



## **A case study of Stroke patients in Senegal: application of Generalized extreme value regression model**

**Aba Diop<sup>1,\*</sup>, El Hadji Dème<sup>2</sup>, Idrissa Sy<sup>3</sup> and Aliou Diop<sup>2</sup>**

<sup>1</sup>ERESMA, Université Alioune Diop, Bambey, Sénégal

<sup>2</sup>LERSTAD, Université Gaston Berger, Saint-Louis, Sénégal

<sup>3</sup>Hôpital Aristide Le Dantec, Dakar, Sénégal

Received on March 30, 2021; Accepted on June 1, 2021

Copyright © 2021, The African Journal of Applied Statistics (AJAS) and The Statistics and Probability African Society (SPAS). All rights reserved

**Abstract.** Logistic regression model is widely used in many studies to investigate the relationship between a binary response variable  $Y$  and a set of potential predictors  $X$ . The binary response may represent, for example, the occurrence of some outcome of interest ( $Y = 1$  if the outcome occurred and  $Y = 0$  otherwise). When the dependent variable  $Y$  represents a rare event, the logistic regression model shows relevant drawbacks. In order to overcome these drawbacks we propose the Generalized Extreme Value (*GEV*) regression model. In particular, we suggest the quantile function of the *GEV* distribution as link function. Strokes are a serious pathology and a neurological emergency involving the vital prognosis and the functional prognosis. In Senegal, strokes account for more than 30% of hospitalizations and are responsible for nearly two thirds of mortality. In this work, we use the *GVE* regression model for binary data to determine the risk factors leading to stroke and to develop a predictive model of life-threatening outcomes in central Senegal.

**Key words:** generalized extreme value model; logistic regression model; stroke.

**AMS 2010 Mathematics Subject Classification Objects :** 62P10; 62F03.

---

\*Corresponding author: Aba Diop ([aba.diop@uadb.edu.sn](mailto:aba.diop@uadb.edu.sn))

El Hadji Dème: [elhadjidemeufrsat@gmail.com](mailto:elhadjidemeufrsat@gmail.com)

Idrissa Sy: [IDY.SY.10@hotmail.com](mailto:IDY.SY.10@hotmail.com)

Aliou Diop: [aliou.diop@ugb.edu.sn](mailto:aliou.diop@ugb.edu.sn)

**Résumé.** (Abstract in French) Le modèle de régression logistique est utilisé dans beaucoup de domaine pour étudier la relation entre une variable réponse binaire  $Y$  et un ensemble de prédicteurs  $X$ . La réponse peut représenter, par exemple, l'occurrence d'un certain résultat d'intérêt ( $Y = 1$  si le résultat s'est produit et  $Y = 0$  sinon). Lorsque la variable dépendante  $Y$  représente un événement rare, la régression logistique présente des inconvénients notables. Afin de surmonter ces inconvénients, nous proposons le modèle de régression basé sur la loi généralisée des valeur extrêmes (*GVE*). En particulier, nous proposons le quantile de la loi généralisée des valeurs extrêmes comme fonction de lien. Les accidents vasculaires cérébraux (*AVC*) sont une pathologie grave et une urgence neurologique mettant en jeu le pronostic vital. Au Sénégal, les *AVC* représentent plus de 30% des hospitalisations et sont responsables de près de deux tiers de la mortalité. Dans ce travail, nous utilisons le modèle de régression *GVE* sur des données binaires pour déterminer les facteurs conduisant à un *AVC* et de développer un modèle prédictif de l'engagement du pronostic vital au Sénégal-central.

**The authors.**

**Aba DIOP**, Ph.D., is Assistant Professor at University Alioune Diop of Bambey, SENEGAL

**El Hadji Dème**, Ph.D., is Professor of Mathematics and Statistics at University Gaston Berger of Saint-Louis, SENEGAL

**Idrissa Sy**, Engineer, is a Researcher in Epidemiology, Aristide Le Dantec hospital of Dakar, SENEGAL

**Aliou Diop**, Ph.D. and Doctorate of Sciences, is a Full Professor in Mathematics and Statistics at University Gaston Berger of Saint-Louis, SENEGAL

## 1. Introduction

Stroke is a public health issue, a serious pathology and a neurological emergency involving both vital and functional prognosis [Biousse \(1994\)](#), [Bogousslavsky et al. \(1993\)](#). The term "stroke" suggests an unstoppable event and does not suggest that it is a sudden complication of a chronic vascular disease that has been evolving quietly for many years. This definition allows us to say that, stroke is a frequent and serious pathology that causes many disabilities. (see [Bousser and Mas \(2000\)](#)).

