

REPUBLIQUE DU BURUNDI

**MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE**

UNIVERSITE DU LAC TANGANYIKA (ULT)

INSTITUT INTERFACULTAIRE DE STATISTIQUE

**« ANALYSE STATISTIQUE DES FACTEURS INFLUENÇANT LES
PATHOLOGIES MENTALES AU BURUNDI »**

**Rapport de stage effectué au Centre Neuro-Psychiatrique de Kamenge
du 04 Avril au 04 Mai 2022 avec des données de l'année 2020**

Par

ISHIMWE Lina Carolle

et

IRAKOZE Chany Bertilla

**Sous l'encadrement de
Mme Olyne NIYOMWUNGERE**

**Rapport de stage présenté et défendu
en vue de l'obtention du Diplôme de
Baccalauréat en Statistique**

Bujumbura, Décembre 2023

DEDICACES

À mon père ;

À ma mère ;

À mes sœurs.

À tous ceux qui me sont chers.

ISHIMWE Lina Carolle

À mon père ;

À ma mère ;

À mes grands frères et sœurs ;

Et à tous qui me portent au cœur.

IRAKOZE Chany Bertilla

REMERCIEMENTS

A la fin de ce travail, nous aimerions porté nos remerciements à toutes ces personnes qui, d'une manière générale ou particulière, ont contribué dans l'exécution des différentes tâches qui ont abouti à la réalisation de ce document.

En premier lieu, nous rendons grâce à Dieu pour sa bonté accordée jusqu'à cet instant.

En deuxième lieu, nos remerciements vont également à l'endroit de nos parents, qui sans leur soutien infaillible, nous ne serions en aucun cas parvenues à cette étape. Qu'ils trouvent ce document comme le couronnement de leurs efforts.

En troisième lieu, nos remerciements s'adressent plus particulièrement au Directeur Général du Centre Neuro-Psychiatrique de Kamenge, le Frère Herménégide NDUWIMANA pour avoir accepté notre présence au sein du service du Système d'Information Sanitaire de cette fondation. Nous remercions vivement aussi Monsieur Fulgence NDAYIKUNDA, le responsable du service du Système d'Information Sanitaire qui nous a accueillies et nous a guidées tout au long du stage.

En quatrième lieu, nous adressons nos sincères remerciements à Madame Olyne NIYOMWUNGERE qui, en tant qu'encadreuse, s'est toujours montrée à l'écoute et disponible tout au long de la réalisation de ce rapport.

En cinquième lieu, nous remercions l'Université du Lac Tanganyika qui a mis au point des programmes actualisés et aussi les professeurs de l'Institut Interfacultaire de Statistique pour les connaissances qu'ils nous ont inculquées.

En dernier lieu, nous remercions tous nos camarades, nos amis avec qui, nous avons cueilli les roses et les épines, nous ne dirons jamais assez merci.

LISTE DES ABBREVIATIONS

AIC	: Akaike Information Criterion
AUC	: Area Under the Curve
BIC	: Bayesian Information Criterion
C.N.P.K	: Centre Neuro-Psychiatrique de Kamenge
DHIS2	: District Health Information System
ETS	: Ecole Technique Secondaire de Kamenge
FIV	: Facteur d'Inflation de la Variance
HL	: Hosmer et Lemeshow
IRLS	: Iterative Reweighted Least Square
OMS	: Organisation Mondiale de la Santé
OR	: Odds Ratio
ROC	: Receiver Operating Characteristic
SCR	: Somme des Carrés des Résidus
SIS	: Système d'Information Sanitaire
TFP	: Taux de Faux Positifs
TVP	: Taux de vrais Positifs
ULT	: Université du Lac Tanganyika

TABLE DES MATIERES

DEDICACES.....	i
REMERCIEMENTS.....	ii
LISTE DES ABBREVIATIONS	iii
TABLE DES MATIERES.....	iv
LISTE DES TABLEAUX	vii
LISTE DES FIGURES.....	viii
CHAPITRE 0 : INTRODUCTION GENERALE.....	1
0.1. Contexte et Généralité sur l'épilepsie et les troubles mentaux	1
0.2. Problématique	2
0.3. Objectifs.....	2
0.3.1. Objectif principal	2
0.3.2. Objectif spécifique	3
0.4. Hypothèses de recherche.....	3
0.5. Méthodologie de recherche	3
0.6. Intérêt du sujet.....	3
0.7. Délimitation	4
0.8. Organisation du travail.....	4
CHAPITRE I : PRESENTATION DU LIEU DE STAGE	5
I.1.Présentation et Description du CNPK.....	5
I.1.1. Description du lieu de stage	5
I.1.2. Historique.....	5
I.1.3. Situation géographique	6
I.1.4. La vision du CNPK.....	7
I.1.5. Les Objectifs du CNPK.....	7
I.2. Déroulement du stage et activités accomplies	7
CHAPITRE II : CADRE THÉORIQUE ET METHODOLOGIE DE L'ETUDE.....	8
II.1. Cadre théorique de l'étude	8
II.1.1. Définitions des concepts	8
II.1.1.1. Maladies mentales ou troubles mentaux.....	8
II.1.1.2. Psychologie (Vie de l'esprit).....	9
II.1.1.3. Schizophrénie.....	9
II.1.1.4. Le trouble bipolaire	9

II.1.1.5. Epilepsie	10
II.1.2. Certains facteurs pouvant influencés les troubles mentaux.....	10
II.2. Méthodologie de l'étude adoptée.....	10
II.2.1. Type d'étude et population d'étude.....	10
II.2.2. Description des données	11
II.2.2.1. Définitions des variables	11
II.2.2.2. Choix du modèle	11
II.2.3. Statistique descriptive.....	11
II.2.4. Méthode utilisée : Modèle logistique	12
II.2.4.1. Définition de la Régression Logistique	12
II.2.4.2. Hypothèse de la régression logistique	14
II.2.4.3. Détermination de variables	14
II.2.4.3.1. Test du Chi-deux	14
II.2.4.3.2. Test de Fisher	15
II.2.4.4. Estimation des paramètres du modèle	16
II.2.4.4.1. Odds Ratio	17
II.2.4.4.2. Sélection de variables	17
II.2.4.4.3. Critères de multicollinéarité.....	18
II.2.4.4.4. Significativité globale du modèle.....	18
II.2.4.5. Evaluation de la performance du modèle	19
II.2.4.5.1. La Courbe de ROC	19
II.2.4.5.2. Le Critère AUC	21
II.2.4.6. Validation du modèle	21
II.2.5. Echantillonnage.....	22
CHAPITRE III : PRÉSENTATION ET INTERPRETATION DES RESULTATS.....	23
III.1. Analyse descriptive	23
III.1.1. Répartition de l'échantillon selon le Diagnostic	23
III.1.2. Comportement des facteurs sociodémographiques en fonction du Diagnostic.....	24
III.2. Analyse Inférentielle	28
III.2.1. Analyse primaire des données	28
III.2.1.1. Tableaux de contingence	28
III.2.1.2. Facteurs influençant les maladies mentales.....	30
III.2.2. Estimation des paramètres du modèle.....	32
III.2.2.1. Régression multiple.....	32

III.2.2.2. Sélection de variables	34
III.2.2.3. Critère de multicollinéarité	35
III.2.3. Evaluation de la performance du modèle	35
III.2.4. Validation du modèle	36
III.2.4.1. Test de rapport de vraisemblance	36
III.2.5. Prédiction du modèle	38
CHAPITRE IV : CORRELATION ENTRE LES ELEMENTS CONSTITUTIFS DES	
UNITÉS D’ENSEIGNEMENT ET LE STAGE	39
IV.1. Initiation à l’informatique.....	39
IV.2. Initiation à la recherche	39
IV.3. Calcul statistique sur ordinateur(R).....	39
IV.4. Statistique descriptive	39
IV.5. Traitement de données d’enquête	40
IV.6. Organisation d’une enquête	40
IV.7. Modèle linéaire	40
IV.8. Statistique sectorielle	40
IV.9. Théorie et Méthode Statistique	40
IV.10. Introduction au logiciel de programmation R.....	40
CONCLUSION GÉNÉRALE	41
REFERENCES.....	43
ANNEXES	44

LISTE DES TABLEAUX

Tableau 1 : Interprétation des valeurs du critère AUC	21
Tableau 2 : Tableau de contingence entre sexe et diagnostic.....	28
Tableau 3 : Tableau de contingence entre âge et diagnostic	28
Tableau 4 : Tableau de contingence entre région et diagnostic.....	29
Tableau 5: Tableau de contingence entre Religion et diagnostic	30
Tableau 6 : Tableau des facteurs influençant les maladies mentales.....	30
Tableau 7 : Test de Fisher	31
Tableau 8: Tableau de la régression multiple.....	32
Tableau 9 : Tableau des odds ratios	33
Tableau 10:Tableau de Sélection des AIC.....	34
Tableau 11: Tableau des FIV	35
Tableau 12: Tableau du test de rapport de vraisemblance 1	36
Tableau 13 : Tableau de rapport vraisemblance 2.....	37
Tableau 14 : Tableau prédictive	38

LISTE DES FIGURES

Figure 1 : Courbe de ROC	20
Figure 2 : Répartition de l'échantillon selon le Diagnostic.....	23
Figure 3 : Contribution de la variable Sexe.....	24
Figure 4 : Contribution de la variable Age.....	25
Figure 5 : Contribution de la variable Région.....	26
Figure 6 : Contribution de la variable Religion	27
Figure 7 : Courbe de ROC.....	36

CHAPITRE 0 : INTRODUCTION GENERALE

0.1. Contexte et Généralité sur l'épilepsie et les troubles mentaux

D'un point de vue philosophique, le cerveau est l'organe qui perçoit, qui pense, et qui agit [1]. Au niveau sociologique, le cerveau est le chef d'orchestre de l'organisme qu'il gère tout en se gérant lui-même. Il est responsable de nos comportements, et donc de nos interactions avec les individus qui composent la société [1].

Scientifiquement, le cerveau constitue un enjeu majeur car il renferme encore des mystères tant dans son développement, dans son fonctionnement normal et pathologique que dans ses facultés d'adaptation [1]. Il est composé de 100 milliards de cellules nerveuses (les neurones) qui constituent un réseau câblé très précis [3].

Selon les observations faites dans la base de données d'où nous avons découvert que deux des pathologies mentales sont les plus visibles chez les burundais. C'est ainsi que nous avons opté de les étudier et de mettre en évidence leur causes.

Aussi précieux qu'est cet organe, il n'en est pas moins exposé aux aléas de la vie. Nous verrons ici deux catégories de maladies qui touchent le cerveau : les maladies psychiatriques et neurologiques.

Les troubles mentaux (psychiatriques ou psychologiques) impliquent des perturbations de la pensée, de l'émotion et/ou du comportement. De petites perturbations de ces aspects de la vie sont fréquentes, mais lorsque ces troubles entraînent une détresse importante de la personne et/ou interfèrent avec sa vie quotidienne, ils sont considérés comme une maladie mentale ou un trouble mental. Les affections mentales peuvent être temporaires ou de longue durée. Il existe de nombreux types de troubles mentaux, désignés aussi sous le nom de problèmes de santé mentale [2].

En 2019, selon l'OMS, une personne sur huit dans le monde présente un trouble mental soit 970 millions de personnes présentait un trouble mental, les troubles anxieux et les troubles dépressifs étant les plus courants. La plupart des individus présentant des troubles mentaux n'ont pas accès à des soins efficaces. Nombre d'entre eux sont également victimes de stigmatisation ou de discrimination et subissent des violations de leurs droits [5].

L'épilepsie est une affection neurologique, chronique du cerveau qui touche 50 millions de personnes dans le monde.

Elle se caractérise par des crises récurrentes se manifestant par de brefs épisodes de tremblements involontaires touchant une partie du corps (crises partielles). Elles s'accompagnent parfois d'une perte de conscience et du contrôle de la vessie et de l'évacuation intestinale [4].

Ces crises résultent de décharges électriques excessives dans un groupe de cellules cérébrales. Ces décharges peuvent se produire dans différentes parties du cerveau. Les crises peuvent varier en intensité, allant de brèves pertes d'attention ou de petites secousses musculaires à des convulsions sévères et prolongées [4].

Voyants combien ces maladies sont banalisées voire même tabou tout en détruisant la vie de nos proches, ça nous a incité à nous pencher sur ce cas tout en abordant le thème "Quels sont les facteurs influençant les pathologies mentales au Burundi?".

