

ECONOMETRIA

QUARTO PROBLEM SET

Esercizio 1

Considerate un modello dove la variabile da spiegare è di tipo binario:

$$y_i = \begin{cases} 1 & \text{se l'individuo è laureato} \\ 0 & \text{se l'individuo non è laureato} \end{cases}$$

Il modello lineare che spiega y_i in funzione, per esempio, del livello del reddito e di altre variabili esplicative, è:

$$y_i = x_i' \beta + \varepsilon_i$$

- Considerando che $E(\varepsilon_i|x_i) = 0$, derivate $E(y_i|x_i)$.
- ε_i è distribuito normalmente?
- ε_i è omoschedastico?
- $x_i' \hat{\beta}_{OLS}$ assume valori solo compresi nell'intervallo $[0, 1]$?

Esercizio 2

Supponiamo di avere dati su un campione di famiglie e di essere interessati a determinare le variabili rilevanti nella scelta di acquisto di una barca per le vacanze estive.

- Che tipo di modello teorico può essere considerato? Che tipo di modello stimabile può essere considerato?
- Scrivere un modello teorico di tipo lineare.
- Scrivete il modello stimabile.
- Come sono gli effetti marginali?

Esercizio 3

Considerate un modello in cui stimiamo un modello che vuole spiegare le scelte lavorative di alcune donne intervistate, considerando l'età, lo stato civile, il livello di istruzione. La variabile dipendente è:

$$y_i = \begin{cases} 1 & \text{se la donna lavora} \\ 0 & \text{se la donna non lavora} \end{cases}$$

Il modello stimato è:

$$y_i = \beta_0 + \beta_1 age_i + \beta_2 married_i + \beta_3 educ_i + \beta_4 children_i + \varepsilon_i$$

Ecco la tabella delle statistiche descrittive e l'output di un logit.

```
. sum work age education married
```

Variable	Obs	Mean	Std. Dev.	Min	Max
work	2000	.6715	.4697852	0	1
age	2000	36.208	8.28656	20	59
education	2000	13.084	3.045912	10	20
married	2000	.6705	.4701492	0	1

```
. logit work age married children education
```

```
Iteration 0: log likelihood = -1266.2225
Iteration 1: log likelihood = -1046.086
Iteration 2: log likelihood = -1028.5151
Iteration 3: log likelihood = -1027.9154
Iteration 4: log likelihood = -1027.9144
```

```
Logistic regression                               Number of obs   =       2000
                                                    LR chi2(4)      =       476.62
                                                    Prob > chi2     =       0.0000
Log likelihood = -1027.9144                       Pseudo R2      =       0.1882
```

	work	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age		.0579303	.007221	8.02	0.000	.0437774 .0720833
married		.7417775	.1264704	5.87	0.000	.4939001 .9896549
children		.7644882	.0515287	14.84	0.000	.6634938 .8654827
education		.0982513	.0186522	5.27	0.000	.0616936 .1348089
_cons		-4.159247	.3320397	-12.53	0.000	-4.810033 -3.508462

- a) Commentate la significatività dei coefficienti?
- b) Quale test è utilizzato per la significatività della regressione?
- c) Quale statistica è utilizzata per valutare la bontà del modello?

Esercizio 4

Consideriamo lo stesso modello dell'esercizio precedente. Questa volta usiamo una stima probit. Ecco l'output della stima e degli effetti marginali.

```

probit work age married children education

Iteration 0:  log likelihood = -1266.2225
Iteration 1:  log likelihood = -1040.0608
Iteration 2:  log likelihood = -1027.2398
Iteration 3:  log likelihood = -1027.0616
Iteration 4:  log likelihood = -1027.0616

Probit regression                               Number of obs   =       2000
                                                LR chi2(4)      =       478.32
                                                Prob > chi2     =       0.0000
Log likelihood = -1027.0616                    Pseudo R2       =       0.1889
-----+-----
      work |          Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      age  |   .0347211   .0042293     8.21  0.000   .0264318   .0430105
  married |   .4308575   .074208     5.81  0.000   .2854125   .5763025
  children |   .4473249   .0287417    15.56  0.000   .3909922   .5036576
  education | .0583645   .0109742     5.32  0.000   .0368555   .0798735
      _cons |  -2.467365   .1925635    -12.81  0.000  -2.844782  -2.089948
-----+-----

. mfx compute

Marginal effects after probit
      y = Pr(work) (predict)
      = .71835948
-----+-----
variable |          dy/dx   Std. Err.      z    P>|z|     [ 95% C.I.  ]     X
-----+-----
      age  |   .011721   .00142     8.25  0.000   .008935   .014507   36.208
  married* |   .150478   .02641     5.70  0.000   .098716   .20224    .6705
  children |   .1510059   .00922    16.38  0.000   .132939   .169073   1.6445
  educat~n |   .0197024   .0037     5.32  0.000   .012442   .026963   13.084
-----+-----

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

```

- a) Che differenza c'è tra una stima logit e probit?

b) Qual è l'effetto marginale dell'istruzione sulla variabile dipendente?

c) Secondo voi, il modello è significativo?

Esercizio 5

Ecco lo stesso modello ristimato con il logit:

diverso rispetto al caso dei stima probit?

b) Che cosa rappresentano gli odds-ratio?

c) Che cosa rappresenta β_j ? Qual è il valore di β_j nel caso di variazioni dell'istruzione sulla variabile dipendente?

