UNIVERSIDADE DE BRASÍLIA
DEPARTAMENTO DE ESTATÍSTICA

Disciplina: Modelos Lineares Generalizados
Curso: Graduação em Estatística
Código: EST0011
Semestre: 2023.2
Prof. Frederico Machado Almeida

Lista de Exercícios #4

Observações:

- Questões para entregar: 1, 2(b,c), 3 e 4.
- Demais questões são apenas para estudar.
- Prazo de entrega: **17/11/2023**

Q01. Suponha que $Y_i \sim exp\,(\phi)$, para $i = 1, \cdots, 5$. Isto é,

$$f(y_i|\phi) = \frac{1}{\phi}\exp\left(-\frac{y_i}{\phi}\right), \text{ com } y_i > 0 \text{ e } \phi > 0.$$

Suponha que o vetor observado foi $\mathbf{y} = (2.8, 3.5, 2.4, 1.9, 3.0)^{\top}$ com média $\bar{y} = 2.72$.

(a) Implemente manualmente o algoritmo de Newton-Raphson e itere até obter a convergência. Feito isso, apresente uma tabela contendo as iterações, as estimativas (para cada iteração), e a tolerância.

(b) Como o valor obtido se compara com o estimador de máxima verossimilhança? Isto é, $\hat{\phi} = \bar{y}$?

Q02. Assuma que $y_1, y_2, \cdots, y_n$ é uma amostra observada de Y. Para cada distribuição abaixo, apresente a função desvio escalonado total.

(a) Distribuição Weibull: $f(y_i|\alpha,\beta) = \frac{\alpha}{\beta}\left(\frac{y_i}{\beta}\right)^{\alpha-1}\exp\left[-\left(\frac{y_i}{\beta}\right)^{\alpha}\right]$, com $y_i > 0$; $\alpha, \beta > 0$.

(b) Distribuição Binomial: $Y_i \sim$ Binom (n_i, p_i) $P(Y_i = y_i|\alpha,\beta) = \begin{pmatrix} n_i \\ y_i \end{pmatrix} p_i^{y_i}\left(1-p_i\right)^{n_i-y_i}$, com $y_i = 0, 1, 2, \cdots, n_i$ e $0 < p_i < 1$.

(c) Distribuição Beta: $f(y_i|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y_i^{\alpha-1}\left(1-y_i\right)^{\beta-1}$, com $0 < y_i < 1$, e $\Gamma(\cdot)$ denota a função gama.

Q03. Considere a função gama apresentada na questão anterior. Conforme mostrado na `Lista 3`, a distribuição gama é membro da família exponencial com ligação canônica. Usando a função de ligação logarítmica (frequentemente usada como função de ligação canônica para a distribuição Poisson), obtenha:

(a) A função escore $\mathrm{U}\left(\beta_j\right)$ e as entradas genéricas (j, s) da matriz de informação de Fisher, digamos $\mathcal{I}_{js}\left(\boldsymbol{\beta}\right)$.

(b) A matriz dos pesos $W(\boldsymbol{\beta})$ e a matriz $G(\boldsymbol{\beta}) = \text{diagonal}\left\{g'(\mu_i), \text{ com } i = 1, \cdots, n\right\}$.

(c) A variável dependente ajustada na t-ésima iteração do Fisher Scoring, i.e., $z_i^{(t)}$.

Q04. Os dados apresentados na Tabela são referentes ao tempo até a morte y_i (em semanas) de dezassete pacientes que sofrem de leucemia, e x_i denota a contagem inicial dos glóbulos brancos, em uma escala de $\log_{10}$.

i	y_i	x_i	i	y_i	x_i
1	65	3,36	10	143	3,85
2	156	2,88	11	56	3,97
3	100	3,63	12	26	4,51
4	134	3,41	13	22	4,54
5	16	3,78	14	01	5,00
6	108	4,02	15	01	5,00
7	121	4,00	16	05	4,72
8	04	4,23	17	65	5,00
9	39	3,73			

(a) Obtenha o gráfico de dispersão de y_i contra x_i. O diagrama apresenta alguma tendência?