Stroke is a sudden loss of brain function due to an infarction or haemorrhage. There are two pathological entities:

**Ischemic strokes**, also called cerebral infarction or softening of the brain, are most often the result of a thrombus (a clot that forms in an artery), an embolus (a foreign body, most often a clot, which, carried by the circulation, will obstruct the artery downstream), or a narrowing of the artery caused by atherosclerosis (thickening of the inner lining of the arterial wall). Atherosclerosis is the leading cause of ischemic stroke (50-60% of cases).

**Hemorrhagic strokes** are caused by an effusion of blood into the brain tissue. Their cause is usually high blood pressure or, much more rarely, vascular malformations (angioma, aneurysm), clotting disorders or complications of anticoagulant treatment [Grillo \*et al.\* \(2006\)](#).

Logistic regression is widely used to model binary response data in medical studies ([Agresti \(2002\)](#), [Celeux and Nakache \(1994\)](#) and [Menard \(2002\)](#)). An example of a binary response variable is the infection status (infected *vs* uninfected) with respect to some disease. A logistic regression model can be used to investigate the relationship between the infection status and various potential predictors [Hilbe\(2009\)](#), [Nakache and Confais \(2003\)](#). If  $Y_i$  denotes the infection status for the  $i$ -th individual in a sample of size  $n$  ( $Y_i = 1$  if the individual is infected, and  $Y_i = 0$  otherwise), and  $\mathbf{X}_i$  denotes the corresponding ( $p$ -dimensional, say) predictor, the logistic regression model expresses the relationship between  $Y_i$  and  $\mathbf{X}_i$  in term of the conditional probability  $\mathbb{P}(Y_i = 1|\mathbf{X}_i)$  of infection, as:

$$\log \left( \frac{\mathbb{P}(Y_i = 1|\mathbf{X}_i)}{1 - \mathbb{P}(Y_i = 1|\mathbf{X}_i)} \right) = \beta^\top \mathbf{X}_i,$$

where  $\beta \in \mathbb{R}^p$  is an unknown parameter to be estimated. An extensive literature has been devoted so far to statistical inference in logistic regression models (see [Agresti \(2002\)](#), [Celeux and Nakache \(1994\)](#), [Nakache and Confais \(2003\)](#) and [Collet\(2003\)](#)). Estimation and testing procedures for this class of models are now well established and are available in standard statistical softwares [Hilbe\(2009\)](#).

When the logistic regression model is employed, it is assumed that the response curve between the covariates and the probability is symmetric. This assumption may not always be true, and it may be severely violated when the number of observations in the two response categories are significantly different from each other. This unbalance is not uncommon when we consider binary rare events data (i.e. binary dependent variables with a very small number of ones) which happens with only a small probability. Applying a nonflexible link function to the data with this special feature may result in link misspecification. Consequences of link misspecification have been studied by a number of authors in the literature (see, for example [Czado and Santner \(1992\)](#) and [King and Zeng \(2001\)](#)). In particular, for independent binary observations, [Czado and Santner \(1992\)](#) show that falsely assuming a logistic link leads to a substantial increase in the bias and mean squared error of the parameter estimates as well as the predicted probabilities, both asymptotically and in finite samples. The logistic regression shows same important drawbacks in rare events studies: the probability of rare event is underestimated and the logit link is a symmetric function, so the response curve approaches zero as the same rate it approaches one.

In this context, focusing on binary rare events data, [Calabrese and Osmetti \(2013\)](#) used GEV regression model, that uses the quantile function of the GEV distribution as link function, for analyzing default probabilities and find that its predictive

performance is better than that of the logistic regression predictive model. They point out that logistic regression could underestimate the probability of default of rare events since the link function is a symmetric function. Wang and Dey (2010) used this model for an application to B2B electronic payments system adoption.

The rest of this paper is organized as follows. In Section 2, we describe the GEV regression model adapted to this setting. In Section 3, we apply the GEV regression model to a real data on stroke patients. A discussion, recommendation and some perspectives are given respectively in Sections 4 and 5.

## 2. Generalized extreme value regression model

Let  $(Y_1, \mathbf{X}_1), \dots, (Y_n, \mathbf{X}_n)$  be independent and identically distributed copies of the random vector  $(Y, \mathbf{X})$  defined on the probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . For every individual  $i = 1, \dots, n$ ,  $Y_i$  is a binary response variable indicating say, the infection status with respect to some disease (that is,  $Y_i = 1$  if the  $i$ -th individual is infected, and  $Y_i = 0$  otherwise). Let  $\mathbf{X}_i = (1, X_{i2}, \dots, X_{ip})'$  be random vectors of predictors or covariates.

