



**Institut Supérieure des Sciences de la Population(ISSP)**

**Filière** **: Licence Professionnelle Analyse Statistique ( S4 )**

**Thème : Modèle logit conditionnel**

**Professeur :**

**Dr Israël SAWADOGO**

**ETUDIANTS :**

**IMA Hamidou**

**KOUDA N. Toussaint**

**SAWADOGO Boniface**

**Table des matières**

[Introduction 2](#_Toc169186876)

[I. Spécification du modèle 2](#_Toc169186877)

[I.1. Contexte & Justification 2](#_Toc169186878)

[I.2. Définition 3](#_Toc169186879)

[II. Méthode d’estimation du modèle Logit Conditionnel 4](#_Toc169186880)

[II.1. Définition de la vraisemblance 4](#_Toc169186881)

[II.2. Procédure de détermination des paramètres 5](#_Toc169186882)

[II.3. Estimation des coefficients 6](#_Toc169186883)

[II.4. Test d’hypothèses 7](#_Toc169186884)

[III. Domaines d’applications 7](#_Toc169186885)

[III.1. Dans le domaine du marketing 8](#_Toc169186886)

[III.2. Dans le domaine du transport 8](#_Toc169186887)

[III.3. Dans le domaine de la médecine 8](#_Toc169186888)

[III.4. Économie du travail 9](#_Toc169186889)

[IV. Avantages et limites du modèle logit conditionnel 9](#_Toc169186890)

[IV.1. Avantages 9](#_Toc169186891)

[IV.2. Limites du modèle 10](#_Toc169186892)

[V. Cas pratique 11](#_Toc169186893)

[Conclusion 11](#_Toc169186894)

# Introduction

Depuis l'innovation du modèle logit par McFadden en 1973, la régression logistique conditionnelle s'est imposée comme un instrument efficace pour modéliser les préférences parmi diverses options en se basant sur des attributs spécifiques. Dans le domaine de la recherche scientifique, ce modèle a été largement adopté pour analyser les comportements, les préférences et les décisions humaines. Bien qu'elle partage des similitudes avec le modèle logit multinomial non ordonné, la régression logistique conditionnelle se distingue par des aspects fondamentaux. En effet, alors que le modèle logit multinomial utilise un unique vecteur de variables indépendantes pour chaque observation, le modèle logit conditionnel considère des valeurs distinctes des variables indépendantes pour chaque alternative. Cela implique que les caractéristiques associées à chaque alternative sont prises en compte de manière individuelle, permettant ainsi une compréhension plus approfondie des facteurs qui influencent les choix. Dans ce rapport, nous allons examiner en détail les principes, les méthodes d’estimation des paramètres et les applications du modèle logit conditionnel, dans le but d'approfondir notre compréhension des mécanismes de prise de décision humaine.

## Spécification du modèle

### Contexte et Justification

Le modèle logit conditionnel est utilisé pour analyser des données où chaque individu a plusieurs choix possibles, et chaque choix a des caractéristiques spécifiques. Cette méthode est particulièrement utile pour l'analyse conjointe et pour étudier des habitudes de transport, par exemple, où chaque individu choisit un seul moyen de transport parmi plusieurs options. Dans ce contexte, le modèle logit conditionnel permet de modéliser les préférences des individus pour chaque alternative et d'estimer les paramètres du modèle qui influencent ces préférences.

Le modèle logit conditionnel est justifié par le fait que les choix des individus sont influencés par des facteurs non observés, ce qui rend difficile de déduire des propriétés du comportement de choix de la population à partir des modèles individuels. Les facteurs non observés peuvent inclure des préférences personnelles, des habitudes, des croyances, etc. qui varient d'un individu à l'autre. Cependant, en modélisant les caractéristiques des alternatives plutôt que celles des individus, le modèle logit conditionnel permet de prendre en compte ces facteurs non observés et d'obtenir des résultats plus précis.

