Université Joseph Ki-Zerbo
Institut Supérieur des Sciences de la Population
Licence Professionnelle en Analyse Statistique - 2e année

Économétrie des variables qualitatives

Analyse des déterminants du type de logement des ménages urbains au Burkina Faso

Réalisé par :
NIAMPA Abdoul Fataho
SAWADOGO Pengdwendé Orianne-Aurele
YAMEOGO Saïdou

Enseignant :
Dr. Boyam Fabrice YAMEOGO

Juin 2025

Table des matières

1 Introduction

Dans un contexte de croissance démographique et d'urbanisation rapide au Burkina Faso, la question du logement en milieu urbain est devenue un enjeu stratégique tant pour les autorités publiques que pour les acteurs du secteur privé. Les ménages urbains vivent dans des habitats variés, allant des maisons modernes aux constructions traditionnelles en banco ou en matériaux précaires, en fonction de leurs caractéristiques économiques, sociales et culturelles. Dans ce cadre, l'entreprise immobilière ImmoFaso S.A, spécialisée dans le développement de logements urbains, cherche à mieux comprendre les facteurs qui influencent le type de logement occupé par les ménages afin d'adapter son offre immobilière à la demande réelle. Elle souhaite identifier les profils de ménages à cibler pour chaque type de logement disponible. Pour répondre à cette problématique, ce travail propose d'utiliser une régression logistique multinomiale, appliquée aux données de l'Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM 2018). La variable d'intérêt, le type de logement, comporte initialement huit modalités, que nous regroupons en quatre grandes catégories selon leur niveau de confort et de modernité, afin de faciliter l'analyse et l'interprétation. L'analyse s'appuie sur huit variables explicatives pertinentes, choisies parmi les dimensions économiques, démographiques et relatives aux conditions d'habitat. L'objectif est d'identifier les principaux déterminants du type de logement, afin de fournir à ImmoFaso S.A. des recommandations opérationnelles pour mieux segmenter le marché urbain et proposer des logements adaptés aux profils des ménages burkinabè.

2 Analyse descriptive préliminaire

Avant de procéder à la construction des variables synthétiques et à la modélisation, une analyse descriptive des données a été réalisée pour explorer les caractéristiques des ménages urbains et identifier les relations potentielles entre les variables. Cette analyse s'appuie sur la base `data_urban`, qui contient les données brutes avant tout regroupement ou création de variables composites.

2.1 Variables quantitatives

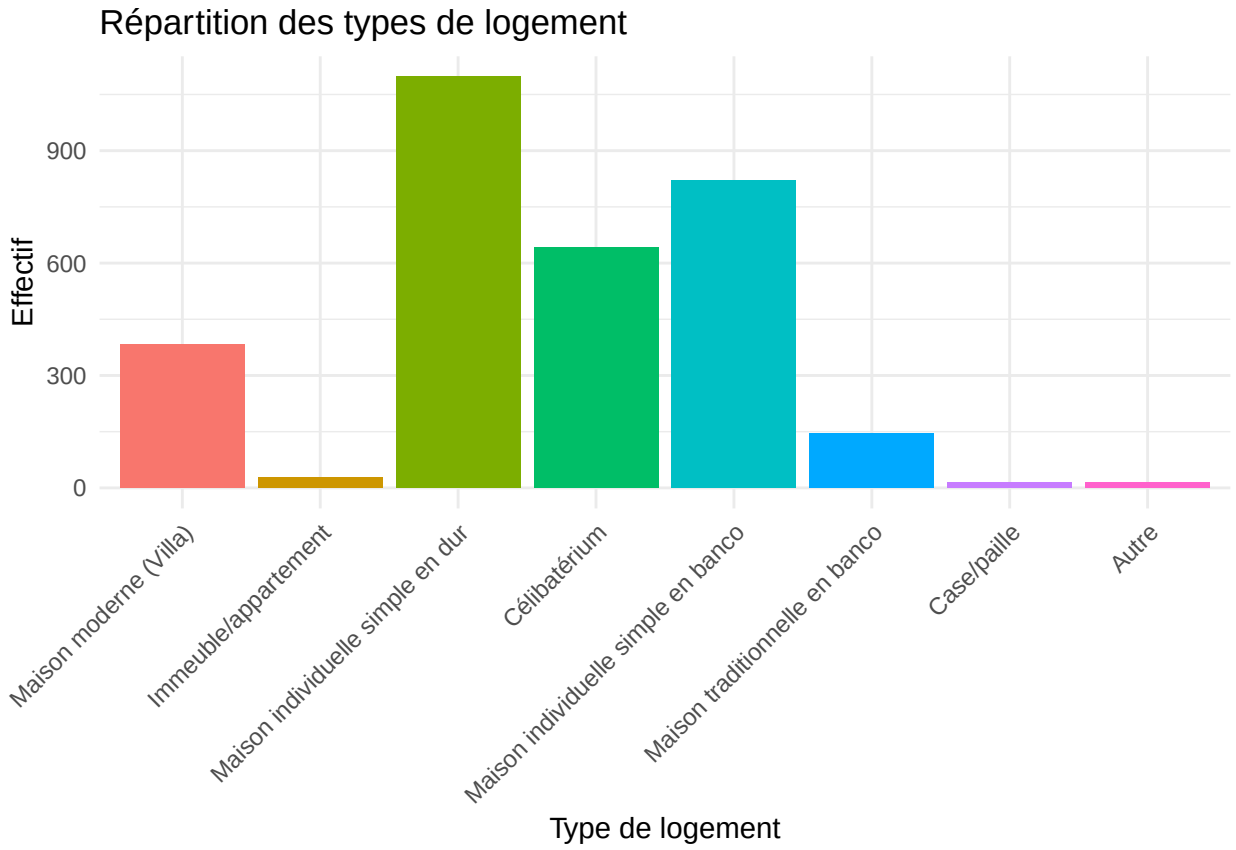
L'analyse statistique descriptive des variables quantitatives constitue une étape essentielle pour explorer les caractéristiques de base des données collectées. Elle permet de résumer de manière synthétique les principales tendances et la dispersion des variables d'intérêt. Le tableau ci-dessous présente les indicateurs statistiques clés: moyenne, médiane, écart-type, minimum, maximum et asymétrie pour trois variables quantitatives extraites de l'enquête : le revenu total du ménage (dtot), la taille du ménage (hhsize) et l'âge du chef de ménage (hage). Ces mesures fournissent un aperçu global de la structure économique, démographique et sociale des ménages enquêtés.

