

BIBLIOTECA DI SCIENZE STATISTICHE

SERVIZIO BIBLIOTECARIO NAZIONALE

BID P050942088 BID

ACQ. 645/103 INV. 83740

COLL. 5-coll. VI P. 6/2001

**IL GRADO DI INVALIDITÀ
PERMANENTE NELLA POLIZZA
INFORTUNI: UNA VALUTAZIONE DI
VARIABILI PREDITTIVE MEDIANTE
MODELLO A RISCHI PROPORZIONALI
DI COX**

B. Scarpa

2001.6

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

Aprile 2001

BIBLIOTECA DI SCIENZE STATISTICHE
CENTRO BIBLIOTECARIO NAZIONALE
via Lovatich 10
00144 ROMA
COL. 2-666 V. CLASSE 1904

IL GRADO DI VALIDITÀ
PERMANENTE NEL POLIZI
INFORME UNA VALUTAZIONE IN
VARIABILI TRIDIMENSIONALI
MODELLO A RISCHI PROPORZIONALI
DI COX

B. Scapè

1981-8

Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 241-213
35121 Padova

Aprile 2001

Il grado di invalidità permanente nella polizza infortuni: una valutazione di variabili predittive mediante modello a rischi proporzionali di Cox

Bruno Scarpa

Dipartimento di Scienze Statistiche

Università di Padova

Introduzione

Quando avviene un sinistro legato ad una polizza infortuni di invalidità permanente, viene associata ad ogni infortunio una misura della gravità del sinistro: il *grado di invalidità permanente*.

Chiaramente il costo del sinistro per la compagnia di assicurazione è direttamente legato al *grado di invalidità permanente*, che è quindi la variabile fondamentale per determinare la tariffa di tale garanzia in polizza (Daboni, 1993).

Per comprendere meglio questa variabile, è utile studiare la relazione tra il *grado di invalidità permanente* e altre caratteristiche degli assicurati che possono essere rilevanti per gli infortuni. Un tale studio può risultare, inoltre, utile per raggruppare le variabili in classi omogenee in quanto a rischio di invalidità permanente e per aggiornare la tariffa in uso in una particolare compagnia.

Il grado di invalidità permanente, come rilevato dalle compagnie di assicurazione, oggi, in Italia, 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 2% nei dati analizzati) ed è unicamente tra questi che, ovviamente, viene rilevato il grado di invalidità (normalmente per i assicurati non sinistrati si ipotizza grado di invalidità 0).

L'altra caratteristica interessante riguarda la franchigia che caratterizza le tariffe finora vigenti. La garanzia invalidità permanente viene, infatti, venduta con una franchigia di qualche punto percentuale (1%, 3%, 5%,?). Chi, cioè, ha un sinistro che assegna un'invalidità permanente inferiore alla franchigia non viene risarcito; tipicamente, quindi, tale sinistro non

viene neppure denunciato. Tale caratteristica dei dati a disposizione non fornisce, quindi, una informazione accurata per stimare le probabilità di avere un basso grado di invalidità permanente. Questa probabilità può essere utile per diversi scopi, sia di carattere conoscitivo (conoscere la distribuzione di tutti i gradi di invalidità) sia di carattere decisionale (la conoscenza di tali probabilità potrebbe indurre a cambiare in qualche maniera la tariffa in termini di franchigie).

Questo lavoro, nel tenere conto delle caratteristiche dei dati a disposizione, si propone almeno due obiettivi.

In prima battuta si vuole proporre un nuovo metodo per stimare la distribuzione di probabilità del grado di invalidità permanente a partire da un portafoglio clienti di una compagnia di assicurazione; un secondo (anche se non secondario) obiettivo raggiunto dal presente lavoro riguarda l'ipotesi di una tariffa (puramente) tecnica per coprire i costi dell'invalidità permanente.

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 invalidità permanente* (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 invalidità permanente*.

Il modello proposto non sembra avere problemi nel tener conto di informazioni parziali che sono fornite per esempio dalla percentuale di franchigia. 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 franchigia; 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 effettuare le stime si utilizzano i dati provenienti dal portafoglio di una compagnia di assicurazione.

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 invalidità permanente* 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.