### Définition

En partant de l'hypothèse que le choix d'un individu dépend de l'utilité qu'il retire de chaque alternative. L'utilité d'une alternative est composée d'une partie observable, qui dépend des caractéristiques de l'alternative, et d'une partie non observable, qui représente les facteurs non mesurés qui influencent le choix de l'individu.

Soit 𝑈𝑖𝑗 l'utilité de l'alternative *j* pour l'individu *i*. On peut écrire :   
Où :

* Est la partie observable de l'utilité, qui dépend des caractéristiques de l'alternative 𝑗et des paramètres du modèle ,
* est la partie non observable de l'utilité, qui suit une loi de Gumbel.

La loi de Gumbel est définie par la fonction de répartition suivante :

Où *a* et *b* sont des paramètres de la loi. Sa fonction de densité est :

Cette loi est utilisée pour modéliser des phénomènes extrêmes tels que :

* Les valeurs maximales de variables aléatoires suivant une loi normale
* Les temps de retour de phénomènes naturels tels que les crues de rivières
* Les températures maximales
* Les coûts de sinistres en assurance

En faisant l'hypothèse que l'individu i choisit l'alternative qui lui procure la plus grande utilité, on peut montrer que la probabilité de choisir l'alternative 𝑦 s'écrit *P*(*Y*=*y*) =

​​En supposant que = , on obtient la formule de la probabilité du modèle logit conditionnel devient :

Où :

* est la probabilité de choisir l'alternative k,
* est le vecteur des paramètres du modèle, avec la constante et le coefficient de la première variable explicative, est le coefficient de la deuxièmevariable explicative, etc.
* est le vecteur caractéristique des alternatives de l’individu i ,
* mest le nombre total d'alternatives.

En normalisant les vecteurs de caractéristiques par le vecteur de caractéristique de la première modalité tel que on peut alors écrire :

## Méthode d’estimation du modèle Logit Conditionnel

Plusieurs méthodes peuvent être utilisées pour l’estimation des paramètres du modèle logit conditionnel dont les importantes sont notamment celle du maximum de vraisemblance, les méthodes semi-paramétriques et les méthodes non paramétriques. Dans le cadre de notre étude, nous nous contenterons sur la méthode du maximum de vraisemblance qui est une méthode paramétrique. Ainsi, pour estimer les paramètres de du modèle, on cherche à maximiser la fonction de vraisemblance.

### Définition de la vraisemblance

Soit un échantillon de la loi de probabilité On appelle vraisemblance du paramètre , l’application :

Où

Pour un modèle logit conditionnel, la probabilité de sélection peut être définie de la façon suivante :

La vraisemblance associée à un modèle logit conditionnel à n modalités et d’un vecteur  à k+1 paramètres s’écrit :

### Procédure de détermination des paramètres

Pour cette partie, nous montrerons comment procéder pour déterminer les estimateurs des paramètres.

Pour estimer les paramètres du modèle, on cherche à maximiser la fonction de vraisemblance.

On sait que :

Avec si l’individu i choisi l’alternative j et sinon.

Posons

Pour retrouver le vecteur d’estimateur, il suffit résoudre le problème suivant :

### Estimation des coefficients

Contrairement à la régression linéaire, il n’existe pas de solution analytique exacte pour estimer les paramètres . L’une des méthodes d’optimisation qu’on utilise généralement pour trouver les valeurs de est la méthode de l’algorithme de Newton-Raphson qui maximise la fonction de vraisemblance.

#### Algorithme de Newton-Raphson

Il consiste à effectuer une approximation quadratique de la fonction à maximiser, en chacun des points de l’itération. Dans ce cas, si le hessien est défini négatif, on obtient le maximum de l’approximation par la condition du premier ordre sur une forme quadratique, dont on peut calculer facilement l’expression analytique parce qu’elle est linéaire. La succession de maxima ainsi obtenue donne la solution du problème. Le développement limité au second ordre de ℓ(θ) au voisinage de θ(p) est égal à :

Avec :

la valeur du paramètre à la p-ième itération.

