

BURKINA FASO

Unité-Progrès-Justice

MINISTERE DE L'ECONOMIE, DE LA FINANCE ET DE LA PROSPECTIVE

**INSTITUT NATIONAL DE LA STATISTIQUE ET DE LA DEMOGRAPHIQUE
(INSD)**






ISSP-UJKZ



**POJET D'ECONOMETRIE DES VARIABLES
QUALITATIVES**

RAPPORT DE PROJET

Groupe 6 :

 **BATIONO Martinien**
 **CISSE Oumarou**
 **NANA Hervé**

Enseignant :

Dr. Fabrice Boyam YAMEOGO

Année académique 2023-2024

Table des matières

| | |
|---|----|
| Introduction..... | 2 |
| Chargement des packages importants pour l'analyse..... | 2 |
| Chargement de la base et traitement..... | 2 |
| Chargement de la base..... | 2 |
| Traitement de la base..... | 3 |
| Création de la variable hhid..... | 3 |
| Recodage de la variable dépendante..... | 3 |
| Recodage des variables qualitatives..... | 5 |
| Modélisation..... | 9 |
| Les variables prises en compte dans la modélisation | 9 |
| Modélisation et interprétation | 10 |
| Confort élevé..... | 11 |
| Confort faible..... | 12 |
| Confort moyen..... | 12 |
| Interprétation..... | 13 |
| PseudoR2 de McFadden et Nagelkerke | 14 |
| Matrice de confusion | 14 |
| Significativité globale des variables explicatives..... | 15 |
| Calcul du taux d'erreur | 15 |
| Conclusion | 16 |

Introduction

Dans un contexte de développement urbain dynamique, la Côte d'Ivoire fait face à des défis importants en matière de logement. Afin de répondre efficacement aux besoins des ménages, il est crucial d'adopter une approche fondée sur des données concrètes pour déterminer les priorités en matière de logement. L'objectif de notre étude est de cibler, à partir de variables explicatives pertinentes, le type de logement à proposer en priorité aux ménages ivoiriens.

Chargement des packages importants pour l'analyse

Ces différents packages sont importants et requis dans l'analyse.

```
library(nnet)
library(haven)
library(dplyr)

library(questionr)
library(car)

library(gtsummary)
library(caret)

library(DescTools)
```

Chargement de la base et traitement

Chargement de la base

La base contient les données de l'Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM) réalisée en 2018 en Côte d'Ivoire. Elle est constituée de 12992 observations réparties en 106 variables.

```
base_EHCVM <-
read_dta("C:/Users/cisse/Desktop/Projet_Var_Quali_2024_LPAS2/s11_me_CIV2018.d
ta")

names(base_EHCVM)

## [1] "vague" "grappe" "menage" "s11q00"
## [5] "s11q01" "s11q02" "s11q03__1" "s11q03__2"
## [9] "s11q03__3" "s11q04" "s11q05" "s11q06"
## [13] "s11q07" "s11q08" "s11q09a" "s11q09b__0"

summary(base_EHCVM)
```

Dans la base, il y a des variables qualitatives et quantitatives. Parmi ces variables, d'autres n'ont pas d'observations (NA) et n'interviennent pas dans l'analyse. Nous allons donc traiter la base.

Traitement de la base

Création de la variable hhid

Dans la base il n'y a pas d'identifiant unique. Nous allons donc en créer en se basant sur certaines variables déjà présentes dans la base notamment grappe et menage.

```
## Création de la variables hhid
base_EHCVM$hhid= base_EHCVM$grappe*100 + base_EHCVM$menage

## Vérification des doublons et Le nombre d'observations
anyDuplicated(base_EHCVM$hhid)

## [1] 0

nrow(base_EHCVM)

## [1] 12992
```

Recodage de la variable dépendante

Nous allons maintenant recoder la variable dépendante s11q01 dans une nouvelle variable nommée Typl_logement qui sera utilisée pour la regression. Comme critère de regroupement, nous avons utilisé le confort. En effet, les individus pour le choix des habitations recherchent plus le confort. Pour l'étude, il y a trois catégories:

- Confort élevé (villa et appartement) :
 - Villa : elles sont généralement des habitations spacieuses, bien construites avec des matériaux de qualité supérieure, et souvent équipées d'aménagements modernes (jardin, piscine, etc.). Elles offrent un niveau de confort élevé.
 - Appartement : Les appartements, en particulier ceux situés dans des immeubles modernes, sont souvent bien conçus, avec des équipements, la sécurité, etc., ce qui les classe également dans le confort élevé.
- Confort moyen (cour commune, maison isolée, bande par particulier et bande par société immobilière) :
 - Cour commune : Les maisons en cour commune sont généralement de taille moyenne et partagent des espaces extérieurs avec d'autres habitations. Elles offrent un confort modéré mais peuvent manquer de certains équipements privés.