Comme ces pathologies mentales sont moins abordées, nous allons apporter des chiffres qui feront travailler les méninges de plus d'un afin de constater l'urgence qu'il y a à les prévenir.

0.2. Problématique

Nombreux sont ceux qui ignorent l'existence des pathologies mentales, qui minimisent leurs ampleurs et les dégâts qu'elles peuvent occasionner lorsqu'elles sont traitées tardivement.

Vu le petit nombre de centres psychiatriques dont le Burundi dispose, le manque de données fiables ainsi que l'absence crue de sensibilisation en ce qui concerne ces maladies, nous désirons apporter notre contribution à travers ce travail afin de sensibiliser la population sur l'existence de ces pathologies pour ainsi les inciter à se prévenir.

Pour faire face à ce problème, nous sommes parties de la question suivante : << **Quels sont les facteurs influençant les pathologies mentales au Burundi?** >>

0.3. Objectifs

Notre travail vise deux objectifs qui sont entre autre l'objectif principal et l'objectif spécifique:

0.3.1. Objectif principal

L'objectif principal de notre travail est d'identifier quelques facteurs liés aux risques des pathologies mentales et leur proportion dans notre pays;

0.3.2. Objectif spécifique

L'objectif spécifique de notre travail est de prédire la probabilité d'être atteint de pathologies mentales ou non connaissant les autres facteurs ;

0.4. Hypothèses de recherche

Les hypothèses issues de notre travail sont :

- H 1 : L'Age est un facteur lié aux pathologies mentales ;
- H 2 : Le Sexe influence les pathologies mentales ;
- H 3 : La région de résidence influence les pathologies mentales ;
- H 4 : La Religion est un facteur à risque pour les pathologies mentales.

0.5. Méthodologie de recherche

Pour aboutir aux objectifs de notre travail, l'approche méthodologique s'articule sur quatre principales étapes. En premier lieu, il a été question d'effectuer une recherche documentaire pour découvrir et comprendre le sujet par la consultation des articles en ligne, livres et sites Web. L'étape suivante est le traitement des données et l'analyse descriptive afin de détecter les informations qu'on peut tirer des données avant de passer à l'étape qui constitue le noyau de cette étude à savoir la modélisation par le modèle de régression logistique.

0.6. Intérêt du sujet

Le sujet relève un intérêt sur trois aspects :

- En premier lieu pour nous, car il nous permet de mettre en pratique les matières apprises durant notre cursus académique et aussi d'apprendre davantage en ce qui concerne la santé mentale ;
- En second lieu, notre travail relève un intérêt académique pour l'ULT puisque ce travail sera entre autres un outil de documentation pour celui ou celle intéressé(e) par le sujet traité dans ce travail ;
- Et enfin, notre travail relève un intérêt institutionnel puisque les résultats issus des analyses réalisées dans ce travail seront utiles dans l'élaboration de techniques de prise en charge des

patients atteints de pathologies mentales. De plus, le CNPK étant un hôpital de référence dans le traitement des maladies mentales, ce travail sera un appui dans la bonne marche et l'aboutissement de leurs buts.

0.7. Délimitation

Notre travail a été délimité dans le temps et dans l'espace :

- Au niveau temporel, il a été réalisé en parallèle avec un stage académique au CNPK qui s'est déroulé du 04/Avril au 03/Mai 2022 et qui a porté sur les données de l'année 2020 ;
- Au niveau spatial, il a été réalisé au CNPK qui est un centre spécialisé dans le diagnostic et traitement des maladies mentales plus précisément le service du SIS ;

0.8. Organisation du travail

Outre l'introduction et la conclusion générales, notre travail s'articule sur quatre grands points à savoir :

La Présentation du lieu de stage, le Cadre Théorique et Méthodologie de l'étude, la Présentation et Interprétation des résultats, et enfin la Corrélation entre les éléments constitutifs d'unités d'enseignement et le stage effectué.

CHAPITRE I : PRESENTATION DU LIEU DE STAGE

Dans ce chapitre, nous verrons tout ce qui est en rapport avec La C.N.P.K

I.1.Présentation et Description du CNPK

I.1.1. Description du lieu de stage

Le Centre Neuropsychiatrique de Kamenge (CNPK) offre les services de médecine générale, de laboratoire, de radiographie mais il est surtout spécialisé dans le traitement des maladies mentales. Le CNPK englobe des bureaux administratifs, le service des urgences, des bâtiments d'hospitalisations, les réfectoires pour les malades hospitalisés, une chapelle, des salles de jeux pour les malades, une bibliothèque, ainsi que des espaces verts de détente.

Le CNPK offre un service médical, un service d'hospitalisation, un service de psychologie, un service social, ainsi qu'un service d'ergothérapie qui aide surtout lors des séances de détente pour les malades

I.1.2. Historique

Le C.N.P.K est né grâce à l'initiative du père José VERKEST, un père blanc, missionnaire au Burundi depuis 1940. Après avoir été amputé d'une jambe en 1965, il prit la décision de se consacrer au bien-être des personnes vivantes avec un handicap physique en créant un centre pour les personnes handicapées à la paroisse Kiganda en province de MURAMVYA.

Dans le but de chercher à soulager la souffrance des malades mentaux et leurs familles, le père écrivit une lettre au Frère Supérieur Général des Frères de la charité. Il avait déjà soumis ce désir à Monseigneur André MAKARAKIZA, (alors) archevêque de GITEGA, qui lui conseilla de contacter les Frères de la Charité. Monseigneur André MAKARAKIZA envoya une lettre au Frère Supérieur Général le 26 Août 1977.

Par la suite, le Frère Arthur SWITEN, Provincial des Frères de la charité écrivit une lettre à Monseigneur André MAKARAKIZA pour lui demander de contacter les autorités gouvernementales afin de constituer une synergie, d'où l'initiative de ce projet le 10 Janvier 1979. Le Ministère de la Santé Publique et le Frère Supérieur ont tenu une réunion à cet effet.

L'hôpital RWIBAGA qui avait été proposé par le Gouvernement fut rejeté car cet endroit n'était pas propice à cause de son mauvais climat, du problème d'eau et d'électricité.

C'est le terrain des Frères de la Charité qui sera choisi comme un milieu favorable pour abriter l'hôpital neuropsychiatrique.

L'an 1980 connaîtra le démarrage des consultations en ambulatoire avec nomination du premier psychiatre qui répondait au nom de Dr Paul LEGENTIL.

C'est en 1980 que le centre commença à accueillir en hospitalisation les cas les plus graves notamment ceux qui étaient hospitalisés à l'Hôpital Prince Régent Charles.

Le Centre Neuropsychiatrique de Kamenge a initié le plan quinquennal 1978-1981. Le Ministère de la Santé Publique s'était fixé l'objectif de promouvoir la médecine sociale et la médecine pour tous. C'est à la fois une médecine intégrale et intégrée parce qu'elle compte pour tout l'homme, tant au niveau de son aspect physique, mental et de son hygiène corporelle.

En effet, le CNPK a commencé à l'intérieur du pays sous forme de branches, le CNP de NGOZI, le CSM de GITEGA. Cela facilite aux malades mentaux qui vivent dans les régions qui sont les plus éloignées de Bujumbura d'accéder aux soins psychiatriques. Il faut noter qu'actuellement, il y a également la construction d'une pédopsychiatrie pour séparer les patients adultes des enfants pour mieux suivre et traiter les enfants malades psychiquement.

En 1983 : Une équipe itinérante démarre des consultations psychiatriques dans quelques provinces sanitaires périphériques notamment : GITEGA et MURAMVYA au centre du pays et KAYANZA au nord.

Fin 1984 : L'effectif des consultations s'élève déjà à 600 cas enregistrés.

En 1987 : Par manque de fonds et de personnel, l'ouverture des hospitalisations initialement prévues n'a pas lieu et les activités des équipes mobiles sont supprimées.

Le 17/7/1987 : Un premier Neuropsychiatre Burundais (Dr BARANCIRA) fut nommé au centre.

En 1998 : Un accord de principe est donné par le Ministre de la Santé publique pour une installation d'une pharmacie des médicaments usuels et psychotropes au CNPK. [6]

I.1.3. Situation géographique

Le Centre Neuropsychiatrique de Kamenge (CNPK) est une institution de santé située au Nord dans la Province de Bujumbura Mairie. Il est dans la Commune Ntahangwa dans la Zone de Gihosha, Quartier Muyaga à presque 300m de l'ETS Kamenge, avenue Mwakiro N° 19.

I.1.4. La vision du CNPK

La vision du CNPK s'inspire du but de la politique nationale du développement sanitaire 2011-2015 qui est d'Améliorer l'Etat de santé de la population, non seulement parce que c'est un droit humain, mais aussi pour permettre la relance de l'économie et la diminution de la pauvreté en maintenant le capital humain en bonne santé [6].

I.1.5. Les Objectifs du CNPK

Les Objectifs du CNPK correspondent aux objectifs généraux de la stratégie nationale de la santé mentale suivante.

- Actualiser et adapter la régistration existante en alignement avec les normes internationales des droits de l'homme et en promouvoir l'application effective en santé mentale,
- Améliorer la qualité et l'accessibilité des services de santé mentales,
- Offrir à la population des soins de santé mentale de qualité dans les structures de soins primaires,
- Accroître les ressources humaines en santé mentale et améliorer leurs compétences,
- Intensifier les campagnes d'information, d'éducation et de sensibilisation du public et promouvoir le travail en synergie des intervenants en santé mentale [6].

I.2. Déroulement du stage et activités accomplies

Le stage effectué a duré un mois à compter du 04/Avril au 03/Mai 2022 au CNPK dans les services du SIS. Durant notre période de stage, nous avons pu avoir un aperçu sur les logiciels utilisés dans le milieu sanitaire notamment: Open Clinic qui joue le rôle d'enregistrement des malades et DHIS2 qui joue le rôle de transmission d'informations sur les maladies (nombre de malades, tranches d'âge,...). Ça a été pour nous une occasion de voir et comprendre l'apport des Statistiques dans le monde sanitaire.

Nous avons également eu l'opportunité de visiter quelques services du CNPK comme les services d'hospitalisation l'occasion pour nous d'avoir un contact avec les malades.

CHAPITRE II : CADRE THÉORIQUE ET METHODOLOGIE DE L'ETUDE

Dans ce chapitre, nous verrons deux points qui sont : Le cadre théorique de notre travail ainsi que sa méthodologie de recherche

II.1. Cadre théorique de l'étude

Ce chapitre comprend l'ensemble des concepts utilisés pour ce travail.

Quelques concepts ci-dessous et mots clés utilisés tout au long des analyses sont définis dans ce cadre. Le travail cherche à déterminer les facteurs influençant les troubles mentaux. Pour les estimations, les variables utilisées considèrent que les caractéristiques sociodémographiques et le facteur non modifiable comme l'âge influence les troubles mentaux.

II.1.1. Définitions des concepts

II.1.1.1. Maladies mentales ou troubles mentaux

La maladie mentale touche au psychisme de l'individu et affecte ses pensées ou son comportement. Les origines de ces maladies sont diverses et varient d'une personne à l'autre mais l'on peut se retrouver des facteurs communs que sont le stress environnemental, ou les antécédents génétiques. Dans la majorité des cas, les maladies mentales peuvent être soignées par le biais d'une thérapie.

Une maladie mentale, parfois appelée trouble mental, est un ensemble de manifestations cliniques d'ordre psychologique et comportemental qui ne sont pas toujours aisées à identifier. Ces signes incluent les troubles de l'humeur, une altération de la perception de la réalité ou l'anxiété, la souffrance, etc.

Ces maladies mentales sont caractérisées notamment par l'instabilité, l'impulsivité ou par l'accomplissement d'actes antisociaux.

Synonymes : maladie dégénérative, maladie nerveuse, troubles nerveux, folie, maladie psychique, troubles mentaux, troubles psychiatriques, aliénation, troubles psychiques[5].

II.1.1.2. Psychologie (Vie de l'esprit)

Esprit ou psyché, en particulier considéré dans sa globalité. Le psychisme désigne ce qui n'est pas somatique, il est attaché au corps mais jusqu'à une certaine limite.

C'est un ensemble de phénomènes participant à l'activité psychique, consciente ou inconsciente, considérée globalement ou partiellement.