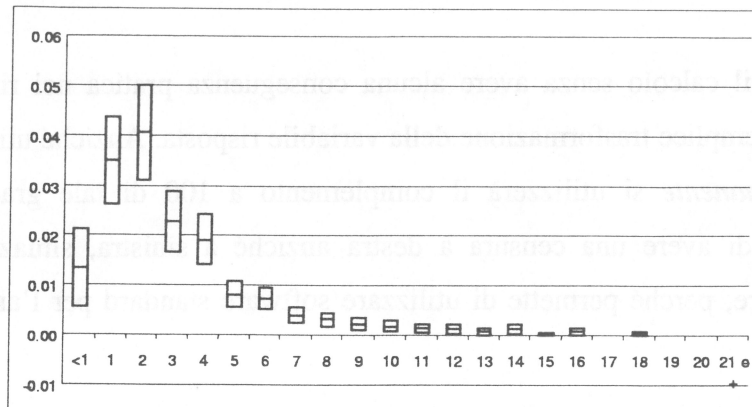
Analisi preliminare

Per conoscere un po' meglio le variabili a disposizione e le relazioni tra loro, vengono effettuate alcune semplici analisi descrittive e grafiche.

Per stimare la distribuzione della percentuale di invalidità permanente di solito si utilizza il semplice stimatore basato sulla proporzione di assicurati sinistrati per classe di percentuale di invalidità rispetto a tutti i sinistri chiusi.

Tale stimatore presenta però un problema non indifferente se si osserva che ciascuna polizza presenta percentuali di franchigia differenti. Le stime delle invalidità permanenti inferiori al 5-7% allora risultano sottostimate o comunque non corrette. E' però possibile cercare di sfruttare al meglio tutte le informazioni disponibili utilizzando una stima, basata sullo stimatore attuariale *life-table*, della distribuzione di probabilità, che viene mostrata nella figura 1.

Figura 1 Stima della funzione di densità della percentuale di permanente con le bande di confidenza asintotiche.



E' stata analizzata con il medesimo metodo la stessa distribuzione anche al variare delle modalità delle variabili di interesse.

Le variabili utilizzate nel seguito sono *sex*, *età*, *zona geografica* e *categoria di rischio* (una classificazione di rischio in uso dalla compagnia di interesse).

Le due variabili *età* e *zona geografica* vengono raggruppate in classi omogenee per meglio adattarsi al problema in esame: è questa la via usuale nella creazione di tariffe.

Il raggruppamento delle età

Per decidere come raggruppare in classi le età si è valutata la distribuzione marginale della percentuale di infortunio permanente al variare delle età e si è deciso di suddividere in quattro classi tale variabile facendo seguito a ciò che l'evidenza empirica suggerisce.

Le età sono state così suddivise: 0-14; 15-24; 25-65; >66.

Il raggruppamento delle province

Per determinare una nuova variabile che raggruppa le province omogenee rispetto al rischio, al variare della percentuale di invalidità permanente, è stato adattato un modello a rischi proporzionali, con le variabili *sex*, *categoria di rischio* ed *età* raggruppata in classi assieme con la variabile *provincia* che così è stata ordinata in termini di rischio di invalidità permanente al netto delle altre variabili¹.

Le province sono state raggruppate quindi in 3 zone a seconda della loro rischiosità:

¹ Ovviamente la scelta della provincia di riferimento influisce sull'ordinamento e sulle stime, si è scelto allora di stimare il modello in tre casi distinti, tenendo come province di riferimento Ascoli, Napoli e Genova: una nella parte bassa, una nella parte media e una nella parte alta della classifica della rischiosità.

Gruppo 1: EN, VC, OR, SR, BZ, NA, BR, CL, AG, MN, AT, UD, NO, SO, VI, RG, PN, TE, SA, FE, PA

Gruppo 2: AO, PD, AL, FG, PV, ME, CS, SS, CN, RC, VA, CB, CE, PE, BA, MC, LE, PZ, RE, TO, CT, CO, VR, TA, VE, BL, PC, NU, AV, RI, PT, TN, SI, CA, RO, PI

Gruppo 3: TV, BS, PR, MI, FR, AP, BG, LI, AN, FO, FI, MT, CH, TR, BN, IS, RA, PS, MO, VT, PG, RM, CZ, IM, TS, CR, LT, AR, GR, LU, SV, MS, BO, AQ, GO, SP, GE.

