

# APDT rendu final

Edouard PHILIPP Raphaël WALKER

```
Coeff_de_redressement Mode MODE_vpoond.vppass mode_ML.TC.VPOOND.VPPASS NBR_VP_DISPO permis_VP_1_oui_0_non.
:      188.65  VP          VPOOND          VPOOND          VPOOND          2          1
:      188.65  VP          VPOOND          VPOOND          VPOOND          2          1
:      188.65  VP          VPPASS          VPPASS          VPPASS          2          1
:      188.65  VP          VPPASS          VPPASS          VPPASS          2          1
:      188.65  VP          VPOOND          VPOOND          VPOOND          2          1

nbr_pers_18ans_et... motorisation_nbVP.nbi8. motorisation SEXE STATUT AGE age_0.17ans age_18.25ans
:      2          1          0.715  1          1  42          0          0
:      2          1          0.562  1          1  42          0          0
:      2          1          0.715  2          1  34          0          0
:      2          1          0.562  2          1  34          0          0
:      2          1          0.715  1          1  28          0          0

age_26.35ans age_36.45ans age_46.55ans age_60et.ans tpsgen.VP tpsgen.TC densite tpsgenMot tpsgen not
:      0          1          0          0          29.384  34.236 254.120 24.47874 41.05650
:      0          1          0          0          24.646  32.747 48.557 18.40381 48.85409
:      1          0          0          0          29.384  34.236 254.120 24.47874 41.05650
:      1          0          0          0          24.646  32.747 48.557 18.40381 48.85409
:      1          0          0          0          21.155  41.644 25.686 29.77546 29.58741

tpsgenvp tpsgento
:      29.384  34.236
:      24.646  32.747
:      29.384  34.236
:      24.646  32.747
:      21.155  41.644
```

Illustration 1 : Échantillon des données disponibles

Lors de ce TP nous avons essayé d'estimer au mieux le choix modal réel à partir de plusieurs paramètres. On s'intéressera ici à l'utilité des TC. Pour cela, nous utilisons un modèle Logit, ce modèle est assez simple et est notamment caractérisé par l'hypothèse d'indépendance des choix non pertinents ou IIA.

Cette hypothèse correspond au fait que le ratio de probabilité de choix entre 2 modes ne sera pas modifié par l'apparition d'un 3<sup>e</sup> mode. On peut aussi l'interpréter en disant que les élasticités croisées sont les mêmes pour tous les modes, c'est à dire que si un mode perd en utilité, tous les autres gagneront le même pourcentage de probabilité de choix. Cette hypothèse est réaliste dans certains cas mais fautive dans d'autres. L'hypothèse d'indépendance des choix non pertinents n'a d'avantages que lorsqu'elle reflète la réalité.

Ici, on ne s'intéresse qu'à 2 choix de modes de déplacement, les TC et la VP. On suppose donc que les autres choix sont non-pertinents, l'IIA reflète donc la réalité. Un modèle logit est donc adapté à ce TP.

Nous avons choisi d'approximer l'utilité des transports en commun en effectuant plusieurs régressions linéaires. On obtiendra donc des utilités fonctions linéaires de plusieurs paramètres. Nous utiliserons plusieurs outils statistiques comme le  $R^2$  de MacFadden et le logarithme de la vraisemblance pour comparer l'efficacité de nos régressions, on pourra ainsi déterminer laquelle de nos fonctions d'utilité est la plus performante.

## Première régression : tpгенXmot, tpген\_mot, densité

Nous avons commencé par évaluer le modèle utilisé lors de la séance précédent.

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

Nous demandons donc à R de nous faire une régression linéaire en fonction de tpгенXmot, tpген\_mot, et de la densité.

La probabilité de choisir les TC est sensée augmenter si le temps de parcours en TC diminue et si la motorisation diminue, le coefficient lié au paramètre tempsTC\*motorisation est donc sensé être négatif. On retrouve bien ce résultat dans le résultat de la régression comme le montre le tableau ci-dessous.