The proposed link function based on the GEV distribution (see Wang and Dey (2010)) is given by

$$G(x|\mu, \sigma, \tau) = \begin{cases} \exp \left[ - \left\{ 1 + \tau \frac{x-\mu}{\sigma} \right\}_+^{-1/\tau} \right] & \text{if } \tau \neq 0 \\ \exp \left[ - \exp \left\{ \frac{x-\mu}{\sigma} \right\} \right] & \text{if } \tau = 0 \end{cases} \quad (1)$$

where  $\mu \in \mathbb{R}$  is the location parameter,  $\sigma \in \mathbb{R}_+$  is the scale parameter,  $\tau \in \mathbb{R}$  is the shape parameter and  $x_+ = \max(x, 0)$ . A more detailed discussion on the extreme value distributions can be found in Coles (2001) and Kotz and Nadarajah (2000). Its importance as a link function arises from the fact that the shape parameter  $\tau$  purely controls the tail behavior of the distribution.

For a binary response variable  $Y_i$  and the vector of explanatory variables  $\mathbf{x}_i$ , let  $\pi(\mathbf{x}_i) = \mathbb{P}(Y_i = 1 | \mathbf{X}_i = \mathbf{x}_i)$  the conditional probability of infection. Since we consider the class of Generalized Linear Models, we suggest the GEV cumulative distribution function proposed by Calabrese and Osmetti (2013) as the response curve given by

$$\begin{aligned} \pi(\mathbf{x}_i) &= 1 - \exp\{[(1 - \tau(\beta_1 + \beta_2 \mathbf{x}_{i2} + \dots + \beta_p \mathbf{x}_{ip}))_+]^{-1/\tau}\} \\ &= 1 - \text{GEV}(-\beta' \mathbf{x}_i; \tau) \end{aligned} \quad (2)$$

where  $\beta = (\beta_1, \dots, \beta_p)' \in \mathbb{R}^p$  is an unknown regression parameter measuring the association between potential predictors and the risk of infection (for a susceptible individual) and  $\text{GEV}(x; \tau)$  represents the cumulative probability at  $x$  for the GEV distribution with a location parameter  $\mu = 0$ , a scale parameter  $\sigma = 1$ , an unknown shape parameter  $\tau$ . For  $\tau \rightarrow 0$ , the previous model (2) becomes the response curve of the log-log model, for  $\tau > 0$  and  $\tau < 0$  it becomes the Fréchet and Weibull response curve respectively, a particular case of the GEV one.

The link function of the GEV model is given by

$$\frac{1 - [\log(1 - \pi(x_i))]^{-\tau}}{\tau} = \beta' \mathbf{x}_i =: \eta(\mathbf{x}_i) \quad (3)$$

that represents a noncanonical link function. Note the parameters  $\mu$  and  $\sigma$  are set to fixed constants for model identifiability. Wang and Dey (2010) showed that the GEV link model specified in 2 is negatively skewed for  $\tau < \log(2) - 1$  and positively skewed for  $\tau > \log(2) - 1$ . The link function is approximately symmetric at  $\tau = \log(2) - 1$ .

The likelihood function for the unknown  $p$ -dimensional parameter  $\beta$  from the independent sample  $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$  is as follows:

$$L_n(\beta) = \prod_{i=1}^n [1 - \text{GEV}(-\beta' \mathbf{x}_i; \tau)]^{y_i} \times [\text{GEV}(-\beta' \mathbf{x}_i; \tau)]^{1-y_i}. \quad (4)$$

We define the maximum likelihood estimator  $\hat{\beta}_n$  as the solution of the  $p$ -dimensional score equation

$$\dot{l}_n(\beta) = \frac{\partial l_n(\beta)}{\partial \beta} = 0, \quad (5)$$

where  $l_n(\beta) = \log L_n(\beta)$  is the log-likelihood function.

**Remark 1.** Wang and Dey (2010) favor a Bayesian analysis since Bayesian methods do not depend on the regularity assumptions required by the asymptotic theory of maximum likelihood. In particular, in the unusual situation where  $\tau < -0.5$  and the classical theory of maximum likelihood breaks down, Bayesian inference provides a variable alternative. Then, we use in this setting the *bgeva* function of the statistical software R.

