

```
. *CORRIGE DE L'EXERCICE CHAPITRE 12-EXERCICE 1
. *Dalila Chenaf-Nicet  Université de Bordeaux.
.
. * Tout d'abord s'assurer d'avoir ouvert le fichier de données stata C12EX1 avant d'ouvrir
le do.fileC12EX1 qui est le fichier programme.
.
. * Une fois le fichier de données ouvert ainsi que le do-file appuyer sur Run (Execute en
haut à droite de la barre de menu du dofile)
. * pour démarrer le programme. Il sera exécuté dans son intégralité.
.
. * Toutefois il est possible en sélectionnant les parties du programme de l'exécuter pas à
pas afin de voir apparaître pas à pas les
. * différents résultats.
.
.
. *
. * CHAPITRE 12 EXERCICE 2
. * Estimation de modèle logit probit binaire
.
.
. * Question 1 : Estimation logit du modèle
.
. logit reussite genre necono nenfants nmicro
```

```
Iteration 0:  log likelihood = -41.555491
Iteration 1:  log likelihood = -15.493662
Iteration 2:  log likelihood = -15.051909
Iteration 3:  log likelihood = -15.037515
Iteration 4:  log likelihood = -15.037494
Iteration 5:  log likelihood = -15.037494
```

```
Logistic regression              Number of obs   =           60
                                LR chi2(4)        =          53.04
                                Prob > chi2         =          0.0000
Log likelihood = -15.037494      Pseudo R2       =          0.6381
```

reussite	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
genre	-3.761718	1.437072	-2.62	0.009	-6.578326	-.9451089
necono	.6320616	.2395646	2.64	0.008	.1625237	1.1016
nenfants	-.682523	.3788703	-1.80	0.072	-1.425095	.0600492
nmicro	.1553216	.188916	0.82	0.411	-.214947	.5255902
_cons	-3.265634	2.020062	-1.62	0.106	-7.224882	.6936146

```
.
. * A la lecture du résultat nous constatons que la variable micro n'est pas significative
(prob 0.41), que nenfant à une
. * prob de 0.07 et donc que la variable est faiblement significative. Seules genre est
necono sont significatives
.
. * Nous refaisons une estimation en enlevant les variables non significatives. Nous
utilisons "fitstat" pour obtenir un certain
. * nombre de statistiques utiles à l'analyse du Logit
.
. logit reussite genre necono nenfants
```

```
Iteration 0:  log likelihood = -41.555491
Iteration 1:  log likelihood = -15.754671
Iteration 2:  log likelihood = -15.388903
Iteration 3:  log likelihood = -15.376703
Iteration 4:  log likelihood = -15.376685
Iteration 5:  log likelihood = -15.376685
```

```
Logistic regression              Number of obs   =           60
                                LR chi2(3)        =          52.36
                                Prob > chi2         =          0.0000
Log likelihood = -15.376685      Pseudo R2       =          0.6300
```

reussite	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
----------	-------	-----------	---	------	----------------------	--

genre	-3.634605	1.410947	-2.58	0.010	-6.400011	-.8691996
necono	.6958568	.2317894	3.00	0.003	.2415579	1.150156
nenfants	-.7467423	.3789427	-1.97	0.049	-1.489456	-.0040283
_cons	-2.859277	1.910378	-1.50	0.134	-6.60355	.8849959

. fitstat

Measures of Fit for logit of reussite

Log-Lik Intercept Only:	-41.555	Log-Lik Full Model:	-15.377
D(56):	30.753	LR(3):	52.358
		Prob > LR:	0.000
McFadden's R2:	0.630	McFadden's Adj R2:	0.534
Maximum Likelihood R2:	0.582	Cragg & Uhler's R2:	0.776
McKelvey and Zavoina's R2:	0.842	Efron's R2:	0.692
Variance of y*:	20.774	Variance of error:	3.290
Count R2:	0.883	Adj Count R2:	0.759
AIC:	0.646	AIC*n:	38.753
BIC:	-198.530	BIC':	-40.075

. * Log-lik intercept (LR) est le log Likelihood ; Log-lik full model (LU) est le Restr. Log Likelihood.

. * On peut aussi calculer avec $LU/n = -0.25$ le Avg. Log Likelihood qui n'est pas donné par Stata mais que l'on calcule aisément.

