

## Sleep Disorder Modelling Test: Theories and applications based on the multinomial logit model

*Abdellatif ELLOUMI, Olfa Ben Amor Zegnani, and Abeljelil FARHAT*

Faculty of Economic and Management Sciences, Mahdia, Tunisia

Copyright © 2018 ISSR Journals. This is an open access article distributed under the *Creative Commons Attribution License*, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**ABSTRACT:** This study attempts to model a qualitative endogenous variable whose number exceeds two terms. We address the proper technique, which is the multinomial logit technique\model, to measure the impairment in sleep disorder in terms of the IAH. Then, a succinct description of the available data is provided. The main results achieved, applying the multinomial logistic regression, are expected and explained. The two variables (sex of the patient and Clinical snoring), exhibit very high coefficients in exponentials, which demonstrate their importance in distinguishing between the different categories (L, M and S). We propose a procedure for calculating rates of good rankings. In applying this procedure, we get a moderate rate of high good ranking. This positive output can be explained by the construction of the estimated values using the multi-class method in which these values were constructed using the conditional probabilities.

**KEYWORDS:** Biostatistics, help to the medical decision, sleep disorder, modelling of the qualitative variables, Multinomial Logistic Regression.

### 1 INTRODUCTION

In this paper, we attempt to model a qualitative endogenous variable whose number exceeds two terms. Henceforth, it is about a multinomial model or more particularly a model with a qualitative multinomial variable-i.e., polytomous. In the literature, and according to Amemiya (1981), three categories of multinomial models arise:

- multinomial models of intrinsic order ;
- multinomial models of natural sequence and
- Unordered multinomial models

In practice, and according to Hurlin (2003), the third aforementioned category (Unordered multinomial models) are the most frequently used models in the literature. Within this category, other various model sub-types are identified.

The most commonly used models are the multinomial logit model and the conditional logit model of McFadden (see McFadden (1984)) which include an extension of the binary logit studied in the previous chapter. These two models are easy to use, but they are problematic with regard to the assumption of the independence of irrelevant alternatives. Obviously, having encountered such a limitation, researchers (such as Bouthelier & Daganzo (1979) and Zeileis & Croissant (2010)) have developed alternative models, for instance the hierarchical logit model or the multinomial probit model. However, it must be noted that these latter require rather complex estimation techniques.

Given the inherently endogenous issue, we are going to present and apply the multinomial logit model ordered on a database that includes 256 patients (Tunisians) aged between 9 and 81 years old and observed from 2001 à 2010.

## 2 DATABASE DESCRIPTION AND METHODOLOGY EMPLOYED

Our database was built in 2012 in the department of functional Explorations of the Nervous System in the University Health Centre "Sahloul", Sousse, Tunisia<sup>1</sup>.

### 2.1 DATA BASE VARIABLES

The definitions of the variables are mostly extracted from Baguet et al. (2003) and Brun et al. (2013) studies.

- Patient : name of the patient.
- Year : year of the exam.
- Age : patient's age during the exam.
- Sex : patient's gender (M ou F).
- Medical history : a medical list of probable diseases for each patient, presented with binary variables (0 or 1) as follows:
  - Diabetes : the diabetes is a disease in which the use and stocking of alimentary sugar is impaired by the human body resulting in an elevated levels of glucose in the blood: i.e. hyperglycemia.
  - Hypertension: a cardiovascular disease defined by a high arterial blood pressure.
  - Dyslipidemia (Dys) : which is an abnormal high or low concentration of lipids (triglycerides, cholesterol and/or fat phospholipids) in blood.
  - Cerebrovascular accident (CVA): the medical term known as Cerebrovascular accident (CVA) and referred to as a stroke. This disease is a sudden neurological impairment caused by a blockage or the rupture of a blood vessel in the brain.
  - Coronary artery disease (Core) : which means the impedance or blockage of one or more arteries that supply blood to the heart, usually due to the obstruction or hardening of the arteries.
  - Others: other medical records, such as depression, nasal obstruction, breast cancer, osteoarthritis...
- Weight: patient's weight in kg.
- Height : patient's height in cm.
- body mass index (BMI) : in kg/m<sup>2</sup>.  $IMC = \frac{Weight}{(Height)^2}$
- Clinical signs of snoring: a binary variable indicating whether the patient snores or not (0 or 1).
- Complaint about excessive daytime sleepiness: a binary variable indicating whether the patient is complaining about excessive daytime sleepiness. In these cases, some patients fall asleep on a chair or while driving...
- The Epworth Sleepiness Scale (ESS): It's a variable constructed on the basis of responses to 8 questions scaled from 0 to 3. The ESS score (the sum of 8 item scores, 0-3) can range from 0 to 24. If the score is less than 10, that person's sleep propensity is considered to be normal.
- Migraine: morning headaches on awakening , mainly because not getting enough sleep.
- Asthenia (Asth) : physical weakness and loss of strength..
- the number of apneas (Nap) : frequency of breathing interruption
- Hypopnea rate: low respiratory rate.
- The apnea-Hypopnea Index (AHI) : is an index used to indicate the severity of sleep apnea.

