

BIBLIOTECA DI SCIENZE STATISTICHE

SERVIZIO BIBLIOTECARIO

BID P0V0942084 BID

ACQ. 674 / '03 INV. 83739

COLL. 5-coll. WP. 7/2001

**LA TARIFFA INCENDIO E FURTO
PER LE AUTO: UN MODELLO DI
COX**

B. Scarpa

2001.7

**Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 241-243
35121 Padova**

Aprile 2001

BIBLIOTECA DI SCIENZE
IN VECCHIA BIBLIOTECA
LIB. 60.201.10.10
ACC. 67.1.10.10
CON 2.000.10.10

LA TARIFFA INCENDIO E FURTO
PER L'AUTO: UN MODELLO DI

(OX)

B. Scarp

2001.7

Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 241-243
35121 Padova

Aprile 2001

La tariffa Incendio e Furto per le auto: un modello di Cox

Bruno Scarpa

Dipartimento di Scienze Statistiche
Università di Padova

Ottobre 2000

Introduzione

La determinazione di una tariffa per la garanzia Incendio e Furto delle auto viene generalmente fatta valutando l'impatto di alcune variabili di personalizzazione sul tasso puro o su altri indicatori di rischio come la frequenza dei sinistri o indicatori di controllo come il rapporto sinistri a premi (Daboni, 1993).

In generale tale impatto viene misurato attraverso procedure descrittive basate su semplici tabelle o attraverso costruzione di modelli statistici parametrici che descrivano al variare delle variabili di personalizzazione il tasso puro o frequenza e costo medio congiuntamente. Un tale approccio considera indistintamente sinistri di differente grado, considerando sostanzialmente analoghi i rischi di avere un danno totale o uno parziale. Può essere interessante cercare di studiare in maniera più approfondita la relazione tra il diverso tipo di rischio e le variabili esplicative.

Una possibilità consiste nell'utilizzare alcune tecniche basate sulla cosiddetta "analisi di sopravvivenza" per modellare il grado medio di danno. Tali metodologie consentono di valutare in maniera approfondita e dettagliata la tariffa in vigore, e permettono di misurare il rischio in modo da quantificarlo e quotarlo.

La variabile *grado di danno* per ogni polizza può essere modificata in maniera tale da comprendere gli scoperti e i minimi. Si può così ottenere un modello che raccolga in se tutte le componenti importanti per la descrizione del problema tariffario per l'incendio e furto delle auto.

Si ottiene così una tabella di tassi puri al variare delle modalità assumibili dalle variabili rilevanti per descrivere tale rischio, che potrebbero essere applicati al valore commerciale, ottenendo così il premio puro come da tariffa.

Le variabili e i dati

La variabile principale che si utilizza in questa analisi e che caratterizza il metodo per ottenere la tariffa qui proposto è il cosiddetto *grado di danno* per polizza.

Tale variabile misura la percentuale di oggetto assicurato che subisce il danno. Si definisce infatti a priori il valore assicurato del veicolo e si misura poi il costo del sinistro.

Per comprendere meglio questa variabile è utile studiare la relazione tra il *grado di danno* e altre caratteristiche degli assicurati che possono essere rilevanti per l'incendio e il furto delle auto. Un tale studio può risultare, inoltre, utile per raggruppare le variabili in classi omogenee in quanto a rischio di incendio e furto e, ad una compagnia di assicurazioni, per aggiornare la tariffa in vigore.

Analogamente ad altre garanzie assicurative (cfr. Scarpa, 2000), il grado di danno per incendio o furto di un'auto presenta due caratteristiche principali che certamente influenzano l'analisi statistica.

Da una parte, tra gli assicurati, la percentuale di sinistrati è molto bassa (meno del 3% nei dati analizzati) ed è unicamente tra questi che, ovviamente, viene rilevato il grado di danno (è possibile assegnare grado di danno 0 a tutte le polizze non sinistrate).

