



UNIVERSITÉ D'ABOMEY-CALAVI



ÉCOLE NATIONALE D'ÉCONOMIE APPLIQUÉE ET DE MANAGEMENT

DÉPARTEMENT DE LA STATISTIQUE

Mémoire de fin de formation en vue de l'obtention du diplôme de Licence

Formation : Statistique

Option : Statistiques Économique et Sectorielle (SES)

Facteurs explicatifs de l'hypertension artérielle chez les personnes âgées de 30 à 64 ans au Bénin

Rédigé par :

Conrad S. AGOSSOU

&

Espérancia A. D. AYELO

Directeur de mémoire :

Dr M. Nicodème ATCHADE,

Statisticien Economètre,

Enseignant Chercheur à l' UNSTIM

Maître Assistant CAMES

Directeur de stage :

M. Elisée DAHOU,

Ingénieur Statisticien Economiste,

Responsable de la Section Etudes et Statistiques

à la CNSS

40ème promotion

L'Ecole Nationale d'Economie Appliquée et de Management (ENEAM) n'entend donner ni approbation ni improbation aux opinions émises dans ce document. Les raisonnements et arguments reviennent donc aux auteurs.

Approbation

*Nous certifions que le présent mémoire a été réalisé par ses auteurs et exempt de tout plagiat.
Il est arrivé à terme et peut être soutenu devant un jury.*

Cotonou le.....

Signatures

Directeur de mémoire :

Dr M. Nicodème ATCHADE

Directeur de stage :

M. Elisée DAHOU

Dédicace

Nous dédions ce travail à nos chers parents tout en les remerciant pour leurs sacrifices et leurs clairvoyances qui nous serviront tout au long de notre vie, à nos frères et sœurs pour leur soutien et leur affection, tout en leur souhaitant la réussite dans tout ce qu'ils entreprennent. A tous les membres des familles AGOSSOU et AYELO. A toute personne honorant ce mémoire en l'utilisant honnêtement aux fins sacrées de la recherche scientifique tout en lui souhaitant bonne réussite.

Conrad AGOSSOU et Espérancia AYELO

Remerciements

L'aboutissement de ce travail n'a été possible que grâce à une assistance soutenue, une collaboration franche, permanente et effective de nombreuses personnes. Nous voulons remercier :

- ✍ Monsieur Albert HONLONKOU, Ph.D., Maître de conférences CAMES, Professeur Agrégé des sciences économiques, Directeur de l'ENEAM ;
- ✍ Monsieur Théophile K. DAGBA, Ph.D., Maître de Conférences, Directeur-adjoint, Chargé des affaires académiques de l'ENEAM ;
- ✍ Professeur Rosaline D. WOROU HOUNDEKON, ancienne Directrice de l'Ecole Nationale d'Economie Appliquée et de Management (ENEAM) ;
- ✍ Dr M. Nicodème ATCHADE, Statisticien Economètre, Enseignant-chercheur à l'Ecole Nationale Supérieure de Génie Mathématique et Modélisation de l'UNSTIM, notre Directeur de mémoire, pour sa patience, ses observations, ses orientations et ses conseils ;
- ✍ Dr Barthélémy SENOU, Chef du département Statistique de l'Ecole Nationale d'Economie Appliquée et de Management (ENEAM) ;
- ✍ Tous les enseignants de l'ENEAM pour leurs enseignements et leurs conseils ;
- ✍ Tout le personnel de la CEC et de la CNSS pour leur accueil et leur conseils ;
- ✍ Les membres du jury qui ont accepté de consacrer leur précieux temps pour l'appréciation de ce travail ;
- ✍ Tous nos camarades de la 40ème promotion de la filière Statistique de l'ENEAM ;
- ✍ Toutes les personnes qui directement ou indirectement ont contribué à la réussite de ce travail.

Liste des sigles et acronymes

AIC	Akaïke Information Criterion
AME	Average Marginal Effect
ANOVA	Analysis Of Variance
BIC	Bayesian Information Criterion
CCPF	Caisse de Compensation de Prestations Familiales
CCPFAT	Caisse de Compensation des Prestations Familiales et Accidents de Travail
CDSS	Caisse Dahoméenne de Sécurité Sociale
CNAF	Caisse Nationale d'Allocation Familiale
CNSS	Caisse Nationale de Sécurité Sociale
DTS	Diplôme de Technicien Supérieur
ENEAM	Ecole Nationale d'Economie Appliquée et de Management
FRM	Fondation pour la Recherche Médicale
HTA	Hypertension artérielle
IMC	Indice de Masse Corporelle
INStaD	Institut National de la Statistique et de la Démographie
IPRAO	Institut de Prévoyance et de Retraite de l'Afrique Occidentale
JAMA	Journal of the American Medical Association
MCV	Maladies cadiovasculaires
MFTP	Ministère du Travail et de la Fonction Publique
mmHg	Millimètre de mercure
MNT	Maladie Non Transmissible
OBSS	Office Béninois de Sécurité Sociale
OMS	Organisation Mondiale de la Santé
PIB	Produit Intérieur Brut
PRFI	Pays à Revenu Faible et Intermédiaire
SNIGS	Système National d'Information et de Gestion Sanitaire
UNSTIM	Université Nationale des Sciences, Technologies, Ingénierie et Mathématiques

Sommaire

Approbation	ii
Dédicace	iii
Remerciements	iv
Liste des acronymes	v
Résumé	ix
Abstract	x
Introduction	1
1 Cadre institutionnel de l'étude	3
2 Cadre théorique et méthodologique de l'étude	8
3 Résultats, analyses et interprétations	23
Conclusion	36
Références	36
Annexe	a

Liste des tableaux

2.1 Variables explicatives de l'étude	21
3.1 Résultats de l'analyse descriptive bivariée	24
3.2 Tests de Khi deux	26
3.3 Résultats des tests du choix de modèle	27
3.4 Résultats des tests du choix de la meilleure combinaison des variables	27
3.5 Résultats de l'estimation du probit	28
3.6 Pertinence du modèle probit	29
3.7 Qualité du modèle	29
3.8 Effets marginaux	29
3.9 Tests du choix de modèle	32
3.10 Résultats de l'estimation du probit selon le genre	32
3.11 Pertinence du modèle probit	34
3.12 Qualité du modèle	34
3.13 Vérification des hypothèses	35

Table des figures

2.1 Organigramme de la modélisation	22
3.1 Prévalence de l'hypertension artérielle	23
3.2 Répartition de la population selon l'âge	23
3.3 Répartition de la population selon le genre	24

Résumé

Cette étude a pour but de déterminer les facteurs explicatifs de l'hypertension artérielle (HTA) chez les personnes âgées de 30 à 64 ans au Bénin. Elle se fonde sur les données issues de l'Enquête Démographique de santé 2017-2018 (EDSB-V) réalisée par l'INStaD(ex INSAE). A cet effet, une modélisation économétrique (modèle Probit) a été effectuée grâce aux logiciels EXCEL et R v4.0.5 après l'analyse descriptive (univariée et bivariée) des variables considérées étayée par l'estimation des effets marginaux.

De l'analyse descriptive, il ressort que la prévalence de l'hypertension artérielle dans la population d'étude est de 17,90%. Des résultats obtenus de l'analyse économétrique, il ressort que l'âge, le genre, la région, le niveau d'instruction, le niveau de vie et l'état diabétique sont les facteurs qui influencent la prévalence de l'hypertension artérielle. De façon spécifique, il ressort que l'âge, la région, le niveau de vie et le statut diabétique ont un effet significatif sur l'HTA au seuil de 5% et se révèlent donc être des facteurs explicatifs de l'HTA que ce soit chez les hommes ou chez les femmes. Par contre, on remarque que l'instruction et l'état matrimonial se trouvent être un facteur explicatif de l'HTA au niveau des hommes uniquement. Les effets marginaux nous ont permis d'étayer nos analyses en nous donnant la probabilité pour chaque catégorie de personnes d'avoir l'hypertension artérielle.

Au regard de ces résultats, nous suggérons donc aux autorités sanitaires d'éduquer et de sensibiliser les populations sur les maladies non transmissibles et aux facteurs à risque de l'hypertension artérielle.

Abstract

The aims of this study are to determine the explanatory factors of high blood pressure among people aged 30 to 64 years in Benin. It is based on data from the 2017-2018 Demographic Health Survey (EDSB-V) conducted by INStaD(ex INSAE). For this purpose, an econometric modeling was performed using EXCEL and R v4.0.5 software after descriptive analysis (univariate and bivariate) of the variables considered, supported by the estimation of marginal effects.

The descriptive analysis showed that the prevalence of hypertension in the study population was 17.90%. The results of the econometric analysis showed that age, gender, region, education level, standard of living, and diabetes status were the factors that influenced the prevalence of hypertension. Specifically, age, region, standard of living, and diabetic status were found to have a significant effect on hypertension at the 5% level and were therefore found to be explanatory factors for hypertension in both men and women. On the other hand, education and marital status were found to be explanatory factors for hypertension in men only. The marginal effects allowed us to support our analyses by giving us the probability of each category of persons having hypertension.

In view of these results, we suggest that the health authorities educate and sensitize the population about noncommunicable diseases and the risk factors for hypertension.

Introduction

Les maladies cardiovasculaires sont un problème majeur de santé publique de plus en plus important dans le monde, une tendance qui se produit à un rythme alarmant dans les pays à revenu faible et intermédiaire (PRFI), avec une charge de morbidité en forte augmentation. Selon la FRM, ces maladies regroupent les pathologies qui touchent le cœur et l'ensemble des vaisseaux sanguins, comme l'athérosclérose, les troubles du rythme cardiaque, l'infarctus du myocarde, l'insuffisance cardiaque, les accidents vasculaires cérébraux ou encore l'hypertension artérielle qui est l'objet de notre étude. D'après l'OMS, l'hypertension artérielle (HTA) et l'hypertension clinique sont des facteurs de risque majeurs pour les maladies cardiovasculaires et les maladies rénales chroniques, et une cause majeure de mortalité dans le monde.

Entre 2000 et 2010, la proportion de la population mondiale souffrant d'hypertension artérielle ou d'hypertension non contrôlée avait légèrement diminué de 2,6% dans les pays à revenu élevé (Mills et al., 2016). Cependant, de fortes augmentations dues à la croissance et au vieillissement de la population ont été enregistrées par l'Organisation Mondiale de la Santé (OMS) au cours de la dernière décennie, la plus forte augmentation étant de 30 % en Afrique. La prévalence la plus faible d'hypertension artérielle a été notée dans la région des Amériques, à 18 %, tandis que l'estimation mondiale chez les adultes âgés de 18 ans et plus était d'environ 22 % en 2014 (JAMA, 2003). Ce qui indique que les régions africaines sont touchées dans une proportion très inquiétante.

Au Bénin, malgré la disponibilité des médicaments antihypertenseurs à faible coût, les données suggèrent que moins de la moitié des personnes hypertendues ont été traitées et que moins d'un tiers d'entre elles obtiennent une prise en charge adaptée. Malgré les réformes mises en place pour lutter contre les maladies non transmissibles, notamment par l'installation des Programmes Nationaux de Lutte contre les MNT (PNLMNT) et du Système National d'Information et de Gestion Sanitaire (SNIGS), l'hypertension artérielle est un mal qui demeure et qui persiste. L'hypertension touche 25,9% des adultes béninois selon une enquête nationale menée en 2015 par le ministère de la santé, selon la méthodologie par étapes de l'Organisation mondiale de la santé ; environ 68,2 % n'avaient jamais fait mesurer leur tension artérielle (TA) et plus de la moitié des personnes ayant une TA élevée n'étaient pas au courant de leur statut (STEPS, 2015). Tous ces faits montrent qu'un diagnostic tardif de l'hypertension et un mauvais contrôle de la TA augmentent la fréquence des complications hypertensives.

Face à tous ces constats, l'hypertension artérielle est une maladie qui mérite une attention particulière. C'est à cette préoccupation que la présente étude intitulée « Facteurs explicatifs de l'hypertension artérielle chez les personnes âgées de 30 à 64 ans au Bénin » se donne pour mission de déterminer la prévalence et les facteurs explicatifs de l'hypertension artérielle au Bénin. Le premier chapitre présente le cadre institutionnel de l'étude. Le deuxième chapitre est consacré au cadre théorique et méthodologique de l'étude. Il présente la problématique, les objectifs, les hypothèses de l'étude, la revue de littérature puis la méthodologie utilisée. Le troisième chapitre expose les résultats de l'analyse descriptive et de l'analyse économétrique des données. Ce dernier chapitre se termine

par les préconisations opérationnelles, la conclusion et les références bibliographiques.

1 | Cadre institutionnel de l'étude

La Caisse Nationale de Sécurité Sociale (CNSS) nous a servi de cadre pour le stage pratique qui nous a permis de relever des observations et de dégager le thème de notre étude. Nous présentons alors dans ce chapitre, cette institution à travers ses structures et ses mécanismes de fonctionnement ainsi que les différentes activités menées au cours du stage.

1.1 Environnement du stage

1.1.1 Présentation et historique de la CNSS

Le système de sécurité sociale qui apparaît aujourd'hui n'est pas né en un jour. Il s'est construit sur l'émanation d'un processus et est le résultat d'une maturité historique. Au Bénin, la sécurité sociale s'est faite comme tout phénomène social, qui très souvent, est le résultat d'une lente évolution.

Etablissement public à caractère social, la Caisse Nationale de Sécurité Sociale (CNSS) est dotée d'une personnalité civile et d'une autonomie financière. Placée sous tutelle du Ministère du Travail et de la Fonction Publique (MFTP), elle est gérée par un Directeur Général et administrée par un Conseil d'Administration (CA). Elle exerce une importante action sociale en faveur des travailleurs salariés. Elle mène des actions de prévention des risques professionnels au niveau des entreprises et unités de production. Ainsi, elle procède à l'information, et à la sensibilisation des employeurs et des travailleurs, au contrôle et à la surveillance des conditions de santé et de sécurité en milieu du travail, à des enquêtes destinées à situer les responsabilités et surtout à détecter les causes pour édicter les mesures préventives des accidents de travail et des maladies professionnelles. La CNSS garantit les droits aux travailleurs béninois ayant exercé leurs activités professionnelles à l'étranger et aux expatriés qui ont travaillé au Bénin.

Cette institution a été créée le 26 Janvier 1956 sous la dénomination de Caisse de Compensation de Prestations Familiales (CCPF) gérant la branche des prestations familiales et de maternité instituée au profit des travailleurs salariés en activité. Son financement est assuré par les cotisations des employeurs.

Depuis sa création, l'institution a connu progressivement les mutations ci-après :

- ✍ En 1959, un système de répartition des accidents de travail et des maladies professionnelles a été mis en place avec la première mutation de la dénomination de l'institution qui devient Caisse de Compensation des Prestations Familiales et Accidents de Travail (CCPFAT).
- ✍ En 1970, la branche de retraite précédemment basée à Dakar au Sénégal pour la sous-région et gérée par l'Institut de Prévoyance et de Retraite de l'Afrique Occidentale (IPRAO) a été dé-localisée par ordonnance N°70-17 du 25 mars 1970 afin de prendre en charge au plan national la branche d'assurance Vieillesse, invalidité et décès au profit des travailleurs salariés. Ce fut la deuxième mutation de dénomination soit Caisse Dahoméenne de Sécurité Sociale (CDSS).
- ✍ En 1971, le pouvoir public, sans motif d'ordre technique évident, décide par ordonnance N°71-10 du 19 mars 1971 de scinder la CDSS en (02) deux organismes distincts, ce fut la troisième

mutation à savoir :

- la Caisse Nationale de Sécurité Sociale (CNSS), chargée des branches des pensions et des risques professionnels.
 - la Caisse Nationale d'Allocation Familiale (CNAF), chargée de la branche des prestations familiales.
- ✍ En 1973, le pouvoir public décide à nouveau de la réunification des (02) deux organismes par l'ordonnance N°73-3 du 17 janvier 1973 portant création, organisation et fonctionnement de l'Office Béninois de Sécurité Sociale (OBSS), ordonnance dont les dispositions sont modifiées par l'ordonnance N°76-6 du 26 janvier 1976. Ce fut alors la quatrième mutation de dénomination.

L'Office Béninois de Sécurité Sociale (OBSS) dans sa gestion sans une véritable maîtrise des paramètres de performance et d'équilibre des branches, a connu à partir de l'année 1987 une grave crise financière l'ayant conduit au Programme d'Ajustement Structurel (PAS) où, sur la base d'un plan de redressement assez drastique, les experts du Fonds Monétaire International (FMI) et de la Banque Mondiale exigèrent entre autres conditionnalités, l'autonomie intégrale de l'organisme qui ne devrait plus être pris pour un office de l'Etat avec immixtion permanente du pouvoir public dans sa gestion, d'où la nécessité de nouvelles réformes conduisant à l'adoption d'un Code de Sécurité en lieu et place des différents textes législatifs éparses et propres à chaque branche de prestations gérées.

Ainsi, en 1993 un Code de Sécurité Sociale a été délibéré et adopté par l'Assemblée Nationale en sa séance du 23 Juin 1998 puis en sa séance du 09 Juillet 2002, pour une mise en conformité avec la constitution suite à la décision DC 98-097 du 11 Décembre 1998 et promulgué par le Président de la République en 2003 sous forme de loi 98-019 du 21 mars 2003 portant code de Sécurité Sociale en République du Bénin et comporte 149 articles.

