



Dipartimento di Statistica  
"Giuseppe Parenti"

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze – [www.ds.unifi.it](http://www.ds.unifi.it)

W O R K I N G P A P E R 2 0 0 4 / 0 1

Regressione di Poisson:  
un nuovo metodo  
di trasformazione dei residui

Giuseppe Rossi, Laura Marchini  
Marco Marchi



Università degli Studi  
di Firenze

*Biostatistics*

# Regressione di Poisson: un nuovo metodo di trasformazione dei residui

Giuseppe Rossi<sup>1</sup>, Laura Marchini<sup>1</sup>, Marco Marchi<sup>2</sup>

<sup>1</sup> Unità di Epidemiologia e Biostatistica, Sezione di Epidemiologia,  
Istituto di Fisiologia Clinica CNR, Pisa

<sup>2</sup> Dipartimento di Statistica, Università di Firenze

**RIASSUNTO.** I residui sono comunemente utilizzati per verificare, mediante analisi grafica e numerica, l'adeguatezza dei modelli di regressione. Nei modelli lineari generalizzati i residui vengono trasformati in residui normali standard con distribuzione approssimativamente Normale a media nulla e varianza unitaria. Per la regressione di Poisson sono disponibili differenti forme funzionali di residui normali standard. In questo lavoro viene proposta una nuova tipologia di residui normali standard e confrontata con quelle più comunemente utilizzate. Le prestazioni sono illustrate e valutate mediante simulazioni. Infine è mostrata un'applicazione a dati reali nel campo dell'epidemiologia ambientale.

**PAROLE CHIAVE:** Regressione di Poisson; Residui standardizzati; Trasformazioni normalizzanti.

## 1. Introduzione

La regressione di Poisson gioca un ruolo importante nell'analisi statistica di dati sperimentali ed osservazionali, in particolare nel campo dell'epidemiologia ambientale. Dato che un modello malspecificato può compromettere seriamente la validità e l'efficienza delle analisi effettuate, la verifica del modello diviene essenziale.

L'analisi dei residui permette di verificare la validità delle assunzioni alla base dei metodi di indagine (Box e Cox, 1964) e, mediante tecniche grafiche, numeriche e miste (Anscombe e Tuckey, 1963), di valutare l'adeguatezza di un modello, testarne la bontà di adattamento, studiare l'effetto di potenziali nuove covariate o gli effetti non lineari di quelle già presenti nel modello, individuare eventuali valori anomali e valutare l'indipendenza delle osservazioni. Metodi di verifica del modello quando la variabile risposta segue una distribuzione di Poisson sono stati illustrati ampiamente da Breslow (1996).

Per osservazioni che presentano una distribuzione di Poisson la variabile casuale  $\mathbf{Y}$  è assunta avere la forma  $E(\mathbf{Y}) = \exp(\mathbf{X}\boldsymbol{\beta})$ , dove  $\mathbf{X}$  rappresenta la matrice delle covariate e  $\boldsymbol{\beta}$  il vettore dei parametri incogniti. Il residuo  $R_i$  è definito come  $Y_i - \hat{\mu}_i$ , dove  $\hat{\mu}_i = \hat{\mu}_i(\mathbf{b})$  rappresenta la

frequenza attesa, sulla base di un modello dipendente dai parametri stimati  $b_1, \dots, b_p$ , delle osservazioni di Poisson,  $Y_i$ .

Per applicare i metodi standard di analisi dei residui a dati di tipo poissoniano è opportuno definire i residui in modo da ottenere variabili identicamente distribuite, normali, a media nulla e varianza unitaria. Cox e Snell (1968) hanno dato una definizione generale dei residui e derivato alcune proprietà asintotiche nel caso di distribuzioni discrete.

In questo lavoro viene proposta una nuova forma di residui normali standardizzati ottenuta apportando un aggiustamento alla trasformazione avanzata da Rossi, Lampugnani e Marchi (1999) per la normalizzazione di dati poissoniani. Al fine di evidenziarne le proprietà, i residui proposti sono stati confrontati con le migliori tipologie di residui presenti in letteratura ed elencate nella Sezione 2. Per comparare le diverse prestazioni è stato effettuato uno studio mediante simulazioni. I risultati sono illustrati nella Sezione 3.

Infine, nella Sezione 4 è riportata un' applicazione a dati reali relativi ad uno studio di epidemiologia ambientale volto ad indagare l'associazione tra patologia respiratoria ed inquinamento dell'aria.

