ME 720 - Modelos Lineares Generalizados
Primeiro semestre de 2016
Prova I
Data: 28/04/2016

Nome: ______________________________ RA: ________________

Leia atentamente as instruções abaixo:

- **Coloque seu nome completo e RA em todas as folhas que você recebeu, inclusive nesta.**
- **Utilize somente o espaço delimitado para cada questão/item (veja se existe limite do número de linhas que você pode escrever).**
- **Leia atentamente cada uma das questões.**
- **Enuncie, claramente, todos os resultados que você utilizar.**
- **Justifique, adequadamente, seus desenvolvimentos sem, no entanto, escrever excessivamente.**
- **O(a) aluno(a) só poderá sair da sala após as 16h30, mesmo que já tenha finalizado a prova. Após a saída do(a) primeiro(a) aluno(a) não será permitido a entrada de nenhum(a) outro(a) aluno(a).**
- **Não é permitido empréstimo de material.**
- **Não serão dirimidas dúvidas de quaisquer natureza, após os 20 minutos iniciais.**
- **Resolva a prova, preferencialmente, à caneta (azul ou preta), e procure ser organizado(a).**
- **Contestações a respeito da nota/correção, só serão consideradas se estiverem por escrito.**
- **A nota do aluno(a) será** $\frac{NP}{NT} \times 10$, **em que NP é o número de pontos obtidos na prova e NT é o numero total de pontos da prova.**
- **Os resultados numéricos finais devem ser apresentados com, somente, duas casas decimais, a não ser que seja solicitado um número diferente de casas.**
- **A prova terá duração de 120 minutos, das 16h às 18h, improrrogáveis.**

Faça uma excelente Prova!!

1. Sejam $Y_i \overset{ind.}{\sim}$ Bernoulli(μ_i), logito$(\mu_i) = \ln\left(\dfrac{\mu_i}{1-\mu_i}\right) = \beta x_i, \beta \in (-\infty, \infty)$ e x_i (não aleatórias e conhecidas), $i = 1, 2, ..., n$. Lembre-se de que $\mu_i = \dfrac{e^{\beta x_i}}{1+e^{\beta x_i}} = \dfrac{1}{1+e^{-\beta x_i}}$. Responda os itens:

   a) Obtenha a função escore e a informação de Fisher associadas ao modelo, simplificando-as o máximo possível, e apresente a equação que deve ser resolvida para que se obtenha o estimador de máxima verossimilhança (emv) de β. (250 pontos)

   b) Obtenha a expressão da razão de chances (RC), em relação ao aumento em uma unidade no valor da covariável (x_i) e a denote por τ. Além disso, obtenha o respectivo emv e sua distribuição assintótica utilizando o método Delta. (250 pontos)

2. O conjunto de dados analisado foi extraído do censo do IBGE de 2000 e apresenta, para cada unidade da federação, o número médio de anos de estudo e a renda média mensal (em reais) do chefe ou chefes do domicílio. O presente objetivo é estudar o relacionamento da renda média mensal (Y_i) em função do número médio de anos de estudo (x_i), i=1,2,...,27. Para isso quatro modelos foram ajustados (veja a descrição deles abaixo) em que $\overline{x} = \dfrac{1}{27}\sum_{i=1}^{27} x_i$. Considere que a aproximação do desvio pela distribuição de $\chi^2_{(n-p)}$ é adequada para os quatro modelos. Alguns resultados relativos ao ajuste deles encontram-se na Tabela 1 e nas Figuras 1 e 2. Responda os itens abaixo:

   (Modelo 1) : $Y_i \overset{ind.}{\sim} N(\mu_i, \sigma^2), \mu_i = \beta_0 + \beta_1(x_i - \overline{x})$

   (Modelo 2) : $Y_i \overset{ind.}{\sim} N(\mu_i, \sigma^2), \mu_i = \beta_0 + \beta_1(x_i - \overline{x}) + \beta_2(x_i - \overline{x})^2$

   (Modelo 3) : $Y_i \overset{ind.}{\sim} \text{gama}(\mu_i, \phi), \ln(\mu_i) = \beta_0 + \beta_1(x_i - \overline{x})$

