

Econometria

Prof. Valerio Potì

Esercizi Empirici III

ESERCIZI EMPIRICI 11.1 e 11.2

Premessa

Si è ipotizzato che il divieto di fumo nei luoghi di lavoro induca i fumatori a smettere riducendo le loro opportunità di fumare. Negli esercizi empirici 11.1 e 11.2, si stimerà l'effetto del divieto di fumo nei luoghi di lavoro usando dati su un campione di lavoratori statunitensi dal 1991 al 1993 (file *Smoking.xls*). I dati contengono informazioni che indicano se gli individui fossero o meno soggetti al divieto di fumo nel luogo di lavoro, se fumassero, e altre caratteristiche individuali.

Una descrizione dettagliata del dataset è fornita nel file *Smoking_Description.doc*.

Il file *Smoking.xls* fornisce dati su 10.000 lavoratori ma, ai fini dell'esercizio, si utilizzino i dati relativi a un sotto-campione costituito dalle prime 3500 osservazioni (corrispondenti ai primi 3500 lavoratori) del dataset cui i quesiti stessi si riferiscono.

Nello svolgere gli esercizi empirici 11.1 e 11.2, si noti che le variabili binarie che misurano il livello di istruzione (*hsdrop*, *hsgrad*, *colsome* and *colgrad*) siano mutualmente esclusive. Ad esempio, per un individuo con una laurea triennale (*college graduate*), solo la variabile *colgrad* assume valore pari a 1 e tutte le altre assumono valore pari a zero mentre per un individuo che ha frequentato un programma di Master o laurea specialistica (*postgraduate*) tutte le variabili assumono valore pari a zero per evitare collinearità perfetta con la costante.

ESERCIZIO EMPIRICO 11.1

- a) Si stimi la probabilità di fumare per (i) tutti i lavoratori, (ii) i lavoratori colpiti dal divieto di fumo nei luoghi di lavoro, e (iii) i lavoratori non colpiti dal divieto di fumo nei luoghi di lavoro.
- b. Qual è la differenza nella probabilità di fumare tra i lavoratori colpiti dal divieto di fumo nei luoghi di lavoro e i lavoratori non colpiti dal divieto? Si usi un modello di probabilità lineare per determinare se questa differenza è statisticamente significativa.
- c. Si stimi un modello di probabilità lineare con smoker come variabile dipendente e con i seguenti regressori: smkban, female, age, age2, hsdrop, hsgrad, colsome, colgrad, black e hispanic. Si confronti l'effetto del divieto di fumo stimato da questa regressione con la risposta data al punto (b). Si suggerisca, basandosi sulla sostanza di questa regressione, una ragione che spieghi la variazione nell'effetto stimato del divieto di fumo tra (b) e (c).
- d. Si verifichi l'ipotesi che la probabilità di fumare non dipenda dal livello d'istruzione nella regressione del punto (c). La probabilità di fumare aumenta o diminuisce con il livello d'istruzione?
- e. Si verifichi l'ipotesi che la probabilità di fumare non dipenda dal livello d'istruzione nella regressione del punto (c). La probabilità di fumare aumenta o diminuisce con il livello d'istruzione?
- f. Basandosi sulla regressione del punto (c), c'è una relazione lineare tra age e la probabilità di fumare? Si rappresenti la relazione tra la probabilità di fumare e age per $18 \leq \text{age} \leq 65$ per un laureato, bianco, non ispanico, non soggetto a divieto di fumo nei luoghi di lavoro.