## 2. Residui e nuovo metodo di trasformazione

Caratteristica principale dei residui è il loro incremento in valore assoluto all'aumentare della discrepanza tra valori osservati e predetti. I residui inoltre devono essere facilmente interpretabili e presentare una distribuzione normale standard sotto un modello correttamente specificato, assumendo l'assenza di valori anomali. Per la regressione di Poisson le più comuni tipologie di residui normali standardizzati presenti in letteratura (Mc Cullagh e Nelder, 1989; Hardin e Hilbe, 2001) sono:

- Residui di Pearson: 
$$r_p = \frac{y - \hat{\mu}}{\sqrt{\hat{\mu}}} \quad (2.1)$$

- Residui in Radice: 
$$r_r = 2(\sqrt{y} - \sqrt{\hat{\mu}}) \quad (2.2)$$

- Residui di Anscombe: 
$$r_A = \left\{ \frac{\left[ 3/2(y^{2/3} - \hat{\mu}^{2/3}) \right]}{\hat{\mu}^{1/6}} \right\} \quad (2.3)$$

- Residui di Devianza:  $r_D = \text{sgn}(y - \hat{\mu}) \{2(y \log(y / \hat{\mu}) - y + \hat{\mu})\}^{1/2}$  (2.4)

Per la trasformazione in radice (2.2) sono stati proposti vari aggiustamenti al fine di migliorarne le prestazioni (Bartlett, 1936; Anscombe, 1948). I risultati più soddisfacenti sono stati ottenuti

utilizzando la trasformazione corretta  $Y = \sqrt{y + \frac{1}{4}}$ , asintoticamente distribuita come una Norm

$(\sqrt{\mu}, 1/4)$  (Brown, Zhang e Zhao, 2001).

Per aumentare la convergenza dei residui di devianza a una distribuzione Norm (0,1), McCullagh e Nelder (1983) hanno proposto di aggiungere alla formulazione iniziale (2.4) il termine  $\frac{\rho_3(\theta)}{6}$ ,

dove  $\rho_3(\theta) = E_\theta \left\{ \left[ \frac{(y - \mu)}{SD_\theta(y)} \right]^3 \right\}$ ,  $\mu = E_\theta(y)$ .

Per una distribuzione di Poisson( $\mu$ ) risulta  $\rho_3(\theta) = (1/\sqrt{\mu})$ .

Pierce e Schafer (1986) hanno ripreso questo aggiustamento utilizzandolo per i residui di Anscombe (1953).

In questo lavoro vengono prese in esame le seguenti tipologie di residui:

1) Residui di Pearson:  $r_P = \frac{y - \hat{\mu}}{\sqrt{\hat{\mu}}}$  (2.1)

2) Residui in Radice aggiustati:  $r_{ra} = 2(\sqrt{y + 1/4} - \sqrt{\hat{\mu}})$  (2.5)

3) Residui di Anscombe aggiustati:  $r_{Aa} = \left\{ \frac{[3/2(y^{2/3} - \hat{\mu}^{2/3})]}{\hat{\mu}^{1/6}} \right\} + \frac{1}{6\sqrt{\hat{\mu}}}$  (2.6)

4) Residui di Devianza aggiustati:  $r_{Da} = \text{sgn}(y - \hat{\mu}) \{2(y \log(y / \hat{\mu}) - y + \hat{\mu})\}^{1/2} + \frac{1}{6\sqrt{\hat{\mu}}}$  (2.7)

5) Residui Proposti: 
$$r_{Ra} = \frac{1}{2\sqrt{\hat{\mu}}} y + \sqrt{y+1/4} - \frac{3\sqrt{\hat{\mu}}}{2} \quad (2.8)$$