II.1.1.3. Schizophrénie

La schizophrénie d'après Hanus M. (1975.p66), est une affection grave et chronique, atteint avec une fréquence à peu près égale les sujets jeunes de deux sexes. Elle se rencontre dans toutes les races et dans toutes les cultures. << Ses caractéristiques au début sont : la discordance idéo-affective, incohérence, l'ambivalence, l'autisme, les hallucinations et les idées délirantes désigne généralement ce que l'on appelle la psychose du jeune adulte entre 15 et 25 ans. On ne connaît pas jusqu'à présent son étiologie et elle entraîne la discordance et l'incohérence qui font que le sujet ne reconnaît plus les éléments de son environnement (dépersonnalisation-déréalisation). Le début peut être aigu dans 35 à 40% des cas et se présente sous forme de bouffées délirantes aiguës et d'états dysthymiques >>.

Exceptionnellement, elle peut apparaître parfois pendant l'enfance ou après 3 ans on parle alors de schizophrénie à but tardif selon le cas. Cette maladie semble être le résultat d'une combinaison de problèmes, mettant en cause une certaine vulnérabilité génétique et des facteurs propres à l'environnement [5].

II.1.1.4. Le trouble bipolaire

Le trouble bipolaire est un trouble qui se manifeste chez le malade sous deux modalités ou deux pôles : un pôle triste et un pôle gai. C'est pourquoi il est aussi appelé le trouble maniaco-dépressif ou la psychose maniaco-dépressive.

Les troubles bipolaires se caractérisent par les épisodes au cours desquels l'humeur et le niveau d'activité de la personne sont nettement perturbés. Ces perturbations consistent, à certains moments, en une élévation de l'humeur et une augmentation de l'énergie et de l'activité (manie), et à d'autres moments, en une dégradation de l'humeur et une baisse de l'énergie et de l'activité (dépression). La récupération est classiquement complète entre les épisodes. Les personnes ne présentant que des épisodes maniaques sont aussi classées comme atteintes d'un trouble bipolaire [16].

II.1.1.5. Epilepsie

L'épilepsie est une maladie neurologique qui se caractérise par un fonctionnement anormal de l'activité du cerveau. Elle se traduit par la répétition de crises imprévisibles et souvent très brèves [10].

Ces crises résultent de décharges électriques excessives dans un groupe de cellules cérébrales. Ces décharges peuvent se produire dans différentes parties du cerveau. Les crises peuvent varier en intensité, allant de brèves pertes d'attention ou de petites secousses musculaires à des convulsions sévères et prolongées. Leur fréquence est également variable, de moins d'une fois par an à plusieurs fois par jour [4].

II.1.2. Certains facteurs pouvant influencés les troubles mentaux

On pense actuellement que la maladie mentale est causée par une interaction complexe de facteurs, notamment : [2]

- Génétique
- Biologiques (facteurs psychiques)
- Psychologiques
- Environnementaux (y compris des facteurs sociaux et culturels)

II.2. Méthodologie de l'étude adoptée

Nous aurons recours aux statistiques descriptives et inférentielles . Cette dernière permettra d'affiner des relations entre une ou plusieurs variables dépendantes et d'autres variables (indépendantes) à travers des fonctions mathématiques.

Nous disposons d'un échantillon représentatif de la population mère, c'est-à-dire l'image simplifiée de cette dernière et d'en préciser les critères et les règles de sélection. Nous préciserons également les critères et les règles de sélection utilisés.

II.2.1. Type d'étude et population d'étude

Cette étude est de type transversal à visée exploratoire et analytique faite dans le Service d'Information Sanitaire à la C.N.P.K.

La population cible est composée de personnes diagnostiquées de l'Epilepsie, de la Schizophrénie et de Troubles bipolaires durant la période du 1er janvier 2020 jusqu'au 31 décembre 2020. Cette population est faite de personnes venant de tout le pays et de tout âge.

II.2.2. Description des données

La base de données autour de laquelle se sont faites nos analyses dans ce travail comporte 5 variables: l'age, le sexe, la region, la religion et le diagnostic.

II.2.2.1. Définitions des variables

- Sexe : ayant deux modalités féminin et masculin.
- Religion : ayant 6 modalités Adventiste, Catholique, Islam, Pentecôte, Protestant, Témoins de Jéhovah.
- Région : ayant 6 modalités Bujumbura Mairie, Centre, Est, Nord, Ouest, Sud.
- Age : réparti en classes de 0-9 ans, 10-19 ans, 20-49 ans, 50 et plus.
- Diagnostic : est soit les troubles mentaux, soit l'épilepsie donc une variable binaire

II.2.2.2. Choix du modèle

Le choix du modèle a été guidé par la condition de la variable d'intérêt qui est une variable qualitative, le choix s'est donc porté sur **le Modèle Logistique** qui est dans ce cas adapté pour faire les analyses.

II.2.3. Statistique descriptive

L'analyse descriptive nous a servi dans la description et la structuration des données.

Elle nous a permis de (montrer) présenter la relation qui existe entre les différentes données et d'en tirer une information statistique.

Afin de produire des statistiques relatives à ce travail, nous aurons à effectuer des analyses descriptives au travers des tableaux et des graphiques.

II.2.4. Méthode utilisée : Modèle logistique

Le modèle logistique est typiquement utilisé avec des variables à expliquer binaire, or une telle situation se rencontre fréquemment dans le domaine des sciences humaines et sociales ou en médecine. Il suit les étapes suivantes :

- Hypothèse de la Régression Logistique
- Estimation des paramètres du modèle
- Evaluation de la performance du modèle
- Validation du modèle
- Prédiction du modèle

II.2.4.1. Définition de la Régression Logistique

La régression logistique binaire est une technique permettant de modéliser une variable dépendante qualitative binaire prenant ses valeurs dans $\{0,1\}$ à l'aide d'une combinaison linéaire des variables explicatives (qualitative ou quantitative) avec la fonction de lien logit.

La régression logistique ordinaire ou régression logistique binaire vise à expliquer une variable d'intérêt binaire (c'est-à-dire de type « oui/non » ou « vrai/faux ») et traite beaucoup de cas dont la prévision de présence/absence d'une maladie. Les variables explicatives qui seront introduites dans le modèle peuvent être quantitatives ou qualitatives.

En tant que procédure non paramétrique, la régression logistique présente l'avantage de ne pas exiger de contraintes quant à la normalité des distributions des variables [15].

Cette technique est utilisée pour des études ayant pour but de vérifier si des variables indépendantes peuvent prédire une variable dépendante dichotomique. Contrairement à la régression multiple et l'analyse discriminante, cette technique n'exige pas une distribution normale des prédictors ni l'homogénéité des variances [14].

Soit Y une variable à valeurs dans $\{0, 1\}$ à expliquer par p variables explicatives

$X = (1, X_1, \dots, X_p)'$. Le modèle logistique propose une modélisation de

la loi de $Y | X = x$ par une loi de Bernoulli de paramètre $p_\beta(x) = P_\beta(Y = 1 | X = x)$ telle que :

fonction de lien logit :

$$\log \frac{p_\beta(x)}{1-p_\beta(x)} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p = x' \beta, (1.1)$$

ou encore

$$\text{logit } p_\beta(x) = x' \beta$$

logit désignant la fonction bijective et dérivable de $]0, 1[$

dans $\mathbb{R} : p \mapsto \log(p/(1-p))$.

L'égalité (1.1) peut également s'écrire

$$p_\beta(x) = P_\beta(Y = 1 | X = x) = \frac{\exp(x' \beta)}{1 + \exp(x' \beta)}.$$

Remarque 1.1 Dans un modèle logistique, nous effectuons deux choix pour définir le modèle :

1. le choix d'une loi pour $Y | X=x$ ici la loi de Bernoulli ;
2. le choix de la modélisation de $P(Y = 1 | X = x)$ par
logit $P_\beta(Y = 1 | X = x) = x' \beta$.

La fonction logit est bijective et dérivable. Elle est appelée fonction de lien.

Remarquons également que

$$E_\beta [Y | X=x] = P_\beta(Y = 1 | X = x)$$

$$V_\beta (Y | X=x) = P_\beta(Y = 1 | X = x)(1 - P_\beta(Y = 1 | X = x))$$

ce qui implique que la variance n'est pas constante et varie selon x [12].

Définir le modèle revient à spécifier la loi de ε

Si ε suit une loi logistique, c'est à dire de fonction de répartition

$$F_\varepsilon(x) = \frac{\exp(x)}{1 + \exp(x)}; \text{ alors le modèle est le modèle logistique.}$$

Si ε suit une loi normale centrée réduite alors le modèle est le modèle probit [13].

II.2.4.2. Hypothèse de la régression logistique

La régression logistique introduit l'hypothèse fondamentale suivante :

$$\ln \left[\frac{P(X|Y=+)}{P(X|Y=-)} \right] = b_0 + b_1 X_1 + \dots + b_p X_p,$$

Cette hypothèse couvre une large palette de lois de distribution des données :

- La loi normale (comme pour l'analyse discriminante) ;
- Les lois exponentielles ;
- Les lois discrètes ;
- Les lois Beta, les lois Gamma et les lois de Poisson ;
- Un mélange de variables explicatives binaires (0/1) et continues, cette propriété est très importante car elle rend opérationnelle la régression logistique dans de très nombreuses configurations.

La régression logistique est une méthode semi-paramétrique car l'hypothèse porte uniquement sur le rapport de ces probabilités. Elle est moins restrictive. Son champ d'action est donc théoriquement plus large [12].

II.2.4.3. Détermination de variables

II.2.4.3.1. Test du Chi-deux

Le test du Chi-2 permet de tester l'indépendance des variables à partir du tableau de Contingence. Toutefois, celui-ci ne constitue pas un indicateur commode de la force de l'association entre deux variables car il n'est pas normé. L'expression du Chi-2 est la suivante

$$X_t^2 = \sum_{j=1}^c \sum_{i=1}^1 \frac{(n_{ij} - n_{ij}^t)^2}{n_{ij}^t}$$

Cette quantité suit une loi du χ^2 à $(1-1) (c-1)$ degrés de liberté. Si le χ^2 observé est supérieur au X_t^2 théorique, à un certain seuil de signification, on peut alors rejeter l'hypothèse nulle et ainsi conclure sur le lien entre les deux variables [10].

H_0 : Les données sont indépendantes

H_1 : Les données sont corrélées (liées)

p-value $\leq 5\%$: On rejette H_0

II.2.4.3.2. Test de Fisher

Le test de Fisher permet de tester la significativité de l'ensemble des coefficients d'un modèle.

Soit le modèle général :

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \dots + a_{k-1} x_{(k-1)t} + \varepsilon_t \quad \text{pour } t = 1, 2, \dots, T.$$

Les hypothèses du test de Fisher sont les suivantes :

$$H_0: a_1 = a_2 = \dots = a_{k-1} = 0 \text{ (la constante } a_0 \text{ est non nul)}$$

\Rightarrow L'ensemble des coefficients du modèle est non significatif

$$H_1 : \text{il existe au moins un coefficient non nul.}$$

La statistique de test sous H_0 vraie est :

$$f = \frac{(SCR_c - SCR_{nc}) / (dl_c - dl_{nc})}{SCR_{nc} / dl_{nc}} \sim F(dl_c - dl_{nc}, dl_{nc}) = F(p, q)$$

où SCR_c = SCR du modèle contraint (modèle lorsque H_0 est vérifiée)

SCR_{nc} = SCR du modèle non contraint (modèle lorsque H_1 est vérifiée)

dl_c = degré de liberté du modèle contraint = $T - 1$ (car il n'y a qu'une seule variable explicative qui est non nul qui est le terme constant a_0)

dl_{nc} = degré de liberté du modèle non contraint = $T - k$ (car il y a k variables explicatives au maximum dans le modèle).

La règle de décision est la suivante :

Si $f > f^*(p, q)$ où $f^*(p, q)$ est la valeur donnée par la table de Fisher pour p et q donnés et pour un risque fixé

\Rightarrow On accepte H_1 : il existe au moins un coefficient non nul.

Ce test est peu utilisé car lorsqu'il indique qu'il y a au moins un coefficient non nul, il ne précise pas lesquels. Il est moins précis que le test de Student [20].

II.2.4.4. Estimation des paramètres du modèle

Pour estimer les paramètres de la régression logistique par la méthode du maximum de vraisemblance, nous devons tout d'abord déterminer la loi de distribution de $P(Y=X)$.