Table 1: Résumé statistique des variables quantitatives

Variable	Moyenne	Mediane	Ecart_type	Minimum	Maximum	Asymetrie
dtot	2689561.06	2143829	2062566.85	192662.2	18610512	2.46
hhsize	5.62	5	3.43	1.0	51	2.20
hage	46.10	44	14.12	18.0	99	0.60

Les données montrent que le revenu des ménages est en moyenne élevé mais très inégal, avec quelques ménages aux revenus extrêmement hauts. La taille moyenne des ménages est grande, autour de 5 à 6 personnes, avec certains ménages exceptionnellement nombreux. L'âge du chef de ménage est plus homogène, autour de 45 ans en moyenne, avec une distribution relativement équilibrée.

2.2 Variables qualitatives



La "Maison individuelle simple en dur" est le type de logement le plus fréquent, regroupant plus de 900 ménages, ce qui traduit une forte préférence pour ce mode d'habitat. Elle est suivie par la "Maison individuelle simple en banco" et les "Immeubles/appartements", avec respectivement environ 800 et 700 ménages, reflétant une diversité entre habitat traditionnel et moderne. En revanche, les "Maisons modernes (Villa)", les "Célibatériums", les "Cases/paille" et la catégorie "Autre" sont peu représentés, indiquant une adoption limitée de ces formes d'habitat, probablement en raison de contraintes économiques, culturelles ou liées à l'urbanisation.

2.3 Analyse croisée

Le tableau croisé présente la répartition des types de logement selon les régions du Burkina Faso, basée sur les données de l'EHCVM 2018. Cette analyse met en lumière les variations régionales des préférences résidentielles, révélant les habitats les plus courants dans chaque zone géographique. Elle offre ainsi un outil précieux pour comprendre les dynamiques territoriales du logement et orienter les politiques d'aménagement ainsi que les stratégies commerciales des acteurs immobiliers comme ImmoFaso S.A.

Table 2: Tableau croisé : Types de logement dans les 3 régions les plus et les moins peuplées en termes de ménages recensés

typ_logement_label	Centre	Hauts Bassins	Est	Centre-Sud	Nord	Plateau-Central
Maison moderne (Villa)	117	18	49	13	8	21
Immeuble/appartement	8	6	2	2	0	2
Maison individuelle simple en dur	206	110	64	61	49	61
Célibatérium	143	145	40	20	50	3
Maison individuelle simple en banco	54	33	77	83	67	72
Maison traditionnelle en banco	8	10	5	12	17	8
Case/paille	0	1	2	1	0	0
Autre	4	0	0	0	0	0

Le tableau met en évidence de fortes disparités régionales dans les types de logement. La région du **Centre**, la plus urbanisée, concentre les logements modernes tels que les **villas**, les **immeubles** et les **célibatériums**, contrairement aux régions moins peuplées comme le **Centre-Sud**, l'**Est** ou le **Plateau-Central**, où dominent les logements plus rudimentaires, notamment les **maisons en banco**. Le type de logement le plus courant dans toutes les régions reste la **maison individuelle simple en dur**, reflétant un compromis entre durabilité et accessibilité. Les **cases en paille** sont quasiment absentes, signalant leur disparition progressive. Ces différences traduisent des niveaux variés d'urbanisation et de développement socio-économique entre les régions.

3 Construction de variables synthétiques

Afin d'enrichir l'analyse et de réduire la dimensionnalité des informations disponibles, plusieurs variables composites ont été construites à partir des données originales.

3.1 Regroupement de la variable dépendante typ_logement

La variable initiale typ_logement comporte 8 modalités. Pour faciliter l'interprétation du modèle multinomial et garantir une taille suffisante dans chaque classe, nous avons regroupé ces modalités

en 4 catégories suivant leur rapprochement et représentativité dans la base de données :

3.2 Score de possession de biens (niveau de confort)

Un score synthétique de confort des ménages a été construit en deux étapes via des analyses en composantes principales (ACP). D'abord, un score de possession de sept biens durables (télévision, réfrigérateur, cuisinière, voiture, fer à repasser, ordinateur, décodeur) a été créé à partir de la première composante principale. Ensuite, ce score a été combiné avec l'accès à l'électricité et à l'eau potable, et une seconde ACP a produit un score global de confort. Ce dernier, continu, a été segmenté en trois niveaux (faible, moyen, élevé) selon les terciles, facilitant son usage dans les analyses statistiques.

