

Concorso pubblico, per titoli ed esami, per la copertura di n. 1 posto di Tecnologo di II livello – posizione economica D5 a tempo determinato per la durata di 18 mesi, eventualmente rinnovabile, con regime di impegno a tempo pieno presso il Dipartimento di Economia, Management e Diritto dell’Impresa dell’Università degli Studi di Bari Aldo Moro, nell’ambito del Piano Nazionale di Ripresa e Resilienza, Missione 4 “Istruzione e ricerca” Componente 2 Investimento 1.3 “Partenariati estesi alle università, ai centri di ricerca, alle aziende per il finanziamento di progetti di ricerca di base” finanziato dall’Unione europea – NextGenerationEU, per il Progetto “GRINS – Growing Resilient, INclusive and Sustainable” (codice progetto PE0000018_CUP H93C22000650001), indetto con DDG n. 457 del giorno 16/04/2024

ESTRATTO DEL VERBALE DELL’ESPLETAMENTO DELLA PROVA SCRITTA

La Commissione, presa visione di quanto previsto dall’art. 7 del bando di concorso, predispone le seguenti tracce:

Traccia n. 1

I dati riportati sono stati raccolti da alcuni cittadini che indicano se hanno mai utilizzato il servizio di bike sharing urbano (risposta SI-NO). Nel questionario sono state raccolte anche informazioni relative a reddito (reddito), età (age) e genere (female). Utilizzando questo set di dati, possiamo prevedere l’utilizzo del sistema di bike sharing rispetto ad alcune variabili socio-economiche. La nostra variabile di risultato è binaria (bike sharing SI=1; bike sharing NO=0) e utilizzeremo un modello probit.

Statistiche descrittive

Variable	Obs	Mean	Std. Dev.	Min	Max
Bike-sharing	400	.3175	.4660867	0	1
reddito	400	5870.7	115.5165	2200	8000
female	400	.1625	.3693709	0	1
age	400	40.3899	.3805668	18	70

Statistiche descrittive della variabile bike sharing

Bike sharing	Freq.	Percent	Cum.
0	273	68.25	68.25
1	127	31.75	100.00
Total	400	100.00	

Stima econometrica probit (variabile dipendente bike sharing)

probit bike sharing reddito female age

Iteration 0: log likelihood = -249.98826

Iteration 1: log likelihood = -238.97735

Iteration 2: log likelihood = -238.94339

Iteration 3: log likelihood = -238.94339

Probit regression Number of obs = 400

 LR chi2(3) = 22.09

 Prob > chi2 = 0.0001

Log likelihood = -238.94339

 Pseudo R2 = 0.0442

Bike sharing| Coef. Std. Err. z P>|z| [95% Conf. Interval]

-----+-----
reddito | .0015244 .0006382 2.39 0.017 .0002736 .0027752
female | .2730334 .1795984 1.52 0.128 -.078973 .6250398
age | .4009853 .1931077 2.08 0.038 .0225012 .7794694
_cons | -2.797884 .6475363 -4.32 0.000 -4.067032 -1.528736

- 1) Illustrare per quale motivo il modello utilizzato nell'esempio è un modello probit e non OLS. Illustrare inoltre le principali differenze rispetto a modelli multinomial logit.
- 2) Utilizzando il modello probit, è possibile interpretare i coefficienti direttamente come per i modelli OLS?
- 3) Come si interpreta il coefficiente “_cons” che rappresenta la costante del modello? Utilizzando le statistiche descrittive riportate in alto, descrivere la variabile dipendente utilizzata nel modello (bike sharing).
- 4) Il modello probit richiede un presupposto di omoschedasticità? Gli errori devono essere normalmente distribuiti? Cosa rappresenta lo Stimatore di massima verosimiglianza (Maximum Likelihood)?
- 5) Perché nella stima non è presente il valore R^2 ?
- 6) Ipotizzando di cambiare il contesto di analisi con un focus sulla sostenibilità territoriale, mantenendo l'utilizzo di un modello probit, quale variabile dipendente dicotomica utilizzerebbe per l'analisi? Quali variabili indipendenti? Quali fonti statistiche considererebbe? Ricorda alcuni lavori scientifici recenti che hanno utilizzato questi modelli per lo studio della sostenibilità territoriale?

Traccia n. 2

In questo esempio viene studiato il salario degli operatori turistici (\ln_w) in termini di anni di scolarizzazione completati ($grade$), età attuale (age) ed età al quadrato ($c.age\#c.age$), numero di anni di lavoro maturati ($experience$) ed esperienza al quadrato ($c.experience\#c.experience$), anni attuali di permanenza nel lavoro attuale ($tenure$) e permanenza in carica al quadrato ($c.tenure\#c.tenure$). L'esempio riporta le informazioni per i lavoratori per gli ultimi 10 anni. Il modello utilizzato è un modello panel a effetti fissi.

```
. xtreg ln_w grade age c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp tenure
> c.tenure#c.tenure 2.race not_smsa south, fe
note: grade omitted because of collinearity.
note: 2.race omitted because of collinearity.
Fixed-effects (within) regression      Number of obs   =   28,091
Group variable: idcode                 Number of groups =    4,697
R-squared:                             Obs per group:
  Within = 0.1727                      min =          1
  Between = 0.3505                      avg =          6.0
  Overall = 0.2625                      max =          15
                                         F(8, 23386)    =   610.12
corr(u_i, Xb) = 0.1936                  Prob > F       =    0.0000
```

ln_wage	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
grade	0	(omitted)				
age	.0359987	.0033864	10.63	0.000	.0293611	.0426362
c.age#c.age	-.000723	.0000533	-13.58	0.000	-.0008274	-.0006186
ttl_exp	.0334668	.0029653	11.29	0.000	.0276545	.039279
c.ttl_exp# c.ttl_exp	.0002163	.0001277	1.69	0.090	-.0000341	.0004666
tenure	.0357539	.0018487	19.34	0.000	.0321303	.0393775
c.tenure# c.tenure	-.0019701	.000125	-15.76	0.000	-.0022151	-.0017251