Y est une variable binaire définie dans $\{+, -\}$ (ou $\{1,0\}$ pour simplifier les écritures). Pour un individu ω , on modélise la probabilité à l'aide de la loi binomiale $B(1; \pi)$, avec

$$P[Y(\omega)/X(\omega)] = \pi(\omega)^{y(\omega)} \times (1 - \pi(\omega))^{(1-y(\omega))}$$

Cette modélisation est cohérente avec ce qui a été dit précédemment, en effet :

- Si $y(\omega) = 1$, alors $P[Y(\omega) = 1/X(\omega)] = \pi$;
- Si $y(\omega) = 0$, alors $P[Y(\omega) = 0/X(\omega)] = 1 - \pi$; [12]

Le caractère binaire de la variable à expliquer rend la méthode des moindres carrés impossible à mettre en œuvre dans ce contexte [13].

On rappelle que les estimateurs des moindres carrés du modèle linéaire gaussien coïncident avec les estimateurs du maximum de vraisemblance [13].

C'est par cette approche que sont estimés les paramètres du modèle logistique à partir d'un n-échantillon $(x_1; Y_1), \dots, (x_n; Y_n)$ (les variables aléatoires Y_i sont indépendantes) [13].

La vraisemblance d'un échantillon s'écrit alors sous cette forme suivante : [12]

$$L = \prod_i p_i^{y_i} (1 - p_i)^{1-y_i}$$

On note que la vraisemblance correspond à la probabilité d'obtenir l'échantillon à partir d'un tirage dans la population. Elle varie donc de 0 et 1. La méthode du maximum de vraisemblance consiste à produire les paramètres $\beta = (\beta_0, \dots, \beta_J)$ de la régression logistique qui rendent maximum la probabilité d'observer cet échantillon [12].

L'estimateur du maximum de vraisemblance, s'il existe, annule le gradient de la vraisemblance (et celui de la log-vraisemblance).

Pour le cas d'une régression à p variables logistique, la vraisemblance est non linéaire en $\beta = (\beta_1, \dots, \beta_p)^T$ et n'admet de solution explicite. Il faut donc utiliser des algorithmes d'optimisation itératifs pour maximiser la vraisemblance notamment l'algorithme IRLS [12].

L'algorithme de Newton- Raphson est aussi l'une des méthodes numériques les plus utilisées

pour optimiser la log-vraisemblance. Il démarre avec une initialisation quelconque du vecteur de paramètre . Pour passer de l'étape i à l'étape $(i + 1)$, il se rapproche de la solution finale $\hat{\beta}$ en utilisant la formule suivante [12].

$$\beta^{i+1} = \beta^i - \left(\frac{\partial LL}{\partial \beta} \cdot \frac{\partial LL}{\partial \beta'} \right)^{-1} \cdot \frac{\partial LL}{\partial \beta}$$

II.2.4.4.1.Odds Ratio

Le rapport des probabilités est appelé « Odds » (cote). On définit l'Odd comme étant le rapport : $Odd = \frac{p}{1-p}$ où p traduit, par exemple, la probabilité de réussir une action et $1 - p$, la probabilité d'échouer.

Il est important de noter que les coefficients de la régression logistique ne présentent pas les *odds* mais les *odds ratio (OR)*. Ces derniers traduisent les chances que la variable y prenne la modalité j *versus* la modalité de référence $y=0$, lorsque $x=\text{mod}1$, *versus* $x=\text{mod}2$.

Si on note $p0$ la probabilité de réussir cette action pour une femme et $p1$ la probabilité de réussir cette même action pour un homme, alors l'OR associé au genre est égal au rapport :

$$OddRatio = \frac{\frac{P1}{1-P1}}{\frac{P0}{1-P0}}$$

Si l'OR prend la valeur 1, cela traduit le fait que la probabilité de réussir l'action est la même pour les hommes que pour les femmes. Une valeur supérieure à 1 indique par contre que les hommes ont plus de chance de réussir l'action que les femmes.

Dans le modèle logistique, les coefficients calculés sont en fait égaux au logarithme népérien de l'OR. Il faut donc appliquer la fonction exponentielle aux coefficients de la régression afin de pouvoir analyser les Odds Ratio [15].

II.2.4.4.2. Sélection de variables

On peut comparer des modèles entre eux, ou procéder à une sélection automatique du modèle, en se basant aux critères d'information tels que l'AIC (critère d'Akaike) ou le BIC (critère de Schwarz), à l'aide des algorithmes ci-après :

- FORWARD : c'est une sélection qui part du modèle trivial et qui rajoute une à une des variables explicatives jusqu'à ce que l'on déclenche la règle d'arrêt.
- BACKARD : Celle-ci part du modèle complet et inclue la totalité des descripteurs, puis enlève une à une les variables non significatives.

- STEPWISE : elle alterne FORWARD et BACKWARD. Elle consiste à vérifier si chaque ajout de variable ne provoque pas l'explicative qui aurait été intégré.

Le modèle avec la plus petite valeur du critère d'information est celui à retenir.

Toutes les variables ne sont pas pertinentes pour la prédiction. Opérer une sélection de variables permet de simplifier le modèle et en améliorer la lisibilité, tout en conservant ses qualités prédictives [12].

II.2.4.4.3. Critères de multicollinéarité

Pour mesurer la multicollinéarité, vous pouvez étudier la structure de corrélation des variables de prédiction. Vous pouvez également examiner les facteurs d'inflation de la variance (FIV) des coefficients de régression du modèle. Les FIV mesurent l'accroissement de la variance d'un coefficient de régression estimé s'il existe une corrélation entre vos prédicteurs. Si tous les FIV sont égaux à 1, il n'existe pas de multicollinéarité, mais si certains FIV sont supérieurs à 1, les prédicteurs sont corrélés. Lorsqu'un FIV est supérieur à 5, le coefficient de régression de ce terme n'est pas correctement estimé [19].

II.2.4.4.4. Significativité globale du modèle

En régression classique, l'hypothèse de nullité de chaque coefficient est testée au moyen de la statistique de STUDENT. Le test équivalent en régression logistique s'appelle le test de WALD. Cette statistique, notée W , est obtenue en rapportant l'estimation β du maximum de vraisemblance à une estimation de son erreur standard (se) : [11]

$$W = \beta / (\text{sê}(\beta))$$

Evaluer la significativité du modèle revient à tester la nullité simultanée de tous les coefficients relatifs aux variables explicatives dans le modèle. Les hypothèses nulle et alternative se définissent comme suit :

$$H_0 : \beta_1 = \beta_2 = \dots \beta_p = 0$$

$$H_1 : \forall i \in \{1, 2, \dots, p\} \exists \text{ au moins } \beta_i \neq 0$$

La statistique du test (W) suit une loi du khi-2 à p degré de liberté sous H_0 . Elle s'écrit : [12]

$$W_{(p)} = \hat{\beta}_{(p)}' \times \hat{\Sigma}_{(p)}^{-1} \times \hat{\beta}_{(p)}.$$

II.2.4.5. Evaluation de la performance du modèle

Nous nous consacrons à ce que l'on appellerait des méthodes d'évaluation externes 1, basées sur les prédictions $\hat{y}'(\omega)$ et/ou les probabilités a posteriori $\hat{\pi}(\omega)$ fournies par le classifieur. A aucun moment nous n'exploitons des informations spécifiques (internes) à la régression logistique (log vraisemblance). Ces méthodes d'évaluation externes sont les suivantes [12].

II.2.4.5.1. La Courbe de ROC

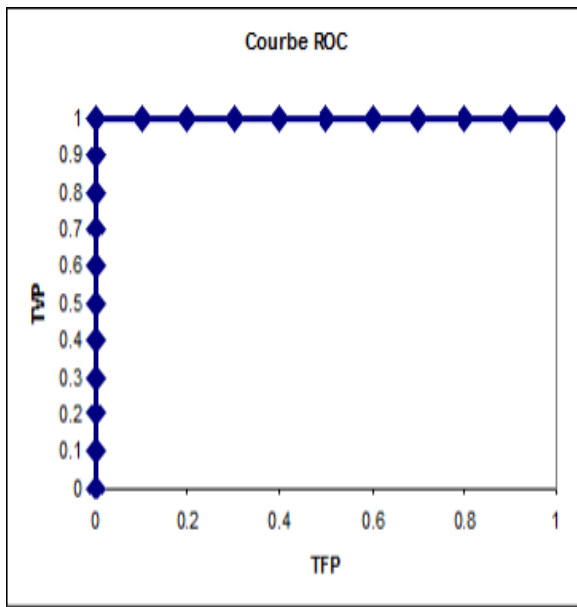
La courbe ROC permet de mettre en relation le taux de vrais positifs TVP (la sensibilité, le rappel) et le taux de faux positifs TFP ($\text{TFP} = 1 - \text{Spécificité}$) reçus à l'aide du modèle de prédiction dans un graphique de nuage de points.

Le taux de vrais positifs traduit la proportion des individus qui ont été correctement prédites parmi toutes les observations positives tandis que le taux de faux négatifs montre la proportion des observations incorrectement prédites positives parmi les observations négatives,

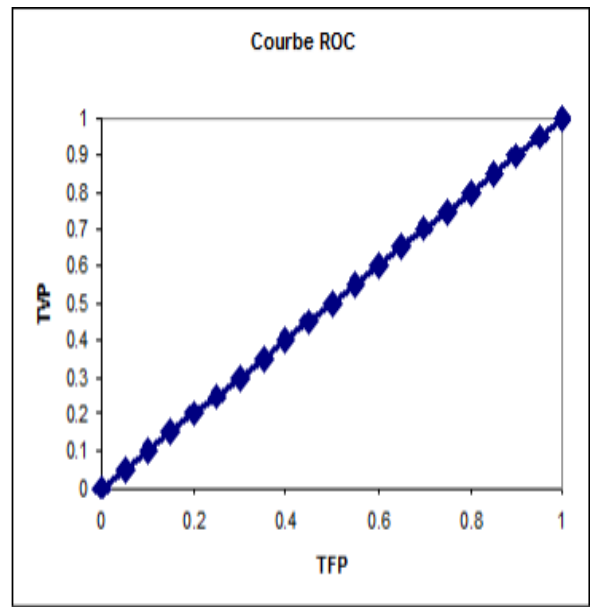
La courbe ROC correspond au graphique nuage de points qui relie les couples (TVP, TFP). Le premier point est forcément (0, 0), le dernier est (1, 1).

Deux cas extrêmes peuvent se présenter:

- La courbe ROC est collée aux extrémités Ouest et Nord du repère. Les scores sont totalement inopérants, dans ce cas les positifs et les négatifs sont mélangés.
- La courbe ROC se confond avec la première bissectrice [12].



[A] Discrimination parfaite. Tous les positifs sont situés devant les négatifs lorsque l'on trie le tableau selon un score décroissant.



[B] Pas de discrimination. Les positifs et les négatifs sont mélangés c.-à-d. présentent des scores en moyenne identiques.

Figure 1 : Courbe de ROC

II.2.4.5.2. Le Critère AUC

Il est possible de caractériser numériquement la courbe ROC en calculant la surface située sous la courbe. C'est le critère AUC. Elle exprime la probabilité de placer un individu positif devant un négatif. Ainsi, dans le cas d'une discrimination parfaite, les positifs sont sûrs d'être placés devant les négatifs, nous avons $AUC = 1$. A contrario, si le classifieur attribue des scores au hasard, il y a autant de chances de placer un positif devant un négatif que l'inverse, la courbe ROC se confond avec la première bissectrice, nous avons $AUC = 0.5$. C'est la situation de référence, notre classifieur doit faire mieux. On propose généralement différents paliers pour donner un ordre d'idées sur la qualité de la discrimination.

Valeur de l'AUC	Commentaire
$AUC = 0.5$	Pas de discrimination.
$0.7 \leq AUC < 0.8$	Discrimination acceptable
$0.8 \leq AUC < 0.9$	Discrimination excellente
$AUC \geq 0.9$	Discrimination exceptionnelle

Tableau 1 : Interprétation des valeurs du critère AUC

Pour calculer l'AUC, nous pouvons utiliser une bête intégration numérique, la méthode des trapèzes par exemple. Nous verrons plus loin que sa valeur peut être obtenue autrement, en faisant le parallèle avec le test de Mann-Whitney.

Au final, il apparaît que le critère AUC est un résumé très commode. Il permet, entre autres, les comparaisons rapides entre les classifieurs. Mais il est évident que si l'on souhaite analyser finement leur comportement, rien ne vaut la courbe ROC [12]

II.2.4.6. Validation du modèle

❖ Test de rapport de vraisemblance

Le test de rapport de vraisemblance est un test d'hypothèse qui compare l'adéquation de l'ajustement de deux modèles afin de déterminer celui qui offre le meilleur ajustement pour les données échantillons : un modèle non contraint dont tous les paramètres sont libres et son modèle contraint correspondant avec moins de paramètres pour tester l'hypothèse nulle. [17]

Le test du rapport de vraisemblance est basé sur le ratio des deux fonctions de vraisemblance obtenues des modèles avec ou sans les restrictions paramétriques.