(b) Uma possível especificação para $\mu_i = \mathbb{E}(Y_i)$ é,

$$\mathbb{E}(Y_i) = \exp(\beta_0 + \beta_1 x_i). \tag{1}$$

Como pode ser visto, esta especificação permite que $\mathbb{E}(Y_i)$ seja sempre positivo para todo β_j, com $j = 0, 1$ e todo x_i. Qual função de ligação seria apropriada para este caso? Justifique a sua resposta.

(c) A distribuição Exponencial tem sido frequentemente usada para modelar tempos de sobrevivência. Sua fdp é dada por: $f(y_i; \mu) = \frac{1}{\mu}\exp\left(-\frac{y_i}{\mu}\right)$. Observe que, esta distribuição é um caso particular da distribuição Gama (veja Q01-(e), Lista 3), com $\nu = 1$. Mostre que $\mathbb{E}(Y_i) = \mu_i$ e $Var(Y_i) = \mu_i^2$.

(d) Usando algum software estatístico, ajuste o modelo de regressão, assumindo que os dados seguem uma distribuição Exponencial, com $\mathbb{E}(Y_i)$ dada segundo a equação (1).

(e) Para o modelo obtido em (d): (i) compare os valores observados: y_i e os estimados $\hat{y}_i = \exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)$; (ii) use os resíduos padronizados $r_i = \frac{y_i - \hat{y}_i}{\hat{y}_i}$ para avaliar a qualidade de ajuste do modelo.

(f) Interprete os valores dos coeficientes de regressão, $\hat{\beta}_0$ e $\hat{\beta}_1$.

Q05. Sabe-se que a concentração de colesterol no soro sanguíneo aumenta com a idade, mas é menos claro se o nível de colesterol também está associado ao peso corporal. As variáveis de interesse são: o nível de colesterol sérico de trinta e dois pacientes (milhões de moles por litro), idade (anos) e o índice de massa corporal (peso dividido pela altura ao quadrado, onde o peso foi medido em quilogramas e a altura em metros).
Com base no banco de dados descrito, foi ajustado o modelo de regressão Gama, com função de ligação inversa (`por default`). Toda informação sobre o ajuste do modelo está apresentada no Apêndice.

Q02. Assuma que $y_1, y_2, \cdots, y_n$ é uma amostra observada de Y. Para cada distribuição abaixo, apresente a função desvio escalonado total.

(b) Distribuição Binomial: $Y_i \sim$ Binom(n_i, p_i) $P(Y_i = y_i|\alpha,\beta) = \binom{n_i}{y_i} p_i^{y_i}(1-p_i)^{n_i-y_i}$, com $y_i = 0, 1, 2, \cdots, n_i$ e $0 < p_i < 1$.

(c) Distribuição Beta: $f(y_i|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y_i^{\alpha-1}(1-y_i)^{\beta-1}$, com $0 < y_i < 1$, e $\Gamma(\cdot)$ denota a função gama.

(b) Distribuição Binomial: $Y_i \sim$ Binom(n_i, p_i) $P(Y_i = y_i|\alpha,\beta) = \binom{n_i}{y_i} p_i^{y_i}(1-p_i)^{n_i-y_i}$, com $y_i = 0, 1, 2, \cdots, n_i$ e $0 < p_i < 1$.