. * LR(3df) est la LR statistic. Vous avez aussi les critères d'information (AIC, BIC)

. * a) Interprétation statistique de la régression logit

. ***** Les coefficients sont tous significativement différents de zéro, hormis le terme constant. La statistique de la

. * Log vraisemblance LR = 53,35 que l'on compare à un Khi2 à 3ddl, nous rejetons H0, le modèle comporte au moins une variable explicative

. * Le pseudo R2=MacFadden= $0.63 = 1 - (LU/LR) = 1 - (-15.377)/(-41.55) = 1 - 0.37 = 0.63$. Le modèle est validé

. * Le logiciel Stata propose une table permettant d'appréhender les qualités prévisionnelles du modèle sur l'échantillon

. * en comparant la probabilité estimée pour un individu i d'être = 1 ($P(dep=1)$) au seuil arbitraire de 50% à la valeur observée

. * des $y_i=0$ ou des $y_i=1$

. logit reussite genre necono nenfants

Iteration 0: log likelihood = -41.555491
 Iteration 1: log likelihood = -15.754671
 Iteration 2: log likelihood = -15.388903
 Iteration 3: log likelihood = -15.376703
 Iteration 4: log likelihood = -15.376685
 Iteration 5: log likelihood = -15.376685

Logistic regression	Number of obs	=	60
	LR chi2(3)	=	52.36
	Prob > chi2	=	0.0000
Log likelihood = -15.376685	Pseudo R2	=	0.6300

reussite	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
genre	-3.634605	1.410947	-2.58	0.010	-6.400011 - .8691996
necono	.6958568	.2317894	3.00	0.003	.2415579 1.150156
nenfants	-.7467423	.3789427	-1.97	0.049	-1.489456 -.0040283
_cons	-2.859277	1.910378	-1.50	0.134	-6.60355 .8849959

. estat classification

Logistic model for reussite

----- True -----			
Classified	D	~D	Total
+	27	3	30
-	4	26	30
Total	31	29	60

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

True D defined as reussite != 0

Sensitivity	$\Pr(+ D)$	87.10%
Specificity	$\Pr(- \sim D)$	89.66%
Positive predictive value	$\Pr(D +)$	90.00%
Negative predictive value	$\Pr(\sim D -)$	86.67%
False + rate for true ~D	$\Pr(+ \sim D)$	10.34%
False - rate for true D	$\Pr(- D)$	12.90%
False + rate for classified +	$\Pr(\sim D +)$	10.00%
False - rate for classified -	$\Pr(D -)$	13.33%
Correctly classified		88.33%

.
 . * Dans l'échantillon on a 29 individus qui ont raté leur examen ($y_i=0$), le modèle indique que 26 individus ont une probabilité estimée de réussite inférieure à 50%. Donc dans 89,66% des cas les échecs ont été correctement prévus par le modèle.
 . * Pour les individus (31 au total) pour lesquels $y_i=1$, le modèle indique que 27 individus ont une probabilité estimée de réussite à l'examen supérieure à 50%. Donc dans 87,1% des cas les réussites sont correctement prévues.

. * Dans ce modèle le taux d'erreur de prévision est faible

. * b) Interprétation économique.

. * Le modèle s'écrit (en arrondissant les coefficients estimés)

. * $\ln(\pi/(1-\pi)) = -3.63 \text{ genre} + 0.70 \text{ necono} - 0.75 \text{ nenfants} - 2.86 + \epsilon_i$

. * Le nombre de frères et sœurs agit négativement sur le taux de réussite. Les étudiants issus de familles nombreuses ont un taux de réussite plus faible.

. * La note d'économétrie est un facteur positif de réussite. Enfin les garçons réussissent en général moins bien que les filles.

. * Question 2 : Estimation d'un modèle Probit et du modèle linéaire général.

. * Le modèle probit

. probit reussite genre necono nenfants

Iteration 0: log likelihood = -41.555491
 Iteration 1: log likelihood = -15.835087
 Iteration 2: log likelihood = -15.544835
 Iteration 3: log likelihood = -15.542149
 Iteration 4: log likelihood = -15.542148

Probit regression	Number of obs	=	60
	LR chi2(3)	=	52.03
	Prob > chi2	=	0.0000
Log likelihood = -15.542148	Pseudo R2	=	0.6260

reussite	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
genre	-1.824203	.6504425	-2.80	0.005	-3.099047	-.549359
necono	.3631482	.1102326	3.29	0.001	.1470962	.5792002
nenfants	-.4281971	.2192259	-1.95	0.051	-.857872	.0014778