1. Per quale motivo è stato utilizzato un modello panel e non un modello *pooled* OLS? Quali sono le principali differenze?
2. Quali tipologie di panel conosce? In base a cosa sceglie una tipologia rispetto ad un'altra? Esistono test statistici a supporto della scelta?
3. Per quale motivo nel modello sono stati considerati i coefficienti "c.age#c.age", "c.experience#c.experience", "c.tenure#c.tenure"? Come si interpretano questi coefficienti?
4. Qual è il modo migliore per minimizzare la distorsione da variabili omesse? Che cosa si intende per errori di misura e distorsione da errori nelle variabili? Definire i concetti di eteroschedasticità e omoschedasticità. Inoltre, in caso di presenza di eteroschedasticità, come il modello panel può tenere conto di questo effetto?
5. Per quale motivo la variabile "grade" è stata omessa automaticamente dal modello? Per quale motivo non può essere stimata?
6. Ipotizzando di cambiare il contesto di analisi con quello della sostenibilità territoriale, mantenendo la struttura del database panel, quale variabile dipendente utilizzerebbe per l'analisi? Quali variabili indipendenti? Quali fonti statistiche considererebbe? Ricorda alcuni lavori scientifici recenti che hanno utilizzato questi modelli per lo studio della sostenibilità territoriale?

Traccia n. 3

Nell'esempio riportato, i dati sono stati raccolti da 200 turisti. La misura del risultato in questa analisi è la preferenza tra diverse strutture ricettive – hotel, B&B o ostelli – da cui vedremo quali relazioni esistono con il reddito dei turisti (reddito), l'età (*age*) e il genere (*female*). La nostra variabile di risposta, *strutture ricettive*, è una variabile categorica. I livelli di strutture ricettive non hanno un ordinamento naturale e nell'esempio il gruppo di riferimento è *Hotel*. I coefficienti mostrati nella tabella sono rappresentati in termini di log-odds multinomiali. Il modello utilizzato in questo esempio è un modello multinomial logit.

Statistiche descrittive della variabile strutture ricettive

tab strutture ricettive

	Freq.	Percent	Cum.
Ostelli	47	23.50	23.50
Hotel	95	47.50	71.00
B&B	58	29.00	100.00
Total	200	100.00	

Multinomial logit

mlogit strutture ricettive reddito age female

Iteration 0: log likelihood = -210.58254
Iteration 1: log likelihood = -194.75041
Iteration 2: log likelihood = -194.03782
Iteration 3: log likelihood = -194.03485
Iteration 4: log likelihood = -194.03485

Multinomial logistic regression Number of obs = 200
 LR chi2(6) = 33.10
 Prob > chi2 = 0.0000
Log likelihood = -194.03485 Pseudo R2 = 0.0786

Strutture ric.	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
-----+-----					
Ostelli					
reddito	-.0235647	.0209747	-1.12	0.261	-.0646744 .017545
age	-.0389243	.0195165	-1.99	0.046	-.0771759 -.0006726
female	.8166202	.3909813	2.09	0.037	.050311 1.582929
_cons	1.912256	1.127256	1.70	0.090	-.2971258 4.121638
-----+-----					
B&B					
reddito	.022922	.0208718	1.10	0.272	-.0179861 .0638301
age	.0430036	.0198894	2.16	0.031	.0040211 .081986
female	-.032862	.3500153	-0.09	0.925	-.7188793 .6531553
_cons	-4.057323	1.222939	-3.32	0.001	-6.45424 -1.660407
-----+-----					

(strutture ricettive==hotel is the base outcome)

Iteration Loga

Iteration 0: log likelihood = -210.58254

Iteration 1: log likelihood = -194.75041

Iteration 2: log likelihood = -194.03782

Iteration 3: log likelihood = -194.03485

Iteration 4: log likelihood = -194.03485

- 1) Per quale motivo il modello adottato in questo esempio è un multinomial logit e non un modello logit o OLS?
- 2) Cosa indica il Likelihood Ratio (LR) Chi-Square test ($LR\ chi2(6)$)?
- 3) Come si interpretano le variabili *reddito*, *age* e *female* per la categoria *Ostelli* rispetto a *Hotel*?
- 4) Quali sono le assunzioni del modello Multinomial? Cosa indica l'assunzione dell'indipendenza nel modello Multinomial Logit? A cosa servono i *relative risk ratio* nei modelli multinomial logit? Come si calcolano?
- 5) Qual è la differenza tra i modelli multinomial logit e ordered logit? Fornire degli esempi.
- 6) Ipotizzando di voler analizzare il settore dei trasporti e di voler impiegare una variabile dipendente adatta per l'utilizzo di modelli di multinomial logit. Quale variabile utilizzerebbe per l'analisi? Quali variabili indipendenti? Quali fonti statistiche considererebbe? Ricorda alcuni lavori scientifici recenti

È stata sorteggiata la traccia n. 3.

Il Segretario della Commissione
Dott.ssa Oriana ANTONUCCI