La maximisation de cette forme quadratique par rapport à donne la condition du premier ordre :

De plus la dérivée seconde est égale à , qui est définie négative lorsque l’objectif est concave au point . Dans ce cas, on a bien un maximum local donné par les conditions du premier ordre. Dans l’ensemble, l’itération est donnée par:

#### Propriétés statistiques des estimateurs

La méthode du maximum de vraisemblance du modèle logit conditionnel peut être présentée dans des conditions très générales pour fournir des estimateurs asymptotiquement efficaces et normalement distribués. Les exemples suggèrent que l’approximation est raisonnablement bonne, même sur des échantillons assez petits. Ces résultats peuvent être utilisés pour construire des limites de confiance approximatives sur un grand échantillon et tester des hypothèses pour les paramètres.

### Test d’hypothèses

Les tests d’hypothèses couramment utilisés pour évaluer la significativité des paramètres estimés dans un modèle logit conditionnel :

#### Test du rapport de vraisemblance (Likelihood Ratio Test)

Ce test compare la vraisemblance du modèle complet (avec tous les coefficients) à celle d’un modèle restreint (avec certains coefficients fixés à zéro). Autrement dit le meilleur modèle est le modèle contenant toute les variables explicatives et le modele restrient, est celui qui ne contient que quelques variables explicatives ou aucune variable explicative.

Soient :

Soient les hypothèses suivantes :

Soient la fonction de vraisemblance du modèle restreint et la fonction de de vraisemblance du modèle complet.

La statistique de ce test est définie comme suit :

R suit une distribution de à (k-l) ddl au seuil où k est le nombre de variable explicative de modèle complet et l le nombre de variable explicative de modèle restreint.

Règle de décision du meilleur modèle.

* Si ≤ , on ne peut donc pas rejette l’hypothèse et concluons que les paramètres supplémentaires n’améliorent pas significativement le modèle et donc le modèle restreint est meilleur comparativement au modèle complet.
* Sinon rejette l’hypothèse et concluons que les paramètres supplémentaires améliorent significativement le modèle et donc le modèle complet est meilleur comparativement au modèle restreint.

#### Test du rapport de cotes (Odds Ratio Test)

Notons que ce test utilisé dans les modèles de régression logistique pour évaluer l'impact des variables explicatives sur la probabilité d’événement binaire et/ou le choix d’une alternative.

L a statistique de ce test est :

Ce test évalue l’hypothèse nulle selon laquelle l’odds ratio associé à une variable explicative est égal à 1. Si la valeur p est inférieure au niveau de signification, on rejette l’hypothèse nulle et conclut que l’odds ratio est significativement différent de 1.

Formulation d’hypothèse :

Règle de décision :

* Si , on accepte cela signifie que la variable explicative n'a pas d'effet sur la probabilité de choix d’une alternative car le coefficient de la variable explicative n’est pas significativement différent de 0.
* Si , on ne peut donc pas accepte cela signifie que la variable explicative a un effet significatif sur la probabilité de choix d’une alternative autrement la variable explicative dimunie la probabilité de choix d’une alternative donnée de (1-).
* Si , on ne peut donc pas accepte cela signifie que la variable explicative a un effet significatif sur la probabilité de choix d’une alternative autrement la variable explicative augmente la probabilité de choix d’une alternative donnée de ().

#### Test du score (Score Test)

Ce test est basé sur les scores (dérivées partielles) du modèle logit. Il évalue l’hypothèse nulle que les coefficients sont nuls.

Hypothèse :

La statistique de ce test est :

La statistique de score suit approximativement une distribution , où k est le nombre de paramètres à estimer.