3.3 Score de qualité du logement

À ce niveau, un score simple de qualité du logement a été construit à partir de trois variables binaires : type de mur, de toit et de sol. Le score qualite_logement est obtenu par somme directe (valeurs entre 0 et 3).

3.4 Indice socioéconomique par ACM

Ici, un indice synthétique de position sociale a été calculé par analyse des correspondances multiples (ACM) à partir du niveau d'éducation (heduc) et du statut dans la population active (hcsp) :

3.5 Sous-ensemble des variables candidates

La sélection des variables explicatives est cruciale pour construire un modèle économétrique solide. Elle consiste à identifier, selon la littérature, le contexte théorique et les données disponibles, les facteurs susceptibles d'influencer la variable dépendante. Dans cette étude, les variables ont été regroupées par thèmes (économique, sociodémographique, habitat, géographique) pour structurer l'analyse et faciliter l'interprétation. Le choix s'appuie sur la qualité des données, leur pertinence conceptuelle, l'absence de multicolinéarité et l'amélioration de l'ajustement, garantissant ainsi la rigueur méthodologique et la validité des résultats.

Table 3: Tableau : Justification des variables explicatives retenues

Variable	Description	Justification
type_logement	Type de logement (4 catégories)	Variable cible à prédire
statut_occupation	Occupation : propriétaire, locataire…	Lien direct avec le logement
age_chef	Âge du chef de ménage (cat.)	Influence les besoins en logement
taille_menage	Nombre de membres du ménage	Impacte la taille et le confort requis

Variable	Description	Justification
statut_matrimonial	État matrimonial du chef	Indicateur de stabilité sociale
niveau_education	Niveau d'instruction	Détermine les opportunités économiques
cat_socio_prof	Catégorie socioprofessionnelle	Mesure le statut socioéconomique
depenses_menage	Dépenses totales du ménage	Pouvoir d'achat du ménage
actif_derniere_sem	Actif durant la dernière semaine	Indique la participation économique
qualite_logement	Score sur les matériaux (sol, mur…)	Reflète la qualité du logement
niveau_confort	Score équipement (TV, frigo, etc.)	Indicateur de niveau de vie
electricite	Accès à l'électricité	Présence d'infrastructures modernes
region_residence	Région de résidence	Effets géographiques urbains
score_socioeco	Score éducation + activité	Score synthétique socioéconomique

4 Sélection des variables explicatives

4.1 Detection et traitement des valeurs manquantes

Avant d'estimer un modèle, il est essentiel d'examiner la complétude des données. La présence de valeurs manquantes peut biaiser les résultats ou réduire la puissance statistique si elles ne sont pas traitées de manière adéquate. Dans cette section, nous commençons par identifier les variables comportant des données manquantes, en utilisant un résumé global par variable.

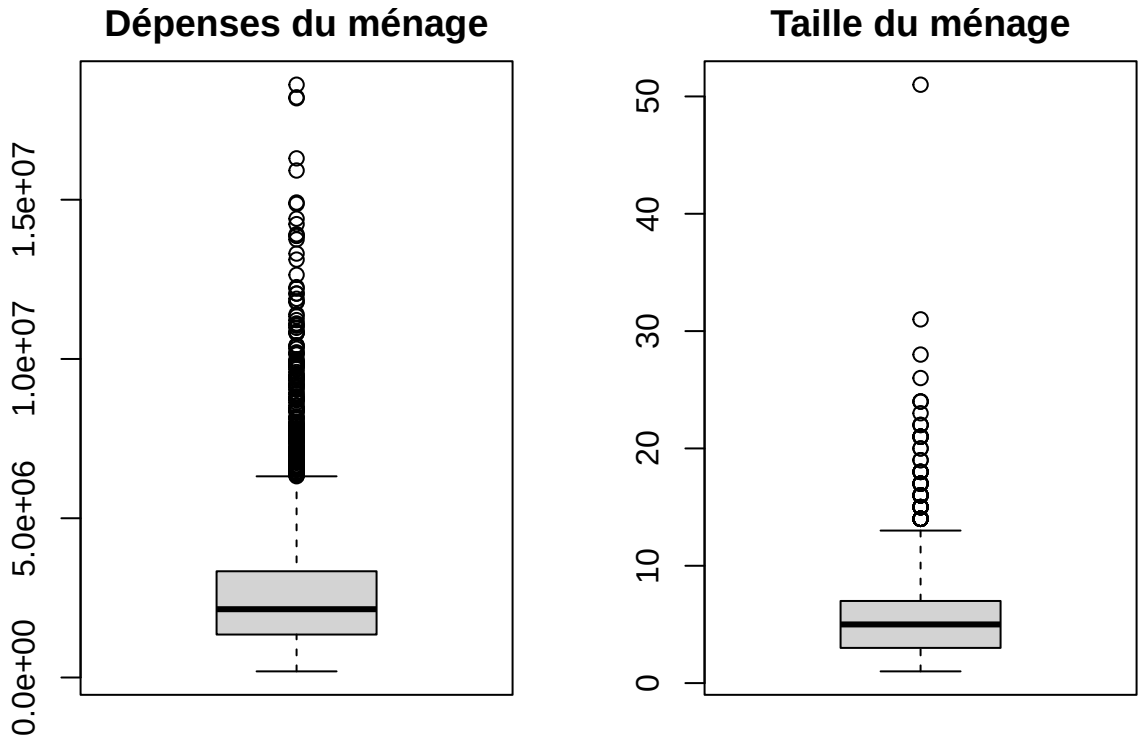
```
## # A tibble: 14 x 3
##    variable           n_miss pct_miss
##    <chr>               <int>    <num>
##  1 cat_socio_prof        359     11.4
##  2 type_logement           0      0
##  3 statut_occupation       0      0
##  4 age_chef                0      0
##  5 taille_menage           0      0
##  6 statut_matrimonial      0      0
##  7 niveau_education        0      0
##  8 depenses_menage         0      0
##  9 actif_derniere_sem      0      0
## 10 qualite_logement        0      0
## 11 niveau_confort          0      0
## 12 electricite             0      0
## 13 region_residence        0      0
## 14 score_socioeco          0      0
```

