



ROYAUME DU MAROC
*-*_*-*_*-*
HAUT COMMISSARIAT AU PLAN
*-*_*-*_*-*_*-*_**

**INSTITUT NATIONAL
DE STATISTIQUE ET D'ECONOMIE APPLIQUEE**



INSEA

Projet de Fin d'Etudes

La santé maternelle au Maroc : Niveaux, Tendances et Facteurs explicatifs

Préparé par : **M. Mustapha OUL-HAJ**
M. Tarik ENNACHAT

Sous la direction de : **M. Mustapha BERROUYNE** (INSEA)
Mme Nassira EL IDRISI (CERED)

Soutenu publiquement comme exigence partielle en vue de l'obtention du

Diplôme d'Ingénieur d'Etat

Filière : *Statistique - Démographie*

Devant le jury composé de :

- M. Mustapha BERROUYNE (INSEA)
 - M. Abdellali FADLALLAH (INSEA)
 - Mme. Nassira EL IDRISI (CERED)



RÉSUMÉ



Résumé

La vision de cette présente étude est de donner les niveaux et les tendances des grands axes qui englobent la santé maternelle au Maroc ainsi que l'analyse des déterminants de la planification familiale comme l'une de ces principales matières. Nous avons utilisé une analyse des données sur un échantillon de 11 069 femmes enquêtées non-célibataires et âgées de 15 à 49 ans qui nous permettra de vérifier la significativité de chaque variable explicative retenue par notre étude.

Les données sont tirées de l'Enquête Nationale sur la Population et la Santé Familiale (ENPSF) réalisée par le Ministère de la santé en 2011, cette enquête a touché les différents volets de la santé maternelle et infantile, tout en collectant les informations relatives aux caractéristiques socio-démographiques des ménages et à leurs conditions d'habitation.

L'accès universel à la planification familiale (PF) est une préoccupation majeure du gouvernement marocain et ceci dans le but d'améliorer la santé maternelle et infantile. But auquel l'ENPSF à travers son programme de PF adhère pleinement. De ce fait, cette étude a pour objectif d'identifier les facteurs qui influencent significativement l'utilisation de la contraception par les femmes marocaines.

Nous avons effectué une analyse bi-variée pour évaluer les corrélations entre les variables étudiées. Les résultats ont montré que les variables « Milieu de résidence » et « Situation matrimoniale » sont liées à la planification familiale. Nous avons fait appel à l'analyse de la régression logistique, en plus de la méthode d'arbre de classification afin d'identifier les variables explicatives les plus significatives pour construire nos modèles. À partir de ces deux modèles, les résultats obtenus montrent que l'âge de la femme, le milieu de résidence, la situation matrimoniale, ainsi que le nombre des enfants vivants, influent à la non-participation de la planification familiale.

Mots clés : santé maternelle, planification familiale, femme, caractéristique sociodémographique, contraception.

A decorative horizontal line featuring four symmetrical scrollwork flourishes at its ends and a central decorative element in the middle.

DÉDICACE

Dédicace

Je dédie ce travail

À ALLAH, qui m'a donné la puissance et la perspective nécessaire pour élaborer ce travail.

À mon cher père Ahmed, tu es là pour m'éduquer et me guider. Avec toi, je sens en sécurité et je ne manque de rien. Je veux te remercier pour tout ce que tu fais.

À ma très chère mère Aicha source de tendresse, ma force, mon idole, la personne que j'aime le plus au monde, celle qui m'a donné la vie. Je te remercie pour ton amour et tes prières tout au long de ma construction.

À mon cher frère Achraf, tu es mon bras droit, je te souhaite que du bonheur et de la réussite AMINE.

À mon cher petit frère Yasser, mon âme, je te remercie pour ton bonheur et ta tendresse.

À mes chers camarades :

Abdelali OUALLA

Khalid KASSI

Mohammed ID AZZI

Tarik ENNACHAT, mon binôme

Aimane HAFID

Othman EL OUADI

À toutes les personnes qui ont participé à son élaboration et à tous ceux que j'ai omis de citer.

MUSTAPHA

Dédicace

C'est avec profonde gratitude, que j'aimerai dédier ce travail:

À la mémoire de ma grand-mère qui m'a soutenue tout au long de mes années d'études. Que Dieu la garde dans son vaste paradis.

*À ma très chère mère **LAILA***

Aucune dédicace ne saurait exprimer mon respect mon amour éternel et ma considération pour les sacrifices que vous avez consenti pour mon instruction et mon bien être. Je vous remercie pour tout le soutien et l'amour que vous me portez depuis mon enfance et j'espère que votre bénédiction m'accompagne toujours.

Que ce modeste travail soit l'exaucement de vos vœux tant formules, le fruit de vos innombrables sacrifices. Puisse Dieu, le très haut, vous accorder santé, bonheur et longue vie.

À nos encadrants **Mme. Nassira EL IDRISI** et **M. Mustapha BERROUYNE** pour tous ses précieux conseils, pour ses écoutes actives, ses disponibilités.

Je n'oublie surtout pas mon ami et mon binôme **Mustapha OUL-HAJ** pour son motivation pendant l'élaboration de notre travail.

TARIK



REMERCIEMENTS



Remerciements

Tout d'abord, nous tenons à remercier le bon Dieu le tout-puissant de nous avoir donné la force et le courage de mener à bien ce modeste travail.

Nous adressons nos profondes gratitude à M. BERROUYNE Mustapha, notre encadrant interne de ce projet, pour l'encadrement et l'accompagnement tout au long de ce travail, ainsi que pour son orientation, son soutien et sa patience.

Nos remerciements vont également à Madame EL IDRISI Nassira, notre encadrante au sein de l'organisme d'accueil, pour l'initiation à la démarche liée à la réalisation du présent rapport, au temps alloué, aux conseils et aux encouragements.

Nous ne manquons pas de remercier Madame BENLGOULLI Malika, la responsable des stagiaires pour nous avoir accordé sa confiance dans ce projet et pour toute l'aide qu'il nous a apporté afin de réussir ce travail.

Mille mercis à toute l'équipe du CERED, pour son accueil, sa disponibilité et sa bonne humeur permanente.

TABLE DES MATIÈRES

Résumé	3
Dédicace.....	5
Remerciement	8
Table de matière	10
Liste des abréviations	13
Liste des tableaux	14
Liste des figures	15
Liste des graphiques	16
Introduction général.....	17
 Chapitre I : Santé maternelle : une revue de littérature.....	19
Introduction.....	20
1. Histoire et évolution du concept de santé maternelle	20
2. Définition de la santé maternelle	21
3. Définition des concepts relatifs à la santé maternelle	21
3.1. Mortalité maternelle	21
3.2. Morbidité maternelle.....	23
3.3. Planification familiale	24
4. Revue de littérature : Les déterminants de la santé maternelle	25
4.1. Indicateurs et utilisation des services de santé maternelle	25
4.2. Indicateurs liés à la demande.....	26
4.3. Indicateurs liés à l'offre	27
5. Cadre conceptuel.....	27
Conclusion	28
 Chapitre II : Source de données et méthodologie.....	29
Introduction.....	30
1. Source de données	30
1.1. Population cible	30
1.2. Objectifs	30
1.3. Financement de l'enquête.....	31
1.4. Questionnaire.....	31
1.4.1. Questionnaire ménage	31
1.4.2. Questionnaire femme.....	31
1.5. Echantillonnage	32
1.6. Evaluation de qualité des données	33
1.6.1. Indice de MYERS	33
1.6.2. Indice combiné des Nations Unies	34

2. Choix des variables	34
3. Méthodologie.....	35
3.1. Analyse descriptive (analyse bi-variée)	36
3.2. Arbre de classification	37
3.2.1. Construction d'un arbre de classification	38
3.2.2. Validation du modèle d'arbre de classification.....	40
3.3. Analyse explicative.....	42
3.4.Outils d'analyse	46
Conclusion	46
Chapitre III : Niveaux et tendances de la santé maternelle au Maroc	47
Introduction.....	48
1. Niveaux et tendances de la mortalité maternelle	48
1.1 Niveau de la mortalité maternelle.....	48
1.2 Tendance de la mortalité maternelle.....	49
2. Niveaux et tendances de la morbidité maternelle	49
2.1 . Niveau de la morbidité maternelle.....	49
2.1.1. Selon l'âge de la femme.....	50
2.1.2. Selon le milieu de résidence.....	50
2.1.3. Selon le quantile de bien-être	50
2.2 Tendance de la mortalité maternelle	51
3. Niveaux et tendances de la planification familiale	52
3.1. Niveau de la planification familiale	52
3.1.1. Selon l'âge de la femme.....	52
3.1.2. Selon le milieu de résidence.....	53
3.1.3. Selon le quantile de bien-être	53
3.1.4. Selon la situation professionnelle de la femme	53
3.2. Tendance de la planification familiale	54
Conclusion	55
Chapitre IV : Présentations des résultats de l'analyse explicative	57
Introduction.....	58
1. Résultats d'analyse descriptive.....	58
1.1. Milieu de résidence.....	58
1.2. Niveau scolaire le plus atteint	58
1.3. Situation matrimoniale.....	58
1.4. Travail	59
1.5. Nombre de fois marié.....	59
2. Résultats des arbres de classification	59
3. Résultats de régression logistique.....	63

3.1. Rappel des variables retenues.....	63
3.2. Modélisation de la planification familiale.....	63
3.3. Interprétation des résultats	67
3.4. Comparaison entre modèle de régression logistique et celui d'arbre de classification.....	67
Conclusion	68
Conclusion générale	69
Bibliographie	71
Annexes	74

LISTE DES ABRÉVIATIONS

CART	Classification And Regression Trees
CERED	Centre d'Etudes et des Recherches Démographiques
CHAID	Chi-Square Automatic Interaction Detection
DIU	Dispositif Intra Utérin
ENPSF	Enquête National sur la Population et la Santé familiale
FNUP	Fonds des Nations Unies pour la Population
HCP	Haut Commissariat au Plan
MAMA	Méthode de l'Allaitement Maternal et de l'Aménorrhée
ODD	Objectifs de Développement Durable
OMS	Organisation Mondiale pour la Santé
PAPFAM	Pan Arab Project For Family Health
UNICEF	Fonds des Nations Unies pour l'Enfance

LISTE DES TABLEAUX

N°	Intitulé	Page
1	Les valeurs de l'indice de Myers pour les deux sexes	33
2	Répulsion et attraction détectés par chiffre terminal	34
3	Variables retenues	35
4	Evolution du Ratio de Mortalité Maternelle par milieu de résidence entre 2010-2016	48
5	Répartition en % des femmes âgées de 15-49 ans qui n'a aucun cancer et celle qui souffrent d'un cancer au moment de l'enquête au niveau national	49
6	Pourcentage des femmes non célibataires âgées de 15-49 ans qui souffrent d'une maladie cancéreuse selon le type de cancer et groupe d'âges	50
7	Pourcentage des femmes non célibataires âgées de 15-49 ans qui souffrent d'une maladie cancéreuse selon le type de cancer et le milieu de résidence	50
8	Pourcentage des femmes non célibataires âgées de 15-49 ans qui souffrent d'une maladie cancéreuse selon le type de cancer et le quantile de bien-être	51
9	Répartition en pourcentage des femmes mariées âgées de 15 à 49 ans qui utilisent au moment de l'enquête une méthode contraceptive selon la méthode utilisée et le groupe âges	52
10	Répartition en pourcentage des femmes mariées âgées de 15 à 49 ans qui utilisent au moment de l'enquête une méthode contraceptive selon la méthode utilisée et le milieu de résidence	53
11	Répartition en pourcentage des femmes mariées âgées de 15 à 49 ans qui utilisent au moment de l'enquête une méthode contraceptive selon la méthode utilisée et le quantile de bien-être	53
12	Répartition en pourcentage des femmes mariées âgées de 15 à 49 ans qui utilisent au moment de l'enquête une méthode contraceptive selon la méthode utilisée et la situation professionnelle	54
13	Matrice de confusion	62
14	Résultats du modèle global de régression logistique	63
15	Résultat final du modèle de régression logistique	65
16	Résultat final détaillé du modèle de régression logistique	65
17	Matrice de confusion	65
18	Comparaison des deux modèles	68

LISTE DES FIGURES

N°	Intitulé	Page
1	Cadre conceptuel des déterminants de la santé maternelle	28
2	Schéma d'un arbre	37
3	Répartition et taille des trois sous-bases créées	41
4	Schéma type de travail en analyse prédictive	59
5	Indicateurs de performances	62
6	Courbe ROC du modèle final	66

LISTE DES GRAPHIQUES

N°	Intitulé	Page
1	Evolution du RMM entre 2000-2017	49
2	Répartition des décès des femmes âgées de 15-49 ans (en %) selon la cause, selon différentes enquêtes réalisées au Maroc depuis 2003, ENPSF-2018	51
3	Répartition en % des femmes âgées de 15-49 ans qui utilisent au moment de l'enquête une méthode contraceptive au niveau national	52
4	Evaluation des taux de la prévalence contraceptive selon que la méthode utilisée est moderne ou traditionnelle, selon différentes enquêtes réalisées au Maroc depuis 1997, ENPSF- 2018	54
5	Répartition (en %) de Femmes mariées âgées de 15 à 49 ans qui utilisent au moment de l'enquête une méthode contraceptive selon la méthode utilisée, selon enquêtes réalisées au Maroc depuis 2011, ENPSF-2018	55

INTRODUCTION GÉNÉRALE

La santé maternelle, qualifiée comme moteur du développement social et économique, est une composante importante de la santé globale. Elle demeure encore l'une des préoccupations majeures des acteurs en charge des politiques de santé dans les pays en développement. En outre, au niveau africain, elle présente un enjeu majeur pour bénéficier des effets positifs de la croissance démographique à l'horizon 2050 (Rapport OMS, 2016). En 1985, Starrs a déclaré que plus d'un demi-million de femmes décédaient chaque année de complications obstétricales. Dans la même année, Rosenfield et Maine pointent du doigt le peu de priorité accordée aux mères dans les programmes de santé maternelle et néonatale.

Les estimations mondiales pour l'année 2017 indiquaient 295 000 décès maternels, soit 35 % de moins qu'en 2000 (451 000). Le taux de mortalité maternelle (TMM) mondial en 2017 est estimé à 211 décès maternels pour 100 000 naissances vivantes, ce qui représentait une diminution de 38% par rapport à l'année 2000. Le taux annuel moyen de réduction du TMM mondial au cours de la période 2000-2017 était de 2,9 % ; cela signifie qu'en moyenne, le TMM mondial a diminué de 2,9 % chaque année entre 2000 et 2017 (OMS, 2019).

En matière de santé maternelle, le Maroc a fait d'importants progrès au cours de la période 1990-2018, à travers les efforts de l'Etat et de ses partenaires dans la mise en œuvre d'un ensemble de plans d'action visant à réduire la mortalité maternelle. Entre 2000 et 2017, le taux de mortalité maternelle (TMM) au Maroc a chuté de plus de la moitié, passant de 188 à 70 (pour 100 000 naissances vivantes). Ainsi, les réductions annuelles moyennes étaient d'environ 6% pour le TMM. Les réductions annuelles pour ce taux ont dépassé les réductions moyennes mondiales de 3 %. Selon le rapport de l'enquête confidentielle sur les décès maternels au Maroc en 2010, les causes principales de la mortalité maternelle sont les hémorragies (33%) et les prééclampsies/éclampsie (18%).

Pour définir la santé maternelle, il faut évoquer trois problématiques, à savoir : la mortalité maternelle, la morbidité maternelle et la planification familiale. Comme ce thème concerne plusieurs dimensions, nous focaliserons notre travail sur la planification familiale.

Les résultats des diverses ENPSF (celles menées en 2003, 2011 et 2018), indiquent que le problème de santé maternelle demeure un défi majeur pour l'État marocain, où il y a encore des femmes qui ignorent la planification familiale. Cela peut affecter leur santé comme celle de leurs enfants. Cette étude, réalisée au cours d'un stage accueilli par le Centre d'Etudes et de Recherches Démographiques, consiste à répondre aux questions suivantes : **Quel est le profil des femmes qui ne participent pas à la planification familiale? Et pour quelles raisons ces femmes ne sont-elles pas impliquées dans cette politique?**

L'objectif de notre travail consiste à analyser cette dimension pour orienter les actions visant l'amélioration de la santé de la femme marocaine.

Outre l'introduction et la conclusion, le présent rapport s'articule autour de quatre chapitres. Le premier aborde les approches théoriques explicatives de la santé maternelle et synthétise les travaux empiriques sur la santé maternelle à travers la revue de littérature. Le deuxième chapitre présente les questions méthodologiques et les sources des données. Dans le chapitre trois, nous identifions les niveaux et tendances de la santé maternelle au Maroc selon les caractéristiques de la femme. Le dernier chapitre portera sur le profil des femmes non-utilisatrices de la contraception et les facteurs explicatifs derrière cette non utilisation.

*Chapitre I : Santé
maternelle : une revue
de littérature*

Introduction

La bonne santé et la survie des mères ont des conséquences directes sur la survie et la bonne santé de leurs enfants. Si la mère va bien, son nouveau-né est plus susceptible de survivre et de bien se développer. Par contre, une maman malade peut transmettre sa maladie et ses carences à son enfant (pendant la grossesse, au cours ou après l'accouchement). La question de la santé maternelle reste l'un des principaux soucis des acteurs en charge des politiques de santé tant dans les pays développés que ceux en développement. Au niveau de l'ONU, son intérêt a conduit à mettre en œuvre des objectifs de développement durable (ODD), en application depuis 2016. Le troisième objectif de développement durable est de réduire le taux mondial de mortalité maternelle à moins de 70 pour 100 000 naissances vivantes d'ici 2030 et d'assurer l'accès de tous aux services de santé sexuelle et reproductive. À l'échelle africaine, la santé maternelle reste un défi majeur pour profiter des effets positifs de la croissance démographique d'ici 2050 (Rapport de l'OMS, 2016). A cet égard, plusieurs programmes ont été mis en place pour améliorer la santé maternelle, y compris la planification familiale (PF).

1. Histoire et évolution du concept de santé maternelle

Le terme « santé reproductive » est apparu à la fin des années 1980, sous l'impulsion des organismes internationaux (Organisation Mondiale de la Santé (OMS), Fonds des Nations unies pour la Population (FNUP)), ainsi que de mouvements et organisations non-gouvernementales centrés sur les droits des femmes. En fait, cela correspond à une certaine vision du développement social et sain de la population.

Les questions relatives à la reproduction ont été abordées, pendant de nombreuses années, grâce à un programme intitulé « santé maternelle et infantile ». Ces programmes portent principalement sur la lutte contre la morbidité et la mortalité maternelles et infantiles. Ainsi, la question de la planification familiale était rarement incluse. Cette position reflétait une vision pronataliste des autorités sanitaires et des gouvernements, pour des raisons religieuses et démographiques (forte fécondité par compensation d'une mortalité infantile élevée) ou bien politiques (besoin de main d'œuvre). De ce point de vue, une évolution s'est opérée depuis les années 1960 : les programmes de l'Organisation mondiale de la santé, qui a joué un rôle moteur à cet égard, fournissent un bon exemple du concept de « santé maternelle et infantile » dans la régulation de la fécondité vers la santé reproductive.

En 1951, à la demande du gouvernement indien, l'Organisation mondiale de la santé a mené des recherches sur la planification familiale, en particulier sur les méthodes naturelles de régulation de la fécondité. Malgré ce besoin, la planification familiale est restée un sujet tabou compte tenu de la réticence des États membres lors de la Conférence mondiale sur la population en 1953.

Mais dès 1965, la division de la Santé Familiale a créée une unité appelée "Reproduction humaine", soulignant la nécessité de comprendre les aspects socioculturels et économiques de la reproduction humaine. En 1968, une déclaration à l'Assemblée de l'OMS dirait que "chaque famille devrait avoir accès à des informations et à des conseils sur les questions de planification familiale, y compris la fécondité et l'infertilité" (KESSLER, 1992 : 45).

À partir des années 1990, les conditions démographiques, sanitaires et socio-économiques (programme d'ajustement structurel) des pays du tiers-monde ont fait évoluer la conception de la reproduction humaine. Ces questions ne se limitent plus à la morbidité-mortalité maternelle et infantile ou à la planification familiale. La pandémie du sida et la recrudescence des maladies sexuellement transmissibles ont nécessité des programmes de prévention pour les hommes et les femmes d'âges différents.

2. Définition de la santé maternelle

Selon Pierre Fournier (2007), la santé maternelle est « *une priorité à long terme pour la santé mondiale, et les politiques mises en place pendant plusieurs décennies par les systèmes des pays les plus pauvres sont largement axées sur les mères et les enfants. On s'étonnera alors que, pour certains indicateurs très sensibles aux normes sociales (contraception) ou à la qualité des services (mortalité maternelle), les résultats soient aussi disparates. Certains pays ont fait des progrès remarquables, tandis que d'autres font peu de progrès, stagnent ou régressent en cas de conflit de longue durée ou de conflit civil. Malheureusement, ces situations sont le plus souvent observées en Afrique australe. Même s'il existe de nombreuses preuves quant aux types de stratégies à utiliser pour améliorer la santé maternelle, ils établies de nombreux contraintes dans les différents domaines.* » [19].

L'Organisation mondiale de la santé (OMS) a introduit le concept de « santé reproductive », pour désigner un ensemble des problèmes liés à la santé féminine tels que la mortalité et la morbidité maternelle, et planification familiale. Elle n'est pas seulement considérée comme une absence de maladie ou de trouble dans le processus reproductif, mais plutôt comme une condition par laquelle ce processus s'achève dans un état de complet bien-être physique, psychologique et social. Cela signifie que les individus aient la possibilité de procréer, que les femmes puissent mener à bien leur grossesse et accoucher sans risques et que la reproduction ait une issue heureuse (survie des enfants et leur croissance naturelle). Pour évoquer ce thème, il convient ici d'utiliser le terme « santé maternelle », qui est plus explicite et met en évidence son premier problème : le problème de la mortalité maternelle.

3. Définition des concepts relatifs à la santé maternelle

3.1. Mortalité maternelle

L'OMS a défini ce terme comme suit : « *La mort de toute femme succombant à n'importe quelle cause, que ce soit pendant la grossesse, l'accouchement ou les 42 jours suivant son issue, sans tenir compte de la durée ou du siège de la gestation.* »

En plus de cette définition générale, d'autres définitions spécifiques sont indiquées comme suit (OMS, 2004) :

- **Décès par cause obstétricale directe** : décès qui résultent des complications obstétricales (grossesse, accouchement et post-partum), d'interventions, d'omissions, d'un traitement incorrect ou d'un enchaînement d'événements résultant de l'un des facteurs ci-dessus.