Esercizio 6

Osservate le seguenti statistiche descrittive nel caso di un dataset panel.

fatal= tasso di mortalità in incidenti in autostrada
 beertax= tassa in caso di birra
 unrate= tasso di disoccupazione
 perincK= reddito personale in migliaia di dollari
 state= stato US
 year=anno

```
. xtsum fatal beertax unrate perincK state year
```

Variable		Mean	Std. Dev.	Min	Max	Observations
fatal	overall	2.040444	.5701938	.82121	4.21784	N = 336
	between		.5461407	1.110077	3.653197	n = 48
	within		.1794253	1.45556	2.962664	T = 7
beertax	overall	.513256	.4778442	.0433109	2.720764	N = 336
	between		.4789513	.0481679	2.440507	n = 48
	within		.0552203	.1415352	.7935126	T = 7
unrate	overall	7.346726	2.533405	2.4	18	N = 336
	between		1.953377	4.1	13.2	n = 48
	within		1.634257	4.046726	12.14673	T = 7
perincK	overall	13.88018	2.253046	9.513762	22.19345	N = 336
	between		2.122712	9.95087	19.51582	n = 48
	within		.8068546	11.43261	16.55782	T = 7
state	overall	30.1875	15.30985	1	56	N = 336
	between		15.44883	1	56	n = 48
	within		0	30.1875	30.1875	T = 7
year	overall	1985	2.002983	1982	1988	N = 336
	between		0	1985	1985	n = 48
	within		2.002983	1982	1988	T = 7

a) Cosa indicano N, n e T?

b) In quale caso la variabilità all'interno del gruppo è bassissima od addirittura pari a zero? Come mai?

c) In quale caso la variabilità tra i gruppi è bassissima od addirittura pari a zero? Come mai?

Esercizio 7

Ecco la stima panel con effetti fissi, random e la stima between, considerando le variabili dell'esercizio precedente.

In questo modello abbiamo regredito gli incidenti in autostrada sulle altre variabili:

$$y_{it} = x'_{it}\beta + \varepsilon_{it} + \alpha_i$$

dove α_i sono gli effetti individuali fissi.

```
. xtreg fatal beertax unrate perincK, fe
```

```
Fixed-effects (within) regression      Number of obs   =      336
Group variable: state                  Number of groups =       48

R-sq:  within = 0.1102                  Obs per group:  min =       7
      between = 0.1862                      avg =      7.0
      overall  = 0.1338                      max =       7

                                          F(3,285)       =     11.76
corr(u_i, Xb) = -0.6481                  Prob > F       =     0.0000
```

fatal	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
beertax	-.4051889	.189984	-2.13	0.034	-.7791386	-.0312392
unrate	-.0267553	.0105621	-2.53	0.012	-.0475449	-.0059657
perincK	.0082151	.0214774	0.38	0.702	-.0340593	.0504895
_cons	2.330946	.3833095	6.08	0.000	1.576469	3.085423
sigma_u	.66643095					
sigma_e	.18349926					
rho	.92952737 (fraction of variance due to u_i)					

```
F test that all u_i=0:      F(47, 285) =     44.03      Prob > F = 0.0000
```

```
xtreg fatal beertax unrate perincK, re
```

```
Random-effects GLS regression      Number of obs   =      336
Group variable: state             Number of groups =      48

R-sq:  within = 0.0763            Obs per group:  min =      7
        between = 0.0588          avg =           7.0
        overall = 0.0587         max =           7

Random effects u_i ~ Gaussian     Wald chi2(3)    =      25.71
corr(u_i, X) = 0 (assumed)       Prob > chi2    =      0.0000
```

fatal	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
beertax	.0490308	.119353	0.41	0.681	-.1848968	.2829583
unrate	-.0434029	.0100291	-4.33	0.000	-.0630595	-.0237463
perincK	-.0325237	.0196911	-1.65	0.099	-.0711176	.0060701
_cons	2.785583	.358486	7.77	0.000	2.082963	3.488203
sigma_u	.44361297					
sigma_e	.18349926					
rho	.85389513	(fraction of variance due to u_i)				

```
. xtreg fatal beertax unrate perincK, be
```

```
Between regression (regression on group means) Number of obs   =      336
Group variable: state             Number of groups =      48

R-sq:  within = 0.0852            Obs per group:  min =      7
        between = 0.3672          avg =           7.0
        overall = 0.2495         max =           7

sd(u_i + avg(e_i.)) = .449002     F(3,44)        =      8.51
                                                Prob > F       =      0.0001
```

fatal	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
beertax	.1275125	.1538482	0.83	0.412	-.1825481	.4375732
unrate	.0152232	.0399585	0.38	0.705	-.0653078	.0957542
perincK	-.1345182	.0403863	-3.33	0.002	-.2159115	-.0531249
_cons	3.730294	.7997585	4.66	0.000	2.118487	5.342101

a) Secondo voi le stime sono diverse nei tre casi? Se sì, in che caso avete dei risultati contrastanti?

b) Spiegate in breve le differenze tra le tre stime

c) Quale sceglieresti?

Esercizio 8

a) Sempre riferendosi al caso precedente, come possiamo testare la possibilità di usare un random effect piuttosto che uno stimatore ad effetti fissi? Spiegate brevemente il test.

b) Ecco l'output del test:

```
----- Coefficients -----
      |      (b)      (B)      (b-B)      sqrt(diag(V_b-V_B))
      |      fix      ran      Difference      S.E.
-----+-----
beertax |  -.4051889   .0490308   -.4542197   .1478133
  unrate |  -.0267553  -.0434029   .0166477   .0033129
 perincK |   .0082151  -.0325237   .0407388   .0085754
-----
```

```
      b = consistent under Ho and Ha; obtained from xtreg
      B = inconsistent under Ha, efficient under Ho; obtained from xtreg
```

Test: Ho: difference in coefficients not systematic

```
      chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
              =      25.62
      Prob>chi2 =      0.0000
      (V_b-V_B is not positive definite)
```

Commentate.