L'altra caratteristica interessante riguarda le informazioni presenti nei data set disponibili nelle aziende che riflettono le caratteristiche delle tariffe finora vigenti. La garanzia incendio e furto dell'auto viene infatti venduta con scoperto e minimo. Chi, cioè, ha un sinistro che assegna un danno inferiore al massimo tra scoperto e minimo non viene risarcito; tipicamente, quindi, tale sinistro non viene neppure denunciato. Tale caratteristica dei dati a disposizione non fornisce una informazione accurata per stimare le probabilità (o le frequenze) di avere un basso grado di danno. Questa probabilità può essere utile per diversi scopi, sia di carattere conoscitivo (conoscere la distribuzione di tutti i gradi di danno) sia di carattere decisionale (la conoscenza di tali probabilità potrebbe indurre a cambiare in qualche maniera la tariffa in termini di scoperti).

In prima battuta si vuole stimare la distribuzione probabilistica del grado di danno per le polizze in un portafoglio di una compagnia assicurativa; un secondo (anche se non secondario) obiettivo raggiunto dal presente lavoro riguarda l'ipotesi di una tariffa (puramente) tecnica per coprire i costi di incendio e furto delle auto.

Come "sottoprodotto" dell'analisi si ottiene anche una riclassificazione di alcune variabili che vengono raggruppate in classi più o meno omogenee in quanto a rischio.

Gli strumenti che vengono utilizzati si basano su un approccio inferenziale al problema che cerca di prevedere attraverso un modello statistico la variabile risposta *grado di danno* (noto solamente a sinistro avvenuto) usando alcune variabili note al momento di stipulare la polizza che in qualche modo aiutano a stimare il rischio (nelle compagnie di assicurazione si parla di "personalizzazione delle tariffe", Klugman e Hogg, 1984, Lemaire, 1995).

Vista la particolare struttura della variabile risposta in esame è sembrato utile cercare di utilizzare un approccio semiparametrico in cui non si fanno assunzioni particolarmente forti sulla distribuzione della variabile risposta. Il modello a rischi proporzionali di Cox (cfr. ad es. Cox e Oakes, 1984) fornisce uno strumento utile per fare inferenza lasciando suggerire alle osservazioni la forma stessa della distribuzione della probabilità del *grado di danno*.

Il modello proposto non sembra avere problemi nel tener conto di informazioni parziali che sono fornite per esempio dalla percentuale di scoperto/minimo. In tal caso infatti è noto che tutti gli assicurati che non hanno denunciato alcun sinistro non hanno avuto infortuni con grado di invalidità permanente superiore alla loro scoperto/minimo; non è però noto se quelle polizze non hanno riportato davvero nessun sinistro o se ne hanno avuto uno di basso grado. Una informazione parziale del tipo appena descritto può essere vista come una forma di *censura* (Cox e Oakes, 1984) e può essere trattata e utilizzata in maniera particolarmente semplice da stimatori attuariali tipo *life-table* (Lawless, 1982) e dai modelli che vengono utilizzati nel seguito.

Per rendere più agevole il calcolo senza avere alcuna conseguenza pratica nei risultati si preferisce utilizzare una semplice trasformazione della variabile risposta. Anziché utilizzare il *grado di danno* si utilizzerà il complemento a 100 di tale grado. Tale trasformazione permette di avere una censura a destra anziché a sinistra, situazione più maneggevole, in particolare, perché permette di utilizzare software standard per l'analisi dei dati di durata.

Nella tariffa di molte compagnie di assicurazione, le variabili che vengono utilizzate per la personalizzazione sono la zona territoriale (12 zone), la classe di rischio (4 classi) e il tipo di alimentazione (diesel o benzina).

Per una prima valutazione del nuovo metodo proposto si vuole provare la procedura utilizzando le stesse variabili come variabili esplicative del modello.

Anziché utilizzare tutte le polizze presenti negli archivi aziendali si preferisce, in prima battuta, estrarre un campione casuale dalle polizze del 1997 e provare il modello su tale nuovo insieme di dati.

Si è scelto di ottenere un campione di 168.187 polizze dell'anno 1997. Di queste 168.187 polizze solo 2.282 hanno avuto un incendio o un furto con costo superiore al loro scoperto e minimo, le rimanenti 165.905 (circa il 98,64%) polizze vengono così considerate censure.

Analisi preliminare

Per conoscere meglio le relazioni tra le variabili a disposizione vengono calcolati alcuni test non-parametrici che verificano l'ipotesi che, marginalmente, le variabili esplicative influiscano nel grado di danno (in particolare sembra adatto ai dati disponibili il test *Log Rank*).