Cette loi consacre la quatrième et la dernière mutation puis, l'institution retrouve désormais la dénomination universelle de Caisse Nationale de Sécurité Sociale (CNSS).

1.1.2 Organisation et missions de la CNSS

✍ Organisation de la CNSS

Elle est administrée par un Conseil d'Administration tripartite (Employeurs, Travailleurs, Etat). Elle est gérée par un Directeur Général assisté d'un Directeur Adjoint. Conformément à la décision n°069/08/CNSS/DG/SP-C du 31 mars 2008, la CNSS comprend une Administration Centrale dont le siège est à Cotonou et des Structures décentralisées que sont les agences régionales.

Administration Centrale

La Caisse est gérée par un Directeur Général assisté d'un Directeur Général Adjoint qui le supplée en cas d'absence ou d'empêchement.

Le système administratif central de la CNSS est composé d'une Direction Générale et de six (06) Directions Techniques. Il s'agit de :

- ✍ **La Direction Générale** Elle est responsable de la mise en œuvre de la politique de la CNSS et de sa gestion. Les organes ci-après lui sont rattachés :

- Le Secrétariat Particulier,
- Un poste d'Attaché du Directeur Général,
- La Cellule du Médecin Conseil,
- La Cellule des Affaires Juridiques,
- La Cellule des Relations Publiques et de la Communication,
- La Cellule Informatique,
- La Cellule des Etudes et de la Coopération,
- Les agences régionales.

Les Directions techniques

Il s'agit de :

— Direction de l'Audit Interne et de l'Inspection (DAII)

Chargée du contrôle à priori et à posteriori de toutes les structures de la Caisse, de la conduite des enquêtes à la suite des malversations ou faits assimilés commis par des tiers ou par des agents de la Caisse et de l'exécution de toutes les tâches ou missions confiés par la Direction Générale.

— Direction des Ressources Humaines (DRH)

Chargée de la gestion des ressources humaines à travers le recrutement, la formation professionnelle et la gestion de la paie du personnel.

— Direction du Budget et du Patrimoine (DBP)

Chargée de l'élaboration et de l'exécution de budget autonome de la Caisse, de l'approvisionnement, du matériel et de la gestion immobilière.

— Direction des Prestations (DP)

Elle s'occupe de la gestion de toutes les branches de prestations, de la carrière des travailleurs affiliés et de la prévention des risques professionnels.

— Direction du Recouvrement (DR)

Elle veille à l'immatriculation des employeurs, à l'affiliation des travailleurs, à la gestion des comptes cotisants, au recouvrement des cotisations, au contrôle des employeurs, ainsi qu'au précontentieux du recouvrement des cotisations.

— Direction Financière et Comptable (DFC)

Elle a pour attribution la tenue de la comptabilité, la gestion de la trésorerie et des placements et le paiement des dépenses techniques et administratives de la Caisse.

Structures décentralisées

La Caisse dispose également des agences régionales installées au chef-lieu des six (06) anciens départements : Cotonou, Porto-Novo, Lokossa, Abomey, Parakou, Natitingou. Elles sont chargées essentiellement :

- de l'étude des dossiers de prestations à court et à long terme ;
- de la liquidation des droits et du paiement en matière de prestations à court terme ;
- du paiement des prestations à long terme liquidées par la direction générale ;

— du recouvrement des cotisations sociales.

✎ **Mission de la CNSS**

La Caisse a pour mission de servir sous conditions prédéfinies, des prestations aux travailleurs soumis au code du travail ainsi qu'aux membres de leur famille, et de les garantir contre les risques susceptibles de réduire ou de supprimer leur capacité de gain.

L'accomplissement de cette mission nécessite des ressources financières dont la plus grande partie est constituée des cotisations des employeurs et des travailleurs.

La CNSS exerce diverses activités à savoir des activités principales désignées par les prestations techniques et des activités secondaires dénommées autres prestations.

1.2 Contexte du stage

1.2.1 Présentation de la Cellule des Etudes et de la Coopération

La cellule au sein de laquelle s'est déroulé notre stage est la Cellule Etudes et Coopérations (CEC). Elle comprend 2 sections qui sont : la section études et statistiques et la section coopération. La section études et statistiques est chargée de :

- ✍ assurer les études de toute nature ;
- ✍ procéder aux enquêtes, sondages, échantillonnages et autres recherches devant conduire à des études de tous genres ;
- ✍ concevoir les supports et la collecte des données statistiques ;
- ✍ analyser et exploiter les données statistiques ;
- ✍ préparer tous les supports devant servir aux évaluations actuarielles ;
- ✍ élaborer le rapport d'activités de la caisse ;
- ✍ élaborer les annuaires statistiques de la caisse.

Quant à la Section Coopération, elle est chargée de :

- ✍ assurer l'élaboration et la mise au point des projets de textes ;
- ✍ collecter auprès des directions des informations ayant trait à leurs attributions ;
- ✍ assurer les relations de coopération technique en matière de sécurité sociale avec les organismes internationaux et les institutions étrangères de sécurité sociale et cela en liaison avec les services concernés de la caisse ;
- ✍ étudier et mettre au point les projets de conventions et accords en matière de sécurité sociale.

1.2.2 Activités menées lors du stage

Les toutes premières semaines de notre stage furent consacrées à la recherche d'un thème pour notre étude à travers des entretiens avec le maître de stage et les fouilles de différentes bibliothèques numériques comme traditionnelles. Au cours du stage, nous avons participé à la collecte des données relatives aux pensions, rentes et indemnités journalières des femmes enceintes à l'agence CNSS de

Cotonou, dans l'optique de la préparation d'une base de données devant servir à l'élaboration de l'annuaire statistique. Nous avons aussi exécuté d'autres activités telles que la saisie des documents dans Word ainsi que le traitement et l'analyse des données dans Excel. Toutefois, il faut noter que la majeure partie du temps pendant notre stage a été consacrée à la recherche des informations et des données ayant permis la rédaction de notre mémoire.

2 | Cadre théorique et méthodologique de l'étude

Ce chapitre expose une réflexion théorique sur les facteurs explicatifs de l'hypertension artérielle. Il définit la démarche méthodologique adoptée et montre essentiellement l'intérêt scientifique du sujet, en posant le problème, les objectifs et hypothèses de recherche appuyés par les études déjà réalisées dans ce sens.

2.1 Problématique et intérêt de l'étude

Dans cette section, nous allons exposer l'origine de notre problème d'étude et expliquer l'intérêt qu'il y a de l'étudier.

2.1.1 Problématique

À l'échelle mondiale, les maladies cardiovasculaires (MCV) sont considérées comme les principales causes de décès ou d'années de vie ajustées sur l'incapacité, l'hypertension jouant un rôle central dans les MCV (Forouzanfar et al., 2017). L'hypertension artérielle représente environ 13,5 % des décès annuels dans le monde. En 2016, environ 17,9 millions de personnes sont décédées de maladies cardiovasculaires, ce qui représente 31 % du nombre total de décès dans le monde, dont 9,4 millions de décès ont été attribués à l'hypertension (OMS, 2019). Dans le monde, environ 1,13 milliard de personnes souffrent d'hypertension, et les deux tiers d'entre elles vivent dans des pays à revenu faible ou intermédiaire (PRFI) (OMS, 2019).

Selon l'OMS (2022), les MNT deviennent de plus en plus la principale cause de mortalité en Afrique subsaharienne. En effet, en Afrique, entre 50% et 88% des décès survenus sont dus à des MNT alors qu'en 2019, les MNT représentaient 37% de la mortalité. En Afrique, l'accent est mis sur les maladies infectieuses ; transmissibles au détriment de l'hypertension artérielle qui est méconnue ou mal soignée et qui est pourtant un tueur silencieux (Dauphinot et al., 2008). C'est dans la Région africaine que la prévalence de l'hypertension est la plus élevée puisqu'elle touche 46% des adultes âgés de 25 ans et plus. Dans l'ensemble, les pays à revenu élevé ont une prévalence de l'hypertension plus faible (35%) que celle enregistrée dans d'autres groupes de pays (40%) (OMS, 2011). Près de 80 % des décès par maladies cardio-vasculaires surviennent dans des pays à revenu faible ou intermédiaire. Or, ce sont ces pays qui peuvent le moins se permettre de supporter les conséquences sociales et économiques d'une mauvaise santé. Les taux de mortalité actuels standardisés selon l'âge sont plus élevés dans les pays à revenu faible que dans les pays développés (OMS, 2011).

Au Bénin, d'après une étude menée sur la prise en charge des hypertendus dans la ville de Cotonou en 2011, il est montré que 4,8% des médecins généralistes ne connaissent pas la définition de l'HTA et seulement 61% pouvaient décrire les conditions de mesure de la pression artérielle (Houenassi et al., 2016).

L'hypertension artérielle est un vecteur d'accentuation de la pauvreté des ménages. En effet, les décès prématurés, les incapacités, les bouleversements personnels et familiaux, la perte de revenu et les dépenses de santé liées à l'hypertension représentent un lourd fardeau pour les familles, les communautés et les finances publiques. Dans les pays à revenu faible ou intermédiaire, de nombreuses personnes ne font pas soigner leur hypertension car cela revient beaucoup trop cher. Les familles doivent alors faire face à des dépenses de santé exorbitantes, souvent sur le long terme dans le cas de complications de l'hypertension, ce qui fait plonger des millions de personnes dans la pauvreté (OMS, 2011). Au Bénin, à l'issue de l'enquête STEP 2007, la prévalence de l'HTA était de 27,5% et est passée à 25,9% en 2015 soit une baisse de 1,7 point de pourcentage en 8 ans. Il y a eu certes une certaine baisse, mais pas dans une proportion importante. Ce qui montre la persistance de l'HTA malgré toutes les mesures mises en place. Bon nombre d'auteurs se sont intéressés au sujet de l'HTA que ce soit en Afrique ou dans le monde. Certaines études se rejoignent et d'autres divergent par rapport aux facteurs qui peuvent expliquer la survenance de l'HTA.

Notons que l'HTA a un effet non négligeable sur la production des pays en ce sens qu'il contribue à une baisse de cette dernière. D'après l'OMS, le PIB des pays pauvres connaît une baisse de 4% chaque année du fait des principales maladies non transmissibles avec les maladies cardiovasculaires surtout l'HTA représentant près de la moitié de ce coût (OMS, 2013). En effet, cet impact sur la production des pays pauvres pourrait être attribué à la réduction de la main d'œuvre due au nombre de décès et des cas de paralysies récurrents imputables à l'HTA. Outre l'impact économique ci-dessus abordé, notons que la prévalence incessante voire grandissante des maladies non transmissibles entraînera des dépenses publiques importantes en santé pour ces pays.

Il s'avère donc nécessaire de s'interroger sur les facteurs explicatifs de l'hypertension artérielle au Bénin. Pour ce faire, il s'agira de répondre à l'interrogation suivante : **quels sont les facteurs explicatifs de l'hypertension artérielle chez les individus âgés de 30 à 64 ans au Bénin ? Au vu de ces facteurs, quel est le profil le plus susceptible d'avoir l'HTA au Bénin ?**

2.1.2 Intérêt de l'étude

L'hypertension artérielle est aujourd'hui l'un des problèmes ayant les plus lourdes conséquences au niveau économique et financier dans le monde. L'incidence croissante de l'hypertension artérielle signifie que les budgets de tous les pays sont menacés à cause de la perte de productivité due aux décès prématurés et au temps de travail diminué.

A ce titre, identifier les facteurs explicatifs pertinents en ce qui concerne les cas d'hypertension artérielle représente aujourd'hui une source d'intérêt croissante. Ces facteurs permettent de renseigner les patients sur leur maladie et de leur permettre de prendre les dispositions adéquates au plus tôt. Cette étude sera très utile pour une meilleure orientation du programme de sensibilisation contre l'hypertension au Bénin.

2.2 Objectifs et hypothèses de l'étude

2.2.1 Objectifs de l'étude

L'objectif général de cette étude est d'analyser les facteurs explicatifs de l'hypertension artérielle chez les personnes âgées de 30 à 64 ans au Bénin. De façon spécifique, il s'agira de :

- ✓ déterminer les facteurs explicatifs de l'HTA chez les individus âgés de 30 à 64 ans au Bénin ;
- ✓ déterminer le profil de l'individu le plus susceptible d'avoir l'HTA au Bénin ;
- ✓ déterminer si les personnes souffrant de l'HTA constituent un ensemble homogène selon le genre.

2.2.2 Hypothèses

Pour atteindre ces objectifs, trois hypothèses sont formulées :

- ✓ les facteurs sociodémographiques et économiques, et les facteurs métaboliques sont les facteurs qui influent sur la prévalence de l'HTA ;
- ✓ les femmes d'âges avancés, mariées, riches et diabétique sont les plus susceptibles d'avoir l'HTA ;
- ✓ les personnes souffrant de l'HTA constituent un ensemble hétérogène selon le genre.

2.3 Revue de littérature

2.3.1 Définitions des concepts

Hypertension artérielle : l'hypertension artérielle (HTA) correspond à une augmentation anormale de la pression du sang sur la paroi des artères. Selon l'OMS, On considère qu'une personne souffre d'hypertension lorsque l'on constate à plusieurs reprises, et pas le même jour, une tension artérielle systolique supérieure ou égale à 140 mm Hg et/ou une tension artérielle diastolique supérieure ou égale à 90 mm Hg.

Pression artérielle systolique : la pression systolique correspond à la pression artérielle mesurée lors de la phase de la systole, autrement dit c'est la pression artérielle maximale au sommet de la phase d'éjection systolique (ABBOU, 2016)

Pression artérielle diastolique : la pression diastolique correspond à la tension artérielle mesurée lors de la phase de relâchement du cœur, autrement dit c'est la pression artérielle minimale avant l'ouverture des valvules aortiques (ABBOU, 2016)

Facteur de risque : selon l'OMS (2013) « *c'est tout attribut, caractéristique ou exposition d'un sujet qui augmente la probabilité de développer une maladie ou de souffrir d'un traumatisme* ».

Prévalence : « *rapport du nombre de cas morbides à l'effectif total d'une population, sans distinction entre les cas nouveaux et les cas anciens, à un moment ou pendant une période donnée* » (Dictionnaire Larousse en ligne, 2020).

2.3.2 Revue empirique

L'hypertension artérielle (HTA) est un problème mondial majeur de santé publique selon l'OMS et sa prévalence varie considérablement d'un contexte à l'autre (OMS, 2010). En 2008, selon l'OMS, la prévalence mondiale de l'hypertension chez les adultes âgés de 25 ans et plus était d'environ 40 % et était plus élevée dans la Région africaine (46 %).

L'OMS (2013) regroupe les facteurs explicatifs de l'hypertension artérielle en trois principaux, que sont les facteurs sociaux, les facteurs comportementaux et les facteurs métaboliques. Pour commencer, parmi les facteurs sociaux, l'accent est mis sur l'urbanisation, le vieillissement, le revenu,

l'éducation et le logement. Au sein des facteurs comportementaux, on enregistre : la mauvaise alimentation, le tabagisme, le manque d'activité physique, la sédentarité et l'usage nocif d'alcool. Pour terminer, les facteurs métaboliques identifiés sont l'obésité, le diabète et l'hyperlipidémie (Panorama mondial de l'hypertension, 2013).

En Algérie, une étude menée à Tlemcen sur les facteurs de risques des maladies cardiovasculaires sur un échantillon de 1300 personnes a fait ressortir que 36,2 % [IC 95 % : 32,2-40,4] des personnes étaient hypertendues (Yahia-Berrouiguet et al., 2011). On constatait que cette prévalence était plus importante au niveau des hommes que des femmes et augmentait avec l'âge. Cependant, on note que parmi les personnes ayant une tension artérielle élevée, seulement 35% se connaissaient hypertendues et cette connaissance était significativement plus importante dans la sous-population des femmes que des hommes. Lang et al. (1983) dans une étude menée sur 1491 personnes ont montré l'association entre la consommation de tabac et de café et la survenue l'hypertension artérielle. À travers une analyse multivariée intégrant les variables telles que le sexe, l'âge, l'IMC, l'exercice physique, le milieu de résidence, la consommation de tabac et de café, ils ont obtenu comme résultats que la consommation de cigarette et de café avait un effet promoteur significatif sur la pression artérielle.

En 2007 en France, l'étude Mona Lisa (Wagner et al., 2007) incluant 4825 personnes montrait également que comparés aux femmes, les hommes présentaient une plus grande prévalence pour l'hypertension, soit (47% contre 35%) et que cette prévalence augmentait avec l'âge. Cependant, les femmes ont un suivi médical plus contrôlé comparé aux hommes (39% contre 24%), mais qui baisse au fur et à mesure qu'elles vieillissent. Par contre, Frérot et al. (1999) ont obtenu en France dans une étude menée sur 14492 personnes d'âge supérieur ou égal à 20 ans une prévalence de l'HTA de 16,5% avec respectivement une prévalence de 18,8% au niveau des femmes et de 14 % au niveau des hommes. Ceci indique que les femmes sont plus touchées par l'HTA que les hommes. Plusieurs facteurs sont déterminants dans la survenance de l'HTA d'après l'étude menée par Frérot et al. (1999) dans laquelle un accent particulier a été mis sur les facteurs socioéconomiques. Les variables telles que le niveau d'étude, l'occupation actuelle, le revenu du ménage, la zone géographique (les régions), la taille du ménage, l'IMC, la consommation du tabac, le diplôme, le milieu social, l'interaction entre l'âge et la variable retraite avaient un lien avec la prévalence de l'HTA. Deux modèles de régression logistique ont été estimés : un pour les hommes et un autre pour les femmes. Globalement, dans les deux modèles, l'âge, le revenu et l'IMC avaient un effet significatif sur la prévalence de l'HTA. De façon spécifique, le niveau d'étude et l'occupation de l'individu avaient un effet significatif sur l'HTA chez des hommes contrairement au niveau des femmes où on retrouve plutôt la zone géographique et la taille du ménage qui se trouvent avoir un effet significatif sur l'HTA. Dans cette étude de Frérot et al. (1999), on remarque que la consommation du tabac, ainsi que le milieu social et le diplôme ne déterminent pas l'HTA comme plusieurs études l'ont soutenu.

