

Analyse de la régression

Aymen Bouhou

Introduction :

Le Haut-Commissariat au Plan a mené une enquête sur la consommation et les dépenses des ménages, cette enquête contient des dépenses annuelles du ménage, et parmi les dépenses étudiées on trouve **la dépense des légumes secs en conserve** (Pois chiches, pois cassés, lentilles vertes, brunes, beluga ou corail, haricots blancs, rouges, noir...), leur consommation revêt un grand intérêt pour les consommateurs qui aiment prendre soin de leur santé, et qui ont besoins des sources nutritionnelles d'énergie mais n'ont pas le temps ou les moyens de préparer un repas de légumes secs normales, alors ils prennent ces légumes en conserve comme alternative.

Choix des variables :

Il est donc intéressant de suivre l'évolution de la consommation de ces produits, nous essayerons de citer des facteurs qui peuvent expliquer la consommation des légumes secs en conserve à partir des informations collecter sur les ménages de la part de l'HCP, Le premier sera le milieu de vie est ce qu'il s'agit d'un milieu rural ou urbain car la culture de consommation des deux milieu sont différentes les urbain consomme plus les conserves que les ruraux donc cela peut ne donner des indications. Le deuxième facteur sera la région, nous savons qu'on consomme de légumes secs pour de l'énergie et ils ne protègent du froid donc leurs consommation sera plus grand en régions froides et aussi les cultures et les plats des différentes régions des Maroc sont très différentes. On passe à la taille de ménage elle fera aussi de la différence, plus de membres de famille plus de consommation, on prendra l'âge comme facteur car les matières conserver sont déconseiller pour les gens âgées donc la consommation variera par rapport au différentes catégorie d'âges. Le facteur suivant l'état matrimonial, les horaires des repas pour les personnes mariées sont exactes dont ils seront préparés d'avance premièrement pour économiser les budgets car comme on sait le prix des conserves est un peu élevé par rapport au légumes secs normaux, donc la consommation des boîtes conserves sera faible, au contraire les célibataires ou les divorcés leurs horaires de repas sont flexibles, donc ils tendent vers la consommation des conserves plus. On peut noter que l'âge est aussi un facteur important car il est déconseillé pour les personnes âgées de prendre les produits conservés donc leurs consommation sera inférieure à celle des jeunes personnes. Le niveau scolaire reflète la culture de la personne et son mode de vie donc il pourra nous donner une idée s'il consomme les légumes secs ou non et ce qu'elles sont en conserve ou non, il donne aussi des idées sur le salaire du chef du ménage et la consommation augmente avec le salaire. On ajoute un autre facteur le sexe du chef de ménage, comme on le sait la plupart des familles marocaines sont traditionnel, genre la mari travaille et la femme prend soins de la maison donc la consommation des conservé sera moins par rapport à une famille dont le chef de ménage est une femme. Le dernier facteur sera la situation dans la profession habituelle durant les dernières 12 mois du chef de ménage, une chef de ménage salarié ne dépense pas de la même manière d'un chef de ménage inactif, et il ne dépense pas comme un chef de ménage d'un employeur donc ce facteur sera important pour avoir une idée sur la dépense en légumes secs en conserve, on vient de déterminer les facteurs qui pourront expliquer la consommation des légumes secs en conserve, on passe maintenant à l'analyse statistique en utilisant le logiciel Stata, on a choisi les variables qui peuvent expliquer la dépense sur les légumes secs en conserve le but maintenant est de savoir à quel point ces variables peuvent nous donner une idée ou une estimation de la dépense alors on prend une

base de données contenant un échantillon de 15470 observations prise du site du Haut-commissariat du plan.