Le choix du mode TC est plus probable si le temps de parcours en voiture augmente et si la motorisation diminue. On s'attend donc à retrouver un coefficient positif devant le paramètre tempsVP/motorisation. De même, on trouve un coefficient positif pour la densité

Les résultats obtenus sont donc cohérents, en accord avec nos hypothèses. Intéressons-nous maintenant aux indicateurs statistiques.

```
> CM<- mlogit.data(choix_modal, varying = c(19:20), choice = "Mode", shape = "wide")
> ml.choix<- mlogit(Mode ~ 0|tpгенXmot+tpген_mot+densite,reflevel='VP',CM )
> summary(ml.choix)

Call:
mlogit(formula = Mode ~ 0 | tpгенXmot + tpген_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 function values within tolerance limits

Coefficients :
              Estimate Std. Error t-value Pr(>|t|)
TC:(intercept) -2.19417147  0.10995316 -19.956 < 2.2e-16 ***
TC:tpгенXmot    -0.04739041  0.00312468 -15.166 < 2.2e-16 ***
TC:tpген_mot     0.03203769  0.00210769  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)
```

*Illustration 2 : Indicateurs statistiques correspondant à la 1ère régression.*

La probabilité de l'hypothèse nulle est extrêmement faible (<2,2\*10<sup>-16</sup>) pour tous les paramètres que nous avons testés, on peut donc rejeter l'hypothèse nulle. Nous allons donc garder ces variables dans la régression suivante. La valeur du R<sup>2</sup> de McFadden est assez faible, le log de vraisemblance aussi, nous allons essayer d'améliorer ces résultats statistiques en faisant de nouvelles régressions.

## Deuxième régression : permis\_VP\_1\_oui\_0\_non, tpgenXmot, tpgen\_mot, densité

Nous avons vu avec la dernière régression que les temps généralisés, la motorisation et la densité influencent le choix modal, mais qu'ils ne suffisent pas à la modéliser suffisamment au vue des indicateurs statistiques.

Nous avons donc choisi d'ajouter un paramètre à la régression linéaire pour la compléter et la rendre plus performante. Comme peu de trajets se font en tant que passager (par opposition à conducteur) d'une voiture, il nous semble logique que la possession d'un permis voiture ou non influence le choix modal.

Nous avons donc ajouté à la régression un paramètre qui prend la valeur 1 si l'individu concerné possède le permis en 0 sinon. Un individu qui a le permis a plus de chance d'utiliser les VP qu'un individu sans permis, le coefficient attaché au paramètre permis est donc sensé être négatif.

```
> ml.choix2<- nlogit(Mode ~ 0 | permis_VP_1_oui_0_non.+tpgenXmot+tpgen_mot+densite,reflevel='VP',CM )
> summary(ml.choix2)

Call:
nlogit(formula = Mode ~ 0 | permis_VP_1_oui_0_non. + 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:0s
g'(-H)^-1g = 0.00308
successive function values within tolerance limits

Coefficients :
              Estimate Std. Error t-value Pr(>|t|)
TC:(intercept)  0.47748450  0.14899048   3.2048  0.001952 **
TC:permis_VP_1_oui_0_non. -3.22597902  0.11138479 -28.9570 < 2.2e-16 ***
TC:tpgenXmot    -0.04719457  0.00398538 -13.9407 < 2.2e-16 ***
TC:tpgen_mot     0.03490629  0.00232933  14.9856 < 2.2e-16 ***
TC:densite      0.00411221  0.00025043  16.4205 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -2951.2
McFadden R^2:  0.29796
Likelihood ratio test : chisq = 1897 (p.value = < 2.22e-16)
```

*Illustration 3 : Indicateurs statistiques correspondant à la 2<sup>e</sup> régression.*

On trouve comme on s'y attendait un coefficient négatif devant le paramètre de permis. De plus, la t-value nous permet de rejeter l'hypothèse nulle sur ce paramètre. On obtient un meilleur résultat pour le R<sup>2</sup> et pour le log de vraisemblance que précédemment. Cette régression est meilleure que le précédent d'un point de vue statistique.

Cependant, ajouter un paramètre à une régression améliore de manière artificielle la valeur du R<sup>2</sup>, il faudra donc éviter de rajouter trop de paramètres. Ici comme le log de vraisemblance a augmenté de manière considérable (plus de 10), l'amélioration par rapport à la régression précédente n'est vraisemblablement pas seulement due à un effet de volume.

