

# **ETUDE DE COMPORTEMENT DES CONTRIBUABLES MINIERES SUR LEUR PARESSE A PAYER LES IMPOTS : CAS DE LA REGION VAKINANKARATRA A MADAGASCAR**

1- RAJAONARISOA Ando Nomenjanahary,  
[Ecole Doctorale Ingénierie et Géoscience \(INGE\)](#)

2- RAKOTONINDRAINY

Professeur Titulaire O L'IESAV

Université d'Antananarivo

3-Docteur HERITAHINA RAMBELOSON

Maitre de Conférences

[Ecole Doctorale Ingénierie et Géoscience \(INGE\)](#)

[Université d'Antananarivo](#)

4- Docteur [RANDRIANJA Kanto Volahasina](#)

Maitre de Conférences

[Ecole Doctorale Ingénierie et Géoscience \(INGE\)](#)

- Université d'Antananarivo

## **RESUME**

Les études menées dans le cadre de la présente thèse ont permis de faire dégager d'une part les impacts et points positifs de la fiscalité Malagasy, et d'autre part ses impacts et points négatifs ainsi que leurs causes. Les études ont considéré également l'impact négatif le plus important qui est la fuite ou la paresse des contribuables miniers à payer leurs impôts. Des analyses sur ce comportement de paresse ont été effectuées afin de pouvoir trouver les facteurs qui l'expliquent. Les modélisations successives, par régression logistique, ont permis de trouver les priorités des facteurs explicatifs. Une fois trouvés, ces facteurs constituent les problèmes urgents à résoudre.

Les recherches ont abouti ainsi à aider la prise de décision dans la politique de rehaussement de nombre des contribuables dans le secteur minier au niveau de la région Vakinankaratra. Et par conséquent, les investigations et les observations faites contribueront à l'augmentation des recettes fiscales en provenance du secteur mine, conduisant de ce fait, à la valorisation des ressources minières de la région Vakinankaratra.

**Mots clés : Fiscalité, Impact, Contribuables dans le secteur mine, Régression logistique**

## **ABSTRACT**

In this thesis, we have conducted studies in order to discover positive impacts and positive points of the Malagasy tax system in one hand, and to find out its negative impacts and its negative points in other hand. The most harmful negative impact is the laziness or the escape of mining taxpayers to pay their taxes. The analysis of the behaviour of laziness was carried out in order to find the factors which explain this behaviour. The successive modelling, by using logistic regression, has allowed to find the priority of the problems. The explanatory factors constitute the problem to be resolved.

The researches result in helping to make decision in increasing the number of mining taxpayers in the region of Vakinankaratra. Consequently, the carried out investigation and observation contribute to the rise of public takings from the mining sector.

**Key words : Tax system, Impact, mining taxpayers, logistic regression**

## INTRODUCTION

L'adage célèbre Malagasy dit que la richesse est comme les poils du nez : si on en arrache beaucoup, ça fait mal, si on en arrache peu, ça fait mal aussi. Cette maxime populaire et vieille est généralement vraie car le paiement de l'impôt, qui est tiré de la richesse, est douloureux pour certains individus. Mais il existe aussi des contribuables qui accomplissent leur devoir de citoyen (ou plus précisément le paiement de l'impôt) en toute honnêteté. La fuite devant l'impôt ou la paresse des contribuables, quelque soit la raison, figure parmi les problèmes qui causent la hausse des activités informelles. Trouver la raison expliquant cette paresse permettrait à l'Administration de corriger certains éléments de la fiscalité pour rendre cette dernière plus attractive. C'est dans cette optique que focalise la présente publication.

En résumé, nos objectifs sont alors :

- d'expliquer par modélisation, la fuite ou la paresse des contribuables à payer leurs impôts par le choix de plusieurs variables explicatives les plus pertinentes, et par l'utilisation des outils informatiques
- de trouver un autre modèle plus parcimonieux dont les variables explicatives sont limitées en nombre pour retenir les variables indépendantes les plus pertinentes. Ces dernières constituent ainsi les problèmes urgents à résoudre pour remédier à la fuite ou à la paresse des contribuables pour payer leurs impôts.

## 1. PRESENTATION DE LA REGION VAKINANKARATRA

### 1.1. Situation géographique de la région de Vakinankaratra

La région Vakinankaratra est l'une des vingt-deux régions de Madagascar. Elle est située dans la Province d'Antananarivo, au centre de l'île. Elle est délimitée à l'Est par les régions d'Alaotra-Mangoro et d'Atsinanana, à l'Ouest par la région de Menabe, au Nord par les régions d'Analamanga, d'Itasy et de Bongolava ; et au Sud par la région d'Amoron'i Mania. Le chef-lieu de la région est Antsirabe I.



**Figure 1.** Localisation de la région de Vakinankaratra à Madagascar.

### 1.2. Démographie

La concentration humaine (la plus forte densité) se trouve surtout à Antsirabe. La densité globale de la population est de l'ordre de 78,9 habitants au km<sup>2</sup>.

| Nom District | Population résidente | Superficie en km <sup>2</sup> | Densité hab/ km <sup>2</sup> | % Population résidente |
|--------------|----------------------|-------------------------------|------------------------------|------------------------|
| Ambatolampy  | 25 000               | 1 709                         | 14,6                         | 1,65                   |
| Antanifotsy  | 300 040              | 3 425                         | 87,6                         | 19,80                  |
| Antsirabe I  | 186 633              | 180                           | 1 036,9                      | 12,32                  |
| Antsirabe II | 337 543              | 2 769                         | 121,9                        | 22,27                  |
| Betafo       | 338 636              | 4 607                         | 73,5                         | 22,35                  |
| Faratsiho    | 181 202              | 2 015                         | 89,9                         | 11,96                  |
| Mandoto      | 146 333              | 4 500                         | 32,5                         | 9,66                   |
| TOTAL        | 1 515 387            | 19 205                        | 78,9                         | 100                    |

**Tableau 1.** Répartition spatiale de la population  
Source : Préfecture Vakinankaratra (2002)

### 1.3. Les bureaux fiscaux dans la région de Vakinankaratra

Les bureaux fiscaux de la région Vakinankaratra ne sont pas au même nombre que les districts. Ainsi, certains districts ont été regroupés pour un seul bureau fiscal.