Formellement, notons

- $L_u = \pi_{i=1}^n f_u(y_i)$ la vraisemblance du modèle non-restreint (*unrestricted*), et
- $L_r = \pi_{i=1}^n f_r(y_i)$ la vraisemblance du modèle restreint (*restricted*).

Ainsi, ce rapport des vraisemblance équivaut à $\frac{L_u}{L_r}$. Nous savons qu'au lieu de travailler avec la fonction de vraisemblance, il est plus pratique de travailler avec les log-vraisemblances. Ainsi, le test du rapport de vraisemblance peut s'exprimer comme:

$$\text{TRV} = 2 \ln \left(\frac{L_u}{L_r} \right) = 2(l_u - l_r),$$

avec $l_u = \ln(L_u)$ la fonction de logvraisemblance du modèle non-restreint, et la fonction de $l_r = \ln(L_r)$ logvraisemblance du modèle restreint. Puisque le modèle sans restriction inclut celui avec restrictions, l_u sera toujours plus grand (ou égal) que l_r . Ainsi, le test du rapport de vraisemblance sera toujours positif, et aura une distribution suivant une X_q^2 où q est le nombre de restrictions paramétriques.

Une valeur du test TRV petit signifie que le modèle restreint est presque aussi bon que le modèle non-restreint, ce qui justifie ainsi l'utilisation du modèle le plus simple, i.e. le modèle restreint [18].

II.2.5. Echantillonnage

Sachant qu'il y a un grand nombre de personnes souffrant de maladies mentales ayant été traité à la C.N.P.K durant ces dernières années et que leur système d'enregistrement des patients a été informatisé dès 2020, nous avons opté d'étudier sur les patients traités durant la période de 2020-2021. Notre étude se porte sur des patients souffrant d'épilepsie et de troubles mentaux. Notre échantillon est donc ces patients traités durant cette période.

CHAPITRE III : PRÉSENTATION ET INTERPRÉTATION DES RESULTATS

Dans ce chapitre, il sera question de mettre en évidence la relation entre la variable à expliquer (Le diagnostic) et les autres variables. Afin d'identifier et de relever la relation existante entre la variable d'intérêt et les autres variables indépendantes des méthodes statistiques seront utilisés.

Le but est de savoir les facteurs associés aux pathologies mentales et d'estimer le traité au CNPK dans la période de 2020-2021.

III.1. Analyse descriptive

III.1.1. Répartition de l'échantillon selon le Diagnostic

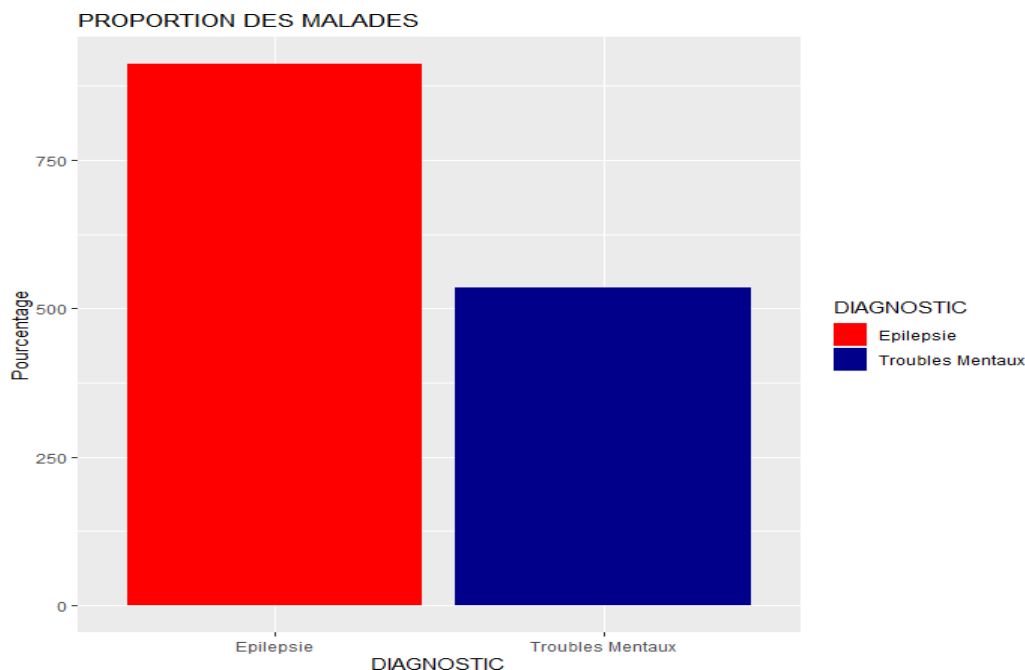
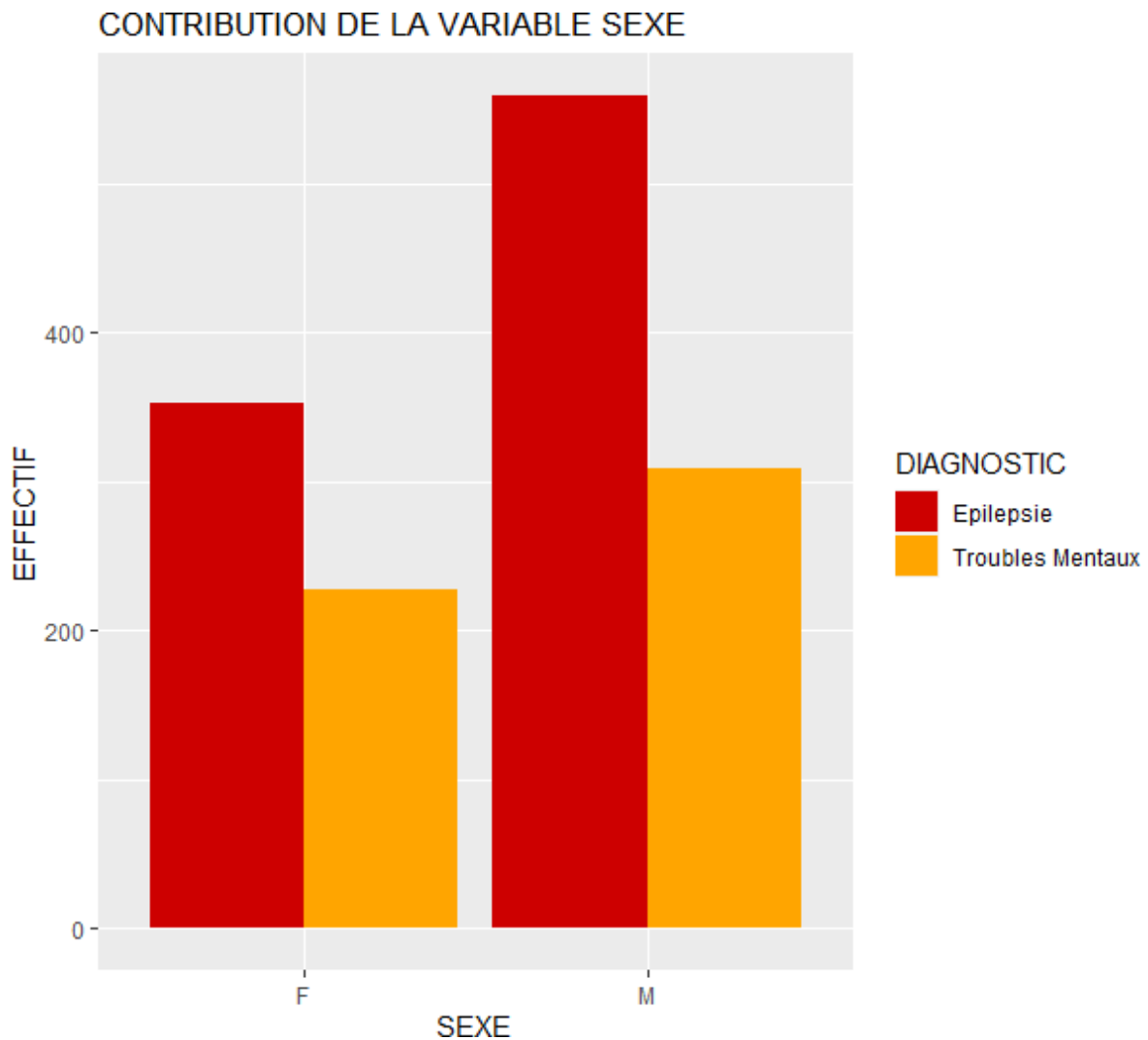


Figure 2 : Répartition de l'échantillon selon le Diagnostic

Le graphique renseigne sur le nombre de malades diagnostiqués épileptiques et celui des malades avec des troubles mentaux dans l'échantillon. L'effectif des épileptiques (912 malades soit 63.02%) est plus élevé que celui de ceux ayant des troubles mentaux (535 malades soit 36.97%).

III.1.2. Comportement des facteurs sociodémographiques en fonction du Diagnostic



Ce graphique montre que l'effectif des malades diagnostiqués de troubles mentaux est faible à ceux diagnostiqués de l'Epilepsie que ce soit pour les Hommes ou pour les Femmes. L'Epilepsie est plus présent chez les Hommes que chez les Femmes. En effet, on enregistre 559 cas sur 912 chez les Hommes soit 61.29% contre 353 chez les Femmes soit 38.7%.

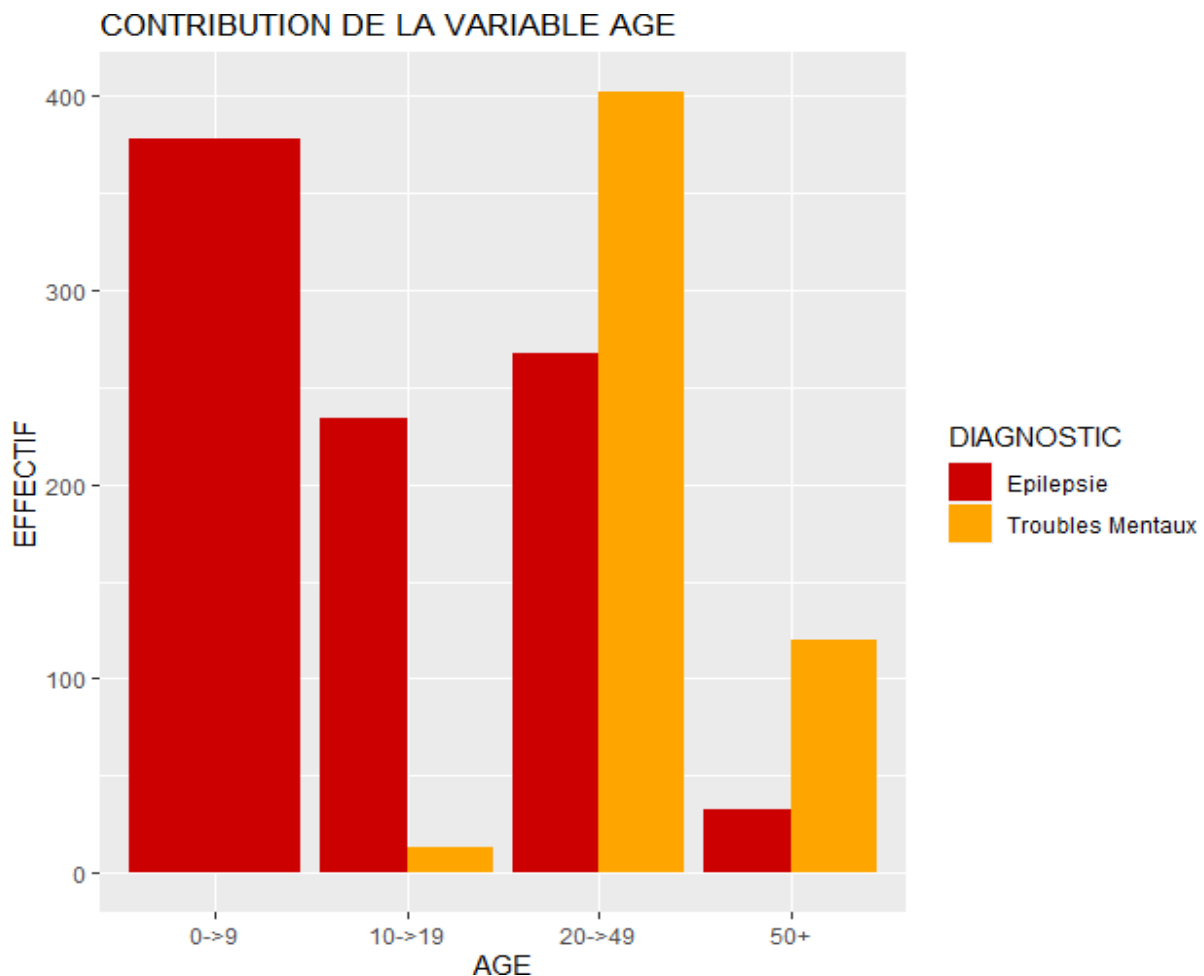


Figure 4 : Contribution de la variable Age

Ce graphique nous montre que c'est chez les individus ayant entre 20-49 ans qu'on observe plus de cas de Troubles mentaux, tandis que chez les individus en bas âge on n'en observe aucun. Cela est probablement dû au manque d'informations de la part des parents qui retarde le diagnostic en n'envoyant pas leurs enfants se faire consulté. En effet, 402 cas sur un total de 535 cas soit 75.1% de Troubles Mentaux au total se sont fait remarqué chez les patients ayant entre 20-49 ans. On peut également voir que le plus de cas d'Epilepsie sont enregistrés chez les patients qui sont entre 0-9 ans. En effet, on remarque qu'un total de 378 patients sur 912 cas soit 45,44% ayant entre 0-9 ans a été enregistré.