### 3. A study on Stroke data in Senegal

#### 3.1. Data description

In this section, we consider an application on Stroke data in central Senegal. Stroke is a sudden neurological deficit of vascular origin caused by an infarct or haemorrhage in the brain (see Bogousslavsky *et al.* (1993) and Biousse (1994) for more details). We consider here a database of size  $n = 162$ . The data was collected in the context of a prospective and analytical study, carried out on a period of **8 months** from **april 5** to **november 30, 2016** at Medical Imagery Service of both **Matlaboul Fawzeini Hospital** in **Touba** and **Elhadj Ibrahima Niass regional hospital** in **Kaolack** located in central Senegal. Patients with

**Table 1.** Nature of covariates

Covariate	Nature	Abbreviation
<b>Age</b>	age of stroke patients	Age
<b>Sex</b>	sex of stroke patients	Sex
<b>Stroke type</b>	Ischemic stroke or Hemorrhagic stroke	Stroke-Typ
<b>Cardiopathy</b>	coronary insufficiency	Cardiopathy
<b>Diabetes</b>	insufficient insulin production by the pancreas	Diabetes
<b>Hypertension</b>	abnormal increase of blood pressure on artery walls	Hypertension
<b>Hemiplegia</b>	total or partial paralysis of one half (left or right) of the body	Hemiplegia
<b>Disturbance of consciousness</b>	any disturbance of vigilance and conscious thinking	Disturb-cons
<b>Motor deficiency</b>	affected mobility of the upper and/or lower limbs	Motor-def
<b>Severity Cerebral commitment</b>	displacement of parts of the nervous structure contained in the cranium through an orifice	SC-commitment
<b>Intraventricular haemorrhages</b>	bleeding into the ventricles of the brain	Ivh
<b>Hospital Admission Delay</b>	delay between the first symptoms and admission to hospital	Delay
<b>Vital Prognosis</b>	vital prognosis engaged	Prognosis

Age	[24 – 40[	[40 – 60[	[60 – 95[
Size.	12	55	95
Freq. (%)	7.4	34	58.6

Sex	Female	Male
Size.	73	89
Freq. (%)	45.1	59.4

CT confirmation of stroke were included in the study. We will therefore investigate the factors which may explain an unfavourable evolution of their health status.

In Sénégal, stroke is the most frequent neurological disease. Known for their high mortality and morbidity rates, they account for more than 30% of hospital admissions and nearly two-thirds of the loss of human life (see [Sene-Diouf et al.](#) and [Touré et al.](#)).

We consider the following covariates in the dataset:

Hypertension	No	Yes
Size.	47	115
Freq. (%)	29	71

Stroke-Type	Normal	Ischemic	Hemorrhagic
Size.	25	98	39
Freq. (%)	15.4	60.5	24.1

Cardiopathy	No	Yes
Size.	143	19
Freq. (%)	88.3	11.7

Diabetes	No	Yes
Size.	122	40
Freq. (%)	75.3	24.7

Motor-def	No	Yes
Size.	27	135
Freq. (%)	16.7	83.3

Dist-cons	No	Yes
Size.	88	74
Freq. (%)	54.3	45.7

Hemiplegia	No	Yes
Size.	63	99
Freq. (%)	38.9	61.1

SC-commitment	No	Yes
Size.	112	50
Freq. (%)	69.1	30.1

### 3.2. Results

In this study, the dependent variable is the evolution of the health status of stroke patients (vital prognosis). We denote  $Y$  the binary variable defined as follows:

$$Y_i = \begin{cases} 1 & \text{if the vital prognosis evolves favourably} \\ 0 & \text{if the vital prognosis evolves unfavourably} \end{cases}$$

Our aim is to

1. Identify the factors leading to stroke;
2. Evaluate the scores of new cases and predict the vital prognosis in these new cases.