Les bureaux fiscaux sont répartis comme suit :

- Le Centre Fiscal Antsirabe A qui est chargé des contribuables du district d'Antsirabe I ayant un chiffre d'affaire de moins de 200 millions Ariary.
- Le Centre Fiscal Antsirabe B qui s'occupe des contribuables du district d'Antsirabe II ayant un chiffre d'affaire de moins de 200 millions Ariary.
- Le Centre Fiscal Betafo qui se charge de recevoir les contribuables des districts Betafo et Mandoto ayant un chiffre d'affaire de moins de 200 millions Ariary.
- Le Centre Fiscal Ambatolampy qui est responsable des contribuables du district d'Ambatolampy ayant un chiffre d'affaire de moins de 200 millions Ariary.
- Le Centre Fiscal Antanifotsy qui est chargé de recevoir les contribuables du district Antanifotsy ayant un chiffre d'affaire de moins de 200 millions Ariary.
- Le Service Régional des Entreprises de Vakinankaratra qui s'occupe des contribuables ayant un chiffre d'affaire de plus de 200 millions Ariary et de moins de 4 milliards Ariary.
- La Direction des Grandes Entreprises qui se charge de recevoir les contribuables ayant un chiffre d'affaire de plus de 4 milliards Ariary.
- 

### 1.4. Les opérateurs miniers répertoriés dans l'annuaire fiscal à Vakinankaratra

Pour voir un aperçu général de l'évolution du nombre des contribuables légaux (qui paient leurs impôts) du secteur minier situés dans la région du Vakinankaratra, il nous est indispensable de présenter le tableau suivant :

| Année | Nombre |
|-------|--------|
| 2016  | 243    |
| 2017  | 301    |
| 2018  | 320    |
| 2019  | 381    |

**Tableau 2.** Nombre d'opérateurs miniers répertoriés dans le Fisc de Vakinankaratra  
Source : base de données NIFONLINE de la Direction Générale des Impôts

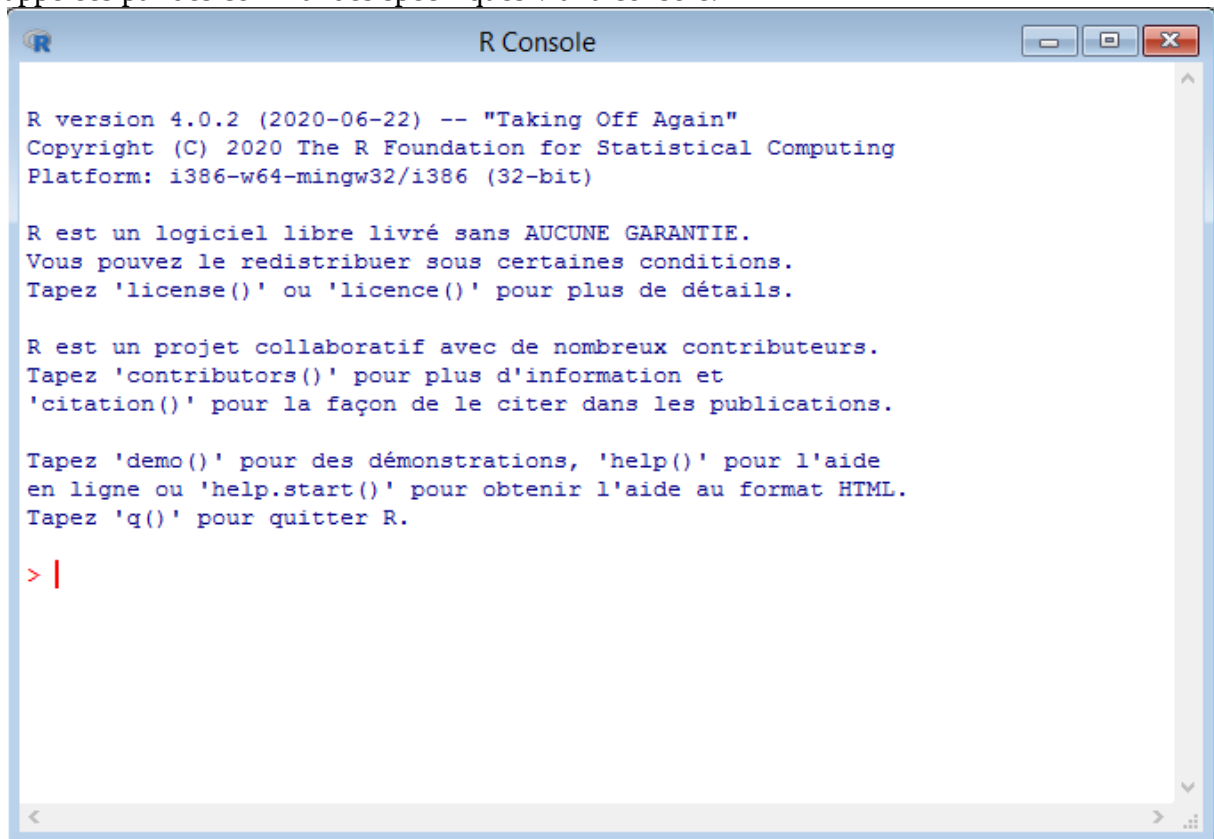
## 2. METHODOLOGIE, MODELISATION

### 2.1. Présentation du logiciel R

R est un langage de programmation et un logiciel libre destiné aux statistiques et à la science des données. Il fait partie de la liste des paquets GNU.C'est un logiciel libre distribué selon les termes de la licence GNU GPL.

R est un langage interprété où les utilisateurs utilisent une interface en ligne de commande.

R fonctionne avec plusieurs fenêtres sous Windows. En particulier nous distinguons la fenêtre R console, c'est-à-dire la fenêtre principale où sont réalisées par défaut les entrées de commandes et sorties de résultats en mode texte. À celle-ci peuvent s'ajouter un certain nombre de fenêtres facultatives telles que les fenêtres graphiques et les fenêtres d'informations (historique des commandes, aide, visualisation de fichiers, . . .), toutes appelées par des commandes spécifiques via la console.



```
R Console

R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.

R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

> |
```

Figure N-1 : R Console

### 2.2. Choix des variables explicatives et variable expliquée

Le choix de la variable d'intérêt ou variable expliquée a été effectué sur la base du problème de la faiblesse des nombres des contribuables voulant payer leurs impôts. Pourquoi ce nombre faible ? Ainsi, cette variable d'intérêt est la suivante :

« Paresse du contribuable à payer son impôt », dénommée PARESSE qui peut être 0 ou 1. 1 si le contribuable n'est pas enthousiaste à payer son impôt, 0 dans le cas contraire.

