delta \text{âge}} = \frac{12,40844}{1043,96} \times \frac{35}{1} = 0,416$

On a calculé l'élasticité pour une variation unitaire de l'âge

Cela implique que pour une variation de l'âge de 1 an

Le montant des cotisations annuelles varie de 0,416%

Partie 2:

© Théo Jalabert



1) Le modèle pour Sécurité $Y_i = \begin{cases} 1 & \text{si } Y_i^* > 0 \\ 0 & \text{sinon} \end{cases}$ avec $Y_i^* = X_i\beta + \epsilon_i$ va être
Modèle à choix discuté (modèle de Probit)

2) On a $\frac{\partial P[Y_i]}{\partial X_i} = \beta_i f(X_i\beta)$ (formule) $X_i\beta =$

3)a) On a pris les $\hat{\beta}$ d'où le \neq .

b) $\beta_{jumeau} \times \text{élasticité}_{Q5}$

c) Ça l'augmente de $6,9 \times 10^{-8}$.

d) Oui car la variable 1 sera nulle donc on Sécurité le coeff β_{jumeau} qui est négatif

Puis la finit du sport $\beta_{sport} > 0$

4) Prédic^o: 8165 pas en bonne santé

1381 pas en mauvaise santé

Obsv^e: 6381 "

3165 "

erreur de 17% pour la prédic^o bonne santé

62,6%

Mauvaise santé

Done Ok~

Partie 2 :

© Théo Jalabert

1) Le modèle qui a été estimé pour l'école :

$$Y_i = \begin{cases} 1 & \text{si } Y_i^* > 0 \\ 0 & \text{Sinon} \end{cases} \quad Y_i = \text{état de santé}$$

$$Y_i^* = X_i \beta + \varepsilon_i \quad \text{Va faire les } X_i \text{ sont les Va et } \beta = (\beta_j) \text{ les coeff}$$

On a utilisé un modèle à choix discrét et plus particulièrement le modèle Probit

2) On a $\frac{\partial P(Y_i=1)}{\partial x_j} = \beta_j f(X_i \beta)$

Ici, $\beta_0 = -0.033978$ car c'est une femme dans $\text{Sexe}^* \text{age} = 0$

$$\text{et } X_i \beta = 103,96 \times 0,0000206 - 35 \times 0,033978 - 0,089303 - 3 \times 0,0432489 + 2,178364$$

$\overset{\text{cas de la personne}}{=} 0,79119$

$$\Rightarrow \frac{\partial P(Y_i=1)}{\partial x_0} = -0,033978 \times \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \times 0,79119^2\right)$$

$$= -0,01$$

Donc la proba d'être considéré en bonne santé diminue de 1% avec l'âge.
(i.e cette femme a 1% moins de chance d'être considérée en bonne santé qu'une personne ayant au moins que 35 ans).

3) a) X

b) $\frac{\partial P(Y_i=1)}{\partial X_{june1}} = \beta_{june1} f(X_i \beta)$ avec $X_i \beta = \sum_{k=0}^K X_{ik} \beta_k$

L'effet marginale de la Va June1 est $-0,0301754$

\Rightarrow cela signifie qu'une personne qui gagne ce 3% de chaque mois d'écho considère comme en bonne santé qu'une personne qui ne gagne pas.

c) On suppose maintenant un P de 1%

Donc la proba d'être en bonne santé P de 6.9×10^{-6}

\Rightarrow Donc la proba d'être en bonne santé si le taux de l'obésité amineille augmente de 1% est $6.9 \times 10^{-6} = 0,00069\%$

d) l'effet marginal de

Jusne 1	meilleur
	meilleur
	pas

fruit	meilleur
	meilleur
	pas

Sport	meilleur
	meilleur
	pas

Sur l'état de santé est de effet

\Rightarrow Sport améliore l'état de santé à proba

mais Jusne et manger des fruits/légumes non

Donc globalement oui les bons comportent améliorent l'état de santé des assurés

4) Mauvaises prél : $718 + 1302 = 2620$

Donc le taux de mauvaises prélev est de $\frac{2620}{10146} = 0,258 \text{ Lo.5}$

Donc le modèle est jugé de bonne qualité.