Université d'Abomey-Calavi

072 B.P. 50 Cotonou, République du Bénin

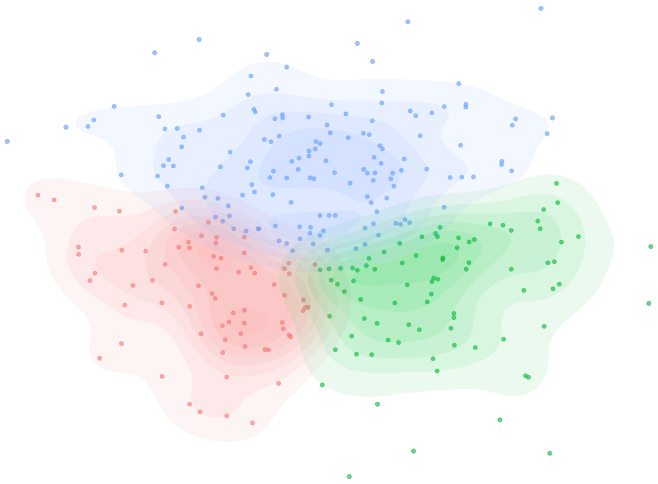
Chaire Internationale en Physique, Mathématiques et Applications

(CIPMA-CHAIRE UNESCO)

CYCLE : Master 2 **FILIÈRE :** Statistique Appliquée aux Vivants

Déterminants des conditions socio-économiques de vie des ménages

Modèle Polytomique : Sorted or not



Réalisé par :

- GANGNON Junior
- SESSOU G. Pascal
- YESSIFOU Chabi André
- CODJO Eliab

Sous la direction de :

Dr. ATCHADE Nicodème

Année Académique
2025-2026

TABLE DES MATIÈRES

Table des matières

LISTE DES FIGURES

Table des figures

LISTE DES TABLEAUX

Liste des tableaux

INTRODUCTION

Le niveau socio-économique des ménages constitue un déterminant central du bien-être individuel et collectif. Il regroupe des dimensions multiples : revenus, éducation, conditions de logement, accès aux services et patrimoine et influe fortement sur l'accès à la santé, à l'éducation et aux opportunités économiques. Comprendre les facteurs associés à l'appartenance à des catégories socio-économiques distinctes (par exemple Faible, Moyen, Élevé) permet d'orienter des politiques publiques ciblées visant à réduire les inégalités et à promouvoir un développement inclusif.

De nombreuses études ont mis en évidence que les Conditions de vie des ménages dépend fortement du capital humain, de la composition du ménage et des conditions d'emploi. Par exemple, [1] montre que, au Bénin, l'alphabétisation et le niveau d'instruction des chefs de ménages réduisent significativement le risque de pauvreté monétaire. De même, une analyse récente fondée sur l'enquête EMOP au Mali [2] indique que la taille du ménage et le sexe du chef influent sur la consommation par ménage, les ménages dirigés par une femme affichant en moyenne des dépenses plus faibles. Enfin, le rapport du PNUD Bénin [3] souligne l'importance du capital humain et de la structure de l'emploi pour expliquer les faibles revenus et la prévalence de la pauvreté.

Cependant, ces effets peuvent varier selon les contextes locaux et selon la manière dont on mesure le statut socio-économique.Dans le cadre d'une application pratique du cours de Modèle Paramétrique qui est un cours fondamental pour les futurs ingénieurs que nous somme ce présent document s'est donc intéressé à un ensemble de données compilées par compilées par l'Ecole Nationale de Statistiques, de Planification et de Démograpie (ENSPD) en 2022. Il prend en compte plusieurs variables économiques, sanitaires et démographiques dont notre variable d'intêret a plusieurs modalités ***Faible / Moyen / Élevé*** de niveau de vie.

Modèle polytomique (sorted or not)
Déterminants du niveau socio-économique

1 Présentation du jeu de données

1.1 Contenu du jeu de données

Le jeu de données que nous avons utilisé provient de l'enquête ménage effectué en 2022 par l'ENSPD. Il contenait 1954 observations (donc ménages) qui avaient été observés autour de 30 à 40 variables. Pour notre étude et vu la direction de notre travail nous nous sommes retouvés à 8 principales variables d'intêret.

1.2 Présentation des variables

1.2.1 Variable dépendante : Conditions de vie du ménage

La variable d'intérêt de cette étude est la condition de vie du ménage, une variable catégorielle ordonnée ou polytomique qui traduit la situation économique globale des ménages. Dans la base, cette variable prend plusieurs modalités telles que *Faible*, *Moyen* et *Élevé*. Elle constitue la variable dépendante dans le modèle de régression polytomique visant à identifier les déterminants socioéconomiques des conditions de vie des ménages.

1.2.2 Variables explicatives

Les variables explicatives sélectionnées reflètent les caractéristiques démographiques, sociales et économiques du chef de ménage et de son environnement. Elles comprennent :

Variables démographiques et ou culturelles :

— **Sexe du chef de ménage** : Variable dichotomique (Masculin/Féminin) permettant d'analyser l'effet du genre sur la varible d'intêret.
— **Âge du chef de ménage** : Variable continue ou catégorisée reflétant l'expérience et le cycle de vie du ménage.
— **Taille du ménage** : Nombre de personnes composant le ménage, indicateur de la charge démographique.
— **Ethnie** : indicateur de différences culturelles .
— **Religion** : indicateur de croyances.
— **Statut matrimoniale** : indicateur relationelle

Variables socio-économiques :

— **Niveau d'instruction du chef de ménage** : Variable catégorielle ordonnée (Aucun/Primaire/Secondaire/Supérieur) mesurant le capital humain.

Ces variables permettent de capturer les multiples dimensions qui influencent les conditions de vie des ménages et constituent les prédicteurs du modèle polytomique.

1.2.3 Tableau récapitulatif des variables

Le tableau ci-dessous suivis de son code présente de manière synthétique l'ensemble des variables retenues dans cette étude, regroupées par catégories thématiques.

Il synthétise les principales variables mobilisées pour expliquer la variable d'intêret. Ces variables couvrent un large spectre de déterminants démographiques, sociaux, éducatifs et générationnels, permettant une analyse multidimensionnelle des facteurs influençant le statut socio-économique des ménages.

Création du tableau des variables

```
\begin{lstlisting}[language=R,basicstyle=\ttfamily\small]
library(knitr)
library(kableExtra)

# Création du dataframe
variables_df <- data.frame(
  Categorie = c(
    "Caractéristiques démographiques",
    "",
    "Caractéristiques sociales",
    "",
    "",
    "Caractéristiques liées à l'éducation"
  ),
  Variable = c(
    "Sexe_chef_menage",
    "Groupe_age_chef",
    "Ethnie",
    "Religion",
    "Statut_matrimonial",
    "Niveau_instruction"
  ),
  Description = c(
    "Genre du chef de ménage",
    "Tranche d'âge du chef",
    "Appartenance ethnique",
    "Religion du chef",
    "Statut marital",
    "Niveau d'instruction"
  ),
  stringsAsFactors = FALSE
)
\end{lstlisting}
```

Table 1 – Tableau récapitulatif des variables

Catégorie	Variable	Description
Caractéristiques démographiques	Sexe_chef_menage	Genre du chef de ménage
	Groupe_age_chef	Tranche d'âge du chef
Caractéristiques sociales	Ethnie	Appartenance ethnique
	Religion	Religion du chef
	Statut_matrimonial	Statut marital
Caractéristiques liées à l'éducation	Niveau_instruction	Niveau d'instruction

Source : Enquête ENSPD 2022

Un certain nombre d'indicateurs permettent d'approcher les conditions de vie des ménages. On peut citer :

L'approche monétaire

L'approche monétaire est de loin la plus répandue. Elle tient compte des revenus du ménage et cherche à fixer un seuil plus adapté selon la taille du ménage.

L'approche subjective

Elle intègre l'opinion de la personne enquêtée sur sa propre situation financière et son bien-être, du nombre d'enfants qu'il y a dans le ménage.

L'approche par les biens matériaux

Elle est élaborée à partir de nombreux indicateurs comme le manque de bien-être matériel. Pour les tenants de cette approche, ce qui permet de définir la catégorie des pauvres ce n'est pas le manque de tel ou tel bien matériel élémentaire mais plutôt le cumul des handicaps.

Notre choix s'est porté sur **l'approche par les biens matériaux** que nous jugeons adaptable au contexte, car les revenus ont toujours fait l'objet d'une mauvaise déclaration pour des raisons fiscales ou autres.

L'identification de la construction de notre indicateur prend en compte aussi bien les caractéristiques de l'habitat que les biens d'équipement du ménage : la possession de biens tels que l'électricité, le téléphone, la radio, le téléviseur, le réfrigérateur, le vélo, la moto, la voiture, le mode d'approvisionnement en eau de boisson et le type de toilette.

La méthode utilisée pour la construction de l'indicateur de niveau de vie est la **Classification Hiérarchique Ascendante (CAH)** basée sur l'Analyse Factorielle des Correspondances Multiples (AFCM) à partir du logiciel R. À l'issue de cette méthode d'analyse, nous avons obtenu trois classes d'individus. Ces différentes classes ont été recodées en trois modalités permettant de catégoriser les individus en *Faible*, *Moyen* et *Élevé*.

1.3.2 Inspection de la variable dépendante

Avant d'appliquer un modèle polytomique, il est essentiel de vérifier la répartition des observations entre les différentes modalités de la variable dépendante. Un déséquilibre extrême entre les catégories pourrait affecter la qualité des estimations et la robustesse du modèle. Le tableau ci-dessous présente la distribution des ménages selon leur niveau de vie.

Distribution de la variable dépendante

```
\begin{lstlisting}[language=R,basicstyle=\ttfamily\small]
# ============================================================
# CHARGEMENT DES BIBLIOTHÈQUES
# ============================================================
library(tidyverse)
library(stargazer)
library(magrittr)
library(margins)
library(caret)
library(readxl)

# ============================================================
# IMPORTATION DES DONNÉES
# ============================================================
directory <- "D:/SAV/SAV2_2025/modele_parametrique/TP_polytomique/Tp_poly/Poly2/Modele_polytomic/"
Base_polytomique <- read_excel(paste0(directory, 'Base_polytomique.xlsx'))

# ============================================================
# CALCUL DES STATISTIQUES DE DISTRIBUTION
# ============================================================
effectifs <- table(Base_polytomique$Conditions_de_vie)
pourcentages <- round(prop.table(effectifs) * 100, 2)

# Création du dataframe
distribution_df <- data.frame(
  Conditions_de_vie = names(effectifs),
  Effectif = as.numeric(effectifs),
  Pourcentage = paste0(pourcentages, " %"),
  stringsAsFactors = FALSE
)

# Ajout d'une ligne Total
total_row <- data.frame(
  Conditions_de_vie = "Total",
  Effectif = sum(distribution_df$Effectif),
  Pourcentage = "100",
  stringsAsFactors = FALSE
)

distribution_df <- rbind(distribution_df, total_row)
\end{lstlisting}
```

Table 2 – Distribution des ménages selon les conditions de vie

Conditions de vie	Effectif	Pourcentage
Eleve	530	27.12 %
Faible	173	8.85 %
Moyen	1251	64.02 %
Total	**1954**	**100**

Source : Enquête ENSPD 2022

La variable dépendante “Conditions de vie” prend trois modalités ordonnées (Élevé, Moyen, Faible). Dans l'échantillon (n = 1954), la catégorie Moyen est largement majoritaire (1251 ménages, 64.02 %), la catégorie Élevé compte 530 ménages (27.12 %) et la catégorie Faible n'est que 173 ménages (8.85 %).

De prime abord une variable est ordinale quand ses modalités peuvent être rangées sur une échelle du « plus petit » au « plus grand » selon un critère logique. Ici Faible, Moyen, Élevé correspondent clairement à des niveaux croissants du niveau de vie : on peut dire qu'« Élevé » est plus que « Moyen », et « Moyen » est plus que « Faible ». C'est donc un ordre naturel qui oriente en première analyse vers un modèle polytomique ordinal (modèle des odds proportionnels).

Cependant, on note un déséquilibre marqué entre modalités (la classe Moyen domine). Ce déséquilibre peut influer sur l'estimation et la puissance pour détecter des effets concernant la petite modalité Faible ; il impose de vérifier la robustesse des résultats et l'hypothèse d'odds proportionnels.

D'après la littérature deux choix s'offrent à nous :

— Partir sur un tirage aleatoire au niveau de chaque modalités afin d'equilibrer les effectifs
— Effectuer des tests comme le test de **Brant** pour juger de la pertinence du modele ordonné. Si l'hypothèse est violée, nous estimerons un modèle multinomial (non ordonné) qui est plus adéquat dans ce cas.

Pour notre travail nous estimerons les deux modèles s'il y a lieu avec et nous les comparerons dans le cas ou l'hypothèse de proportionnalité sera violée (le test de Brant).

2 Présentation des modèles statistiques

Dans le cadre de l'analyse des déterminants socio-économiques des Conditions de vie des ménages, nous mobilisons des modèles de régression polytomique. Ces modèles sont particulièrement adaptés lorsque la variable dépendante est catégorielle et possède plus de deux modalités. Dans notre cas, la variable « Conditions de vie » prend trois modalités ordonnées : *Faible*, *Moyen* et *Élevé*. Deux approches principales se distinguent selon que l'on considère ou non l'ordre naturel des catégories : le modèle polytomique ordonné et le modèle polytomique non ordonné. Le choix entre ces deux modèles dépend de la nature de la variable dépendante et de l'hypothèse de proportionnalité.

2.1 Modèle polytomique ordonné (Ordered Logit)

2.1.1 Principe du modèle

Le modèle polytomique ordonné, également appelé modèle logit proportionnel ou modèle à cotes proportionnelles, est approprié lorsque la variable dépendante présente un ordre naturel entre ses catégories. Dans notre étude, **Conditions de vie** possède une hiérarchie claire : Faible < Moyen < Élevé. Ce modèle suppose que l'effet des variables explicatives est constant d'une catégorie à l'autre (hypothèse des cotes proportionnelles).