En 2008, aux Antilles Guyane, la prévalence de l'HTA était de 19% au sein des salariés avec une progression de cette fréquence avec l'Âge (IMANO, 2008). Il a été constaté que l'excès pondéral et la sédentarité augmentent la probabilité d'être hypertendue. Aussi, certains facteurs tels que le sexe, la prise de tabac et les facteurs socioéconomiques (le niveau d'éducation, les catégories socioprofessionnelles et le niveau de vie) ont-ils un impact significatif sur la prévalence de l'hypertension. Par ailleurs, les facteurs psychosociaux jouent également un rôle dans le développement de l'HTA

avec notamment le manque de confiance dans le futur. L'étude INTERHEART a montré l'importance des facteurs psychosociaux dans la prévalence des maladies cardiovasculaires. En effet, les déterminants psychosociaux affectaient significativement et augmentaient la probabilité de la survenance de l'infarctus du myocarde et la part du risque imputable à l'HTA était de 19,7% (Rosengren et al., 2004).

Fourcade et al. (2007) ont fait ressortir le rôle que joue l'urbanisation dans l'élévation de la tension artérielle à travers une synthèse des écrits concernant l'HTA en Afrique subsaharienne. Ils ont montré également l'association significativement positive entre l'obésité et la pression artérielle systolique et diastolique. Cooper et al. (2001) cité par Fourcade et al. (2007) dans l'étude menée en Afrique de l'ouest ont montré que l'obésité était associée au risque d'HTA.

Dans une étude menée au Maroc sur 525 diabétiques de type 24, Berraho et al. (2009) ont constaté une prévalence de l'hypertension de 70,4 %, soit 67,7 % chez les hommes et 72,0 % chez les femmes. Les variables telles que l'âge, le niveau d'étude, l'indice de masse corporelle, l'activité physique, l'ancienneté du diabète et le statut tabagique étaient liées à l'HTA chez ces sujets et une régression logistique confirmait l'effet positif de l'âge, de l'IMC, de la durée du diabète chez les personnes de sexe féminin et uniquement l'âge chez les hommes.

Dans une étude réalisée au Congo chez les adultes, Katchunga et al. (2011) ont montré l'influence de l'âge, du surpoids et du diabète sucré sur l'hypertension. Toujours au Congo, dans une étude menée au sein des militaires, 53,5% des personnes enquêtées étaient hypertendues (Sa et Jl, 2014). Cependant, les facteurs qui étaient liés à l'HTA dans cette étude sont l'âge, l'IMC, l'obésité abdominale et la consommation d'alcool. Une analyse multivariée avait révélé comme facteur de risque l'âge, la consommation de boissons alcoolisées et l'IMC rejoignant ainsi donc l'étude menée précédemment.

L'étude menée au Dakar par Macia et al. (2015) renseignait sur quelques facteurs influençant le risque d'HTA. Ces analyses indiquent que le niveau d'éducation, la situation matrimoniale, l'obésité et l'âge étaient associés à l'HTA. Cependant, les facteurs ayant un effet significatif sur l'hypertension dans le modèle logistique étaient l'âge et l'IMC. On constate que le risque d'être hypertendu augmente avec l'âge. Il en est de même quand on passe d'un individu de poids normal à un individu en surpoids.

Au Bénin, une étude menée dans les départements du Borgou et de l'Alibori dans une population de 15 à 72 ans d'âge, a montré une prévalence de 31,1% (Séraphin et al., 2016). Ils ont identifié dans leur étude réalisée dans les départements du Borgou et de l'Alibori comme facteurs de risque de l'HTA l'obésité, la consommation excessive de tisane, la consommation de tabac, la protéinurie et l'hématurie. Dans cette étude contrairement à d'autres études, l'âge, la consommation d'alcool et les variables concernant la qualité de l'alimentation ne sont pas déterminants dans la survenance de l'HTA.

2.3.3 Revue méthodologique

Diverses méthodes ont été utilisées dans la littérature pour évaluer les facteurs de risque associés à l'HTA.

En effet Cooper et al. (2001) ont comparé les taux d'HTA en Afrique de l'Ouest, dans les Caraïbes et aux Etats-Unis sur la base d'un protocole collaboratif hautement standardisé en utilisant des données issues d'enquête.

(Rosengren et al., 2004) ont réalisé une étude de cas-contrôle sur des patients ayant subit un

infarctus et sur des individus témoins appariés selon l'âge et appariés selon le sexe. Ces individus provenant de 262 centres en Asie, en Europe, au Moyen-Orient, en Afrique, Australie et Amérique du Nord et du Sud.

Frérot et al.(1999) ont estimé deux modèles de régression logistique, un pour les hommes et un autre pour les femmes.

Dans l'étude Mona Lisa (Wagner et al., 2007), les différences dans la prévalence et la prise en charge de l'HTA selon le groupe d'âge et le centre ont été estimées avec un test de Chi deux pour les variables catégorielles et une analyse de variance pour les variables continues. Une relation dose-effet entre les classes d'âge et la prévalence de l'HTA a été recherchée avec un test de tendance d'Armitage.

2.4 Méthodologie

Sous cette rubrique, nous allons présenter l'approche méthodologique ayant permis d'analyser les données collectées dans le cadre de cette étude. Il sera question de décrire notre méthodologie d'analyse, de spécifier nos modèles et enfin d'exposer notre procédure d'estimation et d'analyse.

2.4.1 Nature et sources des données

Cette étude se fonde sur les données issues de l'Enquête Démographique de santé 2017-2018 (EDSB-V) réalisée par l'INStaD. En effet, l'EDSB-V est une enquête s'inscrivant dans le cadre du programme DHS fournissant des données relatives aux indicateurs portant sur la santé de l'enfant, la santé maternelle, la fécondité, les maladies transmissibles et non transmissibles, les habitudes de consommation, le paludisme, etc. Plus explicitement, nous utilisons la base des femmes (Women's file (BJIR71SV)) et celle des hommes (Men's file (BJMR71SV)).

2.4.2 Population d'étude

La collecte des données a été menée auprès des femmes dont l'âge est compris entre 30 et 49 ans pour les femmes, et auprès des hommes dont l'âge est compris entre 30 et 64 ans. En effet, au cours de l'EDSB-V, la pression artérielle a été uniquement mesurée chez ces personnes.

Élaboré à partir de la méthode d'échantillonnage stratifié à deux degrés, l'échantillon considéré dans le cadre de l'EDSB-V est un échantillon représentatif tant au niveau départemental que sur le plan national. Le pays est subdivisé en 12 zones d'études assimilables aux douze départements et la stratification au niveau départemental tient compte de la nature de la zone notamment rurale ou urbaine.

Les unités primaires représentent les zones de dénombrement mises en place lors du troisième Recensement Général de la Population et de l'Habitation effectué au Bénin (RGPH3). Les unités secondaires représentent les ménages. Ainsi, 555 zones de dénombrement ont été tirées au premier degré par la méthode d'échantillonnage systématique avec une probabilité équivalente à la taille de la Zone. Au second degré, 26 ménages ont été tirés dans chaque zone de dénombrement retenue au premier degré en tenant compte du fait que ce soit en zone rurale ou en zone urbaine par la méthode d'échantillonnage systématique à probabilités égales.

Suite à l'échantillonnage, 14435 ménages ont été retenus comprenant 6528 ménages provenant de zones urbaines de 251 grappes et 7907 ménages provenant de zones rurales de 304 grappes. Toutes les personnes de sexe féminin de 15 à 49 ans résidant habituellement dans le ménage et présentes la veille de l'enquête ont été considérées dans l'échantillon.

Un sous échantillon de la moitié des ménages a été constitué et dans ce dernier, les hommes ayant entre 15 et 64 ans pouvaient être enquêtés. Les personnes de 30 à 49 ans au niveau des femmes et de 30 à 64 ans au niveau des hommes étaient éligibles pour la prise de mesure concernant la pression artérielle dans ce sous-échantillon.

Cependant, dans le cadre de notre étude, suite à la constitution de la base de données nécessaire pour effectuer nos analyses, et après avoir sélectionné de manière exhaustive tous les individus de la tranche d'âge de 30 à 64 ans pour lesquels la pression artérielle fut mesurée, la taille de l'échantillon a été réduite à 6776 personnes.

Parmi les 14 435 ménages retenus après l'échantillonnage pour l'EDSB-V, 14293 ont été identifiés lors de l'enquête. Sur ces 14293 ménages identifiés, 14156 ont été enquêtés, ce qui indique un taux de réponse de 99%. Que ce soit en milieu urbain ou rural, le taux de réponse n'est pas trop différent. Concernant les femmes, le taux de réponse s'élevait à 98%. En effet, parmi les 16 233 femmes de 15 à 49 ans retenues pour être enquêtées sur les 14 156 ménages identifiés, seulement 15 928 d'entre elles ont pu être enquêtées avec succès.

De même, au niveau des hommes, le taux de réponse était de 98%. 7 595 hommes furent enquêtés sur les 7 782 hommes retenus initialement.

2.4.3 Méthodes et outils d'analyse

2.4.3.1 Analyse descriptive

Les méthodes descriptives regroupent des techniques d'analyse univariée et bivariée. L'analyse univariée permet de dresser un portrait de la population étudiée à partir de sa répartition suivant la variable d'intérêt de l'étude et les principales variables explicatives retenues. L'analyse bivariée a pour but d'identifier les associations entre deux variables et d'apprécier la significativité statistique de cette association à partir d'un seuil donné. Dans le cadre de cette étude, elle permettra non seulement de voir, au seuil de 5% à l'aide du test de Khi-deux, la significativité ou non de l'association entre la variable dépendante et chacune des variables explicatives, mais aussi de mesurer les variations du phénomène selon ces dernières.

2.4.3.2 Analyse économétrique

Modèle de régression Probit

Considérons une variable Y binaire et X l'ensemble des variables explicatives. On cherche à expliquer les valeurs de Y grâce à X c'est-à-dire à estimer la probabilité que $Y_i = 1$ sachant X_i (ou que $Y_i = 0$). On remarque alors que :

$$Pr(Y_i = 1|X_i) = Pr(X_i\theta + \varepsilon_i \geq 0|X_i) = Pr(X_i\theta \geq -\varepsilon_i|X_i) = F_{-\varepsilon}(X_i\theta) = F_{\varepsilon}(X_i\theta). \quad (2.1)$$

F est donc la fonction de répartition d'une gaussienne centrée réduite, usuellement notée Φ :

$$F(X_i \theta) = \Phi(X_i \theta) = \int_{-\infty}^{X_i \theta} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \quad (2.2)$$

La densité correspondante, usuellement notée ϕ , est :

$$f(X_i \theta) = \phi(X_i \theta) = \frac{e^{-(X_i \theta)^2}}{\sqrt{2\pi}} \quad (2.3)$$

Estimation

Supposons un ensemble de données $\{y_i, x_i\}_{i=1}^n$ contient n unités statistiques indépendantes correspondant au modèle probit. Pour l'observation unique, conditionnellement au vecteur des entrées de cette observation, nous avons :

$$Pr(y_i = 1|x_i) = \Phi(x_i \theta). \quad (2.4)$$

$$Pr(y_i = 0|x_i) = 1 - \Phi(x_i \theta) \quad (2.5)$$

La probabilité d'une seule entrée est alors :

$$L(\theta, y_i, x_i) = \Phi(x_i \theta)^{y_i} [1 - \Phi(x_i \theta)]^{(1-y_i)} \quad (2.6)$$

Etant donné que les observations sont indépendantes et distribuées de manière identique, la vraisemblance de l'ensemble de l'échantillon, ou la vraisemblance conjointe, sera égale au produit des vraisemblances des observations individuelles :

$$L(\theta, Y, X) = \prod (\Phi(x_i \theta)^{y_i} [1 - \Phi(x_i \theta)]^{(1-y_i)}) \quad (2.7)$$

La fonction conjointe de log-vraisemblance est donc :

$$L(\theta, Y, X) = \sum (y_i \ln \Phi(x_i \theta) + (1 - y_i) \ln (1 - \Phi(x_i \theta))) \quad (2.8)$$

Effets marginaux

Soit l'équation $Prob(Y = 1|X) = F(X\beta) = F(\beta_0 + \sum_{i=1}^p \beta_i X_i)$ On appelle **effet marginal** la quantité :

$$ef_j = \frac{\partial Prob(Y = 1|X)}{\partial X_j} \quad (2.9)$$

$$ef_j = \beta_j \frac{\partial F}{\partial X_j} = \beta_j f(X\beta) \quad (2.10)$$

Ainsi donc, si X_j varie de $X_j + 1$ alors, la probabilité varie de ef_j . C'est donc un outil de surveillance de l'évolution de la probabilité prédictive.

Détection des valeurs anormales

La détection de valeurs anormales dans les données est cruciale car elles peuvent avoir une influence négative dans les estimations et, à fortiori, dans les prévisions (effet levier de la fonction de régression). Les méthodes les plus utilisées sont la méthode des résidus normalisés de Pearson et le critère des distances de Cook. Dans le cas de notre étude, nous allons utiliser le critère des distances de

Cook.

◆ Critères des distances de Cook

Pour tout $i = 1, \dots, n$, on définit la distance de Cook de la i -ème observation par :

$$d_i = \frac{[W^{1/2}X(X^tWX)^{-1}X^tW^{1/2}]_{i,i}}{(p+1)(1-[W^{1/2}X(X^tWX)^{-1}X^tW^{1/2}]_{i,i})}(e_i^*)^2. \quad (2.11)$$

où

$$W = \text{diag}(\hat{p}(x_1)(1-\hat{p}(x_1)), \dots, \hat{p}(x_n)(1-\hat{p}(x_n))). \quad (2.12)$$

Si $d_i > 1$, on envisage l'anormalité de la i -ème observation. Le seuil $4/(n-(p+1))$ est parfois utilisé.

Pertinence du modèle :

1. La règle du pouce

Si le modèle de régression probit est pertinent, la déviance D ne doit pas être trop éloignée de $\mathbb{E}(\hat{D})$, laquelle est proche de $\nu = n - (p + 1)$ (voir égale) :

$$\frac{D}{\nu} \simeq 1 \quad (2.13)$$

Si tel est le cas, le modèle est pertinent.

2. Test de Hosmer-Lemeshow

Pour évaluer la pertinence du modèle de régression probit, on préconise le test de Hosmer-Lemeshow. La p -valeur associée utilise la loi du Khi-deux : si p -valeur > 0.05 , on admet que le modèle est bien adapté aux données.

3. Test des résidus de la déviance

Le test des résidus de la déviance est similaire à celui des résidus de Pearson, mais avec les déviiances résiduelles :

pour tout $i \in 1, \dots, n$, la i -ème déviance résiduelle est la réalisation de :

$$\hat{D}_i = \text{Sign}(Y_i - \hat{p}(x_i)) \sqrt{2(Y_i \ln(\frac{Y_i}{\hat{p}(x_i)}) + (1 - Y_i) \ln(\frac{1 - Y_i}{1 - \hat{p}(x_i)}))}. \quad (2.14)$$

et la déviance est la réalisation de :

$$\hat{D} = \sum_{i=1}^n \hat{D}_i^2. \quad (2.15)$$

Soient χ_{obs}^2 cette réalisation et $K \sim \chi^2(\nu)$.

Alors la p -valeur associée est :

$$p\text{-valeur} = P(K \geq \chi_{obs}^2). \quad (2.16)$$

4. Test des résidus de Pearson

Le test des résidus de Pearson permet d'évaluer la pertinence du modèle.

On considère les hypothèses :

$$H_0 : p(x) = \text{probit}^{-1}(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)$$

$$H_1 : p(x) \neq \text{probit}^{-1}(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p).$$

On calcule :

$$\chi^2_{obs} = \sum_{i=1}^n (e_i^{\circ})^2. \quad (2.17)$$

On considère une $var K \sim \chi^2(\nu)$.

Alors la p -valeur associée est :

$$p\text{-valeur} = P(K \geq \chi^2_{obs}). \quad (2.18)$$

Si $p\text{-valeur} > 0.05$, alors on admet que le modèle est bien adapté aux données (Chesneau, 2020).

Significativité de la régression

1. Test de Wald

Soit $j \in 0, \dots, p$. Le test de Wald permet d'évaluer l'influence de X_j sur Y .

