



HAL
open science

Analyse des données en coupe instantanée : Application sur le logiciel Stata

Nathan Mbende, Jean-Claude Nkashama Mukenge, Jonathan Tshitolo

► To cite this version:

Nathan Mbende, Jean-Claude Nkashama Mukenge, Jonathan Tshitolo. Analyse des données en coupe instantanée : Application sur le logiciel Stata. 2024. hal-04362269

HAL Id: hal-04362269

<https://hal.science/hal-04362269>

Preprint submitted on 5 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyse des données en coupe instantanée : Application au logiciel stata.

Nathan Mbende(nathanmbende4@gmail.com)
Jean-Claude Nkashama(jackonkashama@gmail.com)
Jonathan Tshitolo(jonathantshitolo70@gmail.com)

,

JANVIER 2024



Table des matières

1	Données en coupe instantanée	8
2	Données Quantitatives et Qualitatives	8
2.1	Données quantitatives	8
2.1.1	Types des données quantitatives	8
2.2	Données qualitatives	8
2.2.1	Types des données qualitatives	9
3	Console Stata 17	10
4	Analyses univariées	12
4.1	Analyse descriptive	13
4.2	Tableau des fréquences	15
5	Analyses bivariées	18
5.1	Analyse de relation	18
5.2	Analyse de corrélation	21
5.3	Analyse de comparaison	25
6	Modèles de régression	35
6.1	Régression Linéaire Multiple et simple	37
6.2	Régression Multiple et simple à variables discontinues	46
6.2.1	Logit/Probit	47
6.2.2	Ologit Multiple	59
6.3	Mlogit Multiple	63
7	Appendice 1 : Force et direction du coefficient de corrélation	68
8	Appendice 2 :La commande edit	68
9	Appendice 3 : Importer fichier Excel	70
10	Appendice 4 : La commande asdoc	73
11	Appendice 5 : Install.	80
12	Appendice 6 : Describe	81
13	Appendice 7 : Help	82
14	Appendice 8 : Option detail	83
15	Appendice 9 : Tabstat,stats	86
16	Appendice 10 : Ameans	87
17	Appendice 11 : Analyse graphique de la distribution normale	88

18 Appendice 12 : Tab2,firstonly	94
19 Appendice 13 : Tau de kendall's	95
20 Appendice 14 : Test post hoc de Welch	96
21 Appendice 15 : Test post hoc Dunett's	97
22 Appendice 16 : Ologit simple	98
23 Appendice 17 : Mlogit simple	101
24 Appendice 18 : Oprobit simple	103
25 Appendice 19 : Oprobit Multiple	106
26 Appendice 20 : Mprobit Simple	112
27 Appendice 21 : Mprobit Multiple	114

Liste des tableaux

1	Exemple d'une variable qualitative ordinale	9
2	Présentation des variables	18
3	Coefficient	68

Table des figures

1	Présentation stata	10
2	Tableur stata	10
3	Les moyens de collage	11
4	L'option Variables names	11
5	Collage réussi	12
6	Analyse descriptive	13
7	Histogramme	14
8	Tableau des fréquences	15
9	Diagramme à barres	16
10	Diagramme circulaire	17
11	Chi2	19
12	Fisher exact	20
13	Normalité Shapiro	21
14	Normalité Skwness et Kurtosis	22
15	Corrélation de Pearson	23
16	Corrélation de Spearman	24
17	Kolmogorov Smirnov	25
18	Moyennes sous-groupes	26
19	Test de levène	27
20	Student	28
21	Kolmogorov Smirnov	29
22	Moyennes sous groupes	29
23	Mann Whitney	30
24	Moyennes sous-groupes	31
25	Anova	32
26	Moyennes sous-groupes	33
27	Kruskal Wallis	34
28	MCO	38
29	Normalité des résidus	39
30	Normalité des résidus	39
31	Breush-Pagan	40
32	Test de White	41
33	Spécification du modèle	42
34	Test de multicolinéarité	43
35	MCO simple	44
36	Linktest	45
37	Logit	47
38	Effets marginaux	49

39	Prédiction du modèle	50
40	Logit et effets marginaux	51
41	Prédiction du modèle	52
42	Probit	53
43	Effets marginaux	55
44	Prédiction	56
45	Probit et effet marginal	57
46	Prédiction	58
47	Ologit	59
48	Effets marginaux	60
49	Effets marginaux suite	61
50	Effets marginaux suite et fin	61
51	Spécification	62
52	Mlogit	63
53	Mlogit suite	64
54	Effets marginaux	65
55	Effets marginaux suite	66
56	Effets marginaux fin	67
57	Commande edit	68
58	Commande edit	69
59	Importer le fichier	70
60	Importer le fichier	71
61	Importer le fichier	71
62	Importation réussie	72
63	Asdoc	73
64	Asdoc	74
65	Asdoc	75
66	Asdoc	76
67	Asdoc	77
68	Asdoc	78
69	Aperçu word	79
70	Install	80
71	Describe	81
72	Help	82
73	Detail	83
74	Tabstat	86
75	Ameans	87
76	Normalité Histogramme	88
77	Normalité Histogramme	89
78	Normalité Histogramme	90
79	Kernel density	91
80	Kernel density	92
81	Kernel density	93
82	tab2,firstonly	94
83	Tau de kendall's	95
84	Post hoc Welch	96

85	Dunett's	97
86	Ologit simple	98
87	Effets Marginaux	99
88	Spécification du test	100
89	Mlogit simple	101
90	Effets Marginaux	102
91	Oprobit Simple	103
92	Effets Marginaux	104
93	Spécification du modèle	105
94	Oprobit	106
95	Effets Marginaux	107
96	Effets Marginaux suite	108
97	Effets Marginaux suite	109
98	Effets Marginaux fin	110
99	Spécification du modèle	111
100	Mprobit Simple	112
101	Effets Marginaux	113
102	Mprobit	114
103	Mprobit suite	115

Résumé

Un document pratique permettant de faire le pont entre les théories nécessaires, importantes et la pratique des notions liées à l'analyse des données en coupé instantanée sur le logiciel Stata. En commençant par les analyses univariées, bivariées et multivariées pour les variables qualitatives et quantitatives.

Le document donne aussi une introduction pour la prise en main et quelques astuces pour mieux utiliser Stata avec une approche de quelques exercices d'interprétation pour faire participer le lecteur.

Le document aborde à tour de rôle les analyses ci-après : Analyse descriptive, tableau des fréquences, Analyse graphique (Histogramme, diagramme circulaire et diagramme en barres) test de Chi2, test de Cramer, test de Gamma, test Fisher exact, test de normalité (Shapiro Wilk, test de Skwness et Kurtosis ainsi que les analyses graphiques), test de Pearson, test de Spearman, test de Tau de Kendall, test de Kolmogorov Smirnov, test de Levène, test de student, test post hoc de Welch, test de Mann Whitney, test de Bartlett, test d'Anova, test post hoc de Dunett's, test de Kruskal Wallis, Méthode des moindres carrés ordinaires, régression Logit avec ses formes (Logit, ologit et mlogit) et la régression Probit (probit, oprobit et mprobit)..

1 Données en coupe instantanée

¹On appelle données en coupe instantanée, transversale, les données qui sont collectées au même moment ou approximativement au même moment.

2 Données Quantitatives et Qualitatives

Les tests statistiques dépendent de la nature des données, d'où l'importance capitale de maîtriser et distinguer clairement les différentes natures des données existantes.

Il existe deux types des données statistiques. L'on parle alors des données quantitatives et qualitatives.

2.1 Données quantitatives

²Les données quantitatives ont des valeurs numériques et présentent un caractère dénombrable. I.E. calculable, admettant des opérations mathématiques ayant un sens. Elles sont des deux types.

2.1.1 Types des données quantitatives

L'on a les données quantitatives discrètes et continues.

La compréhension la plus simple des données discrète, ces sont des données quantitatives n'admettant pas des virgules.

Exemples : les nombres des personnes, salles, l'âge, etc.

Les données continues sont des quantitatives admettant des virgules, des quantitatives décimales.

Exemple : La taille d'une personne, son poids, la superficie, etc.

2.2 Données qualitatives

³Les données qualitatives incluent des labels ou des noms utilisés pour identifier une caractéristique de chaque élément.

Elles peuvent aussi utiliser des codes numériques.

Elles n'admettent pas les opérations mathématiques car les résultats issus de ces opérations seront vides de sens mais admettent les opérateurs de comparaison.

⁴Elles contiennent des valeurs qui expriment une qualité, un état, une condition, un statut unique et exclusif.

Un autre jargon utilisé pour les données qualitatives est les données catégorielles.

Elles sont de deux types.

1. Anderson, Sweeney et Williams., (2010), *Statistiques pour l'économie et la gestion*, édition 3, de boeck.

2. Anderson, Sweeney et Williams., (2010), *Statistiques pour l'économie et la gestion*, édition 3, de boeck.

3. Anderson, Sweeney et Williams., (2010), *Statistiques pour l'économie et la gestion*, édition 3, de boeck.

4. Musangu M. (2018), *Elements de statistique descriptive*, G1 Faculté d'administration des affaires et sciences économiques, UPC, Kinshasa.

2.2.1 Types des données qualitatives

Nous avons deux types des données qualitatives, d'une part les nominales et de l'autre part les ordinales. Toujours selon l'approche simpliste, une donnée qualitative est dite nominale lorsque⁵l'on utilise des nombres, symboles, noms comme des simples étiquettes pour différencier des groupes auxquels les différentes observations appartiennent.

Elles n'acceptent pas des opérateurs de comparaison.

Exemples : La variable état matrimonial contient des éléments tels que célibataire, marié(e)s, divorcé(e)s et veuf(ve)s. dans notre base des données, nous décidons d'utiliser le code 2 pour tout célibataire, 4 pour les marié(e)s, 6 pour les divorcé(e)s et 8 pour les veuf(ve)s. ces codes numériques utilisés sont juste pour des raisons de distinction des différents groupes de la variable état matrimonial.

Une donnée qualitative est dite ordinale lorsqu'elle suit un ordre bien déterminer qui respecte les critères préalablement établis et utilisent les nombres et symboles.

Elle admet les opérateurs de comparaison.

Exemple : les mentions que l'on retrouve au sein des universités.

TABLE 1 – Exemple d'une variable qualitative ordinale

Mention
Plus grande distinction
Grande distinction
Distinction
Satisfaction
Ajourner
Non accessible à la filière

Source : L'auteur, inspiré

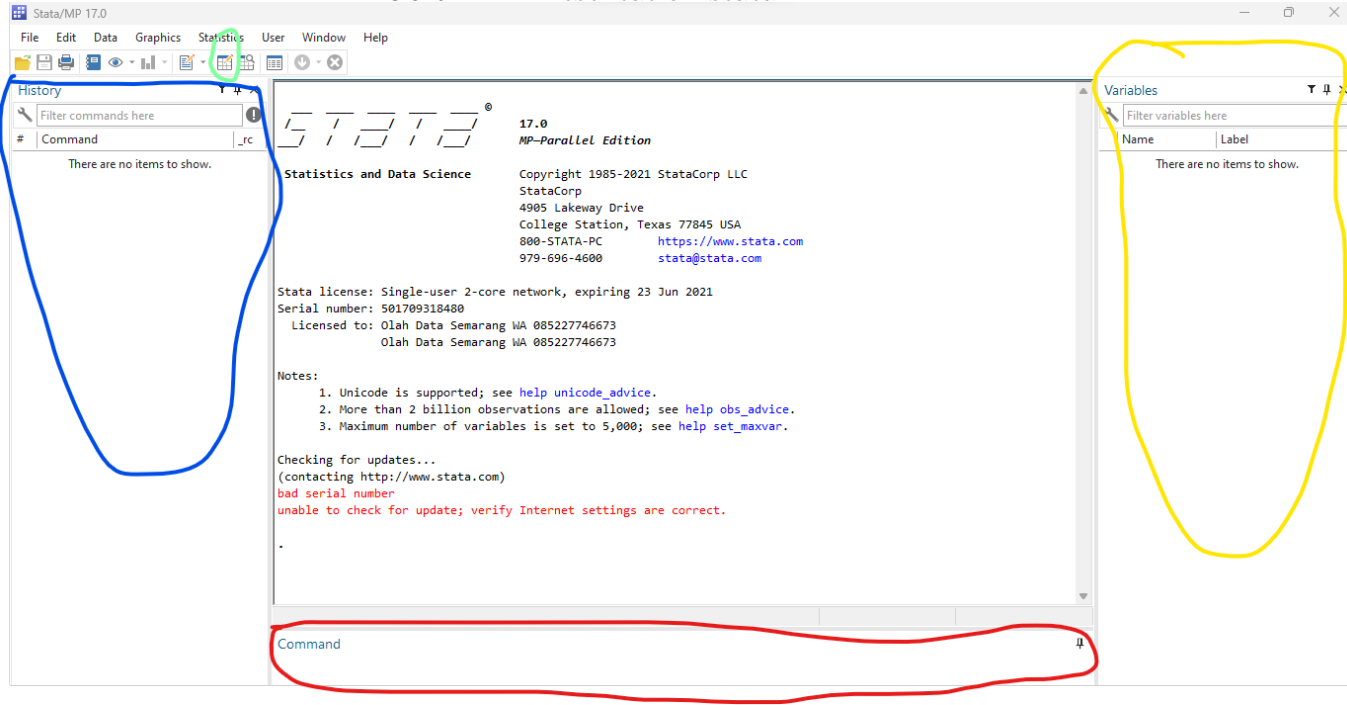
Commentaire : l'on peut effectuer des opérations de comparaison telles que la mention grande distinction est $>$ à la mention satisfaction.

5. Musangu M. (2018), *Elements de statistique descriptive*, G1 Faculté d'administration des affaires et sciences économiques, UPC, Kinshasa.

3 Console Stata 17

Voici à quoi ressemble la console Stata V.17.

FIGURE 1 – Présentation stata

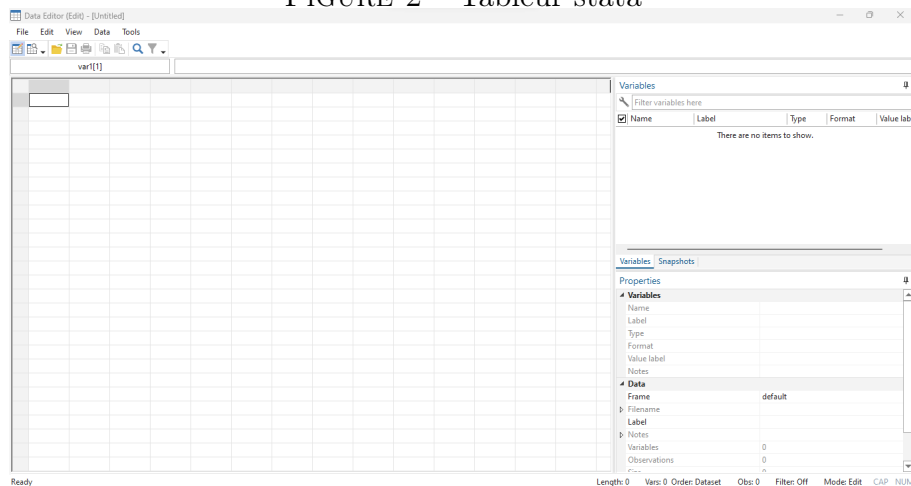


Commentaire :

La partie encerclée en rouge, c'est l'espace réservé aux commandes stata, en bleu, c'est pour garder et consulter votre historique des commandes, en jaune, c'est l'espace ou apparaîtra les variables une fois insérer et en vert, l'espace pour coller une base des données Excel.

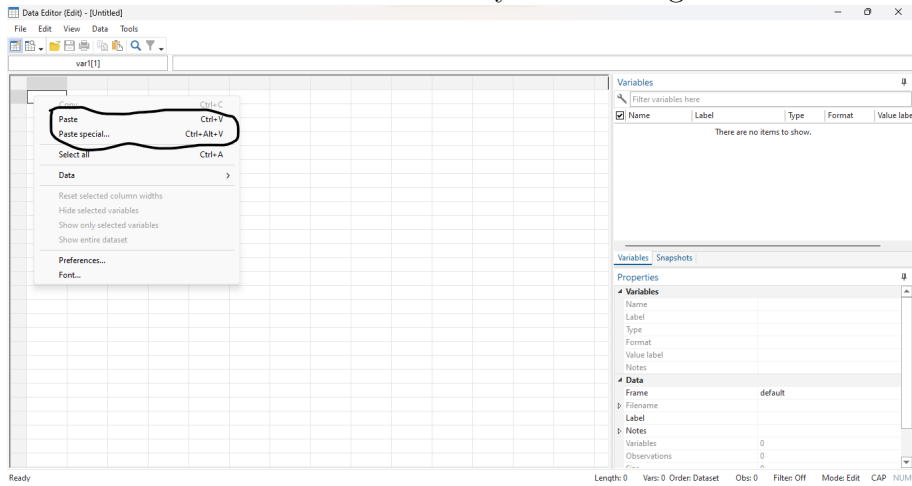
En cliquant sur la partie verte, stata ouvre une autre fenêtre pour vous permettre de coller votre base des données en format Excel.

FIGURE 2 – Tableur stata



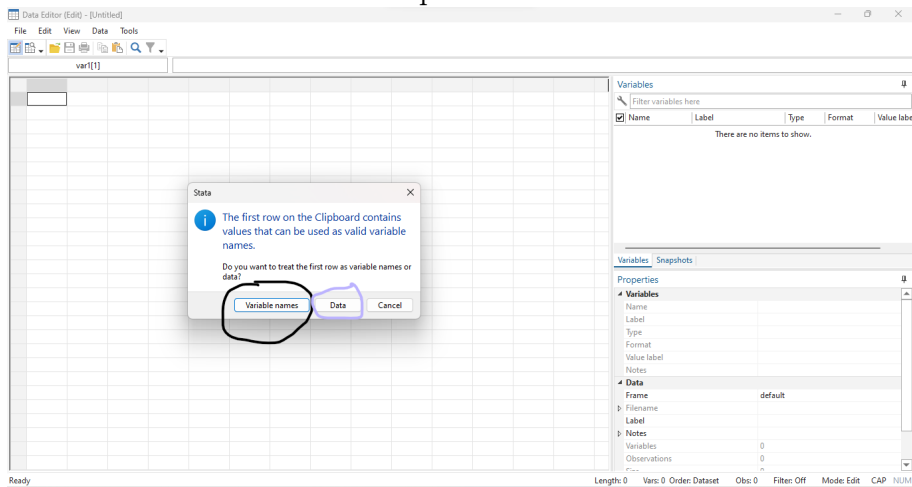
Pour coller la base des données, stata propose deux moyens, le moyen le plus rapide est juste de cliquer sur paste.

FIGURE 3 – Les moyens de collage



il faut ensuite, opter pour variables names.

FIGURE 4 – L'option Variables names

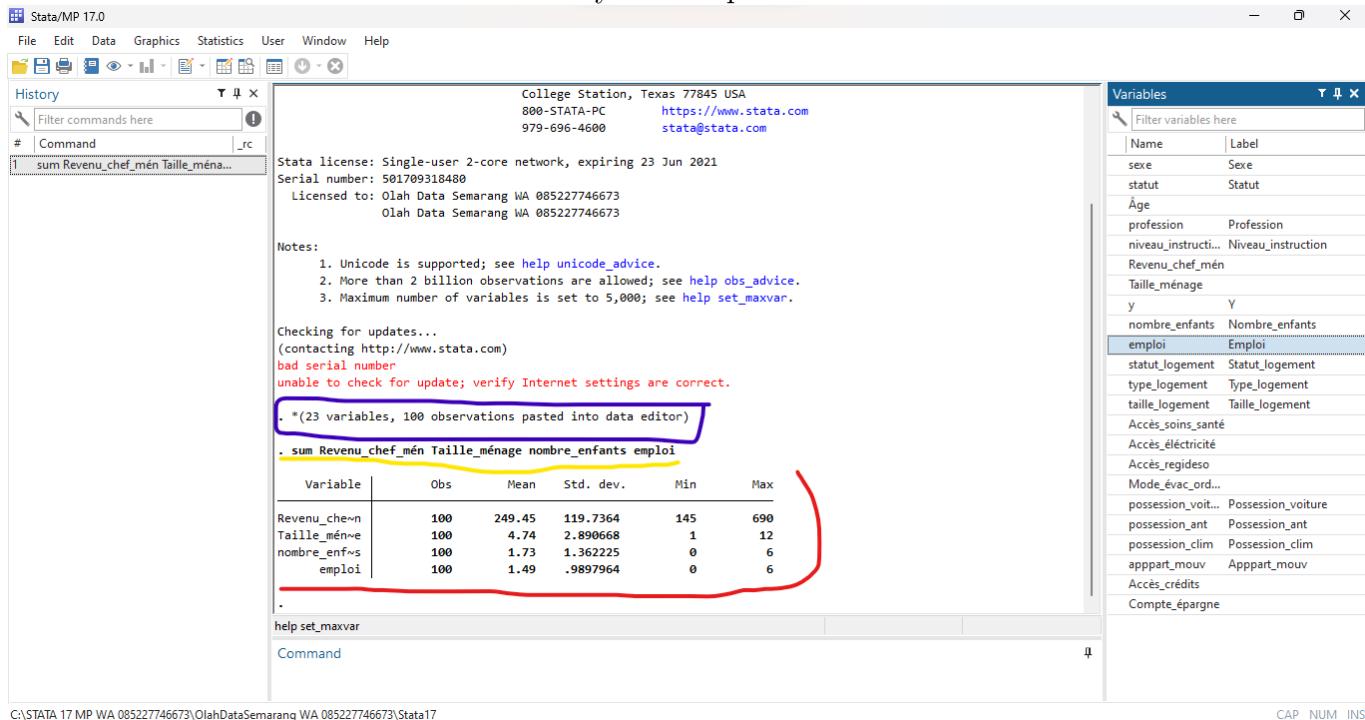


4.1 Analyse descriptive

C'est l'analyse univariée des données quantitatives, elle permet de résumer une variable avec quelques indicateurs de tendance centrale et de la variabilité telles que la moyenne, l'écart-type, le minimum, maximum. et elle renseigne aussi sur le nombre des observations.

La commande est sum suivi de la variable étudiée ou des variables étudiées.

FIGURE 6 – Analyse descriptive



Commentaire :

Le revenu mensuel des chefs de ménage est 249,45\$, avec un écart type qui montre que les valeurs sont très dispersés autour de la moyenne. Le revenu minimal est 145 et maximal 690\$ pour nos 100 ménages enquêtés.

En moyenne, nous avons des ménages composés de 5 personnes (4,74 étant une variable quantitative discrète) les données sont d'une façon étroite autour de la moyenne. La taille minimum de ménage est de une personne et maximum 12 personnes dans un ménage pour les 100 ménages observés.

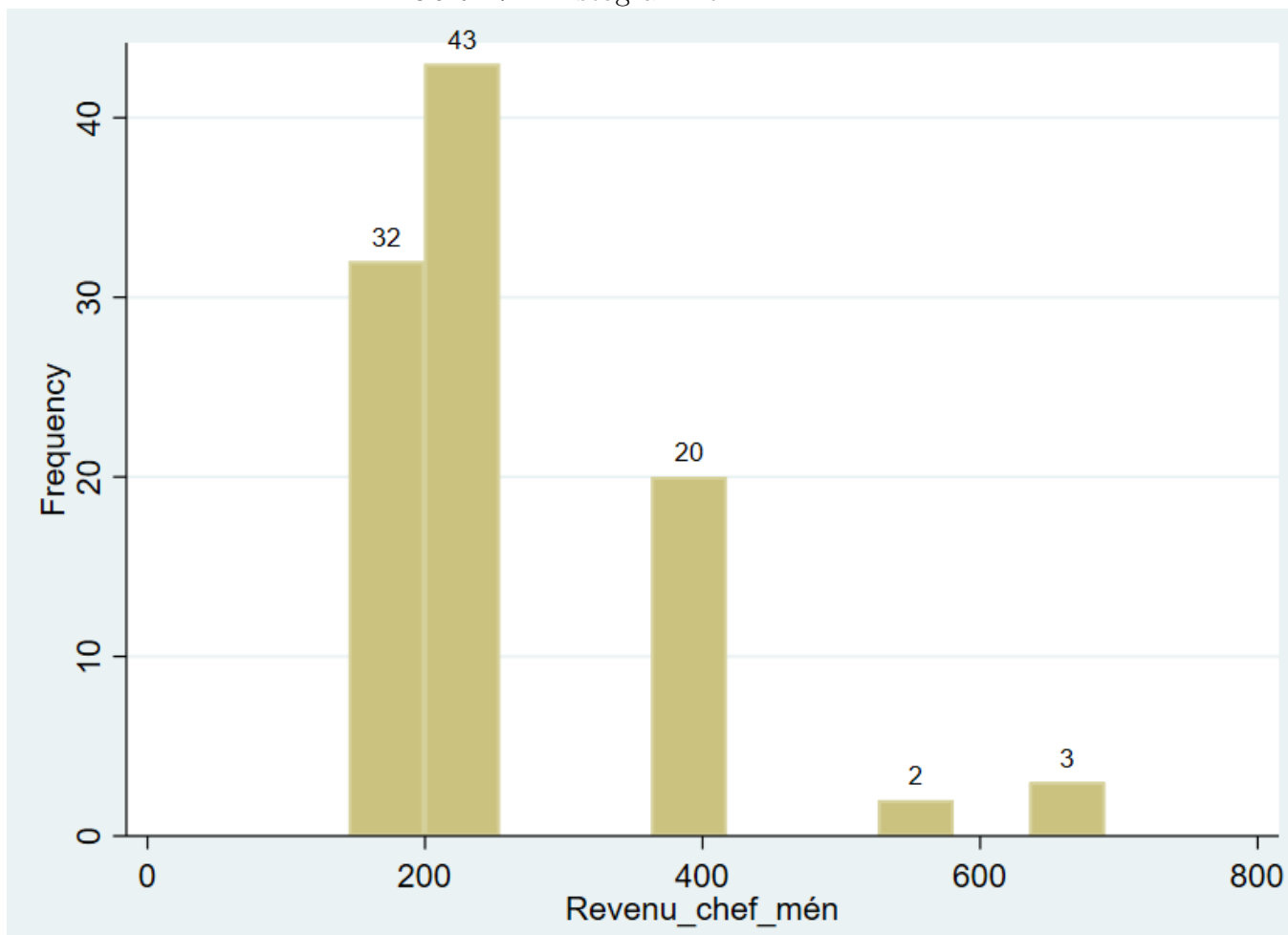
Par ménage, nous avons en moyenne 2 enfants (1.73) , avec un écart type qui montre que les données sont autour de la moyenne d'une façon étroite, le nombre d'enfant minimum est de 0 ce qui traduit qu'il y'a des ménages sans enfants et le maximum est de 6 pour nos 100 ménages.