Un regard plus poussé sur les valeurs des coefficients nous fait remarquer qu'ils n'ont globalement pas changé de la première régression à la deuxième à l'exception de celui de la constante qui a changé de signe. Nous remarquons par ailleurs que l'on ne peut pas rejeter l'hypothèse nulle pour la constante.

### 3è régression, ratio voiture/nbadultes pour la motorisation

La régression que nous avons fait par la suite se place dans la continuité de la précédente. Néanmoins, de ce cas, on utilise non pas le fait d'avoir le permis pour refléter l'influence de l'accès à la voiture des individus mais le ratio du nombre de véhicules du ménage sur le nombre de personnes de plus de 18 ans dans le ménage.

Nous voulons ainsi repérer grâce aux indicateurs statistiques fournis par RR, laquelle des deux méthodes est la plus adaptée à notre cas.

```
> ml.choix3<- mlogit(Mode ~ tpsgen | permis_VP_1_oui_0_non. + motorisation_nbVP.nb18. + densite, relevel="VP", CH )
> summary(ml.choix3)

Call:
mlogit(formula = Mode ~ tpsgen | permis_VP_1_oui_0_non. + motorisation_nbVP.nb18. +
  densite, data = CH, relevel = "VP", method = "nr", print.level = 0)

Frequencies of alternatives:
      VP      TC
0.84948 0.15052

nr method
6 iterations, 0h:0m:0s
g'(-H)^-lg = 4.81E-07
gradient close to zero

Coefficients:
              Estimate Std. Error t-value Pr(>|t|)
TCs (intercept)  2.43430229  0.14343357  16.972 < 2.2e-16 ***
tpsgen          -0.04571119  0.00395782  -11.550 < 2.2e-16 ***
TCspermis_VP_1_oui_0_non. -2.05412449  0.12022487  -17.086 < 2.2e-16 ***
TCsmotorisation_nbVP.nb18. -3.03529983  0.12410091  -24.458 < 2.2e-16 ***
TCdensite       0.00385482  0.00029539   12.876 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -2671.2
McFadden R^2:  0.30971
Likelihood ratio test : chisq = 2396.9 (p.value = < 2.22e-16)
```

*Illustration 4: Indicateurs statistiques correspondant à la 3è régression.*

On note que le log de vraisemblance est meilleur dans ce cas (-2671 contre -2951), de même que le  $R^2$  de MacFadden (0.31 contre 0.24). On conclut donc des deuxième et troisième régressions que statistiquement, le ratio du nombre de véhicules sur celui des personnes de plus de 18 ans est un meilleur indicateur que la possession du permis pour traduire l'influence de l'accès individuel à la voiture. Nous garderons donc le ratio du nombre de véhicule sur le nombre de personnes de plus de 18 ans dans nos prochaines régressions.



## 5è Régression : comparaison temps généralisés génériques/spécifique

Dans cette régression, nous nous sommes intéressés à la forme optimale pour le temps généralisé dans les régressions. Ainsi, il s'agit de déterminer si la variable spécifique temps généralisé est préférable ou d'avoir les deux variables génériques, temps généralisé pour la voiture personnelle et les transports en commun.

On note entre la régressions ci-dessous et la régression ml.choix3, une variable temps généralisé spécifique l'autre avec deux variables génériques, que la régression avec les deux variables génériques donne de meilleurs résultats statistiquement.

En effet, le log de likelihood est plus élevé dans ce cas (-2645 contre -2671). Cette différence de log de likelihood est supérieure à 10. Un écart supérieur à 10 indique que la régression avec deux variables est préférable à celle avec une variable spécifique et qu'il ne s'agit pas simplement d'un effet de volume lié au fait d'avoir une variable en plus. De même, au regard des  $R^2$  de MacFadden (0.3164 contre 0.3097), les deux variables génériques sont préférables.