Détection des points aberrants :

Comme on remarque que le nombre des ménages observés est très grand donc on trouve bien sûr qu'il existe des observations qui sont distantes des autres observations, ceci créera un problème car ces points la peuvent affecter la crédibilité de l'analyse, ils peuvent être due à la variabilité de la mesure, ou ils peuvent indiquer une erreur expérimentale, la **figure 1** permet de mieux visualiser ces points aberrants grâce à la boîte de moustache, la meilleure décision à prendre est d'éliminer ces observations pour améliorer la qualité de notre estimation, nous utilisons la distance de Cook qui est utilisée dans l'analyse de régression pour trouver des valeurs aberrantes influentes dans l'ensemble des variables prédictive, c'est un moyen d'identifier les points qui affectent négativement notre modèle de régression, alors cette méthode nous permet de détecter les points qui peuvent influencer notre modèle avec leurs forts effets de levier, la **figure 2** présente une données aberrantes ainsi celle détectée par cette méthode ainsi que le code utilisé pour la détecter, on passe à supprimer tous les observations aberrantes, on trouve que la qualité de la régression linéaire augmente la **figure 3** compare notre R^2 adj avant est après l'élimination des points aberrants.

La multicollinéarité :

Maintenant on détecte la multicollinéarité car elle peut augmenter la variance des coefficients de régression et les rendre instables et difficiles à interpréter, on mesure cette multicollinéarité entre les variables qualitatives par l'analyse de variance (ANOVA) comme la montre la **figure 4**, on trouve que p-value de la variable sexe est supérieure à 5% donc on doit rejeter l'hypothèse qu'elle ne soit pas colinéaire avec les autres variables, mais on les autres variables explicatives du modèle tant que leurs p-value sont tous largement inférieure à 5% ce qui signifie l'absence de la multicollinéarité, il reste une seule variable qui est quantitative « la taille de ménage » c'est la seule variable explicative quantitative, donc c'est pas la peine de calculer la matrice de corrélation et on la garde tous simplement, maintenant on doit trouver une remède pour éliminer multicollinéarité entre les différentes modalités des variables qualitatives on doit d'abord coder les variables qualitatives et répartir ces modalités comme des sous variables, le logiciel stata permet de coder ces variables heureusement par une seule commande, cette opération est pour trouver la meilleure combinaison possible de variables explicatives pour notre modèle suivant le critère d'information bayésien (BIC) qui est un critère de sélection des modèles parmi un ensemble fini de modèles, le modèle ayant la valeur BIC la plus basse est préféré, après cette opération on retient 24 variables donc on a éliminé 17 variables, la **figure 5** montre le modèle final retenue.

L'estimation :

L'étape suivante est d'estimer le modèle du coup on prend notre dépense des légumes sec en conserve, et les coefficients estimés on les interprète sous conditions que les autres variables sont stables, le premier paramètre nous donne que la dépense des légumes sec en conserve deux ménages qui ont les mêmes caractéristiques sauf le premier habite en milieu urbain et le deuxième habite en milieu rural est différente, le ménage qui habite le milieu urbain dépense un 77 MAD de plus que le ménage qui habite le milieu rural, le deuxième critère de répartition est la région, on

