

MODELLI STATISTICI II

Modelli Lineari Generalizzati

Laura Ventura

Dipartimento di Scienze Statistiche
Università degli Studi di Padova
ventura@stat.unipd.it

a.a. 2008–2009

1

Programma del corso

1	Introduzione	7
2	Modelli lineari (LM)	10
2.1	Richiami	12
2.2	Comandi in R	25
2.3	Esempi	26
2.4	Carenze dei LM	32
3	Modelli lineari generalizzati (GLM)	38
3.1	Esempi	39
3.2	Introduzione	49
3.3	La famiglia esponenziale	52
3.4	Momenti	58
3.5	La funzione legame	65
3.6	GLM: Specificazione completa	66
3.7	Specificazione di un GLM in R	77

2

3.8	Funzioni di legame canoniche	78
-----	--	----

4 Inferenza nei GLM **82**

4.1	Stima dei parametri di un GLM	82
4.2	Informazione di Fisher	86
4.3	Legame canonico	87
4.4	Algoritmi iterativi	91
4.5	La stima del parametro ϕ	98
4.6	Risultato di <code>glm()</code> in R	100

5 Adeguatezza dei modelli **113**

5.1	Devianza	115
5.2	Confronto di modelli annidati	125
5.3	Bontà d'adattamento	137
5.4	Residui	143

6 Applicazioni alle tabelle di frequenza **154**

6.1	Introduzione	157
-----	------------------------	-----

3

6.2	Schemi di campionamento	164
6.3	Modelli log-lineari	171
6.4	Modelli per tabelle con dimensione maggiore di due	174
6.5	Stima dei parametri	175
6.6	Specificazione in R	177
6.7	La devianza per un modello Poisson log-lineare	180

7 Quasi-verosimiglianza **203**

7.1	Modello di quasi-verosimiglianza	206
7.2	Inferenza	210
7.3	I comandi in R	212

4

Informazioni

1. Docente: **Laura Ventura**
Dipartimento di Scienze Statistiche,
Via C. Battisti, 241
e-mail: ventura@stat.unipd.it
<http://homes.stat.unipd.it/ventura/>
Ricevimento: Martedì e Giovedì dalle 10.30 alle 12.00.
2. Il materiale didattico è disponibile in rete (dalla pagina della Facoltà o dalla pagina personale).
3. Modalità di esame: Esame Scritto in Aula Asid.

Alcuni testi di riferimento

1. Agresti, A. (2007), *An Introduction to Categorical Data Analysis*, 2nd Edition, J. Wiley, New York.
2. Agresti, A. (2002), *Categorical data analysis*, 2nd Edition, J. Wiley, New York.
3. Azzalini, A. (2001), *Inferenza Statistica: una Presentazione basata sul Concetto di Verosimiglianza*, Springer-Italia, Milano. (Cap. 6)
4. Bortot, P., Ventura, L., Salvan, A. (2000), *Inferenza Statistica: Applicazioni con S-Plus e R*, Cedam, Padova. (Cap. 5)
5. Dobson, A.J. (2002), *An Introduction to Generalized Linear Models*, 2nd Edition, Chapman & Hall, London.
6. McCullagh, P., Nelder, J.A. (1989), *Generalized Linear Models*, Chapman & Hall, London.
7. Pace, L., Salvan, A. (2001), *Introduzione alla Statistica - II. Inferenza, Verosimiglianza, Modelli*, Cedam, Padova. (Cap. 8 e 10)
8. Piccolo, D. (1998), *Statistica*, Il Mulino, Bologna.

1 Introduzione

- **Dati:** In molte applicazioni si raccolgono dati, sulle n unità statistiche, relativi a diverse variabili:

unità nel campione	variabile d'interesse	primo regressore	...	ultimo regressore
1	Y_1	x_{11}	...	x_{1p}
2	Y_2	x_{21}	...	x_{2p}
...
i	Y_i	x_{i1}	...	x_{ip}
...
n	Y_n	x_{n1}	...	x_{np}

- **Modelli Statistici:** Utili nello studio delle relazioni tra misurazioni su più variabili fatte sullo stesso gruppo di unità statistiche.

• Variabili:

- ⇒ risposta o dipendente: Y ;
- ⇒ esplicative o indipendenti: x_1, x_2, \dots, x_p .

• Scala delle variabili:

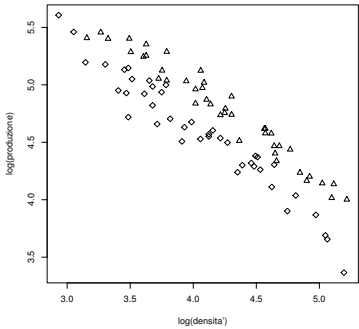
- ⇒ nominale (es. rosso, blu, verde,...);
- ⇒ dicotomica (es. maschio, femmina);
- ⇒ ordinale (es. elementare, media, superiore...);
- ⇒ discreta (es. numero di visite in un anno);
- ⇒ continua (es. peso dopo una dieta).

variabili esplicative	variabile risposta Y	
	binaria	continua
nominali	tabelle di contingenza regressione logistica modelli log-lineari	tabelle di contingenza modelli log-lineari <i>t</i> -test analisi della varianza
continue	regressione logistica	regressione multipla
miste	regressione logistica	analisi della covarianza regressione multipla

2 Modelli lineari (LM)

Un esempio: Cipolle bianche

Produzione, in grammi per pianta, di una specie particolare di cipolle bianche, coltivate in diversi appezzamenti di due località dell’Australia del Sud. Per ogni appezzamento è riportata la densità delle piante per metro quadro (su scala logaritmica).



Domande:

- La produzione (in una particolare località) è in relazione con la densità delle piante?
- Se in una località la densità è pari a 4 piante per metro quadro, quanto si pensa che possa essere la produzione?

Per fortuna i modelli statistici ci permettono di rispondere a queste domande!!
Infatti l’analisi dei dati è rivolta a determinare se e come i diversi valori delle x_j per le varie unità statistiche influenzano la variabile risposta Y .

2.1 Richiami

- I modelli lineari (LM) studiano le relazioni fra una *variabile risposta* Y e una o più *variabili esplicative* x_1, x_2, \dots, x_p ($p \geq 1$).
- La variabile Y è rappresentabile tramite la relazione

$$\begin{aligned} Y_i &= \mu_i + \epsilon_i \\ &= \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i, \end{aligned}$$

con $\mu_i = E(Y_i)$ o, utilizzando la notazione matriciale,

$$Y = X\beta + \epsilon,$$

- dove:
- $X\beta$ è la componente sistematica del modello.
 - ϵ è la componente stocastica del modello.

In particolare, si ha

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix},$$

con:

- $\mapsto \mathbf{Y}$ = vettore $(n \times 1)$ delle risposte;
- $\mapsto X$ = matrice $(n \times p)$ di regressione contenente i valori delle variabili esplicative;
- $\mapsto \boldsymbol{\beta}$ = vettore $(p \times 1)$ dei parametri (coefficienti) di regressione;
- $\mapsto \boldsymbol{\epsilon}$ = vettore $(n \times 1)$ delle componenti della variabile errore.

13

- La variabile Y è sempre una variabile numerica.
- Le variabili esplicative x_j possono essere:
 - \mapsto variabili quantitative (numeriche);
 - \mapsto variabili categoriali (fattori).
- Una variabile esplicativa numerica introduce un solo parametro nel modello.
- Una variabile esplicativa categoriale introduce nel modello un parametro per ciascun livello del fattore.

14

Esempio: Cipolle bianche

La formula di R

`log(produzione) ~ log(densita') + localita'`

in cui `log(densita')` (x) è una variabile numerica e `localita'` (L) è un fattore con due livelli, esprime un modello del tipo

$$Y_i = \beta_1 x_i + \beta_2 I_{(L_i=1)} + \beta_3 I_{(L_i=2)} + \epsilon_i, \quad (1)$$

dove $I_{(L_i=j)}$ è la variabile indicatore che assume valore 1 se $(L_i = j)$, $j = 1, 2$, e 0 altrimenti.

In forma matriciale:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} x_1 & 1 & 0 \\ x_2 & 1 & 0 \\ \vdots & & \vdots \\ x_k & 0 & 1 \\ \vdots & & \vdots \\ x_n & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

15

- Nessun modello ha una parameterizzazione unica. Ad esempio, il modello (1) può essere scritto nella forma:

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 I_{(L_i=2)} + \epsilon_i.$$

Il modello è equivalente, anche se l'interpretazione dei parametri cambia. In questo caso il modello ha la forma matriciale:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 & 0 \\ 1 & x_2 & 0 \\ \vdots & & \vdots \\ 1 & x_k & 1 \\ \vdots & & \vdots \\ 1 & x_n & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

16

- Il modello espresso dalla formula

$$\log(\text{produzione}) \sim \log(\text{densita}') + \text{localita}'$$

è additivo, nel senso che l'effetto di $\log(\text{densita}')$ è lo stesso per ciascun livello di $\text{localita}'$. Però, ci sono situazioni in cui l'effetto può essere diverso. In questo caso è necessario assumere un modello con interazioni tra variabili. Ad esempio, la formula

$$\log(\text{produzione}) \sim \log(\text{densita}') * \text{localita}'$$

esprime il modello di regressione

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 I_{(L_i=2)} + \beta_3 I_{(L_i=2)} x_i + \epsilon_i.$$

Il modello contiene un'intercetta e un coefficiente angolare diversi per ogni valore del fattore $\text{localita}'$.

Inferenza nel LM

- Per l'inferenza nel modello $Y = X\beta + \epsilon$, si assume che:

- X è una matrice non stocastica di rango pieno p ;
- Le componenti di ϵ sono indipendenti a media 0 e varianza comune σ^2 (omoschedasticità), ossia

$$E(\epsilon_i) = 0, \text{var}(\epsilon_i) = \sigma^2, \text{cov}(\epsilon_i, \epsilon_j) = 0;$$

- Per le procedure di inferenza (stima puntuale, verifica di ipotesi o intervalli di confidenza), l'assunzione usuale è che le componenti di ϵ seguono una distribuzione normale, ossia

$$\epsilon_i \sim N(0, \sigma^2).$$

- Per ottenere uno stimatore del parametro β si può ricorrere al *metodo dei minimi quadrati* (o SMV):

1. Si sceglie il valore $\hat{\beta}$ di β che minimizza

$$\sum_{i=1}^n (Y_i - \mu_i)^2 = (Y - X\beta)^T (Y - X\beta);$$

2. Si trova

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

- È possibile mostrare che lo stimatore ai minimi quadrati soddisfa le seguenti proprietà:

1. $E(\hat{\beta}) = \beta$ e $\text{var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$;
2. $\hat{\beta} \sim N_p(\beta, \sigma^2 (X^T X)^{-1})$ (o esatta);
3. $\hat{\beta}$ è lo stimatore lineare non distorto con varianza minima (Gauss-Markov).

- La varianza σ^2 viene stimata con la *varianza residua corretta*

$$S^2 = \frac{SQE}{n - p},$$

dove

$$\begin{aligned} SQE &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \\ &= (Y - \hat{Y})^T (Y - \hat{Y}) \\ &= (Y - X\hat{\beta})^T (Y - X\hat{\beta}) \end{aligned}$$

è la *somma dei quadrati dei residui*.

- Sotto ipotesi di normalità segue che

$$\hat{\beta} \sim N_p(\beta, \sigma^2 (X^T X)^{-1}) \quad \text{e} \quad \frac{(n-p)S^2}{\sigma^2} \sim \chi_{n-p}^2,$$

con $\hat{\beta}$ e S^2 stocasticamente indipendenti. Questo è un risultato molto importante per l'inferenza (stima intervallare e verifica di ipotesi) sui parametri del modello di regressione.

Scomposizione di somme di quadrati

- È utile ricordare la scomposizione della devianza empirica dei valori Y in:

$$SQT = SQR + SQE ; \tag{2}$$

ossia, la somma dei quadrati totale (*devianza totale*) in somma dei quadrati di regressione (*devianza spiegata dal modello*) e somma dei quadrati residua (*devianza dei residui*). Lo studio delle componenti della (2) è fondamentale nei LM, in quanto il confronto tra SQR e SQT è un indicatore della bontà di adattamento del modello.

- Un indicatore normalizzato - con valori in $[0, 1]$ - di bontà di adattamento del modello ai dati è il coefficiente di determinazione del modello

$$R^2 = 1 - SQE/SQT .$$

- Sia \mathcal{F}_1 il modello minimale con la sola intercetta ($p = 1$). Sia \mathcal{F}_p il modello corrente a p parametri e sia \mathcal{F}_{p_0} un modello ridotto, con $1 < p_0 < p$. Allora, la varianza spiegata dal modello corrente \mathcal{F}_p può essere suddivisa per le componenti del modello stesso, come si mostra nella Tabella 1 sul **prospetto di analisi della varianza (anova)**.

Fonte di variabilità	g.l.	SQ	test su miglioramento
totale	n	SQT	
costante	1	$n\bar{Y}^2$	
miglioramento con \mathcal{F}_{p_0} rispetto a \mathcal{F}_1	$p_0 - 1$	SQR_{p_0}	$\frac{SQR_{p_0}/(p_0 - 1)}{SQE_{p_0}/(n - p_0)}$ $\sim F_{p_0 - 1, n - p_0}$
miglioramento con \mathcal{F}_p rispetto a \mathcal{F}_{p_0}	$p - p_0$	$SQR_p - SQR_{p_0}$	$\frac{(SQR_p - SQR_{p_0})/(p - p_0)}{SQE_p/(n - p)}$ $\sim F_{p - p_0, n - p}$
residui di \mathcal{F}_p	$n - p$	SQE_p	

- La perdita di bontà di adattamento del modello \mathcal{F}_{p_0} rispetto a \mathcal{F}_p può essere valutata attraverso la statistica

$$F = \frac{(SQE_{p_0} - SQE_p)/(p - p_0)}{SQE_p/(n - p)} \sim F_{p - p_0, n - p} .$$

- È un test di nullità di un sottoinsieme di coefficienti di regressione (confronto tra modelli). Per $1 < p_0 < p$, l’ipotesi nulla di semplificazione del modello è

$$H_0 : \beta_{p_0 + 1} = \beta_{p_0 + 2} = \dots = \beta_p = 0 .$$

Analisi dei residui

- Per una verifica della bontà di adattamento di un modello è necessario completare lo studio effettuando un controllo empirico del modello tramite un’analisi dei residui.
- Nei LM i residui standardizzati sono definiti come

$$\frac{y_i - \hat{y}_i}{\sqrt{s^2 h_{ii}}} , \tag{3}$$

dove h_{ii} è l’ i -esimo elemento sulla diagonale di $I_n - X(X^T X)^{-1} X^T$.

- Strumenti grafici: diagramma dei residui, grafico dei residui rispetto alle variabili esplicative, grafico dei residui rispetto ai valori stimati dal modello;
- Verifica dell’ipotesi di normalità: diagramma Q-Q normale dei residui.

2.2 Comandi in R

In R si usano le funzioni:

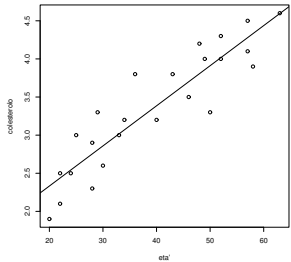
- `lm(formula)` per adattare un modello di regressione;
- `summary` per ottenere il risultato convenzionale di una analisi di regressione;
- `anova` per effettuare un'analisi della varianza;
- `residuals` per ottenere i residui del modello;
- `predict` per ottenere valori previsti dal modello per nuovi dati;
- `fitted` per ottenere i valori stimati con il modello di regressione;
- `plot` per ottenere grafici collegati al modello di regressione.

2.3 Esempi

A. Dati sul colesterolo (lipid)

Relazione fra colesterolo ed età, basato su un campione di $n = 24$ pazienti.

```
cor(colesterolo,eta') = 0.907
plot(colesterolo,eta')
```



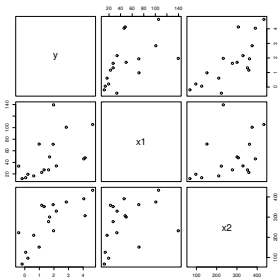
```
lm(colesterolo ~ eta')
mu-hat = 1.28 + 0.05 x eta'
```

B. Dati sul Tamigi (thames)

Valori dell'ossigeno (y) contro i valori di clorofilla (x_1) e la lucidità (x_2) in diverse locazioni sul Tamigi.

```
> cor(thames)
      y      x1      x2
y    1.0000000  0.5442528  0.7703502
x1   0.5442528  1.0000000  0.3916644
x2   0.7703502  0.3916644  1.0000000

> pairs(thames)
```



```
> summary(lm(y ~ x1+x2))

Call:
lm(formula = y ~ x1 + x2)

Residuals:
    Min       1Q   Median       3Q      Max
-1.4982 -0.4457  0.0969  0.3940  2.1206

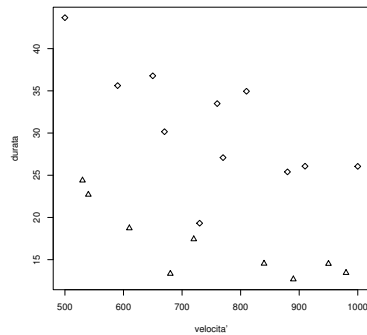
Coefficients:
            Estimate Std. Error t val Pr(>|t|)
(Intercept) -1.338418  0.628919  -2.13  0.05157
x1           0.011776  0.006932   1.699  0.11147
x2           0.009077  0.002326   3.903  0.00159

Residual standard error: 0.9419 on 14 degrees of freedom
Multiple R-Squared: 0.6629, Adjusted R-squared: 0.6148
F-statistic: 13.77 on 2 and 14 DF, p-value: 0.0004945
```

Dati su macchine industriali (lathe)

Durata di vita (Y) di due tipi, A e B, di macchine industriali rispetto alla velocità di operazione (x).

```
> cor(speed, lifetime)
-0.4313059
> cor(speed[1:9], lifetime[1:9])
-0.838278
> cor(speed[10:20], lifetime[10:20])
-0.6737355
```



29

```
> summary(lm(lifetime~speed*make))

Call:
lm(formula = lifetime ~ speed * make)

Residuals:
    Min       1Q   Median       3Q      Max
-5.1750 -1.4999  0.4849  1.7830  4.8652

Coefficients:
              Estimate Std. Error t val Pr(>|t|)
(Intercept)  32.774760  4.633472  7.073 2.63e-06
speed        -0.020970  0.006074  -3.452 0.00328
makeB        23.970593  6.768973  3.541 0.00272
speed:makeB  -0.011944  0.008842  -1.351 0.19553

Residual standard error: 2.968 on 16 degrees of freedom
Multiple R-Squared: 0.9105,    Adjusted R-squared: 0.8937
F-statistic: 54.25 on 3 and 16 DF,  p-value: 1.319e-08
```

```
> summary(lm(lifetime~speed+make))

Call:
lm(formula = lifetime ~ speed + make)

Residuals:
    Min       1Q   Median       3Q      Max
-5.552684 -1.786752 -0.001612  1.839470  4.983763

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  36.98560    3.51038   10.536 7.16e-09 ***
speed        -0.02661    0.00452   -5.887 1.79e-05 ***
makeB        15.00425    1.35967   11.035 3.59e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.039 on 17 degrees of freedom
Multiple R-Squared: 0.9003,    Adjusted R-squared: 0.8886
F-statistic: 76.75 on 2 and 17 DF,  p-value: 3.086e-09
```

30

```
> anova(lm(lifetime~speed+make), lm(lifetime~speed*make))
```

Analysis of Variance Table

```
Model 1: lifetime ~ speed + make
Model 2: lifetime ~ speed * make
  Res.Df  RSS Df Sum of Sq  F Pr(>F)
1     17 157.055
2     16 140.976  1    16.079 1.8248 0.1955
```

31

2.4 Carenze dei LM

In molte situazioni non è possibile ricondurre la relazione tra le variabili alla forma prevista dal LM, ossia

$$Y_i = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i,$$

$i = 1, \dots, n$. Infatti:

(1) Può succedere che la relazione sia del tipo

$$Y_i = h(x_{i1}, \dots, x_{ip}) + \epsilon_i,$$

con $h(\cdot)$ non lineare nei parametri.

↪ Applicazioni in ambito econometrico, fisico, medico, botanico, ... Ad esempio:

$$Y_i = h(x_i, \beta) + \epsilon_i, \text{ con}$$

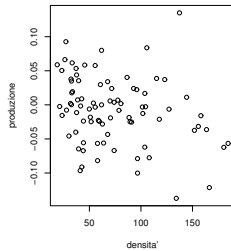
Y_i = rendimento (produzione) per pianta

x_i = densità delle piante

$$h(x_i, \beta) = \frac{1}{\beta_0 + \beta_1 x_i + \beta_2 x_i^2}$$

$$\epsilon_i \sim N(0, \sigma^2) \quad \text{indipendenti}$$

32



⇒ Inferenza simile ai LM
 ⇒ In R esiste il pacchetto `nls`

33

(2) Il campo di variazione di Y è noto e può non essere compatibile con quello dei LM normali. Quando Y è discreta, o dicotomica, o nominale, o positiva. Ad esempio:

Y = numero di strumenti ad alta tecnologia posseduti dallo studente (PC, videocamera, cellulare, antenna parabolica,...)