ESERCIZIO EMPIRICO 11.2

- a) Si stimi un modello Probit usando gli stessi regressori dell'Esercizio empirico 11.1 (c).
- b) Si verifichi l'ipotesi che il coefficiente di $Smkban$ sia zero in questa regressione Probit rispetto all'alternativa che sia diverso da zero, al livello di significatività del 5%. Si confrontino la statistica t e le conclusioni tratte con quelle della domanda dell'esercizio empirico 11.1 (d) basandosi sul modello di probabilità lineare.
- c) Si verifichi l'ipotesi che la probabilità di fumare non dipenda dal livello d'istruzione in questo modello Probit. Si confrontino i propri risultati con quelli della domanda dell'Esercizio empirico 11.1 (e) basati sul modello di probabilità lineare.
- d) Il signor A è uomo, non ispanico, ha 20 anni e ha abbandonato la scuola superiore. Usando la regressione Probit di (a) e presumendo che il signor A non sia soggetto a divieto di fumo nel luogo di lavoro, si calcoli la probabilità che i signor A fumi. Si effettui nuovamente il calcolo assumendo che sia invece soggetto a divieto di fumo nel luogo di lavoro. Qual è l'effetto del divieto di fumo sulla probabilità di fumare?
- e) Si ripeta (d) per la signora B, donna bianca, 40 anni, laureata.
- f) Si ripetano (d) ed (e) usando il modello di probabilità lineare dell'Esercizio empirico 11.1 (c)
- g) Basandosi sulle risposte a (d)-(f), il modello Probit e il modello di probabilità lineare sono differenti? Se lo sono, quale ha più senso? Gli effetti stimati sono elevati dal punto di vista pratico?
- h) Rimangono importanti minacce alla validità interna?

ESERCIZIO EMPIRICO 11.1

a) Si stimi la probabilità di fumare per (i) tutti i lavoratori, (ii) i lavoratori colpiti dal divieto di fumo nei luoghi di lavoro, e (iii) i lavoratori non colpiti dal divieto di fumo nei luoghi di lavoro.

(i) Dalla *Tabella 11.1.a (colonna (1))* si individua che la probabilità di fumare per i lavoratori, considerati nel complesso, è pari a 20%.

(ii) Dopo aver introdotto nel modello Least Squares anche la variabile Smkban si può osservare, dalla *Tabella 11.1.a (colonna (2))*, una probabilità di fumare in presenza o meno di un divieto di fumo di 24% e di circa 18%.

$$0.238632 - 0.055594 \times 1 = 0.183038, \text{ ovvero il } 18.30\%.$$

(Smkban = 1 comporta il divieto di fumare nei luoghi di lavoro).

(iii) Infine, la probabilità che un lavoratore non soggetto al divieto di fumare nel luogo di lavoro fumi è:

$$0.238632 - 0.055594 \times 0 = 0.238632, \text{ ovvero il } 23.86\%.$$

(Smkban = 0 implica il permesso di fumare nei luoghi di lavoro).

b. Qual è la differenza nella probabilità di fumare tra i lavoratori colpiti dal divieto di fumo nei luoghi di lavoro e i lavoratori non colpiti dal divieto? Si usi un modello di probabilità lineare per determinare se questa differenza è statisticamente significativa.

La differenza che vi è tra i lavoratori colpiti dal divieto di fumo nei luoghi di lavoro e quelli invece non colpiti dal divieto è di -5.56%.

$$(0.238632 - 0.055594 \times 1) - (0.238632 - 0.055594 \times 0) = 18.30\% - 23.86\% = -5.56\% \\ (-0.055594)$$

La differenza è statisticamente significativa perché il coefficiente che determina questa differenza è stimato con il metodo least-squares, la cui probabilità di ipotesi nulla è rifiutata a qualunque livello. Il valore di R-squared è di 0,004459 e, nonostante sia un valore molto piccolo, esso è significativo; a dimostrazione della significatività del valore di R-squared si può guardare quello della statistica F che permette di ipotizzare la bontà del modello (si veda *Tabella 11.1.a colonna (2)*).

c. Si stimi un modello di probabilità lineare con smoker come variabile dipendente e con i seguenti regressori: smkban, female, age, age2, hsdrop, hsgrad, colsome, colgrad, black e hispanic. Si confronti l'effetto del divieto di fumo stimato da questa regressione con la risposta data al punto (b). Si suggerisca, basandosi sulla sostanza di questa regressione, una ragione che spieghi la variazione nell'effetto stimato del divieto di fumo tra (b) e (c).