Pour gérer les données manquantes, nous allons utiliser l'imputation multiple avec le package mice. La variable catégorielle cat_socio_prof a été imputée par régression polytomique (polyreg), adaptée aux modalités multiples. La matrice des prédicteurs a été ajustée pour éviter la colinéarité et exclure la variable cible afin d'éviter la fuite d'information. L'imputation multiple permet d'obtenir des estimations plus fiables que la suppression simple des données manquantes. La base finale est issue de la combinaison des 5 jeux de données imputés.

On peut désormais confirmer qu'aucune valeur manquante ne subsiste sur les variables retenues pour le modèle. Une fois cette vérification achevée, l'analyse se poursuivra avec le traitement des valeurs aberrantes, afin d'assurer la robustesse des résultats.

4.2 Traitement des valeurs aberrantes

Avant d'engager l'analyse statistique, il est important d'identifier d'éventuelles valeurs aberrantes (ou outliers) susceptibles de biaiser les résultats. Ces valeurs extrêmes peuvent être dues à des erreurs de saisie, à des situations atypiques ou à une forte hétérogénéité dans la population observée. À cette fin, des boîtes à moustaches (boxplots) ont été utilisées pour visualiser la distribution des variables quantitatives clés, notamment les dépenses totales du ménage et la taille du ménage. Ces représentations permettent de repérer visuellement les observations éloignées des quartiles (valeurs au-delà de 1,5 fois l'écart interquartile).



En observant les deux, on remarque une forte disparité entre les depenses et les tailles des ménages. On remarque en effet la présence de valeurs atypiques. Ainsi nous allons utiliser le critère IQR (Interquartile Range) qui s'avère particulièrement efficace. Ce critère repose sur l'analyse de la distribution d'une variable, en identifiant les observations qui s'éloignent fortement de la zone centrale (le cœur de la distribution).

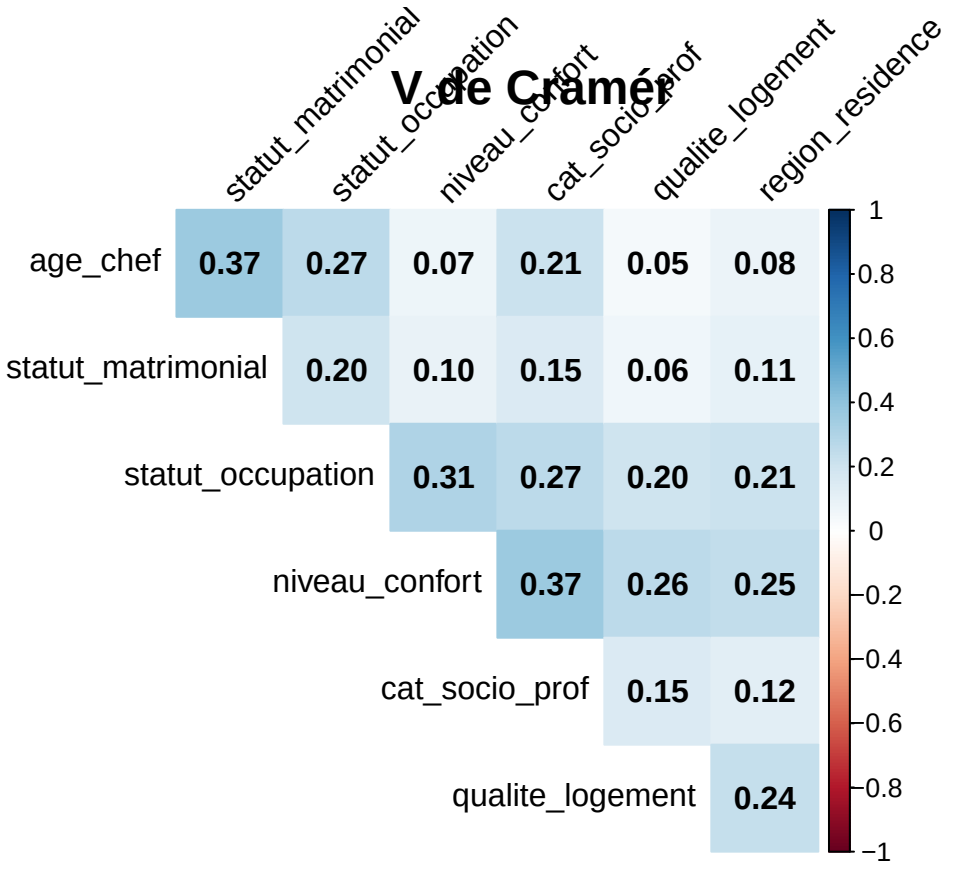
4.3 Test de correlation entre les variables quantitatives