2.1.2 Formulation générale

Soit Y la variable dépendante représentant les **Conditions de vie** avec J modalités ordonnées ($j = 1, 2, \ldots, J$). Le modèle polytomique ordonné est défini par :

$$P(Y \leq j|X) = \frac{1}{1 + \exp(-(\alpha_j - X'\beta))} \tag{1}$$

où :

- $P(Y \leq j|X)$ est la probabilité cumulée que Y soit inférieur ou égal à la catégorie j étant donné X
- α_j sont les seuils spécifiques à chaque catégorie, avec $\alpha_1 < \alpha_2 < \ldots < \alpha_{J-1}$
- X est le vecteur des variables explicatives

— β est le vecteur des coefficients associés aux variables explicatives (commun à toutes les catégories)

La probabilité d'appartenir à une catégorie spécifique j est alors :

$$P(Y = j|X) = P(Y \leq j|X) - P(Y \leq j-1|X) \tag{2}$$

2.1.3 Application à notre étude

Dans le cadre de notre analyse des déterminants des **Conditions de vie**, le modèle s'écrit :

$$\begin{aligned} P\big(\text{Niveau\_vie} \leq j \mid X\big) &= \frac{1}{1+\exp\big(-(\alpha_j - X^\top\beta)\big)} \\ \text{avec} \quad X^\top\beta &= \beta_1\,\text{Taille\_menage} + \beta_2\,\text{Sexe\_chef} + \beta_3\,\text{Groupe\_age} \\ &\quad + \ \beta_4\,\text{Ethnie} + \beta_5\,\text{Religion} + \ \beta_6\,\text{Statut\_matrimonial} \\ &\quad + \ \beta_7\,\text{Niv\_instruction}. \end{aligned} \tag{3}$$

où :

— $j \in \{1, 2, 3\}$ représente les modalités : 1 = Faible, 2 = Moyen, 3 = Élevé
— α_1 et α_2 sont les deux seuils à estimer
— Les variables explicatives sont :
  — **Taille_menage** : Nombre de personnes dans le ménage
  — **Sexe_chef** : Sexe du chef de ménage (0 = Féminin, 1 = Masculin)
  — **Groupe_age** : Catégorie d'âge du chef de ménage
  — **Ethnie** : Appartenance ethnique du chef de ménage
  — **Religion** : Religion du chef de ménage
  — **Statut_matrimonial** : Statut marital du chef de ménage
  — **Niv_instruction** : Niveau d'instruction du chef de ménage

2.1.4 Hypothèse des cotes proportionnelles

L'hypothèse fondamentale du modèle ordonné est celle des **cotes proportionnelles** (proportional odds assumption). Elle stipule que l'effet d'une variable explicative sur le logarithme des cotes est le même pour toutes les transitions entre catégories adjacentes. Formellement :

$$\frac{P(Y \leq j|X)}{P(Y > j|X)} = \exp(\alpha_j - X'\beta) \tag{4}$$

Cette hypothèse sera testée à l'aide du test de Brant. Si elle est violée, le modèle polytomique non ordonné sera privilégié.

2.2 Modèle polytomique non ordonné (Multinomial Logit)

2.2.1 Principe du modèle

Le modèle polytomique non ordonné, ou modèle logit multinomial, est utilisé lorsque les catégories de la variable dépendante n'ont pas d'ordre naturel, ou lorsque l'hypothèse des cotes proportionnelles du modèle ordonné est violée. Ce modèle permet aux effets des variables explicatives de varier d'une catégorie à l'autre. Dans ce modèle, on choisit une catégorie de référence (généralement la première ou la dernière) et on modélise les probabilités relatives de chaque autre catégorie par rapport à cette référence.

2.2.2 Formulation générale

Soit Y la variable dépendante avec J modalités. En prenant la catégorie $j = 1$ comme référence, le modèle s'écrit :

$$P(Y = j|X) = \frac{\exp(X'\beta_j)}{\sum_{k=1}^{J} \exp(X'\beta_k)} \quad \text{pour } j = 1, 2, \ldots, J \tag{5}$$

avec la contrainte $\beta_1 = 0$ pour la catégorie de référence. Le rapport de probabilités (odds ratio) entre la catégorie j et la catégorie de référence est :

$$\frac{P(Y = j|X)}{P(Y = 1|X)} = \exp(X'\beta_j) \tag{6}$$

d'où :

$$\log\left(\frac{P(Y = j|X)}{P(Y = 1|X)}\right) = X'\beta_j = \beta_{0j} + \beta_{1j}X_1 + \beta_{2j}X_2 + \ldots + \beta_{pj}X_p \tag{7}$$

2.2.3 Application à notre étude

En considérant le niveau de vie « Faible » comme catégorie de référence, le modèle s'écrit :

$$\begin{aligned}\log\left(\frac{P(\text{Niveau\_vie} = \text{Moyen})}{P(\text{Niveau\_vie} = \text{Faible})}\right) = {} & \beta_{0,\text{Moyen}} + \beta_{1,\text{Moyen}}\text{Taille\_menage} + \\ & + \beta_{2,\text{Moyen}}\text{Sexe\_chef} + \beta_{3,\text{Moyen}}\text{Groupe\_age} \\ & + \beta_{4,\text{Moyen}}\text{Ethnie} + \beta_{5,\text{Moyen}}\text{Religion} \\ & + \beta_{6,\text{Moyen}}\text{Statut\_matrimonial} \\ & + \beta_{7,\text{Moyen}}\text{Niv\_instruction}\end{aligned} \tag{8}$$

$$\log\left(\frac{P(\text{Niveau\_vie} = \text{Élevé})}{P(\text{Niveau\_vie} = \text{Faible})}\right) = \beta_{0,\text{Élevé}} + \beta_{1,\text{Élevé}}\text{Taille\_menage} + \beta_{2,\text{Élevé}}\text{Sexe\_chef} + \beta_{3,\text{Élevé}}\text{Groupe\_age} + \beta_{4,\text{Élevé}}\text{Ethnie} + \beta_{5,\text{Élevé}}\text{Religion} + \beta_{6,\text{Élevé}}\text{Statut\_matrimonial} + \beta_{7,\text{Élevé}}\text{Niv\_instruction} \quad (9)$$

Chaque catégorie (Moyen et Élevé) possède son propre vecteur de coefficients β_j, permettant aux effets des variables explicatives de différer selon la modalité de la variable dépendante.

2.2.4 Interprétation des coefficients

Dans le modèle multinomial, un coefficient $\beta_{kj} > 0$ indique que l'augmentation de la variable X_k accroît la probabilité d'appartenir à la catégorie j plutôt qu'à la catégorie de référence. L'odds ratio correspondant est $\exp(\beta_{kj})$.

2.3 Choix du modèle approprié

Le choix entre le modèle ordonné et le modèle non ordonné repose sur plusieurs critères :

1. **Test de l'hypothèse des cotes proportionnelles** : Le test de Brant permet de vérifier si l'hypothèse des cotes proportionnelles est respectée. Si p-value < 0.05, l'hypothèse est rejetée et le modèle non ordonné est préférable.
2. **Nature de la variable dépendante** : Si l'ordre des catégories est fondamental pour l'interprétation, le modèle ordonné est plus pertinent.
3. **Critères d'ajustement** : Les critères AIC (Akaike Information Criterion) et BIC (Bayesian Information Criterion) permettent de comparer la qualité d'ajustement des modèles. Le modèle avec les valeurs les plus faibles est préféré.

Dans notre analyse, nous commencerons par estimer le modèle ordonné, puis nous testerons l'hypothèse des cotes proportionnelles. En cas de violation, nous estimerons le modèle non ordonné et comparerons les deux approches pour retenir celle qui décrit le mieux les déterminants des **Conditions de vie** des ménages.

3 Mise en œuvre des modèles

La mise en œuvre des modèles polytomiques pour l'analyse des déterminants du niveau de vie des ménages suit une démarche méthodologique rigoureuse. Cette section présente les différentes étapes, depuis la préparation des données jusqu'à la validation finale du modèle retenu.

3.0 Synthèse de la démarche méthodologique

Le tableau ci-dessous récapitule les étapes de mise en œuvre des modèles polytomiques :

Étape	Description	Outils/Tests
1	Préparation des données	Traitement des valeurs manquantes, détection des valeurs aberrantes
2	Analyse descriptive	Statistiques descriptives, tableaux croisés
3	Estimation du modèle ordonné	Maximum de vraisemblance (fonction `polr`)
4	Test des cotes proportionnelles	Test de Brant
5	Estimation du modèle non ordonné (si nécessaire)	Maximum de vraisemblance (fonction `multinom`)
6	Comparaison des modèles	AIC, BIC, test du rapport de vraisemblance
7	Validation du modèle	Matrice de confusion, pseudo R^2
8	Interprétation des résultats	Coefficients, odds ratios, effets marginaux

Table 3 – Synthèse des étapes de mise en œuvre des modèles

3.1 Préparation et exploration des données

3.1.1 Vérification de la qualité des données

Avant toute modélisation, il est essentiel de vérifier la qualité des données :

— **Traitement des valeurs manquantes** : Identification des variables comportant des données manquantes et choix d'une stratégie de traitement (suppression, imputation, etc.)
— **Détection des valeurs aberrantes** : Analyse des valeurs extrêmes qui

pourraient influencer les estimations
— **Vérification de la cohérence** : Contrôle de la cohérence logique entre les variables (par exemple, âge et statut matrimonial)

3.1.2 Analyse descriptive de la variable dépendante

L'analyse de la distribution de la variable « Niveau de vie » permet de :

— Vérifier l'équilibre entre les catégories (détection d'un éventuel déséquilibre majeur)
— Confirmer l'ordre naturel des modalités : Faible < Moyen < Élevé
— Calculer les effectifs et pourcentages pour chaque catégorie

3.1.3 Analyse descriptive des variables explicatives

Pour chaque variable explicative, on examine :

— **Variables continues** (Taille du ménage, Âge) : Calcul des statistiques descriptives (moyenne, médiane, écart-type, minimum, maximum)
— **Variables catégorielles** (Sexe, Ethnie, Religion, etc.) : Tableau de fréquences et proportions pour chaque modalité
— **Analyse bivariée** : Croisement de chaque variable explicative avec la variable dépendante pour identifier des associations préliminaires (si besoin se fait sentir)

3.2 Estimation des modèles

3.2.1 Estimation du modèle polytomique ordonné

Le modèle polytomique ordonné est estimé en premier lieu, car il est plus parcimonieux et adapté à la nature ordinale de la variable dépendante. L'estimation se fait par la méthode du maximum de vraisemblance. **Étapes d'estimation :**

1. Spécification du modèle avec toutes les variables explicatives
2. Estimation des coefficients β et des seuils α_j
3. Analyse de la significativité des coefficients (test de Wald, p-values)

Sous R :

```
library(MASS)
modele_ordonne <- polr(Conditions_vie ~ Taille_menage + Sexe_chef +
                       Groupe_age + Ethnie + Religion +
                       Statut_matrimonial + Niv_instruction,
                       data = donnees, Hess = TRUE)
summary(modele_ordonne)
```

3.2.2 Test de l'hypothèse des cotes proportionnelles

Le test de Brant permet de vérifier si l'hypothèse des cotes proportionnelles est respectée. Cette hypothèse suppose que l'effet des variables explicatives est constant pour toutes les transitions entre catégories.

Hypothèses du test :

— H_0 : L'hypothèse des cotes proportionnelles est respectée (le modèle ordonné est approprié)
— H_1 : L'hypothèse est violée (le modèle non ordonné est préférable)

Sous R :

```
library(brant)
brant_test <- brant(modele_ordonne)
print(brant_test)
```

Règle de décision :

— Si p-value > 0.05 : On conserve le modèle ordonné
— Si p-value < 0.05 : On estime le modèle non ordonné

3.2.3 Estimation du modèle polytomique non ordonné

Si le test de Brant conduit au rejet de l'hypothèse des cotes proportionnelles, on estime le modèle polytomique non ordonné (multinomial logit).

Étapes d'estimation :

1. Définition de la catégorie de référence (généralement « Faible »)
2. Estimation des $(J-1)$ vecteurs de coefficients
3. Analyse de la significativité des coefficients pour chaque équation
4. Calcul des effets marginaux ou odds ratios pour faciliter l'interprétation

Sous R :

```
library(nnet)
modele_non_ordonne <- multinom(Conditions_vi ~ Taille_menage + Sexe_chef
                               Groupe_age + Ethnie + Religion +
                               Statut_matrimonial + Niv_instruction,
                               data = donnees)
summary(modele_non_ordonne)
```

3.3 Comparaison et sélection du modèle

3.3.1 Critères d'information

La comparaison des modèles repose sur plusieurs critères d'information :

— **AIC (Akaike Information Criterion)** : Mesure la qualité du modèle en pénalisant la complexité

$$AIC = -2\log(L) + 2k \tag{10}$$

où L est la vraisemblance maximale et k le nombre de paramètres

— **BIC (Bayesian Information Criterion)** : Similaire à l'AIC mais avec une pénalité plus forte pour les modèles complexes

$$BIC = -2\log(L) + k\log(n) \tag{11}$$

où n est la taille de l'échantillon

Règle de décision : Le modèle avec les valeurs AIC et BIC les plus faibles est préféré.

3.3.2 Test du rapport de vraisemblance

Pour comparer les modèles emboîtés, on utilise le test du rapport de vraisemblance :

$$LR = -2[\log(L_0) - \log(L_1)] \sim \chi^2(df) \tag{12}$$

où L_0 est la vraisemblance du modèle contraint (ordonné) et L_1 celle du modèle libre (non ordonné).

3.3.3 Pseudo R-carré

Plusieurs mesures de pseudo R-carré permettent d'évaluer la qualité globale

de l'ajustement :

— **Pseudo R² de McFadden** :

$$R^2_{McFadden} = 1 - \frac{\log(L_M)}{\log(L_0)} \tag{13}$$

où L_M est la vraisemblance du modèle complet et L_0 celle du modèle nul
— **Pseudo R² de Nagelkerke** : Version ajustée du R² de Cox et Snell, variant entre 0 et 1

3.4 Validation du modèle retenu

3.4.1 Tests de significativité globale

Test de significativité globale du modèle :

Le test du rapport de vraisemblance compare le modèle complet au modèle nul (sans variables explicatives) :

— H_0 : Tous les coefficients sont nuls (le modèle n'apporte aucune information)
— H_1 : Au moins un coefficient est non nul

Un p-value < 0.05 indique que le modèle est globalement significatif.

3.4.2 Tests de significativité individuelle

Pour chaque variable explicative, on teste :

— $H_0 : \beta_k = 0$ (la variable n'a pas d'effet sur la variable dépendante)
— $H_1 : \beta_k \neq 0$ (la variable a un effet significatif)

Le test de Wald permet de déterminer la significativité de chaque coefficient. Les variables non significatives peuvent être retirées du modèle.

3.4.3 Analyse des prédictions

Matrice de confusion :

La matrice de confusion compare les catégories observées aux catégories prédites par le modèle. Elle permet de calculer :

— **Taux de bon classement global** : Proportion d'observations correctement classées
— **Taux de bon classement par catégorie** : Sensibilité pour chaque modalité

Sous R :

```
# Pour le modèle ordonné
predictions <- predict(modele_ordonne, type = "class")
table_confusion <- table(Observé = donnees$Niveau_vie,
                         Prédit = predictions)
taux_classement <- sum(diag(table_confusion)) / sum(table_confusion)
```

3.4.4 Vérification des hypothèses

Pour le modèle ordonné :

— Vérification de l'hypothèse des cotes proportionnelles (test de Brant)
— Absence de multicolinéarité entre les variables explicatives (VIF - Variance Inflation Factor)

Pour le modèle non ordonné :

— Indépendance des alternatives non pertinentes (IIA - Independence of Irrelevant Alternatives)
— Test de Hausman-McFadden pour vérifier l'hypothèse IIA
— Absence de multicolinéarité

3.5 Interprétation des résultats

3.5.1 Interprétation des coefficients

Pour le modèle ordonné :

Le signe du coefficient indique le sens de l'effet :

— $\beta_k > 0$: La variable augmente la probabilité d'appartenir aux catégories supérieures de niveau de vie
— $\beta_k < 0$: La variable diminue cette probabilité

Pour le modèle non ordonné : Chaque coefficient β_{kj} indique l'effet de

la variable X_k sur la probabilité d'appartenir à la catégorie j par rapport à la catégorie de référence.