- **Décès par cause obstétricale indirecte** : décès qui résultent d'une maladie préexistante ou d'une affection apparue au cours de la grossesse sans qu'elle soit due à des causes obstétricales directes, mais qui a été aggravée par les effets physiologiques de la grossesse.
- **Morts maternelles tardives** : décès qui résultent de causes obstétricales directes ou indirectes, survenus plus de 42 jours mais moins d'un an après la terminaison de la grossesse.
- **Morts maternelles liées à la grossesse** : décès survenus au cours de la grossesse ou dans un délai de 42 jours après sa terminaison, quelle que soit la cause de la mort.
- **Morts maternelles fortuites ou accidentelles** : décès survenus lors de la grossesse ou dans le post-partum, mais dont les causes n'ont aucun rapport avec celle-ci.

La mortalité maternelle s'exprime sous forme des différents indicateurs (OMS, 2007) :

- **Taux de mortalité maternelle (TMM)** : il se calcule en mettant au numérateur le nombre de décès maternels, et au dénominateur le nombre de naissances vivantes.
- **Ratio de mortalité maternelle** : nombre de décès maternels pour 1 000 naissances vivantes. Il reflète le risque de décès maternel lorsqu'une femme est enceinte.
- **Risque de décès maternel sur la durée de vie adulte** : il tient compte de la probabilité de grossesse et de décès suite à cette grossesse, cumulée pendant la période de reproduction (35 ans par convention).
- **La proportion de décès maternels parmi les décès de femmes en âge de procréer** : cet indicateur est utilisé par l'OMS pour estimer la mortalité maternelle dans les pays qui utilisent des enquêtes spécifiques (méthode des sœurs).

La simple déclaration de cette définition met en lumière les difficultés rencontrées par les chercheurs dans ce domaine, même dans les pays développés où les certificats médicaux de causes de décès apparaissent comme une source fiable d'information. Selon les études faites en France (Bouvier-Colle, 1991) et aux États-Unis (Ellenbrock, 1988), un tiers des décès maternels ne sont pas comptabilisés comme tels.

Pour mesurer la mortalité maternelle, il faut beaucoup d'informations qui ne sont pas toujours disponibles (cause du décès, état de grossesse, âge exact au moment du décès...). Néanmoins, ces données proviennent des sources d'information suivantes :

- Déclaration dans les registres d'état civil, et enquêtes confidentielles ou ponctuelles sur les cas de décès ;
- Enquêtes en population ;
- Enquêtes utilisant la méthode des sœurs ;
- Études de mortalité pendant la période reproductive ;
- Autopsies verbales ;
- Recensements ;

Ces difficultés entraînent, dans la plupart des cas, une sous-estimation des décès maternels. En fait, dans de nombreux cas, seules les données hospitalières sont accessibles.

Il est important de savoir les différentes causes de mortalité pour établir des priorités. Chaque cause fait l'objet d'une ou de plusieurs interventions très différentes : chirurgie pour les dystocies majeures et les ruptures utérines, contraception pour la prévention des avortements clandestins, etc. En ce sens, l'OMS dans Au-delà des chiffres (OMS, 2004) met au point des méthodes pour déterminer les causes des décès maternels. Il est maintenant possible de trouver des façons de prévenir ces décès en améliorant l'accessibilité et la qualité des soins et des services de santé.

Les causes directes (troubles hypertensifs, dystocie, infection puerpérale, hémorragie, complications des avortements, grossesse extra-utérine) représenteraient la vaste majorité des décès maternels. Par exemple, en Afrique subsaharienne, le VIH-sida accroît la proportion des causes indirectes, ramenant les causes directes à environ 70 %. En Asie et en Amérique latine, les causes directes avoisinent les 90 % (Ronmans et Graham, 2006). En ce qui concerne les autres causes indirectes, les plus fréquentes seraient l'anémie, l'embolie pulmonaire, les hépatites, le paludisme, la tuberculose, les maladies cardio-vasculaires, le tétanos, mais aussi les accidents de la route et notamment les violences sociales envers les femmes. Dans certaines parties du monde, la violence contre les civils (insécurité, conflits...) est la principale cause de mortalité maternelle.

D'après l'OMS, la mortalité maternelle est très élevée, avec environ 830 femmes meurent chaque jour dans le monde du fait de complications liées à la grossesse ou à l'accouchement. En 2015, 303 000 femmes sont décédées pendant ou après la grossesse ou l'accouchement. Pendant la période 1990-2015, le taux mondial de mortalité maternelle n'a diminué que de 2,3 % par an. Néanmoins, la diminution de la mortalité maternelle s'est accélérée depuis 2000. Dans certains pays, la baisse annuelle de la mortalité maternelle entre 2000 et 2010 s'est située au-dessus de 5,5 %, un taux nécessaire pour atteindre les objectifs du Millénaire pour le développement.

Le taux de mortalité maternelle (TMM) du Maroc a diminué de plus de moitié entre 2000 et 2017, passant d'environ 188 à 70 (pour 100 000 naissances vivantes). En conséquence, la mortalité maternelle a diminué en moyenne d'environ 6 % par an [5].

3.2. Morbidité maternelle

La morbidité maternelle serait sans doute encore plus révélatrice de l'inégalité des femmes dans les pays en développement par rapport à celles des pays industrialisés. Les données sur la morbidité sont limitées, incomplètes et pas fiables, en particulier en raison de l'absence de « définitions internationalement admises » (Ronmans, 2009). En Afrique de l'Ouest, où 20 326 femmes enceintes ont été suivies, la morbidité maternelle grave était de 6,17 cas pour 100 femmes, soit trente fois plus fréquente que la mortalité (Prual et al, 2000). En Australie, l'étude portant sur plus de 500 000 naissances a montré une incidence de morbidité grave de 12,5 % et un rapport morbidité/décès maternelle de 283 (Roberts et al., 2009). Par contraste, une étude sur la morbidité aiguë sévère de la mère en Europe et en Amérique du Nord, l'incidence a montré 10 fois inférieure à celle de l'Afrique occidentale (van Roosmalen et Zwart, 2009). Le ratio morbidité/décès maternels variait entre 50 et 110. Dans le monde, environ 300 millions de femmes sont atteintes de maladies ou de séquelles évitables dues à la grossesse ou à un accouchement (de Bernis, 2005).

Nous présentons ci-dessous quelques chiffres significatifs retrouvés dans la littérature :

Morbidité immédiate :

- Grossesse ectopique : les données hospitalières dans les pays en développement africains indiquent un taux de létalité de 1 % à 3 %, soit dix fois plus que dans les pays industrialisés (Goyaux, 2003) ;
- Rupture utérine : prévalence de 0,1 % à 1 % des accouchements sur l'utérus cicatriciel ou non sans distinction. Des études menées au Nigeria, au Ghana, en Ethiopie et au Bangladesh indiquent que près de 75% des cas de rupture utérine se produisent sur un utérus sans cicatrice avec taux de mortalité maternelle variant de 1 % à 13 % et taux de mortalité périnatale variant de 74 % à 92 % (Hofmeyr et coll, 2005).

Séquelles :

- Infection chronique de tractus génital et stérilité secondaires à un avortement clandestin : elles touchent respectivement 5 % et 2 % des femmes en âge de procréer (SMO, 2004) ;
- Fistule vésico-vaginales : entre 654 000 et 2 millions de femmes en sont atteintes (Asie - Afrique subsaharienne) (de Bernis, 2005), (Stanton et al, 2007) ;

3.3. Planification familiale

Le concept de planification familiale vise à faire connaître l'avenir aux familles. Il s'agit de l'ensemble des méthodes et des services qui permettent aux couples et aux individus de choisir le nombre d'enfants qu'ils veulent, ainsi que le moment et l'espacement des naissances. En effet, elle comprend les méthodes de contraception modernes telles que la pilule, les progestatifs injectables, les implants, les systèmes vaginaux dits de barrière, et les préservatifs féminins et masculins. Les services comprennent les soins de santé, le soutien, l'information et l'éducation liés à la santé sexuelle et reproductive.

Selon l'OMS, la planification familiale est définie comme un mode de vie que les personnes et les couples adoptent librement, en se basant sur des connaissances précises, des attitudes et des décisions réfléchies afin d'améliorer l'état de santé et le bien-être du groupe familial et de contribuer ainsi efficacement au développement social du pays.

Les femmes et couples ayant accès aux services de la planification familiale, et plus particulièrement les méthodes modernes de contraception sont plus efficaces pour éviter les grossesses non désirées et les naissances dans l'espace. Grâce à ces méthodes, les femmes sont beaucoup moins susceptibles de décéder à l'accouchement ou d'être exposées à des problèmes sanitaires à long terme. De plus, la planification familiale améliore également les chances de survie de l'enfant et son état de santé général.

Le Programme National de Planification Familiale a été créé au Maroc sous l'entièr responsabilité du ministère de la Santé en 1966. Ce programme a gagné beaucoup de terrain à partir des années 70 et notamment des années 80, grâce à l'initiative pilote et à l'exécution continue à l'échelle nationale du programme des visites à domicile de motivation systématique (VDMS). De plus, en 1988, l'expansion du programme au secteur privé par l'initiative Al Hilal a encouragé l'implication des prestataires privés. Aujourd'hui, les services de planification familiale sont disponibles par divers canaux pour l'ensemble de la population marocaine.

Les résultats immédiats de ces initiatives sont d'importantes améliorations des indicateurs démographiques et de santé. À titre d'exemple, l'utilisation de la contraception (pourcentage de femmes mariées utilisant une méthode contraceptive) a gagné 23 points entre 1980 et 1992, passant de 19 % à 42 % et 16 points entre 1992 et 1997, passant de 42 % à 58 %.

4. Revue de littérature : Les déterminants de la santé maternelle

Les soins de santé maternelle font référence aux consultations prénatales, à l'accouchement assisté par un professionnel de la santé et aux soins postnatals. La raison de ce choix est le manque d'attention portée aux soins postnatals en raison du manque de services adéquats dans la plupart des pays en développement. Même si l'importance des soins pré-natals est plus débattue, ils constituent, avec l'accouchement assisté, un moyen efficace de lutter contre les niveaux élevés de mortalité maternelle et néonatale dans les pays en développement (le cas du Maroc).

4.1. Indicateurs et utilisation des services de santé maternelle

Les indicateurs socioculturels font partie des déterminants de la fécondité, de la mortalité ou de l'utilisation des services de santé au Maroc, même si leur impact n'est pas toujours quantifié. En général, les écrits sur les déterminants contextuels de l'utilisation ou des résultats des services de santé au Maroc expliquent une partie importante des différences qui ne sont pas expliquées par des indicateurs socioculturels qui limitent les rôles et le statut des femmes ; diffusion d'un concept de grossesse et d'utilisation des services de santé maternelle ou de fertilité et finalement empêcher les femmes de se faire soigner.

- Indicateurs concernant l'utilisation des services, à la grossesse et à l'accouchement :**

Dans plusieurs groupes ethniques d'Afrique, la grossesse est considérée comme un processus normal dans la vie d'une femme et ne nécessite pas de traitement médical (Beninguisse 2003). Ce concept ou indicateur accepté peut dissuader les femmes de demander des soins. Certaines communautés vont plus loin, louant les naissances uniques. Cette manière explicite de féliciter une femme en lui accordant plus de respect lorsqu'elle n'utilise pas les services médicaux crée une norme qui peut dissuader d'autres femmes de rechercher des soins formels.

4.2. Indicateurs liés à la demande

Les facteurs liés à la demande, elle est basée sur les variables qui sont capables d'être modifiées par les politiques selon le sens d'Ensor et Cooper (2004).

- Emploi de la femme :**

L'emploi reflète une certaine autonomie des femmes pour pouvoir exercer des activités indépendantes dans plusieurs secteurs d'activités. Le côté économique des emplois qui nous concernent ne peut être assurée que si les femmes disposent également une partie de revenus discrétionnaires. D'autre part, l'emploi confère aux femmes une certaine autonomie et modernité, la femme autonome peut prendre en toute liberté la décision de recourir aux soins ou pas et éviter tous les obstacles que cela lui impose. Elle peut accéder à l'information nécessaire et pouvoir aux frais des soins.

- État socio-économique du ménage :**

La plupart des études montrent des effets positifs du statut socio-économique des Ménages mesurés par le revenu, les dépenses de consommation ou les biens matériels utilisés pour les services de santé maternelle. D'autres variables liées à la capacité des ménages à payer leurs dépenses, telles que les services d'assurance et la présence d'institutions de microfinance, sont tout aussi importantes pour accéder à la capacité financière des ménages.

- Niveau d'instruction :**

Le niveau d'instruction des femmes est Considéré comme l'un des principaux facteurs socioculturels favorisant l'usage de Santé maternelle par exemple, une femme plus éduquée a un meilleur accès et une meilleure compréhension de l'information, est plus ouverte aux questions de santé qui la concernent, est plus « moderniste » dans ces décisions, est plus connectée aux réseaux sociaux, et jouit d'une plus grande autonomie.

Le niveau d'éducation d'un conjoint est également important dans la recherche de soins pour les femmes. Cependant, l'éducation des conjoints semble jouer un rôle moins important dans la décision de recourir aux soins que le niveau d'éducation des femmes. Enfin, dans les pays en développement, le niveau d'éducation de la communauté a un effet significatif sur les comportements de santé, et cet effet va au-delà des effets individuels de l'éducation.

- Expériences passées avec les services de santé maternelle :**

L'utilisation des services dans le passé peut affecter les soins médicaux de différentes manières. Cette utilisation peut créer une habitude chez les femmes Peut la familiariser sans effort avec les exigences d'un recours formel aux soins et de l'aguerrir par rapport aux difficultés d'un tel recours.

4.3. Indicateurs liés à l'offre

Les facteurs que nous considérons ici ne concernent que ceux liés aux services de santé maternelle. Concernant la prestation de services, Simkhada et al. (2008) soutiennent que cet aspect continue également de recevoir relativement peu d'attention, principalement en raison d'un manque de données appropriées. En tant que structure de prestation de services, les services de santé doivent être disponibles, accessibles, abordables et de haute qualité pour être acceptés par les femmes.

- **L'accès géographique :**

L'accessibilité géographique, déterminée par la distance aux centres de santé, est un aspect parmi eux, il a reçu le plus d'attention de la part des chercheurs. À cette fin, ils ont montré que la distance aux centres de santé affectait négativement l'utilisation des services prénatals ont été fournis lors de la première visite, mais pas lors des visites suivantes au Népal. L'impact négatif de l'éloignement des centres médicaux sur les soins prénatals et l'accouchement assisté a été démontré en Thaïlande et au Vietnam. De même, Gage (2007) trouve qu'au Mali, les femmes qui vivent entre 10 et 15 km d'un centre de santé ont en moyenne 60 % moins de chances d'accoucher auprès d'un personnel qualifié par rapport aux femmes qui vivent plus près.

- **Les Coûts des services :**

Les coûts des services affectent indirectement l'utilisation, ensuite peut aussi s'apprécier par l'effet qu'une abolition de ces coûts engendre sur le recours aux soins. Toutes les formes des coûts des services de santé maternelle, peuvent collecter majoritairement des revenus des ménages et commandent en partie leur décision de recourir aux soins surtout chez les ménages à faible revenu.

- **La Qualité des services :**

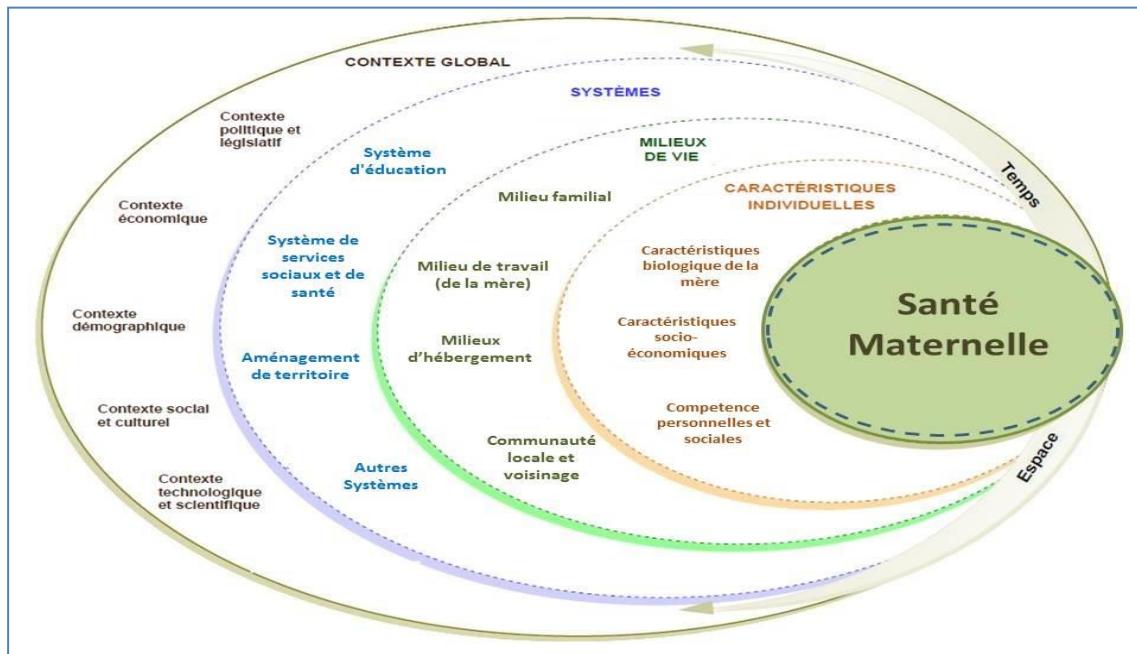
Généralement, un service de qualité est susceptible de briser les résistances qui limitent son utilisation et d'assurer une meilleure efficacité des interventions, tout en ajoutant la perception des clients.

Les services de santé de bonne qualité peuvent apporter davantage d'information aux femmes (surtout les femmes les moins éduquées) et ainsi promouvoir une plus grande utilisation des services (Montgomery et al. 2003).

5. Cadre conceptuel

Pour concevoir notre modèle conceptuel (Figure n° 1), nous nous sommes inspirés du cadre conceptuel issue de résultat d'une réflexion commune, ministère de la Santé et des Services sociaux du Québec (mars 2010), qui explique les inégalités socio-économiques de recours aux soins de santé maternelle, à partir des caractéristiques individuelles passant par le milieu de vie et les différents systèmes vers un contexte global en respectant l'échelle spatio-temporel [21].

Figure n° 1: Cadre conceptuel des déterminants de la santé maternelle

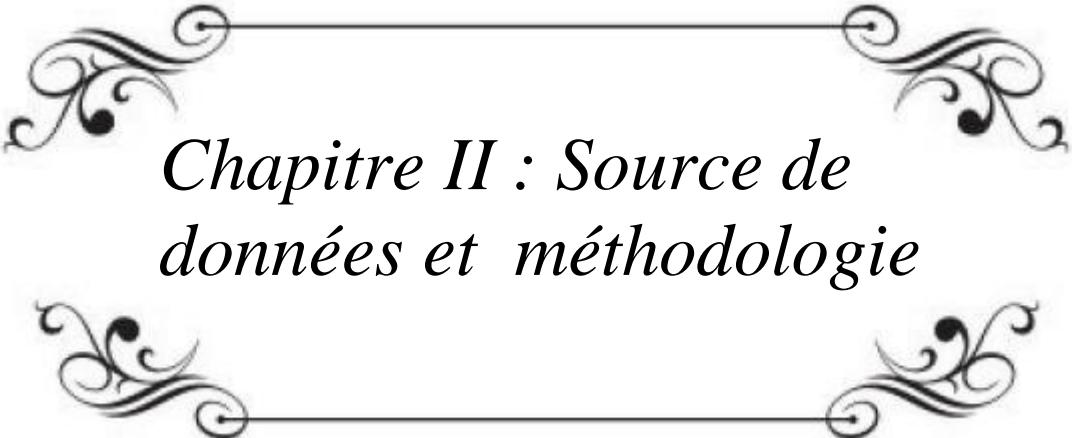


Source : Adapte de résultat d'une réflexion commune sur la santé en global, ministère de la Santé et des Services sociaux du Québec (mars 2010)

La santé maternelle est déterminée par l'accumulation d'un ensemble de facteurs tels que les milieux de vie, les systèmes d'éducation, de services sociaux et de santé et les caractéristiques biologiques et socio-démographiques de la femme. Notons que la santé maternelle est influencée directement par les caractéristiques individuelles et indirectement par le milieu de vie et les systèmes.

Conclusion

Dans ce chapitre, après avoir présenté certains des principaux concepts de notre étude, nous avons par la suite défini un cadre conceptuel que nous a inspiré la revue de littérature. À travers celle-ci, il en ressort que les facteurs socioculturels, socioéconomiques et sociodémographiques qui caractérisent les femmes sont des déterminants empiriques du non-recours à la planification familiale. Par l'étude de l'histoire du concept de santé maternelle, on souligne que ce terme est un sujet complexe, compte tenu de la divergence des idées soulevées par les divers chercheurs. C'est pourquoi, nous focaliserons sur la planification familiale comme une variable à expliquer. Le prochain chapitre sera consacré à la méthodologie et aux sources de données.



*Chapitre II : Source de
données et méthodologie*

Introduction

Les données qui seront utilisées dans le cadre de notre étude sont issues de l'enquête nationale sur la population et la santé familiale réalisée en 2011. Dans ce chapitre, nous allons présenter les objectifs de cette enquête, la population cible, la détermination de la taille de l'échantillon, outils et technique de collecte des données et logiciels utilisés pour notre étude. La dernière partie de ce chapitre portera sur les variables retenues d'après le schéma conceptuel présenté dans le chapitre précédent ainsi que la méthodologie suivie, comme les arbres de classification et la régression logistique pour atteindre nos objectifs.

1. Source de données

Les données utilisées dans ce travail viennent de la base de données de l'enquête nationale sur la population et la santé familiale (ENPSF) réalisée par le ministère de la Santé du Maroc (MS) en 2011.

1.1. Population cible

Notre étude a visé les femmes marocaines en activité génitale et âgées de 15 à 49 ans.