	_cons	-1.491466	1.108776	-1.35	0.179	-3.664627	.6816959

```
. fitstat
```

Measures of Fit for probit of reussite

Log-Lik Intercept Only:	-41.555	Log-Lik Full Model:	-15.542
D(56):	31.084	LR(3):	52.027
		Prob > LR:	0.000
McFadden's R2:	0.626	McFadden's Adj R2:	0.530
Maximum Likelihood R2:	0.580	Cragg & Uhler's R2:	0.773
McKelvey and Zavoina's R2:	0.830	Efron's R2:	0.681
Variance of y*:	5.878	Variance of error:	1.000
Count R2:	0.900	Adj Count R2:	0.793
AIC:	0.651	AIC*n:	39.084
BIC:	-198.199	BIC':	-39.744

. * Les valeurs des coefficients sont de même signe mais de valeurs différentes par rapport au modèle Logit car la spécification

. * n'est pas la même. Cependant on peut retrouver les valeurs du Logit en multipliant les coefficients par 1.81.

. * Le modèle MCO

```
. regress reussite genre necono nenfants
```

Source	SS	df	MS	Number of obs	=	60
Model	9.15983132	3	3.05327711	F(3, 56)	=	29.36
Residual	5.82350201	56	.103991107	Prob > F	=	0.0000
				R-squared	=	0.6113
				Adj R-squared	=	0.5905
Total	14.9833333	59	.253954802	Root MSE	=	.32248

reussite	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
genre	-.2955676	.0871459	-3.39	0.001	-.4701418 -.1209934
necono	.0666722	.0150106	4.44	0.000	.0366023 .0967421
nenfants	-.0784989	.0360422	-2.18	0.034	-.1507002 -.0062977
_cons	.2172432	.2042165	1.06	0.292	-.1918516 .6263379

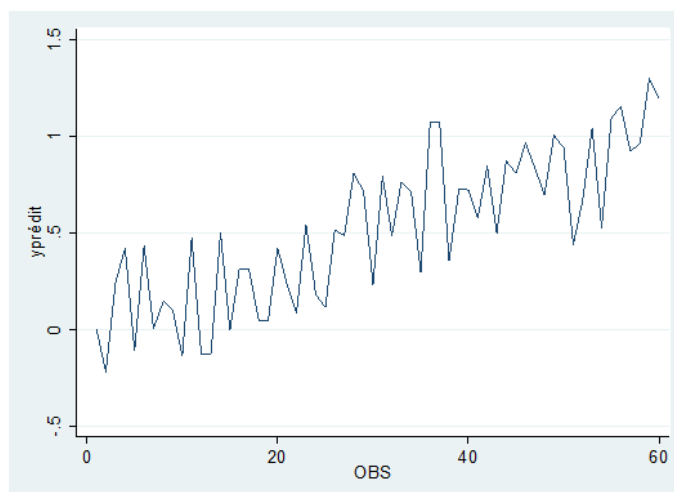
. * Les coefficients sont affectés du même signe que dans les modèles probit et logit. Cependant le graphe des valeurs ajustées ci-après

. * indique que les valeurs sont parfois négatives ou supérieures à 1 ce qui est incompatible avec la valeur d'une probabilité.

```
. predict e, re
```

```
. gen yprédit=reussite-e
```

```
. twoway (line yprédit obs)
```



```
.
. * Question 3 : comment lire les résultats
.
. * soit 1 étudiant dont les caractéristiques sont les suivantes :
. *   - Nenfants = 1 , necono= 12, genre= masculin. Le modèle nous donne alors comme
résultats:
. *  $\text{Log}(\text{Pi}/(1-\text{Pi})) = -0,75 \cdot \text{nenfant} + 0,70 \cdot \text{noécono} - 3.63 \cdot \text{genre} - 2.86 + \text{résidu}$ 
. *  $\text{Log}(\text{Pi}/(1-\text{Pi})) = -0,75 \cdot 1 + 0,70 \cdot 12 - 3.63 \cdot 1 - 2.86 = 1.109$ 
. *  $\exp(1.109) = \text{Pi}/(1-\text{Pi}) = 3.033$  et donc  $P = 75\%$  de chance de réussir à son examen de licence
.
end of do-file

. graph save Graph "P:\nicet001\Desktop\exercices
bourbonnais\Chapitre12.EX1-EX2-EX3\C12EX1\Graph1.gph"
(file P:\nicet001\Desktop\exercices bourbonnais\Chapitre12.EX1-EX2-EX3\C12EX1\Graph1.gph
saved)

. exit, clear
```