Après la gestion des valeurs manquantes, nous allons maintenant nous interesser à la corrélation entre les variables quantitatives. En cas de forte corrélation détectée, nous allons procéder à la suppression d'une des variables pour conserver la robustesse du modèle. Tout d'abord nous allons réaliser le test de Shapiro-Wilk pour voir la distribution des données des variables quantitatives afin de nous orienter sur le test de corrélation approprié.

Étant donné que les variables ne suivent pas une distribution normale(d'après le test de Shapiro-Wilk), le test de Spearman a été privilégié pour évaluer la relation entre la taille du ménage et les dépenses du ménage. Les résultats montrent une corrélation positive modérée ($r = 0.301$), statistiquement significative. Cela signifie qu'en général, les dépenses des ménages tendent à augmenter avec leur taille, bien que cette relation ne soit pas très forte. Ce résultat est cohérent avec l'idée selon laquelle les besoins croissent avec le nombre de personnes dans un ménage. Le test ne revèle pas de corrélation forte entre les dépenses et la taille des ménages donc pour la suite les deux variables seront toutes consevées.

4.4 Test de correlation entre les variables qualitatives

Afin d'évaluer l'association entre variables qualitatives, le V de Cramer a été utilisé. Cet indicateur, dérivé du test du chi², permet de mesurer la force de l'association entre deux variables catégorielles, indépendamment de leurs dimensions. Nous allons visuliser ces resultats afin de choisir les variables à la fois pertinentes et moins corrélées.



Les résultats révèlent des liens généralement faibles à modérés, aucun coefficient ne dépassant 0.40. Les relations les plus notables concernent age_chef et niveau_confort (V = 0.37), ainsi que statut_occupation (V = 0.27), mais ces niveaux restent en deçà des seuils critiques de redondance. La variable score_socioeco, dérivée d'une ACP, est par construction indépendante. De plus, region_residence montre peu d'association avec les autres variables. Ainsi, les variables age_chef, score_socioeco, niveau_confort, qualite_logement, region_residence et statut_occupation peuvent être intégrées dans une régression logistique multinomiale sans risque notable de colinéarité, en couvrant des dimensions complémentaires du profil des ménages.

5 Regression logistique

Dans le but d'expliquer les déterminants du type de logement des ménages, une régression logistique multinomiale a été estimée. La variable dépendante type_logement est catégorielle à plus de deux modalités, justifiant l'usage de ce modèle. Les variables explicatives incluent des caractéristiques démographiques (taille_menage, age_chef), économiques (score_socioeco, depenses_menage), résidentielles (region_residence, statut_occupation), ainsi qu'un indicateur de confort (niveau_confort) et de qualité de l'habitat (qualite_logement). Par ailleurs, la variable continue depenses_menage a été log-transformée afin de réduire l'asymétrie de sa distribution et améliorer l'ajustement du modèle.

```r
Base_finale = DataBase

# Application du logarithme
Base_finale$depenses_menage= log(Base_finale$depenses_menage + 1)

# Regression
model_3 <- multinom(type_logement ~ taille_menage + age_chef + score_socioeco + niveau_confort
                    data = Base_finale)
summary(model_3)
```

Interprétation des résultats: Le modèle logit multinomial montre que le type de logement est influencé principalement par le score socioéconomique, la qualité du logement, le statut d'occupation, ainsi que la région de résidence. Certaines modalités de ces variables présentent des effets marqués sur les probabilités d'occuper un type de logement particulier. L'ajustement du modèle est globalement satisfaisant (AIC = 4658,2), et les résultats confirment l'importance des facteurs sociaux et territoriaux dans le ciblage des logements à proposer aux ménages.

6 Diagnostic et validation du modèle

Afin de garantir la robustesse et la fiabilité du modèle multinomial estimé, il est essentiel de procéder à une série de validations statistiques. Ces étapes permettent de s'assurer que le modèle est correctement spécifié, qu'il fournit des prédictions cohérentes et qu'il respecte les hypothèses de base.

6.1 Test du rapport de vraisemblance (Likelihood Ratio Test - LRT)

Le test du rapport de vraisemblance (Likelihood Ratio Test – LRT) permet d'évaluer la performance globale du modèle en comparant le modèle complet (incluant les variables explicatives) à un modèle nul (ne contenant que l'intercept). L'objectif est de tester si les variables explicatives introduites dans le modèle apportent une amélioration significative de l'ajustement par rapport à un modèle sans information. Plus précisément, il s'agit de tester l'hypothèse nulle suivante :

$$H_0 : \text{Tous les coefficients des variables explicatives sont nuls}$$

Si la statistique du LRT est significative (p-value < 0.05), on peut conclure que le modèle complet est globalement meilleur que le modèle nul, ce qui justifie la présence des variables explicatives.