prend comme repère la région Tanger-Tétouan, on constate qu'un ménage qui habite à la région Marrakech, sa dépense des légumes secs en conserve est moins de 496 MAD par rapport à un ménage qui habite la région repère (Tanger-Tétouan) sachant que la seule différences entre ces ménages est la région , un ménage qui habite la région Laayoune-Sakia El Hamra dépense moins qu'un ménage en région repère (Tanger-Tétouan) de 646 MAD en légumes secs en conserve, un autre ménage qui habite la région de Beni Mellal on trouve qu'il consomme moins qu'un ménage qui habite à Tanger en en légumes secs en conserve de 350 MAD, un ménage qui habite la région Souss-Massa-Deraa sa consommation est inférieure de 353 MAD d'un ménage de la région repère (Tanger-Tétouan) en légumes secs en conserve, un autre ménage qui habite la région Guelmim-Oued Noun consomme de légumes secs en conserve moins qu'un ménage qui habite à Tanger de 533 MAD, un ménage qui habite la région Daraa-Tafilalt sa consommation de légumes secs en conserve est inférieure à celle d'un ménage de la région de Tanger de 252 MAD, un autre ménage qui habite à Casablanca consomme moins d'un ménage s'il habitait à Tanger de 214 MAD en légumes secs en conserve , un ménage qui habite la région oriental consommera moins de légumes secs en conserve qu'il habitait à Tanger de 211 MAD, Si ce ménage de Tanger habitait la région de Dakhla-Ouad Dahab il consommera moins de légumes secs en conserve de 207 MAD, s'il déménage a Fez-Meknès il consommera moins de 137 MAD par rapport s'il a reste à Tanger, et un ménage de Rabat-Sale consomme moins de légumes secs en conserve d'un ménage de Tanger-Tétouan de 116 MAD, le critère suivant sera la taille de ménage en moyenne le nombre de membres de famille augmente d'une seule personne la consommation de légumes secs en conserve augmente de 41 MAD, on prend aussi le critère de l'état matrimonial on trouve qu'une personne marie consomme les légumes secs en conserve plus qu'un personne célibataire de 69 MAD, le critère d'âge fait aussi la différence car un chef de ménage de tranche d'âge 35-39 sa consommation de légumes secs conservés est moins de 194 MAD d'un chef de ménage de tranche d'âge de 15-19, et s'il de tranche d'âge entre 40 ans et 45 ans il consommera moins d'un chef de ménage d'âge entre 15 ans et 19 ans de 89 MAD, et un chef de ménage de tranche d'âge entre 70 et 74 consomme plus d'un chef de ménage d'âge entre 15 ans et 19 ans de 37 MAD de légumes secs en conserve, le critère de niveau scolaire fait aussi une différence, le niveau scolaire primaire du chef de ménage augmente la consommation de 75 MAD par rapport à un chef de ménage sans niveau scolaire, un niveau d'étude supérieur collège augmente la consommation des légumes secs de 50 MAD par rapport à un chef de ménage sans niveau scolaire, un niveau de secondaire qualifiant l'augmente avec 102 MAD par rapport à un chef de ménage sans niveau scolaire ,et enfin un niveau d'étude supérieure augmente la consommation des légumes secs en conserve de 69 MAD par rapport à un chef de ménage sans niveau scolaire, le dernier critère de différence est la situation professionnel dans les 12 derniers mois, une personne inactif consomme plus d'un chômeur en légumes secs en conserve de 35 MAD , en même temps un chef de ménage indépendant consomme plus qu'un chômeur de 127 MAD en légumes secs en conserve, et on trouve que les chefs de ménage qui n'ont pas déclaré leurs situation professionnel pendant les 12 derniers mois ont une dépense moins de 132 MAD de légumes secs en conserve d'un chef de ménage chômeur.

Inférence :

Maintenant on a le modèle final obtenue par le critère d'information bayésien contient 24 variables, on doit savoir si chacune variable joue un rôle explicatif dans le modèle, on effectue un test de Student ou test de significativité du coefficient de la variable explicative, on trouve que p-

value pour chacune des variables explicatives du modèle retenu est supérieur à 5% donc tous les variables sont significatives.

Pour tester la significativité globale du test on utilise la test de Fisher qui permet de déterminer si l'ensemble des variables explicatives prises simultanément, l'hypothèse de base est que les variables prises ne permettent pas d'expliquer les variations de la dépense, la p-value du test est largement inférieur à 5%, alors l'hypothèse donc il existe au moins un coefficient significatif.

Le R^2 et le R^2 ajusté nous donnent une idée du nombre de points de données qui se trouvent dans la ligne de l'équation de régression. Cependant, le R^2 ajusté nous indique le pourcentage de variation expliqué par les seules variables indépendantes qui affectent réellement la variable dépendante est dans notre cas on a trouvé R^2_{adj} égale à 15,73% donc on peut seulement expliquer partiellement la dépense en légumes secs en conserve mais il ne permet pas la prédiction de ces dépenses.

