Analisi Statistica per le Imprese

Prof. L. Neri

Dip. di Economia Politica e Statistica

5.2 Metodi regressivi: modello logit

Perchè il modello logit in campo aziendale

Tra i metodi quantitativi di analisi di marketing strategico ci occupiamo del modello di regressione logistica (definito anche logit), consiste nella creazione di un modello non lineare che individui le principali caratteristiche in base alle quali poter effettuare

- una previsione delle vendite
- identificare il potenziale di mercato
- studiare il comportamento del cliente
- valutazioni sulla soddisfazione dei consumatori
- una segmentazione del mercato

La regressione logistica è utilizzata per studiare la relazione esistente tra una variabile dipendente y (qualitativa) e una o più variabili indipendenti x che possono essere sia qualitative che quantitative. La variabile y è una variabile le cui modalità rappresentano due o più alternative mutuamente esclusive.

Per esempio un' analista potrebbe essere interessato a studiare:

- il grado di soddisfazione dei clienti (da non soddisfatti a completamente soddisfatti)
- le cause della scelta di un determinato prodotto (scelta prodotto A, non scelta di A)
- stato di salute dell'azienda (sana/in crisi)
- opinione di una certa categoria di consumatori sul prodotto M (pessimo, discreto, buono, ottimo, eccellente)
- il riscontro positivo/negativo ad un'offerta promozionale
- la propensione all'acquisto di un certo prodotto (bassa, media, alta)

Obiettivi

Gli obiettivi possono essere molteplici:

- individuare tra le variabili indipendenti quelle a maggiore potere esplicativo, che vanno quindi interpretate come determinanti del possesso o meno dell'attributo: a seconda che siano correlate positivamente o negativamente con il fenomeno studiato possono essere considerate come fattori di rischio o come fattori di protezione;
- ricercare la combinazione lineare delle variabili indipendenti che meglio discrimina fra il gruppo delle unità che possiedono l'attributo e quello delle unità che non lo posseggono;
- stimare la probabilità del possesso dell'attributo per una nuova unità statistica su cui è stato osservato il vettore di variabili x e, fissato per tale probabilità un valore soglia, classificare come appartenente alla categoria che possiaede l'attributo o all'altra.

Tipologie di modelli

- logit per variabili binarie (dicotomiche)
- logit multinomiale - si applica quando la variabile dipendente ha più di due categorie (argomento non trattato nel corso)
- logit ordinale - si applica quando la variabile dipendente è su scala ordinale (argomento non trattato nel corso)

Logit per variabili binarie

Si specificano tali modelli quando la variabile dipendente può assumere solo due valori che rappresentano il successo o l'insuccesso, generalmente la presenza o assenza di un attributo di interesse.
y è binaria e assume solo due valori che per convenienza codifichiamo con 0 e 1 per l'$i-esima$ unità statistica $(i = 1...n)$ potremmo assumere, ad esempio:

$$\begin{cases} y_i\!=\!1\; se\, il\, cliente\, acquista; \\ y_i\!=\!0\; se\, il\, cliente\, non\, acquista \end{cases} \tag{1}$$

in questo modo potremo stimare, sulla base della conoscenza dei valori assunti dalle variabili esplicative, la probabilità che si verifichi l'acquisto di un prodotto (piuttosto che stimare un modello di regressione per valutare e prevedere le vendite del prodotto).

Esempio

Nell'esempio precedente avremo allora una matrice dei dati del tipo:

cliente	acquisto/non acquisto	y_i
1	yes	1
2	no	0
3	no	0
4	yes	1

Distribuzione di Bernoulli

y_i è una realizzazione di una v.c. Y_i che può assumere solo due valori uno e zero con probabilità π_i e $1-\pi_i$ rispettivamente. La distribuzione di Y_i è detta distribuzione di Bernoulli con parametro π_i, che può essere scritta in forma compatta come

$$P(Y_i = y_i) = \pi_i^{y_i} (1-\pi_i)^{1-y_i} \tag{2}$$

per $y_i = 0, 1$.