We ran a generalized extreme value regression analysis of the model defined as follows:

$$\mathbb{P}(Y_i = 1|x) = 1 - GEV(-(\beta_1 + \beta_2 \times \mathbf{Age} + \beta_3 \times \mathbf{Sex} + \beta_4 \times \mathbf{Stroke-typ} + \beta_5 \times \mathbf{Cardiopathy} + \beta_6 \times \mathbf{Diabetes} + \beta_7 \times \mathbf{Hypertension}) + \beta_8 \times \mathbf{Hemiplegia} + \beta_9 \times \mathbf{Disturb-cons} + \beta_{10} \times \mathbf{Motor-def} + \beta_{11} \times \mathbf{SC-commitment} + \beta_{12} \times \mathbf{Ivh} + \beta_{13} \times \mathbf{Delay}; \tau).$$

Ivh	No	Yes
Size.	126	36
Freq. (%)	77.8	22.2

Prognosis	Unfavorable	Favorable
Size.	46	116
Freq. (%)	28.4	71.6

Delay (hours)	< 6	]6 - 24]	]24 - 48]	]48 - 72]	> 72 [ [
Size.	54	50	30	13	15
Freq.	33.3	30.9	18.5	8	9.3

We use the backward stepwise variable selection method (using Wald testing at level 10%) to choose the significant covariates to be retained in the final model. The final results of these fitting procedure are given in Table 2.

**Table 2.** Stroke data analysis

parameter	estimate	stand. error	p-value
<i>Intercept</i>	4.381	0.889	8.33e-07
<b>Age</b>	0.4026	0.2083	0.053
<b>Delay</b>	-0.618	0.1112	2.65e-08
		<b>Clinical signs</b>	
<b>Disturb-cons</b>	0.635	0.298	0.033
<b>Motor-def</b>	-3.599	0.605	2.72e-09
		<b>Severity signs</b>	
<b>SC-commitment</b>	0.502	0.304	0.098
<b>Ivh</b>	1.383	0.4007	0.0005

The adjusted probability of vital prognosis for the  $i$ -th individual is given by

$$\mathbb{P}(Y_i = 1|x) = 1 - GEV(-(4.381 + 0.4026 \times \mathbf{Age} - 0.618 \times \mathbf{Delay} + 0.635 \times \mathbf{Disturb-cons} - 3.599 \times \mathbf{Motor-def} + 0.502 \times \mathbf{SC-commitment} + 1.383 \times \mathbf{Ivh}; \tau).$$

#### 4. Discussion

In this paper, we have considered the problem of estimating the generalized extreme value regression model for binary response data using stroke data in central Senegal.

Age has a significant influence on the vital prognosis of stroke patients. Older patients are more at risk than other patents. Indeed, age favours the degradation of functioning of blood vessels. This imbalance is also highlighted by the synchronicity of cardiovascular risk factors, which increases the vulnerability to stroke (Diagne *et al.* (2016), Zuber and Mas (1994)).

For clinical signs, the most evocative are disturbances of consciousness and motor deficiency. These are all disturbances of vigilance, awakening and conscious thought and affected mobility of the upper and/or lower limbs. In fact, more patients in our study population with consciousness disturbances or motor deficiency had an unfavorable evolution of their state of health.

The scanner must be performed urgently, at best within the first six hours after the onset of symptoms. We were also interested in the time between the first signs



of stroke and admission to the health facility and the time between admission and CT scan. The first was more significant for unfavorable evolution of the vital prognosis of stroke patients. This explains why this delay increases the vital prognosis because it reduces the possibilities of functional rehabilitation.

This delay would be linked to various factors:

- The relatively high cost of scan for patients who are often elderly and without financial resources;
- Difficulty and lateness in transporting patients to hospital;
- Some patients were first guided to traditional practitioners before being transferred to hospital.

For severity signs, their appearance seriously engages the vital prognosis:

- Severity Cerebral commitment which is a fearsome mechanical complication of intracranial hypertension leading to the expulsion of some parts of the brain through the intracranial orifices.
- Intraventricular hemorrhages, which are bleeds that enter the ventricles of the brain.

## 5. Recommendations and Perspectives

Considering the diversity of risk factors and the high cost of medical care, awareness must be raised to enable assistance to be provided at the first signs of a stroke and, above all, to collect more data that would enable the development of guidance and decision-making tools.

In regression model for binary data, it is usually of interest to estimate the probability of infection  $\pi(x)(\mathbf{x}) = \mathbb{P}(Y = 1 | \mathbf{X} = \mathbf{x})$ , for some given value  $\mathbf{x}$  of the covariates and to investigate its properties. Another issue of interest deals with the inference in the generalized extreme value regression model in a high-dimensional setting, when the predictor dimension is much larger than the sample size (this problem arises, for example, in genetic studies where high-dimensional data are generated using microarray technologies).

**Acknowledgment.** The authors acknowledge from the Epidemiology department of *Aristide Le Dantec Hospital* (Hôpital Aristide Le Dantec de Dakar) of Senegal. They also acknowledge grants from the editor team of the African Journal of Applied Statistic for the handling of this paper and the Africa Statistika Journal.