Il dominio della Y è $\{0, 1, 2, \dots\}$

Obiettivo è modellare la Y in funzione di alcune variabili esplicative sugli studenti:

x_1 = età

x_2 = attività del capofamiglia

x_3 = sesso

x_4 = reddito dello studente

⇒ Il dominio della Y è noto e non è compatibile con quello dei LM normali.

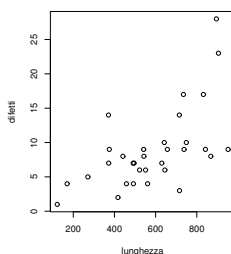
34

(3) La varianza del termine di errore, e quindi anche della Y , può non essere costante. L'evidenza empirica talvolta evidenzia un legame tra varianza e valore medio.

• Esempio:

Y = numero di difetti per rotolo di stoffa

x = lunghezza (in mt) del rotolo di stoffa



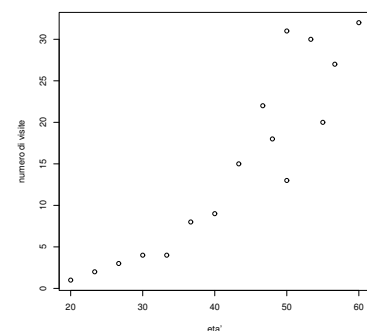
⇒ Il valor medio dei difetti cresce con la lunghezza, ma anche la variabilità del numero di difetti aumenta con la lunghezza (e quindi con il suo stesso valore medio).

35

(4) Spesso si ha a che fare con distribuzioni di Y molto lontane dalla normalità. Talvolta, è più realistico modellare la Y con distribuzioni discrete (per esempio, con dati binari o conteggi).

Esempio

Numero di visite mediche (Y) in un anno in funzione dell'età (x).



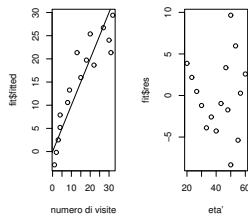
36

```
> fit <- lm(numero ~ eta)
> summary(fit)

Residuals:
    Min       1Q   Median       3Q      Max
-8.3435 -2.9102 -0.3435  2.7761  9.6565

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -19.00932  4.25990   -4.462  0.000536 ***
eta          0.80706   0.09733    8.292  9e-07 ***

Residual standard error: 4.711 on 14 degrees of freedom
Multiple R-Squared: 0.8308,    Adjusted R-squared: 0.8188
F-statistic: 68.76 on 1 and 14 DF,  p-value: 8.998e-07
```



$$\hat{\mu} = -19 + 0.80 \times \text{eta}$$

$$\hat{\mu}(20) = -19 + 0.80 \times 20 = 6$$

$$\hat{\mu}(10) = -19 + 0.80 \times 10 = -2$$

37

3 Modelli lineari generalizzati (GLM)

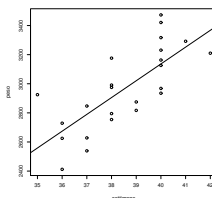
- Ampia classe di modelli di regressione con l'obiettivo di estendere il LM classico con errori normali, per spiegare la relazione fra una variabile risposta e una o più variabili esplicative.
- Classe di modelli che include quelli lineari e consente di trattare con funzioni $h(\cdot)$ non lineari e variabili non normali.
- Struttura - sia di costruzione che di analisi - simile a quella dei LM.
- Le applicazioni includono: regressione binomiale, regressione di Poisson, tabelle di frequenza, analisi dei dati di sopravvivenza.
- Funzioni per adattare modelli della forma GLM esistono in R (ed altri pacchetti) $\Rightarrow \text{glm}()$.
- L'inferenza sui parametri di regressione è basata sulla funzione di verosimiglianza.

38

3.1 Esempi

(A) *Situazione tipo: Regressione lineare.* Il termine *regressione* deriva dalla famosa applicazione compiuta nel 1886 dal biologo e statistico Galton, che esaminò le altezze dei figli (Y) in funzione delle altezze dei genitori (x). In realtà la prima applicazione di un modello simile avvenne nel 1877 e riguardava lo studio delle altezze di due successive piante di piselli (Darwin era cugino di Galton).

Esempio: Peso alla nascita (birthweight)



Y_i = peso alla nascita

x_i = settimane di gestazione per bimbo i

$\Rightarrow Y_i \sim N(\mu_i, \sigma^2)$, con $\mu_i = \beta_0 + \beta_1 x_i$.

La regressione lineare stimata è

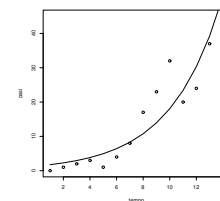
$\Rightarrow \hat{\mu} = -1484.0 + 115.5x$.

39

(B) *Situazione tipo: Regressione di Poisson.* Numero di "eventi" per uno specifico fenomeno che si verificano in un preciso intervallo di tempo (o un'area di opportunità, come una lunghezza o superficie, . . .). La variabile di interesse è una variabile conteggio che può assumere solo valori interi non negativi.

Esempio: AIDS

Casi di mortalità a causa dell'AIDS in Australia per periodi di tre mesi fra il 1983 e il 1986.



La Y è *discreta* e assume solo valori non negativi. Dal grafico si nota che il valore medio di Y cresce con il tempo t . Il grafico mostra anche che la variabilità di Y aumenta con t , e quindi con il suo stesso valor medio.

40

Trattandosi di un conteggio, l'utilizzo di una v.c. di Poisson è preferibile a quello di una v.c. normale.

- 1. Un modello possibile per Y_i (numero di decessi nel periodo i) è la distribuzione di Poisson:
 $Y_i \sim \text{Poisson}(\mu_i)$.
- 2. Un modello lineare per spiegare μ_i in funzione del tempo non è appropriato.

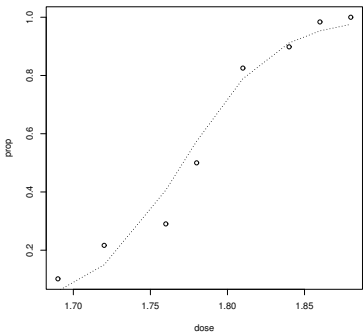
(C) Situazione tipo: Regressione binomiale. Numero di "successi" rispetto al totale di prove effettuate. Ad esempio, il numero di cavie morte rispetto al totale di cavie sottoposte ad una specifica sperimentazione chimica. La variabile di interesse è una proporzione e può assumere solo valori in $[0, 1]$.

Esempio: Dati da una prova clinica (clinicaltrial)

Esperimento per stimare l'effetto di un nuovo farmaco. La variabile Z_i indica il numero di pazienti che rispondono positivamente alla dose x_i del nuovo farmaco.

Dose x_i	1.69	1.72	1.76	1.78	1.81	1.84	1.86	1.88
Numero di pazienti m_i	59	60	62	56	63	59	62	60
Numero positivo z_i	6	13	18	28	52	53	61	60

Il grafico mostra la proporzione di risposte positive, $Y_i = Z_i/m_i$, per ogni dose.



Per questi dati sembra naturale un modello con

$$Z_i \sim \text{Bin}(m_i, \mu_i)$$

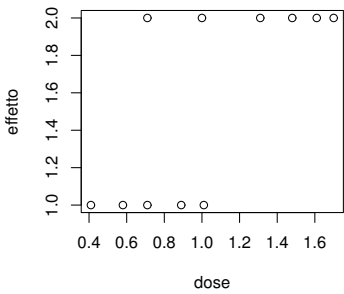
e $\mu_i = h(x_i)$.

Si lavora con $Y_i = Z_i/m_i \in [0, 1]$, ove $\mu_i = E(Y_i)$.

(D) Situazione tipo: Regressione binomiale ($m_i = 1$). In uno studio di natura medico-farmaceutica può interessare appurare se dopo il trattamento con un dato farmaco il "paziente" guarisce o meno. La variabile risposta di tipo "successo-insuccesso" è dicotomica.

Esempio: Dati su topi (mice)

x_i = dose di un certo veleno
 $Y_i = 1$ per una risposta positiva, altrimenti $Y_i = 0$.



Per questi dati sembra naturale un modello con:

1. $Y_i \sim \text{Bernoulli}(\mu_i)$, con $\mu_i = \text{Pr}(Y_i = 1) = E(Y_i)$.
2. Modello non-lineare per $h(\cdot)$, con $\mu_i = h(x_i)$.

Questo è un altro caso particolare di regressione binomiale per $m_i = 1$.

NOTA

- I modelli per dati binari sono quelli più frequentemente adoperati nell'ambito dei GLM.
- La Y è dicotomica: presente/assente, successo/insuccesso, funzionante/guasto, vivo/morto, ...

• Contesti:

- *Medicina*: il paziente guarisce/o no dopo la cura;
- *Biologia*: la cellula reagisce/o no dopo il trattamento;
- *Botanica*: il seme germoglia/o no dopo essere stato fertilizzato;
- *Economia*: l'azienda fallisce/o no dopo aver ricevuto una sovvenzione; il disoccupato trova lavoro/o no dopo aver partecipato a un corso di formazione;
- *Scienze Sociali*: l'intervistato è favorevole/o no ad una data proposta;
- *Marketing*: il consumatore compra/o no un prodotto dopo una campagna pubblicitaria;
- *Calcio*: una squadra vince una partita/o no dopo un infortunio di un giocatore.

(E) Situazione tipo: Tabelle di frequenza. La variabile osservata sono le frequenze nelle celle di una tabella di contingenza.

Esempio: Melanoma maligno (skin)

Tabella di frequenza:

Tipo di tumore	Testa/collo	Tronco	Estremità	Totale
Hutchinson	22	2	10	34
Superficiale	16	54	115	185
Nodulare	19	33	73	125
Indeterminante	11	17	28	56
Totale	68	106	226	400

Chiaramente LM normali non sono adeguati. Vedremo che una forma di regressione di Poisson funziona bene.

ALTRI MODELLI

• Modelli per dati di sopravvivenza

La Y è la durata di sopravvivenza a partire da un certo tempo (es. resistenza di un certo macchinario, durata della sopravvivenza dopo un intervento, ...) in funzione di x_1, \dots, x_p variabili esplicative \Rightarrow per questi dati sembra naturale un modello con

$$h(y, x_1, \dots, x_p; \beta)$$

e con Y_i con distribuzione esponenziale, gamma, weibull...

3.2 Introduzione

- I GLM sono abbastanza flessibili per permettere la modellazione di tutti gli esempi precedenti (... e tanti altri ...).

- Anche se la classe è molto più ampia, scopriremo che le idee della:

- formulazione del modello;
- stima dei parametri;
- test ANOVA, ecc.

sono simili al caso del LM.

49

- Innanzitutto, riconsideriamo gli elementi che definiscono il LM normale. Il LM classico

$$Y = X\beta + \varepsilon$$

è basato sulle tre componenti:

1. **Componente casuale:** $Y_i \sim N(\mu_i, \sigma^2)$, indipendenti, con $\mu_i = \mathbf{x}_i^T \beta$ e \mathbf{x}_i^T i -esima riga di X , $i = 1, 2, \dots, n$;
2. **Componente sistematica (predittore lineare):** $\eta_i = \sum_{j=1}^p x_{ij} \beta_j = \mathbf{x}_i^T \beta$;
3. **Legame** tra valor medio e predittore lineare: $\mu_i = \eta_i$.

- I parametri ignoti sono β e σ^2 . Nei LM questi appartengono a parti separate del modello.

50

- La classe dei GLM si ottiene generalizzando le componenti 1. e 3., mantenendo l'ipotesi di indipendenza. In particolare:

- Consideriamo come distribuzione possibile per Y_i non solo la normale, ma qualunque altra distribuzione che appartiene alla *famiglia esponenziale*, mantenendo l'ipotesi di indipendenza;
- Consideriamo altre forme di legame tra predittore lineare $\eta_i = \mathbf{x}_i^T \beta$ e valor medio μ_i del tipo:

$$g(E(Y_i)) = g(\mu_i) = \eta_i = \mathbf{x}_i^T \beta \quad (4) \\ \Rightarrow \mu_i = g^{-1}(\mathbf{x}_i^T \beta),$$

con $g(\cdot)$ funzione monotona e derivabile (detta *funzione di legame*, da *link function*).

- Mantenere invariata la 2. vuole dire che possiamo conservare tutte le idee per la costruzione dei modelli dalla teoria dei LM.

- La (4) è particolarmente conveniente qualora il dominio della Y non coincide con \mathbb{R} .

51

3.3 La famiglia esponenziale

- Introduciamo una famiglia da usare nella 1. che comprenda non solo la distribuzione normale, ma anche altre distribuzioni: Poisson, binomiale, esponenziale, gamma, ecc.

- È una classe di modelli privilegiata in quanto:

- costituisce un serbatoio di modelli utili per le applicazioni;
- le procedure inferenziali basate sulla verosimiglianza sono generali, semplici e accurate;
- l'algoritmo numerico di stima è lo stesso per tutti i modelli della classe.

52

- Si dice che la variabile Y appartiene alla famiglia esponenziale se ha funzione di densità (o probabilità) del tipo

$$f(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \right\}, \quad (5)$$

dove θ e ϕ sono parametri scalari ignoti, $b(\cdot)$ e $c(\cdot)$ sono funzioni note la cui scelta individua una particolare distribuzione, e il dominio di Y non dipende da θ o ϕ . Scriveremo $Y \sim EF(b(\theta), \phi)$.

- θ è il *parametro naturale* della famiglia esponenziale.
- ϕ è un parametro di dispersione o di scala (in molti casi è fissato).

Esempio: Distribuzione di Poisson

- Importante per descrivere fenomeni di tipo "conteggio".
- Se $Y \sim \text{Poisson}(\lambda)$, la sua funzione di probabilità è

$$\begin{aligned} f(y; \lambda) &= \frac{e^{-\lambda} \lambda^y}{y!} \\ &= \exp\{y \log \lambda - \lambda - \log y!\}, \end{aligned}$$

per $y = 0, 1, \dots$

- Questa ha la forma richiesta nella (5) con $\theta = \log \lambda$ parametro naturale, $\phi = 1$, $b(\theta) = \lambda = e^\theta$ e $c(y, \phi) = -\log y!$.
- Scriveremo $Y \sim EF(e^\theta, 1)$.

Esempio: Distribuzione Binomiale

- Importante per descrivere risultati di "prove" (successi su prove indipendenti).
- Sia $Z \sim \text{Bin}(m, \pi)$; se consideriamo $Y = Z/m$, la sua funzione di probabilità è

$$\begin{aligned} f_Y(y; \pi) &= f_Z(my; \pi) = \binom{m}{my} \pi^{my} (1 - \pi)^{m - my} \\ &= \exp\{\log \binom{m}{my} + my \log \pi + (m - my) \log(1 - \pi)\} \\ &= \exp \left[m \left\{ y \log \frac{\pi}{1 - \pi} + \log(1 - \pi) \right\} + \log \binom{m}{my} \right], \end{aligned}$$

per $y = 0, 1/m, 2/m, \dots, 1$.

- Questa ha la forma richiesta nella (5) con $\theta = \log \frac{\pi}{1 - \pi}$ parametro naturale, $\phi = 1/m$,

$$b(\theta) = -\log(1 - \pi) \Big|_{\pi = \frac{e^\theta}{1 + e^\theta}} = \log(1 + e^\theta)$$

$$\text{e } c(y, \phi) = \log \binom{m}{my} = \log \binom{1/\phi}{y/\phi}.$$

- Scriveremo $Y \sim EF(\log(1 + e^\theta), 1/m)$.

Esempio: Distribuzione Gamma

- Importante per descrivere risposte positive (es. durate).

- Se $Y \sim \text{Gamma}(\nu, \lambda)$, con $\nu > 0$ fissato, la sua funzione di densità è

$$\begin{aligned} f(y; \nu, \lambda) &= \frac{y^{\nu-1} \exp\{-y/\lambda\}}{\lambda^\nu \Gamma(\nu)} \\ &= \exp\{(\nu - 1) \log y - y/\lambda - \nu \log \lambda - \log \Gamma(\nu)\} \\ &= \exp\{-y/\lambda - \nu \log \lambda + (\nu - 1) \log y - \log \Gamma(\nu)\} \end{aligned}$$

per $y > 0$.

- Questa ha la forma richiesta nella (5) con $\theta = -1/\lambda$ parametro naturale, $\phi = 1$,

$$b(\theta) = \nu \log \lambda \Big|_{\lambda = -1/\theta} = -\nu \log(-\theta)$$

$$\text{e } c(y, \phi) = (\nu - 1) \log y - \log \Gamma(\nu).$$

- Scriveremo $Y \sim EF(-\nu \log(-\theta), 1)$.

Esempio: Distribuzione Normale

- Caso del LM.

- Se $Y \sim N(\mu, \sigma^2)$, la sua funzione di densità è

$$\begin{aligned} f(y; \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y - \mu)^2\right\} \\ &= \exp\left\{-\frac{y^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right\} \\ &= \exp\left\{\frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right\}. \end{aligned}$$

- Questa ha la forma richiesta nella (5) con $\theta = \mu$ parametro naturale, $\phi = \sigma^2$,

$$b(\theta) = \mu^2/2|_{\mu=\theta} = \theta^2/2$$

$$\text{e } c(y, \phi) = -\frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2).$$

- Scriveremo $Y \sim EF(\theta^2/2, \sigma^2)$.

57

3.4 Momenti

- Le funzioni $b(\cdot)$ e $c(\cdot)$ sono importanti nella valutazione e interpretazione dei momenti della distribuzione.

- Sia $Y \sim f(y; \theta)$, con $E(Y) = \int yf(y; \theta)dy$ e $\text{var}(Y) = E(Y^2) - E(Y)^2$.

- Ricordiamo due risultati di base sulle derivate della funzione di log-verosimiglianza.

- Siano $L(\theta) = L(\theta; y)$ la verosimiglianza, e $\ell(\theta) = \ell(\theta; y) = \log L(\theta) = \log f(y; \theta)$ la log-verosimiglianza per θ . Allora,

$$E(\ell'_*(\theta)) = E\left(\frac{d}{d\theta}\ell(\theta; Y)\right) = 0$$

e

$$i(\theta) = \text{var}(\ell'_*(\theta)) = E[(\ell'_*(\theta))^2]$$

$$= E(-\ell''_{**}(\theta)) = E\left(-\frac{d^2}{d\theta^2}\ell(\theta; Y)\right)$$

sotto opportune condizioni di regolarità.

58

- Se Y è una realizzazione di una distribuzione appartenente alla famiglia esponenziale, la log-verosimiglianza per θ risulta:

$$\ell(\theta) = \log f(y; \theta) = \frac{y\theta - b(\theta)}{\phi} + c(y, \phi).$$

Allora, per ϕ fissato, si ha:

$$\ell'_*(\theta) = \frac{d\ell}{d\theta} = \frac{y - b'(\theta)}{\phi} \quad (\Rightarrow \hat{\theta} : y = b'(\theta))$$

$$\ell''_{**}(\theta) = \frac{d^2\ell}{d\theta^2} = \frac{-b''(\theta)}{\phi}.$$

Da queste espressioni segue che

$$\boxed{E(Y) = \mu = b'(\theta)}$$

in quanto

$$E(Y/\phi) - E(b'(\theta)/\phi) = 0$$

$$\Rightarrow E(Y) - b'(\theta) = 0$$

$$\Rightarrow E(Y) = \mu = b'(\theta).$$

59

Inoltre segue che

$$\text{var}\left(\frac{Y - b'(\theta)}{\phi}\right) = \frac{b''(\theta)}{\phi} \Rightarrow$$

$$\boxed{\text{var}(Y) = \phi b''(\theta)}$$

in quanto

$$\text{var}\left(\frac{Y - b'(\theta)}{\phi}\right) = \frac{\text{var}(Y)}{\phi^2}$$

e

$$E(-\ell''_{**}(\theta)) = E\left(\frac{b''(\theta)}{\phi}\right) = \frac{b''(\theta)}{\phi}$$

$$\Rightarrow \frac{\text{var}(Y)}{\phi^2} = \frac{b''(\theta)}{\phi}$$

Di solito, si pone $V(\mu) = b''(\theta)$, quindi

$$\boxed{\text{var}(Y) = \phi V(\mu)}.$$

60

• La funzione $V(\mu)$ è detta *funzione di varianza* ed è intesa come funzione del valor medio μ anche se appare scritta come funzione di θ (basta invertire la relazione tra μ e θ data da $\mu = b'(\theta)$).

• Poichè la varianza della variabile risposta è legata al valor medio μ tramite una funzione di varianza $V(\mu)$, nei GLM si permettono specifiche forme di eteroschedasticità.

• FANTASTICO!!! Diventa assai semplice calcolare i momenti (media e varianza) di una v.c. con distribuzione appartenente alla famiglia esponenziale.

Esempio: Distribuzione di Poisson

Abbiamo $Y \sim EF(e^\theta, 1)$. Quindi

$$b(\theta) = e^\theta \quad \text{e} \quad \phi = 1 \, .$$

Di conseguenza

$$\Rightarrow E(Y) = \mu = b'(\theta) = e^\theta = \lambda$$

e

$$\Rightarrow var(Y) = b''(\theta) = e^\theta = \lambda \, .$$

Allora, $V(\mu) = \mu$. In altre parole, all’aumentare del valore atteso di Y , aumenta anche la sua varianza.