3.5.2 Calcul et interprétation des odds ratios

L'odds ratio permet une interprétation plus intuitive des effets :

$$OR_k = \exp(\beta_k) \tag{14}$$

Interprétation :

- $OR_k > 1$: Une augmentation d'une unité de X_k multiplie les chances d'appartenir à une catégorie supérieure par OR_k
- $OR_k < 1$: Une augmentation de X_k réduit ces chances
- $OR_k = 1$: La variable n'a pas d'effet

3.5.3 Effets marginaux

Les effets marginaux permettent de quantifier l'impact d'une variation d'une variable explicative sur la probabilité d'appartenir à chaque catégorie de la variable dépendante. **Sous R :**

```
library(effects)
effets <- effect("Niv_instruction", modele_ordonne)
plot(effets)
```

Ces effets sont particulièrement utiles pour présenter les résultats de manière accessible aux décideurs. Cette démarche méthodologique garantit la rigueur de

l'analyse et la robustesse des résultats obtenus. Le choix final entre le modèle ordonné et le modèle non ordonné repose sur les tests statistiques et les critères d'ajustement, tout en tenant compte de la parcimonie et de l'interprétabilité du modèle.

4 Résultats

Cette section présente les résultats de l'estimation des modèles polytomiques appliqués aux conditions de des ménages. Nous procédons selon une démarche rigoureuse : estimation des deux modèles (ordonné et non ordonné), tests statistiques, comparaison, sélection du modèle le plus approprié, et validation des hypothèses.

Préparation et description générale des données

```
\begin{lstlisting}[language=R,basicstyle=\ttfamily\small]
# ============================================================
# CHARGEMENT DES PACKAGES
# ============================================================
library(MASS)
library(nnet)
library(brant)
library(car)
library(lmtest)
library(pROC)
library(DescTools)
library(effects)
library(dplyr)
library(knitr)
library(kableExtra)
library(ggplot2)
library(gridExtra)
library(tibble)

# ============================================================
# CHARGEMENT DE LA BASE DE DONNÉES
# ============================================================
donnees_raw <- Base_polytomique

# Statistiques générales
Nombre_variables <- ncol(donnees_raw)
Nombre_Individus <- nrow(donnees_raw)
Valeurs_Manquantes <- sum(is.na(donnees_raw))

# Création du dataframe récapitulatif
distribution_df <- data.frame(
  Nombre_Individus = as.numeric(Nombre_Individus),
  Nombre_variables = as.numeric(Nombre_variables),
  Valeurs_Manquantes = as.numeric(Valeurs_Manquantes),
  stringsAsFactors = FALSE
)
\end{lstlisting}
```

Table 4 – Description générale de la Base de données

Nombre Individus	Nombre variables	Valeurs Manquantes
1954	**9**	**0**

Source : Enquête ENSPD 2022

Le Tableau 4 présente une description générale de la base de données utilisée dans le cadre de cette étude. L'échantillon est composé de 1954 individus (ménages), ce qui constitue une taille d'échantillon substantielle permettant d'assurer la robustesse des analyses statistiques. La base de données comporte 9 variables qui renseignent sur les caractéristiques socio-économiques et démographiques des ménages enquêtés. Un aspect particulièrement notable de cette base de données est l'absence de valeurs manquantes (0 valeur manquante), ce qui témoigne de la qualité du processus de collecte des données et facilite grandement les analyses statistiques subséquentes. Cette complétude des données permet d'éviter les biais potentiels liés aux valeurs manquantes et garantit que l'ensemble des observations peut être exploité dans la modélisation des déterminants des conditions socio-économiques de vie des ménages. Les données proviennent de l'enquête 2022 menée par l'ENSPD (sur la Situation des Personnes et Ménages a Parakou).

4.1.1 Nettoyage et recodage des variables

Table 5 – Synthèse des variables pour l'analyse

Variable	Type	Modalités
Condition_vie	**Ordinale**	**Faible, Moyen, Eleve**
Sexe_chef	Nominale	Masculin, Feminin
Taille_menage	Ordinale	1 personne, 2-3 personnes, 4-5 personnes, 6-8 personnes, 9+ personnes
Groupe_age	Ordinale	15-29 ans, 30-44 ans, 45-59 ans, 60+ ans
Ethnie	Nominale	9 modalités
Religion	Nominale	4 modalités
Statut_matrimonial	Nominale	Celibataire, Marie(e) monogame, Marie(e) polygame, Divorce(e), Veuf/Veuve, Union libre
Niv_instruction	Ordinale	Aucun, Primaire, Secondaire, Superieur

4.2 Analyse descriptive

Statistiques descriptives des variables quantitatives

```
\begin{lstlisting}[language=R,basicstyle=\ttfamily\small]
# ============================================================
# CALCUL DES STATISTIQUES DESCRIPTIVES
# ============================================================

stats_quantitatives <- donnees_complete %>%
  summarise(
    # Taille du ménage
    Taille_Moyenne = round(mean(Taille_menage_brut, na.rm = TRUE), 2),
    Taille_Min = min(Taille_menage_brut, na.rm = TRUE),
    Taille_Max = max(Taille_menage_brut, na.rm = TRUE),
    Taille_EcartType = round(sd(Taille_menage_brut, na.rm = TRUE), 2),
    Taille_Mediane = median(Taille_menage_brut, na.rm = TRUE),

    # Âge du chef de ménage
    Age_Moyenne = round(mean(Age_chef, na.rm = TRUE), 2),
    Age_Min = min(Age_chef, na.rm = TRUE),
    Age_Max = max(Age_chef, na.rm = TRUE),
    Age_EcartType = round(sd(Age_chef, na.rm = TRUE), 2),
    Age_Mediane = median(Age_chef, na.rm = TRUE)
  )

# Reformater pour un tableau vertical
stats_tableau <- data.frame(
  Variable = c("Taille du menage", "Age du chef de menage"),
  Moyenne = c(stats_quantitatives$Taille_Moyenne,
              stats_quantitatives$Age_Moyenne),
  Mediane = c(stats_quantitatives$Taille_Mediane,
              stats_quantitatives$Age_Mediane),
  Ecart_type = c(stats_quantitatives$Taille_EcartType,
                 stats_quantitatives$Age_EcartType),
  Minimum = c(stats_quantitatives$Taille_Min,
              stats_quantitatives$Age_Min),
  Maximum = c(stats_quantitatives$Taille_Max,
              stats_quantitatives$Age_Max)
)
\end{lstlisting}
```

Table 6 – Statistiques descriptives des variables quantitatives

Variable	Moyenne	Mediane	Ecart-type	Min	Max
Taille du menage	2.77	2	1.89	1	23
Age du chef de menage	34.69	30	14.64	15	95

Le Tableau 6 présente les statistiques descriptives des deux variables quantitatives de notre étude. Concernant la taille du ménage, on observe qu'en moyenne, les ménages de l'échantillon comptent 2,77 personnes, avec une médiane de 2 personnes. Cette proximité entre la moyenne et la médiane suggère une distribution relativement symétrique. L'écart-type de 1,89 indique une variabilité modérée dans la composition des ménages. La taille des ménages varie de 1 personne (ménages unipersonnels) à 23 personnes (ménages de très grande taille), reflétant ainsi une grande diversité dans la structure familiale des ménages enquêtés. Quant à l'âge du chef de ménage, la moyenne se situe à 34,69 ans, tandis que la médiane est de 30 ans. L'écart entre ces deux indicateurs suggère une légère asymétrie positive de la distribution, avec la présence de quelques chefs de ménage d'âge plus avancé qui tirent la moyenne vers le haut. L'écart-type de 14,64 ans témoigne d'une dispersion importante des âges. L'échantillon couvre une large tranche d'âge, allant de 15 ans (jeunes chefs de ménage) à 95 ans (chefs de ménage très âgés), ce qui permet d'analyser les conditions socio-économiques à différentes étapes du cycle de vie. Ces statistiques révèlent une hétérogénéité importante au sein de l'échantillon, tant en termes de structure des ménages que d'âge des chefs de ménage, ce qui est essentiel pour une analyse approfondie des déterminants des conditions de vie.

Visualisation des distributions

```
\begin{lstlisting}[language=R,basicstyle=\ttfamily\small]
# ============================================================
# GRAPHIQUES DE DISTRIBUTION
# ============================================================

par(mfrow = c(1, 2))

# Graphique 1 : Taille du ménage
boxplot(donnees_complete$Taille_menage_brut,
        main = "Taille du menage",
        ylab = "Nombre de personnes",
        col = "#3498DB",
        boxwex = 0.8)
points(1, mean(donnees_complete$Taille_menage_brut, na.rm = TRUE),
       pch = 18, col = "red", cex = 2)

# Graphique 2 : Âge du chef de ménage
boxplot(donnees_complete$Age_chef,
        main = "Age du chef de menage",
        ylab = "Age (annees)",
        col = "#E67E22")
points(1, mean(donnees_complete$Age_chef, na.rm = TRUE),
       pch = 18, col = "red", cex = 2)
\end{lstlisting}
```

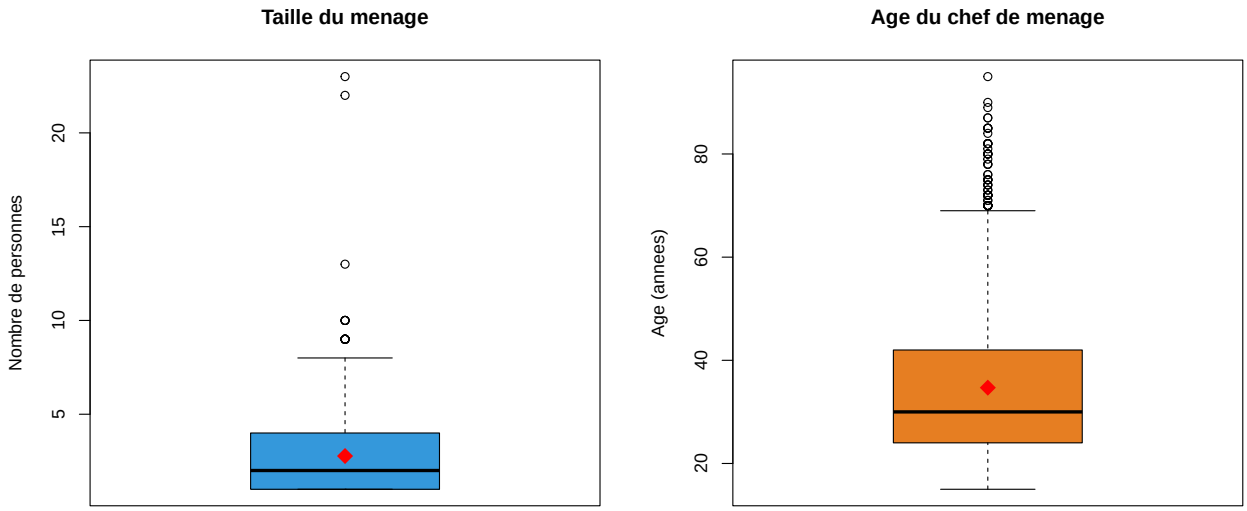


FIGURE 1 – Distribution des variables quantitatives

Visualisation des distributions

```
\begin{lstlisting}[language=R,basicstyle=\ttfamily\small]
par(mfrow = c(1, 1))
\end{lstlisting}
```

La Figure 1 présente la distribution des deux variables quantitatives de l'étude à travers des diagrammes en boîtes (boxplots). Le point rouge représente la moyenne de chaque variable. Pour la taille du ménage (graphique de gauche), on observe une distribution asymétrique avec une médiane proche de 3 personnes. La boîte à moustaches révèle que 50% des ménages comptent entre environ 2 et 5 personnes. La présence de plusieurs valeurs aberrantes (points au-dessus de la moustache supérieure) indique l'existence de ménages de très grande taille, allant jusqu'à plus de 20 personnes. La moyenne (point rouge) se situe légèrement au-dessus de la médiane, confirmant l'asymétrie positive de cette distribution, due à ces ménages de grande taille qui tirent la moyenne vers le haut. Concernant l'âge du chef de ménage (graphique de droite), la distribution apparaît également asymétrique avec une médiane autour de 30 ans. On note la présence de plusieurs valeurs extrêmes, notamment des chefs de ménage très âgés (au-delà de 80 ans). L'écart entre la moyenne (point rouge) et la médiane confirme l'asymétrie positive observée dans les statistiques descriptives, liée à la présence de chefs de ménage d'âge avancé. Ces visualisations mettent en évidence l'hétérogénéité importante des ménages enquêtés, tant en termes de composition que d'âge du chef de ménage, justifiant ainsi l'intérêt d'une analyse multivariée pour identifier les déterminants des conditions socio-économiques.