Par ménage, nous avons en moyenne 1 personnes ayant un emploi, l'écart type traduit une distribution étroite des données autour de la moyenne, il existe des ménages où personne ne travaille d'où le minimum de 0 et en maximum des ménages dans lequel 6 personnes travaillent pour nos 100 ménages.

Nous pouvons aussi générés des graphiques. Il en existe plusieurs tels que l'histogramme de fréquence, le polygone de fréquence, l'ogive, etc. Nous allons présenter ici l'histogramme de fréquence.

la commande est `histogram Revenu chef mén,frequency addlabel`

FIGURE 7 – Histogramme



Commentaire :

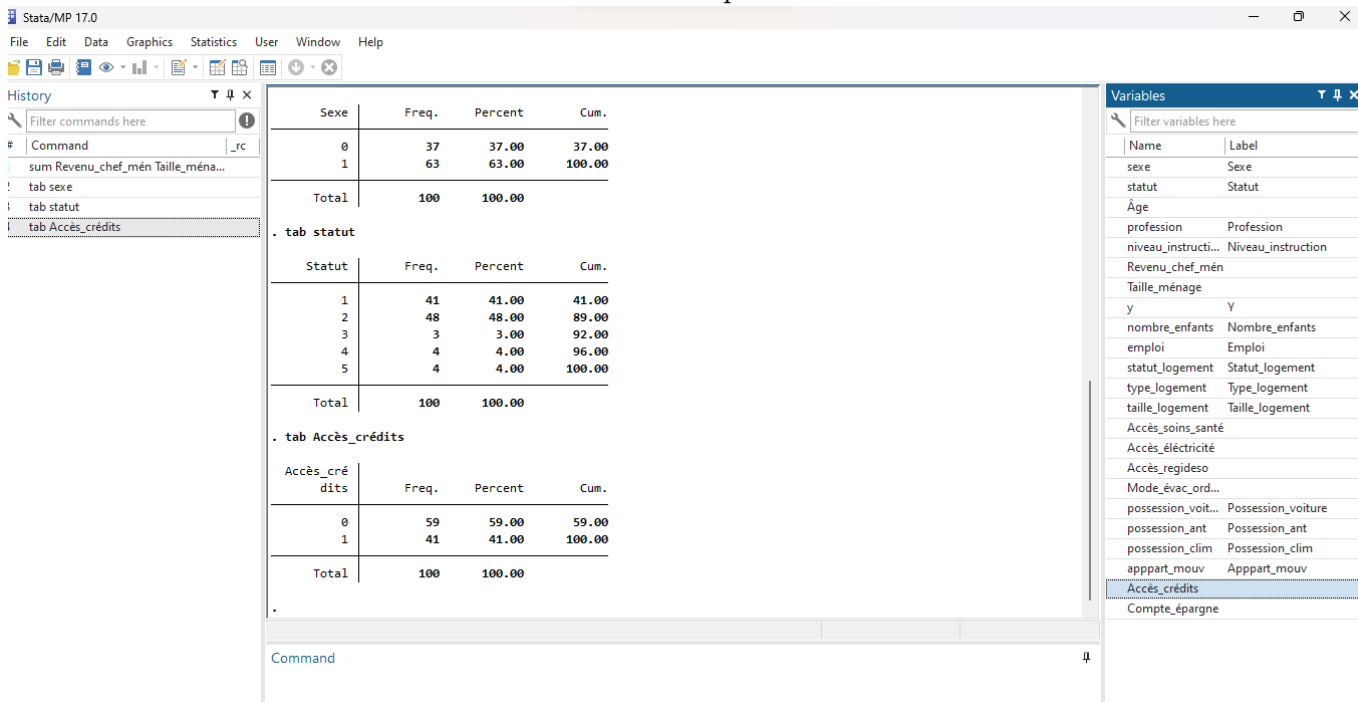
Nous avons 32 chefs des ménages qui ont un revenu compris entre $[0-200[$, 63 chefs des ménages qui ont un revenu compris entre $[200-400[$, 2 chefs des ménages qui ont un revenu compris entre $[400-600[$ et 3 chefs des ménages ayant un revenu compris entre $[600-800[$.

4.2 Tableau des fréquences

C'est l'analyse univariée des données qualitatives, comme les valeurs numériques ou codes d'un questionnaire sont arbitraires, les calculs associés se résument à des comptages. Ce comptage nous fournit des éléments tels que les fréquences absolues, relatives et en pourcentage.

La commande est `tab` suivie de la variable étudiée. Pour générer plusieurs tables de fréquence à la fois, la commande est `tab1` suivie de toutes les variables.

FIGURE 8 – Tableau des fréquences



Commentaire :

Pour le sexe, genre, 1 représente les hommes et 0 les femmes. 1 et 0 sont donc des codes numériques. Nos résultats montrent que nous avons enquêté 100 ménages dont 63 étaient représentés par des hommes et 37 par des femmes, ce qui se traduit par la présence de 63% d'hommes et 37% de femmes.

Pour le statut, 2 correspond aux personnes mariées donc la majorité de notre échantillon soit 48% sont des mariées.

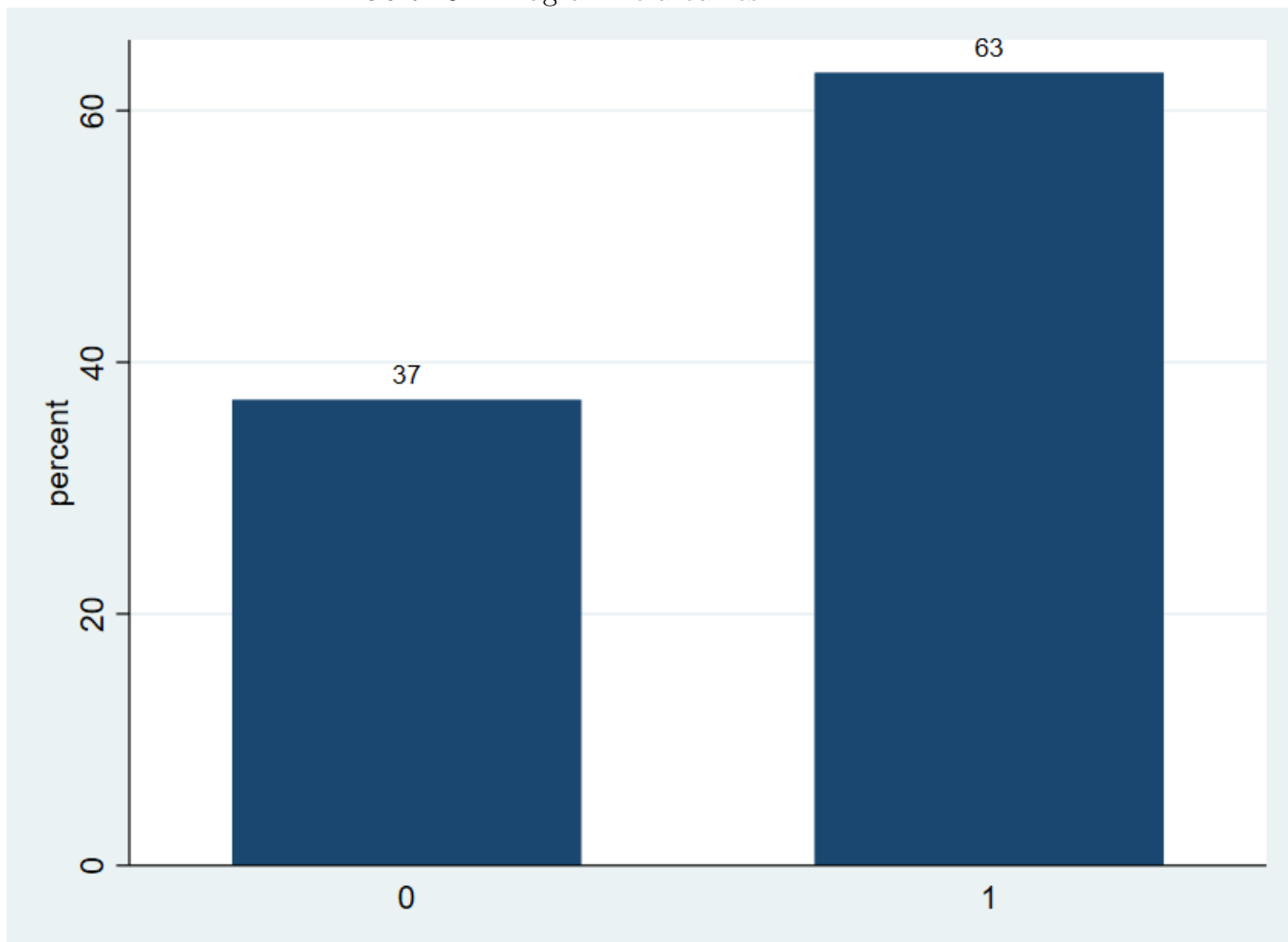
Pour l'accès au crédit, 59% des ménages enquêtés n'ont pas accès aux crédits contre 41% qui en ont.

L'on peut aussi faire une analyse graphique des variables qualitatives. Nous en présentons deux.

Le diagramme en bâton ou en barres et le diagramme circulaire.

Pour le diagramme en barre, la commande est `graph bar, over(sexe) label(total)`

FIGURE 9 – Diagramme à barres

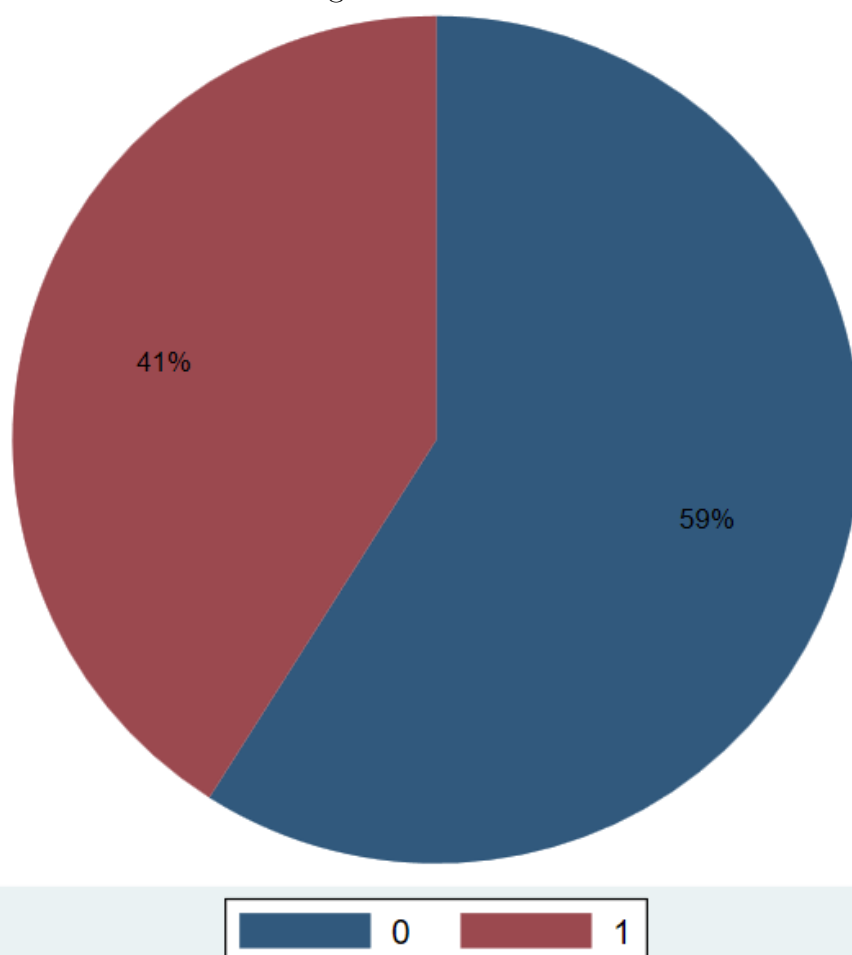


Commentaire :

Le graphique illustre bien ce que l'on a déjà vu (figure7), notre échantillon est composé à 37% des femmes contre 63% d'hommes.

Pour le diagramme circulaire, la commande est `graph pie, over(Accès Crédits) plabel(_all percent)`

FIGURE 10 – Diagramme circulaire



Commentaire :

Comme nous le savons déjà, nous avons dans notre échantillon 59% des personnes qui ont déjà eu accès aux crédits contre 41% qui n'ont jamais eu accès aux crédits.

5 Analyses bivariées

Nous présentons dans cette section 10 tests et un ensemble de 6 tests servant de pré-requis.

L'analyse bivariée permet de tester la relation existante entre deux variables. il peut s'agir de deux variables qualitatives ou quantitatives.

cette analyse bivariée permet aussi de faire des comparaisons des groupes en présence de deux variables, l'une étant qualitative et l'autre quantitative.

Ces différents tests reposent sur des hypothèses et possèdent des critères de décision.

Voici les variables qui seront utilisées.

TABLE 2 – Présentation des variables

Nom de la variable	Type de la variable	Mesure et description
La pauvreté	Qualitative binaire	1=pauvres 0= non pauvres
Profession	Qualitative multimodale	1=employé chez les privés 2= fonctionnaire 3=profession libérale 4=Retraités 5=Chômeurs
Avoir un compte épargne	Qualitative binaire	1=Oui 0=Non
Revenu chef de ménage	Quantitative	Revenu mesurer mensuellement
Taille de ménage	Quantitative	Nombre des personnes composant le ménage
âge	Quantitative	Mesurer en année
Niv instruction	Qualitative multimodale	1= primaire et secondaire 2=universitaire 3= post universitaire

5.1 Analyse de relation

L'on appelle analyse de relation, liaison ou tableau de contingence, l'ensemble des tests effectués pour tester la relation entre deux variables qualitatives.

Ces tests reposent sur les hypothèses suivantes :

- H0 : indépendance entre la variable à expliquer et celle qui explique.
- H1 : dépendance entre la variable qui explique et celle qui est expliquée.

Les critères de décision :

- Si la probabilité associée au test soit la P-value est supérieur au seuil de 0.05 soit 5% : acceptation de l'hypothèse nulle (H0) ; rejet de l'hypothèse une (H1).
- Si la probabilité associée au test de soit P-value est inférieur au seuil de 0.05% : rejet de l'hypothèse nulle(H0) donc il y a acceptation de l'hypothèse une (H1).

Présentation des tests.

1. Test de Chi2 ou Khi2 :

Ce test est utilisé lorsque tous les effectifs théoriques sont supérieurs à 5.

Il permet de tester l'association entre deux variables mais sans en donner l'information sur l'intensité de l'association.

Pour mesurer l'intensité, il faut faire appel au test v de cramer ou gamma.

le V de cramer mesure l'intensité lorsque les deux variables qualitatives sont nominales et le gamma, lorsque les deux variables sont ordinales.

Illustration : On veut savoir s'il existe une relation entre la pauvreté et la profession.

La commande est `tab pauvreté profession,chi2 V`

FIGURE 11 – Chi2

tab pauvreté profession,chi2 V

Y	Profession					Total
	1	2	3	4	5	
0	29	16	13	2	4	64
1	2	15	13	3	3	36
Total	31	31	26	5	7	100

Pearson chi2(4) = 17.4167 Pr = 0.002

Cramér's V = 0.4173

Commentaire :

La probabilité associée au test de chi2 est égale à 0,002 ce qui est inférieure au seuil de 0,05 signifiant donc qu'il existe une relation entre la pauvreté et la profession occupée ce qui nous pousse donc au rejet de H0 et l'acceptation de H1.

Le V de cramer nous fait comprendre que cette relation est positive avec une intensité de 41,73%.

2. Test de fisher exact :

⁶ Le test exact de fisher est utilisé pour déterminer s'il existe ou non une association significative entre deux variables catégorielles. Il est généralement utilisé comme alternative au test de chi2 lorsque l'on a au minimum un effectif théorique inférieur à 5 dans un tableau 2*2.

Illustration : On veut savoir s'il existe une relation entre la pauvreté et le fait d'avoir un compte épargne.

La commande est `tab pauvreté compte épargne,exact`

FIGURE 12 – Fisher exact

tab pauvreté Compte_épargne,exact

Y	Compte_épargne		Total
	0	1	
0	62	2	64
1	34	2	36
Total	96	4	100

Fisher's exact = 0.617
1-sided Fisher's exact = 0.455

Commentaire :

La probabilité associée au test de fisher est supérieure au seuil de 0,05 ce qui traduit l'absence d'une relation entre les deux variables étudiées par conséquent, l'on accepte le H0.

6. Muayila P. (2023), *Introduction aux techniques d'enquêtes*, L2 LMD Faculté d'administration des affaires et sciences économiques, UPC, Kinshasa.

5.2 Analyse de corrélation

Cette analyse permet d'établir la corrélation entre deux variables quantitatives.

Nous présentons ici deux tests, celui de rang de spearman et le test de pearson. Le choix entre ces deux tests est basé sur la normalité de la distribution des données.

Les données peuvent-être normalement distribuées ou pas.

Avant d'aborder les deux tests de corrélation, il est donc primordial d'aborder les différents tests permettant d'étudier la distribution des données.

Énonçons la règle de décision et les hypothèses :

Une population, un échantillon ou encore des données sont dites normalement distribuées lorsque la probabilité associée au test est supérieure au seuil de 0,05 ou de 5%.

Les hypothèses sont :

- H0 : l'échantillon suit une loi normale.
- H1 : l'échantillon ne suit pas une loi normale.

Les tests de normalité

1. Shapiro Wilk : Ce test de normalité est effectué lorsque la taille de l'échantillon est inférieure à 50.

La commande est swilk suivie de la variable ou des variables.

Exemple :

FIGURE 13 – Normalité Shapiro

```
. swilk nombre_enfants
```

```
Shapiro-Wilk W test for normal data
```

Variable	Obs	W	V	z	Prob>z
nombre_enfants	45	0.95236	2.063	1.535	0.06243

Commentaire :

La probabilité associée au test est de 0.06 qui est supérieure au seuil de 0.05, par conséquent la variable nombre d'enfants est normalement distribuée.

2. Skwness et Kurtosis : ce test de normalité est effectué lorsque la taille de l'échantillon est supérieure à 50.

La commande est sktest suivie de la variable ou des variables.

Exemple :

FIGURE 14 – Normalité Skwness et Kurtosis

```
. sktest Revenu_chef_mén Taille_ménage age
```

Skewness and kurtosis tests for normality

Variable	Obs	Pr(skewness)	Pr(kurtosis)	Joint test	
				Adj chi2(2)	Prob>chi2
Revenu_chef_mén	100	0.0000	0.0003	32.21	0.0000
Taille_ménage	100	0.0743	0.0959	5.72	0.0571
age	100	0.2598	0.1150	3.86	0.1452

Commentaire :

La variable revenu chef de ménage n'est pas normalement distribuée car sa probabilité est de 0,0000 qui est inférieure au seuil de 0,05.

La variable taille de ménage et l'âge sont normalement distribuées car leurs probabilités respectives sont supérieures au seuil de 0,05.

Abordons maintenant nos deux tests de corrélation.

1. Test de corrélation de Pearson.

Le test de corrélation de Pearson est le test à utiliser lorsque les deux variables quantitatives étudiées sont normalement distribuées.

Illustration : on veut tester s'il existe une corrélation entre la taille de ménage et l'âge du chef de ménage. Nous savons déjà que ces deux variables sont normalement distribuées comme nous renseigne la figure 11.

La commande est donc `pwcorr taille de ménage âge, sig star(5)`.

FIGURE 15 – Corrélation de Pearson

```
. pwcorr Taille_ménage age, sig star(5)
```

	Taille_ménage	age
Taille_ménage	1.0000	
age	-0.0402	1.0000
	0.6914	

Commentaire :

La corrélation entre la taille de ménage et l'âge est non significative car la probabilité est 0,6914 qui est supérieure au seuil de 0,05.

la valeur au-dessus de la probabilité indique l'intensité et la direction de la corrélation.⁷ Quoique la corrélation n'est pas significative, elle est néanmoins négative et à une force modérée.

7. La section 7.1. permet de mieux comprendre la direction et la force du coefficient de corrélation

2. Test de rang de Spearman

Ce test est utilisé lorsque les deux variables quantitatives ne sont pas normalement distribuées ou au moins l'une d'entre elles ne l'est pas.

Illustration : On veut tester la corrélation entre le revenu du chef de ménage et la taille de ménage. Nous savons que les deux variables sont anormalement distribuées en se référant à la figure 11.

La commande est `spearman Revenu_chef_mén Taille_ménage,stats(rho obs p) star(0.05) matrix`

FIGURE 16 – Corrélation de Spearman

```
. spearman Revenu_chef_mén Taille_ménage,stats(rho obs p) star(0.05) matrix
```

Key
<i>rho</i>
<i>Number of obs</i>
<i>Sig. level</i>

	Revenu	Taille
Revenu	1.0000 100	
Taille	0.1288 100 0.2017	1.0000 100

Commentaire :

La corrélation entre le revenu du chef de ménage et la taille de ménage n'est pas significative car la probabilité associée au test est 0,2017 est supérieure au seuil de 0,05. quoique pas significative, la corrélation est positive à hauteur de 12,88%.

5.3 Analyse de comparaison

Les tests présentés dans cette section sont utilisés lorsque les échantillons ne sont pas appariés.

Cette analyse permet de faire des comparaisons des moyennes, médianes ou des variances lorsque les données concernent au moins deux populations.

Le choix du test de comparaison dépend :

- Du nombre de la population ou des sous-groupes de la variable qualitative, souvent notée K . L'on a des cas où K est égal à 2 (variable qualitative binaire ou bimodale) et k supérieur à 2 (variable qualitative multimodale ou plurimodale).
- De la distribution de la variable quantitative que nous avons déjà abordé à la page 13 lorsque K est supérieure à 2 et de la distribution de la variable quantitative dans les sous-groupes de la variable qualitative lorsque K est égal à 2.
- De l'homogénéité des variances, I.E. l'égalité des variances.

1. Premier cas :
 $K=2$.

- Si la distribution de la variable quantitative dans les sous-groupes de la variable qualitative est normale et que les variances sont homogènes, l'on fait le test de **Student**, à noter que pour le test de Student, le test de normalité à effectuer est le test de Kolmogorov Smirnov, ce dernier permet de voir la distribution de la variable quantitative dans les sous-groupes d'une variable qualitative binaire.

Illustration : on veut savoir si les ménages qui épargnent sont des petites taille comparer aux ménages qui n'épargnent pas.

Première étape : tester la normalité des sous-échantillons (groupes).

La commande est `ksmirnov Taille_ménage,by(Compte_épargne)`

FIGURE 17 – Kolmogorov Smirnov

```
. ksmirnov Taille_ménage,by( Compte_épargne)

Two-sample Kolmogorov-Smirnov test for equality of distribution functions

Smaller group          D          p-value
-----
0                      0.0938      0.935
1                     -0.1875      0.763
Combined K-S          0.1875      0.999

Note: Ties exist in combined dataset;
      there are 12 unique values out of 100 observations.
```

Commentaire :

les probabilités de sous-échantillons sont respectivement 0.935 et 0.763 étant

toutes deux supérieures au seuil de 0.05, les deux sous-échantillons sont normalement distribués.