On considère les hypothèses :

$$H_0 : \beta_j = 0 \quad \text{contre} \quad H_1 : \beta_j \neq 0.$$

On calcule la réalisation z_{obs} de :

$$Z_* = \frac{\hat{\beta}_j}{\hat{\sigma}(\hat{\beta}_j)}.$$

On considère une variable

$$Z \sim N(0, 1), \quad (2.19)$$

Alors la p -valeur associée est

$$p\text{-valeur} = \mathbb{P}(|Z| \geq |z_{obs}|). \quad (2.20)$$

Ce test repose sur le fait que l'estimateur du maximum de vraisemblance a pour loi asymptotique la loi normale.

2. Test de Déviance

Soit $j \in 0, \dots, p$. Le test de la déviance vise à évaluer l'influence (ou la contribution) de X_j sur Y . La p -valeur associée utilise la loi du Khi-deux.

3. Test de lr (likelihood ratio)

On considère les hypothèses :

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

H_1 : il y a au moins un coefficient non nul.

Pour ce faire, on utilise le test du rapport de vraisemblance (lr pour likelihood ratio) (asymptotique). La p -valeur associée utilise la loi du Khi-deux.

Rapports des côtes ou odds ratio

On appelle rapport des côtes (ou odds ratio) de 2 valeurs x_* et x_0 de $X = (X_1, \dots, X_p)$ le réel :

$$RC(x_*, x_0) = \frac{\frac{p(x_*)}{1-p(x_*)}}{\frac{p(x_0)}{1-p(x_0)}} = \frac{p(x_*)(1-p(x_0))}{(1-p(x_*)p(x_0)}. \quad (2.21)$$

Soient $j \in 1, \dots, p$ et $e_j = (0, \dots, 0, 1, 0, \dots, 0)$ (le 1 se situe à la j -ème composante).

Si X_j augmente d'une unité, alors le rapport des côtes est :

$$RCj = RC(x + e_{j,x}) = \exp(\beta_j), \quad x \in \mathbb{R}^p. \quad (2.22)$$

Par conséquent,

- si $RC_j > 1$, l'augmentation d'une unité de X_j entraîne une augmentation des chances que $Y = 1$ se réalise ;
- si $RC_j = 1$, l'augmentation d'une unité de X_j n'a pas d'impact sur Y ;
- si $RC_j < 1$, l'augmentation d'une unité de X_j entraîne une augmentation des chances que $Y = 0$ se réalise.

Un estimateur de RC_j est :

$$\hat{RC}_j = \exp(\hat{\beta}_j). \quad (2.23)$$

Avec l'observation de celui-ci, on peut interpréter l'influence de X_j sur $Y = 1$ en la comparant à 1.

Qualité du modèle

Pour évaluer la qualité du modèle de régression logistique, on préconise :

1. Le taux d'erreur

On appelle le taux d'erreur (de prédiction) le réel :

$$t = \frac{1}{n} \left(\sum_{i=1}^n 1_{y_i=0 \cap \tilde{y}_i=1} \right) + \left(\sum_{i=1}^n 1_{y_i=1 \cap \tilde{y}_i=0} \right). \quad (2.24)$$

Ce taux est la proportion des modalités prédites qui diffèrent des modalités observées (c'est aussi la somme des 2 valeurs non-diagonales de la matrice de confusion divisée par n).

Plus t est proche de 0, meilleure est la qualité prédictive du modèle.

On convient que la qualité prédictive du modèle est mauvaise lorsque $t > 0.5$.

2. La courbe ROC

Soit $\tau \in [0, 1]$. On appelle i -me prédiction du groupe au niveau τ la réalisation $\tilde{y}_i(\tau)$ de :

$$\hat{Y}_i(\tau) = \begin{cases} 1 & \text{si } \hat{p}(xi) \geq \tau, \\ 0 & \text{sinon.} \end{cases}$$

À partir de ces prédictions, on définit :

- la fréquence de fausse alarme ("1-specificity") :

$$ffa(\tau) = \frac{\sum_{i=1}^n 1_{y_i=0 \cap \tilde{y}_i=1}}{\sum_{i=1}^n 1_{y_i=0}}; \quad (2.25)$$

- la fréquence de bonne détection ("sensitivity") :

$$fbd(\tau) = \frac{\sum_{i=1}^n 1_{y_i=1 \cap \tilde{y}_i=1}}{\sum_{i=1}^n 1_{y_i=1}}. \quad (2.26)$$

On appelle courbe ROC, la courbe passant par les points :

$$\{(f f a(\tau), f b d(\tau)); \tau \in [0, 1]\}.$$

Plus la courbe longe les axes $x = 0$ et $y = 1$, meilleur est le modèle. Autrement dit, plus l'aire sous la courbe ROC est proche de 1, meilleur est le modèle.

Sélection des variables du modèle

Dans le cadre de la modélisation, il est important que les variables choisies expliquent au mieux notre variable d'intérêt. Le choix des variables explicatives s'est fait suivant deux approches. La première consiste à sélectionner les variables liées à la variable d'intérêt à partir du test de Khi-deux. La seconde approche est l'approche pas à pas.

1. Sélection des variables par le test de Khi-deux

Cette approche consiste à utiliser les variables qui sont liées à la variable d'intérêt au moyen du test d'indépendance de Khi-deux au seuil de 5%. Les hypothèses de ce test sont les suivantes :

$$H_0 : \text{les deux variables sont indépendantes}$$

$$H_1 : \text{les deux variables ne sont pas indépendantes}$$

2. Approche pas à pas

En statistique, la régression pas à pas est une méthode d'ajustement de modèles de régression dans laquelle le choix des variables prédictives est effectué par une procédure automatique. À chaque étape, une variable est prise en compte pour l'addition ou la soustraction de l'ensemble des variables explicatives en fonction d'un critère pré-spécifié. Habituellement, on utilise les critères AIC, BIC, C_p de Mallows.

Le critère AIC est défini par :

$$AIC = -2 \log \tilde{L} + 2k \quad (2.27)$$

où \tilde{L} est la vraisemblance maximisée et k le nombre de paramètres dans le modèle.

Le Bayesian information criterion (BIC) est défini par :

$$BIC = -2 \log(\tilde{L}) + k \log(n) \quad (2.28)$$

où \tilde{L} est la vraisemblance maximisée, k le nombre de paramètres dans le modèle et n le nombre de variables dans le modèle.

Le critère C_p de Mallows est défini par :

$$C_p = \frac{SC_{res}}{\hat{\sigma}^2} - (n - 2p) \quad (2.29)$$

où SC_{res} est la somme des carrés des résidus, $\hat{\sigma}^2$ est le carré moyen de l'erreur pour le modèle incluant tous les prédicteurs, n le nombre d'observations et p le nombre de termes dans le modèle, constante incluse.

◆ Procédures courantes de régression pas à pas

- La sélection directe

Elle consiste à commencer sans variable dans le modèle, à tester l'ajout de chaque variable à l'aide d'un critère d'ajustement du modèle choisi ; à ajouter la variable (le cas échéant) dont l'inclusion donne l'amélioration la plus statistiquement significative de l'ajustement ; et à répéter ce processus jusqu'à ce qu'aucun n'améliore le modèle de manière statistiquement significative.

- L'approche en arrière

On part d'un modèle de régression linéaire multiple avec toutes les variables X_1, \dots, X_p et on étudie leur significativité. On retire la moins significative (donc celle qui a la plus grande p -valeur). Puis, on refait une régression linéaire multiple avec les variables restantes et on retire de nouveau la moins significative. On réitère ce processus jusqu'à n'avoir que des variables significatives.

- L'approche en avant

On considère les p régressions simples possibles avec une seule variable explicative X_1 ou X_2 ou \dots X_p et on étudie leur significativité. On retient la plus significative (donc celle qui a la plus petite p -valeur). On fait alors les $p-1$ régressions linéaires multiples contenant la variable retenue et une seule autre variable parmi les $p-2$ restantes. On garde alors la plus significative parmi ces dernières. On réitère ce processus jusqu'à ce qu'aucune variable ne soit retenue.

Ainsi, deux modèles sont obtenus à partir de ces approches. Le choix du meilleur modèle sera fait par comparaison à partir du test ANOVA : si p -valeur > 0.05 , alors les variables étudiées ne contribuent pas significativement au modèle.

◆ Spécification des variables

L'hypertension est la variable expliquée dans cette étude. Au cours de la collecte des données de l'EDSB-V, les enquêteurs ont mesuré la pression artérielle dans un sous-échantillon d'un ménage sur deux de toutes les femmes de 30-49 ans et de tous les hommes de 30-64 ans. Trois mesures ont été prises au cours de l'interview (au début, au milieu et à la fin) et une valeur moyenne a été calculée.

On considère qu'une personne souffre d'hypertension lorsque l'on constate une tension artérielle systolique moyenne supérieure ou égale à 140 mmHg et/ou une tension artérielle diastolique moyenne supérieure ou égale à 90 mmHg. Cependant, certains facteurs que nous appellerons facteurs externes tels que : la prise d'un repas, la prise de boissons caféinées, la consommation de cigarette ou la pratique d'une activité sportive intensive 30 minutes avant la prise de la pression pourraient être à l'origine d'une forte tension artérielle observée. Aussi, une tension artérielle normale pourrait être observée chez un individu lors de la prise, alors que ce dernier souffre d'hypertension. Ainsi, un individu sera déclaré hypertendu si :

- il a déjà été une fois diagnostiqué comme étant hypertendu par un professionnel ;
- ou s'il a une pression systolique/diastolique anormale et qu'aucun des facteurs externes n'est observé.

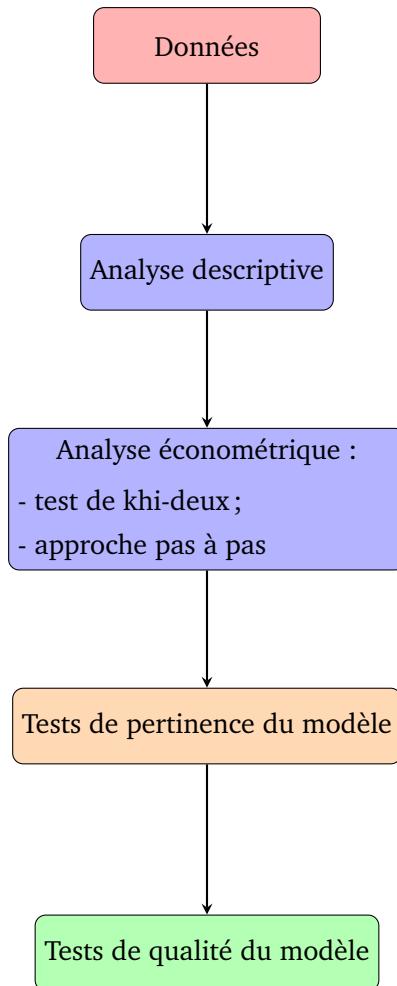
$$\text{hypertension} = \begin{cases} 1 & \text{si l'individu est hypertendu;} \\ 0 & \text{sinon.} \end{cases}$$

TABLEAU 2.1 – Variables explicatives de l'étude

N°	Noms	Description	Modalités
Facteurs sociodémographiques et économiques			
1	Âge	Indique l'âge de l'individu	1 : 30-34 ans 2 : 35-39 ans 3 : 40-44 ans 4 : 45-49 ans 5 : 50-54 ans 6 : 55-59 ans 7 : 60-64 ans
2	Genre	Indique le sexe de l'individu	1 : Masculin 2 : Féminin
3	Milieu de résidence	Indique le milieu de résidence de l'individu	1 : Urbain 2 : Rural
4	Région	Indique la région de l'individu	1 : Nord 2 : Sud 3 : Centre
5	Instruction	Indique le niveau d'éducation de l'individu	1 : Sans instruction 2 : Primaire 3 : Secondaire 4 : Supérieure
6	Etat matrimonial	Indique le statut matrimonial de l'individu	1 : Jamais en union 2 : Marié 3 : Veuf 4 : Divorcé 5 : Séparé
7	Occupation	Indique si l'individu a une occupation ou non	1 : Non occupé 2 : Occupé
8	Niveau de vie	Indique le niveau de vie de l'individu	1 : Très pauvre 2 : Pauvre 3 : Moyen 4 : Riche 5 : Très riche
Facteurs métaboliques			
9	Diabète	Indique si l'individu est diabétique ou non	1 : Non 2 : Oui

Source : Réalisé par les auteurs à partir des données.

FIGURE 2.1 – Organigramme de la modélisation



2.5 Limites de l'étude

Lors de la réalisation de notre étude, nous avons rencontré plusieurs difficultés. En effet, nous n'avons pas pu considérer tous les facteurs pouvant expliquer l'HTA (l'IMC, l'obésité, la fréquence d'activité physique, la prédisposition familiale, les facteurs psychosociaux, les facteurs comportementaux). Il faut aussi noter que nous n'avons pas pu utiliser des données récentes pour notre étude. Nos données datent de 2017.

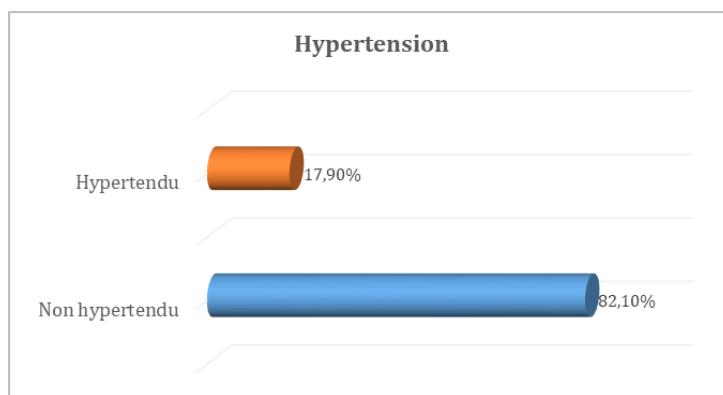
3 | Résultats, analyses et interprétations

Ce chapitre expose les résultats obtenus après le traitement, l'analyse des données et leurs interprétations. Il présente également les préconisations opérationnelles qui découlent de la présente étude.

3.1 Analyse descriptive

3.1.1 Analyse univariée

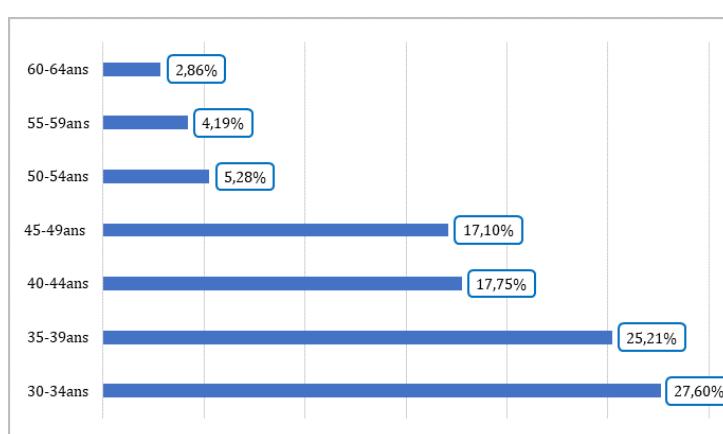
FIGURE 3.1 – Prévalence de l'hypertension artérielle



Source : Auteurs, 2022

De l'analyse de la figure 3.1, on note une prévalence de l'hypertension de 17,9% (soit 1213 personnes). Un pourcentage certes faible, mais pas négligeable.

FIGURE 3.2 – Répartition de la population selon l'âge

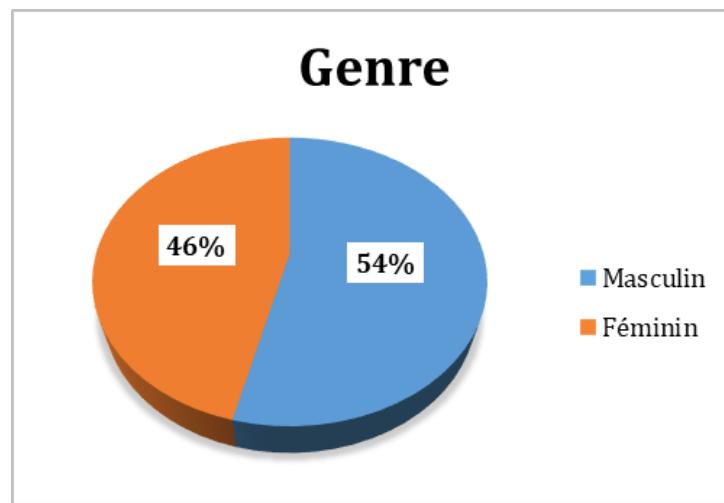


Source : Auteurs, 2022

Notre échantillon (6776 individus) est constitué d'individu de 30 à 64 ans (30 à 49 ans pour les femmes, et 30 à 64 ans pour les hommes).

Il ressort de l'analyse de la figure 3.2 que le groupe d'âge le plus représenté est celui de 30 à 34 ans (27,60%) et celui le moins représenté est le groupe d'âge de 60 à 64 ans (2,86%). Après la catégorie des individus de 30 à 34 ans s'en suit dans l'ordre celle de 35 à 39 ans, celle de 40 à 44 ans, celle de 45 à 49 ans, celle de 50 à 54 ans et celle de 55 à 59 ans.

FIGURE 3.3 – Répartition de la population selon le genre



Source : Auteurs, 2022

De la figure 3.3, on constate que plus de la moitié des individus de notre échantillon sont de genre masculin (54%), et les femmes constituent 46% de l'échantillon.