Esempio: Distribuzione Binomiale

Abbiamo $Y \sim EF(\log(1 + e^\theta), 1/m)$. Allora

$$b(\theta) = \log(1 + e^\theta) \quad \text{e} \quad \phi = 1/m \, .$$

Di conseguenza

$$E(Y) = \mu = b'(\theta) = \frac{e^\theta}{1 + e^\theta} = \pi$$

e

$$var(Y) = \phi b''(\theta) = \frac{1}{m} \frac{e^\theta}{(1 + e^\theta)^2} = \frac{1}{m} \pi(1 - \pi) \, .$$

Allora, $V(\mu) = \mu(1 - \mu)$.

• Altre importanti distribuzioni del tipo $EF(b(\theta), \phi)$ sono indicate nella Tabella 2 che segue, assieme agli altri elementi caratteristici.

Distribuzione	Tabella 2			
	Normale $N(\mu, \sigma^2)$	Poisson $Po(\mu)$	Binomiale / m $Bin(m, \mu)/m$	Gamma $Ga(\omega, \omega/\mu)$
Supporto	$(-\infty, \infty)$	$\{0, 1, 2, \dots\}$	$\{0, 1/m, 2/m, \dots, 1\}$	$(0, \infty)$
ϕ	σ^2	1	$\frac{m-1}{m}$	ω^{-1}
$b(\theta)$	$\theta^2/2$	$\exp(\theta)$	$\log(1 + e^\theta)$	$-\log(-\theta)$
$c(y, \phi)$	$-\left(\frac{y^2}{2\phi} + \frac{\log(2\pi\phi)}{2}\right)$	$-\log y!$	$\log\left(\frac{m}{my}\right)$	$\omega \log(\omega y) - \log y$
$\mu(\theta)$	θ	$\exp(\theta)$	$e^\theta/(1 + e^\theta)$	$-\log \Gamma(\omega)$
$V(\mu)$	1	μ	$\mu(1 - \mu)$	$-1/\theta$
legame canonico	identità	logaritmo	logit	μ^2 reciproco

3.5 La funzione legame

• La specificazione di un GLM viene completata attraverso la scelta della relazione fra il valor medio $\mu_i = E(Y_i)$ e le variabili esplicative (predittore lineare $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$). In particolare, assumiamo che il legame fra μ_i , la media di Y_i , e \mathbf{x}_i , il vettore di covariate, abbia la forma

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta} ,$$

⇒

$$\mu_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta}) ,$$

con $g(\cdot)$ funzione nota monotona e derivabile. La funzione $g(\cdot)$ è detta *funzione legame*, o *legame*, tra μ_i e η_i .

• In questa maniera l’effetto delle covariate è lineare, però il legame fra μ_i e η_i è in genere non-lineare.

In forma schematica:

componente casuale	predittore lineare	legame
$Y_i \sim EF(b(\theta_i), \phi)$ con $b'(\theta_i) = \mu_i$	$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$	$g(\mu_i) = \eta_i$

• Riassumendo, le quantità:

- ↦ *parametro naturale* θ_i
- ↦ *media* μ_i
- ↦ *predittore lineare* η_i

sono legate tra loro attraverso le relazioni:

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

e

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

3.6 GLM: Specificazione completa

- Siano Y_i le osservazioni e \mathbf{x}_i le variabili esplicative, $i = 1, 2, \dots, n$. Un GLM è caratterizzato dalle seguenti componenti:
 - 1. **Componente casuale:** $Y_i \sim EF(b(\theta_i), \phi)$, indipendenti, con $E(Y_i) = \mu_i = b'(\theta_i)$, $i = 1, 2, \dots, n$;
 - 2. **Componente sistematica (predittore lineare):** $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$, ove \mathbf{x}_i è un vettore di costanti e $\boldsymbol{\beta}$ un vettore di parametri;
 - 3. **Legame:** esiste una funzione legame $g(\cdot)$ tale per cui $g(\mu_i) = \eta_i \Leftrightarrow \mu_i = g^{-1}(\eta_i)$, $i = 1, 2, \dots, n$.
- I parametri ignoti sono $\boldsymbol{\beta}$, e a volte ϕ .
 - A differenza del LM tradizionale, nel caso dei GLM la precisa separazione della variabile risposta Y_i in due componenti, sistematica e casuale, non è più possibile.

Esempio: Regressione normale

Abbiamo

$$Y_i \sim N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2) , \quad i = 1, 2, \dots, n,$$

indipendenti.

In questo caso: $\theta_i = \eta_i = \mu_i$ e $g(\cdot)$ è la funzione identità.

Nel caso di normalità è equivalente scrivere $Y_i \sim N(\mu_i, \sigma^2)$ oppure $Y_i = \mu_i + \epsilon_i$, con $\epsilon_i \sim N(0, \sigma^2)$. In questo caso abbiamo una separazione completa tra componente sistematica μ_i (che dipende da \mathbf{x}_i e $\boldsymbol{\beta}$) e componente casuale ϵ_i (che dipende solo da σ^2).

Esempio: Regressione di Poisson

Abbiamo

$$Y_i \sim \text{Poisson}(\mu_i), \quad i = 1, 2, \dots, n,$$

indipendenti.

Si cerca una qualche funzione di legame $g(\cdot)$ per cui valga la relazione $g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$. Ad esempio, potrebbe essere la funzione logaritmo, nel qual caso la positività di $\mu_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}}$ è assicurata.

La scelta:

$$\mu_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}} = e^{\eta_i},$$

ossia $\eta_i = \log \mu_i$ è detta *regressione poissoniana* e il modello ha legame logaritmico, con $\theta_i = \eta_i (= \log \mu_i)$.

69

Esempio: Regressione di Poisson

Si desidera studiare la relazione tra numero di visite mediche (Y) ed età (x). La variabile risposta è un conteggio, che assume solo valori non negativi. Un modello log-lineare esprime un modello lineare per il logaritmo di μ_i , del tipo

$$\log \mu_i = \beta_0 + \beta_1 x_i.$$

Se in particolare si ottengono i valori stimati $\hat{\beta}_0 = -0.49$ e $\hat{\beta}_1 = 0.06$, ciò significa che

$$\hat{\mu}_i = \exp(-0.49 + 0.06x_i).$$

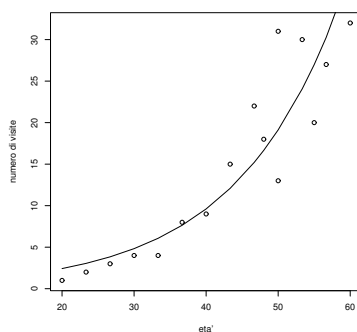
70

Per un (una) paziente di 20 anni, il numero medio stimato di visite in un anno è:

$$\hat{\mu} = \exp(-0.49 + 0.06 \times 20) \cong 2.$$

Per un (una) paziente di 10 anni, il numero medio stimato di visite in un anno è:

$$\hat{\mu} = \exp(-0.49 + 0.06 \times 10) \cong 1.1.$$



71

Esempio: Regressione binomiale

- Abbiamo visto i legami naturali per il modello normale ed il modello di Poisson.

- Consideriamo modelli per dati che esprimono il numero di "successi" rispetto al totale di prove effettuate.

- Cosa si deve usare per la regressione Binomiale?

- È un modello che rientra nei GLM, con

$$Z_i \sim \text{Bin}(m_i, \pi_i).$$

⇒ Si vuole modellare la probabilità π_i in funzione di certi valori delle variabili esplicative.

72

- Poichè il valor medio di Z_i dipende anche da m_i e il nostro obiettivo è studiare la relazione tra le variabili esplicative e la probabilità di successo π_i , è più naturale utilizzare come variabile risposta non il numero di successi ma la proporzione di successi:

$$Y_i = Z_i/m_i .$$

- In questo modo, $E(Y_i) = \mu_i = \pi_i$ è la probabilità in questione. Il parametro μ_i deve essere contenuto nell'intervallo $[0, 1]$. Ha allora senso scegliere una funzione legame che rispetta questo vincolo. Una forma naturale è

$$\mu_i = \Psi(\eta_i) ,$$

ove $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ e $\Psi(\cdot)$ è una funzione di ripartizione. La funzione legame è allora

$$g(\mu) = \Psi^{-1}(\mu) .$$

73

- Scelte standard per $\Psi(\cdot)$:

1. $\Psi(\eta) = \Phi(\eta)$, funzione di ripartizione della distribuzione normale standard. In questo caso $g(\mu) = \Phi^{-1}(\mu)$ non ha una forma analitica. Questo modello prende il nome di *regressione probit*. In R: `binomial(link=probit)`.

2. $\Psi(\eta) = \frac{e^\eta}{1+e^\eta}$, per cui

$$g(\mu) = \Psi^{-1}(\mu) = \log \frac{\mu}{1-\mu} .$$

Questo modello prende il nome di *regressione logit* o *logistica*. In R:

`binomial(link=logit)`.

3. $\Psi(\eta) = 1 - \exp(-e^\eta)$, per cui $g(\mu) = \Psi^{-1}(\mu) = \log\{-\log(1-\mu)\}$. La funzione legame è detta "complementary log-log". In R: `binomial(link=cloglog)`.

- In genere, l'inferenza non cambia molto con un cambiamento di legame. Però, l'interpretazione dei modelli risulta più facile con una scelta di legame appropriata.

74

- La funzione *logistica* presenta alcuni vantaggi
 1. è interpretabile in termini di log-rapporto delle probabilità di successo e insuccesso;
 2. essa comporta una considerevole semplicità nelle analisi statistiche;
 3. permette di trattare dati raccolti retrospettivamente.
- Il modello binomiale con legame logit prende il nome di modello *proportional odds* (*rapporto di quote*). In questo caso

$$\eta = \Psi^{-1}(\mu) = \log \frac{\mu}{1-\mu} ,$$

e questa scrittura implica che un cambiamento additivo in η conduce a un cambiamento moltiplicativo negli *odds*, $\mu/(1-\mu)$.

75

Esempio

Per studiare l'insorgenza di malattie cardiovascolari ($Y = 1$) in relazione all'abitudine al fumo (x), il modello logit è:

$$\mu = Pr(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

ovvero

$$\text{logit}(\mu) = \log \frac{\mu}{1-\mu} = \beta_0 + \beta_1 x .$$

Se in particolare si ottiene un valore stimato $\hat{\beta}_1 = 0.7$, ciò significa che il passaggio dalla modalità Fumo = NO ($x = 0$) a Fumo = SI ($x = 1$) comporta un aumento moltiplicativo del log-rapporto di quote pari a 0.7, ovvero del rapporto di quote pari a $e^{0.7} \cong 2$. In altre parole, per i fumatori la probabilità di ammalarsi rispetto a quella di non ammalarsi è in un rapporto di 2 a 1 rispetto ai non fumatori.

76

3.7 Specificazione di un GLM in R

La specificazione in R è un'estensione naturale del modo in cui si specificano e si adattano i LM. Ora si usa la funzione `glm()`, al posto della funzione `lm()`, nel seguente modo:

`glm(formula, family,)`

dove

- *formula*: specifica la variabile risposta e descrive le variabili esplicative da includere nel predittore lineare (come in `lm()`).
- *family*: determina il modello probabilistico di riferimento e la funzione legame. Ad esempio, `normal(link=identity)` o `poisson(link=log)`.

⇒ Qual è la funzione legame di default in R?

Esempio: Poisson

Per la distribuzione di Poisson, si ha

$$b(\theta) = e^\theta.$$

Quindi, l'inversa di $b'(\cdot)$ è la funzione $\log(\cdot)$. Dunque,

$$\eta = g(\mu) = \log \mu$$

è il legame canonico per la distribuzione Poisson. Segue che scrivere `poisson` per *family* nella funzione `glm()` di R è equivalente a scrivere `poisson(link=log)`.

3.8 Funzioni di legame canoniche

- In ogni GLM c'è una scelta della funzione di legame, che gode di particolari proprietà.
- Ricordiamo le relazioni tra θ , μ e η :

$$\eta = g(\mu) \quad \text{e} \quad \mu = b'(\theta).$$

- Tra tutte le funzioni di legame $g(\cdot)$ è privilegiata la $g(\mu) = \theta(\mu)$, secondo la quale il parametro naturale θ della famiglia esponenziale risulta combinazione lineare delle variabili esplicative. Formalmente,

$$\eta = g(\mu) = g(b'(\theta)) = \theta,$$

ossia $g(\cdot)$ è l'inversa di $b'(\cdot)$. Questa scelta di $g(\cdot)$ prende il nome *legame canonico* per la famiglia esponenziale.

- Con questa scelta, il modello ha proprietà statistiche desiderabili. Inoltre, per ciascuna distribuzione il legame canonico è la funzione di default in R.

Esempio

```
##### visite mediche in funzione dell'eta' ####

fit <- glm(numero ~ eta,family=poisson)

summary(fit)
Call:
glm(formula = numero ~ eta, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4869  -0.9250  -0.4152   0.4447   2.4897

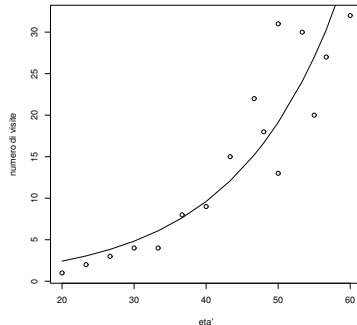
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.49196    0.36608  -1.344    0.179
eta           0.06886    0.00721   9.550 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 136.494  on 15  degrees of freedom
Residual deviance: 19.265  on 14  degrees of freedom
AIC: 89.853

Number of Fisher Scoring iterations: 4

> muhat <- predict(fit,type="response")
> plot(x,y,xlab="eta",ylab="numero di visite")
> lines(x,muhat,type="l")
```



81

4 Inferenza nei GLM

4.1 Stima dei parametri di un GLM

- La stima con il metodo dei minimi quadrati non funziona bene nei GLM. Come nel LM, si potrebbe scegliere quel valore $\tilde{\beta}$ che minimizza

$$\sum_{i=1}^n (Y_i - \mathbf{x}_i^T \beta)^2.$$

- Problemi:

- Y_i e $\mathbf{x}_i^T \beta$ sono definiti su scale differenti;
- si ignora il legame che rapporta Y_i e $\mathbf{x}_i^T \beta$ sulla stessa scala;
- nei GLM non c'è più il termine di errore.

- Le stime ai minimi quadrati non coincidono più con le SMV.

82

⇒ MA ALLORA COSA SI FA NEI GLM???

- La procedura di massima verosimiglianza è preferibile (più flessibile, migliori proprietà, ecc.). Nei LM normali coincide con la stima con il metodo dei minimi quadrati.
- Una proprietà delle famiglie esponenziali è che sono sufficientemente regolari per assicurare che la SMV è la soluzione (unica) dell'equazione di verosimiglianza $\ell_*(\beta) = 0$.
- Supponiamo ϕ noto.
- Per l'ipotesi di indipendenza tra le componenti, la log-verosimiglianza $\ell(\beta)$ ha la forma

$$\begin{aligned} \ell(\beta) &= \sum_{i=1}^n \log f(y_i; \theta_i) = \\ &= \sum_{i=1}^n \left\{ \frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi) \right\} = \sum_{i=1}^n \ell_i(\beta), \end{aligned}$$

ove θ_i è funzione di β attraverso la relazione

$$g(\mu_i) = g(b'(\theta_i)) = \eta_i = \mathbf{x}_i^T \beta.$$

83

- Per ottenere la SMV di β è necessario risolvere le equazioni di verosimiglianza:

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial \ell_i}{\partial \beta_j} = 0, \quad \text{per } j = 1, 2, \dots, p.$$

- Calcoliamo

$$\begin{aligned} \frac{\partial \ell_i}{\partial \beta_j} &= \frac{\partial \ell_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \\ &= \frac{\partial \ell_i}{\partial \theta_i} \left(\frac{\partial \mu_i}{\partial \theta_i} \right)^{-1} \left(\frac{\partial \eta_i}{\partial \mu_i} \right)^{-1} \frac{\partial \eta_i}{\partial \beta_j}, \end{aligned}$$

i cui termini possono essere riscritti come

$$\begin{aligned} \frac{\partial \ell_i}{\partial \theta_i} &= \frac{y_i - b'(\theta_i)}{\phi} = \frac{y_i - \mu_i}{\phi}, \\ \frac{\partial \mu_i}{\partial \theta_i} &= b''(\theta_i) = \frac{\text{var}(Y_i)}{\phi}, \\ \frac{\partial \eta_i}{\partial \mu_i} &= g'(\mu_i), \\ \frac{\partial \eta_i}{\partial \beta_j} &= x_{ij}. \end{aligned}$$

84

- Quindi abbiamo

$$\begin{aligned}\frac{\partial \ell_i}{\partial \beta_j} &= \frac{y_i - \mu_i}{\phi} \frac{\phi}{\text{var}(Y_i)} \frac{1}{g'(\mu_i)} x_{ij} \\ &= \frac{(y_i - \mu_i) x_{ij}}{\phi V(\mu_i) g'(\mu_i)}.\end{aligned}$$

- Le equazioni di verosimiglianza per β sono dunque

$$\sum_{i=1}^n \frac{(y_i - \mu_i)}{V(\mu_i) g'(\mu_i)} x_{ij} = 0,$$

$j = 1, 2, \dots, p$, dove $\mu_i = g^{-1}(x_i^T \beta)$.

- Questo rimane valido per qualsiasi valore di ϕ .
- In genere le equazioni di verosimiglianza per β sono non lineari nei parametri e devono essere risolte con algoritmi numerici (iterativi). A questo scopo torna utile l'informazione di Fisher...

4.2 Informazione di Fisher

- Consideriamo le derivate seconde di ℓ_i :

$$\begin{aligned}-E\left(\frac{\partial^2 \ell_i}{\partial \beta_j \partial \beta_k}\right) &= E\left(\frac{\partial \ell_i}{\partial \beta_j} \frac{\partial \ell_i}{\partial \beta_k}\right) \\ &= E\left(\left(\frac{(Y_i - \mu_i) x_{ij}}{\phi V(\mu_i) g'(\mu_i)}\right) \left(\frac{(Y_i - \mu_i) x_{ik}}{\phi V(\mu_i) g'(\mu_i)}\right)\right) \\ &= \frac{x_{ij} x_{ik}}{\phi^2 (V(\mu_i))^2 (g'(\mu_i))^2} E((Y_i - \mu_i)^2) \\ &= \frac{x_{ij} x_{ik}}{\phi V(\mu_i) (g'(\mu_i))^2},\end{aligned}$$

che costituiscono l'elemento (j, k) -esimo della matrice di informazione attesa. In notazione matriciale,

$$i(\beta) = \frac{X^T W X}{\phi},$$

con $W = \text{diag}(w_1, \dots, w_n)$ e

$$w_i = \frac{1}{V(\mu_i) (g'(\mu_i))^2}.$$

4.3 Legame canonico

- L'adozione di una funzione di legame canonica ($\eta_i = g(\mu_i) = g(b'(\theta_i)) = \theta_i$) dà luogo ad alcune semplificazioni nell'inferenza basata sulla log-verosimiglianza $\ell(\beta)$.

- Per quanto riguarda la derivata prima, se la funzione legame è quella canonica si ha $g'(\mu_i) = 1/V(\mu_i)$ e risulta

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{(y_i - \mu_i) x_{ij}}{\phi V(\mu_i) g'(\mu_i)} = \frac{(y_i - \mu_i) x_{ij}}{\phi}.$$

- Questo implica che le equazioni di verosimiglianza si semplificano grandemente, risultando

$$\sum_{i=1}^n y_i x_{ij} = \sum_{i=1}^n x_{ij} \mu_i.$$

In notazione matriciale, $X^T Y = X^T \mu$. Queste equazioni sono coerenti con la struttura generale delle equazioni di verosimiglianza nelle famiglie esponenziali.

- Un'altra interessante proprietà riguarda l'informazione di Fisher. In corrispondenza di un legame canonico informazione attesa e informazione osservata coincidono e hanno (j, k) -esimo elemento

$$\frac{x_{ij} x_{ik} V(\mu_i)}{\phi}.$$

In notazione matriciale,

$$i(\beta) = j(\beta) = \frac{X^T V X}{\phi},$$

con $V = \text{diag}(V(\mu_i))$.

- Osservazione: Il risultato generale di normalità asintotica dello SMV fornisce l'approssimazione

$$\hat{\beta} \sim N_p(\beta, i(\beta)^{-1}),$$

per n elevato. Una stima consistente per $i(\beta)$ è $i(\hat{\beta})$. Se ϕ è ignoto, anche questo parametro va stimato.

Esempio: normale con legame identità

Abbiamo $g(\mu) = \mu$, per cui $g'(\mu) = 1$. Inoltre, $V(\mu) = 1$, $\phi = \sigma^2$ e $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$. Le equazioni di verosimiglianza diventano

$$\sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}) x_{ij}}{\sigma^2} = 0.$$

Semplificando σ^2 , le equazioni si riducono alle equazioni normali dei minimi quadrati:

$$X^T(\mathbf{y} - X\boldsymbol{\beta}) = 0$$

o, equivalentemente,

$$\begin{aligned} X^T X \boldsymbol{\beta} &= X^T \mathbf{y} \\ \Rightarrow \hat{\boldsymbol{\beta}} &= (X^T X)^{-1} X^T \mathbf{y}. \end{aligned}$$

89

Esempio: Poisson con legame logaritmico

Abbiamo $g(\mu) = \log \mu$, per cui $g'(\mu) = 1/\mu$. Inoltre, $V(\mu) = \mu$, $\phi = 1$ e $\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$. Le equazioni di verosimiglianza diventano

$$\sum_{i=1}^n (y_i - e^{\mathbf{x}_i^T \boldsymbol{\beta}}) x_{ij} = 0,$$

che sono non lineari in $\boldsymbol{\beta}$. Dunque, una soluzione esplicita in genere non esiste.