Table 7 – Conditions de vie selon le sexe du chef de ménage

Sexe	Conditions de vie	Effectif	\%
Masculin	Faible	131	8.81
Masculin	Moyen	957	64.36
Masculin	Eleve	399	26.83
Feminin	Faible	42	8.99
Feminin	Moyen	294	62.96
Feminin	Eleve	131	28.05

Table 8 – Conditions de vie selon la taille du ménage

Taille	Conditions de vie	Effectif	\%
1 personne	Faible	50	8.85
1 personne	Moyen	383	67.79
1 personne	Eleve	132	23.36
2-3 personnes	Faible	75	8.87
2-3 personnes	Moyen	553	65.37
2-3 personnes	Eleve	218	25.77
4-5 personnes	Faible	25	6.49
4-5 personnes	Moyen	236	61.30
4-5 personnes	Eleve	124	32.21
6-8 personnes	Faible	19	13.77
6-8 personnes	Moyen	73	52.90
6-8 personnes	Eleve	46	33.33
9+ personnes	Faible	4	20.00
9+ personnes	Moyen	6	30.00
9+ personnes	Eleve	10	50.00

Tests d'indépendance du Chi-deux

```
\begin{lstlisting}[language=R,basicstyle=\ttfamily\small]
# ============================================================
# TESTS D'INDÉPENDANCE DU CHI deux
# ============================================================

calc_chi2 <- function(var) {
  tryCatch({
    test <- chisq.test(table(donnees_complete$Condition_vie, var))
    return(data.frame(
      Statistique = round(as.numeric(test$statistic), 2),
      Degres_liberte = as.numeric(test$parameter),
      p_value = round(as.numeric(test$p.value), 4),
      Significatif = dplyr::case_when(
        test$p.value < 0.001 ~ "*** (p<0.001)",
        test$p.value < 0.01  ~ "** (p<0.01)",
        test$p.value < 0.05  ~ "* (p<0.05)",
        test$p.value < 0.10  ~ ". (p<0.10)",
        TRUE ~ "Non significatif"
      ),
      stringsAsFactors = FALSE
    ))
  }, error = function(e) {
    return(data.frame(
      Statistique = NA,
      Degres_liberte = NA,
      p_value = NA,
      Significatif = "Erreur calcul",
      stringsAsFactors = FALSE
    ))
  })
}

# Calculer les tests pour chaque variable
tests_chi2 <- bind_rows(
  calc_chi2(donnees_complete$Sexe_chef) %>% mutate(Variable = "Sexe chef"),
  calc_chi2(donnees_complete$Taille_menage) %>% mutate(Variable = "Taille menage"),
  calc_chi2(donnees_complete$Groupe_age) %>% mutate(Variable = "Groupe d'age"),
  calc_chi2(donnees_complete$Statut_matrimonial) %>%
    mutate(Variable = "Statut matrimonial"),
  calc_chi2(donnees_complete$Niv_instruction) %>%
    mutate(Variable = "Niveau instruction"),
  calc_chi2(donnees_complete$Religion) %>% mutate(Variable = "Religion"),
  calc_chi2(donnees_complete$Ethnie) %>% mutate(Variable = "Ethnie")
) %>%
  dplyr::select(Variable, Statistique, Degres_liberte, p_value, Significatif)
\end{lstlisting}
```

Table 9 – Tests d'independance du Chi-deux entre variables et conditions de vie

Variable	Chi-deux	ddl	p-value	Significativite
Sexe chef	0.32	2	0.8532	Non significatif
Taille menage	30.22	8	0.0002	*** (p<0.001)
Groupe d'age	143.54	6	0.0000	*** (p<0.001)
Statut matrimonial	21.06	10	0.0207	* (p<0.05)
Niveau instruction	66.83	6	0.0000	*** (p<0.001)
Religion	8.11	6	0.2299	Non significatif
Ethnie	31.19	16	0.0127	* (p<0.05)

Note : Seuils de significativite : *** p inferieur 0.001, ** p inferieur 0.01, * p inferieur 0.05, . p inferieur 0.10

Variables significativement associees aux conditions de vie (p inferieur 0.05)

- **Taille menage***** (p = 0.0002)
- **Groupe d'age***** (p = 0)
- **Statut matrimonial*** (p = 0.0207)
- **Niveau instruction***** (p = 0)
- **Ethnie*** (p = 0.0127)

Variables non significatives (p superieur ou egal 0.05)

- **Sexe chef** (p = 0.8532)
- **Religion** (p = 0.2299)

Legende : *** p inferieur 0.001, ** p inferieur 0.01, * p inferieur 0.05, $^{\cdot}$ p inferieur 0.10

graphiques-distribution generale

```
\begin{lstlisting}[language=R,basicstyle=\ttfamily\small]
# Créer une mosaïque de graphiques
library(patchwork)

# Tableau de fréquences
freq_condition_vie <- donnees_complete %>%
  count(Condition_vie) %>%
  mutate(
    Pourcentage = round(n / sum(n) * 100, 2),
    Pct_cumule = cumsum(Pourcentage)
  )

tab_sexe <- donnees_complete %>%
  count(Sexe_chef, Condition_vie) %>%
  group_by(Sexe_chef) %>%
  mutate(Pourcentage = round(n / sum(n) * 100, 2)) %>%
  ungroup()

# Graphique 1 : Conditions de vie (camembert) avec labels
g1 <- ggplot(freq_condition_vie, aes(x = "", y = Pourcentage, fill = Condition_vie)) +
  geom_col(width = 1, color = "white") +
  coord_polar("y") +
  scale_fill_manual(values = c("Faible" = "#E74C3C", "Moyen" = "#F39C12", "Eleve" = "#27AE60")) +
  geom_text(aes(label = paste0(Pourcentage, "%")),
            position = position_stack(vjust = 0.5),
            color = "white", size = 5, fontface = "bold", check_overlap = TRUE) +
  labs(title = "Conditions de vie", fill = "") +
  theme_void() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))

# Graphique 2 : Sexe (camembert) avec labels
g2 <- donnees_complete %>%
  count(Sexe_chef) %>%
  mutate(pct = round(n/sum(n)*100, 1)) %>%
  ggplot(aes(x = "", y = pct, fill = Sexe_chef)) +
  geom_col(width = 1, color = "white") +
  coord_polar("y") +
  scale_fill_manual(values = c("Masculin" = "#3498DB", "Feminin" = "#E91E63")) +
  geom_text(aes(label = paste0(pct, "%")),
            position = position_stack(vjust = 0.5),
            color = "white", size = 5, fontface = "bold", check_overlap = TRUE) +
  labs(title = "Sexe du chef", fill = "") +
  theme_void() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))

# Graphique 3 : Niveau instruction (barres) avec labels au-dessus
g3 <- donnees_complete %>%
  count(Niv_instruction) %>%
  mutate(pct = round(n/sum(n)*100, 1)) %>%
  ggplot(aes(x = Niv_instruction, y = pct, fill = Niv_instruction)) +
  geom_col() +
  geom_text(aes(label = paste0(pct, "%")),
            vjust = -0.3, size = 4, fontface = "bold") +
  scale_fill_brewer(palette = "YlOrRd") +
  labs(title = "Niveau d'instruction", x = "", y = "%") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"),
        legend.position = "none",
        axis.text.x = element_text(angle = 45, hjust = 1))

# Graphique 4 : Taille ménage (barres) avec labels au-dessus
g4 <- donnees_complete %>%
```

```
  count(Taille_menage) %>%
  mutate(pct = round(n/sum(n)*100, 1)) %>%
  ggplot(aes(x = Taille_menage, y = pct, fill = Taille_menage)) +
  geom_col() +
  geom_text(aes(label = paste0(pct, "%")),
            vjust = -0.3, size = 4, fontface = "bold") +
  scale_fill_brewer(palette = "Blues") +
  labs(title = "Taille du menage", x = "", y = "%") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"),
        legend.position = "none",
        axis.text.x = element_text(angle = 45, hjust = 1))

# Combiner les graphiques
(g1 | g2) / (g3 | g4) +
  plot_annotation(title = "Vue d'ensemble de la distribution des variables",
                  theme = theme(plot.title = element_text(hjust = 0.5, face = "bold", size = 16)))
\end{lstlisting}
```

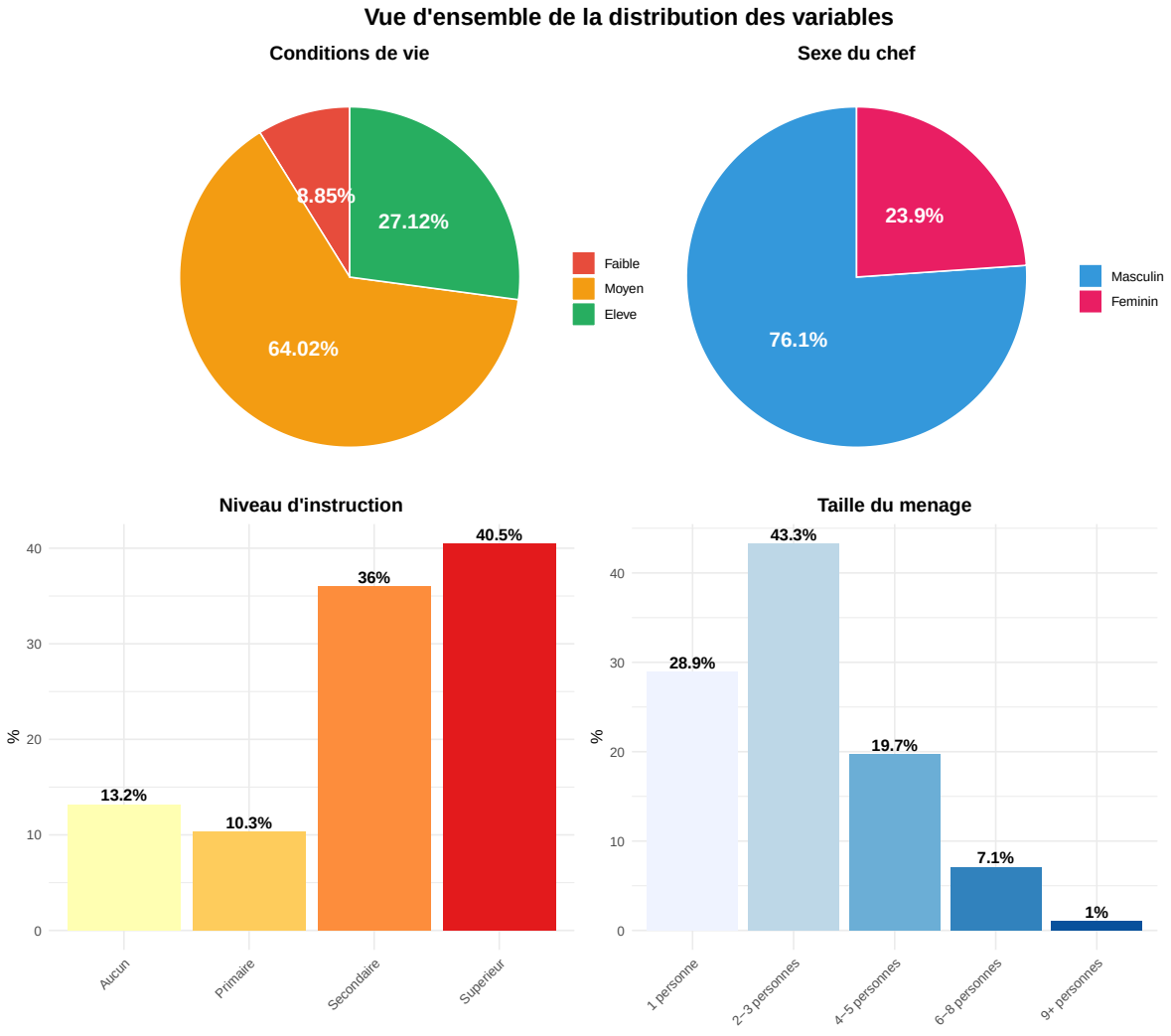


FIGURE 2 – Vue d'ensemble de la distribution des variables

Tout est prêt pour les estimations

Déterminants des conditions socio-économique de la vie

4.3 Estimation du modèle polytomique ordonné

Modele full

```
\begin{lstlisting}[language=R,basicstyle=\ttfamily\small]
donnees_complete <- donnees_complete %>%
  mutate(
    Groupe_age = factor(as.character(Groupe_age),
                        levels = c("15-29 ans","30-44 ans","45-59 ans","60+ ans"),
                        ordered = FALSE),
    Niv_instruction = factor(as.character(Niv_instruction),
                             levels = c("Aucun","Primaire","Secondaire","Superieur"),
                             ordered = FALSE),
    Taille_menage = factor(as.character(Taille_menage),
                           levels = c("1 personne",
                                      "2-3 personnes",
                                      "4-5 personnes",
                                      "6-8 personnes",
                                      "9+ personnes"),
                           ordered = FALSE)
  )

# 1) Définir la formule complète
formule_full <- as.formula("Condition_vie ~ Taille_menage +
                           Groupe_age +
                           Ethnie +
                           Statut_matrimonial +
                           Niv_instruction")

# 2) Ajuster le modèle complet (point de départ pour stepAIC)
modele_full <- polr(formule_full, data = donnees_complete, Hess = TRUE, method = "logistic")

# 3) Lancer stepAIC directement (direction both)
set.seed(123) # reproductibilité
modele_step <- MASS::stepAIC(modele_full, direction = "both", trace = FALSE)
\end{lstlisting}
```

Table 10 – Modele final (selection stepAIC)

Terme	Estimate	Std Error	t value	p-value	Signif	OR	IC 95 pct
Groupe_age30-44 ans	1.0446	0.1169	8.9383	0.00000	***	2.842	2.260 - 3.574
Groupe_age45-59 ans	2.0694	0.1638	12.6314	0.00000	***	7.920	5.745 - 10.919
Groupe_age60+ ans	2.3122	0.1917	12.0624	0.00000	***	10.096	6.934 - 14.700
EthnieBariba et apparentes	0.0023	0.2360	0.0095	0.99240		1.002	0.631 - 1.592
EthnieDendi et apparentes	0.4204	0.2835	1.4831	0.13800		1.523	0.874 - 2.654
EthnieEtranger	0.2827	0.3168	0.8925	0.37210		1.327	0.713 - 2.468
EthnieFon et apparentes	0.0992	0.2358	0.4207	0.67400		1.104	0.696 - 1.753
EthnieOttamari et apparentes	-0.4893	0.2709	-1.8064	0.07086		0.613	0.361 - 1.042
EthniePeulh et apparentes	0.3308	0.3527	0.9379	0.34830		1.392	0.697 - 2.779
EthnieYao et apparentes	-0.2427	0.3864	-0.6280	0.53000		0.785	0.368 - 1.673
EthnieYoruba/Nago et apparentes	0.0025	0.2438	0.0104	0.99170		1.003	0.622 - 1.617
Niv_instructionPrimaire	0.2491	0.2028	1.2282	0.21940		1.283	0.862 - 1.909
Niv_instructionSecondaire	1.2408	0.1649	7.5260	0.00000	***	3.458	2.503 - 4.777
Niv_instructionSuperieur	2.0203	0.1741	11.6029	0.00000	***	7.541	5.361 - 10.608
Faible|Moyen	-0.5249	0.2653	-1.9787	0.04785	*		
Moyen|Eleve	3.2584	0.2771	11.7600	0.00000	***		

Avant d'interpréter les coefficients et de tirer des conclusions politiques ou économiques, il est nécessaire de s'assurer que le modèle estimé est bien adapté aux données. En particulier, le modèle logit ordinal repose sur l'hypothèse des cotes proportionnelles (proportional odds) : l'effet d'une covariable sur le log-odds est supposé identique pour toutes les coupures entre catégories (Faible Moyen et Moyen Élevé). Si cette hypothèse est violée, les coefficients fournis par le modèle polr ne décrivent pas correctement l'effet des variables et l'on risque des interprétations trompeuses.

C'est pourquoi nous effectuons systématiquement le test de Brant sur le modèle ordonné estimé

si la p-value globale superieur à 0.05, l'hypothèse des cotes proportionnelles n'est pas rejetée et le modèle ordonné (polr) est jugé approprié : on peut alors interpréter les coefficients et présenter les odds ratios et effets marginaux ;

si la p-value globale inférieur ou égale à 0.05, l'hypothèse est rejetée : nous n'interprétons pas le modèle ordonné tel quel. Dans ce cas, nous estimons un modèle non ordonné (multinomial) ou, si pertinent, un modèle partiellement parallèle permettant que certaines variables violent la contrainte tandis que d'autres la respectent.