1.2. Objectifs

L'ENPSF s'est donné comme objectifs de :

- Calculer des taux démographiques et notamment des taux de fécondité et de mortalité infanto-juvénile au niveau national et par milieu de résidence (urbain-rural) ;
- Évaluer la couverture vaccinale des enfants ainsi que le suivi des grossesses et des accouchements par milieu de résidence et par région ;
- Mesurer les taux d'utilisation de la contraception par méthode, par milieu de résidence et par région ;
- Étudier les tendances en matière de fécondité, y compris les besoins contraceptifs non satisfaits ;
- Évaluer la fréquence de certaines maladies chroniques, la santé familiale et la santé reproductive ;
- Apprécier la prévalence de la diarrhée et d'autres maladies chez les enfants de moins de 5 ans et le recours aux soins ;
- Évaluer l'état nutritionnel des enfants à l'aide des mesures ;

- Estimer la prévalence des maladies chroniques chez les membres du ménage, plus particulièrement le diabète et l'hypertension.

1.3. Financement de l'enquête

L'ENPSF a été mis en œuvre par le MS avec l'aide technique du Projet Pan Arabe pour la Promotion de la Famille (PAPFAM). Cette enquête était financée essentiellement par le budget de l'État (**76,3 %**). Des contributions financières à hauteur de **23,7 %** ont été apportées par l'UNICEF (**7,1 %**), le FNUAP (**6,1 %**), l'OMS (**6,1 %**) et le PAPFAM (**4,4 %**) [14].

1.4. Questionnaire

Deux questionnaires ont servi à la collecte de données dans le cadre de l'ENPSF-2011, il s'agit d'un questionnaire de ménage et d'un questionnaire individuel adressé aux femmes de 15 à 49 ans qui ne sont pas célibataires et qui résident habituellement dans les ménages interrogés.

1.4.1. Questionnaire ménage

Le questionnaire du ménage tient compte de tous les membres du ménage et est structuré en quatre parties qui s'ajoutent aux données d'identification du ménage [14] :

Première partie : liste des membres de la famille, leur sexe, leur âge, leur état matrimonial, leur niveau de scolarité, leur profession, ainsi que s'ils bénéficient d'une couverture médicale.

Deuxième partie : informations sur la prévalence des maladies chroniques parmi les membres du ménage enquêté.

Troisième partie : caractéristiques du logement, par exemple, type de logement, accès à l'eau, à l'électricité et à l'assainissement, éléments de confort, etc.

Quatrième partie : mesures anthropométriques (mesure le poids et la taille de tous les enfants de moins de 5 ans qui sont membres du ménage ayant fait l'objet de l'enquête.). Le questionnaire du ménage fournit les informations permettant d'établir les dénominateurs pour le calcul des taux démographiques (mortalité, natalité, fécondité) et permet d'identifier les femmes éligibles à un entretien individuel.

1.4.2. Questionnaire femme

Le questionnaire destiné aux femmes est utilisé pour consigner de l'information sur les femmes éligibles (âgées de 15 à 49 ans) qui ne sont pas célibataires et qui sont résidentes habituelles du ménage.

Il est divisé en sept parties, chacune traitant d'un thème :

Caractéristiques socio-démographiques de l'enquêtée : cette section contient des questions sur l'âge de l'enquêtée, sa scolarisation et sa capacité à lire, à écrire et à utiliser l'ordinateur, son activité socioéconomique, son mariage et les caractéristiques de son mari sans parler de la relation possible avec celui-ci et du régime du mariage (monogame ou polygame).

Reproduction et survie des enfants : elle comprend des questions liées aux antécédents de naissance au cours des cinq années précédent l'enquête (enfants vivants, enfants décédés, morts-nés, fausse couche), l'état de grossesse au moment de l'enquête.

Soins de la maternité de la dernière naissance vivante survenue durant les 5 dernières années : cette section comporte 39 questions relatives aux consultations prénatales, à l'assistance au moment de l'accouchement et aux soins postnatals.

Santé et nutrition de la dernière naissance vivante survenue les 5 dernières années : cette section contient 67 questions sur l'allaitement maternel (fréquence et durée, type d'allaitement et introduction des aliments de complément), elle s'est intéressée également à la vaccination de l'enfant et à la prévalence d'un certain nombre de maladie (diarrhée, IRA et la fièvre) et à leur traitement.

Santé reproduction de la femme : les 35 questions posées ont permis d'avoir un aperçu de la prévalence de certaines maladies associées à l'activité reproductive de la femme et de mettre la lumière sur le comportement des femmes ayant eu un problème de santé reproductive (recours aux soins, motifs du non-recours).

SIDA et autres infections sexuellement transmissibles : 15 questions ont été posées pour évaluer les connaissances des femmes interrogées sur les IST-SIDA et pour clarifier les attitudes des femmes envers les personnes vivant avec le SIDA.

Contraception et préférences en matière de la fécondité : cette section permet de collecter des informations relatives à la connaissance, spontanée ou non, des méthodes contraceptives, à leur utilisation actuelle et future et leurs sources d'approvisionnement. Elle permet également de comprendre les intentions de fécondité des femmes enquêtées et leurs opinions sur certains sujets touchant la vie de leurs filles ou des filles en général (35 questions).

1.5. Échantillonnage

L'échantillon de cette enquête était un échantillon stratifié d'environ 16 000 ménages avec un taux de sondage égal à 0,243 %. L'échantillon est représentatif au niveau national, par milieu de résidence et par région [14].

L'échantillon de l'Enquête Démographique à Passages Répétés, réalisée par le Haut Commissariat au Plan entre 2009 et 2010 qui a touché plus de 100 000 ménages, a constitué la base de sondage pour l'enquête ENPSF-2011. Environ 640 Grappes ont été sélectionnées et en moyenne 25 ménages ont été tirés dans chaque grappe.

Le plan de sondage adopté est un sondage probabiliste stratifié et à trois degrés :

- **Premier degré** : tirage d'un échantillon probabiliste d'Unités Primaires (UP) au niveau de chacune des strates en tenant compte des différents niveaux de représentativité des résultats.
- **Deuxième degré** : tirage probabiliste d'une seule Unité Secondaire (US) de chacune des UP choisies.
- **Troisième degré** : tirage d'un échantillon probabiliste de ménages de chacune des US.

1.6. Évaluation de la qualité des données

En général, les enquêtes et recensements menés dans les pays en développement sont souvent entachés d'erreurs. Ces erreurs peuvent provenir des omissions ou des doubles comptes, comme elles peuvent être dues à la mal déclaration des âges basée l'attraction ou la répulsion des âges ronds. Plus l'ampleur des erreurs est importante, plus les résultats obtenus par leur utilisation sont biaisés. D'où provient la nécessité d'évaluer toute base de données avant de l'utiliser. Dans la suite de ce paragraphe, nous allons évaluer la qualité des données de l'ENSPF 2011, en se basant graphiquement sur la courbe de la structure par âge de la population enquêtés et celle des rapports d'âge, et analytiquement sur l'indice de Myers et l'indice combiné des nations unies (les étapes de calculs sont détaillées dans l'annexe n° 2) [16].

1.6.1. Indice de MYERS

On ne peut pas directement comparer les effectifs totaux successifs des femmes dont les âges compris entre 15 et 49 ans se terminant respectivement par chacun de ces chiffres de 0 à 9 du fait de la décroissance normale des effectifs avec l'âge.

Myers a donc proposé un indice portant son nom qui pourrait mesurer l'attraction ou la répulsion des âges se terminant par chacun des chiffres entre 0 à 9.

En se basant sur cet indice, on constate que la déclaration des âges pour les deux sexes est relativement bonne et que les âges des hommes sont mieux déclarés que ceux des femmes.

Tableau n° 1: Les valeurs de l'indice de Myers pour les deux sexes

	Sexe masculin	Sexe féminin
I _M	4,67	4,91

Source : Conception des auteurs à partir des données de l'ENPSF 2011

Tableau n° 2: Répulsion et attraction détectés par chiffre terminal

Attraction/répulsion		
	Sexe masculin	Sexe féminin
0	Attractifs	Attractifs
1	Attractifs	Attractifs
2	Répulsifs	Répulsifs
3	Répulsifs	Répulsifs
4	Répulsifs	Répulsifs
5	Attractifs	Attractifs
6	Attractifs	Attractifs
7	Répulsifs	Répulsifs
8	Répulsifs	Répulsifs
9	Répulsifs	Répulsifs

Source : Conception des auteurs à partir des données de l'ENPSF 2011

Nous avons retrouvé que les âges se terminant respectivement par 0, 1,5 et 6 sont attractifs, alors que les âges se terminant par 2, 3, 4, 7,8 et 9 sont répulsifs pour les deux sexes.

1.6.2. Indice combiné des Nations Unies

Contrairement à l'indice de MYERS, l'indice combiné des Nations Unies se calcule sur la base de répartition des groupes d'âge. Il tente de mesurer la régularité des répartitions par sexe et par âge. Le procès de calcul de cet indice est décrit dans l'annexe n° 2.

Le calcul de cet indice donne la valeur de 25,9 qui est comprise entre 20 et 40, nous pouvons dire alors que les données de l'ENPSF 2011 sont de bonne qualité.

2. Choix des variables

Afin d'atteindre notre objectif global d'évaluer l'impact de la planification familiale sur la santé maternelle, nous nous entendons sur la prédominance des facteurs économiques, sociodémographiques comme principaux facteurs qui déterminent la décision d'utiliser des méthodes contraceptives. À cet effet, le tableau n° 3 ci-dessous présente les variables retenues pour notre étude.

Tableau n° 3: Variables retenues

Variable	Symbol	Modalités
Milieu de résidence	QTYPE	1 : Urbain 2 : Rural
Age	Q103	Aucun
Niveau de scolarisation le plus atteint	Q105	1 : Primaire 2 : Collège 3 : Lycée 4 : Supérieur 5 : Alphabétisation
Situation matrimoniale	Q122	1 : Mariée 2 : Veuve 3 : Divorcée 4 : Séparée
Nombre de fois mariée	Q123	1 : Une fois 2 : Plus d'une fois
Nombre total des enfants en vie	Q221A	Aucun
Nombre de fois total des consultations	Q305A	Aucun
Travail	Q110B	1 : Oui 2 : Non
Planification familiale	Q701	1 : Oui 2 : Non

Source : Conception des auteurs à partir des données de l'ENPSF 2011

3. Méthodologie

L'analyse des données consiste à transformer les données collectées pour les condenser d'une manière qui facilite l'interprétation, sans perdre la plupart des informations qu'elles contiennent. Toutefois, ce processus de transformation nécessite le recours à des méthodes statistiques particulières basées sur l'objet de l'étude et la nature des variables concernées par l'analyse.

À partir des objectifs et des variables de notre étude, nous utiliserons surtout les méthodes statistiques suivantes :

- L'analyse descriptive qui regroupe l'analyse uni-variée et bi-variée ;
- Les arbres de classifications qui est une méthode à la fois descriptive et explicative ;
- L'analyse explicative : la régression logistique.

3.1. Analyse descriptive (analyse bi-variée)

L'analyse bi-variée permet d'évaluer les corrélations entre les variables, ce qui est impossible avec une analyse uni-variée (tri à plat). Comme certaines variables ont une nature qualitative, nous utiliserons la statistique de Khi-2 mesurée avec un seuil de 5 % afin d'évaluer la significativité des relations d'interdépendance entre la variable dépendante et chacune des variables explicatives à partir d'un tableau de contingence.

La variable dépendante dans notre cas a deux modalités (Oui/Non), si nous supposons que la variable indépendante X a p modalités, le tableau de contingence s'écrit sous la forme suivante :

	Y=1	Y=0	n_{i.}
X₁	n₁₁	n₁₀	n_{1..}
X₂	n₂₁	n₂₀	n_{2..}
.	.	.	.
X_j	n_{j1}	n_{j0}	n_{j..}
.	.	.	.
X_p	n_{p1}	n_{p0}	n_{p..}
n_{.j}	n_{.1}	n_{.0}	n = n_{..}

Avec :

$\frac{n_{1.}}{n}$: est la probabilité qu'un individu ait la modalité 1 de Y ;

$\frac{n_{j.}}{n}$: est la probabilité qu'un individu ait la modalité j de X ;

$\frac{n_{j.}}{n} \times \frac{n_{1.}}{n}$: est la probabilité qu'un individu ait simultanément les modalités 1 de Y et j de X.

Le test de Khi-deux consiste à mesurer l'écart entre l'effectif observé n_{ij} et l'effectif théorique $\frac{n_{i.} \times n_{j.}}{n}$, et d'en déduire l'existence et l'intensité d'une liaison mathématique.

Les hypothèses du test de Khi-deux sont :

$$H_0 : n_{ij} = \frac{n_{i.} \times n_{j.}}{n} \quad vs \quad H_1 : n_{ij} \neq \frac{n_{i.} \times n_{j.}}{n}$$

La statistique du test est donnée par :

$$U = \sum_{j=1}^h \sum_{i=1}^k \frac{\left(n_{ij} - \frac{n_{i.} * n_{j.}}{n} \right)^2}{\frac{n_{i.} * n_{j.}}{n}} \sim \chi^2_{(p-1); \alpha}$$

Nous fixons le seuil de significativité à 5 %, si nous trouvons :

- Si $U > \chi^2_{0,05 ; (p-1)}$ alors nous rejetons l'hypothèse H_0 c'est-à-dire qu'il existe une association entre la variable d'intérêt et la variable explicative en question.
- Si $U < \chi^2_{0,05 ; (p-1)}$ alors nous acceptons l'hypothèse H_0 ce qui signifie que la variable d'intérêt est indépendante de la variable explicative choisie.

3.2. Arbres de classification

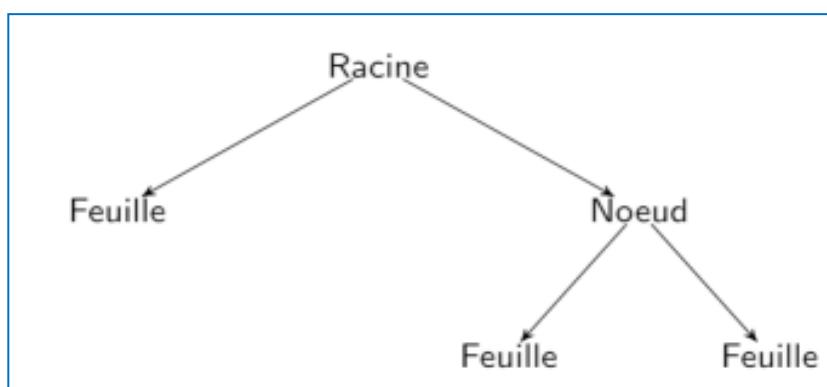
Nous utiliserons une méthode de classification qui fait partie des méthodes qui se sont essentiellement envolées dans le cadre des approches d'apprentissage automatique (machine learning) en intelligence artificielle.

Les arbres de décision ou sont la modélisation d'une classification. Ils apprennent à partir d'observations (exemples). Un exemple est représenté par une gamme d'attributs ainsi qu'une classe associée, il faut connaître la classe parce que les arbres décisionnels travaillent à la classification en mode supervisé.

Le principe d'un arbre de décision ou de classification est de pouvoir donner une prédiction d'une variable de sortie (variable dépendante) en fonction de différentes variables d'entrée, ici les variables explicatives. Ainsi, il permet de classifier la population cible et de déterminer les variables les plus différenciées par rapport à la variable cible, c'est-à-dire sélectionner à partir des variables explicatives celle qui présente le plus de corrélation avec la variable à expliquer.

Un arbre de classification est composé de nœuds en arborescence, le nœud à base de l'arbre est appelé la racine, chacun des nœuds sous la racine est soit une feuille ou un sous-arbre, comme sur le schéma récapitulatif suivant :

Figure n° 2: Schéma d'un arbre



3.2.1. Construction d'un arbre de classification

Il existe plusieurs algorithmes de construction d'arbres tels que ID3, C4.5, CART et CHAID.

Les algorithmes de construction d'arbre de classification permettent de créer des arbres de décision avec une taille la plus petite que possible, ceci afin d'établir des règles de décision simples. Plus un arbre de décision est grand, plus les règles sont compliquées. Les algorithmes de construction d'arbres choisissent les attributs toujours par rapport aux classes.

Les arbres de classification proviennent d'un jeu d'apprentissage, un jeu d'apprentissage est une matrice, dont les lignes représentent les exemples et les colonnes représentent les caractéristiques de ces exemples. Pour construire des arbres de décision, il faut suivre 3 étapes [19] :

La première étape : consiste à vérifier si on doit faire un nœud terminal pour représenter les exemples du jeu d'apprentissage. Pour faire un nœud terminal, il faut tenir compte du fait que tous les exemples du jeu d'apprentissage appartiennent à la même classe ou tous les attributs ont été utilisés pour les nœuds précédents. Cette étape permet d'arrêter l'expansion de la branche de l'arbre ;

La deuxième étape : consiste à trouver l'attribut pour représenter le nœud de l'arbre ;

La troisième étape : consiste à éclater le jeu d'apprentissages pour créer les branches du nœud, chacune des branches du nœud prend une des différentes valeurs que l'attribut du nœud peut prendre. Pour chacune des branches qu'on aura créées, il faut recommencer le processus en prenant les exemples correspondants à la branche.

Dans notre étude, nous allons choisir seulement quelques algorithmes de construction d'arbre dans l'ordre suivant :

a) Algorithme **CART (Classification And Regression Trees)**

Concernant l'algorithme CART, la variable dépendante peut être nominale ou ordinaire ou bien continue. Mais les variables explicatives doivent être soit nominales soit continues.

La construction de l'arbre CART se fait comme suit :

- i) Pour chaque variable explicative, on calcule deux mesures standard d'impureté, l'indice d'entropie et l'indice de GINI.

$$GINI = \sum_{j=1}^J p_j (1 - p_j)$$

$$Entropie = - \sum_{j=1}^J p_j \log p_j$$

Où p_j : la proportion d'observation dans la classe j.

ii) Choisir la meilleure variable explicative est celle de l'indice d'entropie le plus petit ou même de gain informationnel le plus élevé.

$$\text{Gain informationnel} = I(\text{nœud}) - \sum_{j=1}^J P(b) I(b)$$

Où $I(\text{nœud})$: indice d'impureté pour le nœud.

b : le nombre de branches qui divisent le nœud parent.

$P(b)$: la proportion d'observation dans la branche b.

b) Algorithme CHAID (Chi-Square Automatic Interaction Detection)

CHAID est l'une des plus anciennes méthodes d'arbre de classification, initialement proposée par Kass (1980 ; après Ripley, 1996, l'algorithme CHAID est une version modifiée de l'algorithme THAID développé par Morgan et Messenger, 1973.).

L'algorithme utilisé pour construire des arbres de décision (non-binaires) :

- Pour des problèmes de classification, lorsque la variable dépendante est de nature catégorielle en utilisant le test du χ^2 pour déterminer la prochaine meilleure répartition à chaque niveau.
- Pour les problèmes de régression quand la variable dépendante est continue, le programme va en fait calculer des tests de FISHER.

La procédure de l'algorithme CHAID :

L'Accouplement des Modalités : en répartissant la distribution des différents variables explicatives en un certain nombre de catégories d'effectifs sensiblement égaux, ensuite on va examiner les catégories afin de déterminer pour chacun, le couple de catégories les moins différentes significativement par rapport à la variable dépendante à l'aide de test chi² (*Chi-deux de Pearson*).

Sélectionnez la variable de division : on va sélectionner alors pour chaque X la variable avec la p_value ajustée la plus basse pour la division, c'est-à-dire le prédicteur qui produit la division la plus significative.

CHAID traite les attributs discrets avant les attributs continus.

Y / X	x_1	x_l	x_L	Σ
y_1				
		⋮		
y_k	...	n_{kl}	...	$n_{k.}$
		⋮		
y_K				
Σ		$n_{.l}$		n

Les lignes représentent la répartition des classes et les colonnes représentent la répartition des valeurs de cet attribut, la dernière ligne et la dernière colonne représente la sommation.

$$\chi^2 = \sum_{k=1}^K \sum_{l=1}^L \frac{\left(n_{kl} - \frac{n_{k\cdot} * n_{\cdot l}}{n} \right)^2}{\frac{n_{k\cdot} * n_{\cdot l}}{n}}$$

Comparaison des valeurs observées avec les valeurs théoriques lorsque Y et X sont indépendants (produit des marges).

Pour chaque nœud, on reprend les mêmes procédures précédentes.

CHAID utilise une technique de pré élagage, c'est-à-dire qu'on arrête la croissance de l'arbre lorsque le test remplit une certaine condition.

Si la plus petite p_value ajustée du prédicteur (Bonferroni) est supérieure à une valeur α de division, le processus de division se termine, le nœud correspondant est le nœud terminal.

3.2.2. Validation du modèle d'arbre de classification

Pour valider une classification, il est important de prendre en considération les erreurs de classification, elles sont la conséquence de la création d'une feuille qui contient plus qu'une classe. Autrement dit, lorsque l'algorithme est invoqué, le jeu d'apprentissage contient des exemples qui appartenaient à au moins deux classes différentes et que l'ensemble de tests à effectuer était vide. Le calcul de l'erreur de classification d'un arbre commence par les feuilles de la racine (du bas vers le haut de l'arbre) [2].

L'erreur apparente d'une feuille représente le nombre d'exemples mal classifiés par cette feuille, celui de l'arbre est la somme des erreurs de toutes les feuilles. La formule de calcul du taux d'erreur de classification est la suivante :

$$\text{Taux d'erreur} = \frac{\text{Nbr erreurs}}{\text{Nbr exemple}}$$

Pour éviter le sur-apprentissage, c'est-à-dire qu'on a deux nœuds qui contiennent un seul élément, on supprime les sous-arbres les moins significatifs de l'arbre afin de les remplacer par des feuilles avec un seuil d'erreur acceptable, de cette façon, on garde de bonnes performances prédictives (généralisation).

La phase d'élagage (pruning) consiste à enlever les feuilles les moins significatives de l'arbre. Cette phase est exécutée après la construction de l'arbre. On considère que la racine de l'arbre comprend au moins deux feuilles. Tant qu'il existe un sous-arbre que l'on peut remplacer par une feuille, sans faire croître l'estimation de l'erreur réelle, alors on élague ce sous-arbre. On doit savoir si les nœuds fils sont des feuilles ou des sous-arbres. Si tous les nœuds sont des feuilles, on remplace la racine du sous-arbre par une feuille, si l'erreur de la nouvelle feuille est plus petite que celle de l'ancien sous-arbre.

Pour créer la feuille de remplacement, on crée la liste des possibilités d'éléments des feuilles du sous-arbre et on prend la valeur la plus fréquente pour le nom de l'étiquette et le(s) autre(s) dans le tableau Étiquette des autres possibilités. On remonte ensuite jusqu'à la racine jusqu'à ce qu'on ne puisse plus remplacer un sous-arbre par une feuille.

