



RÉPUBLIQUE DU BENIN



UNIVERSITE DE PARAKOU (UP)

ECOLE NATIONALE DE STATISTIQUE DE PLANIFICATION ET DE DEMOGRAPHIE

GROUPE 24 : Introduction aux logiciels R, STATA, SPSS

MASTER 1

THEME : R – Régression Logistique : Logit, Probit et Tobit

Membre du groupe :

1. SABI KOTE Bani Kpénté
2. ILEKOYO Remedino

Sous la supervision de : Dr SODJINOUE Epiphane

JUIN 2025

Table des matières

Introduction	2
I. Régression Logistique (Logit)	3
1. Définition et objectifs	3
2. Forme mathématique	3
3. Estimation dans R.....	3
4. Interprétation.....	3
5. Applications	3
II. Régression Probit.....	3
1. Définition	3
2. Forme mathématique	3
3. Estimation dans R.....	4
4. Comparaison logit vs probit.....	4
5. Applications	4
III. Régression Tobit.....	4
1. Définition	4
2. Forme mathématique	4
3. Estimation dans R.....	4
4. Interprétation.....	5
5. Applications	5
IV. Comparaison des trois modèles	5
Conclusion.....	6

Introduction

En statistiques, lorsqu'on cherche à modéliser une variable dépendante qualitative ou censurée, les modèles de régression classiques ne sont pas adaptés. Les modèles logit, probit et tobit permettent de mieux gérer ces cas spécifiques. Dans le logiciel R, ces modèles sont largement utilisés en sciences sociales, économie, médecine et évaluation de projet. Cet exposé vise à présenter de manière détaillée les modèles logit, probit et tobit, leurs fondements, leur mise en œuvre dans R et les situations dans lesquelles ils sont pertinents.

I. Régression Logistique (Logit)

1. Définition et objectifs

La régression logistique est utilisée lorsque la variable dépendante est binaire (0/1, succès/échec, oui/non). Elle permet d'estimer la probabilité d'un événement en fonction d'un ou plusieurs prédicteurs.

2. Forme mathématique

On modélise la probabilité p que $Y = 1$: $\log(p / (1 - p)) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$ Cette équation est appelée fonction logit.

3. Estimation dans R

Fonction principale : `glm()`

```
model_logit <- glm(Y ~ X1 + X2, family = binomial(link =  
"logit"), data = dataset)  
summary(model_logit)
```

4. Interprétation

Les coefficients représentent les log-odds. On peut transformer les coefficients en odds ratios avec :

```
exp(coef(model_logit))
```

5. Applications

- Études médicales : facteurs influençant une maladie.
- Sciences sociales : probabilité de voter, d'abandonner un programme, etc.

II. Régression Probit

1. Définition

La régression probit est aussi utilisée pour modéliser une variable binaire. Elle suppose que la probabilité de succès suit une distribution normale cumulative.

2. Forme mathématique

$P(Y = 1|X) = \Phi(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)$

Φ est la fonction de répartition de la loi normale standard.

3. Estimation dans R

```
model_probit <- glm(Y ~ X1 + X2, family = binomial(link  
= "probit"), data = dataset)  
summary(model_probit)
```

4. Comparaison logit vs probit

- Logit utilise la fonction logistique ; probit utilise la normale standard.
- Résultats similaires, mais logit est plus interprétable (odds ratios).

5. Applications

- Modèles de choix discrets.
- Économie, marketing, psychologie.

III. Régression Tobit

1. Définition

Le modèle Tobit est utilisé quand la variable dépendante est censurée (partiellement observée). Par exemple, des valeurs en dessous ou au-dessus d'un seuil ne sont pas observées précisément (comme les revenus censurés à 0).

2. Forme mathématique

$Y^* = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$, avec $\varepsilon \sim N(0, \sigma^2)$

Mais on observe :

$Y = Y^*$ si $Y^* > L$

$Y = L$ si $Y^* \leq L$ (souvent $L = 0$)

3. Estimation dans R

```
library(AER)  
model_tobit <- tobit(Y ~ X1 + X2, left = 0, data = dataset)  
summary(model_tobit)
```

4. Interprétation

Le modèle estime les effets sur la variable latente Y^* . Plus complexe à interpréter, mais très utile pour données censurées.

5. Applications

- Analyse de revenus (non négatifs).
- Durée de chômage (censurée à 0).

IV. Comparaison des trois modèles

Modèle	Type de variable dépendante	Fonction de lien	Cas typique
Logit	Binaire (0/1)	Logistique	Participation à un programme
Probit	Binaire (0/1)	Normale	Choix d'un produit
Tobit	Continue censure	Identité (latente)	Revenus censurés à 0

Conclusion

Les modèles logit, probit et tobit sont essentiels pour l'analyse de données non-linéaires ou censurées. Chacun répond à un besoin spécifique : binaire pour logit/probit, censuré pour tobit. Dans R, leur implémentation est accessible via les fonctions `glm()` et `tobit()`. Maîtriser ces modèles permet de mieux comprendre des phénomènes complexes dans de nombreux domaines, notamment en suivi-évaluation, où l'analyse de comportements ou de résultats partiels est fréquente.