Les variables explicatives les plus pertinentes qui peuvent expliquer la paresse sur le comportement du contribuable à payer l'impôt sont citées ci-après :

- Le contribuable n'accepte et n'approuve pas la collaboration entre lui et le député pour définir la loi fiscale ; variable dénommée RDEP qui est binaire. 1 si le contribuable n'approuve pas, 0 si il approuve

- Le contribuable n'accepte pas ce que stipule la constitution selon laquelle l'impôt à payer dépend de la capacité contributive du citoyen, dénommée CONSTIT. 1 si il n'accepte pas, 0 si il accepte
- Informatisation de la fiscalité défavorable au contribuable, variable INFO qui est binaire aussi. 1 si informatisation désapprouvée par le contribuable, 0 dans le cas contraire.
- Textes en grand nombre confus pour le contribuable, variable NBTXT. 1 si textes en grand nombre ne sont pas appréciés par le contribuable, 0 dans le cas opposé.
- Facilité d'accès des textes, variable FACTXT. 1 si le contribuable estime que les textes sont difficiles d'accès, 0 dans le cas inverse.
- Légalité des pratiques administratives, les pratiques administratives des agents des impôts sont-elles légales vis-à-vis du contribuable ?, variable PRATLOI : 1 si le contribuable estime que les pratiques sont illégales, 0 dans le cas contraire.
- Durée de traitement des dossiers au niveau des bureaux, variable DURTRAIT : 1 si le contribuable n'apprécie pas la durée de traitement de dossier, 0 dans le cas inverse.
- Amende et pénalisation acceptées par le contribuable, variable AMPE : 1 si le contribuable trouve que les amendes et pénalisations sont trop lourdes, 0 sinon.
- Taux, calcul de l'impôt non acceptable pour le contribuable, variable TCALC : 1 si taux et calcul inacceptable pour l'enquête, 0 sinon.
- Obligations fiscales non tolérées par le contribuable, variable OBLIG : 1 si le contribuable ne tolère pas, 0 sinon.
- Les barèmes et les impôts minimums sont élevés pour le contribuable, variable BARIMIN : 1 si le contribuable les juge hauts, 0 sinon.
- Les contrôles fiscaux ne sont pas supportés par le contribuable, variable CTRLFISC : 1 si le contribuable ne supporte pas les contrôles fiscaux, 0 sinon.
- Utilisation des recettes fiscales pour les dépenses publiques, non approuvée par le contribuable, variable UTILITEIMPOT : 1 si l'utilisation des recettes fiscales n'est pas juste pour le contribuable, 0 sinon.
- Nombre des bureaux fiscaux, variable NBBUREAU : 1 si le nombre des bureaux fiscaux n'est pas suffisant pour le contribuable, 0 dans le cas contraire.
- Nombre des agents des impôts, variable AGENT : 1 si le nombre des agents fiscaux n'est pas satisfaisant pour le contribuable, 0 sinon.
- Recrudescence du secteur informel décourageante, variable SECTINF : 1 si le contribuable désespère face au secteur informel, 0 sinon.
- Contribuable victime d'un va-et-vient, variable VA et VI : 1 si l'enquête est déjà victime, 0 sinon.

Dans cette étude alors, nous allons chercher à évaluer si les variables explicatives sont significativement liées à la paresse des contribuables à payer leurs impôts.

### 2.3. La régression logistique binaire

Rappelons que notre variable expliquée est dichotomique c'est-à-dire à deux modalités. Ensuite, les données manipulées sont nombreuses. Le choix de la régression logistique se justifie en fait sur ces deux raisons.

Mathématiquement, l'équation de notre modèle est :

$$PARESSE = P(X) = \frac{e^{(\sum \beta X)}}{1 + e^{\sum \beta X}}$$

Où les coefficients  $\beta$  seront à déterminer par les modélisations avec le logiciel R et X les variables explicatives.

## 2.4. Analyse multi variée et analyse uni variée

Le choix de notre analyse est fondamental. L'analyse uni variée n'est pas indiquée car elle ne tient pas compte des relations entre les variables explicatives. On opte plutôt sur l'analyse multi variée car les données de sortie (paramètres) en analyse multi variée font l'objet d'un calcul qui prend en compte les relations entre les variables. Le logiciel R est complet car il intègre dans ses bibliothèques des commandes qui font ressortir les données de sortie en analyse multi variée. Il faut signaler également que des commandes du logiciel R permettent de calculer les paramètres (données de sortie) en analyse uni variée. Pour modéliser, il est nécessaire de préciser que nous ordonnons le logiciel R pour faire sortir les paramètres (ou données de sortie), comme les coefficients  $\beta$ , les OR ou Odds-Ratios, les p-values, les intervalles de confiance,...

## 2.5. Odds-Ratios (OR) ET P-Value

Les OR sont calculés par  $OR = e^{\beta}$  où  $\beta$  sont les coefficients calculés de notre modèle.

Dans un test statistique, la **valeur-p** (en anglais *p-value* pour *probability value*), parfois aussi appelée **p-valeur**, est la probabilité pour un modèle statistique donné sous l'hypothèse nulle d'obtenir la même valeur ou une valeur encore plus extrême que celle observée.

La p-value est utilisée pour quantifier la significativité statistique d'un résultat dans le cadre d'une hypothèse nulle. L'idée générale est de prouver que l'hypothèse nulle n'est pas vérifiée car dans le cas où elle le serait le résultat observé serait fortement improbable. Il s'agit à ce titre d'une extension du principe de preuve par l'absurde.

Un résultat statistiquement significatif est un résultat qui serait improbable si l'hypothèse nulle (qui représente en général la norme) était vérifiée. Il en résulte donc que l'hypothèse nulle ne s'applique pas au résultat observé et donc que le cas étudié diffère de manière notable de la norme et ainsi présente un intérêt particulier.

Le statisticien RONALD FISHER a introduit les termes de significativité, d'hypothèse nulle, et l'utilisation de la valeur-p. Selon lui, l'hypothèse nulle ne peut jamais être acceptée, mais peut seulement être rejetée par le test statistique. Dans cette approche, la valeur-p est considérée comme une mesure d'à quel point les données plaident contre l'hypothèse nulle. Les seuils suivants sont généralement pris pour référence :

- $p \leq 0,01$  : très forte présomption contre l'hypothèse nulle
- $0,01 < p \leq 0,05$  : forte présomption contre l'hypothèse nulle
- $0,05 < p \leq 0,1$  : faible présomption contre l'hypothèse nulle
- $p > 0,1$  : pas de présomption contre l'hypothèse nulle

## 2.6. Constitution de la base de données

Le nombre de contribuables miniers enquêtés est de 301 individus. Il est difficile d'estimer le nombre exact des personnes exerçant dans les mines dans la région de Vakinankaratra, vu que bon nombre de personnes se trouvent illicitement dans le secteur informel. Toutefois, on a recherché le maximum de personnes enquêtées pour retrouver le maximum d'avis.

Les données inscrites manuellement sur les fiches d'enquêtes devraient être transcrites en un tableau de données traitables par le logiciel R. Le tableau constitue notre base de données. Un fichier sur Excel en format csv a été édité et chaque ligne du tableau représente l'enregistrement d'une personne enquêtée.