3.1.2 Analyse bivariée

TABLEAU 3.1 – Résultats de l'analyse descriptive bivariée

Variables	Modalités	Hypertension	
		0(%)	1(%)
Facteurs sociodémographiques et économiques			
Âge	30-34ans	1635(87,43)	235(12,57)
	35-39ans	1441(84,37)	267(15,63)
	40-44ans	962(79,97)	241(20,03)
	45-49ans	899(77,57)	260(22,43)
	50-54ans	260(96,65)	9(3,35)
	55-59ans	215(75,70)	69(24,30)
	60-64ans	151(77,4)	43(22,16)
Genre	Masculin	3071(84,16)	578(15,84)
	Féminin	2492(79,69)	635(20,31)
Milieu de résidence	Urbain	2390(78,72)	646(21,28)

Variables	Modalités	Hypertension	
		0(%)	1(%)
Région	Rural	3173(84,84)	567(15,16)
	Nord	2262(90,52)	237(9,48)
	Sud	163(23,80)	522(76,20)
	Centre	161(26,18)	454(73,82)
Instruction	Sans instruction	3434(84,39)	635(15,61)
	Primaire	1121(80,47)	272(19,53)
	Secondaire	772(74,88)	259(25,12)
	Supérieure	236(83,39)	47(16,61)
Etat matrimonial	Jamais en union	150(87,21)	22(12,79)
	Marié	4192(84,06)	795(15,94)
	Vie avec partenaire	56(16,82)	277(83,18)
	Veuf	132(73,33)	48(26,67)
	Divorcé	64(75,29)	21(24,71)
Occupation	Séparé	169(77,17)	50(22,83)
	Non occupé	303(83,93)	58(16,07)
	Occupé	5260(82,00)	1155(18,00)
Niveau de vie	Très pauvre	1227(89,82)	139(10,18)
	Pauvre	1130(85,67)	189(14,33)
	Moyen	1021(84,52)	187(15,48)
	Riche	1020(78,58)	278(21,42)
	Très riche	1165(73,50)	420(26,50)
Facteurs métaboliques			
Diabète	Non	5514(82,51)	
	Oui	49(52,69)	
		44(47,31)	

Source : Auteurs, 2022

Le tableau 3.1 présente les profils ligne de chaque variable explicative avec notre variable d'intérêt. Nos conclusions sont basées sur le profil ligne le plus élevé parmi les modalités d'une variable explicative au niveau des hypertendus. Il ressort de ce tableau que :

- la prévalence de l'hypertension artérielle est plus élevée chez les personnes de la tranche d'âge de 55 à 59 ans (24,30%). Cependant, la prévalence de l'HTA est plus faible chez les plus jeunes de notre échantillon notamment les personnes âgées de 30 à 34 ans (11%). Cette tendance porte à croire que les plus âgées ont une plus forte prévalence à l'HTA comparé aux plus jeunes ;
- les personnes de sexe féminin souffrent plus d'hypertension que les hommes (20,31% contre 15,84%) ;
- les personnes vivant en milieu urbain présentent une plus forte prévalence d'HTA comparée à ceux vivant en milieu rural (21,2% contre 15,16%) ;
- les personnes résidant au sud et au centre du Bénin présentent une plus forte prévalence à l'hypertension artérielle ;
- les personnes avec une éducation scolaire de niveau secondaire sont les plus susceptibles d'avoir de l'hypertension artérielle, s'ensuit les personnes avec un niveau d'instruction primaire. Par contre la prévalence est plus faible au niveau des individus sans niveau d'éducation

(15,61%). Cela porte à croire que cette prévalence a tendance à croître au fur et à mesure qu'on évolue vers des niveaux d'éducation plus élevés ;

- la prévalence de l'hypertension artérielle est plus forte chez les personnes vivant avec un partenaire (83,18%) et plus faible chez ceux qui n'ont jamais été en union (12 ,79%) ;
- les personnes ayant une occupation souffrent plus d'hypertension ;
- les personnes riches et très riches ont une plus forte prévalence à l'hypertension artérielle (21,42% et 26,50%). Cette prévalence est par contre faible chez les individus très pauvres et pauvres (10,18% et 14,33% respectivement). On remarque donc une augmentation de la prévalence de l'HTA avec la catégorie socioéconomique ;
- la prévalence de l'hypertension artérielle est plus forte chez les personnes diabétiques.

3.2 Test de pré-estimation

TABLEAU 3.2 – Tests de Khi deux

Test de khi-deux	
Variables explicatives	p-value
Age	< 2,2e – 16
Genre	2,036e – 06
Milieu de résidence	8,008e – 11
Région	< 2,2e – 16
Instruction	1,132e – 16
Etat matrimonial	1,207e – 12
Occupation	0,3875
Niveau de vie	< 2,2e – 16
Diabète	2,605e – 13

Source : Auteurs, 2022

Les résultats des tests de Khi-deux de la variable expliquée « hypertension » par les variables explicatives indiquent que toutes les variables explicatives sont statistiquement liées à la variable d'intérêt au seuil de 5% excepté la variable « occupation ». De tous ces résultats, seules les variables présentant une liaison significative avec la variable dépendante ont été prises en compte dans la suite des analyses.

3.3 Analyse économétrique

3.3.1 Estimation du modèle commun

3.3.1.1 Choix du modèle

Nous avons estimé un modèle logit et un modèle probit puis, nous avons calculé le critère AIC pour chacun de ces modèles afin de choisir le meilleur modèle pour notre étude.

Modèle A : logit(hypertension~age + genre + milieu de residence + region + instruction + occupation + etat matrimoniale + niveau de vie + diabete)

Modèle B : probit(hypertension~age + genre + milieu de residence + region + instruction + occupation + etat matrimoniale + niveau de vie + diabete)

TABLEAU 3.3 – Résultats des tests du choix de modèle

	AIC
Modèle A	5924,997
Modèle B	5922,56

Source : Auteurs, 2022

En comparant l'AIC de chaque modèle, on remarque que le modèle B a l'AIC le plus faible (5922,56). Ainsi, nous estimerons un modèle probit.

3.3.1.2 Choix des variables explicatives

Comme annoncé dans la méthodologie, nous avons procédé à deux estimations pour donner la meilleure prédiction que possible des facteurs explicatifs de l'HTA dans notre population. Le modèle 1 est constitué de toutes les variables explicatives liées à notre variable d'intérêt à partir du test de Khi-deux ; le modèle 2 est obtenu à partir de la régression pas à pas.

Modèle 1 : hypertension~age + genre + milieu de residence + region + instruction + etat matrimoniale + occupation + niveau de vie + diabete

Modèle 2 : hypertension~age + genre + milieu de residence + region + instruction + etat matrimoniale + niveau de vie + diabete

TABLEAU 3.4 – Résultats des tests du choix de la meilleure combinaison des variables

	Resid. Dev	AIC
Modèle 1	5872,6	5922,56
Modèle 2	5873,8	5921,822

Source : Auteurs, 2022

En comparant l'AIC de chaque modèle, on remarque que le modèle 2 a l'AIC le plus faible (5921,822). Ainsi, le meilleur modèle retenu pour la prédiction de notre variable d'intérêt est le modèle 2.

3.3.1.3 Résultats de la régression Probit

TABLEAU 3.5 – Résultats de l'estimation du probit

Variables	Coefficients	Odds Ratio	p-value
Age (référence = 30-34ans)			
35-39ans	0,1357	1,1453	0,01277*
40-44ans	0,3246	1,3834	2,285e-08***
45-49ans	0,4076	1,5033	2,608e-12***
50-54ans	0,6715	1,9572	7,577e-15***
55-59ans	0,6545	1,9242	1,039e-11***
60-64ans	0,5935	1,8104	1,765e-07***
Genre (référence = Masculin)			
Féminin	0,2967	1,3454	5,372e-12***
Milieu de résidence (référence = Urbain)			
Rural	-0,0676	0,9346	0,11986
Région (référence = Nord)			
Sud	0,3394	1,4041	1,768e-10***
Centre	0,4488	1,5665	<2,2e-16***
Instruction (référence = Sans instruction)			
Primaire	-0,0019	0,9980	0,9687
Secondaire	0,1483	1,1599	0,0117*
Supérieure	-0,1826	0,8330	0,0838.
Etat matrimonial (référence = Jamais en union)			
Marié	0,0438	1,0448	0,7341
Vie avec partenaire	0,2391	1,2701	0,0741.
Veuf	0,1832	1,2010	0,2706
Divorcé	0,2562	1,2921	0,2000
Séparé	0,1259	1,1342	0,4317
Niveau de vie (référence = Très pauvre)			
Pauvre	0,1411	1,1515	0,0312*
Moyen	0,1620	1,1758	0,0151*
Riche	0,3488	1,4173	1,025e-07***
Très riche	0,4442	1,5593	8,378e-10***
Diabète (référence = Non)			
Oui	0,7160	2,0464	1,430e-07***

Notes : (ns) Non significatives ; (.) 10% ; (*) 5% ; (**) 1% ; (*** 0,1%

Source : Auteurs, 2022

A la lecture de ce tableau, on constate que les coefficients des modalités "genre féminin", "région sud", "région centre", "instruction secondaire", "niveau de vie pauvre", "niveau de vie moyen", "niveau de

vie riche", "niveau de vie très riche", "diabète oui" et toutes les modalités de la variable âge sont significatives au seuil de 5% et positivement liées à la probabilité d'avoir l'HTA.

3.3.1.4 Pertinence du modèle

TABLEAU 3.6 – Pertinence du modèle probit

Tests	Critère de significativité
Règle de pouce	D/v = 0,8699
Test de Hosmer-Lemeshow	Pr(Chi2)= 0,433
Test de résidus de Pearson	Pr(Chi2)=0,5483
Test de résidus de la déviance	Pr(Chi2) \cong 1

Source : Auteurs, 2022

On remarque que la valeur D/v est très proche de 1. De plus, les probabilités de Khi-deux des tests d'Hosmer-Lemeshow, des résidus de Pearson et des résidus de la déviance sont toutes supérieures au seuil de 5%, donc le modèle est bien adapté aux données. Ainsi, le modèle est pertinent.

3.3.1.5 Qualité du modèle

TABLEAU 3.7 – Qualité du modèle

Tests	Valeurs
Taux d'erreur	0,1779
Aire sous la courbe ROC	0,6944

Source : Auteurs, 2022

Le taux d'erreur du modèle est inférieur à 0,5 ; donc le modèle est de bonne qualité. De plus, l'aire sous la courbe ROC (0,90) est supérieure à 0,60 ; le modèle est donc d'une bonne qualité prédictive.

3.3.1.6 Effets marginaux

TABLEAU 3.8 – Effets marginaux

Variables	AME(%)
Age (référence = 30-34ans)	
35-39ans	2,78
40-44ans	7,29
45-49ans	9,51
50-54ans	17,44
55-59ans	16,89
60-64ans	14,96
Genre (référence = Masculin)	

Variables	AME(%)
Féminin	7,20
Milieu de résidence (référence = Urbain)	
Rural	-1,63
Région (référence = Nord)	
Sud	7,61
Centre	10,58
Instruction (référence = Sans instruction)	
Primaire	-0,05
Secondaire	3,76
Supérieure	-4,01
Etat matrimonial (référence = Jamais en union)	
Marié	1
Vie avec partenaire	5,94
Veuf	4,44
Divorcé	6,41
Séparé	2,98
Niveau de vie (référence = Très pauvre)	
Pauvre	2,94
Moyen	3,41
Riche	8,03
Très riche	10,68
Diabète (référence = Non)	
Oui	21,85

Notes : (ns) Non significatives ; (.) 10% ; (*) 5% ; (**) 1% ; (***) 0,1%

Source : Auteurs, 2022

Il ressort de ce tableau que :

- la probabilité qu'un individu ait l'HTA augmente de 2,78% lorsqu'il est âgé de 35 à 39 ans ; elle augmente de 7,29% lorsqu'il est âgé de 40 à 44 ans et de 9,51% lorsqu'il est âgé de 45 à 49 ans. Pour les personnes âgées 50 à 54 ans, elle augmente de 17,44%, et de 16,89% pour les personnes âgées de 55 à 59 ans. Pour ce qui est des individus âgés de 60 à 64 ans, la probabilité qu'ils soient hypertendus augmente de 14,96%. Notons que l'âge avancé fait augmenter la probabilité d'avoir l'HTA ;
- la probabilité d'être hypertendu augmente de 21,85% lorsque l'individu est diabétique ;
- la probabilité qu'un individu ait l'HTA augmente de 7,20% lorsqu'il est de sexe féminin ;
- la probabilité qu'un individu ait l'HTA augmente de 3,76% lorsqu'il a une éducation de niveau secondaire et diminue de 4% lorsqu'il a une éducation de niveau supérieure. Ce pourrait être dû au fait que les personnes avec une éducation de niveau supérieure sont plus informées sur l'hypertension et prennent donc des précautions pour l'éviter ;

- la probabilité qu'un individu ait l'HTA augmente de 2,94% lorsqu'il a un niveau de vie pauvre et de 8,03% lorsqu'il a un niveau de vie riche. Elle augmente par contre de 3,41% lorsqu'il a un niveau de vie moyen et de 10,68% lorsqu'il a un niveau de vie très riche. De plus on remarque que plus l'individu à un niveau de vie élevé, plus la probabilité d'avoir l'HTA augmente ;
- la probabilité qu'un individu ait l'HTA augmente de 7,61% lorsqu'il habite au sud du Bénin et de 10,58% lorsqu'il habite au centre du Bénin.

3.3.1.7 Interprétations des résultats

La prévalence de l'HTA dans notre population (17,9%) est presque similaire à celle trouvée en Côte d'Ivoire par Sackou et al. (2020), qui était de 18,3%.

Selon les résultats de l'estimation du tableau 3.5, nous remarquons que :

- les personnes de la tranche d'âge de 35 à 39 ans ont 1,14 fois plus de risque d'avoir l'HTA par rapport aux personnes de la tranche d'âge de 30 à 34 ans. Les personnes de la tranche d'âge de 40 à 44 ans ont 1,38 fois plus de risque d'avoir l'HTA par rapport aux personnes de la tranche d'âge de 30 à 34 ans. Les personnes de la tranche d'âge de 45 à 49 ans ont 1,50 fois plus de risque d'avoir l'HTA par rapport aux personnes de la tranche d'âge de 30 à 34 ans. Les personnes de la tranche d'âge de 50 à 54 ans ont 1,95 fois plus de risque d'avoir l'HTA par rapport aux personnes de la tranche d'âge de 30 à 34 ans. Les personnes de la tranche d'âge de 55 à 59 ans ont 1,92 fois plus de risque d'avoir l'HTA par rapport aux personnes de la tranche d'âge de 30 à 34 ans. Les personnes de la tranche d'âge de 60 à 64 ans ont 1,81 fois plus de risque d'avoir l'HTA par rapport aux personnes de la tranche d'âge de 30 à 34 ans. De plus, nous remarquons que plus l'individu vieillit plus il est susceptible d'avoir l'HTA. Cela pourrait s'expliquer par le fait que plus l'âge est avancé, plus le corps se fatigue et le coeur n'a plus la même vigueur pour faire circuler le sang. Ces résultats confirment les travaux effectués par Yahia-Berrouiguet et al. (2011) en Algérie et Ratovoson et al. (2014) au Madagascar. Des résultats similaires étaient également obtenus par Wagner et al. (2007) en France suite à l'étude Mona lisa ;
- les personnes de sexe féminin ont 1,34 fois plus de risque d'avoir l'HTA par rapport aux personnes de sexe masculin. Cela pourrait s'expliquer par les changements hormonaux fréquents chez la femme, la grossesse ou la ménopause. Des résultats similaires ont été obtenus par Frérot et al. (1999) en France ;
- les personnes vivant au sud du Bénin ont 1,40 fois plus de risque d'avoir l'HTA par rapport aux personnes vivant au nord du Bénin. Les personnes vivant au centre du Bénin ont 1,56 fois plus de risque d'avoir l'HTA par rapport aux personnes vivant au nord du Bénin. Cela pourrait s'expliquer par le fait que les parties sud et centre du bénin sont plus urbanisées que la partie nord du Bénin ;
- les personnes ayant une éducation scolaire de niveau secondaire ont 1,15 fois plus de risque d'avoir l'HTA par rapport aux personnes n'ayant aucune éducation. Cela peut s'expliquer par le fait que les personnes n'ayant pas d'éducation ont moins de chance d'avoir un niveau de vie riche ou moyen que les personnes ayant une éducation de niveau secondaire ;
- les personnes pauvres ont 1,15 fois plus de risque d'avoir l'HTA par rapport aux personnes très

pauvre. Les personnes ayant un niveau de vie moyen ont 1,17 fois plus de risque d'avoir l'HTA par rapport aux personnes très pauvres. Les personnes riches ont 1,41 fois plus de risque d'avoir l'HTA par rapport aux personnes très pauvres. Les personnes très riches ont 1,55 fois plus de risque d'avoir l'HTA par rapport aux personnes très pauvres. Notons que plus la condition socioéconomique de l'individu est élevée et plus il présente de risque d'être hypertendu. Ce résultat pourrait s'expliquer dans notre contexte par le fait que les personnes d'un niveau de vie élevé ont des habitudes alimentaires contenant plus de gras, ou font moins d'activité physique compte tenu de l'existence de moyens de déplacement ;

- les personnes ayant le diabète ont 2,04 fois plus de risque d'avoir l'HTA par rapport aux personnes n'ayant pas de diabète. Cela peut s'expliquer par la mauvaise hygiène de vie, l'augmentation de la sédentarité, la prise de poids ou encore la mauvaise alimentation des personnes diabétiques.