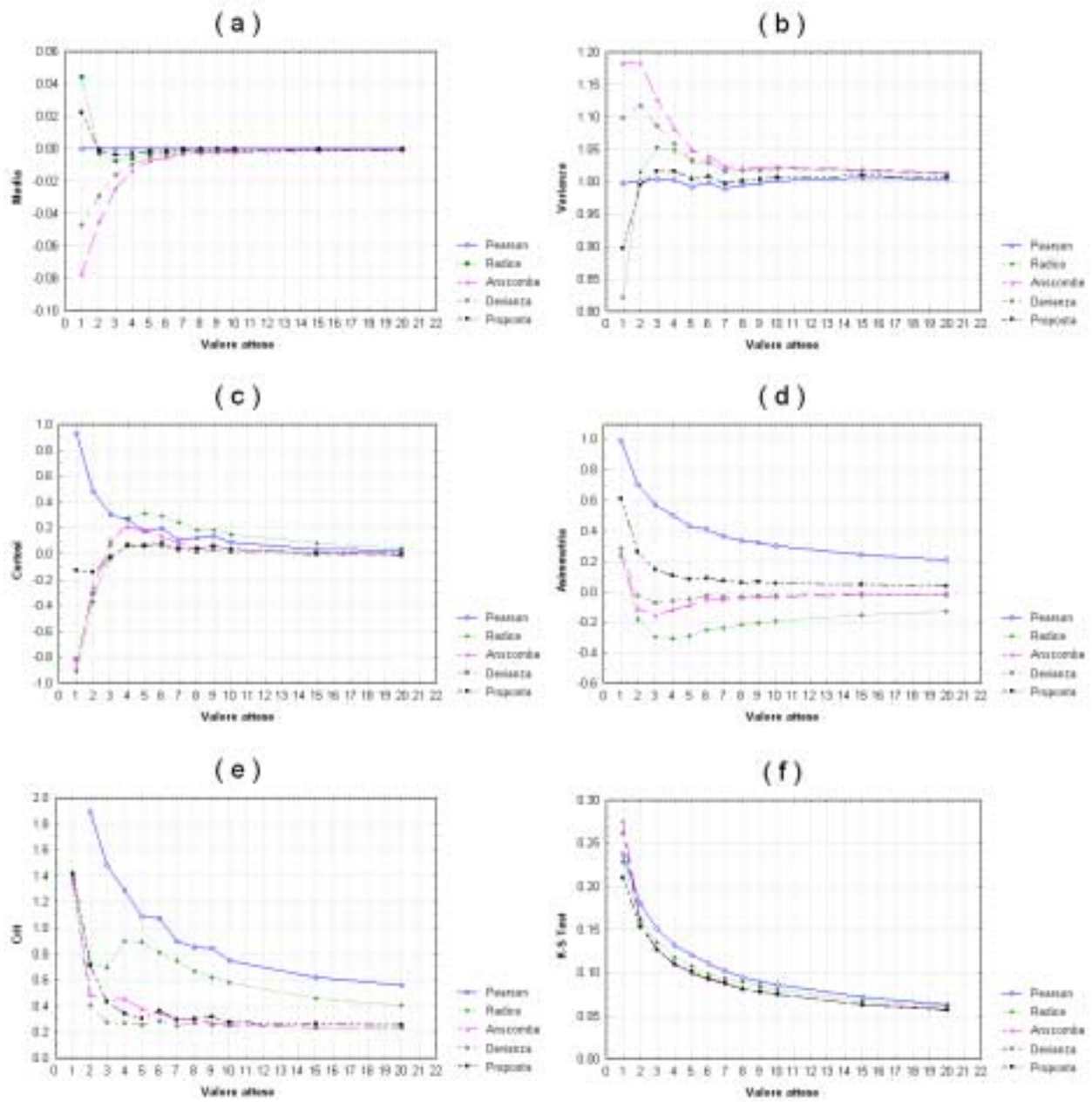
I residui proposti derivano da un aggiustamento apportato alla trasformazione normalizzante avanzata da Rossi, Lampugnani e Marchi (1999) asintoticamente distribuita come una devziata normale standard. I residui proposti sono ottenuti dalla semisomma dei residui di Pearson (2.1) e dei residui in Radice aggiustati (2.5).

### 3. Simulazioni

Allo scopo di verificare la velocità di convergenza alla distribuzione Normale Standard delle varie tipologie di residui in relazione alla media attesa della distribuzione di Poisson, sono stati simulati campioni da una distribuzione di Poisson con diverse medie attese, in particolare pari a 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20. Per ognuna delle medie attese sono stati simulati 100 campioni, ciascuno di numerosità 1000, e ad ogni campione sono state applicate tutte le trasformazioni normalizzanti prese in esame (2.1;2.5-2.8). Per valutare la bontà con cui la distribuzione dei residui ottenuti approssima una Normale Standard sono stati utilizzati indici di posizione, dispersione e forma della distribuzione, la statistica test di Kolmogorov-Smirnov per la normalità e valori di probabilità, calcolati sui dati simulati (S-PLUS 2000 software). Di ciascun indice, a parità di trasformazione e di media attesa, è stata calcolata la media aritmetica dei valori assunti nei 100 campioni.

In Figura 1 sono riportati, per ciascuna trasformazione e media attesa, i valori medi di media, varianza, curtosi, asimmetria, di un criterio (Reed e Wu, 1974) che combina i due indici di forma (  $\text{crit} = 2*|\text{asimmetria}| + |\text{curtosi}|$  ) e della statistica test di Kolmogorov-Smirnov per la normalità. I dati relativi ai grafici sono riportati in Tabella 1.

