

# Détermination de la loi de génération sur les données de 1995

---

Le motif de déplacement que nous allons traiter est le motif « travail » pour les émissions et les attractions. Notre but est de déterminer la meilleure loi de génération, en utilisant des régressions linéaires successives, et cela en testant les différentes variables proposées dans les fichiers fournis.

Il nous faudra avant tout minimiser le problème de l'hétéroscédasticité. Aussi, pour les variables qui concernent une catégorie de la population, on peut diminuer l'hétéroscédasticité en les ramenant à des taux, en les divisant par la population totale. Par exemple on considère le taux de scolaires dans le primaire dans une population donnée plutôt que la donnée brute des scolaires dans le primaire.

Malheureusement, le problème de l'hétéroscédasticité restera présent pour les variables qu'on ne peut transformer sous forme de taux. Par exemple, le nombre d'emplois dans le tertiaire ou encore les places d'étudiants,...

Afin d'évaluer la qualité des différentes régressions linéaires que nous allons réaliser nous allons suivre des résultats qui sortent de ces mêmes régressions linéaires :

- Le  $R^2$  qui correspond au taux de points expliqués par la droite de régression.  
Plus  $R^2$  est grand, plus la régression linéaire est "proche" du nuage de point et plus la variable prise en compte est explicative du motif. Cependant, si on prend en compte trop de variables en même temps, le  $R^2$  sera grand sans que toutes les variables soient nécessairement explicatives, il faut ainsi prendre en compte cet état de fait.
- Le  $R^2$  corrigé permettra de comparer deux régressions linéaires qui ne comportent pas le même nombre de variables.
- Le F de Fisher qui correspond au rapport entre les carrés expliqués et les carrés non expliqués.  
Si le F est inférieur à 5%, on peut considérer que la régression est significative pour au moins une variable à 95%.
- Le t de Student permet de savoir si une variable doit être conservée ou non. Ainsi si le t de Student est supérieur à 5%, la variable n'est pas utile au modèle.

On obtient les valeurs de ces différents coefficients en sortie de la régression linéaire, selon leurs signes et leurs valeurs on détermine si notre choix des variables de la régression linéaire est pertinent pour le motif. Si ce n'est pas le cas, on écartera ce choix de variables.

# Emissions

## 1ère régression

Pour cette première régression, on a décidé de sélectionner toutes les variables qui semblent pouvoir être explicatives pour les émissions pour le motif "travail". Notre choix s'est donc fait sur les variables suivantes :

- Le « taux des actifs ayant un emploi »
- Une nouvelle variable qu'on introduit, le « taux d'emplois hors de la zone » qui correspond au pourcentage d'emploi se trouvant hors de la zone (1- « nombre d'emploi de la zone »/ « nombre d'emploi total »)
- La « motorisation ».

	A	B	C	D	E	F	G	H	I	J	K
1	RAPPORT DETAILLE										
2											
3	Statistiques de la régression										
4	Coefficient de	0,559812051									
5	Coefficient de	0,313423122									
6	Coefficient de	0,10552406									
7	Erreur-type	19983,81848									
8	Observations	25									
9											
10	ANALYSE DE VARIANCE										
11		Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F					
12	Regression	3	4010712129	1336904043	5,021512444	0,00847005					
13	Résidus	22	8785760024	399353001,1							
14	Total	25	12796478153								
15											
16		Coefficients	Erreur-type	Statistique t	Probabilité	on pour seul de confiance = 05,0%					
17	Constante	156714,7346	60241,21018	2,601453957	0,016292974	31782,11128	281847,558	31782,11122	281647,358		
18	Variable X 1	-111460,1725	143257,8209	-0,778030005	0,444833917	-408558,7091	182638,3642	-408558,709	182638,3642		
19	Variable X 2	0	0	65555	#NOMBRE!	0	0	0	0		
20	Variable X 3	-110834,1891	38027,95301	-2,871131762	#NOMBRE!	-190336,2339	-30731,1423	-190336,239	-30732,1425		

Cette régression linéaire semble prometteuse puisque le  $R^2$  est raisonnablement (0,31), la valeur critique du F de Fisher est lui très bon (0.00885). Néanmoins, les valeurs des coefficients que sort la régression linaire du tableur sont incohérentes.

