

MODELLI STOCASTICI DELLA CLASSE GLM

Si possono considerare GLM con distribuzione specificata o modelli con quasi-verosimiglianza, questi ultimi sono modelli di tipo semiparametrico. Illustriamo il loro uso come:

- **strumento di stima e controllo dell'adattamento**, per ottenere un modello adeguato a descrivere il fenomeno, sulla base dei dati, e per ricavare le stime;
- **strumento di previsione e di valutazione della dispersione**, per ottenere dal modello stimato previsioni e valutazioni di rischiosità (in particolare, ai fini di arrivare al calcolo di margini per il rischio e capitali di rischio).

I GLM sono una **classe flessibile di modelli** con cui si studia l'influenza di un insieme di variabili esplicative sulla distribuzione di una variabile risposta (modelli di regressione).

In ambito assicurativo sono largamente utilizzati nella tariffazione *a priori*, ma anche per trattare altri problemi, incluso quello della valutazione delle riserve.

I GLM sono uno strumento operativo. Sono disponibili diversi *software* statistici (SAS, R, S-Plus, GLIM, EMBLEM,...) che forniscono le stime e le analisi sulla bontà dei modelli.

Possono essere uno **strumento alternativo** per ottenere le stesse valutazioni fornite da diversi metodi tradizionali di stima delle riserve.

Molti metodi stocastici di stima delle riserve sinistri proposti in letteratura, possono essere visti come particolari GLM.

I GLM sono **applicabili a dati individuali**, cioè a informazioni a livello di singolo sinistro consentendo di utilizzare al meglio l'informazione disponibile.

Il ricorso a dati raggruppati (per periodo di origine del sinistro e per periodo di differimento del pagamento) non è necessario.

RICHIAMI SUI MODELLI LINEARI GENERALIZZATI (GLM)

IPOTESI DEI GLM

Modelli di regressione che generalizzano i lineari.

Con riferimento a n **unità statistiche**, si dispone di un insieme di osservazioni

$$\{(y_i, \mathbf{x}_i), i = 1, \dots, n\},$$

– y_i **valore di una grandezza di interesse**,

– \mathbf{x}_i **vettore delle determinazioni** assunte da un insieme di **variabili esplicative**.

Il vettore dei valori

$$\mathbf{y} = (y_1, \dots, y_n)'$$

è visto come valore osservato del vettore aleatorio

$$\mathbf{Y} = (Y_1, \dots, Y_n)'$$

delle **variabile risposta**.

Per il vettore delle variabili risposta $\mathbf{Y} = (Y_1, \dots, Y_n)'$ è formulata una ipotesi probabilistica che mette in relazione la distribuzione di \mathbf{Y} con i vettori delle determinazioni delle variabili esplicative.

In sintesi, un GLM è definito dalle seguenti ipotesi.

- **Ipotesi probabilistiche.** Le variabili risposta Y_1, \dots, Y_n sono stocasticamente indipendenti, con distribuzioni appartenenti ad una medesima famiglia esponenziale lineare.
- **Ipotesi strutturali.** Il legame tra la speranza matematica μ_i della variabile risposta Y_i ed il vettore \mathbf{x}_i delle determinazioni delle variabili esplicative, relativi all' i -esima osservazione, è

$$g(\mu_i) = \mathbf{x}_i' \boldsymbol{\beta},$$

dove $\boldsymbol{\beta}$ è un **vettore di parametri** e g una **funzione di collegamento**, invertibile. Si ha quindi

$$E(Y_i) = \mu_i = g^{-1}(\mathbf{x}_i' \boldsymbol{\beta}).$$

Si ha inoltre

$$\text{var}(Y_i) = \frac{\phi}{\omega_i} V(\mu_i),$$

dove ϕ è un **parametro di dispersione** e V una funzione che caratterizza la famiglia delle distribuzioni delle variabili risposta, detta **funzione di varianza**.

Dai **dati** che sono le determinazioni y_i delle variabili risposta ed i vettori \mathbf{x}_i delle determinazioni delle variabili esplicative, si **stimano** il vettore $\boldsymbol{\beta}$ dei **parametri di regressione ed il parametro di dispersione** ϕ .

Consideriamo più in dettaglio gli elementi del modello.

Distribuzione del vettore delle variabili risposta

- Y_1, \dots, Y_n sono **stocasticamente indipendenti, con distribuzioni appartenenti ad una medesima famiglia esponenziale lineare,**

- Y_i ha distribuzione del tipo $f(y; \theta_i, \phi, \omega_i) = \exp\left\{\frac{\omega_i}{\phi} [y\theta_i - b(\theta_i)]\right\} c(y, \phi, \omega_i)$

dove $\theta_1, \dots, \theta_n, \phi$ sono parametri, $\omega_i > 0$ è un peso assegnato. Inoltre, i supporti delle distribuzioni di Y_1, \dots, Y_n non dipendono dai parametri.

Si noti che

- la **funzione cumulante b** e il **parametro di dispersione ϕ** non dipendono da i ,
- dipendono, in generale da i , il **parametro canonico θ_i** e il **peso ω_i** .

La funzione cumulante b individua la famiglia esponenziale. Le distribuzioni delle variabili Y_i sono tutte della stessa famiglia, per esempio tutte Poisson oppure tutte gamma.

Il parametro canonico è collegato con la speranza matematica della distribuzione:

$$E(Y_i) = \mu_i = b'(\theta_i).$$

La famiglia esponenziale lineare si può individuare anche con **la funzione di varianza** $V(\mu) = b''(b^{-1}(\mu))$.

Famiglie esponenziali di distribuzioni, tra le più usate in ambito attuariale, sono casi particolari di famiglie con **funzioni di varianza di tipo potenza**, ovvero del tipo $V(\mu) = \mu^p$. In particolare, le famiglie normale, Poisson, gamma, gaussiana inversa, corrispondono a $p = 0, 1, 2, 3$, rispettivamente.

Sussiste il seguente **legame tra speranza matematica e varianza**,

$$\text{var}(Y_i) = \frac{\phi}{\omega_i} b''(\theta_i) = \frac{\phi}{\omega_i} V(\mu_i).$$

Pertanto, **la specificazione di una particolare struttura per la speranza matematica**, $\mu_i = g^{-1}(x'_i \beta)$, **implica una struttura anche per la varianza.**

Con riferimento ai **pesi**, a parità di ϕ e $V(\mu_i)$, $\text{var}(Y_i)$ è tanto maggiore quanto minore è ω_i .

I pesi possono allora essere utilizzati per incorporare nel modello informazioni sull'affidabilità delle singole osservazioni.

Le variabili esplicative

Per le n unità statistiche sono disponibili un insieme di caratteristiche **osservabili a priori, giudicate influenti sulle distribuzioni delle variabili risposta**.

Tali caratteristiche possono essere **variabili numeriche** oppure variabili qualitative o numeriche con determinazioni in livelli, dette **variabili di classificazione** o **fattori**.

Una variabile C di classificazione con ℓ livelli o modalità, c_1, \dots, c_ℓ , può essere codificata con le variabili **0-1**,

$$X_i = \begin{cases} 1 & \text{se } C = c_i \\ 0 & \text{altrimenti} \end{cases} \quad i = 1, \dots, \ell.$$

Riesce

$$C = c_i \Leftrightarrow X_i = 1 \quad \text{e} \quad X_j = 0, \text{ per } j \neq i$$

Le variabili X_1, \dots, X_ℓ sono le variabili **indicatrici** dei livelli.

Poiché

$$\sum_{i=1}^{\ell} X_i = 1$$

sono sufficienti $\ell - 1$ variabili indicatrici per descrivere C : la rimanente si ottiene per complemento a uno.

Si ottengono così le **variabili esplicative** del modello.

La matrice di regressione

A codificazione avvenuta, sia m il numero delle variabili esplicative del modello; X_1, \dots, X_m .

E' detta **matrice di regressione** (*design matrix*) la matrice \mathbf{X} in cui per l'osservazione i -esima, si ha un vettore riga con $x_{i0} = 1$, e x_{i1}, \dots, x_{im} determinazioni delle variabili esplicative

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1m} \\ \vdots & \vdots & & \vdots \\ 1 & x_{i1} & \dots & x_{im} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{nm} \end{bmatrix}$$

\mathbf{X} ha n righe e $m+1$ colonne. Indichiamo con p il numero delle colonne. Supponiamo che $n > p$ e che le p colonne siano **linearmente indipendenti**.

Il previsore lineare

Il vettore $\mathbf{x}'_i = (1, x_{i1}, \dots, x_{im})$ delle determinazioni delle variabili esplicative influisce sulla distribuzione di Y_i tramite il **previsore lineare relativo all' i -esima unità statistica**

$$\eta_i = \beta_0 + x_{i1}\beta_1 + \dots + x_{im}\beta_m = \mathbf{x}'_i \boldsymbol{\beta}$$

dove $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_m)'$ è un vettore di parametri, **comuni a tutte le unità statistiche**.

I **parametri di regressione $\boldsymbol{\beta}$** sono considerati **certi**, ma **non noti**.

La componente β_0 è detto **intercetta**.

Il vettore

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta},$$

di componenti η_1, \dots, η_n è detto **previsore lineare** e rappresenta **la componente sistematica** del modello. E' **funzione lineare dei parametri $\beta_0, \beta_1, \dots, \beta_m$** .

In alcuni casi, può essere opportuno non considerare in \mathbf{X} la prima colonna di termini unitari. Allora nell'espressione di η_i non compare l'addendo β_0 .

La funzione di collegamento

E' una funzione $g: D \subset \mathfrak{R} \rightarrow \mathfrak{R}$ strettamente monotona e derivabile che **mette in relazione le componenti del previsore lineare con le speranze matematiche delle variabili risposta**

$$\eta_i = g(\mu_i).$$

Segue che

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\mathbf{x}'_i \boldsymbol{\beta}).$$

Il dominio D deve essere tale che $g^{-1}(\mathbf{x}'_i \boldsymbol{\beta})$ dia valori ammissibili per μ_i .

$$\text{Funzione di collegamento identica, } g(\mu) = \mu, \quad \Rightarrow \quad \mu_i = \mathbf{x}'_i \boldsymbol{\beta} = \sum_{j=0}^m x_{ij}\beta_j$$

modello additivo per le speranze matematiche delle variabili risposta.

$$\text{Funzione di collegamento logaritmica, } g(\mu) = \log(\mu), \quad \Rightarrow \quad \mu_i = e^{\mathbf{x}'_i \boldsymbol{\beta}} = \prod_{j=0}^m e^{x_{ij}\beta_j}$$

modello moltiplicativo per le speranze matematiche delle variabili risposta.

$$\text{Funzione di collegamento potenza, } g(\mu) = \begin{cases} \mu^\gamma & \gamma \neq 0 \\ \log \mu & \gamma = 0. \end{cases}$$

$$\text{Funzione canonica di collegamento, } g(\mu) = b^{-1}(\mu), \quad \Rightarrow \quad \eta_i = g(\mu_i) = \theta_i$$

mette direttamente in collegamento il previsore lineare con il parametro canonico.

STIMA DEI PARAMETRI NEI GLM

I parametri delle distribuzioni delle variabili risposta sono

- i **parametri canonici** $\theta_1, \dots, \theta_n$,
- il **parametro di dispersione** ϕ .

In alcuni casi ϕ è noto, per esempio nella distribuzione di Poisson $\phi = 1$.

Nei GLM, la **stima dei parametri** $\theta_1, \dots, \theta_n$ si ottiene **stimando il vettore dei parametri di regressione** β .

Dato β , rimangono determinati i parametri canonici:

- dal vettore dei parametri β , data la matrice X , $\Rightarrow \eta_i = \mathbf{x}'_i \beta$
- da η_i , data la funzione di collegamento $g(\cdot)$ $\Rightarrow \mu_i = g^{-1}(\eta_i)$
- da μ_i , data la funzione cumulante $b(\cdot)$ $\Rightarrow \theta_i = b^{-1}(\mu_i)$

quindi

$$\theta_i = b^{-1}(g^{-1}(\mathbf{x}'_i \beta)), \quad i = 1, \dots, n.$$

Il parametro vettoriale β è usualmente stimato con il **metodo della massima verosimiglianza (ML)**.

Motivo: proprietà dei corrispondenti stimatori, da cui discendono alcuni risultati sulle distribuzioni delle statistiche che sono utilizzate per l'inferenza.

Per y_1, \dots, y_n **valori osservati** delle variabili risposta Y_1, \dots, Y_n , la **log-verosimiglianza** è

$$\begin{aligned} \underline{\ell}(\boldsymbol{\theta}, \phi; \mathbf{y}) &= \log \underline{L}(\boldsymbol{\theta}, \phi; \mathbf{y}) = \sum_{i=1}^n \left\{ \frac{\omega_i}{\phi} [y_i \theta_i - b(\theta_i)] + \log c(y_i, \phi, \omega_i) \right\} \\ &= \sum_{i=1}^n \underline{\ell}_i(\theta_i, \phi; y_i) \end{aligned}$$

Indicata con $\ell(\beta)$ la log-verosimiglianza vista come funzione di β , le stime di ML si ottengono risolvendo le **equazioni di verosimiglianza** o **equazioni score**

$$\left\{ \frac{\partial \ell(\beta)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial \ell_i(\beta)}{\partial \beta_j} = 0, \quad j = 0, \dots, m, \right.$$

che, in funzione di $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$, sono

$$\sum_{i=1}^n x_{ij} \frac{\omega_i}{\phi} (y_i - \mu_i) \frac{1}{g'(\mu_i)V(\mu_i)} = 0, \quad j = 0, \dots, m,$$

$$\mu_i = g^{-1}(\mathbf{x}'_i \beta).$$

Osservazione importante: la stima di β non dipende da ϕ .

La soluzione dei sistemi avviene per via numerica, con i metodi di **Newton-Raphson** o **scoring di Fisher**, ed è fornita dai *software* statistici.

Distribuzione asintotica

Dalle **proprietà asintotiche** degli stimatori di ML, se n è “sufficientemente grande”, si può supporre che lo stimatore di massima verosimiglianza di β abbia distribuzione normale

$$\tilde{\beta} \approx N(\hat{\beta}, [\mathcal{I}(\hat{\beta})]^{-1}),$$

dove $\mathcal{I}(\beta) = E \left[- \frac{\partial^2 \tilde{\ell}}{\partial \beta_j \partial \beta_h} \right]_{j,h}$, matrice di informazione di Fisher,

o anche

$$\tilde{\beta} \approx N(\hat{\beta}, [-H(\hat{\beta})]^{-1}),$$

dove $H(\beta) = \left[\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_h} \right]_{j,h}$, matrice hessiana.

GLM PER LE RISERVE

Variabili risposta in un GLM per le riserve

Se i dati sono relativi alle **osservazioni riassunte in un triangolo run-off**, è naturale indicare con Y_{ij} la variabile risposta relativa all'osservazione nella cella (i, j) e può essere

- il pagamento incrementale relativo alla cella: pagamenti per sinistri di origine i effettuati con differimento j , P_{ij} ; pagamenti per sinistri di origine i chiusi con differimento j , P_{ij}^F ; pagamenti incrementali espressi in unità monetarie dell'anno di riferimento, P_{ij}^* ,
- il pagamento incrementale relativo ad una cella, rapportato ad una misura di esposizione, P_{ij} / ω_{ij} ,
- il numero di pagamenti N_{ij} ,
- il numero di sinistri denunciati D_{ij} .