**Figura 1** Valori medi degli indici per trasformazione e media attesa della distribuzione di Poisson



**Tabella 1** Valori medi degli indici per trasformazione e media attesa della distribuzione di Poisson

Indice	Trasformazione	Media attesa											
		1	2	3	4	5	6	7	8	9	10	15	20
media	<b>Pearson</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	<b>Radice</b>	0.044	-0.002	-0.007	-0.006	-0.003	-0.003	-0.001	-0.001	-0.001	-0.002	-0.001	-0.001
	<b>Anscombe</b>	-0.077	-0.045	-0.024	-0.013	-0.008	-0.005	-0.003	-0.003	-0.002	-0.002	-0.001	-0.001
	<b>Devianza</b>	-0.047	-0.029	-0.016	-0.010	-0.005	-0.004	-0.003	-0.002	-0.002	-0.002	-0.001	-0.001
	<b>Proposta</b>	0.022	-0.001	-0.004	-0.003	-0.002	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	0.000
varianza	<b>Pearson</b>	0.997	1.001	1.004	1.002	0.993	0.999	0.990	0.995	0.998	1.002	1.006	1.004
	<b>Radice</b>	0.822	1.015	1.053	1.049	1.031	1.029	1.016	1.017	1.017	1.020	1.017	1.012
	<b>Anscombe</b>	1.182	1.184	1.127	1.083	1.048	1.040	1.023	1.022	1.022	1.023	1.019	1.014
	<b>Devianza</b>	1.099	1.118	1.086	1.059	1.034	1.030	1.016	1.017	1.017	1.020	1.017	1.012
	<b>Proposta</b>	0.897	0.995	1.016	1.016	1.004	1.008	0.998	1.001	1.004	1.008	1.009	1.006
curtosi	<b>Pearson</b>	0.930	0.483	0.304	0.262	0.178	0.194	0.108	0.120	0.134	0.086	0.036	0.030
	<b>Radice</b>	-0.911	-0.373	0.069	0.280	0.313	0.291	0.246	0.180	0.185	0.148	0.082	0.042
	<b>Anscombe</b>	-0.873	-0.257	0.107	0.212	0.173	0.137	0.083	0.048	0.062	0.033	0.004	-0.006
	<b>Devianza</b>	-0.815	-0.309	-0.032	0.069	0.062	0.059	0.029	0.015	0.036	0.013	-0.008	-0.013
	<b>Proposta</b>	-0.127	-0.139	-0.024	0.062	0.061	0.079	0.044	0.039	0.060	0.031	0.004	-0.002
asimmetria	<b>Pearson</b>	0.989	0.706	0.575	0.502	0.436	0.411	0.363	0.335	0.324	0.300	0.245	0.208
	<b>Radice</b>	0.242	-0.179	-0.295	-0.306	-0.288	-0.249	-0.236	-0.217	-0.201	-0.192	-0.149	-0.130
	<b>Anscombe</b>	0.237	-0.112	-0.151	-0.118	-0.086	-0.049	-0.046	-0.039	-0.030	-0.030	-0.019	-0.018
	<b>Devianza</b>	0.282	-0.027	-0.072	-0.061	-0.051	-0.027	-0.032	-0.031	-0.023	-0.025	-0.016	-0.017
	<b>Proposta</b>	0.610	0.261	0.147	0.108	0.084	0.090	0.070	0.063	0.065	0.057	0.050	0.040
crit	<b>Pearson</b>	2.909	1.894	1.479	1.293	1.090	1.075	0.901	0.853	0.844	0.753	0.623	0.558
	<b>Radice</b>	1.396	0.732	0.697	0.896	0.891	0.812	0.75	0.667	0.623	0.580	0.458	0.402
	<b>Anscombe</b>	1.347	0.485	0.441	0.460	0.382	0.353	0.301	0.296	0.272	0.266	0.240	0.238
	<b>Devianza</b>	1.379	0.406	0.275	0.269	0.258	0.282	0.253	0.271	0.254	0.253	0.235	0.236
	<b>Proposta</b>	1.422	0.707	0.434	0.348	0.304	0.364	0.298	0.302	0.317	0.277	0.266	0.260
K-S test	<b>Pearson</b>	0.237	0.180	0.151	0.132	0.121	0.110	0.103	0.095	0.089	0.086	0.072	0.063
	<b>Radice</b>	0.229	0.162	0.135	0.118	0.108	0.098	0.093	0.089	0.085	0.080	0.067	0.062
	<b>Anscombe</b>	0.276	0.158	0.127	0.109	0.099	0.092	0.087	0.082	0.078	0.075	0.063	0.058
	<b>Devianza</b>	0.261	0.158	0.127	0.109	0.099	0.092	0.087	0.082	0.078	0.075	0.063	0.058
	<b>Proposta</b>	0.210	0.153	0.127	0.111	0.101	0.093	0.088	0.081	0.078	0.075	0.063	0.057

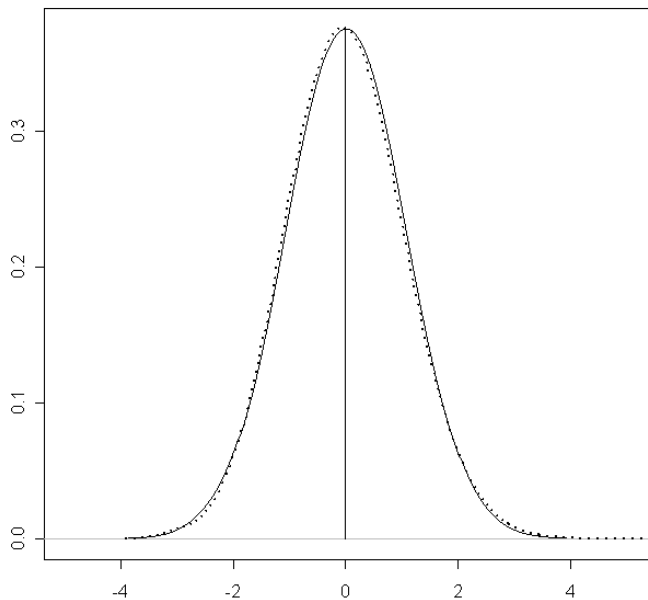
Come si può osservare in Figura 1a-b, la trasformazione proposta presenta, insieme a quella di Pearson, i valori di media e varianza più prossimi ai valori attesi di 0 e 1, già a bassi valori di media attesa della distribuzione di Poisson. La qualità di tutte le procedure grafiche utilizzate nell'analisi dei residui inoltre non può prescindere dalla forma della distribuzione e in particolare