Le variabili utilizzate nel seguito (scelte come primo esempio) sono *tipo di alimentazione*, *settore territoriale* e *classe di rischio RAS*.

Le variabili esplicative

In tabella 1 si trovano le statistiche test e i livelli di significatività osservati per i test Log Rank, di Wilcoxon e di rapporto di verosimiglianza marginali sulla diversità del rischio al variare delle diverse variabili esplicative considerate.

Tabella 1. Testi non-parametrici univariati

		Log Rank			Wilcoxon			-2log(LR)		
		Statistic a	Gradi di libertà	p-val ue	Statistic a	Gradi di libertà	p-value	Statistic a	Gradi di libertà	p-val e
Tipo alimentazione	di	169.755 9	1	0.000 1	86.4470	1	0.0001	45.9076	1	0.0001
Settore territoriale		659.128 2	10	0.000 1	514.245 8	10	0.0001	267.459 2	10	0.0001
Classe di rischio		15.2444	3	0.000 1	406.106 3	3	0.0001	218.777 0	3	0.0001

Come è noto, l'assunzione fondamentale per il modello di Cox è che le *funzioni di rischio*, $h(x)$, (cfr. Cox e Oakes, 1984) siano proporzionali al variare delle modalità delle variabili esplicative. Per verificare, almeno approssimativamente, tale assunto si è controllato che i

grafici delle trasformate logaritmiche della *funzione di sopravvivenza* e della sua iterazione (log-log) al variare delle varie modalità, non si incrociano.

Il modello

Le analisi preliminari condotte, rilevando una sostanziale diversità di comportamento, almeno marginale, tra i diversi segmenti della popolazione di assicurati in esame, aiutano a determinare un modello che possa descrivere in maniera adeguata la relazione tra il 'rischio' (la probabilità di avere un incendio o un furto dell'auto con un fissato costo percentuale) e le variabili esplicative scelte.

In prima approssimazione ci si sofferma solo su poche e semplici variabili esplicative, anche se sembra possibile aggiungere al modello, senza grosse difficoltà, alcune altre informazioni presenti nei data set aziendali.

Nell'ipotesi che le *funzioni di rischio (hazard function)*¹ rispetto al grado di danno siano proporzionali al variare delle modalità delle variabili esplicative considerate, si può scrivere un modello di Cox e stimare attraverso un approccio di verosimiglianza parziale i parametri del modello senza ipotizzare alcuna forma predefinita per la distribuzione della variabile *grado di danno* (per maggiori informazioni al riguardo si veda per esempio Cox e Oakes, 1984).

Nella tabella 2 vengono presentate le stime dei parametri, gli *standard error*, il test di Wald sulla significatività di questi parametri e il relativo livello di significatività osservato e il rapporto dei rischi (odds ratio), misura quest'ultima che fornisce in maniera immediata il rapporto tra il grado di danno con o senza la modalità considerata.

¹ Si osservi che *funzione di rischio* è un nome tecnico che traduce l'*hazard function* inglese, che non ha nulla a che fare col concetto di rischio che si utilizza nel mondo attuariale ed assicurativo, per quel che interessa nel presente contesto indica semplicemente una particolare probabilità limite al variare del grado di danno. Una tale funzione caratterizza completamente una variabile aleatoria, come la funzione di ripartizione o la funzione caratteristica, ma è di facile calcolo per variabili aleatorie positive.

Tabella 2. Stime dei parametri con relativi *standard error* e test di significatività

Variabile	Gradi di libertà	Stima dei parametri	Standard Error	Chi-quadro di Wald	Pr> Chi-quadro	Rapporto dei rischi
Tipo alimentazione	di 1	0.379570	0.08098	21.96847	0.0001	1.462
Medio rischio	1	0.387381	0.06407	36.55833	0.0001	1.473
Rischio	1	0.928835	0.07044	173.87611	0.0001	2.532
Alto rischio	1	1.304507	0.10286	160.84898	0.0001	3.686
Settore territoriale 2	1	0.240158	0.15549	2.38560	0.1225	1.271
Settore territoriale 3	1	0.195183	0.18674	1.09244	0.2959	1.216
Settore territoriale 4	1	0.087854	0.13610	0.41666	0.5186	1.092
Settore territoriale 5	1	0.477045	0.18479	6.66414	0.0098	1.611
Settore territoriale 7	1	0.747688	0.16174	21.36968	0.0001	2.112
Settore territoriale 8	1	0.691882	0.13399	26.66376	0.0001	1.997
Settore territoriale 9	1	1.087115	0.14085	59.56761	0.0001	2.966
Settore territoriale 10	1	1.401326	0.13042	115.44245	0.0001	4.061
Settore territoriale 11	1	1.330129	0.21744	37.41991	0.0001	3.782
Settore territoriale 12	1	1.626676	0.16225	100.52093	0.0001	5.087