Si è scelto di stimare la probabilità di fumare utilizzando il modello OLS e considerando i seguenti regressori: Smkban, Female, Age, Age², Hsdrop, Hsgrad, Colsome, Colgrad, Black e Hispanic.

Nella *Tabella 11.1.a (colonna (3))* si riportano i risultati.

L'effetto del divieto di fumo sulla probabilità di fumare è di -4%; il peso del coefficiente del regressore Smkban è inferiore rispetto alla precedente stima effettuata al punto b, dove esso era precisamente di -6%. (si veda **Tabella 11.1.a colonna (2)**).

Tale discrepanza tra i due coefficienti potrebbe essere spiegata dall'introduzione di nuove informazioni contenute nelle variabili aggiunte attraverso la regressione effettuata in **tabella 11.1.a colonna (3)**.

In quest'ultima regressione effettuata troviamo un migliore grado di correlazione, difatti R-squared è dieci volte più grande del precedente e inoltre i coefficienti, oltre ad essere significativi complessivamente (statistica F), lo sono anche singolarmente a qualunque livello. Tuttavia le uniche eccezioni è l'intercetta (si veda **Tabella 11.1.a colonna (3)**).

d. Si verifichi l'ipotesi che la probabilità di fumare non dipenda dal livello d'istruzione nella regressione del punto (c). la probabilità di fumare aumenta o diminuisce con il livello d'istruzione?

L'ipotesi che il coefficiente di Smkban sia zero è rigettata a livello di significatività di 10% e di 5% e perciò tale coefficiente è statisticamente significativo, come si può evincere dal valore della probabilità della statistica t in **Tabella 11.1.a colonna (3)**.

e. Si verifichi l'ipotesi che la probabilità di fumare non dipenda dal livello d'istruzione nella regressione del punto (c). La probabilità di fumare aumenta o diminuisce con il livello d'istruzione?

Il peso dei singoli regressori diminuisce con l'aumentare del livello di istruzione. Non è trascurabile la loro presenza all'interno del modello Least Squares in quanto si tratta di regressori statisticamente significativi a qualunque livello. Dai risultati ottenuti, possiamo notare che il loro coefficiente diminuisce con l'aumentare del livello di istruzione, per tanto la probabilità diminuisce.

Nella regressione di cui al punto c, la probabilità che l'individuo fumi dipende dal livello d'istruzione. Questo poiché i coefficienti dei regressori legati all'istruzione (Hsdrop, Hsgrad, Colsome e Colgrad) sono tutti statisticamente significativi (al livello di significatività del 99%), in quanto hanno rispettivamente valori di T-Statistic pari a 7,77 11,13 8,51 3,02. Per esempio, il coefficiente della variabile Colgrad è 0,054499 quindi la probabilità di fumare per un laureato è 5,45%. Il coefficiente della variabile Hsdrop è 0,230781 quindi la probabilità di fumare per un individuo che ha abbandonato la scuola superiore è 23,08%, ovvero quasi cinque volte maggiore rispetto ad un laureato. Inoltre, poiché i coefficienti sono positivi e diminuiscono all'aumentare del livello di istruzione, la probabilità di fumare diminuisce all'aumentare del grado di istruzione.

f. Basandosi sulla regressione del punto (c), c'è una relazione lineare tra age e la probabilità di fumare? Si rappresenti la relazione tra la probabilità di fumare e age per $18 \leq \text{age} \leq 65$ per un laureato, bianco, non ispanico, non soggetto a divieto di fumo nei luoghi di lavoro.

La relazione tra la probabilità di fumare e l'età non è lineare. Il coefficiente del regressore Age² è piccolo, ma statisticamente significativo. La sua presenza è necessaria in quanto se questo regressore fosse omesso sembrerebbe che la probabilità di fumare aumenti linearmente con l'avanzare dell'età e ciò non è plausibile. Per descrivere la forma funzionale che descrive questo effetto si può utilizzare

la forma quadratica, che permette di conservare la correzione. Se provassimo ad implementare il medesimo modello ma senza considerare la variabile quadratica, otterremmo che il coefficiente Age diventa negativo.