Pour évaluer les sous-arbres, on utilise deux méthodes :

a) Échantillon de validation

La procédure de validation du modèle d'arbre de classification se divise en 3 étapes :

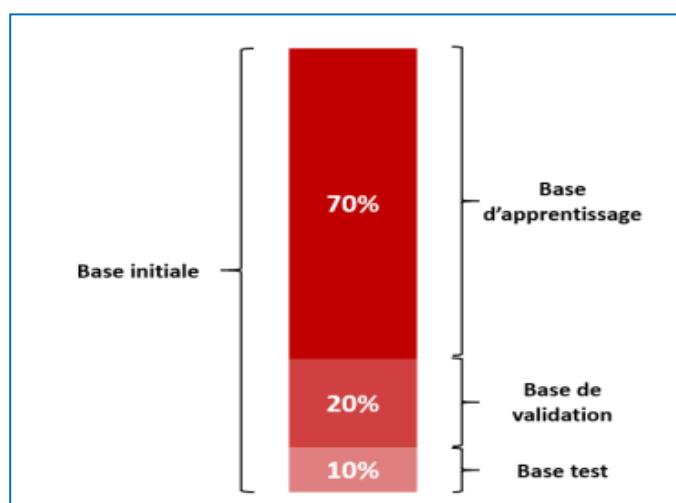
Étape 1 (Apprentissage): consiste à construire un modèle (arbre maximal) sur un échantillon pour lequel on connaît la variable cible, il est préférable que la base d'apprentissage soit composée d'environ 70 % des individus de la base initiale ;

Étape 2 (Validation): consiste à vérifier le modèle sur un deuxième échantillon pour lequel on connaît la variable cible et la comparer avec la valeur prédite par le modèle. Cette base compose environ 20 % de la base initiale et son rôle est d'élaguer notre arbre ;

Étape 3 (Test): consiste à tester la puissance explicative de notre modèle, une fois ce dernier créé. Nous allons dans un premier temps ignorer la variable à expliquer des individus de cette base, et nous allons tenter de la prédire avec nos différents modèles. Chacun donnera donc ses prédictions pour les individus de la base. Nous allons dans un second temps comparer ces prédictions aux vraies valeurs de sortie de nos individus. C'est pour cela qu'il est primordial que la base de test ne serve pas à la création de l'arbre, sinon ce dernier pourrait s'adapter à la base de test afin de réduire son erreur, ce qui fausserait la comparaison. De plus, notons que cette base de test est composée de 10 % de la base initiale.

Le schéma ci-dessous résume la répartition de nos observations parmi les trois sous-bases en question.

Figure n° 3: Répartition et taille des trois sous-bases créées



b) Validation croisée

Pour effectuer une validation croisée, on stratifie la population en n tailles égales, on en choisit les n-1 premiers échantillons pour faire le jeu d'apprentissage et l'autre échantillon pour le jeu de test. On construit un arbre de décision à partir du jeu d'apprentissages et on teste les règles avec le jeu de tests.

Ensuite, on répète le processus jusqu'à chacun des échantillons soit testé. À la fin du processus, on aura n ensembles de règles de décision. Pour avoir une approche plus générale, on peut combiner les règles de façon à diminuer les conflits entre les règles. La validation croisée est une autre technique pour éviter le sur-apprentissage [11, 22].

3.3. Analyse explicative

La planification familiale étant une variable qualitative dichotomique, la régression logistique binaire est la méthode d'analyse multi-variée explicative appropriée. En régression logistique, la variable dépendante de nature dichotomique s'explique par une série d'autres variables qui peuvent être qualitatives ou quantitatives. Il est à noter que la régression logistique utilise la méthode de maximum de vraisemblance pour estimer les paramètres du modèle. Compte tenu de la linéarité du modèle, ces paramètres sont estimés par itération : il s'agit donc d'une méthode probabiliste. En effet, la régression logistique estime la probabilité « P » pour qu'une femme n'utilise pas les méthodes contraceptives alors que « 1-P » est la probabilité de l'échec de cet événement [9].

Le modèle utilisé pour modéliser ce type de variable est celui du **logit** pour sa simplicité, puisque nous connaissons sa fonction de répartition et sa densité qui s'écrivent de la façon suivante :

$$F(x_i, \beta) = \frac{e^{-x_i\beta}}{1 + e^{-x_i\beta}} \quad \text{et} \quad f(x_i, \beta) = \frac{e^{-x_i\beta}}{(1 + e^{-x_i\beta})^2}$$

On note les variables explicatives par $X_i = (1, X_{i1}, \dots, X_{ip})$ et les paramètres à estimer par $\beta = (\beta_0, \beta_1, \dots, \beta_p)$.

La méthode d'estimation utilisée est celle du maximum de vraisemblance. La vraisemblance s'écrit :

$$L(\beta) = \prod_{i=1}^N F(x_i, \beta)^{y_i} (1 - F(x_i, \beta))^{1-y_i}$$

En pratique, il est habituellement plus simple de chercher à maximiser le log de la vraisemblance (ce qui est pareil, car le log est une fonction en croissance.) et on nomme cette fonction la log-vraisemblance qui est :

$$l(\beta) = \log L(\beta) = \sum_{i=1}^N y_i \log F(x_i, \beta) + \sum_{i=1}^N (1 - y_i) \log(1 - F(x_i, \beta))$$

Pour la régression logistique, il suffit de remplacer F par sa valeur et de résoudre l'équation résultant de la condition de la maximisation de premier ordre :

$$\frac{\partial l(\beta)}{\partial x_i} = 0$$

Par analogie avec la régression linéaire, la valeur de la log-vraisemblance est estimée à $\hat{\beta}$, $l(\hat{\beta})$ augmente toujours lorsque des régressseurs sont ajoutées, de sorte qu'il ne sera pas en mesure d'être utilisé comme un outil de sélection des variables. Les critères d'information sont des fonctions de la log-vraisemblance, mais insèrent une pénalité pour le nombre de coefficients $\hat{\beta}$:

- Le critère d'information d'Akaike (**AIC**)

$$AIC = -2l(\hat{\beta}) + 2 * (p + 1)$$

- Le critère d'information Bayésien (**BIC**)

$$BIC = -2l(\hat{\beta}) + \log(n) * (p + 1)$$

Où p : le nombre des variables explicatives x_i

Et n : la taille d'échantillon étudié

Après l'estimation du modèle, il faut examiner sa qualité. Pour ceci, un certain nombre de tests sont utilisés, notamment le test de khi-2 de Pearson pour la significativité globale du modèle, le LR-test et le test de Hosmer et Lemeshow pour la qualité de l'ajustement, la courbe ROC pour le pouvoir discriminant du modèle et bien d'autres.

- Test de χ^2 de Pearson

Soient n_m : la fréquence de l'état x_m ,

$z_m = \sum_{i=1}^n \{y_i = 1/x_m\}$: le nombre d'observations de l'état x_m appartenant à G_1 et $\hat{\pi}_0(m)$: la fréquence relative de l'état ($Y=1|x_m$) estimée sous H_0 .

On a : sous H_0 : $\hat{z}_m = n_m \hat{\pi}_0(m) = n_m \hat{\pi}(m, \hat{\beta}_0)$ où $\hat{\beta}_0$ est l'estimation des paramètres du modèle logistique sous H_0 .

L'écart entre les 2 modèles (le modèle saturé et le modèle défini par H_0) est mesuré par la quantité :

$$E = \sum_{m=1}^M r^2(z_m, \hat{z}_m)$$

Avec $r(z_m, \hat{z}_m)$ est le résidu de Pearson défini comme suit :

$$r(z_m, \hat{z}_m) = \frac{z_m - \hat{z}_m}{\sqrt{[n_m \hat{\pi}_0(m)(1 - \hat{\pi}_0(m))]}}$$

Sous H_0 : E suit asymptotiquement une χ^2 à $(M-q)$ degré de liberté. Où q représente le nombre de paramètres à estimer sous H_0 .

Règle de décision :

Au niveau de signification α : on rejette $H_0 \Leftrightarrow E > \chi^2_{M-q,1-\alpha}$

- Test du rapport de vraisemblance (**LR**)

$$LR = -2(l_0(\hat{\beta}) - l(\hat{\beta})) \xrightarrow{N \rightarrow \infty} \chi^2_M$$

Où $l_0(\hat{\beta})$ est la log-vraisemblance évaluée sous l'hypothèse nulle H_0 et $l(\hat{\beta})$ la log-vraisemblance sous l'hypothèse H_1 .

Soit le test suivant :

$$\begin{aligned} H_0: \beta_1 &= \beta_2 = \dots = \beta_M = 0 \\ &\text{Vs} \\ H_1: \exists \beta_k &\neq 0 \end{aligned}$$

Règle de décision :

Au niveau de signification α : on rejette H_0 si $LR > \chi^2_{M-1,1-\alpha}$

- Test de Wald (**W**)

$$W_j = \frac{\hat{\beta}_j^2}{\hat{\sigma}_j^2}$$

Avec : $\hat{\sigma}_j^2$ la valeur lue sur la diagonale principale de la matrice de variance-covariance des coefficients estimé.

Soit le test suivant :

$$\begin{aligned} H_0: \beta_1 &= \beta_2 = \dots = \beta_M = 0 \\ &\text{Vs} \\ H_1: \exists \beta_j &\neq 0 \end{aligned}$$

Règle de décision :

Au niveau de signification α : on rejette H_0 si $W_j > \chi^2_{M-1,1-\alpha}$

- Test de Hosmer et Lemeshow (**HL**)

Le test d'adéquation de l'ajustement de Hosmer-Lemeshow permet de comparer les effectifs attendus des événements et non événements aux effectifs observés afin d'évaluer l'ajustement du modèle aux données. Cette statistique est calculée à partir des tableaux 2 x g des effectifs observés et théoriques, où g représente le nombre de groupes. Elle est comparée à une distribution du khi-deux à g-2 degrés de liberté.

La statistique de (HL) est :

$$C^2 = \sum_{j \in [1;g]} \frac{(O_j - N_j \hat{\pi}_0(j))^2}{N_j \hat{\pi}_0(j)(1 - \hat{\pi}_0(j))}$$

Où N_j représente le nombre d'observations du j-ième groupe, O_j représente le nombre de réponses du j-ième groupe et représente la moyenne des probabilités prévues dans le j-ième groupe.

Le test est le suivant :

$$\begin{aligned} H_0 &: \text{Le modèle est adéquat} \\ &\text{Vs} \\ H_1 &: \text{Le modèle n'est pas adéquat} \end{aligned}$$

Règle de décision :

À un niveau de signification $\alpha = 5\%$, on accepte H_0 si la p_value est supérieure à 0,05. Sinon, si la p_value est inférieure à 0,05 le modèle est rejeté.

- Pseudo R² de McFadden

$$R_{MF}^2 = 1 - \frac{l(\hat{\beta})}{l_0(\hat{\beta})}$$

Où $l_0(\hat{\beta})$ est la log-vraisemblance du modèle nul et $l(\hat{\beta})$ la log-vraisemblance du modèle comparé.

Analogie de R² de la régression linéaire, elle mesure combien (%) de déviance est expliquée par le modèle.

- La courbe ROC

En guise d'indicateur de la capacité de discrimination du modèle, on utilisera la courbe ROC.

La courbe ROC (de l'anglais Receiver Operator Characteristic curve) est un moyen d'exprimer la relation entre la sensibilité et la spécificité d'un test diagnostique.

La sensibilité est définie comme la probabilité de classer l'individu dans la catégorie $y = 1$ étant donné qu'il est effectivement observé dans celle-ci :

$$\text{Sensibilité} = P(\text{Test+} | y = 1)$$

La spécificité est définie comme la probabilité de classer l'individu dans la catégorie $y = 0$ étant donné qu'il est effectivement observé dans celle-ci :

$$\text{Spécificité} = P(\text{Test-} | y = 0)$$

La surface sous cette courbe nous permet d'évaluer la précision du modèle pour discriminer les outcomes positifs ($y = 1$) des outcomes négatifs ($y = 0$).

On retiendra comme règle de décision :

- Si l'aire ROC = 0.5, il n'y'a pas de discrimination.
- Si l'aire $0.7 \leq \text{ROC} \leq 0.8$, la discrimination est acceptable.
- Si l'aire $0.8 \leq \text{ROC} \leq 0.9$, la discrimination est excellente.
- Si l'aire $\text{ROC} \geq 0.9$, la discrimination est exceptionnelle.

Après avoir examiné la qualité du modèle, nous pouvons interpréter les résultats du modèle en se basant sur les rapports de chances ou bien les Odds Ratio (OR). En effet, nous traduisons les rapports de chance de l'utilisation des méthodes de la contraception. Lorsque l'OR est supérieur à 1, on indiquera que l'événement a plus de chance de se produire sinon on dira que l'événement a moins de chance de se réaliser.

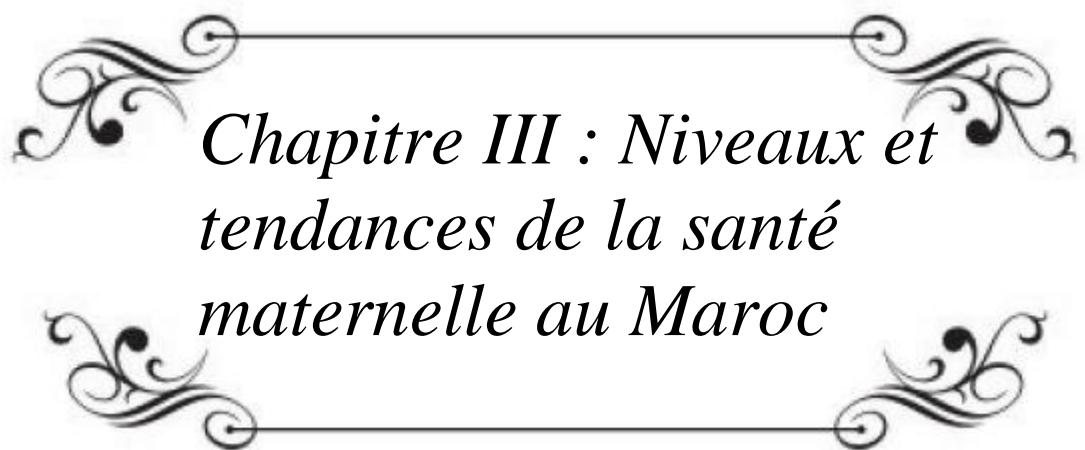
Nous interpréterons aussi les signes des coefficients des variables explicatives. Une variable dont le coefficient a un signe positif, influence positivement non-utilisation de la planification familiale, autrement dit la probabilité de n'utiliser pas les méthodes contraceptives croît avec cette variable. D'autre part, si le signe est négatif, alors cette variable affecte négativement la non-utilisation de la planification familiale. C'est-à-dire que la chance de n'utiliser pas la planification familiale décroît avec cette variable.

3.4. Outils d'analyse

En ce qui concerne l'analyse et le traitement des données, nous aurons recours à certains logiciels. Microsoft Office Excel 2007 est utile pour créer des tableaux et des graphiques. Au niveau de l'analyse bi-varié, l'analyse explicative particulièrement la régression et la construction des arbres de classification, nous utiliserons le langage de programmation PYTHON qui présente de nombreuses caractéristiques intéressantes telles qu'il est multi-plateforme, c'est-à-dire qu'il fonctionne sur de nombreux systèmes d'exploitation : Windows, Mac OS X ainsi qu'il est gratuit, de haut niveau, bien interprété et orienté objet. Ansai, nous utiliserons le logiciel SPSS qui est un logiciel entièrement intégré pour réaliser tous les tests statistiques habituellement utilisées en sciences sociales et en psychologie.

Conclusion

Dans ce chapitre, nous avons identifié parmi les variables qualitatives qui caractérisent les femmes, celles qui ont un lien avec non-utilisation des méthodes contraceptives. Nous avons ensuite déterminé les profils des femmes non célibataires et non utilisatrices de la contraception. Après cela, nous avons construit un modèle de régression logistique en utilisant des variables caractérisant les femmes comme variables indépendantes. Ceci permettra d'établir ceux qui influencent significativement le non-recours aux méthodes contraceptives par les femmes marocaines non célibataires.



*Chapitre III : Niveaux et
tendances de la santé
maternelle au Maroc*

Introduction

Dans ce troisième chapitre, nous cherchons à identifier les niveaux et tendances de la santé maternelle au Maroc selon les caractéristiques de la femme telles que l'âge, le milieu de résidence, le niveau d'instruction et l'état matrimonial.

1. Niveaux et tendances de la mortalité maternelle

1.1. Niveau de la mortalité maternelle

Le niveau de mortalité maternelle est aujourd'hui une mesure qui fournit des informations sur le développement humain et social de chaque pays. Il s'agit d'un indicateur particulièrement révélateur de la situation des femmes en particulier, de l'accès des femmes aux soins de santé et de la façon dont le système de santé répond à leurs besoins. Cet indicateur est généralement mesuré par le Ratio de Mortalité Maternelle (RMM), qui nécessite des données sur les décès maternels et les naissances vivantes survenus dans le pays durant une période déterminée.

Tableau n° 4: Evolution du Ratio de Mortalité Maternelle par milieu de résidence entre 2010-2016

	RMM 2009-2010	RMM 2015-2016	Taux de variation 2010-2016	TRAM* 2010-2016
Urbain	73	44,6	-39%	7,9%
Rural	148	111,1	-25%	4,7%
National	112	72,6	-35%	7%

(*) TRAM : Taux de réduction annuel moyen

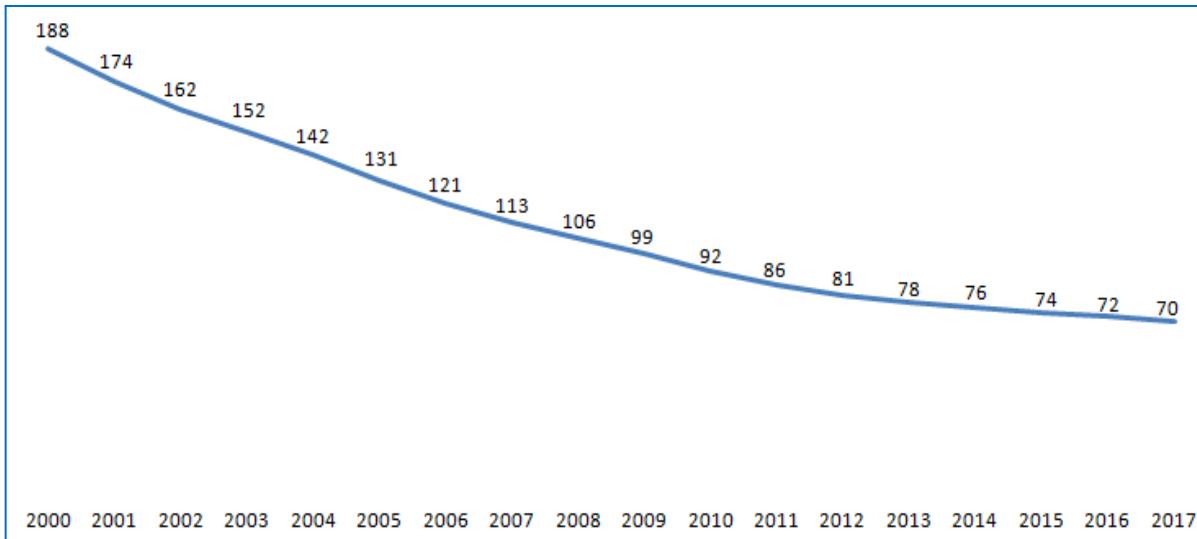
Source : ENPSF 2018

Il ressort de ce tableau que :

- Le niveau de mortalité maternelle était de 72,6 pour 100 000 naissances vivantes sur la période 2015-2016, soit une baisse de 35 % entre 2010 et 2016, avec une baisse annuelle moyenne de 7 % ;
- Les niveaux de mortalité maternelle en milieux urbain et rural ont atteint respectivement 44,6 et 111,1 pour 100 000 naissances vivantes. Le rythme de réduction est plus rapide dans les zones urbaines, à 7,9 %, contre 4,7 % dans les zones rurales.

1.2. Tendance de la mortalité maternelle

Graphique n° 1: Evolution du RMM entre 2000-2017



Source : Conception des auteurs à partir des données de Banque Mondiale

Pour l'ensemble de la période 2000-2017, nous remarquons une diminution du ratio de mortalité maternelle, c'est-à-dire que c'est en 2000 qu'on enregistre la valeur la plus élevée (188 pour 100 000 naissances vivantes) et en 2017, il a atteint 70 pour 100 000 naissances vivantes.

2. Niveaux et tendances de la morbidité maternelle

2.1. Niveau de la morbidité maternelle

Dans la suite de cette partie, nous allons présenter le niveau de la morbidité maternelle au Maroc selon trois caractères sociodémographiques des femmes.

Tableau n° 5: Répartition en % des femmes âgées de 15-49 ans qui n'a aucun cancer et celle qui souffrent d'un cancer au moment de l'enquête au niveau national

	Pourcentage qui n'a aucun cancer	Pourcentage de celle qui souffrent d'un cancer	Effectif
National	99,7	0,3	11069

Source : Conception des auteurs à partir des données de l'ENPSF 2011

Le tableau montre que la répartition en pourcentage des femmes non célibataires dont leurs âges varient entre 15 et 49 qui souffrent d'une maladie cancéreuse au moment de l'enquête ou l'avoir eu par le passé est faible est de 0.3 % (300/100 000) des femmes enquêtées.

2.1.1. Selon l'âge de la femme

Les données du tableau donnent les estimations des femmes qui souffrent d'un ou des types de Cancer selon le groupe d'âges. On constate globalement que le pourcentage de celles qui souffrent d'un cancer s'augmente avec le groupe d'âge jusqu'à 0,7 % entre 40 et 49 avec le type de cancer le plus fréquent est le cancer de sein avec 0,5 % et 0,4 % pour les groupes d'âge 40-44 et 45-49 respectivement.

Tableau n° 6: Pourcentage des femmes non célibataires âgées de 15-49 ans qui souffrent d'une maladie cancéreuse selon le type de cancer et le groupe d'âges

Groupe d'âges	Type de Cancer									Effectif
	% de celles qui souffrent d'un cancer	Cancer du Sein	Cancer du col de l'utérus	Cancer de l'ovaire	Cancer de l'œsophage	Cancer de la vésicule biliaire	Cancer Oral	Tumeur Cérébrale	Tumeur Aisselles	
15-19	0	0	0	0	0	0	0	0	0	397
20-24	0	0	0	0	0	0	0	0	0	1265
25-29	0,1	0	0	0	0	0	0	0,1	0	1793
30-34	0,2	0	0	0	0	0,1	0	0	0	1997
35-39	0,2	0,1	0	0	0,2	0	0	0	0	2000
40-44	0,7	0,5	0,2	0	0	0	0,1	0	0	1901
45-49	0,7	0,4	0,2	0	0	0	0	0	0,1	1715

Source : Conception des auteurs à partir des données de l'ENPSF 2011

2.1.2. Selon le milieu de résidence

La répartition en milieu rural et urbain montre que le deuxième possède un pourcentage des femmes qui souffrent d'un cancer égale à 0,4 % ce qui présente le double de celle de milieu rural, on constate aussi que la prévalence du cancer du sein est presque 4 fois plus élevée en urbain qu'en rural.

