

2014

APDT

Robin Leconte &  
Yohan Planche

# **ANALYSE DES CHOIX MODAUX**

Explication des choix modaux en fonction des variables explicatives. Analyse statistique et régressions linéaires.

## Estimation du modèle de choix modal agrégé utilisé lors de la séance 7

$$V_{TC} - V_{VP} = \alpha + \beta * tg_{VP} / mot + \theta * (tg_{TC} * mot) + \lambda * densité$$

### Résultats et interprétation

```
Call:
mlogit(formula = Mode ~ 0 | tpgenXmot + tpgen_mot + densite,
       data = CM, reflevel = "VP", method = "nr", print.level = 0)

Frequencies of alternatives:
      VP      TC
0.84948 0.15052

nr method
6 iterations, 0h:0m:1s
g'(-H)^-1g = 0.000284
successive fonction values within tolerance limits

Coefficients :
              Estimate Std. Error t-value Pr(>|t|)
TC:(intercept) -2.19417147  0.10995316 -19.956 < 2.2e-16 ***
TC:tpgenXmot   -0.04739041  0.00312468 -15.166 < 2.2e-16 ***
TC:tpgen_mot    0.03203769  0.00210765  15.201 < 2.2e-16 ***
TC:densite      0.00349533  0.00023302  15.000 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -3440.4
McFadden R^2: 0.11093
Likelihood ratio test : chisq = 858.5 (p.value = < 2.22e-16)
```

La régression linéaire nous permet d'estimer le modèle utilisé pour le choix modal. Il s'agit de voir quelle formule correspond au mieux à l'utilité des TC et donc quels paramètres influencent le choix de telle ou telle alternative modale.

Les coefficients de la régression sont acceptables que lorsqu'ils sont cohérents. L'utilité des TC évolue de façon contraire à la motorisation et au temps généralisé TC, mais évolue conjointement avec le temps généralisé VP et la densité ; comme l'atteste l'analyse des signes des différents coefficients.

Pour toutes les régressions effectuées, nous comparerons le R<sup>2</sup> de McFadden et le log de vraisemblance pour trouver la fonction d'utilité la plus performante. Dans le cas de modèles ne différant que par l'ajout d'une variable, nous pourrions utiliser le test du ratio de vraisemblance.

## Comparaison de différentes fonctions d'utilité

Estimation avec un paramètre spécifique ou un paramètre générique

$$V_{ic} - V_{vp} = \beta * (tpsgen_{TC} - tpsgen_{VP}) + \text{mot} + \text{densité}$$

ou

$$V_{ic} - V_{vp} = (\alpha_{TC} - \alpha_{VP}) * tpsgentc + (\beta_{TC} - \beta_{VP}) * tpsgenvp + \text{mot} + \text{densité}$$

Nous comparons ici deux fonctions. D'un point de vu global, elles contiennent les mêmes informations, à la différence que les temps généralisés sont agrégés dans le premier cas et désagrégés dans le second. Il s'agit de comparer ces modèles pour la base de notre travail. Les 3 modèles utilisés ici contiennent chacun les 4 mêmes variables. Les R2, s'ils sont différents entre eux, témoigneront donc d'une différence significative de performance entre modèles.

Principaux résultats des régressions sous R:

Fonction	de base	spécifique	générique
<b>R2</b>	0.11093	0.27526	0.26915
<b>LL</b>	-3440.4	-2804.5	-2828.2
<b>chi2</b>	858.5	2130.3	2083
<b>Statistiques t</b>	OK	OK	OK
<b>Signe des coefficients</b>	OK	OK	OK

Les trois fonctions d'utilité sont acceptables puisque la probabilité du t de Student permet de rejeter l'hypothèse nulle et que les coefficients des variables sont de signes cohérents.

La statistique R2 nous permet de choisir le meilleur modèle. Ici, il s'agit de la fonction spécifique – model désagrégé – qui offre le meilleur R2 : 27% contre 26 et 11%.

De fait, nous utiliserons ce modèle de base pour la suite du travail : model auquel nous ajouterons successivement des variables explicatives afin de déterminer quelles variables sont influentes.

## Incidences de quelques variables socio-démographiques sur les résultats du modèle

Il s'agit dans cette partie de tester l'influence de variables telle que l'âge ou le sexe, indépendamment d'autres. Pour cela, il suffit d'ajouter cette variable à la fonction d'utilité et d'observer son effet sur les statistiques de la régression linéaire. Par compositions successives, nous pourrions ainsi définir la fonction utilité qui nous semble la plus performante.