$$IAH = \frac{Nap + Nhyp}{temps} \times 60.$$

In our sample study, we divided the endogenous variable (The apnea-Hypopnea Index (AHI)) into four sub-categories according to the sleeping disorder issue:

- N : None, when  $IAH < 5$ .
- M : Mild, when  $5 \leq IAH < 15$ .
- M : Moderate, when  $15 \leq IAH < 30$ .
- S: Severe, when  $IAH \geq 30$ .

---

<sup>1</sup>For further information, see professor "Mohamed DOGUI", head of department or doctor "Ridha BEN CHEIKH".

The first sub-category will be considered as a reference. We are basing our threshold support on the work of Meslier (2011).

## 2.2 DESCRIPTIVE STATISTICS OF THE DATABASE SELECTED VARIABLES.

Our sample encompasses 155 men and 101 women. The table below summarises the description of the remaining variables according to the number of the patients per category:

*Table 1. Descriptive statistics of variables*

Modality	Diab	HTA	Dys	AVC	Cor	Ant	IAH	Ronf
+ (oui)	38	67	24	5	9	82	152	214
- (non)	218	189	232	251	247	174	104	42

In order to avoid perfect multicollinearity amongst explanatory variables, we have already taken into consideration this criteria while selecting our variables. We have maintained a severe threshold of 50%, which means that only exogenous variables with the correlation coefficient less than or equal to 0,5 are maintained. The following table illustrates the correlation matrix of the exogenous variables that are maintained in this study.

*Table 2. The correlation matrix of the exogenous variables*

Modality	Diab	HTA	Dys	AVC	Cor	Ant	Sexe	Ronf
Diab	1	0.28	0.20	-0.06	0.16	0.07	-0.20	0.07
HTA	0.28	1	0.33	0.17	0.22	0.01	-0.17	0.14
Dys	0.2	0.33	1	0.15	0.08	0.04	0.01	0.03
AVC	-0.06	0.17	0.15	1	0.13	-0.04	-0.06	-0.01
Cor	0.16	0.22	0.08	0.13	1	0.01	-0.02	0.03
Ant	0.07	0.01	0.04	-0.04	0.01	1	-0.01	-0.17
Sexe	-0.2	-0.17	0.01	-0.06	-0.02	-0.01	1	0.07
Ronf	0.07	0.14	0.03	-0.01	0.03	-0.17	0.07	1

We notice that there are no correlated variables to one other. The highest correlation coefficient is  $r^2(\text{Dys};\text{HTA}) = 0,33 < 0,5$ . Henceforth, we can conclude that all of the maintained exogenous variables are independent of one another within our model. To further confirm these results, we have conducted multicollinearity tests of Farrar and Glauber (1968) (see Bourbonnais and Terraza (2008)). Having calculated the transformation of the determinant of the matrix of linear correlation coefficients, it was possible to derive a test statistic, which has  $H_0$  (Chi square)(absence of a multicollinearity problem). It is about a test statistic of Farrar-Glauber. In our case, the P-value = 0,013756 < 0,05. We accept the null hypothesis of then dependence of variables.

The selection results of the variables in our sample were unexpected. We anticipate that other variables would be more discriminatory in the model, mainly the age and the weight of the patient. However, it is not the case with our sample.

## 3 THE CHOICE OF METHODOLOGY

According to Bourbonnais and Terraza (2008), when the dependant variable is qualitative or categorical, the linear regression model is inappropriate. The existence of several categories would automatically create variance classes resulting in heteroscedasticity error problems. The third assumption of Gauss is violated. Hence, another alternative should be sought which is « the logistic regression ». In this paper, we present and employ the logit multi-classes model.