Tableau n° 7: Pourcentage des femmes non célibataires âgées de 15-49 ans qui souffrent d'une maladie cancéreuse selon le type de cancer et le milieu de résidence

Milieu	Type de Cancer									Effectif
	% de celle qui souffrent d'un cancer	Cancer du Sein	Cancer du col de l'utérus	Cancer de l'ovaire	Cancer de l'œsophage	Cancer de la vésicule biliaire	Cancer Oral	Tumeur Cérébrale	Tumeur Aisselles	
Urbain	0,4	0,2	0,1	0	0	0	0	0	0	6113
Rural	0,2	0,1	0	0	0,1	0	0	0	0	4956

Source : Conception des auteurs à partir des données de l'ENPSF 2011

2.1.3. Selon le quantile de bien-être

Le tableau montre que 20 % des femmes riches présentent la moitié des cas observés, ici encore, très probablement, ce sont les facilités d'accès aux moyens de dépistage et les conditions socio-économiques favorables des femmes qui expliqueraient cette proportion élevée.

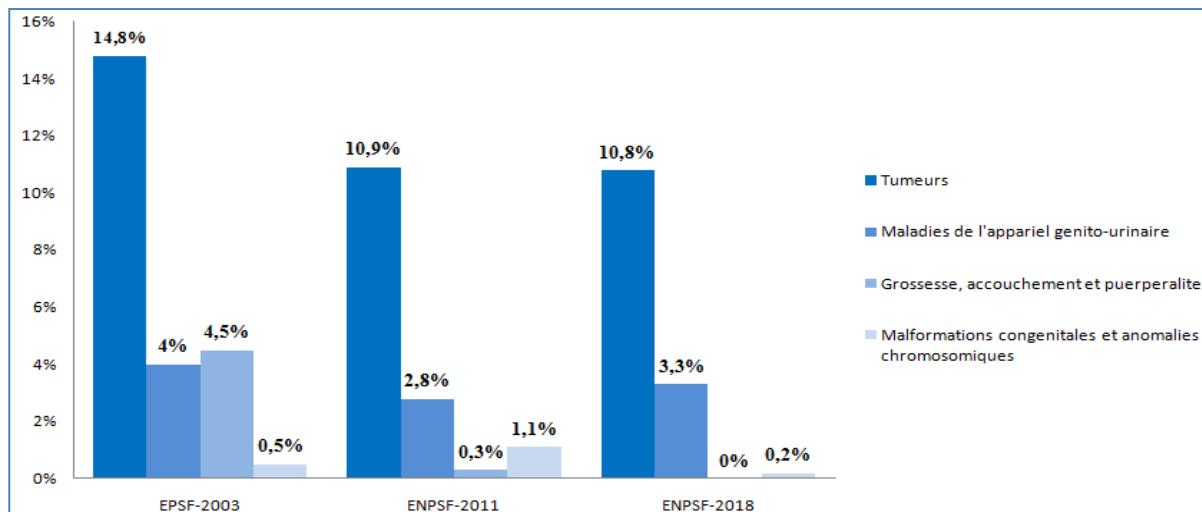
Tableau n° 8: Pourcentage des femmes non célibataires âgées de 15-49 ans qui souffrent d'une maladie cancéreuse selon le type de cancer et le quantile de bien-être

Quantile de bien être	Type de Cancer									Effectif
	% de celle qui souffrent d'un cancer	Cancer du Sein	Cancer du col de l'utérus	Cancer de l'ovaire	Cancer de l'œsophage	Cancer de la vésicule biliaire	Cancer Oral	Tumeur Cérébrale	Tumeur Aisselles	
Le plus pauvre	0,3	0,1	0,1	0	0,2	0	0	0	0	2172
Second	0,2	0	0,1	0	0	0	0	0	0	2229
Moyen	0,2	0,1	0,1	0	0	0	0	0	0	2333
Quatrième	0,2	0,1	0	0	0	0,1	0	0	0	2234
Le plus riche	0,6	0,4	0	0	0	0	0	0,1	0,1	2101

Source : Conception des auteurs à partir des données de l'ENPSF 2011

2.2. Tendance de la morbidité maternelle

Graphique n° 2: Répartition des décès des femmes âgées de 15-49 ans (en %) selon la cause, selon différentes enquêtes réalisées au Maroc depuis 2003, ENPSF-2018



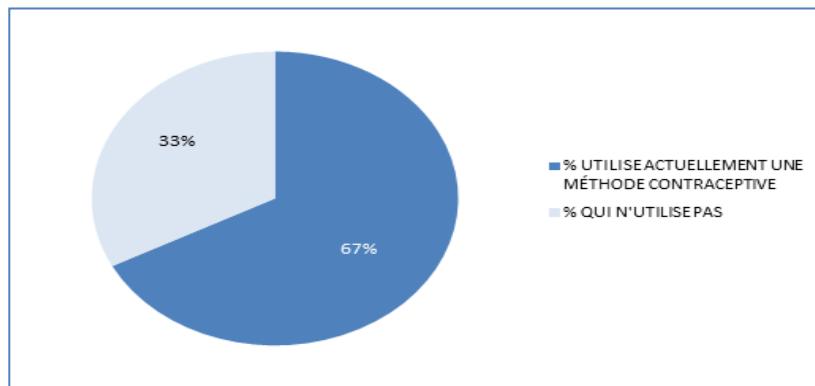
Source : Conception des auteurs à partir des données de l'ENPSF de 2004, 2011 et 2018

Le graphique n° 2 montre que la proportion des décès des femmes âgées de 15-49 ans à cause des tumeurs a diminuée entre 2003 et 2018 au Maroc. En effet, cette proportion est passée de 14,8 % en 2003 à 10,9 % en 2011 puis à 10,8 % en 2014. Quant aux femmes décès à cause des maladies de l'appareil génito-urinaire, la proportion des femmes mortes a connue d'abord une diminution avant d'augmenter pour atteindre 3,3 %. En effet cette proportion est d'abord passée de 4 % en 2003 à 2,8 % en 2011 avec une baisse relative de 30 %. Puis avec une croissance de 15 %, cette proportion est passée de 2,8 % en 2011 à 3,3 % en 2018. Ainsi, on remarque que la proportion des femmes décédées à cause de grossesse, accouchement et puérperalité a connu une baisse sur l'ensemble de la période 2003-2018. En effet, cette proportion a diminué et est passée de 4 % en 2003 à 0,3 % en 2011 puis à 0 % en 2018. Concernant la catégorie des femmes décédées à cause de malformations congénitales et d'anomalies chromosomiques, la proportion des décès des femmes a connue d'abord une augmentation avant de diminuer pour atteindre 0,2 %. En effet, cette proportion est d'abord passée de 0,5 % en 2003 à 1,1 % en 2011 avec une hausse relative de 54 %. Puis avec une décroissance rapide de 82 %, cette proportion est passée de 1,1 % en 2011 à 0,2 % en 2018.

3. Niveaux et tendances de la planification familiale

3.1. Niveau de la planification familiale

Graphique n° 3 : Répartition en % des femmes âgées de 15-49 ans qui utilisent au moment de l'enquête une méthode contraceptive au niveau national



Source : Conception des auteurs à partir des données de l'ENPSF 2011

Le graphique circulaire ci-dessus montre que les deux tiers (67 %) des femmes enquêtées utilisent actuellement une méthode contraceptive.

3.1.1. Selon l'âge de la femme

Les données concédées par le tableau ainsi le diagramme ci-dessous nous représentent les estimations des femmes qui utilisent actuellement une méthode contraceptive selon le groupe d'âges et les différentes méthodes contraceptives. On remarque que la méthode contraceptive la plus utilisée est la pilule (48,4 %), ce qui représente environ 72 % de l'ensemble des méthodes contraceptives. Alors que le DIU est faiblement utilisé avec un taux de 4,2 % seulement. Concernant les méthodes traditionnelles, les plus utilisées sont l'abstinence périodique, le retrait et la MAMA avec des proportions respectives de 4,4 %, 4,1 % et 1,9 %.

Tableau n° 9 : Répartition en pourcentage des femmes mariées âgées de 15 à 49 ans qui utilisent au moment de l'enquête une méthode contraceptive selon la méthode utilisée et le groupe d'âges

Groupe d'âges	Méthode de Contraception Actuelle (Méthode principale)									
	Pourcentage qui Utilise actuellement une méthode contraceptive	% méthode Moderne	Pilule	DIU	Autres méthodes Modernes	MAMA	Abstinence périodique	Retrait	Autres	Effectif des femmes
15-19	37,8	32,6	32,1	0,5	0	2,6	1,4	1	0,3	384
20-24	59	52,8	50	1,9	0,9	1,5	2,1	2,4	0,2	1191
25-29	67,2	59	53,9	2,9	2,1	2,5	2,8	2,8	0,1	1688
30-34	73,6	62,4	54,2	4	4,3	2,7	4,1	4,3	0,1	1877
35-39	76,5	64,8	53,1	7,4	4,3	2,3	5,1	4,2	0,1	1832
40-44	73	59	47,2	5,3	6,6	1,4	6,6	5,6	0,3	1716
45-49	56,3	44	33,4	3,6	6,9	0,4	5,8	5,6	0,5	1492

Source : Conception des auteurs à partir des données de l'ENPSF 2011

3.1.2. Selon le milieu de résidence

Auprès des femmes mariées, la prévalence contraceptive, en milieu Urbain pour l'ensemble des méthodes utilisées est supérieure à celle du milieu rural : 69 % contre 65,5 %.

Tableau n° 10 : Répartition en pourcentage des femmes mariées âgées de 15 à 49 ans qui utilisent au moment de l'enquête une méthode contraceptive selon la méthode utilisée et le milieu de résidence

Milieu	Méthode de Contraception Actuelle (Méthode principale)									Effectif des femmes
	Pourcentage qui Utilise actuellement une méthode contraceptive	% méthode Moderne	Pilule	DIU	Autres méthodes Modernes	MAMA	Abstinence périodique	Retrait	Autres	
Urbain	68,9	57	46	6	5	1,1	5,8	4,8	0,2	5519
Rural	65,5	56,4	51,2	2,1	3,1	2,8	2,8	3,3	0,3	4661

Source : Conception des auteurs à partir des données de l'ENPSF 2011

3.1.3. Selon le quantile de bien-être

En ce qui concerne le quantile de richesse, les taux de prévalence contraceptive pour toutes les méthodes contraceptives vont de près de 65 % chez les femmes des ménages les plus pauvres à un peu plus de 71 % chez les femmes des ménages les plus riches.

Tableau n° 11 : Répartition en pourcentage des femmes mariées âgées de 15 à 49 ans qui utilisent au moment de l'enquête une méthode contraceptive selon la méthode utilisée et le quantile de bien-être

Quantile de bien être	Méthode de Contraception Actuelle (Méthode principale)								
	Pourcentage qui Utilise actuellement une méthode contraceptive	% méthode Moderne	Pilule	DIU	Autres méthodes Modernes	MAMA	Abstinence périodique	Retrait	Autres
Le plus pauvre	64,9	54,5	48,9	2,2	3,5	3,6	2,9	3,7	0,3
Second	65,4	56,7	51,3	2,3	3,2	2,3	3,3	3	0,1
Moyen	67,5	59	51,4	3,7	3,8	1	3,5	3,5	0,4
Quatrième	68	56,2	46,9	5,4	3,9	1,7	5,3	4,8	0,1
Le plus riche	71,3	57,3	43,2	7,6	6,5	0,8	7,3	5,7	0,2

Source : Conception des auteurs à partir des données de l'ENPSF 2011

3.1.4. Selon la situation professionnelle de la femme

On note sur le tableau n° 12 que la prévalence contraceptive pour toutes les méthodes augmente avec le niveau d'instruction : 67,1 % chez les femmes sans aucun diplôme scolaire : 67,3 % chez les femmes ayant un diplôme d'études de base, chez la plupart des femmes, 70,5 % d'entre elles sont scolarisées. En outre, plus le niveau d'études est élevé, plus l'utilisation du DIU.

Tableau n° 12: Répartition en pourcentage des femmes mariées âgées de 15 à 49 ans qui utilisent au moment de l'enquête une méthode contraceptive selon la méthode utilisée et la situation professionnelle

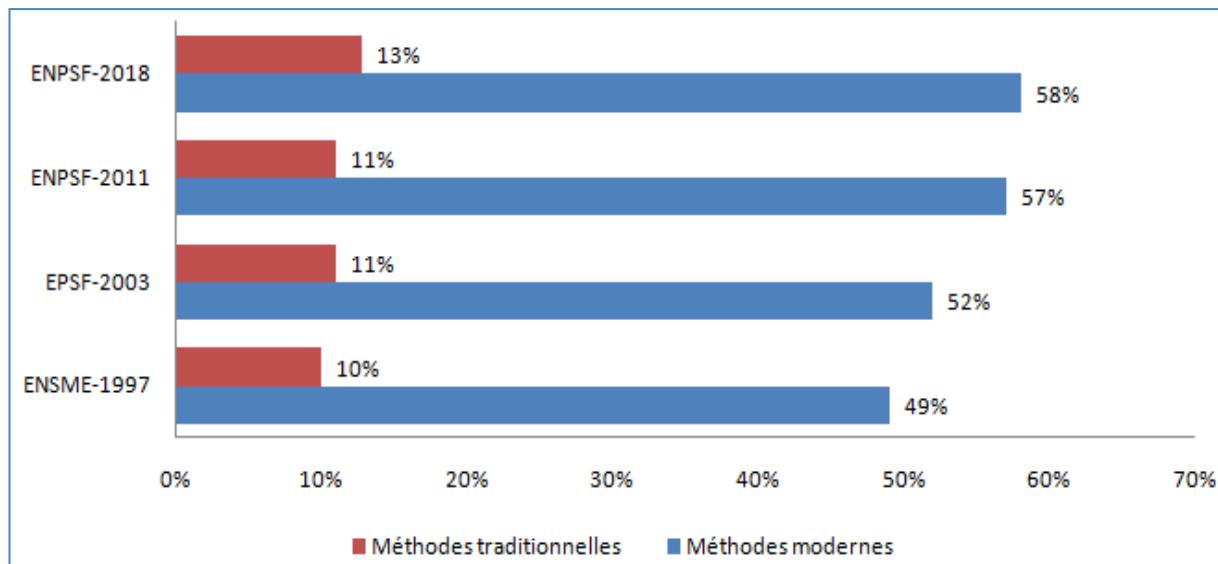
Situation Professionnelle	Méthode de Contraception Actuelle (Méthode principale)									
	Pourcentage qui Utilise actuellement une méthode contraceptive	% méthode Moderne	Pilule	DIU	Autres méthodes Modernes	MAMA	Abstinence périodique	Retrait	Autres	Effectif
Sans certificat	67,1	56,9	49,9	3,1	4	2,1	3,7	4,2	0,2	7324
Fondamental	67,3	56,3	46,1	6,2	4	1,5	5,9	3,5	0,2	2191
Secondaire et +	70,5	56,5	39,2	11	6,4	0,7	7,6	5,5	0,1	649
Travail actuellement	67,1	55,6	44,3	5,7	5,6	0,5	6,5	4,2	0,3	1145

Source : Conception des auteurs à partir des données de l'ENPSF 2011

3.2. Tendance de la planification familiale

L'utilisation des méthodes contraceptives modernes ou traditionnelles a augmenté chez les femmes âgées de 15 à 49 ans au Maroc entre 1997 et 2018. En effet, en 1997, la prévalence de la contraception moderne est de 49 % et a atteint la valeur 58 % en 2018. Concernant la méthode traditionnelle, le taux de la prévalence contraceptive en 1997 est de 10 % et a atteint 11 % en 2018. Le graphique n° 3 montre que la méthode contraceptive moderne est utilisée par les femmes que celle traditionnelle.

Graphique n° 4 : Evaluation des taux de la prévalence contraceptive selon que la méthode utilisée est moderne ou traditionnelle, selon différentes enquêtes réalisées au Maroc depuis 1997, ENPSF- 2018

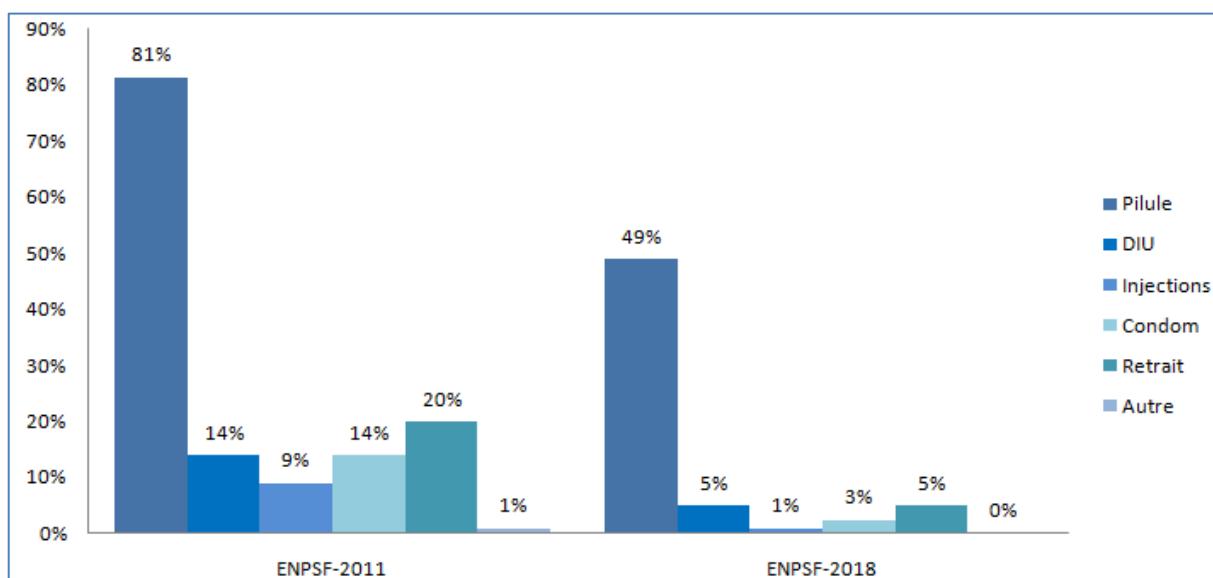


Source : Conception des auteurs à partir des données de l'ENPSF de 2011 et 2018

Il ressort de l'analyse du graphique n° 5 que :

- L'utilisation de pilule a atteint 81 % en 2011 et a baissé jusqu'à 49 % en 2018 ;
- La proportion des femmes utilisatrices du DIU est d'abord passée de 14 % en 2011 à 5 % en 2018 ;
- L'utilisation des injections a connu une baisse, en passant de 9 % en 2011 à 1 % en 2018 ;
- Le recours du condom a connu une diminution, en passant de 14% en 2011 à 3 % en 2018 ;
- L'utilisation de méthode traditionnelle « retrait » a baissé de 20 % en 2011 à 5 % en 2018 ;
- Les autres méthodes contraceptives ne sont plus utilisées en 2018.

Graphique n° 5 : Répartition (en %) de Femmes mariées âgées de 15 à 49 ans qui utilisent au moment de l'enquête une méthode contraceptive selon la méthode utilisée, selon les enquêtes réalisées au Maroc depuis 2011, ENPSF-2018



Source : Conception des auteurs à partir des données de l'ENPSF de 2011 et 2018

Conclusion

L'Agence Internationale de Recherche sur le Cancer évalue le taux de la prévalence sur 5 ans de 47,3 pour 100 000 femmes à travers le Maroc. Évidemment, les comparaisons sont encore difficiles à faire dans un certain nombre de limites : taille d'échantillon est relativement petite (11 069 femmes ayant été enquêtées), âge des femmes limité à 49 ans, la prévalence s'étendant aux femmes plus précoces et naturellement le biais inhérent à la perception de la maladie.

Par rapport aux enquêtes précédentes, la prévalence contraceptive en 2011 (67,4 %) a montré une tendance à la hausse et était de 4 points de pourcentage supérieur à celle de 2004 (63 %). 56,7 % de méthodes modernes (52 % selon EPSF 2003-2004) et 10,6 % de méthodes traditionnelles (11 % selon EPSF 2003-2004).

*Chapitre IV : Présentation
des résultats de l'analyse
explicative*

Introduction

Après avoir établi le cadre théorique de notre étude et présenté les données à notre disposition. Ce chapitre comprend trois sections. La première est consacrée à l'analyse bi-variée du croisement de la variable dépendante avec chacune des variables explicatives. La deuxième est consacrée au modèle d'arbre de classification qui détermine le profil des femmes qui n'ont pas subi l'événement « Planification familiale ». La troisième quant à elle, est consacrée à la régression logistique binaire pour comprendre les raisons qui amènent les mères n'a pas adopté les méthodes de contraceptions.

1. Résultats d'analyse descriptive

Dans le cadre d'analyse descriptive bi-variée, on utilise généralement les tableaux croisés pour caractériser les relations entre les deux séries d'observation considérées. La base de données qu'on va utiliser est issue de l'enquête nationale sur la population et la santé familiale 2011. On va croiser la variable **planification familiale** avec les variables explicatives qualitatives, pour connaître la nature de la relation qui existe entre eux.

1.1. Milieu de résidence

Selon le tableau n° 1 en annexe 3, 58,3 % des femmes qui utilisent des méthodes contraceptives sont des citadins contre 41,7 % des femmes ruraux.

D'après le résultat du test de khi-deux, on rejette l'hypothèse nulle et on peut dire qu'il y a de relation entre le milieu de résidence et la planification familiale.

1.2. Niveau scolaire le plus atteint

Concernant le niveau scolaire des femmes qui utilisent des méthodes contraceptives, 65,8 % sont au niveau collégial, suivi de type « Primaire » et « Lycée » avec 21,5 % et 6,5 % respectivement (voir le tableau n° 2 en annexe 3).

D'après le test de khi-deux, on accepte l'hypothèse nulle et on en déduit qu'il n'y a pas la relation entre le niveau scolaire de la femme et la planification familiale.