90

4.4 Algoritmi iterativi

- Le equazioni di verosimiglianza per i GLM non ammettono, in genere, soluzione esplicita ed è necessario risolverle con metodi iterativi.
- Nei GLM c'è la possibilità di affrontare con un unico algoritmo il problema della soluzione delle equazioni di verosimiglianza: questo algoritmo agisce risolvendo una successione di problemi di stima di minimi quadrati.
- Un metodo iterativo, in generale, prevede di partire con un valore iniziale $\hat{\boldsymbol{\beta}}^{(0)}$ e ottenere una sequenza $\hat{\boldsymbol{\beta}}^{(1)}, \hat{\boldsymbol{\beta}}^{(2)}, \dots$, secondo uno schema di aggiornamento della $\hat{\boldsymbol{\beta}}^{(t+1)}$ tramite la $\hat{\boldsymbol{\beta}}^{(t)}$, fino a quando il valore

$$\|\hat{\boldsymbol{\beta}}^{(t+1)} - \hat{\boldsymbol{\beta}}^{(t)}\|$$

è sufficientemente piccolo ($< \epsilon$).

91

Newton-Raphson

- Indichiamo con

$$\boldsymbol{\ell}_*(\boldsymbol{\beta}) = \left(\frac{\partial \ell}{\partial \beta_1}, \dots, \frac{\partial \ell}{\partial \beta_p} \right)^T$$

il vettore *score*. Si vuole risolvere l'equazione

$$\boldsymbol{\ell}_*(\boldsymbol{\beta}) = 0.$$

- Il metodo di Newton-Raphson (che viene giustificato tramite uno sviluppo di Taylor) è basato sulla regola di aggiornamento alla $(t+1)$ -esima iterazione

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(t)} + (i^{(t)})^{-1} \boldsymbol{\ell}_*^{(t)}, \quad (6)$$

con $\boldsymbol{\ell}_*^{(t)} = \boldsymbol{\ell}_*(\hat{\boldsymbol{\beta}}^{(t)})$ e con la matrice Hessiana sostituita dall'informazione attesa e con segno cambiato e calcolata in $\hat{\boldsymbol{\beta}}^{(t)}$, ossia

$$i_{jk}^{(t)} = E \left(-\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k} \right),$$

per $j, k = 1, 2, \dots, p$. Questa modificazione prende

92

il nome di metodo *scoring* di Fisher. Così facendo si mantiene la convergenza dell'algoritmo e le espressioni risultano semplificate (poi se la funzione legame è quella canonica le due espressioni coincidono).

- L'espressione (6) è equivalente a

$$(i^{(t)})\hat{\beta}^{(t+1)} = (i^{(t)})\hat{\beta}^{(t)} + \ell_*^{(t)}.$$

Per la definizione dell'informazione attesa, il membro di destra può essere scritto come

$$(i^{(t)})\hat{\beta}^{(t)} + \ell_*^{(t)} = X^T W^{(t)} s^{(t)},$$

ove s è il vettore con componenti

$s_i = x_i^T \beta + (y_i - \mu_i) g'(\mu_i)$. Infine, si può arrivare alla scrittura

$$X^T W^{(t)} X \hat{\beta}^{(t+1)} = X^T W^{(t)} s^{(t)},$$

che ha la stessa forma delle equazioni normali nel LM. Solo che nei GLM va risolta con metodi iterativi.

Esempio

```
> # stima di un modello di regressione logistica usando R
> # dati: rotenone (veleni)

> attach(rotenone)
> rotenone
  Poison Logdose Kill.1 Kill.2
1      1    1.01    44      6
2      1    0.89    42      7
3      1    0.71    24     22
.....
15     3    1.00    27     19
16     3    0.71    22     24
17     3    0.40     7     40

> # stimiamo un GLM binomiale/logit,
> # chiedendo le SMV ad ogni iterazione:

> Kill <- cbind(Kill.1,Kill.2)
> Poison <- as.factor(Poison)
> for(it in 1:5) {
+   fit <- glm(Kill ~ Poison + Logdose,binomial,maxit=it)
+   print(c(fit$dev,fit$coef)) }
(Intercept) Poison2 Poison3 Logdose
32.53      -3.10     -1.51     -0.64    4.58
(Intercept) Poison2 Poison3 Logdose
31.97      -3.25     -1.60     -0.68    4.81
(Intercept) Poison2 Poison3 Logdose
31.97      -3.26     -1.60     -0.69    4.82
(Intercept) Poison2 Poison3 Logdose
31.97      -3.26     -1.60     -0.69    4.82
(Intercept) Poison2 Poison3 Logdose
31.97      -3.26     -1.60     -0.69    4.82
Warning messages:
.....
```

- Allora, l'iterazione di Newton-Raphson è

$$\hat{\beta}^{(t+1)} = (X^T W^{(t)} X)^{-1} X^T W^{(t)} s^{(t)}. \tag{7}$$

- Questo argomento mostra che ogni passo dell'algoritmo è equivalente a una stima ai minimi quadrati ponderati, sebbene i valori e i pesi cambino ad ogni passo. Per questo prende il nome di *Algoritmo dei Minimi Quadrati Pesati Iterati*.

- L'algoritmo ha due passi principali

1. Dato $\hat{\beta}^{(t)}$, si calcolano $s^{(t)}$ e $W^{(t)}$;
2. Si ottiene quindi $\hat{\beta}^{(t+1)}$ tramite la (7).

Per avviare l'algoritmo una scelta naturale e generalmente conveniente dei valori iniziali è $s^{(0)} = g(Y_i)$ (a volte con qualche modifica...) e $W^{(0)}$ pari alla matrice identità.

```
> # solitamente
> fit <- glm(Kill ~ Poison + Logdose,binomial)
> summary(fit)

Call:
glm(formula = Kill ~ Poison + Logdose, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7725  -1.0948   0.5153   1.4039   2.2419

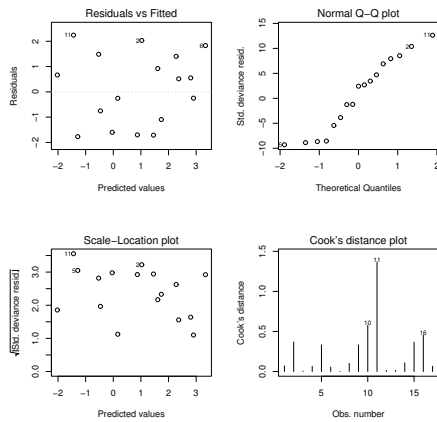
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.2654      0.2797  -11.673  < 2e-16 ***
Poison2      -1.6034      0.2656   -6.036 1.58e-09 ***
Poison3      -0.6911      0.2309   -2.994  0.00276 **
Logdose       4.8277      0.3395   14.222  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 350.069  on 16  degrees of freedom
Residual deviance: 31.971  on 13  degrees of freedom
AIC: 98.955

Number of Fisher Scoring iterations: 4

> fit$coef
(Intercept)      Poison2      Poison3      Logdose
-3.2653728  -1.6033907  -0.6910894   4.8277172
>
> par(mfrow=c(2,2))
> plot(fit)
```

97

4.5 La stima del parametro ϕ

• Nel caso del LM, la stima di β avviene indipendentemente dal valore della varianza σ^2 . C'è un fenomeno identico per il parametro di dispersione ϕ nei GLM.

• La soluzione delle equazioni di verosimiglianza per β , date da

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

è la stessa per qualsiasi scelta di ϕ . Di conseguenza, la stima di β ha la stessa forma se ϕ è noto oppure no.

• Nelle situazioni che richiedono una stima di ϕ , si potrebbe ricorrere alla SMV. Nella pratica è più comune utilizzare uno stimatore alternativo, numericamente più stabile della SMV, e più robusto rispetto a scostamenti dal modello.

98

• Ricordiamo che $\text{var}(Y_i) = \phi V(\mu_i)$. In altre parole,

$$\frac{E((Y_i - \mu_i)^2)}{V(\mu_i)} = \phi.$$

Questo suggerisce la seguente stima per ϕ :

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}, \quad (8)$$

ove

$$\hat{\mu}_i = g^{-1}(\mathbf{x}_i^T \hat{\beta})$$

sono i valori di μ_i stimati. Questi vengono messi a disposizione dall'algoritmo iterativo per β .

• In generale, $\hat{\phi}$ è consistente.

• Nel caso del LM con legame identità, $\hat{\phi}$ coincide con l'usuale espressione di S^2 ; questa connessione spiega il termine $n-p$ al denominatore di $\hat{\phi}$.

99

4.6 Risultato di `glm()` in R

• Il risultato della funzione `glm` è una lista che contiene varie informazioni sul modello stimato. Queste sono estraibili attraverso opportune funzioni. Le principali sono:

- `coef`: valori stimati dei parametri;
- `summary`: sintesi dei risultati della stima;
- `deviance`: per la devianza;
- `resid`: residui del modello;
- `predict`: valori previsti dal modello;
- `anova`: analisi della devianza;
- `plot`: analisi grafiche.

100

Stime dei parametri

Con `glm` i parametri β vengono stimati con il metodo della massima verosimiglianza (SMV).

Normalità asintotica

Per n moderatamente grande, la distribuzione stimata approssimata dello SMV $\hat{\beta}$ è

$$\hat{\beta} \sim N_p(\beta, [i(\hat{\beta})]^{-1}) ,$$

con

$$i(\hat{\beta}) = \frac{X^T \hat{W} X}{\phi} ,$$

con \hat{W} calcolato nella SMV $\hat{\beta}$. Con questo si ottengono le varianze e le covarianze per $\hat{\beta}$. Le varianze stimate sono gli elementi della diagonale della matrice $\hat{\phi}(X^T \hat{W} X)^{-1}$. Inoltre, i termini fuori dalla diagonale contegono le covarianze.

Funzioni dei parametri

- Sia $\tau = h(\beta)$ una funzione del vettore dei parametri β .
- Nel caso in cui l'interesse dell'inferenza è focalizzato su τ , ad esempio $\tau = \beta_2 - \beta_1$, la SMV di τ è $\hat{\tau} = h(\hat{\beta})$, ad esempio $\hat{\tau} = \hat{\beta}_2 - \hat{\beta}_1$ (Proprietà di Equivarianza della SMV).
- Per costruire un intervallo di confidenza per τ o per effettuare una verifica di ipotesi vale l'approssimazione

$$\hat{\tau} \sim N_r \left(h(\beta), (\nabla h(\beta))^T (i(\beta))^{-1} \nabla h(\beta) \right) ,$$

con $r = \dim(\tau)$ e $\nabla h(\beta)^T = (\frac{\partial h}{\partial \beta_1}, \dots, \frac{\partial h}{\partial \beta_p})$.

Intervalli di confidenza

- Sfruttando la distribuzione approssimata dello SMV $\hat{\beta}$, si ottiene per una componente β_j di β l'intervallo con livello di confidenza approssimato $1 - \alpha$ con estremi

$$\hat{\beta}_j \pm z_{1-\alpha/2} \sqrt{\phi[(X^T \hat{W} X)^{-1}]_{j,j}} .$$

- Quando ϕ è ignoto, viene sostituito con la sua stima.
- Nel modello normale, in cui il parametro ϕ viene stimato, il quantile $z_{1-\alpha/2}$ viene sostituito con $t_{n-p, 1-\alpha/2}$, ossia con il quantile dalla distribuzione t con $n - p$ gradi di libertà che corrisponde ai gradi di libertà nella stima di ϕ . Comunque i due approcci sono asintoticamente equivalenti (per $n - p$ sufficientemente grande).
- Anche i test di nullità di un coefficiente di regressione, per l'ipotesi nulla $H_0 : \beta_j = 0$, possono basarsi sulla distribuzione approssimata dello SMV $\hat{\beta}$.

Esempio

```
glm(formula = Kill ~ Poison + Logdose, family = binomial)

....

Coefficients:
              Estimate      Std. Error z value      Pr(>|z|)
(Intercept)   -3.2654         0.2797    -11.673    < 2e-16 ***
Poison2        -1.6034         0.2656     -6.036    1.58e-09 ***
Poison3        -0.6911         0.2309     -2.994    0.00276 **
Logdose         4.8277         0.3395     14.222    < 2e-16 ***
```

Il modello stimato è

$$\text{logit}(\mu_i) = \hat{\beta}_1 + \hat{\beta}_2 \text{Poison2}_i + \hat{\beta}_3 \text{Poison3}_i + \hat{\beta}_4 \text{Logdose}_i .$$

Si potrebbe lavorare con il modello alternativo ma equivalente

$$\text{logit}(\mu_i) = \alpha_1 \text{Poison1}_i + \alpha_2 \text{Poison2}_i + \alpha_3 \text{Poison3}_i + \beta_4 \text{Logdose}_i ,$$

con $\alpha_1 = \beta_1, \alpha_2 = \beta_2 + \beta_1$ e $\alpha_3 = \beta_3 + \beta_1$.

Ad esempio, $\alpha_2 = h(\beta) = \beta_1 + \beta_2$ e quindi $\hat{\alpha}_2 = h(\hat{\beta}) = \hat{\beta}_1 + \hat{\beta}_2$ e

$$\begin{aligned} \widehat{Var}(\hat{\alpha}_2) &= \left[\frac{\partial h(\beta)}{\partial \beta_1} \dots \frac{\partial h(\beta)}{\partial \beta_4} \right] \widehat{Var}(\hat{\beta}) \left[\frac{\partial h(\beta)}{\partial \beta_1} \dots \frac{\partial h(\beta)}{\partial \beta_4} \right]^T \\ &= [1 \ 1 \ 0 \ 0] \widehat{Var}(\hat{\beta}) [1 \ 1 \ 0 \ 0]^T \\ &= \widehat{Var}(\hat{\beta}_1) + \widehat{Var}(\hat{\beta}_2) + 2\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) , \end{aligned}$$

dove $\widehat{Var}(\hat{\beta}) = i(\hat{\beta})^{-1}$ ha come elemento (i, j) , $\widehat{Cov}(\hat{\beta}_i, \hat{\beta}_j)$.

Esempio: Distribuzione Gamma

- Se $Y \sim \text{Gamma}(\nu, \lambda)$, con $\nu > 0$ fissato, la sua funzione di densità è

$$\begin{aligned} f(y; \nu, \lambda) &= \frac{y^{\nu-1} \exp\{-y/\lambda\}}{\lambda^\nu \Gamma(\nu)} \\ &= \exp\{(\nu-1) \log y - y/\lambda - \nu \log \lambda - \log \Gamma(\nu)\} \\ &= \exp\{-y/\lambda - \nu \log \lambda + (\nu-1) \log y - \log \Gamma(\nu)\} \end{aligned}$$

per $y > 0$.

- Alternativamente, possiamo usare la parametrizzazione $\mu = \nu\lambda = E(Y)$. Quindi

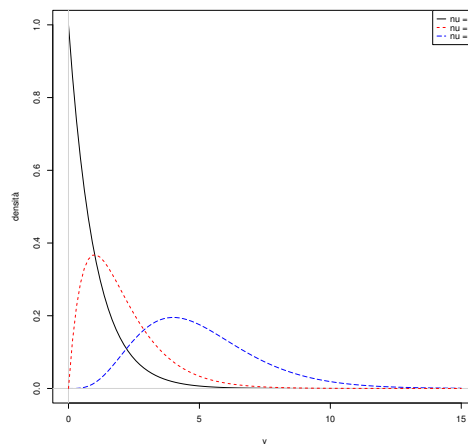
$$f(y; \mu, \nu) = \exp\{\nu(-y/\mu - \log \mu) + \nu \log \nu y - \log y - \log \Gamma(\nu)\}$$

- Questa densità si può scrivere in forma esponenziale

$$f(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right\},$$

dove $\theta = -1/\mu$, $\phi = 1/\nu$, $b(\theta) = -\log(-\theta)$ e $c(y, \phi) = c(y, \nu) = \nu \log \nu y - \log y - \log \Gamma(\nu)$ (si veda la tabella dei lucidi).

Densità Gamma con parametro di scala 1



- Adatta per modellare variabili continue positive, come ad esempio dati di durata o sopravvivenza.
- Notiamo che $b'(\theta) = -1/\theta = \mu$, $b''(\theta) = 1/\theta^2 = \mu^2$ e quindi $\text{Var}(Y) = \phi\mu^2$.
- Quindi la varianza è funzione (quadratica) della media.
- Però il coefficiente di variazione è costante:

$$\frac{\text{Var}(Y)^{1/2}}{E(Y)} = \frac{\sqrt{\phi}\mu}{\mu} = \sqrt{\phi}.$$

- In modo simile a quanto accade nella Normale, il parametro di dispersione è in genere ignoto e viene quindi stimato.
- La funzione legame canonica è il reciproco ($1/\mu$). Altre possibilità sono il logaritmo (molto usato) e la funzione identità.

Un esempio: scimpanzè e parole...

Scimp.	Parole									
	1	2	3	4	5	6	7	8	9	10
1	178	60	177	36	225	345	40	2	287	14
2	78	14	80	15	10	115	10	12	129	80
3	99	18	20	25	15	54	25	10	476	55
4	297	20	195	18	24	420	40	15	372	190

- La tabella riporta il tempo (in minuti) impiegato da 4 scimpanzè per imparare ciascuna di 10 parole.
- Vista la natura della variabile risposta si può pensare di modellare la variabile risposta con un modello lineare generalizzato per dati Gamma.
- Ci sono due possibili predittori, entrambi di tipo qualitativo.

```
> library(SMPracticals)
> data(chimps)
> chimps
  chimp word  y
1     1    1 178
2     1    2  60
3     1    3 177
4     1    4  36
5     1    5 225
6     1    6 345
7     1    7  40
8     1    8   2
9     1    9 287
10    1   10  14
11    2    1  78
12    2    2  14

.....

38    4    8  15
39    4    9 372
40    4   10 190
```

```
> chimps.glm <- glm(y~chimp+word,family=Gamma,data=chimps)
> summary(chimps.glm)

Call:
glm(formula = y ~ chimp + word, family = Gamma, data = chimps)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4855   -0.3889   -0.0162    0.2529    1.2732

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.004799   0.002309    2.08  0.047 *
chimp2       0.007206   0.003890    1.85  0.075 .
chimp3       0.003088   0.002653    1.16  0.255
chimp4      -0.000517   0.001570   -0.33  0.745
word2        0.028722   0.013146    2.18  0.038 *
word3        0.002111   0.003631    0.58  0.566
word4        0.035523   0.015615    2.27  0.031 *
word5        0.007921   0.005642    1.40  0.172
word6       -0.001575   0.002529   -0.62  0.538
word7        0.027797   0.012810    2.17  0.039 *
word8        0.095412   0.037406    2.55  0.017 *
word9       -0.002457   0.002335   -1.05  0.302
word10       0.005233   0.004696    1.11  0.275
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for Gamma family taken to be 0.53063)

Null deviance: 60.378  on 39  degrees of freedom
Residual deviance: 17.081  on 27  degrees of freedom
AIC: 424

Number of Fisher Scoring iterations: 7
```

```
> chimps.glm2 <- glm(y~chimp+word,family=Gamma(log),data=chimps)
> summary(chimps.glm2)

Call:
glm(formula = y ~ chimp + word, family = Gamma(log), data = chimps)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6766   -0.4510   -0.0488    0.3074    1.2847

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.458      0.375   14.54 2.7e-14 ***
chimp2       -0.973      0.295   -3.31 0.00269 **
chimp3       -0.808      0.295   -2.74 0.01068 *
chimp4       -0.113      0.295   -0.38 0.70479
word2        -1.771      0.466   -3.80 0.00074 ***
word3        -0.365      0.466   -0.78 0.44007
word4       -1.821      0.466   -3.91 0.00056 ***
word5       -1.102      0.466   -2.37 0.02541 *
word6         0.279      0.466    0.60 0.55461
word7       -1.725      0.466   -3.70 0.00096 ***
word8       -2.555      0.466   -5.49 8.3e-06 ***
word9         0.811      0.466    1.74 0.09301 .
word10       -0.514      0.466   -1.10 0.27972
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for Gamma family taken to be 0.43367)

Null deviance: 60.378  on 39  degrees of freedom
Residual deviance: 14.972  on 27  degrees of freedom
AIC: 418.4

Number of Fisher Scoring iterations: 12
```

- Nel primo modello si è usato il legame canonico ($1/\mu$): all’aumentare del predittore lineare aumenta il reciproco della media (e quindi la media diminuisce...).
- Nel secondo caso, si è usato il legame logaritmo ($\log \mu$): all’aumentare del predittore lineare aumenta il logaritmo della media (e quindi anche la media)
- In quest’ultimo caso l’interpretazione dei coefficienti è più agevole.
- Il parametro di dispersione è stimato come:

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

Ad esempio, nel primo modello si ha

```
> mu <- fitted(chimps.glm)
> sum((chimps$y-mu)^2/mu^2)/chimps.glm$df.residual
[1] 0.53063
> summary(chimps.glm)$dispersion
[1] 0.53063
```

5 Adeguatezza dei modelli

• Fasi, in successione logica e cronologica, di un processo di modellazione:

1. **Specificazione del modello:** individuazione della EF opportuna, del predittore lineare e della funzione di legame per i dati oggetto di studio;
2. **Stima del modello:** adattamento del modello specificato per i dati oggetto di studio e relativa selezione delle variabili esplicative da includere nel modello;
3. **Controllo del modello finale:** valutazione della "bontà" del modello adattato tramite analisi dei residui e analisi della devianza.