- se $y_i = 1$ otteniamo $P((Y_i = 1)) = \pi_i$
- se $y_i = 0$ otteniamo $P((Y_i = 0)) = 1-\pi_i$

Media e Varianza

E' facilmente verificabile che il valore atteso e la varianza di Y_i sono

$$E(Y_i) = \pi_i \tag{3}$$

$$V(Y_i) = \pi_i(1 - \pi_i) \tag{4}$$

- media e varianza dipendono da $\pi_i \Longrightarrow$ ogni fattore che influenza la probabilità di successo altera la media e la varianza

Questo suggerisce che un modello lineare che assume che i predittori influenzino la media ma che la varianza sia costante è inadeguato per studiare dati di tipo binario

Il modello di probabilità lineare

Il prossimo step per definire il nostro modello riguarda la parte sistematica

- vorremmo che le probabilità π_i dipendessero da un vettore di covariate osservate x_i
- l'idea più semplice potrebbe essere di specificare π_i come funzione lineare delle covariate cioè:

$$\pi_i = x_i^{'}\beta \tag{5}$$

dove β è un vettore di coefficienti che deve essere stimato.

- il modello così definito prende il nome di modello di probabilità lineare (linear probability model), tale modello può essere stimato con OLS

La trasformazione del modello

Un problema che si presenta con il modello sopra definito è il seguente:

- π_i giace tra 0 e 1, ma il predittore lineare $x_i'\beta$ può assumere qualsiasi valore sull'asse reale,

quindi non c'è garanzia che i valori predetti dal modello siano compresi tra 0 e 1.

- una semplice soluzione a tale problema è quello di "trasformare" la probabilità e specificare tale "trasformazione" della probabilità come funzione delle covariate.

Dalle probabiltà agli odds

Questa trasformazione può essere fatta in due passi:

- passiamo dal concetto di probabilità π_i al concetto di odd definito come il rapporto tra la probabilità di successo ed il suo complemento:

$$odds_i = \frac{\pi_i}{1-\pi_i} = \frac{P(Y_i = 1)}{1 - P(Y_i = 1)} = \frac{P(Y_i = 1)}{P(Y_i = 0)} \tag{6}$$

A titolo esemplificativo:

- se la probabilità di un evento è 1/2, l'odds sarà 1 su 1, ovvero successo e insuccesso sono equiprobabili
- se la probabilità di un evento è 1/3, l'odds sarà 1 su 2, ovvero il successo e meno probabile dell'insuccesso
- se la probabilità di un evento è 2/3, l'odds sarà 2 su 1, ovvero il successo è più probabile dell'insuccesso

La trsformazione logit

Prendiamo il logaritmo dell'odd, calcolando il logit o log-odd

$$\eta_i = logit\,(\pi_i) = log\frac{\pi_i}{1-\pi_i} \tag{7}$$

che ha l'effetto di rimuovere la restrizione del campo di variazione della probabilità, infatti

- se la probabilità tende a zero, l'odds tende a zero e il logit tende a $-\infty$
- se la probabilità tende a uno, l'odds e il logit tendono a$+\infty$
- in conclusione il logit proietta le probabilità da $[0,1]$ su tutto l'asse reale

Il modello di regressione logistica

Supponiamo di disporre di n osservazioni independenti $y_1 \dots . y_n$, e che l'$i-esima$ osservazione possa essere trattata come una realizzazione di un v.c. Y_i. Assumiamo che Y_i abbia una distribuzione Binomiale

$$Y_i \sim B(1; \pi_i) \tag{8}$$

questo definisce la struttura stocastica del modello che studiamo.

Il modello di regressione logistica

Si definisce il modello di regressione logistica assumendo che il logit della probabilità, piuttosto che la probabilità stessa, sia funzione lineare delle covariate

$$logit\,(\pi_i) = x_i^{'}\beta \tag{9}$$

dove x_i è un vettore $h \times 1$ di covariate e β è un vettore $h \times 1$ di coefficienti di regressione

- l'espressione precedente definisce la struttura sistematica del modello
- il modello definito è un modello lineare generalizzato con risposta binomiale e trasformazione (link) logit

I coefficienti di regressione

I coefficienti di regressione possono essere interpretati in modo analogo a quelli del modello di regressione lineare, ma ricordando che sul lato sinistro della relazione c'è un logit, quindi

- β_j rappresenta il cambiamento nel logit della probabilità di successo associato ad un cambiamento unitario nel $j-esimo$ predittore lineare tenendo costanti gli altri predittori (variabili indipendenti)

Il modello in termini di probabilità

Specificando il modello in termini della probabilità π_i abbiamo una forma del modello più complicata

$$\pi_i = P\left(Y_i = 1|x_i\right) = \frac{exp\left(x_i^{'}\beta\right)}{1 + exp\left(x_i^{'}\beta\right)} \tag{10}$$

mentre sul lato sinistro compare la probabilità, sul lato destro c'è una funzione non lineare dei predittori e non c'è un modo semplice per esprimere l'effetto sulla probabilità per un incremento unitario di un predittore matenendo costanti le altre variabili