La relazione tra la probabilità di fumare per un soggetto laureato con un'età compresa tra i 18, bianco, non ispanico, non soggetto a divieto di fumo nei luoghi di lavoro è la seguente:

$$\text{Smoker} = -0,040144 + (-0,036561) \times 0 + (0,008622) \times \text{Age} + (-0,000105) \times \text{Age}^2 + 0.230781 \times 0 + 0.235466 \times 0 + 0.160597 \times 0 + 0.054499 \times 1 + (-0,051196) \times 0 + (-0,037490) \times 0 + (-0,075809) \times 0$$

$$\text{Smoker} = -0,040144 + (0,008622) \times \text{Age} + (-0,000105) \times \text{Age}^2 + 0.054499$$

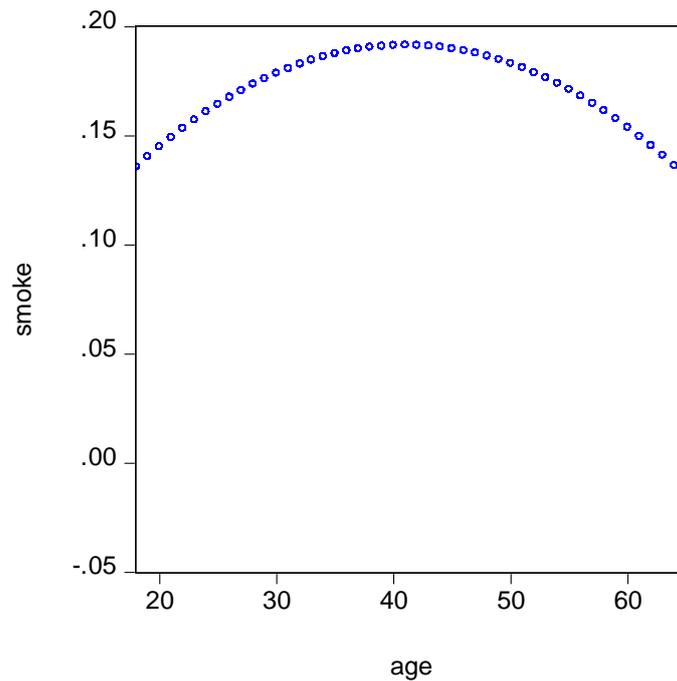


Figura 1. Nella figura è rappresentata la relazione tra la probabilità di fumare e Age, per $18 \leq \text{Age} \leq 65$ per un laureato, bianco, non ispanico, non soggetto a divieto di fumo nei luoghi di lavoro.

TABELLE ESERCIZIO 11.1

Con statistiche t:			
	(1)	(2)	(3)
Variabile dipendente: Smoker			
Smkban		-0,06 (-3,96)	-0,04 (-2,53)
Age			0,009 (3,03)
Age ²			-0,0001 (-3,26)
Hsdrop			0,23 (7,77)
Hsgrad			0,23 (11,13)
Colsome			0,16 (8,51)
Colgrad			0,05 (3,02)
Black			-0,05 (-2,09)
Hispanic			-0,08 (-3,80)
Female			-0,04 (-2,75)
Intercetta	0,20 (29,2)	0,24 (21,46)	-0,04 (-0,63)
F-statistic		15,67 (0,00)	16,65 (0,00)
SER	0,40	0,40	0,39
R ²	0,00	0,0044	0,045
Adjusted R ²	0,00	0,0041	0,043
n	3500	3500	3500

Tabella 11.1.a. La tabella riporta tre diverse regressioni, realizzate con il metodo OLS classico: in colonna (1) la variabile Smoker è stata regredita unicamente sull'intercetta per considerare i lavoratori nel loro complesso; in colonna (2) è stata effettuata la medesima regressione (1) con l'aggiunta della variabile Smkban per poter differenziare gli effetti di un divieto di fumo sulla probabilità di fumare; infine, in colonna (3) si riporta la regressione di Smoker su diverse variabili, ovvero: Smkban, Age, Age², Hsdrop, Hsgrad, Colsome, Colgrad, Black, Hispanic, Female, oltre ad una costante (Intercetta). Per ciascuna variabile, in tutte le regressioni, oltre al valore del coefficiente, si riporta il valore della statistica t tra parentesi. Per la statistica F, si riporta in parentesi il valore p. Il campione oggetto di studio contiene n = 3500 osservazioni. Si riportano inoltre: la statistica F per la significatività congiunta dei regressori, la deviazione standard campionaria dei residui e due misure di bontà di adattamento del modello ai dati (R² e adjusted R²).