## 2.7. MODELISATION SUR LE LOGICIEL R

Nous allons nous intéresser à la modélisation sur le logiciel R. Nous allons commencer par le chargement de la base de données qui a été saisie sur Microsoft Excel et enregistrée sous format csv dans le fichier bdd.csv. La commande correspondante est la suivante :

```
>bd <- read.csv2 ("bdd.csv")
```

Vérifions si la base de données est vraiment chargée :

```
>head(bd)
```

Résultat :

|   | id | PARESSE | RDEP | CONSTIT | INFO | NBTXT | FACTXT | PRATLOI | DURTRAIT | AMPE | TCALC | OBLIG |
|---|----|---------|------|---------|------|-------|--------|---------|----------|------|-------|-------|
| 1 | 1  | 1       | 1    | 1       | 0    | 0     | 1      | 1       | 1        | 1    | 1     | 1     |
| 2 | 2  | 1       | 0    | 0       | 1    | 0     | 1      | 0       | 0        | 1    | 1     | 1     |
| 3 | 3  | 1       | 0    | 0       | 1    | 1     | 0      | 0       | 0        | 0    | 1     | 0     |
| 4 | 4  | 1       | 1    | 1       | 1    | 1     | 1      | 0       | 1        | 1    | 1     | 1     |
| 5 | 5  | 1       | 1    | 0       | 1    | 0     | 0      | 1       | 0        | 0    | 0     | 0     |
| 6 | 6  | 1       | 1    | 0       | 1    | 0     | 1      | 1       | 1        | 1    | 0     | 1     |

|   | BARIMIN | CTRLFISC | UTILITEIMPOT | NBBUREAU | AGENT | SECTINF | VAetVI |
|---|---------|----------|--------------|----------|-------|---------|--------|
| 1 | 1       | 1        | 1            | 1        | 0     | 1       | 0      |
| 2 | 0       | 0        | 1            | 1        | 0     | 1       | 0      |
| 3 | 1       | 0        | 1            | 1        | 1     | 1       | 0      |
| 4 | 1       | 1        | 1            | 1        | 1     | 1       | 1      |
| 5 | 0       | 0        | 1            | 0        | 1     | 1       | 0      |
| 6 | 1       | 1        | 1            | 0        | 1     | 1       | 1      |

Le résultat de la commande sur R nous montre les six premières lignes de la base chargée.

On va faire maintenant le résumé de la base bd :

```
>summary(bd)
```

Résultat :

| id          | PARESSE       | RDEP           | CONSTIT       |
|-------------|---------------|----------------|---------------|
| Min. : 1    | Min. :0.000   | Min. :0.0000   | Min. :0.000   |
| 1st Qu.: 76 | 1st Qu.:0.000 | 1st Qu.:0.0000 | 1st Qu.:0.000 |
| Median :151 | Median :1.000 | Median :1.0000 | Median :0.000 |
| Mean :151   | Mean :0.691   | Mean :0.6346   | Mean :0.495   |
| 3rd Qu.:226 | 3rd Qu.:1.000 | 3rd Qu.:1.0000 | 3rd Qu.:1.000 |
| Max. :301   | Max. :1.000   | Max. :1.0000   | Max. :1.000   |

| INFO           | NBTXT          | FACTXT         | PRATLOI        |
|----------------|----------------|----------------|----------------|
| Min. :0.0000   | Min. :0.0000   | Min. :0.0000   | Min. :0.0000   |
| 1st Qu.:0.0000 | 1st Qu.:0.0000 | 1st Qu.:0.0000 | 1st Qu.:0.0000 |
| Median :1.0000 | Median :0.0000 | Median :0.0000 | Median :0.0000 |
| Mean :0.5748   | Mean :0.4751   | Mean :0.4983   | Mean :0.4884   |
| 3rd Qu.:1.0000 | 3rd Qu.:1.0000 | 3rd Qu.:1.0000 | 3rd Qu.:1.0000 |
| Max. :1.0000   | Max. :1.0000   | Max. :1.0000   | Max. :1.0000   |

| DURTRAIT       | AMPE           | TCALC          | OBLIG         |
|----------------|----------------|----------------|---------------|
| Min. :0.0000   | Min. :0.0000   | Min. :0.0000   | Min. :0.000   |
| 1st Qu.:0.0000 | 1st Qu.:0.0000 | 1st Qu.:0.0000 | 1st Qu.:0.000 |
| Median :0.0000 | Median :0.0000 | Median :1.0000 | Median :0.000 |
| Mean :0.4618   | Mean :0.4983   | Mean :0.6246   | Mean :0.485   |
| 3rd Qu.:1.0000 | 3rd Qu.:1.0000 | 3rd Qu.:1.0000 | 3rd Qu.:1.000 |
| Max. :1.0000   | Max. :1.0000   | Max. :1.0000   | Max. :1.000   |

| BARIMIN        | CTRLFISC       | UTILITEIMPOT   | NBBUREAU       |
|----------------|----------------|----------------|----------------|
| Min. :0.0000   | Min. :0.0000   | Min. :0.0000   | Min. :0.0000   |
| 1st Qu.:0.0000 | 1st Qu.:0.0000 | 1st Qu.:0.0000 | 1st Qu.:0.0000 |
| Median :0.0000 | Median :0.0000 | Median :1.0000 | Median :1.0000 |
| Mean :0.4983   | Mean :0.4917   | Mean :0.6379   | Mean :0.5183   |
| 3rd Qu.:1.0000 | 3rd Qu.:1.0000 | 3rd Qu.:1.0000 | 3rd Qu.:1.0000 |
| Max. :1.0000   | Max. :1.0000   | Max. :1.0000   | Max. :1.0000   |

| AGENT          | SECTINF        | VAetVI         |
|----------------|----------------|----------------|
| Min. :0.0000   | Min. :0.0000   | Min. :0.0000   |
| 1st Qu.:0.0000 | 1st Qu.:0.0000 | 1st Qu.:0.0000 |
| Median :0.0000 | Median :1.0000 | Median :1.0000 |
| Mean :0.4585   | Mean :0.6478   | Mean :0.5116   |
| 3rd Qu.:1.0000 | 3rd Qu.:1.0000 | 3rd Qu.:1.0000 |
| Max. :1.0000   | Max. :1.0000   | Max. :1.0000   |

On constate que les données dichotomiques ne sont pas traitées comme des variables qualitatives, pour les changer et afin que le logiciel les reconnaisse comme données qualitatives, on saisit les commandes suivantes :

```
> for (i in 2:19) {
+ bd[,i] <- factor(bd[,i])}
```

Et on revérifie avec la commande summary :

```
>summary(bd)
```

Résultat :

|          | id   | PARESE | RDEP  | CONSTIT | INFO  | NBTXT | FACTXT | PRATLOI | DURTRAIT | AMPE  | TCALC | OBLIG | BARIMIN | CTRLFISC | UTILITEIMPOT | NBBUREAU | AGENT | SECTINF | VAetVI |
|----------|------|--------|-------|---------|-------|-------|--------|---------|----------|-------|-------|-------|---------|----------|--------------|----------|-------|---------|--------|
| Min.     | : 1  | 0: 93  | 0:110 | 0:152   | 0:128 | 0:158 | 0:151  | 0:154   | 0:162    | 0:151 | 0:113 | 0:155 | 0:151   | 0:153    | 0:109        | 0:145    | 0:163 | 0:106   | 0:147  |
| 1st Qu.: | 76   | 1:208  | 1:191 | 1:149   | 1:173 | 1:143 | 1:150  | 1:147   | 1:139    | 1:150 | 1:188 | 1:146 | 1:150   | 1:148    | 1:192        | 1:156    | 1:138 | 1:195   | 1:154  |
| Median   | :151 |        |       |         |       |       |        |         |          |       |       |       |         |          |              |          |       |         |        |
| Mean     | :151 |        |       |         |       |       |        |         |          |       |       |       |         |          |              |          |       |         |        |
| 3rd Qu.: | 226  |        |       |         |       |       |        |         |          |       |       |       |         |          |              |          |       |         |        |
| Max.     | :301 |        |       |         |       |       |        |         |          |       |       |       |         |          |              |          |       |         |        |

Il faut vérifier maintenant les conditions pour une analyse multivariée et une régression logistique.