On remarquera également que chez les patients ayant entre 20-49 ans, il y'a non seulement plus de cas de Troubles Mentaux mais aussi que le nombre de patients diagnostiqué épileptique est non négligeable; on enregistre 267 cas sur un total de 912 soit 32.10% cas d'Epilepsie.

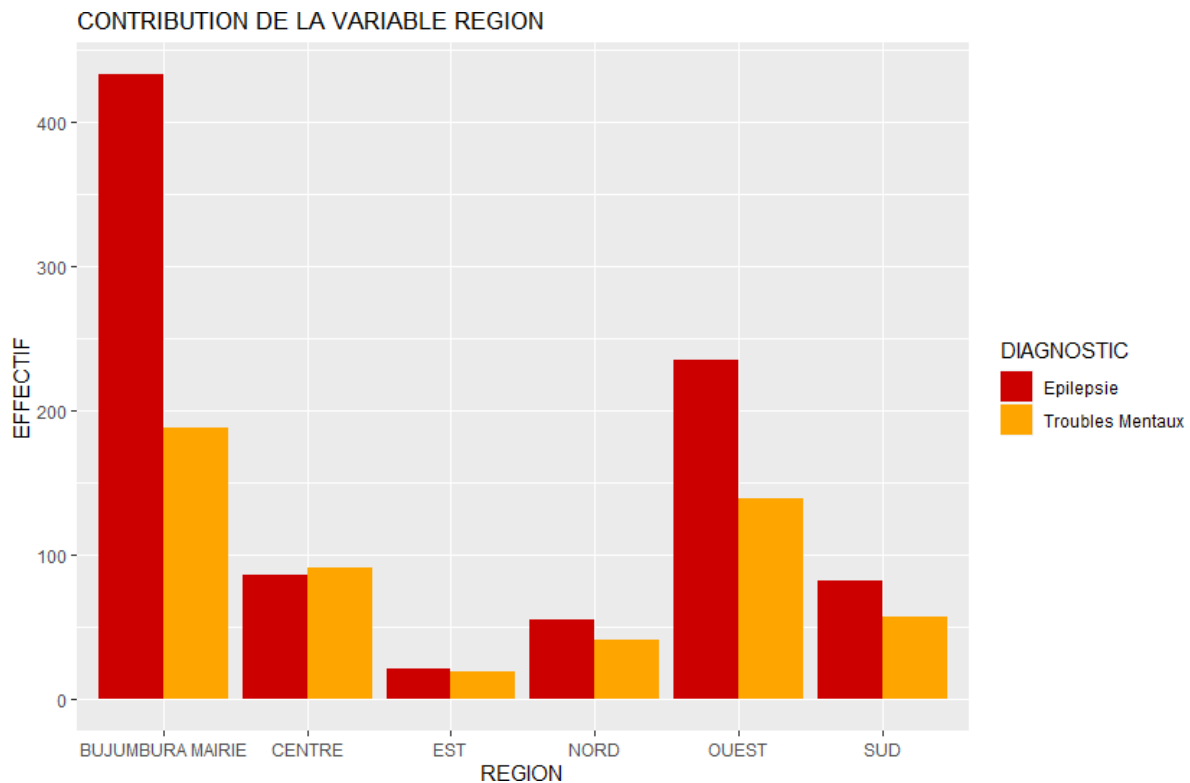


Figure 5 : Contribution de la variable Région

Selon ce graphique, dans la région de Bujumbura Mairie, se sont fait remarquer non seulement plus de cas d'Epilepsie mais aussi plus de cas de Troubles Mentaux. Ce constat prend sa source dans le fait qu'en plus de sa densité élevée, les techniques de dépistage et de traitement de ces maladies sont plus accessibles grâce à la proximité d'un centre spécialisé dans ce domaine: le CNPK.

C'est à l'EST du pays que se trouvent le moins de patients diagnostiqués à l'Epilepsie et aux Troubles Mentaux. A BUJUMBURA MAIRIE, 433 cas soit 47,47% contre 21 cas seulement à l'EST soit 2,3% sur un total de 912 patients épileptiques ont été enregistré. Mais encore, sur 535 cas de Troubles Mentaux, 188 cas soit 35,14% étaient présents à BUJUMBURA MAIRIE contre 19 cas soit 3,55% à l'EST du pays.

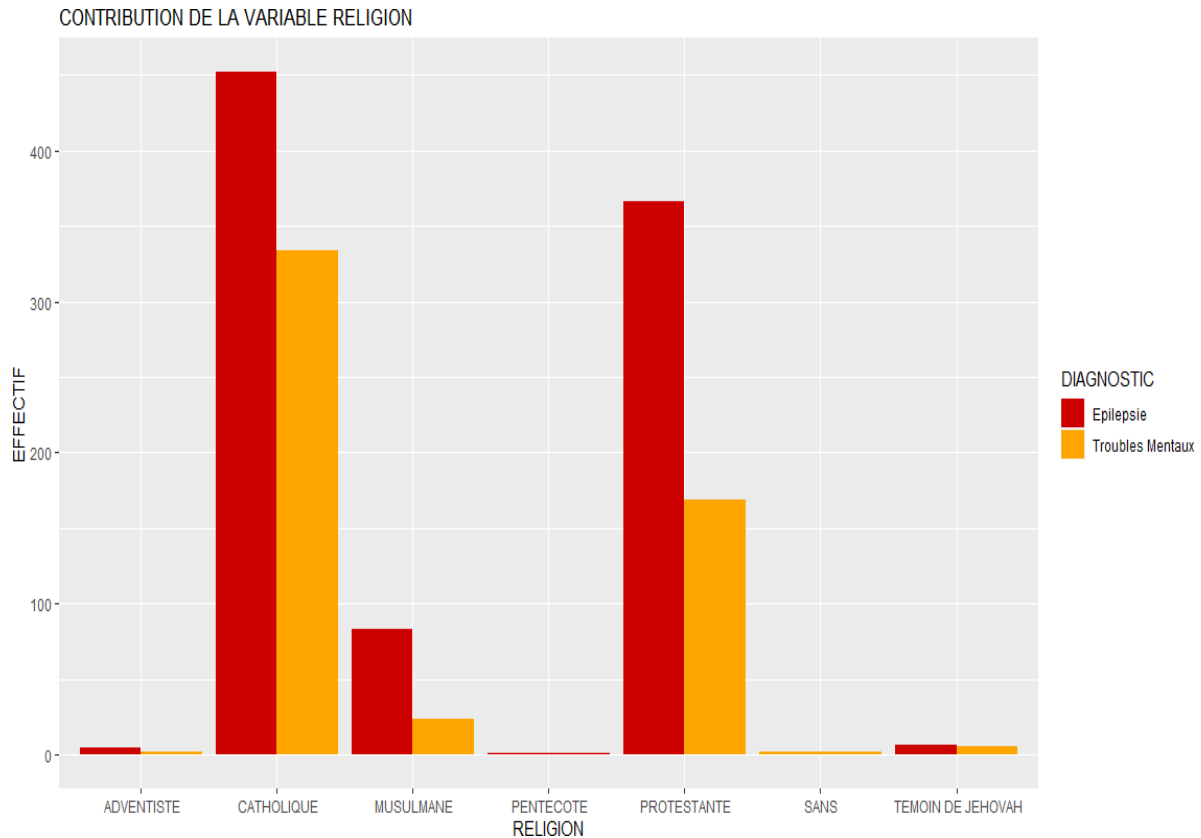


Figure 6 : Contribution de la variable Religion

Cette représentation montre que sur le point de vue de la Religion, les patients Catholiques et Protestants sont les plus diagnostiqués d'Epilepsie et de Troubles Mentaux. En revanche, peu de malades ont été enregistrés chez les Adventistes et les Témoins de Jéhovah. 452 cas soit 49,56% chez les Catholiques contre 6 cas soit 0,65% chez les Témoins de Jéhovah sur un total de 912 cas d'Epilepsie et 334 cas soit 62,42% chez les Catholiques contre 5 cas de Troubles Mentaux chez les Témoins de Jéhovah soit 0,93 %.

III.2. Analyse Inférentielle

III.2.1. Analyse primaire des données

III.2.1.1. Tableaux de contingence

Le tableau de contingence ou tableau croisé est un moyen de description des variables qui permet de mettre en relation deux variables et de comparer les différentes modalités de ces variables.

DIAGNOSTIC SEXE	Epilepsie	Troubles Mentaux
F	0.609	0.391
M	0.645	0.355

Tableau 2 : Tableau de contingence entre sexe et diagnostic

Le tableau de contingence entre les deux variables SEXE et DIAGNOSTIC montre que l'Epilepsie est plus diagnostiquée que les Troubles Mentaux chez les Hommes et les Femmes avec un pourcentage de 64.5% de patients diagnostiqués épileptiques contre 35.5% de Troubles Mentaux chez les Hommes tandis que chez les Femmes seulement 39.1% de patients diagnostiquées de Troubles Mentaux.

DIAGNOSTIC AGE	Epilepsie	Troubles Mentaux
0->9ans	1.000	0.000
10->19	0.947	0.053
20->49	0.399	0.600
50+	0.215	0.784

Tableau 3 : Tableau de contingence entre âge et diagnostic

Le tableau ci-dessus met en relation les variables AGE et DIAGNOSTIC et nous montre que les sujets âgés entre 10 et 19 ans sont plus diagnostiqués d'Epilepsie que de Troubles Mentaux soit 94.7% d'épileptiques contre 5.3% de cas de Troubles Mentaux tandis que pour les patients

âgés de 50 ans et plus sont plus diagnostiqués de Troubles Mentaux que d'Epilepsie.

Par ailleurs, les sujets âgés de 0 à 9 mois présents dans la population étudiée ne sont diagnostiqués que d'Epilepsie parce qu'à cet âge il est difficile voire même impossibles de repérer des signes pouvant mener au diagnostic des troubles mentaux.

DIAGNOSTIC REGION	Epilepsie	Troubles Mentaux
BUJUMBURA MAIRIE	0.697	0.303
CENTRE	0.486	0.514
EST	0.525	0.475
NORD	0.573	0.427
OUEST	0.629	0.372
SUD	0.590	0.410

Tableau 4 : Tableau de contingence entre région et diagnostic

La mise en relation entre les variables "REGION" et "DIAGNOSTIC" montre une forte présence d'Epilepsie à Bujumbura Maire soit 69.7% de cas enregistrés cela étant probablement dû à sa forte densité. Les patients venant du Centre du Pays sont plus diagnostiqués de Troubles Mentaux avec un pourcentage de 51.4% contre 48.6% pour l'Epilepsie

DIAGNOSTIC	Epilepsie	Troubles Mentaux
RELIGION		
ADVENTISTE	0.667	0.333
CATHOLIQUE	0.575	0.425
ISLAM	0.783	0.217
PENTECOTE	1.000	0.000
PROTESTANT	0.684	0.316
SANS	0.000	1.000
TEMOIN DE JEHOVAH	0.545	0.455

Tableau 5: Tableau de contingence entre Religion et diagnostic

Le tableau de contingence entre les variables "RELIGION" et "DIAGNOSTIC" montre une forte présence d'Epilepsie chez les patients de religion "Pentecôte" que les Troubles Mentaux. Par ailleurs, les patients "Sans Religion" sont les plus diagnostiqués de Troubles Mentaux avec un taux de 100% et avec le taux de Diagnostic d'Epilepsie le plus faible 0%.