Per verificare l'ipotesi nulla globale che tutti i parametri assumono valore zero, ed avere così una misura generale di adattamento del modello sono stati effettuati i soliti tre test: il rapporto di verosimiglianza, il test basato sullo *score* e il test di Wald.

Criterio	Senza esplicative	Con esplicative	Chi-quadro del modello
-2 log L	52608.610	51606.146	1002.464 con 14 gl (p=0.0001)
Score	.	.	1137.967 con 14 gl (p=0.0001)
Wald	.	.	1045.255 con 14 gl (p=0.0001)

La verifica del modello è stata fatta semplicemente con alcune analisi grafiche. I residui della devianza (cfr. Kalbfleish e Prentice, 1980) sono stati disegnati al variare del grado di invalidità previsto dal modello. La sostanziale casualità di tali residui e l'assenza di pattern riconoscibili induce a ritenere il modello almeno in prima battuta soddisfacente. Anche il grafico dei residui di martingala (Therneau e al., 1990) è stato fatto, e, anche in questo caso, non si riconoscono forme particolarmente allarmanti.

Il modello stimato viene allora considerato valido e può essere utilizzato, innanzitutto, per stimare la distribuzione del grado di danno per la garanzia incendio e furto delle auto in un portafoglio analogo a quello analizzato, ma può anche essere utile per prevedere tale distribuzione in un qualsiasi altro portafoglio (può essere utile ad una compagnia di assicurazione per determinare in che direzione si vuole andare, quali clienti vuole attirare e quali scoraggiare).

Una tariffa personalizzata

Un secondo utilizzo di un modello appena visto può essere quello di proporre una tariffa personalizzata con scoperti (o anche senza).

Determinare una tariffa per l'Incendio e Furto significa stimare il tasso di rischio per ciascun segmento della popolazione in esame.

Il modello stimato nella sezione precedente fornisce una stima dell'intera distribuzione del grado di danno per Incendio o Furto dell'auto (e non solo della parte osservabile a causa degli scoperti), al variare delle variabili di segmentazione considerate. La conoscenza di tale distribuzione permette di ottenere delle misure globali di posizione come media o mediana.

Un tale indicatore medio (in particolare se si considera il valore atteso) misura, allora, per ciascun assicurato il suo costo atteso.

È possibile ottenere una stima della tariffa determinando diverse ipotesi di percentuale di scoperto o di minimo da applicare agli assicurati.

In particolare, se si vuole tener conto di questi scoperti e minimi, è sufficiente calcolare il valore atteso della variabile che si ottiene assegnando valore zero anche ai gradi di danno inferiori al nuovo scoperto o minimo.

In questa maniera per ogni segmento di popolazione determinato dalle variabili di personalizzazione si ottiene un tasso da applicare al valore assicurato di ciascuno.

Per esempio, si è simulata una tariffa con gli scoperti tuttora in vigore (cfr. tabella 3, la prima riga di ciascuna cella indica la percentuale di scoperto applicata, la seconda riga indica l'ammontare di minimo applicato; il costo del sinistro per l'assicurazione è al netto del massimo tra percentuale di scoperto applicata al valore assicurato e ammontare del minimo di tariffa).

In tabella 4 sono riportati i tassi per mille per i diversi segmenti di popolazione (per auto alimentate a Benzina – tabella 4a – e a Diesel – tabella 4b). Come variabili di segmentazione oltre al tipo di carburante si è considerato il livello di rischio e la zona geografica.