*Harrell et al.* recommandent d'utiliser *10 fois plus d'événements que de variables* étudiées pour toute analyse multivariée. Pour notre cas, il y a 18 variables ;  $10 \times 18 = 180$ .

Le nombre d'évènement 1 de PARESE doit être donc supérieur à 180.

On entre sur R la commande suivante pour vérifier :

```
>table(bd$PARESE)
```

Résultat :

```
0 1
93 208
```

Il y a 208 évènements de 1 pour la variable PARESE, donc la condition est satisfaite.

Par ailleurs, il faut aussi justifier l'absence de surdispersion.

La surdispersion est généralement évaluée par le **ratio de la déviance résiduelle sur le nombre de degrés de libertés** du modèle.

$$\phi^{\wedge} = \text{devianceresiduelle} / \text{nddl}$$

Si ce ratio est **supérieur à 1**, alors il y a **surdispersion**.

Pour cela, nous appliquons tout de suite la fonction régression logistique sur bd :

```
> reglog <- glm(PARESE ~
RDEP+CONSTIT+INFO+NBTXT+FACTXT+PRATLOI+DURTRAIT+AMPE+TCALC+OBLIG+BARIMIN+CTRLFISC+UTILITEIMPOT+
NBBUREAU+AGENT+SECTINF+VAetVI, family = binomial, data = bd)
```

Et ensuite voir les résultats de la commande summary :

```
> summary(reglog)
```

Résultat :

```
Call:
glm(formula = PARESSE ~ RDEP + CONSTIT + INFO + NBTXT + FACTXT +
     PRATLOI + DURTRAIT + AMPE + TCALC + OBLIG + BARIMIN + CTRLFISC +
     UTILITEIMPOT + NBBUREAU + AGENT + SECTINF + VAetVI, family = binomial,
     data = bd)
```

Deviance Residuals:

| Min     | 1Q      | Median | 3Q     | Max    |
|---------|---------|--------|--------|--------|
| -2.8166 | -0.7581 | 0.4160 | 0.7202 | 1.8358 |

Coefficients:

|               | Estimate | Std. Error | z value | Pr(> z ) |     |
|---------------|----------|------------|---------|----------|-----|
| (Intercept)   | -2.07932 | 0.69615    | -2.987  | 0.00282  | **  |
| RDEP1         | 0.22544  | 0.31104    | 0.725   | 0.46857  |     |
| CONSTIT1      | -0.65418 | 0.31597    | -2.070  | 0.03841  | *   |
| INFO1         | 1.41381  | 0.31696    | 4.461   | 8.17e-06 | *** |
| NBTXT1        | -0.09575 | 0.30332    | -0.316  | 0.75224  |     |
| FACTXT1       | 0.16329  | 0.30214    | 0.540   | 0.58888  |     |
| PRATLOI1      | -0.13010 | 0.30630    | -0.425  | 0.67103  |     |
| DURTRAIT1     | 0.90031  | 0.31405    | 2.867   | 0.00415  | **  |
| AMPE1         | -0.61380 | 0.30500    | -2.012  | 0.04417  | *   |
| TCALC1        | -0.07448 | 0.31928    | -0.233  | 0.81555  |     |
| OBLIG1        | 1.03259  | 0.31618    | 3.266   | 0.00109  | **  |
| BARIMIN1      | -0.12660 | 0.30444    | -0.416  | 0.67753  |     |
| CTRLFISC1     | 0.60169  | 0.30824    | 1.952   | 0.05094  | .   |
| UTILITEIMPOT1 | 1.25994  | 0.31488    | 4.001   | 6.30e-05 | *** |
| NBBUREAU1     | -0.22043 | 0.30261    | -0.728  | 0.46635  |     |
| AGENT1        | 0.48353  | 0.30356    | 1.593   | 0.11119  |     |
| SECTINF1      | 1.32494  | 0.30607    | 4.329   | 1.50e-05 | *** |
| VAetVI1       | 0.07688  | 0.30271    | 0.254   | 0.79951  |     |

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 372.20 on 300 degrees of freedom  
 Residual deviance: 282.61 on 283 degrees of freedom  
 AIC: 318.61

Number of Fisher Scoring iterations: 5

Ici,  $\hat{\phi} = 282.61/283 = 0.9986219 < 1$ ,

Par conséquent, il n'y a pas de surdispersion.

Revenons maintenant aux résultats de la régression logistique auparavant, les coefficients ont été calculés, ce qui nous donne comme modèle :

$$PARESSE = P(X) = \frac{e^{\sum \beta X}}{1 + e^{\sum \beta X}}$$

Avec  $\sum \beta X = -2,08 + 0,225RDEP - 0,654CONSTIT + 1,41INFO - 0,095NBTXT + 0,163FACTXT - 0,13PRATLOI + 0,9DURTRAIT - 0,613AMPE - 0,074TCALC + 1,032OBLIG - 0,126BARIMIN + 0,601CTRLFISC + 1,26UTILITEIMPOT - 0,22NBBUREAU + 0,483AGENT + 1,325SECTINF + 0,076VAetVI$

Or, il faut rechercher un modèle plus simple pour expliquer la variable PARESSE. Autrement dit, il faut rechercher parmi les variables explicatives celles qui expliquent vraiment la variable PARESSE. Ce qui revient à rechercher les priorités des problèmes qui découragent les contribuables à payer les impôts. Ainsi, il faut éliminer certaines variables explicatives dont l'influence n'est pas importante pour la variable expliquée PARESSE.

Pour cela, nous éliminerons manuellement les variables explicatives en tenant compte de leurs p-values.