Validation du modèle :

La validation peut consister à analyser la qualité de l'ajustement de la régression, à analyser si les résidus de la régression sont aléatoires et à vérifier si la performance prédictive du modèle se détériore sensiblement, on commence par le test de normalité pour savoir si les résidus suivent une loi normale pour cela on utilise la droite de Henry qui est une présentation graphique permettant d'évaluer la normalité d'une distribution, on trouve que la représentation des fréquences réelles cumulées n'épouse pas la trajectoire de la droite de Henry. En conclusion, la forme de distribution de cette variable ne décrit pas exactement une courbe Gaussienne comme le montre la **figure 6**, donc l'hypothèse de normalité n'est pas vérifiée, il y a pas mal d'autres méthode pour vérifie la normalité, on peut utiliser le test de Shapiro-Wilk qui examine si une variable est normalement distribuée dans une population, la **figure 7** montre que la p-value de ce test tend vers 0 donc les résidus ne sont pas normalement distribués, on essaye de régler ce problème avec la transformation de Box Cox qui est une transformation qui permet de rendre une variable dépendante non normale en une forme normale, mais malheureusement il y en a des ménages avec une dépense nulle en légumes secs en conserve donc l'utilisation de la transformation de Box Cox est impossible pour notre analyse alors il y a pas de solution pour que la dépense prend une forme normale.

L'homoscédasticité, ou tout simplement on peut l'appeler l'homogénéité des variances, est une hypothèse de comparaison de variances égales ou similaires dans des différents groupes. Il s'agit d'une hypothèse très importante des tests statistiques paramétriques, la **figure 8** contient le test de Breusch-Pagan de l'homoscédasticité qui vérifie si la variance des erreurs d'une régression dépend des valeurs de notre variables explicatives, la p-value de ce test est inférieur à 5% donc on trouve que les variances des résidus des variables examinées sont largement différentes, donc on parle d'un cas d'hétéroscédasticité, la solution de ce problème n'est rien qu'une inférence robuste, cette méthode utilise des estimateurs MCO cohérents, et calcule une variance alternative pour rendre les estimateur constant et comme ça notre problème est réglé, la **figure 9** nous montre la démarche de ce test.

L'autocorrélation est le problème suivant à régler, la procédure est d'étudier le degré de corrélation des mêmes variables entre deux intervalles de temps successifs, on mesure la façon dont la version décalée de la valeur d'une variable est liée à la version originale de celle-ci dans une série temporelle, pour la constater on utilise le test de Durbin et Watson qui permet de tester ce qu'on a cité précédemment et de détecter une autocorrélation des erreurs d'ordre un comme la

montre la **figure 10**, le test nous donne la valeur 1,64 on compare cette valeur a la région critique du test par le tableau de Durbin et Watson , on trouve qu'elle est proche de 2, donc heureusement l'autocorrélation n'existe pas.

