

APDT rendu numéro 4

Edouard PHILIPP
Raphaël WALKER

Question 1 :

Un modèle logit comporte généralement 3 variables explicatives : le temps généralisé, le taux de motorisation et l'offre de stationnement. La différence d'utilité entre deux déplacements selon un mode est calculée de la manière suivante :

deltaV=

$$\left[k_m + \frac{ttc_{ij}}{\tau t c_m / mot_i} - \frac{tvp_{ij}}{\tau vp_m * mot_i} - \frac{d_j}{\delta_m} \right]$$

avec ttc_{ij} , le temps généralisé TC entre les zones i et j ;
 tvp_{ij} , le temps généralisé VP entre les zones i et j ;
 mot_i , la motorisation de la zone i ;
 d_j , la densité de la zone j, exprimée en population + emplois à l'ha ;
 k_m , $\tau t c_m$, τvp_m et δ_m , paramètres de répartition pour le motif m.

On a donc 3 termes d'utilités :

$$\frac{ttc_{ij}}{\tau t c_m / mot_i}$$

C'est-à-dire le temps généralisé en TC * la motorisation divisé par un paramètre pour le motif m.

Ce terme correspond à l'utilité des TC. Il est pondéré par le taux de motorisation qui influe fortement sur l'attractivité du mode TC via les populations captives.

$$\frac{tvp_{ij}}{\tau vp_m * mot_i}$$

C'est le temps généralisé en VP / la motorisation et un paramètre de répartition pour le motif m.

Il correspond à l'utilité des VP. Il est pondéré par le taux de motorisation qui augmente l'attractivité des VP par la facilité et la capacité d'usage.

$$\frac{d_j}{\delta_m}$$

C'est la densité de la zone j divisée par un paramètre correctif pour le motif m.

Ce terme prend en compte la probabilité de présence d'offres de stationnement (plus un espace est dense, moins il y aura de places à priori) et l'attractivité des TC. En effet, dans les zones peu denses, les TC auront des interstations plus élevées, le temps d'accès à une station sera plus élevé et les

fréquences et capacités des lignes seront plus faibles. Les TC sont donc moins attractifs pour l'utilisateur dans les zones peu denses.

La part des TC est calculée selon la formule suivante :

$$\%TC = \frac{1}{1 + \exp\left(k_m + \frac{ttc_{ij}}{\tau t c_m / mot_i} - \frac{tvp_{ij}}{\tau vp_m * mot_i} - \frac{d_j}{\delta_m}\right)}$$

On retrouve bien une formulation de type logit :

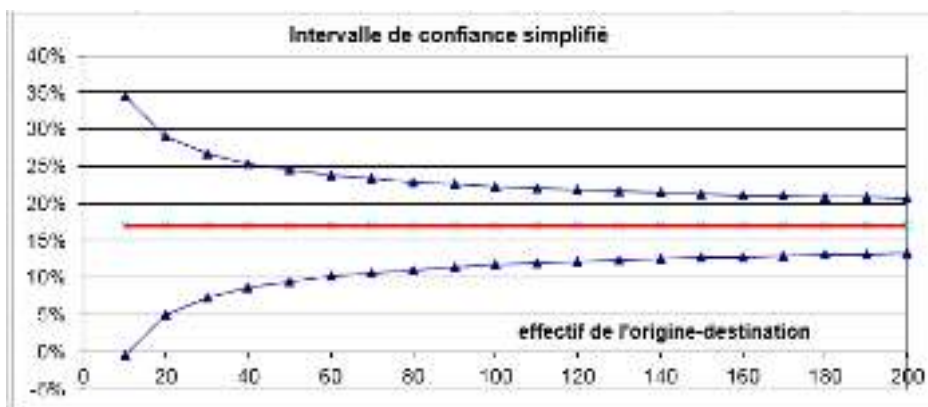
$$Pn(i) = 1 / (1 + \exp(-\mu * \Delta V)) \quad \text{où } \Delta V = k_m + \frac{ttc_{ij}}{\tau t c_m / mot_i} - \frac{tvp_{ij}}{\tau vp_m * mot_i} - \frac{d_j}{\delta_m}$$

On peut aussi voir sous la forme: $(V_{VP} - V_{TC}) = \log\left(\frac{1 - \%TC}{\%TC}\right)$

Question 2 :

Les effectifs d'enquêtés posent un problème de fiabilité statistique. En effet, beaucoup d'OD ne sont représentés que par un petit nombre de déplacements. Par exemple 5% des OD sont représentées par 5 déplacements ou moins, et 11% par 10 déplacements ou moins. On en arrive à un intervalle de confiance très large (entre 0% et 35%) si l'on conserve tout ces déplacements.