Enfin, du point de vue des coefficients, le choix des deux variables génériques est renforcé par le fait que leurs coefficients ne sont pas opposés (0.08 et -0.05). Le fait d'avoir ces deux variables apporte ainsi une information. De plus, le signe de ces coefficients est cohérent, une augmentation du temps pour la voiture personnelle augment la fréquence du mode TC alors qu'une augmentation du temps des TC diminue la fréquence de ce mode. On poursuit ainsi avec les deux variables génériques pour les temps généralisés VP et TC.

```
> ml.choix3b1a4 = mlchoice3(Choice = 0|transport+personnelle, VP, TC, mode, motorisation_nbVT.nb18, edeptime, reflevel="VP", PM = 1)
> summary(ml.choix3b1a4)

Call:
mlchoice3b1a4 = Mode = 0 | transport + personnelle + mode VP TC mode +
  motorisation_nbVT.nb18. + edeptime, data = CM, reflevel = "VP",
  method = "nl", print.level = 0)

Frequencies of alternatives:
      VP      TC
0.34948 0.35052

no method
0 iterations, 0h:0m:0s
q1-H1^2-tq = 4.31E-07
gradient close to zero

Coefficients:
              FAMILIAR  Std. Error t-Statistic   Prob[>|t|]
TC:(intervenir)    1.25535527    0.16230867    11.132 < 2.2e-16 ***
TC:transport       0.07546027    0.00001203    13.040 < 2.2e-16 ***
TC:transport      -0.04906712    0.00090004   -10.553 < 2.2e-16 ***
TC:transport VP 1 ou 0 non- -2.09210887    0.12169175   -17.206 < 2.2e-16 ***
TC:motorisation_nbVT.nb18. -3.06551720    0.12170224   -20.583 < 2.2e-16 ***
TC:edeptime       0.00421120    0.00030422    11.344 < 2.2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

log-likelihood: -2645.3
Maximum H^2: 0.81545
likelihood ratio test : chisq = 2360 (p-value = < 2.22e-16)
```

Illustration 6: Indicateurs statistiques correspondant à la 5è régression.

## 6è régression, on ajoute le sexe :

Nous nous demandons à présent si le sexe est un facteur déterminant dans le choix d'un mode de transport. Dans la régression suivant, on ajoute le sexe comme facteur, un individu a une valeur dans le paramètre sexe de 1 ou 2 selon qu'il soit homme ou femme.

Ce paramètre est à retenir. En effet, le log de vraisemblance augmente de plus de 10, le  $R^2$  de MacFadden passe à 0.33 et la probabilité du t de Student est bonne à  $2 \times 10^{-16}$ . L'amélioration de la régression n'est ainsi pas due à un effet de volume et on retient ainsi ce paramètre.

```
> ml.choix7C<- mlogit(Mode ~0|(tpogenvp+tpagente+permis_VF_1_oui_0_ non. +SEXE+motorisation_nbVF.nb15.+densite, data=VF1, CH )
> summary(ml.choix7C)

Call:
mlogit(formula = Mode ~ 0 | (tpogenvp + tpagente + permis_VF_1_oui_0_ non. +
SEXE + motorisation_nbVF.nb15. + densite, data = CH, relevel = "VF",
method = "nr", print.level = 0)

Coefficients of alternatives:
VF      TC
0.26948 0.15052

nr method
V Likelihood, H0=0.000
q | N* lg = 2.05E 05
successive function values within tolerance limits

Coefficients:
              Estimate Std. Error t-value Pr(>|t|)
TC:(Intercept)    0.4942284    0.20678286   2.391  0.0188 *
TC:tpagenvp      -0.0020400    0.00612805  -0.330  <2e-16 ***
TC:tpagente     -0.04807059    0.00401301  -12.177  <2e-16 ***
TC:permis_VF_1_oui_0_ non. -1.02438428    0.10491801  -14.001  <2e-16 ***
TC:SEXE          0.78895880    0.07430328   10.603  <2e-16 ***
TC:motorisation_nbVF.nb15.  3.28802416    0.12926080  25.419  <2e-16 ***
TC:densite       0.00127267    0.00020775   6.134  <2e-16 ***
---
Signal. codes:  0 'none' 1 'small' 2 'mod' 3 'big' 4 'very big' 5 'huge' 6 'massive' 7 'insane' 8 'untenable' 9 'unstable'

Log Likelihood: -2883.6
McFadden R^2:  0.33079
Likelihood ratio test:  chiSq = 3560.3 (p-value = < 2.23e-16)
```

*Illustration 7: Indicateurs statistiques correspondant à la 6è régression.*

Notons que le coefficient affecté au paramètre sexe est positif, les femmes enquêtées prennent donc plus souvent les TC que les hommes.