Aussi nous avons décidé donc forcer la constante à 0 ce qui paraît cohérent, au moins pour les variables "taux des actifs ayant un emploi" et la « taux d'emplois hors de la zone ».

## Régression 1 forcée à l'origine

	A	B	C	D	E	F	G	H	I	J	K
1	RAPPORT DÉTAILLÉ										
2											
3	Statistiques de la régression										
4	Coefficient de dé	0,878223158									
5	Coefficient de dé	0,757288348									
6	Coefficient de dé	0,688769105									
7	Erreur-type	22726,39214									
8	Observations	25									
9											
10	ANALYSE DE VARIANCE										
11		Degré de liberté	Somme des carrés	Moyenne des carrés	F	sur critique de F					
12	Régression	3	3545227518	11817609173	22,88084561	8,06537E-07					
13	Résidus	22	13362665797	516484809							
14	Total	25	48815493386								
15											
16		Coefficients	Erreur-type	Statistique t	Probabilité	sur seuil de	pour seuil de	sur seuil de	sur seuil de	confiance = 95,0%	
17	Constante	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
18	Variable X 1	156705,4285	156709,388	0,999400836	0,32846684	-168478,568	481887,400	-168478,568	481887,400		
19	Variable X 2	37279,29454	75552,38351	0,493425143	0,626603548	-119406,759	193965,3479	-119406,759	193965,3479		
20	Variable X 3	-96812,1148	47368,88653	-2,043799854	0,053127783	-185848,758	1434,528662	-185848,758	1434,528662		

Cette régression linéaire forcée à zéro semble également prometteuse puisque le  $R^2$  est bon (0,75), la valeur critique du F de Fisher est lui très bonne (8.0 E-07).

On décide de retirer la variable « motorisation » de la régression linéaire suivante.

## Régression 2

Pour la régression suivante on travaille donc avec les deux variables suivantes :

- Le « taux des actifs ayant un emploi »
- Une nouvelle variable qu'on introduit, le « taux d'emplois hors de la zone » qui correspond au pourcentage d'emploi se trouvant hors de la zone (1- « nombre d'emploi de la zone » / « nombre d'emploi total »)

	A	B	C	D	E	F	G	H	I	J	K
1	RAPPORT DÉTAILLÉ										
2											
3	Statistiques de la régression										
4	Coefficient d	0,885171702									
5	Coefficient d	0,7601689									
6	Coefficient d	0,760181254									
7	Erreur-type	11307,80879									
8	Observation:	25									
9											
10	ANALYSE DE VARIANCE										
11		Degré de liberté	Somme des carrés	Moyenne des carrés	F	sur critique de F					
12	Régression	2	9883414281	4991707141	99,0384158	5,79403E-08					
13	Résidus	22	2813083972	127867453,3							
14	Total	24	12796478153								
15											
16		Coefficients	Erreur-type	Statistique t	Probabilité	sur seuil de	pour seuil de	sur seuil de	sur seuil de	confiance = 95,0%	
17	Constante	646854,1398	70688,78457	9,150732237	5,91578E-09	500254,5733	793453,7063	500254,5733	793453,7063		
18	Variable X 1	-49902,8529	81273,31773	-0,61401123	0,543303506	-218433,813	118648,1066	-218433,813	118648,1066		
19	Variable X 2	-613830,919	72108,98369	-8,51254103	3,07387E-08	-763375,803	464286,036	-763375,803	464286,036		

Cette régression linéaire semble bonne puisque le  $R^2$  est bon (0,78), la valeur critique du F de Fisher est lui très bonne (5.79 E-07). On décide de la forcée à l'origine.