   (Modelo 4) : $Y_i \overset{ind.}{\sim} \text{gama}(\mu_i, \phi), \ln(\mu_i) = \beta_0 + \beta_1(x_i - \overline{x}) + \beta_2(x_i - \overline{x})^2$

   a) Qual dos quatro modelos você escolheria para analisar o conjunto de dados? Justifique sua escolha do modo mais amplo possível, com base nos resultados apresentados. Seus comentários não podem ultrapassar 20 linhas. (200 pontos)

   b) Para os modelo 3 interprete os parâmetros e^{β_0} e e^{β_1}, em termos do problema. Repita o procedimento para o modelo 4, considerando os parâmetros e^{β_0} e $e^{-\beta_1/(2\beta_2)}$. (150 pontos)

   c) Encontre a distribuição assintótica do emv de $e^{-\beta_1/(2\beta_2)}$, com base no método Delta, e um respectivo IC (95%) assintótico para ele. (200 pontos)

Tabela 1: Estatísticas de comparação de modelos, desvio estimado (e respectivo p-valor): Questão 2

Modelo	AIC	BIC	desvio	p-valor (desvio)
1	315,26	319,15	25,00	0,4624
2	298,66	303,85	24,00	0,4616
3	288,13	292,02	27,02	0,3547
4	290,09	295,28	27,02	0,3034