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 variabili esplicative considerate.

Si è aggiunta anche una riga che considera le zone geografiche usate da un'altra compagnia concorrente, in modo da confrontare le statistiche test ottenute per le due diverse classificazioni territoriali.

Tabella 1. Testi non-parametrici univariati

| | Log Rank | | | Wilcoxon | | | -2log(LR) | | |
|---------------------------------------|----------------|---------------------|-------------|------------|---------------------|---------|--------------|---------------------|-------------|
| | Statistic a | Gradi di libertà | p-val ue | Statistica | Gradi di libertà | p-value | Statistica | Gradi di libertà | p-valu e |
| Sesso | 29.9936 | 1 | 0.000 | 36.1478 | 1 | 0.0001 | 54.3266 | 1 | 0.0001 |
| Zona geografica | 91.0631 | 2 | 0.000 | 76.3238 | 2 | 0.0001 | 128.585 | 2 | 0.0001 |
| Classi di Età | 15.2444 | 3 | 0.001 | 17.1158 | 3 | 0.0007 | 20.3945 | 3 | 0.0001 |
| Zona geografica altra compagnia | 134.372 1 | 2 1 | 0.000 1 | 115.3152 | 2 | 0.0001 | 114.925 3 | 2 | 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 infortunio con un determinato grado) 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 disponibili alle compagnie (per esempio la differenza città-provincia o lo stato civile dell'assicurato?).

Nel valutare il modello con le 4 variabili *sesso, categoria di rischio, classe di età e classe territoriale* congiuntamente considerate si è osservato che la variabile categoria di rischio (che tra l'altro dovrebbe essere la più discriminante) in realtà ha un'influenza sul grado medio (ma anche sulla frequenza, visto che il modello considera il grado medio dell'assicurato in generale e non solo del sinistrato) molto ridotta. Solamente la categoria ad alto rischio sembra aumentare in maniera considerevole e significativa il grado dell'invalidità permanente; gli assicurati appartenenti alle altre 3 categorie, invece, sembrano sostanzialmente comportarsi allo stesso modo in quanto a rischio di invalidità permanente. Queste tre categorie vengono allora raggruppate in un unico gruppo.

Nell'ipotesi allora che le *funzioni di rischio (hazard function)* rispetto al grado di invalidità permanente, 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 invalidità permanente* (cfr. per esempio Cox e Oakes, 1984).

In 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 invalidità con o senza la modalità considerata.

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 |
|----------------|------------------|---------------------|----------------|--------------------|----------------|---------------------|
| SESSO | 1 | 0.228772 | 0.04004 | 32.64138 | 0.0001 | 1.257 |
| CATEGORIA | 1 | 1.084855 | 0.44771 | 5.87157 | 0.0154 | 2.959 |
| RISCHIO | | | | | | |
| ZONA 2 | 1 | 0.541625 | 0.06362 | 72.48547 | 0.0001 | 1.719 |
| ZONA 3 | 1 | 1.001978 | 0.05835 | 294.87182 | 0.0001 | 2.724 |
| ETÀ 15-24 | 1 | 1.557065 | 0.57968 | 7.21497 | 0.0075 | 4.745 |
| ETÀ 25-65 | 1 | 1.443706 | 0.57771 | 6.24513 | 0.0125 | 4.236 |
| ETÀ >66 | 1 | 1.620805 | 0.58408 | 7.70048 | 0.0055 | 5.057 |

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 | 90898.630 | 90411.218 | 487.412 con 7 gl (p=0.0001) |
| Score | | | 452.251 con 7 gl (p=0.0001) |
| Wald | | | 428.541 con 7 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 nel grafico dei residui di martingala (Therneau e al., 1990) 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 invalidità permanente per infortunio in un portafoglio analogo a quello analizzato, ma può anche essere utile per prevedere tale distribuzione in un qualsiasi altro portafoglio (tipicamente può essere utile per determinare in che direzione andare, chi cercare di attirare e chi rigettare).

Una tariffa personalizzata

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

Predisporre una tariffa per l'invalidità permanente nella polizza infortuni, significa prevedere la percentuale di infortuni per ogni segmento di popolazione considerato.