## Régression 3 à l'origine

	A	B	C	D	E	F	G	H	I	J	K
1	RAPPORT DETAILLÉ										
2											
3	Statistiques de la régression										
4	Coefficient d	0,968959219									
5	Coefficient d	0,938851968									
6	Coefficient d	0,892746402									
7	Erreur-type	11150,61281									
8	Observations	25									
9											
10	ANALYSE DE VARIANCE										
11	Degré de liberté des carrés des cov F sur critique de F										
12	Régression	2	41954222506	21977111253	176,6605094	2,80684E-14					
13	Résidus	23	2881270810	124469078,7							
14	Total	25	46815493216								
15											
16	Coefficients Erreur-type Statistique t Probabilité pour une queue pour une queue pour une queue pour une queue pour une queue pour une queue pour une queue = 95,0%										
17	Constante	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
18	Variable X 1	1,037489686	0,114680333	9,482791473	2,06218E-09	0,850255342	1,32472403	0,850255342	1,32472403		
19	Variable X 2	11680,08603	3569,983228	3,355782161	0,002736176	4595,013058	19365,15001	4595,013058	19365,15001		
20											

La régression est bonne : elle nous donne un  $R^2$  bon (0,938), une valeur critique du F de Fisher est lui très bonne (2.8 E-14).

## Conclusion

Au vu des résultats des différentes régressions linéaires réalisées on décide de retenir les variables comme étant des variables explicatives des déplacements "émissions" pour le motif "travail" "taux d'actif ayant un emploi". Ces deux variables sont :

- Le « nombre d'emplois »
- La nouvelle variable introduite, le « taux d'emplois hors de la zone »

## Attraction

### 1<sup>ère</sup> Régression

Nous étudierons 3 variables.

La première variable naturelle à prendre en compte est le nombre d'emplois de la zone considérée (Variable x1). En effet, il semble cohérent qu'au plus une zone possède d'emplois au plus elle est attractive en terme de travail. La seconde variable est le nombre de scolaires (toute catégories confondues) ; de façon à voir si l'éducation est une source significative d'emploi ou non (Variable x2). Enfin, nous considérerons la variable population pour voir si cette variable influe sur l'attraction (Variable x3).

Remarque : nous passons directement les variables 2 et 3 en taux afin de supprimer les problèmes d'hétéroscédasticité.

Statistiques de la régression										
4	Coefficient de détermination multip	0,978817424								
5	Coefficient de détermination R^2	0,958083549								
6	Coefficient de détermination R^2	0,952095484								
7	Erreur-type	6766,494145								
8	Observations	25								
9										
ANALYSE DE VARIANCE										
	Degré de liberté	Somme des carrés	MS	F	Seuil critique de F					
12	Régression	3	21976857530	7325619177	159,9988707	1,27247E-14				
13	Résidus	21	961494303,2	45785443,01						
14	Total	24	22938351833							
15										
	Coefficients	Erreur-type	Statistique t	Probabilité	pour seuil de	pour seuil de	pour seuil de	pour seuil de	pour seuil de	
17	Constante	636,7069777	2710,862477	0,234872474	0,816582749	-5000,84016	6274,254116	-5000,84016	6274,254116	
18	Variable X 1	1,670872255	0,16313657	10,2421686	1,37215E-09	1,331611186	2,010133325	1,331611186	2,010133325	
19	Variable X 2	31638,18964	319050,695	0,111700712	0,912121498	-627864,053	699140,4322	-627864,053	699140,4322	
20	Variable X 3	-104512,59	292989,4471	-0,35671111	0,724867318	-713817,501	504792,3201	-713817,501	504792,3201	

Le R<sup>2</sup> est tout à fait convenable. Mais la simple vue des coefficients montre que le résultat n'est pas bon. Nous allons donc forcer le passage à l'origine :