Se i **dati** sono “**individuali**”, informazioni a livello di singolo sinistro, la variabile risposta Y_i può essere

- il costo ultimo di un sinistro i ,
- l'indicatore dell'evento “il sinistro i è chiuso senza seguito”.

Variabili esplicative in un GLM per le riserve

Se i dati sono relativi alle **osservazioni riassunte in un triangolo run-off**, si possono considerare come variabili esplicative

- l'anno di origine,
- l'anno di differimento,
- l'anno di pagamento.

L'anno di differimento è una variabile numerica. L'anno di origine e di pagamento sono codificate con valori numerici $0, 1, \dots, t; 0, 1, \dots, 2t$.

Se i **dati** sono "**individuali**", con riferimento al **costo** ultimo **di un sinistro**, si possono considerare come variabili esplicative,

- le **caratteristiche tariffarie** dell'assicurato,
- il **tipo di danno** (cose, persone),
- il differimento nella **denuncia** del sinistro,
- il differimento nella **liquidazione** del sinistro,
- per lesioni a persone, il **tipo di ferite** riportate, le **età dei feriti**, il ricorso o meno alla **ospedalizzazione**, il **luogo dell'incidente**, i tipi di trattamenti...

ESEMPIO. Illustriamo ora alcune elaborazioni numeriche realizzate in **SAS** su dati di un portafoglio di una Compagnia italiana.

I dati comprendono

- il triangolo *run-off* dei **pagamenti incrementali** P_{ij} relativi agli anni di origine (qui anni di accadimento) dal 1990 al 2003,
- per ogni anno di origine, il **numero di sinistri denunciati nell'anno** ω_i .

Consideriamo come variabili esplicative l'**anno di accadimento** e l'**anno di differimento**, entrambe come **variabili di classificazione**. Per la struttura di regressione si ha

Parametri di regressione: μ intercetta, $\alpha_0, \dots, \alpha_t$ per gli anni di origine, β_0, \dots, β_t per gli anni di differimento,

$$\boldsymbol{\beta} = (\mu, \alpha_0, \alpha_1, \dots, \alpha_t, \beta_0, \beta_1, \dots, \beta_t)',$$

Matrice di regressione: X , con $1+(t+1)+(t+1)$ colonne e $n = \frac{(t+1)(t+2)}{2}$ righe. Il rango di X è $p = 1+2t$.

Per evitare problemi di indeterminazione dei parametri, si deve porre un vincolo, per esempio $\alpha_0 = \beta_0 = 0$: si scelgono così l'anno di origine 0 e l'anno di differimento 0 come livelli di riferimento.

Previsori lineari: $\eta_{ij} = \mathbf{x}_{ij}' \boldsymbol{\beta} = \mu + \alpha_i + \beta_j$.

Funzione di collegamento, scegliamo $g = \log$. Segue

$$E(Y_{ij}) = \mu_{ij} = e^{\eta_{ij}} = e^{\mu} e^{\alpha_i} e^{\beta_j} = K a_i b_j,$$

un modello moltiplicativo con un fattore, a_i , che dipende solo dall'anno di accadimento e un fattore, b_j , che dipende solo dall'anno di differimento.

Modello Poisson-logaritmo per le riserve

- **variabili risposta:** $Y_{ij} = P_{ij}$, con **distribuzioni di Poisson**
- **variabili esplicative:** anno di accadimento (*annoacc*), anno di differimento (*annodiff*),
 $\eta_{ij} = \mathbf{x}_{ij}' \boldsymbol{\beta} = \mu + \alpha_i + \beta_j$,
- **funzione di collegamento:** $g = \log$.

Nell'output di SAS si hanno:

- informazioni sul modello
- informazioni sull'accostamento ai dati
- stime dei parametri e analisi delle stime

MODELLI STOCASTICI DELLA CLASSE DEI GLM

17

```

The GENMOD Procedure

Model Information

Data Set          WORK.DATIV
Distribution       Poisson
Link Function     Log
Dependent Variable pagamenti
Observations Used 105

Class Level Information

Class      Levels  V alues
-----
annoacc    14      1  2  3  4  5  6  7  8  9  10 11 12 13 99
annodiff   14      1  2  3  4  5  6  7  8  9  10 11 12 13 99

Criteria For Assessing Goodness Of Fit

Criterion          DF          Value          V alue/DF
-----
Deviance           78    1 4260424.598    182825.9564
Scaled Deviance   78    1 4260424.598    182825.9564
Pearson Chi-Square 78    1 4447110.935    185219.3710
Scaled Pearson X2  78    1 4447110.935    185219.3710
Log Likelihood                    9218270255.4

```

MODELLI STOCASTICI DELLA CLASSE DEI GLM

18

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald	95% Confidence Limits	Chi-Square	Pr > ChiSq
Intercept	1	16.1578	0.0002	16.1574	16.1581	7.225E9	<.0001
annoacc	1	0.1685	0.0002	0.1681	0.1690	473998	<.0001
annoacc	2	0.3099	0.0002	0.3095	0.3104	1688300	<.0001
annoacc	3	0.1911	0.0002	0.1906	0.1916	605889	<.0001
annoacc	4	0.2646	0.0002	0.2641	0.2650	1191696	<.0001
annoacc	5	0.4359	0.0002	0.4354	0.4364	3451576	<.0001
annoacc	6	0.5303	0.0002	0.5299	0.5308	5257320	<.0001
annoacc	7	0.5360	0.0002	0.5355	0.5364	5290452	<.0001
annoacc	8	0.6910	0.0002	0.6906	0.6915	9217844	<.0001
annoacc	9	0.7318	0.0002	0.7314	0.7323	1.034E7	<.0001
annoacc	10	0.8491	0.0002	0.8487	0.8496	1.427E7	<.0001
annoacc	11	0.7718	0.0002	0.7713	0.7722	1.114E7	<.0001
annoacc	12	0.7203	0.0002	0.7198	0.7207	8877816	<.0001
annoacc	13	0.7465	0.0003	0.7460	0.7471	6821783	<.0001
annoacc	99	0.0000	0.0000	0.0000	0.0000	.	.
annodiff	1	-0.0887	0.0001	-0.0889	-0.0886	847441	<.0001
annodiff	2	-1.1171	0.0001	-1.1173	-1.1168	6.342E7	<.0001
annodiff	3	-1.7941	0.0002	-1.7945	-1.7937	8.443E7	<.0001
annodiff	4	-2.3215	0.0003	-2.3220	-2.3209	7.675E7	<.0001
annodiff	5	-2.3994	0.0003	-2.4000	-2.3988	6.61E7	<.0001
annodiff	6	-2.7240	0.0004	-2.7247	-2.7232	5.33E7	<.0001
annodiff	7	-2.2936	0.0003	-2.2942	-2.2929	4.818E7	<.0001
annodiff	8	-3.0068	0.0005	-3.0078	-3.0058	3.435E7	<.0001
annodiff	9	-3.2233	0.0006	-3.2246	-3.2221	2.552E7	<.0001
annodiff	10	-3.3294	0.0008	-3.3309	-3.3279	1.93E7	<.0001
annodiff	11	-3.6627	0.0010	-3.6647	-3.6607	1.255E7	<.0001
annodiff	12	-3.1271	0.0010	-3.1291	-3.1251	9564465	<.0001
annodiff	13	-2.3422	0.0010	-2.3442	-2.3402	5294849	<.0001
annodiff	99	0.0000	0.0000	0.0000	0.0000	.	.
Scale	0	1.0000	0.0000	1.0000	1.0000	.	.

NOTE: The scale parameter was held fixed.

Tendenzialmente,

- i parametri legati agli anni di origine hanno andamento crescente,
- i parametri legati al differimento hanno andamento decrescente.

In questo modello, si ha $E(P_{ij}) = \exp(\mu + \alpha_i) \exp(\beta_j)$. Posto $B = \sum_{j=0}^t e^{\beta_j}$, si può interpretare

- $B \exp(\mu + \alpha_i)$ costo complessivo atteso dei sinistri dell'anno di origine i ,
- $\exp(\beta_j)/B$ aliquota di tale costo che viene pagata con differimento j .

La crescita degli α_i comporta crescita dei costi complessivi al crescere dell'anno di origine.

La decrescenza dei β_j comporta che è decrescente, come è naturale, l'aliquota del costo totale che è pagata nei diversi anni di differimento.

Il GLM produce le stime del **metodo della catena**.

INFERENZA STATISTICA PER I GLM

BONTA' DI ADATTAMENTO AI DATI

Indicazioni di tipo globale.

Si basano sul confronto tra il **modello corrente** c e il **modello saturo** f , che ha un numero di parametri pari al numero di osservazioni e che fornisce come valori stimati μ_i^* i valori osservati y_i .

Supponiamo dapprima ϕ noto.

Devianza scalata o statistica del log-rapporto di verosimiglianza

Si confrontano le **massime verosimiglianze** ottenibili con il modello corrente c e con il modello saturo f . Intuitivamente, il modello c spiega bene i dati se la corrispondente ML non è molto diversa da quella ottenibile con il modello f .

Indicate con $\hat{\beta}$ e β^* le stime di ML del parametro di regressione per i modelli c ed f ,

$$\begin{aligned} S(c, f) &= -2 \log(L_c(\hat{\beta})/L_f(\beta^*)) = -2(\log L_c(\hat{\beta}) - \log L_f(\beta^*)) \\ &= -2(\ell_c(\hat{\beta}) - \ell_f(\beta^*)) \end{aligned}$$

che si può scrivere

$$S(c, f) = S(\hat{\mu}; \mathbf{y}) = -2 \sum_{i=1}^n \omega_i \int_{y_i}^{\hat{\mu}_i} \frac{y_i - s}{V(s)} ds.$$

Devianza

$$\begin{aligned} D(c, f) &= \phi S(c, f) \\ &= D(\hat{\mu}; \mathbf{y}) = \phi S(\hat{\mu}; \mathbf{y}) = -2 \sum_{i=1}^n \omega_i \int_{y_i}^{\hat{\mu}_i} \frac{y_i - s}{V(s)} ds \end{aligned}$$

La devianza scalata è proporzionale a $1/\phi$, **la devianza non dipende da ϕ** .

Statistica chi-quadrato di Pearson

E' un'altra misura dello scostamento tra modello corrente e modello saturo, definita dalla

$$X^2 = \sum_{i=1}^n \omega_i \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}.$$

Anche di questa statistica si considera una **versione scalata**, di valore

$$\frac{X^2}{\phi} = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\frac{\phi}{\omega_i} V(\hat{\mu}_i)}.$$

Nel **modello di regressione lineare normale**, sotto le ipotesi del modello corrente, le precedenti statistiche hanno distribuzione $\chi^2(n-p)$, per ogni n . Si possono allora effettuare *test* formali di verifiche di ipotesi.

Per i GLM diversi dal modello lineare normale non si hanno, in generale, risultati esatti sulle distribuzioni delle due statistiche. Basandosi su proprietà asintotiche si assume che per n elevato, nelle ipotesi del modello corrente, valga l'approssimazione $\chi^2(n-p)$. **L'approssimazione in generale non è buona.**

Una **regola pratica** per avere indicazioni di bontà dell'adattamento realizzato dal modello:

$$\frac{S(c, f)}{n-p} \quad \frac{X^2/\phi}{n-p}$$

se sono molto **maggiori di 1** possono indicare un adattamento non soddisfacente.

Se ϕ non è noto è sostituito con una sua stima ottenuta con uno stimatore consistente.

La stima del parametro ϕ

La stima di ϕ , quando non è noto, si può ottenere, oltre che con il metodo della massima verosimiglianza, anche ricorrendo a stimatori consistenti:

$$\hat{\phi} = \frac{D(c, f)}{n-p} \quad \hat{\phi} = \frac{X^2}{n-p}$$

Esempio. Analisi sull'adattamento del Modello di Poisson

Una prima **indicazione sulla bontà dell'adattamento ai dati** realizzato dal modello si ottiene guardando alla tabella seguente tratta dall'*output* di SAS

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	78	14260424.598	182825.9564
Scaled Deviance	78	14260424.598	182825.9564
Pearson Chi-Square	78	14447110.935	185219.3710
Scaled Pearson X2	78	14447110.935	185219.3710
Log Likelihood		9218270255.4	

Considerando per le statistiche “Scaled Deviance” e “Scaled Pearson X2” la determinazione di Value/DF, si osserva che molto maggiore di 1, può essere una indicazione di un cattivo adattamento del modello. E' quanto accade nel nostro caso: indicazione di **sovradisersione**.

CONTROLLO DEL MODELLO STIMATO

Analisi informali mediante i residui possono

- evidenziare **scostamenti sistematici** tra valori stimati e osservati e quindi essere utilizzate per esplorare l'adeguatezza della funzione di varianza, della funzione di collegamento, delle variabili esplicative presenti nel previsore lineare,
- evidenziare **singole osservazioni** che si discostino dalla maggior parte delle rimanenti.

Residui ordinari (raw residuals)

$$r_i = y_i - \hat{\mu}_i.$$

Residui di Pearson

$$r_{Pi} = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)/\omega_i}}.$$

Il residuo fornisce indicazione sull'impatto dell'osservazione i -esima sul valore della statistica chi-quadrato di Pearson:

$$\sum_{i=1}^n r_{Pi}^2 = X^2$$

Residui della devianza (deviance residuals)

$$r_{Di} = \text{sign}(y_i - \hat{\mu}_i) \sqrt{d_i},$$

dove d_i è l'addendo i -esimo nell'espressione della devianza e

$$\text{sign}(y_i - \hat{\mu}_i) = \begin{cases} 1 & \text{se } y_i - \hat{\mu}_i \geq 0 \\ -1 & \text{se } y_i - \hat{\mu}_i < 0. \end{cases}$$

Il residuo r_{Di} ha lo stesso segno di $y_i - \hat{\mu}_i$ e fornisce indicazione sull'impatto dell'osservazione i -esima nell'analisi della devianza

$$\sum_{i=1}^n r_{Di}^2 = D(c, f).$$

Si considerano anche **versioni standardizzate** dei precedenti due tipi di residui.

Residui di Pearson studentizzati

$$r_{Pi}^{(s)} = \frac{y_i - \hat{\mu}_i}{\sqrt{\frac{\hat{\phi}}{\omega_i} V(\hat{\mu}_i)(1 - h_{ii})}}$$

Residui della devianza studentizzati

$$r_{Di}^{(s)} = \text{sign}(y_i - \hat{\mu}_i) \sqrt{\frac{d_i}{\hat{\phi}(1 - h_{ii})}}$$

dove h_{ii} elemento diagonale della matrice *hat*.

Da proprietà asintotiche degli stimatori di ML (valide nel caso di **dati raggruppati**)

- se il modello è adeguato,
- se i dati sono raggruppati e se nell'ambito di ciascun gruppo c'è un numero “sufficientemente” elevato di osservazioni,

i residui dovrebbero mostrare un **andamento** analogo a quello di **osservazioni di numeri aleatori con distribuzione** approssimativamente **normale di media nulla e varianza costante, unitaria per le versioni standardizzate**.