Deuxième étape : croiser la variable quantitative à la variable qualitative afin de dégager les moyennes des sous-groupes.

La commande est `tabstat Taille_ménage,by(Compte_épargne) stat (n mean min max)`

FIGURE 18 – Moyennes sous-groupes

```
. tabstat Taille_ménage,by( Compte_épargne) stat (n mean min max)
```

```
Summary for variables: Taille_ménage
```

```
Group variable: Compte_épargne
```

Compte_épargne	N	Mean	Min	Max
0	96	4.760417	1	12
1	4	4.25	1	7
Total	100	4.74	1	12

Commentaire :

Nous voyons qu'en moyenne les chefs de ménage qui épargnent ont des ménages qui sont constitués en moyenne de 4.25 personnes ~ 4 tandis que ceux qui n'épargnent pas ont généralement des ménages plus grands 4.7 ~ 5.

Troisième étape : vérifier l'homogénéité des variances.

La commande est `robvar Taille ménage,by(Compte épargne)`

FIGURE 19 – Test de levène

```
. robvar Taille_ménage,by( Compte_épargne)
```

Compte_épa rgne	Summary of Taille_ménage		
	Mean	Std. dev.	Freq.
0	4.7604167	2.9082271	96
1	4.25	2.7537853	4
Total	4.74	2.8906677	100

W0	=	0.06388804	df(1, 98)	Pr > F =	0.80098246
W50	=	0.06206939	df(1, 98)	Pr > F =	0.803776
W10	=	0.05612649	df(1, 98)	Pr > F =	0.8132209

Commentaire :

La probabilité étant supérieure au seuil de 0.05, les variances de deux sous-échantillons sont donc égales.

Quatrième étape : faire le test de Student.

La commande est `ttest Taille_ménage,by(Compte_épargne)`

FIGURE 20 – Student

```
. ttest Taille_ménage,by( Compte_épargne)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
0	96	4.760417	.2968197	2.908227	4.171155	5.349678
1	4	4.25	1.376893	2.753785	-.1318869	8.631887
Combined	100	4.74	.2890668	2.890668	4.166429	5.313571
diff		.5104167	1.481748		-2.430064	3.450897

diff = mean(0) - mean(1) t = 0.3445
H0: diff = 0 Degrees of freedom = 98

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.6344 Pr(|T| > |t|) = 0.7312 Pr(T > t) = 0.3656

Commentaire :

Le t-test est non significative, car la probabilité est supérieure au seuil ($p=0.6344$). nous ne pouvons donc pas parler d'une différence significative entre les ménages qui épargnent et celles qui n'épargnent pas.

Note : Il existe des cas où la condition de normalité est satisfaite mais pas celle de l'homogénéité, dans ce cas, il faut faire un test post hoc, il en existe plusieurs.

Nous optons pour le test⁸ post hoc de Welch. La commande est `ttest variable quantitative,by(variable qualitative) unequal`.

8. Voir Appendice 19

- Si la distribution de la population est non normale ou anormale et que les variances sont non-homogènes, l'ont fait le test de **Mann Whitney**.

Illustration : on veut savoir si les ménages des non pauvres sont constitués de moins des personnes que les ménages des pauvres.

Première étape : vérifier la normalité par le test de kolmogorov Smirnov.
La commande est `ksmirnov Taille ménage,by(pauvreté)`

FIGURE 21 – Kolmogorov Smirnov

Two-sample Kolmogorov-Smirnov test for equality of distribution functions

Smaller group	D	p-value
0	0.6510	0.000
1	0.0000	1.000
Combined K-S	0.6510	0.000

Note: Ties exist in combined dataset;
there are 12 unique values out of 100 observations.

Commentaire :

En voyant les probabilités, le premier sous-échantillon n'est pas normalement distribué.

Deuxième étape : comparer les moyennes en croisant les deux variables, hormis la commande utilisée au niveau du Student, l'on peut procéder par une autre commande, `mean Taille ménage,over(pauvreté)`

FIGURE 22 – Moyennes sous groupes

`. mean Taille_ménage,over(pauvreté)`

Mean estimation		Number of obs = 100		
	Mean	Std. err.	[95% conf. interval]	
<code>c.Taille_ménage@pauvreté</code>				
0	3.328125	.2782231	2.77607	3.88018
1	7.25	.3571825	6.541272	7.958728

Commentaire :

L'on voit qu'en moyenne, les ménages non pauvres sont composés de 3 personnes et les ménages pauvres de 7 personnes.

Troisième étape : le test de Mann Whitney.

La commande est `ranksum Taille ménage,by(pauvreté)`

FIGURE 23 – Mann Whitney

```
. ranksum Taille_ménage,by( pauvreté)
```

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test
```

pauvreté	Obs	Rank sum	Expected
0	64	2339	3232
1	36	2711	1818
Combined	100	5050	5050

```
Unadjusted variance 19392.00
```

```
Adjustment for ties -261.47
```

```
Adjusted variance 19130.53
```

```
H0: Taille~e(pauvreté==0) = Taille~e(pauvreté==1)
```

```
z = -6.456
```

```
Prob > |z| = 0.0000
```

```
Exact prob = 0.0000
```

Commentaire :

La probabilité étant inférieure au seuil de 0.05, elle vient confirmer la tendance qui s'était dégagée au niveau des moyennes des sous-groupes donc il existe une différence significative, I.E. les ménages des non pauvres sont constitués de moins des personnes que les ménages des pauvres.

2. Deuxième cas :

K est supérieur à 2 et les échantillons sont indépendants.

- Si la distribution de la population est normale et que les variances sont homogènes, l'on fait le test **d'Anova**.

Illustration : l'hypothèse que l'on teste est que la taille de ménage serait différente selon le niveau d'instruction du chef de ménage.

Première étape : la vérification de la normalité (voir la figure 14).

Deuxième étape : croiser la variable quantitative à la variable qualitative afin de dégager les moyennes des sous-groupes.

la commande est `tabstat Taille_ménage,by(niveau_instruction) stats(n mean max min)`

FIGURE 24 – Moyennes sous-groupes

```
. tabstat Taille_ménage,by( niveau_instruction) stats(n mean max min)
```

Summary for variables: Taille_ménage
Group variable: niveau_instruction

niveau_instruction	N	Mean	Max	Min
1	19	5.789474	12	1
2	72	4.541667	11	1
3	9	4.111111	7	1
Total	100	4.74	12	1

Commentaire :

L'on voit que les chefs de ménage avec un haut niveau d'instruction ont en moyenne moins d'enfants (4.11 pour les post-universitaire contre 4.54 pour les universitaire et 5.78 pour ceux qui se sont limités au niveau primaire et secondaire.)

Troisième étape : Faire le test d'Anova et vérifier l'homogénéité des variances.

La commande est `. oneway Taille_ménage niveau_instruction`

FIGURE 25 – Anova

```
. oneway Taille_ménage niveau_instruction
```

Source	Analysis of variance			F	Prob > F
	SS	df	MS		
Between groups	38.8021053	3	12.9340351	1.57	0.2005
Within groups	788.437895	96	8.21289474		
Total	827.24	99	8.3559596		

```
Bartlett's equal-variances test: chi2(2) = 5.0044 Prob>chi2 = 0.082
```

```
note: Bartlett's test performed on cells with positive variance:  
1 single-observation cells not used
```

Commentaire :

Le test est non significatif car la probabilité est supérieure au seuil de 0.05, par conséquent nos variances sont homoscedastiques et l'on peut procéder au test d'Anova. Contrairement à nos attentes, le test montre que la différence n'est pas significative, la probabilité (0.2005) étant supérieure au seuil de 0.05 par conséquent la taille de ménage ne diffèrent pas selon le niveau d'instruction.

Note : Il existe des cas où la condition de normalité est satisfaite mais pas celle de l'homogénéité, dans ce cas, il faut faire un test post hoc, il en existe plusieurs.

Nous optons pour le test⁹ post hoc de Dunnett's.

La commande est `pwmean Revenu chef_mén, over(profession) mcompare(dunnett)effect`.

9. Voir Appendice 20

- Si la distribution de la population est non normale ou anormale et que les variances sont non-homogènes, l'ont fait le test de **Kruskal Wallis**.

Illustration : L'on veut savoir si le revenu mensuel des chefs de ménage diffèrent selon la profession de ce dernier.

Première étape : la vérification de la normalité (voir la figure 14).

Deuxième étape : comparer les moyennes en croisant les deux variables.

la commande est `mean Revenu_chef_mén,over(profession)`

FIGURE 26 – Moyennes sous-groupes

```
. mean Revenu_chef_mén,over( profession)
```

Mean estimation		Number of obs = 100			
		Mean	Std. err.	[95% conf. interval]	
c.Revenu_chef_mén@profession					
	1	256.6129	20.61864	215.701	297.5248
	2	263.0645	25.59053	212.2874	313.8417
	3	266.7308	21.10585	224.8522	308.6094
	4	161	16	129.2525	192.7475
	5	156.4286	11.42857	133.7518	179.1053

Commentaire :

Le tableau indique des différences des moyennes, l'on peut clairement voir que les chômeurs et les retraités sont ceux qui ont le revenu le plus bas.

Troisième étape : le test de Kruskal Wallis proprement dit.

La commande est `kwallis Revenu_chef_mén,by(profession)`

FIGURE 27 – Kruskal Wallis

```
. kwallis Revenu_chef_mén,by( profession)  
  
Kruskal-Wallis equality-of-populations rank test
```

profes~n	Obs	Rank sum
1	31	1642.00
2	31	1610.50
3	26	1524.50
4	5	120.00
5	7	153.00

```
chi2(4) = 13.341  
Prob = 0.0097
```

```
chi2(4) with ties = 15.164  
Prob = 0.0044
```

Commentaire :

La probabilité du test est inférieure au seuil, par conséquent, le revenu diffère selon les types de profession.

6 Modèles de régression

La régression permet d'étudier l'impact d'une variable indépendante sur la variable dépendante ou les impacts de plusieurs variables indépendantes sur la variable dépendante.

Dans cette section, nous illustrons la régression linéaire simple et multiple (MCO), la régression logistique simple et multiple puis la régression probit simple et multiple.

Le choix du modèle se base sur la nature de la variable dépendante.

Nous allons utilisés ces variables :

Questions	Modalités
Sexe du chef de ménage	0=Femme ; 1=Homme
Statut matrimoniale	1=Célibataire
	2=mariée monogame
	3=Mariée polygame
	4=Divorcé
	5=Veuf
Age	En année
Profession du chef de ménage	1=employé chez les privés
	2=Fonctionnaire
	3=Profession libérale et
	Indépendant
	4=Retraité et chômage
	5=Autres
Niveau d'instruction	1=Primaire et secondaire
	2=Universitaire
	3=Post-universitaire
Revenu du chef de ménage	Montant en usd
Taille du ménage	Nombre de personne au ménage
Emploi	Nombre de personnes avec un emploi
Statut logement	1=Propriétaire
	2=Locataire
	3=Autres
Type logement	1=Maison
	2=Appartement
	3=Autre
Taille logement	Nombre de pièce
Accès aux soins de santé	1=Hôpital
	2=Traditionnelle
	3=Automédication
Accès à l'électricité de la Snel	1=jamais ; 2=régulier ; 3=Irrégulier
Accès eau de la Regideso	1=jamais ; 2=régulier ; 3=Irrégulier
Mode d'évacuation des ordures	1=Dépotoir/incinération/enlèvement
	2=Enfouissement
	3=Services spéciales
	4=Rue et autres
Possession d'une voiture ou moto	0=Non ; 1=Oui
Possession d'une antenne avec décodeur	0=Non ; 1=Oui
Possession d'une climatisation et machine à lavage	0=Non ; 1=Oui
Appartenance à un mouvement associative ou culturel	0=Non ; 1=Oui
Accès au crédit	0=Non ; 1=Oui
Possession compte épargne	0=Non ; 1=Oui

6.1 Régression Linéaire Multiple et simple

Cette régression a pour but de modéliser les relations entre une variable à expliquer quantitative et une ou plusieurs variables explicatives qui peuvent-être quantitatives ou qualitatives sous la forme d'une fonction linéaire.

Ce modèle a des hypothèses qui doivent-être respectées pour l'appliquer :

- La condition de linéarité des relations entre la variable dépendante et ses variables indépendantes.(Cette hypothèse n'est pas toujours évaluée car elle est minimisable)
- Dans le cas de la régression multiple, il faut s'assurer que les variables explicatives ne présentent pas une trop forte multi-colinéarité.
- Les erreurs doivent-être homoscédastiques.
- Les résidus doivent-être indépendants.
- Les résidus doivent-être normalement distribués.

Illustration : On veut savoir quelles sont les variables qui influent sur la taille de ménage.

Première étape : Estimer le modèle.

La commande est `reg variable dépendante variables indépendantes`

FIGURE 28 – MCO

Source	SS	df	MS	Number of obs	=	100
Model	611.694206	21	29.1282955	F(21, 78)	=	10.54
Residual	215.545794	78	2.76340762	Prob > F	=	0.0000
				R-squared	=	0.7394
				Adj R-squared	=	0.6693
Total	827.24	99	8.3559596	Root MSE	=	1.6624

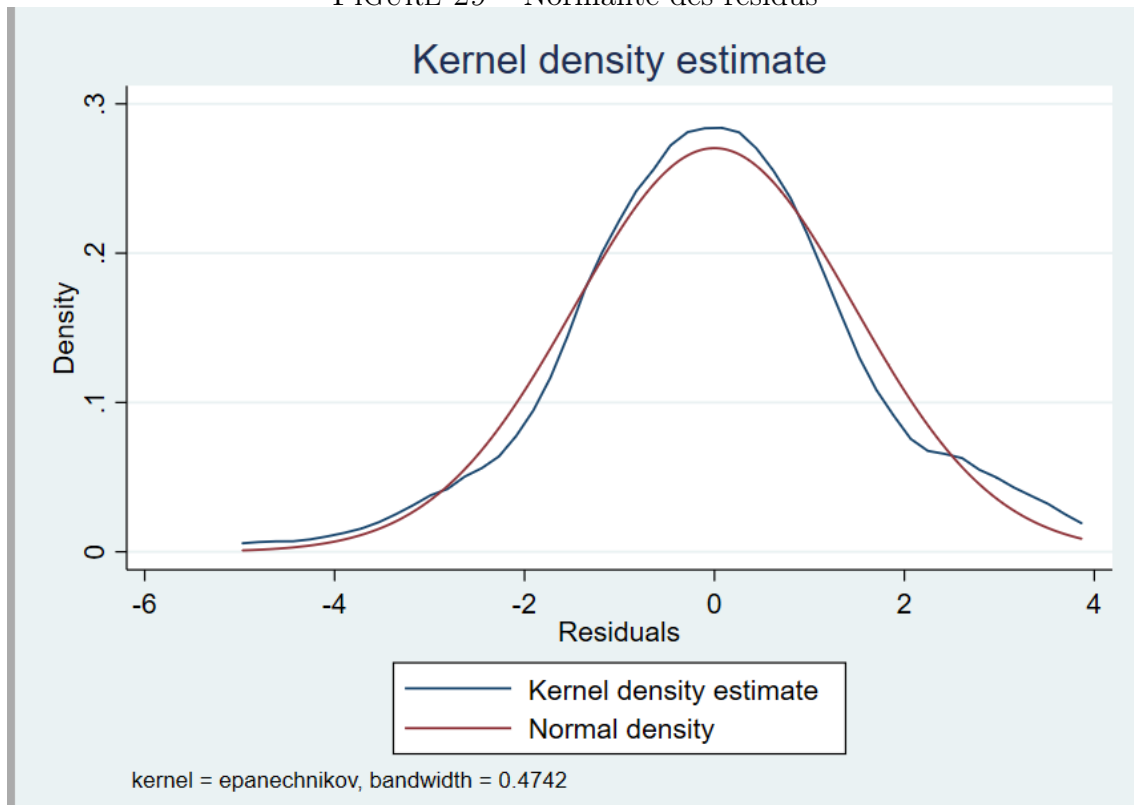
Taille_ménage	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
sexe	-.0956348	.3743989	-0.26	0.799	-.8410058	.6497362
statut	.2252796	.1918457	1.17	0.244	-.1566557	.607215
age	.0150857	.0152989	0.99	0.327	-.0153721	.0455435
profession	.1705788	.1756832	0.97	0.335	-.1791797	.5203372
niveau_instruction	-.4491257	.3446129	-1.30	0.196	-1.135197	.2369458
Revenu_chef_mén	.0061523	.0021369	2.88	0.005	.0018981	.0104064
pauvreté	3.438082	.5081558	6.77	0.000	2.426421	4.449742
emploi	.4940019	.2051159	2.41	0.018	.0856474	.9023564
statut_logement	.3402283	.3348763	1.02	0.313	-.3264593	1.006916
type_logement	-.1507461	.3466042	-0.43	0.665	-.8407821	.53929
taille_logement	.6255973	.1523351	4.11	0.000	.3223214	.9288732
Accès_soins_santé	.3339547	.2974022	1.12	0.265	-.2581277	.926037
Accès_électricité	-.3227283	.3357464	-0.96	0.339	-.991148	.3456914
Accès_regideso	-.0861266	.3406887	-0.25	0.801	-.7643857	.5921325
Mode_évac_ordures	-.2505644	.1735516	-1.44	0.153	-.5960791	.0949503
possession_voiture	.3479614	.4368498	0.80	0.428	-.5217397	1.217662
possession_ant	1.229044	.5179375	2.37	0.020	.1979096	2.260178
possession_clim	-.8001819	.4317686	-1.85	0.068	-1.659767	.0594034
appart_mouv	-.4430568	.4062944	-1.09	0.279	-1.251927	.3658133
Accès_crédits	-.3066964	.4235735	-0.72	0.471	-1.149966	.5365736
Compte_épargne	-1.09677	.9115829	-1.20	0.233	-2.911592	.7180522
_cons	-1.254841	1.901538	-0.66	0.511	-5.040513	2.53083

Commentaire :

Les résultats montrent que le modèle est globalement significatif car la probabilité associée au test est de 0.0000 qui est inférieure à 1%. La variance expliquée est de 73.94%. les variables retenues expliquent à 66.93% la variable dépendante mais ces 66.93% sont expliqués que par 6 variables. les variables ayant des probabilités inférieures à 10%.

Deuxième étape : Vérifier la normalité des résidus
 La commande est predict residuols,res
 kdensity residuols,normal

FIGURE 29 – Normalité des résidus



Commentaire :
 le graphique montre que les résidus sont normalement distribués.

L'on peut aussi voir cette distribution par le test de Skwness et Kurtosis.

FIGURE 30 – Normalité des résidus

`. sktest residuols`

Skewness and kurtosis tests for normality

Variable	Obs	Pr(skewness)	Pr(kurtosis)	Joint test	
				Adj chi2(2)	Prob>chi2
residuols	100	0.7966	0.2428	1.46	0.4810

Commentaire :

la probabilité étant supérieure au seuil de 0.05, les résidus sont normalement distribués.

Troisième étape : il faut tester l'homoscédasticité des erreurs avec le test de Breush-Pagan.

La commande est `hettest`

FIGURE 31 – Breush-Pagan

```
. hettest
```

```
Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
```

```
Assumption: Normal error terms
```

```
Variable: Fitted values of Taille_ménage
```

```
H0: Constant variance
```

```
chi2(1) = 14.83
```

```
Prob > chi2 = 0.0001
```

Commentaire :

La probabilité du test étant inférieure au seuil de 0.05, nos erreurs sont hétérocédastiques, il faudrait donc corriger cela.

Quatrième étape : Correction de l'hétéroscédasticité avec le test de White.

La commande est `reg variable dépendante variables indépendantes,robust`

FIGURE 32 – Test de White

Linear regression		Number of obs	=	100
		F(21, 78)	=	16.26
		Prob > F	=	0.0000
		R-squared	=	0.7394
		Root MSE	=	1.6624

Taille_ménage	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
sexe	-.0956348	.3858299	-0.25	0.805	-.8637631	.6724936
statut	.2252796	.1777709	1.27	0.209	-.128635	.5791943
age	.0150857	.0166544	0.91	0.368	-.0180707	.048242
profession	.1705788	.1608929	1.06	0.292	-.1497344	.490892
niveau_instruction	-.4491257	.3218238	-1.40	0.167	-1.089828	.1915762
Revenu_chef_mén	.0061523	.0031611	1.95	0.055	-.000141	.0124456
pauvreté	3.438082	.466006	7.38	0.000	2.510335	4.365828
emploi	.4940019	.1754438	2.82	0.006	.1447202	.8432836
statut_logement	.3402283	.2786003	1.22	0.226	-.2144224	.894879
type_logement	-.1507461	.3350113	-0.45	0.654	-.8177023	.5162102
taille_logement	.6255973	.1885692	3.32	0.001	.2501849	1.00101
Accès_soins_santé	.3339547	.3443137	0.97	0.335	-.3515212	1.019431
Accès_électricité	-.3227283	.3392727	-0.95	0.344	-.9981685	.3527118
Accès_regideso	-.0861266	.2815028	-0.31	0.760	-.6465556	.4743024
Mode_évac_ordures	-.2505644	.1617491	-1.55	0.125	-.5725821	.0714533
possession_voiture	.3479614	.3462029	1.01	0.318	-.3412757	1.037199
possession_ant	1.229044	.4717921	2.61	0.011	.2897779	2.16831
possession_clim	-.8001819	.3565628	-2.24	0.028	-1.510044	-.0903199
appart_mouv	-.4430568	.3884711	-1.14	0.258	-1.216443	.3303297
Accès_crédits	-.3066964	.4165311	-0.74	0.464	-1.135946	.5225534
Compte_épargne	-1.09677	.6452694	-1.70	0.093	-2.381403	.1878628
_cons	-1.254841	2.062239	-0.61	0.545	-5.360443	2.85076

commentaire :

Après correction du problème d'hétéroscédasticité, les résultats montrent que le modèle est globalement significatif car la probabilité associée au test est de 0.0000 qui est inférieure à 1%. La variance expliquée est de 73.94%. la variable dépendante est expliquée que par 7 variables. les variables ayant des probabilités inférieures à 10%.

Cinquième étape : Vérifier si le modèle est bien spécifié.

La commande est linktest

FIGURE 33 – Spécification du modèle

```
. linktest
```

Source	SS	df	MS	Number of obs	=	100
Model	612.913739	2	306.45687	F(2, 97)	=	138.70
Residual	214.326261	97	2.20954908	Prob > F	=	0.0000
Total	827.24	99	8.3559596	R-squared	=	0.7409
				Adj R-squared	=	0.7356
				Root MSE	=	1.4865

Taille_ménve	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
_hat	1.159174	.2225227	5.21	0.000	.7175276	1.60082
_hatsq	-.0157286	.0211711	-0.74	0.459	-.0577474	.0262903
_cons	-.304889	.5212222	-0.58	0.560	-1.339371	.7295928

Commentaire :

On s'intéresse à la p-value du _hatsq qui a une valeur supérieure au seuil de 5%, ce qui veut dire que le modèle est bien spécifié.

Sixième étape : Détection de la multicolinéarité avec le test de vif (Variance inflation factor).

La commande est `vif` ou `estat vif`

FIGURE 34 – Test de multicolinéarité

`. vif`

Variable	VIF	1/VIF
Revenu_chef	2.35	0.426387
pauvreté	2.15	0.464482
taille_log	1.91	0.524681
possession	1.82	0.549008
possession	1.72	0.580144
Accès_élec	1.66	0.601582
statut_log	1.63	0.612376
possession	1.61	0.620932
Accès_créd	1.57	0.636725
type_logem	1.57	0.637015
profession	1.49	0.671970
emploi	1.48	0.677203
Accès_soin	1.47	0.678611
Mode_évac	1.43	0.700406
appart_mouv	1.41	0.710539
niveau_ins	1.37	0.731967
Accès_regio	1.26	0.793612
age	1.26	0.796624
statut	1.24	0.809432
sexe	1.18	0.845733
Compte_épar	1.15	0.866007
Mean VIF	1.56	

Commentaire :

Etant donné que les vif sont tous inférieurs à 10, cela renseigne qu'il n'existe aucune multicolinéarité.

Pour répondre à la question posée au niveau de l'illustration, les variables qui influent sur la taille des ménages sont le revenu du chef des ménages, la pauvreté, le nombre des personnes ayant un emploi, la taille du logement, la possession de la climatisation, la possession antenne et le compte épargne.

Pour une régression simple, il faudrait refaire les mêmes étapes.
 Illustration : on veut savoir si la taille de ménage est influencé par la taille du logement et si cette influence est positive ou négative.
 L'on va aller plus directement étant donné que nous avons déjà vu les étapes.