Remarque méthodologique : le test de Brant peut être sensible aux modalités rares et à la spécification des variables. Si seule une ou deux variables violent la contrainte, une solution pragmatique consiste à estimer un modèle partiellement

non parallèle (ou à reformuler la variable).

Test de Brant en mode silence

```
\begin{lstlisting}[language=R,basicstyle=\ttfamily\small]
# ----------------------------
# Exécuter brant() SANS afficher sa sortie et extraire p-values
# Résultats disponibles :
#  - df_pvalues : data.frame(Terme, p_value, p_display)
#  - p_global   : p-value Omnibus (numérique ou NA)
# ----------------------------

# exécuter brant() en silence (capture stdout et messages)
tempf <- tempfile()
con <- file(tempf, open = "w")
sink(con)                 # redirige stdout
sink(con, type = "message") # redirige messages
brant_res <- tryCatch(
  brant::brant(modele_step),
  error = function(e) e
)
sink(type = "message")    # restaurer message sink
sink()                    # restaurer stdout
close(con)
unlink(tempf)             # supprimer fichier temporaire
\end{lstlisting}
```

Table 11 – P-values et significativité

Variable	p-value	Signif.
Omnibus	<0.001	***
Groupe_age30-44 ans	<0.001	***
Groupe_age45-59 ans	0.0114	*
Groupe_age60+ ans	0.0111	*
EthnieBariba et apparentes	0.5570	
EthnieDendi et apparentes	0.7592	
EthnieEtranger	0.7763	
EthnieFon et apparentes	0.3766	
EthnieOttamari et apparentes	0.4434	
EthniePeulh et apparentes	0.3425	
EthnieYao et apparentes	0.3851	
EthnieYoruba/Nago et apparentes	0.6997	
Niv_instructionPrimaire	0.5523	
Niv_instructionSecondaire	0.1298	
Niv_instructionSuperieur	0.9701	

Test de Brant et hypothèse de proportionnalité des odds

Le **test de Brant** (ou test de parallélisme des pentes) a été appliqué afin de vérifier l'hypothèse de proportionnalité des odds, condition essentielle dans les modèles de régression logistique ordinale de type *proportional odds*.

Résultats globaux

La p-value globale obtenue est $<$ **0,001**, indiquant un rejet significatif de l'hypothèse nulle de proportionnalité des coefficients.
Autrement dit, **l'hypothèse de pentes parallèles n'est pas respectée** pour l'ensemble du modèle.

Analyse par variable

L'examen des p-values spécifiques par variable permet d'identifier les prédicteurs contribuant à cette non-proportionnalité.
Dans le présent modèle, la variable **Groupe_age 30–44 ans** présente également une **p-value $<$ 0,001**, suggérant qu'elle participe à la violation de l'hypothèse de pentes parallèles.

Conséquences et options

Plusieurs stratégies peuvent être envisagées pour traiter cette violation :

1. **Modélisation partielle**
   Permettre des pentes non parallèles pour certaines variables.
   — Exemple : `VGAM::vglm()` ou `ordinal::clm()` avec l'option `nominal = TRUE`.
2. **Modèle multinomial**
   Recourir à un modèle multinomial (`nnet::multinom()`), qui ne repose pas sur l'hypothèse de proportionnalité des odds.
3. **Simplification du modèle**
   Parfois, il est possible de simplifier le modèle ou de regrouper certaines catégories de la variable réponse, si cela est conceptuellement justifié.

modele multinomiale

```
\begin{lstlisting}[language=R,basicstyle=\ttfamily\small]
# ---------------------------
# MODELE MULTINOMIAL + stepAIC + tableau compact (OR, IC95, p-values)
# Version corrigée pour éviter l'erreur sur les noms de colonnes (z_value / p_value)
# ---------------------------

# 0) S'assurer que la réponse est un FACTEUR non ordonné (référence = premier niveau)
donnees_complete <- donnees_complete %>%
  mutate(
    Condition_vie = relevel(factor(as.character(Condition_vie), ordered = FALSE),
                            ref = "Faible"),
    Groupe_age = factor(as.character(Groupe_age),
                        levels = c("15-29 ans","30-44 ans","45-59 ans","60+ ans"),
                        ordered = FALSE),
    Niv_instruction = factor(as.character(Niv_instruction),
                             levels = c("Aucun","Primaire","Secondaire","Superieur"),
                             ordered = FALSE),
    Taille_menage = factor(as.character(Taille_menage),
                           levels = c("1 personne","2-3 personnes","4-5 personnes","6-8 personnes"
    ,"9+ personnes"),
                           ordered = FALSE)
  )



# 1) Formule (même que pour polr)
formule_full <- as.formula("Condition_vie ~ Taille_menage + Groupe_age + Ethnie + Statut_
    matrimonial + Niv_instruction")

# 2) Ajuster le modèle multinomial complet (point de départ pour stepAIC)
set.seed(123)
modele_full_multinom <- nnet::multinom(formule_full, data = donnees_complete, Hess = TRUE, trace =
     FALSE)


# 3) Lancer stepAIC (direction both) pour sélection
set.seed(123)
modele_step_multinom <- MASS::stepAIC(modele_full_multinom, direction = "both", trace = FALSE)
\end{lstlisting}
```

Table 12 – Modèle multinomial final (sélection stepAIC)

Terme	Estimate	Std. Error	z value	p-value	Signif.	OR	IC 95%
Eleve — (Intercept)	-2.0893	0.3247	-6.4347	0.0000000	***	0.124	0.065 – 0.234
Eleve — Taille_menage2-3 personnes	-0.1342	0.2294	-0.5853	0.5584000		0.874	0.558 – 1.371
Eleve — Taille_menage4-5 personnes	0.1721	0.2972	0.5790	0.5626000		1.188	0.663 – 2.127
Eleve — Taille_menage6-8 personnes	-0.5865	0.3524	-1.6644	0.0960400		0.556	0.279 – 1.110
Eleve — Taille_menage9+ personnes	-1.1069	0.6749	-1.6401	0.1010000		0.331	0.088 – 1.241
Eleve — Groupe_age30-44 ans	1.5428	0.2282	6.7597	0.0000000	***	4.677	2.990 – 7.316
Eleve — Groupe_age45-59 ans	3.1011	0.3544	8.7508	0.0000000	***	22.223	11.095 – 44.509
Eleve — Groupe_age60+ ans	3.4043	0.4019	8.4712	0.0000000	***	30.094	13.690 – 66.154
Eleve — Niv_instructionPrimaire	0.3734	0.3437	1.0864	0.2773000		1.453	0.741 – 2.849
Eleve — Niv_instructionSecondaire	1.9051	0.2911	6.5446	0.0000000	***	6.720	3.798 – 11.890
Eleve — Niv_instructionSuperieur	3.6158	0.3356	10.7754	0.0000000	***	37.182	19.262 – 71.774
Moyen — (Intercept)	1.0512	0.2514	4.1819	0.0000289	***	2.861	1.748 – 4.683
Moyen — Taille_menage2-3 personnes	-0.0280	0.2010	-0.1394	0.8892000		0.972	0.656 – 1.442
Moyen — Taille_menage4-5 personnes	0.1993	0.2691	0.7405	0.4590000		1.221	0.720 – 2.069
Moyen — Taille_menage6-8 personnes	-0.7298	0.3135	-2.3282	0.0199000	*	0.482	0.261 – 0.891
Moyen — Taille_menage9+ personnes	-1.9869	0.6857	-2.8975	0.0037610	**	0.137	0.036 – 0.526
Moyen — Groupe_age30-44 ans	0.2945	0.2008	1.4670	0.1424000		1.343	0.906 – 1.990
Moyen — Groupe_age45-59 ans	0.9696	0.3299	2.9394	0.0032890	**	2.637	1.381 – 5.034
Moyen — Groupe_age60+ ans	1.0169	0.3721	2.7327	0.0062810	**	2.765	1.333 – 5.733
Moyen — Niv_instructionPrimaire	0.0263	0.2724	0.0965	0.9231000		1.027	0.602 – 1.751
Moyen — Niv_instructionSecondaire	0.6387	0.2346	2.7225	0.0064800	**	1.894	1.196 – 3.000
Moyen — Niv_instructionSuperieur	1.7312	0.2806	6.1686	0.0000000	***	5.648	3.258 – 9.789

Le **Tableau 12** présente les résultats du modèle multinomial final obtenu après une sélection stepwise basée sur le critère AIC. Le modèle inclut les coefficients estimés, leurs erreurs standard, les statistiques z, les p-values, les odds ratios (OR) et leurs intervalles de confiance à 95%. La catégorie de référence pour la variable dépendante est **IC95** (niveau de vie faible).

Facteurs associés à un niveau de vie élevé (vs Faible)

Plusieurs variables se révèlent significativement associées à un niveau socio-économique élevé :

L'âge du chef de ménage (tous significatifs ***) :

Un chef de 30-44 ans a 4,68 fois plus de chances d'avoir un niveau de vie élevé qu'un chef de 15-29 ans Un chef de 45-59 ans a 22,22 fois plus de chances Un chef de 60 ans et plus a 30,09 fois plus de chances

Message simple : Plus le chef de ménage est âgé, plus ses chances d'être riche augmentent énormément. À 60 ans, vous avez 30 fois plus de chances qu'à 20 ans ! Le niveau d'instruction (significatif pour secondaire*** et supérieur***) :

Avoir le niveau primaire ne change pas significativement les chances (OR = 1,45, non significatif) Avoir le niveau secondaire multiplie les chances par 6,72 (***) Avoir le niveau supérieur multiplie les chances par 37,18 (***)

Message simple : L'éducation est très importante. Quelqu'un avec un diplôme universitaire a 37 fois plus de chances d'être riche qu'une personne sans éducation. C'est le facteur le plus puissant !

La taille du ménage :

Aucun effet significatif observé pour cette variable

Facteurs associés à un niveau de vie moyen (vs Faible)

La taille du ménage (significatif seulement pour grandes familles) :

Les ménages de 6-8 personnes ont 0,48 fois les chances d'être "Moyen" (*), soit 2 fois moins de chances Les ménages de 9+ personnes ont 0,14 fois les chances d'être "Moyen" (**), soit 7 fois moins de chances

Message simple : Plus la famille est nombreuse, plus il est difficile d'avoir un niveau de vie moyen. Les très grandes familles (9+ personnes) ont 7 fois moins de chances. L'âge du chef de ménage (significatif pour âges avancés) :

Les chefs de 45-59 ans ont 2,64 fois plus de chances d'être "Moyen" (**) Les chefs de 60+ ans ont 2,77 fois plus de chances d'être "Moyen"** (**)

Message simple : Les personnes plus âgées ont plus de chances d'être au moins dans la classe moyenne (et encore plus dans la classe élevée comme vu avant). Le niveau d'instruction (significatif pour secondaire** et supérieur***) :

Le niveau primaire ne change rien (non significatif) Le niveau secondaire multiplie les chances par 1,89 (**) Le niveau supérieur multiplie les chances par 5,65 (***)

Message simple : L'éducation augmente aussi les chances d'être dans la classe moyenne. Mais les personnes très éduquées vont surtout vers la classe élevée (OR = 37 vu plus haut). Ce qu'il faut retenir

L'éducation supérieure est le facteur le plus puissant : Elle multiplie par 37 vos chances d'avoir un bon niveau de vie. C'est énorme ! L'âge joue beaucoup : À 60 ans, vous avez 30 fois plus de chances d'être riche qu'à 20 ans. Il faut du temps pour accumuler des ressources. Les grandes familles sont très désavantagées :

Avoir 9 personnes ou plus dans le ménage divise par 7 vos chances d'avoir un niveau de vie moyen. Plus il y a de bouches à nourrir, plus c'est difficile.

Interprétation générale

Ces résultats mettent en évidence trois déterminants majeurs des conditions socio-économiques des ménages : **l'éducation du chef de ménage** (facteur le plus déterminant), **l'âge/expérience** et la **composition du ménage**. L'effet particulièrement fort de l'instruction supérieure souligne l'importance du capital humain dans l'amélioration des conditions de vie. La significativité statistique élevée (*** pour p $<$ 0,001) de la plupart des variables confirme la robustesse du modèle.

4.4 Comparaison des modèles

4.4.1 Critères d'information (AIC et BIC)

Comparaison initiale des modèles Les métriques globales (log-vraisemblance, AIC, BIC) et le pseudo-R^2 (McFadden) pour les deux spécifications (modèle ordinal PO estimé par `polr` et modèle multinomial estimé par `multinom`) sont présentées dans le Tableau~14 ci-dessous.

Table 13 – Comparaison : logLik / AIC / BIC / McFadden / Nagelkerke

Critere	modele_step (m1)	modele_step_multinom (m2)
logLik	-1505.890	-1494.384
AIC	3043.779	3032.769
BIC	3133.022	3155.477
McFadden	0.0976	0.1045
Nagelkerke	0.1876	0.1997

Table 14 – Test du rapport de vraisemblance (LR) et p-value globale du test de Brant

Comparaison	LR_stat	df	p_value
m1 vs m2 (LR)	23.011	6	<0.001
Test de Brant (global)	—	—	<0.001

Pour choisir entre le modèle ordonné (m1, `modele_step`) et le modèle multinomial (m2, `modele_step_multinom`), nous avons comparé plusieurs critères : log-vraisemblance, AIC, BIC, pseudo-R^2 (McFadden et Nagelkerke) et le test du rapport de vraisemblance (LR). Les résultats principaux sont résumés dans le tableau ci-dessous :

— logLik : m1 = -1505.890 ; m2 = -1494.384.
— AIC : m1 = 3043.779 ; m2 = 3032.769 (AIC favorise m2).
— BIC : m1 = 3133.022 ; m2 = 3155.477 (BIC favorise m1).
— McFadden : m1 = 0.0976 ; m2 = 0.1045. (McFadden $= 1 - \frac{\ell_{mod}}{\ell_{null}}$, valeur plus élevée = meilleur pouvoir explicatif relatif.)
— Nagelkerke : m1 = 0.1876 ; m2 = 0.1997.

Le test du rapport de vraisemblance donne LR = 23.011 avec df = 6 ($p < 0.001$), indiquant que le modèle multinomial améliore significativement l'ajustement par rapport au modèle ordonné, sous la condition que les modèles soient nestés. Par ailleurs, le test de Brant (global) donne $p < 0.001$, ce qui signifie un rejet de l'hypothèse d'odds proportionnels — autrement dit, la contrainte d'effet constant des covariables entre modalités n'est pas vérifiée.

Au vu de ces éléments, et en particulier du rejet très significatif de l'hypothèse d'odds proportionnels, nous retenons le modèle multinomial (m2) pour l'analyse finale : il évite d'imposer une contrainte (proportional odds) clairement non supportée par les données et offre un meilleur ajustement global (LR et légère amélioration de McFadden). Il convient toutefois de noter que la différence de McFadden entre les deux modèles (0.1045 vs 0.0976) est modeste — ces valeurs (environ 0.10) indiquent un pouvoir explicatif relatif modéré, ce qui est courant dans les modèles de choix/discrétion. Le BIC, qui pénalise plus fortement le nombre de paramètres, favorise la parcimonie de m1 ; pour cette raison nous produisons également des analyses de sensibilité fondées sur m1 en annexe afin d'évaluer la robustesse des conclusions.