Analisi grafiche mediante i residui

- Per evidenziare **scostamenti sistematici** tra valori osservati e valori stimati:
 - grafici dei residui al variare dei valori attesi stimati, per es. il grafico delle coppie $(\hat{\mu}_i, r_{Di}^{(s)})$, $i = 1, \dots, n$; per una migliore visualizzazione $(f(\hat{\mu}_i), r_{Di}^{(s)})$, es. $f(\mu) = 2\sqrt{\mu}$ per Poisson, $f(\mu) = 2 \log \mu$ per gamma;
 - grafici dei residui al variare dei valori stimati del previsore lineare, per es. il grafico delle coppie $(\hat{\eta}_i, r_{Di}^{(s)})$, $i = 1, \dots, n$;
 - grafici dei residui al variare delle possibili determinazioni di una variabile esplicativa, per ogni variabile.

Se il modello è adeguato i punti del grafico dovrebbero disporsi in una “**banda**” **orizzontale** attorno all'asse delle ascisse.

Deviazioni sistematiche tipiche:

- una **curvatura nella media** può indicare una scelta non adeguata della funzione di collegamento, che potrebbe non essere stata inserita qualche variabile esplicativa rilevante, che potrebbe essere opportuno trasformare variabili esplicative (es. log o quadratiche),
- **variazioni sistematiche della banda dei valori dei residui** al variare dei valori stimati, può indicare inadeguatezza della funzione di varianza.

I grafici sono poco indicativi se le variabili risposta hanno poche determinazioni.

- Per verificare l'**adeguatezza della funzione di varianza**:

- grafico delle coppie $(\hat{\mu}_i, |r_{Di}^{(s)}|)$, $i = 1, \dots, n$.

Se il modello è adeguato il grafico non dovrebbe evidenziare **andamenti tendenziali**.

La presenza di un *trend* può essere indicativa di una scelta non adeguata della funzione di varianza. Un *trend* crescente può indicare che la funzione di varianza cresce troppo lentamente rispetto alla media, viceversa un *trend* decrescente.

Per modelli con funzione di varianza di tipo potenza $V(\mu) = \mu^p$, si può analizzare la bontà dell'adattamento, al variare di p .

- Per verificare l'adeguatezza della funzione di collegamento,

– grafico delle coppie $(\hat{\eta}_i, \hat{z}_i^*)$, $i = 1, \dots, n$, con

$$\hat{z}_i^* = g(\hat{\mu}_i) + g'(\hat{\mu}_i)(y_i - \hat{\mu}_i).$$

\hat{z}_i^* è il valore in y_i dell'approssimante lineare di g , relativamente a $\hat{\mu}_i$ (*pseudodati*).

Se il modello è adeguato, il grafico dovrebbe mostrare un andamento **approssimativamente lineare**.

Grafici che evidenzino andamenti che si discostino in modo sistematico da un andamento lineare, suggeriscono di modificare la scelta della funzione di collegamento.

Per modelli con funzione collegamento di tipo potenza,

$$g(\mu) = \begin{cases} \mu^\gamma & \text{se } \gamma \neq 0 \\ \log \mu & \text{se } \gamma = 0 \end{cases}$$

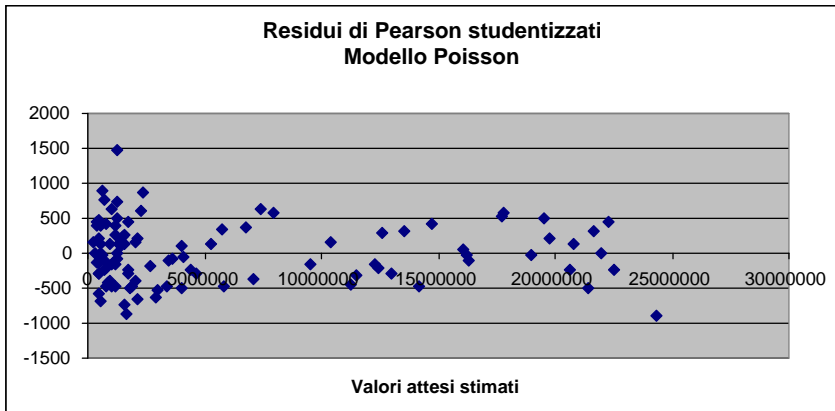
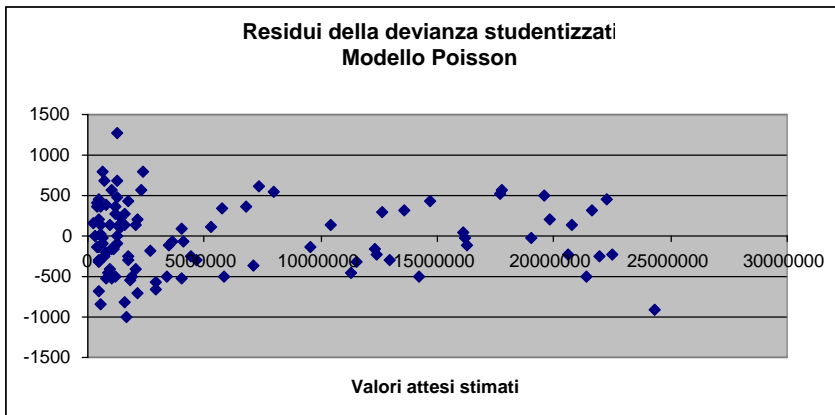
si può analizzare la bontà dell'adattamento, al variare di γ .

- Per evidenziare la presenza di osservazioni “anomale”, i cosiddetti *outliers*,

- i grafici dei residui al variare dell'indice dell'osservazione, es. $(i, r_{Di}^{(s)})$, $i = 1, \dots, n$,
- il grafico delle coppie $(\hat{\mu}_i, y_i)$, $i = 1, \dots, n$.

L'effetto dell'inserimento o meno di alcuni tipi di osservazioni, in particolare di quelle evidenziate come anomale dai grafici precedenti, può essere tratta stimando il modello con l'osservazione e senza l'osservazione e verificando l'effetto sulla stima dell'intercetta e degli altri parametri di regressione.

Esempio. I residui studentizzati della devianza e di Pearson per il *ModelloPoisson* sono riportati nei grafici seguenti



E' evidente che i residui osservati non corrispondono alle determinazioni di un campione casuale di variabili di media 0 e scarto quadratico medio 1. Si hanno indicazioni che il modello non è adatto a descrivere la dispersione dei dati.

Esempio. Modello gamma-logaritmo per le riserve

- **variabili risposta** $Y_{ij} = P_{ij}$, con **distribuzioni gamma**,
- **variabili esplicative**: anno di accadimento (*annoacc*), anno di differimento (*annodiff*),
 $\eta_{ij} = x_{ij}' \boldsymbol{\beta} = \mu + \alpha_i + \beta_j$,
- **funzione di collegamento**: $g = \log$.

The GENMOD Procedure

Model Information

Data Set WORK.DATIV
 Distribution Gamma
 Link Function Log
 Dependent Variable pagamenti
 Observations Used 105

Class Level Information

Class Levels V alues
 annoacc 14 1 2 3 4 5 6 7 8 9 10 11 12 13 99
 annodiff 14 1 2 3 4 5 6 7 8 9 10 11 12 13 99

Criteria For Assessing Goodness Of Fit

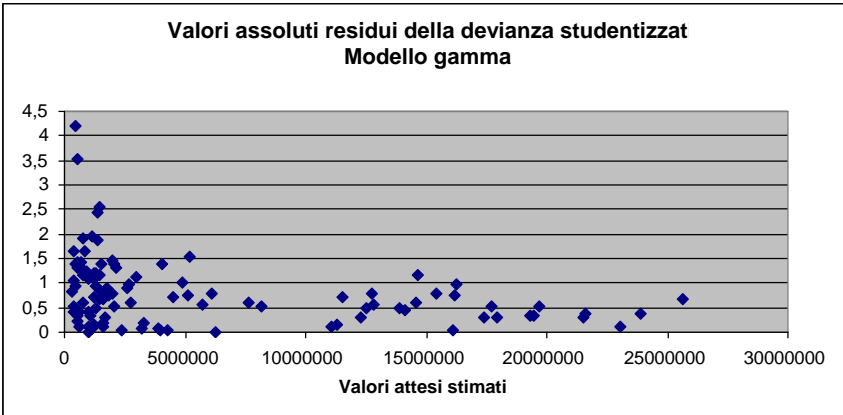
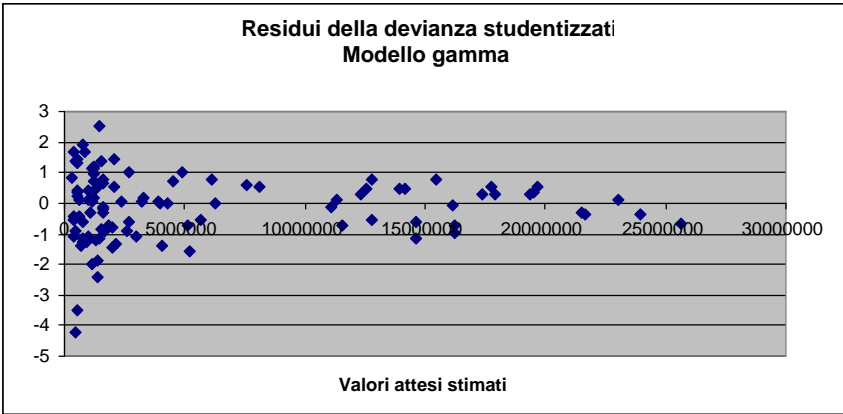
Criterion	DF	Value	Value/DF
Deviance	78	9.9323	0.1273
Scaled Deviance	78	88.4751	1.1343
Pearson Chi-Square	78	8.7563	0.1123
Scaled Pearson X2	78	78.0000	1.0000
Log Likelihood		-1573.3582	

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald	95% Confidence Limits	Chi-Square	Pr > Chi
Intercept	1	16.3646	0.1400	16.0903	16.6390	1 3666.4	<.0001
annoacc	1	0.2396	0.1320	-0.0191	0.4983	3.29	0.0695
annoacc	2	0.2364	0.1377	-0.0334	0.5062	2.95	0.0859
annoacc	3	-0.0378	0.1415	-0.3151	0.2395	0.07	0.7893
annoacc	4	-0.0195	0.1477	-0.3091	0.2700	0.02	0.8947
annoacc	5	0.1012	0.1534	-0.1995	0.4020	0.44	0.5095
annoacc	6	0.1889	0.1590	-0.1227	0.5005	1.41	0.2348
annoacc	7	0.3370	0.1657	0.0122	0.6617	4.14	0.0420
annoacc	8	0.4124	0.1759	0.0677	0.7572	5.50	0.0190
annoacc	9	0.4317	0.1873	0.0646	0.7988	5.31	0.0212
annoacc	10	0.6940	0.2032	0.2959	1.0922	11.67	0.0006
annoacc	11	0.6255	0.2270	0.1806	1.0704	7.59	0.0059
annoacc	12	0.5253	0.2675	0.0009	1.0496	3.85	0.0496
annoacc	13	0.5396	0.3631	-0.1721	1.2513	2.21	0.1372
annoacc	99	0.0000	0.0000	0.0000	0.0000	.	.
annodiff	1	-0.1047	0.1315	-0.3625	0.1530	0.63	0.4258
annodiff	2	-1.1463	0.1355	-1.4119	-0.8808	71.57	<.0001
annodiff	3	-1.7918	0.1395	-2.0653	-1.5183	164.87	<.0001
annodiff	4	-2.2928	0.1438	-2.5747	-2.0109	254.07	<.0001
annodiff	5	-2.4195	0.1498	-2.7131	-2.1259	260.93	<.0001
annodiff	6	-2.7622	0.1560	-3.0679	-2.4564	313.52	<.0001
annodiff	7	-2.3163	0.1632	-2.6362	-1.9965	201.46	<.0001
annodiff	8	-3.0787	0.1727	-3.4171	-2.7403	317.93	<.0001
annodiff	9	-3.3611	0.1859	-3.7255	-2.9966	326.74	<.0001
annodiff	10	-3.4670	0.2027	-3.8644	-3.0697	292.44	<.0001
annodiff	11	-3.8490	0.2256	-4.2911	-3.4069	291.13	<.0001
annodiff	12	-3.3855	0.2673	-3.9093	-2.8616	160.41	<.0001
annodiff	13	-2.5491	0.3631	-3.2608	-1.8374	49.28	<.0001
annodiff	99	0.0000	0.0000	0.0000	0.0000	.	.
Scale	0	8.9078	0.0000	8.9078	8.9078		

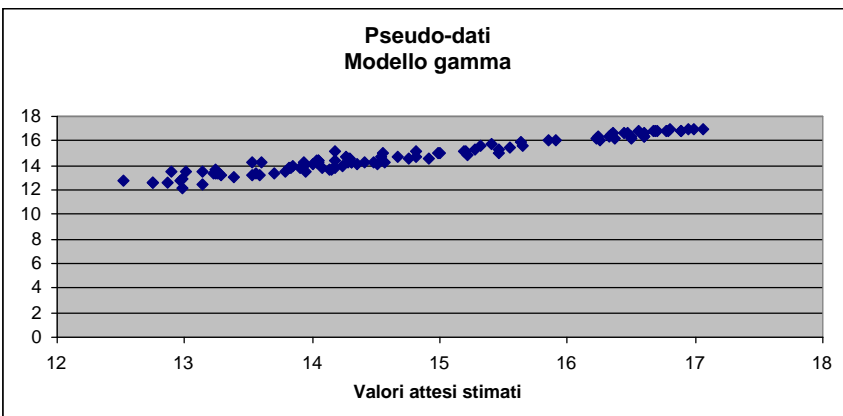
NOTE: The Gamma scale parameter was estimated by DOF/Pearson's Chi-Square

Sull'andamento dei parametri valgono considerazioni analoghe a quelle fatte per il modello Poisson-logaritmo.



Si ha una indicazione che la funzione di varianza cresce troppo rapidamente rispetto alla media.

Grafico degli “pseudo-dati” per indicazioni sulla funzione di collegamento



Stime delle riserve

Anno i	Riserva stimata Poisson	Riserva stimata Gamma	in %
1	1.183.644	1.271.726	-0,074
2	1.985.355	1.818.789	0,084
3	2.086.149	1.649.241	0,209
4	2.730.687	2.081.508	0,238
5	3.881.736	2.858.850	0,264
6	5.140.570	3.841.105	0,253
7	6.963.868	6.227.337	0,106
8	9.494.382	7.993.236	0,158
9	11.853.086	9.941.329	0,161
10	15.714.940	15.602.114	0,007
11	18.288.287	18.601.479	-0,017
12	24.367.054	23.508.407	0,035
13	45.101.492	43.279.371	0,040
Totale	148.791.251	138.674.494	0,068

VERIFICHE DI IPOTESI E CONFRONTO TRA MODELLI

Problema. Consideriamo il confronto tra due modelli, M_0 e M_1 , con

- distribuzioni delle variabili risposta appartenenti alla **medesima famiglia esponenziale lineare**,
- la **medesima funzione di collegamento**,
- **diversa struttura di regressione**: il modello M_1 con p parametri; il modello vincolato M_0 con $p-s$ parametri, un sottovettore del vettore dei parametri del modello M_1 .

Rispetto a M_1 , M_0 è detto **modello ridotto** o **annidato** o **sottomodello**.

Obiettivo del confronto: **valutare la significatività** delle variabili che compaiono nel modello M_1 e non nel modello M_0 , tenuto conto delle variabili che compaiono nel modello M_0 .

Si può effettuare una **verifica di ipotesi**: H_0 contro H_1 : non H_0 .

Esempi.