FIGURE 35 – MCO simple

```
. reg Taille_ménage taille_logement,robust
```

```
Linear regression                Number of obs   =       100
                                F(1, 98)        =       35.80
                                Prob > F             =       0.0000
                                R-squared            =       0.2830
                                Root MSE         =       2.4601
```

Taille_ménage	Robust				
	Coefficient	std. err.	t	P> t	[95% conf. interval]
taille_logement	1.015686	.1697508	5.98	0.000	.6788206 1.352551
_cons	1.164787	.6089873	1.91	0.059	-.0437288 2.373302

```
. predict residuols2,res
```

```
. kdensity residuols2,normal
```

```
. sktest residuols2
```

```
Skewness and kurtosis tests for normality
```

Variable	Obs	Pr(skewness)	Pr(kurtosis)	Joint test	
				Adj chi2(2)	Prob>chi2
residuols2	100	0.0212	0.9907	5.22	0.0736

Commentaire :
 pour le commentaire du modèle, on peut se référer à l'illustration précédente.

La seule nouveauté est d'insister sur un fait, la commande reg accompagné du robust corrige directement les éventuels problèmes d'hétéroscédasticité, l'on ne peut pas faire le test de Breusch-Pagan après avoir déjà corrigé l'hétéroscédasticité. à la place de robust, l'on peut utiliser vce(r). La commande serait donc reg variable dépendante variable indépendante, vce(r)

FIGURE 36 – Linktest

. linktest

Source	SS	df	MS	Number of obs	=	100
Model	257.376904	2	128.688452	F(2, 97)	=	21.90
Residual	569.863096	97	5.87487728	Prob > F	=	0.0000
				R-squared	=	0.3111
				Adj R-squared	=	0.2969
Total	827.24	99	8.3559596	Root MSE	=	2.4238

faillie_ménve	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
_hat	2.465479	.7536376	3.27	0.001	.9697173	3.961241
_hatsq	-.1387441	.0697567	-1.99	0.050	-.2771918	-.0002964
_cons	-3.504275	1.930447	-1.82	0.073	-7.335678	.3271274

Commentaire :

Le modèle est acceptable, car le hatsq est égal à la valeur du seuil.

Il ne peut pas avoir des problèmes de multicollinéarité car l'on a qu'une seule variable indépendante.

6.2 Régression Multiple et simple à variables discontinues

Ces modèles de régression sont générés lorsque la variable dépendante est une qualitative.

L'on fait généralement le choix entre le modèle logit et le probit. le choix est théoriquement basé sur la distribution des erreurs qui peuvent suivre la loi logistique ou la loi normale. Dans la pratique, les deux modèles donnent des résultats similaires sauf pour les grands échantillons et il n'existe pas des tests économétriques qui peuvent aider à faire le choix entre les deux.

Ces modèles ont des variantes. dans ce document, nous exploiterons :

- Le logit/probit : lorsque la variable dépendante est binaire et nominale.
- Ologit/Oprobit : lorsque la variable dépendante est ordinale.
- Mlogit/Mprobit : lorsque la variable dépendante est multimodale et nominale.

6.2.1 Logit/Probit

Illustration : On veut savoir quelles sont les déterminants de la pauvreté.
Première étape : Estimer le modèle logit

La commande logit est logit variable dépendante variables indépendantes,ro¹⁰l'option ro permet de corriger d'éventuels problèmes d'hétéroscédasticité.

FIGURE 37 – Logit

Logistic regression		Number of obs = 100				
Log pseudolikelihood = -18.154567		Wald chi2(21) = 44.83				
		Prob > chi2 = 0.0018				
		Pseudo R2 = 0.7222				
pauvreté	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
sexe	3.13607	1.493156	2.10	0.036	.2095373	6.062602
statut	.250232	.5534159	0.45	0.651	-.8344432	1.334907
profession	.9792319	.631625	1.55	0.121	-.2587303	2.217194
niveau_instruction	.6792286	.8358627	0.81	0.416	-.9590322	2.317489
Revenu_chef_mén	-.0531765	.0161451	-3.29	0.001	-.0848202	-.0215327
nombre_enfants	1.863816	.6871478	2.71	0.007	.5170315	3.210601
emploi	2.739096	1.123601	2.44	0.015	.5368776	4.941314
statut_logement	-1.084365	1.420846	-0.76	0.445	-3.869173	1.700442
type_logement	.622639	1.03503	0.60	0.547	-1.405982	2.65126
taille_logement	2.011259	.5950068	3.38	0.001	.8450667	3.177451
Accès_soins_santé	3.18628	1.205953	2.64	0.008	.822656	5.549904
Accès_électricité	-.9128737	1.414513	-0.65	0.519	-3.685269	1.859521
Accès_regideso	-.0510668	1.009638	-0.05	0.960	-2.02992	1.927787
Mode_évac_ordures	-1.456473	.4643778	-3.14	0.002	-2.366637	-.5463093
possession_voiture	-4.060645	1.886325	-2.15	0.031	-7.757774	-.3635157
possession_ant	3.8145	1.958152	1.95	0.051	-.0234064	7.652407
possession_clim	2.015522	1.678378	1.20	0.230	-1.274038	5.305082
appart_mouv	3.365102	1.755158	1.92	0.055	-.0749451	6.805149
Accès_crédits	-6.230359	2.539274	-2.45	0.014	-11.20725	-1.253473
Compte_épargne	1.660679	1.809762	0.92	0.359	-1.886388	5.207746
age	.0736026	.0508958	1.45	0.148	-.0261512	.1733565
_cons	-11.11448	6.523008	-1.70	0.088	-23.89934	1.670382

Commentaire :

Le modèle est globalement significatif car sa probabilité est inférieure au seuil

10. Muayila sondage

de 0.05, ce qui veut dire qu'au moins une variable indépendante retenue explique la variable dépendante.

la colonne $P > |Z|$ permet d'identifier les variables étant statistiquement et significativement associées à la variable dépendante.

En effet, les variables peuvent-etre statistiquement et significativement associées à la variable dépendante lorsqu'elles ont des p-values inférieures à 1, 5 et 10%.

Au seuil de 1%, sont statistiquement et significativement associées à la pauvreté les variables revenu chef ménage, les nombres d'enfants scolarisés, les nombres des personnes ayant un emploi, la taille du logement, l'accès aux soins de santé, le mode d'évacuation des ordures et l'accès au crédit.

Au seuil de 5%, les variables sexe et possession voiture.

Au seuil de 10%, les variables possession d'une antenne avec décodeur et appartenir à un mouvement associative ou culturel.

Nous avons donc 11 variables qui déterminent la pauvreté.

L'on peut interpréter le signe du coefficient.

Les variables revenu chef de ménage, mode d'évacuation des ordures, possession d'une voiture et l'accès au crédit sont les seules variables qui diminuent la la probabilité d'être pauvre. et les 7 autres variables statistiquement et significativement associées à la variable pauvreté augmentent les chances ou la probabilité d'être pauvre.

Deuxième étape : Estimer les effets marginaux.

La commande est `mfx compute`

FIGURE 38 – Effets marginaux

`. mfx compute`

Marginal effects after logit

$y = \text{Pr}(\text{pauvreté})$ (predict)

= .05431888

variable	dy/dx	Std. err.	z	P> z	[95% C.I.]	X
sexe*	.1469968	.11697	1.26	0.209	-.082256 .376249	.63
statut	.012854	.0319	0.40	0.687	-.049677 .075385	1.82
profes~n	.0503015	.04052	1.24	0.215	-.029125 .129728	2.26
niveau~n	.0348908	.0439	0.79	0.427	-.051158 .12094	1.89
Revenu~n	-.0027316	.00187	-1.46	0.145	-.006402 .000939	249.45
nombre~s	.0957412	.08342	1.15	0.251	-.067762 .259244	1.73
emploi	.1407028	.10168	1.38	0.166	-.05859 .339996	1.49
statut~t	-.0557021	.07879	-0.71	0.480	-.21013 .098726	1.76
type_l~t	.0319839	.0468	0.68	0.494	-.059739 .123707	1.33
taille~t	.103315	.07297	1.42	0.157	-.03971 .24634	3.52
Accès~nté	.1636739	.11085	1.48	0.140	-.053592 .38094	1.14
Accès~ité	-.0468928	.072	-0.65	0.515	-.188018 .094233	2.25
Accès_~o	-.0026232	.05225	-0.05	0.960	-.105032 .099786	2.2
Mode_é~s	-.0748166	.07192	-1.04	0.298	-.215774 .066141	1.99
posses~e*	-.31369	.1644	-1.91	0.056	-.635903 .008523	.52
posses~t*	.1106351	.07076	1.56	0.118	-.028048 .249319	.79
posses~m	.103534	.08783	1.18	0.238	-.068602 .27567	.5
apppar~v*	.3005955	.16641	1.81	0.071	-.025561 .626752	.38
Accès_~s*	-.4234728	.15552	-2.72	0.006	-.728294 -.118652	.41
Compte~e*	.1694895	.25049	0.68	0.499	-.321471 .66045	.04
age	.0037808	.00351	1.08	0.281	-.003096 .010658	41.05

Commentaire :

Les effets marginaux mesurent comment un petit changement dans une variable indépendante affecte la probabilité de l'événement.

Dans notre cas, les effets marginaux permettent de voir comment un petit changement d'une variable indépendante peut affecter la probabilité d'être pauvre.

¹¹Par exemple : toutes les autres variables restants constantes, l'augmentation d'une unité du revenu diminue la probabilité d'être pauvre d'environ 0.27% ce qui reste minime comme impact.

La variable accès au crédit diminue la probabilité d'être pauvre d'environ 42.34%. i.e. avoir un accès au crédit élevé diminue la probabilité d'être pauvre. La variable nombre d'enfants augmentent la probabilité d'être pauvre d'environ 9.5%. i.e. avoir un nombre plus élevé d'enfant augmente la probabilité d'être pauvre.

Troisième étape : voir le tableau de prédiction du modèle.

La commande est lstat

FIGURE 39 – Prédiction du modèle

```

. lstat
Logistic model for pauvreté

```

Classified	True		Total
	D	~D	
+	31	3	34
-	5	61	66
Total	36	64	100

```

Classified + if predicted Pr(D) >= .5
True D defined as pauvreté != 0

```

Sensitivity	Pr(+ D)	86.11%
Specificity	Pr(- ~D)	95.31%
Positive predictive value	Pr(D +)	91.18%
Negative predictive value	Pr(~D -)	92.42%
False + rate for true ~D	Pr(+ ~D)	4.69%
False - rate for true D	Pr(- D)	13.89%
False + rate for classified +	Pr(~D +)	8.82%
False - rate for classified -	Pr(D -)	7.58%
Correctly classified		92.00%

Commentaire :

Le tableau indique que pour les personnes pauvres, 31 cas sur 36 ont été bien prédit, donc 5 cas mal prédit cela veut dire qu'il y'a 5 ménages classés comme pauvres mais qui ne sont pas pauvres. et pour les personnes non pauvres, 61 cas sur 64 ont été bien prédit, donc 3 cas mal prédit, ce qui veut dire qu'il

11. le lecteur peut s'exercer en interprétant les autres variables

y'a 3 ménages classés comme non-pauvres mais qui sont en réalité pauvres. le modèle a correctement classé 92% des observations.

¹²On peut faire une régression logistique simple : on veut savoir si le revenu est un déterminant de la pauvreté.

FIGURE 40 – Logit et effets marginaux

. logit pauvreté Revenu_chef_mén,ro

```
Iteration 0:  log pseudolikelihood = -65.341819
Iteration 1:  log pseudolikelihood = -57.404463
Iteration 2:  log pseudolikelihood = -56.88125
Iteration 3:  log pseudolikelihood = -56.878529
Iteration 4:  log pseudolikelihood = -56.878529
```

Logistic regression

```
Number of obs = 100
Wald chi2(1) = 11.41
Prob > chi2 = 0.0007
Pseudo R2 = 0.1295
```

Log pseudolikelihood = -56.878529

pauvreté	Robust		z	P> z	[95% conf. interval]	
	Coefficient	std. err.				
Revenu_chef_mén	-.0106974	.0031665	-3.38	0.001	-.0169036	-.0044911
_cons	1.84234	.7010422	2.63	0.009	.4683224	3.216357

. mfx compute

Marginal effects after logit

```
y = Pr(pauvreté) (predict)
= .30446642
```

variable	dy/dx	Std. err.	z	P> z	[95% C.I.]	X
Revenu~n	-.0022653	.00059	-3.82	0.000	-.003429 - .001102	249.45

12. Partant de l'illustration précédente, le lecteur peut s'exercer en interprétant les résultats de cette illustration

FIGURE 41 – Prédiction du modèle

`. lstat`

Logistic model for pauvreté

Classified	True		Total
	D	~D	
+	18	14	32
-	18	50	68
Total	36	64	100

Classified + if predicted $\Pr(D) \geq .5$

True D defined as pauvreté != 0

Sensitivity	$\Pr(+ D)$	50.00%
Specificity	$\Pr(- \sim D)$	78.13%
Positive predictive value	$\Pr(D +)$	56.25%
Negative predictive value	$\Pr(\sim D -)$	73.53%

False + rate for true ~D	$\Pr(+ \sim D)$	21.88%
False - rate for true D	$\Pr(- D)$	50.00%
False + rate for classified +	$\Pr(\sim D +)$	43.75%
False - rate for classified -	$\Pr(D -)$	26.47%

Correctly classified	68.00%
----------------------	---------------

On va illustrer le modèle probit avec la même illustration du modèle logit.
Première étape : Estimer le modèle probit

La commande est `probit variable dépendante variables indépendantes,ro`

FIGURE 42 – Probit

Probit regression		Number of obs = 100				
		Wald chi2(21) = 58.24				
		Prob > chi2 = 0.0000				
Log pseudolikelihood = -17.887077		Pseudo R2 = 0.7263				
pauvreté	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
sexe	1.818658	.7791614	2.33	0.020	.2915297	3.345786
statut	.1664134	.2654257	0.63	0.531	-.3538114	.6866381
profession	.5898148	.2921495	2.02	0.043	.0172122	1.162417
niveau_instruction	.3564002	.4094903	0.87	0.384	-.4461861	1.158987
Revenu_chef_mén	-.0304559	.0072095	-4.22	0.000	-.0445862	-.0163256
nombre_enfants	1.062167	.3241102	3.28	0.001	.4269222	1.697411
emploi	1.63773	.5747716	2.85	0.004	.5111982	2.764261
statut_logement	-.608337	.6526514	-0.93	0.351	-1.88751	.6708363
type_logement	.3129441	.563654	0.56	0.579	-.7917974	1.417686
taille_logement	1.180674	.2841121	4.16	0.000	.6238249	1.737524
Accès_soins_santé	1.825743	.552365	3.31	0.001	.7431271	2.908358
Accès_électricité	-.5323763	.6085622	-0.87	0.382	-1.725136	.6603837
Accès_regideso	-.0552149	.5539597	-0.10	0.921	-1.140956	1.030526
Mode_évac_ordures	-.885307	.2626197	-3.37	0.001	-1.400032	-.3705818
possession_voiture	-2.351097	.8763114	-2.68	0.007	-4.068636	-.6335586
possession_ant	2.153229	.9531368	2.26	0.024	.2851149	4.021342
possession_clim	1.149078	.8468108	1.36	0.175	-.5106409	2.808796
appart_mouv	1.938941	.7977816	2.43	0.015	.3753173	3.502564
Accès_crédits	-3.579181	1.270489	-2.82	0.005	-6.069294	-1.089068
Compte_épargne	.9332939	.8157364	1.14	0.253	-.66552	2.532108
age	.0440635	.0245908	1.79	0.073	-.0041336	.0922607
_cons	-6.45199	3.059844	-2.11	0.035	-12.44917	-.4548063

Commentaire :

Le modèle est globalement significatif car sa probabilité est inférieure au seuil de 0.05, ce qui veut dire qu'au moins une variable indépendante retenue explique la variable dépendante.

la colonne $P > |Z|$ permet d'identifier les variables étant statistiquement et significativement associées à la variable dépendante.

En effet, les variables peuvent-être statistiquement et significativement associées à la variable dépendante lorsqu'elles ont des p-values inférieures à 1, 5 et 10%.

Au seuil de 1%, sont statistiquement et significativement associées à la pauvreté les variables revenu chef ménage, les nombres d'enfants scolarisés, les nombres des personnes ayant un emploi, la taille du logement, l'accès aux soins de santé, le mode d'évacuation des ordures et l'accès au crédit.

Au seuil de 5%, les variables sexe, profession, possession voiture, possession d'une antenne avec décodeur et appartenir à un mouvement associative ou culturel.

Au seuil de 10%, la variable âge

Nous avons donc 13 variables qui déterminent la pauvreté.

L'on peut interpréter le signe du coefficient.

Les variables revenu chef de ménage, mode d'évacuation des ordures, possession d'une voiture et l'accès au crédit sont les seules variables qui diminuent la probabilité d'être pauvre. et les 9 autres variables statistiquement et significativement associées à la variable pauvreté augmentent les chances ou la probabilité d'être pauvre.

Deuxième étape : Estimer les effets marginaux.

La commande est `mfx compute`

FIGURE 43 – Effets marginaux

```
Marginal effects after probit
y = Pr(pauvreté) (predict)
= .04988547
```

variable	dy/dx	Std. err.	z	P> z	[95% C.I.]	X
sexe*	.1626399	.10126	1.61	0.108	-.035835 .361115	.63
statut	.0171318	.03034	0.56	0.572	-.042324 .076588	1.82
profes~n	.0607198	.04079	1.49	0.137	-.019227 .140667	2.26
niveau~n	.0366904	.04344	0.84	0.398	-.04846 .121841	1.89
Revenu~n	-.0031353	.00174	-1.80	0.072	-.006546 .000275	249.45
nombre~s	.1093471	.07267	1.50	0.132	-.033083 .251778	1.73
emploi	.1685997	.09775	1.72	0.085	-.022982 .360182	1.49
statut~t	-.0626266	.06888	-0.91	0.363	-.197628 .072375	1.76
type_l~t	.0322167	.05549	0.58	0.562	-.07654 .140974	1.33
taille~t	.1215471	.07027	1.73	0.084	-.016182 .259276	3.52
Accès~nté	.1879551	.09885	1.90	0.057	-.005786 .381696	1.14
Accès~vité	-.0548066	.06644	-0.82	0.409	-.185031 .075418	2.25
Accès_~no	-.0056842	.05703	-0.10	0.921	-.117465 .106096	2.2
Mode_é~ns	-.0911399	.0643	-1.42	0.156	-.217162 .034883	1.99
posses~ve*	-.3332393	.12134	-2.75	0.006	-.57107 -.095409	.52
posses~nt*	.1158722	.06717	1.73	0.085	-.015779 .247523	.79
posses~m	.1182943	.08677	1.36	0.173	-.051764 .288353	.5
apppar~v*	.319994	.14624	2.19	0.029	.033378 .60661	.38
Accès_~s*	-.429079	.13844	-3.10	0.002	-.700413 -.157746	.41
Compte~ve*	.1804676	.2203	0.82	0.413	-.251311 .612246	.04
age	.0045362	.00314	1.45	0.148	-.001613 .010685	41.05

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Commentaire :

Les effets marginaux mesurent comment un petit changement dans une variable indépendante affecte la probabilité de l'événement.

Dans notre cas, les effets marginaux permettent de voir comment un petit changement d'une variable indépendante peut affecter la probabilité d'être pauvre.

¹³Par exemple : toutes les autres variables restants constantes, l'augmentation d'une unité du revenu diminue la probabilité d'être pauvre d'environ 0.31% ce qui reste minime comme impact.

La variable accès au crédit diminue la probabilité d'être pauvre d'environ 42.90%. i.e. avoir un accès au crédit élevé diminue la probabilité d'être pauvre. La variable nombre d'enfants augmentent la probabilité d'être pauvre d'environ 10.93%. i.e. avoir un nombre plus élevé d'enfant augmente la probabilité d'être pauvre.

Troisième étape : voir le tableau de prédiction du modèle.

La commande est `lstat`

FIGURE 44 – Prédiction

```
. lstat
```

Probit model for pauvreté

Classified	True		Total
	D	~D	
+	30	3	33
-	6	61	67
Total	36	64	100

Classified + if predicted $\Pr(D) \geq .5$
True D defined as pauvreté != 0

Sensitivity	$\Pr(+ D)$	83.33%
Specificity	$\Pr(- \sim D)$	95.31%
Positive predictive value	$\Pr(D +)$	90.91%
Negative predictive value	$\Pr(\sim D -)$	91.04%
False + rate for true ~D	$\Pr(+ \sim D)$	4.69%
False - rate for true D	$\Pr(- D)$	16.67%
False + rate for classified +	$\Pr(\sim D +)$	9.09%
False - rate for classified -	$\Pr(D -)$	8.96%
Correctly classified		91.00%

¹⁴Commentaire :

Le tableau indique que pour les personnes pauvres, 30 cas sur 36 ont été bien prédit, donc 6 cas mal prédit cela veut dire qu'il y'a 6 ménages classés comme pauvres mais qui ne sont pas pauvres. et pour les personnes non pauvres, 61

13. le lecteur peut s'exercer en interprétant les autres variables

14. Comme on peut bien le voir, le modèle logit et probit ont presque les mêmes résultats, par conséquent, vous retrouverez le `oprobit` et `mprobit` dans les appendices.

cas sur 64 ont été bien prédit, donc 3 cas mal prédit, ce qui veut dire qu'il y'a 3 ménages classés comme non-pauvres mais qui sont en réalité pauvres. le modèle a correctement classé 91% des observations.

¹⁵On peut faire une régression Probit simple : on veut savoir si le revenu est un déterminant de la pauvreté.

FIGURE 45 – Probit et effet marginal

pauvreté		Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
Revenu_chef_mén		-.0063081	.0017972	-3.51	0.000	-.0098305	-.0027857
_cons		1.082872	.4115764	2.63	0.009	.2761968	1.889547

variable	dy/dx	Std. err.	z	P> z	[95% C.I.]	X
Revenu~n	-.0022311	.00059	-3.79	0.000	-.003386 - .001077	249.45

15. Partant de l'illustration précédente, le lecteur peut s'exercer en interprétant les résultats de cette illustration

FIGURE 46 – Prédiction

. lstat

Probit model for pauvreté

Classified	True		Total
	D	~D	
+	18	14	32
-	18	50	68
Total	36	64	100

Classified + if predicted $\Pr(D) \geq .5$

True D defined as pauvreté != 0

Sensitivity	$\Pr(+ D)$	50.00%
Specificity	$\Pr(- \sim D)$	78.13%
Positive predictive value	$\Pr(D +)$	56.25%
Negative predictive value	$\Pr(\sim D -)$	73.53%
False + rate for true ~D	$\Pr(+ \sim D)$	21.88%
False - rate for true D	$\Pr(- D)$	50.00%
False + rate for classified +	$\Pr(\sim D +)$	43.75%
False - rate for classified -	$\Pr(D -)$	26.47%
Correctly classified		68.00%

la variable profession est statistiquement et significativement associée à la variable niveau d'instruction au seuil de 5%
 et aucune variable au seuil de 10%
 la variable accès soins de santé a un impact négatif et les deux autres variables influencent positivement.

Deuxième étape : avoir les effets marginaux.
 La commande est `margins,dydx(*)`

FIGURE 48 – Effets marginaux

		Delta-method				
		dy/dx	std. err.	z	P> z	[95% conf. interval]
Accès_soins_santé						
	_predict					
	1	.1399229	.0520663	2.69	0.007	.0378749 .2419709
	2	-.0918502	.0382932	-2.40	0.016	-.1669034 -.016797
	3	-.0480727	.0271555	-1.77	0.077	-.1012965 .0051511
profession						
	_predict					
	1	-.0295767	.0273579	-1.08	0.280	-.0831972 .0240437
	2	.0194152	.0184236	1.05	0.292	-.0166944 .0555248
	3	.0101615	.0102527	0.99	0.322	-.0099333 .0302564
statut						
	_predict					
	1	.0180941	.0404589	0.45	0.655	-.0612038 .097392
	2	-.0118776	.0273279	-0.43	0.664	-.0654393 .0416842
	3	-.0062165	.0134577	-0.46	0.644	-.0325932 .0201602
sexe						
	_predict					
	1	-.0585023	.0700532	-0.84	0.404	-.195804 .0787994
	2	.0384029	.0488396	0.79	0.432	-.0573209 .1341268
	3	.0200994	.0231919	0.87	0.386	-.0253558 .0655546
Revenu_chef_mén						
	_predict					
	1	-.0007181	.00032	-2.24	0.025	-.0013453 -.0000908
	2	.0004714	.0002618	1.80	0.072	-.0000417 .0009845
	3	.0002467	.0001201	2.05	0.040	.0000113 .0004821

¹⁷Commentaire :

Dans cette première figure, la variable accès aux soins de santé augmente la probabilité d'avoir un niveau d'instruction primaire et secondaire par rapports

17. Le lecteur peut s'exercer en interprétant les deux autres variables statistiquement et significativement associées à la variable niveau d'instruction et à l'aide d'une base des données, il peut s'exercer pour le modèle oprobit

aux autres niveaux d'études d'environ 13.99‰.

elle diminue à la probabilité d'avoir un niveau d'instruction universitaire par rapports aux autres niveaux d'études d'environ 9.18‰.

Elle diminue également la probabilité d'avoir un niveau d'instruction post universitaire par rapports aux autres niveaux d'études d'environ 4.87‰.