113

- L'adattare un modello ai dati può essere visto come un modo di sostituire un insieme di dati y con un insieme di valori stimati $\hat{\mu}$, derivati da un modello dipendente da un numero relativamente limitato di parametri.
- In generale, i valori stimati $\hat{\mu}$ non coincidono esattamente con i valori osservati y .
- Sorge quindi naturale il problema di stabilire quanto *discrepanti* siano $\hat{\mu}$ e y poiché, mentre una piccola discrepanza può essere tollerata, una grande discrepanza in genere non lo è.
- Misure di discrepanza, o di bontà d'adattamento, possono essere definite in vario modo. Nel seguito, considereremo prevalentemente una misura basata sulla log verosimiglianza, che chiameremo **devianza**.
- Questa misura potrà essere utilizzata sia per valutare la bontà d'adattamento di un particolare modello ai dati (3.), sia per confrontare due o più modelli annidati (2.).

114

5.1 Devianza

• Consideriamo il problema di confrontare due GLM annidati, indicati con M_C e M_R , tali che $M_R \subset M_C$. In particolare, il modello corrente M_C contiene p parametri:

$$\eta_{i,MC} = \beta_0 + \beta_1 x_{1i} + \dots + \beta_{p_0} x_{p_0 i} + \dots + \beta_p x_{pi} ,$$

e il modello ridotto M_R contiene p_0 parametri, con $p > p_0$:

$$\eta_{i,MR} = \beta_0 + \beta_1 x_{1i} + \dots + \beta_{p_0} x_{p_0 i} .$$

• Si consideri la partizione di β in $\beta = (\beta_{MR}, \beta_{MRC})$, con $\beta_{MR} = (\beta_1, \dots, \beta_{p_0})$ e $\beta_{MRC} = (\beta_{p_0+1}, \dots, \beta_p)$. Si supponga di voler verificare

$$H_0 : \beta_{MRC} = 0 \quad \text{contro} \quad H_1 : \beta_{MRC} \neq 0 .$$

• Come criterio per confrontare M_C e M_R si vuole usare il rapporto di verosimiglianza

$$W = 2\{\ell(\hat{\beta}) - \ell(\hat{\beta}_{MR})\} .$$

115

COSA SUCCEDEVA NEL LM???

• Nei LM con errori normali, con σ^2 noto, il rapporto di verosimiglianza è funzione della devianza (somma dei quadrati dei residui) $D = SQE = \sum_i (y_i - \hat{\mu}_i)^2$ associata a ciascuno dei due modelli. In particolare per confrontare due modelli annidati ($M_R \subset M_C$), il rapporto di verosimiglianza suggerisce di rifiutare H_0 per valori elevati della statistica

$$W = 2\{\ell(\hat{\beta}) - \ell(\hat{\beta}_{MR})\} = \frac{D_{MR} - D}{\sigma^2} ,$$

ove $D_{MR} = SQE_{H_0}$ e $D = SQE$ rappresentano la somma dei quadrati dei residui con riferimento al modello ridotto e corrente, rispettivamente.

• Sotto H_0 tale statistica ha distribuzione $\chi^2_{p-p_0}$.

116

E NEI GLM COSA SI FA????

- In analogia ai LM con errori normali, per i GLM si vuole esprimere il rapporto di verosimiglianza in modo che mantenga chiara la connessione tra le due classi di modelli. A tale scopo abbiamo bisogno di estendere il concetto di **devianza** all'ambito dei GLM.

- Partiamo dal test del log-rapporto di verosimiglianza. La log-verosimiglianza per un GLM è

$$\ell(\beta) = \sum_{i=1}^n \ell_i(\beta),$$

con

$$\ell_i(\beta) = \frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi).$$

- Nel caso di modelli annidati, risulta che la statistica

$$W = 2\{\ell(\hat{\beta}) - \ell(\hat{\beta}_{MR})\}$$

ha sotto H_0 distribuzione approssimata $\chi^2_{p-p_0}$.

- L'analogia formale con il LM normale può essere evidenziata introducendo la verosimiglianza associata al modello *saturo* o *massimale*. Il modello saturo è definito come:

→ un GLM basato sulla stessa distribuzione del modello corrente;

→ un GLM con la stessa funzione legame del modello corrente;

→ un GLM con numero di parametri pari al numero di osservazione, tipicamente n ;

- Le funzioni di log-verosimiglianza per il modello saturo e per il modello corrente possono essere valutate nelle rispettive SMV (date da $\tilde{\theta}$ e da $\hat{\theta}$). Se il modello corrente si adatta bene ai dati, $\ell(\tilde{\theta})$ dovrebbe essere equivalente a $\ell(\hat{\theta})$. Se invece il modello corrente si adatta male, allora $\ell(\hat{\theta})$ dovrebbe essere molto più piccola di $\ell(\tilde{\theta})$.

- Formalizzando, la quantità

$$D(y; \hat{\theta}) = 2\phi\{\ell(\tilde{\theta}) - \ell(\hat{\theta})\} = \phi \sum_{i=1}^n D_i$$

con

$$D_i = 2\{y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)\},$$

è detta *funzione di devianza* del modello.

- La funzione

$$D^*(y; \hat{\theta}) = \frac{D(y; \hat{\theta})}{\phi} = \sum_{i=1}^n D_i \quad (9)$$

è la *devianza scalata* e coincide con il log rapporto di verosimiglianza.

- Sia D che D^* sono sempre non negative.

- La quantità (9) è piccola quando il modello rappresenta bene i dati, mentre se la (9) è grande il modello corrente non si adatta bene ai dati. In questo senso la *devianza* è equivalente alla SQE nel LM.

- Si osservi che $\ell(\tilde{\theta})$ rappresenta la log-verosimiglianza ottenuta ponendo $\mu_i = b'(\theta_i) = y_i (\Leftrightarrow (\partial \ell_i / \partial \theta_i) = 0)$, ovvero adattando il modello di regressione saturo con $p = n$ parametri. Tale modello ha tanti parametri quante sono le osservazioni.

- La differenza delle log-verosimiglianze tra modello saturo e quello in esame, data dalla (9) rappresenta una misura della diminuzione della bontà di adattamento dovuta al passaggio dal modello saturo a quello corrente con $p < n$ variabili esplicative. Pertanto, la devianza $D(y; \hat{\theta})$ ha lo stesso ruolo della somma dei quadrati dei residui (devianza residua SQE) nel LM lineare classico.

- Dunque, $\ell(\tilde{\theta})$ fornisce un valore di riferimento per la log-verosimiglianza. Tale modello non è di utilità pratica, dal momento che non sintetizza i dati, ma serve come termine di confronto per il modello effettivamente in esame.

Esempio: Normale con legame identità

- $Y_i \sim N(\mu_i, \sigma^2)$, $b(\theta) = \frac{\theta^2}{2}$, $\theta = \mu = b'(\theta)$ e $\phi = \sigma^2$.
- $\ell(\theta) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu_i)^2$
- Per il modello saturo si ha $\tilde{\mu}_i = y_i$, e quindi

$$\ell(\tilde{\theta}) = -\frac{n}{2} \log \sigma^2 .$$

- Per il modello corrente si ha $\hat{\mu}_i = x_i^T \hat{\beta}$, e quindi

$$\ell(\hat{\theta}) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \hat{\mu}_i)^2$$

- Allora, la devianza scalata è

$$D^*(y; \hat{\theta}) = 2 \left(\ell(\tilde{\theta}) - \ell(\hat{\theta}) \right) = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\sigma^2} ,$$

che coincide con la SQE del modello corrente, divisa per σ^2 .

Esempio: Poisson

- $Y_i \sim \text{Poisson}(\mu_i)$, $b(\theta_i) = e^{\theta_i} = b'(\theta_i) = \mu_i$, $\phi = 1$
- $\theta_i = \log \mu_i$ e $g(\mu_i) = x_i^T \beta$ (ad es., $g(\mu) = \log \mu$)
- $\ell(\theta) = \sum_{i=1}^n y_i \log \mu_i - \sum_{i=1}^n \mu_i$
- Per il modello saturo si ha $\tilde{\mu}_i = y_i$, e quindi

$$\ell(\tilde{\theta}) = \sum_{i=1}^n y_i \log y_i - \sum_{i=1}^n y_i .$$

- Per il modello corrente $\hat{\mu}_i = g^{-1}(x_i^T \hat{\beta})$, e quindi

$$\ell(\hat{\theta}) = \sum_{i=1}^n y_i \log \hat{\mu}_i - \sum_{i=1}^n \hat{\mu}_i .$$

- Quindi

$$\begin{aligned} D(y; \hat{\theta}) &= 2 \left(\ell(\tilde{\theta}) - \ell(\hat{\theta}) \right) \\ &= 2 \left(\sum_{i=1}^n y_i \log \frac{y_i}{\hat{\mu}_i} - \sum_{i=1}^n y_i + \sum_{i=1}^n \hat{\mu}_i \right) \\ &= 2 \sum_{i=1}^n y_i \log \frac{y_i}{\hat{\mu}_i} \cong \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} . \end{aligned}$$

- Se il modello ha l'intercetta $\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{\mu}_i$.

Esempio: Binomiale

- $Y_i \sim \text{Bin}(1, \pi_i)$, con $\pi_i = \text{Pr}(Y_i = 1) = E(Y_i) = \mu_i$
- $g(\mu_i) = x_i^T \beta$
- $\ell(\theta) = \sum_{i=1}^n (y_i \log \pi_i + (1 - y_i) \log(1 - \pi_i))$
- Per il modello saturo si ha $\tilde{\mu}_i = y_i$ e

$$\ell(\tilde{\theta}) = \sum_{i=1}^n (y_i \log y_i + (1 - y_i) \log(1 - y_i)) .$$

- Per il modello corrente $\hat{\mu}_i = g^{-1}(x_i^T \hat{\beta})$ e quindi

$$\ell(\hat{\theta}) = \sum_{i=1}^n (y_i \log \hat{\pi}_i + (1 - y_i) \log(1 - \hat{\pi}_i)) .$$

- $$\begin{aligned} D(y; \hat{\theta}) &= 2 \left(\ell(\tilde{\theta}) - \ell(\hat{\theta}) \right) \\ &= 2 \sum_{i=1}^n \left(y_i \log \frac{y_i}{\hat{\pi}_i} + (1 - y_i) \log \frac{1 - y_i}{1 - \hat{\pi}_i} \right) \\ &= 2 \sum_{i=1}^n o_i \log \frac{o_i}{e_i} , \end{aligned}$$

con o_i = frequenze osservate (y_i e $1 - y_i$) e e_i = frequenze stimate ($\hat{\pi}_i$ e $1 - \hat{\pi}_i$).

Esempio: Gamma

- $Y_i \sim \text{Gamma}(\nu, \mu_i)$, $b(\theta) = -\log(-\theta)$,
 $\mu = b'(\theta) = -1/\theta$ e $\phi = 1/\nu$.
- $g(\mu_i) = x_i^T \beta$
- $\ell(\theta) = \frac{1}{\phi} \sum_{i=1}^n (-\frac{y_i}{\mu_i} - \log \mu_i)$

- Per il modello saturo si ha $\tilde{\mu}_i = y_i$, e quindi

$$\ell(\tilde{\theta}) = \frac{1}{\phi} \sum_{i=1}^n (-1 - \log y_i) .$$

- Per il modello corrente si ha $\hat{\mu}_i = g^{-1}(x_i^T \hat{\beta})$, e quindi

$$\ell(\hat{\theta}) = \frac{1}{\phi} \sum_{i=1}^n \left(-\frac{y_i}{\hat{\mu}_i} - \log \hat{\mu}_i \right)$$

- Allora, la devianza scalata è

$$\begin{aligned} D^*(y; \hat{\theta}) &= 2 \left(\ell(\tilde{\theta}) - \ell(\hat{\theta}) \right) \\ &= \frac{2}{\phi} \sum_{i=1}^n \left\{ \frac{(y_i - \hat{\mu}_i)}{\hat{\mu}_i} - \log \frac{y_i}{\hat{\mu}_i} \right\} . \end{aligned}$$

5.2 Confronto di modelli annidati

- Nel caso di modelli annidati M_C e M_R , il test del rapporto di verosimiglianza è

$$\begin{aligned} W &= 2 \left\{ \ell(\hat{\beta}) - \ell(\hat{\beta}_{MR}) \right\} \\ &= \frac{D(Y, \hat{\theta}_{MR}) - D(Y, \hat{\theta})}{\hat{\phi}}, \end{aligned}$$

e ha distribuzione approssimata $\chi^2_{p-p_0}$ sotto H_0 .

- Il test della validità del modello ridotto viene dal confronto di

$$W = \frac{D(Y, \hat{\theta}_{MR}) - D(Y, \hat{\theta})}{\hat{\phi}}$$

con i quantili della distribuzione $\chi^2_{p-p_0}$. Si rifiuta H_0 se la statistica assume valori elevati (*p-value* piccolo).

125

- In R si possono confrontare due modelli annidati attraverso il comando

```
anova(modello.ridotto, modello)
```

che fornisce la tabella di analisi della devianza. È possibile chiedere il *p-value*, scegliendo l'opportuna distribuzione di riferimento: `test='Chisq'` per ϕ noto o `test='F'` per ϕ ignoto.

- Nel caso di confronti sequenziali di modelli annidati, in cui i vari termini (variabili o interazioni di variabili) entrano sequenzialmente nel modello, possiamo giudicare l'apporto di un singolo termine in base alla corrispondente riduzione di devianza, analogamente all'analisi della varianza nei modelli lineari normali. Bisogna tener conto che l'esito di tali confronti può dipendere dall'ordine con cui i vari termini entrano nel modello. Infatti, ciascuna riduzione di devianza è relativa all'apporto del relativo termine *al netto* di tutti i termini già presenti nel modello.

- È possibile ottenere la tabella di analisi (sequenziale) della devianza di un modello con il comando `anova(modello)`.

127

- Nel caso in cui ϕ non sia noto, si sostituisce la sua stima consistente

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

(vedi (8) a pagina 99), oppure, meno frequentemente, con la stima alternativa (pur sempre consistente)

$$\tilde{\phi} = \frac{D(Y, \hat{\theta})}{(n-p)}$$

- Sotto opportune condizioni

$$W/(p-p_0) = \frac{\{D(Y, \hat{\theta}_{MR}) - D(Y, \hat{\theta})\}/(p-p_0)}{\hat{\phi}}$$

ha distribuzione $F_{p-p_0, n-p}$. Nel caso normale, tale distribuzione è esatta. Più in generale, come ad esempio nel caso gamma, la distribuzione è approssimata.

126

Esempio: Poisson

```
# Numero di visite mediche (y) in funzione dell'eta' (x)
```

```
> fit <- glm(y~x, family=poisson)
> summary(fit)
```

```
Call:
glm(formula = y ~ x, family = poisson)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.487  -0.925  -0.415   0.445   2.490
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.49196    0.36608  -1.34    0.18
x             0.06886    0.00721   9.55 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for poisson family taken to be 1)
```

```
Null deviance: 136.494 on 15 degrees of freedom
Residual deviance: 19.265 on 14 degrees of freedom
AIC: 89.85
```

```
Number of Fisher Scoring iterations: 4
```

128


```
> anova(fit,test='Chisq')
Analysis of Deviance Table

Model: poisson, link: log

Response: y

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                                15      136.5
x      1      117.2      14       19.3  2.6e-27
```

```
> fit1 <- glm(y~x+I(x^2),family=poisson)

> anova(fit,fit1,test='Chisq')
Analysis of Deviance Table

Model 1: y ~ x
Model 2: y ~ x + I(x^2)
      Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1           14      19.26
2           13      12.58 1       6.68    0.01
```

```
> summary(fit1)

Call:
glm(formula = y ~ x + I(x^2), family = poisson)

Deviance Residuals:
      Min       1Q   Median       3Q      Max
-2.0116  -0.3954   0.0421   0.3014   1.8836

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.839883    1.472314  -2.61  0.00911 **
x             0.225615    0.065363   3.45  0.00056 ***
I(x^2)       -0.001746    0.000713  -2.45  0.01439 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 136.494 on 15 degrees of freedom
Residual deviance: 12.584 on 13 degrees of freedom
AIC: 85.17

Number of Fisher Scoring iterations: 4
```

```
> anova(fit1,test='Chisq')
Analysis of Deviance Table

Model: poisson, link: log

Response: y

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                                15      136.5
x      1      117.2      14       19.3  2.6e-27
I(x^2)  1         6.7      13       12.6  9.7e-03

> fitp <- glm(y~x+I(x^2)+I(x^3)+I(x^4),family=poisson)
>
> anova(fitp,test='Chisq')
Analysis of Deviance Table

Model: poisson, link: log

Response: y

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                                15      136.5
x      1      117.2      14       19.3  2.6e-27
I(x^2)  1         6.7      13       12.6  9.7e-03
I(x^3)  1         0.1      12       12.5    0.8
I(x^4)  1         0.3      11       12.2    0.6
```

Esempio: Binomiale

```
> attach(rotenone)
> Kill <- cbind(Kill.1,Kill.2)
> Poison <- as.factor(Poison)
> fit <- glm(Kill ~ Poison + Logdose,binomial)
> summary(fit)

Call:
glm(formula = Kill ~ Poison + Logdose, family = binomial)

Deviance Residuals:
      Min       1Q   Median       3Q      Max
-1.772  -1.095   0.515   1.404   2.242

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.265      0.280  -11.67 < 2e-16 ***
Poison2        -1.603      0.266   -6.04  1.6e-09 ***
Poison3        -0.691      0.231   -2.99  0.0028 **
Logdose         4.828      0.339   14.22 < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 350.069 on 16 degrees of freedom
Residual deviance: 31.971 on 13 degrees of freedom
AIC: 98.96

Number of Fisher Scoring iterations: 4
```

```
> anova(fit,test='Chisq')
Analysis of Deviance Table

Model: binomial, link: logit

Response: Kill

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                16      350
Poison    2         18      14      332 1.2e-04
Logdose   1        300      13        32 3.1e-67

> anova(glm(Kill~Logdose+Poison,binomial),test='Chisq')
Analysis of Deviance Table

Model: binomial, link: logit

Response: Kill

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                16      350
Logdose    1        278      15        72 1.8e-62
Poison     2         40      13        32 2.3e-09
```

Esempio: Gamma

```
>library(SMPracticals)
>data(chimps)
>attach(chimps)
>chimps
      chimp word   y
1         1    1 178
2         1    2  60
3         1    3 177
4         1    4  36
5         1    5 225
6         1    6 345
7         1    7  40
8         1    8   2
9         1    9 287
10        1   10  14
11        2    1  78
12        2    2  14

....

38        4    8  15
39        4    9 372
40        4   10 190
```

```
> chimps.glm1 <- glm(y~chimp+word,family=Gamma(log))
> anova(chimps.glm1,test='F')
Analysis of Deviance Table

Model: Gamma, link: log

Response: y

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev    F Pr(>F)
NULL                39      60.4
chimp    3         6.9      36      53.4 5.34 0.0051 **
word     9        38.5      27      15.0 9.85 1.7e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
```

```
> chimps.glm2 <- glm(y~word+chimp,family=Gamma(log))
> aod <- anova(chimps.glm2,test='F')
> aod
Analysis of Deviance Table

Model: Gamma, link: log

Response: y

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev    F Pr(>F)
NULL                39      60.4
word     9        39.2      30      21.2 10.04 1.4e-06 ***
chimp    3         6.2      27      15.0  4.78 0.0085 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

> # test F = rid. di devianza/gdl/stima par di dispersione

> phihat <- summary(chimps.glm2)$dispersion
> F <- aod$Deviance[3]/aod$Df[3]/phihat
> F
[1] 4.7812
> 1-pf(F,3,27)
[1] 0.0084594
```

5.3 Bontà d’adattamento

- La devianza ha alcune semplici proprietà che indicano la sua utilità anche come misura di bontà d’adattamento:
 - Se un modello si adatta perfettamente ai dati, $\tilde{\mu} = y$ e la devianza è zero, $D(y; \tilde{\theta}) = 0$; altrimenti la devianza è positiva.
 - La massimizzazione della verosimiglianza per ogni modello è equivalente alla minimizzazione della devianza. Quindi la massima verosimiglianza fornisce il miglior adattamento ai dati secondo il criterio della devianza.
- Come possiamo calibrare la devianza?
- La devianza stessa può essere vista come la differenza di devianze tra il modello corrente e il modello saturo, $D(y; \hat{\theta}) = D(y; \hat{\theta}) - D(y; \tilde{\theta})$.
- Ogni modello si può pensare annidato all’interno del modello saturo.