dalla simmetria. Poiché in una distribuzione gaussiana i valori di asimmetria e curtosi standardizzati sono pari a zero, la Figura 1c-d-e evidenzia come i residui in Radice e quelli di Pearson risultino inadeguati nella forma.

La trasformazione proposta presenta, insieme a quella di Anscombe e di Devianza, le prestazioni migliori sia in termini di criterio che di statistica test di K-S.

In Figura 2 vengono messe a confronto le curve di densità relative a due campioni di numerosità 50.000. Il primo (linea continua) è stato simulato partendo da una distribuzione Normale Standard, il secondo (linea tratteggiata) è stato ottenuto applicando la trasformazione proposta a dati simulati partendo da una distribuzione di Poisson a media 5.

**Figura 2** Curve di densità per la trasformazione proposta (linea tratteggiata) e per una Normale Standard (linea continua), basate su 50.000 dati simulati.



Per completezza e per evidenziare meglio le caratteristiche delle varie trasformazioni, sono state calcolate le probabilità nelle due code di una variabile di Poisson con media 7.96, 4.96, 1.86 e confrontate con quelle ottenute tramite le trasformazioni esaminate basate sulla distribuzione Normale (Tabella 2).



**Tabella 2** Probabilità per una distribuzione di Poisson a media  $\mu$  (coda di sinistra e di destra) e approssimazioni Normali basate sulle diverse trasformazioni

$\mu = 7.96$													
y	0	1	2	3	4	5	11	12	13	14	15	16	17
<b>Poisson</b>	0.0003	0.0031	0.0142	0.0435	0.1019	0.1949	0.1802	0.1091	0.0619	0.0331	0.0166	0.0079	0.0036
<b>Pearson</b>	0.0024	0.0068	0.0173	0.0393	0.0802	0.1470	0.0802	0.0393	0.0173	0.0068	0.0024	0.0022	0.0007
<b>Radice</b>	0.0000	0.0003	0.0041	0.0208	0.0643	0.1445	0.1433	0.0873	0.0508	0.0283	0.0151	0.0078	0.0039
<b>Anscombe</b>	0.0000	0.0009	0.0064	0.0247	0.0671	0.1425	0.1406	0.0821	0.0449	0.0231	0.0112	0.0051	0.0022
<b>Devianza</b>	0.0000	0.0011	0.0068	0.0252	0.0675	0.1428	0.1407	0.0823	0.0451	0.0233	0.0113	0.0052	0.0023
<b>Proposta</b>	0.0001	0.0017	0.0087	0.0289	0.0719	0.1458	0.1420	0.0816	0.0435	0.0215	0.0099	0.0042	0.0017

$\mu = 4.96$										
y	0	1	2	3	7	8	9	10	11	12
<b>Poisson</b>	0.0070	0.0418	0.1280	0.2707	0.2320	0.1292	0.0655	0.0304	0.0130	0.0052
<b>Pearson</b>	0.0129	0.0377	0.0919	0.1894	0.1790	0.0862	0.0349	0.0118	0.0034	0.0008
<b>Radice</b>	0.0003	0.0133	0.0729	0.1980	0.1759	0.0985	0.0517	0.0257	0.0121	0.0055
<b>Anscombe</b>	0.0005	0.0171	0.0746	0.1904	0.1743	0.0920	0.0442	0.0194	0.0079	0.0029
<b>Devianza</b>	0.0011	0.0180	0.0753	0.1906	0.1744	0.0922	0.0445	0.0197	0.0080	0.0030
<b>Proposta</b>	0.0023	0.0229	0.0820	0.1937	0.1779	0.0921	0.0426	0.0176	0.0065	0.0022