(1) Per valutare la significatività di un'unica variabile numerica, la condizione espressa dall'ipotesi nulla riguarda **un unico parametro**

$$H_0 : \beta_k = 0.$$

(2) Per valutare **globalmente la significatività di alcune variabili** corrispondenti ai parametri $\beta_{j_1}, \dots, \beta_{j_s}$, si formula l'ipotesi nulla

$$H_0 : \beta_{j_1} = 0, \dots, \beta_{j_s} = 0.$$

Rientra in quest'ultimo esempio il confronto tra modello corrente e modello saturo, in cui M_0 è il modello corrente e M_1 il modello saturo.

Problema. Ridiscutere la suddivisione in livelli delle variabili esplicative di classificazione.

Esempio.

(3) Consideriamo una variabile di classificazione C_k , codificata mediante un vettore di variabili indicatrici. Per valutare se **mantenere separati i due livelli** j_h, j_g , si formula l'ipotesi nulla

$$H_0 : \beta_{j_h}^{(k)} = \beta_{j_g}^{(k)}.$$

Gli esempi sono casi particolari di verifiche di ipotesi in cui l'ipotesi nulla esprime un **vincolo lineare** per il vettore dei parametri di regressione.

Più in generale, possiamo considerare verifiche di ipotesi per β del tipo

$$H_0 : L\beta = \xi \quad \text{contro} \quad H_1 : L\beta \neq \xi,$$

dove L è una matrice $s \times p$, con $s \leq p$, di rango pieno s .

Negli esempi, ξ è il **vettore nullo** e la matrice L

(1) $H_0 : \beta_k = 0$

$$L = (0, \dots, 0, \beta_k, 0, \dots, 0),$$

(2) $H_0 : \beta_{j_1} = 0, \dots, \beta_{j_s} = 0$

$$L = \begin{bmatrix} & \beta_{j_1} & & \beta_{j_2} & & \beta_{j_s} & & & \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{bmatrix}$$

(3) $H_0 : \beta_{j_h}^{(k)} = \beta_{j_g}^{(k)}$

$$L = (0, \dots, 0, \beta_{j_h}^{(k)}, 0, \dots, 0, \beta_{j_g}^{(k)}, 0, \dots, 0),$$

Per i *test* si considerano le seguenti statistiche.

La **statistica del log-rapporto di verosimiglianza**, di valore

$$\lambda = -2(\ell(\hat{\boldsymbol{\beta}}^{(0)}) - \ell(\hat{\boldsymbol{\beta}})),$$

dove

ℓ è la log-verosimiglianza del modello M_1 ,

$\hat{\boldsymbol{\beta}}$ è il punto di massimo di ℓ ,

$\hat{\boldsymbol{\beta}}^{(0)}$ il punto di massimo di ℓ , vincolato dalla condizione $L\boldsymbol{\beta} = \boldsymbol{\xi}$.

Intuitivamente, se la massima log-verosimiglianza non vincolata $\ell(\hat{\boldsymbol{\beta}})$ è significativamente maggiore di quella vincolata $\ell(\hat{\boldsymbol{\beta}}^{(0)})$, e quindi λ è **“grande”**, si rifiuta l'ipotesi H_0 , se λ è **“piccolo”**, si accetta l'ipotesi H_0 .

Dai risultati asintotici sulle stime di massima verosimiglianza, si ha

$$\text{sub } H_0 \quad \tilde{\lambda} \overset{a}{\sim} \chi^2(s).$$

con s rango della matrice L . Nel caso dei modelli annidati, $s = df_0 - df_1$, differenza tra i numeri di gradi di libertà del sottomodello M_0 e del modello M_1 .

La **statistica di Wald**, di valore

$$w = (L\hat{\boldsymbol{\beta}} - \boldsymbol{\xi})' (L\mathcal{J}(\hat{\boldsymbol{\beta}})^{-1}L')^{-1} (L\hat{\boldsymbol{\beta}} - \boldsymbol{\xi}),$$

dove $\mathcal{J}(\hat{\boldsymbol{\beta}})$ è la matrice d'informazione di Fisher calcolata in $\hat{\boldsymbol{\beta}}$.

Il valore w è una misura della distanza tra

$L\hat{\boldsymbol{\beta}}$, stima non vincolata del vettore $L\boldsymbol{\beta}$,

$\boldsymbol{\xi}$ valore vincolato di $L\boldsymbol{\beta}$.

Intuitivamente, se la distanza tra $L\hat{\boldsymbol{\beta}}$ e $\boldsymbol{\xi}$ è elevata, e quindi se w è **“grande”**, si rifiuta l'ipotesi H_0 , se w è **“piccolo”**, si accetta l'ipotesi H_0 .

Dai risultati asintotici sulle stime di massima verosimiglianza, riesce

$$\text{sub } H_0 \quad \tilde{w} \overset{a}{\sim} \chi^2(s).$$

con s rango della matrice L .

Esempio (1) Nel test con $H_0 : \beta_k = 0$,

$$[L_{\mathcal{J}}(\hat{\beta})]^{-1} L' = v_{kk}$$

è l'elemento di posto (k,k) dell'inversa della matrice d'informazione di Fisher calcolata in $\hat{\beta}$.

v_{kk} è una stima della varianza asintotica di $\tilde{\beta}_k$.

Per w si ha

$$w = \left(\hat{\beta}_k / \sqrt{v_{kk}} \right)^2.$$

Dai risultati asintotici sulle stime di massima verosimiglianza, riesce

$$\text{sub } H_0 \quad \tilde{w} \stackrel{a}{\sim} \chi^2(1).$$

Esempio. Riprendiamo l'output per il modello con distribuzioni gamma.

Parameter	DF	Estimate	Analysis Of Parameter Estimates			Chi-Square	Pr > Chi
			Standard Error	Wald	95% Confidence Limits		
Intercept	1	16.3646	0.1400	16.0903	16.6390	1 3666.4	<.0001
annoacc	1	0.2396	0.1320	-0.0191	0.4983	3.29	0.0695
annoacc	2	0.2364	0.1377	-0.0334	0.5062	2.95	0.0859
annoacc	3	-0.0378	0.1415	-0.3151	0.2395	0.07	0.7893
annoacc	4	-0.0195	0.1477	-0.3091	0.2700	0.02	0.8947
annoacc	5	0.1012	0.1534	-0.1995	0.4020	0.44	0.5095
annoacc	6	0.1889	0.1590	-0.1227	0.5005	1.41	0.2348
annoacc	7	0.3370	0.1657	0.0122	0.6617	4.14	0.0420
annoacc	8	0.4124	0.1759	0.0677	0.7572	5.50	0.0190
annoacc	9	0.4317	0.1873	0.0646	0.7988	5.31	0.0212
annoacc	10	0.6940	0.2032	0.2959	1.0922	11.67	0.0006
annoacc	11	0.6255	0.2270	0.1806	1.0704	7.59	0.0059
annoacc	12	0.5253	0.2675	0.0009	1.0496	3.85	0.0496
annoacc	13	0.5396	0.3631	-0.1721	1.2513	2.21	0.1372
annoacc	99	0.0000	0.0000	0.0000	0.0000	.	.
annodiff	1	-0.1047	0.1315	-0.3625	0.1530	0.63	0.4258
annodiff	2	-1.1463	0.1355	-1.4119	-0.8808	71.57	<.0001
annodiff	3	-1.7918	0.1395	-2.0653	-1.5183	164.87	<.0001
annodiff	4	-2.2928	0.1438	-2.5747	-2.0109	254.07	<.0001
annodiff	5	-2.4195	0.1498	-2.7131	-2.1259	260.93	<.0001
annodiff	6	-2.7622	0.1560	-3.0679	-2.4564	313.52	<.0001
annodiff	7	-2.3163	0.1632	-2.6362	-1.9965	201.46	<.0001
annodiff	8	-3.0787	0.1727	-3.4171	-2.7403	317.93	<.0001
annodiff	9	-3.3611	0.1859	-3.7255	-2.9966	326.74	<.0001
annodiff	10	-3.4670	0.2027	-3.8644	-3.0697	292.44	<.0001
annodiff	11	-3.8490	0.2256	-4.2911	-3.4069	291.13	<.0001
annodiff	12	-3.3855	0.2673	-3.9093	-2.8616	160.41	<.0001
annodiff	13	-2.5491	0.3631	-3.2608	-1.8374	49.28	<.0001
annodiff	99	0.0000	0.0000	0.0000	0.0000	.	.
Scale	0	8.9078	0.0000	8.9078	8.9078	.	.

NOTE: The Gamma scale parameter was estimated by DOF/Pearson's Chi-Square

Molti parametri non sono significativamente diversi da zero. E' effettuata la verifica di ipotesi

$$H_0 : \beta_j = 0 \quad \text{contro} \quad H_1 : \beta_j \neq 0,$$

con la **statistica di Wald**, per verificare la significatività del parametro β_j

- un **valore basso** della statistica (*Chi-Square*) rispetto ad un valore critico della distribuzione $\chi^2(1)$,
- un **valore elevato** del *p-value* ($Pr > ChiSq$),

porterebbero a giudicare **statisticamente non significativo** mantenere $\beta_j \neq 0$.

ULTERIORI ANALISI PER IL CONTROLLO DI UN MODELLO

Accenniamo ad altri elementi che possono essere considerati per effettuare analisi su

- singoli parametri,
- previsore lineare,
- il modello nel suo complesso.

Questi elementi si possono ottenere come *output* dei *software* statistici.

Analisi su singoli parametri

Per ogni parametro stimato $\hat{\beta}_k$, indicazioni sulla sua significatività si possono ricavare da:

Standard error:
$$\hat{\sigma}_k = \sqrt{v_{kk}}$$

dove $v_{kk} = \text{var}(\tilde{\beta}_k)$ è l'elemento di posto (k,k) di $\mathcal{J}(\hat{\beta})^{-1}$.

Un valore elevato di $\hat{\sigma}_k$ può indicare non affidabilità della stima $\hat{\beta}_k$.

Intervallo di confidenza per β_k con livello $1-\alpha$:
$$\hat{\beta}_k \pm z_{1-\alpha/2} \hat{\sigma}_k$$

dove $z_{1-\alpha/2}$ è il quantile di ordine $1-\alpha/2$ della $N(0,1)$.

Un intervallo ampio può indicare non attendibilità della stima. Un intervallo che contenga lo zero può indicare non significatività della variabile corrispondente a β_k .

Analisi sul previsore lineare

Per effettuare analisi sul valore stimato del previsore lineare $\hat{\eta}_i = \mathbf{x}'_i \hat{\boldsymbol{\beta}}$ relativo all'osservazione i -esima:

Standard error:

$$\hat{\sigma}_{\eta_i} = \sqrt{\mathbf{x}'_i \mathcal{J}(\hat{\boldsymbol{\beta}})^{-1} \mathbf{x}_i}.$$

Stima dello scarto quadratico medio dello stimatore di η_i .

Intervallo di confidenza per η_i con livello $1-\alpha$:

$$\hat{\eta}_i \pm z_{1-\alpha/2} \hat{\sigma}_{\eta_i},$$

dove $z_{1-\alpha/2}$ è il quantile di ordine $1-\alpha/2$ della $N(0,1)$.

Intervallo di confidenza per μ_i :

$$g^{-1}(\hat{\eta}_i \pm z_{1-\alpha/2} \hat{\sigma}_{\eta_i}),$$

per la monotonia della funzione di collegamento, se gli estremi dell'intervallo, $\hat{\eta}_i \pm z_{1-\alpha/2} \hat{\sigma}_{\eta_i}$, appartengono al dominio di g^{-1} .

MODELLI CON QUASI-VEROSIMIGLIANZA

La classe dei GLM e le relative tecniche per l'inferenza statistica sono state estese in diversi modi per aumentare ulteriormente la flessibilità e l'applicabilità dei modelli.

I modelli con quasi-verosimiglianza (QL), sono **modelli semiparametrici**, nei quali si **specificano solamente le strutture dei primi due momenti delle distribuzioni delle variabili risposta** e non anche una particolare forma di distribuzione.

In questo caso **si ottengono stime del valore atteso e della varianza** delle variabili risposta, ma **si perde**, in generale, la possibilità di ottenere una **distribuzione stimata**.

Ipotesi di un modello QL o GLM semiparametrico o quasi-GLM

- **Ipotesi probabilistiche.** Le variabili risposta Y_1, \dots, Y_n sono **stocasticamente indipendenti** e

$$\text{var}(Y_i) = \frac{\phi}{\omega_i} V(\mu_i), \quad i = 1, \dots, n,$$

dove

- $E(Y_i) = \mu_i$,
- V è una funzione della speranza matematica, detta **funzione di varianza**,
- $\phi > 0$ è un **parametro di dispersione**,
- $\omega_i > 0$ è un **peso** assegnato.

- **Ipotesi strutturali.** Il legame tra la speranza matematica μ_i della variabile risposta Y_i ed il vettore \mathbf{x}_i delle determinazioni delle variabili esplicative, è

$$\mu_i = g^{-1}(\mathbf{x}'_i \boldsymbol{\beta}), \quad i = 1, \dots, n,$$

dove

- $\boldsymbol{\beta}$ vettore di **parametri**,
- g **funzione di collegamento**.

La struttura dei primi due momenti è analoga a quella dei GLM, ma ci si svincola dall'ipotesi che le distribuzioni siano specificate ed appartengano ad una famiglia esponenziale lineare.

Stima del parametro $\boldsymbol{\beta}$

Si risolve il **sistema di Wedderburn** o di **quasi-verosimiglianza**,

$$\sum_{i=1}^n x_{ij} \frac{\omega_i}{\phi} (y_i - \mu_i) \frac{1}{g'(\mu_i) V(\mu_i)} = 0, \quad j = 0, \dots, m,$$

formalmente uguale al sistema delle equazioni di verosimiglianza in un GLM.

E' il sistema delle condizioni del primo ordine per la funzione

$$Q(\boldsymbol{\mu}, \phi; \mathbf{y}) = \sum_{i=1}^n \omega_i \int_{y_i}^{\mu_i} \frac{y_i - s}{\phi V(s)} ds, \quad \text{con } \mu_i = g^{-1}(\mathbf{x}'_i \boldsymbol{\beta}),$$

detta **quasi-(log)-verosimiglianza**.

E' un sistema di equazioni di verosimiglianza solo se esiste una distribuzione della famiglia esponenziale lineare con funzione di varianza V , e si decide di scegliere tali distribuzioni per le variabili risposta.

Le stime che si ottengono risolvendo il sistema sono dette **stime di massima quasi-verosimiglianza** e si prova che soddisfano proprietà, come la consistenza e la normalità asintotica, analoghe a quelle delle stime di massima verosimiglianza.

Adattamento ai dati

Nell'approccio della quasi-verosimiglianza, indicazioni sulla bontà dell'adattamento ai dati si ricavano dalla statistica chi-quadrato di Pearson e dalla **quasi-devianza** la cui espressione è la seguente

$$d(\hat{\boldsymbol{\mu}}, \mathbf{y}) = -2 \sum_{i=1}^n \omega_i \int_{y_i}^{\hat{\mu}_i} \frac{y_i - s}{V(s)} ds.$$

Per il suo calcolo basta conoscere la speranza matematica e la funzione di varianza.