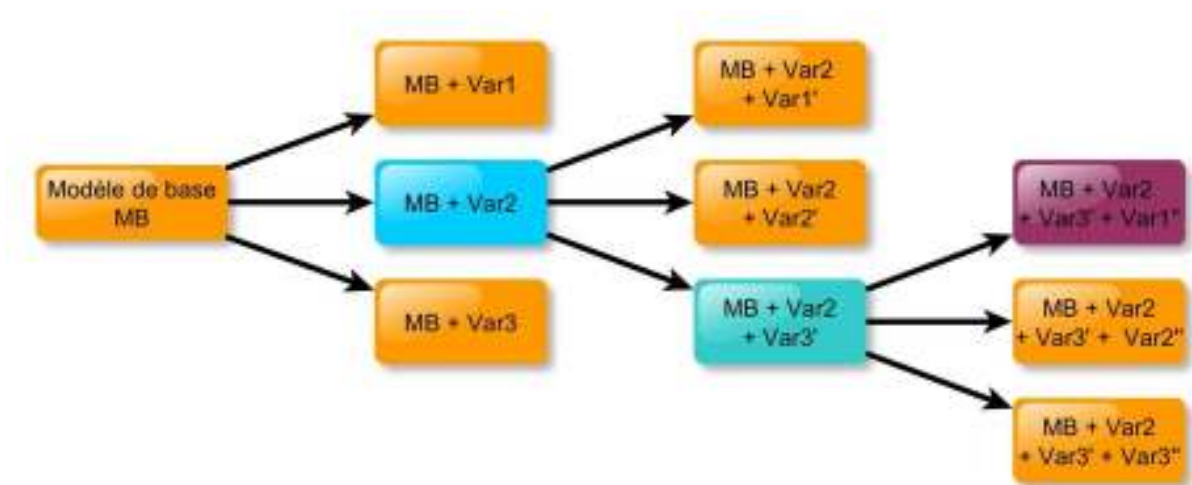
### Démarche

Comme nous l'avons décrit précédemment, la fonction d'utilité utilisant des paramètres spécifiques donne les meilleurs résultats. Elle sert ici de modèle de base, à laquelle on ajoute les variables « sexe », « âge » ou « permis » soit tour à tour soit toutes ensemble.

Nous avons utilisé le modèle de base auquel nous avons ajouté différentes variables. L'idée étant de faire plusieurs itérations. A chaque étape nous avons comparé les régressions, nous permettant de conserver la variable offrant le meilleur R2 (sous réserve de la cohérence des autres paramètres).

Pour tester l'apport de chaque nouvelle variable, nous utiliserons le test du ratio de vraisemblance.