Règle de décision :

* Si Score > , on rejette ​ et on conclut que les coefficients des variables explicatives sont significativement différents de zéro. Cela signifie que les variables explicatives ont un effet significatif sur la probabilité de choisir une alternative.
* Sinon, on ne rejette pas ​ et on conclut que les coefficients des variables explicatives ne sont pas significativement différents de zéro. Cela signifie que les variables explicatives n'ont pas d'effet significatif sur la probabilité de choisir une alternative.
* Si la valeur p est faible, on rejette l’hypothèse nulle. Un score élevé indique que l’alternative est plus attrayante pour l’individu, tandis qu’un score bas indique le contraire.

## Domaines d’applications

Le modèle logit conditionnel comme définie plus haut, ne modélise pas les caractéristiques de l’individu mais celle des différentes alternatives c’est-à-dire les différents choix. Toute fois note les domaines d’application de ce modèle est énorme. Parmi eux, l’on peut citer :

### Dans le domaine du marketing

Ce modèle permet d’effectuer des analyses statistiques permettant de recueillir les préférences des clients sur les caractéristiques d'un produit ou un service, c’est-à-dire comprendre les choix de produits en fonction des prix, des avis des clients, et des caractéristiques des produits. Cette étude à quel point ces facteurs influencent les décisions d'achat des consommateurs dans un environnement de commerce. Notons par exemple l’étude de **Pradeep K. Chintagunta**, **Dipak C. Jain** et **Narasimhan J. Vilcassim** "**Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data**" utilise le modèle logit conditionnel pour analyser les choix de marques des consommateurs

### Dans le domaine du transport

Si on cherche à étudier des habitudes de transport, par exemple, on aura quatre types de transports (voiture / train / avion / vélo), chacun de ces types de transport à des caractéristiques (son prix, son coût environnemental…), mais un individu ne choisira qu’un seul des quatre moyens de transport. Notons que ce type d’étude a déjà été aborder notamment par Daniel McFadden dans étude de "**The Measurement of Urban Travel Demand**" utilise bien le modèle logit conditionnel pour analyser les choix de mode de transport comme la voiture, le train et le bus. Les variables prises en compte sont plutôt le coût du voyage, le temps de trajet et les caractéristiques sociodémographiques des voyageurs.

### Dans le domaine de la médecine

Notons que cette utilisation dans ce domaine tire tous son sens dans fait qu’elle permet d’analyser les choix de traitements médicaux par les patients. Cette étude examine comment les coûts des traitements, les effets secondaires et l'efficacité attendue influencent les décisions des patients. Comme illustration d’utilisation de ce modèle dans la médecine l’on mentionne l'étude de **Kenneth E. Train** et **Wesley W. Wilson** "**Estimation of the Effect of Medical Treatment Choices on Health Outcomes**" utilise le modèle logit conditionnel pour analyser les choix de traitements médicaux par les patients. L'étude se concentre plutôt sur l'estimation de l'effet des choix de traitement sur les résultats de santé.

### Économie du travail

L'étude de **Richard Blundell** et **Thomas MaCurdy** "**Labor Supply: A Review of Alternative Approaches**" passe en revue différentes approches pour modéliser l'offre de travail. L'étude a une portée plus générale sur les différentes méthodes d'analyse de l'offre de travail.

## Avantages et limites du modèle logit conditionnel

### Avantages

L’un des grands avantages du modèle logit conditionnel est sa capacité à prédire la probabilité associée à une nouvelle modalité introduite parmi les modalités préexistantes. En effet, la probabilité associée à une nouvelle modalité s’exprime comme suit :

Où

𝑚+1 représente l’indice de la nouvelle modalité ajoutée aux 𝑚 modalités existantes.

est le vecteur de paramètres estimé avec 𝑚 premières modalités.

est le vecteur de caractéristique associée à la modalité 𝑗 ; valeur normalisée telle que . Quant à , il représente la valeur estimée du vecteur de caractéristique associée à la modalité 𝑚 + 1 , normalisée telle que .

La valeur estimée est généralement une valeur hypothétique qu’on attribue à la nouvelle modalité.