Esempio: Fuma fuma...

```
> library(SMPracticals)
> data(smoking)
> attach(smoking)
> smoking
```

	age	smoker	alive	dead
1	18-24	1	53	2
2	18-24	0	61	1
3	25-34	1	121	3
4	25-34	0	152	5
5	35-44	1	95	14
6	35-44	0	114	7
7	45-54	1	103	27
8	45-54	0	66	12
9	55-64	1	64	51
10	55-64	0	81	40
11	65-74	1	7	29
12	65-74	0	28	101
13	75+	1	0	13
14	75+	0	0	64

- Quindi, saremmo tentati di pensare che la devianza stessa possa avere distribuzione approssimata χ^2_{n-p} (eventualmente moltiplicato per ϕ).
- Nel modello lineare normale questo è vero (ed è esatta). Però in generale è falso perché l’approssimazione $\chi^2_{p-p_0}$ per la differenza di devianze di modelli annidati presuppone che $n \rightarrow \infty$, ma che p e p_0 siano fissati; nel modello saturo $p = n$.
- Tuttavia, con una diversa argomentazione, si può vedere che questo risultato è ancora vero per modelli Poisson con medie stimate grandi ($\hat{\mu}_i > 5$), modelli binomiali con m_i (quindi **non** con $m_i = 1$!) grandi e modelli Gamma con ϕ piccolo (e quindi con parametro di forma grande).
- In generale comunque, una devianza molto più grande dei suoi gradi di libertà dovrebbe far sorgere qualche sospetto sulla bontà del modello.

```
> fit1 <- glm(cbind(dead,alive)~smoker,family=binomial)
> summary(fit1)
```

Call:
glm(formula = cbind(dead, alive) ~ smoker, family = binomial)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-9.05	-5.67	-1.87	5.78	12.17

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.7805	0.0796	-9.80	<2e-16 ***
smoker	-0.3786	0.1257	-3.01	0.0026 **

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 641.5 on 13 degrees of freedom
Residual deviance: 632.3 on 12 degrees of freedom
AIC: 683.3

Number of Fisher Scoring iterations: 4

```
> fit2 <- glm(cbind(dead,alive)~smoker+age,family=binomial)
> summary(fit2)

Call:
glm(formula = cbind(dead, alive) ~ smoker + age,
family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-7.25e-01 -2.28e-01  4.87e-05  1.91e-01  6.82e-01

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.860     0.594   -6.50  8.0e-11 ***
smoker          0.427     0.177    2.41  0.01576 *
age25-34        0.120     0.687    0.17  0.86118
age35-44        1.341     0.629    2.13  0.03287 *
age45-54        2.113     0.612    3.45  0.00056 ***
age55-64        3.181     0.601    5.30  1.2e-07 ***
age65-74        5.088     0.620    8.21 < 2e-16 ***
age75+       27.807    11293.143  0.0025 0.99804
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 641.4963  on 13  degrees of freedom
Residual deviance:  2.3809  on  6  degrees of freedom
AIC: 65.38

Number of Fisher Scoring iterations: 20
> 1-pchisq(2.3809,6)
[1] 0.88155
```

141

```
> anova(fit1,fit2,test='Chisq')
Analysis of Deviance Table

Model 1: cbind(dead, alive) ~ smoker
Model 2: cbind(dead, alive) ~ smoker + age
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1         12         632
2          6          2    6      630  8.2e-133
```

- La devianza del primo modello è molto alta rispetto ai gradi di libertà, il che indica un pessimo adattamento del modello ai dati.
- Al contrario, il secondo modello ha una devianza bassa rispetto ai gradi di libertà, indicando un buon adattamento del modello.
- Essendo il modello binomiale con m_i tutti piuttosto elevati, possiamo confrontare il suo valore con la distribuzione χ^2_6 .
- NOTA: il fatto che l'effetto del fumo sulla salute cambi direzione da un modello all'altro è dovuto al fatto che l'associazione tra due variabili può cambiare sostanzialmente se non si tiene conto di variabili legate ad entrambe (paradosso di Simpson).

142

5.4 Residui

- Adattato un GLM è importante confrontare i valori stimati $\hat{\mu}_i$ con le risposte y_i . Il modo più semplice per farlo è attraverso l'ispezione grafica dei residui.
- Ricordiamo che nel caso dei LM, i residui sono uno strumento per valutare l'adeguatezza di un modello stimato.
- Esistono diverse estensioni di residui per i GLM.
- L'estensione diretta del concetto di residuo standardizzato è data da

$$r_{Pi} = \frac{Y_i - \hat{\mu}_i}{\sqrt{\hat{\phi}V(\hat{\mu}_i)}}, \quad (10)$$

che è detto *residuo di Pearson*. La (10) è analoga alla definizione di residui nei LM, si veda la (3), in cui viene stimato il termine di errore ϵ_i .

143

- Poichè nei GLM tale ϵ_i non esiste, ha senso considerare il contributo alla devianza. Infatti nel LM classico la devianza dei residui (SQE) è pari a

$$SQE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - x_i^T \hat{\beta})^2,$$

mentre nel GLM l'analogia quantità è data dalla devianza. Ricordiamo che la devianza è

$$D(y, \hat{\theta}) = \sum_{i=1}^n D_i.$$

Valori grandi di D_i corrispondono a dati che non vanno bene per il modello. Definiamo allora

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

che prende il nome di *residuo di devianza* per il modello.

- Uno sviluppo di Taylor mostra che $r_{Pi} \approx r_{Di}$.
- Esistono anche altri tipi di residui...

144

- Se il modello è valido, (molto) approssimativamente i residui (di qualsiasi tipo), riscalati per $\hat{\phi}$ se necessario, dovrebbero seguire la distribuzione $N(0, 1)$. Pertanto,
 - ↳ grafici che controllano se i residui seguono una distribuzione normale (utile anche per identificare valori anomali);
 - ↳ grafici dei residui contro i valori stimati $\hat{\mu}_i$;
 - ↳ grafici dei residui contro le variabili esplicative (corretta specificazione della formula del modello),
- possono essere usati per controllare la bontà del modello stimato.
- Quando la risposta è dicotomica, l'esame dei residui è "difficile".

Esempio: Poisson

```

> fit <- glm(y ~ x, poisson)

> resid(fit,type="response")
> resid(fit,type="pearson")
> resid(fit,type="deviance")

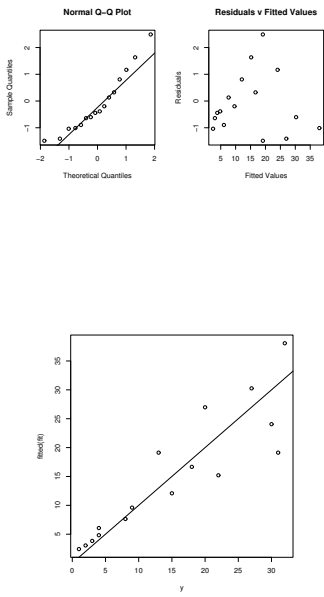
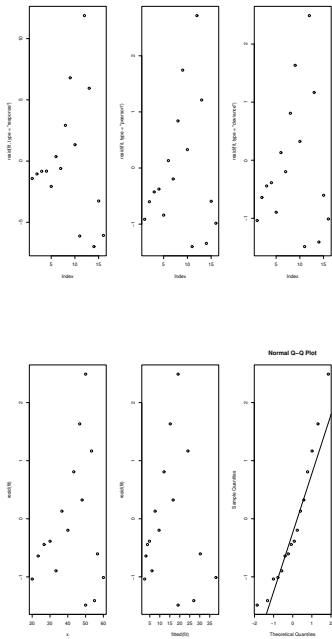
> plot(resid(fit,type="response"))
> plot(resid(fit,type="pearson"))
> plot(resid(fit,type="deviance"))

> plot(x,resid(fit))
> plot(fitted(fit),resid(fit))
> qqnorm(resid(fit))
> qqline(resid(fit))

> resid.check(fit)

> plot(y,fitted(fit))
> abline(0,1)
# fitted(fit) = predict(fit,type="response")

```



Esempio: Binomiale

```
> seed
      fert x
1  0.0000000 0
2  0.1315789 0
.....
19 2.3684211 1
20 2.5000000 1

> attach(seed)

> fit.seed <- glm(x~fert,binomial)

> summary(fit.seed)

Call:
glm(formula = x ~ fert, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7477  -0.3797   0.1143   0.3694   1.7220

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.231      1.651  -1.957   0.0504
fert           3.811      1.687   2.259   0.0239
```

```
(Dispersion parameter for binomial
family taken to be 1)

Null deviance: 25.898  on 19  degrees of freedom
Residual deviance: 12.412 on 18  degrees of freedom
AIC: 16.412

Number of Fisher Scoring iterations: 6

> qchisq(0.95,18)
[1] 28.8693

> 1-pchisq(12.412,18)
[1] 0.8252536

> anova(fit.seed,test="Chisq")
Analysis of Deviance Table

Model: binomial, link: logit

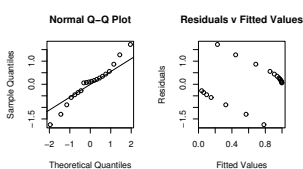
Response: x

Terms added sequentially (first to last)

              Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                                19    25.8979
```

```
fert  1  13.4861      18    12.4118    0.0002

> resid.check(fit.seed)
```



• La natura binaria della variabile risposta rende difficile l’analisi dei residui. È importante verificare che nessun residuo risulti particolarmente elevato.

```
• Nei casi di modelli binari può essere utile
verificare le capacità predittive del modello stimato:
si calcola il numero di casi in cui vi è un accordo tra
valori previsti e valori osservati.

> table(x,predict(fit.seed,type="response")>=0.5)

x    FALSE  TRUE
0     5     2
1     2    11
```

• In 5+11=16 casi si hanno conclusioni corrette, mentre in 2+2=4 casi si commette un errore nella previsione.

• Esistono in realtà modi più raffinati per verificare le capacità predittive del modello, tipo le tecniche di validazione incrociata (*cross validation*) e/o le curve ROC...

- Se si vuole stimare la probabilità, dando anche un intervallo di confidenza di livello approssimato 0.95, che un seme germogli con `fert=1`:

```
> predict(fit.seed,data.frame(fert=1),
type="response",se=T)
```

```
$fit
[1] 0.641119
```

```
$se.fit
[1] 0.1718472
```

```
$residual.scale
[1] 1
```

```
> 0.641119-1.96* 0.1718472
[1] 0.3042985
> 0.641119+1.96* 0.1718472
[1] 0.9779395
```

Esempio: Guasti di compressori

- In un fissato periodo di tempo, si è registrato il numero di guasti degli anelli di tenuta dei pistoni in ciascuno dei tre componenti di quattro compressori a vapore simili.
- Insieme di dati in cui la numerosità complessiva non è prefissata.

compressore	componente			totale
	A	B	C	
1	17	17	12	46
2	11	9	13	33
3	11	8	19	38
4	14	7	28	49
totale	53	41	72	166

6 Applicazioni alle tabelle di frequenza

Esempio: Russare e patologie cardiache

- In un’indagine compiuta su $n = 2484$ soggetti si sono rilevati il “russare notturno” (con 4 modalità) e la “presenza/assenza” di patologie cardiache.
- Insieme di dati in cui la numerosità complessiva n è prefissata.

patologie cardiache	russare notturno				totale
	mai	a volte	spesso	sempre	
presenti	24	35	21	30	110
assenti	1355	603	192	224	2374
totale	1379	638	213	254	2484

Esempio: Indagine di mercato

- In un’indagine compiuta sul livello di soddisfazione delle informazioni radiofoniche si sono rilevate tre variabili: Periodo, Razza dell’intervistato, Livello di soddisfazione.

periodo	razza	risposta		
		buono	sufficiente	scarso
A	bianco	81	23	5
	nero	325	253	54
B	bianco	224	144	24
	nero	600	636	158

6.1 Introduzione

- I dati da analizzare sono organizzati nella forma di tabelle di frequenza. Le osservazioni sono classificate in base alle modalità di una o più variabili (qualitative, ma anche quantitative).
- La “variabile osservata” sono quindi le frequenze nelle classi di una tabella di contingenza. La tabella riporta la frequenza assoluta di ciascuna classe.
- Consideriamo il caso bidimensionale con una tabella di frequenza di dimensione $I \times J$ con classi (celle) (i, j) . I dati si presentano nella forma

$y = (y_{11}, \dots, y_{IJ})$,

dove $y_{ij}, i = 1, \dots, I$ e $j = 1, \dots, J$, è il numero di unità statistiche classificate nella (i, j) –esima classe.

- Può essere di interesse valutare l’indipendenza delle due variabili rilevate.
- Si può usare il tradizionale test X^2 :

$$X^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(y_{ij} - \hat{y}_{ij})^2}{\hat{y}_{ij}},$$

ove \hat{y}_{ij} , per $i = 1, \dots, I$ e $j = 1, \dots, J$, è la frequenza attesa sotto l’ipotesi di indipendenza, espressa da

$$\hat{y}_{ij} = \frac{y_{i+} y_{+j}}{n},$$

con $y_{i+} = \sum_{j=1}^J y_{ij}, y_{+j} = \sum_{i=1}^I y_{ij}$ e $n = \sum_{i=1}^I \sum_{j=1}^J y_{ij}$.

- Sotto l’ipotesi nulla, la distribuzione approssimata di X^2 è χ^2 con $(I - 1)(J - 1)$ gradi di libertà.

Esempio: Russare e patologie cardiache

- In un’indagine compiuta su 2484 soggetti si sono rilevati il russare notturno e la presenza/assenza di patologie cardiache.

```
> xmat <- matrix(c(24,35,21,30,1355,603,192,224),
                 byrow=T,nrow=2)
> xmat
      [,1] [,2] [,3] [,4]
[1,]   24   35   21   30
[2,] 1355  603  192  224
> snore <- as.table(xmat)
> snore
      A      B      C      D
A    24    35    21    30
B  1355   603   192   224

> dimnames(snore)[[1]] <- c('si','no')
> dimnames(snore)[[2]] <- c('mai','avolte',
                             'spesso','sempre')
> snore
      mai avolte spesso sempre
si    24     35     21     30
no  1355    603    192    224
```

- Ad occhio, i dati sembrano suggerire una dipendenza tra il russare notturno e la presenza assenza di patologie cardiache. In effetti questo è confermato dal test X^2 di Person.

```
> prop.table(snore,2)
      mai avolte spesso sempre
si  0.017404 0.054859 0.098592 0.118110
no  0.982596 0.945141 0.901408 0.881890

> chisq.test(snore)

Pearson's Chi-squared test

data:  xmat
X-squared = 72.7821, df = 3, p-value = 1.082e-15
```


Un pò di intuito...

- Nell'esempio precedente sul russare e le patologie cardiache, nell'ipotesi di indipendenza, le frequenze sono:

$$\hat{y}_{ij} = \frac{y_{i+} y_{+j}}{n}.$$

Prendendo il logaritmo si ottiene

$$\begin{aligned} \log \hat{y}_{ij} &= \log y_{i+} + \log y_{+j} - \log n \\ &= \theta + \alpha_i + \beta_j, \end{aligned} \quad (11)$$

dando luogo ad una forma additiva nei parametri.

- Il parametro θ rappresenta l'effetto costante, e gli α_i e β_j sono gli effetti principali delle due variabili. Siccome valgono i vincoli sui totali, questo induce dei vincoli sui parametri, del tipo $\sum_{i=1}^I \alpha_i = 0 = \sum_{j=1}^J \beta_j$, lasciando in totale $1 + (I - 1) + (J - 1)$ parametri liberi.
- La scrittura (11) rivela una stretta analogia con la notazione dell'analisi della varianza a due criteri di classificazione.

- Nella pratica l'indipendenza si incontra in poche situazioni e per saggiare l'ipotesi di indipendenza abbiamo bisogno di un modello più generale di riferimento. Allora è ragionevole inserire un "termine di interazione" al modello di indipendenza, ossia

$$\log \hat{y}_{ij} = \theta + \alpha_i + \beta_j + \delta_{ij}, \quad (12)$$

con il vincolo $\sum_{i=1}^I \delta_{ij} = 0 = \sum_{j=1}^J \delta_{ij}$. Questa scrittura è la più generale per le tabelle di dimensione 2 ed ha $(I - 1)(J - 1)$ parametri in più di quello espresso dalla (11).

- Interpretando le \hat{y}_{ij} come frequenze attese, allora la (12) può essere "letta" come

$$\log(E(Y_{ij})) = \theta + \alpha_i + \beta_j + \delta_{ij}.$$

6.2 Schemi di campionamento

- Per l'inferenza si deve individuare il meccanismo probabilistico generatore dei dati che può essere di tipi diversi, a seconda del modo in cui i dati sono stati raccolti.
- I modelli adatti a descrivere dati di conteggio sono differenti a seconda dello schema di campionamento, della presenza o meno di variabili esplicative e della loro natura.
- I dati riassunti in una tabella di frequenza possono quindi essere generati da diversi schemi di campionamento. In dettaglio, la numerosità complessiva n può essere prefissata, oppure può ritenersi realizzazione di una variabile casuale.

- (a) *Osservazione diretta del fenomeno (Poisson)*. Per i dati di conteggio con numerosità non prefissata, noi osserviamo il fenomeno per un dato periodo e classifichiamo gli eventi. Allora, è ragionevole assumere che le y_{ij} della tabella siano realizzazioni di variabili casuali indipendenti con distribuzione di Poisson con media μ_{ij} , $i = 1, \dots, I$ e $j = 1, \dots, J$. Ossia

$$Y_{ij} \sim \text{Poisson}(\mu_{ij}),$$

perciò

$$Pr(\mathbf{Y} = \mathbf{y}) = \prod_{i=1}^I \prod_{j=1}^J \frac{e^{-\mu_{ij}} \mu_{ij}^{y_{ij}}}{y_{ij}!}, \quad (13)$$

con $\{\mu_{ij}\}$ parametri incogniti da stimare.

- (b) *Osservazione per un numero fissato di eventi (Multinomiale)*. Si decide preliminarmente di raccogliere i dati relativi a n unità, invece di fissare il tempo di osservazione del fenomeno. La distribuzione dei dati cambia. Per $n = \sum_{i=1}^I \sum_{j=1}^J y_{ij}$ fissato, il modello statistico appropriato è la distribuzione multinomiale $Mn_d(n, \pi)$, con $d = I \times J$, $\pi = (\pi_{11}, \dots, \pi_{IJ})$ e funzione di probabilità:

$$Pr(\mathbf{Y} = \mathbf{y}) = \frac{n!}{\prod_{i=1}^I \prod_{j=1}^J y_{ij}!} \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{y_{ij}}, \quad (14)$$

con $0 < \pi_{ij} < 1$ e $\sum_{i=1}^I \sum_{j=1}^J \pi_{ij} = 1$.

- (c) *Un totale di riga o di colonna prefissato (Prodotto-Multinomiale)*. Non si vincola solo il numero totale n di osservazioni, ma si impongono valori prefissati a tutta una riga o una colonna di frequenze marginali. Il modello statistico sarà allora il prodotto delle funzioni di probabilità multinomiali relative a ciascuna riga (o colonna).

- Esistono definizioni simili degli schemi di campionamento per tabelle con più di due dimensioni.

- *Osservazione*. Vi è un importante collegamento tra i modelli (13) e (14). Assunto il modello di Poisson, la distribuzione della statistica $n = \sum_{i=1}^I \sum_{j=1}^J y_{ij}$ è una Poisson con parametro $\mu = \sum_{i=1}^I \sum_{j=1}^J \mu_{ij}$. Allora la distribuzione di \mathbf{y} condizionata a n ha funzione di probabilità multinomiale $Mn_d(n, \pi)$, con $\pi_{ij} = \mu_{ij}/\mu$, $i = 1, \dots, I$ e $j = 1, \dots, J$, in accordo con il modello (14).

Dimostrazione.

$$\begin{aligned} Pr(\mathbf{Y} = \mathbf{y} | n) &= \frac{\prod_{i=1}^I \prod_{j=1}^J e^{-\mu_{ij}} \mu_{ij}^{y_{ij}} / y_{ij}!}{e^{(\sum_{i=1}^I \sum_{j=1}^J \mu_{ij})} (\sum_{i=1}^I \sum_{j=1}^J \mu_{ij})^n / n!} \\ &= \frac{n!}{\prod_{i=1}^I \prod_{j=1}^J y_{ij}!} \prod_{i=1}^I \prod_{j=1}^J \left(\frac{\mu_{ij}}{\mu} \right)^{y_{ij}}. \end{aligned}$$

- Guardando i dati è impossibile individuare il modello giusto: si deve sapere il modo in cui sono stati raccolti i dati. Infatti, sappiamo che a seconda del modo in cui i dati sono stati raccolti abbiamo forme diverse della verosimiglianza.

⇒ **Risultato fondamentale.** Si può mostrare che l'inferenza basata sulla verosimiglianza è essenzialmente la stessa in ciascun caso di raccolta dei dati. Ciò consente di stimare modelli con distribuzione dei dati di tipo multinomiale, usando la verosimiglianza di tipo Poisson.

- In particolare, i valori dei parametri che massimizzano le verosimiglianze (SMV) sono gli stessi, così come le derivate seconde delle log-verosimiglianze (e quindi anche gli errori standard delle stime coincidono).

- Anche per il caso di una marginale fissata, è possibile sviluppare considerazioni analoghe alle precedenti e constatare l'equivalenza delle verosimiglianze.