Après le test, on obtient les résultats suivants :

- **Deviance résiduelle du modèle nul** : 7416.191

- **Deviance résiduelle du modèle complet** : 4502.224
- **Différence de déviance (statistique LRT)** : 2913.966
- **Degrés de liberté** : 75
- **p-value** : < 0.0001

Interprétation : La statistique G^2 est très élevée et la p-value est pratiquement nulle, ce qui indique un rejet clair de l'hypothèse nulle. Autrement dit, les variables explicatives introduites dans le modèle améliorent **significativement** l'ajustement par rapport au modèle nul. Le modèle `model_3` présente donc une **validité globale satisfaisante** pour expliquer les différences observées dans le type de logement.

6.2 Pseudo R² de McFadden

Le pseudo R^2 de McFadden est une mesure d'ajustement global utilisée pour les modèles de régression logistique, notamment multinomiaux. Contrairement au R^2 classique de la régression linéaire, il mesure la performance relative du modèle complet par rapport à un modèle nul (sans prédicteurs).

La formule est la suivante :

$$R^2_{\text{McFadden}} = 1 - \frac{\log L_{\text{modèle}}}{\log L_{\text{modèle nul}}}$$

Des valeurs proches de 0 indiquent un mauvais ajustement. En pratique, un R^2 de McFadden compris entre 0.2 et 0.4 est généralement considéré comme acceptable.

Résultats :

```
## fitting null model for pseudo-r2
## # weights:  8 (3 variable)
## initial  value 3992.527760
## final  value 3708.095450
## converged
```

```
##           llh       llhNull            G2      McFadden          r2ML
## -2251.1122349 -3708.0954496  2913.9664293     0.3929196     0.6364338
##          r2CU
##     0.6888921
```

Interprétation : Le pseudo R^2 de McFadden obtenu (0.393) indique un bon niveau d'ajustement global du modèle multinomial model_3. Cela confirme que les variables explicatives apportent une contribution substantielle à la prédiction du type de logement. Ce résultat est cohérent avec les conclusions du test du rapport de vraisemblance.

6.3 Validation croisée (Cross-validation)

La validation croisée est une méthode utilisée pour évaluer la **capacité prédictive** d'un modèle statistique sur des données nouvelles, non utilisées lors de l'estimation. Elle permet de tester la **robustesse** et la **généralisation** du modèle, en réduisant les risques de surapprentissage (overfitting). Dans cette étude, la validation croisée permet de vérifier si le modèle multinomial `model_3` conserve un bon pouvoir prédictif en dehors de l'échantillon utilisé pour son estimation.

Résultats:

Table 4: Accuracies par fold et accuracy moyenne (validation croisée 5 folds)

Fold	Accuracy
Fold 1	0.6551
Fold 2	0.6632
Fold 3	0.6957
Fold 4	0.6875
Moyenne :	0.6754

Interprétation : Les résultats montrent une performance stable du modèle à travers les différentes partitions du jeu de données. L'accuracy moyenne d'environ 67.3 % indique que le modèle parvient à prédire correctement le type de logement dans plus de deux tiers des cas sur des données non utilisées lors de l'entraînement.

6.4 Test de significativité de Wald

Table 5: Tests de significativité des coefficients du modèle multinomial (test de Wald) - 10 premières lignes

Variable	Modalite	Estimate	Std_Error	z_value	p_value
Célibatérium	(Intercept)	16.4759	2.8328	5.8161	0.0000
Logement traditionnel	(Intercept)	43.7579	3.8095	11.4864	0.0000
Maison individuelle en dure	(Intercept)	24.1388	3.5468	6.8058	0.0000
Célibatérium	age_chefadulte	-0.0496	0.2926	-0.1694	0.8655
Logement traditionnel	age_chefadulte	-0.2375	0.3082	-0.7705	0.4410
Maison individuelle en dure	age_chefadulte	0.2559	0.2814	0.9093	0.3632
Célibatérium	age_chefâgé	-0.7391	0.3422	-2.1600	0.0308
Logement traditionnel	age_chefâgé	-0.6283	0.3300	-1.9042	0.0569
Maison individuelle en dure	age_chefâgé	-0.0882	0.3018	-0.2923	0.7700
Célibatérium	depenses_menage	-1.8297	0.2752	-6.6493	0.0000

Interpretation: Le modèle multinomial révèle que plusieurs facteurs démographiques, socioéconomiques et liés au logement influencent significativement le choix du type de logement par rapport à la catégorie de référence. L'âge du chef de ménage, les dépenses, le niveau de confort et la qualité du logement jouent un rôle important, tout comme la région de résidence et le statut d'occupation. Par exemple, l'âge diminue la probabilité de choisir un célibatérium, tandis que des dépenses plus élevées réduisent la probabilité de certains logements. Le score socioéconomique montre une hiérarchie claire dans les choix, et la taille du ménage favorise certains types de logement tout en limitant d'autres. Ces résultats confirment que les caractéristiques individuelles et contextuelles sont déterminantes pour expliquer la diversité des types de logement choisis.

6.5 Multicolinéarité

La multicolinéarité survient lorsque plusieurs variables explicatives d'un modèle de régression sont fortement corrélées, ce qui complique l'estimation précise des coefficients, augmente les erreurs standard et diminue la fiabilité des tests statistiques. Dans un modèle multinomial, elle peut biaiser l'interprétation des effets individuels et déstabiliser les coefficients estimés. Pour la détecter, on utilise notamment le facteur d'inflation de la variance (VIF), où une valeur élevée (au-delà de 5 ou 10) indique un risque de multicolinéarité problématique. L'analyse de cette multicolinéarité est essentielle pour assurer la robustesse et la fiabilité du modèle.