## 7è régression : on rajoute la motorisation<sup>2</sup>

Nous avons voulu vérifier si l'effet de la motorisation peut être limitée à un simple effet linéaire. On sais en effet que dans la réalité la motorisation joue un rôle primordial, de plu, nous avons été interpellé par le fait que le coefficient attribué à la motorisation est toujours de loin le plus élevé dans nos régressions (sa valeur absolue est 2 fois plus élevée que celle du 2è coefficient le plus élevé). Nous avons donc voulu vérifier si un modélisation de son effet par une parabole est plus pertinente qu'un modélisation une droite.

Cette régression complète donc la régression précédente en ajoutant le paramètre pseudomot\_sq, il s'agit du paramètre traduisant la motorisation individuelle (nombre de VP / nombre de personnes de plus de 18 ans) élevé au carré. On effectue ainsi un développement limité de l'effet de la motorisation à l'ordre deux.

L'ajout de ce paramètre est pertinent, le log de likelihood passe de -2590 à -2577 soit une variation de plus de 10 donc non liée à un effet de volume, et le R<sup>2</sup> de MacFadden passe de 0.331 à 0.334. La probabilité du t de Student de cette variable est bonne, -2,2.e-16

```
> ml.choix10<- mlogit(Mode ~ 0 | spagevvp + spagevnc + pseudomot_sq + permie + SENE + motorisation_nbVP.nb18, + @constab, data = CM,
+ summary(ml.choix10))

Call:
mlogit(formula = Mode ~ 0 | spagevvp + spagevnc + pseudomot_sq +
      permie + SENE + motorisation_nbVP.nb18, + @constab, data = CM,
      reflevel = "VF", method = "hr", print.level = 0)

Frequencies of alternatives:
      VF      TC
0.04840 0.10050

no method
7 iterations, Ch:Gm:Gm
g*|-H|-1g = 1.18e-08
successive function values within tolerance limits

Coefficients:
              TC:(intercept)      0.03480180      0.20057079      3.0444      0.000020 **
              TC:spagevvp      0.08178800      0.08814e23      18.8088 < 2.2e-16 ***
              TC:spagevnc     -0.04873240      0.00402220     -12.1131 < 2.2e-16 ***
              TC:pseudomot_sq      0.52300340      0.04000020      13.0190 < 3.2e-16 ***
              TC:permie      -1.78864186      0.12e882e5     -19.2498 < 2.2e-16 ***
              TC:SENE      0.70000228      0.07333221      10.4296 < 2.2e-16 ***
              TC:motorisation_nbVP.nb18. -0.99040001      0.10476425     -34.8710 < 3.2e-16 ***
              TC:@constab      0.00828388      0.00018008      10.6770 < 2.2e-16 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood: -2577
McFadden R^2: 0.33405
```

Avec ce développement à l'ordre 2 de la motorisation, on a un coefficient d'ordre 2 positif et d'ordre 1 négatif. La motorisation a donc une influence sous forme de parabole. La variation dans les faibles valeurs de motorisation est plus importante que dans les valeurs élevées. Cela traduit une importance plus grande dans un ménage sur les déplacements de la première voiture du ménage par rapport au fait d'avoir une deuxième voiture.

La constante de cette régression a une valeur relativement faible et une t-value peu élevé, elle est cependant suffisante pour que l'on puisse rejeter l'hypothèse nulle et conserver la constante.

## **Conclusion :**

En conclusion, pour la modélisation du choix modal via le modèle Logit, nous avons fait appel à différents paramètres. Cette modélisation est une approximation qu'il convient de relativiser, notamment du fait de la spécificité de la population observée et d'éventuels biais dans le recueil des données. Néanmoins, différents résultats sont apparus au fil de la modélisation. L'influence des différents temps généralisés est importante. Ainsi les temps généralisés des deux modes ont une influence sur les comportements de déplacement et pas seulement le temps généralisé le plus faible. On peut ainsi penser que les usagers ont des sensibilités différentes aux variations de temps généralisés entre les deux modes. Le sexe et la densité sont importants aussi dans le choix modal. Enfin la motorisation a un rôle primordial sur le choix modal. Plus précisément, cette influence est très importante pour des valeurs de motorisation faible, la première voiture du ménage ayant un impact clef sur le choix modal.