Definindo a log-verossimilhança:

$$\mathcal{L}(\beta; y) = \sum_{i=1}^{N}\left[y_i \log p_i - y_i \log(1-p_i) + n_i \log(1-p_i) + \log\binom{n_i}{y_i}\right]$$

Como para o modelo saturado temos que todos os p_i's são distintos e o MLE é $\hat{p}_i = y_i/n_i$, o valor máximo da log-verossimilhança será:

$$\mathcal{L}(b_{max}; y) = \sum_{i=1}^{N}\left[y_i \log(y_i/n_i) - y_i \log\left(\frac{n_i - y_i}{n_i}\right) + n_i \log\left(\frac{n_i - y_i}{n_i}\right) + \log\binom{n_i}{p_i}\right]$$

$\forall$ $p < N$ parâmetros, seja $\hat{p}_i$ os MLE's das probabilidades estimadas e $\hat{y}_i = n_i \hat{p}_i$ os valores ajustados. Logo, a log-verossimilhança avaliada nesses valores será:

$$\mathcal{L}(b; y) = \sum_{i=1}^{N}\left[y_i \log(\hat{y}_i/n_i) - y_i \log\left(\frac{n_i - \hat{y}_i}{n_i}\right) + n_i \log\left(\frac{n_i - \hat{y}_i}{n_i}\right) + \log\binom{n_i}{p_i}\right]$$

Logo a função desvio será: $D = 2[\mathcal{L}(b_{max}, y) - \mathcal{L}(b, y)]$

$$D = \sum_{i=1}^{N}\left[y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i)\log\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right)\right]$$

Como o desvio escalonado é dado como função do desvio:

$D^* = \phi \cdot D$; onde $\phi = \hat{\phi} = 1$ para o modelo binomial, temos que $D^* = D = \sum_{i=1}^{N} \left[y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i) \log\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right) \right]$

(c) Distribuição Beta: $f(y_i|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y_i^{\alpha-1}(1-y_i)^{\beta-1}$, com $0 < y_i < 1$, e $\Gamma(\cdot)$ denota a função gama.

reparametrizando utilizando a média e a dispersão: $\mu = \frac{\alpha}{\alpha+\beta}$, $\phi = \alpha + \beta$

$\longrightarrow \mu = \frac{\alpha}{\phi} \longrightarrow \alpha = \mu \cdot \phi$; $\beta = \phi - \alpha = \phi(1-\mu)$, temos

que $E(y) = \frac{\alpha}{\alpha+\beta} = \mu$; $Var(\mu) = \frac{V(\mu)}{1+\phi}$; logo: $\phi = \frac{\mu_t(1-\mu_t) - 1}{Var(y_t)}$

onde $V(\mu) = \mu(1-\mu)$, onde μ é a resposta média e ϕ pode ser interpretado como o parâmetro de precisão no sentido de que, para μ fixo, $\phi \uparrow \rightarrow Var(y) \downarrow$. Reescrevemos a densidade como:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\,\Gamma((1-\mu)\phi)} \, y^{\mu\phi - 1} (1-y)^{(1-\mu)\phi - 1} \quad , \; 0 < y < 1$$

estabelecendo μ_t e ϕ desconhecida o modelo será obtido assumindo que y_t é escrito da forma:

$g(\mu_t) = \sum_{i=1}^{k} x_{ti} \beta_i = \eta_t$, onde $(\beta_1, \ldots, \beta_k)^T$ é o vetor de regressores desconhecidos $(\beta \in \mathbb{R}^k)$ e $(x_{t1}, \ldots, x_{tk})$ as observações em k covariáveis $(k < n)$

definindo a linking function $g(\cdot)$ como a logit-link, temos:

$$\mu_t = \frac{e^{x_t^T \beta}}{1 + e^{x_t^T \beta}} \quad ; \text{ onde } x_t^T = (x_{t1}, \ldots, x_{tk}) \ ; \ t = 1, \ldots, n$$

A log-verossimilhança será:

$$\mathcal{L}(\beta, \phi) = \sum_{i=1}^{n} \mathcal{L}_t(\mu_t, \phi)$$

equação 1

onde
$$\mathcal{L}_t(\mu_t, \phi) = \log \Gamma(\phi) - \log \Gamma(\mu_t \phi) - \log \Gamma((1-\mu_t)\phi) + (\mu_t \phi - 1) \log y_t + \{(1-\mu_t)\phi - 1\} \log(1 - y_t)$$