FIGURE 49 – Effets marginaux suite

pauvreté						
_predict						
1	-.045613	.0818025	-0.56	0.577	-.2059429	.1147169
2	.0299419	.0545659	0.55	0.583	-.0770052	.1368891
3	.015671	.0283131	0.55	0.580	-.0398216	.0711637
nombre_enfants						
_predict						
1	.0339147	.0264871	1.28	0.200	-.0179992	.0858285
2	-.0222628	.0177707	-1.25	0.210	-.0570927	.0125672
3	-.0116519	.0104616	-1.11	0.265	-.0321563	.0088525
emploi						
_predict						
1	.0701199	.0399172	1.76	0.079	-.0081163	.1483561
2	-.0460291	.0288087	-1.60	0.110	-.1024931	.0104349
3	-.0240908	.0160339	-1.50	0.133	-.0555166	.007335
Accès_crédits						
_predict						
1	-.279843	.0773554	-3.62	0.000	-.4314569	-.1282292
2	.1836986	.0535197	3.43	0.001	.0788019	.2885953
3	.0961445	.0533922	1.80	0.072	-.0085023	.2007913
Compte_épargne						
_predict						
1	-.1578522	.2560246	-0.62	0.538	-.6596511	.3439468
2	.1036196	.1653525	0.63	0.531	-.2204654	.4277046
3	.0542326	.0945523	0.57	0.566	-.1310865	.2395517

FIGURE 50 – Effets marginaux suite et fin

age						
_predict						
1	.0025804	.003146	0.82	0.412	-.0035857	.0087466
2	-.0016939	.0021531	-0.79	0.431	-.005914	.0025262
3	-.0008866	.0010812	-0.82	0.412	-.0030058	.0012327

FIGURE 51 – Spécification

. linktest

Iteration 0: log likelihood = -72.71388
 Iteration 1: log likelihood = -59.973265
 Iteration 2: log likelihood = -58.939679
 Iteration 3: log likelihood = -58.924965
 Iteration 4: log likelihood = -58.924963

Ordered logistic regression

Number of obs = 100
 LR chi2(2) = 27.58
 Prob > chi2 = 0.0000
 Pseudo R2 = 0.1896

Log likelihood = -58.924963

niveau_instruction	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
_hat	1.471357	.4354655	3.38	0.001	.6178604	2.324854
_hatsq	-.1759767	.1297676	-1.36	0.175	-.4303166	.0783631
/cut1	-.6762958	.3184948			-1.300534	-.0520574
/cut2	4.627312	.7045126			3.246493	6.008131
/cut3	6.530982	1.15866			4.260049	8.801914

Commentaire :

Etant donné que la probabilité du _hatsq est supérieure au seuil de 0.05, notre modèle est bien spécifié.

6.3 Mlogit Multiple

¹⁸Illustration : On veut voir dans cette base des données, quelles sont les variables qui peuvent expliquer l'accès aux soins de santé du chef de ménage, cette illustration est juste à titre exemplatif afin de garder la même base des données.

Première étape : estimer le modèle mlogit.

La commande est mlogit variable dépendante variable indépendante,ro

FIGURE 52 – Mlogit

Multinomial logistic regression		Number of obs = 100				
Log pseudolikelihood = -21.084638		Wald chi2(22) = 713.43				
		Prob > chi2 = 0.0000				
		Pseudo R2 = 0.3005				
Accès_soins_santé		Robust Coefficient	std. err.	z	P> z	[95% conf. interval]
1		(base outcome)				
2	profession	.4327414	.9775579	0.44	0.658	-1.483237 2.34872
	statut	.3297554	.7743752	0.43	0.670	-1.187992 1.847503
	sexe	.6058333	1.830844	0.33	0.741	-2.982554 4.194221
	niveau_instruction	-.676114	1.158888	-0.58	0.560	-2.947494 1.595266
	Revenu_chef_mén	.0028003	.0067379	0.42	0.678	-.0104057 .0160063
	pauvreté	.5125893	1.412655	0.36	0.717	-2.256164 3.281342
	nombre_enfants	-1.375393	.8211169	-1.68	0.094	-2.984752 .2339671
	emploi	.2420737	.395169	0.61	0.540	-.5324433 1.016591
	Accès_crédits	-.4901594	.7198251	-0.68	0.496	-1.900991 .9206719
	Compte_épargne	-12.46963	1.233459	-10.11	0.000	-14.88717 -10.0521
	age	-.1086469	.0465978	-2.33	0.020	-.1999769 -.0173169
	_cons	.5782558	1.802317	0.32	0.748	-2.954221 4.110733

Commentaire :

Le modèle est globalement significatif car sa probabilité est inférieure au seuil de 0.05.

La machine a choisi automatiquement comme modalité de base(base outcome) l'hôpital, i.e. les coefficients associés aux autres modalités sont interprétés par rapport à l'hôpital.

Pour la modalité traditionnelle, il y'a 3 variables qui sont statistiquement et significativement associées à la variable dépendante, l'on a les nombres d'enfants scolarisés (au seuil de 10‰), avec un coefficient négatif qui veut dire que les chefs des ménages ayant beaucoup d'enfants scolarisés ont une probabilité faible de choisir le mode traditionnel par rapport à l'hôpital.

18. Vous retrouverez le Mlogit simple dans les appendices.

la variable compte épargne (au seuil de 1%) avec un coefficient négatif qui veut dire que les chefs des ménages ayant un compte épargne ont une probabilité faible de choisir le mode traditionnel par rapport à l'hôpital.

La variable âge (au seuil de 5%) avec un coefficient négatif qui veut dire qu'à mesure que l'âge du chef de ménage avance, celui-ci a une probabilité faible de choisir le mode traditionnel par rapport à l'hôpital.

FIGURE 53 – Mlogit suite

3						
profession	1.277681	.615441	2.08	0.038	.0714384	2.483923
statut	-.7561652	1.274001	-0.59	0.553	-3.253161	1.740831
sexe	.7288412	1.763436	0.41	0.679	-2.72743	4.185113
niveau_instruction	-3.063968	1.402466	-2.18	0.029	-5.812751	-.3151841
Revenu_chef_mén	.0145149	.0077255	1.88	0.060	-.0006268	.0296566
pauvreté	2.336941	2.202894	1.06	0.289	-1.980651	6.654533
nombre_enfants	-.8389589	.7876202	-1.07	0.287	-2.382666	.7047484
emploi	-.3506544	.5844979	-0.60	0.549	-1.496249	.7949405
Accès_crédits	2.066909	1.259714	1.64	0.101	-.402085	4.535903
Compte_épargne	-10.84144	1.353678	-8.01	0.000	-13.4946	-8.188283
age	-.0826427	.0526778	-1.57	0.117	-.1858893	.0206039
_cons	-1.887523	3.394169	-0.56	0.578	-8.539972	4.764927

Note: 4 observations completely determined. Standard errors questionable.

Commentaire

Pour la modalité automédication, il y'a 4 variables qui sont statistiquement et significativement associées à la variable dépendante, l'on a La variable profession (au seuil de 5%) avec un coefficient positif qui veut¹⁹ dire. niveau d'instruction (au seuil de 5%), avec un coefficient négatif qui veut dire que les chefs des ménages ayant un niveau élevé ont une probabilité faible de choisir le mode automédication par rapport à l'hôpital.

Le revenu du chef de ménage (au seuil de 10%), avec un coefficient positif. la variable compte épargne (au seuil de 1%) avec un coefficient négatif qui veut dire que les chefs des ménages ayant un compte épargne ont une probabilité faible de choisir le mode automédication par rapport à l'hôpital.

19. le lecteur peut s'exercer avec cette interprétation et celle sur le revenu

Deuxième étape : avoir les effets marginaux.

La commande est `margins, dydx(*)`

FIGURE 54 – Effets marginaux

		Delta-method				
		dy/dx	std. err.	z	P> z	[95% conf. interval]
profession						
	_predict					
	1	-.0053208	.0054339	-0.98	0.327	-.0159711 .0053296
	2	.0027566	.0050337	0.55	0.584	-.0071093 .0126225
	3	.0025642	.0023479	1.09	0.275	-.0020376 .007166
statut						
	_predict					
	1	-.0005979	.0040263	-0.15	0.882	-.0084892 .0072934
	2	.0021231	.0031376	0.68	0.499	-.0040266 .0082727
	3	-.0015252	.0023763	-0.64	0.521	-.0061825 .0031322
sexe						
	_predict					
	1	-.005331	.0092502	-0.58	0.564	-.0234611 .012799
	2	.003873	.0080321	0.48	0.630	-.0118696 .0196156
	3	.001458	.0038469	0.38	0.705	-.0060817 .0089978
niveau_instruction						
	_predict					
	1	.0104468	.0068424	1.53	0.127	-.002964 .0238575
	2	-.004293	.0058241	-0.74	0.461	-.015708 .0071219
	3	-.0061537	.0043117	-1.43	0.154	-.0146046 .0022971

Commentaire :

²⁰La variable profession diminue la probabilité de choisir l'hôpital comme mode d'environ 0.53%, elle augmente la probabilité de choisir le mode traditionnel d'environ 0.27% et celle de choisir l'automédication d'environ 0.25%.

20. Le lecteur peut s'exercer en commentant les variables niveau d'instruction, revenu et compte épargne et exécuter le `mprobit` avec une base des données adéquates.

FIGURE 55 – Effets marginaux suite

Revenu_chef_mén							
	_predict						
	1	-.0000469	.0000633	-0.74	0.459	-.000171	.0000772
	2	.0000178	.0000593	0.30	0.765	-.0000985	.000134
	3	.0000292	.0000197	1.48	0.139	-9.42e-06	.0000677
pauvreté							
	_predict						
	1	-.0079481	.0119173	-0.67	0.505	-.0313056	.0154093
	2	.0032545	.0109018	0.30	0.765	-.0181126	.0246217
	3	.0046936	.0045277	1.04	0.300	-.0041805	.0135677
nombre_enfants							
	_predict						
	1	.0104728	.0077243	1.36	0.175	-.0046665	.0256121
	2	-.0088033	.0073689	-1.19	0.232	-.0232461	.0056395
	3	-.0016695	.002354	-0.71	0.478	-.0062833	.0029443
emploi							
	_predict						
	1	-.0008475	.0030947	-0.27	0.784	-.0069129	.005218
	2	.0015559	.0029245	0.53	0.595	-.0041759	.0072877
	3	-.0007084	.0010499	-0.67	0.500	-.0027661	.0013493
Accès_crédits							
	_predict						
	1	-.0009955	.0048457	-0.21	0.837	-.0104929	.008502
	2	-.0031681	.0037399	-0.85	0.397	-.010498	.0041619
	3	.0041635	.0030405	1.37	0.171	-.0017957	.0101227

FIGURE 56 – Effets marginaux fin

Compte_épargne							
	_predict						
	1	.1014138	.1091014	0.93	0.353	-.112421	.3152485
	2	-.0797706	.1074675	-0.74	0.458	-.290403	.1308617
	3	-.0216432	.0179979	-1.20	0.229	-.0569184	.013632
age							
	_predict						
	1	.00086	.0009784	0.88	0.379	-.0010576	.0027776
	2	-.0006952	.000968	-0.72	0.473	-.0025924	.0012021
	3	-.0001648	.0001231	-1.34	0.181	-.000406	.0000764

7 Appendice 1 : Force et direction du coefficient de corrélation

TABLE 3 – Coefficient

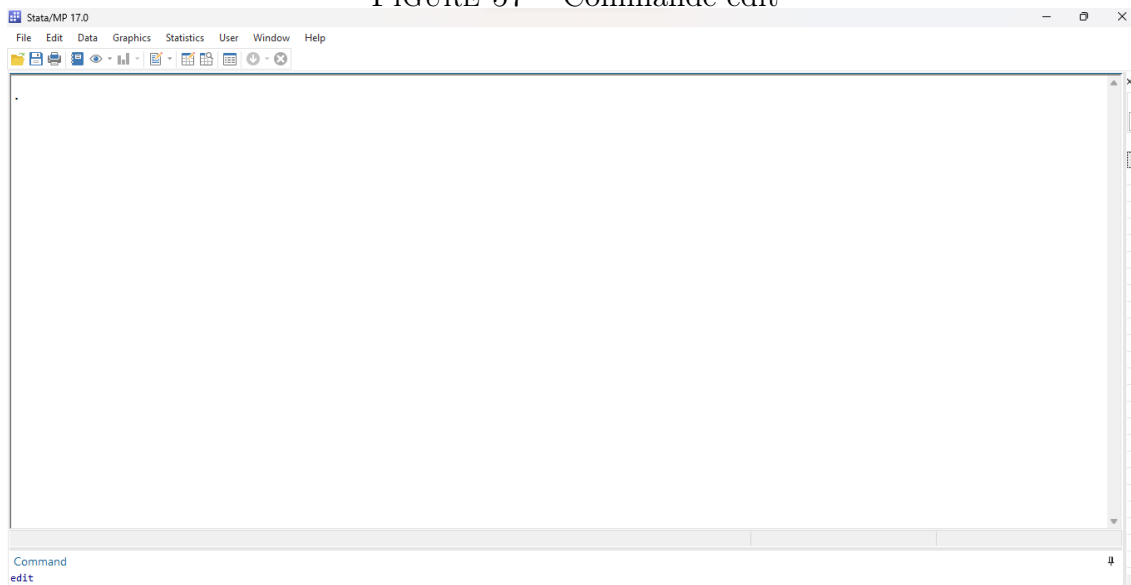
Valeur du coefficient de corrélation	Force	Direction
Supérieur à 0,5	Fort	Positif
Entre 0,3 et 0,5	Modéré	Positif
Entre 0 et 0,3	Faible	Positif
0	Aucun	Aucun
Entre 0 et -0,3	Faible	Négatif
Entre -0,3 et -0,5	Modéré	Négatif
Moins de -0,5	Fort	Négatif

Source : <https://datatab.fr/tutorial/pearson-correlation>, consulté le 10/12/2023.

8 Appendice 2 : La commande edit

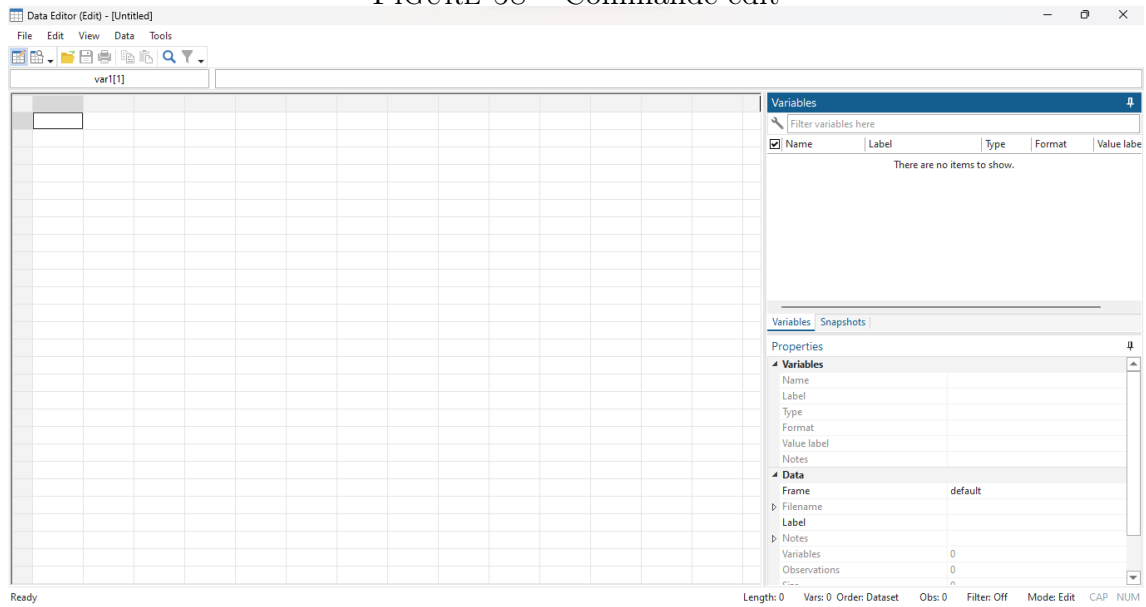
La commande edit permet d'ouvrir la fenêtre permettant de coller la base des données Excel.

FIGURE 57 – Commande edit



Après avoir exécuter la commande, voici la fenêtre.

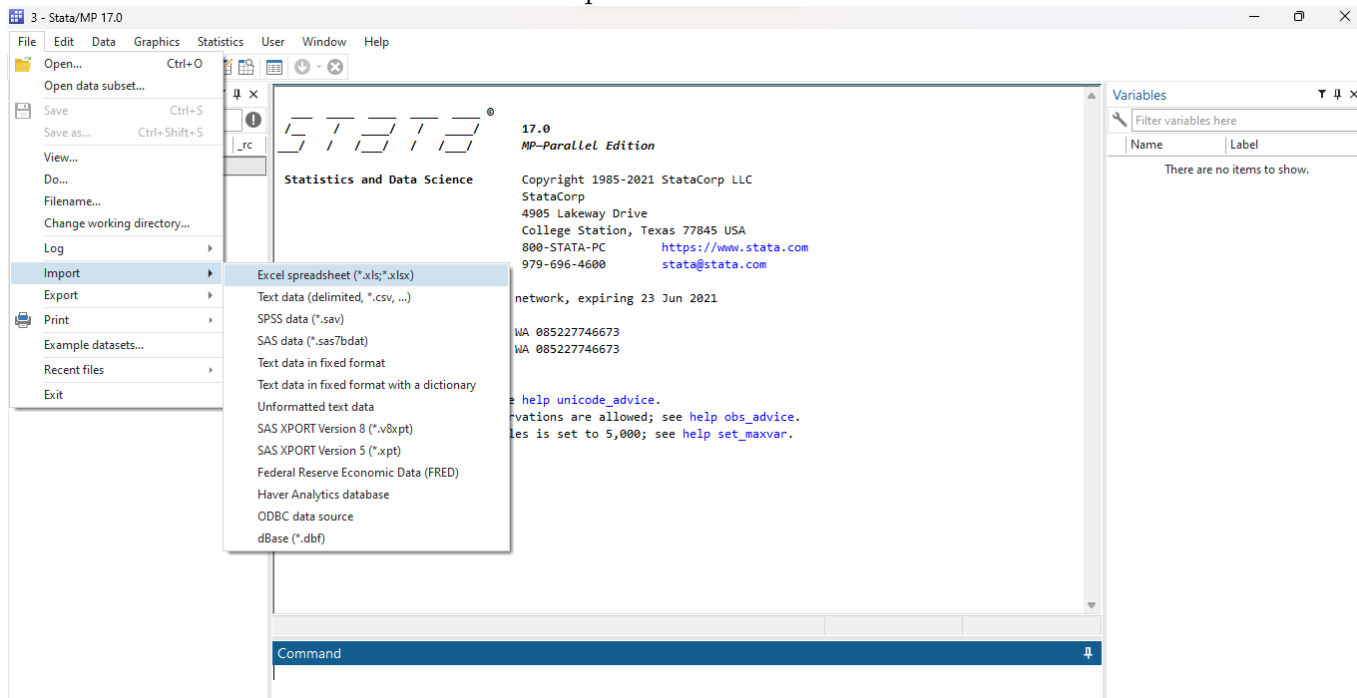
FIGURE 58 – Commande edit



9 Appendice 3 : Importer fichier Excel

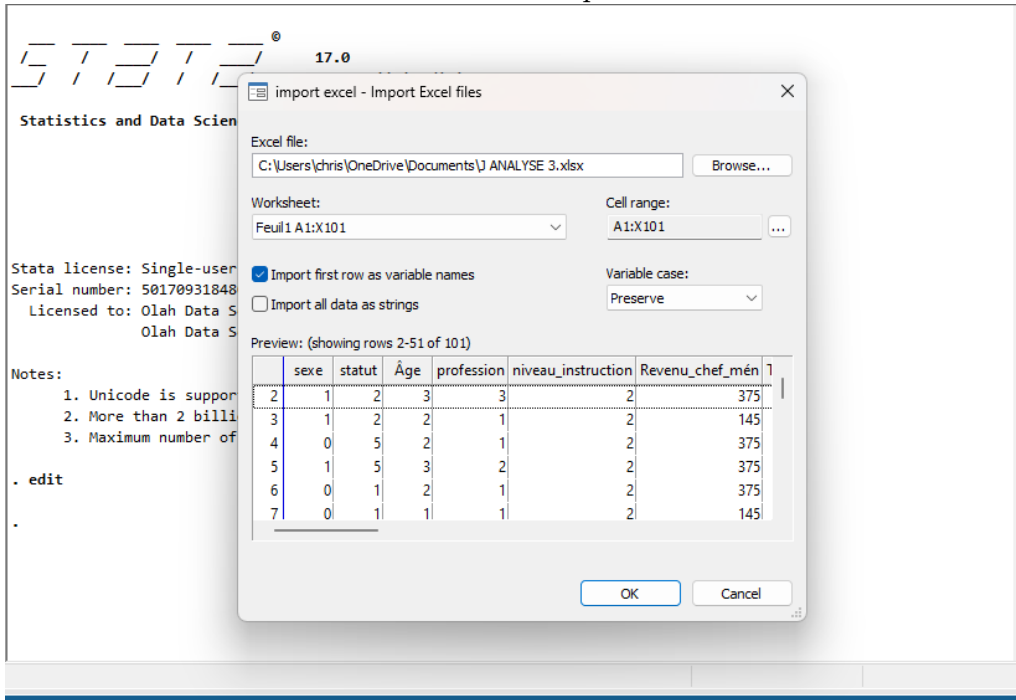
On peut aussi importer directement un fichier Excel sur Stata sans copier-coller.

FIGURE 59 – Importer le fichier



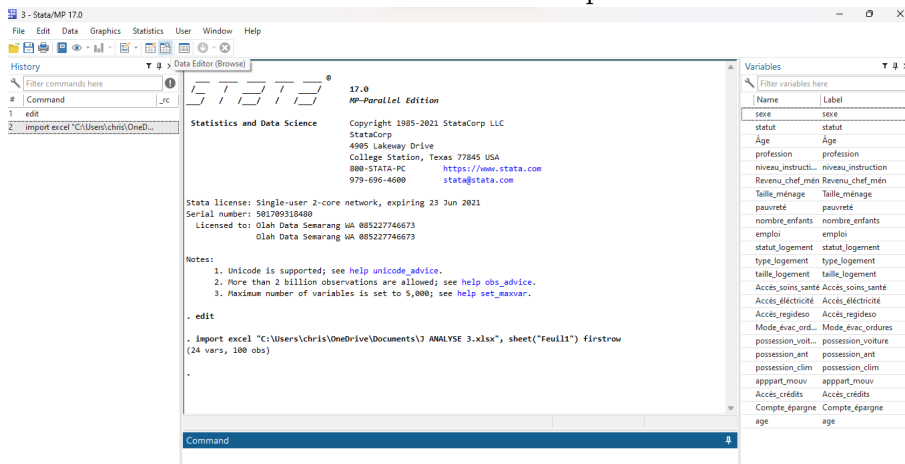
Ensuite, il faut surtout cocher import first row as variables names, cette option fait comprendre au logiciel que la première ligne fait référence aux noms des variables.

FIGURE 60 – Importer le fichier



Il faut cliquer sur data editor(browser) pour avoir un aperçu des variables importées.

FIGURE 61 – Importer le fichier



Voici l'aperçu du data editor (browser)

FIGURE 62 – Importation réussie

The screenshot shows the Data Editor (Browse) window for a dataset named 'sexe[1]'. The main window displays a table with 26 rows and 12 columns. The columns are: sexe, statut, Âge, profession, niveau_ins-n, Revenu_che-n, Taille_ménage, pauvreté, nombre_enf-s, emploi, and statut_loj. The data is as follows:

Row	sexe	statut	Âge	profession	niveau_ins-n	Revenu_che-n	Taille_ménage	pauvreté	nombre_enf-s	emploi	statut_loj
1	1	2	3	3	2	375	8	0	3	2	
2	1	2	2	1	2	145	2	0	3	1	
3	0	5	2	1	2	375	3	0	1	1	
4	1	5	3	2	2	375	5	0	2	2	
5	0	1	2	1	2	375	4	0	2	2	
6	0	1	1	1	2	145	3	0	1	1	
7	0	1	2	1	2	225	2	0	1	1	
8	1	2	3	3	2	375	3	0	2	3	
9	1	2	1	2	2	225	6	1	2	1	
10	1	2	3	2	2	375	6	0	1	1	
11	0	2	2	3	2	225	4	0	1	1	
12	1	2	2	2	2	375	5	0	1	1	
13	1	2	2	3	2	225	11	1	3	2	
14	1	2	1	1	2	225	4	0	1	1	
15	1	2	2	2	2	375	6	0	1	1	
16	0	2	3	2	1	145	8	1	2	2	
17	0	1	1	1	2	225	2	0	1	1	
18	1	1	1	1	2	225	1	0	0	1	
19	1	2	2	3	2	225	6	1	1	2	
20	0	2	3	2	1	145	8	1	2	2	
21	0	1	1	4	2	145	9	1	5	2	
22	0	1	1	1	2	225	1	0	0	1	
23	0	2	2	1	2	375	6	0	2	1	
24	1	1	1	2	2	145	11	1	0	1	
25	0	1	1	5	2	145	7	1	3	1	
26	0	1	1	5	2	145	4	1	2	2	

The Variables panel on the right lists the following variables:

- sexe: Type byte, Format %10.0g
- statut: Type byte, Format %10.0g
- Âge: Type byte, Format %10.0g
- profession: Type byte, Format %10.0g
- niveau_instructi...: Type byte, Format %10.0g
- Revenu_chef_mén: Type int, Format %10.0g
- Taille_ménage: Type byte, Format %10.0g
- pauvreté: Type byte, Format %10.0g
- nombre_enfants: Type byte, Format %10.0g

The Properties panel shows the following data:

- Frame: default
- Filename: [empty]
- Label: [empty]
- Notes: [empty]
- Variables: 24
- Observations: 100
- File: [empty]

At the bottom, the status bar indicates: Vars: 24 Order: Dataset Obs: 100 Filter: Off Mode: Browse CAP NUM

10 Appendice 4 : La commande asdoc

²¹La commande asdoc permet d'enregistrer, exporter directement les résultats en format word.

