

TD4 – Chapitre Répartition modale

Dans ce TD, nous travaillons sur le calage de la répartition modale, tout en essayant d'assurer une bonne précision des données. Le motif sur lequel nous travaillons est « autres motifs ».

Question 1 – Interprétation du modèle de choix modal

- Suite à ce que nous avons vu en cours, voici les 2 formes du modèle Logit et Probit :

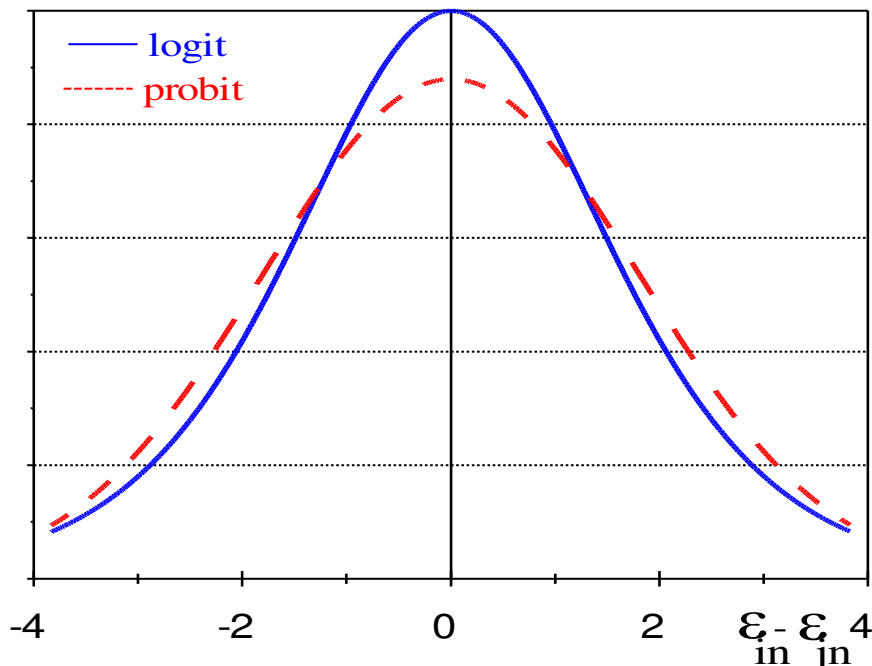


Figure 1 - Fonctions de densité logistique Probit et Logit

Les caractéristiques du modèle Probit sont principalement : les choix des individus impliqués dans le modèle sont indépendants (pris comme hypothèse), moins de restriction quant aux hypothèses sur les résidus. Sa formulation mathématique est annoncée comme « complexe » cependant. C'est pourquoi le modèle Logit est aussi utilisé fréquemment.

Les hypothèses sont également nombreuses : l'hypothèse IID (*Independently and Identically Distributed*) permettant d'avoir une covariance nulle et une variance identique pour toutes les alternatives. La mise en place mathématique est donc plus aisée, et c'est pour cela qu'il sera retenu.

- Termes de la différence d'utilité

Le modèle retenu est celui développé par le SEMALY et le LET sur l'agglomération lyonnaise et est le

$$\text{suivant : } \%TC = \frac{1}{1 + \exp\left(k_m + \frac{ttc_j}{\tau c_m / \text{mot}_i} - \frac{tv p_j}{\tau v p_m * \text{mot}_i} - \frac{d_j}{\delta_m}\right)}$$

avec tt_{ij} , le temps généralisé TC entre les zones i et j ;
 tv_{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'hectare ;
 k_m , τ_{cm} , τ_{vm} et δ_m , les paramètres de répartition modale pour le motif m .

Les trois termes explicatifs sont les suivants :

$$\%TC = \frac{1}{1 + \exp\left(k_m + \underbrace{\frac{tt_{ij}}{\tau_{cm}/mot_i}} - \underbrace{\frac{tv_{ij}}{\tau_{vm} * mot_i}} - \underbrace{\frac{d_j}{\delta_m}}\right)}$$

Le terme $\frac{tt_{ij}}{\tau_{cm}/mot_i}$ augmente suivant le temps généralisé TC et la motorisation. Si il augmente, alors l'exponentielle également, et le %TC diminue. Ceci paraît normal, on est moins enclin à utiliser le TC si son temps généralisé est important.

Quand le terme $\frac{tv_{ij}}{\tau_{vm} * mot_i}$ augmente, le terme en exponentielle diminue (à cause du signe négatif). Donc quand le temps généralisé VP augmente, la part TC augmente, ce qui est compréhensible.

Si le terme de densité $\frac{d_j}{\delta_m}$ augmente, alors l'exponentielle diminue en raison du signe négatif devant ce terme. Le %TC augmente alors, ce qui apparaît normal.