### 3.1 THE MULTINOMIAL LOGIT MODEL

In this study, we present and put onto practice the ordered multinomial model which is exposed and explained by Zeileis and Croissant (2010). The choice of the « ordered » type is dictated by the nature of our database. The endogenous variable ("The apnea-Hypopnea Index (AHI)") is a measure that ranges from 0 to 150. We can hereafter deconstruct it in intervals. The

polytomic univariate ordered model is a model whose variable includes many modalities and a preferred order among those different modalities. We suppose that the modalities are the same for all individuals:

$$m_i = m, \forall i = 1, \dots, N$$

A model is written under the following form:

$$y_i = \begin{cases} 0 & \text{si } y_i^* < c_1 \\ 1 & \text{si } c_1 \leq y_i^* < c_2 \\ \vdots \\ m & \text{si } y_i^* > c_m \end{cases}, \forall i = 1, \dots, N \quad (1)$$

With  $c_{j+1} > c_j$  and the latent variable  $y_i^*$  is defined by :

$$y_i^* = x_i\beta + \varepsilon_i \forall i = 1, \dots, N \quad (2)$$

Saliba et al. (2007) confirm that, if the distribution function  $F(x)$  follows the logistic function, the model is necessarily a multinomial ordered logit model. Whereas, if the distribution function  $F(x)$  follows the standard normal distribution, the model in question is naturally an ordered probit multinomial model.

In practice , such class structuring on  $y_i^*$  does make sense only if the number of classes is relatively weak (in our case 4). Subsequently, the distribution is certainly logistic.

Our study is based on the latter definition, we can derive the distribution of observed qualitative variable  $y_i$  which would serve us later for the plausibility functional construction. For  $i = 1, \dots, N$  we have :

$$\left\{ \begin{array}{l} \mathbf{Prob}(y_i = 0) = \mathbf{Prob}(y_i^* < c_1) = F\left(\frac{c_1 - x_i\beta}{\sigma_\varepsilon}\right) \\ \mathbf{Prob}(y_i = 1) = \mathbf{Prob}(c_1 \leq y_i^* < c_2) = F\left(\frac{c_2 - x_i\beta}{\sigma_\varepsilon}\right) - F\left(\frac{c_1 - x_i\beta}{\sigma_\varepsilon}\right), \forall i = 1, \dots, N \\ \vdots \\ \mathbf{Prob}(y_i = m) = \mathbf{Prob}(y_i^* > c_m) = 1 - F\left(\frac{c_m - x_i\beta}{\sigma_\varepsilon}\right) \end{array} \right. \quad (3)$$

Which could be written in the following general form:

$$\mathbf{Prob}(y_i = j) = F\left(\frac{c_{j+1} - x_i\beta}{\sigma_\varepsilon}\right) - F\left(\frac{c_j - x_i\beta}{\sigma_\varepsilon}\right) \quad (4)$$

Où  $i = 1, \dots, N$  et  $j = 1, \dots, m$

Conventionally, with,  $c_0 = -\infty$  et  $c_{m+1} = +\infty$ .

Then, in order to construct the plausibility associated with the sample  $y$ , and according to the work of Balakrishnan (1990) and Guisan et al. (2002), we call upon, as we used to, to Lagrangian as follows :

$$\mathcal{L}(y, \beta, c_1, \dots, c_m, \sigma_\varepsilon) = \prod_{i=1}^N \prod_{j=1}^{m_i} \left[ F\left(\frac{c_{j+1} - x_i\beta}{\sigma_\varepsilon}\right) - F\left(\frac{c_j - x_i\beta}{\sigma_\varepsilon}\right) \right]^{y_{ij}}$$

Where we are able to define the binary variable  $c_{ij}$  as follows :

$$y_{ij} = \begin{cases} 1 & \text{if } y_i = j \\ 0 & \text{ot otherwise} \end{cases}, \forall i = 1, \dots, N \text{ et } \forall j = 1, \dots, m \quad (6)$$

Generally, we can only identify the following parameters:

$$\tilde{\beta} = \frac{\beta}{\sigma_\varepsilon} \text{ et } \tilde{c}_j = \frac{c_j}{\sigma_\varepsilon}$$

Thereafter, we can explain the plausibility in terms of the following parameters:

$$\mathcal{L}(y, \tilde{\beta}, \tilde{c}_1, \dots, \tilde{c}_m) = \prod_{i=1}^N \prod_{j=1}^{m_i} \left[ F(\tilde{c}_{j+1} - x_i\tilde{\beta}) - F(\tilde{c}_j - x_i\tilde{\beta}) \right]^{y_{ij}} \quad (7)$$