## 2<sup>ème</sup> Régression

Statistiques de la régression										
4	Coefficient d	0,99150124								
5	Coefficient d	0,983074708								
6	Coefficient d	0,9360815								
7	Erreur-type	6619,599208								
8	Observations	25								
9										
ANALYSE DE VARIANCE										
	Degré de liberté	Somme des carrés	MS	F	Seuil critique de F					
12	Régression	3	55993346935	18664448978	425,9432912	5,83256E-19				
13	Résidus	22	964020060,8	43819093,67						
14	Total	25	56957366996							
15										
	Coefficients	Erreur-type	Statistique t	Probabilité	pour seuil de	pour seuil de	pour seuil de	pour seuil de	pour seuil de	
17	Constante	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
18	Variable X 1	1,661569868	0,154820192	10,73225561	3,28671E-10	1,340492443	1,982647294	1,340492443	1,982647294	
19	Variable X 2	60650,44791	294224,285	0,206136784	0,838579779	-549533,373	670834,2686	-549533,373	670834,2686	
20	Variable X 3	-112144,942	284860,4918	-0,39368373	0,697604169	-702909,444	478619,5601	-702909,444	478619,5601	

Ici aussi nous avons un R<sup>2</sup> excellent. Mais les deux dernières variables donnent encore des résultats incohérents. La seule variable légitime (entre les trois étudiées) ne peut qu'être le nombre d'emplois.

### 3<sup>ème</sup> Régression (Retenue)

Statistiques de la régression								
Coefficient de détermination multiple	0,991346715							
Coefficient de détermination R <sup>2</sup>	0,982768349							
Coefficient de détermination R <sup>2</sup>	0,941101882							
Erreur-type	6194,885518							
Observations	25							
ANALYSE DE VARIANCE								
	Degré de liberté des carrés des cov	F	sur critique de F					
Régression	1 55975897525 55975897525	1368,785867	5,35145E-22					
Résidus	24 961469471,1 40894561,3							
Total	25 56957366096							
Coefficients		Erreur-type	Statistique t	Probabilité	pour seul de	pour seul de	pour seul de	pour seul de
Constante	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
Variable X 1	1,582489382	0,042773321	36,99710619	1,11357E-12	1,491209287	1,670768877	1,491209287	1,670768877

Les résultats de cette régression, via les différentes valeurs observées, sont tout à fait convenables : c'est bien le nombre d'emplois d'une zone qui rend cette dernière attractive !

### Comparaison avec l'EMD 1985

Nous reprenons ici, la précédente régression avec les données issues de l'enquête ménage déplacement de 1985 :

Statistiques de la régression								
Coefficient d	0,989234088							
Coefficient d	0,978584081							
Coefficient d	0,936917414							
Erreur-type	7567,953645							
Observations	25							
ANALYSE DE VARIANCE								
	Degré de liberté des carrés des cov	F	sur critique de F					
Régression	1 62810116806 62810116806	1096,661695	6,54443E-21					
Résidus	24 1374574137 57273922,37							
Total	25 64184690943							
Coefficients		Erreur-type	Statistique t	Probabilité	pour seul de	pour seul de	pour seul de	pour seul de
Constante	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
Variable X 1	1,685702069	0,050903129	33,11588283	1,51526E-21	1,580643173	1,790760964	1,580643173	1,790760964

Sans surprise, les résultats de cette régression sont cohérents. Pourtant, comment expliquer la différence de la pente des régressions ? (1,58 en 1995 contre 1,69 en 1985)

Cette différence signifie qu'en 1985, il y a plus de déplacements en attraction qu'en 1995 pour le même nombre d'emplois (tout du moins pour les mêmes proportions).

On peut l'expliquer par la longueur des chaînes : en 1985 nous étions plus mobiles et nous ne soucions moins de la longueur de nos déplacements. On peut envisager que certains actifs rentraient chez eux lors de la pause déjeuner par exemple ; comptant ainsi un trajet supplémentaire.