#### Illustration du protocole :

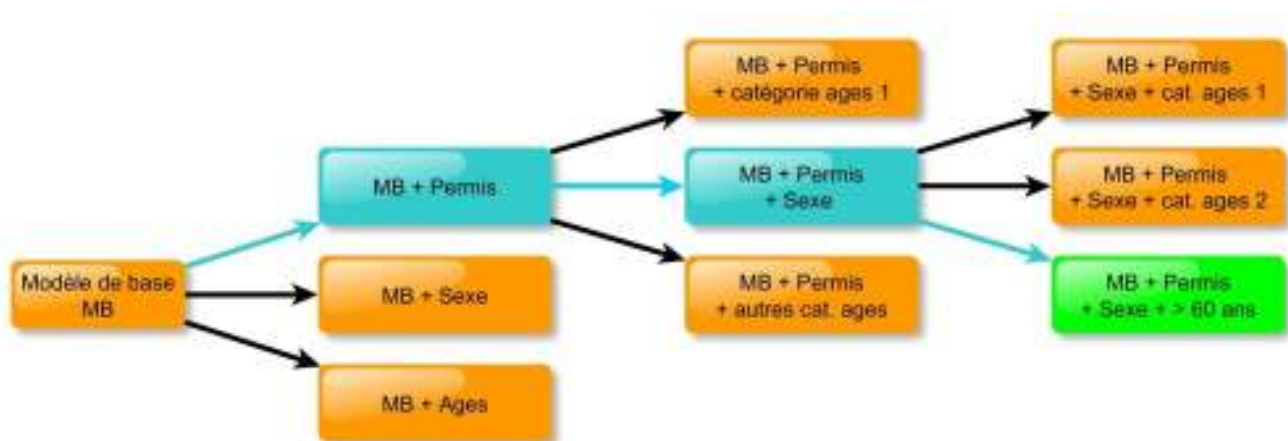


Ci-dessous ce trouve le tableau récapitulatif des différents tests effectués. Nous avons volontairement conservé uniquement les tests pertinents. D'autres essais ont été faits, avec les âges par exemple, mais les résultats n'étant pas concluants nous les avons exclus. Les autres variables testées, comme l'âge absolu ou les autres catégories d'âges, menaient à l'obtention de t de Student non satisfaisants et ont de ce fait été éliminées dans la définition d'un modèle optimal.

Résultats synthétisés :

Fonction	Modèle de base MB	MB + permis	MB +sexe	MB +permis + sexe	MB + Permis + sexe + >60ans
R2	0.27526	0.31643	0.30109	0.33079	0.33146
LL	-2804.5	-2645.2	-2704.5	-2589.6	-2587
chi2	2130.3	2449	2330.3	2560.1	2565.3
Statistiques t	OK	OK	OK	OK	OK
Signe des coefficients	OK	OK	OK	OK	OK
ratio de vraisemblance		318.6000	948.4000	111.2000	5.2000
chi2 à K-a degrés de liberté		3.8415	3.8415	3.8415	3.8415
TEST du gain de vraisemblance		82.93724204	246.885375	28.9473362	1.353652412

Schéma de synthèse :



La configuration optimale est donc : **modèle désagrégé (tps généralisé, motorisation et densité) + Permis de conduire + Sexe + Catégorie âges > 60 ans**. Autrement dit, ces 4 variables expliquent à elles seules 33 % des choix modaux. La régression linéaire étant la suivante :

```

Coefficients :
                Estimate Std. Error t-value Pr(>|t|)
TC:(intercept)  0.52648757  0.20770171  2.5348  0.01125 *
TC:tpsgentc     -0.04980034  0.00401796 -12.1953 < 2e-16 ***
TC:tpsgenvp     0.08179596  0.00613490  13.3329 < 2e-16 ***
TC:densite      0.00330016  0.00030795  10.7167 < 2e-16 ***
TC:motorisation_nbVP.nb18. -3.27522671  0.12937334 -25.3161 < 2e-16 ***
TC:permis_VP_1_oui_0_non. -1.84611075  0.12501751 -14.7668 < 2e-16 ***
TC:SEXE         0.78297467  0.07499606  10.4402 < 2e-16 ***
TC:age_60et.ans -0.63932655  0.29685087  -2.1537  0.03126 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

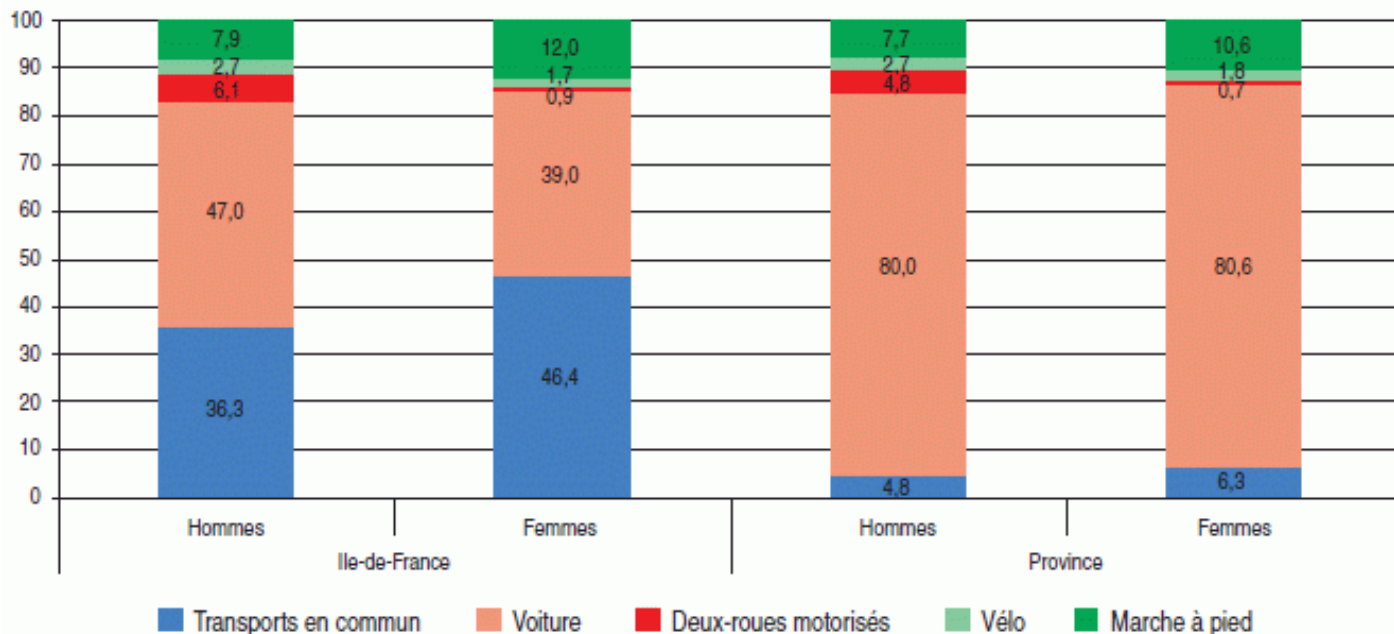
log-Likelihood: -2587
McFadden R^2: 0.33146
likelihood ratio test : chisq = 2565.3 (p.value = < 2.22e-16)