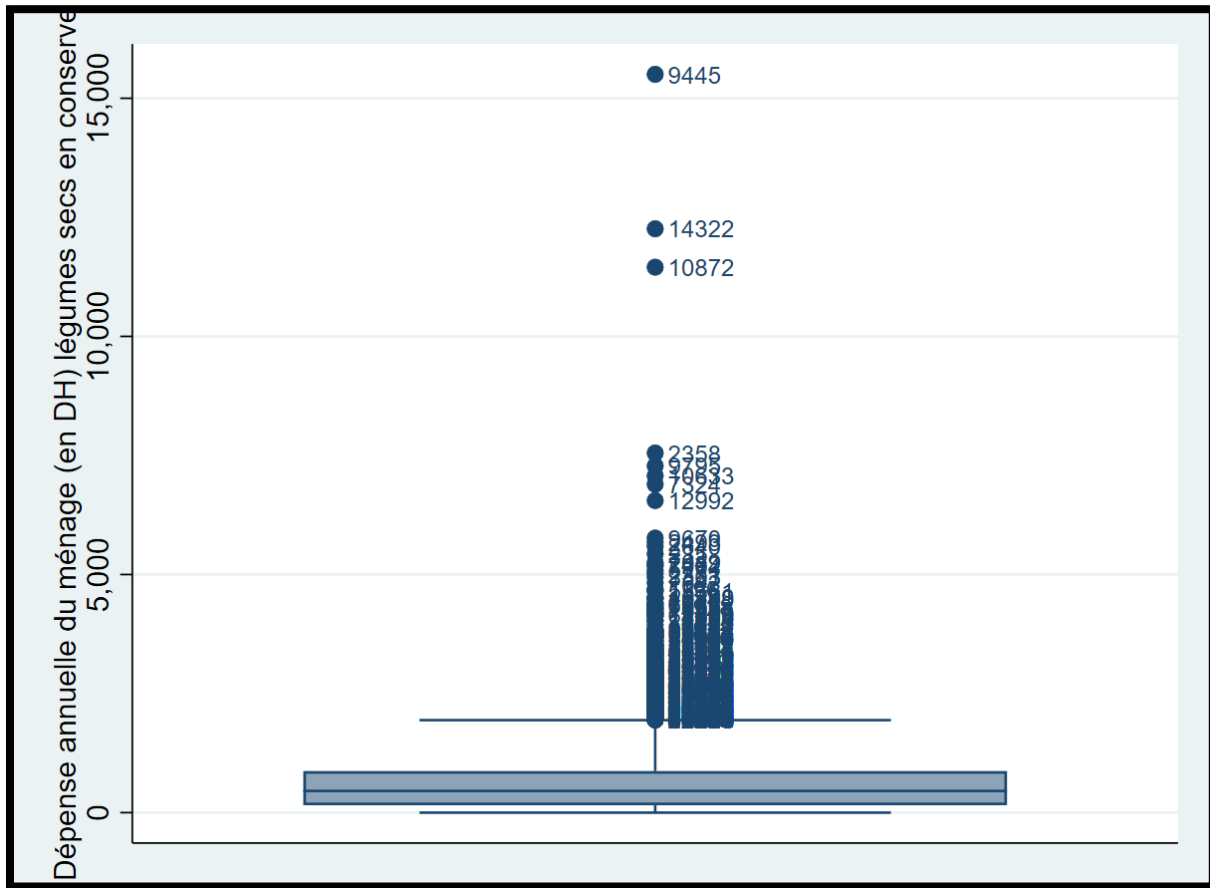
La dernière station de notre analyse est le test de Ramsey, ce test vérifie si les combinaisons des valeurs ajustées aident à expliquer la variable de dépense, l'hypothèse de base est que les variables expliquent la dépense donc on tester est savoir à quel point cette hypothèse est vraie la **figure 11** nous donne que la p-value du test de Ramsey est inférieure a 5%, donc on conclue que notre modèle est mal spécifié, il pourrait être mieux approché par un polynôme ou une autre fonction non linéaire.

Conclusion :

On a terminé notre démarche de régression linéaire, on a s'il y en a une relation linéaire entre la dépense en légumes secs en conserve, les variables qu'on a précisé au début, c'est le temps de récapituler les résultats, la qualité du modèle choisis est très faible il ne permet pas d'avoir une idée claire sur la dépense donc on ne nous permet pas estimer avec une grande précision la dépense d'un ménage en légumes secs en conserve met il permet de l'expliquer envers 16% , de plus il y en a plusieurs problèmes de normalité ainsi que l'hétéroscédasticité, le test de Ramsey a confirmé que le modèle est mal spécifié, peut être qu'un modèle non linéaire permettra d'expliquer est estimer la dépense en légumes secs en conserves mieux que celui-ci ou ajouter d'autres facteurs qui peuvent aider le modèle pour être plus spécifique.