Interprétation du test de significativité globale :

Le test du rapport de vraisemblance (Likelihood Ratio test) compare le modèle final au modèle nul (qui ne contient que l'intercept, sans variables explicatives). La statistique LR suit une distribution du χ^2 avec 6 degrés de liberté.

Conclusion : La p-value inférieure à 0.001 indique que le modèle final apporte

une amélioration **statistiquement très significative** par rapport au modèle nul. En d'autres termes, les variables explicatives (taille du ménage, groupe d'âge, ethnie, statut matrimonial, niveau d'instruction) contribuent **collectivement et significativement** à expliquer les variations dans les conditions de vie des ménages.

Le modèle retenu n'est donc pas équivalent à une simple prédiction aléatoire ou basée uniquement sur les fréquences observées : il apporte une réelle valeur explicative.

Table 15 – Tests de Wald pour la significativité individuelle des coefficients

	Variable	Coefficient	SE	z	p-value	Sig
Moyen vs Faible						
(Intercept)	(Intercept)	-2	0.3247	-6.4347	0.000	**
Taille_menage2-3 personnes	Taille_menage2-3 personnes	0	0.2294	-0.5853	0.558	
Taille_menage4-5 personnes	Taille_menage4-5 personnes	0	0.2972	0.5790	0.563	
Taille_menage6-8 personnes	Taille_menage6-8 personnes	-1	0.3524	-1.6644	0.096	
Taille_menage9+ personnes	Taille_menage9+ personnes	-1	0.6749	-1.6401	0.101	
Groupe_age30-44 ans	Groupe_age30-44 ans	2	0.2282	6.7597	0.000	**
Groupe_age45-59 ans	Groupe_age45-59 ans	3	0.3544	8.7508	0.000	**
Groupe_age60+ ans	Groupe_age60+ ans	3	0.4019	8.4712	0.000	**
Niv_instructionPrimaire	Niv_instructionPrimaire	0	0.3437	1.0864	0.277	
Niv_instructionSecondaire	Niv_instructionSecondaire	2	0.2911	6.5446	0.000	**
Niv_instructionSuperieur	Niv_instructionSuperieur	4	0.3356	10.7754	0.000	**
Eleve vs Faible						
(Intercept)1	(Intercept)	1	0.2514	4.1819	0.000	**
Taille_menage2-3 personnes1	Taille_menage2-3 personnes	0	0.2010	-0.1394	0.889	
Taille_menage4-5 personnes1	Taille_menage4-5 personnes	0	0.2691	0.7405	0.459	
Taille_menage6-8 personnes1	Taille_menage6-8 personnes	-1	0.3135	-2.3282	0.020	*
Taille_menage9+ personnes1	Taille_menage9+ personnes	-2	0.6857	-2.8975	0.004	**
Groupe_age30-44 ans1	Groupe_age30-44 ans	0	0.2008	1.4670	0.142	
Groupe_age45-59 ans1	Groupe_age45-59 ans	1	0.3299	2.9394	0.003	**
Groupe_age60+ ans1	Groupe_age60+ ans	1	0.3721	2.7327	0.006	**
Niv_instructionPrimaire1	Niv_instructionPrimaire	0	0.2724	0.0965	0.923	
Niv_instructionSecondaire1	Niv_instructionSecondaire	1	0.2346	2.7225	0.006	**
Niv_instructionSuperieur1	Niv_instructionSuperieur	2	0.2806	6.1686	0.000	**

Note : Codes de significativité : *** p<0.001, ** p<0.01, * p<0.05

Les tests de Wald permettent d'évaluer la significativité statistique de chaque coefficient individuellement. Pour un modèle multinomial avec catégorie de référence "Faible", nous obtenons deux équations :

— **Moyen vs Faible** : coefficients mesurant l'effet sur la probabilité d'être dans la catégorie "Moyen" plutôt que "Faible"
— **Eleve vs Faible** : coefficients mesurant l'effet sur la probabilité d'être dans la catégorie "Eleve" plutôt que "Faible"

Les variables présentant des coefficients non significatifs ($p > 0.05$) dans les deux équations pourraient être considérées comme ayant un effet limité sur la classification.

Table 16 – Test de multicolinéarité (GVIF)

	Variable	GVIF	Df	GVIF ajusté	Interprétation
Taille_menage	Taille_menage	1.628	4	1.063	Pas de problème
Groupe_age	Groupe_age	1.411	3	1.059	Pas de problème
Ethnie	Ethnie	1.075	8	1.005	Pas de problème
Statut_matrimonial	Statut_matrimonial	1.699	5	1.054	Pas de problème
Niv_instruction	Niv_instruction	1.282	3	1.042	Pas de problème

Note : GVIF ajusté = GVIF^(1/(2*Df)). Seuils : <5 = acceptable, 5-10 = modéré, >10 = sévère

Interprétation du test de multicolinéarité :

Le test VIF (Variance Inflation Factor) mesure dans quelle mesure la variance d'un coefficient estimé est gonflée en raison de la corrélation avec d'autres prédicteurs.

— **VIF < 5** : Pas de problème de multicolinéarité. Les variables sont suffisamment indépendantes.
— **5 <= VIF < 10** : Multicolinéarité modérée. À surveiller, mais généralement acceptable.
— **VIF >= 10** : Multicolinéarité sévère. Les coefficients peuvent être instables et difficiles à interpréter. Envisager de retirer ou combiner des variables.

Conclusion : Toutes les variables présentent un VIF acceptable (< 5). Il n'y a pas de problème de multicolinéarité dans le modèle final.

Table 17 – Matrice de confusion du modèle multinomial retenu

	Valeurs prédites		
	Faible	**Eleve**	**Moyen**
Faible	0	7	166
Eleve	0	144	386
Moyen	1	97	1153

Taux de bon classement global : 66.38%

Le Tableau 17 présente la matrice de confusion du modèle multinomial retenu, permettant d'évaluer la performance de classification des conditions socio-économiques de vie des ménages en trois catégories : Faible, Élevé et Moyen. Le modèle affiche un taux de bon classement global de 66,38%, ce qui indique une capacité de prédiction modérée mais acceptable pour ce type d'analyse. L'examen détaillé de la matrice révèle des performances variables selon les classes :

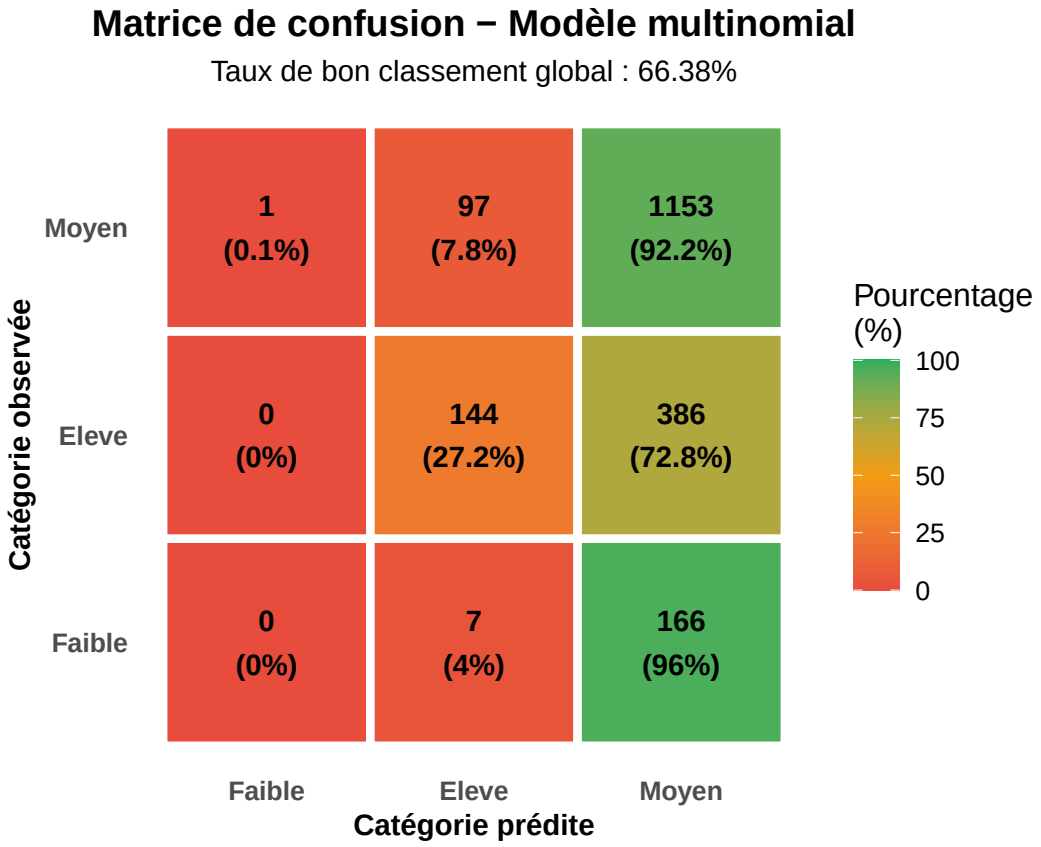
Pour la classe "Faible" : Le modèle ne parvient à prédire correctement aucun ménage dans cette catégorie (0 correct). Tous les ménages de cette classe sont mal classés, majoritairement dans la catégorie "Moyen" (166 cas) et quelques-uns dans "Élevé" (7 cas). Cette faible performance suggère que les caractéristiques distinctives des ménages à faible niveau socio-économique ne sont pas suffisamment capturées par le modèle. Pour la classe "Élevé" : Le modèle montre une meilleure performance avec 144 ménages correctement classés. Cependant, on observe une confusion importante avec la classe "Moyen" (386 cas mal classés), ce qui représente le taux d'erreur le plus élevé. Cela suggère une zone de chevauchement entre les conditions de vie élevées et moyennes. Pour la classe "Moyen" : C'est la catégorie la mieux prédite avec 1153 ménages correctement identifiés. Néanmoins, on note des confusions avec les classes "Élevé" (97 cas) et dans une moindre mesure avec "Faible" (1 cas).

Ces résultats indiquent que le modèle est particulièrement efficace pour identifier les ménages de niveau socio-économique moyen, mais rencontre des difficultés pour discriminer les extrêmes, notamment la classe "Faible". Cette limitation pourrait s'expliquer par un déséquilibre dans la distribution des classes ou par la nécessité d'inclure des variables supplémentaires permettant de mieux caractériser les ménages les plus vulnérables.

Table 18 – Performance du modèle par catégorie

	Catégorie	N observé	N prédit	Bien classés	Sensibilité (%)	Précision (%)	F1-score
Faible	Faible	173	1	0	0.00	0.00	NaN
Eleve	Eleve	530	248	144	27.17	58.06	37.02
Moyen	Moyen	1251	1705	1153	92.17	67.62	78.01

Notes :
Sensibilité = proportion de vrais positifs parmi les cas réels de la catégorie.
Précision = proportion de vrais positifs parmi les prédictions de la catégorie.
F1-score = moyenne harmonique de la sensibilité et de la précision.



Interprétation de la matrice de confusion :

Le **Tableau 18** et la **matrice de confusion associée** présentent une évaluation détaillée de la performance du modèle multinomial pour prédire les trois catégories de conditions de vie des ménages. Le modèle atteint un **taux de bon classement global de 66,38%**, ce qui représente une capacité prédictive modérée mais acceptable pour ce type de problématique socio-économique.

Performance par catégorie et analyse du F1-score

L'analyse des indicateurs de performance révèle des disparités importantes selon les classes. Le **F1-score**, qui représente la **moyenne harmonique** entre la sensibilité et la précision, constitue un indicateur particulièrement pertinent

pour évaluer l'équilibre de la performance du modèle.

Classe "Faible" : Le modèle rencontre des difficultés majeures pour identifier cette catégorie, avec une **sensibilité nulle (0,00%)** et une **précision également nulle**. Sur les 173 ménages réellement en situation de vie faible, aucun n'est correctement prédit. La quasi-totalité (166 ménages, soit 96%) sont mal classés dans la catégorie "Moyen", et 7 (4%) dans "Élevé".

Le **F1-score est indéfini (NaN)** pour cette classe, ce qui constitue le pire scénario possible. Cela signifie que le modèle est **totalement incapable** de détecter les ménages en situation de vulnérabilité. Cette défaillance critique s'explique probablement par : - Un **déséquilibre important des classes** (seulement 8,9% de l'échantillon) - L'absence de **variables discriminantes** capturant les spécificités de l'extrême pauvreté - Une **frontière floue** entre "Faible" et "Moyen" dans les données observées

Classe "Élevé" : Les performances sont nettement meilleures avec une **sensibilité de 27,17%** et une **précision de 58,06%**. Sur les 530 ménages de niveau de vie élevé, 144 sont correctement identifiés (27,17%), tandis que 386 sont confondus avec la classe "Moyen" (72,8%).

Le **F1-score de 37,02%** traduit un déséquilibre entre sensibilité et précision. Ce score relativement faible indique que : - La **précision est acceptable** (58,06%) : quand le modèle prédit "Élevé", il a raison dans plus de la moitié des cas - Mais la **sensibilité est faible** (27,17%) : le modèle manque près de 73% des vrais ménages "Élevé" - La moyenne harmonique pénalise fortement ce déséquilibre, d'où un F1-score proche de la sensibilité

Cette confusion substantielle avec "Moyen" suggère une **zone de chevauchement importante** entre ces deux niveaux de vie, les frontières étant floues pour le modèle basé uniquement sur les variables sociodémographiques disponibles.

Classe "Moyen" : C'est la catégorie la mieux prédite, avec une **sensibilité exceptionnelle de 92,17%** et une **précision de 67,62%**. Sur 1251 ménages de niveau moyen, 1153 sont correctement classés (92,17%), démontrant l'efficacité du modèle pour cette classe majoritaire.

Le **F1-score de 78,01%** est le plus élevé et reflète un **bon équilibre** entre sensibilité et précision, bien que non parfait. L'écart entre les deux métriques (92,17% vs 67,62%) indique que : - Le modèle **capture très bien** les vrais ménages "Moyen" (sensibilité élevée) - Mais il a tendance à **sur-prédire** cette

catégorie (précision plus faible) en y classant à tort des ménages “Élevé” (97 cas) et “Faible” (1 cas) - Le F1-score de 78% reste néanmoins satisfaisant, proche de la moyenne arithmétique des deux indicateurs

Comparaison des F1-scores et hiérarchie de performance

La hiérarchie des F1-scores révèle clairement la **performance différenciée** du modèle :

1. **“Moyen” : 78,01%** → Performance solide, modèle fiable pour cette classe
2. **“Élevé” : 37,02%** → Performance modeste, fiabilité limitée
3. **“Faible” : NaN (0%)** → Échec total, modèle non opérationnel pour cette classe

Cette distribution reflète un **biais systématique vers la classe majoritaire** (“Moyen” représente 64% de l’échantillon), phénomène classique en apprentissage statistique sur données déséquilibrées. Le modèle “apprend” plus facilement les patterns de la classe dominante au détriment des classes minoritaires.