Econométrie des variables qualitatives

- Maison isolée : Bien que séparée des autres habitations, la maison isolée peut varier en termes de taille et d'équipements, la plaçant souvent dans une catégorie de confort moyen.
- Bande par particulier et bande par société immobilière : Ces types de logement sont des constructions souvent standardisées, offrant un confort correct, mais sans les luxes associés aux villas ou appartements.
- Confort faible (case, banco, baraque) :
 - Case : Les cases sont des habitations traditionnelles souvent construites avec des matériaux locaux comme la paille, offrant un confort basique.
 - Banco : Le banco est un matériau traditionnel (terre crue), utilisé pour construire des maisons qui, bien que fonctionnelles, manquent souvent d'équipements modernes, d'où leur classification dans le confort faible.
 - Baraque : Les baraques sont souvent des constructions précaires, parfois en tôle ou en bois, offrant un confort très limité.
- Autre type (autre) :
 - Autre : Cette catégorie regroupe les types de logements qui ne rentrent pas dans les classifications ci-dessus, souvent en raison de leur rareté ou de caractéristiques spécifiques qui ne permettent pas de les placer dans les autres catégories.

Transformation en caractères

```
base_EHCVM$s11q01 = as.character(base_EHCVM$s11q01)
```

Recodage de base\$s11q01 en base\$typl_logement

```
base_EHCVM$typl_logement <- as.character(base_EHCVM$s11q01)
base_EHCVM$typl_logement[base_EHCVM$s11q01 == "1"] <- "Confort élevé"
base_EHCVM$typl_logement[base_EHCVM$s11q01 == "2"] <- "Confort élevé"
base_EHCVM$typl_logement[base_EHCVM$s11q01 == "3"] <- "Confort moyen"
base_EHCVM$typl_logement[base_EHCVM$s11q01 == "4"] <- "Confort moyen"
base_EHCVM$typl_logement[base_EHCVM$s11q01 == "5"] <- "Confort moyen"
base_EHCVM$typl_logement[base_EHCVM$s11q01 == "6"] <- "Confort moyen"
base_EHCVM$typl_logement[base_EHCVM$s11q01 == "7"] <- "Confort faible"
base_EHCVM$typl_logement[base_EHCVM$s11q01 == "8"] <- "Confort faible"
base_EHCVM$typl_logement[base_EHCVM$s11q01 == "9"] <- "Autre type"
table(base_EHCVM$typl_logement)
```

Vu la grandeur de la base, nous allons sélectionner certaines variables pertinentes pour l'étude et les mettre dans une nouvelle base qui servira pour la suite.

Sélection des variables spécifiques

```
base <- base_EHCVM %>%
  select(hhid,typl_logement,s11q02, s11q04, s11q05, s11q06,s11q07,s11q08,
s11q13, s11q15,s11q18, s11q19, s11q20, s11q21, s11q22, s11q23, s11q27a,
s11q27b, s11q28,s11q34, s11q35, s11q36, s11q38, s11q42, s11q43, s11q46,
```

```
s11q50,  
s11q53__1,s11q53__2,s11q53__3,s11q53__4,s11q53__5,s11q53__6,s11q53__7,s11q53__8,  
s11q54, s11q55, s11q56, s11q60)
```

```
# Affichage des premières lignes pour vérifier la sélection  
head(base)
```