En effet, reprenons l’exemple du choix entre les modes de transport définis par les modalités suivantes : 1-Marche à pied et associés, 2- Vélo, 3- Voiture personnelle, 4- Bus. Supposons maintenant que dans cette localité on veuille introduire un nouveau mode de transport, le tramway (codé par 5). Puisque la modalité était inexistante, on ne peut donc pas déterminer avec précisions les durées de trajet et même le coût financier à cette nouvelle modalité. Il faut alors s’inspirer d’autres expériences à partir d’autres localités afin d’avoir une idée approximative sur les durées de trajet et du coût financier. Ce qui permet alors d’obtenir et par ricochet .

Par ailleurs, tout comme le modèle logit multinomial, le modèle logit conditionnel respecte la condition dite IIA (Independance of Irrelevant Alternative). En effet, en calculant le rapport de probabilité entre deux alternatives 𝑗 et 𝑙, on trouve :

Cette condition montre que les disparités entre deux réponses quelconques ne dépendent que de et du vecteur de paramètres . Selon cette condition IAA, l’introduction d’une nouvelle modalité ne modifie pas le rapport de probabilité entre deux modalités quelconques.

### Limites du modèle

Une des limites du modèle logit conditionnel est que la condition IAA peut être discutable dans beaucoup de situations. Prenons par exemple le cas d’un pays A dont les habitants pour se rendre dans un pays B ont le choix entre trois types de trajets :

1. le trajet terrestre assuré exclusivement par autocar par une compagnie nommée LandTrans,
2. le trajet maritime assuré par bateau de croisière exclusivement par une compagnie nommée SeaTrans,
3. le trajet aérien assuré par avion de ligne par une compagnie nommée AirTransOne.

Supposons par ailleurs qu’un tiers des voyageurs choisisse chaque mode de transport. Dans ce cas, la probabilité associée à chaque modalité est 0,33 et le rapport de probabilité entre les modalités deux à deux vaut 1. Supposons maintenant qu’une nouvelle compagnie de transport aérien soit introduite nommée AirTransTwo. Puisque AirTransOne AirTransTwo sont supposées proposer des services identiques, ces deux compagnies doivent donc avoir les mêmes probabilités d’être choisie. Dans ces conditions pour que la condition IAA soit toujours vérifiée (rapports de probabilité non influencés par l’ajout d’une modalité supplémentaire), il faudrait que les parts de marché de chaque modalité soient de 0.25 (soit un quart des passagers pour chaque modalité). Ce qui apparait peu réaliste dans la mesure où la présence de deux compagnies sur le même segment peut tirer les prix à la baisse dans ce segment et inciter les voyageurs à reporter leur choix sur ce mode de transport. La question peut également se poser lorsqu’on introduit un nouveau mode de transport terrestre en l’occurrence le transport ferroviaire assurée par une compagnie de nommée RailwayTrans. Il devient moins sûr que les probabilités initiales soient maintenues après cette introduction. D’une manière générale, lorsque l’hypothèse IAA n’est pas vérifiée, il faut alors penser à des modèles alternatifs qui ne sont pas fondés sur cette hypothèse comme par exemple les modèles multinomiaux séquentiels.

## Cas pratique

# Conclusion

Pour conclure, notre étude collective a mis en évidence l’efficacité du modèle de régression logistique conditionnelle dans l’analyse des préférences et des décisions humaines. Grâce à l’application de ce modèle, nous avons réussi à déterminer les éléments clés qui orientent les choix individuels et à prédire les probabilités de certaines décisions. Les résultats obtenus attestent de la robustesse du modèle et de sa capacité à être appliqué dans divers secteurs, tels que le marketing, la santé et l’économie. Nous avons également prouvé que le modèle peut être utilisé pour prendre des décisions plus informées en tenant compte des caractéristiques spécifiques de chaque alternative. En définitive, notre étude a démontré que le modèle de régression logistique conditionnelle est un outil précieux pour comprendre les mécanismes de prise de décision humaine et optimiser le processus décisionnel.