Pour commencer, on lance notre première commande, ici, il s'agit d'un tableau des fréquences.

La commande asdoc tab variable, save(nom du document que l'on veut créer.doc)

FIGURE 63 – Asdoc

```
. asdoc tab pauvreté , save(Analyse.doc)
```

pauvreté	Freq.	Percent	Cum.
0	64	64.00	64.00
1	36	36.00	100.00
Total	100	100.00	

Click to Open File: [Analyse.doc](#)

21. Le lecteur peut se servir de cette section pour s'exercer en interprétant les différents résultats.

Ensuite, on lance les restes des analyses des fréquences.

La commande devient `asdoc tab1 variables, save(nom du document.doc)`

FIGURE 64 – Asdoc

`. asdoc tab1 sexe statut profession, save (Analyse.doc)`
 (File Analyse.doc already exists, option `append` was assumed)

sexe	Freq.	Percent	Cum.
0	37	37.00	37.00
1	63	63.00	100.00
Total	100	100.00	

statut	Freq.	Percent	Cum.
1	41	41.00	41.00
2	48	48.00	89.00
3	3	3.00	92.00
4	4	4.00	96.00
5	4	4.00	100.00
Total	100	100.00	

profession	Freq.	Percent	Cum.
1	31	31.00	31.00
2	31	31.00	62.00
3	26	26.00	88.00
4	5	5.00	93.00
5	7	7.00	100.00
Total	100	100.00	

Click to Open File: [Analyse.doc](#)

On choisi quelques variables pour illustrer asdoc avec les analyses bivariées.

La commande est `asdoc tab2 variable dépendante variables indépendantes, chi222firstonly`

FIGURE 65 – Asdoc

```
. asdoc tab2 pauvreté sexe Accès_crédits, chi2 firstonly
(File Analyse.doc already exists, option append was assumed)
-> tabulation of pauvreté by sexe
```

pauvreté	sexe		Total
	0	1	
0	25	39	64
1	12	24	36
Total	37	63	100

Pearson chi2(1) = 0.3244 Pr = 0.569

```
-> tabulation of pauvreté by Accès_crédits
```

pauvreté	Accès_crédits		Total
	0	1	
0	35	29	64
1	24	12	36
Total	59	41	100

Pearson chi2(1) = 1.3668 Pr = 0.242

Click to Open File: [Analyse.doc](#)

22. L'option `firstonly` fait comprendre à la machine que seule la première variable est la dépendante. Sans cette option, la machine lancera des analyses avec chaque variable prise comme la dépendante à tour de rôle

as doc et el modèle de régression.

La commande est `asdoc logit pauvreté sexe statut Revenu_chef_mén,ro`

FIGURE 66 – Asdoc

```
. asdoc logit pauvreté sexe statut Revenu_chef_mén,ro
(File Analyse.doc already exists, option append was assumed)

Iteration 0:  log pseudolikelihood = -65.341819
Iteration 1:  log pseudolikelihood = -55.92373
Iteration 2:  log pseudolikelihood = -55.027358
Iteration 3:  log pseudolikelihood = -55.01578
Iteration 4:  log pseudolikelihood = -55.015771
Iteration 5:  log pseudolikelihood = -55.015771

Logistic regression                                Number of obs =   100
                                                    Wald chi2(3) =  12.26
                                                    Prob > chi2   =  0.0065
Log pseudolikelihood = -55.015771                Pseudo R2     =  0.1580
```

pauvreté	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
sexe	.4929391	.4755373	1.04	0.300	-.4390969	1.424975
statut	.4236335	.2595136	1.63	0.103	-.0850038	.9322708
Revenu_chef_mén	-.0128463	.0037514	-3.42	0.001	-.020199	-.0054936
_cons	1.223992	.8078018	1.52	0.130	-.3592707	2.807254

Click to Open File: [Analyse.doc](#)

.

As doc et les effets marginaux.

La commande est asdoc mfx compute

FIGURE 67 – Asdoc

Marginal effects after logit

y = Pr(pauvreté) (predict)

= .28926448

variable	dy/dx	Std. err.	z	P> z	[95% C.I.]	X
sexe*	.0983608	.09278	1.06	0.289	-.083493 .280214	.63
statut	.087095	.05352	1.63	0.104	-.017794 .191984	1.82
revenu~n	-.0026411	.00067	-3.96	0.000	-.003948 -.001335	249.45

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Click to Open File: [Analyse.doc](#)

As doc et le modèle de spécification.

La commande est asdoc linktest

FIGURE 68 – Asdoc

```
. asdoc lstat
(File Analyse.doc already exists, option append was assumed)
```

Logistic model for pauvreté

Classified	True		Total
	D	~D	
+	16	11	27
-	20	53	73
Total	36	64	100

Classified + if predicted $\Pr(D) \geq .5$

True D defined as pauvreté $\neq 0$

Sensitivity	$\Pr(+ D)$	44.44%
Specificity	$\Pr(- \sim D)$	82.81%
Positive predictive value	$\Pr(D +)$	59.26%
Negative predictive value	$\Pr(\sim D -)$	72.60%

False + rate for true ~D	$\Pr(+ \sim D)$	17.19%
False - rate for true D	$\Pr(- D)$	55.56%
False + rate for classified +	$\Pr(\sim D +)$	40.74%
False - rate for classified -	$\Pr(D -)$	27.40%

Correctly classified	69.00%
----------------------	--------

Click to Open File: [Analyse.doc](#)

Un aperçu du modèle logit et effets marginaux sur word grâce à la commande asdoc.

FIGURE 69 – Aperçu word

Logistic regression						
pauvret	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]
sexe	.493	.476	1.04	.3	-.439	1.425
statut	.424	.26	1.63	.103	-.085	.932
Revenu_chef_m	-.013	.004	-3.42	.001	-.02	-.005
Constant	1.224	.808	1.52	.13	-.359	2.807
Mean dependent var		0.360	SD dependent var			0.482
Pseudo r-squared		0.158	Number of obs			100
Chi-square		12.264	Prob > chi2			0.007
Akaike crit. (AIC)		118.032	Bayesian crit. (BIC)			128.452

*** $p < .01$, ** $p < .05$, * $p < .1$

Marginal effects after logit
 $y = \text{Pr}(\text{pauvret})$ (predict)
 $= .28926448$

variable	dy/dx	Std.	err.	z	P>z	[
sexe*	0.098	0.093	1.060	0.289	-0.083	0.280
statut	0.087	0.054	1.630	0.104	-0.018	0.192
Revenu~n	-0.003	0.001	-3.960	0.000	-0.004	-0.001

11 Appendice 5 : Install.

Il existe un certain nombre des packages(programmes) qui ne sont pré-installés lorsque nous téléchargeons stata mais que nous pouvons selon nos besoins installer.

La commande est `ssc install nom du package`

FIGURE 70 – Install

```
. . .  
. ssc install jb6  
checking jb6 consistency and verifying not already installed...  
installing into c:\ado\plus\...  
installation complete.
```

```
.  
Command
```

12 Appendice 6 : Describe

Le logiciel donne la possibilité de voir les packages existant, non encore téléchargés afin que l'on puisse selon notre besoin télécharger.

La commande est `ssc describe` première lettre de la liste des packages.

Par exemple, je voudrais voir la liste de tous les programmes commençant par `b` que contient `stata` et que je peux télécharger.

`SSC describe b`

FIGURE 71 – Describe

```
. ssc describe b

http://fmwww.bc.edu/repec/bocode/b/
(no title)

PACKAGES you could -net describe-:
  b1x2          module to account for changes when X2 is added to a base model with X1
  babibplot     module to plot two graph types which are rooted in Bland-Altman plots using journal and
                paper percentiles
  backrasch     module to implement a backward procedure with a Rasch model
  backup        module to make daily backup of important files (Windows only)
  bacondecomp   module to perform a Bacon decomposition of difference-in-differences estimation
  baing         module to determine and estimate the number of common factors following Bai and Ng
  balancetable  module to build a balance table and print it in a LaTeX file or an Excel file
  bandplot      module to plot summary statistics of responses for bands of predictors
  baplot        module to produce Bland-Altman plots
  barplot       module to plot varlist against xvar
  barplot2      module to produce bar plot with optional error bars
  basetable     module to compare a set of risk factors or effects with respect to a categorical variable
  batcher       module to parallelise tasks
  batplot       module to produce Bland-Altman plots accounting for trend
  bayerhanck    module to compute test for non-cointegration
  bayesmixedlogit module to perform Bayesian estimation of mixed logit models
  bayesmixedlogitwtp module for Bayesian estimation of mixed logit model in willingness-to-pay (WTP) space
  bayesmlogit   module to perform Bayesian estimation of mixed logit models
  bcii          module to to estimate the number needed to treat (NNT) and confidence intervals for
                patients improving, or 'benefiting' (either improvements gained or deteriorations
                prevented), in a random
```

Commentaire : cette figure ne fait pas état de tous les packages commençant par `b`, la liste continue.

13 Appendice 7 : Help

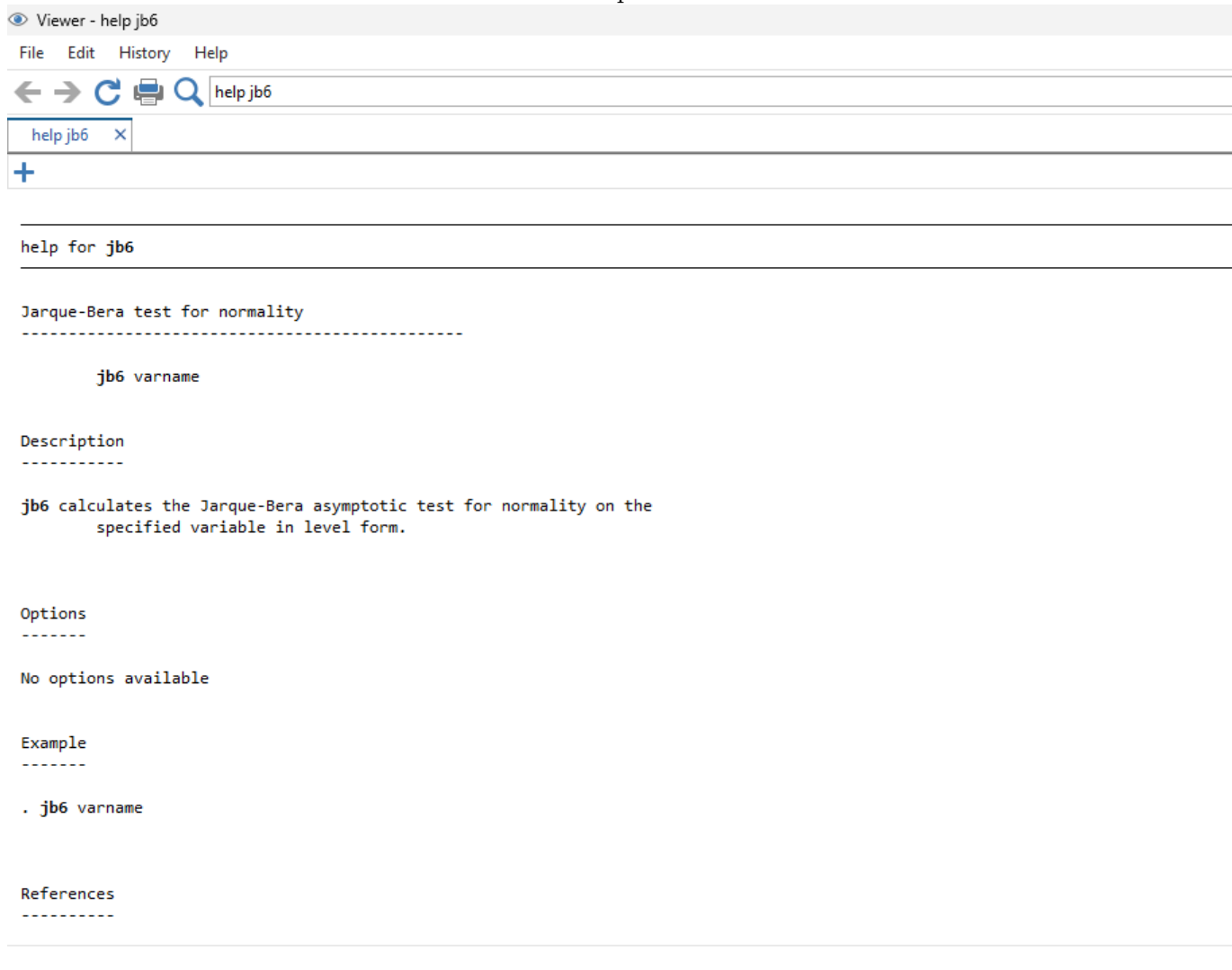
Nous pouvons demander de l'aide au sujet d'un package afin d'avoir des informations nécessaires.

La commande est `help nom du package`.

Nous nous référons à la figure 70 se trouvant à la page 80 pour demander au logiciel de nous expliquer son utilisation.

```
help jb6
```

FIGURE 72 – Help



14 Appendice 8 : Option detail

L'on peut vouloir faire une analyse descriptive approfondies et pour ce faire, il faudrait aller au-déla de la comande sum qui nous offre juste la moyenne, l'écart-type, le minimum et le maximum.

L'un des moyens que l'on peut utiliser est la commande detail.

La commande est sum nos différentes variables quantitatives, detail

FIGURE 73 – Detail

Revenu_chef_mén				
	Percentiles	Smallest		
1%	145	145		
5%	145	145		
10%	145	145	Obs	100
25%	145	145	Sum of wgt.	100
50%	225		Mean	249.45
		Largest	Std. dev.	119.7364
75%	300	550		
90%	375	650	Variance	14336.82
95%	462.5	690	Skewness	1.72717
99%	690	690	Kurtosis	6.264017
Taille_ménage				
	Percentiles	Smallest		
1%	1	1		
5%	1	1		
10%	1	1	Obs	100
25%	2	1	Sum of wgt.	100
50%	4.5		Mean	4.74
		Largest	Std. dev.	2.890668
75%	7	11		
90%	8	11	Variance	8.35596
95%	10.5	11	Skewness	.4249188
99%	11.5	12	Kurtosis	2.353478
.				
Command				

Commentaire :

Avec cette commande, on a des indicateurs qui se rajoutent, tels que les percentiles, la médiane, la variance, les smallest et largest, le coefficient d'asymétrie qu'on appelle le *skwness* et le coefficient d'aplatissement que l'on appelle le *kurtosis*.

Expliquons à présent chaque élément :

1. ²³Le percentile : il fournit les informations sur la manière dont les observations sont réparties dans l'intervalle la plus petite (minimum) et la plus grande valeur (maximum).

Nous voyons pour le revenu que le 90^{èm} percentile est égal à 375, ce qui nous fait comprendre que seulement 10% de notre échantillon ont un revenu supérieur ou égal à 375. Pour la taille de ménage, l'on constate que 75% de notre échantillon ont une taille de ménage inférieure ou égale à 7 personnes.

2. ²⁴La médiane : il s'agit ici du milieu d'une distribution, elle correspond à la valeur de la distribution qui partage l'effectif en deux sous-effectifs de même taille.

Nous voyons que la médiane pour le revenu est de 225 ce qui traduit que 50% de notre échantillon ont un revenu inférieur à 225 et le 50% restant ont un revenu supérieur à 225. Pour la taille de ménage, 50% de nos ménages sont composés de moins de 5 personnes et le 50% restant sont composés de plus de 5 personnes.

3. Smallest : le premier smallest correspond au minimum.
4. Largest : le dernier largest correspond au maximum.
5. Le *skwness* : ce coefficient nous fournit les informations sur l'asymétrie de la distribution, grâce à elle, l'on peut analyser la normalité d'une distribution. Nous pouvons dire que l'asymétrie indique à quel point notre distribution sous-jacente s'écarte de la distribution normale puisque la distribution normale a une asymétrie de 0.
En général, nous avons trois types d'asymétrie.

- (a) Symétrique : lorsque l'asymétrie est proche de 0 et que la moyenne est presque la même que la médiane
- (b) Inclinaison négative : Lorsque la queue gauche de l'histogramme de la distribution est plus longue et que la majorité des observations sont concentrées sur la queue droite. Dans ce cas, nous pouvons également utiliser le terme « biaisé à gauche » ou « arrière gauche ». Et la médiane est supérieure à la moyenne.
- (c) Inclinaison positive : Lorsque la queue droite de l'histogramme de la distribution est plus longue et que la majorité des observations sont

23. Musangu M. (2018), *Elements de statistique descriptive*, G1 Faculté d'administration des affaires et sciences économiques, UPC, Kinshasa.

24. Musangu M. (2018), *Elements de statistique descriptive*, G1 Faculté d'administration des affaires et sciences économiques, UPC, Kinshasa.

concentrées sur la queue gauche. Dans ce cas, nous pouvons également utiliser le terme « biaisé à droite » ou « à queue droite ». Et la médiane est inférieure à la moyenne.

Valeurs d'asymétrie et interprétation :

Il existe de nombreuses approches différentes de l'interprétation des valeurs d'asymétrie. Une règle de base stipule que :

- (a) Symétrique : valeurs comprises entre -0,5 et 0,5
- (b) Données asymétriques modérées : valeurs comprises entre -1 et -0,5 ou entre 0,5 et 1
- (c) Données fortement biaisées : valeurs inférieures à -1 ou supérieures à 1.

Nous avons pour le revenu un coefficient de skewness de 1.72 qui fait référence à une asymétrie fortement biaisées, ce qui fait comprendre que le revenu n'est pas normalement distribué. Tandis que pour la taille de ménage, le coefficient est de 0.42 compris entre -0.5 et 0.5, nous avons donc une symétrie, cette variable suit la loi normale, elle est normalement distribuée.

6. Le kurtosis : En statistique, nous utilisons la mesure de l'aplatissement pour décrire la « queue » de la distribution car elle décrit la forme de celle-ci. C'est aussi une mesure du « pic » de la distribution. Une distribution de kurtosis élevée a un pic plus net et des queues plus grosses plus longues, tandis qu'une distribution de kurtosis faible a un pic plus arrondi et des queues plus courtes et plus fines.

Voyons les trois principaux types de kurtosis :

- (a) Mésokurtique : c'est la distribution normale
- (b) Leptokurtic : Cette distribution a des queues plus grosses et un pic plus net. Le kurtosis est « positif » avec une valeur supérieure à 3
- (c) Platykurtic : La distribution a un pic plus bas et plus large et des queues plus minces. Le kurtosis est « négatif » avec une valeur inférieure à 3

La valeur kurtosis du revenu est 6.26, ce qui laisse comprendre que cette variable est leptokurtic, donc elle a des queues plus grosses et un pic plus net. Celle de la variable taille de ménage est 2.35 est mésokurtique, elle est normalement distribuée.

15 Appendice 9 : Tabstat,stats

Cette commande est comme detail, elle permet de faire une analyse descriptive plus poussée.

La commande est tabstat nos variables, stats(mean variance sd min max median sk k).

FIGURE 74 – Tabstat

```
. tabstat Âge emploi nombre_enfants taille_logement ,stats(mean variance sd min max median sk
```

Stats	Âge	emploi	nombre~s	taille~t
Mean	1.78	1.49	1.73	3.52
Variance	.6177778	.979697	1.855657	2.292525
SD	.7859884	.9897964	1.362225	1.514109
Min	1	0	0	1
Max	3	6	6	9
p50	2	1	1	3
Skewness	.406238	1.786655	.97901	.7074225
Kurtosis	1.740485	7.66465	3.45307	3.702822

Commentaire :

tous ces indicateurs ont déjà été interprétés (page 82-83), prière au lecteur de s'exercer en interprétant ces données.

16 Appendice 10 : Ameans

En effet, les commandes `sum` et `mean` ne donnent que la moyenne arithmétique, c'est la moyenne la plus utilisée mais il est à noter qu'utiliser cette moyenne face à certaines variables, données est un biais, une erreur pourtant commise par plusieurs.

Hormis la moyenne arithmétique, le logiciel `stata` propose les moyennes harmonique et géométrique.

En effet,²⁵ la moyenne harmonique est utilisée dans des domaines où il existe des liens de proportionnalité inverse.²⁶ *Pour une distance donnée plus la vitesse est élevée, le temps du trajet est court.*

²⁷ La moyenne géométrique : est la moyenne permettant de calculer les taux moyens, notamment les taux moyens annuels. Son utilisation est justifiée uniquement lorsque les valeurs ont un caractère multiplicatif.

Nous les implémentons à titre illustratif .

La commande est `ameans` variables

FIGURE 75 – Ameans

```
. amean Revenu_chef_mén Taille_ménage emploi nombre_enfants taille_logement Âge
```

Variable	Type	Obs	Mean	[95% conf. interval]	
Revenu_chef_mén	Arithmetic	100	249.45	225.6917	273.2083
	Geometric	100	227.8221	210.0356	247.1149
	Harmonic	100	211.3238	197.2298	227.5871
Taille_ménage	Arithmetic	100	4.74	4.166429	5.313571
	Geometric	100	3.729593	3.20797	4.336033
	Harmonic	100	2.742727	2.344379	3.304156
emploi	Arithmetic	100	1.49	1.293603	1.686397
	Geometric	94	1.399551	1.272105	1.539764
	Harmonic	94	1.278332	1.189539	1.38145
nombre_enfants	Arithmetic	100	1.73	1.459705	2.000295
	Geometric	86	1.698443	1.503601	1.918533
	Harmonic	86	1.475129	1.339043	1.642004
taille_logement	Arithmetic	100	3.52	3.219568	3.820432
	Geometric	100	3.188579	2.906357	3.498208
	Harmonic	100	2.833116	2.548938	3.18861
Âge	Arithmetic	100	1.78	1.624043	1.935957
	Geometric	100	1.611821	1.473966	1.762569
	Harmonic	100	1.463415	1.350233	1.597306

25. Musangu M. (2018), *Elements de statistique descriptive*, G1 Faculté d'administration des affaires et sciences économiques, UPC, Kinshasa.

26. Exemple

27. Musangu M. (2018), *Elements de statistique descriptive*, G1 Faculté d'administration des affaires et sciences économiques, UPC, Kinshasa.