Recodage des variables qualitatives

Nous allons recoder les variables qualitatives et regrouper certaines de leurs modalités pour une bonne présentation et ainsi faciliter l'analyse.

```
base$s11q04=as.character(base$s11q04)  
  
## Recodage de base$s11q04 en base$statut_occupation  
base$statut_occupation <- as.character(base$s11q04)  
base$statut_occupation[base$s11q04 == "1"] <- "Propriétaire"  
base$statut_occupation[base$s11q04 == "2"] <- "Propriétaire"  
base$statut_occupation[base$s11q04 == "3"] <- "Propriétaire"  
base$statut_occupation[base$s11q04 == "4"] <- "Propriétaire"  
base$statut_occupation[base$s11q04 == "5"] <- "Locataire"  
base$statut_occupation[base$s11q04 == "6"] <- "Logé gratuitement"  
base$statut_occupation[base$s11q04 == "7"] <- "Logé gratuitement"  
base$statut_occupation[base$s11q04 == "8"] <- "Autre"  
base$statut_occupation = as.factor(base$statut_occupation)  
  
base$s11q19 = as.character(base$s11q19)  
  
## Recoding base$s11q19 into base$murs_exterieurs  
base$murs_exterieurs <- as.character(base$s11q19)  
base$murs_exterieurs[base$s11q19 == "1"] <- "Moderne"  
base$murs_exterieurs[base$s11q19 == "2"] <- "Traditionnelles"  
base$murs_exterieurs[base$s11q19 == "3"] <- "Moderne"  
base$murs_exterieurs[base$s11q19 == "4"] <- "Banco amélioré/ semi-dur"  
base$murs_exterieurs[base$s11q19 == "5"] <- "Planches/toles"  
base$murs_exterieurs[base$s11q19 == "6"] <- "Traditionnelles"  
base$murs_exterieurs[base$s11q19 == "7"] <- "Traditionnelles"  
base$murs_exterieurs[base$s11q19 == "8"] <- "Autre"  
base$murs_exterieurs = as.factor(base$murs_exterieurs)  
  
base$s11q20 = as.character(base$s11q20)  
  
## Recodage de base$s11q20 en materiau_toit  
base$materiau_toit <- as.character(base$s11q20)  
base$materiau_toit[base$s11q20 == "1"] <- "Dalle en ciment"  
base$materiau_toit[base$s11q20 == "2"] <- "Couvert en terre"  
base$materiau_toit[base$s11q20 == "3"] <- "Tôles"  
base$materiau_toit[base$s11q20 == "4"] <- "Traditionnel"  
base$materiau_toit[base$s11q20 == "5"] <- "Traditionnel"  
base$materiau_toit[base$s11q20 == "6"] <- "Traditionnel"
```

```

base$materiau_toit[base$s11q20 == "7"] <- "Traditionnel"
base$materiau_toit[base$s11q20 == "8"] <- "Autres"

base$materiau_toit=as.factor(base$materiau_toit)

base$s11q21 =as.character(base$s11q21)

## Recodage de base$s11q21 en base$revetement_sol
base$revetement_sol <- as.character(base$s11q21)
base$revetement_sol[base$s11q21 == "1"] <- "Carreaux/Marbre"
base$revetement_sol[base$s11q21 == "2"] <- "Ciment/Béton"
base$revetement_sol[base$s11q21 == "3"] <- "Terre battue"
base$revetement_sol[base$s11q21 == "4"] <- "Terre battue"
base$revetement_sol[base$s11q21 == "5"] <- "Autre"
base$revetement_sol=as.factor(base$revetement_sol)

base$s11q22 = as.character(base$s11q22)

## Recodage de BASE_QUALI$s11q22 en BASE_QUALI$s11q22_rec
base$eau_courant <- as.character(base$s11q22)
base$eau_courant[base$s11q22 == "1"] <- "Oui"
base$eau_courant[base$s11q22 == "2"] <- "Non"
table(base$eau_courant)

base$eau_courant=as.factor(base$eau_courant)

base$s11q34<- as.character(base$s11q34)
## Recodage de Base$s11q34 en Base$s11q34_rec
base$connect_electricite <- as.character(base$s11q34)
base$connect_electricite[base$s11q34 == "1"] <- "Oui"
base$connect_electricite[base$s11q34 == "2"] <- "Oui"
base$connect_electricite[base$s11q34 == "3"] <- "Oui"
base$connect_electricite[base$s11q34 == "4"] <- "Non"
base$connect_electricite = as.factor(base$connect_electricite)
table(base$connect_electricite)

base$s11q35<-as.character(base$s11q35)

## Recodage de s11q35
base$electricite_loyer <- as.character(base$s11q35)
base$electricite_loyer[base$s11q35 == "1"] <- "Oui"
base$electricite_loyer[base$s11q35 == "2"] <- "Non"
base$electricite_loyer=as.factor(base$electricite_loyer)

base$s11q36<-as.character(base$s11q36)

## Recodage de s11q36
base$type_abonnement <- base$s11q36
base$type_abonnement[base$s11q36 == "1"] <- "Compteur classique"
base$type_abonnement[base$s11q36 == "2"] <- "Compteur avec carte prépayée"
base$type_abonnement[base$s11q36 == "3"] <- "Les deux"