Ainsi donc, nous pouvons dire que les déterminants de l'HTA chez la population de 30 à 64 ans au Bénin sont : l'âge, le genre, la région, le niveau d'instruction, le niveau de vie et le diabète.

3.3.2 Estimation des modèles selon le genre

3.3.2.1 Choix des variables explicatives

Comme pour le modèle général, nous avons d'abord estimé un modèle logit et un modèle probit, puis choisi le meilleur modèle à partir du critère AIC : le modèle probit. Pour l'estimation du probit, nous avons procédé à deux différentes estimations pour donner la meilleure prédition possible des facteurs explicatifs de l'HTA. Le modèle 1 est constitué de toutes les variables explicatives ; le modèle 2 est obtenu à partir de la régression pas à pas.

TABLEAU 3.9 – Tests du choix de modèle

Modèles	Modèle homme		Modèle femme	
	Resid.Dev	AIC	Resid.Dev	AIC
Modèle 1	2897,8	2945,798	2944,6	2986,56
Modèle 2	2898,6	2942,563	2954,5	2980,495

Source : Auteurs, 2022

Dans les deux cas, en comparant l'AIC de chaque modèle, on remarque que le modèle 2 a l'AIC le plus faible (2942,563 et 2980,495). Ainsi, le meilleur modèle retenu pour la prédition de notre variable d'intérêt est le modèle 2 dans les deux cas.

3.3.2.2 Résultats de la régression

TABLEAU 3.10 – Résultats de l'estimation du probit selon le genre

Variables	Modèle homme			Modèle femme		
	Coefficient	Odds Ratio	p-value	Coefficient	Odds Ratio	p-value
Age (référence = 30-34ans)						

Variables	Modèle homme			Modèle femme		
	Coefficient	Odds Ratio	p-value	Coefficient	Odds Ratio	p-value
35-39ans	0.1239	1.1319	0.1516	0.1449	1.1560	0.0381*
40-44ans	0.2430	1.2750	0.0068**	0.3847	1.4692	4,055e-07***
45-49ans	0.3805	1.4630	1,966e-05***	0.4339	1.5432	1,039e-08***
50-54ans	0.6327	1.8828	3,871e-11***	-	-	-
55-59ans	0.6277	1.8734	2,521e-09***	-	-	-
60-64ans	0.5693	1.7670	2,785e-06***	-	-	-
Occupation (référence = Occupé)						
Inoccupé	-	-	-	0.1692	1.1844	0.1031
Milieu de résidence (référence = Urbain)						
Rural	-	-	-	-0.0979	0.9067	0.1109
Région (référence = Nord)						
Sud	0.3432	1.4094	6,012e-06***	0.3422	1.4080	2,592e-06***
Centre	0.4531	1.5732	6,473e-11***	0.4089	1.5052	4,569e-09***
Instruction (référence = Sans instruction)						
Primaire	0.0675	1.0698	0.3238	-	-	-
Secondaire	0.2529	1.2878	0.0009***	-	-	-
Supérieure	-0.2430	0.7842	0.0653.	-	-	-
Etat matrimonial (référence = Jamais en union)						
Marié	0.1979	1.2189	0.2540	-	-	-
Vie avec partenaire	0.4586	1.5819	0.0110*	-	-	-
Veuf	0.3018	1.3522	0.2886	-	-	-
Divorcé	0.5083	1.6625	0.0646	-	-	-
Séparé	0.0966	1.1014	0.6720	-	-	-
Niveau de vie (référence = Très pauvre)						
Pauvre	0.1654	1.1799	0.0781.	0.0845	1.0882	0.3553
Moyen	0.2041	1.2264	0.0301*	0.0770	1.0800	0.4165
Riche	0.3421	1.4079	0.0002***	0.3232	1.3816	0.0003***
Très riche	0.4040	1.4978	5,914e-05***	0.4818	1.6190	3,185e-07***
Diabète (référence = Non)						
Oui	0.9375	2.5535	1,521e-06***	0.4788	1.6142	0.0126*

Notes : (ns) Non significatives ; (.) 10% ; (*) 5% ; (**) 1% ; (***) 0,1%

Source : Auteurs, 2022

3.3.2.3 Pertinence du modèle

Dans le modèle homme comme dans le modèle femme, on remarque que la valeur D/v est très proche de 1. De plus, les probabilités de Khi-deux des tests d'Hosmer-Lemeshow, des résidus de Pearson et des résidus de la déviance sont toutes supérieures au seuil de 5%, donc le modèle est bien adapté aux données. Ainsi, le modèle est pertinent.

TABLEAU 3.11 – Pertinence du modèle probit

Tests	Critère de significativité	
	Modèle homme	Modèle femme
Règle de pouce	D/v = 0.7991627	D/v = 0.9487781
Test de Hosmer-Lemeshow	Pr(Chi2)=0.950	Pr(Chi2)=0.469
Test de résidus de Pearson	Pr(Chi2)=0.4395	Pr(Chi2)=0.5330
Test de résidus de la déviance	Pr(Chi2)= 1	Pr(Chi2)=0.9797

Source : Auteurs, 2022

3.3.2.4 Qualité du modèle

TABLEAU 3.12 – Qualité du modèle

Tests	Valeurs	
	Modèle homme	Modèle femme
Taux d'erreur	0.1573	0.2024
Aire sous la courbe ROC	0,7142	0.6712

Source : Auteurs, 2022

Dans le modèle homme comme dans le modèle femme, le taux d'erreur du modèle estimé est inférieur à 0,5 ; donc le modèle est de bonne qualité. De plus, l'aire sous la courbe ROC des deux modèles (0,7142 et 0,6712) est supérieure à 0,60 ; le modèle est donc d'une bonne qualité prédictive.

3.3.2.5 Interprétations des résultats

Quand on passe aux modèles selon le genre, tout comme au niveau du modèle global, l'âge, la région, le niveau de vie et le statut diabétique ont un effet significatif sur l'HTA au seuil de 5% et se révèlent être des facteurs explicatifs de l'HTA que ce soit chez les hommes ou chez les femmes. Cependant, on remarque que l'instruction et l'état matrimonial se trouvent être un facteur explicatif de l'HTA au niveau des hommes uniquement. Les femmes ayant entre 35 et 49 ans, résidant au sud ou au centre, qui sont riches ou très riches et diabétiques présentent plus de risque d'avoir l'HTA. Les hommes de 30 à 69 ans, mariés avec un niveau d'instruction d'au moins le secondaire et ayant les mêmes caractéristiques que les femmes présentent plus de risque d'avoir l'HTA.

3.4 Vérification des hypothèses

TABLEAU 3.13 – Vérification des hypothèses

Hypothèses	Conclusion
H1 : les facteurs sociodémographiques et économiques, et les facteurs métaboliques sont les facteurs qui influent sur la prévalence de l'HTA	Validée
H2 : les femmes d'âges avancés, mariées, riches et diabétiques sont les plus susceptibles d'avoir l'HTA	Validée
H3 : les personnes souffrant de l'HTA constituent un ensemble hétérogène selon le genre	Validée

Source : Auteurs, 2022

Au vu des résultats des analyses, toutes nos hypothèses sont validées.

3.5 Préconisations opérationnelles

Au terme de cette étude et au regard des variables entrant dans l'explication des facteurs de la prévalence de l'hypertension artérielle, des actions concrètes doivent être mises en œuvre par les autorités politiques et les populations du Bénin pour mieux prévenir et réduire les risques et les décès qu'occasionne cette maladie. Ainsi donc, nous formulons les suggestions suivantes.

A l'endroit du Ministère de la Santé, il s'agira de :

- ✍ mettre en place des programmes d'éducation et de sensibilisation sur l'hypertension artérielle sur toute l'étendue du territoire ;
- ✍ former du personnel de santé à cet effet ;
- ✍ mettre à disposition des fonds à ce sujet ;
- ✍ lancer des campagnes de dépistage sur toute l'étendue du territoire.

A l'endroit de la population du Bénin, il s'agira de :

- ✍ prêter attention à sa santé cardiaque ;
- ✍ ne pas négliger sa santé cardiaque surtout quand on est diabétique et en âge avancé ;
- ✍ se rapprocher du personnel de santé dès qu'on a des signes inhabituels même les plus banales ;
- ✍ veiller à son hygiène de vie (alimentation saine et équilibré, activité physique régulière...)
- ✍ faire des consultations trimestrielles dans des centres de santé pour s'assurer de son bien être.

Conclusion

L'hypertension artérielle est un facteur de risque majeur et grandissant, et si elle n'est pas contrôlée, elle peut entraîner des problèmes cardiovasculaires majeurs plus tard dans la vie. Tout comme les autres pays, le Bénin est confronté à cette maladie.

Afin de dégager et d'analyser les principaux facteurs explicatifs de l'hypertension artérielle dans notre population d'étude, nous avons exécuté une régression probit binaire sur 6776 individus à l'aide du logiciel R. Les résultats de cette analyse révèlent que l'âge, le genre, la région, le niveau d'instruction, le niveau de vie et l'état diabétique sont des facteurs qui influencent la prévalence de l'hypertension artérielle. Aussi, les modèles selon le genre montrent-ils que ces facteurs explicatifs varient quand on passe des hommes aux femmes. L'analyse des effets marginaux révèle que l'état diabétique et l'âge avancé sont les facteurs principaux de la présence de l'hypertension artérielle dans notre population d'étude.

Par ailleurs, le manque de données sur certaines variables comme l'IMC, l'obésité, l'hérédité de l'hypertension, l'alcoolisme, le tabagisme et bien d'autres variables ne nous a pas permis de tirer des conclusions sur ces facteurs à risque présents dans d'autres études.

Au vu de ces résultats, nous suggérons aux décideurs d'éduquer et de sensibiliser les populations sur les facteurs à risque de l'hypertension artérielle. Nous leur suggérons aussi de rendre le diagnostic et le traitement de l'hypertension artérielle facile et accessible pour toute la population. La population quant à elle doit veiller particulièrement à la santé cardiaque des personnes âgées et diabétiques.

Références

- [1] Berraho, M., Bendahhou, K., Benslimane, A., El Achhab, Y., El Rhazi, K., et Nejjari, C. (2009). « Etude des connaissances des facteurs de risque comportementaux des cancers chez la population marocaine ». *Revue d'Épidémiologie et de Santé Publique*. <https://doi.org/10.1016/j.respe.2009.02.053> .
- [2] Chobanian AV et al. Le septième rapport du Comité national mixte sur la prévention, la détection, l'évaluation et le traitement de l'hypertension artérielle : le rapport JNC 7. *JAMA* 2003.
- [3] Christophe Chesneau (2020), « Modèles de régression ». Université de Caen
- [4] Dauphinot, V., Naudin, F., Guéguen, R., Perronnin, M., et Sermet, C. (2008). « Écarts entre morbidité déclarée et morbidité diagnostiquée ». *Revue française des affaires sociales*.
- [5] Fondation pour la Recherche Médicale . « Écarts entre morbidité déclarée et morbidité diagnostiquée ». *Revue française des affaires sociales*.
- [6] Forouzanfar, M. H. et al. « Les maladies cardiovasculaires ». <https://www frm.org/recherches-maladies-cardiovasculaires#:text=Les%20maladies%20cardiovasculaires%20regroupent%20les,encore%20les%20accidents%20vasculaires%20cérébraux>.
- [7] Fourcade, L., Paule, P., et Mafart, B. (2007). « Hypertension artérielle en Afrique subsaharienne actualité et perspectives ». *Médecine Tropicale*.
- [8] Frérot, L., Pape, A. L., et Sermet, C. (1999). « L'hypertension artérielle en France : Prévalence et prise en charge thérapeutique ». Centre de recherche d'étude et de documentation de la santé.
- [9] Houenassi MD, Codjo LH, Dokou D, et al. « Prise en charge de l'hypertension artérielle à Cotonou ville, Bénin : connaissances, attitudes et pratiques des médecins généralistes ». *Cardio-vasculaire J Afr. 2016*. Publié le 23 août 2016.
- [10] Inamo, J. (2008.). « Aspects Epidémiologiques de l'Hypertension Artérielle aux Antilles-Guyane ». France : Université de Toulouse. <http://thesesups.ups-tlse.fr/300/1> .
- [11] Katchunga, P.B., M'buyamba-Kayamba, J.R., Masumbuko, B.E., et al. (2011) « Hypertension in the Adult Congolese Population of Southern Kivu : Results of the Vitaraa Study ». *La Presse*

Médicale. <https://doi.org/10.1016/j.lpm.2010.10.036> .

- [12] Lang, T., Bureau, J. F., Degoulet, P., Salah, H., et Benattar, C. (1983). « Blood pressure, coffee, tea and tobacco consumption : An epidemiological study in Algiers ». *European Heart Journal*. <https://doi.org/10.1093/oxfordjournals.eurheartj.a061531> .
- [13] Dictionnaire larousse en ligne, <https://www.larousse.fr/dictionnaires/francais/pr%C3%A9valence/63858#text=Rapport%20du%20nombre%20de%20cas,ou%20pendant%20une%20p%C3%A9riode%20donn%C3%A9es>.
- [14] Macia, E., Duboz, P., et Gueye, L. (2015). « L'hypertension artérielle à Dakar : Prévalence, connaissance, traitement et contrôle ». *Bulletin de la Société de pathologie exotique*. <https://doi.org/10.1007/s13149-014-0369-2> .
- [15] Mills KT, Bundy JD, Kelly TN, et al. « Disparités mondiales de la prévalence et du contrôle de l'hypertension : analyse systématique d'études basées sur la population dans 90 pays ». *Circulation* . 2016;134(6) :441-450. doi :10.1161/CIRCULATIONAHA.115.018912
- [16] Ministère de la Santé du Bénin. Rapport final de l'enquête STEPS au Bénin en 2015. <http://www.who.int/chp/steps/2008-STEPS-Report-Benin.pdf> (27 novembre 2020).
- [17] Genève, Organisation mondiale de la Santé, 2011. (Résumé d'orientation en français : Rapport sur la situation mondiale des maladies non transmissibles 2010). <http://www.who.int/nmh/publications/ncd-report2010/fr/> .
- [18] OMS (2013). Panorama mondial de l'Hypertension : <https://apps.who.int/iris/bitstream/handle/10665/85334/WHO-DCO-WHD-2013.2-fre.pdf> .
- [19] OMS (2022).Les décès dus aux maladies non transmissibles sont en hausse en Afrique : <https://www.afro.who.int/fr/news/les-deces-dus-aux-maladies-non-transmissibles-sont-en-hausse-en-afrique> .
- [20] Ratovoson, R., Rasetarinera Rabarisoa, O., Rogier, C., Piola, P., & Pacaud, P. (2014). L'hypertension artérielle chez les adultes en milieu rural à Moramanga, Madagascar. *Revue d'Épidémiologie et de Santé Publique*, 62(5), 229.<https://doi.org/10.1016/j.respe.2014.06.187>.
- [21] Rosengren, A., Hawken, S., Ôunpuu, S., Sliwa, K., Zubaid, M., Almahmeed, W. A., Blackett, K. N., Sitthi-amorn, C., Sato, H., et Yusuf, S. (2004). « Association of psychosocial risk factors with risk of acute myocardial infarction in 11 119 cases and 13 648 controls from 52

- countries (the INTERHEART study) : Case-control study ». *The Lancet*, 364(9438), 953-962. [https://doi.org/10.1016/S0140-6736\(04\)17019-0](https://doi.org/10.1016/S0140-6736(04)17019-0) .
- [22] Sa, M., et Jl, L. (2014). « Facteurs de risque de l'hypertension artérielle chez les militaires congolais de la garnison de Kinshasa ». *Revue Médicale des Grands Lacs*, 3(1), 63-76.
- [23] Sackou, J. K., Tiadé, M. L., Hounsa, A. A., Malik, S. K., Coulibaly, M., Desquith, A. A., Kadjo, F. K., Agoua, S. A., Oga, S., & Kouadio, L. K. (2020). Prevalence and factors associated with hypertension in Anonkoi 3, a peri-urban area in Abidjan (Côte d'Ivoire). *Journal of Public Health in Africa*, 10(2), 90-94. <https://doi.org/10.4081/jphia.2019.1106>.
- [24] Séraphin, A., Jacques, V., Leopold, A. B., Precieux, O., Houétonjji, C. L., Albert, D. C., Adébayo, A., et Léonard, F. (2016). « Prévalence De La Pression Artérielle Élevée Et Les Facteurs Associés Dans Les Départements Du Borgou Et De l'Alibori Au Bénin En 2014 ». *European Scientific Journal, ESJ*, 12(33), 235. <https://doi.org/10.19044/esj.2016.v12n33p235> .
- [25] Wagner, A., Dallongeville, J., Ferrières, J., Amouyel, P., et Haas, B. (2007). « État des lieux sur l'hypertension artérielle en France en 2007 : L'étude Mona Lisa ». 4. BEH thématique, 49-50.
- [26] World Health Organization. Global action plan on physical activity 2018-2030 : more active people for a healthier world. World Health Organization, 2019.
- [27] Yahia-Berrouiguet, A., Benyoucef, M., Meguenni, K., et Brouri, M. (2011). « Enquête sur la prévalence des facteurs de risque de maladies cardiovasculaires à Tlemcen (Algérie) ». *Médecine des Maladies Métaboliques*, 5(4), 42-48. [https://doi.org/10.1016/S1957-2557\(11\)70071-2](https://doi.org/10.1016/S1957-2557(11)70071-2) .