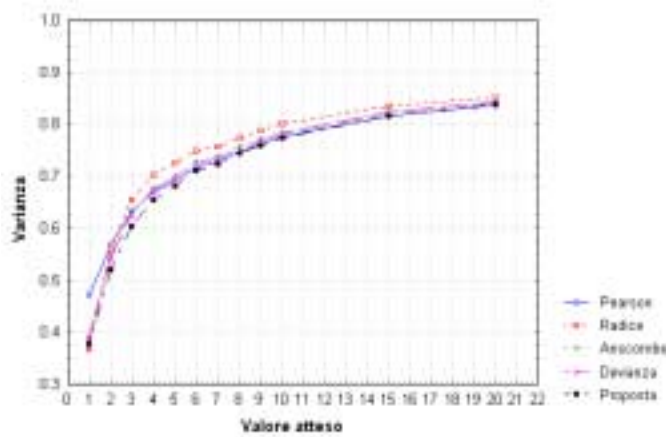
$\mu = 1.86$									
y	0	1	3	4	5	6	7	8	9
<b>Poisson</b>	0.1556	0.4452	0.2855	0.1186	0.0409	0.0120	0.0031	0.0007	0.0002
<b>Pearson</b>	0.0863	0.2641	0.2016	0.0583	0.0106	0.0012	0.0000	0.0000	0.0000
<b>Radice</b>	0.0420	0.3115	0.1900	0.0814	0.0318	0.0115	0.0039	0.0013	0.0004
<b>Anscombe</b>	0.0272	0.2844	0.1867	0.0688	0.0211	0.0055	0.0012	0.0002	0.0000
<b>Devianza</b>	0.0354	0.2844	0.1870	0.0693	0.0216	0.0058	0.0014	0.0003	0.0001
<b>Proposta</b>	0.0611	0.2874	0.1957	0.0691	0.0188	0.0040	0.0007	0.0001	0.0000

Tra le varie tipologie di residui quelli di Pearson risultano i migliori nell'adattamento alla coda di sinistra della distribuzione e i peggiori nell'adattamento alla coda di destra. Situazione diametralmente opposta si osserva per i residui in Radice. I residui Proposti, di Devianza e di Anscombe, molto simili tra loro, presentano un analogo accostamento in entrambe le code della distribuzione, risultando pertanto i più equilibrati.

Le medesime probabilità sono state calcolate anche per i residui corretti per la continuità. I residui in questo caso presentano valori di probabilità che approssimano meglio quelli esatti

basati sulla distribuzione di Poisson ma risultano inadeguati in termini di varianza, con valori distanti dall'unità (Figura 3).

**Figura 3** Valori medi di varianza per media attesa della distribuzione di Poisson e per le trasformazioni corrette per la continuità



#### 4. Applicazione a dati reali

In epidemiologia ambientale è pratica corrente condurre studi di tipo ecologico per analizzare, a livello aggregato di popolazione, gli effetti a breve termine dell'inquinamento atmosferico sulla salute (Katsouyanni et al., 2001; Rossi et al., 1999). Un metodo per verificare l'esistenza di un tale effetto è quello di correlare i livelli giornalieri di inquinamento con quelli degli indicatori di salute mediante l'analisi di serie temporali.

Uno studio effettuato nella città di Torino ha indagato l'associazione tra inquinamento giornaliero dell'aria e numero di ricoveri giornalieri non programmati per patologia respiratoria nella coorte di censimento 1991 durante il periodo 1995-1999.

I dati ambientali, così come i dati meteorologici relativi a temperatura e umidità, sono stati ottenuti dalle stazioni di monitoraggio presenti nella città.

L'associazione tra il numero giornaliero di ricoveri per patologia respiratoria e gli inquinanti dell'aria è stata verificata utilizzando modelli additivi generalizzati (GAM).

Per ovviare ai problemi di implementazione dei GAM, riscontrati recentemente da diversi autori, sono stati utilizzati criteri restrittivi per la stima dei parametri come suggerito da Dominici et al. (2002).

L'analisi è stata effettuata mediante la regressione di Poisson includente funzioni di smoothing (Loess) del tempo, della temperatura media e dell'umidità relativa, variabili indicatrici per i giorni della settimana, per i giorni di festa, per i giorni di epidemia di influenza e un termine lineare per l'inquinante (particolato di diametro inferiore a 10  $\mu\text{m}$ , PM10).

La media dei ricoveri giornalieri per patologia respiratoria è risultata pari a 7.9. La variazione percentuale nel numero medio giornaliero di ammissioni ospedaliere per un incremento di 10  $\mu\text{g}/\text{m}^3$  nella media dei valori dello stesso giorno e del giorno precedente di PM10 è risultata pari a 0.7% (I.C.<sub>0,05</sub> = [0.12, 1.28]).

Sulla base del modello stimato sono stati calcolati i residui di Pearson (2.1), in Radice (2.5), di Devianza (2.7) e Proposti (2.8). Sono stati omessi i residui di Anscombe (2.6) in quanto del tutto simili ai residui di Devianza (2.7). Non sono state osservate autocorrelazioni nei residui.

Le prestazioni delle differenti forme di residui sono messe in evidenza dai Normal Probability Plots (Figura 4) e dalle statistiche descrittive riportate in Tabella 5. Queste ultime mostrano come le prestazioni migliori in termini di media e varianza siano date dai residui di Pearson e dai residui Proposti. I residui di Pearson mostrano però una marcata asimmetria positiva, come evidenziato dal relativo Normal Probability Plot.