Finally, it is sufficient to maximise the plausibility-log function in  $\tilde{\beta}, \tilde{c}_1, \dots, \tilde{c}_m$  to find the maximum likelihood estimators. The estimators properties are identical with those studied in the dichotomous univariate model. Add to that, we also assume the threshold parameters  $c_j$ :

$$\hat{\beta} = \arg \max_{\{\tilde{\beta}\}} [\ln \mathcal{L}(y, \tilde{\beta}, \tilde{c}_1, \dots, \tilde{c}_m)] \quad (8)$$

$$\hat{c}_j = \arg \max_{\{\tilde{c}_j\}} [\ln \mathcal{L}(y, \tilde{\beta}, \tilde{c}_1, \dots, \tilde{c}_m)] \quad (9)$$

$$\text{With } L(y, \tilde{\beta}, \tilde{c}_1, \dots, \tilde{c}_m) = \sum_{i=1}^N \sum_{j=1}^{m_i} y_{ij} \ln [F(\tilde{c}_{j+1} - x_i \tilde{\beta}) - F(\tilde{c}_j - x_i \tilde{\beta})] \quad (10)$$

Where  $F(\cdot)$  is a given distribution function.

The very first applications of the ordered qualitative multinomial models were with Gurland et al. (1960) used for an insecticidal dosage as well as David and Legg (1975) for a real estate acquisition. In his article, Amemiya (1981) identifies two good examples of economic problems to which this modelling is adjusted.

The first is applied in the favourite/popular domain of the earlier applications of the qualitative models namely the bio-econometrics. It is about Gurland et al's study (1960). The researchers have sought to model the individual tolerance in  $y_i^*$  product depending on an insecticide dosage  $x_i$  sprayed on an  $i^{\text{th}}$  individual. The researchers (ibid.) assume that the observed variable  $y_i$ , which indicates the insect status, can take three probable values. The number of modalities is:  $m = 3$  for  $\forall i = 1, \dots, N$ . The possible values are:

$$y_{i=} \begin{cases} 0 & \text{si } y_i^* > x_i + \delta \\ 1 & \text{si } x_i < y_i^* \leq x_i + \delta \\ 2 & \text{si } y_i^* \leq x_i \end{cases}$$

As the values undertaken by the multinomial variable are:  $(y_i = 0, 1, 2)$ , belonging to one of these modalities is fixed by size criteria (based on thresholds). This is clearly an ordered polytomic model.

The second example, which is broader in response to exogenous variables, is the fruit of David and Legg's labour (1975). The researchers have looked for modelling the real estate price based on a certain number of exogenous variables (the size of the property, the purchaser's age, the purchaser's income, number of years of the purchaser's education...). The authors present the data as follows: If a property is acquired, its price is unobservable. Only if it belongs to one of the three categories of the possible prices can it be observable. Given the inherent nature of the used data, the model is presented in the following way:

$$y_{i=} \begin{cases} 0 & \text{si } y_i^* < c_1 \\ 1 & \text{si } c_1 \leq y_i^* < c_2 \\ 2 & \text{si } y_i^* > c_2 \end{cases}$$

Once more, the values assumed by the property price are:  $(y_i = 0, 1, 2)$ . And belonging to one of these modalities is fixed by size criteria (based on thresholds). It is clearly an ordered polytomic model.

#### 4 RESULTS

In the table below, we present the list of the maintained variables in the logistic regression based on the categories.

Table 3. List of the maintained variables in the logistic regression based on the categories

Variable	Modality	P : (IAH < 5)	L : (5 ≤ IAH < 15)	M : (15 ≤ IAH < 30)	S : (IAH ≥ 30)
Diab	0	90	21	11	96
	1	15	4	7	12
HTA	0	83	20	10	76
	1	22	5	8	32
Dys	0	98	24	16	94
	1	7	1	2	14
AVC	0	103	25	18	105
	1	2	0	0	3
Cor	0	104	24	15	104
	1	1	1	3	4
Ant	0	56	19	12	87
	1	49	6	6	21
Sexe	0	55	14	8	24
	1	50	11	10	84
Ronf	0	41	1	0	0
	1	64	24	18	108