Ainsi, on exécute les commandes suivantes sur R pour voir les p-values du modèle en analyse multi variée et uni variée :

```
> library(finalfit)
> dep <- "PARESSE"
> exp <-
c("RDEP", "CONSTIT", "INFO", "NBTXT", "FACTXT", "PRATLOI", "DURTRAIT", "AMPE",
"TCALC", "OBLIG", "BARIMIN", "CTRLFISC", "UTILITEIMPOT", "NBBUREAU", "AGENT",
"SECTINF", "VAetVI")
> res <- bd %>%
+ finalfit(dep, exp)
> dep
```

Les résultats des commandes sont les suivantes :

| Dependent: PARESSE | 0           | 1          | OR (univariable)          | OR (multivariable)        |
|--------------------|-------------|------------|---------------------------|---------------------------|
| RDEP               | 0 36 (32.7) | 74 (67.3)  | -                         | -                         |
|                    | 1 57 (29.8) | 134 (70.2) | 1.14 (0.69-1.89, p=0.602) | 1.25 (0.68-2.31, p=0.469) |
| CONSTIT            | 0 43 (28.3) | 109 (71.7) | -                         | -                         |
|                    | 1 50 (33.6) | 99 (66.4)  | 0.78 (0.48-1.27, p=0.323) | 0.52 (0.28-0.96, p=0.038) |
| INFO               | 0 56 (43.8) | 72 (56.2)  | -                         | -                         |
|                    | 1 37 (21.4) | 136 (78.6) | 2.86 (1.73-4.76, p<0.001) | 4.11 (2.24-7.79, p<0.001) |
| NBTXT              | 0 46 (29.1) | 112 (70.9) | -                         | -                         |
|                    | 1 47 (32.9) | 96 (67.1)  | 0.84 (0.51-1.37, p=0.482) | 0.91 (0.50-1.65, p=0.752) |
| FACTXT             | 0 48 (31.8) | 103 (68.2) | -                         | -                         |
|                    | 1 45 (30.0) | 105 (70.0) | 1.09 (0.67-1.78, p=0.737) | 1.18 (0.65-2.14, p=0.589) |
| PRATLOI            | 0 48 (31.2) | 106 (68.8) | -                         | -                         |
|                    | 1 45 (30.6) | 102 (69.4) | 1.03 (0.63-1.68, p=0.917) | 0.88 (0.48-1.60, p=0.671) |
| DURTRAIT           | 0 57 (35.2) | 105 (64.8) | -                         | -                         |
|                    | 1 36 (25.9) | 103 (74.1) | 1.55 (0.95-2.57, p=0.083) | 2.46 (1.34-4.62, p=0.004) |
| AMPE               | 0 42 (27.8) | 109 (72.2) | -                         | -                         |
|                    | 1 51 (34.0) | 99 (66.0)  | 0.75 (0.46-1.22, p=0.246) | 0.54 (0.29-0.98, p=0.044) |
| TCALC              | 0 40 (35.4) | 73 (64.6)  | -                         | -                         |
|                    | 1 53 (28.2) | 135 (71.8) | 1.40 (0.84-2.30, p=0.191) | 0.93 (0.49-1.73, p=0.816) |
| OBLIG              | 0 61 (39.4) | 94 (60.6)  | -                         | -                         |
|                    | 1 32 (21.9) | 114 (78.1) | 2.31 (1.40-3.87, p=0.001) | 2.81 (1.53-5.30, p=0.001) |
| BARIMIN            | 0 48 (31.8) | 103 (68.2) | -                         | -                         |
|                    | 1 45 (30.0) | 105 (70.0) | 1.09 (0.67-1.78, p=0.737) | 0.88 (0.48-1.60, p=0.678) |
| CTRLFISC           | 0 53 (34.6) | 100 (65.4) | -                         | -                         |
|                    | 1 40 (27.0) | 108 (73.0) | 1.43 (0.88-2.35, p=0.154) | 1.83 (1.00-3.38, p=0.051) |
| UTILITEIMPOT       | 0 51 (46.8) | 58 (53.2)  | -                         | -                         |
|                    | 1 42 (21.9) | 150 (78.1) | 3.14 (1.89-5.25, p<0.001) | 3.53 (1.92-6.62, p<0.001) |
| NBBUREAU           | 0 40 (27.6) | 105 (72.4) | -                         | -                         |
|                    | 1 53 (34.0) | 103 (66.0) | 0.74 (0.45-1.21, p=0.231) | 0.80 (0.44-1.45, p=0.466) |
| AGENT              | 0 55 (33.7) | 108 (66.3) | -                         | -                         |
|                    | 1 38 (27.5) | 100 (72.5) | 1.34 (0.82-2.21, p=0.246) | 1.62 (0.90-2.97, p=0.111) |
| SECTINF            | 0 54 (50.9) | 52 (49.1)  | -                         | -                         |
|                    | 1 39 (20.0) | 156 (80.0) | 4.15 (2.49-7.02, p<0.001) | 3.76 (2.08-6.93, p<0.001) |
| VAetVI             | 0 43 (29.3) | 104 (70.7) | -                         | -                         |
|                    | 1 50 (32.5) | 104 (67.5) | 0.86 (0.53-1.40, p=0.546) | 1.08 (0.60-1.96, p=0.800) |

Les données en analyse uni variée ne tiennent pas compte des influences entre les variables, celles de l'analyse multi variée sont pourtant importantes pour nous pour éliminer les variables à enlever.

Les variables dont les p-values multi variés sont supérieurs à 0,2 qui sont encadrés seront à éliminer :