Annexe

Cette section contient les sorties brutes de nos analyses sous le logiciel R ainsi que d'autres résultats de notre travail.

Annexe 1 : Test de Khi-deux

```
> chisq.test(age,hypertension)
Pearson's Chi-squared test
data: age and hypertension
X-squared = 94.267, df = 6, p-value < 2.2e-16

> chisq.test(genre,hypertension)
Pearson's Chi-squared test with Yates' continuity correction
data: genre and hypertension
X-squared = 22.561, df = 1, p-value = 2.036e-06

> chisq.test(milieu_residence,hypertension)
Pearson's Chi-squared test with Yates' continuity correction
data: milieu_residence and hypertension
X-squared = 42.256, df = 1, p-value = 8.008e-11

> chisq.test(region,hypertension)
Pearson's Chi-squared test
data: region and hypertension
X-squared = 193.14, df = 2, p-value < 2.2e-16

> chisq.test(instruction,hypertension)
Pearson's Chi-squared test
data: instruction and hypertension
X-squared = 53.982, df = 3, p-value = 1.132e-11

> chisq.test(etat_mat,hypertension)
Pearson's Chi-squared test
data: etat_mat and hypertension
X-squared = 64.844, df = 5, p-value = 1.207e-12
```

```

> chisq.test(occupation,hypertension)

Pearson's Chi-squared test with Yates' continuity correction

data: occupation and hypertension
X-squared = 0.74668, df = 1, p-value = 0.3875

> chisq.test(niveau_vie,hypertension)

Pearson's Chi-squared test

data: niveau_vie and hypertension
X-squared = 162.38, df = 4, p-value < 2.2e-16

> chisq.test(diabetes,hypertension)

Pearson's Chi-squared test with Yates' continuity correction

data: diabetes and hypertension
X-squared = 53.486, df = 1, p-value = 2.605e-13

> chisq.test(cigarette,hypertension)

Pearson's Chi-squared test with Yates' continuity correction

data: cigarette and hypertension
X-squared = 2.2353, df = 1, p-value = 0.1349

```

Annexe 2 : Choix du modèle

Estimation du modèle logit

```

> summary(modele_log)

Call:
glm(formula = hypertension ~ age + genre + milieu_residence +
    region + instruction + etat_mat + occupation + niveau_vie +
    diabetes, family = binomial(logit), data = BASE)

Deviance Residuals:
    Min      1Q      Median      3Q      Max
-1.5442 -0.6780 -0.5046 -0.3373  2.6178

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) -3.54617  0.30134 -11.768 < 2e-16 ***
age35-39ans  0.25717  0.10002  2.571  0.0101 *  
age40-44ans  0.59396  0.10474  5.671 1.42e-08 ***
age45-49ans  0.73320  0.10452  7.015 2.30e-12 ***
age50-54ans  1.20290  0.15071  7.981 1.45e-15 ***
age55-59ans  1.19362  0.16879  7.072 1.53e-16 ***
age60-64ans  1.10839  0.20018  5.537 3.08e-08 ***
genreFeminin 0.54280  0.07835  6.928 4.28e-12 ***
milieu_residenceRural -0.11735  0.07767 -1.511  0.1308
regionsud    0.61260  0.09748  6.284 3.29e-10 ***
regioncentre 0.80264  0.09080  8.839 < 2e-16 ***
instructionPrimaire -0.00726  0.08942 -0.081  0.9353
instructionSecondaire 0.25613  0.10253  2.498  0.0125 *  
instructionSupérieure -0.29592  0.18822 -1.572  0.1159
etat_matMarié(e)    0.09268  0.24051  0.385  0.7000
etat_matVie avec un partenaire 0.43077  0.24778  1.739  0.0821 .
etat_matVeuf(vé)    0.34613  0.29868  1.159  0.2465
etat_matDivorce(e)   0.44561  0.35640  1.250  0.2112
etat_matSeparé(e)   0.23670  0.29071  0.814  0.4155
occupationoccupé    0.17750  0.15566  1.140  0.2541
niveau_viePauvre    0.24914  0.12288  2.027  0.0426 *  
niveau_vieMoyen     0.29930  0.12423  2.409  0.0160 *  
niveau_vieRiche     0.63280  0.12029  5.260 1.44e-07 ***
niveau_vieTrès riche 0.79690  0.13090  6.088 1.14e-09 ***
diabetesOui        1.17464  0.22403  5.243 1.58e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6368 on 6775 degrees of freedom
Residual deviance: 5875 on 6751 degrees of freedom
AIC: 5925

Number of Fisher Scoring iterations: 5

```

Estimation du modèle probit

```

> summary(modele_prob)

Call:
glm(formula = hypertension ~ age + genre + milieu_residence +
    region + instruction + etat_mat + occupation + niveau_vie +
    diabetes, family = binomial(probit), data = BASE)

Deviance Residuals:
    Min      1Q   Median      3Q      Max
-1.5127 -0.6825 -0.5079 -0.3221  2.7015

Coefficients:
                                         Estimate Std. Error z value Pr(>|z|)
(Intercept)                      -2.005858  0.161873 -12.392 < 2e-16 ***
age35-39ans                      0.134055  0.054549  2.458  0.0140 *
age40-44ans                      0.322690  0.058113  5.553 2.81e-08 ***
age45-49ans                      0.406890  0.058275  6.982 2.90e-12 ***
age50-54ans                      0.673441  0.086367  7.797 6.32e-15 ***
age55-59ans                      0.659240  0.096274  6.848 7.51e-12 ***
age60-64ans                      0.606410  0.114033  5.318 1.05e-07 ***
genre_feminin                     0.303438  0.043438  6.986 2.83e-12 ***
milieu_residenceRural            -0.069227  0.043489 -1.592  0.1114
region_sud                         0.336955  0.053249  6.328 2.49e-10 ***
region_centre                      0.445017  0.049123  9.059 < 2e-16 ***
instruction_Primaire              -0.001974  0.050562 -0.039  0.9689
instruction_Secondaire            0.149663  0.058915  2.540  0.0111 *
instruction_Superieure            -0.178615  0.105646 -1.691  0.0909 .
etat_mat_Mariee(e)                0.035652  0.129171  0.276  0.7825
etat_mat_Vie_avec_un_partenaire  0.229754  0.134035  1.714  0.0865 .
etat_mat_Veuf(ve)                 0.175973  0.166318  1.058  0.2900
etat_mat_Divorce(e)                0.248625  0.200051  1.243  0.2139
etat_mat_Separe(e)                 0.118517  0.160235  0.740  0.4595
occupation_occupé                  0.096553  0.086521  1.116  0.2644
niveau_vie_pauvre                  0.140965  0.065550  2.150  0.0315 *
niveau_vie_moyen                  0.160241  0.066709  2.402  0.0163 *
niveau_vie_riche                   0.348049  0.065554  5.309 1.10e-07 ***
niveau_vie_tres_riche              0.443435  0.072398  6.125 9.07e-10 ***
diabetesoui                        0.716479  0.136175  5.261 1.43e-07 ***
...
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6368.0  on 6775  degrees of freedom
Residual deviance: 5872.6  on 6751  degrees of freedom
AIC: 5922.6

Number of Fisher Scoring iterations: 5

```

Critère AIC des deux modèles

```

> modele_log$aic
[1] 5924.997
> modele_prob$aic
[1] 5922.56
>

```

Annexe 3 : Sélection des variables du modèle

Probit avec les variables liées par Khi-deux

```

> summary(modele_prob)

Call:
glm(formula = hypertension ~ age + genre + milieu_residence +
    region + instruction + etat_mat + occupation + niveau_vie +
    diabete, family = binomial(probit), data = BASE)

Deviance Residuals:
    Min      1Q   Median      3Q      Max
-1.5127 -0.6825 -0.5079 -0.3221  2.7015

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.005858  0.161873 -12.392 < 2e-16 ***
age35-39ans  0.134055  0.054549  2.458  0.0140 *
age40-44ans  0.322690  0.058113  5.553 2.81e-08 ***
age45-49ans  0.406890  0.058275  6.982 2.90e-12 ***
age50-54ans  0.673441  0.086367  7.797 6.32e-15 ***
age55-59ans  0.659240  0.096274  6.848 7.51e-12 ***
age60-64ans  0.606410  0.114033  5.318 1.05e-07 ***
genrefeminin 0.303438  0.043436  6.986 2.83e-12 ***
milieu_residenceRural -0.069227  0.043489 -1.592  0.1114
regionsud  0.336955  0.053249  6.328 2.49e-10 ***
regioncentre 0.445017  0.049123  9.059 < 2e-16 ***
instructionPrimaire -0.001974  0.050562 -0.039  0.9689
instructionSecondaire 0.149663  0.058915  2.540  0.0111 *
instructionSuperieure -0.178615  0.105646 -1.691  0.0909 .
etat_matMarié(e)  0.035652  0.129171  0.276  0.7825
etat_matVie avec un partenaire 0.229754  0.134035  1.714  0.0865 .
etat_matVeuf(e)  0.175973  0.166318  1.058  0.2900
etat_matDivorce(e) 0.248625  0.200051  1.243  0.2139
etat_matSeparé(e) 0.118517  0.160235  0.740  0.4595
occupationoccupé  0.096553  0.086521  1.116  0.2644
niveau_viePauvre  0.140965  0.065550  2.150  0.0315 *
niveau_vieMoyen  0.160241  0.066709  2.402  0.0163 *
niveau_vieRiche  0.348049  0.065554  5.309 1.10e-07 ***
niveau_vieTres riche 0.443435  0.072398  6.125 9.07e-10 ***
diabeteOui  0.716479  0.136175  5.261 1.43e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6368.0 on 6775 degrees of freedom
Residual deviance: 5872.6 on 6751 degrees of freedom
AIC: 5922.6

Number of Fisher Scoring iterations: 5

```

Méthode Step

```

> step.model <- step(modele_prob, direction = "both", k=2)
Start:  AIC=5922.56
hypertension ~ age + genre + milieu_residence + region + instruction +
    etat_mat + occupation + niveau_vie + diabete

              Df Deviance    AIC
- occupation  1  5873.8 5921.8
<none>          5872.6 5922.6
- milieu_residence 1  5875.1 5923.1
- instruction  3  5886.5 5930.5
- etat_mat      5  5890.6 5930.6
- diabete       1  5900.3 5948.3
- niveau_vie    4  5918.6 5960.6
- genre         1  5922.0 5970.0
- region        2  5958.4 6004.4
- age           6  5991.7 6029.7

Step:  AIC=5921.82
hypertension ~ age + genre + milieu_residence + region + instruction +
    etat_mat + niveau_vie + diabete

              Df Deviance    AIC
<none>          5873.8 5921.8
- milieu_residence 1  5876.2 5922.2
+ occupation      1  5872.6 5922.6
- instruction     3  5887.9 5929.9
- etat_mat        5  5892.1 5930.1
- diabete         1  5901.6 5947.6
- niveau_vie      4  5920.0 5960.0
- genre           1  5922.0 5968.0
- region          2  5961.6 6005.6
- age             6  5992.1 6028.1
> |

```

Annexe 4 : Régression probit

```

> summary(modele_probit)

Call:
glm(formula = hypertension ~ age + genre + milieu_residence +
    region + instruction + etat_mat + niveau_vie + diabetes,
    family = binomial(probit), data = BASE)

Deviance Residuals:
    Min      1Q  Median      3Q      Max
-1.5081 -0.6842 -0.5083 -0.3190  2.7046

Coefficients:
                                         Estimate Std. Error z value Pr(>|z|)
(Intercept)                         -1.922744  0.142935 -13.452 < 2e-16 ***
age35-39ans                          0.135738  0.054509  2.490  0.0128 *
age40-44ans                          0.324601  0.058080  5.589 2.28e-08 ***
age45-49ans                          0.407669  0.058260  6.997 2.61e-12 ***
age50-54ans                          0.671530  0.086376  7.774 7.58e-15 ***
age55-59ans                          0.654540  0.096242  6.801 1.04e-11 ***
age60-64ans                          0.593552  0.113652  5.223 1.76e-07 ***
genreFeminin                         0.296755  0.043037  6.895 5.37e-12 ***
milieu_residenceRural                -0.067604  0.043466 -1.555  0.1199
regionsud                            0.339459  0.053205  6.380 1.77e-10 ***
regioncentre                         0.448842  0.049011  9.158 < 2e-16 ***
instructionPrimaire                  -0.001983  0.050559 -0.039  0.9687
instructionSecondaire                 0.148362  0.058900  2.519  0.0118 *
instructionSupérieure                -0.182624  0.105653 -1.729  0.0839 .
etat_matMarié(e)                   0.043856  0.129123  0.340  0.7341
etat_matvie avec un partenaire      0.239158  0.133938  1.786  0.0742 .
etat_matveuf(ve)                   0.183217  0.166326  1.102  0.2707
etat_matDivorce(e)                  0.256267  0.199986  1.281  0.2000
etat_matSeparé(e)                  0.125971  0.160221  0.786  0.4317
niveau_viePauvre                    0.141130  0.065537  2.153  0.0313 *
niveau_vieMoyen                     0.162010  0.066681  2.430  0.0151 *
niveau_vieRiche                     0.348820  0.065541  5.322 1.03e-07 ***
niveau_vieTrès riche                0.444273  0.072386  6.138 8.38e-10 ***
diabetesOui                         0.716098  0.136105  5.261 1.43e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6368.0 on 6775 degrees of freedom
Residual deviance: 5873.8 on 6752 degrees of freedom
AIC: 5921.8

Number of Fisher Scoring iterations: 5

```

Odds.ratio

```

> odds.ratio(modele_probit)
Waiting for profiling to be done...
                                         OR      2.5 % 97.5 %      p
(Intercept)                         0.14621 0.10999 0.1923 < 2.2e-16 ***
age35-39ans                          1.14538 1.02946 1.2745  0.01277 *
age40-44ans                          1.38348 1.23476 1.5501 2.285e-08 ***
age45-49ans                          1.50331 1.34147 1.6848 2.608e-12 ***
age50-54ans                          1.95723 1.65168 2.3168 7.577e-15 ***
age55-59ans                          1.92426 1.59189 2.3214 1.039e-11 ***
age60-64ans                          1.81041 1.44497 2.2589 1.765e-07 ***
genreFeminin                         1.34549 1.23692 1.4640 5.372e-12 ***
milieu_residenceRural                0.93463 0.85838 1.0178  0.11986
regionsud                            1.40419 1.26602 1.5578 1.768e-10 ***
regioncentre                         1.56650 1.42323 1.7249 < 2.2e-16 ***
instructionPrimaire                  0.99802 0.90383 1.1015  0.96872
instructionSecondaire                 1.15993 1.03344 1.3013  0.01177 *
instructionSupérieure                0.83308 0.67570 1.0229  0.08389 .
etat_matMarié(e)                   0.04483 0.81653 1.3524  0.73413
etat_matvie avec un partenaire      1.27018 0.98265 1.6592  0.07417 .
etat_matveuf(ve)                   1.20108 0.86898 1.6683  0.27065
etat_matDivorce(e)                  1.29210 0.87373 1.9084  0.20005
etat_matSeparé(e)                  1.13425 0.83143 1.5562  0.43173
niveau_viePauvre                    1.15157 1.01296 1.3097  0.03129 *
niveau_vieMoyen                     1.17587 1.03177 1.3405  0.01511 *
niveau_vieRiche                     1.41739 1.24710 1.6120 1.025e-07 ***
niveau_vieTrès riche                1.55936 1.35387 1.7973 8.378e-10 ***
diabetesOui                         2.04643 1.57004 2.6655 1.430e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>

```

Annexe 5 : Tests de pertinence

▷ Règle de pouce

```
> #REGLE DU POUCE
> library(stats)
> deviance(modele_probit)/df.residual(modele_probit)###0,8703103#PROCHE DE 1
[1] 0.8699381
> |
```

▷ Hosmer-Lemeshow

```
> #TEST DE HOSMER-LEMESHOW
> library(performance)
> performance_hosmer(modele_probit)
# Hosmer-Lemeshow Goodness-of-Fit Test
Chi-squared: 8.007
df: 8
p-value: 0.433
Summary: model seems to fit well.
> |
```

▷ Tests de Pearson et test de déviance

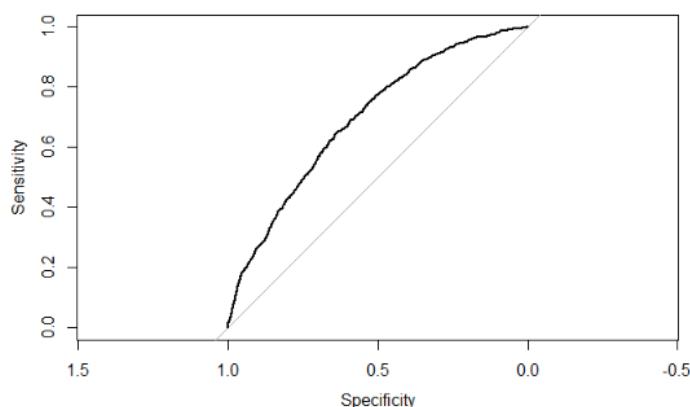
```
> #TEST DES RESIDUS DE PEARSON & TEST DES RESIDUS DE LA DEVIANCE
> s2=sum(residuals(modele_probit,type="pearson")^2)
> dd1=df.residual(modele_probit)
> pvaleur=1-pchisq(s2,dd1)
> pvaleur
[1] 0.5483187
>
> pvaleuri1=1-pchisq(deviance(modele_probit),df.residual(modele_probit))
> pvaleuri
[1] 1
> |
```