While applying the multinomial logistic regression using the "nnet" package under a free-licence of the software "R" (see Venables and Ripley (2002)), we end up with a procedure convergence after 40 iterations and we obtain the following results:

- First, the plausibility log is 222,831284 starting from a value of 354,891356.
- For the « mild » category, marked by "cat = L", we obtain :

$$\ln\left(\frac{p(L)}{p(P)}\right) = -\frac{3.385}{(1,102)} + \frac{0.026}{(0,688)} Diab - \frac{0.438}{(0,645)} HTA - \frac{0.314}{(1,225)} Dys - \frac{13.122}{(1,4 \times 10^{-6})} Avc + \frac{3.175}{(2,298)} Cor - \frac{0.919}{(0,536)} Ant - \frac{0.0535}{(0,496)} Sexe + \frac{2.819}{(1,087)} Ronf \quad (11)$$

For the « moderate » category, marked by "cat = M", we obtain :

$$\ln\left(\frac{p(M)}{p(P)}\right) = -\frac{25.179}{(0,309)} + \frac{0.907}{(0,684)} Diab + \frac{0.611}{(0,647)} HTA + \frac{0.031}{(1,016)} Dys - \frac{13.097}{(2,4 \times 10^{-6})} Avc + \frac{4.157}{(2,328)} Cor - \frac{0.505}{(0,595)} Ant + \frac{0.934}{(0,618)} Sexe + \frac{23.03}{(0,309)} Ronf \quad (12)$$

For the « mild » category, marked by "cat = L", we obtain:

$$\ln\left(\frac{p(S)}{p(P)}\right) = -\frac{21.369}{(0,171)} - \frac{0.436}{(0,537)} Diab + \frac{0.374}{(0,429)} HTA + \frac{0.869}{(0,682)} Dys + \frac{0.472}{(1,678)} Avc + \frac{3.016}{(2,312)} Cor - \frac{1.399}{(0,382)} Ant + \frac{1.659}{(0,381)} Sexe + \frac{21.117}{(0,171)} Ronf \quad (13)$$

Where : (.) : The estimated standard deviation of the estimated coefficient, Residual Deviance = 445.6626 et AIC = 499.6626

We take as an example, the « snoring » variable; the increase in this variable unit is associated with the increase of odds ratio in the « L » class compared to "P" class of 2, 819 units.

In other words, a patient who has a mild sleep disorder is 2,819 times more likely to snore than a patient who has no sleep disorder problems.

According to Colonna (2006), the ratio reporting the probability of choosing a results' category rather than choosing a reference category is often labelled the relative risk. We have applied the coefficient exponentials of our model to see these risk ratio. We note that the two variables (sex and snoring) have very high exponential coefficients, which proves their importance in distinguishing between the different categories (L, M et S).

We can also use the predicted probabilities to better understand the model. We can calculate the assumed probabilities for each of our results' levels using the function "Fitted" under "R" (see Chambers (2008)). We can start by generating the assumed probabilities for the observations in our set of data package and in the frontlines display:

**Table 4. Extrait des probabilités prédites**

Number	P	L	M	S
1	0.1689417	0.09103106	0.05014657	<b>0.6898807</b>
2	0.6718369	0.15211878	0.04728530	0.1287590
3	0.6432986	0.09402770	0.08338925	0.1792845
4	0.4318418	0.09402770	0.1916263	0.2813199
5	0.3408814	0.04727244	0.1124888	0.4993573
6	<b>0.9935014</b>	0.00649859	$3.276385e^{-11}$	$2.881610e^{-10}$

By examining this output result sample, we take for instance the first and the last (1<sup>st</sup> and 6<sup>th</sup> patient in the database):

- The first is a patient whose IAH = 47, so he belongs to class "S". This explains the high probability "0.6898807" to belong to this category rather than others.
- The last patient whose IAH = 0, belongs to class "P", which explains the high probability "0.9935014" to belong to this category rather than others.

In the literature, we have found a criterion for rate construction of top ranking to this method. Henceforth, we suggest the following procedure:

1. Using the function "Fitted" under "R" (see Chambers and Hastie (1992)). We obtain the probability of belonging to each group.
2. Comparing the probability for each patient while keeping the highest probability.
3. We suggest that the patient belongs to the class whose probability is the highest.
4. Thus we know the actual and estimated classes.

We can sum up the results in the table below:

**Table 5. Rate of top rankings**

Class	P	L	M	S	Total
Actual	105	25	18	108	256
Estimated	91	23	15	96	225
TBC / Class	86.67%	92%	83.33%	88.89%	87.89%

We note that we have a good rate of moderate top rankings (87,89%).