Effetti marginali

Per poter utilizzare in maniera corretta il modello logit è importante saper interpretare β_j.
Con riferimento al modello espresso in termini di probabilità, per valutare l'effetto sulla probabilità di successo della variazione di una covariata continua, ad esempio x_j, si ricorre a

$$\frac{\partial \pi_i}{\partial x_{ij}} = \frac{\partial P(Y_i = 1|x_i)}{\partial x_{ij}} = \beta_j \pi_i (1 - \pi_i) = \beta_j \frac{exp\left(x_i^{'}\beta\right)}{\left[1 + exp\left(x_i^{'}\beta\right)\right]^2} \qquad (11)$$

Quindi l'effetto sulla probabilità di successo della variazione di una delle covariate:

- dipende dal valore assunto da tutte le covariate x_i
- ma coincide con il segno del corrispondente coefficienteβ_j

Può essere interessante valutare tale variazione in corrispondenza di particolari valori delle covariate (spesso si sceglie a tale scopo il vettore dei valori medi, $\bar{x}$)

Effetti marginali

Se x_{ij} è una variabile binaria, l'effetto parziale della variazione di x_{ij} da zero a uno, mantenendo tutte le altre variabili esplicative costanti, è semplicemente dato da:

$$P(y=1|x_{i1},x_{i2},..x_{ij}=1,..x_{ik})-P(y=1|x_{i1},x_{i2},..x_{ij}=0,..x_{ik})$$

Interpretazione in termini di odds ratio (OR)

Consideriamo l'odds-ratio per una variabile dicotomica

Ragioniamo in termini di odds che per l' $i-esima$ unità è dato da

$$\frac{\pi_i}{1-\pi_i} = exp\left(x_i^{'}\beta\right) \tag{12}$$

Si consideri un modello con due esplicative: x_1 continua ed x_2 dicotomica.

$$odds(x_2=1) = \frac{P(y=1|x_1,x_2=1)}{1-P(y=1|x_1,x_2=1)} = exp\left(\beta_0+\beta_1 x_1+\beta_2\right)$$

$$odds(x_2=0) = \frac{P(y=1|x_1,x_2=0)}{1-P(y=1|x_1,x_2=0)} = exp\left(\beta_0+\beta_1 x_1\right)$$

Interpretazione in termini di odds ratio (OR)

Allora,

$$OR = \frac{odds(x_2 = 1)}{odds(x_2 = 0)} = exp\,(\beta_2)$$

Supponiamo per esempio che $exp\,(\beta_2) = 2$ poiché tale valore rappresenta il rapporto tra la propensione al successo riferita ad $x_2 = 1$ e la propensione al successo riferita ad $x_2 = 0$, possiamo affermare che le unità caratterizzate da $x_2 = 1$ hanno una propensione al successo doppia rispetto alle unità caratterizzate da $x_2 = 0$.

interpretazione in termini di odds ratio (OR)

Consideriamo ora l'odds ratio per una variabile continua.

Se x_1 è continua, espressa in una data unità di misura, si ha che l'OR corrispondente ad un incremento unitario della variabile è uguale al caso dicotomico, cioè $exp\,(\beta_1)$

Se, ai fini interpretativi, è più interessante considerare un incremento di c unità $(c \neq 1)$ piuttosto che un incremento unitario della variabile, allora il logaritmo dell'odds ratio corrispondente è uguale a $exp\,(c\beta_1)$

Interpretazione in termini di odds ratio (OR)

Consideriamo ora l'odds ratio per una variabile x categorica o ordinale a k modalità

Si ricordi che le k modalità della variabile sono espresse nel modello attraverso $k-1$ variabili dummy. Se nel calcolo degli odds ratio il gruppo delle unità portatrici della modalità corrispondente all'annullamento di tutte le dummy viene preso come "gruppo di riferimento" quel tipo di codifica garantisce che il logaritmo dell'odds ratio del gruppo delle unità che portano l'$i-esima$ modalità della variabile x rispetto al gruppo di riferimento è (per $i = 1, \ldots, k-1$) pari a $\beta_{1,i}$ e quindi l'odds ratio di questo gruppo rispetto al gruppo di riferimento è uguale a $exp(\beta_{1,i})$