Question 2 – Analyse de la précision des données

Nous cherchons ici le seuil minimum des effectifs de l'enquête sur les 25 zones. Il faut trouver un compromis entre un grand nombre d'OD et un intervalle de confiance satisfaisant.

- 25 zones au seuil 30

Nombre d'O-D dont l'effectif est supérieur au seuil		96
Nombre de chaînes incluses dans ces O-D		11,121
ponds de ces O-D dans le total des chaînes enquêtées		77.1%

03 février 2014

Les résultats donnés sont donc de 96 OD, soit 77 % environ des chaînes enquêtées, et l'intervalle de confiance de 17,5%. Le R^2 est également de 55%.

- 25 zones au seuil 40

Nombre d'O-D dont l'effectif est supérieur au seuil		67
	Nombre de chaînes incluses dans ces O-D	10,109
	pois de ces O-D dans le total des chaînes enquêtées	70.1%

Les résultats sont de 67 OD, soit 70 % des chaînes enquêtées. L'intervalle de confiance est de 15,1 %, et le R^2 est de 54%.

- 25 zones au seuil 50

Nombre d'O-D dont l'effectif est supérieur au seuil		55
	Nombre de chaînes incluses dans ces O-D	9,572
	pois de ces O-D dans le total des chaînes enquêtées	66.3%

Il y'a ici 55 observations, soit 2/3 environ des OD totales. Le R^2 est de 57% et l'intervalle de confiance est égal à 13 ,5%.

Le seuil à 40 observations apparait comme un bon compromis entre les critères de l'intervalle de confiance et du nombre d'OD suffisamment grand.

Question 3 – Calage du modèle de répartition modale sur le découpage en 25 zones en sélectionnant les OD dont l'effectif est supérieur à un seuil donné

Suite à la question 2, nous utilisons les seuils de 30, 40 puis 50 Origine-Destinations.

- Seuil 30 OD

RAPPORT DÉTAILLÉ					
Statistiques de régression					
Coefficient de détermination multiple	0.747043455				
Coefficient de détermination R^2	0.552628194				
Coefficient de détermination R^2	0.535975075				
Ereur-type	1.201691853				
Observations	96				
ANALYSE DE VARANCE					
	Degré de liberté	Somme des carrés	Moyenne des carrés	f	Valeur critique de f
Régression	3	164.5061437	54.83504791	37.57675395	6.04444E-16
Résidus	92	134.5302723	1.461372525		
Total	95	299.0364161			
	Coefficients	Ereur-type	Statistique t	Probabilité	Limite inférieure pour des% de confiance = 99%
Constante	4.277192041	0.633555292	6.75103446	1.25121E-05	3.018894557
Variable X 1	0.041994942	0.014754772	2.845193559	0.005455093	0.012690592
Variable X 2	-0.052945027	0.012021648	-4.33491280	1.04671E-05	-0.085835916
Variable X 3	-0.004135305	0.02094788	-1.883501463	5.7048E-05	-0.005321572

03 février 2014

 $R^2 = 55\%$

Valeur critique du F de Fisher = 6,04E-16. Cette valeur est très faible donc on peut s'assurer que les coefficients de la régression ne sont pas nuls.

Les 3 probabilités du t de Student sont inférieures à 5%, donc les trois variables de temps généralisé TC*motorisation, temps généralisé VP/motorisation et densité sont importantes. Les valeurs respectives sont 0,0419 ; -0,062 ; - 0,0044.

- Seuil 40 OD

RAPPORT DÉTAILLÉ					
Statistiques de la régression					
Coefficient de détermination multiple		0.739054794			
Coefficient de détermination R ²		0.546201989			
Coefficient de détermination R ²		0.52459256			
Ereur-type		1.270680225			
Observations		67			
ANALYSE DE VARIANCE					
	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	3	122.4344914	40.81149714	25.27609529	7.44775E-11
Résidus	63	101.7215788	1.614628235		
Total	66	224.1560702			
	Coefficients	Ereur-type	Statistique t	Probabilité	Limite inférieure pour seuil de confiance = 95%
Constante	3.826993643	0.873484205	4.381296906	4.55097E-05	2.081474743
Variable X 1	0.063701237	0.025654687	2.483025306	0.015702029	0.012434436
Variable X 2	-0.058181488	0.014705595	-3.956418564	0.000195679	-0.087568274
Variable X 3	-0.004670253	0.001363151	-3.426070943	0.001083061	-0.007394293

 $R^2 = 0,546$ (plutôt faible).

F de Fisher inférieur à 5% et les t de Student sont inférieurs à 5%.