Tale percentuale può essere ottenuta utilizzando i risultati del modello adattato nella sezione precedente.

Avere una stima dell'intera distribuzione del grado di invalidità permanente (e non solo della parte osservabile a causa della franchigia), infatti, 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, per ciascun assicurato, il suo costo atteso per la compagnia (Daboni, 1993).

In particolare, se si vuole tener conto della franchigia che si prevede di applicare alle polizze future, è sufficiente calcolare il valore atteso della variabile che si ottiene assegnando valore zero anche ai gradi di invalidità permanente inferiori alla nuova franchigia. Tale variabile e la sua distribuzione è facilmente calcolabile a partire dalle stime del modello appena ottenuto.

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 franchigia al 5% (fortemente!) personalizzata. In tabella 3 sono riportati i tassi per mille per i diversi segmenti di popolazione. Le prime 4 righe indicano le classi per le 4 variabili di segmentazione, Sesso (M e F), Categoria di rischio (1,2,3,4), Gruppo territoriale (1,2,3, cfr. analisi preliminare) e Classe di età (1:0-14; 2:15-24; 3:25-65; 4:>66). L'ultima riga presenta la stima dei tassi di rischio ottenuti calcolando la media della variabile grado di invalidità permanente stimato utilizzando il modello presentato nella sezione precedente dopo aver assegnato valore nullo ai gradi di invalidità permanente inferiori al 5%.

Si osservi come i tassi risultano molto diversi a seconda dei segmenti considerati, ottenendo così dei valori che tengono conto congiuntamente degli effetti delle varie variabili considerate e differenziano la popolazione in maniera chiara tra i diversi livelli di rischio di infortunio.

Tabella 3. Tassi per mille per i diversi segmenti di popolazione

| | | | | | | | | | | | | |
|------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Sesso | M | M | M | M | M | M | M | M | M | M | M | M |
| Categoria | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Gruppo | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 |
| Classe età | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Tasso | 0.16702 | 0.78977 | 0.70546 | 0.84151 | 0.28688 | 1.35323 | 1.20917 | 1.44159 | 0.45418 | 2.13510 | 1.90868 | 2.27386 |

| | | | | | | | | | | | | |
|------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Sesso | M | M | M | M | M | M | M | M | M | M | M | M |
| Categoria | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Gruppo | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 |
| Classe età | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Tasso | 0.49332 | 2.31723 | 2.07173 | 2.46767 | 0.84627 | 3.94665 | 3.53193 | 4.20039 | 1.33740 | 6.17532 | 5.53376 | 6.56698 |

| | | | | | | | | | | | | |
|------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Sesso | F | F | F | F | F | F | F | F | F | F | F | F |
| Categoria | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Gruppo | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 |
| Classe età | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Tasso | 0.20990 | 0.99168 | 0.88592 | 1.05657 | 0.36048 | 1.69785 | 1.51741 | 1.80848 | 0.57057 | 2.67586 | 2.39286 | 2.84920 |

| | | | | | | | | | | | | |
|------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Sesso | F | F | F | F | F | F | F | F | F | F | F | F |
| Categoria | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Gruppo | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 |
| Classe età | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Tasso | 0.61970 | 2.90338 | 2.59667 | 3.09121 | 1.06225 | 4.93355 | 4.41771 | 5.24884 | 1.67802 | 7.69509 | 6.90194 | 8.17858 |

Conclusioni e prospettive

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 agevole ottenere delle misure di variabilità delle stime dei tassi ottenuti, in modo da poter verificare l'effettiva diversità tra i diversi tassi in una scala di misura appropriata e utilizzarli per applicarli in una tariffa effettiva
- È possibile inserire altre variabili esplicative nel modello che possano aiutare maggiormente a prevedere il 'rischio'

- È necessario 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.
- È possibile provare ad estendere la procedura adottata anche a tariffe diverse dagli infortuni, come per esempio l'incendio e furto delle auto, dove scoperto e minimo fanno le funzioni della franchigia, e possono quindi essere considerate come variabili di censura.

In questo modo si potrebbe ottenere una tariffa incendio e furto direttamente sul tasso, senza bisogno di comporre i due modelli per frequenza e costo medio come oggi usualmente si fa (Lemaire, 1995).

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.

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