## 5 CONCLUSION

In this paper, we have adopted the multinomial logit technique whilst measuring the impairment in sleep disorder in terms of the IAH. The main results obtained are expected and explained.

The two variables (patient's gender and snoring) have very high exponential coefficients, which proves their importance in distinguishing between the different categories (L, M et S).

We suggest a calculating procedure of good ranking rates. While applying the latter, we obtain a moderate rate of high good ranking. This positive outcome can be further explained by the construction of estimated values through the multi-class method in which we construct the former using conditional probabilities.

## REFERENCES

- [1] T. Amemiya, "Qualitative response models : A survey", *Journal of Economic Literature*, vol. 19, no. 4, pp1483–1536, 1981.
- [2] C. Hurlin, "Econométrie des variables qualitatives. chapitre 2. Modèles multinomiaux. modèles logit multinomiaux ordonnés et non ordonnés", Technical report, Université d'Orléans, 2003.
- [3] D. McFadden, "Chapter 24 econometric analysis of qualitative response models" vol. 2 of *Handbook of Econometrics*, pp 1395–1457. Elsevier, 1984.
- [4] F. Bouthelier and C. F. Daganzo, "Aggregation with multinomial probit and estimation of disaggregate models with aggregate data : A new methodological approach", *Transportation Research Part B : Methodological*, vol. 13, no. 2, pp 133–146, 1979.
- [5] A. Zeileis, and Y. Croissant, "Extended model formulas in r : Multiple parts and multiple responses", *Journal of Statistical Software*, vol. 34, no. 1, pp 1–13, 2010.
- [6] J. P. Baguet, J. L. Pépin, L. Hammer, P. Lévy and J. M. Mallion, "Conséquences cardiovasculaires du syndrome d'apnées obstructives du sommeil", *La Revue de Médecine Interne*, vol. 24 no. 8, pp 530–537, 2003.
- [7] J. F. Brun, C. Fedou, M. Desplan, A. Avignon and J. Mercier, "Évaluation d'un dépistage du syndrome d'apnées et hypopnées du sommeil chez des patients diabétiques et/ou obèses par RU sleeping", *Médecine du Sommeil*, vol. 10 no. 2, pp 62–63, 2013.
- [8] N. Meslier, "Syndrome d'apnées et hypopnées obstructives du sommeil et diabète de type 2. interactions réciproques", *Médecine du Sommeil*, vol. 8, no. 1, pp 21–31, 2011.
- [9] R. Bourbonnais and M. Terraza, "Analyse des séries temporelles, 2e édition", Dunod, Paris, 2008.
- [10] B. Saliba and A. Paraponaris and B. Ventelou, "Situations à l'égard du travail des personnes atteintes de maladies chroniques", *Revue d'Épidémiologie et de Santé Publique*, vol. 55, no. 4, pp 253–263, 2007.
- [11] N. Balakrishnan, "Approximate maximum likelihood estimation for a generalized logistic distribution", *Journal of Statistical Planning and Inference*, vol. 26, no. 2, pp 221–236, 1990.
- [12] A. Guisan and T. C. J. Edwards and T. Hastie, "Generalized linear and generalized additive models in studies of species distributions : setting the scene", *Ecological Modelling*, vol. 157, no. 2–3, pp 89–100, 2002.
- [13] J. Gurland and I. Lee and P. Dahm, "Polychotomous quantal response in biological assay", *Biometrics*, pp 382–388, 1960.
- [14] J. M. David and W. E. Legg, "An application of multivariate probit analysis to the demand for housing", *Journal of Business and Economic Statistics*, pp 295–300, 1975.
- [15] W. N. Venables and B. D. Ripley, "Modern Applied Statistics with S. Fourth edition", Springer, New York, 2002.
- [16] M. Colonna, "Influence des paramètres a priori dans l'estimation bayésienne de risques relatifs. analyse spatiale du cancer de la vessie dans l'agglomération grenobloise", *Revue d'Épidémiologie et de Santé Publique*, vol. 54, no. 6, pp 529 – 542, 2006.
- [17] J. Chambers, "Software for Data Analysis : Programming with R", Springer, New York, 2008.
- [18] J. M. Chambers and T. J. Hastie, "Statistical models in s", Technical report, Wadsworth & Brooks-Cole, 1992.