6.3 Modelli log-lineari

- Per il modello di Poisson, le frequenze attese nelle classi sono:

$$E(Y_{ij}) = \mu_{ij} .$$

Nell'ipotesi di indipendenza tra le due variabili della tabella, avremmo

$$\mu_{ij} = \mu_{i+} \mu_{+j} .$$

- Per il modello multinomiale le frequenze attese nelle classi sono:

$$E(Y_{ij}) = n\pi_{ij} .$$

Se le variabili fossero indipendenti, avremmo

$$\pi_{ij} = \pi_{i+} \pi_{+j} .$$

- Un'analogia argomentazione può essere sviluppata anche per il prodotto di multinomiali.

- Nelle tabelle di contingenza tutte le ipotesi usuali possono essere formulate come modelli moltiplicativi per le frequenze attese nelle classi.

- Obiettivo è lo studio della relazione tra $E(Y_{ij})$ e il predittore lineare η_{ij} , del tipo

$$\eta_{ij} = \theta + \alpha_i + \beta_j ,$$

ove gli α_i e β_j sono gli *effetti principali*.

- In ciascun caso, una struttura *moltiplicativa* per la media μ_{ij} fornisce una semplificazione naturale. Questo corrisponde a un modello additivo per $\log \mu_{ij}$. Perciò, quello che ci interessa è la possibilità che il *modello log-lineare*

$$\log \mu_{ij} = \theta + \alpha_i + \beta_j$$

(dove $\sum_{i=1}^I \alpha_i = 0 = \sum_{j=1}^J \beta_j$) fornisca una buona descrizione dei dati.

- Per questi modelli la funzione logaritmo è il legame naturale tra μ_{ij} e la combinazione lineare dei parametri. Da qui il nome *modello log-lineare*. Inoltre, la funzione logaritmo è il legame canonico per la distribuzione di Poisson. Si capisce quindi che i modelli log-lineari sono quelli più comuni per l'analisi delle tabelle di frequenza. Di fatto, i modelli log-lineari sono rilevanti anche negli altri schemi di campionamento.

- Il modello massimale può essere scritto come

$$\log \mu_{ij} = \theta + \alpha_i + \beta_j + \delta_{ij}$$

e quindi l'ipotesi di indipendenza corrisponde all'ipotesi di assenza di interazione, ossia $\delta_{ij} = 0$.

6.4 Modelli per tabelle con dimensione maggiore di due

- Nelle tabelle di dimensione $I \times J$, il termine di interazione conduce al modello saturo.
- Nelle tabelle di dimensione più elevata, i termini di interazione possono essere necessari per specificare il modello. Ad esempio, per un modello 3-dimensionale, si considera il modello log-lineare:

$$\log \mu_{ijk} = \theta + \alpha_i + \beta_j + \gamma_k + \delta_{ij} + \epsilon_{ik} + \phi_{jk} + \psi_{ijk}$$

con i vincoli richiesti, per esempio $\epsilon_{i+} = 0 \forall i$ e $\psi_{+jk} = 0 \forall j, k$. Questo modello è saturo, però un modello del tipo

$$\log \mu_{ijk} = \theta + \alpha_i + \beta_j + \gamma_k + \delta_{ij} + \epsilon_{ik}$$

assume una struttura di indipendenza nel modello.

6.5 Stima dei parametri

- L'inferenza può essere basata sul modello di Poisson in quanto l'inferenza basata sulla verosimiglianza è essenzialmente la stessa in ciascun caso di raccolta dei dati.
- Ciò consente di stimare modelli con distribuzione dei dati di tipo multinomiale, usando la verosimiglianza di tipo Poisson. In particolare, i valori dei parametri che massimizzano le due verosimiglianze (SMV) sono gli stessi, così come le derivate seconde delle log-verosimiglianze (e quindi anche gli errori standard delle stime coincidono).
- A condizione che i parametri che corrispondono alle frequenze marginali fissate siano sempre incluse nel modello, anche per il caso di una marginale fissata è possibile sviluppare considerazioni analoghe alle precedenti e constatare l'equivalenza delle verosimiglianze.

Esempio

Ad esempio, se avessimo una tabella $I \times J \times K$ per i tre fattori

Eta×Sesso×Fumare

e avessimo ottenuto le osservazioni campionarie separatamente per ogni combinazione Eta×Sesso, allora y_{ij+} sarebbe fisso per ciascun (i, j) , e dovremmo includere Eta*Sesso nel modello.

6.6 Specificazione in R

• La situazione di riferimento corrisponde a una distribuzione di Poisson per i dati e un modello log-lineare prevede un legame logaritmico per il previsore lineare.

• Dunque, questo è un modello GLM con (family=poisson(log)). Un esempio generico sarebbe

```
fit <- glm(Freq ~ Riga + Colonna,poisson)
```

dove Freq sono le frequenze osservate nella tabella, e Riga e Colonna sono i fattori che classificano righe e colonne della tabella.

• Per quanto detto sulle verosimiglianze, l’analisi è identica per le situazioni multinomiale e prodotto-multinomiale. È solo necessario che i totali del modello siano in accordo con i totali fissati per il disegno. La regola è che dobbiamo includere nel predittore lineare tutti i fattori e le interazioni che corrispondono ai totali marginali fissi.

```
> snore
      mai avolte spesso sempre
si   24      35      21      30
no 1355     603     192     224

> snore2 <- data.frame(snore)
> snore2
  Var1  Var2 Freq
1   si   mai   24
2  no   mai 1355
3   si avolte   35
4  no avolte  603
5   si spesso   21
6  no spesso  192
7   si sempre   30
8  no sempre  224

> names(snore2) <- c('malattie','russare','Freq')
> snore2
  malattie russare Freq
1        si      mai   24
2        no      mai 1355
3        si avolte   35
4        no avolte  603
5        si  spesso   21
6        no  spesso  192
7        si  sempre   30
8        no  sempre  224
```

```
> mod1 <- glm(Freq~.,poisson,data=snore)
> mod1
Call: glm(formula = Freq ~ ., family = poisson,
  data = snore)

Coefficients:
(Intercept)  Var1no Var2avolte Var2spesso Var2sempre
      4.112    3.072      -0.771      -1.868      -1.692

Degrees of Freedom: 7 Total (i.e. Null);  3 Residual
Null Deviance:      3930
Residual Deviance: 65.9  AIC: 128
> mod2 <- glm(Freq~russare+malattie,
family=poisson,data=snore2)
> mod2
Call: glm(formula = Freq ~ russare + malattie,
  family = poisson, data = snore2)

Coefficients:
(Intercept)  russareavolte russarespesso russaresempre
      4.112      -0.771      -1.868      -1.692
malattieno
      3.072

Degrees of Freedom: 7 Total (i.e. Null);  3 Residual
Null Deviance:      3930
Residual Deviance: 65.9  AIC: 128
```

6.7 La devianza per un modello Poisson log-lineare

• Conveniamo che le frequenze della tabella siano indicate con y_i , per $i = 1, \dots, d$, e d numero totale di classi.

• Abbiamo

$$D_i = 2 \left(\ell_i(\tilde{\theta}) - \ell_i(\hat{\theta}) \right) = 2 \{ y_i (\log y_i - \log \hat{\mu}_i) - y_i + \hat{\mu}_i \} ,$$

con $b(\theta) = \exp(\theta) = \mu$ e $\theta_i = \log \mu_i$. Dunque

$$D_i = 2 \left\{ y_i \log \frac{y_i}{\hat{\mu}_i} - y_i + \hat{\mu}_i \right\} .$$

Da uno sviluppo del termine logaritmico risulta

$$D_i \cong \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} .$$

- La statistica

$$\sum_i \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

è conosciuta come la statistica X^2 di Pearson, uno strumento familiare per condurre test di indipendenza in una tabella a due dimensioni.

- Abbiamo allora una nuova giustificazione per questa procedura: il test di Pearson è un'approssimazione del test del rapporto di verosimiglianza per la bontà di stima di un modello Poisson log-lineare.

Esempio: Indagine di mercato

```
> opinion
  Freq Colore Periodo Risposta
1    81      B      A    buono
2    23      B      A    suff
3     5      B      A    scar
4   325      N      A    buono
5   253      N      A    suff
6    54      N      A    scar
7   224      B      B    buono
8   144      B      B    suff
9    24      B      B    scar
10  600      N      B    buono
11  636      N      B    suff
12  158      N      B    scar

> attach(opinion)

> # modello saturo

> fit <- glm(Freq ~ Colore*Periodo*Risposta,poisson)
```

```
> anova(fit)

Analysis of Deviance Table
Model: poisson, link: log
Response: Freq
Terms added sequentially (first to last)

              Df Deviance Resid. Df Resid. Dev
NULL                                11      2319.50
Colore              1    986.38      10    1333.12
Periodo             1    445.39       9     887.73
Risposta            2    805.63       7      82.10
Colore:Periodo       1     18.06       6      64.04
Colore:Risposta      2     39.73       4      24.31
Periodo:Risposta     2     20.81       2       3.50
Colore:Periodo:Risposta 2       3.50       0 -6.797e-27

> qchisq(0.95,2)
[1] 5.991465
```

```
> fit1 <- glm(Freq ~ Colore+Periodo+Risposta+
Colore*Periodo+Colore*Risposta+Periodo*Risposta,
poisson)
> # oppure
> fit1 <- update(fit,~.-Colore:Periodo:Risposta)
> # oppure
> fit1 <- glm(Freq ~ .^2,poisson,data=opinion)

> anova(fit1)
Analysis of Deviance Table
Model: poisson, link: log
Response: Freq
Terms added sequentially (first to last)

              Df Deviance Resid. Df Resid. Dev
NULL                                11      2319.50
Colore              1    986.38      10    1333.12
Periodo             1    445.39       9     887.73
Risposta            2    805.63       7      82.10
Colore:Periodo       1     18.06       6      64.04
Colore:Risposta      2     39.73       4      24.31
Periodo:Risposta     2     20.81       2       3.50
```

```
> summary(fit1)

Call:
glm(formula = Freq ~ Colore + Periodo + Risposta +
Colore * Periodo + Colore * Risposta +
Periodo * Risposta, family = poisson)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      4.30613    0.10274  41.913 < 2e-16
ColoreN           1.49854    0.11006  13.616 < 2e-16
PeriodoB          1.13562    0.11271  10.076 < 2e-16
Rispostascar     -2.74189    0.23663 -11.587 < 2e-16
Rispostasuff     -0.90268    0.12218  -7.388 1.49e-13
ColoreN:PeriodoB -0.55484    0.11978  -4.632 3.62e-06
ColoreN:Rispostascar 0.93028    0.20958   4.439 9.05e-06
ColoreN:Rispostasuff 0.60306    0.10794   5.587 2.31e-08
PeriodoB:Rispostascar 0.48735    0.16270   2.995 0.00274
PeriodoB:Rispostasuff 0.38041    0.09362   4.064 4.83e-05
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 2319.4972  on 11  degrees of freedom
Residual deviance:    3.5002  on  2  degrees of freedom
AIC: 101.53

Number of Fisher Scoring iterations: 4
```

```
# utilizzo di una procedura "piu' automatica" --->
#####

> op.back <- glm(Freq~Colore*Periodo*Risposta,
family=poisson)

# sfoltiamo il modello (analisi backward):
#####

> drop1(op.back)
Single term deletions

Model:
Freq ~ Colore * Periodo * Risposta
                Df Deviance   AIC
<none>                -2.4e-15 102.0
Colore:Periodo:Risposta  2         3.5 101.5

> 1-pchisq(3.50,2)
[1] 0.1737739

# si puo' eliminare l'interazione di ordine tre
#####

> op.back1 <- update(op.back,
.~.-Colore:Periodo:Risposta)
```

```
# proviamo a togliere un'altra interazione
#####

> drop1(op.back1,test='Chisq')
Single term deletions

Model:
Freq ~ Colore + Periodo + Risposta + Colore:Periodo
+ Colore:Risposta + Periodo:Risposta
                Df Deviance   AIC   LRT Pr (Chi)
<none>                3.5 101.5
Colore:Periodo      1    26.4 122.4  22.9 1.7e-06 ***
Colore:Risposta     2    48.0 142.1  44.5 2.1e-10 ***
Periodo:Risposta    2    24.3 118.3  20.8 3.0e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

> 1-pchisq(24.3-3.5,2)
[1] 3.0432e-05

# stop
#####
```

Esempio: metrobus di Padova

```
> attach(Bus)
> Bus
      Freq Sesso Localita Opinione
1      9      F    Urban      Pro
2     35      F    Extra      Pro
3     12      F    Urban     Anti
4     10      F    Extra     Anti
5     25      M    Urban      Pro
6     60      M    Extra      Pro
7     20      M    Urban     Anti
8     15      M    Extra     Anti

# modello saturo
#####
> fit <- glm(Freq~Sesso*Localita*Opinione,poisson)
# proviamo a semplificare il modello
#####
> drop1(fit,test='Chisq')
Single term deletions

Model:
Freq ~ Sesso * Localita * Opinione
                Df Deviance   AIC   LRT Pr (Chi)
<none>                -3.8e-15 54.3
Sesso:Localita:Opinione  1         0.3 52.6  0.3    0.59
```

```
> fit1 <- update(fit,~.-Sesso:Localita:Opinione,
poisson)
> anova(fit1)
Analysis of Deviance Table

Model: poisson, link: log

Response: Freq

Terms added sequentially (first to last)

              Df Deviance Resid. Df Resid. Dev
NULL              7       77.0
Sesso             1       15.9
Localita          1       15.9
Opinione          1       28.6
Sesso:Localita    1        0.6
Sesso:Opinione    1        0.3
Localita:Opinione 1       15.4
```

```
# proviamo a semplificare ancora
#####

> drop1(fit1,test='Chisq')
Single term deletions

Model:
Freq ~ Sesso + Localita + Opinione + Sesso:Localita +
Sesso:Opinione + Localita:Opinione

              Df Deviance  AIC  LRT Pr(Chi)
<none>              0.3 52.6
Sesso:Localita      1      1.3 51.6  1.0   0.32
Sesso:Opinione      1      1.0 51.4  0.7   0.40
Localita:Opinione   1     15.7 66.0 15.4 8.9e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1
```

```
# togliamo l'interazione Sesso:Opinione
#####
> fit2 <- update(fit1,~.-Sesso:Opinione)
> fit2

Call:  glm(formula = Freq ~ Sesso + Localita +
Opinione + Sesso:Localita+Localita:Opinione,
family = poisson)

Coefficients:
(Intercept)                SessoM
          2.2380              0.5108
LocalitaUrban              OpinionePro
          0.0826              1.3350
SessoM:LocalitaUrban LocalitaUrban:OpinionePro
          0.2513             -1.2744

Degrees of Freedom: 7 Total (i.e. Null);  2 Residual
Null Deviance:      77
Residual Deviance: 1.01  AIC: 51.4

# semplifichiamo ancora...
#####
```

```
> drop1(fit2,test='Chisq')
Single term deletions

Model:
Freq ~ Sesso + Localita + Opinione + Sesso:Localita+
Localita:Opinione

              Df Deviance  AIC  LRT Pr(Chi)
<none>              1.0 51.4
Sesso:Localita      1      1.6 50.0  0.6 0.43657
Localita:Opinione   1     16.0 64.3 15.0 0.00011 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

# togliamo l'interazione Sesso:Localita
#####

fit3 <- update(fit2,~.-Sesso:Localita)

# ancora??
#####
```



```
> drop1(fit3,test='Chisq')
Single term deletions

Model:
Freq ~ Sesso + Localita + Opinione +
Localita:Opinione
              Df Deviance   AIC   LRT Pr(Chi)
<none>                1.6 50.0
Sesso             1      17.5 63.9 15.9 6.7e-05 ***
Localita:Opinione 1      16.6 62.9 15.0 0.00011 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

# non posso piu' semplificare: mi fermo
#####
```

```
# modello finale
#####
> summary(fit3)

Call:
glm(formula = Freq ~ Sesso + Localita + Opinione + Localita:Opinione,
    family = poisson)

Deviance Residuals:
    1     2     3     4     5     6     7     8 
-0.924  0.221  0.190  0.371  0.640 -0.165 -0.143 -0.285 

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)      2.183      0.223   9.78 < 2e-16 ***
SessoM           0.598      0.153   3.90 9.6e-05 ***
LocalitaUrban     0.247      0.267   0.92 0.35506
OpinionePro       1.335      0.225   5.94 2.9e-09 ***
LocalitaUrban:OpinionePro -1.274      0.333  -3.82 0.00013 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 77.0243  on 7  degrees of freedom
Residual deviance: 1.6151  on 3  degrees of freedom
AIC: 49.96

Number of Fisher Scoring iterations: 4
```

```
# tabella stimata
#####
> ftable(xtabs(fitted(fit3)~Opinione+Sesso+Localita))
              Localita  Extra  Urban
Opinione Sesso
Anti      F              8.871 11.355
          M             16.129 20.645
Pro       F             33.710 12.065
          M             61.290 21.935

# tabella osservata
#####

> ftable(xtabs(Freq~Opinione+Sesso+Localita))
              Localita  Extra  Urban
Opinione Sesso
Anti      F              10     12
          M             15     20
Pro       F             35      9
          M             60     25
```

```
# analisi in avanti (forward) --->
#####

> bus.for <- glm(Freq~Sesso+Localita+Opinione,
family=poisson)
> bus.for

Call:  glm(formula = Freq ~ Sesso + Localita +
Opinione, family = poisson)

Coefficients:
      (Intercept)      SessoM  LocalitaUrban  OpinionePro 
        2.569         0.598         -0.598          0.817 

Degrees of Freedom: 7 Total (i.e. Null);  4 Residual
Null Deviance:      77
Residual Deviance: 16.6  AIC: 62.9

> 1-pchisq(16.6,4)
[1] 0.002311206

#si puo' fare meglio...
#####
```

```
> add1(bus.for,~.^2,test='Chisq')
Single term additions

Model:
Freq ~ Sesso + Localita + Opinione
              Df Deviance   AIC   LRT Pr(Chi)
<none>                16.6 62.9
Sesso:Localita      1      16.0 64.3   0.6 0.43657
Sesso:Opinione      1      16.3 64.6   0.3 0.55667
Localita:Opinione   1       1.6 50.0 15.0 0.00011 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

# aggiungiamo l'interazione Localita:Opinione
#####
```

```
> bus.for1 <- glm(Freq~Sesso+Localita*Opinione,
family=poisson)
> bus.for1

Call:  glm(formula = Freq ~ Sesso +
Localita * Opinione, family = poisson)

Coefficients:
              (Intercept)                SessoM
                2.183                0.598
OpinionePro  LocalitaUrban:OpinionePro
                1.335                -1.274
Degrees of Freedom: 7 Total (i.e. Null);  3 Residual
Null Deviance:      77
Residual Deviance: 1.62   AIC: 50
# ancora?
#####
> add1(bus.for1,~.^2,test='Chisq')
Single term additions

Model:
Freq ~ Sesso + Localita * Opinione
              Df Deviance   AIC   LRT Pr(Chi)
<none>                1.6 50.0
Sesso:Localita      1      1.0 51.4   0.6   0.44
Sesso:Opinione      1      1.3 51.6   0.3   0.56
# ci possiamo fermare...
#####
```

Esempio: Uccelli e farfalle

- Nell’ambito di uno studio sulle capacità di apprendimento degli uccelli tropicali, un (crudele) sperimentatore ha raccolto dei dati sulla “risposta” di un *rufous-tailed jacamar* a otto specie di farfalle: *phrissa boisduvalli* (Ab), *Phoebis argante* (Pa), *Dryas iulia* (Di), *Pierella luna* (Pl), *Consul fabius* (Cf), *Siproeta stelenes* (Ss).
- In particolare, ha colorato con diversi colori (Unpainted, Brown, Yellow, Blue, Green, Red, Orange, Black) la parte inferiore delle ali della farfalla e ha poi inserito la farfalla nella gabbia dell’uccello.
- La risposta dell’uccello ha tre modalità: non attacco (N), assaggia ma non mangia (S), mangia (E).
- Evidentemente, i margini della tabella relativi alla specie e al colore sono determinati dallo sperimentatore: quindi ogni modello dovrà contenere l’interazione `species:colour`.

```
> jacamar.tab <- xtabs(Freq~fate+species+colour,data=jacamar2)
> ftable(jacamar.tab)
              colour Unpainted Brown Yellow Blue Green Red Orange Black
fate species
N   Ab              0      7      7      6      3      4      4      4
   Pa              6      2      4      0      1      0      6      0
   Di              1      1      5      0      5      6      4      1
   Pl              4      2      2      4      6      4      7      4
   Cf              0      0      0      0      0      0      0      7
   Ss              0      0      0      0      0      3      1      0
S   Ab              0      1      2      0      0      0      2      0
   Pa              1      1      0      0      1      0      0      0
   Di              0      0      0      0      0      0      1      0
   Pl              1      2      0      0      0      0      0      2
   Cf              0      0      0      0      0      0      0      1
   Ss              0      0      0      1      0      0      1      1
E   Ab             14      2      1      0      1      0      0      0
   Pa              0      0      2      0      0      0      0      0
   Di              2      1      1      1      0      0      1      1
   Pl              5      4      5      3      2      2      1      2
   Cf              0      3      1      1      1      1      2      0
   Ss              1      1      3      1      3      1      1      0

> margin.table(jacamar.tab,c(2,3))
              colour
species Unpainted Brown Yellow Blue Green Red Orange Black
Ab      14      10      10      6      4      4      6      4
Pa       7       3       6      0      2      0      6      0
Di       3       2       6      1      5      6      6      2
Pl      10       8       7      7      8      6      8      8
Cf       0       3       1      1      1      1      2      8
Ss       1       1       3      2      3      4      3      1
```

```
# modello saturo
#####
> jac1 <- glm(Freq~colour*species*fate,data=jacamar2,
family=poisson)

# proviamo a eliminare l'interazione a tre
#####
> drop1(jac1,test='Chisq')
Single term deletions

Model:
Freq ~ colour * species * fate
             Df Deviance AIC LRT Pr(Chi)
<none>             5.1e-10 491
colour:species:fate 70          91 442  91   0.049 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1
Warning message:
In glm.fit(x[, jj, drop = FALSE], y, wt, offset = obj) :
  fitted rates numerically 0 occurred

# volendo si puo' togliere l'interazione di ordine 3
# (ma anche no...)
#####
```

```
> jac2 <- update(jac1,~.-colour:species:fate)
> l-pchisq(jac2$deviance,jac2$df.resid)
[1] 0.049068
> ftable(xtabs(round(fitted(jac2))^fate+species+colour,data=jacamar2))
      colour Unpainted Brown Yellow Blue Green Red Orange Black
fate species
N   Ab      4      5      6      5      3      4      5      3
   Pa      5      2      5      0      2      0      5      0
   Di      1      1      4      1      4      6      5      2
   Pl      1      3      3      4      5      5      6      6
   Cf      0      0      0      0      0      1      1      5
   Ss      0      0      0      0      1      2      1      0
S   Ab      1      2      1      0      0      0      1      1
   Pa      1      1      1      0      0      0      1      0
   Di      0      0      0      0      0      0      0      0
   Pl      0      1      0      0      0      0      1      2
   Cf      0      0      0      0      0      0      0      1
   Ss      0      0      0      0      0      0      1      1
E   Ab      9      3      3      1      1      0      0      0
   Pa      1      0      0      0      0      0      0      0
   Di      2      1      2      0      1      0      1      0
   Pl      8      4      4      3      3      1      1      0
   Cf      0      3      1      1      1      0      1      2
   Ss      1      1      2      1      2      2      1      0
```

- Esiste sicuremante una dipendenza del “destino” della farfalla dalla specie e dal colore. C’è anche una certa evidenza di un effetto di interazione tra specie e colore nel determinare il destino della farfalla.
- Ci sono zeri strutturali nei dati (bisognerebbe tenerne conto...).