Annexe (tous les sorties des logiciels sont accompagnées par le code stata exécuté) :

- **Figure1**(Boite a moustache) :



- **Figure2** (code de détection des points aberrants et une observation aberrante):
 - predict d, cooks
 - list N_ménage DAM_G072 Milieu Région_12 Taille_ménage Sexe_CM Age_quin_CM Etat_matrimonial_CM Niveau_scolaire_agreg_CM Situation_profession_agreg_CM if d>4/(15470-41-1)
 - drop if d>4/(15470-41-1)

15460.	N_ménage 15031	DAM_G072 0	Milieu Urbain	Région_12 Dakhla-Oued	Taille~ge 7	Sexe_CM Masculin	Age_quin~M 40-45	Etat_matr~M Marié
	Niveau_sco~M Secondaire q				Situation~M Salarié			

- **Figure3** (regression avant et après l'élimination des points aberrants):

Avant l'élimination des points aberrants:

```
. reg DAM_G072 i.Milieu i.Région_12 Taille_ménage i.Sexe_CM i.Age_quin_CM i.Etat_matrimonial_CM i.Niveau_scolaire_agreg_CM i.Situation_profession_agreg_CM
```

Source	SS	df	MS	Number of obs	=	15,470
Model	777148433	41	18954839.8	F(41, 15428)	=	49.87
Residual	5.8645e+09	15,428	380120.419	Prob > F	=	0.0000
				R-squared	=	0.1170
				Adj R-squared	=	0.1147
Total	6.6416e+09	15,469	429352.011	Root MSE	=	616.54

Après l'élimination des points aberrants:

```
. reg DAM_G072 i.Milieu i.Région_12 Taille_ménage i.Sexe_CM i.Age_quin_CM i.Etat_matrimonial_CM i.Niveau_scolaire_agreg_CM i.Situation_profession_agreg_CM
```

Source	SS	df	MS	Number of obs	=	14,875
Model	575084528	41	14026451.9	F(41, 14833)	=	69.40
Residual	2.9978e+09	14,833	202105.561	Prob > F	=	0.0000
				R-squared	=	0.1610
				Adj R-squared	=	0.1586
Total	3.5729e+09	14,874	240212.204	Root MSE	=	449.56

- **Figure4** (test ANOVA) :

```
. anova DAM_G072 Milieu Région_12 Sexe_CM Age_quin_CM Etat_matrimonial_CM Niveau_scolaire_agreg_CM Situation_profession_agreg_CM
```

```

      Number of obs =      14,875      R-squared      =    0.1345
      Root MSE      =      456.58      Adj R-squared =    0.1322

```

Source	Partial SS	df	MS	F	Prob>F
Model	4.805e+08	40	12013617	57.63	0.0000
Milieu	7151862.5	1	7151862.5	34.31	0.0000
Région_12	3.452e+08	11	31380351	150.53	0.0000
Sexe_CM	1614676.1	1	1614676.1	7.75	0.0054
Age_quin_CM	21621609	12	1801800.7	8.64	0.0000
Etat_matrimonial_CM	15087235	3	5029078.5	24.12	0.0000
Niveau_scolaire_agreg_CM	12573787	6	2095631.2	10.05	0.0000
Situation_profession_agreg_CM	8081580.6	6	1346930.1	6.46	0.0000
Residual	3.092e+09	14,834	208465.12		
Total	3.573e+09	14,874	240212.2		

- **Figure5** (Modèle final d'après le BIC):