Con valori-p:			
	(1)	(2)	(3)
Variabile dipendente: Smoker			
Smkban		-0,06 (0,00)	-0,04 (0,01)
Age			0,009 (0,004)
Age ²			-0,0001 (0,002)
Hsdrop			0,23 (0,00)
Hsgrad			0,23 (0,00)
Colsome			0,16 (0,00)
Colgrad			0,05 (0,002)
Black			-0,05 (0,03)
Hispanic			-0,08 (0,002)
Female			-0,04 (0,006)
Intercetta	0,20 (0,00)	0,24 (0,00)	-0,04 (0,55)
F-statistic		15,67 (0,00)	39,02 (0,00)
SER	0,40	0,40	0,39
R ²	0,00	0,0044	0,045
Adjusted R ²	0,00	0,0041	0,043
n	3500	3500	3500

Tabella 21.1.a. La tabella riporta tre diverse regressioni, realizzate con il metodo OLS classico: in colonna (1) la variabile Smoker è stata regredita unicamente sull'intercetta per considerare i lavoratori nel loro complesso; in colonna (2) è stata effettuata la medesima regressione (1) con l'aggiunta della variabile Smkban per poter differenziare gli effetti di un divieto di fumo sulla probabilità di fumare; infine, in colonna (3) si riporta la regressione di Smoker su diverse variabili, ovvero: Smkban, Age, Age², Hsdrop, Hsgrad, Colsome, Colgrad, Black, Hispanic, Female, oltre ad una costante (Intercetta). Per ciascuna variabile, in tutte le regressioni, oltre al valore del coefficiente, si riporta il valore p tra parentesi. Per la statistica F, si riporta in parentesi il valore p. Il campione oggetto di studio contiene n = 3500 osservazioni. Si riportano inoltre: la statistica F per la significatività congiunta dei regressori, la deviazione standard campionaria dei residui e due misure di bontà di adattamento del modello ai dati (R² e adjusted R²).

ESERCIZIO EMPIRICO 11.2

a) Si stimi un modello Probit usando gli stessi regressori dell'Esercizio empirico 11.1 (c).

Attraverso un modello non lineare Probit (il quale si differenzia dal modello OLS, nel quale si cerca di minimizzare il quadrato dei residui) si è stimata la relazione tra la probabilità di fumare (Smoker) e le variabili Smkban, Female, Age, Age², Hsdrop, Hsgrad, Colsome, Colgrad, Black e Hispanic.

Ottenendo come risultato quanto mostrato nella *Tabella 11.2.a*.

Dalla *Tabella 11.2.a* si visualizza lo pseudo R^2 (McFadden), che valuta l'adattamento del modello rispetto alla sola intercetta sulla base di un rapporto di verosimiglianza. Maggiore è il valore della funzione di verosimiglianza rispetto a quello della sola intercetta maggiore è lo pseudo R^2 .

b) Si verifichi l'ipotesi che il coefficiente di Smkban sia zero in questa regressione Probit rispetto all'alternativa che sia diverso da zero, al livello di significatività del 5%. Si confrontino la statistica t e le conclusioni tratte con quelle della domanda dell'esercizio empirico 11.1 (d) basandosi sul modello di probabilità lineare.

Osservando i valori della tabella 11.a, il valore della z-statistic di Smkban è -2,68. Possiamo quindi affermare che il valore di Smkban è significativo e rigettare l'ipotesi nulla per cui Smkban sia uguale a zero (livello di significatività del 5%). In conclusione, i risultati ottenuti risultano simili all'esercizio 11.1 (d) anche se il modello utilizzato non è più lineare.

c) Si verifichi l'ipotesi che la probabilità di fumare non dipenda dal livello d'istruzione in questo modello Probit. Si confrontino i propri risultati con quelli della domanda dell'Esercizio empirico 11.1 (e) basati sul modello di probabilità lineare.