1.3. Situation matrimoniale

Concernant l'état matrimonial des femmes, 92,3 % des femmes qui utilisent des méthodes contraceptives sont des femmes mariées, suivies de type « Divorcée » et « Veuve » avec 3,5 % et 2,3 % respectivement (voir le tableau n° 3 en annexe 3).

D'après le test de khi-deux, on rejette l'hypothèse nulle et on en déduit qu'il y a la relation entre la situation matrimoniale de la femme et la planification familiale.

1.4. Travail

Selon le tableau n° 4 en annexe 3, 13,3 % des femmes qui utilisent des méthodes contraceptives sont des femmes travailleuses.

D'après le résultat du test de khi-deux, on accepte l'hypothèse nulle et on peut dire qu'il n'y a pas de relation entre la variable « Travail » et la planification familiale.

1.5. Nombre de fois mariée

Concernant le nombre de fois que les femmes marient, 94,3 % des femmes qui utilisent des méthodes contraceptives sont des femmes n'ont été mariées qu'une seule fois, tandis que 5,7 % ont été mariées plus d'une fois (voir le tableau n° 5 en annexe 3).

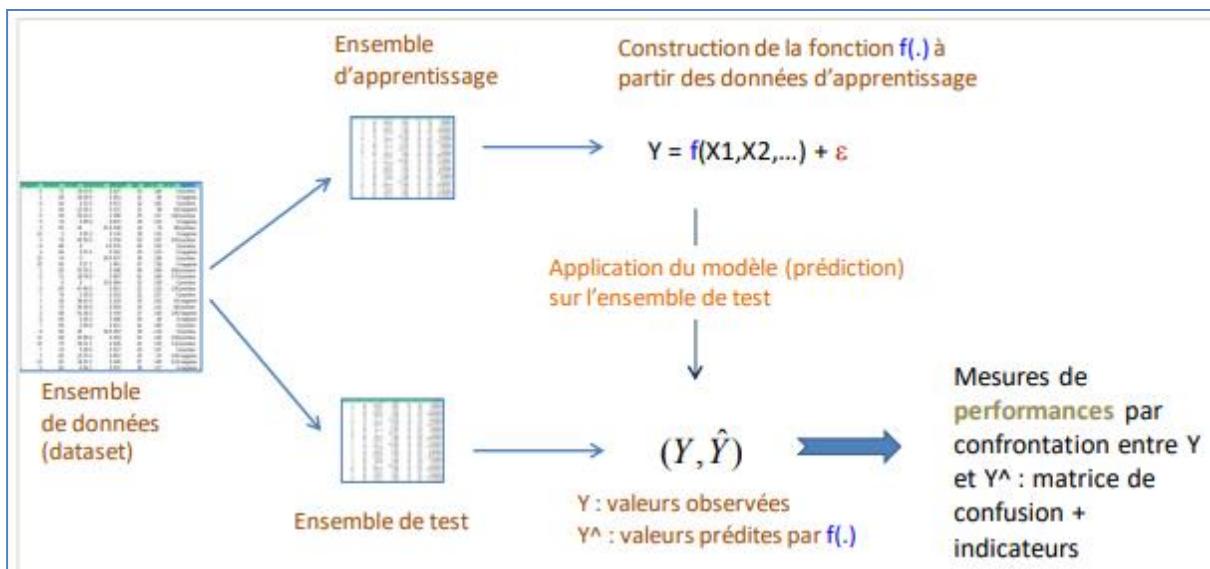
D'après le test de khi-deux, on accepte l'hypothèse nulle et on en déduit qu'il n'y a pas de relation entre la variable « Nombre fois mariée » et la planification familiale.

2. Résultats du modèle d'arbre de classification

Afin de déterminer le profil des femmes qui n'utilisent pas les méthodes contraceptives, nous cherchons à appliquer le schéma type de l'analyse prédictive : nous allons travailler avec les échantillons d'apprentissage et de test, en évaluer les performances sur le second à travers la confrontation des classes observées et prédites.

Nous allons réservier 7 748 observations pour l'apprentissage et 3 321 pour le test, avec un échantillonnage stratifié, c'est-à-dire respectant les proportions des classes dans les deux sous-ensembles.

Figure n° 4 : Schéma type de travail en analyse prédictive



Source : Tutoriel Tanagra Data Mining « Arbres de décision avec Scikit-Learn - Python », février 2020

Nous allons appliquer l'algorithme CART pour classifier les femmes selon leurs situations (utilisatrices des méthodes contraceptives ou pas).

D'après la figure en annexe 4, nous observons que :

- L'arbre est composé de 8 feuilles (sommets n° 8, 9, 10, 11, 12, 13, 14, 15). Il produit donc 8 règles prédictives matérialisées par les chemins partant de la racine aux feuilles ;
- L'effectif de l'échantillon d'apprentissage sur la racine de l'arbre ($n_{\text{Racine}} = n = \text{samples} = 7\,748$) avec 984 "Non" et 6 764 "Oui" (dans l'ordre alphabétique) ;
- Les sommets sont teintés selon la classe majoritaire qu'ils portent, avec plus ou moins d'intensité selon la concentration des effectifs. Ici, visiblement, le bleu est dédié à la réponse "Oui", l'orange à la réponse "Non" ;
- La concentration des classes est calculée à l'aide de l'indice de Gini (on parle aussi de mesure d'impureté.). Pour la racine (sommet n°1), nous avons ($K = 2$, nombre de modalités de la variable "PF") :

$$\begin{aligned} G(\text{Racine}) &= \sum_{k=1}^K \frac{n_{k,\text{Racine}}}{n_{\text{Racine}}} \left(1 - \frac{n_{k,\text{Racine}}}{n_{\text{Racine}}} \right) \\ &= \frac{984}{7748} * \left(1 - \frac{984}{7748} \right) + \frac{6764}{7748} * \left(1 - \frac{6764}{7748} \right) \sim 0,222 \end{aligned}$$

- "Nbr_enfant_vivant" est la variable de segmentation sur la racine, avec la condition "Nbr_enfant_vivant $\leq 1,147$ " ;
- La branche droite de la racine (sommet n°3) correspond à la proposition vraie de la condition, c'est-à-dire "Nbr_enfant_vivant $\leq 1,147$ ". Nous lisons sur le sommet enfant qu'elle correspond à (samples = 4 660) observations, avec 828 "Non" et 3 832 "Oui". Nous avons $G(\text{sommet}) = 0.292$. Plus la valeur de l'indice de Gini est faible, plus les classes sont concentrées sur un sommet ;
- La branche gauche de la racine (sommet n°2) correspond à la négation de la proposition, c'est-à-dire "Nbr_enfant_vivant $> 1,147$ ", elle concerne 3 088 observations avec 156 "Non" et 2 932 "Oui" ;
- Lorsque les variables de segmentation sont quantitatives, il n'est pas rare qu'elles apparaissent plusieurs fois dans l'arbre, mais avec des seuils de découpage différents. C'est le cas de variables "Nbr_enfant_vivant" et "Age" ici.

L'affichage graphique est sympathique, mais devient peu lisible dès lors que la taille de l'arbre augmente. C'est pourquoi, nous limitons sa profondeur à 3. On obtient donc les 8 classes décrivant les différents profils des non-utilisatrices des méthodes contraceptives :

- Classe 1 : représente 0,18 % des femmes enquêtées (sommet n° 8)
 - 71,4 % n'utilisent pas les méthodes contraceptives ;
 - Les femmes ayant un nombre des enfants en vie > 1,147, leurs situations matrimoniales sont mariées et leurs âges ≤ 21,5 ans.
- Classe 2 : représente 1,52 % des femmes enquêtées (sommet n° 9)
 - 14,4 % n'utilisent pas les méthodes contraceptives ;
 - Les femmes ayant un nombre des enfants en vie > 1,147, leurs situations matrimoniales sont mariées et leurs âges > 21,5 ans.
- Classe 3 : représente 1,93 % des femmes enquêtées (sommet n° 10)
 - 15,3 % n'utilisent pas les méthodes contraceptives ;
 - Les femmes ayant un nombre des enfants en vie > 1,147, leurs situations matrimoniales sont mariées et leurs âges ≤ 20,5 ans.
- Classe 4 : représente 36,2 % des femmes enquêtées (sommet n° 11)
 - 3,8 % n'utilisent pas les méthodes contraceptives ;
 - Les femmes ayant un nombre des enfants en vie > 1,147, leurs situations matrimoniales sont mariées et leurs âges > 20,5 ans.
- Classe 5 : représente 5 % des femmes enquêtées (sommet n° 12)
 - 57,1 % n'utilisent pas les méthodes contraceptives ;
 - Les femmes ayant un nombre des enfants en vie ≤ 1,647, et leurs âges ≤ 23,5 ans.
- Classe 6 : représente 1,4 % des femmes enquêtées (sommet n° 13)
 - 4,5 % n'utilisent pas les méthodes contraceptives ;
 - Les femmes ayant un nombre des enfants en vie > 1,647, et leurs âges ≤ 23,5 ans.
- Classe 7 : représente 39,8 % des femmes enquêtées (sommet n° 14)
 - 18,2 % n'utilisent pas les méthodes contraceptives ;
 - Les femmes ayant un nombre des enfants en vie ≤ 1,647, et leurs âges > 23,5 ans.
- Classe 8 : représente 13,8 % des femmes enquêtées (sommet n° 15)
 - 3,6 % n'utilisent pas les méthodes contraceptives ;
 - Les femmes ayant un nombre des enfants en vie > 1,647, et leurs âges > 23,5 ans.

La qualité globale de l'arbre peut être quantifiée par la différence entre l'indice de Gini de la racine (sommet n°1) et la moyenne pondérée des Gini des L feuilles (n° 8, 9, 10, 11, 12, 13, 14, 15), chacune avec un effectif n_l :

$$\Delta_{Arbre} = G(\text{Racine}) - \sum_{l=1}^L \frac{n_l}{n} G(l)$$

Dans notre cas,

$$\Delta_{Arbre} = 0,032$$

Pour évaluer les performances prédictives de l'arbre, nous l'appliquons sur l'échantillon test de taille 3 321. Nous obtenons une prédiction :

Oui	Non
3168	153
Classes observées	Classes prédites

Afin de vérifier, est-ce à juste titre, nous confrontons les classes observées et prédites via la matrice de confusion.

Tableau n° 13 : Matrice de confusion

		Non	Oui
Non	89	333	
Oui	64	2835	

Source : Conception des auteurs à partir de sortie Python

Nous pouvons présenter les indicateurs de performances usuels de la manière suivante :

Figure n° 5 : Indicateurs de performances

	precision	recall	f1-score	support
Non	0.58	0.21	0.31	422
Oui	0.89	0.98	0.93	2899
accuracy			0.88	3321
macro avg	0.74	0.59	0.62	3321
weighted avg	0.86	0.88	0.86	3321

Source : Conception des auteurs à partir de sortie Python

3. Résultats du modèle de la régression logistique

3.1. Rappel des variables retenues

La variable dépendante **Planification familiale** qui définit comme suit :

$$Y = \begin{cases} 0 & \text{si la femme utilise les méthodes contraceptives} \\ 1 & \text{sinon} \end{cases}$$

Nous essayons d'estimer celle-là en fonction des variables explicatives suivantes :

- **L'âge, le statut matrimonial et le nombre d'enfants en vie** de la femme
- **Nombre de fois marié**
- **Nombre de fois des consultations**
- **Niveau d'instruction**
- **Milieu de résidence**
- **Travail**

3.2. Modélisation de la planification familiale

Nous avons effectué une régression linéaire généralisée sur toutes les variables explicatives en utilisant le modèle logistique.

Tableau n° 14 : Résultats du modèle global de régression logistique

Variables	Coef	Std.err	z	P_value
Constante	-2,283	0,371	-6,146	0,000
Milieu	0,259	0,071	3,644	0,000
Age	-0,037	0,004	-8,770	0,000
Niveau_scolarisation	0,065	0,046	1,423	0,155
Situation matrimoniale	0,605	0,048	12,536	0,000
Nbr_fois_marié	0,225	0,146	1,545	0,122
Travail	0,200	0,113	1,782	0,075
Nbr_enfant_vie	-0,176	0,089	-1,985	0,047
Nbr_fois_consultation	-0,009	0,011	-1,821	0,412

Source : Conception des auteurs à partir de sortie Python

On retrouve que les variables **Niveau_scolarisation**, **Nbr_fois_marié**, **Travail** et **Nbr_fois_consultation** ne sont pas significatives. En effet, leurs p_value sont supérieurs à 5 %.

Ensuite, nous testons la nullité simultanée des coefficients en comparant entre le modèle nul et le modèle logistique global.

- **Test du rapport de vraisemblance**

Test du rapport de vraisemblance	
LR	Signif.
240,91	0,00

Source : Conception des auteurs à partir de sortie Python

- **R² de Mc Fadden**

R ² de Mc Fadden	
	0,048
Source : Conception des auteurs à partir de sortie Python	

- **Test de Hosmer et Lemeshow**

Test de Hosmer et Lemeshow	
HL	Signif.
45,46	0,00

Source : Conception des auteurs à partir de sortie Python

- **Critère AIC & BIC**

	AIC	BIC
Modèle nul	5900,47	5907,42
Modèle travail	5675,56	5738,15

Source : Conception des auteurs à partir de sortie Python

Les quatre résultats montrent le rejet d'hypothèse de nullité simultanée des coefficients des variables explicatives (au niveau de 5 %).

En outre, nous vérifions que les coefficients des variables non-significatives peuvent être considérés comme simultanément nuls. Nous pouvons par la suite réaliser les deux tests :

- **Test du rapport de vraisemblance**

Test du rapport de vraisemblance		
LR	Ddl	P_value
7,53	4	0,1102

Source : Conception des auteurs à partir de sortie Python

- **Test de Wald**

Test de Wald		
LR	Ddl	P_value
7,39	4	0,1164

Source : Conception des auteurs à partir de sortie Python

D'après ces deux tests, nous constatons la nullité simultanée des coefficients de (Niveau_scolarisation, Nbr_fois_marié, Travail et Nbr_fois_consultation) au risque 5 %.

Après avoir testé la significativité du modèle logistique global, nous avons obtenus le modèle final comme suit :

Tableau n° 15 : Résultat final du modèle de régression logistique

Variables	Coef	Std.err	z	P_value
Constante	-1,602	0,221	-7,261	0,000
Milieu	0,264	0,070	3,758	0,000
Age	-0,036	0,004	-8,798	0,000
Situation matrimoniale	0,586	0,046	12,684	0,000
Nbr_enfant_vie	-0,167	0,089	-1,881	0,060

Source : Conception des auteurs à partir de sortie Python

Tableau n° 16 : Résultat final détaillé du modèle de régression logistique

Variables	Coef	P_value	Exp(Coef)
Constante	0,394	0,085	1,483
Milieu (1)	0,256	0,000	1,291
Age	-0,039	0,000	0,962
Situation matrimoniale(1)	0,373	0,036	1,451
Situation matrimoniale(2)	-1,086	0,000	0,338
Nbr_enfant_vie	-0,157	0,036	0,855

Source : Conception des auteurs à partir de sortie SPSS

Selon le tableau n° 16, au seuil de 5 %, toutes les variables sont pertinentes puisque leur p_value est inférieure à 5 %.

Tableau n° 17 : Matrice de confusion

		Classes prédictes	
		0	1
Classes observées	0	2891	8
	1	419	3

Source : Conception des auteurs à partir de sortie Python

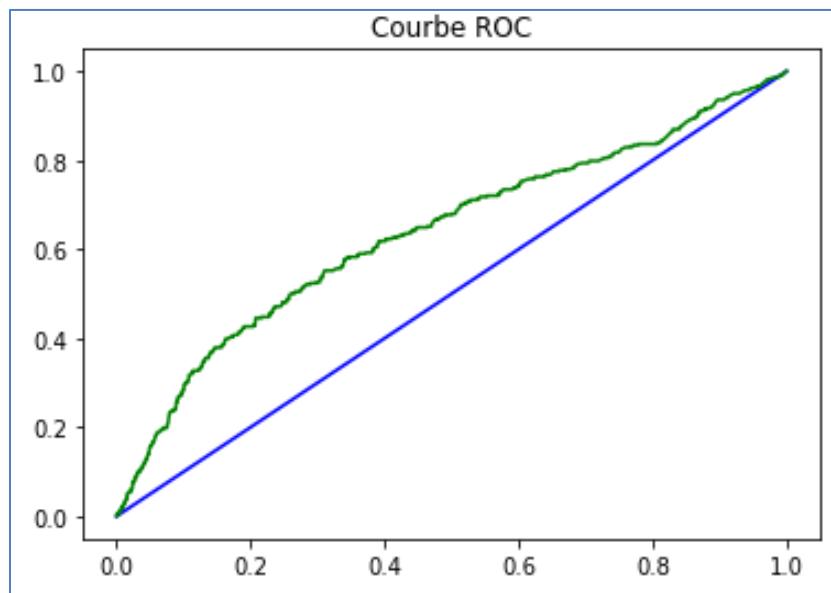
À l'aide de la matrice de confusion, nous pouvons calculer les indices comme la sensibilité et la spécificité.

Sensibilité	0,0071
Spécificité	0,99

À partir de ce résultat, nous concluons que :

- La capacité de connaître les non-utilisatrices des méthodes contraceptives parmi les vraies non-utilisatrices est de l'ordre de 0,0071.
- La capacité de connaître les utilisatrices des méthodes contraceptives parmi les vraies utilisatrices est de l'ordre de 0,99.

Figure n° 6 : Courbe ROC du modèle final



Source : Sortie Python

Le modèle peut être jugé acceptable, car l'aire sous la courbe ROC est de 0,6376. En d'autre terme, le pourcentage de prédiction, c'est-à-dire la capacité de notre modèle à prédire correctement la non-participation de la planification familiale situerait autour de 64 %.

3.3. Interprétation des résultats

Dans la présentation des résultats, il se dégage de cette analyse que le milieu de résidence est un facteur explicatif important de non-utilisation des méthodes contraceptives. Les femmes en milieu rural sont moins susceptibles d'être non-utilisatrice des méthodes contraceptives. En effet, ces femmes ont 1,29 fois moins de risque de ne pas utiliser les méthodes contraceptives que les femmes citadines. Ce résultat pourrait s'expliquer que les femmes en milieu urbain ne participent pas beaucoup à la planification familiale.

Par ailleurs, plus l'âge de la femme augmente et plus la probabilité d'être non-utilisatrice des méthodes contraceptives diminue. Ce qui implique que les femmes qui ne participent pas à la planification familiale (PF) sont moins âgées.

La situation matrimoniale à montrer que le fait d'être marié augmenter la probabilité des femmes de ne pas utiliser les méthodes de planification familiale soit une probabilité 37 % par rapport à ceux qui ne sont pas marié. Ce résultat montre que la situation matrimoniale influence sur le niveau de fécondité.

Concernant la variable « Nbr_enfant_vie », plus le nombre des enfants vivants de la femme augmente et plus la chance de ne pas utiliser les méthodes contraceptives baisse. Ce résultat peut être expliqué par les différents programmes initiés par le gouvernement en matière d'amélioration au service de santé de la reproduction.

La majorité des femmes qui ne bénéficient pas du programme de la planification familiale ont une faible probabilité d'améliorer leur santé maternelle. C'est pourquoi, le gouvernement marocain faut engager avec ses différents partenaires au développement à travers les différents programmes afin de renforcer les capacités des centres de soin en matière de la santé de reproduction.

En résumé, cet effet positif de la planification familiale sur la santé maternelle amélioré la qualité de vie de la population en général, plus particulièrement le capital humain au Maroc.

3.4. Comparaison entre modèle de régression logistique et celui d'arbre de classification

Les deux modèles montrent que les principaux facteurs qui poussent les femmes à ne pas adopter la planification familiale sont le milieu de résidence, l'âge et la situation matrimoniale de la femme et le nombre total des enfants vivants. Pour comparer ces deux modèles, nous avons calculé les indicateurs des performances :

Tableau n° 18 : Comparaison des deux modèles

Modèle	Taux de reconnaissance	Taux d'erreur	Total
Arbre de classification	0,88	0,12	1
Régression logistique	0,87	0,13	1

Source : Conception des auteurs à partir de sortie Python

Nous concluons à partir de ce tableau que le modèle d'arbre de classification permet de prédire bien la non-adaptation de la planification familiale que celui de régression logistique.

Conclusion

Dans le présent chapitre, nous avons effectué une analyse bi-varieée en croisement la variable dépendante et les variables indépendantes. Cette analyse nous a permis de déterminer les variables explicatives les plus significatives pour réaliser nos modèles. Nous avons ensuite construit deux modèles (Arbre de classification et Régression logistique). À partir de ces deux modèles, les résultats obtenus montrent que l'âge de la femme, le milieu de résidence, la situation matrimoniale, ainsi que le nombre des enfants vivants, influent à la non-participation de la planification familiale.

CONCLUSION GÉNÉRALE

L'objectif général de l'étude est d'élucider l'impact des facteurs socio-économiques sur la non-participation à la planification familiale afin de déterminer le profil des femmes qui n'adoptent pas cette politique et les raisons pour lesquelles n'utilisent pas les méthodes contraceptives. Ainsi d'une manière spécifique, il a fallu déterminer les niveaux et tendances de la santé maternelle au Maroc en matière de mortalité et de morbidité maternelle et de planification familiale.

Pour atteindre les objectifs poursuivis dans cette étude, le contexte de cette étude a été décrit dans un premier temps, suivi de la revue de littérature qui a permis d'identifier les facteurs influençant la santé maternelle. Dans cette logique, les données utilisées dans cette étude ont été examinées et la qualité de ces dernières est jugée acceptable dans l'ensemble. Compte tenu des données disponibles, nous avons focalisé notre travail sur la planification familiale en tant que variable cible et l'une des dimensions de la santé maternelle qui a une incidence positive. Les schémas d'analyse (arbre de classification et régression logistique) ont permis de mettre en relation les facteurs socio-économiques et la non-participation à la planification familiale.

Il ressort des analyses des niveaux et tendances de la santé maternelle que le Maroc a fait d'importants progrès en réduction de la mortalité maternelle où le taux de mortalité maternelle a diminué de près de 63 % en dix-sept ans, passant de 188 décès maternels pour 100 000 naissances vivantes en 2000 à 70 en 2017. Des diminutions s'observent au niveau des maladies qui concernent les mères. À titre d'exemples, la proportion des décès des femmes âgées de 15-49 ans à cause des tumeurs est passée de 14,8 % en 2003 à 10,8 % en 2014. Ainsi, la proportion des femmes décédées à cause de grossesse, accouchement et puérperalité a connu une baisse sur l'ensemble de la période 2003-2018. Concernant la planification familiale, en 2011, 67 % des femmes enquêtées utilisent actuellement une méthode contraceptive. La pilule est le contraceptif le plus utilisé (49 %). En outre, on note un faible recours aux injectables (5 %) et au préservatif (3 %). Par ailleurs, les femmes marocaines ont un fort recours aux contraceptifs modernes (58 %).