```

```

base$type_abonnement[base$s11q36 == "4"] <- "Additionneuse"
base$type_abonnement[base$s11q36 == "5"] <- "Sans Compteur"
table(base$type_abonnement)

base$type_abonnement = as.factor(base$type_abonnement)

base$s11q38<-as.character(base$s11q38)

## Recodage de s11q38
base$source_eclairage <- base$s11q38
base$source_eclairage[base$s11q38 == "1"] <- "Electricité réseau"
base$source_eclairage[base$s11q38 == "2"] <- "Electricité (générateur)"
base$source_eclairage[base$s11q38 == "3"] <- "Lampe à pétrole"
base$source_eclairage[base$s11q38 == "4"] <- "Lampe à pile, grosse torche"
base$source_eclairage[base$s11q38 == "5"] <- "Paraffine"
base$source_eclairage[base$s11q38 == "6"] <- "Plaque solaire"
base$source_eclairage[base$s11q38 == "7"] <- "Autre"

base$source_eclairage=as.factor(base$source_eclairage)

base$s11q42<-as.character(base$s11q42)

## Recodage de s11q42
base$duree_moyenne_coupure <- base$s11q42
base$duree_moyenne_coupure[base$s11q42 == "1"] <- "Moins de 15 minutes"
base$duree_moyenne_coupure[base$s11q42 == "2"] <- "De 15 minutes à moins de
30 minutes"
base$duree_moyenne_coupure[base$s11q42 == "3"] <- "De 30 minutes à moins
d'une heure"
base$duree_moyenne_coupure[base$s11q42 == "4"] <- "D'une heure à moins de 3
heures"
base$duree_moyenne_coupure[base$s11q42 == "5"] <- "3 heures et plus"

base$source_eclairage=as.factor(base$source_eclairage)

base$s11q50<-as.character(base$s11q50)

## Recodage de s11q50
base$chaîne_tele <- base$s11q50
base$chaîne_tele[base$s11q50 == "1"] <- "Oui"
base$chaîne_tele[base$s11q50 == "2"] <- "Non"
table(base$chaîne_tele)

base$chaîne_tele=as.factor(base$chaîne_tele)

base$s11q54<-as.character(base$s11q54)
## Recodage de base$s11q54 en base$s11q54_rec
base$mode_ordure <- base$s11q54
base$mode_ordure[base$s11q54 == "1"] <- "Dépotoir public"
base$mode_ordure[base$s11q54 == "2"] <- "Dépotoir public"
base$mode_ordure[base$s11q54 == "3"] <- "Brûlées"
base$mode_ordure[base$s11q54 == "4"] <- "Dépotoir sauvage"

```



```

base$mode_oriture[base$s11q54 == "5"] <- "Dépotoir sauvage"
base$mode_oriture[base$s11q54 == "6"]<- "Autre"
table(base$mode_oriture)

base$mode_oriture=as.factor(base$mode_oriture)

base$s11q55<- as.character(base$s11q55)
## Recodage de Base$s11q55 en Base$s11q55_rec
base$type_sanitaire <- as.character(base$s11q55)
base$type_sanitaire[base$s11q55 == "1"] <- "W.C. intérieur"
base$type_sanitaire[base$s11q55 == "2"] <- "W.C. extérieur"
base$type_sanitaire[base$s11q55 == "3"] <- "W.C. intérieur"
base$type_sanitaire[base$s11q55 == "4"] <- "W.C. extérieur"
base$type_sanitaire[base$s11q55 == "5"] <- "Latrines VIP"
base$type_sanitaire[base$s11q55 == "6"] <- "Latrines dallées"
base$type_sanitaire[base$s11q55 == "7"] <- "Latrines dallées"
base$type_sanitaire[base$s11q55 == "8"] <- "Latrines dallées"
base$type_sanitaire[base$s11q55 == "9"] <- "Latrines dallées"
base$type_sanitaire[base$s11q55 == "10"] <- "W.C. extérieur"
base$type_sanitaire[base$s11q55 == "11"] <- "Aucune toilette"
base$type_sanitaire[base$s11q55 == "12"] <- "Autre"
table(base$type_sanitaire)

base$type_sanitaire=as.factor(base$type_sanitaire)

base$s11q56 = as.character(base$s11q56)

## Recodage de s11q56
base$sanitaire_commun <- as.character(base$s11q56)
base$sanitaire_commun[base$s11q56 == "1"] <- "Oui"
base$sanitaire_commun[base$s11q56 == "2"] <- "Non"
table(base$sanitaire_commun)

base$sanitaire_commun=as.factor(base$sanitaire_commun)

base$s11q60<-as.character(base$s11q60)

## Recodage de s11q60
base$evacuation_eaux_usees <- base$s11q60
base$evacuation_eaux_usees[base$s11q60 == "1"] <- "Puisard (fosse moderne)"
base$evacuation_eaux_usees[base$s11q60 == "2"] <- "Egout"
base$evacuation_eaux_usees[base$s11q60 == "3"] <- "Trou dans la parcelle"
base$evacuation_eaux_usees[base$s11q60 == "4"] <- "Dans la rue/nature"
base$evacuation_eaux_usees[base$s11q60 == "5"] <- "Autre"
table(base$evacuation_eaux_usees)

base$evacuation_eaux_usees=as.factor(base$evacuation_eaux_usees)