Osservando il modello stimato al punto a dell'esercizio, come si nota dalla *Tabella 11.2.a* è possibile notare che i coefficienti relativamente ai livelli di istruzione sono tutti significativi. Questo porta a rigettare l'ipotesi nulla che la probabilità di fumare non dipenda dal grado di istruzione. Osservando inoltre i risultati con quelli ottenuti dall'esercizio empirico 11.1 (e) si nota che viene rispettata quanto già detto, ovvero una diminuzione dei fumatori al crescere del grado di istruzione.

d) Il signor A è uomo, non ispanico, ha 20 anni e ha abbandonato la scuola superiore. Usando la regressione Probit di (a) e presumendo che il signor A non sia soggetto a divieto di fumo nel luogo di lavoro, si calcoli la probabilità che il signor A fumi. Si effettui nuovamente il calcolo assumendo che sia invece soggetto a divieto di fumo nel luogo di lavoro. Qual è l'effetto del divieto di fumo sulla probabilità di fumare?

Sostituendo, nel modello di regressione Probit del punto (a), i valori forniti per il Signor "A", abbiamo calcolato la probabilità che questo fumi. Nello specifico, la probabilità che il signor A sia un fumatore in un luogo di lavoro dove non è soggetto a divieto di fumo è: 36,07%. In coerenza con la specificazione del modello Probit stimato, tale valore è stato ottenuto utilizzando la funzione cumulativa di densità della "distribuzione normale" standardizzata, $F(x)$, sostituendo all'argomento della stessa (la "x") il valore z trovato dalla regressione, alla "media" il valore 0, a "dev_standard" il

valore 1 (in modo tale da ottenere la normale standardizzata) ed a “cumulativo” il parametro “falso” per poter determinare la funzione di probabilità.¹

$$N(-1,9030 + 0,0333 \times 20 - 0,0004 \times 400 + 0,9483 \times 1)$$

La probabilità che il signor A sia un fumatore in un luogo di lavoro dove è soggetto a divieto di fumo è: 33,60%. Risultato trovato attraverso il medesimo procedimento visto in precedenza.

$$N(-1,9030 - 0,1374 \times 1 + 0,0333 \times 20 - 0,0004 \times 400 + 0,9483 \times 1)$$

Possiamo dunque concludere che in un luogo di lavoro senza divieto di fumare, la percentuale di lavoratori fumatori è lievemente più alta. La differenza tra i due è del 4,8% a favore del luogo di lavoro senza divieto di fumare.

e) Si ripeta (d) per la signora B, donna bianca, 40 anni, laureata.

La probabilità della signora B, lavoratrice in un luogo di lavoro dove non c'è divieto di fumare è: 29,57%.

$$N(-1,9030 + 0,0333 \times 40 - 0,0004 \times 1600 + 0,3060 \times 1 + 0,7230 \times 1 + 0,9612 \times 1)$$

La probabilità della signora B in un luogo di lavoro dove c'è il divieto di fumare è invece: 26,84%.

$$N(-1,9030 + 0,137476031067 \times 1 + 0,0333 \times 40 - 0,0004 \times 1600 + 0,3060 \times 1 + 0,7230 \times 1 + 0,9612 \times 1)$$

Inaspettatamente il modello mostra che la probabilità che il soggetto sia una fumatrice, è più alta se questa lavora in un luogo di lavoro soggetto a divieto di fumo rispetto che non. Tale differenza è espressa in misura percentuale per un valore pari al 2.1%.

f) Si ripetano (d) ed (e) usando il modello di probabilità lineare dell'Esercizio empirico 11.1 (c)

Utilizzando il modello di probabilità lineare dell'esercizio 11.1 (c), otteniamo le seguenti probabilità:

- per il Signor “A”: 38,31% in assenza di divieto di fumo e 37,89% con divieto.