En nous fondant sur cette étude, nous avons obtenu plusieurs résultats.

L'analyse à deux variables a révélé que des facteurs socioéconomiques tels que l'âge de la femme, le milieu de résidence et l'état matrimonial sont tous significativement associés à la planification familiale.

L'arbre de classification a permis d'identifier parmi les caractéristiques des femmes, le profil des femmes en union, utilisatrices et non utilisatrices de la contraception. Dans notre cas, nous nous intéressons aux femmes qui ne recourent pas à la contraception, nous constatons que ces femmes sont généralement des femmes mariées dont l'âge est supérieur à 20,5 ans et plus ayant en moyen 1,147 enfant vivant.

Afin d'identifier les déterminants significatifs de la non-utilisation de la contraception, nous avons eu recours au modèle de régression logistique. À partir des résultats obtenus, nous retenons que les facteurs influençant significativement la non-utilisation de la contraception sont l'âge de la femme, le milieu de résidence, la situation matrimoniale, ainsi que le nombre des enfants vivants. Dans ce cadre, les résultats peuvent être résumés comme suit :

- Plus l'âge de la femme augmente et plus la probabilité d'être non-utilisatrice des méthodes contraceptives diminue ;
- Les femmes mariées ont 1,45 fois plus de chances de ne pas utiliser les méthodes de contraception par rapport à ceux qui ne sont pas mariées ;
- Les femmes en milieu rural ont 1,29 fois moins de risque de ne pas utiliser les méthodes contraceptives que les femmes citadines ;
- Plus le nombre des enfants vivants de la femme augmente et plus la chance de ne pas utiliser les méthodes contraceptives baisse.

Au vu de ces résultats, nous formulons les recommandations suivantes :

- Encourager les femmes plus âgées à participer à la planification familiale ;
- Accroître la sensibilisation des femmes à l'utilisation de la contraception moderne, notamment des pilules et des produits injectables, pour une planification efficace des naissances
- Mener des activités de communication auprès des femmes des régions rurales afin de s'assurer qu'elles ont de bonnes croyances au sujet de la planification familiale et des effets de l'utilisation de contraceptifs modernes. Pour ce faire, il faut mener une vaste campagne de sensibilisation avec des messages publicitaires dans les médias afin de montrer l'importance du recours des femmes aux méthodes de contraception moderne ;
- Faire comprendre aux femmes en union l'importance du contrôle de leur procréation et de leur capacité à contrôler leur santé reproductive. Et qu'il existe un ensemble de méthodes efficaces pour éviter la grossesse indésirable sans compromettre la santé de la femme.



BIBLIOGRAPHIE



Bibliographie

1. Bernis (2005), La mortalité maternelle dans le monde
2. Breiman et al, California: Wadsworth international (1984), Classification and Regression Tree
3. Brouwere (2017), La réduction de la mortalité maternelle dans les pays en voie de développement : Un enjeu de société
4. Bruno DUJARDIN, Françoise MIME et Vincent DE BROUWERE (2014), Améliorer la santé maternelle : un guide pour l'action systémique
5. Centre de Santé Reproductive & London School of Hygiene and Tropical Medicine (2020), Maroc : étude exemplaire en santé maternelle et nouveau-née
6. Data Mining : Les modèle CHAID : Des arbres de classification bien adoptés pour la segmentation en Marketing et CRM, <http://www.statsoft.fr/concepts-statistiques/modeles-chaid/modeles-chaid.htm>
7. Données de la Banque Mondiale, Evolution de la mortalité maternelle au Maroc de la période 2000-2017
8. E.Azria, Z.Stewart, C.Gonthier, C.Estellat et C.Deneux-Tharaux (), Social inequalities in maternal health
9. Fadoua BADAOUI (2022), support de cours de modèle linéaire généralisé : Régression logistique
10. J.R. Quinlan, Kluwer Academic publishers (1986), Introduction of Decision Trees
11. Lawrence O.Hall, Nitesh Chawla et Kevin W.Bowyet (1998), Decision Tree Learning on Very Large Data Sets
12. LEKOUKA Fabien Constantin (2021), Impact of the family planning on maternal health in the Republic of Congo, Congo
13. Ministère de la Santé (2005), Rapport de l'Enquête sur la Population et la Santé Familiale 2003/2004
14. Ministère de la Santé (2012), Rapport de l'Enquête Nationale sur la Population et la Santé Familiale 2011
15. Ministère de la Santé (2018), Rapport de l'Enquête Nationale sur la Population et la Santé Familiale 2018
16. Mustapha BERROUYNE (2022), Support de cours des méthodes d'évaluation des données démographiques

Bibliographie

17. OMS (2015), Surveillance des décès maternels et riposte : prévention des décès maternels, informations au service de l'action
18. OMS (2019), Évolution de la mortalité maternelle 2000-2017
19. Pierre Fournier, Slim Haddad et Valéry Ridde (2013), Santé maternelle et accès aux soins de santé en Afrique de l'Ouest
20. Ricco RAKOTOMALALA (2005), Arbres de Décision
21. Santé et services sociaux Québec (2010), Cadre conceptuel de la santé et de ses déterminants
22. Stuart Russell et Peter Norvig (1995), Intelligence Artificial: a modern approach
23. UNICEF (2015), Les objectifs de Développement Durable (ODD) : 15 minutes pour comprendre
24. Université Lyon 2 (2020), Tutoriels Tanagra « Regression Logistique -Python »
25. Université Lyon 2 (2020), « Arbre de décision avec Scikit-Learn-Python »



ANNEXES



Annexe 1 : Présentation de l'organisme d'accueil

Le Centre d'Etudes et des Recherches Démographiques CERED est une division du Haut Commissariat au Plan chargée d'effectuer des études et des recherches qui touchent la démographie. Ce centre est né en 1970 grâce à la coopération entre le FNUAP et le gouvernement marocain. Il est devenu une entité relevant directement du ministère chargé de la population depuis 1996 et du Haut Commissariat au Plan depuis 2003.

Le CERED vise à traiter les thèmes qui touchent le plus la population marocaine. On peut citer à titre d'exemple :

- ✓ La fécondité et ses déterminants
- ✓ La mortalité et ses déterminants
- ✓ La migration interne, et celle internationale
- ✓ Projection de la population
- ✓ Politique de la population
- ✓ Santé de la reproduction
- ✓ Composition et structure des ménages
- ✓ Etat civil et ses composantes
- ✓ Développement du monde rural

Le Centre d'Etudes et de Recherches Démographiques a pour mission de :

- 1- Recueillir et centraliser toutes les informations et études à caractère national ou international, relatives aux problèmes de population ;
- 2- Assurer la diffusion et la vulgarisation des résultats de ses travaux de recherche, études et analyses démographiques et sociales du centre auprès de tous les utilisateurs ;
- 3- Elaborer des projections à court, moyen et long termes en vue de dégager les tendances ainsi que les caractéristiques futures de la population du Maroc ;
- 4- Elaborer des méthodes précises pour la mesure des changements démographiques et des facteurs qui s'y rapportent, en vue de l'établissement de modèles de projection de population aux niveaux national, sectoriel et par catégorie sociodémographique ;
- 5- Préparer, sur la base des études réalisées par le Centre, les dossiers techniques nécessaires à l'élaboration et au suivi de la politique nationale de la population, dans le cadre de la Commission Supérieure de la Population ;
- 6- Mener des enquêtes spécifiques dans les domaines démographique et social ;
- 7- Procéder à des travaux théoriques et appliqués en matière démographique et sociale ;
- 8- Fournir aux tiers les services d'experts en matière démographique et sociale ;

- 9- Préparer et assurer le suivi des programmes de formation continue, de perfectionnement et d'assistance technique dans les domaines démographique et social ;
- 10- Coopérer avec les institutions à caractère similaire et participer aux manifestations traitant de la recherche et des études sur la population.

Annexe 2 : Méthodes d'évaluation de la qualité des données

1. Indice de Myers

L'indice de MYERS mesure l'attraction ou la répulsion des âges se terminant par chacun des chiffres entre 0 à 9. Pour calculer cet indice, nous avons besoin de la répartition de la population par année d'âge et par sexe. Le calcul se fait en faisant les étapes suivantes :

La première étape :

On calcule les sommes S_u des effectifs des personnes de 10 ans et plus dont les âges se terminent respectivement par chacun des chiffres de 0 à 9.

Soit $P(10d+u)$: l'effectif des personnes dont l'âge a pour chiffre des dizaines d et pour chiffre des unités u .

Notons S_u : l'effectif des personnes âgées de 10 ans et plus dont l'âge à comme chiffre des unités u :

$$S_u = \sum_{d=1}^{d_{max}} P(10d + u)$$

La deuxième étape :

Notons aussi S'_u , l'effectif des personnes âgées de 20 ans et plus dont l'âge à comme chiffre des unités u .

On calcule les sommes S'_u de manière suivante :

$$S'_u = \sum_{d=2}^{d_{max}} P(10d + u)$$

La troisième étape :

Les effectifs remaniés de Myers sont les quantités définies par :

$$T = (u + 1) S_u + (9 - u) S'_u$$

La quatrième étape :

On calcule l'effectif remanié total T

$$T = \sum_{u=0}^9 T_u$$

La dernière étape :

L'indice de MYERS vaut alors :

$$I_M = \sum_{u=0}^9 \left| 100 * \frac{T_u}{T} - 10 \right|$$

Selon la conception, I_M se situe entre 0 et 180. Dans le cas où nous avons une bonne déclaration des âges alors : $T_u/T = 10\%$ et $I_M = 0$.

Plus on est proche de la valeur 0 plus la qualité de données est bonne. Si T_u/T est supérieur à 10%, alors on a une attraction aux âges ayant comme chiffre d'unités u. Or dans le cas inverse, on a une répulsion aux âges ayant comme chiffre d'unités u.

2. Indice combiné des Nations Unies

L'indice combiné des Nations Unies a une nature différente des trois indices exposés précédemment. En effet, il se calcule à partir de répartitions par groupes d'âges et non plus par années d'âge, et il tente de mesurer la régularité des répartitions par sexe et par âge. Les étapes de calcul sont les suivantes :

La première étape :

Notons $P_m(i)$ et $P_f(i)$, les effectifs des différents groupes d'âges quinquenaux pour les sexes masculin et féminin, de $i = 0$ (0-4 ans) à $i = 14$ (70-74 ans).

Notons $m(i)$ et $f(i)$, les rapports des groupes d'âges pour chaque sexe de $i = 1$ à 13.

On calcule ces rapports en rapportant l'effectif de chaque groupe d'âges à la moyenne arithmétique des effectifs des groupes d'âges encadrants.

$$m(i) = 100 * \frac{P_m(i)}{\frac{1}{2}[P_m(i-1) + P_m(i+1)]}; \quad f(i) = 100 * \frac{P_f(i)}{\frac{1}{2}[P_f(i-1) + P_f(i+1)]}$$

La deuxième étape :

Puis, on calcule les écarts par rapport à 100 de chacun des rapports précédents, puis les indices du rapport des groupes d'âges :

$$J_m = \frac{1}{13} \sum_{i=1}^{13} |m(i) - 100|; \quad J_f = \frac{1}{13} \sum_{i=1}^{13} |f(i) - 100|$$

La troisième étape :

Ensuite, on calcule les rapports de masculinité :

$$r(i) = 100 * \frac{P_m(i)}{P_f(i)}$$

Et calculons les différences entre rapports de masculinité successifs, la moyenne de valeurs absolus de ces différences est :

$$k = \frac{1}{13} \sum_{i=1}^{12} |r(i) - r(i+1)|$$

La dernière étape :

L'indice combiné des Nations Unies vaut alors :

$$\text{ICNU} = J_m + J_f + 3k$$

Alors :

Si $\text{ICNU} < 20$, les données sont de bonne qualité ;

Si $20 \leq \text{ICNU} < 40$: les données sont de relative bonne qualité et peuvent être ajustées ;

Si $\text{ICNU} \geq 40$, les données seront jugées de très mauvaise qualité.

Annexe 3 : Analyse bi-variée à partir des sorties Python

1. Résultats

Tableau n°1 : Pourcentage des femmes participantes et non participantes à la planification familiale selon le milieu de résidence

		Planification familiale		Total
Milieu de résidence	Urbain	Oui	Non	
	Rural	4032	680	4712
	Total	9663	1406	11069

Khi-2 = 21,85

P_value = $2,94 * 10^{-6}$

Ddl = 1

Tableau n°2 : Pourcentage des femmes participantes et non participantes à la planification familiale selon le niveau scolaire le plus atteint

Niveau scolaire le plus atteint	Planification familiale		Total
	Oui	Non	
Primaire	2083	319	2402
Collège	6358	911	7269
Lycée	636	97	733
Supérieur	402	45	447
Alphabétisation	184	34	218
Total	9663	1406	11069

Khi-2 = 5,54

P_value = 0,23

Ddl = 4

Tableau n°3 : Pourcentage des femmes participantes et non participantes à la planification familiale selon la situation matrimoniale

Situation matrimoniale	Planification familiale		Total
	Oui	Non	
Mariée	8983	1139	10122
Veuve	226	59	285
Divorcée	336	166	502
Séparée	118	42	160
Total	9663	1406	11069

Khi-2 = 249,87

P_value = $6,96 * 10^{-54}$

Ddl = 3

Tableau n°4 : Pourcentage des femmes participantes et non participantes à la planification familiale selon le travail

		Planification familiale		Total
		Oui	Non	
Travail	Oui	1282	191	1473
	Non	8381	1215	9596
Total		9663	1406	11069

Khi-2 = 0,08

P_value = 0,77

Ddl = 1

Tableau n°5 : Pourcentage des femmes participantes et non participantes à la planification familiale selon le nombre de fois mariée

		Planification familiale		Total
		Oui	Non	
Nombre de fois mariée	Une fois	9106	1308	10414
	Plus d'une fois	557	98	655
Total		9663	1406	11069

Khi-2 = 2,99

P_value = 0,08

Ddl = 1

2. Code Python

```
#Importer les librairies utilisées pour notre étude
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import xlrd
#Importer les données
Data= pd.read_excel(r'C:\Users\ADMIN\Desktop\Stage\Women.xlsx')
#Informations sur les variables
print(Data.info( ))
```

#	Column	Non-Null Count	Dtype
0	Milieu	11069 non-null	object
1	Age	11069 non-null	int64
2	Situation matrimoniale	11069 non-null	object
3	Niveau_scolarisation	11069 non-null	object
4	Travail	11069 non-null	object
5	Nbr_fois_mariée	11065 non-null	object
6	Nbr_enfant_vivant	11069 non-null	float64
7	Nbr_fois_consultation	11069 non-null	float64
8	PF	11069 non-null	object

dtypes: float64(2), int64(1), object(6)


```
#Mettre les données sous forme une matrice
Data=pd.DataFrame (Data)
#Générer les tableaux de contingence et calculer les statistiques Khi-2et les p_value
from scipy.stats import chi2_contingency as chi2_contingency

X="Milieu"
Y="PF"
Cont1 = data[[X, Y]].pivot_table(index=X, columns=Y, aggfunc=len).fillna(0).copy().astype(int)
st_chi2, p_value = chi2_contingency(Cont1)

#####
X="Niveau_scolarisation"
Y="PF"
Cont2 = data[[X, Y]].pivot_table(index=X, columns=Y, aggfunc=len).fillna(0).copy().astype(int)
st_chi2, p_value = chi2_contingency(Cont2)
#####

X="Situation_matrimoniale"
Y="PF"
Cont3 = data[[X, Y]].pivot_table(index=X, columns=Y, aggfunc=len).fillna(0).copy().astype(int)
st_chi2, p_value = chi2_contingency(Cont3)
#####

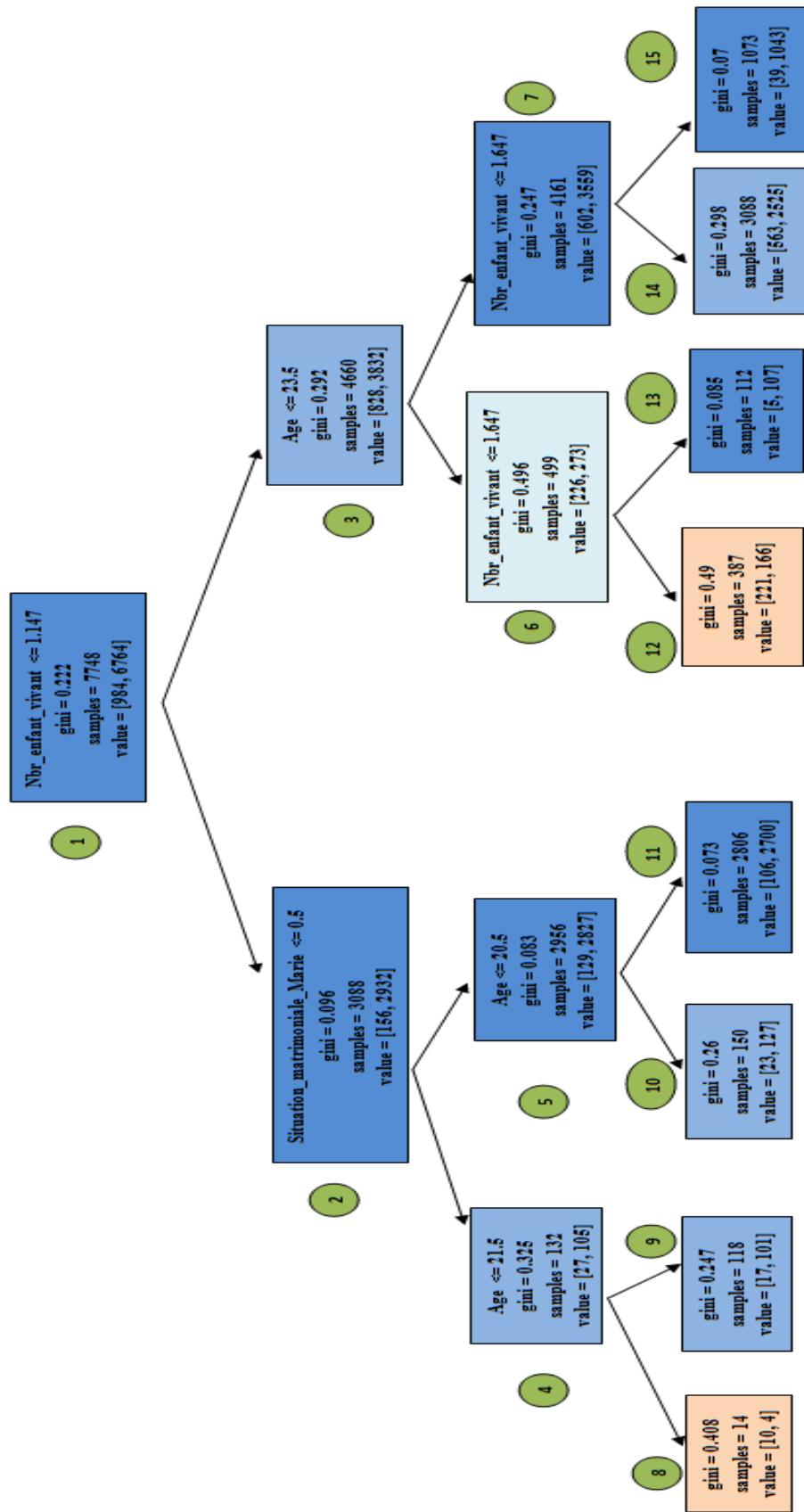
X="Travail"
Y="PF"
Cont4 = data[[X, Y]].pivot_table(index=X, columns=Y, aggfunc=len).fillna(0).copy().astype(int)
st_chi2, p_value = chi2_contingency(Cont4)
#####

X="Nbr_fois_mariée"
Y="PF"
Cont5= data[[X, Y]].pivot_table(index=X, columns=Y, aggfunc=len).fillna(0).copy().astype(int)
st_chi2, p_value = chi2_contingency(Cont5)
```

Source : Nicolas Rangeon, Tutoriel Tanagra « Réaliser un test d'indépendance Khi² en Python avec scipy.stats»

Annexe 3 : Modèle d'arbre de classification à partir des sorties Python

1. Résultats



2. Code Python

```
# Importer les librairies utilisées pour notre étude
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import xlrd
# Importer les données
Data= pd.read_excel(r'C:\Users\ADMIN\Desktop\Stage\Women.xlsx')

# Vérifier la distribution absolue de la variable cible PF
print(Data.PF.value_counts())

Oui      9663
Non     1406
Name: PF, dtype: int64

# La distribution relative
print(Data.PF.value_counts(normalize=True))

Oui      0.872979
Non     0.127021
Name: PF, dtype: float64

# Codage disjonctif des prédictives qualitatives
# Liste des variables qualitatives
lstQuali = [var for var in Data.columns[:-1] if Data[var].dtype == np.object_]
print(lstQuali)

['Milieu', 'Situtation_matrimoniale', 'Niveau_scolarisation', 'Travail',
 'Nbr_fois_mariée']

# Recoder en 0/1 ces variables
dfQualiEncoded = pd.get_dummies(Data[lstQuali])
print(dfQualiEncoded.info())

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11069 entries, 0 to 11068
Data columns (total 15 columns):
 #   Column           Non-Null Count  Dtype  
 ---  -- 
 0   Milieu_Rural      11069 non-null   uint8  
 1   Milieu_Urbain      11069 non-null   uint8  
 2   Situtation_matrimoniale_Divorcé  11069 non-null   uint8  
 3   Situtation_matrimoniale_Mariée  11069 non-null   uint8  
 4   Situtation_matrimoniale_Séparé  11069 non-null   uint8  
 5   Situtation_matrimoniale_Veuve  11069 non-null   uint8  
 6   Niveau_scolarisation_Alphabétisation 11069 non-null   uint8  
 7   Niveau_scolarisation_Collège    11069 non-null   uint8  
 8   Niveau_scolarisation_Lycée    11069 non-null   uint8  
 9   Niveau_scolarisation_Primaire 11069 non-null   uint8  
 10  Niveau_scolarisation_Supérieur 11069 non-null   uint8  
 11  Travail_Non          11069 non-null   uint8  
 12  Travail_Oui          11069 non-null   uint8  
 13  Nbr_fois_mariée_Plus d'une fois 11069 non-null   uint8  
 14  Nbr_fois_mariée_Une fois    11069 non-null   uint8  
dtypes: uint8(15)
```

```

# Constitution du data.frame de travail
# Liste des variables quantitatives
lstQuanti = [var for var in Data.columns[:-1] if Data[var].dtype != np.object_]
print(lstQuanti)

['Age', 'Nbr_enfant_vivant', 'Nbr_fois_consultation']

# Réunir dummies et quantitatives dans le même data frame
DataNew = pd.concat([dfQualiEncoded, Data[lstQuanti]], axis=1)

# Rajouter la variable cible PF
DataNew['PF'] = Data.PF
print(DataNew.info())

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11069 entries, 0 to 11068
Data columns (total 19 columns):
 #   Column           Non-Null Count  Dtype  
---  --  
 0   Milieu_Rural      11069 non-null   uint8  
 1   Milieu_Urbain     11069 non-null   uint8  
 2   Situtation_matrimoniale_Divorcé  11069 non-null   uint8  
 3   Situtation_matrimoniale_Mariée   11069 non-null   uint8  
 4   Situtation_matrimoniale_Séparé   11069 non-null   uint8  
 5   Situtation_matrimoniale_Veuve   11069 non-null   uint8  
 6   Niveau_scolarisation_Alphabétisation 11069 non-null   uint8  
 7   Niveau_scolarisation_Collège    11069 non-null   uint8  
 8   Niveau_scolarisation_Lycée     11069 non-null   uint8  
 9   Niveau_scolarisation_Primaire  11069 non-null   uint8  
 10  Niveau_scolarisation_Supérieur 11069 non-null   uint8  
 11  Travail_Non            11069 non-null   uint8  
 12  Travail_Oui           11069 non-null   uint8  
 13  Nbr_fois_mariée_Plus d'une fois 11069 non-null   uint8  
 14  Nbr_fois_mariée_Une fois       11069 non-null   uint8  
 15  Age                  11069 non-null   float64 
 16  Nbr_enfant_vivant     11069 non-null   float64 
 17  Nbr_fois_consultation 11069 non-null   float64 
 18  PF                  11069 non-null   object  
dtypes: float64(3), object(1), uint8(15)