Ce problème rend pertinent l'usage d'un seuil déterminant un nombre minimum de déplacement pour sélectionner les OD significatives. En effet la figure suivante montre qu'on peut significativement resserrer l'intervalle de confiance en augmentant la valeur seuil. Cependant établir un seuil pose un problème : la disparition d'une partie des OD.



Nous voulons ici établir un seuil permettant d’avoir des données fiables et de garder un pourcentage acceptable du poids des OD dans le total des chaînes enquêtés. Pour cela nous avons étudié les valeurs données par plusieurs seuils.

seuil	20	30	40	50	60
taille de l'intervalle de confiance	24.13%	19.53%	16.84%	15.03%	13.69%
poids des O-D	75.30%	59.70%	49.50%	40.40%	34.50%

Tableau 1 Impact du seuil sur le modèle à 25 zones

seuil	20	30	40	50	60
taille de l'intervalle de confiance	24.13%	19.53%	16.84%	15.03%	13.69%
poids des OD	99.00%	98.00%	97.50%	97.03%	96.43%

Tableau 2 Impact du seuil sur le modèle à 7 zones

Pour obtenir une solution optimale, il faudrait pouvoir trouver une formule $y=f(\text{poids O-D, intervalle})$ nous donnant l’acceptabilité du compromis. Pour cela, il faudrait pondérer l’importance du poids des O-D par rapport à la taille de l’intervalle de confiance. Cette tâche relève du jugement de valeur, nous n’adopterons donc pas cette démarche. Cependant, on peut éliminer les valeurs qui ne nous semblent pas acceptables. C’est avec cette méthode que nous avons procédé.

NOTE : On se basera ici sur les résultats du modèle à 25 zones plutôt que sur celui à 7 zones car ce dernier est beaucoup moins sensible aux variations du seuil.

Nous considérons qu’un intervalle de confiance d’une taille supérieure à 20% n’est pas acceptable. Le seuil choisit doit donc être supérieur à 20.

Nous considérons qu’au moins 40% des O-D doivent être représentés si l’on veut que le résultat soit fiable. Le seuil choisit doit donc être inférieur à 60.

Les valeurs acceptables proposées par le tableau ci-dessus sont donc 30 40 et 50. Nous avons choisi de prendre la valeur du milieu, c’est dire 40. Nous avons donc fixé le seuil à 40 déplacements. Nos résultats prendront donc en compte environ 50% des O-D et l’intervalle de confiance aura une taille d’environ 17%.

Question 3

Les différents résultats de la régression sont présentés dans le tableau suivant :

Seuil	30	Proba t	40	Proba t	50	Proba t	60	Proba t
Coeff R ²	0,750742		0,758406		0,651301		0,68286	
Constante	2,83881464	1,4218E-16	2,49383419	8,4823E-10	2,57768378	1,65E-05	3,353968	0,00011
Variable X1 TgTc*mot	0,05093544	6,7857E-12	0,05944364	3,1071E-09	0,05427238	0,00043449	0,028360	0,187350
Variable X2 TgVp*mot	-0,04113	1,1205E-07	-0,03719	0,00012869	-0,03325	0,03217083	-0,0364	0,038166
Variable X3 densité	-0,00441	7,6873E-11	-0,00419	8,5318E-06	-0,00480	0,00565282	-0,00578	0,00401

Les coefficients de corrélation R² sont compris entre 0.65 et 0.76. Le pourcentage de variance expliquée par la régression se situe donc entre 65 et 76% selon les différents seuils.