- per la Signora “B” risulta 34,97% senza il divieto di fumo nel luogo di lavoro e 34,29% in presenza di tale divieto.

Come era lecito attendersi per entrambi i sessi il divieto di fumare sul luogo di lavoro impatta positivamente sulla probabilità di fumare dell'individuo. Per l'uomo, l'effetto del divieto risulta pari a 0,42; per la donna l'effetto del divieto è 0,68. Si tratta di valori simili tra loro in virtù del fatto che stiamo utilizzando un modello di probabilità lineare.

¹ Nel programma Microsoft Excel, la probabilità che il Signor "A" fumi, ovvero 36,07%, è stato ottenuto utilizzando la funzione cumulativa di densità della “distribuzione normale” standardizzata, F(x), sostituendo all'argomento della stessa (la “x”) il valore z trovato dalla regressione, fornendo nella funzione di Excel come input per la “media” il valore 0 e come input per la “dev_standard” il valore 1 (in modo tale da ottenere la normale standardizzata) ed come input per il campo “cumulativo” il parametro “falso”, così scegliendo la cumulata della funzione di densità.

g) Basandosi sulle risposte a (d)-(f), il modello Probit e il modello di probabilità lineare sono differenti? Se lo sono, quale ha più senso? Gli effetti stimati sono elevati dal punto di vista pratico?

Osservando i risultati ottenuti utilizzando sia il modello Probit sia il modello di proprietà lineare si nota che non vengono mostrate importanti differenze. Le piccole differenze che esistono sono spiegate dalla differente linearità, infatti come già detto nelle precedenti risposte il modello Probit è un modello non lineare.

Il modello Probit dal punto di vista grafico è a forma di S, ciò implica una buona approssimazione nella parte centrale da una retta. La linearità perde il suo significato quando ci si avvicina agli estremi della risposta limitata ovvero 0 ed 1. Le discrepanze che esistono sui risultati sono piccole perché si sta trattando dei dati i cui valori possono essere considerati mediani nell'insieme delle osservazioni.

h) Rimangono importanti minacce alla validità interna?

Minacce possibili alla validità interna sono sicuramente: localizzazione geografica, omissione di variabili come per esempio la tipologia di lavoro, la posizione ricoperta, la durata dei turni di lavoro, la tipologia di contratto sottoscritto (full-time o part-time), tipologia degli studi, caratteristiche fisiche e chimiche delle sigarette, e prezzo delle sigarette.

TABELLE ESERCIZIO 11.2

Variabile dipendente: Smoker	
Smkban	-0,14 (0,007)
Age	0,03 (0,003)
Age ²	-0,0004 (0,002)
Hsdrop	0,95 (0,00)
Hsgrad	0,96 (0,00)
Colsome	0,72 (0,00)
Colgrad	0,31 (0,004)
Black	-0,18 (0,05)
Hispanic	-0,26 (0,0003)
Female	-0,14 (0,005)
Intercetta	-1,90 (0,00)
LR statistic	169,52 (0,00)
SER	0,39
McFadden R ²	0,048
n	3500

Tabella 11.2.a. La tabella riporta la regressione, effettuata con un modello Probit, di Smoker su Smkban, Age, Age², Hsdrop, Hsgrad, Colsome, Colgrad, Black, Hispanic, Female e Intercetta. Per ciascuna variabile, oltre al valore del coefficiente, si riporta il p-value associato tra parentesi. In quanto utilizzate nell'esercizio, ma non indicate in tabella, si riportano di seguito i valori delle statistiche z associate ai coefficienti: Smkban = -2,70, Age = 2,73, Age² = -2,80, Hsdrop = 7,85, Hsgrad = 9,35, Colsome = 7,15, Colgrad = 2,84, Black = -1,99, Hispanic = -3,75, Female = -2,79, Intercetta = -7,13. Il campione oggetto di studio contiene n = 3500 osservazioni. Si riportano inoltre: la statistica F per la significatività congiunta dei regressori, la deviazione standard campionaria dei residui e una misura di adattamento del modello ai dati (McFadden R²).