Stima del parametro ϕ

Si usa lo stimatore di Pearson

$$\hat{\phi} = \frac{X^2}{n-p} = \frac{1}{n-p} \sum_{i=1}^n \omega_i \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)},$$

oppure lo stimatore che si ottiene sostituendo alla devianza la quasi-devianza

$$\hat{\phi} = \frac{d(\hat{\boldsymbol{\mu}}, \mathbf{y})}{n-p}.$$

Verifiche di ipotesi

Si può usare la statistica di Wald

$$w = (\mathbf{L}\hat{\boldsymbol{\beta}} - \boldsymbol{\xi})' (\mathbf{L}\mathcal{I}(\hat{\boldsymbol{\beta}})\mathbf{L}')^{-1} (\mathbf{L}\hat{\boldsymbol{\beta}} - \boldsymbol{\xi}).$$

Esempio. Modello di Poisson con sovradisersione-logaritmo

E' un modello GLM di tipo semiparametrico.

- **variabili risposta** $Y_{ij} = P_{ij}$, **Poisson con sovradisersione**,

Si richiede l'indipendenza delle variabili risposta e si pongono le condizioni

$$E(Y_{ij}) = \mu_{ij} = \exp(\mu + \alpha_i + \beta_j),$$

$$\text{var}(Y_{ij}) = \phi V(\mu_{ij}) = \phi \mu_{ij}, \phi > 1.$$

- **variabili esplicative**: anno di accadimento (*annoacc*), anno di differimento (*annodiff*),

$$\eta_{ij} = \mathbf{x}_{ij}' \boldsymbol{\beta} = \mu + \alpha_i + \beta_j,$$

- **funzione di collegamento**: $g = \log$.

Il modello fornisce le stesse stime dei parametri di regressione del modello Poisson (e quindi le stesse stime delle riserve del metodo della catena).

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	78	14260424.598	182825.9564
Scaled Deviance	78	76.9921	0.9871
Pearson Chi-Square	78	14447110.935	185219.3710
Scaled Pearson X2	78	78.0000	1.0000
Log Likelihood		49769.4718	

L'indicatore sintetico di accettabilità del modello, Value/DF, fornisce ora valori vicino a 1.

Le stime dei parametri sono quelle del Modello di Poisson, cambiano le altre colonne.

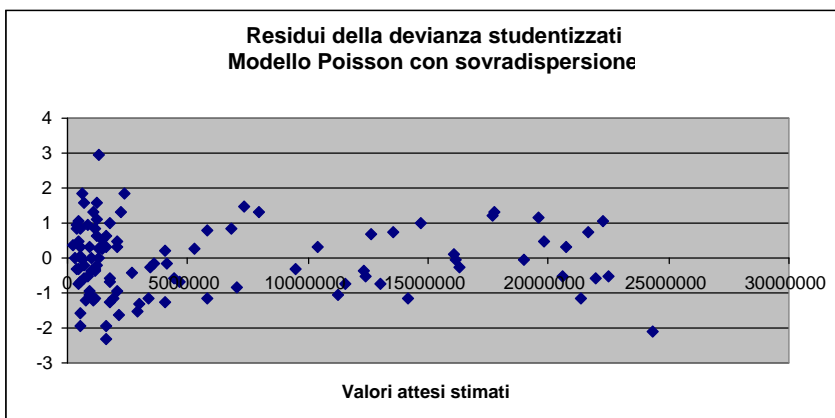
Analysis Of Parameter Estimates

Parameter	D F	Estimate	Standard Error	Wald	95% Confidence Limits	Chi-Square	Pr > ChiSq
Intercept	1	16.1578	0.0818	15.9974	16.3181	39008.2	<.0001
annoacc	1	0.1685	0.1054	-0.0380	0.3750	2.56	0.1097
annoacc	2	0.3099	0.1027	0.1087	0.5111	9.12	0.0025
annoacc	3	0.1911	0.1057	-0.0160	0.3982	3.27	0.0705
annoacc	4	0.2646	0.1043	0.0601	0.4690	6.43	0.0112
annoacc	5	0.4359	0.1010	0.2380	0.6338	18.64	<.0001
annoacc	6	0.5303	0.0995	0.3352	0.7254	28.38	<.0001
annoacc	7	0.5360	0.1003	0.3394	0.7325	28.56	<.0001
annoacc	8	0.6910	0.0980	0.4990	0.8830	49.77	<.0001
annoacc	9	0.7318	0.0980	0.5398	0.9238	55.81	<.0001
annoacc	10	0.8491	0.0968	0.6595	1.0387	77.02	<.0001
annoacc	11	0.7718	0.0995	0.5767	0.9668	60.16	<.0001
annoacc	12	0.7203	0.1040	0.5164	0.9242	47.93	<.0001
annoacc	13	0.7465	0.1230	0.5054	0.9876	36.83	<.0001
annoacc	99	0.0000	0.0000	0.0000	0.0000	.	.
annodiff	1	-0.0887	0.0415	-0.1701	-0.0074	4.58	0.0324
annodiff	2	-1.1171	0.0604	-1.2354	-0.9987	342.42	<.0001
annodiff	3	-1.7941	0.0840	-1.9588	-1.6294	455.86	<.0001
annodiff	4	-2.3215	0.1140	-2.5450	-2.0979	414.38	<.0001
annodiff	5	-2.3994	0.1270	-2.6484	-2.1505	356.89	<.0001
annodiff	6	-2.7240	0.1606	-3.0387	-2.4093	287.77	<.0001
annodiff	7	-2.2936	0.1422	-2.5723	-2.0149	260.13	<.0001
annodiff	8	-3.0068	0.2208	-3.4396	-2.5741	185.47	<.0001
annodiff	9	-3.2233	0.2746	-3.7615	-2.6852	137.81	<.0001
annodiff	10	-3.3294	0.3262	-3.9687	-2.6901	104.19	<.0001
annodiff	11	-3.6627	0.4449	-4.5347	-2.7907	67.78	<.0001
annodiff	12	-3.1271	0.4352	-3.9800	-2.2742	51.64	<.0001
annodiff	13	-2.3422	0.4381	-3.2008	-1.4836	28.59	<.0001
annodiff	99	0.0000	0.0000	0.0000	0.0000	.	.
Scale	0	430.3712	0.0000	430.3712	430.3712		

NOTE: The scale parameter was estimated by the square root of Pearson's Chi-Square/DOF.

Le stime dei parametri di regressione sono ora meno precise. Quasi tutti i parametri risultano significativi al 5%.

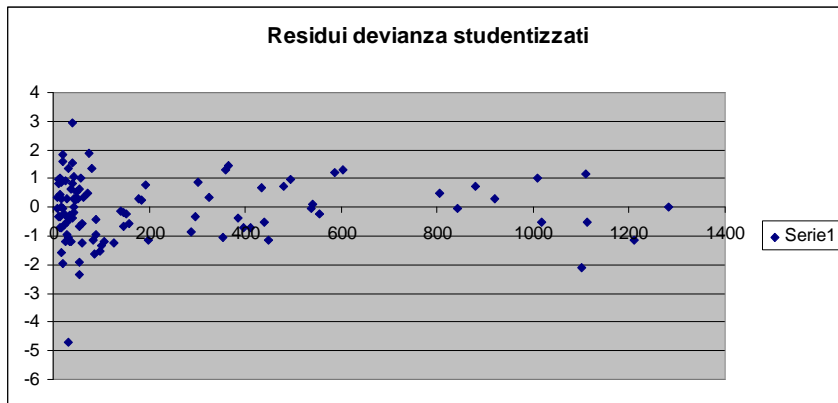
I residui standardizzati di Pearson non segnalano problemi nell'adattamento.



Disponendo dei dati relativi ad una grandezza che fornisca una misura di esposizione, si possono considerare modelli nei quali le variabili risposta sono grandezze rapportate a tale misura.

Modello di Poisson con sovradisersione-logaritmo per pagamenti rapportati ad una misura di esposizione

- **variabili risposta:** i rapporti $Y_{ij} = P_{ij} / w_i$, w_i numero di sinistri dell'anno di origine i , denunciati nell'anno, **modello di Poisson con sovra-dispersione**, con **pesi i numeri di sinistri denunciati** $\omega_{ij} = w_i$,
- **variabili esplicative:** anno di accadimento (*annoacc*), anno di differimento (*annodiff*),
 $\eta_{ij} = x_{ij}' \beta = \mu + \alpha_i + \beta_j$,
- **funzione di collegamento:** $g = \log$.



E' naturale descrivere i pagamenti incrementali P_{ij} come somma dei pagamenti per sinistri trattati con differimento j ,

$$P_{ij} = \sum_{h=1}^{N_{ij}} Z_{ij}^{(h)},$$

dove

- N_{ij} è il numero di pagamenti della cella (i,j) ,
- $Z_{ij}^{(h)}$ è l'importo del pagamento h -esimo.

Se si assegna a P_{ij} una **distribuzione Poisson-composta** con distribuzione dei pagamenti di tipo gamma, allora i numeri aleatori

$$Y_{ij} = P_{ij} / \omega_{ij}$$

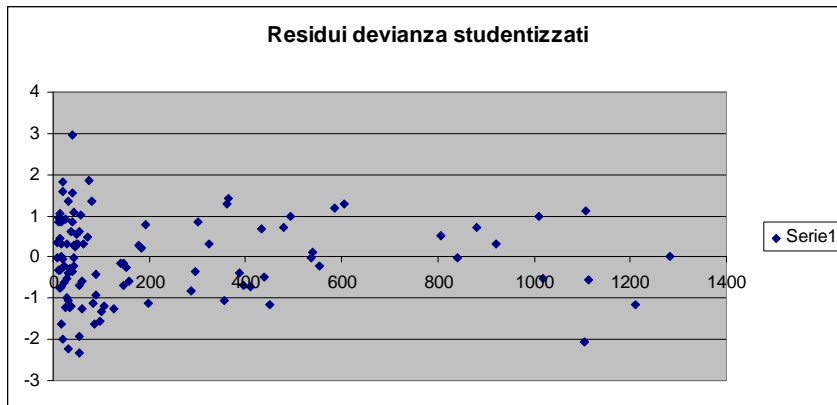
hanno **distribuzione appartenente ad una famiglia esponenziale lineare** con pesi ω_{ij} e **funzione di varianza di tipo potenza**

$$V(\mu) = \mu^p, \quad 1 < p < 2,$$

dove p è legato al parametro di forma della distribuzione gamma (Jorgensen, de Suoza (1994)).

Modello di Poisson-composto-logaritmo per pagamenti rapportati ad una misura di esposizione

- **variabili risposta:** i rapporti $Y_{ij} = P_{ij} / w_i$, w_i numero di sinistri dell'anno di origine i , denunciati nell'anno, **modello di Poisson composta**, con **pesi i numeri di sinistri denunciati** $\omega_{ij} = w_i$: famiglia esponenziale lineare con funzione di varianza di tipo potenza, poniamo $p = 1,01887$.
- **variabili esplicative:** anno di accadimento (*annoacc*), anno di differimento (*annodiff*),
 $\eta_{ij} = x_{ij}' \boldsymbol{\beta} = \mu + \alpha_i + \beta_j$,
- **funzione di collegamento:** $g = \log$.



PROCEDIMENTI DI SELEZIONE DELLE VARIABILI

I modelli di regressione sono spesso utilizzati in situazioni in cui ci sono numerose variabili esplicative potenzialmente influenti sulla valutazione probabilistica delle variabili risposta.

Un modello con molte variabili ha molti parametri. Ma un'importante proprietà di ogni modello statistico è quella della **parsimonia** nel numero dei parametri.

I procedimenti di selezione delle variabili hanno l'obiettivo di **determinare un sottoinsieme di variabili esplicative significative**, in modo che il modello stimato realizzi un buon adattamento ai valori osservati, ma che dipenda da un numero relativamente basso di parametri.

I **procedimenti automatici di selezione** hanno un duplice obiettivo:

- ridurre il numero di modelli da analizzare,
- guidare nella selezione.

I più diffusi sono procedimenti iterativi di tipo *forward*, *backward* e *stepwise*. Si basano sul confronto tra modelli, uno annidato nell'altro.

Procedimento *forward*

- Si parte dal **modello nullo** che ha come unico parametro l'intercetta β_0 .
- **Si inserisce** nel modello **per prima la variabile** maggiormente significativa: quella tra le significative (con $p\text{-value} < \alpha$) **alla quale corrisponde il $p\text{-value}$ più piccolo.**
- Al secondo passo, si confronta il modello così ottenuto con tutti i modelli che contengono, oltre alla variabile selezionata al passo precedente, una nuova variabile e si procede come sopra.
- Il procedimento termina quando non ci sono più variabili significative, cioè quando a partire da un modello selezionato l'introduzione di una ulteriore variabile, qualunque essa sia, comporta un $p\text{-value}$ maggiore o uguale del fissato livello α , oppure quando si è raggiunto un numero prefissato di variabili.

Spesso è usata la statistica $\tilde{\lambda}$, nei modelli con quasi-verosimiglianza la statistica di Wald \tilde{w} .

Procedimento *backward*

- Si parte dal **modello completo** con tutte le variabili inserite.
- **Si toglie** dal modello **per prima la variabile** meno significativa: quella tra le non significative (con $p\text{-value} \geq \alpha$) **alla quale corrisponde il $p\text{-value}$ più elevato.**
- Al secondo passo, si confronta il modello così ottenuto con tutti i modelli che si ottengono eliminando un'ulteriore variabile esplicativa, procedendo come sopra.
- Il procedimento termina quando, a partire da un modello, l'eliminazione di una variabile, qualunque essa sia, comporta un $p\text{-value}$ minore di α .

Procedimento *stepwise*

Combina il procedimento *forward* con il procedimento *backward*.

Con riferimento al generico passo, si considera **un passo *forward*** che conduce ad introdurre nel modello una nuova variabile.

A questo punto, si attua **un passo *backward*** per verificare se la variabile appena inserita renda superflua qualcuna delle variabili precedentemente inserite nel modello.

Il procedimento termina quando si trova un modello uguale ad uno già ottenuto.

STRUTTURE DI REGRESSIONE

Con i dati usualmente disponibili, le variabili esplicative **anno di origine, anno di differimento o sviluppo, anno di pagamento** consentono di tenere conto di diversi aspetti del problema.

Se sono trattate come **variabili di classificazione**, i parametri di regressione sono: l'intercetta μ , $\alpha_0, \dots, \alpha_t$ per gli anni di origine, β_0, \dots, β_t per gli anni di sviluppo, $\gamma_0, \dots, \gamma_t$ per gli anni di pagamento:

$$\boldsymbol{\beta} = (\mu, \alpha_0, \alpha_1, \dots, \alpha_t, \beta_0, \beta_1, \dots, \beta_t, \gamma_0, \gamma_1, \dots, \gamma_t)'$$

La matrice del modello X ha

- $n = (t+1)(t+2)/2$ righe, pari al numero di osservazioni (di celle),
- $1 + 3(t+1)$ colonne di cui solo $p = 3t$ sono linearmente indipendenti.

Il rango di X , pari a p , è il **numero di parametri da stimare**. Per evitare problemi di indeterminatezza, si può fissare, per esempio pari a zero, il valore dei parametri corrispondenti alle colonne di X combinazioni lineari delle altre e calcolare le stime dei rimanenti:

- p **numero dei parametri del modello**
- $n - p$ **numero dei gradi di libertà**.

In particolare, si può porre $\alpha_0 = 0$, $\beta_0 = 0$, $\gamma_0 = 0$ ed ancora, per esempio, $\gamma_t = 0$.