Esempio

Supponiamo di che la variabile x sia una variabile qualitativa con tre modalità che indichiamo per semplicità con A,B, C. Inseriamo nel modello due variabili dummy, che indichiamo con D_B e D_c , quindi scegliamo come riferimento l'attributo A. Indichiamo poi rispettivamente con β_B e β_C i coefficienti stimati con la regressione logistica. Supponiamo poi che $exp\left(\beta_B\right) = 3$ e $exp\left(\beta_C\right) = 1$. Allora possiamo dire che:

- le unità caratterizzate da $x = B$ hanno una propensione al successo tripla rispetto alle unità caratterizzate da $x = A$.
- le unità caratterizzate da $x = C$ hanno una propensione al successo uguale a quella delle unità caratterizzate da $x = A$.

Stima di massima verosimiglianza

Per stimare i parametri del modello si impiega il metodo della massima verosimiglianza e, poiché le equazioni generate dalla massimizzazione della verosimiglianza sono non lineari nei parametri (non ammettono soluzione esplicita), le stime dei coefficienti si ottengono utilizzando procedure numeriche iterative

Scelta del modello

Dopo aver stimato il modello, è necessario valutare la significatività sia del modello nel suo insieme sia con riferimento ai singoli coefficienti. Ci si basa sulle proprietà dello stimatore di massima verosimiglianza che è asintoticamente normale e, di frequente, si impiega il test di tipo LR, per confrontare modelli annidati e scegliere tra questi quello più appropriato nel caso empirico esaminato

Test di ipotesi

Indichiamo con $\widehat{\beta_j}$ lo stimatore di massima verosimiglianza del $j-esimo$ parametro e con $\widehat{var\left(\widehat{\beta_j}\right)}$ la stima della sua varianza possiamo sottoporre a verifica

$$H_0 : \beta_j = 0 \tag{13}$$

che sottopone a verifica la significatività del singolo coefficiente, utilizzando la seguente statistica test

$$z = \frac{\hat{\beta}_j - 0}{\sqrt{\widehat{var\left(\widehat{\beta_j}\right)}}} \tag{14}$$

Per grandi campioni la statistica test tende ad una distribuzione normale standardizzata
se $|z_{obs}| > z_{\alpha/2}$ si rifiuta l'ipotesi nulla al livello di significatività α prescelto

Test di ipotesi

Alternativamente, possiamo considerare il quadrato della statistica test sopra vista che per grandi campioni tende a distribuirsi come un χ^2 con 1 g.l.

$$\chi_1^2 = \left(\frac{\hat{\beta}_j - 0}{\sqrt{\widehat{var\left(\hat{\beta}_j\right)}}} \right)^2 \tag{15}$$

se $\chi^2_{obs} > \chi^2_{\alpha,1}$ si rifiuta l'ipotesi nulla al livello di significatività α prescelto.

Intervallo di confidenza per β_j

Un intervallo di confidenza per β_j al livello di confidenza $(1-\alpha)\%$ è definito da

$$\hat{\beta}_j \pm z_{1-\frac{\alpha}{2}} \sqrt{\widehat{var\left(\widehat{\beta}_j\right)}} \tag{16}$$

dove $z_{1-\frac{\alpha}{2}}$ è il valore critico della normale standardizzata a due code.

gli intervalli di confidenza per gli effetti in scala logit possono essere traslati in intervalli di confidenza per gli OR facendo gli esponenti degli estremi dell'intervallo

pseudo-R^2

Diverse misure di bonta di adattamento sono state proposte per i modelli di scelta binaria, in questo corso introduciamo solo l'indice pseudo-R^2 dato da

$$pseudo - R^2 = 1 - \frac{l\left(\widehat{\theta}\right)}{l\left(\widehat{0}\right)} \tag{17}$$

dove $l\left(\hat{\theta}\right)$ è la log-verosimiglianza del modello specificato e $l\left(\hat{0}\right)$è la log-verosimiglianza del modello stimato con la sola intercetta. tale indice è compreso sempre nell'intervallo $[0,1)$. Vale zero se il modello con la sola intercetta è preferibile al modello stimato, e si avvicina ad 1 al crescere della distanza tra $l\left(\hat{\theta}\right)$ e $l\left(\hat{0}\right)$

Il problema

Supponiamo che un'azienda Gamma sia interessata a valutare l'efficacia che ha avuto uno spot pubblicitario su un suo prodotto lanciato sul mercato. Per questo motivo progetta un'indagine ad hoc. Una volta raccolti i dati il nostro problema quindi consiste nel regredire una variabile dicotomica (acquisto/non acquisto) su un'altra variabile dicotomica che indicizza la visione della pubblicità del prodotto.