Como $D(y; \mu, \phi) = \sum_{i=1}^{n} 2 \cdot (\mathcal{L}_t(\tilde{\mu}_t, \phi) - \mathcal{L}_t(\mu_t, \phi))$; onde $\tilde{\mu}_t$ é o valor de μ_t que satisfaz $\frac{\partial \mathcal{L}_t}{\partial \mu_t} = 0 \rightarrow \phi(y_t^* - \mu_t^*) = 0$ para $\phi \to \infty$ temos

$\mu_t^* = \log \left\{ \frac{\mu_t}{(1 - \mu_t)} \right\}$ e segue que: $\tilde{\mu}_t = y_t$; logo:

$$D(y; \mu, \phi) = \log \Gamma(\phi) - \log \Gamma(y_t \phi) - \log \Gamma((1-y_t)\phi) + (y_t \phi - 1) \log y_t + \{(1-y_t)\phi - 1\} \log(1 - y_t)$$
$$- \left\{ \log \Gamma(\phi) - \log \Gamma(\mu_t \phi) - \log \Gamma((1-\mu_t)\phi) \; (\mu_t \phi - 1) \log y_t + \{(1-\mu_t)\phi - 1\} \log(1 - y_t) \right\}$$

$$\boxed{D = \log \Gamma\left(\frac{\mu_t}{y_t}\right) + \log \Gamma\left(\frac{1-\mu_t}{1-y_t}\right) + \log(y_t)\{\phi(y_t - \mu_t)\} + \log(1 - y_t)\{\phi(\mu_t - y_t)\}}$$

Como $D^* = \hat{\phi} \cdot D$

para obtermos $\hat{\phi}$ temos que satisfazer: $U_\phi(\beta, \phi) = 0$, para isso definimos a matriz informação de Fisher:

$$\frac{\partial \mathcal{L}(\beta, \phi)}{\partial \beta_i} = \sum_{t=1}^{n} \frac{\partial \mathcal{L}_t(\mu_t, \phi)}{\partial \mu_t} \frac{\partial \mu_t}{\partial \eta_t} \frac{\partial \eta_t}{\partial \beta_i}$$

Como $\frac{d \mu_t}{d \eta_t} = \frac{1}{g'(\mu_t)}$:

$$\frac{\partial \mathcal{L}_t(\mu_t, \phi)}{\partial \mu_t} = \phi \left[\log \frac{y_t}{1-y_t} - \{ \psi(\mu_t \phi) - \psi((1-\mu_t)\phi) \} \right]$$

onde $\psi(\cdot)$ caracteriza a função digamma i.e. $\psi(z) = \frac{d}{dz} \log(\Gamma(z))$ $\forall z > 0$, dadas as condições de regularidade, sabe-se que o valor esperado da derivada da equação 1. é zero:

$$\frac{\partial \mathcal{L}(\beta, \phi)}{\partial \beta_i} = \phi \sum_{t=1}^{n} (y_t^* - \mu_t^*) \frac{1}{g'(\mu_t)} x_{ti}$$

onde $y_t^* = \log \left\{ \frac{\mu_t}{(1-\mu_t)} \right\}$ e $\mu_t^* = \psi(\mu_t \phi) - \psi((1-\mu_t) \cdot \phi)$

percebe-se de imediato que $\hat{\phi}$ é intratável pela presença da função digamma, tornando desnecessários os demais calculos das posteriores derivadas parciais. Logo, a titulo de simplificação: $D^* = (U_\phi(\beta, \phi) = 0) \cdot D$

Q03. Considere a função gama apresentada na questão anterior. Conforme mostrado na `Lista 3`, a distribuição gama é membro da família exponencial com ligação canônica. Usando a função de ligação logarítmica (frequentemente usada como função de ligação canônica para a distribuição Poisson), obtenha:

(a) A função escore $\mathrm{U}(\beta_j)$ e as entradas genéricas (j, s) da matriz de informação de Fisher, digamos $\mathcal{I}_{js}(\boldsymbol{\beta})$.