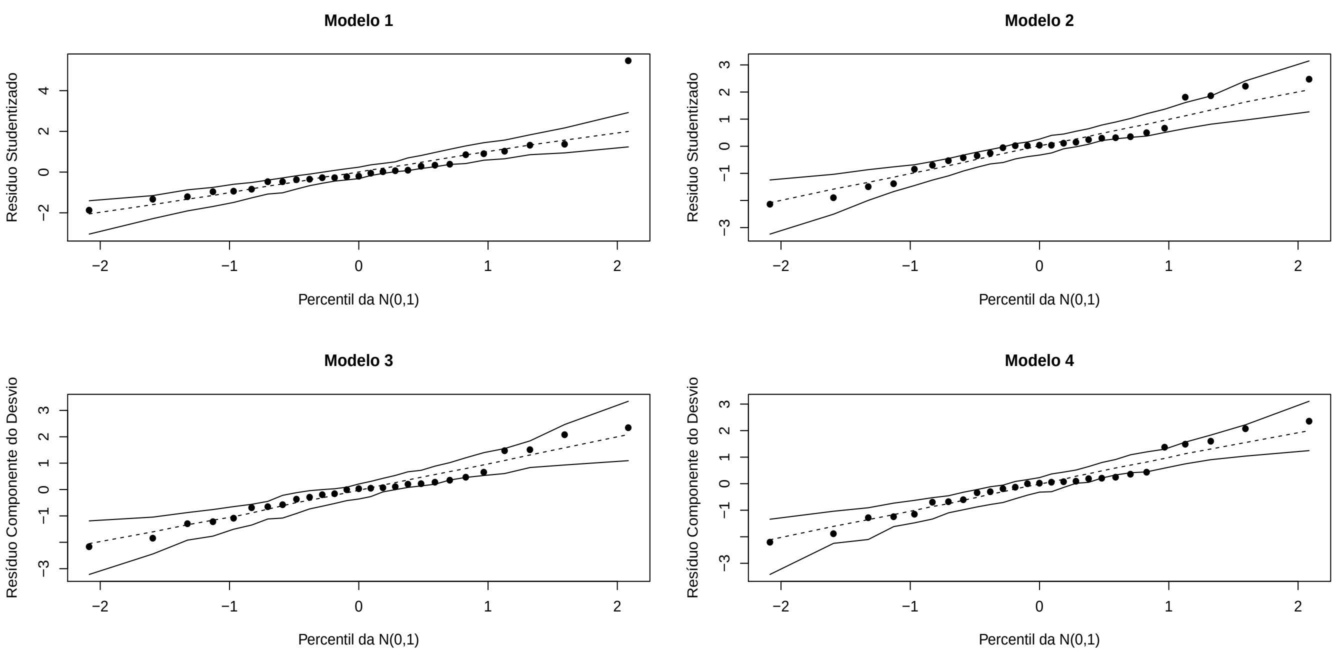

Figura 1: Gráficos de envelope para os quatro modelos: Questão 2

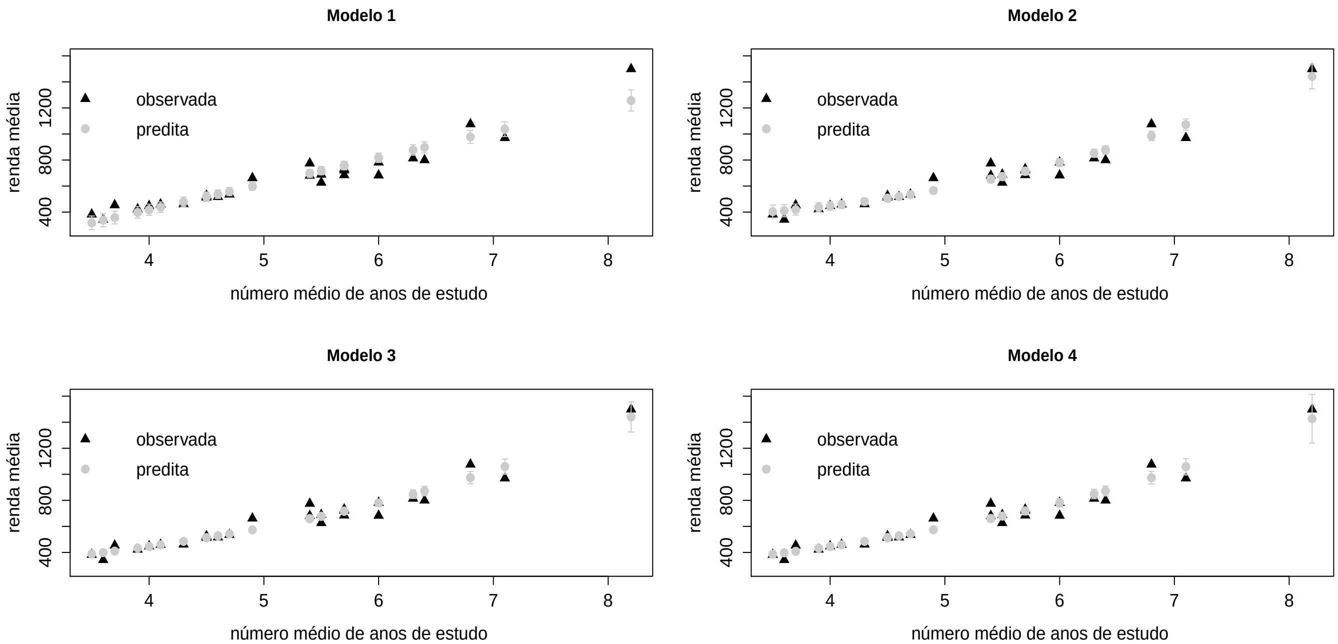

Figura 2: Valores observados e preditos (com IC's (95%)) pelos quatro modelos: Questão 2

3. O conjunto de dados analisado se refere ao percentual de besouros mortos quando expostos à diferentes doses de disulfeto de carbono gasoso (CS_2). Os dados se encontram na Tabela 2. O objetivo consiste em modelar a proporção de insetos mortos em função da dose. Sejam Y_i: número de besouros expostos à dose i de CS_2 que morreram, m_i: número de besouros expostos à dose i de CS_2 e x_i: dose (log da concentração de CS_2) à que os besouros do grupo i foram expostos, $i = 1, 2, ..., 8$. Para analisar os dados considerou-se o seguinte modelo de regressão :

$$\begin{aligned} Y_i &\overset{ind.}{\sim} \text{binominal}(m_i, \mu_i) \\ \ln(-\ln(1-\mu_i)) &= \beta_0 + \beta_1(x_i - \overline{x}), \overline{x} = \frac{1}{8}\sum_{i=1}^{8} x_i. \end{aligned}$$