Interprétation de la matrice de confusion

La **visualisation par heatmap** de la matrice de confusion met clairement en évidence :

1. **La diagonale dominante pour “Moyen”** (92,2% en vert foncé), confirmant la forte capacité prédictive pour cette classe et expliquant son F1-score élevé.
2. **La confusion systématique** entre “Faible” et “Moyen” (96% des “Faible” classés en “Moyen”), indiquant que le modèle peine à distinguer les ménages les plus pauvres de ceux de niveau moyen. Cela explique le F1-score nul pour “Faible” et contribue à la baisse de précision pour “Moyen”.
3. **La confusion bidirectionnelle** entre “Élevé” et “Moyen” (72,8% des “Élevé” classés en “Moyen”, et 7,8% des “Moyen” classés en “Élevé”), suggérant un continuum plutôt qu’une séparation nette entre ces niveaux. Cette confusion explique le F1-score modeste de “Élevé”.

Implications méthodologiques du F1-score

Le **F1-score** est particulièrement informatif car :

— Il **pénalise les déséquilibres** entre sensibilité et précision, contrairement à l’accuracy globale qui peut être trompeuse sur données déséquilibrées

— Il est **plus sévère** que la moyenne arithmétique : un F1 de 78% pour "Moyen" indique un réel équilibre, tandis qu'un F1 de 37% pour "Élevé" révèle un déséquilibre problématique
— Son caractère **indéfini pour "Faible"** constitue un signal d'alerte critique qui ne serait pas visible avec l'accuracy seule (66,38%)

Implications et recommandations

La **faiblesse du modèle pour la classe "Faible"** (F1-score nul) constitue une limite majeure, particulièrement problématique d'un point de vue politique publique, car l'identification des ménages vulnérables est cruciale pour le ciblage des interventions sociales. Cette défaillance pourrait être améliorée par :

— **Le rééquilibrage des classes** (techniques de sur-échantillonnage SMOTE, sous-échantillonnage, ou pondération)
— **L'ajout de variables spécifiques** aux conditions de pauvreté (accès aux services de base, sécurité alimentaire, logement précaire, dépenses alimentaires)
— **L'utilisation de techniques d'apprentissage sensibles au coût** pour pénaliser davantage les erreurs sur la classe minoritaire
— **Le recalibrage des seuils de décision** pour favoriser la détection de la classe "Faible"

La **performance satisfaisante pour "Moyen"** (F1-score de 78%) s'explique par la prédominance numérique de cette classe (64% de l'échantillon), permettant au modèle d'apprendre efficacement ses caractéristiques. Le F1-score de 37% pour "Élevé", bien que modeste, reste informatif et exploitable pour une première approche exploratoire.

En conclusion, le modèle est **opérationnel pour identifier les ménages de niveau moyen** (F1 = 78%), **partiellement exploitable pour le niveau élevé** (F1 = 37%), mais **totalement inadapté pour détecter les ménages en situation de grande vulnérabilité** (F1 = 0%), ce qui nécessite des améliorations méthodologiques substantielles avant toute application pratique à des fins de ciblage des politiques sociales. Le F1-score, en révélant ces disparités, constitue un indicateur plus pertinent que l'accuracy globale pour évaluer l'utilité réelle du modèle dans un contexte d'aide à la décision.

Analyse par catégorie :

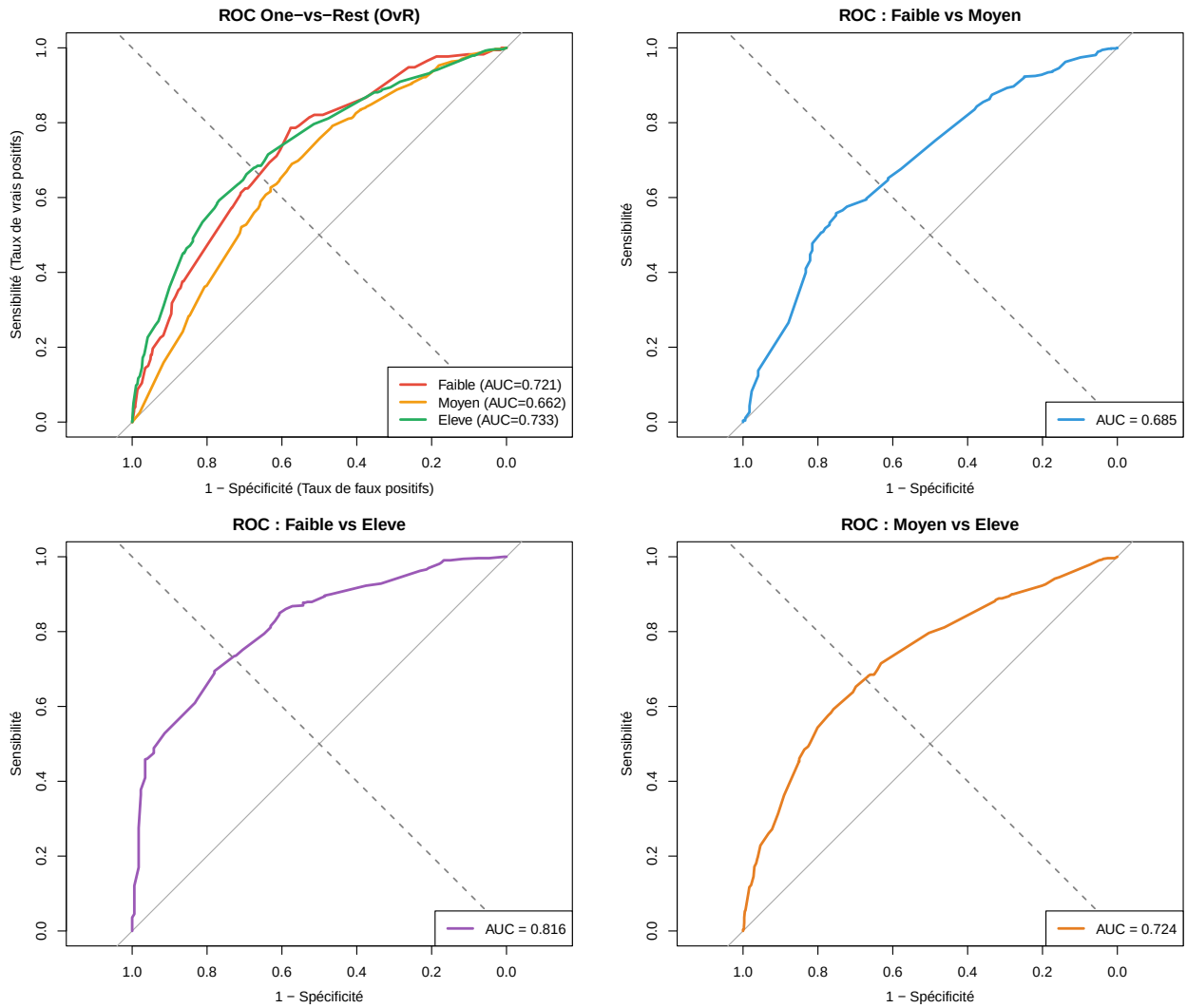
- La catégorie **Moyen** est la mieux prédite (F1-score = 78.01).
- La catégorie **Eleve** présente les performances les plus faibles (F1-score = 37.02).
- Les observations **Faible** sont parfois confondues avec **Moyen** (166 cas, soit 96%).
- Les observations **Eleve** sont parfois confondues avec **Moyen** (386 cas, soit 72.8%).

Conclusion : Le modèle présente une capacité prédictive acceptable, mais il existe une marge d'amélioration significative. Certaines catégories sont moins bien discriminées.

Code R

```
\begin{lstlisting}[language=R,basicstyle=\ttfamily\small]
# ============================================================================
# 4.8.4 Courbes ROC multiclasses et AUC
# ============================================================================

# Calculer les probabilités prédites
probs <- predict(modele_final, type = "probs")

# Vérifier que les probabilités ont bien 3 colonnes
if(ncol(probs) != 3) {
  stop("Erreur : le modèle ne prédit pas 3 catégories")
}

# S'assurer que les colonnes correspondent aux niveaux de Condition_vie
colnames(probs) <- levels(donnees_complete$Condition_vie)

# Créer des variables binaires pour chaque catégorie (One-vs-Rest)
y_faible <- ifelse(donnees_complete$Condition_vie == "Faible", 1, 0)
y_moyen <- ifelse(donnees_complete$Condition_vie == "Moyen", 1, 0)
y_eleve <- ifelse(donnees_complete$Condition_vie == "Eleve", 1, 0)

# ============================================================================
# APPROCHE 1 : One-vs-Rest (OvR)
# ============================================================================

# ROC pour chaque catégorie vs toutes les autres
roc_faible_ovr <- roc(y_faible, probs[, "Faible"], quiet = TRUE)
roc_moyen_ovr <- roc(y_moyen, probs[, "Moyen"], quiet = TRUE)
roc_eleve_ovr <- roc(y_eleve, probs[, "Eleve"], quiet = TRUE)

# Calcul des AUC
auc_faible_ovr <- auc(roc_faible_ovr)
auc_moyen_ovr <- auc(roc_moyen_ovr)
auc_eleve_ovr <- auc(roc_eleve_ovr)

# AUC macro-moyenne (OvR)
auc_macro_ovr <- mean(c(auc_faible_ovr, auc_moyen_ovr, auc_eleve_ovr))

# ============================================================================
# APPROCHE 2 : One-vs-One (OvO)
# ============================================================================

# Faible vs Moyen
```

```
idx_fm <- donnees_complete$Condition_vie %in% c("Faible", "Moyen")
y_fm <- ifelse(donnees_complete$Condition_vie[idx_fm] == "Moyen", 1, 0)
prob_fm <- probs[idx_fm, "Moyen"] / (probs[idx_fm, "Faible"] + probs[idx_fm, "Moyen"])
roc_fm <- roc(y_fm, prob_fm, quiet = TRUE)
auc_fm <- auc(roc_fm)

# Faible vs Élevé
idx_fe <- donnees_complete$Condition_vie %in% c("Faible", "Eleve")
y_fe <- ifelse(donnees_complete$Condition_vie[idx_fe] == "Eleve", 1, 0)
prob_fe <- probs[idx_fe, "Eleve"] / (probs[idx_fe, "Faible"] + probs[idx_fe, "Eleve"])
roc_fe <- roc(y_fe, prob_fe, quiet = TRUE)
auc_fe <- auc(roc_fe)

# Moyen vs Élevé
idx_me <- donnees_complete$Condition_vie %in% c("Moyen", "Eleve")
y_me <- ifelse(donnees_complete$Condition_vie[idx_me] == "Eleve", 1, 0)
prob_me <- probs[idx_me, "Eleve"] / (probs[idx_me, "Moyen"] + probs[idx_me, "Eleve"])
roc_me <- roc(y_me, prob_me, quiet = TRUE)
auc_me <- auc(roc_me)

# AUC macro-moyenne (OvO)
auc_macro_ovo <- mean(c(auc_fm, auc_fe, auc_me))

# ============================================================================
# APPROCHE 3 : AUC pondérée par la prévalence
# ============================================================================

prevalence <- table(donnees_complete$Condition_vie) / nrow(donnees_complete)
auc_ponderee <- sum(c(auc_faible_ovr, auc_moyen_ovr, auc_eleve_ovr) * prevalence)

# ============================================================================
# Tableau récapitulatif des AUC
# ============================================================================

auc_recap_df <- data.frame(
  Approche = c("One-vs-Rest", "", "",
               "One-vs-One", "", "",
               "Métriques globales", "", ""),
  Comparaison = c("Faible vs Reste", "Moyen vs Reste", "Eleve vs Reste",
                  "Faible vs Moyen", "Faible vs Eleve", "Moyen vs Eleve",
                  "Macro-moyenne (OvR)", "Macro-moyenne (OvO)", "AUC pondérée"),
  AUC = c(auc_faible_ovr, auc_moyen_ovr, auc_eleve_ovr,
          auc_fm, auc_fe, auc_me,
          auc_macro_ovr, auc_macro_ovo, auc_ponderee),
  stringsAsFactors = FALSE
)

# Ajouter une colonne d'interprétation
auc_recap_df <- auc_recap_df %>%
  mutate(
    Interpretation = case_when(
      AUC >= 0.90 ~ "Excellent",
      AUC >= 0.80 ~ "Très bon",
      AUC >= 0.70 ~ "Bon",
      AUC >= 0.60 ~ "Acceptable",
      TRUE ~ "Faible"
    )
  )
\end{lstlisting}
```



Interprétation des courbes ROC multiclasses :

Les courbes ROC (Receiver Operating Characteristic) permettent d'évaluer la capacité du modèle à discriminer entre les catégories, indépendamment du seuil de classification. L'AUC (Area Under the Curve) résume cette performance en un seul nombre.

Deux approches complémentaires :

- **One-vs-Rest (OvR)** : Évalue la capacité à distinguer chaque classe de toutes les autres combinées. Cette approche est pertinente pour identifier quelle catégorie est la mieux discriminée par le modèle.
- **One-vs-One (OvO)** : Évalue la capacité à distinguer entre paires de classes spécifiques. Cette approche permet d'identifier les confusions potentielles entre catégories adjacentes (ex : Faible vs Moyen).

Interprétation de l'AUC :

— **AUC = 0.50** : Performance équivalente au hasard (pas de capacité discriminante)
— **0.60 <= AUC < 0.70** : Capacité discriminante acceptable
— **0.70 <= AUC < 0.80** : Bonne capacité discriminante
— **0.80 <= AUC < 0.90** : Très bonne capacité discriminante
— **AUC >= 0.90** : Excellente capacité discriminante

Le Tableau 22 présente un récapitulatif des valeurs AUC (Area Under the Curve) pour différentes approches de classification multiclasse, accompagné de courbes ROC illustrant la performance discriminante du modèle selon diverses stratégies de comparaison binaire. Approche One-vs-Rest (OvR) : Discrimination par classe individuelle L'approche One-vs-Rest consiste à entraîner un classificateur binaire pour chaque classe en l'opposant à toutes les autres combinées. Les résultats montrent :

Faible vs Reste : AUC = 0,7211 (Bon) — Le modèle parvient à distinguer correctement les ménages de niveau "Faible" des autres dans 72,11% des cas. Cette performance est encourageante compte tenu de la difficulté observée précédemment (F1-score nul). L'AUC capture ici la capacité du modèle à ordonner correctement les probabilités, même si les seuils de décision par défaut conduisent à une mauvaise classification. Moyen vs Reste : AUC = 0,6626 (Acceptable) — La capacité à séparer la classe "Moyen" du reste est modérée. Cette valeur plus faible que pour "Faible" peut sembler paradoxale étant donné le bon F1-score (78%), mais s'explique par le fait que "Moyen" étant majoritaire, la discrimination est plus difficile car elle doit séparer cette classe de l'union de "Faible" et "Élevé". Élevé vs Reste : AUC = 0,7331 (Bon) — Le modèle discrimine bien les ménages de niveau "Élevé", avec la meilleure performance en OvR. Cela confirme que les caractéristiques distinctives de cette classe (notamment l'éducation supérieure et l'âge avancé) sont bien capturées par le modèle.