## References

- Agresti, A., 2002. Categorical Data Analysis. Wiley, New York.
- Bioussé V., 1994. Etiologie et mécanisme des accidents vasculaires cérébraux. *AnnRadiol* 37: 11-16.
- Boussier M.G., Mas J.L., 2000. Accident vasculaires cérébraux *Traité de Neurologie*.

- Bogousslavsky J., Bousser M.G., Mas J.L., 1993. Les accidents vasculaires cérébraux. Radiologie ed Doinville, 145.
- Calabrese R., Osmetti S., 2013. Modelling SME Loan Defaults as Rare Events: an Application to Credit Defaults. *Journal of Applied Statistics* 40 (6), 1172-1188.
- Celex G., Nakache J.P., 1994. Analyse Discriminante sur Variables Qualitatives. *Polytechnica*.
- Collet D., 2003. Modelling Binary Data, 2nd ed. Boca Raton, FL: Chapman & Hall/CRC.
- Cordeiro G., de Castro M., 2011. A new family of generalized distributions. *Journal of Statistical Computation and Simulation* 81, 883-898.
- Czado C., Santner T.J., 1992. The effect of link misspecification on binary regression inference. *Journal of Statistical Planning and Inference* 33, 213-231.
- Coles S.G., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer, New York.
- Diagne, N.S. *et al.*, 2016. Vécu post-AVC d'une cohorte sénégalaise: quel facteur de risque pour quel items de qualité de vie. *Journal de Réadaptation Médicale: Pratique et Formation en Médecine Physique et de Réadaptation*, Volume 36, Issue 3, 156-160.
- Dobson A.J., Barnett A.G., 2008. Introduction to Generalized Linear Models (3rd ed.). Chapman and Hall CRC, Boca Raton.
- Famoye F., Singh K.P., 2006. Zero-inflated generalized Poisson regression model with an application to domestic violence data. *Journal of Data Science*, 4, 117-130.
- Grillo P., Velly L., Bruder N., 2006. Accident vasculaire cérébral hémorragique: nouveautés sur la prise en charge. *Annales Françaises d'Anesthésie et de Réanimation*, 25: pp. 868-873.
- Hilbe J.M., 2009. Logistic regression models. Chapman & Hall: Boca Raton.
- Hosmer D.W., Lemeshow S., 2000. Applied logistic regression. Wiley: New York.
- King G., Zeng L., 2001. Logistic Regression in Rare Events Data. *Political Analysis*, 9, 137-163.
- Kotz S., Nadarajah S., 2000. Extreme Value Distributions. Theory and Applications, Imperial College Press, London.
- Manski C.F., Lerman S.R., 1977. The Estimation of Choice Probabilities from Choicebased Samples. *Econometrica* 45 (8).
- Menard S., 2002. Applied Logistic Regression Analysis (Second Edition), Series: Quantitative Applications in the Social Sciences, numéro 106. *Sage Publications*.
- McCullagh P., Nelder J.A., 1989. Generalized Linear Model, Chapman Hall, New York.
- Moghimbeigi A., Eshraghian M.R., Mohammad K., and McArdl, B., 2009. Multilevel zero-inflated negative binomial regression modeling for over-dispersed count data with extra zeros. *Journal of Applied Statistics* 35, 1193-1202.
- Nakache JP, Confais J. Statistique Explicative Appliquée, Partie 2, "Modèle Logistique". *Technip*: pp. 77-168.
- Zabsonre P. *et al.*, 1997. Etude des facteurs de risque de gravité des accidents vasculaires cérébraux chez des noirs ouest africains au Burkina Faso. *Med Trop*.
- Sene-Diouf F. *et al.*, 2007. The management of cerebrovascular events in Senegal. *Revue Neurologique*, volume 163, Issues 8-9, September, p. 823-827.
- Touré K. *et al.*, 2016. Mortalité des patients hospitalisés pour AVC ischémique en neurologie au CHU de Fann à Dakar. *Neurologie-Psychiatrie-Gériatrie*, Vol 17 - Num. 100, p. 230-234. Doi: 10.1016/j.npg.2016.11.006
- Wang X., Dey D.K., 2010. Generalized extreme value regression for binary response data: An application to B2B electronic payments system adoption. *Ann. Appl. Stat.* 4 2000-2023.
- Zuber M., Mas J.L., 1994. Epidémiologie des infarctus cérébraux. *AnnRadiol*, 37, 7-10