

TD2

Econométrie des variables qualitatives 1 Modèles multinomiaux

Exercice : Le choix des marques et ses déterminants

Vous disposez pour 735 personnes de leur choix en termes de marque (Marque1, Marque2, Marque3) ainsi que de leurs caractéristiques et du prix de ces 3 marques. Les données sont disponibles dans le fichier Marque.xls (sous Madoc).

Choice : choix de la personne en termes de marque (Marque1, Marque2, Marque3)

Prix.Marque1 : prix de la marque 1 (en €)

Prix.Marque2 : prix de la marque 2 (en €)

Prix.Marque3 : prix de la marque 3 (en €)

Femme = 1 si la personne interrogée est une femme, 0 sinon

Age : Age de personne interrogée (en année)

Questions :

- 1) Sous quel type de format est enregistrée la base de données? Importer la base sous le logiciel R. Nommer la base : Marque
- 2) Vérifier s'il n'existe pas de valeurs atypiques pour la variable Age
- 3) Transformer la variable de type qualitatif en facteur sous R.
- 4) Quelles sont la ou les variables liée(s) au choix? Quelles sont la ou les variables liée(s) aux individus ?
- 5) Comment appelle-t-on un modèle de choix où on ne prend que les variables liées aux individus ? Faire l'estimation d'un tel modèle avec la fonction mlogit et vglm et interpréter les résultats obtenus. Au préalable, pour la fonction mlogit, transformer la base Marque dans le format adéquat (nom de la nouvelle base : Marqueb). Combien de lignes doit avoir cette nouvelle base ? Quel choix est pris par défaut comme référence pour l'interprétation des résultats ?
- 6) Estimer le modèle avec l'ensemble des variables explicatives en utilisant la fonction mlogit. Comment appelle-t-on ce type de modèle ? Le fait d'avoir rajouter la variable Prix comme variable explicative est-il pertinent ?
- 7) En prenant le modèle retenu à la question 5, calculer les odds-ratios des variables explicatives (via la fonction vglm). Interpréter les résultats.
- 8) En prenant le modèle retenu à la question 5, représenter graphiquement les effets de la variable Age sur la probabilité d'avoir les différentes marques (via la fonction vglm)
- 9) Estimer le modèle de la question 5 (via la fonction mlogit) en tenant compte de l'hétéroscédasticité des erreurs. Doit-on conserver ce modèle ? Comparer les résultats de ce modèle au modèle où les erreurs sont supposées homoscedastiques.
- 10) Supposons maintenant que le choix de la personne se fasse entre la marque 2 et un ensemble de marques constitué par les marques 1 et 3. Pour ce faire, réintroduire le prix comme variable explicative et prendre la marque 2 comme référence. L'hypothèse IIA est-elle vérifiée ?