| Dependent: PARESSE | 0           | 1          | OR (univariable)          | OR (multivariable)        |
|--------------------|-------------|------------|---------------------------|---------------------------|
| RDEP               | 0 36 (32.7) | 74 (67.3)  | -                         | -                         |
|                    | 1 57 (29.8) | 134 (70.2) | 1.14 (0.69-1.89, p=0.602) | 1.25 (0.68-2.31, p=0.469) |
| CONSTIT            | 0 43 (28.3) | 109 (71.7) | -                         | -                         |
|                    | 1 50 (33.6) | 99 (66.4)  | 0.78 (0.48-1.27, p=0.323) | 0.52 (0.28-0.96, p=0.038) |
| INFO               | 0 56 (43.8) | 72 (56.2)  | -                         | -                         |
|                    | 1 37 (21.4) | 136 (78.6) | 2.86 (1.73-4.76, p<0.001) | 4.11 (2.24-7.79, p<0.001) |
| NBTXT              | 0 46 (29.1) | 112 (70.9) | -                         | -                         |
|                    | 1 47 (32.9) | 96 (67.1)  | 0.84 (0.51-1.37, p=0.482) | 0.91 (0.50-1.65, p=0.752) |
| FACTXT             | 0 48 (31.8) | 103 (68.2) | -                         | -                         |
|                    | 1 45 (30.0) | 105 (70.0) | 1.09 (0.67-1.78, p=0.737) | 1.18 (0.65-2.14, p=0.589) |
| PRATLOI            | 0 48 (31.2) | 106 (68.8) | -                         | -                         |
|                    | 1 45 (30.6) | 102 (69.4) | 1.03 (0.63-1.68, p=0.917) | 0.88 (0.48-1.60, p=0.671) |
| DURTRAIT           | 0 57 (35.2) | 105 (64.8) | -                         | -                         |
|                    | 1 36 (25.9) | 103 (74.1) | 1.55 (0.95-2.57, p=0.083) | 2.46 (1.34-4.62, p=0.004) |
| AMPE               | 0 42 (27.8) | 109 (72.2) | -                         | -                         |
|                    | 1 51 (34.0) | 99 (66.0)  | 0.75 (0.46-1.22, p=0.246) | 0.54 (0.29-0.98, p=0.044) |
| TCALC              | 0 40 (35.4) | 73 (64.6)  | -                         | -                         |
|                    | 1 53 (28.2) | 135 (71.8) | 1.40 (0.84-2.30, p=0.191) | 0.93 (0.49-1.73, p=0.816) |
| OBLIG              | 0 61 (39.4) | 94 (60.6)  | -                         | -                         |
|                    | 1 32 (21.9) | 114 (78.1) | 2.31 (1.40-3.87, p=0.001) | 2.81 (1.53-5.30, p=0.001) |
| BARIMIN            | 0 48 (31.8) | 103 (68.2) | -                         | -                         |
|                    | 1 45 (30.0) | 105 (70.0) | 1.09 (0.67-1.78, p=0.737) | 0.88 (0.48-1.60, p=0.678) |
| CTRLFISC           | 0 53 (34.6) | 100 (65.4) | -                         | -                         |
|                    | 1 40 (27.0) | 108 (73.0) | 1.43 (0.88-2.35, p=0.154) | 1.83 (1.00-3.38, p=0.051) |
| UTILITEIMPOT       | 0 51 (46.8) | 58 (53.2)  | -                         | -                         |
|                    | 1 42 (21.9) | 150 (78.1) | 3.14 (1.89-5.25, p<0.001) | 3.53 (1.92-6.62, p<0.001) |
| NBBUREAU           | 0 40 (27.6) | 105 (72.4) | -                         | -                         |
|                    | 1 53 (34.0) | 103 (66.0) | 0.74 (0.45-1.21, p=0.231) | 0.80 (0.44-1.45, p=0.466) |
| AGENT              | 0 55 (33.7) | 108 (66.3) | -                         | -                         |
|                    | 1 38 (27.5) | 100 (72.5) | 1.34 (0.82-2.21, p=0.246) | 1.62 (0.90-2.97, p=0.111) |
| SECTINF            | 0 54 (50.9) | 52 (49.1)  | -                         | -                         |
|                    | 1 39 (20.0) | 156 (80.0) | 4.15 (2.49-7.02, p<0.001) | 3.76 (2.08-6.93, p<0.001) |
| VAetVI             | 0 43 (29.3) | 104 (70.7) | -                         | -                         |
|                    | 1 50 (32.5) | 104 (67.5) | 0.86 (0.53-1.40, p=0.546) | 1.08 (0.60-1.96, p=0.800) |

Par conséquent, on garde les variables suivantes :

CONSTIT, INFO, DURTRAIT, AMPE, OBLIG, CTRLFISC, UTILITEIMPOT, AGENT et SECTINF.

Ensuite, on construit la nouvelle régression logistique avec les nouvelles variables retenues :

```
> reglog2 <- glm(PARESSE ~
```

```
CONSTIT+INFO+DURTRAIT+AMPE+OBLIG+CTRLFISC+UTILITEIMPOT+AGENT+S  
ECTINF, family = binomial, data = bd)
```

Retrouvons les coefficients du nouveau modèle :

```
> summary(reglog2)
```

Résultat :

```

Call:
glm(formula = PARESSE ~ CONSTIT + INFO + DURTRAIT + AMPE + OBLIG +
     CTRLFISC + UTILITEIMPOT + AGENT + SECTINF, family = binomial,
     data = bd)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8889  -0.7462   0.4386   0.7208   1.8013

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -2.0902     0.4870  -4.292 1.77e-05 ***
CONSTIT1      -0.6216     0.3042  -2.043  0.04101 *
INFO1         1.3887     0.3095   4.487 7.21e-06 ***
DURTRAIT1     0.8751     0.3083   2.839  0.00453 **
AMPE1        -0.6259     0.2989  -2.094  0.03624 *
OBLIG1        0.9895     0.3088   3.204  0.00136 **
CTRLFISC1     0.6075     0.3013   2.016  0.04376 *
UTILITEIMPOT1 1.2120     0.3039   3.989 6.64e-05 ***
AGENT1        0.4729     0.2993   1.580  0.11412
SECTINF1      1.3094     0.2992   4.376 1.21e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 372.20  on 300  degrees of freedom
Residual deviance: 284.63  on 291  degrees of freedom
AIC: 304.63

Number of Fisher Scoring iterations: 5

```

En final, l'équation de notre modèle est :

$$PARESSE = P(X) = \frac{e^{(\sum \beta X)}}{1 + e^{\sum \beta X}}$$

Avec  $\sum \beta X = -2,1 - 0,621 \text{CONSTIT} + 1,388 \text{INFO} + 0,875 \text{DURTRAIT} - 0,625 \text{AMPE} + 0,989 \text{OBLIG} + 0,607 \text{CTRLFISC} + 1,21 \text{UTILITEIMPOT} + 0,473 \text{AGENT} + 1,31 \text{SECTINF}$

### 3. RESULTATS, DISCUSSIONS

La fuite des contribuables devant l'impôt favorise l'évasion fiscale. Rappelons qu'il existe des problèmes et des causes qui entraînent cette négligence ou cette paresse ou cette fuite devant l'impôt. En étudiant une population dont les individus enquêtés compte 301 personnes, nous avons pu aborder plusieurs causes expliquant cette fainéantise des contribuables lors des modélisations mathématiques, et l'équation suivante résume la relation liant la paresse, et ses causes multiples :

$$PARESSE = P(X) = \frac{e^{(\sum \beta X)}}{1 + e^{\sum \beta X}}$$

Avec  $\sum \beta X = -2,08 + 0,225 \text{RDEP} - 0,654 \text{CONSTIT} + 1,41 \text{INFO} - 0,095 \text{NBTXT} + 0,163 \text{FACTXT} - 0,13 \text{PRATLOI} + 0,9 \text{DURTRAIT} - 0,613 \text{AMPE} - 0,074 \text{TCALC} + 1,032 \text{OBLIG} - 0,126 \text{BARIMIN} + 0,601 \text{CTRLFISC} + 1,26 \text{UTILITEIMPOT} - 0,22 \text{NBBUREAU} + 0,483 \text{AGENT} + 1,325 \text{SECTINF} + 0,076 \text{VAetVI}$