```

Nous allons renommer les variables quantitatives de sorte qu'elles soient parlantes et changer leur type en numeric.

```
# Renommage des variables dans votre base de données
colnames(base)[which(names(base) == "s11q02")] <- "nombre_pieces"
colnames(base)[which(names(base) == "s11q05")] <- "montant_mensuel_loyer"
colnames(base)[which(names(base) == "s11q15")] <- "montant_logement"

# Changement du type
base$montant_mensuel_loyer=as.numeric(base$montant_mensuel_loyer)
base$montant_logement=as.numeric(base$montant_logement)

## Transformation de la variable en facteur
base$typ1_logement = as.factor(base$typ1_logement)

## Prendre Confort faible comme référence
base$typ1_logement = relevel(base$typ1_logement, ref = "Confort faible")
```

Modélisation

Pour la modélisation, nous avons utilisé la fonction 'multinom' du package 'nnet'. Toutes les variables catégorielles ont été transformées en facteur.

Les variables prises en compte dans la modélisation

Les variables explicatives pour la régression logistique multinomiale sur la variable dépendante **type de logement** (s11q01) sont :

- **Nombre de pièces (nombre_pieces)** : Cette variable est cruciale car elle reflète directement la taille et le confort du logement. Les ménages recherchent souvent des logements en fonction du nombre de pièces pour répondre à leurs besoins d'espace.
- **Eau courante (eau_courant)** : La présence d'eau courante est un indicateur clé de la qualité de vie dans un logement. Les ménages préfèrent des logements qui offrent ce service de base, ce qui influence leur choix.
- **Connexion à l'électricité (connect_electricite)** : L'accès à l'électricité est essentiel pour les ménages, car il affecte non seulement le confort mais aussi la sécurité du logement.
- **Mode d'évacuation des ordures (mode_ordure)** : Cette variable est importante pour évaluer la gestion des déchets, ce qui peut être un facteur décisif dans la perception de la qualité de l'environnement de vie.
- **Type de sanitaire (type_sanitaire)** : Le type de sanitaires disponibles est un indicateur important du niveau de confort et d'hygiène dans un logement, influençant fortement les préférences des ménages.
- **Évacuation des eaux usées (evacuation_eaux_usees)** : Cette variable reflète la gestion des eaux usées dans le logement, un aspect crucial pour le confort et la santé des occupants.

- **Matériau de toit (matériau_toit)** : Le matériau utilisé pour la toiture peut indiquer la durabilité et la qualité du logement, des facteurs importants pour les ménages lors du choix de leur habitation.
- **Statut d'occupation (statut_occupation)** : Le statut d'occupation (propriétaire, locataire, logé gratuitement, etc.) peut influencer la capacité d'un ménage à opter pour un type de logement spécifique, et cette variable est donc essentielle pour comprendre les préférences des ménages.

Ces variables ont été choisies parce qu'elles capturent les aspects clés du confort, de la commodité, et de la qualité de vie associés aux différents types de logements en milieu urbain. En tenant compte de ces facteurs, l'analyse peut aider à identifier les types de logements les plus adaptés aux besoins des ménages, permettant à l'entreprise immobilière de mieux cibler son offre.

Modélisation et interprétation

Modélisation

```
modele = multinom(formula = typ1_logement ~ nombre_pieces + eau_courant +  
connect_electricite + mode_ordure + type_sanitaire + evacuation_eaux_usees +  
matériau_toit + statut_occupation ,maxit= 400, data=base, model= TRUE)
```

```
## # weights: 96 (69 variable)  
## initial value 18010.736340  
## iter 10 value 9376.719865  
## iter 90 value 8216.720273  
## final value 8216.720133  
## converged
```