Les valeurs des probabilités du t de Student sont acceptables (<5%) sauf pour la variable X1 avec un seuil à 60, la régression devant ainsi être rejeté dans ce cas.

Les signes obtenus pour les variables sont cohérents. La variable X1 a un signe positif, en effet lorsque le temps généralisé en TC augmente, cela entraîne une diminution du % de TC dans le modèle logit. De même, le fait que les coefficients soient négatifs pour les deux autres variables est cohérent puisque ces dernières augmentent le % de Tc lorsqu'elles croissent, traduisant une diminution de l'attractivité de la VP.

Du fait des coefficients R² à 75% et 76%, les régressions pour les seuils à 30 et 40 sont les plus satisfaisantes pour ce modèle tandis que pour un seuil supérieur le R² diminue fortement voire la régression doit être rejetée.

FAUX, on peut pas comparer les R² si on a pas le mm jeu de donnée

Il manque l'analyse de stabilité des coefs sur les seuils 30,40,50

Question 4

Pour cette question, les seuils choisis sont plus élevés que pour la précédente puisque la régression sur 7 zones suite à une agrégation, soit de 50 à 80.

Les différents résultats de la régression sont présentés dans le tableau suivant :

Seuil	50	Proba t	60	Proba t	70	Proba t	80	Proba t
Coeff R ²	0,83770787		0,82752955		0,82078116		0,84374453	
Constante	2,24650526	5,5681E-08	2,21734143	2,3989E-07	2,1058173	3,941E-06	2,06514757	6,3425E-06
Variable X1 TgTc*mont	0,05366645	1,3826E-09	0,05432276	6,0142E-09	0,05484812	2,7098E-08	0,05416653	4,87E-08
Variable X2 TgVp*mont	-0,03230	4,0506E-05	-0,03187	0,00011814	-0,02833	0,00151161	-0,02718	0,00675566
Variable X3 densité	-0,00508	1,5164E-07	-0,00513	4,9357E-07	-0,0054	1,7581E-06	-0,00539	6,2591E-06

Les probabilités du t de Student sont acceptables pour les différentes régressions qui ne sont donc pas rejetées directement. De plus, le signe des coefficients est cohérents avec le même logit en suivant les mêmes arguments que pour la question précédente (signe positif pour X1 puisque son augmentation pénalise le % de TC, et signes négatifs pour les coefficients de X2 et X3).

Les coefficients R² sont compris entre 82 et 83% pour la modélisation sur 7 zones. Ainsi un plus grand pourcentage de variance expliqué par les différentes régressions est plus grand dans ce cas que dans le cas à 25 zones (de 75 à 76% de variance expliqué). Cela pourrait amener à préférer un modèle à 7 zones. Néanmoins, cela se ferait au détriment du poids des OD avec des seuils plus élevés dans le cas d'un modèle avec 7 zones.

IL MANQUE

Quelle limite liée à l'agglomération en 7 zones ?

Quelle stabilité des paramètres entre les seuils ?

Question 5 :

En entrant les valeurs trouvées par les régressions précédentes comme valeurs d'initialisation, on prend le seuil à 50 déplacements pour le modèle à 7 zones car il semble être le plus probant au regard des questions précédentes. On lance la macro.

Celle-ci nous donne le résultat suivant :

115	Numéro de l'itération en cours		500
-0,004908278	-0,063697374	0,038592223	4,439529121
0,001285333	0,009979313	0,014834302	0,549447199
0,656510442	0,995042892	#N/A	#N/A
48,41951524	76	#N/A	#N/A
143,8219907	75,24838719	#N/A	#N/A

Ces résultats nous semblent faux et inexploitable. Nous avons lancé la macro en changeant les paramètres de départ plusieurs fois et obtenons toujours le même résultat. Nous ne voyons pas d'où viens l'erreur.

NOTRE Q5 est pourrie faites comme si vous avez rien vu

Conclusion

La méthode à 25 zones donne de moins bon résultats statistiques (R^2 , T de Student) que la méthode à 7 zones. Cependant, diminuer le nombre de zone réduit fortement la pertinence du modèle. On peut même supposer que cette simplification du modèle améliore artificiellement les indicateurs statistiques.