Les valeurs pour les variables sont :

- Temps généralisé TC*motorisation = 0,0637
- Temps généralisé VP/motorisation = - 0,058
- Densité de la zone = - 0,0046

- Seuil 50 OD

RAPPORT DÉTAILLÉ					
Statistiques de la régression					
Coefficient de détermination multiple		0.703401264			
Coefficient de détermination R ²		0.503210067			
Coefficient de détermination R ²		0.563398194			
Ereur-type		1.331122563			
Observations		55			
ANALYSE DE VARIANCE					
	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	3	123.8784334	41.29281114	23.30442329	1.22712E-09
Résidus	51	90.36525115	1.771837277		
Total	54	214.2446846			
	Coefficients	Ereur-type	Statistique t	Probabilité	Limite inférieure pour seuil de confiance = 95%
Constante	4.181888666	1.080962703	3.870830724	0.000305615	2.01356775
Variable X 1	0.054365390	0.020580190	2.638531132	0.005101807	-0.010070149
Variable X 2	-0.063113633	0.018213857	-3.337077235	0.001588002	-0.09627756
Variable X 3	-0.005730188	0.001799515	-3.189158822	0.002405153	-0.009352088

03 février 2014

$$R^2 = 0.578$$

Le F de Fisher est toujours inférieur à 5%, ainsi que les t de Student.

Voici les valeurs pour les variables :

- Temps généralisé TC*/motorisation = 0,05536
- Temps généralisé VP/motorisation = - 0,06011
- Densité de la zone = - 0,0057

Finalement, les valeurs des variables sont bien différentes suivant les 3 seuils d'observations.

Ensuite, les zones seront agrégées suivant seulement 7 zones, avec un poids des OD dans le total des chaînes enquêtées plus important.

Question 4 – Calage du modèle de répartition modale sur le découpage en 7 zones en sélectionnant les OD dont l'effectif est supérieur à un seuil donné

Désormais le modèle est calé selon un découpage en 7 zones. Comme auparavant, les seuils choisis permettent un intervalle de confiance satisfaisant et un nombre représentatif d'OD.

Les seuils choisis sont donc 40, 50 et 60 O-D.

- Seuil 40 OD

Nombre d'O-D dont l'effectif est supérieur au seuil		37
Nombre de chaînes incluses dans ces O-D		14,212
poids de ces O-D dans le total des chaînes enquêtées		98.51%

RAPPORT DETAILLE					
Statistiques de la régression:					
Coefficient de détermination multiple	0.612915309				
Coefficient de détermination R ²	0.6693513				
Coefficient de détermination R ²	0.62887102				
Erreur-type	0.058814821				
Observations	37				
ANALYSE DE VARIANCE					
	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	3	21.038643	7.01288133	21.03725088	6.90201488
Résidus	33	15.98036688	0.484253845		
Total	36	37.01901028			
	Coefficients	Erreur-type	Statistique t	Probabilité	Limite inférieure pour seuil de confiance = 95%
Constante	3.340775001	0.11257041	6.136714681	2.0146E-07	2.00723583
Variable X 1	0.019881471	0.01088884	3.476628927	0.00148871	0.018375316
Variable X 2	-0.038175944	0.008574512	-4.452375063	9.1684E-05	-0.055621591
Variable X 3	-0.025467846	0.00148035	-1.762735055	3.73473E-05	-0.027803542

$R^2 = 0,66$ ce qui montre une amélioration assez significative par rapport aux 25 zones.

Le F et les t de Student sont inférieurs à 5% également.

03 février 2014

Les valeurs des 3 variables sont les suivantes :

- Temps généralisé TC*motorisation = 0,037
- Temps généralisé VP/motorisation = - 0,0381
- Densité de la zone = - 0,0054

- Seuil 50 OD

Nombre d'O-D dont l'effectif est supérieur au seuil		34
Nombre de chaînes incluses dans ces O-D		14,076
pois de ces O-D dans le total des chaînes enquêtées		97.57%

RAPPORT DÉTAILLÉ					
Statistiques de la régression					
Coefficient de détermination multiple			0.81440815		
Coefficient de détermination R^2			0.663280635		
Coefficient de détermination RV2			0.629586099		
Erreur-type			0.72811113		
Observations			34		
ANALYSE DE VARIANCE					
	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	3	30.9039487	10.3013159	19.6886777	3.20691127
Résidu	30	15.71012873	0.523670958		
Total	33	46.66557454			
	Coefficients	Erreur-type	Statistique t	Probabilité	Limite inférieure pour seuil de confiance = 95%
Constante	3.471033812	0.572075485	6.057207589	1.15504E-06	2.307601478
Variable X 1	0.034248788	0.011934749	2.855315196	0.007728653	0.005752242
Variable X 2	-0.038555355	0.009162107	-4.205223232	0.000214235	-0.057276874
Variable X 3	-0.00541608	0.001320223	-4.132349142	0.000116040	-0.008147840

Cette fois, le R^2 est égal à 0,066, ce qui montre une amélioration. De même, les valeurs du F de Fisher et les t de Student sont inférieures à 5%.