Choix du meilleur modèle

```
step(modele)
```

```
##  
## Residual Deviance: 16433.44  
## AIC: 16571.44
```

```
summary(modele)
```

```
## Call:  
## multinom(formula = typ1_logement ~ nombre_pieces + eau_courant +  
## connect_electricite + mode_ordure + type_sanitaire +  
## evacuation_eaux_usees +  
## matériau_toit + statut_occupation, data = base, model = TRUE,  
## maxit = 400)  
##  
## Coefficients:  
## (Intercept) nombre_pieces eau_courantOui  
connect_electriciteOui  
## Autre type -207.661765 -16.3665707 0.2783091  
0.2729751  
## mode_ordureBrûlées mode_ordureDépotoir public
```

```
## Autre type          57.32441          67.4361621
## Confort élevé       1.43479          1.3004049
## Confort moyen       0.44492          0.6222835
##          statut_occupationPropriétaire
## Autre type          1.6429849
## Confort élevé       0.5094706
## Confort moyen       0.3587276
##
## Residual Deviance: 16433.44
## AIC: 16571.44

#ajout des p-values globales
modele %>%
  tbl_regression(exponentiate = TRUE) %>%
  add_global_p(type="II")
```

Confort élevé

- Le nombre de pièces occupées par le ménage est fortement lié à un confort élevé. Un OR aussi élevé suggère que les ménages qui préfèrent plus de pièces ont une probabilité extrêmement plus grande d'avoir un confort élevé, comparé à ceux qui en occupent moins.

- Les ménages ayant de l'eau courante sont 1,65 fois plus susceptibles d'avoir un confort élevé, mais l'intervalle de confiance (0.61, 4.45) traverse 1, ce qui signifie que cette association n'est pas statistiquement significative. En effet, les ménages optant pour un confort élevé souhaitent avoir de l'eau courante.

- Les ménages connectés à l'électricité sont 2,45 fois plus susceptibles d'avoir un confort élevé, mais l'intervalle de confiance (0.85, 7.08) traverse également 1, ce qui réduit la certitude de cette liaison. En effet, les ménages optant pour des logements à confort élevé utilisent plus l'électricité.

- Brûler les ordures est fortement lié à un confort élevé par rapport aux autres modes de gestion des déchets (OR = 65.9), et cela est statistiquement significatif. Les ménages souhaitant un confort élevé préfèrent brûler les ordures. L'utilisation de dépotoirs publics ou sauvages est liée à un confort extrêmement faible (OR = 0.00). - Aucun type de sanitaire n'est significativement lié à un confort élevé, bien que les WC intérieurs aient un OR de 1.63.

- L'évacuation des eaux dans la rue/nature (OR = 3.36) et dans un trou dans la parcelle (OR = 3.15) sont associées à un confort élevé, bien que ces associations ne soient pas statistiquement significatives. Ces ménages (souhaitant avoir des logements à confort élevé) préfèrent évacuer les eaux dans la nature ou dans un trou dans la parcelle. - Les toits traditionnels sont également liés à un confort élevé, avec un OR de 221, ce qui est très significatif. En effet, les ménages ayant des logements à confort élevé optent pour les toits traditionnels pour la couverture de leurs maisons.

- Être propriétaire est lié à un confort élevé, mais l'incertitude est grande (intervalle de confiance large). Le fait d'être propriétaire te permet de vivre dans un confort élevé. En effet, les propriétaires des logements préfèrent que l'entreprise leur donne des logements qui ont un confort élevé.

Confort faible

- Comme pour le confort élevé, il y a une forte association entre le nombre de pièces et un confort faible, avec une incertitude importante. Le choix du nombre de pièces du logement est fortement lié à un faible confort.

- Les ménages ayant de l'eau courante et de l'électricité sont moins susceptibles d'avoir un confort faible ($OR=0.76$ chacun). En effet, les ménages ayant des logements à faible confort préfèrent moins l'eau courante et l'électricité par rapport à ceux ayant des confort élevés. - La brûlure des ordures est aussi associée à un confort faible, bien que moins fortement que pour le confort élevé. Les ménages à faible confort brûlent trop les ordures par rapport aux autres.

- Les WC intérieurs ($OR = 0.05$) sont fortement associés à un confort faible, indiquant que les ménages avec des WC intérieurs sont beaucoup moins susceptibles d'avoir un confort faible. En effet, les ménages qui optent pour des logements à confort faible préfèrent que l'entreprise leur mette dans des logements ayant des WC intérieurs.

- Les toits traditionnels sont encore plus fortement associés à un confort faible que pour un confort élevé. Les ménages à faible confort préfèrent les toits traditionnels. - Les propriétaires préfèrent un confort faible qu'un confort élevé ou moyen. En outre, ils souhaitent que l'entreprise leur donne des logements à confort faible.