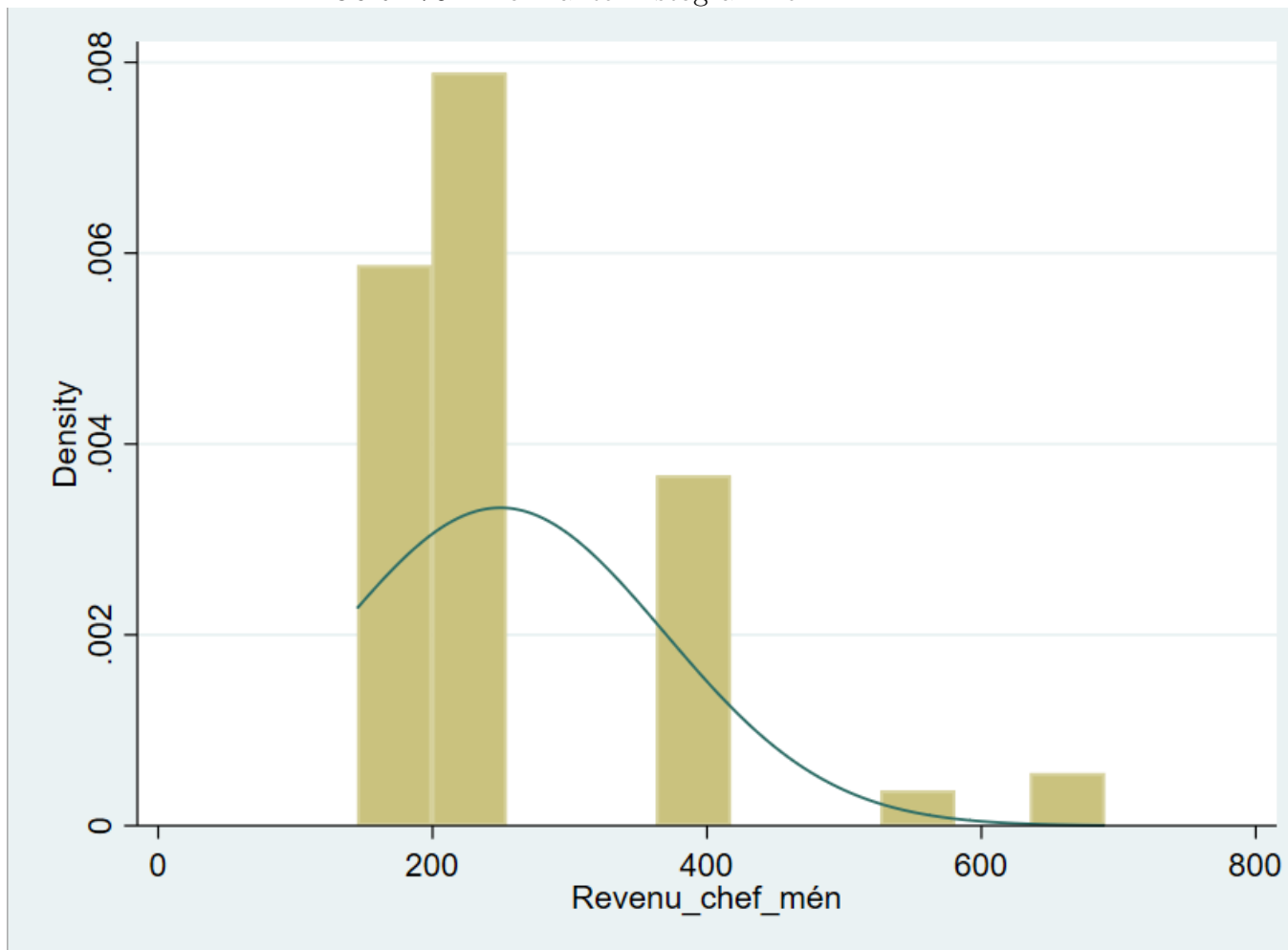
17 Appendice 11 : Analyse graphique de la distribution normale

L'on peut aussi étudier la distribution normale des variables grâce aux graphiques, avec l'approche des histogrammes et de la densité.

1. Histogramme

La commande est `histogram variable, normal`

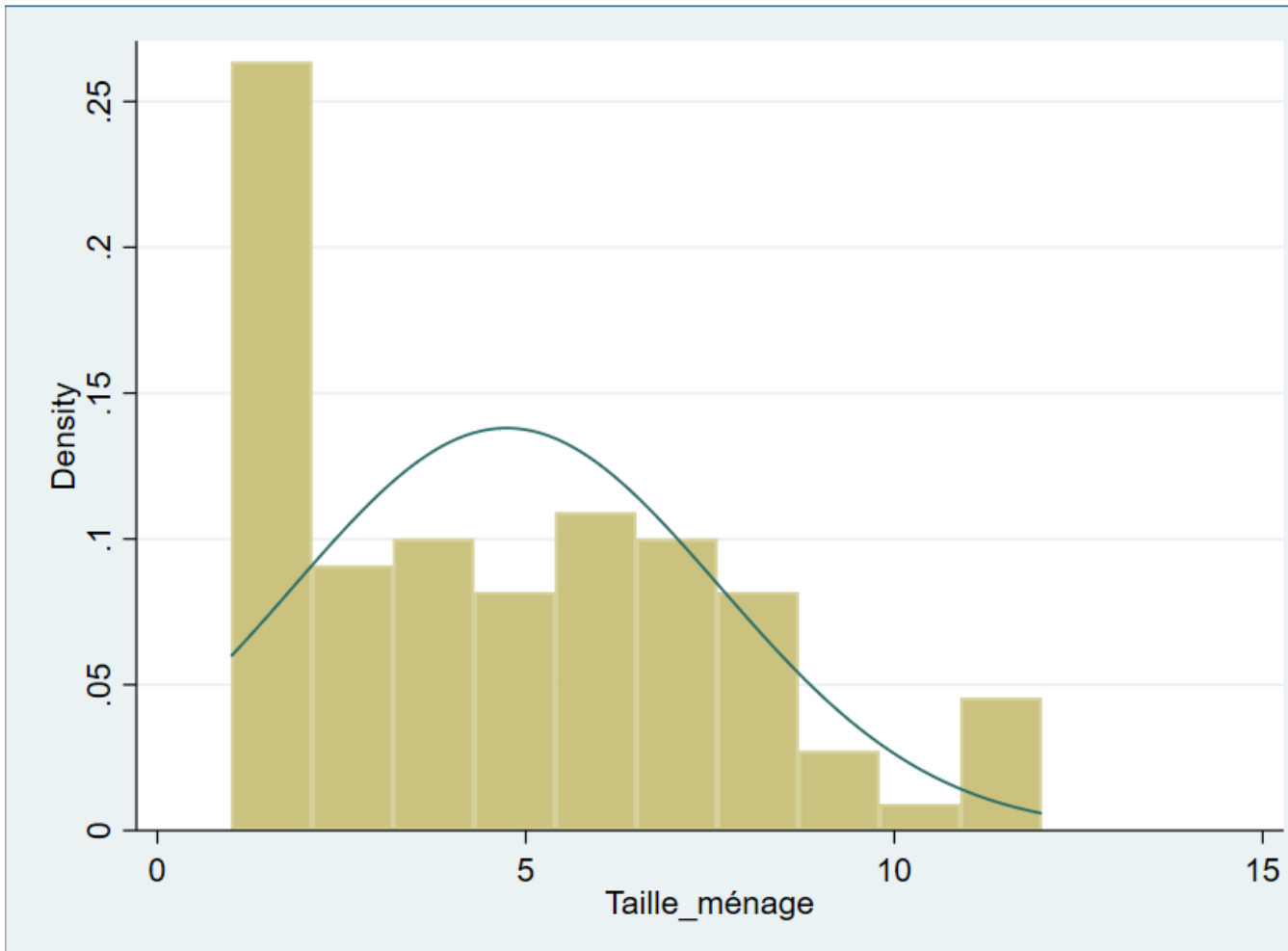
FIGURE 76 – Normalité Histogramme



Commentaire :

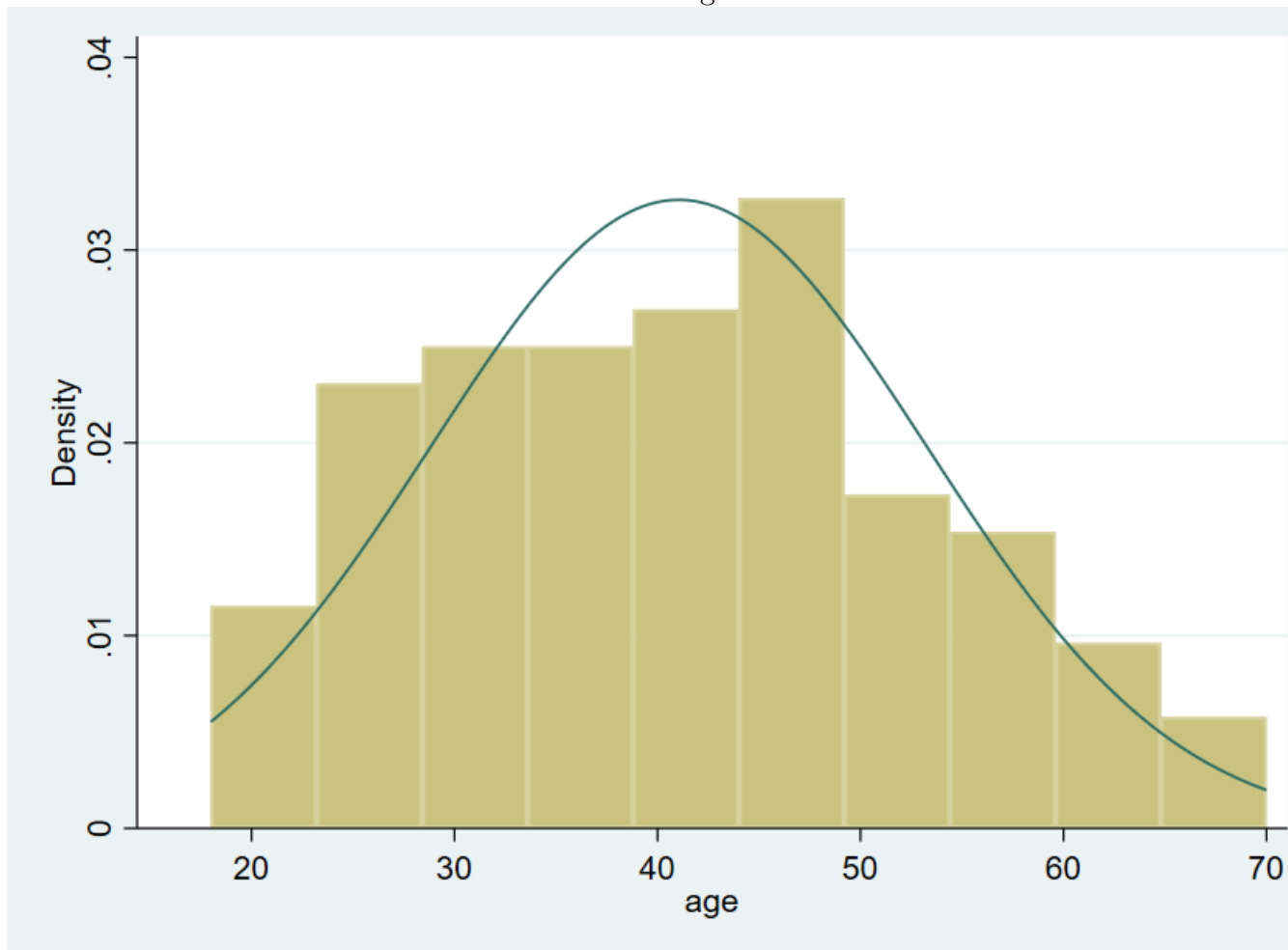
l'on voit que la variable revenu n'est pas normalement distribuée, car sa courbe ne ressemble pas à une cloche et elle a une inclinaison positive, car la plupart des éléments sont à gauche et elle a une queue droite longue.

FIGURE 77 – Normalité Histogramme



Commentaire :
cette variable est normalement distribuée, car sa courbe ressemble à une cloche et la majorité de ses valeurs, données sont concentrées ni à gauche, ni à droite mais au centre.

FIGURE 78 – Normalité Histogramme



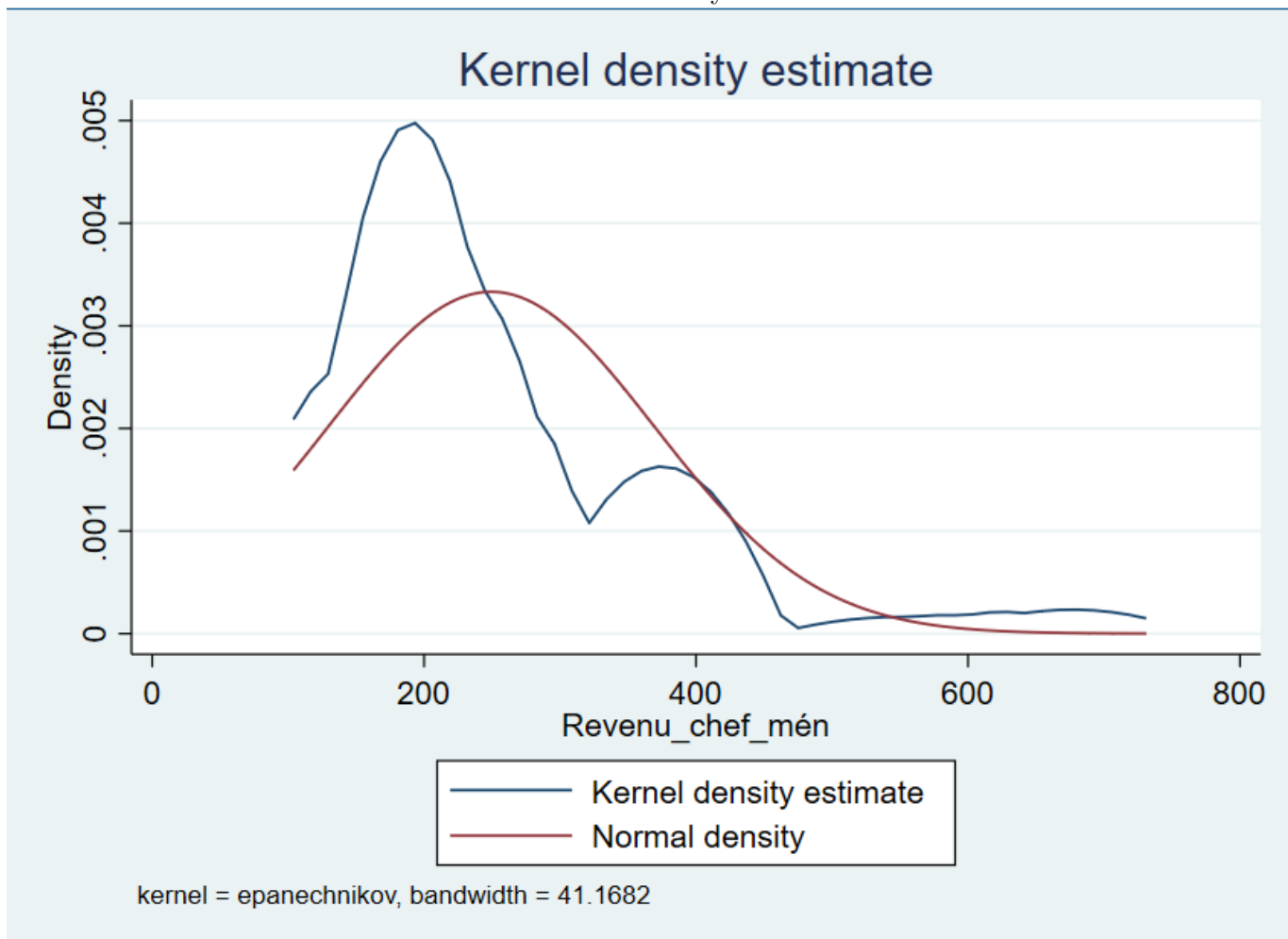
Commentaire :

Cette variable est normalement distribuée, car sa courbe ressemble à une cloche et la majorité de ses valeurs, données sont concentrées ni à gauche, ni à droite mais au centre.

2. Densité : l'on recourt à cette façon de faire lorsque l'histogramme n'est pas assez informatif, elle renseigne aussi sur la concentration des données. une densité plus élevée indique une concentration des données à cet endroit. Nos histogrammes sont très informatifs, nous implémentons donc la kernel density à titre d'exemple.

La commande est `kdensity variable,normal`

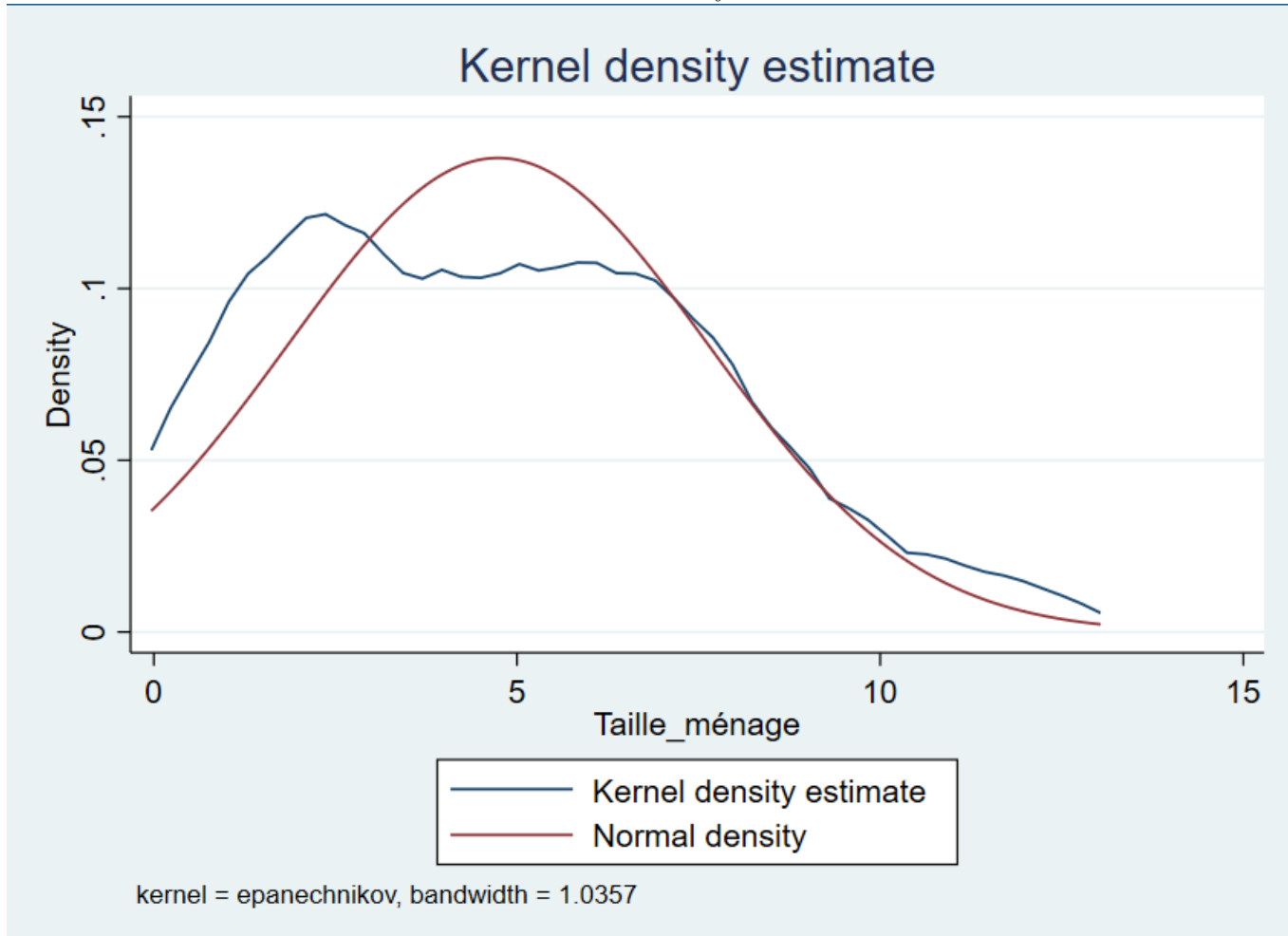
FIGURE 79 – Kernel density



Commentaire :

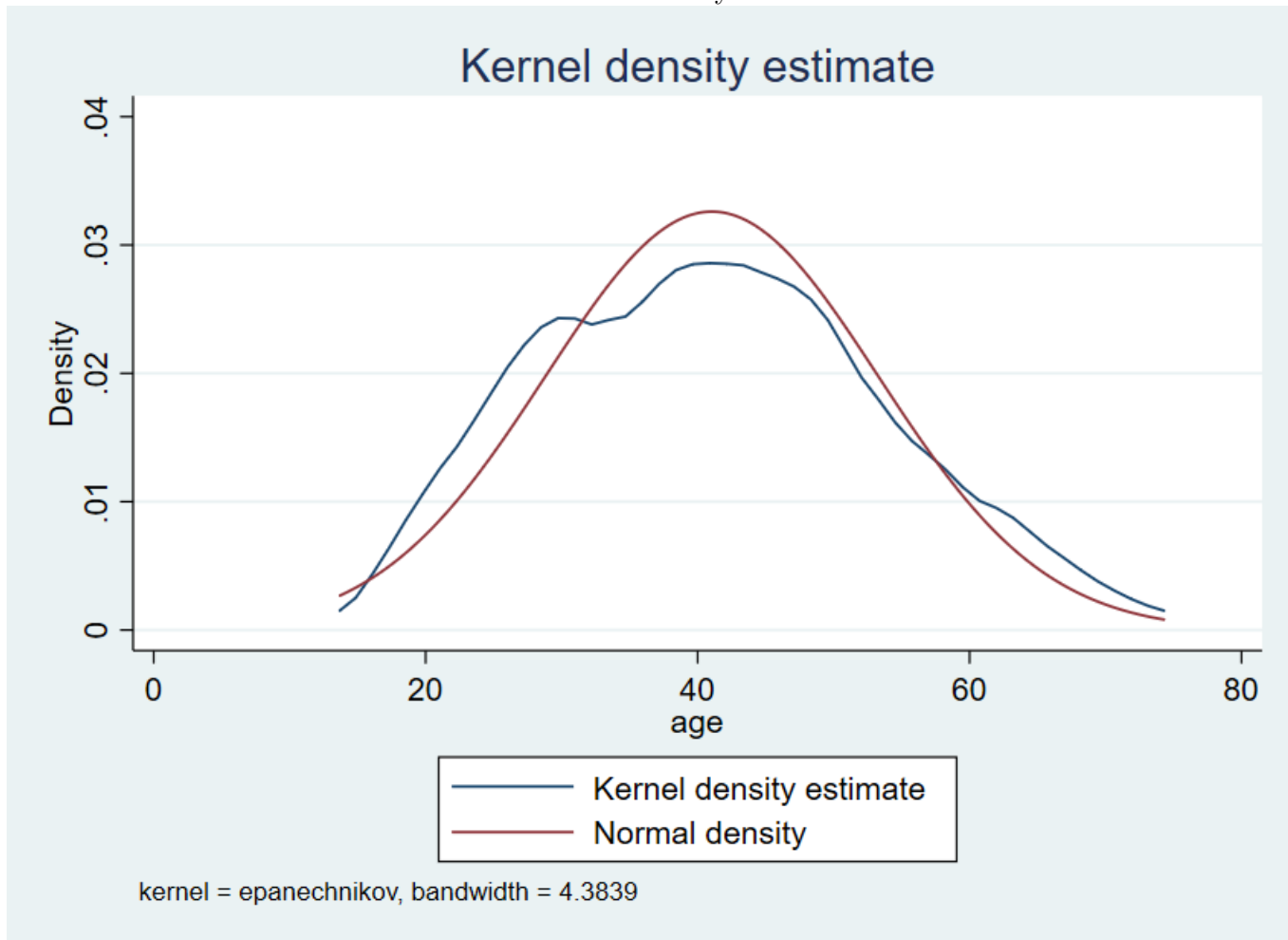
Les données sont concentrées à gauche et la variable n'est pas normalement distribuée.

FIGURE 80 – Kernel density



Commentaire :
Le lecteur peut s'exercer.

FIGURE 81 – Kernel density



Commentaire :
Le lecteur peut s'exercer.

18 Appendice 12 : Tab2,firstonly

La commande tab2 permet d'implémenter l'analyse de relation entre une variable, la dépendante et plusieurs variables indépendantes.

Illustration : l'on veut savoir si la variable pauvreté est en relation avec les variables statuts et professions.

La commande est alors `tab2 pauvreté statut profession,chi2 V28firstonly`

²⁹Commentaire :

FIGURE 82 – tab2,firstonly

```
. tab2 pauvreté profession statut,chi2 V firstonly
```

```
-> tabulation of pauvreté by profession
```

pauvreté	profession					Total
	1	2	3	4	5	
0	29	16	13	2	4	64
1	2	15	13	3	3	36
Total	31	31	26	5	7	100

```
Pearson chi2(4) = 17.4167 Pr = 0.002
```

```
Cramér's V = 0.4173
```

```
-> tabulation of pauvreté by statut
```

pauvreté	statut					Total
	1	2	3	4	5	
0	30	27	1	3	3	64
1	11	21	2	1	1	36
Total	41	48	3	4	4	100

```
Pearson chi2(4) = 4.3926 Pr = 0.355
```

```
Cramér's V = 0.2096
```

```
.
```

28. l'option firstonly fait comprendre au logiciel qu'uniquement la première variable citée qui fait office de la variable dépendante.

29. Le lecteur peut s'exercer

19 Appendice 13 : Tau de kendall's

le taux de kendall est un tel alternatif à celui du test de rang de spearman.

La commande est `ktau Revenu_chef_mén Taille_ménage,stats(tau obs p) star(0.05) pw matrix`

FIGURE 83 – Tau de kendall's

```
. ktau Revenu_chef_mén Taille_ménage, stats(taub obs p) star(0.05) pw matrix
```

Key
<i>tau_b</i>
<i>Number of obs</i>
<i>Sig. Level</i>

	Revenu	Taille
Revenu	1.0000 100	
Taille	0.1089 100 0.1786	1.0000 100

Commentaire :

La corrélation entre le revenu chef ménage et la taille de ménage n'est pas significative, car la probabilité est de 0.1786 qui est très largement supérieur au seuil de 0.05. et le 0.1089 est le coefficient informant de la direction et la force de la corrélation, ce coefficient est interpretable lorsque la corrélation est significative.