Voici la valeur des variables :

- Temps généralisé TC*motorisation = 0,034
- Temps généralisé VP/motorisation = - 0,03856
- Densité de la zone = - 0,00585

- Seuil 60 OD

Nombre d'O-D dont l'effectif est supérieur au seuil		33
Nombre de chaînes incluses dans ces O-D		14,024
pois de ces O-D dans le total des chaînes enquêtées		97.21%

RAPPORT DÉTAILLÉ					
Statistiques de la régression					
Coefficient de	0.813979719				
Coefficient de	0.562480773				
Coefficient de	0.627564991				
Erreur-type	0.733958705				
Observations	33				
ANALYSE DE VARIANCE					
	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	3	30.66309662	10.22103221	18.97367707	6.27242E-07
Résidus	29	16.62216606	0.57316090		
Total	32	46.28526268			
	Coefficients	Erreur-type	Statistique t	Probabilité	Limite inférieure pour seuil de confiance = 95%
Constante	3.549002886	0.611539463	5.803538607	2.73402E-06	2.298354248
Variable X 1	0.033952923	0.012187613	2.785854929	0.009312418	0.009026455
Variable X 2	-0.040668178	0.01065048	-3.818436208	0.000653483	0.062450865
Variable X 3	-0.00576423	0.001366407	-4.249630294	0.000202354	0.000530394

Le R^2 est de 66% cette fois. Les t de Student et le F de Fisher sont inférieurs à 5%.

Voici les valeurs des coefficients des variables :

- Temps généralisé TC*motorisation = 0,03339
- Temps généralisé VP/motorisation = - 0,04066
- Densité de la zone = - 0,0057

Pour cette 2^e question avec agrégat à 7 zones, les proportions des OD dans le total des chaînes enquêtées sont très nettement meilleures que dans la question 3.

Question 5 – Calage du modèle de répartition modale sur le découpage en 25 zones en conservant toute l'information

Il s'agit dans cette cinquième question de choisir un seuil d'affectif minimum, il servira à construire les classes de groupes O-D. On choisit d'examiner deux seuils 120 et 100.

Il faut dans un premier temps mettre en place l'initialisation de la macro « CLAC » que nous allons utiliser pour construire les classes de groupes O-D, pour cela on choisit la « meilleure » des régressions linéaires des questions précédente. Il s'agit de la régression linéaire avec 7 zones et un seuil de 50.

	<i>Coefficients</i>
Constante	3.471033612
Variable X 1	0.034248788
Variable X 2	-0.038565355
Variable X 3	-0.005851689

Figure 2 - Valeurs d'initialisation (régression linéaire avec 7 zones et un seuil de 50)

Notons qu'on fait une moyenne sur les 30 dernières valeurs pour ne pas que les valeurs des premières itérations qui fluctuent beaucoup influent sur le résultat.

Avec l'effectif minimal des classes = 120.

itération	coeff densité	coeff VP	coeff TC	constante	R2	Nombre de classes de la régression
500 ^{ème} itération	-0.0046	-0.0638	0.0472	4.1236	70.02%	71
moyenne sur les 30 dernières itérations	-0.0049	-0.0769	0.0503	4.7211	55.42%	70

Avec l'effectif minimal des classes = 100.

itération	coeff densité	coeff VP	coeff TC	constante	R2	Nombre de classes de la régression
500 ^{ème} itération	-0.0042	-0.0585	0.0478	3.8509	75.07%	81
moyenne sur les 30 dernières itérations	-0.0047	-0.0728	0.0537	4.4825	50.72%	79

C'est le calage avec un effectif minimal de 100 qui est le plus concluant (un R^2 de 50.72% sur les 30 dernières itérations).

Conclusion :

Au regard des résultats des trois méthodes, il nous apparaît tout d'abord que la méthode à 25 zones donne de plus mauvais résultats. Ensuite, il apparaît que la méthode à 7 zones donne de meilleurs résultats, mais on est en droit de se demander la pertinence de la diminution du nombre de zones. Enfin, tentant de conserver toutes les informations on arrive à des résultats satisfaisants.

Il nous paraît alors bien difficile de dire laquelle des méthodes est la meilleure... Tout au plus on peut dire que la moins bonne est semble-t-il la méthode à 25 zones.