```

Essayons maintenant de comprendre pourquoi ces variables plutôt que d'autres expliquent la répartition modale.

- ➔ **Les temps généralisés** : Il s'agit d'une variable explicative évidente. Si un temps de parcours (généralisé) est plus avantageux en VP plutôt qu'en TC, l'individu motorisé aura tendance à utiliser son véhicule personnel. Si l'offre TC est plus performante, l'individu utilisera les transports en commun.
- ➔ **La densité** : Le facteur est important. L'offre TC est meilleure dans les zones denses, proche des centres ville. De fait, les habitants disposent d'un système de transport abouti qu'ils utilisent. A l'inverse, les zones à faible densité – et de fait, en périphérie des agglomérations – ont une offre de TC réduite : l'usage de la VP est privilégié.
- ➔ **La motorisation** : Un individu motorisé peut choisir son mode de transport, alors qu'un individu non-motorisé se reportera essentiellement sur les TC et modes doux.
- ➔ **Le permis de conduire** : Idem. Cependant, un individu titulaire du permis de conduire mais non-motorisé peut toujours jouir des services de location de véhicules ou bien de l'autopartage type « Autolib ».
- ➔ **Le sexe** : Le sexe a une influence sur le choix modal. En effet, les hommes utilisent d'avantage la voiture particulière que les femmes, ayant tendance à se reporter sur les TC et modes doux. Ci-dessous, la répartition des choix modaux en France métropolitaine en fonction du sexe (*Champ : actifs de plus de 15 ans ayant un lieu d'emploi fixe situé dans un rayon de 80 km de leur domicile et résidant en IdF. Source : Insee - SOeS, ENT D 2008*).

## Répartition des déplacements domicile-travail selon le mode principal de transport utilisé, le lieu de résidence et le genre des actifs (en %)



Bien que ces statistiques concernent l'IdF, nous pouvons supposer que cette répartition est transposable, dans une certaine mesure, à l'agglomération lyonnaise.

Remarquons aussi, à titre indicatif, que le constat n'est plus vrai en province. Notamment du fait que l'offre TC qui y est limitée.

➔ **L'âge** : Les 60 ans et plus, ont tendance à utiliser leur voiture particulière plutôt que les transports en commun. Quelles explications ? Cette catégorie d'âge est essentiellement constituée de retraités. Les motifs de déplacements sont donc différents par rapport à ceux des actifs. Tout d'abord, cette catégorie d'âge est moins mobile que celle des actifs. Ensuite, les déplacements sont plutôt orientés loisir, famille et commerces. Et la recherche d'un certain « confort » peut expliquer ce choix modal. En effet les TC sont relativement inconfortables pour certaines personnes âgées.

Pour compléter cette étude il serait nécessaire de déterminer les variables complémentaires qui expliqueraient les 60 % restant. Nous pouvons tout de même donner quelques pistes telles que les revenus, les CSP ou encore une analyse plus fine du maillage des transports en commun.