III.2.1.2. Facteurs influençant les maladies mentales

Test de Chi-2

Facteurs	X ²	Ddl	p. value
Sexe	1.7949	1	0.1803
Age	594.59	3	$<2.2e^{-16}$
Région	32.031	5	$5.859e^{-06}$
Religion	31.918	6	$1.692 e^{-05}$

Tableau 6 : Tableau des facteurs influençant les maladies mentales

Le test du khi-deux a permis de vérifier si les facteurs sont significativement associés à l'Epilepsie et aux Troubles Mentaux. Au seuil de 5%, l'Age ($p < 2.2e^{-16}$), La Région ($5.859e^{-06}$) et La Religion ($1.692 e^{-05}$) sont significativement liés au Diagnostic. En revanche, avec une

p. value=0.1803 ; le Sexe n'est pas significativement lié au Diagnostic. Le test de Khi-deux n'étant pas concluant pour les variables Age et Religion, il a fallu faire recours au test d'indépendance de Fisher.

Test de Fisher

Le test d'indépendance de Fisher aboutit aux mêmes conclusions que le test de Khi-deux pour les facteurs Age, Région et Religion. L'Age ($p=0.0004998$), La Région ($p=0.0004998$) et la Religion ($p=0.0004998$) au seuil de 5% sont significativement liés au Diagnostic. Et on a $p=0.1653$ au seuil de 5% pour le Sexe qui n'est pas significativement lié au Diagnostic.

H_0 : Les données sont indépendantes

H_1 : Les données sont corrélées (liées)

$p\text{-value} \leq 5\%$: On rejette

Facteurs	P-value
Age	0.0004998
Région	0.0004998
Religion	0.0004998
Sexe	0.1653

Tableau 7 : Test de Fisher

III.2.2. Estimation des paramètres du modèle

III.2.2.1. Régression multiple

Facteurs	Modalité	Coefficient
Intercept		-20.3230
Sexe	Masculin	-0.2525
AGE	10-19ans	16.6711
	20-49ans	20.0299
	50+	20.8392
Région	NORD	0.8630
	SUD	0.5360
	EST	1.4234
	OUEST	0.1917
	CENTRE	0.6860
Religion	ADVENTISTE	17.8228
	CATHOLIQUE	0.8099
	PROTESTANTE	0.3254
	PENTECÔTE	0.2209
	MUSULMANE	0.1606
	TEMOINS DE JEHOVAH	0.6542
	SANS	19.7660

Tableau 8: Tableau de la régression multiple

Afin de pouvoir aboutir à des résultats concluants on peut passer par les Odds Ratios.

	OR	2.5 %	97.5 %	p
Intercept	1.4922e-09	6.7403e-146	0	0.970098
Sexe M	7.7684e-01	5.7733e-01	1.0429e+00	0.093890
AGE: 10->19	1.7385e+07	2.3798e+183	1.2026e+109	0.975454
AGE: 20->49	4.9990e+08	6.6544e+161	2.2056e+121	0.970529
AGE: 50+	1.1230e+09	3.6806e+159	1.4409e+125	0.969339
REGION CENTRE	1.9857e+00	1.2551e+00	3.1887e+00	0.003858
REGION NORD	2.3703e+00	1.2596e+00	4.6380e+00	0.009166
REGION EST	4.1514e+00	1.5687e+00	1.2056e+01	0.006115
REGION SUD	1.7092e+00	1.0380e+00	2.8531e+00	0.037276
REGION OUEST	1.2114e+00	8.4891e-01	1.7318e+00	0.291396
RELIGION CATHOLIQUE	2.2477e+00	2.8036e-01	1.9625e+01	0.427089
RELIGION MUSULMANE	1.1743e+00	1.3573e-01	1.0986e+01	0.879905
RELIGION PENTECOTE	1.2472e+00	NA	NA	0.999984
RELIGION PROTESTANTE	1.3846e+00	1.7204e-01	1.2129e+01	0.750180
SANS RELIGION	3.8394e+08	0	NA	0.997924
RELIGION TEMOIN DE JEHOVAH	1.9237e+00	1.6267e-01	2.5013e+01	0.598590

Tableau 9 : Tableau des odds ratios

Le tableau ci-dessus montre les odds ratios des différentes variables qui nous permettent de voir la sensibilité des modalités des différentes variables dans la variable à expliquer. Cette sensibilité se mesure en prenant une modalité qui servira de référence dans la comparaison de ces modalités. Par exemple dans le cas de la variable "Sexe", c'est la modalité "Femme qui servi

de référence. Dans l'interprétation des valeurs de odds ratios, on dira que les individus de sexe Masculin ont 7.7684e-01 fois plus de risque d'avoir un Diagnostic positif que les individus de sexe Féminin.

Afin de vérifier la significativité globale du modèle en testant la nullité simultanée de tous les coefficients relatifs aux variables explicatives dans le modèle, on peut également utiliser le test de Wald. Avec H_0 : le coefficient n'est pas significativement différent de 0, il sera rejeté si et seulement si la p-valeur est inférieur ou égal à 5%. Dans ce cas la p-valeur est égal à $0.046 < 5\%$ ce qui signifie que les coefficients du modèle significativement différents de 0.

III.2.2.2. Sélection de variables

AIC	Modèle sélectionné
1237,34	Sexe + Age + Region +Religion
1208	Sexe + Age + Region
1195,68	Sexe +Age
1190,52	Age

Tableau 10:Tableau de Sélection des AIC

Step: AIC=1190.52

Le modèle final retenu est celui ayant comme variable indépendante Age et comme variable dépendante Diagnostic, car il a le plus bas AIC (1190.5) parmi les modèles considérés.

Cela suggère que, selon le critère d'AIC, le modèle le plus parcimonieux (simple) tout en maintenant une bonne ajustabilité est celui qui inclut uniquement la variable AGE. Les autres variables (SEXE, REGION, RELIGION) ont été retirées lors du processus de sélection.

III.2.2.3. Critère de multicollinéarité

	GVIF	Df	$GVIF^{1/(2*Df)}$
SEXE	1.021111	1	1.010500
AGE	1.041797	3	1.006848
REGION	1.116784	5	1.011107
RELIGION	1.093593	6	1.007484

Tableau 11: Tableau des FIV

Les valeurs de FIV sont proches de 1, et sont inférieurs à 2.5. Il n'y'a pas de multicollinéarité entre les variables de prédiction.

III.2.3. Evaluation de la performance du modèle

❖ Courbe ROC et Critère d'AUC

La courbe ROC et le critère d'AUC permettent d'analyser la performance du modèle de prédiction à travers 2 indicateurs: la sensibilité qui est la capacité du modèle à détecter les malades ainsi que la spécificité qui est la capacité du modèle à détecter les non malades.

Cette méthode consiste à mesurer l'aire sous la courbe de Roc (AUC) qui indiquera à son tour la capacité du modèle à reconnaître les vrais malades.

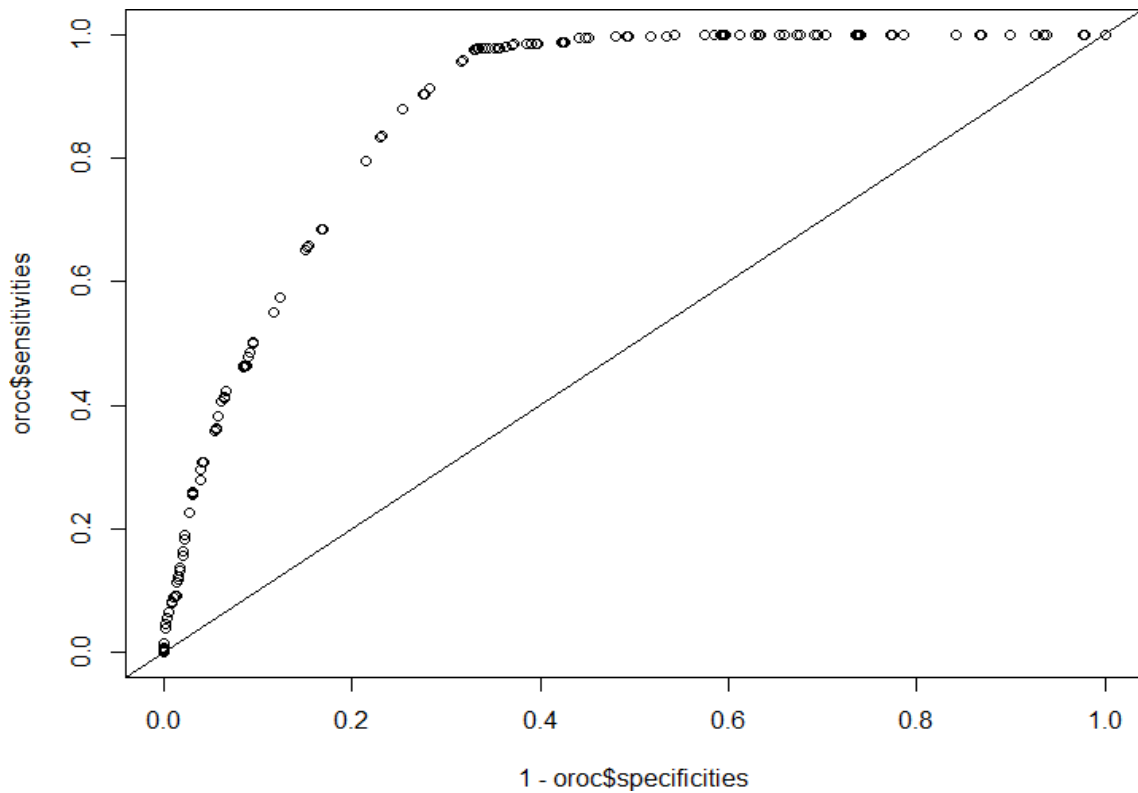


Figure 7 : Courbe de ROC

La courbe ROC permet d'évaluer la performance du modèle à partir du critère de l'AUC. En effet, en mesurant la surface sous la courbe ROC, on évalue le modèle. Avec un AUC= 0.8782 soit 87.82%, la discrimination est excellente.

III.2.4. Validation du modèle

III.2.4.1. Test de rapport de vraisemblance

	Df	LogLik	Df	Chisq	Pr(>Chisq)
1	4	-580.70			
2	16	-560.45	12	40.504	5.928e-05

Tableau 12:Tableau du test de rapport de vraisemblance 1

-L'hypothèse nulle (H_0) dans le test du rapport de vraisemblance est que le modèle réduit est suffisant, c'est-à-dire que la variable "AGE" seule est suffisante pour expliquer la variabilité de la variable de réponse "DIAGNOSTIC".

-Une p-value faible (généralement $< 0,05$) suggère que le modèle complet est statistiquement significativement meilleur que le modèle nul, et notre p-value $5.928e-05 < 0,05$ montre bien que notre modèle est meilleur que le modèle nul. Nous pouvons rejeter l'hypothèse nulle. Ainsi, cela prouve que l'inclusion des variables "SEXE", "AGE", "REGION", et "RELIGION" dans le modèle complet apporte une amélioration significative par rapport au modèle réduit qui n'inclut que la variable "AGE".

	Df	LogLik	Df	Chisq	Pr(>Chisq)
1	15	-561.86			
2	16	-560.45	1	2.8222	0.09297

Tableau 13 : Tableau de rapport vraisemblance 2

-L'hypothèse nulle (H_0) dans le test du rapport de vraisemblance est que le modèle réduit est suffisant, c'est-à-dire que les variables "REGION", "AGE", et "RELIGION" seules sont suffisantes pour expliquer la variabilité de la variable de réponse "DIAGNOSTIC".

-Le p-value est de 0.09297, ce qui est supérieur à un niveau de signification de 0.05. Par conséquent, à un niveau de signification de 0.05, on ne rejette pas l'hypothèse nulle. Cela suggère que le modèle réduit n'est pas significativement moins bon que le modèle complet en termes d'ajustement aux données.