Se la **funzione di collegamento** è il **logaritmo**, si ha un **modello moltiplicativo** per la speranza matematica delle variabili risposta,

$$E(Y_{ij}) = \mu_{ij} = \exp(\mu + \alpha_i + \beta_j + \gamma_{i+j}) = K a_i b_j c_{i+j},$$

dove $K = \exp(\mu)$, $a_i = \exp(\alpha_i)$, $b_j = \exp(\beta_j)$ e $c_k = \exp(\gamma_k)$.

Dalle $\alpha_0 = \beta_0 = \gamma_0 = 0$, segue che K è il valore atteso della variabile risposta corrispondente al primo anno di origine e al differimento nullo.

Poiché la variabile **anno di pagamento** è **di classificazione**, si hanno le stime dei parametri γ_k solo per $k = 1, \dots, t$. Ai fini della previsione occorre disporre delle stime dei parametri per gli anni futuri. In tal caso, si considerano stime $\hat{\gamma}_k$, $k = t+1, \dots, 2t$, ottenute per estrapolazione dai valori stimati, $\hat{\gamma}_1, \dots, \hat{\gamma}_t$. Il problema non si pone se la variabile è trattata come numerica.

L'uso delle tre variabili esplicative come variabili di classificazione può portare a **problemi di sovrapparametrizzazione**. Per esempio: per $t = 13$, $n = 105$ osservazioni e $p = 39$ parametri da stimare.

La **scelta della struttura di regressione** è un aspetto delicato della scelta di un modello. Può essere suggerita da analisi preliminari sviluppate sui dati ed anche da analisi inferenziali di confronto tra modelli, basate su indicatori della bontà dell'accostamento ai dati e su analisi dei residui.

Altre strutture di regressione

Consideriamo alcune strutture di regressione e l'effetto sulla speranza matematica delle variabili risposta, nell'ipotesi che la **funzione di collegamento** sia il **logaritmo**.

Due variabili di classificazione

Il numero dei parametri del modello è $p = 1 + 2t$.

- Anno di origine e anno di differimento

$$E(Y_{ij}) = \exp(\mu + \alpha_i + \beta_j) = K a_i b_j.$$

- Anno di differimento e anno di pagamento

$$E(Y_{ij}) = \exp(\mu + \beta_j + \gamma_{i+j}) = K b_j c_{i+j}.$$

Con la prima struttura di regressione si ritrova l'ipotesi alla base del metodo della catena.

Una variabile di classificazione e una numerica

Il modello ha $p = 2 + t$ parametri.

- Anno di origine di classificazione e anno di differimento numerica

$$E(Y_{ij}) = \exp(\mu + \alpha_i + \beta_j) = K a_i b^j,$$

dove $b = \exp(\beta)$. Il modello traduce l'ipotesi che, in termini attesi, **la quota dell'importo totale di risarcimenti per sinistri dell'anno di origine i , pagata di anno in anno, decresca di una fissata percentuale b** .

- Anno di origine di classificazione e anno di pagamento numerica

$$E(Y_{ij}) = \exp(\mu + \alpha_i + \gamma(i+j)) = K a_i c^{i+j},$$

dove $c = \exp(\gamma)$. Interpretando l'effetto dovuto all'anno di pagamento come effetto inflazionistico, il modello traduce l'ipotesi che **il tasso annuo di inflazione dei costi dei sinistri sia costante**.

I due modelli forniscono la **stessa stima del valore atteso** delle variabili risposta. Dal secondo si ha infatti

$$E(Y_{ij}) = K a_i c^{i+j} = K (a_i c^i) c^j = K A_i B^j,$$

dove $A_i = a_i c^i$ e $B = c$. La struttura a ultimo membro coincide con quella del primo modello.

Particolari andamenti dell'inflazione possono suggerire scelte diverse rispetto alla

$$E(Y_{ij}) = \exp(\mu + \alpha_i + \gamma(i+j)) = Ka_i c^{i+j}.$$

Per esempio, se si osservano andamenti diversi negli anni di pagamento da 0 a T e da T in poi, si può introdurre nel previsore lineare una componente del tipo (Taylor (2000))

$$\gamma_1(i+j) + \gamma_2 \max(0, i+j-T).$$

Ciò equivale a considerare come variabili esplicative numeriche l'anno di pagamento $i+j$ e la sua trasformata $\max(0, i+j-T)$.

- Anno di origine numerica e anno di differimento di classificazione

$$E(Y_{ij}) = \exp(\mu + \alpha_i + \beta_j) = Ka^i b_j,$$

dove $a = \exp(\alpha)$. Il modello traduce l'ipotesi che, in termini attesi, **i risarcimenti totali per anno di origine si incrementino o si riducano di una fissata percentuale in ogni anno**, per esempio per effetto di espansioni o contrazioni di portafoglio.

- Anno di differimento di classificazione e anno di pagamento numerica

$$E(Y_{ij}) = \mu_{ij} = \exp(\mu + \beta_j + \gamma(i+j)) = Kb_j c^{i+j}.$$

Gli ultimi due modelli forniscono la stessa stima per la speranza matematica delle variabili risposta.

Due variabili numeriche

I parametri del modello sono tre.

- Anno di origine e anno di differimento

$$E(Y_{ij}) = \exp(\mu + \alpha_i + \beta_j) = Ka^i b^j.$$

- Anno di differimento e anno di pagamento

$$E(Y_{ij}) = \exp(\mu + \beta_j + \gamma(i+j)) = Kb^j c^{i+j}.$$

I due modelli forniscono le stesse stime delle speranze matematiche μ_{ij} .

Trasformate di una variabile numerica

Differenziazione del primo anno di differimento

Considerando l'anno di differimento come numerica, può essere conveniente distinguere il primo anno di differimento rispetto agli altri.

Infatti, se l'anno di origine è l'anno di sottoscrizione della polizza, **i contratti stipulati nell'anno i hanno esposizione inferiore ad un anno nel primo anno di differimento** (Kaas *et al.* (2001)).

Per esempio, considerando l'anno di origine di classificazione, si ha un modello tale che

$$E(Y_{i0}) = \exp(\mu + \alpha_i + \beta_0) = Ka_i b_0$$
$$E(Y_{ij}) = \exp(\mu + \alpha_i + \beta_j) = Ka_i b^j, \quad \text{per } j \neq 0.$$

Equivale ad introdurre nel modello, oltre al differimento, una variabile esplicativa di tipo indicatore, funzione del differimento, che assume valore 1 se il differimento è 0.

Strutture di tipo Hoerl curve

Se la variabile anno di differimento è trattata come numerica, una struttura di regressione proposta in letteratura prevede di considerare come ulteriore variabile esplicativa il **logaritmo del differimento**.

L'espressione del valore atteso della variabile risposta, considerando l'anno di origine come variabile di classificazione, è

$$E(Y_{ij}) = \exp(\mu + \alpha_i + \beta_j + \delta \log(j+1)) = Ka_i e^{\beta_j} (j+1)^\delta.$$

Il numero di parametri del modello è $p = 1 + t + 2$.

Questa struttura è un caso particolare di *Hoerl curve* (De Jong, Zhenwirth (1983), Zehnwirth (1998), Wright (1990)).

Una forma più generale di *Hoerl curve*, prevede di definire una struttura di regressione del tipo

$$E(Y_{ij}) = \exp(\mu + \alpha_i + \beta_i j + \delta_i \log(j+1)) = Ka_i e^{\beta_i j} (j+1)^{\delta_i}.$$

Equivale a considerare l'anno di origine come variabile di classificazione, l'anno di differimento numerica ed introdurre come variabili esplicative nel modello l'anno di origine e le **interazioni** tra anno di origine e anno di differimento e tra anno di origine e logaritmo dell'anno di differimento.

Esempio. Strutture di regressione

Per analizzare la significatività delle variabili esplicative anno di origine, anno di differimento e anno di pagamento, consideriamo diverse strutture di regressione in un modello GLM semiparametrico con

- **variabili risposta** i rapporti $Y_{ij} = P_{ij} / \omega_i$,
- ω_i **numero di sinistri dell'anno di origine i , denunciati nell'anno**,
- **modello di Poisson con sovra-dispersione**, con **pesi i numeri di sinistri denunciati ω_i** ,
- **funzione di collegamento logaritmo**.

Osservazione.

La stima di ϕ può essere ricavata in modo indipendente dalla stima dei parametri di regressione, in quanto i valori di massima verosimiglianza o di massima quasi-verosimiglianza di tali parametri non dipendono dal parametro di dispersione.

Convien allora stimare ϕ considerando il modello più raffinato, anche se questo può dipendere da un numero molto elevato di parametri, e poi selezionare il miglior modello per la struttura di regressione.

Con lo stimatore di Pearson si ottiene $\hat{\phi} = 161.567$.

Nella Tabella sono riportati, per ciascun modello, il valore della **quasi-devianza D** , il **numero di gradi di libertà df** e la **quasi-devianza scalata $S = D / \hat{\phi}$** .

Modello	η_{ij}	D	df	$S = D / \hat{\phi}$
1	$\mu + \alpha_i + \beta_j + \gamma_{i+j}$	10.720.784	66	66,4
2	$\mu + \alpha_i + \beta_j$	14.260.425	78	88,3
3	$\mu + \beta_j + \gamma_{i+j}$	18.076.648	78	111,9
4	$\mu + \beta_j + \gamma(i+j)$	18.557.627	90	114,9
5	$\mu + \alpha_i + \beta_j$	18.557.627	90	114,9
6	$\mu + \alpha_i + \gamma_{i+j}$	65.946.057	78	408,2
7	$\mu + \alpha_i + \beta_1 \mid j = 0 + \beta_j$	67.834.873	89	419,9
8	$\mu + \alpha_i + \gamma(i+j)$	69.929.014	90	432,8
9	$\mu + \alpha_i + \beta_j$	69.929.014	90	432,8
10	$\mu + \alpha_i + \beta_1 \mid j = 0 + \beta_j$	73.258.370	101	453,4
11	$\mu + \alpha_i + \beta_j$	75.660.230	102	468,3
12	$\mu + \beta_j + \gamma(i+j)$	75.660.230	102	468,3
13	$\mu + \beta_j$	115.428.546	91	714,4
14	$\mu + \beta_j$	171.807.874	103	1063,4
15	$\mu + \alpha_i$	509.021.391	91	3150,5
16	$\mu + \gamma_{i+j}$	913.754.936	91	5655,6
17	$\mu + \gamma(i+j)$	916.212.332	103	5670,8
18	μ	927.524.270	104	5740,8

Modello di **Poisson con sovra-dispersione**, diverse strutture di regressione

Guardando alla colonna delle **quasi-devianze**, si ha

- Il **modello peggiore è il 18**, il *modello nullo*, nel quale si assegna la medesima distribuzione a tutte le variabili risposta.

- La **devianza minima si ha in corrispondenza del modello 1**, che dunque è quello che realizza il miglior accostamento ai dati. Questo modello dipende però da molti parametri in relazione al numero totale di dati disponibili, che è pari a 105.

- Tra i **modelli 2, 3 e 6** che considerano due sole variabili esplicative entrambe di classificazione,

- il **modello 2**, che fornisce le stesse stime del **metodo della catena**, è quello che **realizza il miglior accostamento**,
- il modello 3, che fornisce le stesse stime del **metodo di separazione aritmetico**, conduce ad un valore per la devianza superiore a quello del modello 2,
- il peggiore è il modello 6 con devianza decisamente più elevata rispetto a quella dei due modelli precedenti.

Guardando ai modelli 2 e 6 si ha che, per i nostri dati, avendo inserito nel modello un'intercetta e la variabile anno di origine, **la variabile anno di differimento è più significativa della variabile anno di pagamento**.

- I modelli 4 e 5 conducono allo stesso valore di devianza, deriva dal fatto che i due modelli sono equivalenti.

Analoghe considerazioni sussistono per i modelli 8 e 9 e per i modelli 11 e 12.

Tutti i modelli della Tabella possono essere visti come sottomodelli del modello 1: si ottengono da questo ponendo vincoli su alcuni parametri.

Per il **confronto di due modelli gerarchici nell'approccio semiparametrico** si utilizza la **statistica di Wald**. Fissiamo il livello di significatività al 5%.

Confronto tra il **modello 1**, $\mu + \alpha_i + \beta_j + \gamma_{i+j}$, ed i modelli con due variabili di classificazione. L'*output* di SAS è:

Wald Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
annoacc	12	45.21	<.0001
annodiff	12	374.63	<.0001
annopag	12	22.04	0.0371

- Nell'ultima riga si confronta il **modello 1** col modello che si ottiene eliminando la variabile anno di pagamento (annopag) che è il **modello 2**, $\mu + \alpha_i + \beta_j$. La statistica Chi-Square ha valore 22,04, molto vicino al valore critico della distribuzione $\chi^2(12)$ che è 21,03. Formalmente, questo porterebbe a rifiutare il modello 2 rispetto al modello 1. Il valore della statistica è però molto vicino al valore critico della distribuzione. **Ad un livello di significatività del 3%, il modello 2 non sarebbe rifiutato.**
- **Dal modello 1 al modello 3**, $\mu + \beta_j + \gamma_{i+j}$, la statistica ha valore 45,21, **il modello viene rifiutato**,
- **Dal modello 1 al modello 6**, $\mu + \alpha_i + \gamma_{i+j}$, la statistica ha valore 374,63, **il modello viene rifiutato.**

Gli altri modelli vengono tutti rifiutati, rispetto al modello 1.

Dalle precedenti analisi emerge che, **con il modello di Poisson con sovradisersione**, le strutture di regressione che conducono a **modelli che meglio accostano i dati** sono quelle del **modello 1**, $\mu + \alpha_i + \beta_j + \gamma_{i+j}$, e quelle del **modello 2**, $\mu + \alpha_i + \beta_j$.

Conclusione. La scelta potrebbe ricadere sulla struttura del **modello 2** perché

- **non risulta essere significativamente peggiore del modello 1, pur riducendo notevolmente il numero di parametri da stimare,**
- **è più semplice, anche in vista del problema di previsione.** Infatti se si assegnano ai numeri aleatori Y_{ij} distribuzioni con la struttura di regressione del modello 1, nelle previsioni intervengono i parametri $\gamma_{t+1}, \dots, \gamma_{2t}$ le cui stime non sono fornite dal GLM.

PREVISIONI ED ERRORI DI PREVISIONE

Descrizione generale

Facciamo riferimento ad un generico quadro di numeri aleatori di interesse Y_{ij} , con $i, j = 0, 1, \dots, t$. Appartengono al triangolo superiore i numeri aleatori con $i + j \leq t$, al triangolo inferiore quelli con $i + j > t$.

Ci mettiamo in particolare **nell'ambito dei GLM**. Pertanto, per ogni i, j

- ipotesi di indipendenza dei numeri aleatori Y_{ij} ,
- distribuzioni dei numeri aleatori Y_{ij} appartenenti alla stessa famiglia parametrica o con la stessa funzione di varianza,
- stessa struttura di regressione.

Ci interessano **previsioni dei numeri aleatori del triangolo inferiore o di funzioni di questi**, nonché **valutazioni sulla qualità dei previsori**.

Sia W una grandezza aleatoria funzione delle variabili del triangolo inferiore,

$$W = \bar{f}(Y_{ij} : i + j > t).$$

Per ottenere una **previsione di W** si considera **uno stimatore** \tilde{W} , funzione dei numeri aleatori Y_{ij} del triangolo superiore,

$$\tilde{W} = f(Y_{ij} : i + j \leq t).$$

NB: \tilde{W} e W sono stocasticamente indipendenti.