7 Quasi-verosimiglianza

- Nell’ambito del LM classico il criterio dei minimi quadrati permette di stimare i parametri di regressione senza specificare un vero e proprio modello probabilistico.
- Il criterio dei minimi quadrati richiede la specificazione della relazione tra valore medio della risposta, Y_i , e il predittore lineare, η_i , e la separazione tra valor medio e varianza dell’errore (che non è legata al valor medio):

$$E(Y_i) = \mu_i = \eta_i$$
$$Var(Y_i) = \sigma^2$$

- Anche nell’ambito dei GLM si può seguire questa impostazione, introducendo però l’eventuale relazione tra media e varianza: si deve specificare la funzione di varianza $V(\mu_i)$.
- Infatti, le equazioni di verosimiglianza per β (che non dipendono da ϕ)

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

continuano ad essere delle “buone” equazioni di stima purchè sia $E(Y_i) = \mu_i = g^{-1}(\eta_i)$.

- In altre parole, l'assunzione parametrica

$$Y_i \sim EF(\cdot, \phi)$$

potrebbe anche non essere soddisfatta. Quello che è importante è l'ipotesi sulla media:

$$\mu_i = E(Y_i) = g^{-1}(\eta_i).$$

- Si noti inoltre che l'unica ulteriore caratteristica della distribuzione necessaria per esplicitare l'equazione di stima è la funzione di varianza $V(\mu)$.
- Se la media e la funzione di varianza sono correttamente specificate, lo stimatore $\hat{\beta}$ ottenuto dalle equazioni di stima (standard) appena descritte avrà l'usuale distribuzione approssimata normale, solo con varianza più grande.

7.1 Modello di quasi-verosimiglianza

- Sotto condizioni di regolarità, le equazioni di verosimiglianza per un GLM producono stime per i coefficienti β che mantengono le loro proprietà anche se l'ipotesi che le osservazioni Y_i provengano da una famiglia esponenziale è sostituita dalle più deboli ipotesi sui momenti sino al secondo ordine (*assunzioni del secondo ordine*):

1. $g(\mu_i) = g(E(Y_i)) = \eta_i, \quad i = 1, \dots, n,$
2. $Var(Y_i) = \phi V(\mu_i), \quad i = 1, \dots, n,$
3. Y_i indipendente da Y_j , se $i \neq j$.

- Il modello statistico specificato dalle assunzioni 1–3 è detto *modello di quasi-verosimiglianza*.
- Se $V(\mu) = 1$ e $g(\mu) = \mu$, le ipotesi 1–3 coincidono con le usuali ipotesi del secondo ordine del LM classico.

- Osservazione: le equazioni di quasi-verosimiglianza consentono di definire una *funzione di quasi-verosimiglianza* per β della forma

$$\ell_Q(\beta) = \sum_{i=1}^n \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi V(t)} dt.$$

- La funzione $\ell_Q(\beta)$ non è in generale una log-verosimiglianza, anche se possiede molte delle sue proprietà formali (lo SMV per β ha distribuzione asintotica normale, il TRV basato su $\ell_Q(\beta)$ ha distribuzione approssimata chi-quadrato).
- È possibile definire anche la devianza

$$D_Q(y; \hat{\beta}) = -2\phi \ell_Q(\hat{\beta})$$

e la devianza scalata $D_Q^*(y; \hat{\beta}) = D_Q(y; \hat{\beta})/\phi$.

- Per confrontare modelli annidati, in modo analogo a quanto fatto nei GLM, si può utilizzare la devianza scalata confrontandola con l'usuale distribuzione asintotica nulla: $\chi_{p-p_0}^2$ o $F_{p-p_0; n-p}$ rispettivamente per ϕ noto o stimato (in quest'ultimo caso si usa $D_Q(y; \hat{\beta})/\{\hat{\phi}(p - p_0)\}$).

Perché la quasi-verosimiglianza?

- Il modello di quasi-verosimiglianza è adatto a trattare tanto dati continui quanto dati discreti. In particolare, per dati discreti che rappresentano conteggi o proporzioni e per dati continui che descrivono tempi di attesa, le ipotesi del secondo ordine 1–3 comportano un incremento di flessibilità rispetto alle specificazioni parametriche usuali del GLM con distribuzione Poisson, binomiale e esponenziale.
- Nella pratica vi sono applicazioni in cui il parametro di dispersione si trova in disaccordo con il modello ipotizzato.
- Ad esempio, mentre le famiglie binomiale e Poisson hanno parametro di dispersione ϕ noto ed uguale ad 1, i dati suggeriscono un valore $\phi > 1$.

- Questo è un problema diffuso nell'analisi di dati discreti. Si parla di *sovradisersione* (*overdispersion*) quando la varianza della variabile Y è maggiore di quella teorica. Ad esempio, si ritiene che $\text{var}(Y) = \phi n\pi(1 - \pi) > n\pi(1 - \pi)$, con $\phi > 1$, ove $n\pi(1 - \pi)$ è la varianza di una variabile binomiale.
- In conclusione, il metodo della quasi-verosimiglianza consente quindi di affrontare i *problemi di sovradisersione*: si può infatti specificare $\text{var}(Y_i)$ in modo tale da consentire una maggiore variabilità rispetto a quella imposta dalla famiglia esponenziale di riferimento.
- Dal punto di vista tecnico, la quasi-verosimiglianza permette di modellare anche casi di *sottodispersione* ($\phi < 1$), anche se questi sono piuttosto rari.

7.2 Inferenza

- La procedura di stima dei coefficienti β non dipende da ϕ : quindi $\hat{\beta}$ sarà lo stesso del corrispondente GLM (purché con stessa funzione legame e stessa funzione di varianza).
- Invece variano gli errori standard, essendo la matrice di varianze e covarianze asintotica proporzionale a ϕ , che in questo caso va stimato.
- Una stima per ϕ è (l'usuale):

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

ove

$$\hat{\mu}_i = g^{-1}(\mathbf{x}_i^T \hat{\beta})$$

sono i valori di μ_i stimati. Questi vengono calcolati con la stima $\hat{\beta}$.

- Anche sotto le ipotesi più deboli 1–3:
 1. Continua a valere l'identità dell'informazione: $E(\ell_* \ell_*^T) = -E(\ell_{**})$ (argomentazione chiave delle proprietà asintotiche dello stimatore di massima verosimiglianza);
 2. Si mantiene la consistenza di $\hat{\beta}$;
 3. Continua a valere l'approssimazione

$$\hat{\beta} \sim N_p(\beta, \hat{\phi}(X^T \widehat{W} X)^{-1}),$$

per n elevato. Se n non è elevato può essere preferibile usare l'approssimazione t_{n-p} ;

4. Lo stimatore $\hat{\phi}$ è appropriato per stimare ϕ .
- Si può anche definire l'analogo della devianza, in termini di quasi-devianza (attenzione però a riscalare: vedi pagina 207).

7.3 I comandi in R

- In R il metodo della quasi-verosimiglianza viene selezionato con l'opzione `quasi` della funzione `glm` che sostituisce il nome della famiglia esponenziale.

- La sintassi è:

```
glm(formula, family=quasi(link=legame,
                             variance="funzionedivarianza"))
```

- I nomi delle funzioni di varianza disponibili sono:

<code>constant</code>	per 1;
<code>mu</code>	per μ ; \Rightarrow quasi-poisson;
<code>mu^2</code>	per μ^2 ;
<code>mu^3</code>	per μ^3 ;
<code>mu(1-mu)</code>	per $\mu(1 - \mu)$; \Rightarrow quasi-binomial.

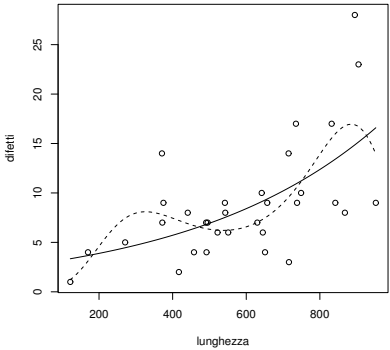
Esempio: Quasi-Poisson

- Per un modello di Poisson, caratterizzato da $\phi = 1$, si può tener conto di una eventuale sovradisersione tramite

$$var(Y_i) = \phi \mu_i, \quad \phi > 1.$$

- La procedura di stima dei coefficienti β non dipende da ϕ ; conseguentemente le stime non cambiano rispetto al caso $\phi = 1$.
- Invece variano gli errori standard, essendo la matrice di varianze e covarianze asintotica proporzionale a ϕ .
- Se ϕ non è noto, una sua stima è $\hat{\phi}$.
- Gli errori standard sono quindi pari a $\sqrt{\hat{\phi}}$ volte quelli del GLM Poisson.

- **Esempio:**
 DD = numero di difetti per rotolo di stoffa
 LL = lunghezza (in mt) del rotolo di stoffa



- Modello:
1. $DD_i \sim Poisson(\mu_i)$, con $\mu_i = E(DD_i)$, $i = 1, \dots, 32$;
 2. $\eta_i = \beta_0 + \beta_1 * LL_i$;
 3. $\log(\mu_i) = \eta_i \Leftrightarrow \mu_i = \exp(\eta_i)$.

```
# stima del modello di regressione di Poisson

> fit.ll <- glm(DD ~ LL,family=poisson)
> summary(fit.ll)

Call:
glm(formula = DD ~ LL, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.741  -1.133  -0.039   0.662   3.074

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.971751   0.212469   4.57  4.8e-06 ***
LL           0.001930   0.000306   6.30  3.0e-10 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 103.714  on 31  degrees of freedom
Residual deviance:  61.758  on 30  degrees of freedom
AIC: 189.1

Number of Fisher Scoring iterations: 4

# test sulla devianza
> 1-pchisq(61.758,30)
[1] 0.0005610522
```

```
# ampliamento della formula del modello

> fit.lll <- glm(DD ~ LL+I(LL^2),family=poisson)

> anova(fit.lll,test='Chisq')
Analysis of Deviance Table

Model: poisson, link: log
Response: DD

Terms added sequentially (first to last)

            Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                31      103.7
LL                   1      42.0    30      61.8  9.3e-11
I(LL^2)              1       0.5    29      61.3    0.5
```

```
# ulteriore ampliamento della formula del modello

> fit.llp <- glm(DD ~ LL+I(LL^2)+I(LL^3)+I(LL^4),
family=poisson)
> anova(fit.llp,test='Chisq')
Analysis of Deviance Table

Model: poisson, link: log

Response: DD

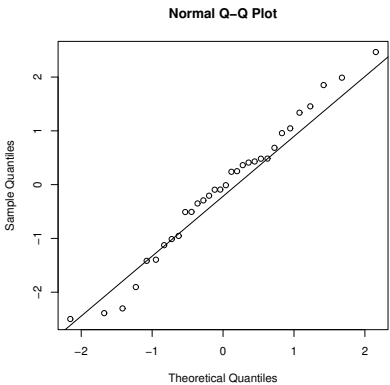
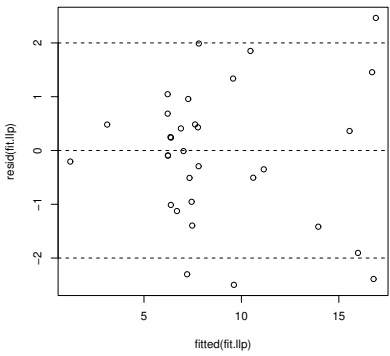
Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                31      103.7
LL                1      42.0      30      61.8  9.3e-11
I(LL^2)          1       0.5      29      61.3    0.5
I(LL^3)          1       1.5      28      59.8    0.2
I(LL^4)          1      10.1      27      49.7  1.5e-03

> 1-pchisq(49.7,27)
[1] 0.0049277
# anche i residui mostrano una maggiore variabilita'
# rispetto a quanto ci si aspetterebbe se il modello fosse
# correttamente specificato

> plot(fitted(fit.llp),resid(fit.llp))
> abline(h=c(-2,0,2),lty='dashed')
> qqnorm(resid(fit.llp)); qqline(resid(fit.llp))
```

217



218

```
# modello di quasi-verosimiglianza
> fitqs <- glm(DD ~ LL,family=quasi(link=log,variance="mu"))
# oppure
> fitqs <- glm(DD ~ LL,family=quasipoisson)
> summary(fitqs)

Call:
glm(formula = DD ~ LL, family = quasipoisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.741  -1.133  -0.039   0.662   3.074

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.971751   0.309503   3.14  0.00378 **
LL           0.001930   0.000446   4.33  0.00016 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for quasipoisson family taken
to be 2.1220)

Null deviance: 103.714  on 31  degrees of freedom
Residual deviance:  61.758  on 30  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4
```

219

```
# ATTENZIONE: La riduzione di devianza non e'
# direttamente interpretabile...

> anova(fitqs,test='F')
Analysis of Deviance Table

Model: quasipoisson, link: log

Response: DD

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev    F Pr(>F)
NULL                31      103.7
LL      1      42.0      30      61.8 19.8 0.00011 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

> F <- (fitqs$null.deviance-fitqs$deviance)/
summary(fitqs)$dispersion
> F
[1] 19.772
> 1- pf(F,1,fitqs$df.res)
[1] 0.00011055
```

220

```
> fitqs2 <- glm(DD ~ LL+I(LL^2)+I(LL^3)+I(LL^4),
family=quasipoisson)
> anova(fitqs2,test='F')
Analysis of Deviance Table

Model: quasipoisson, link: log

Response: DD

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev    F Pr(>F)
NULL                      31      103.7
LL              1       42.0       30      61.8 23.87 4.1e-05 ***
I(LL^2)         1        0.5       29      61.3  0.28  0.599
I(LL^3)         1        1.5       28      59.8  0.83  0.369
I(LL^4)         1       10.1       27      49.7  5.72  0.024 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1
```

```
> summary(fitqs2)

Call:
glm(formula = DD ~ LL + I(LL^2) + I(LL^3) + I(LL^4),
family = quasipoisson)

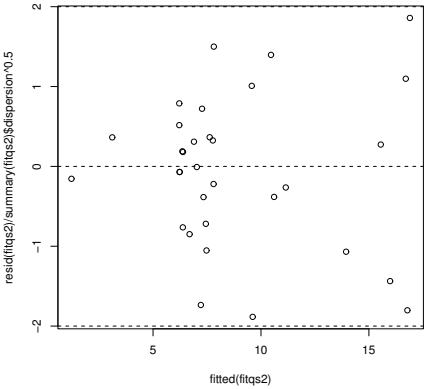
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4985  -0.9682  -0.0515   0.5343   2.4650

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.62e+00  2.75e+00  -1.68   0.104
LL           5.77e-02  2.41e-02   2.40   0.024 *
I(LL^2)      -1.75e-04  7.31e-05  -2.39   0.024 *
I(LL^3)       2.17e-07  9.12e-08   2.38   0.025 *
I(LL^4)      -9.29e-11  4.00e-11  -2.32   0.028 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

(Dispersion parameter for quasipoisson family taken
to be 1.7578)

Null deviance: 103.714  on 31  degrees of freedom
Residual deviance:  49.736  on 27  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4
# residui...
> plot(fitted(fitqs2),
resid(fitqs2)/summary(fitqs2)$dispersion^0.5)
```

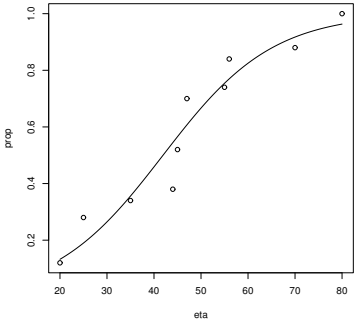


Esempio

```
# Su un campione di persone di eta' diverse e' stato contato
# il numero di individui ciechi. I dati sono il numero di
# individui ciechi (nc), l'eta' (eta) e il numero di
# individui osservati per eta' (ni).

> nc <- c(6,14,17,19,26,35,37,42,44,50)
> ni <- c(50,50,50,50,50,50,50,50,50,50)
> eta <- c(20,25,35,44,45,47,55,56,70,80)

> prop <- nc/ni
> plot(eta,prop)
```




```
# Stima di un modello di regressione logistica:
> fit1 <- glm(cbind(nc,ni-nc)~eta,binomial)

> summary(fit1)

Call:
glm(formula = cbind(nc, ni - nc) ~ eta, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3331  -0.5521  -0.2336   1.3332   1.9363

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.59434    0.38557  -9.322  <2e-16 ***
eta           0.08574    0.00834  10.281  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

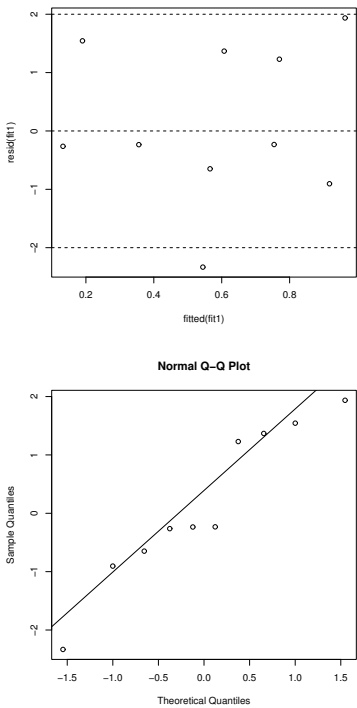
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 185.508  on 9  degrees of freedom
Residual deviance:  16.377  on 8  degrees of freedom
AIC: 56.635

Number of Fisher Scoring iterations: 4

# test sulla devianza e analisi dei residui
> 1-pchisq(16.38,8)
[1] 0.03725306
> plot(fitted(fit1),resid(fit1))
> qqnorm(resid(fit1))
> qqline(resid(fit1))
```

225



226

```
# ampliare il modello in modo polinomiale
# non migliora l'adattamento (provate da soli...)

# modello di quasi-verosimiglianza: stimiamo attraverso
# i dati il parametro di dispersione

> fitq <- glm(cbind(nc,ni-nc)~eta,quasibinomial)
> summary(fitq)

Call:
glm(formula = cbind(nc, ni - nc) ~ eta,
family = quasibinomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.333  -0.552  -0.234   1.333   1.936

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.5943    0.5230  -6.87  0.00013 ***
eta           0.0857    0.0113   7.58  6.4e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for quasibinomial family taken
to be 1.8402)

    Null deviance: 185.508  on 9  degrees of freedom
Residual deviance:  16.377  on 8  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4
```

227

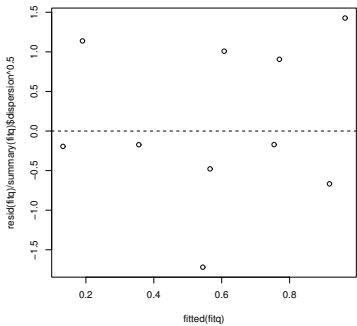
```
# sembra esserci una maggiore variabilita' rispetto a quella
# dal modello binomiale

# la devianza residua riscalata sembra ragionevole
# rispetto ai gradi di liberta'

> 16.37/1.84
[1] 8.89674

# il modello sembra essere piu' adeguato, se si tiene conto
# della sovradisersione

> plot(fitted(fitq),resid(fitq)/summary(fitq)$dispersion^0.5)
```



228