Risultati

Supponiamo di regredire la variabile dicotomica y (acquisto: si=1, no=0) sul regressore x che indica se il cliente ha visto lo spot o meno (visto spot=1 o non visto spot=0). otteniamo i seguenti risultati

	coef.=$\widehat{\beta}$	s.e.	z value	Pr>|z|=p-value
intercetta	-0.9694	0.3441	-2.738	0.00619
x	0.9027	0.4383	2.059	0.03945

Interpretazione risultati

- la pubblicità influenza significativamente la probabilità di acquistare il prodotto; l'ipotesi nulla è rifiutata al livello di significatività del 5% (p-vale<0.05). quindi la visione dello spot ha un'influenza significativa sulle vendite.
- il segno del coeffciente beta per la pubblicità è positivo quindi se il cliente ha visto la pubblicità la probabilità di acquistare il prodotto aumenta
- ma qual'è l'effetto marginale della pubblicità sulla probabilità di acquistare?

Stima della probabiilità di acquisto

Per calcolare direttamente la probabilità stimata di acquisto:

$$P\left(Y_i = 1|x\right) = \frac{exp\left(-0.9694 + 0.9027x\right)}{1 + exp\left(-0.9694 + 0.9027x\right)} \tag{18}$$

che per i due valori assunti dal regressore x assume i seguenti valori

$$P\left(Y_i = 1|x = 1\right) = \frac{exp\left(-0.9694 + 0.9027 \times 1\right)}{1 + exp\left(-0.9694 + 0.9027 \times 1\right)} = 0.4839 \tag{19}$$

$$P\left(Y_i = 1|x = 0\right) = \frac{exp\left(-0.9694\right)}{1 + exp\left(-0.9694\right)} = 0.2750 \tag{20}$$

quindi la probabilità di acquistare è più grande per chi ha visto lo spot.

Effetto marginale di x

Poichè x è una variabile dicotomica il suo effetto marginale sulla probabilità di acquisto è facilmente calcolabile come:

$$P(Y_i = 1|x = 1) - P(Y_i = 1|x = 0) = 0.4839 - 0.2750 = 0.2089$$

Intervallo confidenza

Per costruire un intervallo di confidenza al livello di confidenza αper i singoli coefficienti basta applicare la formula, nell'esempio abbiamo (scelto un livello di confidenza pari a 95%)

$$IC\left(\beta_{95\%}\right) = [0.9027 \pm 1.96 \times 0.4383] \tag{21}$$

Interpretazione OR

Vogliamo comprendere l'effetto che ha x sulla propensione all'acquisto, ragionando in termini di OR

- passo 1: riscriviamo il predittore lineare per il nostro esempio

$$logit(\pi_i) = x_i^{'}\beta = -0.9694 + 0.9027x \tag{22}$$

- passo 2: calcoliamo l'odds associato ad x

$$\frac{\pi_i}{1-\pi_i} = \frac{P(Y_i = 1|x)}{1 - P(Y_i = 1|x)} = exp(-0.9694 + 0.9027x) \tag{23}$$

interpretazione OR

- passo 3: calcoliamo l'odds per i due valori che assume x

per x=1

$$exp(-0.9694 + 0.9027 \times 1) = 0.9361 = \frac{P(Y_i = 1|x = 1)}{1 - P(Y_i = 1|x = 1)} \tag{24}$$

per x=0

$$exp(-0.9694 + 0.9027 \times 0) = 0.3795 = \frac{P(Y_i = 1|x = 0)}{1 - P(Y_i = 1|x = 0)} \tag{25}$$

Odds-ratio (OR)

OR è il rapporto tra i due odds appena calcolati, ovvero

$$odds - ratio = 0.9361/0.3795 = 2.46 = \frac{\frac{P(Y_i=1|x=1)}{1-P(Y_i=1|x=1)}}{\frac{P(Y_i=1|x=0)}{1-P(Y_i=1|x=0)}} = exp\left(0.9027\right) \tag{26}$$

quindi la propensione all'acquisto è circa due volte e mezzo più grande se si è visto lo spot. Notare che senza fare i calcoli potevamo direttamente utilizzare $exp\left(\widehat{\beta}\right)$

intervallo confidenza per OR

per costruire un intervallo di confidenza per gli OR basta fare l'esponente della formula vista in precedenza

$$IC\left(Odds - ratio_{95\%}\right) = exp\left[0.9027 \pm 1.96 \times 0.4383\right] \tag{27}$$