Per avere un termine di confronto, in tabella 5 (a e b) si presentano gli analoghi tassi per una tariffa in cui lo scoperto viene fissato pari al 5% per ogni combinazione di Settore Territoriale e Livello di Rischio.

Ancora per effettuare confronti in tabella 6 sono riportati i tassi osservati per i vari segmenti di popolazione considerati (riportano il rapporto tra costo dei sinistri e valore assicurato per ciascun segmento senza considerare l'effetto congiunto stimato dal modello).

Si osservi come la stima grezza offerta nella tabella 6 sia molto più variabile anche in maniera non monotona (controintuitivo, almeno per quanto riguarda i livelli di rischio).

È interessante anche osservare come il cambio di franchigia (confronto tra le tabelle 4 e 5) comporti una diversità non così significativa (per alto rischio la percentuale di scoperto passa dal 20-25% al 5% con un aumento dei tassi di 5-6 punti percentuale nei casi peggiori).

Tabella 3. Scoperti e Minimi nella tariffa vigente

Scoperti e Minimi nella Tariffa vigente				
Zona	Livelli di Rischio			
	Basso Rischio	Medio Rischio	Rischio	Alto Rischio
1	5 300000	5 300000	10 500000	15 500000
2	5 300000	5 300000	15 500000	20 800000
3	5 300000	5 300000	15 500000	20 800000
4	5 300000	5 300000	15 500000	20 800000
5	5 300000	5 300000	15 500000	20 800000
6	5 300000	10 500000	15 500000	20 800000
7	5 300000	10 500000	15 500000	20 800000
8	10 500000	10 500000	15 500000	20 800000
9	10 500000	10 500000	15 500000	20 800000
10	10 500000	10 500000	15 500000	20 800000
11	15 500000	20 800000	25 1000000	25 1000000
12	15 500000	20 800000	25 1000000	25 1000000

Tabella 4. Tassi per mille per i diversi segmenti di popolazione con scoperti e minimi di tariffa

a) Tipo di alimentazione Benzina

Tasso puro netto stimato				
Livelli di Rischio				
Zona	BR	MR	R	AR
1	1.53045	2.25331	3.33243	4.6081
2	1.94529	2.86365	4.02539	5.6923
3	1.85986	2.73797	3.84873	5.4427
4	1.67082	2.45985	3.45777	4.8902
5	2.4643	3.6270	5.0984	7.2080
6				
7	3.2284	4.09392	6.6773	9.4371
8	2.63042	3.87220	6.3161	8.9273
9	3.90266	5.74309	9.3623	13.2243
10	5.33910	7.85391	12.7948	18.0596
11	4.59590	6.76250	11.3300	16.4514
12	6.17723	9.08571	15.2099	22.0645

b) Tipo di alimentazione Diesel

Tasso puro netto stimato				
Livelli di Rischio				
Zona	BR	MR	R	AR
1	2.2358	3.2910	4.8666	6.7278
2	2.8414	4.1815	5.8778	8.3086
3	2.7167	3.9982	5.6201	7.9448
4	2.4407	3.5924	5.0497	7.1393
5	3.5989	5.2948	7.4427	10.5171
6				
7	4.7135	5.9775	9.7438	13.7620
8	3.8421	5.6541	9.2175	13.0202
9	5.6985	8.3819	13.6526	19.2668
10	7.7931	11.4562	18.6420	25.6939
11	6.9017	9.8681	16.5151	23.9505
12	9.2723	13.2507	22.1497	32.0781

Tabella 5. Tassi per mille per i diversi segmenti di popolazione con scoperti e minimi al 5%

a) Tipo di alimentazione Benzina

Tasso puro netto stimato				
Zona	Livelli di Rischio			
	BR	MR	R	AR
1	1.53045	2.25331	3.86761	5.62365
2	1.94529	2.86365	4.91359	7.14198
3	1.85986	2.73797	4.69826	6.82951
4	1.67082	2.45985	4.2216	6.1377
5	2.4643	3.6270	6.2208	9.0380
6				
7	3.2284	4.7503	8.1425	11.8221
8	3.0535	4.4933	7.7031	11.1858
9	4.5286	6.6604	11.4049	16.5405
10	6.1927	9.1025	15.5660	22.5432
11	5.7690	8.4810	14.5080	21.0185
12	7.7488	11.3835	19.4429	28.1207