Previsione

Il valore osservato dello stimatore è una previsione di W

$$\hat{W} = f(y_{ij} : i + j \leq t).$$

Errore di previsione dello stimatore

Un indicatore della qualità della stima è la radice quadrata del *mean square error of prediction*, non condizionato. Poiché \tilde{W} e W sono stocasticamente indipendenti, si ha

$$MSEP(\tilde{W}) = E(W - \tilde{W})^2 = \text{var}(W) + E[(\tilde{W} - E(W))^2],$$

- $\text{var}(W)$ tiene conto della variabilità intrinseca di W : la sua radice quadrata è un indicatore del **rischio di processo**,
- il secondo addendo tiene conto della variabilità della stima: la sua radice quadrata è un indicatore del **rischio di stima**. Se $E(W) = E(\tilde{W})$, il secondo addendo è $\text{var}(\tilde{W})$.

Una valutazione di $MSEP(\tilde{W})$ si può ottenere considerando stime dei due addendi: $\text{var}(W)$, $E[(\tilde{W} - E(W))^2]$.

Gli stimatori \tilde{W} proposti sono funzioni dello stimatore di massima verosimiglianza del parametro di regressione β .

Previsione ed errore di previsione per Y_{ij} numero aleatorio del triangolo inferiore.

Sia Y_{ij} un numero aleatorio del triangolo inferiore. Si ha

$$E(Y_{ij}) = \mu_{ij} = g^{-1}(\mathbf{x}'_{ij} \boldsymbol{\beta})$$

essendo

\mathbf{x}_{ij} è il vettore delle determinazioni delle variabili esplicative relative a Y_{ij} , $\boldsymbol{\beta}$ il vettore dei parametri di regressione,

g la funzione di collegamento.

Come stimatore di Y_{ij} è naturale considerare

$$\tilde{Y}_{ij} = g^{-1}(\mathbf{x}'_{ij} \tilde{\boldsymbol{\beta}}) = h(\tilde{\boldsymbol{\beta}})$$

con $\tilde{\boldsymbol{\beta}}$ stimatore di massima verosimiglianza di $\boldsymbol{\beta}$.

Infatti, la previsione di Y_{ij} , che è il valore osservato dello stimatore \tilde{Y}_{ij} , è $\hat{Y}_{ij} = \hat{\mu}_{ij} = g^{-1}(\mathbf{x}'_{ij} \hat{\boldsymbol{\beta}})$, con $\hat{\boldsymbol{\beta}}$ stima di massima verosimiglianza di $\boldsymbol{\beta}$, e coincide con la stima della speranza matematica di Y_{ij} , $\hat{E}(Y_{ij}) = \hat{\mu}_{ij}$.

Valutazione dell'errore di previsione

$$MSEP(\tilde{Y}_{ij}) = E(Y_{ij} - \tilde{Y}_{ij})^2 = \text{var}(Y_{ij}) + E[(\tilde{Y}_{ij} - E(Y_{ij}))^2].$$

Per valutare $\text{var}(Y_{ij})$ si può considerare la varianza stimata dal modello

$$\hat{\text{var}}(Y_{ij}) = \frac{\hat{\phi}}{\omega_{ij}} V(\hat{\mu}_{ij}).$$

Per ottenere una stima di $E[(\tilde{Y}_{ij} - E(Y_{ij}))^2]$ ci si può basare sulle due approssimazioni di seguito indicate, che saranno utilizzate ampiamente nel seguito.

- Dalle proprietà asintotiche **dello stimatore di massima verosimiglianza**, se il numero delle osservazioni è sufficientemente elevato, la distribuzione di $\tilde{\beta}$ può essere approssimata con la distribuzione normale multidimensionale

$$\tilde{\beta} \approx N(\hat{\beta}, [\mathcal{I}(\hat{\beta})]^{-1})$$

dove $\mathcal{I}(\hat{\beta})^{-1}$ è l'inversa della matrice matrice di informazione di Fisher calcolata nella stima di massima verosimiglianza di β ;

- **Delta method**. Dalle proprietà asintotiche di una **funzione dello stimatore di massima verosimiglianza**, se il numero delle osservazioni è sufficientemente elevato, la distribuzione di $h(\tilde{\beta})$ può essere approssimata con la normale

$$h(\tilde{\beta}) \approx N(h(\hat{\beta}), J_h(\hat{\beta}) \mathcal{I}(\hat{\beta})^{-1} J_h(\hat{\beta})')$$

dove $J_h(\hat{\beta})$ è la matrice jacobiana della funzione h , calcolata nella stima $\hat{\beta}$ di massima verosimiglianza. Si richiede che la funzione h , scalare o vettoriale, sia differenziabile con matrice jacobiana calcolata in $\hat{\beta}$ di rango massimo. La $J_h(\hat{\beta})'$ indica la trasposta della matrice jacobiana $J_h(\hat{\beta})$.

In forza di questi due risultati si ha

$$\hat{E}(\tilde{Y}_{ij}) = h(\hat{\beta}) = \hat{E}(Y_{ij}),$$

quindi

$$E[(\tilde{Y}_{ij} - E(Y_{ij}))^2] \cong E[(\tilde{Y}_{ij} - \hat{E}(Y_{ij}))^2] = E[(\tilde{Y}_{ij} - \hat{E}(\tilde{Y}_{ij}))^2] \cong E[(\tilde{Y}_{ij} - E(\tilde{Y}_{ij}))^2]$$

e allora $E[(\tilde{Y}_{ij} - E(Y_{ij}))^2]$ si può approssimare con $var(\tilde{Y}_{ij})$.

Per il *delta method*, una stima della varianza dello stimatore è data da

$$\hat{var}(\tilde{Y}_{ij}) = J_h(\hat{\beta}) \mathcal{I}(\hat{\beta})^{-1} J_h(\hat{\beta})'.$$

Si ha dunque

$$MS\hat{E}P(\tilde{Y}_{ij}) = \frac{\hat{\phi}}{\omega_{ij}} V(\hat{\mu}_{ij}) + J_h(\hat{\beta}) \mathcal{I}(\hat{\beta})^{-1} J_h(\hat{\beta})',$$

la cui radice quadrata è l'errore di previsione.

Più in generale, se $\tilde{W} = f(Y_{ij} : i + j \leq t) = h(\tilde{\beta})$, con h regolare, e si ha $\hat{E}(\tilde{W}) = h(\hat{\beta}) = \hat{E}(W)$, si può stimare l'errore di previsione con la radice di

$$MS\hat{E}P(\tilde{W}) = \hat{var}(W) + J_h(\hat{\beta}) [\mathcal{I}(\hat{\beta})]^{-1} J_h(\hat{\beta})'.$$

Previsioni per i pagamenti incrementali e di loro funzioni.

Per ottenere previsioni ed errori di previsione per i pagamenti incrementali P_{ij} del triangolo inferiore e per loro funzioni, consideriamo **GLM** con variabili risposta i **pagamenti incrementali divisi per una misura di esposizione**.

Come variabili esplicative consideriamo *l'anno di origine* e *l'anno di differimento*, trattate come variabili di classificazione. Naturalmente, si possono considerare anche altre strutture di regressione.

Più in dettaglio,

- **variabili risposta:** $Y_{ij} = P_{ij} / w_i$

P_{ij} pagamenti incrementali,

w_i **misure di esposizione**, per esempio numero di polizze o numero di sinistri o premi dell'anno i ,

le misure di esposizione siano anche i pesi delle distribuzioni delle variabili risposta: $\omega_{ij} = w_i$,

- **variabili esplicative:** **anno di origine** e **anno di sviluppo**, di classificazione,
- **funzione di collegamento:** g .

Aspetto stima

Il vettore dei parametri di regressione $\boldsymbol{\beta}$ ed il vettore \mathbf{x}_{ij} delle determinazioni delle variabili esplicative corrispondente a Y_{ij} , $i + j \leq t$, sono dati da

$$\begin{aligned} \boldsymbol{\beta} &= (\mu \quad \alpha_0 \quad \dots \quad \alpha_i \quad \dots \quad \alpha_t \quad \beta_0 \quad \dots \quad \beta_j \quad \dots \quad \beta_t) \\ \mathbf{x}'_{ij} &= (1 \quad 0 \quad \dots \quad 1 \quad \dots \quad 0 \quad 0 \quad \dots \quad 1 \quad \dots \quad 0) \end{aligned}$$

Il previsore lineare è

$$\eta_{ij} = \mathbf{x}'_{ij} \boldsymbol{\beta} = \mu + \alpha_i + \beta_j.$$

Siano

- $\tilde{\boldsymbol{\beta}} = (\tilde{\mu}, \tilde{\alpha}_0, \dots, \tilde{\alpha}_t, \tilde{\beta}_0, \dots, \tilde{\beta}_t)$, lo **stimatore di massima verosimiglianza** dei parametri di regressione basato sulle variabili del triangolo di *run-off*, $Y_{ij} : i + j \leq t$,
- $\hat{\boldsymbol{\beta}} = (\hat{\mu}, \hat{\alpha}_0, \dots, \hat{\alpha}_t, \hat{\beta}_0, \dots, \hat{\beta}_t)$, la **stima di massima verosimiglianza** basata sui dati del triangolo di *run-off*, $y_{ij} : i + j \leq t$,
- $\hat{\phi}$ la stima del parametro di dispersione.

Con una condizione, per esempio $\alpha_0 = \beta_0 = 0$.

Previsioni ed errori di previsione per i pagamenti incrementali

Nel triangolo inferiore $i + j > t$.

Sia \mathbf{x}_{ij} il vettore delle determinazioni delle variabili esplicative per P_{ij} ,

$$\begin{aligned} \boldsymbol{\beta} &= (\mu \quad \alpha_0 \quad \alpha_1 \quad \dots \quad \alpha_i \quad \dots \quad \alpha_t \quad \beta_0 \quad \dots \quad \beta_j \quad \dots \quad \beta_t) \\ \mathbf{x}'_{ij} &= (1 \quad 0 \quad 0 \quad \dots \quad 1 \quad \dots \quad 0 \quad 0 \quad \dots \quad 1 \quad \dots \quad 0) \end{aligned}$$

Per comodità di scrittura, consideriamo anche l'anno di origine 0 e l'anno di differimento 0, ai quali corrisponde sempre la determinazione 0.

- **Stimatore di P_{ij} , $i + j > t$,**

$$\tilde{P}_{ij} = w_i g^{-1}(\mathbf{x}'_{ij} \tilde{\boldsymbol{\beta}}) = w_i g^{-1}(\tilde{\eta}_{ij}) = w_i g^{-1}(\tilde{\mu} + \tilde{\alpha}_i + \tilde{\beta}_j),$$

suggerito dalla $E(P_{ij}) = w_i E(Y_{ij}) = w_i \mu_{ij} = w_i g^{-1}(\mathbf{x}'_{ij} \boldsymbol{\beta})$.

- **Previsione di P_{ij} , $i + j > t$, valore osservato dello stimatore**

$$\hat{P}_{ij} = w_i \hat{\mu}_{ij} = w_i g^{-1}(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j).$$

Stima dell'errore di previsione

Posto $\tilde{P}_{ij} = w_i g^{-1}(\tilde{\mu} + \tilde{\alpha}_i + \tilde{\beta}_j) = h_{ij}(\tilde{\boldsymbol{\beta}})$, con h_{ij} funzione scalare data da $h_{ij}(\boldsymbol{\beta}) = w_i g^{-1}(\mathbf{x}'_{ij} \boldsymbol{\beta})$, \tilde{P}_{ij} è funzione dei numeri aleatori del triangolo superiore tramite $\tilde{\boldsymbol{\beta}}$.

Dal *delta method*

$$\hat{E}(\tilde{P}_{ij}) = h_{ij}(\hat{\boldsymbol{\beta}}) = w_i g^{-1}(\mathbf{x}'_{ij} \hat{\boldsymbol{\beta}}) = \hat{E}(P_{ij}).$$

Dall'espressione generale del quadrato dell'errore di previsione otteniamo

$$M\hat{S}EP(\tilde{P}_{ij}) = \underbrace{\text{vâr}(P_{ij})}_{\substack{\text{quadrato del} \\ \text{rischiodi} \\ \text{processo}}} + \underbrace{\text{vâr}(\tilde{P}_{ij})}_{\substack{\text{quadrato del} \\ \text{rischiodi} \\ \text{stima}}}$$

Rischio di processo. Dalle ipotesi di modello GLM, si ha

$$\text{var}(P_{ij}) = w_i^2 \frac{\phi}{w_i} V(\mu_{ij}).$$

Come valutazione approssimata si può considerare

$$\text{vâr}(P_{ij}) = w_i^2 \frac{\hat{\phi}}{w_i} V(\hat{\mu}_{ij}).$$

Rischio di stima. La valutazione approssimata di $\text{var}(\tilde{P}_{ij})$ ottenuta con il *delta method* è

$$\text{vâr}(\tilde{P}_{ij}) = J_{h_{ij}}(\hat{\boldsymbol{\beta}}) [\mathcal{S}(\hat{\boldsymbol{\beta}})]^{-1} J_{h_{ij}}(\hat{\boldsymbol{\beta}})'$$

La matrice jacobiana $J_{h_{ij}}(\hat{\boldsymbol{\beta}})$ si ottiene derivando $h_{ij}(\boldsymbol{\beta}) = w_i g^{-1}(\mathbf{x}'_{ij} \boldsymbol{\beta})$ rispetto alle componenti del vettore $\boldsymbol{\beta}$ e calcolando le derivate in $\hat{\boldsymbol{\beta}}$. Il risultato è il vettore descritto con lo schema

$$J_{h_{ij}}(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} \mu & \alpha_0 & \dots & \alpha_i & \dots & \alpha_t & \beta_0 & \dots & \beta_j & \dots & \beta_t \\ \left(\frac{w_i}{g'(\hat{\mu}_{ij})}\right) & 0 & \dots & \frac{w_i}{g'(\hat{\mu}_{ij})} & \dots & 0 & 0 & \dots & \frac{w_i}{g'(\hat{\mu}_{ij})} & \dots & 0 \end{pmatrix}$$

che si può scrivere

$$J_{h_{ij}}(\hat{\boldsymbol{\beta}}) = \frac{w_i}{g'(\hat{\mu}_{ij})} \mathbf{x}'_{ij}$$

Allora

$$\begin{aligned} \text{vâr}(\tilde{P}_{ij}) &= J_{h_{ij}}(\hat{\boldsymbol{\beta}}) [\mathcal{S}(\hat{\boldsymbol{\beta}})]^{-1} J_{h_{ij}}(\hat{\boldsymbol{\beta}})' = \left[\frac{w_i}{g'(\hat{\mu}_{ij})} \right]^2 \mathbf{x}'_{ij} [\mathcal{S}(\hat{\boldsymbol{\beta}})]^{-1} \mathbf{x}_{ij} \\ &= \left[\frac{\omega_{ij}}{g'(\hat{\mu}_{ij})} \right]^2 \text{vâr}(\tilde{\eta}_{ij}). \end{aligned}$$

Una *stima dell'errore di previsione* è pertanto

$$\begin{aligned} M\hat{S}EP(\tilde{P}_{ij}) &= \text{vâr}(P_{ij}) + \text{vâr}(\tilde{P}_{ij}) \\ &= \omega_{ij}^2 \frac{\hat{\phi}}{\omega_{ij}} V(\hat{\mu}_{ij}) + \left[\frac{w_i}{g'(\hat{\mu}_{ij})} \right]^2 \mathbf{x}'_{ij} \mathcal{S}(\hat{\boldsymbol{\beta}})^{-1} \mathbf{x}_{ij} \end{aligned}$$

Essa dipende dai valori delle stime $\hat{\mu}_{ij}$, $\hat{\phi}$ e da $\mathcal{S}(\hat{\boldsymbol{\beta}})^{-1}$, stima della matrice delle varianze e covarianze di $\tilde{\boldsymbol{\beta}}$.