Autrement dit, les résultats ne fournissent pas suffisamment de preuves pour rejeter l'hypothèse nulle, indiquant que le modèle réduit avec les variables "REGION", "AGE", et "RELIGION" pourrait être adéquat pour expliquer la variabilité dans la variable de réponse "DIAGNOSTIC".

En conclusion, l'inclusion de la variable SEXE dans le modèle n'apporte pas une amélioration significative en termes d'explication de la variabilité de la variable de réponse "DIAGNOSTIC".

III.2.5. Prédiction du modèle

Individus	Age	Sexe	Région	Religion	Diagnostic	Probabilités
1	20 -> 49	Masculin	Bujumbura-Mairie	Protestante	Troubles mentaux	5.376904e-01
2	20 -> 49	Femme	Bujumbura-Mairie	Protestante	Troubles mentaux	5.656920e-01
3	20 -> 49	Masculin	Bujumbura-Mairie	Catholique	Troubles mentaux	8.992732e-01
4	20 -> 49	Femme	Bujumbura-Mairie	Catholique	Troubles mentaux	7.690978e-01
5	0 -> 9	Homme	Bujumbura-Mairie	Catholique	Epilepsie	6.700834e-01
6	0 -> 9	Femme	Bujumbura-Mairie	Protestante	Epilepsie	5.173828e-09
7	10 -> 19	Homme	Bujumbura-Mairie	Protestante	Epilepsie	3.467305e-02
8	10 -> 19	Femme	Bujumbura-Mairie	Catholique	Epilepsie	1.214313e-01

Tableau 14 : Tableau prédictive

Ce tableau montre la probabilité qu'a un individu d'être diagnostiqué de troubles mentaux ou d'épilepsie dépendamment de ses caractéristiques à un seuil de 5%.

- Individu 1 : un homme ayant un âge compris entre 20 et 49 ans, né à Bujumbura-Mairie, étant de religion catholique, a une probabilité 0,899 d' être diagnostiqué de troubles mentaux. Le modèle estime une probabilité d'environ 89,9% que cet individu souffre de troubles mentaux.
- Individu 8 : une femme ayant un âge compris entre 10-19 ans , né à Bujumbura-Mairie, étant de religion catholique, a une probabilité 0,121 d' être diagnostiqué d'épilepsie. Le modèle estime une probabilité d'environ 12,1% que cet individu ayant ces caractéristiques souffre d'épilepsie.

CHAPITRE IV : CORRELATION ENTRE LES ELEMENTS CONSTITUTIFS DES UNITÉS D'ENSEIGNEMENT ET LE STAGE

Certains éléments des unités d'enseignement ont été très utiles que les autres d'autant plus qu'ils nous ont permis d'effectuer les tâches que nous avons réalisées avec facilité. C'est notamment le cours d'initiation à l'informatique, initiation à la recherche, le cours de calcul statistique sur ordinateur(R), Statistique descriptive, Analyse et traitement de données(SPSS), Organisation d'une enquête, Modèle linéaire, ...

IV.1. Initiation à l'informatique

Ce cours nous a été fortement utile dans la mesure où c'est notamment grâce à ce dernier que nous avons pu être à l'aise avec la manipulation de l'outil indispensable qu'est l'ordinateur lors de notre stage et même après pour la présentation et la rédaction des résultats de notre travail.

IV.2. Initiation à la recherche

L'initiation à la recherche est un cours qui nous a été très utile à mesure que, c'est grâce à ce dernier, que nous avons pu réfléchir sur comment organiser notre rédaction d'une manière scientifique.

IV.3. Calcul statistique sur ordinateur(R)

L'objectif de ce cours est de, manipuler les données, calculer et faire une présentation des graphiques. Ce cours est centré sur le logiciel R, résultant d'un projet de professeurs de l'université d'Auckland en Nouvelle-Zélande (Ross IHAKA et Robert GENTLEMAN, 1993) présentant un environnement permettant de traiter et d'analyser des données. Il nous a été utile pour notamment la présentation des différentes boîtes à moustaches contenu dans notre document.

IV.4. Statistique descriptive

La statistique descriptive nous a beaucoup aidés dans notre travail. L'objectif de la statistique descriptive étant de décrire les données récoltées pour optique d'analyse : grâce à elle nous avons pu faire une bonne description de nos données.

IV.5. Traitement de données d'enquête

Le cours d'analyse et traitement de données nous a été fortement utile, car c'est sur base des notions apprises en classe que nous avons pu faire un nettoyage de notre base de données, calculer des variables à partir d'autres déjà existante et effectuer une analyse introspective.

IV.6. Organisation d'une enquête

Le cours d'organisation d'une enquête nous a beaucoup aidés dans notre travail dans la mesure où c'est grâce aux notions apprises lors de ce cours que nous avons pu comprendre la méthodologie utilisée pour la collecte de données que nous avons utilisées.

IV.7. Modèle linéaire

Le cours de Modèle linéaire nous a beaucoup aidés dans notre travail dans la mesure où, c'est grâce aux notions apprises lors de ce cours que nous avons pu, sur base de certaines hypothèses générées un modèle d'analyse de nos données.

IV.8. Statistique sectorielle

Le cours de statistique sectorielle nous a bien aidés dans le domaine agricole. Il a été vital compte tenu du milieu de travail et aussi de son contenu surtout dans son Module de la Statistique Agricole.

IV.9. Théorie et Méthode Statistique

C'est un cours qui nous a fortement été très utile dans la compréhension du fonctionnement des différents tests statistiques que nous avons eu à utiliser, ainsi la compréhension des notions de généralisation des estimations faites sur base d'un échantillon sur sa population d'origine.

IV.10. Introduction au logiciel de programmation R

Le logiciel R est un langage de programmation qui a pour objet le traitement des données et d'analyse statistique. Ce cours a été d'un apport très agréable dans l'analyse et le traitement de nos données. Aussi, il a été d'une importance capitale dans toutes les analyses effectuées et dans la production des résultats.

CONCLUSION GÉNÉRALE

Cette étude avait pour objectif d'identifier les facteurs influençant les pathologies mentales des patients consultés à la C.N.P.K afin d'orienter les politiques et actions du gouvernement et des intervenants en matière de santé et de sensibiliser la population sur ces maladies.

L'analyse a porté sur les patients hospitalisés (tout âge compris) à travers les données recueillies au sein de la C.N.P.K au service de la SIS durant l'année 2020.

Ce travail avait pour objectif principal de pouvoir prédire le diagnostic d'un patient entre l'Epilepsie et les Troubles Mentaux à travers différentes techniques apprises en Statistique.

Un modèle a été ainsi élaboré.

Afin de pouvoir proposer le meilleur modèle, divers analyses ont été réalisées notamment la mesure de l'influence qu'a chaque variable explicative sur la variable à expliquer ainsi que des tests qui permettent le modèle proposé.

Les résultats des analyses ont permis donc de conclure que l'Age, la Région et la Religion influencent le Diagnostic contrairement au Sexe dont on peut s'en passer.

L'analyse du modèle estimé a donné un AUC égal à 87.82%.

Limites de l'étude

Quelques difficultés ont été relevés lors de ce travail tel que :

- Le manque de renseignements suffisant sur les malades limitant les analyses à très peu de variables
- La récente informatisation du milieu sanitaire qui a limité le champ des données.

Recommandations

De par les résultats de cette analyse, nous avons formulé les recommandations suivantes aux responsables de la santé :

- Conseiller les malades et non-malades à s'occuper de leur santé mentale ;
- Augmenter le budget alloué à la sensibilisation des maladies mentales ;
- Enrichir leur base de données en demandant plus de précision aux malades pour faciliter ceux qui travailleront sur leur base de données à l'avenir.

REFERENCES

Ouvrages généraux

- [1] Institut du cerveau, Pourquoi chercher à comprendre le cerveau?, Fondation reconnue d'utilité publique par décret en date du 13/09/2006 .
- [2] Le Manuel MSD Version pour le grand public présenté par Merck & Co, Inc., Rahway, NJ, États-Unis (connu sous le nom de MSD en dehors des États-Unis et du Canada)
- [3] Institut du cerveau, Fondation reconnue d'utilité publique par décret en date du 13/09/2006 , Anatomie du cerveau : Comment fonctionne le cerveau humain?
- [4] Organisation Mondiale de la Santé 2015, est une agence spécialisée de l'Organisation des Nations unies pour la santé publique créée en 1948, Principaux repères sur l'Epilepsie, publié le 08 juin 2022.
- [5] Organisation Mondiale de la Santé 2015, est une agence spécialisée de l'Organisation des Nations unies pour la santé publique créée en 1948, Troubles mentaux, publié le 08 juin 2022.
- [6] Centre Neuro-Psychiatrique de Kamenge, Association Sans But Lucratif spécialisée dans le domaine de la psychiatrie.
- [7] Pascal Legrand, Denis Bories, Le choix des variables explicatives dans les modèles de régression logistique , CER Groupe ESC Clermont – CRCGM 4 boulevard Trudaine 63 100 Clermont-Ferrand. Page 4
- [8] Pascal Legrand, Denis Bories, Le choix des variables explicatives dans les modèles de régression logistique de Pascal Legrand et de Denis Borie, CER Groupe ESC Clermont – CRCGM 4 boulevard Trudaine 63 100 Clermont-Ferrand.
- [9] Pascal Legrand, Denis Bories, Le choix des variables explicatives dans les modèles de régression logistique de Pascal Legrand et de Denis Bories , CER Groupe ESC Clermont – CRCGM 4 boulevard Trudaine 63 100 Clermont-Ferrand.
- [10] Pascal Legrand, Denis Bories, Le choix des variables explicatives dans les modèles de régression logistique de Pascal Legrand et de Denis Bories , CER Groupe ESC Clermont – CRCGM 4 boulevard Trudaine 63 100 Clermont-Ferrand. Page 7

- [11] F .DUYME et J.J CLAUSTRIAUX , La Régression Logistique Binaire,Faculté universitaire des sciences agronomiques,Unité de Statistique, Informatique et Mathématique appliquées GEMBLOUX,Page 10.
- [12] Ricco Rakotomalala, Pratique de la Régression Logistique : Régression Logistique Binaire et Polytomique, Version 2.0, Page 28
- [13] Laurent Rouvière , Analyse du modèle de Régression logistique, JANVIER 2015,
- [14] Julie Desjardins, L'analyse de Régression Logistique, Tutorial in Quantitative Methods for Psychology 2005, Vol. 1(1), p. 35-41.
- [15] Pascal Legrand, Denis Bories, Le choix des variables explicatives dans les modèles de régression logistique.
- [16] : Guide d'intervention mhGAP.2011.p.42, Prise en charge clinique des troubles mentaux, neurologiques et liés à l'utilisation de substances psychoactives dans les situations d'urgence humanitaire, Catalogue à la source : Bibliothèque de l'OMS.
- [17] : Minitab, Qu' est-ce-que le test de rapport de vraisemblance ?, 2023 Minitab, LLC
- [18] : Bookdown.org, Test de rapport de vraisemblance, Le site Web bookdown.org est un service fourni par RStudio, PBC pour héberger des livres
- [19] : Minitab, Qu' est-ce-que la multicollinéarité ?, 2023 Minitab, LLC
- [20] : Hélène Hamisultane, ECONOMETRIE. Licence. France. 2002. ffccl-01261163

Site Web

1. <https://institutducerveau-icm.org> Consulté le 22 Décembre 2022 à 18h58,
2. <https://www.msmanuals.com> Consulté le 22 Décembre 2022 à 19h40,
3. <https://institutducerveau-icm.org> Consulté le 22 Décembre 2022 à 19h18,
4. <https://www.who.int> (site de l'OMS) Consulté le 22 Décembre 2022 à 19h44,
5. <https://www.who.int> (site de l'OMS) Consulté le 22 Décembre 2022 à 19h51,
6. <https://www.cnpk.bi/index.php/qui-sommes-nous/organigramme>
7. <https://www.actuia.com> Consulté le 08 Mai 2023 à 13h31

8. <https://support.minitab.com/fr-fr/minitab/20/help-and-how-to/quality-and-process-improvement/capability-analysis/supporting-topics/distributions-and-transformations-for-nonnormal-data/what-is-the-likelihood-ratio-test/>
9. <https://support.minitab.com/fr-fr/minitab/21/help-and-how-to/statistical-modeling/regression/supporting-topics/model-assumptions/multicollinearity-in-regression/>
10. <https://bookdown.org/boucherjphilippe/MAT998H/mod%C3%A8les-li%C3%A9s.html>

ANNEXES

Organigramme de l'ASBL-CNPK