b) Tipo di alimentazione Diesel

Tasso puro netto stimato				
Zona	Livelli di Rischio			
	BR	MR	R	AR
1	2.2358	3.2910	5.6455	8.2038
2	2.8414	4.1815	7.1697	10.4133
3	2.7167	3.9982	6.8560	9.9587
4	2.4407	3.5924	6.1615	8.9521
5	3.5989	5.2948	9.0731	13.1691
6				
7	4.7135	6.9318	11.8679	17.2092
8	4.4585	6.5574	11.2292	16.2865
9	6.6089	9.7128	16.6043	24.0383
10	9.0322	13.2628	22.6298	32.6945
11	8.4154	12.3598	21.0993	30.4992
12	11.2957	16.5732	28.2282	40.7053

Tabella 6. Tassi per mille per i diversi segmenti di popolazione osservati

a) Tipo di alimentazione Benzina

Zona	Tasso puro netto stimato			
	BR	MR	R	AR
1	0.0256	1.8687	1.9433	0.4165
2	0.4994	1.5621	2.3233	0.6352
3	1.0906	1.4623	2.2402	2.1363
4	1.6319	1.3921	6.7874	6.5613
5	0.1990	4.2228	1.4714	.
6
7	0.8738	3.0431	3.5881	17.5975
8	3.4969	3.8184	5.6605	14.7935
9	3.7344	6.4461	7.6354	15.3380
10	6.1121	7.8138	13.1600	21.0632
11	8.9387	4.7287	9.5690	65.2622
12	12.3335	16.7271	17.1619	34.9428

b) Tipo di alimentazione Diesel

Zona	Tasso puro netto stimato			
	BR	MR	R	AR
1	.	0.1062	0.2390	3.8559
2	.	3.11165	5.3516	17.1144
3	.	9.75767	0.8360	18.8849
4	.	4.0727	2.8563	20.8377
5	.	7.6086	5.5798	21.7356
6
7	.	7.6086	0.4448	8.7937
8	.	12.4354	6.0048	17.7387
9	.	5.3159	6.9432	15.4221
10	.	16.9825	6.4021	23.1262
11	.	15.4108	8.6078	15.6228
12	.	29.9387	14.7090	20.1165

Conclusione e prospettive

Anche nel caso della garanzia Incendio e furto delle auto i risultati ottenuti sembrano essere soddisfacenti sia come metodologia adottata, sia per i risultati facilmente applicabili, e modulabili a seconda delle esigenze della compagnia.

Nell'ottica di approfondire un tale approccio si possono considerare alcune linee di sviluppo:

- È certamente di interesse cercare di ottenere stime per il modello che siano pesate con i rischi/anno
- È possibile ottenere delle misure di variabilità delle stime dei tassi ottenuti, risulta così possibile verificare l'effettiva diversità tra i diversi tassi in una scala di misura appropriata
- È possibile inserire altre variabili esplicative nel modello che possano aiutare maggiormente a prevedere il 'rischio'
- È utile inoltre effettuare alcune verifiche del modello e della stima della tariffa su un data set diverso da quello usato per ottenere le stime, per avere la sicurezza della bontà della procedura adottata.

Bibliografia

Cox, D.R., Oakes D., 1984, Analysis of Survival Data, Chapman & Hall.

Daboni, L., 1993, Tecniche attuariali delle assicurazioni contro i danni, LINT, Trieste.

Kalbfleisch, J.D., Prentice, R.L., 1980, The Statistical Analysis of Failure Time Data, John Wiley & Sons, Inc. New York

Klugman, S.A., Hogg, R.V., 1984, Loss Distribution, Wiley.

Lawless, J.F., 1982, Statistical Models and Methods for Lifetime Data, Wiley

Lemaire, J., 1995, Bonus–Malus Systems in Automobile Insurance. Kluwer, Boston.

Scarpa, B., 2000, Il grado di invalidità permanente nella polizza infortuni: una valutazione di variabili predittive mediante modello a rischi proporzionali di Cox, Working Paper del Dipartimento di Scienze Statistiche dell'Università di Padova

Therneau, T.M., Grambsch, P.M. e Fleming, T.R., 1990, Martingale–Based Residuals and Survival Models, Biometrika, 77, 147–160