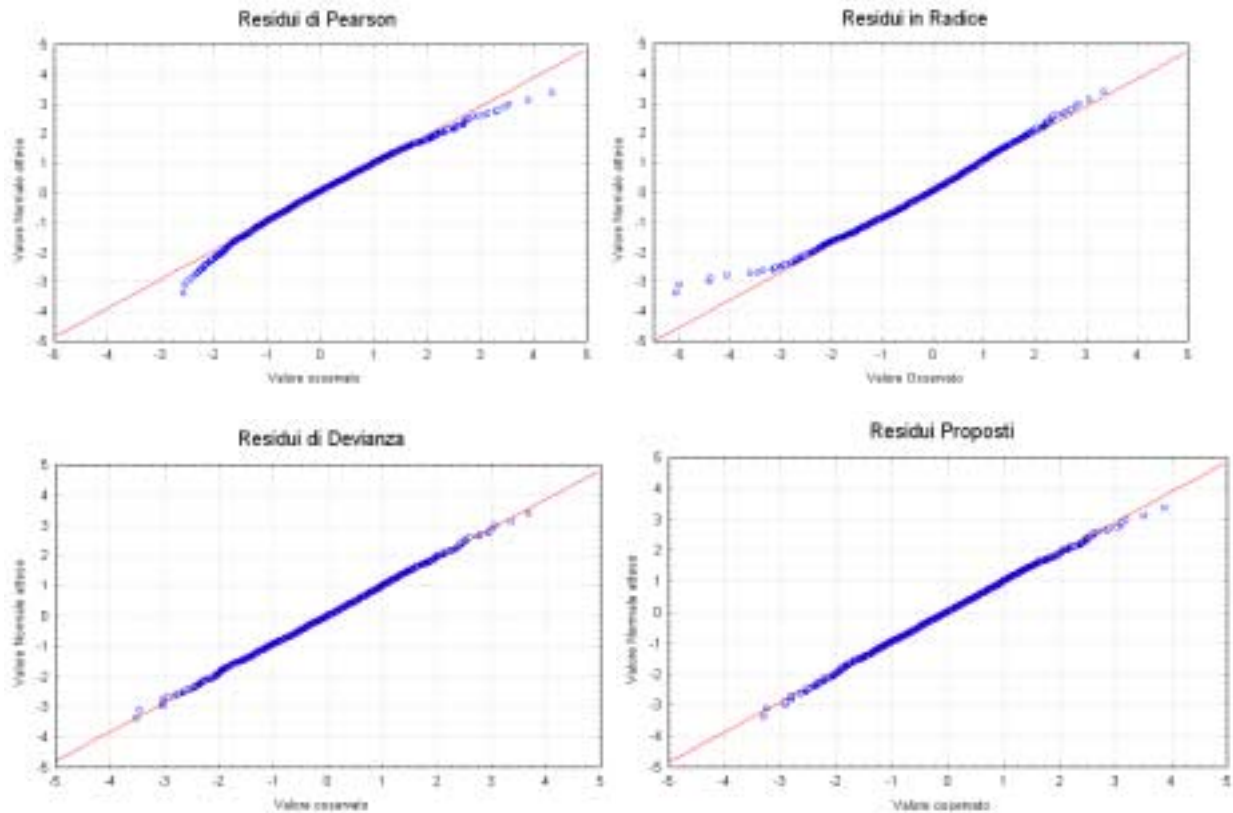
I residui in Radice hanno un andamento soddisfacente nella coda di destra della distribuzione ma inadeguato nella coda di sinistra.

Le prestazioni globalmente migliori, con valori soddisfacenti di media e varianza e con asimmetria e curtosi trascurabili, sono offerte dai residui di Devianza e dai residui Proposti. Questi ultimi, rispetto ai residui di Devianza, presentano un accostamento leggermente peggiore sulla parte finale della coda di destra della distribuzione ma risultano migliori in media e varianza.

**Tabella 5** *Statistiche descrittive dei residui relativi all'applicazione ai dati reali*

<b>Residui</b>	<b>Min.</b>	<b>1°Q.</b>	<b>Media</b>	<b>Mediana</b>	<b>3°Q.</b>	<b>Max.</b>	<b>Var.</b>	<b>Asimm.</b>	<b>Curt.</b>
<b>Pearson</b>	-2.578	-0.717	-0.007	-0.060	0.656	4.300	1.048	0.362	0.179
<b>Radice</b>	-4.066	-0.665	-0.014	0.034	0.703	3.402	1.084	-0.233	0.198
<b>Devianza</b>	-3.520	-0.690	-0.014	-0.002	0.696	3.687	1.081	-0.035	0.034
<b>Proposta</b>	-3.300	-0.690	-0.011	-0.016	0.680	3.875	1.060	0.069	0.059

**Figura 4** Normal Probability Plots dei residui relativi all'applicazione ai dati reali



## 5. Discussione

Nell'ambito della regressione di Poisson la trasformazione Proposta sembra costituire un'alternativa valida per la trasformazione dei residui, competitiva con le migliori trasformazioni presenti in letteratura. I residui Proposti, infatti, in termini di media e varianza danno risultati molto simili a quelli di Pearson ma senza presentare la marcata asimmetria positiva di questi ultimi.