Tali elementi sono forniti nell'*output* dai *software* che trattano i GLM.

Osserviamo ancora che all'espressione dello jacobiano $J_{h_{ij}}(\boldsymbol{\beta})$ si può arrivare considerando h_{ij} come funzione composta di più funzioni e applicando la regola per il calcolo dello jacobiano della composta.

Previsioni ed errori di previsione per i flussi e le riserve

Flussi:
$$F_{t+k} = \sum_{i+j=t+k} P_{ij} = \sum_{i+j=t+k} w_i Y_{ij}, \quad k = 1, \dots, t$$

$$DF_{t+k} = \sum_{i+j=t+k} v_{i+j} P_{ij} = v_{t+k} \sum_{i+j=t+k} w_i Y_{ij}, \quad k = 1, \dots, t$$

Riserve per anno di origine:
$$R_i = \sum_{j \geq t-i+1} P_{ij} = \sum_{j \geq t-i+1} w_i Y_{ij}, \quad i = 1, \dots, t$$

$$L_i = \sum_{j \geq t-i+1} v_{i+j} P_{ij} = \sum_{j \geq t-i+1} v_{i+j} w_i Y_{ij}, \quad i = 1, \dots, t$$

Riserva totale:
$$R = \sum_{(i,j):i+j>t} P_{ij} = \sum_{(i,j):i+j>t} w_i Y_{ij}$$

$$L = \sum_{(i,j):i+j>t} v_{i+j} P_{ij} = \sum_{(i,j):i+j>t} v_{i+j} w_i Y_{ij}$$

Sono tutti numeri aleatori del tipo

$$W = \sum_{(i,j) \in I} v_{i+j} P_{ij}$$

con I conveniente insieme di coppie di indici e $v_{i+j} = 1$ per le grandezze non attualizzate.

Come **stimatore** scegliamo allora

$$\tilde{W} = \sum_{(i,j) \in I} v_{i+j} \tilde{P}_{ij}.$$

Per

$$\tilde{W} = \sum_{(i,j) \in I} v_{i+j} \tilde{P}_{ij} = \sum_{(i,j) \in I} v_{i+j} w_i g^{-1}(\mathbf{x}'_{ij} \tilde{\boldsymbol{\beta}}) = \sum_{(i,j) \in I} v_{i+j} w_i g^{-1}(\tilde{\eta}_{ij}),$$

si ha

$$\hat{E}(\tilde{W}) = \sum_{(i,j) \in I} v_{i+j} \hat{E}(\tilde{P}_{ij}) = \sum_{(i,j) \in I} v_{i+j} w_i g^{-1}(\mathbf{x}'_{ij} \hat{\boldsymbol{\beta}}) = \hat{E}(W).$$

Una stima del MSEP è

$$\begin{aligned} MS\hat{E}P(\tilde{W}) &= \hat{v}ar(W) + \hat{v}ar(\tilde{W}) \\ &= \sum_{(i,j) \in I} v_{i+j}^2 \hat{v}ar(P_{ij}) + J_h(\hat{\boldsymbol{\beta}}) [\mathcal{J}(\hat{\boldsymbol{\beta}})]^{-1} J_h(\hat{\boldsymbol{\beta}})' \end{aligned}$$

con $J_h(\hat{\boldsymbol{\beta}})$ è la matrice jacobiana della funzione

$$h(\boldsymbol{\beta}) = \sum_{(i,j) \in I} v_{i+j} w_i g^{-1}(\mathbf{x}'_{ij} \boldsymbol{\beta}),$$

calcolata nella stima $\hat{\boldsymbol{\beta}}$.

Esempi numerici di calcolo degli errori di previsione

Riserve di generazione e riserva totale

Per le diverse generazioni sono riportati la riserva stimata, la radice quadrato dell'errore quadratico di previsione,

$$\text{MSEP}^{1/2} = \sqrt{\hat{v}ar(R_i) + \hat{v}ar(\tilde{R}_i)}$$

$$\text{Rischio di processo} = \sqrt{\hat{v}ar(R_i)}$$

$$\text{Rischio di stima} = \sqrt{\hat{v}ar(\tilde{R}_i)}$$

Esempio 1. Modello Poisson con sovradisersione, funzione di collegamento logaritmo, struttura di regressione $\mu + \alpha_i + \beta_j$, parametro di dispersione stimato dai dati $\hat{\phi} = 185.219$.

Anno <i>i</i>	Riserva stimata	MSEP ^{1/2}	in %	Rischio di processo	Rischio di stima
1	1.183.644	703.041	0,594	468.224	524.436
2	1.985.355	904.806	0,456	606.404	671.526
3	2.086.149	879.104	0,421	621.607	621.634
4	2.730.687	995.637	0,365	711.179	696.790
5	3.881.736	1.203.941	0,310	847.923	854.694
6	5.140.570	1.379.708	0,268	975.773	975.428
7	6.963.868	1.554.938	0,223	1.135.713	1.062.068
8	9.494.382	1.846.415	0,194	1.326.101	1.284.798
9	11.853.086	2.052.756	0,173	1.481.695	1.420.700
10	15.714.940	2.396.367	0,152	1.706.081	1.682.814
11	18.288.287	2.553.500	0,140	1.840.474	1.770.033
12	24.367.054	3.035.403	0,125	2.124.441	2.168.046
13	45.101.492	5.422.936	0,120	2.890.272	4.588.526
Totale	148.791.251	14.349.199	0,096	5.249.669	13.354.419

Stime delle riserve per il modello di Poisson con sovradisersione

C'è notevole incertezza nella stima delle riserve relative agli anni di origine più lontani, $i = 1, 2$, per i quali le riserve sono relativamente basse: ci sono pochi dati per la stima dei parametri β_j per j elevato.

L'errore di previsione in termini percentuali è poi decrescente. Le stime diventano più precise perché

- intervengono i β_j per j piccolo, per le stime dei quali si dispone di un maggior numero di dati,
- i valori β_j per j elevato hanno relativamente poca influenza.

Il **rischio di processo** e il **rischio di stima** per i diversi anni di origine **sostanzialmente si equivalgono**, tranne quelli relativi all'anno 13.

Per l'anno 13 il rischio di stima è più elevato a causa del fatto che **la previsione di R_{13} dipende** da un numero elevato di pagamenti incrementali e quindi **da un elevato numero di parametri stimati**.

Flussi di pagamenti incrementali

Per le stime dei flussi annui F_{t+k} di pagamenti futuri si possono fare considerazioni simmetriche.

Anno $t+k$	Flusso stimato	RMSEP	in %	Rischio di processo	Rischio di stima
14	42.462.314	3.639.927	0,086	2.804.433	2.320.394
15	23.092.666	2.447.149	0,106	2.068.141	1.308.177
16	16.155.902	2.030.149	0,126	1.729.851	1.062.600
17	13.021.468	1.864.800	0,143	1.553.006	1.032.303
18	11.252.596	1.800.632	0,160	1.443.675	1.076.140
19	9.466.201	1.714.486	0,181	1.324.131	1.089.100
20	8.214.373	1.648.822	0,201	1.233.475	1.094.144
21	6.311.480	1.561.083	0,247	1.081.207	1.126.043
22	5.362.957	1.512.095	0,282	996.656	1.137.149
23	4.659.583	1.508.285	0,324	929.002	1.188.225
24	3.664.461	1.358.827	0,371	823.850	1.080.593
25	3.017.498	1.257.794	0,417	747.596	1.011.508
26	2.109.751	1.132.472	0,537	625.113	944.312

Stime dei flussi di pagamenti per il modello di Poisson con sovredispersione

Si hanno errori di previsione percentualmente più bassi nei primi anni e crescenti nel tempo.

Nella previsione del primo flusso è piuttosto elevato il rischio di processo.

Esempio 2. Modello gamma, collegamento logaritmo, struttura di regressione $\mu + \alpha_i + \beta_j$, $\hat{\phi} = 2.803$.

Rispetto al modello di Poisson con sovra-disperzione: **stime generalmente inferiori. Errore di previsione elevato**, in particolare negli ultimi anni per le riserve, nei primi anni per i flussi: **maggiore variabilità delle stime**. L'errore di previsione della riserva per $i=13$ è del 46%, contro il 12%, l'errore di previsione del primo flusso è del 29%, contro il 9%.

Anno i	Riserva stimata	MSEP ^{1/2}	in %	Rischio di processo	Rischio di stima
1	1.271.726	552.356	0,434	377.089	403.610
2	1.818.789	605.078	0,333	411.742	443.382
3	1.649.241	492.659	0,299	337.605	358.798
4	2.081.508	548.308	0,263	371.075	403.665
5	2.858.850	667.112	0,233	437.313	503.782
6	3.841.105	817.451	0,213	523.022	628.230
7	6.227.337	1.263.031	0,203	823.162	957.941
8	7.993.236	1.759.305	0,220	1.108.703	1.365.991
9	9.941.329	2.119.905	0,213	1.240.716	1.718.901
10	15.602.114	3.613.854	0,232	1.948.405	3.043.626
11	18.601.479	4.888.498	0,263	2.417.845	4.248.699
12	23.508.407	7.756.496	0,330	3.565.562	6.888.395
13	43.279.371	20.057.283	0,463	8.720.755	18.062.199
Totale	138.674.494	23.976.707	0,173	10.143.050	21.725.584

Anno $t+k$	Flusso stimato	MSEP ^{1/2}	in %	Rischio di processo	Rischio di stima
14	41.494.700	12.195.380	0,294	8.615.011	8.631.853
15	22.217.786	4.807.747	0,216	3.488.760	3.308.019
16	15.384.176	2.927.005	0,190	2.162.096	1.972.993
17	12.291.915	2.261.604	0,184	1.706.399	1.484.269
18	10.318.703	1.991.192	0,193	1.479.219	1.332.950
19	8.489.765	1.719.342	0,203	1.260.756	1.169.030
20	7.375.633	1.706.229	0,231	1.181.449	1.231.014
21	5.324.410	1.212.102	0,228	854.005	860.155
22	4.438.559	1.106.804	0,249	768.936	796.086
23	4.026.515	1.235.500	0,307	857.836	889.144
24	3.119.568	1.143.191	0,366	791.636	824.742
25	2.463.202	1.108.799	0,450	747.348	819.089
26	1.729.562	1.135.562	0,657	700.032	894.123

Stime con il modello gamma

RISCHIO DI MODELLO

Il **rischio di modello** si riferisce alla possibilità che il modello scelto per descrivere i pagamenti incrementali non sia accurato.

Confronti tra valutazioni di valore atteso, RMSEP, errori di previsione, di processo e di stima e margine per il rischio per diversi modelli.

Modello	Descrizione	F.ne di varianza $V(\mu)$	Collegamento $g(\mu)$
<i>ODP</i>	QL, Poisson con sovradisersione	μ	$\log\mu$
<i>Gamma</i>	GLM	μ^2	$\log\mu$
(A)	GLM, Poisson composto	$\mu^p, p=1,01887$	$\log\mu$
(B)	QL	$\mu^p, p=0,80633$	$\log\mu$
(C)	GLM, Poisson composto	$\mu^p, p=1,01311$	$\mu^\gamma, \gamma = -0,2$
(D)	QL	$\mu^p, p=0,85828$	$\mu^\gamma, \gamma = -0,08580$

Modello	$\hat{E}(R)$	$RM\hat{S}EP(R)$	Errore di previsione	Rischio di processo	Rischio di stima
<i>ODP</i>	128.185.17	11.374.39	0,0887	0,0356	0,0813
<i>Gamma</i>	120.246.33	23.191.06	0,1929	0,0871	0,1721
(A)	128.149.20	8.662.225	0,0676	0,0273	0,0618
(B)	128.353.22	10.131.39	0,0789	0,0292	0,0733
(C)	103.068.54	6.996.723	0,0679	0,0303	0,0607
(D)	115.364.56	9.006.560	0,0781	0,0309	0,0717

Confronto tra errori con i diversi modelli

SCELTA DI UN MODELLO IN UNA CLASSE DI MODELLI

Per ridurre il rischio di modello si può cercare di **individuare, in una classe di modelli, un modello che si adatti al meglio i dati**, rispetto a qualche misura di adattamento.

Nell'ambito dei GLM o quasi-GLM, a questo fine, **si può agire su diversi elementi**: la **struttura di regressione**, la **distribuzione delle variabili risposta**, la **funzione di collegamento**, i **pesi**.

Struttura di regressione.

Fissata una famiglia esponenziale lineare o, nei modelli con quasi-verosimiglianza, fissato il legame tra la varianza e la speranza matematica, **il confronto tra modelli** con diverse strutture di regressione può essere effettuato tramite il confronto di **devianze** o di **quasi-devianze**.

Per il confronto tra due modelli, uno annidato nell'altro, si può anche effettuare un **test di verifica d'ipotesi**:

- per GLM, basato sulla **statistica del log-rapporto di verosimiglianza** o sulla **statistica di Wald**,
- per QL, basato sulla **statistica di Wald**.

Si sfrutta il risultato che asintoticamente le statistiche hanno, sotto l'ipotesi nulla, distribuzione chi-quadrato con un numero di gradi di libertà pari alla differenza dei gradi di libertà dei due modelli.

Confronto tra modelli con diverse distribuzioni delle variabili risposta, o diversa funzione di varianza, diverso valore del parametro

Si può considerare la **quasi-verosimiglianza estesa** (Nelder, Pregibon (1987))

$$q^+(\boldsymbol{\mu}, \phi; \mathbf{y}) = -\frac{1}{2} \left\{ \frac{D(\boldsymbol{\mu}; \mathbf{y})}{\phi} + \sum_{i=1}^n \log \left[2\pi \frac{\phi}{\omega_i} V(y_i) \right] \right\},$$

dove $D(\boldsymbol{\mu}; \mathbf{y})$ è la devianza nei GLM e la quasi-devianza $d(\boldsymbol{\mu}; \mathbf{y})$ nei modelli QL. Pertanto

$$q^+(\boldsymbol{\mu}, \phi; \mathbf{y}) = -\frac{1}{2} \sum_{i=1}^n \left[\frac{1}{\phi} d_i + \log \left(2\pi \frac{\phi}{\omega_i} V(y_i) \right) \right],$$

con

$$d_i = -2\omega_i \int_{y_i}^{\mu_i} \frac{y_i - s}{s^5} ds.$$

Obiettivo: Rendere elevata q^+ .

Considerando modelli con funzione di varianza di tipo potenza,

$$V(\mu) = \mu^p,$$

e funzione di collegamento di tipo potenza,

$$g(\mu) = \begin{cases} \mu^\gamma & \gamma \neq 0 \\ \log \mu & \gamma = 0, \end{cases}$$

si possono stimare modelli per diversi valori di p , γ e calcolare i corrispondenti valori della funzione q^+ .

Il modello che massimizza la q^+ è quello che realizza un migliore adattamento ai dati, secondo tale misura di adattamento.