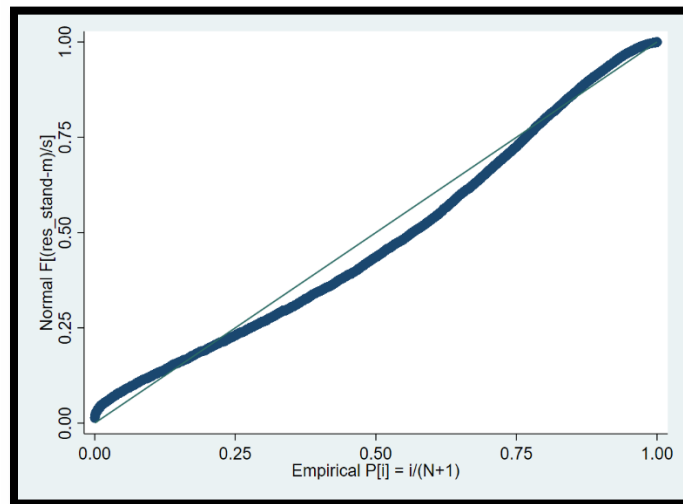
xi:vselect DAM_G072 i.Milieu i.Région_12 Taille_ménage i.Age_quin_CM
i.Etat_matrimonial_CM i.Niveau_scolaire_agreg_CM i.Situation_profession_agreg_CM,
forward bic

Final Model

Source	SS	df	MS	Number of obs	=	14,875
Model	566871768	24	23619657	F(24, 14850)	=	116.68
Residual	3.0060e+09	14,850	202427.242	Prob > F	=	0.0000
				R-squared	=	0.1587
				Adj R-squared	=	0.1573
Total	3.5729e+09	14,874	240212.204	Root MSE	=	449.92

DAM_G072	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
_IRégion_1_7	-496.0076	15.72746	-31.54	0.000	-526.8354	-465.1798
Taille_ménage	41.02626	1.823054	22.50	0.000	37.45284	44.59967
_IRégion_1_11	-646.3224	27.92874	-23.14	0.000	-701.0662	-591.5786
_IRégion_1_5	-350.4631	16.86656	-20.78	0.000	-383.5236	-317.4025
_IRégion_1_9	-353.2162	17.46495	-20.22	0.000	-387.4496	-318.9827
_IRégion_1_10	-533.07	31.49163	-16.93	0.000	-594.7975	-471.3425
_IMilieu_2	-77.28186	8.48181	-9.11	0.000	-93.90726	-60.65647
_IRégion_1_6	-214.7287	14.58458	-14.72	0.000	-243.3163	-186.1411
_IEtat_matr_2	69.7542	10.49729	6.64	0.000	49.17821	90.33018
_IRégion_1_8	-252.4967	21.77765	-11.59	0.000	-295.1836	-209.8098
_IRégion_1_2	-211.5494	17.38304	-12.17	0.000	-245.6223	-177.4765
_IRégion_1_3	-137.7903	15.42713	-8.93	0.000	-168.0294	-107.5512
_IRégion_1_4	-116.6344	15.11703	-7.72	0.000	-146.2656	-87.00312
_IRégion_1_12	-207.7097	30.37318	-6.84	0.000	-267.2449	-148.1745
_INiveau_sc_5	102.8173	16.27053	6.32	0.000	70.92502	134.7095
_IAge_quin_5	-194.1969	42.34396	-4.59	0.000	-277.1963	-111.1975
_INiveau_sc_3	75.27005	12.90201	5.83	0.000	49.98051	100.5596
_IAge_quin_6	-89.43569	22.63174	-3.95	0.000	-133.7967	-45.07468
_INiveau_sc_6	69.14656	16.06073	4.31	0.000	37.66553	100.6276
_ISituation_4	127.3732	32.05301	3.97	0.000	64.54534	190.2011
_ISituation_1	35.61277	9.239529	3.85	0.000	17.50215	53.72339
_INiveau_sc_4	50.7629	13.48479	3.76	0.000	24.33105	77.19475
_IAge_quin_12	37.69156	10.9789	3.43	0.001	16.17155	59.21157
_ISituation_9	-132.7159	40.19465	-3.30	0.001	-211.5024	-53.92943
_cons	526.0083	16.41314	32.05	0.000	493.8365	558.1801