$$f(y|\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta}$$

$$\ell(y|\alpha,\beta) = \frac{-y}{\beta} - \alpha \log \beta + (\alpha-1)\log y - \log\Gamma(\alpha)$$

Com parâmetro canônico $\theta = -\beta^{-1} = \eta$

$$\begin{cases} \frac{d}{d\theta}\alpha \log\left(-\frac{1}{\theta}\right) = -\frac{\alpha}{\theta} = \frac{-\alpha}{-\frac{1}{\beta}} \longrightarrow \mu = \alpha\beta \\ \frac{d}{d\theta} -\frac{\alpha}{\theta} = \alpha\left(-\frac{1}{\theta^2}\right) = \frac{\alpha}{\theta^2} \longrightarrow V(\mu) = -\alpha\beta^2 \Longrightarrow Var(y) = (-1)(-\alpha\beta^2) = \alpha\beta^2 \\ a(\phi) = -1 \end{cases}$$

a) O vetor score $U(\underset{\sim}{\beta}) = \nabla_{\underset{\sim}{\beta}} \ell(\underset{\sim}{\beta}, \phi)$ é obtido portanto:

$$\frac{\partial \ell(\beta_j, \phi)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{a(\phi)} W_i(\beta)\left(\frac{\partial \eta_i}{\partial \mu_i}\right) x_{ij}$$
$$= \sum_{i=1}^{n} (y_i - \alpha\beta) W_i(\beta) \cdot \alpha\beta^2 \cdot x_{ij}$$

(b) A matriz dos pesos $W(\boldsymbol{\beta})$ e a matriz $G(\boldsymbol{\beta}) = \text{diagonal}\left\{g'(\mu_i),\ \text{com } i = 1, \cdots, n\right\}$.

b) $W(\underset{\sim}{\beta}) = \text{diag}\{W_i(\beta),\ i=1,\dots,n\}$ e

$$W_i(\beta) = V^{-1}(\mu_i)\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 = -\frac{1}{\alpha\beta^2}(-\alpha\beta^2)^2 = -\alpha\beta^2$$

$$G(\beta) = \mathrm{diag}\{g'(\mu_i),\ i=1,\dots,n\} = \mathrm{diag}\left\{\frac{\partial \eta_i}{\partial \mu_i},\ i=1,\dots,n\right\}$$

$$= \mathrm{diag}\{-\alpha\beta^2\}_n$$

c) $$\underset{\sim}{z}^{(t)} = \underset{\sim}{\eta}^{(t)} + G(\underset{\sim}{\hat{\beta}}^{(t)})\,(y - \mu^{(t)})$$

$$z_i^{(t)} = \left(-\frac{1}{\beta}\right)^{(t)} + (-\alpha\beta^2)^{(t)}\,(y - (\alpha\beta)^{(t)})$$

(c) A variável dependente ajustada na t-ésima iteração do Fisher Scoring, i.e., $z_i^{(t)}$.

Como temos:

$$z^{(t)} = \eta_i^{(t)} + (y_i - \mu_i^{(t)}) \frac{1}{V(\mu_i^{(t)})[g'(\mu_i^{(t)})]^2}$$

como, $\mu_i = \log(e^{x^T\beta})$ e

$$g(\mu_i) = \sum_{i=1}^{m} x_{ik}\beta_k = \eta_i \longrightarrow \eta_i^{(t)} = x_{tk}\beta_k$$

com $[g'(\mu_i^{(t)})]^{-2} = \frac{1}{(\mu_i^{(t)})^2}$; $V(\mu_i^{(t)})^{-1} = -(\alpha \cdot \beta_k)^{-1}$

logo:

$$z^{(t)} = x_{tk}\beta_k - (y_i - \mu_i^{(t)}) \cdot (\alpha \cdot \beta_k)^{-1} \cdot (\mu_i^{(t)})^{-2}$$