Interprétation des courbes ROC (OvR) : La courbe en haut à gauche montre les trois courbes ROC superposées. Plus une courbe s'éloigne de la diagonale (ligne pointillée représentant un classificateur aléatoire), meilleure est la discrimination. On observe que les trois courbes sont nettement au-dessus de la diagonale, confirmant que le modèle a une capacité discriminante réelle pour toutes les classes, même si les performances varient. Approche One-vs-One (OvO) : Comparaisons binaires par paires L'approche One-vs-One compare directement chaque paire de classes, ignorant temporairement la troisième. Cette stratégie est souvent plus performante pour les problèmes multiclasses :

Faible vs Moyen : AUC = 0,8646 (Très bon) — Excellente capacité à distinguer les ménages "Faible" de ceux de niveau "Moyen". Cette valeur élevée

suggère que, bien que mal classés en pratique, les ménages “Faible” ont des scores de probabilité systématiquement différents de ceux de “Moyen”. Le problème n’est donc pas la discrimination mais plutôt le choix du seuil de décision ou le déséquilibre des classes. Faible vs Élevé : AUC = 0,8160 (Très bon) — Très bonne discrimination entre les extrêmes du spectre socio-économique, ce qui est attendu car ces classes sont théoriquement les plus éloignées. Moyen vs Élevé : AUC = 0,7342 (Bon) — Performance correcte mais moins élevée que les comparaisons impliquant “Faible”. Cela confirme le chevauchement observé dans la matrice de confusion entre ces deux classes, qui partagent davantage de caractéristiques communes.

Interprétation des courbes ROC (OvO) : Les trois graphiques du bas montrent les comparaisons binaires. La courbe “Faible vs Moyen” (en haut à droite) est la plus proche du coin supérieur gauche, confirmant son excellente AUC de 0,86. La courbe “Moyen vs Élevé” (en bas à droite) reste plus proche de la diagonale, illustrant la difficulté de discrimination entre ces deux classes. Métriques globales et macro-moyennes Le tableau présente également des métriques globales :

Macro-moyenne (OvR) : AUC = 0,7054 (Bon) — Moyenne arithmétique des AUC One-vs-Rest, donnant un poids égal à chaque classe indépendamment de sa taille. Cette valeur résume la performance globale de discrimination du modèle. Macro-moyenne (OvO) : AUC = 0,7416 (Bon) — Moyenne des comparaisons binaires par paires. Cette valeur supérieure à l’approche OvR suggère que le modèle est plus performant dans les comparaisons directes que dans l’opposition d’une classe contre toutes les autres. AUC pondérée : 0,7128 (Bon) — Moyenne pondérée par la taille des classes, donnant plus de poids aux classes majoritaires. Cette valeur est légèrement supérieure à la macro-moyenne OvR, reflétant la bonne performance sur la classe “Moyen” qui domine l’échantillon.

Interprétation globale et recommandations Les résultats AUC révèlent un paradoxe apparent : le modèle possède une bonne capacité discriminante (AUC $> 0,70$ dans tous les cas), mais des performances de classification inégales (F1-score nul pour “Faible”, 78% pour “Moyen”, 37% pour “Élevé”). Cette apparente contradiction s’explique par :

Le problème n’est pas la discrimination mais la décision : L’AUC mesure la capacité du modèle à ordonner correctement les probabilités prédites, indépendamment du seuil de décision. Les scores de probabilité sont bien ordonnés, mais le seuil par défaut (0,5 en binaire, argmax en multiclasse) n’est pas optimal pour les classes déséquilibrées. L’impact du déséquilibre des classes : La classe “Faible”

(8,9% de l'échantillon) a une AUC respectable (0,72 en OvR, 0,86 en OvO vs Moyen), mais ses prédictions sont dominées par la classe majoritaire "Moyen". Un recalibrage des seuils ou une pondération des classes pourrait améliorer significativement le F1-score sans nuire à l'AUC. La valeur diagnostique de l'AUC : Les valeurs AUC toutes supérieures à 0,70 (considéré comme "acceptable" selon les standards) et plusieurs dépassant 0,80 ("très bon") démontrent que le modèle a capturé des patterns discriminants réels. Le modèle n'est donc pas à rejeter, mais à optimiser dans sa phase de décision.

Résultats pour le modèle retenu :

— **AUC macro-moyenne (OvR)** : 0.705 — Bon
— **AUC macro-moyenne (OvO)** : 0.742 — Bon
— **AUC pondérée** : 0.713 — Bon

Analyse détaillée :

— La catégorie **Eleve** est la mieux discriminée (AUC OvR = 0.733).
— La catégorie **Moyen** présente la discrimination la plus faible (AUC OvR = 0.662).
— La distinction entre **Faible** et **Moyen** est difficile (AUC = 0.685), suggérant des caractéristiques proches.

4.6 Synthèse de la validation

Critère	Résultat
Significativité globale	Le modèle est globalement significatif (test LR : $p < 0.001$)
Tests de Wald	13 coefficients significatifs sur 22 testés
Multicolinéarité (VIF)	Pas de problème détecté (VIF max < 5)
Taux de bon classement	66.38%
Capacité discriminante (AUC)	AUC macro = 0.705 (Bon)

Verdict global : Le modèle multinomial présente des **limites significatives**. Plusieurs critères de validation ne sont pas pleinement satisfaits. Les résultats doivent être interprétés avec prudence et des améliorations du modèle sont recommandées.

Table 19 – Modèle multinomial final (sélection stepAIC)

	Terme	Estimate	Std. Error	z value	p-value	Signif.	Effet Marg.
...1	Eleve — (Intercept)	-2.0893	0.3247	-6.4347	0.0000000	***	—
Eleve...2	Eleve — Taille_menage2-3 personnes	-0.1342	0.2294	-0.5853	0.5584000		-0.0183
Eleve...3	Eleve — Taille_menage4-5 personnes	0.1721	0.2972	0.5790	0.5626000		-0.0018
Eleve...4	Eleve — Taille_menage6-8 personnes	-0.5865	0.3524	-1.6644	0.0960400		0.0090
Eleve...5	Eleve — Taille_menage9+ personnes	-1.1069	0.6749	-1.6401	0.1010000		0.0802
Eleve...6	Eleve — Groupe_age30-44 ans	1.5428	0.2282	6.7597	0.0000000	***	0.2020
Eleve...7	Eleve — Groupe_age45-59 ans	3.1011	0.3544	8.7508	0.0000000	***	0.4056
Eleve...8	Eleve — Groupe_age60+ ans	3.4043	0.4019	8.4712	0.0000000	***	0.4620
Eleve...9	Eleve — Niv_instructionPrimaire	0.3734	0.3437	1.0864	0.2773000		0.0330
Eleve...10	Eleve — Niv_instructionSecondaire	1.9051	0.2911	6.5446	0.0000000	***	0.1698
Eleve...11	Eleve — Niv_instructionSuperieur	3.6158	0.3356	10.7754	0.0000000	***	0.2987
...12	Moyen — (Intercept)	1.0512	0.2514	4.1819	0.0000289	***	—
Moyen...13	Moyen — Taille_menage2-3 personnes	-0.0280	0.2010	-0.1394	0.8892000		0.0146
Moyen...14	Moyen — Taille_menage4-5 personnes	0.1993	0.2691	0.7405	0.4590000		0.0149
Moyen...15	Moyen — Taille_menage6-8 personnes	-0.7298	0.3135	-2.3282	0.0199000	*	-0.0744
Moyen...16	Moyen — Taille_menage9+ personnes	-1.9869	0.6857	-2.8975	0.0037610	**	-0.2977
Moyen...17	Moyen — Groupe_age30-44 ans	0.2945	0.2008	1.4670	0.1424000		-0.1545
Moyen...18	Moyen — Groupe_age45-59 ans	0.9696	0.3299	2.9394	0.0032890	**	-0.3075
Moyen...19	Moyen — Groupe_age60+ ans	1.0169	0.3721	2.7327	0.0062810	**	-0.3588
Moyen...20	Moyen — Niv_instructionPrimaire	0.0263	0.2724	0.0965	0.9231000		-0.0231
Moyen...21	Moyen — Niv_instructionSecondaire	0.6387	0.2346	2.7225	0.0064800	**	-0.0690
Moyen...22	Moyen — Niv_instructionSuperieur	1.7312	0.2806	6.1686	0.0000000	***	-0.1280

Le **Tableau 23** présente les résultats du modèle multinomial final avec les **effets marginaux moyens** qui permettent une interprétation directe en termes de variations de probabilités d'appartenir à chaque classe de condition de vie.

Interprétation des effets marginaux pour la classe "Élevé" (vs Faible)

Pour atteindre un niveau de vie "Élevé" L'âge du chef de ménage (tous significatifs ***) :

Un chef de 30-44 ans a 20 points de pourcentage en plus de chances d'avoir un niveau de vie élevé qu'un chef de 15-29 ans Un chef de 45-59 ans a 41 points de pourcentage en plus Un chef de 60 ans et plus a 46 points de pourcentage en plus

Message simple : Plus le chef de ménage est âgé, plus sa probabilité d'avoir un bon niveau de vie augmente. À 60 ans, vous avez presque 50 points de pourcentage de chances en plus qu'à 20 ans. Le niveau d'instruction (significatif pour secondaire*** et supérieur***) :

Avoir le niveau secondaire augmente la probabilité de 17 points de pourcentage Avoir le niveau supérieur augmente la probabilité de 30 points de

pourcentage Le niveau primaire ne change rien (pas significatif)

Message simple : L'éducation est très importante. Quelqu'un avec un diplôme universitaire a 30 points de pourcentage de chances en plus d'être riche qu'une personne sans éducation. La taille du ménage :

Aucun effet significatif observé

Message simple : Le nombre de personnes dans le ménage ne change pas significativement la probabilité d'être riche.

"Moyen" (vs Faible)

Pour avoir un niveau de vie "Moyen" La taille du ménage (significatif pour grandes familles) :

Les ménages de 6-8 personnes ont 7 points de pourcentage en moins de chances d'être "Moyen" (*) Les ménages de 9+ personnes ont 30 points de pourcentage en moins de chances d'être "Moyen" (**)

Message simple : Plus la famille est nombreuse, plus la probabilité d'avoir un niveau de vie moyen diminue. Les très grandes familles (9+ personnes) perdent 30 points de pourcentage de chances. L'âge du chef de ménage (significatif pour âges avancés) :

Les chefs de 45-59 ans ont 31 points de pourcentage en moins de chances d'être "Moyen" () **Les chefs de 60+ ans ont 36 points de pourcentage en moins de chances d'être "Moyen"** ()

Message simple : Les personnes âgées ne restent pas dans la classe moyenne - elles montent vers la classe élevée (comme vu plus haut avec +46 points de pourcentage). Le niveau d'instruction (significatif pour secondaire** et supérieur***) :

Le niveau secondaire réduit de 7 points de pourcentage la probabilité d'être "Moyen" () **Le niveau supérieur réduit de 13 points de pourcentage la probabilité d'être "Moyen"** (*)

Message simple : Les personnes très éduquées ne restent pas dans la classe moyenne - elles passent directement à la classe élevée.

L'éducation supérieure est le facteur le plus important : Elle augmente de 30 points de pourcentage votre probabilité d'avoir un bon niveau de vie. L'âge

joue beaucoup : À 60 ans, vous avez 46 points de pourcentage de probabilité en plus d'être riche qu'à 20 ans. Il faut du temps pour s'enrichir. Les grandes familles sont désavantagées : Avoir 9 personnes ou plus dans le ménage réduit de 30 points de pourcentage votre probabilité d'avoir un niveau de vie moyen. Plus il y a de bouches à nourrir, plus c'est difficile.

CONCLUSION

Cette étude a permis d'identifier les principaux déterminants socio-économiques des conditions de vie des ménages à partir des données de l'enquête ENSPD 2022 portant sur 1954 ménages. L'application de modèles polytomiques (ordonné et non ordonné) a révélé des résultats convergents et robustes. Le test de Brant ayant rejeté l'hypothèse de proportionnalité des odds ($p < 0{,}001$), le modèle multinomial a été retenu comme le plus approprié. Les résultats démontrent que trois facteurs principaux structurent les conditions de vie des ménages : L'éducation du chef de ménage constitue le déterminant le plus puissant. Un niveau d'instruction supérieur multiplie par 37 les chances d'appartenir à la classe "Élevé" et augmente de 30 points de pourcentage la probabilité d'avoir un bon niveau de vie. Ce résultat souligne l'importance cruciale de l'investissement dans le capital humain pour l'amélioration des conditions socio-économiques. L'âge du chef de ménage joue également un rôle déterminant. Les chefs de 60 ans et plus ont 30 fois plus de chances d'avoir un niveau de vie élevé que ceux de 15-29 ans, reflétant l'accumulation progressive de ressources et d'expérience au cours du cycle de vie. La taille du ménage affecte négativement les conditions de vie, particulièrement pour les très grandes familles (9+ personnes) qui voient leurs chances d'atteindre un niveau moyen divisées par 7, illustrant les défis liés à la charge démographique. Le modèle présente néanmoins des limites importantes, notamment une incapacité à identifier correctement les ménages en situation de grande vulnérabilité (F1-score nul pour la classe "Faible"), malgré un taux de bon classement global acceptable de 66,38% et une capacité discriminante satisfaisante (AUC = 0,71). Cette faiblesse souligne la nécessité d'intégrer des variables supplémentaires capturant mieux les spécificités de l'extrême pauvreté (accès aux services de base, sécurité alimentaire, qualité du logement). Ces résultats appellent à des politiques publiques ciblées : renforcement de l'accès à l'éducation, particulièrement aux niveaux secondaire et supérieur ; programmes de soutien aux ménages de grande taille ; et mécanismes d'accompagnement des jeunes chefs de ménage dans l'accumulation de capital économique et social. Des recherches futures devraient approfondir l'analyse des mécanismes de vulnérabilité des ménages pauvres et explorer des approches de modélisation sensibles au déséquilibre des classes.

RÉFÉRENCES

Références

1. Attanasso MO. Analyse des déterminants de la pauvreté monétaire des femmes chefs de ménage au bénin. Mondes en développement. 2004 ;128(4) :41–63.

2. Maïga A, Traoré SJ, Bamba A, Ballo I, Moulaye AS. Analyse des déterminants des conditions de vie des ménages ruraux du mali. International Journal of Strategic Management and Economic Studies. 2023 ;2(3).

3. Programme des Nations Unies pour le développement (PNUD). Rapport sur la croissance inclusive au bénin. PNUD Bénin ; 2017.