20 Appendice 14 : Test post hoc de Welch

A titre illustratif, nous allons reprendre l'exemple du test de student juste pour illustrer ce test. L'on ne devrait normalement pas prendre ce cas car l'homogénéité est respecté mais c'est juste pour pouvoir illustrer le test de welch.

La commande est `ttest Taille_ménage ,by(Compte_épargne)unequal`

FIGURE 84 – Post hoc Welch

```
. ttest Taille_ménage ,by( Compte_épargne )unequal
```

Two-sample t test with unequal variances

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
0	96	4.760417	.2968197	2.908227	4.171155	5.349678
1	4	4.25	1.376893	2.753785	-.1318869	8.631887
Combined	100	4.74	.2890668	2.890668	4.166429	5.313571
diff		.5104167	1.408522		-3.759902	4.780735

```
diff = mean(0) - mean(1) t = 0.3624
H0: diff = 0 Satterthwaite's degrees of freedom = 3.28508
```

```
Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.6304 Pr(|T| > |t|) = 0.7391 Pr(T > t) = 0.3696
```

³⁰Commentaire :

30. Le lecteur peut s'exercer en interprétant.

21 Appendice 15 : Test post hoc Dunnett's

A titre illustratif, nous allons reprendre l'exemple du test d'Anova juste pour illustrer ce test. L'on ne devrait normalement pas prendre ce cas car l'homogénéité est respecté mais c'est juste pour pouvoir illustrer le test de Dunnett's.

La commande est `ttest Taille ménage ,by(Compte épargne)unequal`

FIGURE 85 – Dunnett's

```
. pwmean Taille_ménage,over(niveau_instruction) mcompare(dunnett)effect
```

Pairwise comparisons of means with equal variances

Over: niveau_instruction

	Number of comparisons
niveau_instruction	3

Taille_ménage	Contrast	Std. err.	Dunnett t	P> t	Dunnett [95% conf. interval]
niveau_instruction					
2 vs 1	-1.229474	.7360454	-1.67	0.231	-2.986244 .527297
3 vs 1	-1.589474	1.440429	-1.10	0.554	-5.027446 1.848499
5 vs 1	-4.789474	2.940264	-1.63	0.249	-11.80721 2.228258

Commentaire :

Étant donné que les p-values, les probabilités sont toutes supérieures au seuil de 0.05, donc il n'existe pas des différences significatives.

22 Appendice 16 : Ologit simple

Il peut y'arriver que l'on soit en présence uniquement de deux variables et que l'on veule analyser l'impact de x sur y.

Illustration : l'on voudrait savoir si l'accès au crédit des chefs de ménage a une influence sur les niveaux d'instruction.

La commande est `ologit niveau_instruction accès_crédits,ro`

FIGURE 86 – Ologit simple

```

. ologit niveau_instruction Accès_crédits ,ro

Iteration 0:  log pseudolikelihood = -72.71388
Iteration 1:  log pseudolikelihood = -67.952999
Iteration 2:  log pseudolikelihood = -67.641062
Iteration 3:  log pseudolikelihood = -67.64035
Iteration 4:  log pseudolikelihood = -67.64035

Ordered logistic regression                                Number of obs =   100
                                                           Wald chi2(1) =    7.80
                                                           Prob > chi2   =  0.0052
Log pseudolikelihood = -67.64035                          Pseudo R2      =  0.0698
    
```

niveau_instruction	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
Accès_crédits	1.687472	.6040253	2.79	0.005	.5036044	2.87134
/cut1	-.9681736	.2843728			-1.525534	-.4108132
/cut2	3.754699	.5835381			2.610985	4.898413
/cut3	5.633495	1.293955			3.09739	8.1696

³¹Commentaire :

31. Le lecteur peut s'exercer

FIGURE 87 – Effets Marginaux

```
. margins,dydx(*)
```

Average marginal effects
Model VCE: Robust

Number of obs = 100

dy/dx wrt: Accès_crédits

```
1._predict: Pr(niveau_instruction==1), predict(pr outcome(1))
2._predict: Pr(niveau_instruction==2), predict(pr outcome(2))
3._predict: Pr(niveau_instruction==3), predict(pr outcome(3))
4._predict: Pr(niveau_instruction==5), predict(pr outcome(5))
```

	Delta-method				
	dy/dx	std. err.	z	P> z	[95% conf. interval]
Accès_crédits					
_predict					
1	-.2410431	.0836946	-2.88	0.004	-.4050815 -.0770047
2	.1498088	.0634333	2.36	0.018	.0254819 .2741357
3	.0748271	.0452737	1.65	0.098	-.0139078 .1635619
4	.0164073	.0145738	1.13	0.260	-.0121568 .0449713

³²Commentaire :

32. Le lecteur peut s'exercer

FIGURE 88 – Spécification du test

`. linktest`

note: `_hatsq` omitted because of collinearity.

Iteration 0: log likelihood = -72.71388

Iteration 1: log likelihood = -67.952999

Iteration 2: log likelihood = -67.641062

Iteration 3: log likelihood = -67.64035

Iteration 4: log likelihood = -67.64035

Ordered logistic regression

Number of obs = 100

LR chi2(1) = 10.15

Prob > chi2 = 0.0014

Pseudo R2 = 0.0698

Log likelihood = -67.64035

niveau_instruction	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
<code>_hat</code>	1	.3511032	2.85	0.004	.3118504	1.68815
<code>_hatsq</code>	0	(omitted)				
<code>/cut1</code>	-.9681736	.2860748			-1.52887	-.4074772
<code>/cut2</code>	3.754699	.6075907			2.563843	4.945555
<code>/cut3</code>	5.633495	1.100729			3.476106	7.790884

³³Commentaire :

33. Le lecteur peut s'exercer

23 Appendice 17 : Mlogit simple

Illustration : l'on voudrait savoir si l'âge du chef de ménage a une influence sur le mode d'accès aux soins de santé.

La commande est `mlogit accès soins de santé âge,ro`

FIGURE 89 – Mlogit simple

```

Iteration 0:  log pseudolikelihood = -26.515837
Iteration 1:  log pseudolikelihood = -25.291945
Iteration 2:  log pseudolikelihood = -24.189157
Iteration 3:  log pseudolikelihood = -24.063205
Iteration 4:  log pseudolikelihood = -24.051533
Iteration 5:  log pseudolikelihood = -24.049395
Iteration 6:  log pseudolikelihood = -24.048915
Iteration 7:  log pseudolikelihood = -24.0488
Iteration 8:  log pseudolikelihood = -24.048775
Iteration 9:  log pseudolikelihood = -24.04877
Iteration 10: log pseudolikelihood = -24.048769

Multinomial logistic regression               Number of obs =   100
                                             Wald chi2(2)  = 325.11
                                             Prob > chi2   = 0.0000
Log pseudolikelihood = -24.048769           Pseudo R2    = 0.0930

```

Accès_soins_santé	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
1	(base outcome)					
2						
Revenu_chef_mén	-.0027094	.0055585	-0.49	0.626	-.0136039	.0081851
_cons	-2.516755	1.342281	-1.87	0.061	-5.147578	.1140675
3						
Revenu_chef_mén	-.199935	.0110889	-18.03	0.000	-.2216689	-.1782011
_cons	26.33546	2.250471	11.70	0.000	21.92462	30.7463

³⁴Commentaire :

34. Le lecteur peut s'exercer

FIGURE 90 – Effets Marginaux

```
. margins,dydx(*)
```

Average marginal effects Number of obs = 100
 Model VCE: Robust

dy/dx wrt: Revenu_chef_mén

1._predict: Pr(Accès_soins_santé==1), predict(pr outcome(1))
 2._predict: Pr(Accès_soins_santé==2), predict(pr outcome(2))
 3._predict: Pr(Accès_soins_santé==3), predict(pr outcome(3))

	Delta-method					[95% conf. interval]
	dy/dx	std. err.	z	P> z		
Revenu_chef_mén						
_predict						
1	.0036558	.0024719	1.48	0.139	-.0011891	.0085007
2	.0000899	.0001957	0.46	0.646	-.0002937	.0004735
3	-.0037457	.0026051	-1.44	0.150	-.0088517	.0013603

³⁵Commentaire :

35. Le lecteur peut s'exercer

24 Appendice 18 : Oprobit simple

Illustration : l'on voudrait savoir si l'accès au crédit des chefs de ménage a une influence sur les niveaux d'instruction.

La commande est `oprobit niveau_instruction accès_crédits,ro`

FIGURE 91 – Oprobit Simple

```

. oprobit niveau_instruction Accès_crédits ,ro

Iteration 0:  log pseudolikelihood = -72.71388
Iteration 1:  log pseudolikelihood = -68.182071
Iteration 2:  log pseudolikelihood = -68.102093
Iteration 3:  log pseudolikelihood = -68.101888
Iteration 4:  log pseudolikelihood = -68.101888

Ordered probit regression                                Number of obs =   100
                                                         Wald chi2(1)    =   7.98
                                                         Prob > chi2     =  0.0047
Log pseudolikelihood = -68.101888                       Pseudo R2      =  0.0634
    
```

niveau_instruction	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
Accès_crédits	.819707	.2902074	2.82	0.005	.2509109	1.388503
/cut1	-.6163828	.1706717			-.9508933	-.2818724
/cut2	2.034285	.287805			1.470197	2.598372
/cut3	2.808443	.556019			1.718665	3.89822

³⁶Commentaire :

36. Le lecteur peut s'exercer

FIGURE 92 – Effets Marginaux

```
. margins,dydx(*)
```

Average marginal effects

Number of obs = 100

Model VCE: Robust

dy/dx wrt: Accès_crédits

1._predict: Pr(niveau_instruction==1), predict(pr outcome(1))

2._predict: Pr(niveau_instruction==2), predict(pr outcome(2))

3._predict: Pr(niveau_instruction==3), predict(pr outcome(3))

4._predict: Pr(niveau_instruction==5), predict(pr outcome(5))

	Delta-method				[95% conf. interval]	
	dy/dx	std. err.	z	P> z		
Accès_crédits						
_predict						
1	-.2073686	.071547	-2.90	0.004	-.3475981	-.0671391
2	.1188784	.054064	2.20	0.028	.0129148	.224842
3	.0661941	.0376629	1.76	0.079	-.0076238	.140012
4	.0222961	.0165606	1.35	0.178	-.0101621	.0547543

³⁷Commentaire :

37. Le lecteur peut s'exercer

FIGURE 93 – Spécification du modèle

```

. linktest

note: _hatsq omitted because of collinearity.
Iteration 0: log likelihood = -72.71388
Iteration 1: log likelihood = -68.182071
Iteration 2: log likelihood = -68.102093
Iteration 3: log likelihood = -68.101888
Iteration 4: log likelihood = -68.101888

Ordered probit regression                                Number of obs =   100
                                                         LR chi2(1)      =    9.22
                                                         Prob > chi2    =  0.0024
Log likelihood = -68.101888                             Pseudo R2      =  0.0634

```

niveau_instruction	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
_hat	1	.3444985	2.90	0.004	.3247954	1.675205
_hatsq	0	(omitted)				
/cut1	-.6163828	.168554			-.9467425	-.2860231
/cut2	2.034285	.282663			1.480275	2.588294
/cut3	2.808443	.4108911			2.003111	3.613774

³⁸Commentaire :

38. Le lecteur peut s'exercer

25 Appendice 19 : Oprobit Multiple

Illustration : On veut voir dans cette base des données, quelles sont les variables qui peuvent expliquer le niveau d'instruction du chef de ménage, cette illustration est juste à titre exemplatif afin de garder la même base des données.

Première étape : estimer le modèle Oprobit.

La commande est `oprobit variable dépendante variable indépendante,ro`

FIGURE 94 – Oprobit

```
. oprobit niveau_instruction profession statut sexe Revenu_chef_mén pauvreté nombre_enfants emploi Accès_crédits Cor
```

```
Iteration 0:  log pseudolikelihood = -72.71388
Iteration 1:  log pseudolikelihood = -60.897844
Iteration 2:  log pseudolikelihood = -60.43569
Iteration 3:  log pseudolikelihood = -60.434718
Iteration 4:  log pseudolikelihood = -60.434718
```

```
Ordered probit regression                                Number of obs =   100
                                                         Wald chi2(10) =  28.31
                                                         Prob > chi2   = 0.0016
Log pseudolikelihood = -60.434718                       Pseudo R2      = 0.1689
```

niveau_instruction	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
profession	.2773423	.1311365	2.11	0.034	.0203195	.5343651
statut	.1423183	.166824	0.85	0.394	-.1846507	.4692873
sexe	.4039936	.2845087	1.42	0.156	-.1536333	.9616204
Revenu_chef_mén	.001525	.0013138	1.16	0.246	-.0010501	.0041
pauvreté	-.2734736	.3878276	-0.71	0.481	-1.033602	.4866545
nombre_enfants	-.1179118	.1121638	-1.05	0.293	-.3377487	.1019252
emploi	-.2441985	.1616031	-1.51	0.131	-.5609348	.0725379
Accès_crédits	.9941174	.287605	3.46	0.001	.430422	1.557813
Compte_épargne	.6124123	.6941304	0.88	0.378	-.7480582	1.972883
age	-.0073081	.0119342	-0.61	0.540	-.0306987	.0160825
/cut1	-.0994087	.7880923			-1.644041	1.445224
/cut2	2.91194	.8211621			1.302492	4.521388
/cut3	3.818602	.8924215			2.069488	5.567716

³⁹Commentaire :

39. Le lecteur peut s'exercer

FIGURE 95 – Effets Marginaux

```
. margins,dydx(*)
```

Average marginal effects Number of obs = 100
 Model VCE: Robust

dy/dx wrt: **profession statut sexe Revenu_chef_mén pauvreté nombre_enfants emploi Accès_crédits Compte_épargn**

1._predict: Pr(niveau_instruction==1), predict(pr outcome(1))
 2._predict: Pr(niveau_instruction==2), predict(pr outcome(2))
 3._predict: Pr(niveau_instruction==3), predict(pr outcome(3))
 4._predict: Pr(niveau_instruction==5), predict(pr outcome(5))

	Delta-method					
	dy/dx	std. err.	z	P> z	[95% conf. interval]	
profession						
_predict						
1	-.0613314	.0277677	-2.21	0.027	-.1157551	-.0069077
2	.0342298	.0176233	1.94	0.052	-.0003112	.0687708
3	.0205761	.0100805	2.04	0.041	.0008187	.0403335
4	.0065255	.0067801	0.96	0.336	-.0067632	.0198142
statut						
_predict						
1	-.0314722	.0363497	-0.87	0.387	-.1027164	.039772
2	.017565	.0196937	0.89	0.372	-.0210339	.0561639
3	.0105586	.0132019	0.80	0.424	-.0153167	.0364339
4	.0033486	.0050942	0.66	0.511	-.0066359	.013333

⁴⁰Commentaire :

40. Le lecteur peut s'exercer

FIGURE 96 – Effets Marginaux suite

sexe							
_predict							
1	-.089339	.0652632	-1.37	0.171	-.2172525	.0385744	
2	.0498611	.0407314	1.22	0.221	-.0299709	.1296932	
3	.0299724	.0217471	1.38	0.168	-.0126511	.0725958	
4	.0095055	.0102173	0.93	0.352	-.0105201	.0295311	
Revenu_chef_mén							
_predict							
1	-.0003372	.0002925	-1.15	0.249	-.0009105	.0002361	
2	.0001882	.0001783	1.06	0.291	-.0001612	.0005376	
3	.0001131	.0001057	1.07	0.284	-.000094	.0003203	
4	.0000359	.0000343	1.05	0.295	-.0000313	.0001031	
pauvreté							
_predict							
1	.0604759	.0852807	0.71	0.478	-.1066713	.2276231	
2	-.0337523	.0481261	-0.70	0.483	-.1280778	.0605732	
3	-.0202891	.0274955	-0.74	0.461	-.0741792	.0336011	
4	-.0064345	.0120972	-0.53	0.595	-.0301445	.0172755	
nombre_enfants							
_predict							
1	.026075	.0246146	1.06	0.289	-.0221687	.0743186	
2	-.0145527	.0139386	-1.04	0.296	-.041872	.0127665	
3	-.0087479	.0092483	-0.95	0.344	-.0268743	.0093785	
4	-.0027743	.0032105	-0.86	0.388	-.0090669	.0035182	

⁴¹Commentaire :

41. Le lecteur peut s'exercer

FIGURE 97 – Effets Marginaux suite

nombre_enfants						
_predict						
1	.026075	.0246146	1.06	0.289	-.0221687	.0743186
2	-.0145527	.0139386	-1.04	0.296	-.041872	.0127665
3	-.0087479	.0092483	-0.95	0.344	-.0268743	.0093785
4	-.0027743	.0032105	-0.86	0.388	-.0090669	.0035182
emploi						
_predict						
1	.054002	.0353427	1.53	0.127	-.0152684	.1232723
2	-.0301391	.0215614	-1.40	0.162	-.0723988	.0121205
3	-.0181172	.0134848	-1.34	0.179	-.0445468	.0083125
4	-.0057457	.0054342	-1.06	0.290	-.0163965	.0049052
Accès_crédits						
_predict						
1	-.2198388	.0606344	-3.63	0.000	-.3386801	-.1009975
2	.1226946	.0449102	2.73	0.006	.0346721	.2107171
3	.0737538	.037718	1.96	0.051	-.0001721	.1476798
4	.0233904	.0163359	1.43	0.152	-.0086274	.0554081
Compte_épargne						
_predict						
1	-.1354287	.1524334	-0.89	0.374	-.4341926	.1633353
2	.0755843	.0865208	0.87	0.382	-.0939932	.2451619
3	.045435	.0558086	0.81	0.416	-.0639478	.1548179
4	.0144093	.018093	0.80	0.426	-.0210524	.049871

⁴²Commentaire :

42. Le lecteur peut s'exercer

FIGURE 98 – Effets Marginaux fin

age	_predict						
	1	.0016161	.0026185	0.62	0.537	-.003516	.0067482
	2	-.000902	.0014747	-0.61	0.541	-.0037923	.0019883
	3	-.0005422	.0008771	-0.62	0.536	-.0022613	.0011769
	4	-.0001719	.0003308	-0.52	0.603	-.0008202	.0004763

.

⁴³Commentaire :

⁴⁴Commentaire :

43. Le lecteur peut s'exercer

44. Le lecteur peut s'exercer

FIGURE 99 – Spécification du modèle

. linktest

Iteration 0: log likelihood = -72.71388
 Iteration 1: log likelihood = -60.611079
 Iteration 2: log likelihood = -60.160084
 Iteration 3: log likelihood = -60.159537
 Iteration 4: log likelihood = -60.159537

Ordered probit regression

Number of obs = 100
 LR chi2(2) = 25.11
 Prob > chi2 = 0.0000
 Pseudo R2 = 0.1727

Log likelihood = -60.159537

niveau_instruction	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
_hat	1.334828	.5374495	2.48	0.013	.2814458	2.388209
_hatsq	-.1682854	.2367971	-0.71	0.477	-.6323992	.2958284
/cut1	-.0085623	.2664244			-.5307446	.5136199
/cut2	2.969373	.4290032			2.128542	3.810204
/cut3	3.868726	.5674738			2.756498	4.980954

26 Appendice 20 : Mprobit Simple

Illustration : l'on voudrait savoir si l'âge du chef de ménage a une influence sur le mode d'accès aux soins de santé.

La commande est `mprobit accès soins de santé âge,ro`

FIGURE 100 – Mprobit Simple

Multinomial probit regression		Number of obs = 100				
Log pseudolikelihood = -24.104225		Wald chi2(2) = 2254.76				
		Prob > chi2 = 0.0000				
Accès_soins_s~é	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
1	(base outcome)					
2						
Revenu_chef_mén	-.0019338	.0033148	-0.58	0.560	-.0084307	.0045631
_cons	-1.972876	.8171684	-2.41	0.016	-3.574496	-.3712549
3						
Revenu_chef_mén	-.1559864	.0033053	-47.19	0.000	-.1624647	-.1495081
_cons	20.51951

⁴⁵Commentaire :

45. Le lecteur peut s'exercer

FIGURE 101 – Effets Marginaux

```
. margins, dydx(*)
```

```
Average marginal effects
Model VCE: Robust
```

Number of obs = 100

```
dy/dx wrt: Revenu_chef_mén
```

```
1._predict: Pr(Accès_soins_santé==1), predict(pr outcome(1))
2._predict: Pr(Accès_soins_santé==2), predict(pr outcome(2))
3._predict: Pr(Accès_soins_santé==3), predict(pr outcome(3))
```

	Delta-method				
	dy/dx	std. err.	z	P> z	[95% conf. interval]
Revenu_chef_mén					
_predict					
1	.0040339	.0020886	1.93	0.053	-.0000597 .0081276
2	.0004092	.0002444	1.67	0.094	-.0000698 .0008883
3	-.0044432	.0022847	-1.94	0.052	-.008921 .0000347

⁴⁶Commentaire :

46. Le lecteur peut s'exercer

27 Appendice 21 : Mprobit Multiple

Illustration : On veut voir dans cette base des données, quelles sont les variables qui peuvent expliquer l'accès aux soins de santé du chef de ménage, cette illustration est juste à titre exemplatif afin de garder la même base des données.

Première étape : estimer le modèle mprobit.

La commande est `mprobit variable dépendante variable indépendante,ro`

FIGURE 102 – Mprobit

Multinomial probit regression		Number of obs = 100				
Log pseudolikelihood = -11.995569		Wald chi2(10) = .				
		Prob > chi2 = .				
Accès_soins_santé	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
1	(base outcome)					
2						
sexe	-.0724244	.6514595	-0.11	0.911	-1.349262	1.204413
statut	.1355522	.3537666	0.38	0.702	-.5578175	.8289219
profession	.3105359	.3491865	0.89	0.374	-.3738569	.9949288
niveau_instruction	-.5769205	.5653733	-1.02	0.308	-1.685032	.5311908
Revenu_chef_mén	.0025657	.0037585	0.68	0.495	-.0048008	.0099322
pauvreté	.5825477	.7778017	0.75	0.454	-.9419156	2.107011
nombre_enfants	-1.038952	.3940367	-2.64	0.008	-1.81125	-.2666539
emploi	.0689574	.233894	0.29	0.768	-.3894665	.5273813
Accès_crédits	-.3806147	.435604	-0.87	0.382	-1.234383	.4731534
Compte_épargne	-22.63579
age	-.0886901	.030334	-2.92	0.003	-.1481436	-.0292366
_cons	1.36867	1.450985	0.94	0.346	-1.475208	4.212548
3						
sexe	11.42006
statut	6.688565
profession	19.63954
niveau_instruction	-9.890917
Revenu_chef_mén	-.3429548
pauvreté	5.784183

FIGURE 103 – Mprobit suite

3						
	sexe	11.42006
	statut	6.688565
	profession	19.63954
	niveau_instruction	-9.890917
	Revenu_chef_mén	-.3429548
	pauvreté	5.784183
	nombre_enfants	-34.83356
	emploi	5.219533
	Accès_crédits	23.85969
	Compte_épargne	94.01332
	age	-2.823437
	_cons	77.56453

Commentaire :

Cette illustration n'est pas adaptée pour un modèle multinomial probit.

Références

- [1] Anderson, Sweeney et Williams., (2010), *Statistiques pour l'économie et la gestion*, édition 3, de boeck.
- [2] Anderson, Sweeney, Williams et al., (2015), *Statistiques pour l'économie et la gestion*, édition 5, de boeck.
- [3] Muayila P. (2019), *Théorie et pratique de sondage*, L1 Faculté d'administration des affaires et sciences économiques, UPC, Kinshasa.
- [4] Muayila P. (2023), *Introduction aux techniques d'enquêtes*, L2 LMD Faculté d'administration des affaires et sciences économiques, UPC, Kinshasa.
- [5] Musangu M. (2018), *Elements de statistique descriptive*, G1 Faculté d'administration des affaires et sciences économiques, UPC, Kinshasa.
- [6] Nkongo L. et Zumbu D. (2020), *Satistique appliquée aux affaires*, L1 Faculté d'administration des affaires et sciences économiques, UPC, Kinshasa.
- [7] Observatoire Economique et Statistique d'Afrique Subsaharienne .,(2012),*Manuel pratique d'introduction au logiciel stata.*,Afrisat.
- [8] Olivier M., (2009), *l'enquête et ses méthodes*, édition 2, Armand colin.
- [9] Omonga K. (2017), *Statistiques inférencielles*, G2 Faculté d'administration des affaires et sciences économiques, UPC, Kinshasa.
- [10] Regis bourbonnais.,(2015),*Econométrie*,édition 9, Dunod.
- [11] Rohit Dass A., (2022), *Introduction to stata*, Institute of Health Policy Management, and Evaluation Canadian Centre for Health Economics University of Toronto.
- [12] Tremblay M.,Lavallé P., et Tirari M.,(2011),*Pratique et méthode de sondage*,Dunod.
- [13] [https ://datatab.fr/tutorial/pearson-correlation](https://datatab.fr/tutorial/pearson-correlation), consulté le 10/12/2023, à 7h00.