Confort moyen

- Il y a également une forte corrélation entre le nombre de pièces et un confort moyen, avec une grande incertitude parce que l'intervalle de confiance est très large. En effet, l'option pour un confort moyen est fonction du nombre de pièces du logement.

- Les ménages ayant de l'eau courante et de l'électricité sont légèrement plus susceptibles d'avoir un confort moyen. En effet, les ménages préférant des logements à confort moyen souhaitent moins que cette entreprise les oriente dans des logements où il y a l'eau courante et d'électricité.

- Comme pour le confort élevé, brûler les ordures est fortement associé à un confort moyen. Ces ménages préfèrent brûler leurs ordures que les autres modes de gestion des ordures. En effet, ces ménages préfèrent des logements où l'évacuation des ordures repose plus dans la brûlure.

- Les WC intérieurs ne sont pas significativement associés à un confort moyen. En effet, les ménages préférant un confort moyen ne souhaitent pas nécessairement des logements où il y a des WC intérieurs. - Les toits traditionnels sont fortement associés à un confort moyen. Ces ménages optent également pour les toits traditionnels. Opter pour un confort moyen c'est également opter pour des logements à toits traditionnels.

- Les propriétaires ont une corrélation positive avec le confort moyen. En effet, les propriétaires préfèrent plus que cette entreprise les oriente dans des logements à confort moyen. Tous ces résultats montrent des associations (corrélations) très fortes entre certaines caractéristiques du logement (comme le nombre de pièces, le matériau du toit, et

le statut de propriétaire) et les niveaux de confort. Cependant, on enregistre des intervalles de confiance larges et des OR très élevés.

Interprétation

Afin d'évaluer la qualité d'ajustement de notre modèle de régression logistique, nous avons utilisé deux mesures de pseudo- R^2 , à savoir le pseudo- R^2 de McFadden et celui de Nagelkerke.

Calcul du pseudoR2

```
PseudoR2(modele, c("McFadden", "Nagel"))
```

```
##    McFadden Nagelkerke
```

```
## 0.2445971 0.4135695
```

Prédiction

```
predites <- predict(modele, newdata = base)
```

```
predites
```

```
head(predites)
```

```
## [1] Confort moyen  Confort faible Confort moyen  Confort moyen  Confort  
moyen
```

```
## [6] Confort moyen
```

```
## Levels: Confort faible Autre type Confort élevé Confort moyen
```

Matrice de confusion

```
confusion_matrix <- table(predites, Actual = base$typ1_logement)
```

Affichage de la matrice de confusion

```
print(confusion_matrix)
```

```
##              Actual
## predites      Confort faible Autre type Confort élevé Confort moyen
## Confort faible      895          0          21          608
## Autre type          0          0          0           0
## Confort élevé        3          2         512          262
## Confort moyen     1302         28        1207          8152
```