Conclusione

In conclusione il manager dell'azienda Gamma decide di investire ancora in pubblicità perchè la propensione all'acquisto dei clienti che hanno visto la pubblicità è maggiore.

il problema

L'ufficio marketing di una azienda che stipula polizze sulla vita vuole aumentare il suo volume di polizze per questo motivo predispone uno studio tra i propri clienti per capire quali fattori influiscano sul rinnovo della polizza.

- variabile dipendente RINNOVO= y=(1 se si, 0 se no)
- $x_1$età del cliente
- x_2 reddito del cliente
- x_3 collocazione dell'ufficio in cui il cliente si serve (1 se in centro, 0 altrimenti)

risultati del modello logit

	stime	errore standard	χ^2	p-value	OR
intercetta	-8.4349	0.0854	9760.72	$<.0001$	
x_1	0.0223	0.0004	2967.84	$<.0001$	1.023
x_2	0.7431	0.0191	1512.45	$<.0001$	2.102
x_3	0.8237	0.0186	1862.48	$<.0001$	2.279

il modello stimato è quindi

$$logit\,(\pi_i) = -8.43 + 0.0223 x_1 + 0.74 x_2 + 0.82 x_3$$

interpretazione OR

Fissati i valori delle varibili x_1 e x_2 , l'OR per la variabile x_3 è dato da

$$OR_{x_3} = \frac{odds\,(x_3 = 1, x_1 = costante, x_2 = costante)}{1 - odds\,(x_3 = 0, x_1 = costante, x_2 = costante)} = exp(0.8237)$$

questo significa che un individuo che si serve in un ufficio in centro hanno una propensione a rinnovare la polizza 2.3 volte più grande rispetto a chi si serve altrove, mantenendo costanti le altre variabili.

interpretazione OR

La stima del coeff. della variabile x_1 è 0.0223 ciò vuol dire che per un incremento di una unità della variabile età del cliente (ad esempio, se si passa 58 a 59 anni) ci aspettiamo un incremento in logodds di 0.0223. Per un incremento di 5 unità della variabile età del cliente ci aspettiamo un incremento in logodds di 5*0.0223. In altre parole l'odds ratio associato ad incrementi di 5 anni è exp(5*0.0223). Considerato l'incremento di un anno, abbiamo $OR_{x_1} = exp\left(0.0223\right)) = 1.023$ che indica che per ogni anno in più del cliente ci aspettiamo che un incremento nell'odds pari al 2.3% mantenendo fisse le altre variabili. Un ragionamento analogo lo possiamo fare per l'altra variabile continua del modello.

Stima della probabilità e effetto marginale

Supponiamo ora di voler stimare quale sia la probabilità di rinnovare la polizza se il cliente ha 58 anni ed ha un reddito pari 50 (migliaia di euro) e si serve di un ufficio non in centro

$$P(y=1|x_1=58, x_2=50, x_3=0) = \frac{exp(-8.43+0.0223\times 58+0.74\times 50+0.82\times 0)}{1+exp(-8.43+0.0223\times 58+0.74\times 50+0.82\times 0)}$$

qual è l'effetto marginale di x_3 sulla probabilità di rinnovo della polizza?

$$P(y=1|x_1=58, x_2=50, x_3=1) - P(y=1|x_1=58, x_2=50, x_3=0)$$

Conclusione

L'ufficio marketing conclude che

- età, reddito influiscono sul rinnovo della polizza e all'aumentare di questi la probabilità di rinnovo aumenta, quindi sceglieranno di promuovere la polizza a clienti che hanno disponibilità finanziarie e non troppo giovani
- anche l'ubicazione dell'ufficio in centro gioca un ruolo importante quindi in una politica di espansione di uffici si cercheranno sedi nella zona centrale

Econometric analysis / William H. Greene. 6th ed., Upper Saddle River, N.J. : Prentice Hall, 2008

Introduction to econometrics / James H. Stock, Mark W. Watson. 2nd ed., Boston: Pearson Addison Wesley, 2007.

Bracalente B., Cossignani M., Mulas A., 2009, Statistica Aziendale, McGraw-Hill.