- **Figure6** (droite de Henry) :



- **Figure7** (le test de Shapiro-Wilk) :

```
predict res_stand, rstand

pnorm res_stand

swilk res_stand
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
res_stand	14,875	0.95251	331.818	15.698	0.00000

- **Figure8** (test de Breusch-Pagan) :

```
. estat hettest
```

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
 Ho: Constant variance
 Variables: fitted values of DAM_G072

chi2(1)	=	866.74
Prob > chi2	=	0.0000

- **Figure9** (inférence robuste):

```
. regress DAM_G072 _IRégion_1_7 Taille_ménage _IRégion_1_11 _IRégion_1_5 _IRégion_1_9
> Milieu_2 _IEtat_matr_2 _IEtat_matr_4 _IRégion_1_8 _IRégion_1_6 _IRégion_1_2 _IRégion_1_10
> 3 _IRégion_1_4 _IAge_quin_5 _ISituation_9 _INiveau_sc_3 _INiveau_sc_5 _IAge_quin_12
> vce(robust)
```

Linear regression

Number of obs = 14,875
F(21, 14853) = 173.78
Prob > F = 0.0000
R-squared = 0.1561
Root MSE = 450.56

DAM_G072	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
_IRégion_1_7	-494.865	16.05628	-30.82	0.000	-526.3373 -463.3927
Taille_ménage	40.66481	1.939642	20.97	0.000	36.86287 44.46675
_IRégion_1_11	-646.9698	18.85969	-34.30	0.000	-683.9371 -610.0025
_IRégion_1_5	-347.2414	18.00396	-19.29	0.000	-382.5313 -311.9514
_IRégion_1_9	-353.2155	17.82492	-19.82	0.000	-388.1545 -318.2764
_IRégion_1_10	-532.1631	23.69203	-22.46	0.000	-578.6024 -485.7238
_IMilieu_2	-90.25817	8.192221	-11.02	0.000	-106.3159 -74.2004
_IEtat_matr_2	109.5267	12.76552	8.58	0.000	84.50471 134.5487
_IEtat_matr_4	67.82354	14.94392	4.54	0.000	38.5316 97.11548
_IRégion_1_8	-249.7071	22.10623	-11.30	0.000	-293.0381 -206.3762
_IRégion_1_6	-209.6846	16.48457	-12.72	0.000	-241.9964 -177.3728
_IRégion_1_2	-208.2188	19.32971	-10.77	0.000	-246.1074 -170.3302
_IRégion_1_12	-214.2832	30.77607	-6.96	0.000	-274.6081 -153.9583
_IRégion_1_3	-135.1158	17.81158	-7.59	0.000	-170.0287 -100.2029
_IRégion_1_4	-114.191	17.40665	-6.56	0.000	-148.3102 -80.07184
_IAge_quin_5	-166.4138	25.92044	-6.42	0.000	-217.2211 -115.6065
_ISituation_9	-138.0472	27.35663	-5.05	0.000	-191.6696 -84.42481
_INiveau_sc_3	64.55297	12.50276	5.16	0.000	40.04602 89.05992
_INiveau_sc_5	92.01105	17.30826	5.32	0.000	58.08472 125.9374
_IAge_quin_12	37.71047	11.35316	3.32	0.001	15.45688 59.96406
_INiveau_sc_6	66.27747	16.84098	3.94	0.000	33.26706 99.28787
_cons	503.7561	18.13744	27.77	0.000	468.2044 539.3077

- **Figure10** (Durbin et Watson) :

```
. gen t=_n
. tsset t
      time variable: t, 1 to 14875
      delta: 1 unit
. estat dwatson
Durbin-Watson d-statistic( 25, 14875) = 1.640966
```

- **Figure11** (test de Ramsey):

```
. ovtest
Ramsey RESET test using powers of the fitted values of DAM_G072
Ho: model has no omitted variables
      F(3, 14847) =      6.85
      Prob > F =      0.0001
```