Test de la significativité globale des variables

```
Anova(modele, test = "Chisq")
```

```
## Analysis of Deviance Table (Type II tests)
```

```
##
```

```
## Response: typ1_logement
```

```
##              LR Chisq Df Pr(>Chisq)
## nombre_pieces    915.91  3 < 2.2e-16 ***
## eau_courant       36.07  3 7.235e-08 ***
## connect_electricite 185.32  3 < 2.2e-16 ***
## mode_ordure       36.15  9 3.725e-05 ***
```

```
## type_sanitaire      408.43 15 < 2.2e-16 ***
## evacuation_eaux_usees 65.27 12 2.433e-09 ***
## materiau_toit       989.61 12 < 2.2e-16 ***
## statut_occupation   119.26 9  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

PseudoR2 de McFadden et Nagelkerke

Le pseudo- R^2 de McFadden est de 0,2446, ce qui signifie que notre modèle explique environ 24 % des variations observées. Dans le cadre de la régression logistique, il s'agit d'un ajustement modéré et typiquement acceptable. Ce qui indique que notre modèle capture une bonne partie des informations.

D'autre part, le pseudo- R^2 de Nagelkerke, avec une valeur de 0,4136, suggère que notre modèle capture environ 41,36 % de la variation des données. Cette mesure, plus intuitive et ajustée pour varier entre 0 et 1, offre une indication supplémentaire de l'ajustement du modèle. Ainsi, elle renforce l'idée que le modèle fournit une explication significative des variations observées dans les données.

Même si notre modèle n'explique pas tout, il réussit tout de même à capturer une part significative des variations, ce qui en fait un outil utile pour faire des prédictions. Les prédictions du modèle montrent des résultats principalement classés comme "Confort moyen", avec quelques occurrences de "Confort faible" et "Confort élevé". Ce qui signifie que, selon les caractéristiques des données fournies dans notre base, la majorité des logements sont estimés avoir un niveau de confort moyen.

Matrice de confusion

La matrice de confusion nous aide à comprendre comment notre modèle de classification a performé en comparant les prédictions faites avec les véritables catégories des logements. Notre modèle a une tendance marquée à prédire les logements comme étant de "Confort moyen", ce qui représente la classe la plus fréquemment prédite.

De façon plus spécifique la catégorie "Confort moyen", elle a été la plus correctement classifiée, avec 8152 bonnes prédictions. Toutefois, des erreurs ont été commises, avec 28 classifications erronées en "Autre type", 1207 en "Confort élevé", et 1302 en "Confort faible". Bien que la performance générale pour cette catégorie soit satisfaisante.

Pour la catégorie "Confort élevé", le modèle a réussi à classer correctement 512 logements comme tels. Cependant, il a commis des erreurs en les classant comme "Confort moyen" dans 262 cas.

Concernant les logements de "Confort faible", bien que le modèle ait correctement identifié 895 d'entre eux, il y a eu 608 erreurs où ces logements ont été classés à tort comme "Confort moyen".

En examinant les résultats plus en détail, nous constatons que la catégorie "Autre type" n'a été prédite que pour une très petite fraction des cas, et ce uniquement en tant que "Confort faible" ou "Confort moyen".

Significativité globale des variables explicatives

L'ANOVA que nous avons réalisée pour le modèle de classification des types de logements à proposer en priorité aux ménages fourni les résultats suivants:

La variable "nombre_pieces" présente une valeur de chi-carré de 915.90 avec un degré de liberté de 3 et un p-value inférieur à $2.2e-16$. Le nombre de pièces dans un logement a une influence très significative sur le type de logement souhaité, avec un Chi-squared élevé (915.90). Cette variable nous indique que les ménages sont fortement influencés par la taille de leur logement lors de leur choix.

La variable Type sanitaire avec une valeur de chi-carré très élevée de 408.44 et un p-value inférieur à $2.2e-16$ traduit que le type de sanitaires est significativement important pour les ménages. En effet les différences dans les installations sanitaires influencent fortement le choix du type de logement.

la variable Eau courante a une valeur de chi-carré de 36.08 avec un p-value de $7.210e-08$. La disponibilité d'eau courante est également cruciale. Les ménages semblent préférer les logements qui offrent ce confort essentiel.

L'Électricité avec une valeur de chi-carré de 185.33 et un p-value inférieur à $2.2e-16$ montre que la présence d'électricité dans le logement a un effet très marqué. Les logements avec électricité sont clairement favorisés.

En résumé, nos résultats montrent que plusieurs variables explicatives, telles que le nombre de pièces, la connexion à l'électricité, l'accès à l'eau courante, et le type sanitaire, ont un impact significatif sur le type de logement à recommander. Ces variables doivent être prises en compte pour cibler efficacement les logements les plus adaptés aux besoins des ménages.

Calcul du taux d'erreur

```
# Calculer le nombre total d'observations
total_observations <- sum(confusion_matrix)

# Calculer le nombre de mauvaises classifications
incorrect_predictions <- total_observations - sum(diag(confusion_matrix))

# Calcul du taux d'erreur
error_rate <- incorrect_predictions / total_observations

# Afficher le taux d'erreur
print(error_rate)

## [1] 0.2642395

# Calcul de la précision du modèle
accuracy <- mean(predites == base$typ1_logement)
print(accuracy)

## [1] 0.7357605
```


Le modèle présente un taux d'erreur de 26,4 %, ce qui signifie qu'environ un quart des prédictions sont erronées. Sa précision est de 73,6 %, indiquant qu'il réussit à classer correctement près de trois quarts des ménages selon leur type de logement.

Conclusion

Cette étude a permis de cerner de manière précise les types de logements à privilégier en Côte d'Ivoire en fonction des caractéristiques spécifiques des ménages. Nos résultats montrent que, pour répondre efficacement aux besoins des ménages, il est essentiel de se concentrer sur des aspects tels que le nombre de pièces, l'accès à l'eau courante, et les matériaux de construction des logements. La précision de notre modèle, avec un taux d'erreur d'environ 26.4 % et une précision de 73.6 %, souligne l'efficacité de notre approche pour guider les choix de logement en fonction des données disponibles.