

Synthèse Cours Econometrie des variables qualitatives 1.

Chapitres 1-2

- 1) Vérifier que la variable à expliquer porte sur une caractéristique de type qualitatif.
- 2) Nettoyer la base de données et analyse des "corrélal^{on}, indépendan^{ce}"
entre variables
- attention ne supprimer forcément les valeurs atypiques selon les sujets étudiés (estimation avec et sans puis comparaison des résultats)

3)

2 cas de figures

Le phénomène
à étudier est de type binaire
(Oui/Non)

La variable qualitative
a plusieurs modalités
ordonnées

Plusieurs modalités ordonnées → Chapitre 2

- 1) Vérifier qu'il y a assez d'effectifs dans chacune des catégories de Y → si ce n'est pas le cas regrouper Y
- 2) Estimation sous hypothèse d'homoscedasticité des erreurs

→ vérifier que les seuils sont significatifs
→ si ce n'est pas le cas regrouper les catégories de Y

→ vérifier l'égalité des pentes au niveau global
(cglm, cgm R)
glogit2

oui → vérifier hyp d'homos. erreurs
Non → rechercher la variable ou les variables post Pb.

Si Pas Égalité des pentes → estimation possible avec $\text{clm}(R)$ et gologit2 (Stata) d'un modèle relevant cette hypothèse

Par contre, pas possible de prendre en compte l'hétéroscédasticité des erreurs
 • sous Stata, calcul des odd ratios et effets marginaux, sous cmp → odd ratios

si égalité des pentes → • estimation possible avec polr , vglm , $\text{clm}(R)$ et gologit2 (Stata)

- odd ratios, effets marginaux (vglm , gologit2)
- Vérification de l'hétéroscédasticité des erreurs avec $\text{vglm}(R)$, vglm (Stata) avec sources

↓
 • Dans les cas de figure, calcul du R^2 Mac Fadden et Ratio de vraisemblance
 • Possibilité sous polr de calculer la qualité de prédiction du modèle.

2 modalités (Oui/Non) / Y → Chapitre 1.

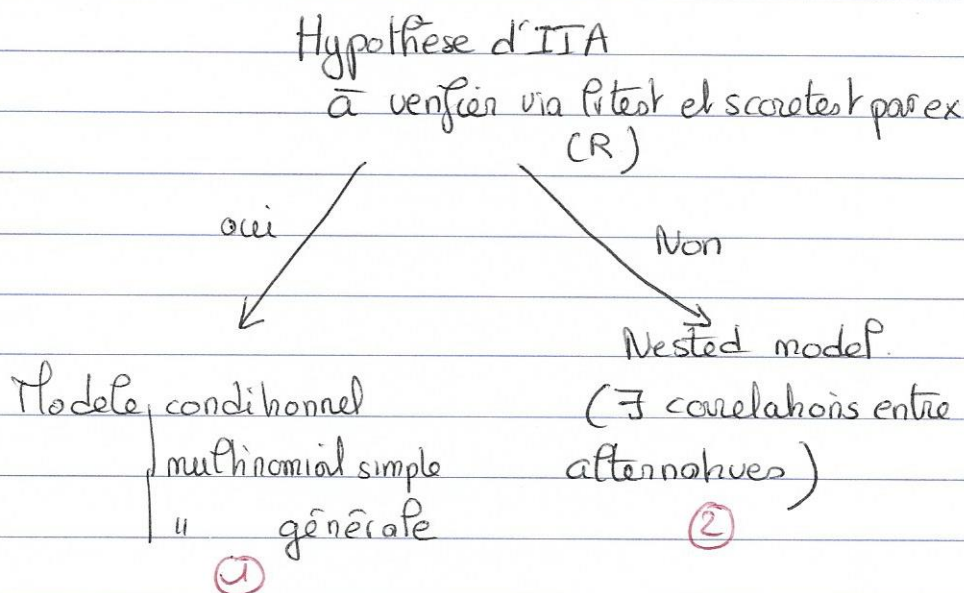
- Estimation du modèle sous hypothèse d'homoscedasticité des erreurs → vérification de la multicolinéarité. (possibilité d'utiliser la méthode du stepwise si "correlation" (idem pour le modèle ordonné avec fonction polr)
- vérification de l'ajustement du modèle (Ratio de vraisemblance)
- si logit → odd ratios
- effets marginaux
- qualité d'ajustement R^2 Mac Fadden
- Tx d'erreur, taux de sensibilité, spécificité
- Observations influençant de manière significative l'estimation ?
- Vérification de l'hypothèse d'homoscedasticité des erreurs si oui, cf ci dessus Sinon, refaire les interprétations ($\text{fitglm}(R)$)

Attention au début mettre toutes les variables explicatives, comme source d'hétéroscédastité

Retenir celles où la p-value $> 0,05$ et renverser l'hypothèse d'homoscédastité des erreurs
(Ne pas oublier de vérifier le vif).

Chapitre 3 : Modèle multinomial non ordonné

→ l'individu est supposé évaluer \neq alternatives (choix) mutuellement exclusifs (et identiques pour les \neq individus)



(1) si estimation d'un modèle conditionnel et générale, la base doit être sous format "long".

R: fonction `mlogit`, data de la library `mlogit`
Stata: `reshape`.

→ calcul odd ratio si `logit` (`mlogit`)

→ effets marginaux

→ R: `lme4` `lmer`.

→ Estimation via `mlogit` ou `aplogit` (stata)

Attention: vglm possible sur un modèle multinomial simple sur base 7 transformée (idem `mlogit`/stata)

→ vérification de l'hypothèse d'homoscédastité des erreurs (`mlogit`) mais sans source de l'hétéroscédastité.

②

Possibilité sous R avec nllogit de tester l'hypothèse IIA
via prtest et scoretest. Si $p \leq 0,05 \rightarrow$ IIA refusée
 \rightarrow Nested model

Fonction nllogit sous stata avec possibilité d'avoir des
variables explicatives \neq pour les 2 niveaux.

Définition au préalable de la structure de l'arbre via
nllogitgen

LR test / IIA donné pour de l'estimation (Idem Rade de
vraisemblance)