```

Subdiviser les données en échantillons d'apprentissage et de test

```

from sklearn.model_selection import train_test_split
DataTrain, DataTest = train_test_split(DataNew, test_size=0.3, random_state=1, stratify=DataNew.PF)

```

Vérifier les dimensions

```

print(DataTrain.shape)
print(DataTest.shape)

```

```
(7748, 19)
(3321, 19)
```

Instanciation de l'arbre

```

from sklearn.tree import DecisionTreeClassifier
arbreFirst = DecisionTreeClassifier(max_depth = 3)

```

Construction de l'arbre

```

arbreFirst.fit(X = DataTrain.iloc[:, :-1], y = DataTrain.PF)

```

Affichage graphique de l'arbre

```

from sklearn.tree import plot_tree
plt.figure(figsize = (100,150))
plot_tree(arbreFirst, feature_names = list(arbreFirst.columns[:-1]), filled=True)
plt.show()

```

```
# Evaluation en test
# Prédiction sur l'échantillon test
pred = arbreFirst.predict (X= DataTest.iloc[:, :-1])

# Distribution des prédictions
print(np.unique(pred, return_counts=True))

(array(['Non', 'Oui'], dtype=object), array([153, 3168], dtype=int64))

# Pour savoir que la distribution est juste, on utilise la matrice de confusion
# Matrice de confusion
from sklearn import metrics
print(metrics.confusion_matrix (DataTest.PF, pred))

# Indicateurs de performances
print(metrics.classification_report (DataTest.PF, pred))
```

Source : Université Lyon 2, Tutoriel Tanagra « Arbre de décision avec Scikit-Learn -Python »

Annexe 4 : Modèle de régression logistique à partir des sorties Python

1. Code Python

```
# Importer les librairies utilisées pour notre étude
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import xlr
import statsmodels as sm
import scipy

# Importer les données
Data= pd.read_excel(r'C:\Users\ADMIN\Desktop\Stage\Women15.xlsx')
# Vérifier la structure de la base de données
print(Data.info())

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11069 entries, 0 to 11068
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Milieu            11069 non-null   int64  
 1   Age               11069 non-null   int64  
 2   Niveau_scolarisation  11069 non-null   int64  
 3   Situation_matrimoniale 11069 non-null   int64  
 4   Nbr_fois_marié      11069 non-null   int64  
 5   Travail            11069 non-null   int64  
 6   Nbr_enfant_vie     11069 non-null   float64 
 7   Nbr_fois_consultation 11069 non-null   float64 
 8   PF                11069 non-null   int64  
dtypes: float64(2), int64(7)

# Subdiviser les données en échantillons d'apprentissage et de test
from sklearn.model_selection import train_test_split
DataTrain, DataTest = train_test_split(DataNew, test_size=0.3, random_state=1, stratify=DataNew.PF)

# Vérifier les dimensions
print(DataTrain.shape)
print(DataTest.shape)

(7748, 9)
(3321, 9)

# Isoler la variable cible (Y) et les variables explicatives (X) dans deux structures distinctes
# Y (PF) est la dernière colonne
yTrain = DataTrain.iloc[:, -1]
# X (les autres) sont les variables qui précédent la dernière
XTrain = DataTrain.iloc[:, :-1]

# Comptage des modalités de Y
print(yTrain.value_counts())

0    2899
1    422
Name: PF, dtype: int64

# 2899 femmes utilisant des méthodes contraceptives et 422 femmes n'en utilisant pas
```

Données X avec la constante

```
from statsmodels.tools import add_constant
XTrainBis = sm.tools.add_constant(XTrain)
```

Vérifier la structure de la base de données

```
print(Data.info())
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 7748 entries, 10512 to 9127
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   const            7748 non-null    float64
 1   Milieu           7748 non-null    int64  
 2   Age              7748 non-null    int64  
 3   Niveau_scolarisation  7748 non-null    int64  
 4   Situation_matrimoniale 7748 non-null    int64  
 5   Nbr_fois_marié    7748 non-null    int64  
 6   Travail          7748 non-null    int64  
 7   Nbr_enfant_vie   7748 non-null    float64 
 8   Nbr_fois_consultation 7748 non-null    float64 
dtypes: float64(3), int64(6)
```

Visualiser les premières lignes de la structure

```
print(XTrainBis.head())
```

	const	Milieu	Age	Niveau_scolarisation	Situation_matrimoniale
10512	1.0	2	49	2	1
8725	1.0	2	49	2	1
10326	1.0	2	35	2	1
3694	1.0	1	28	2	1
1102	1.0	1	29	1	1

	Nbr_fois_marié	Travail	Nbr_enfant_vie	Nbr_fois_consultation
10512	1	2	1.294589	4.540419
8725	1	1	1.000000	4.000000
10326	1	2	3.000000	4.540419
3694	1	2	1.294589	4.540419
1102	1	2	1.000000	2.000000

Régression logistique (variable cible Y et les explicatives X)

Y (PF) est la dernière colonne

```
from statsmodels.api import Logit
lr = Logit(endog=yTrain, exog=XTrainBis)
res = lr.fit()
print(res.summary())
```

Logit Regression Results						
Dep. Variable:	PF	No. Observations:	7748			
Model:	Logit	Df Residuals:	7739			
Method:	MLE	Df Model:	8			
Date:	Fri, 27 May 2022	Pseudo R-squ.:	0.04084			
Time:	19:00:06	Log-Likelihood:	-2828.8			
converged:	True	LL-Null:	-2949.2			
Covariance Type:	nonrobust	LLR p-value:	1.452e-47			
	coef	std err	z	P> z	[0.025	0.975]
const	-2.2832	0.371	-6.146	0.000	-3.011	-1.555
Milieu	0.2590	0.071	3.644	0.000	0.120	0.398
Age	-0.0370	0.004	-8.770	0.000	-0.045	-0.029
Niveau_scolarisation	0.0652	0.046	1.423	0.155	-0.025	0.155
Situation_matrimoniale	0.6054	0.048	12.536	0.000	0.511	0.700
Nbr_fois_marié	0.2256	0.146	1.545	0.122	-0.061	0.512
Travail	0.2008	0.113	1.782	0.075	-0.020	0.422
Nbr_enfant_vie	-0.1767	0.089	-1.985	0.047	-0.351	-0.002
Nbr_fois_consultation	-0.0090	0.011	-0.821	0.412	-0.031	0.012

Prédiction

```
predResub = np.where(res.fittedvalues > 0, 1, 0)
print(predResub)
```

Matrice de confusion

```
from statsmodels.tools import add_constant
XTrainBis = sm.tools.add_constant(XTrain)
```

col_0	0	1
PF		
0	6744	20
1	974	10

Pseudo-R²# R²de McFadden

Log-vraisemblance du modèle

```
print("Log-vraisemblance du modèle : %.4f" % (res.llf))
```

Log-vraisemblance du null modèle

```
print("Log-vraisemblance du null modèle : %.4f" % (res.llnull))
```

```
Log-vraisemblance du modèle : -2828.7804
Log-vraisemblance du null modèle : -2949.2357
```

```
print("R2 de McFadden : %.4f" % (res.prssquared))
```

```
R2 de McFadden : 0.0408
```

Scores fournis par la régression

```
Scores = lr.cdf(res.fittedvalues)
```

Data frame temporaire avec y et les scores

```
df = pd.DataFrame ({ "y":yTrain,"score":Scores})
```

5 intervalles de largeurs égales

```
intv = pd.cut(df.score,bins=5,include_lowest=True)
```

Intégrées dans le df

```
df['intv'] = intv
```

```
print(df)
```

y	score	intv
10512	0 0.076860	(0.0718, 0.0823]
8725	0 0.067251	(0.0289999999999998, 0.0718]
10326	0 0.093748	(0.0918, 0.102]
3694	0 0.122677	(0.113, 0.125]
1102	0 0.119767	(0.113, 0.125]
...
3911	0 0.159596	(0.155, 0.187]
3470	1 0.114927	(0.113, 0.125]
6271	0 0.136301	(0.125, 0.139]
2956	0 0.131024	(0.125, 0.139]
9127	0 0.177806	(0.155, 0.187]

```
[7748 rows x 3 columns]
```

Test de Hosmer-Lemeshow

Effectifs par groupe

```
n_tot = df.pivot_table(index="intv",values="y",aggfunc="count").values[:,0]
print (n_tot)
```

```
[748 677 712 715 714 710 695 645 615 533]
```

Somme des scores par groupes

```
s_scores = df.pivot_table (index='intv',values="score",aggfunc="sum").values[:,0]
print(s_scores)
```

```
[52.28644523 57.57380562 67.33303648 74.76212844 83.21098803 93.0409069
102.68149697 111.98601999 132.18480398 208.94036837]
```

```
# Nombre de positifs par groupe
n_pos = df.pivot_table(index="intv",values="y",aggfunc="sum").values[:,0]
print(n_pos)

[66 66 62 56 59 70 84 119 168 234]

# Nombre de négatifs par groupe
n_neg = n_tot - n_pos
print(n_neg)

[748 677 712 715 714 710 695 645 615 533]

C1 = np.sum((n_pos - s_scores)**2/s_scores)
C2 = np.sum((n_neg - (n_tot - s_scores))**2/((n_tot - s_scores)))

# Statistique de Hosmer-Lemeshow laquelle suite une loi du Khi-2 à 8 degrés de liberté
HL = C1 + C2
print("Statistique de Hosmer-Lemeshow : %.4f" % (HL))

Statistique de Hosmer-Lemeshow : 45.4656

#Probabilité critique
p_value = 1.0 - scipy.stats.chi2.cdf(HL,8)
print(p_value)

3.002873114610338e-07

# Au risque 5%, nous ne pouvons pas rejeter l'hypothèse de compatibilité de notre modèle avec les données

# Test de significativité globale de la régression
# Test du rapport de vraisemblance
# Déviance du modèle travail
dev_modele = (-2) * res.llf
print("Déviance du modèle travail : %.4f" % (dev_modele))

Déviance du modèle travail : 5657.5608

# Déviance du modèle nul
dev_null = (-2) * res.llnull
print("Déviance du modèle nul : %.4f" % (dev_null))

Déviance du modèle nul: 5898.4714

# Statistique du rapport de vraisemblance
LR_stat = dev_null - dev_modele
print("Statistique du rapport de vraisemblance : %.4f" % (LR_stat))

Statistique du rapport de vraisemblance : 240.9106

# P_value du test
p_value = 1.0 - scipy.stats.chi2.cdf(res.llr, res.df_model)
print(p_value)

0.0

# Au risque 5%, nous pouvons rejeter l'hypothèse de nullité des coefficients

# Critères AIC et BIC
# AIC du modèle travail
print("AIC du modèle travail : %.4f" % (res.aic))

AIC du modèle travail : 5675.5608

# AIC du modèle nul (1 seul paramètre estimé, la constante)
print("AIC du modèle nul : %.4f" % (aic_null))
```

```

AIC du modèle nul : 5900.4714
# Apparemment, notre modèle est globalement pertinent

# BIC du modèle travail
print("BIC du modèle travail : %.4f" % (res.bic))

BIC du modèle travail : 5738.1575

# BIC du modèle nul
bic_null = (-2) * res.llnull + np.log(n) * (1)
print("BIC du modèle nul : %.4f" % (bic_null))

BIC du modèle nul : 5907.4266

# Même remarque → notre modèle est globalement pertinent

# Tester un groupe de coefficients
# Passer par la vraisemblance
# Définir la matrice des X sans les 4 variables (Niveau_scolarisation, Nbr_fois_marié, Travail,
# Nbr_fois_consultation)
XTrain_4 = XTrainBis.drop(columns = ['Niveau_scolarisation','Travail','Nbr_fois_marié','Nbr_fois_consultation'])
print(XTrain_4.info())

<class 'pandas.core.frame.DataFrame'>
Int64Index: 7748 entries, 10512 to 9127
Data columns (total 5 columns):
 #   Column           Non-Null Count  Dtype  
---  -- 
 0   const            7748 non-null    float64
 1   Milieu           7748 non-null    int64  
 2   Age              7748 non-null    int64  
 3   Situation_matrimoniale 7748 non-null    int64  
 4   Nbr_enfant_vie   7748 non-null    float64
dtypes: float64(2), int64(3)

# Régression sans les 4 variables
lr_4 = Logit(yTrain, XTrain_4)
# Résultats
res_4 = lr_4.fit()

# Statistique du test du rapport de vraisemblance
LR_4 = (-2) * res_4.llf - (-2) * res.llf
print("Statistique du test du rapport de vraisemblance : %.4f" % (LR_4))

Statistique du test du rapport de vraisemblance : 7.5347

# Degré de liberté = 4 puisque 4 coefficients retirés
Ddl = res_4.df_resid - res.df_resid
print("Degré de liberté du test : %.d" % (Ddl))

Degré de liberté du test : 4

# P_value du test
p_value = 1.0 - scipy.stats.chi2.cdf(LR_4, Ddl)
print("Probabilité critique : %.4f" % (p_value))

Probabilité critique : 0.1102

# Au risque 5%, nous pouvons accepter l'hypothèse nulle c'est-à-dire la nullité des coefficients de
# (Niveau_scolarisation, Nbr_fois_marié, Travail, Nbr_fois_consultation)

# Passer par le test de Wald
# Matrice des coefficients à tester
M = [[0,0,0,1,0,0,0,0],[0,0,0,0,0,1,0,0],[0,0,0,0,0,0,1,0],[0,0,0,0,0,0,0,1]]
# Calculer la statistique du test de Wald
stat_4bis = res.wald_test(M)
print("Statistique du test de Wald : %.4f" % (stat_4bis))

```

```

<Wald test (chi2): statistic = [[7.39557694]], p-value = 0.11640304217817773,
df_denom = 4>

# Nbr_fois_consultation) sont nuls

# Prédiction et matrice de confusion
# Préparation de l'échantillon test
# Par adjonction de la constante
XTest_Bis = add_constant(XTest)
print(XTest_Bis.info())

<class 'pandas.core.frame.DataFrame'>
Int64Index: 3321 entries, 10483 to 2016
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   const            3321 non-null    float64
 1   Milieu           3321 non-null    int64  
 2   Age              3321 non-null    int64  
 3   Niveau_scolarisation  3321 non-null    int64  
 4   Situation_matrimoniale 3321 non-null    int64  
 5   Nbr_fois_marié    3321 non-null    int64  
 6   Travail          3321 non-null    int64  
 7   Nbr_enfant_vie   3321 non-null    float64 
 8   Nbr_fois_consultation 3321 non-null    float64 
dtypes: float64(3), int64(6)

# Calcul de la prédiction sur l'échantillon test
predProbaSm = res.predict(XTest_Bis)
# Convertir en prédiction brute
predSm = np.where(predProbaSm > 0.5, 1, 0)
print(np.unique(predSm, return_counts=True))

(array([0, 1]), array([3310, 11], dtype=int64))

# Sur les 3321 individus en test, 3310 sont associés à la classe « 0 : Oui », 11 à « 1 : Non »

# Matrice de confusion
mcSm = pd.crosstab(yTest, predSm)
print(mcSm)

  col_0      0      1
PF
0        2891    8
1        419     3

# Transformer en matrice Numpy
mcSmNumpy = mcSm.values

# Courbe ROC en test
# Importer le module metrics de la librairie scikit-learn
import sklearn.metrics as metrics

# fpr (false positive rate -- taux de faux positifs) en abscisse
# tpr (true positive rate – taux de vrais positifs) en ordonnée
# pos_label = 1 pour indiquer la modalité cible
fprSm, tprSm, _ = metrics.roc_curve(yTest, predProbaSm, pos_label=1)

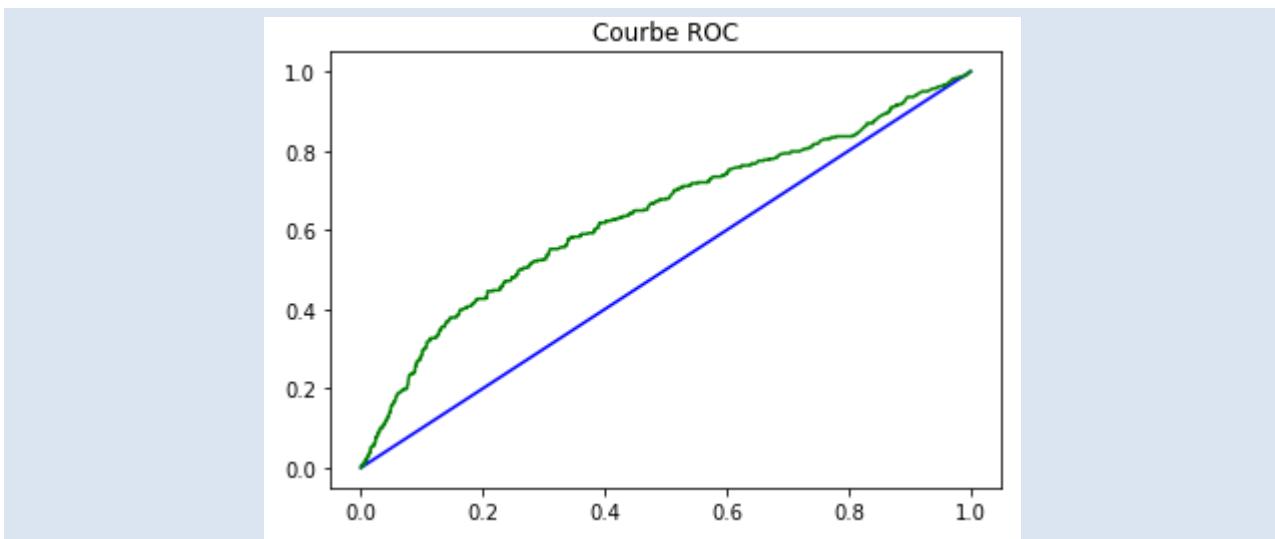
# Graphique
plt.plot(np.arange(0,1,0.1),np.arange(0,1,0.1),'b')

# Rajouter notre diagramme
plt.plot(fprSm, tprSm,"g")

# Titre
plt.title("Courbe ROC")

# Faire apparaître le graphique
plt.show()

```



Valeur de l'AUC

```
aucSm = metrics.roc_auc_score(yTest, predProbaSm)
print("AUC : %.4f" % (aucSm))
```

AUC: 0.6376

Notre modèle final de régression logistique

```
print(res_4.summary())
```

Logit Regression Results						
Dep. Variable:	PF	No. Observations:	7748			
Model:	Logit	Df Residuals:	7743			
Method:	MLE	Df Model:	4			
Date:	Sun, 29 May 2022	Pseudo R-squ.:	0.03957			
Time:	13:29:14	Log-Likelihood:	-2832.5			
converged:	True	LL-Null:	-2949.2			
Covariance Type:	nonrobust	LLR p-value:	2.476e-49			
	coef	std err	z	P> z	[0.025	0.975]
const	-1.6029	0.221	-7.261	0.000	-2.036	-1.170
Milieu	0.2642	0.070	3.758	0.000	0.126	0.402
Age	-0.0362	0.004	-8.798	0.000	-0.044	-0.028
Situation_matrimoniale	0.5861	0.046	12.684	0.000	0.496	0.677
Nbr_enfant_vie	-0.1671	0.089	-1.881	0.060	-0.341	0.007

Source : Université Lyon 2, Tutoriel Tanagra « Regression Logistique -Python »

Annexe 5 : Comparaison entre modèles de régression logistique et d'arbre de classification à partir des sorties Python

```
# Indicateurs de performances (Taux de reconnaissance - Taux d'erreur)
# Arbre de classification
# Taux de reconnaissance
print("Taux de reconnaissance : %.2f" % (metrics.accuracy_score(DataTest.PF, predFirst)))
Taux de reconnaissance : 0.88

# Taux d'erreur
print("Taux d'erreur : %.2f" % (1.0-metrics.accuracy_score(DataTest.PF, predFirst)))

Taux d'erreur : 0.12

# Régression logistique
mcSmNumpy = mcSm.values
# Taux de reconnaissance
accSm = np.sum(np.diagonal(mcSmNumpy))/np.sum(mcSmNumpy)
print("Taux de reconnaissance : %.2f" % (accSm))

Taux de reconnaissance : 0.87

# Taux d'erreur
errSm = 1.0 - accSm
print("Taux d'erreur : %.2f" % (errSm))

Taux d'erreur : 0.13
```

Source : Université Lyon 2, Tutoriels Tanagra « Regression Logistique -Python » & « Arbre de décision avec Scikit-Learn-Python »