Table 6: GVIF et GVIF corrigé pour les variables explicatives

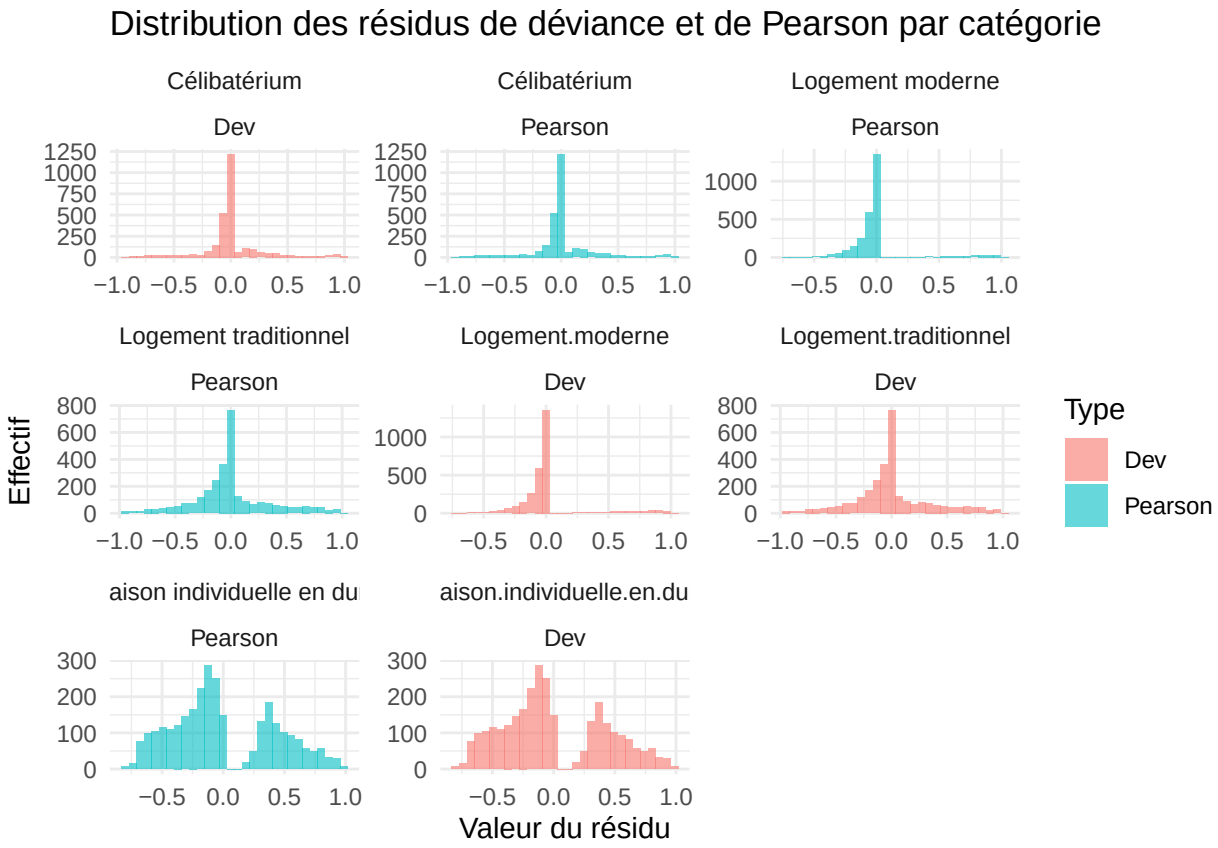
Variable	GVIF	Df	GVIF_corrige
age_chef	1.2491	2	1.0572
score_socioeco	1.6217	1	1.2734
niveau_confort	2.1172	2	1.2063
qualite_logement	1.4716	3	1.0665
region_residence	1.5951	12	1.0196
depenses_menage	1.8448	1	1.3582
statut_occupation	1.9942	3	1.1219

Interprétation : Le tableau présente les valeurs de GVIF (*Generalized Variance Inflation Factor*) ainsi que leur correction tenant compte des degrés de liberté associés à chaque variable explicative. Les valeurs corrigées du GVIF sont toutes inférieures à 2, ce qui indique une absence de multicolinéarité préoccupante entre les variables. Cela signifie que les variables incluses dans le modèle sont suffisamment indépendantes, ce qui garantit la stabilité des coefficients estimés et la fiabilité des tests statistiques.

6.6 Analyse des résidus

L'analyse des résidus permet d'évaluer la qualité de l'ajustement du modèle multinomial.

- Les **résidus de déviance** mesurent la différence entre les observations réelles et les valeurs prédites, en tenant compte de la fonction de vraisemblance du modèle.
- Les **résidus de Pearson** évaluent les écarts standardisés entre les observations et les prédictions, en se basant sur la variance des données. L'étude de ces résidus aide à détecter d'éventuelles observations aberrantes, des problèmes d'ajustement, ou des spécifications manquantes dans le modèle.