Anche la Radice quadrata con la correzione di  $\frac{1}{4}$  dà buoni risultati in media e varianza, ma i suoi indici di forma risultano peggiori rispetto a quelli dei residui Proposti. In termini di probabilità

infatti la trasformazione in Radice, a causa dell'asimmetria negativa di tale trasformata, fitta bene la coda di destra della distribuzione ma non altrettanto bene la coda di sinistra. I residui con un buon accostamento in entrambe le code della distribuzione, nonché molto simili tra loro, sono risultati i residui Proposti, i residui di Devianza e i residui di Anscombe. Tra questi i residui Proposti sembrano essere complessivamente i più equilibrati in termini di media, varianza e forma della distribuzione. In particolare, per una variabile casuale di Poisson con basso valore atteso i risultati suggeriscono la superiorità dei residui Proposti.

## **Bibliografia**

- Anscombe, F.J. (1948). The transformation of Poisson, binomial and negative-binomial data, *Biometrika* **35**, 246-254.
- Anscombe, F.J. (1953). Contribution to the discussion on Professor Hotelling's Paper, *Journal of the Royal Statistical Society, Series B* **15**, 229-230.
- Anscombe, F.J. e Tuckey, J.W. (1963). The examination and analysis of residuals, *Technometrics* **5**, 141-160.
- Bartlett, M.S. (1936a). The square root transformation in analysis of variance, *Journal of the Royal Statistical Society Suppl.* **3**, 68-78.
- Bartlett, M.S. (1947). The use of transformations, *Biometrics* **3**, 39-52.
- Blom, G. (1954). Transformation of the binomial, negative binomial and Poisson distribution, *Biometrika* **41**, 302-316.
- Box, G.E.P. e Cox, D.R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society* **26**, 211-243.
- Breslow, N.E. (1996) Generalized linear models: checking assumptions and strengthening conclusion, *Statistica Applicata* **8**, 23-41.
- Brown, L.D., Zhang, R. e Zhao, L.H. (2001). *Root-Unroot methods for Nonparametric Density Estimation And Poisson random-effects models*, Department of Statistics University of Pennsylvania.
- ([http://ljsavage.wharton.upenn.edu/~lbrown/Papers\\_and\\_Research\\_Reports/papers\\_research.html](http://ljsavage.wharton.upenn.edu/~lbrown/Papers_and_Research_Reports/papers_research.html))

- Brown, L.D. e Zhao, L.H. (2001). *A new test for the Poisson distribution*, Department of Statistics University of Pennsylvania.
- Cox, D.R. e Snell, E.J. (1968). A general definition of residuals, *Journal of the Royal Statistical Society, Series B* **30** (2), 248-275.
- Dominici, F. et al. (2002). On the use of generalized additive models in time-series studies of air pollution and health, *American Journal of Epidemiology* **156**, 1-11.
- Hardin, J. e Hilbe, J. (2001). "*Generalized linear models and extensions*", Stata Press, College Station (Texas).
- Katsouyanni, K. et al. (2001). Confounding and Effect Modification in the Short-Term Effects of Ambient Particles on Total Mortality: Results from 29 European Cities within the APHEA2 Project, *Epidemiology* **12** (5), 521-529.
- Lin, D.Y., Wei, L.J. e Ying, Z. (2002). Model-Checking Techniques Based on Cumulative Residuals, *Biometrics* **58**, 1-12.
- McCullagh e Nelder, (1989). *Generalized linear models*, London: Chapman and Hall.
- Nardi, A. e Schemper, M. (1999). New Residuals for Cox Regression and Their Application to Outlier Screening, *Biometrics* **55**, 523-529.
- Pierce, D.A. e Schafer, D.W. (1986). Residuals in generalized linear models, *Journal of the American Statistical Association* **31**, 977-986.
- Rossi, G., Lampugnani, L. e Marchi, M. (1999). An approximate Cusum procedure for surveillance of health events, *Statistics in medicine* **18**, 2111-2122.
- Rossi, G. et al. (1999). Air pollution and cause-specific mortality in Milan, Italy, 1980-1989. *Arch Environ Health* **54** (3), 158-164.
- Reed, A.H. e Wu, G.T. (1974). Evaluation of a transformation method for estimation of Normal range, *Clinical Chemistry* **20/5**, 576-581.

Copyright © 2004

Giuseppe Rossi, Laura Marchini

Marco Marchi