Annexe 6 : Qualité du modèle

▷ Taux d'erreur

```
> #MARGE D_ERREUR (MATRICE DE CONFONSION/CONFUSION)
> pred.prob.probit = predict(modele_probit, type = "response")
> pred.mod.probit = factor(ifelse(pred.prob.probit > 0.5, "1", "0"))
> mc.probit = table(BASE$hypertension, pred.mod.probit)
> t1 = (mc.probit[1, 2] + mc.probit[2, 1]) / sum(mc.probit)
> t1
[1] 0.1779811
> |
```

▷ Courbe ROC



```
> auc(roc_object)
Area under the curve: 0.6944
> |
```

Annexe 7 : Régression probit selon le genre masculin

```
> summary(modele_probit_H)

Call:
glm(formula = hypertension ~ age + region + instruction + etat_mat +
    niveau_vie + diabetes, family = binomial(probit), data = BASE_HOMME)

Deviance Residuals:
    Min      1Q  Median      3Q      Max
-1.7157 -0.6241 -0.4606 -0.3046  2.6947

Coefficients:
                                         Estimate Std. Error z value Pr(>|z|)
(Intercept)                         -2.13292  0.18830 -11.327 < 2e-16 ***
age35-39ans                          0.12396  0.08645  1.434  0.151617
age40-44ans                          0.24302  0.08984  2.705  0.006828 **
age45-49ans                          0.38056  0.08915  4.269  1.97e-05 ***
age50-54ans                          0.63278  0.09575  6.609  3.87e-11 ***
age55-59ans                          0.62778  0.10533  5.960  2.52e-09 ***
age60-64ans                          0.56930  0.12149  4.686  2.79e-06 ***
regionsud                           0.34322  0.07583  4.526  6.01e-06 ***
regioncentre                         0.45311  0.06936  6.532  6.47e-11 ***
instructionPrimaire                 0.06756  0.06848  0.987  0.323882
instructionSecondaire                0.25295  0.07638  3.312  0.000927 ***
instructionSupérieure                -0.24307  0.13190 -1.843  0.065355 .
etat_matMarié(e)                   0.19798  0.17358  1.141  0.254039
etat_matVie avec un partenaire     0.45867  0.18056  2.540  0.011076 *
etat_matVeuf(ve)                   0.30180  0.28443  1.061  0.288652
etat_matDivorce(e)                 0.50834  0.27511  1.848  0.064630 .
etat_matSeparé(e)                 0.09660  0.22816  0.423  0.672022
niveau_viePauvre                    0.16549  0.09396  1.761  0.078193 .
niveau_vieMoyen                    0.20414  0.09416  2.168  0.030148 *
niveau_vieRiche                     0.34214  0.09418  3.633  0.000280 ***
niveau_vieTres riche                0.40405  0.10061  4.016  5.91e-05 ***
diabetesOui                         0.93750  0.19497  4.808  1.52e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3189.3 on 3648 degrees of freedom
Residual deviance: 2898.6 on 3627 degrees of freedom
AIC: 2942.6

Number of Fisher Scoring iterations: 5
```

Odds.ratio

```
> odds.ratio(modele_probit_H)
Waiting for profiling to be done...
                                         OR      2.5 % 97.5 %      p
(Intercept)                         0.118491 0.080971 0.1691 < 2.2e-16 ***
age35-39ans                          1.131966 0.955502 1.3413 0.1516173
age40-44ans                          1.275096 1.069215 1.5207 0.0068282 **
age45-49ans                          1.463099 1.229151 1.7423 1.966e-05 ***
age50-54ans                          1.882839 1.561066 2.2713 3.871e-11 ***
age55-59ans                          1.873444 1.523859 2.3015 2.521e-09 ***
age60-64ans                          1.767028 1.389987 2.2397 2.785e-06 ***
regionsud                           1.409480 1.216181 1.6345 6.012e-06 ***
regioncentre                         1.573202 1.373398 1.8035 6.473e-11 ***
instructionPrimaire                 1.069892 0.935418 1.2229 0.3238821
instructionSecondaire                1.287822 1.108968 1.4947 0.0009272 ***
instructionSupérieure                0.784216 0.603337 1.0122 0.0653549 .
etat_matMarié(e)                   1.218936 0.879074 1.7365 0.2540392
etat_matVie avec un partenaire     1.581971 1.123303 2.2836 0.0110757 *
etat_matVeuf(ve)                   1.352293 0.766501 2.3655 0.2886520
etat_matDivorce(e)                 1.662535 0.971021 2.8448 0.0646299 .
etat_matSeparé(e)                 1.101414 0.706709 1.7308 0.6720222
niveau_viePauvre                    1.179976 0.981486 1.4201 0.0781926 .
niveau_vieMoyen                    1.226475 1.019991 1.4765 0.0301478 *
niveau_vieRiche                     1.407954 1.171009 1.6956 0.0002804 ***
niveau_vieTres riche                1.497880 1.230679 1.8262 5.914e-05 ***
diabetesOui                         2.553594 1.754515 3.7215 1.521e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

Annexe 8 : Tests de pertinence

▷ Règle de pouce

```
> deviance(modele_probit_H)/df.residual(modele_probit_H)
[1] 0.7991627
> |
```

▷ Hosmer-Lemeshow

```
> performance_hosmer(modele_probit_H)
# Hosmer-Lemeshow Goodness-of-Fit Test

Chi-squared: 2.742
df: 8
p-value: 0.950

Summary: model seems to fit well.
> |
```

▷ Tests de Pearson et test de déviance

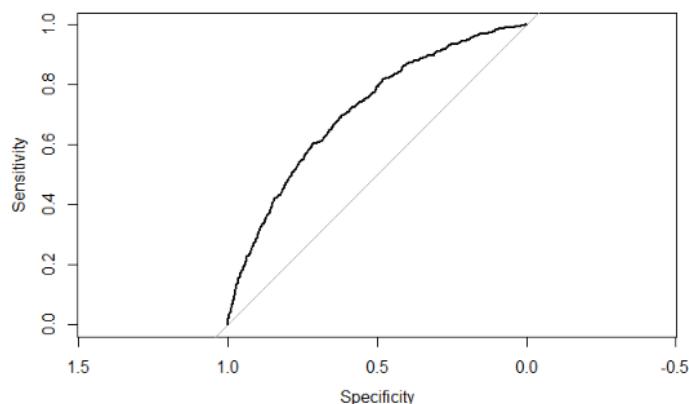
```
> #TEST DES RESIDUS DE PEARSON & TEST DES RESIDUS DE LA DEVIANCE
> s2_H=sum(residuals(modele_probit_H,type="pearson")^2)
> ddi_H=df.residual(modele_probit_H)
> pvalueur_H=1-pchisq(s2_H,ddi_H)
> pvalueur_H
[1] 0.4395641
> pvalueur_H=1-pchisq(deviance(modele_probit_H),df.residual(modele_probit_H))
> pvalueur_H
[1] 1
> |
```

Annexe 9 : Qualité du modèle

▷ Taux d'erreur

```
> #MARGE D_ERREUR (MATRICE DE CONFONSION/CONFUSION)
> pred.prob.probit_H = predict(modele_probit_H, type = "response")
> pred.mod.probit_H = factor(ifelse(pred.prob.probit_H > 0.5, "1", "0"))
> mc.probit_H = table(BASE_HOMME$hypertension, pred.mod.probit_H)
> t1_H = (mc.probit_H[1, 2] + mc.probit_H[2, 1]) / sum(mc.probit_H)
> t1_H
[1] 0.1573034
> |
```

▷ Courbe ROC



```
> auc(roc_object_H)
Area under the curve: 0.7142
> |
```

Annexe 10 : Régression probit selon le genre féminin

```
> summary(modele_probit_F)

Call:
glm(formula = hypertension ~ age + milieu_residence + region +
    occupation + niveau_vie + diabetes, family = binomial(probit),
    data = BASE_FEMME)

Deviance Residuals:
    Min      1Q   Median      3Q      Max
-1.3597 -0.7431 -0.5496 -0.3644  2.4473

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.64435  0.12598 -13.053 < 2e-16 ***
age35-39ans  0.14498  0.06992  2.074 0.038112 *
age40-44ans  0.38474  0.07594  5.066 4.06e-07 ***
age45-49ans  0.43391  0.07580  5.724 1.04e-08 ***
milieu_residenceRural -0.09790  0.06142 -1.594 0.110952
regionsud    0.34222  0.07280  4.701 2.59e-06 ***
regioncentre 0.40897  0.06976  5.862 4.57e-09 ***
occupationoccupé 0.16926  0.10385  1.630 0.103134
niveau_viePauvre 0.08454  0.09146  0.924 0.355315
niveau_vieMoyen 0.07701  0.09479  0.812 0.416545
niveau_vieRiche 0.32326  0.09030  3.580 0.000344 ***
niveau_vieTres riche 0.48182  0.09425  5.112 3.19e-07 ***
diabetesOui   0.47885  0.19198  2.494 0.012621 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3156.0 on 3126 degrees of freedom
Residual deviance: 2954.5 on 3114 degrees of freedom
AIC: 2980.5

Number of Fisher Scoring iterations: 5
```

Odds.ratio

```
> odds.ratio(modele_probit_F)
Waiting for profiling to be done...
            OR  2.5 % 97.5 %      p
(Intercept) 0.19314 0.15024 0.2465 < 2.2e-16 ***
age35-39ans 1.15602 1.00822 1.3257 0.0381125 *
age40-44ans 1.46923 1.26604 1.7051 4.055e-07 ***
age45-49ans 1.54328 1.33056 1.7902 1.039e-08 ***
milieu_residenceRural 0.90674 0.80404 1.0228 0.1109523
regionsud   1.40807 1.22170 1.6237 2.592e-06 ***
regioncentre 1.50526 1.31379 1.7260 4.569e-09 ***
occupationoccupé 1.18442 0.96944 1.4564 0.1031341
niveau_viePauvre 1.08821 0.91027 1.3017 0.3553146
niveau_vieMoyen 1.08005 0.89658 1.3011 0.4165447
niveau_vieRiche 1.38162 1.15857 1.6496 0.0003438 ***
niveau_vieTres riche 1.61902 1.34712 1.9488 3.185e-07 ***
diabetesOui   1.61422 1.10375 2.3509 0.0126206 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

Annexe 11 : Tests de pertinence

▷ Règle de pouce

```
> #REGLE DU POUCE
> deviance(modele_probit_F)/df.residual(modele_probit_F)
[1] 0.9487781
> |
```

▷ Hosmer-Lemeshow

```
> #TEST DE HOSMER-LEMESHOW
> library(performance)
> performance_hosmer(modele_probit_F)
# Hosmer-Lemeshow Goodness-of-Fit Test

Chi-squared: 7.643
df: 8
p-value: 0.469

Summary: model seems to fit well.
> |
```

▷ Tests de Pearson et test de déviance

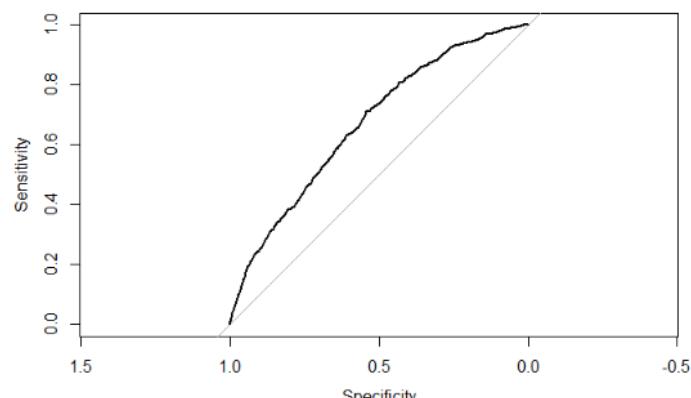
```
> #REGLE DU POUCE
> deviance(modele_probit_F)/df.residual(modele_probit_F)
[1] 0.9487781
> #TEST DES RESIDUS DE PEARSON & TEST DES RESIDUS DE LA DEVIANCE
> s2_F=sum(residuals(modele_probit_F,type="pearson")^2)
> ddi_F=df.residual(modele_probit_F)
> pvalueur_F=1-pchisq(s2_F,ddi_F)
> pvalueur_F
[1] 0.533034
> pvalueur_F=1-pchisq(deviance(modele_probit_F),df.residual(modele_probit_F))
> pvalueur_F
[1] 0.9797323
> |
```

Annexe 12 : Qualité du modèle

▷ Taux d'erreur

```
> #MARGE D_ERREUR (MATRICE DE CONFONSION/CONFUSION)
> pred.prob.probit_F = predict(modele_probit_F, type = "response")
> pred.mod.probit_F = factor(ifelse(pred.prob.probit_F > 0.5, "1", "0"))
> mc.probit_F = table(BASE_FEMME$hypertension, pred.mod.probit_F)
> t1_F = (mc.probit_F[1, 2] + mc.probit_F[2, 1]) / sum(mc.probit_F)
> t1_F
[1] 0.2024304
> |
```

▷ Courbe ROC



```
> auc(roc_object_F)
Area under the curve: 0.6712
> |
```

Annexe 13 : Effets marginaux

factor	AME	SE	z	p	lower	upper
1 age35-39ans	0.0278	0.0112	2.4874	0.0129	0.0059	0.0497
2 age40-44ans	0.0729	0.0134	5.4600	0.0000	0.0467	0.0991
3 age45-49ans	0.0951	0.0140	6.7949	0.0000	0.0677	0.1225
4 age50-54ans	0.1744	0.0258	6.7637	0.0000	0.1238	0.2249
5 age55-59ans	0.1689	0.0289	5.8452	0.0000	0.1123	0.2255
6 age60-64ans	0.1496	0.0337	4.4356	0.0000	0.0935	0.2158
7 chaleureux	0.1235	0.0403	4.5335	0.0000	0.1238	0.1312
8 etat_matDivorce(e)	0.0641	0.0512	1.2513	0.2108	-0.0363	0.1644
9 etat_matMari(e)	0.0100	0.0289	0.3461	0.7293	-0.0467	0.0667
10 etat_matSepar(e)	0.0100	0.0289	0.3461	0.7293	-0.0467	0.0667
11 etat_matVeuf/ve	0.0444	0.0399	1.1140	0.2653	-0.0338	0.1227
12 etat_vie avec un partenaire	0.0594	0.0307	1.9343	0.0531	-0.0000	0.1198
13 genreFemale	0.0720	0.0104	6.5982	0.0000	0.0116	0.6925
14 instructionPrimaire	-0.0005	0.0120	-0.0392	0.9687	-0.0241	0.0231
15 instructionSecondaire	0.0376	0.0154	2.4429	0.0146	0.0074	0.0678
16 instructionSupérieure	-0.0401	0.0216	-1.8577	0.0632	-0.0824	0.0022
17 milieu_residenceRural	-0.0163	0.0105	-1.5503	0.1211	-0.0370	0.0043
18 niveau_vieMoyen	0.0341	0.0140	2.4365	0.0146	0.0027	0.0615
19 niveau_vieRich	0.0384	0.0126	3.1001	0.0308	0.0027	0.0560
20 niveau_vieRich	0.0803	0.0148	5.4250	0.0000	0.0513	0.1093
21 niveau_vieTrs rich	0.1068	0.0173	6.1582	0.0000	0.0728	0.1407
22 regioncentre	0.1058	0.0115	9.2371	0.0000	0.0834	0.1283
23 regionsud	0.0761	0.0119	6.3658	0.0000	0.0526	0.0995

23 rows

Table des matières

Approbation	ii
Dédicace	iii
Remerciements	iv
Liste des acronymes	v
Résumé	ix
Abstract	x
Introduction	1
1 Cadre institutionnel de l'étude	3
1.1 Environnement du stage	3
1.1.1 Présentation et historique de la CNSS	3
1.1.2 Organisation et missions de la CNSS	4
1.2 Contexte du stage	6
1.2.1 Présentation de la Cellule des Etudes et de la Coopération	6
1.2.2 Activités menées lors du stage	6
2 Cadre théorique et méthodologique de l'étude	8
2.1 Problématique et intérêt de l'étude	8
2.1.1 Problématique	8
2.1.2 Intérêt de l'étude	9
2.2 Objectifs et hypothèses de l'étude	9
2.2.1 Objectifs de l'étude	9
2.2.2 Hypothèses	10
2.3 Revue de littérature	10
2.3.1 Définitions des concepts	10
2.3.2 Revue empirique	10
2.3.3 Revue méthodologique	12
2.4 Méthodologie	13
2.4.1 Nature et sources des données	13
2.4.2 Population d'étude	13
2.4.3 Méthodes et outils d'analyse	14
2.5 Limites de l'étude	22
3 Résultats, analyses et interprétations	23

3.1	Analyse descriptive	23
3.1.1	Analyse univariée	23
3.1.2	Analyse bivariée	24
3.2	Test de pré-estimation	26
3.3	Analyse économétrique	26
3.3.1	Estimation du modèle commun	27
3.3.2	Estimation des modèles selon le genre	32
3.4	Vérification des hypothèses	35
3.5	Préconisations opérationnelles	35
Conclusion		36
Références		36
Annexe		a