Les histogrammes des résidus de déviance et de Pearson montrent un bon ajustement global du modèle multinomial, avec des résidus centrés autour de zéro et relativement symétriques, renforçant la validité des estimations. Cependant, une plus grande dispersion des résidus est notée pour les catégories « Maison individuelle en dure » et « Logement traditionnel », indiquant une variabilité accrue et une précision moindre. En revanche, les catégories « Célibatérium » et « Logement moderne » présentent une forte concentration des résidus près de zéro, traduisant une meilleure performance prédictive. Ainsi, bien que le modèle soit performant pour certaines catégories, il pourrait être amélioré pour d'autres en enrichissant les variables explicatives ou en explorant d'autres approches de modélisation.

7 Recommandations opérationnelles pour ImmoFaso S.A.

Sur la base de l'analyse économétrique du modèle multinomial appliqué aux données de l'EHCVM (2018), plusieurs recommandations concrètes peuvent être formulées à l'intention de l'entreprise **ImmoFaso S.A.**, en vue d'adapter son offre immobilière aux besoins différenciés des ménages urbains au Burkina Faso.

D'abord, la **segmentation du marché** selon les caractéristiques des ménages s'impose. Les ménages à **haut score socioéconomique** (éducation et emploi stables) privilégient les logements modernes. Il est donc conseillé à ImmoFaso d'investir dans des appartements bien équipés, situés en zone urbaine, avec des services de qualité (eau, électricité, sécurité). À l'inverse, les ménages à **faibles revenus** optent davantage pour les logements traditionnels ou les célibatériums. Pour répondre à cette demande, l'entreprise peut développer une offre de logements à bas coût dans les zones périurbaines, utilisant des matériaux améliorés (ex. : briques stabilisées), et proposer des formules locatives souples. Les **ménages de grande taille** devraient quant à eux être ciblés par des logements spacieux, notamment des maisons en dur à plusieurs pièces avec cour intérieure.

Ensuite, l'**analyse régionale** met en lumière des spécificités spatiales. Par exemple, la **région du Nord** montre une forte demande pour les célibatériums, tandis que la **région du Sahel** est davantage associée aux logements traditionnels. ImmoFaso peut donc adapter son offre localement : studios pour jeunes travailleurs dans la région du Nord, maisons traditionnelles améliorées en région du Sahel. Cette orientation géographique permettrait d'optimiser les investissements immobiliers en fonction des préférences régionales.

Par ailleurs, les **niveaux de confort** et la **qualité perçue du logement** influencent fortement les choix résidentiels. Il convient donc d'intégrer des équipements modernes (toilettes intérieures, accès à l'électricité, carrelage) dans les projets destinés aux ménages à confort moyen ou élevé, et d'améliorer la durabilité des logements traditionnels sans augmenter drastiquement les coûts pour les ménages modestes. Des solutions intermédiaires peuvent inclure l'utilisation de toits en tôle de qualité et l'accès collectif à l'eau.

Concernant le **statut d'occupation**, des stratégies différenciées doivent être mises en œuvre. Les célibatériums peuvent être proposés avec des taux de location flexibles pour les travailleurs mobiles, tandis que l'accès à la propriété pour les ménages modestes pourrait être facilité via des crédits adaptés ou des programmes de construction par étapes, où les ménages bâtissent leur logement progressivement.

D'un point de vue analytique, bien que le modèle présente une performance correcte (pseudo R^2 de McFadden = 0.39 et une précision moyenne de 67.3 % en validation croisée), certaines catégories comme les « maisons en dur » et « logements traditionnels » sont moins bien prédites. Il est donc recommandé de **collecter des données complémentaires** (qualitatives et quantitatives) sur les préférences des ménages, les freins économiques ou culturels à l'accession au logement, et les infrastructures locales. ImmoFaso pourrait également tester des modèles alternatifs (forêts

aléatoires, modèles avec interactions) pour affiner la compréhension des déterminants du choix de logement.

Enfin, les **diagnostics du modèle** confirment la robustesse de l'analyse : l'absence de multicolinéarité (GVIF < 2) assure la fiabilité des estimations, et la distribution des résidus suggère que le modèle est globalement bien ajusté. Des efforts peuvent cependant être faits pour réduire la variabilité observée dans certaines modalités, notamment via une **mise à jour des données** (postérieures à 2018), afin de tenir compte des évolutions récentes du contexte urbain et des modes d'habitation.

En résumé, ces recommandations opérationnelles visent à permettre à **ImmoFaso S.A.** de mieux adapter son offre aux profils variés des ménages burkinabè, en combinant des stratégies d'investissement ciblées, une segmentation socioéconomique fine, et des approches innovantes en matière de logement abordable et durable.

8 Conclusion

Cette étude, basée sur les données de l'EHCVM 2018, a analysé les facteurs déterminant le type de logement en milieu urbain au Burkina Faso à l'aide d'un modèle logit multinomial pondéré. Les résultats révèlent que des variables telles que le score socioéconomique, la qualité du logement, le statut d'occupation, la taille du ménage, les dépenses et la région influencent significativement les choix résidentiels. Le modèle présente un bon ajustement (pseudo $R^2 = 0{,}393$; précision = 67,3 %), malgré quelques limites sur certaines catégories. Sur cette base, des recommandations ciblées sont formulées pour ImmoFaso S.A., notamment la segmentation de l'offre, le développement de logements adaptés aux différents profils et une stratégie d'investissement géographique différenciée. Ce travail constitue un appui stratégique pour une politique immobilière inclusive, adaptée à la diversité du marché urbain burkinabè.