Ainsi, il existe plusieurs causes et problèmes favorisant la paresse. Pour trouver les problèmes prioritaires, il faut trouver un autre modèle plus parcimonieux ayant un nombre plus limité en variables explicatives. Après l'application de quelques conditions, nous avons trouvé l'équation suivante :

$$PARESSE = P(X) = \frac{e^{(\sum \beta X)}}{1 + e^{\sum \beta X}}$$

Avec  $\sum \beta X = -2,1 - 0,621 \text{CONSTIT} + 1,388 \text{INFO} + 0,875 \text{DURTRAIT} - 0,625 \text{AMPE} + 0,989 \text{OBLIG} + 0,607 \text{CTRLFISC} + 1,21 \text{UTILITEIMPOT} + 0,473 \text{AGENT} + 1,31 \text{SECTINF}$

Les variables retenues qui constituent nos problèmes urgents à résoudre sont alors :

- CONSTIT : certains contribuables miniers de la région Vakinankaratra n'acceptent pas ce que stipule la constitution selon laquelle l'impôt à payer dépend de la capacité contributive du citoyen. Ce comportement d'inacceptation explique la paresse des contribuables à se rapprocher du bureau du Fisc.
- INFO : l'informatisation qui aide beaucoup à alléger les tâches de l'Administration fiscale d'une part et à simplifier partiellement les obligations des contribuables présente quand même des inconvénients poussant certains contribuables à ne pas effectuer leurs obligations.
- DURTRAIT : la durée de traitement actuel des dossiers ne convient pas à des contribuables qui exigent de plus en plus de rapidité dans le traitement de leurs dossiers.
- AMPE : le taux et le calcul des pénalités et amendes sont trop lourds pour certains individus qui leur obligent à fuir et à laisser tomber leurs déclarations fiscales.
- OBLIG : les obligations fiscales sont trop nombreuses, parfois compliquées et parfois difficile à suivre. Ainsi, des contribuables abandonnent et reculent devant des textes trop complexes.
- CTRLFISC : le contrôle fiscal, qui a été créé pour récupérer les argents de l'Etat mais dissimulés par certains contribuables malintentionnés, connaît lui aussi des inconvénients qui démotivent les gens à payer leurs impôts.
- UTILITEIMPOT : certains contribuables sont découragés en observant les façons dont l'Etat utilise les recettes en provenance des impôts. Ces gens ne constatent pas réellement les investissements et les dépenses d'utilisation des recettes publiques.
- AGENT : les contribuables requièrent de plus en plus des accueils plus professionnels et plus chaleureux lorsqu'ils visitent les bureaux fiscaux, raison pour laquelle ils n'apprécient guère le nombre faible des agents des impôts.
- SECTINF : vu le nombre abondant des individus dans le secteur informel, des contribuables se découragent et deviennent de plus en plus démotivés à se conformer à la loi.

Les variables explicatives restant dans le modèle plus parcimonieux constituent les priorités. Il est nécessaire de remédier à ces problèmes urgents pour augmenter les recettes en provenance du secteur mine dans la région étudiée.

Les objectifs définis ont été atteints en ce qui concerne nos études car nous avons pu expliquer la paresse des contribuables à payer leurs impôts à partir de plusieurs variables explicatives. Et finalement, les modélisations ont conduit à un modèle parcimonieux dont les variables explicatives sont limitées en nombre.

Par ailleurs, même si nous avons cherché à trouver les maximums de facteurs expliquant la paresse des contribuables miniers, les limites de nos études qui ont été réalisées résident dans le nombre encore insuffisant des personnes enquêtées. Cette insuffisance aura des conséquences sur le nombre de variables explicatives trouvées aussi. En effet, il pourrait encore y avoir des facteurs non mentionnés dans nos études qui influencent indirectement les comportements des contribuables. Toutefois, nous estimons que, vu la taille des échantillons

dans nos études qui représentent la presque totalité des contribuables répertoriés dans le Fisc au niveau de la région Vakinankaratra, nos études présentent quand même des intérêts.

## **CONCLUSION**

Pour conclure, il est rappelé que les études menées dans le cadre de la présente publication considèrent la fuite ou la paresse des contribuables miniers à payer leurs impôts. Des analyses sur ce comportement de paresse ont été effectuées afin de pouvoir trouver les facteurs qui l'expliquent. Les modélisations successives ont permis de trouver les priorités des facteurs explicatifs. Une fois trouvés, ces facteurs constituent les problèmes urgents à résoudre pour l'Etat.

Les recherches ont abouti ainsi à aider la prise de décision dans la politique de rehaussement de nombre des contribuables dans le secteur minier au niveau de la région Vakinankaratra. Et par conséquent, les investigations et les observations faites contribueront à l'augmentation des recettes fiscales en provenance du secteur mine, conduisant de ce fait, à la valorisation des ressources minières de la région Vakinankaratra.

## BIBLIOGRAPHIE ET WEBOGRAPHIE

- [1]-Andrew Kramar, Simone Mathoulin-Pélissier, et al. :2011. Méthodes biostatistiques appliquées à la recherche clinique en cancérologie
- [2]-Peduzzi, P. ; J. Concato; E. Kemper; TR Holford; AR Feinstein : 1996. *Une étude de simulation du nombre d'événements par variables dans l'analyse de régression logistique*
- [3]-Ministère des Finances et du Budget : 2017. Code général des Impôts suivant la loi de finances 2018
- [4]-Hosmer, David : 2013. *Régression logistique appliquée.*
- [5]-Le Comité Consultatif Constitutionnel, et al. : 2010. Constitution de la quatrième république de Madagascar
- [6]-Sébastien Déjean. 2004. Présentation du logiciel R
- [7]-Agresti, Alan :2002. Catégorielle d'analyse des données.
- [8]-Laurent Rouvière : 2004. Régression logistique avec R
- [9]-Balakrishnan, N., Marcel Dekker : 1991. *Manuel de la distribution logistique.*
- [10]-Cornillon P. & Matzner-Løber E. : 2007. Régression : Théorie et Application.
- [11]-Gouriéroux, Christian : 2000. « Le simple Dichotomie » . *Économétrie des variables dépendantes qualitatives.*
- [12]-Hilbe, Joseph M. : 2009. *Les modèles de régression logistique.*
- [13]-Droesbeke J., Lejeune M. & Saporta J. : 2007. Modèles Statistiques pour Données Qualitatives.
- [14]-Hosmer D. & Lemeshow S. :2000. Applied Logistic Regression.
- [14]-Amemiya, Takeshi :1985. *Modèles de réponse qualitative. Économétrie avancée .*
- [16]-Greene, William H. : 2003. *Analyse économétrie, cinquième édition*
- [17]-Howell, David C. : 2010. *Méthodes statistiques pour la psychologie, 7e éd . Belmont*
- [18]-Berry, Michael JA; Linoff, Gordon :1997. *Techniques d'exploration de données pour le marketing, les ventes et le soutien à la clientèle.*
- [19]-[https://fr.qwe.wiki/wiki/Logistic\\_regression](https://fr.qwe.wiki/wiki/Logistic_regression) : consulté le 19/02/2019
- [20]-[https://fr.wikipedia.org/wiki/R\\_\(langage\)](https://fr.wikipedia.org/wiki/R_(langage)): consulté le 19/02/2019