Alguns resultados relativos ao ajuste do modelo em questão se encontram na Tabela 3 e nas Figuras 3 e 4. O desvio estimado do modelo foi $D(\boldsymbol{y}; \widetilde{\boldsymbol{\mu}}) = 3,44$. Admita que a aproximação do desvio pela distribuição $\chi^2_{(n-p)}$ é apropriada. Responda os itens:

a) O que você pode afirmar com relação à qualidade de ajuste do modelo? Comente, da forma mais completa possível, com base nos resultados apresentados. Você utilizaria tal modelo para analisar os dados? Justifique, adequadamente, sua resposta. Seus comentários não podem ultrapassar 20 linhas. (200 pontos)

b) Obtenha a expressão da razão de chances (RC), em relação ao aumento em uma unidade da covariável (x_i) e a denote por τ. Com base nos resultados apresentados, obtenha uma estimativa pontual para esse parâmetro. (100 pontos)

c) A notação usual para uma dose letal de 100p% é dada por DL_{100p}, ou seja, a probabilidade de um inseto morrer uma vez exposto à essa dose é dada por $p = F(\beta_0 + \beta_1 DL_{100p})$, em que $F(.)$ é uma função de ligação adequada. Para o modelo em questão, obtenha a forma do emv de DL_{100p} para um valor de p fixado, bem como uma estimava pontual considerando p =0,90, com base nos resultados apresentados. (200 pontos)

Tabela 2: Dados relativos à Questão 3

Dose: $log_{10}CS_2$	n^o Besouros expostos	n^o Besouros mortos
1,6907	59	6
1,7242	60	13
1,7552	62	18
1,7842	56	28
1,8113	63	52
1,8369	59	53
1,8610	62	61
1,8839	60	60

Tabela 3: Estimativas e testes de hipótese dos parâmetros do modelo ajustado: Questão 3

Parâmetro	Estimativa	EP	IC(95%)	Estat. Z_t	p-valor
β_0	-0,04	0,08	[-0,20 ; 0,11]	-0,54	0,5914
β_1	22,04	1,80	[18,51 ; 25,57]	12,25	$< 0,0001$

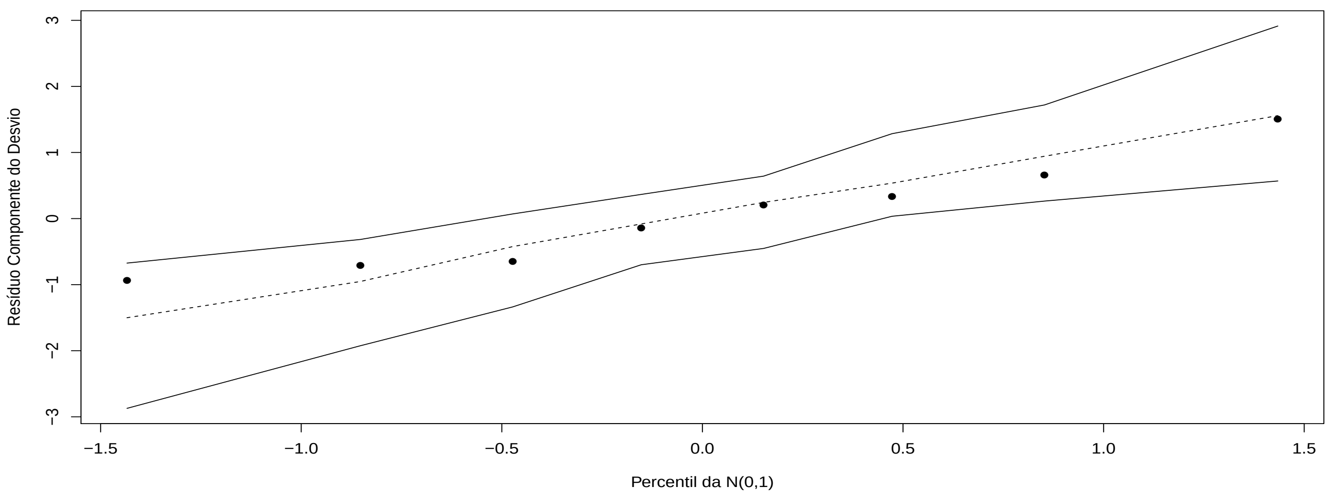

Figura 3: Gráficos de envelope para o modelo ajustado: Questão 3

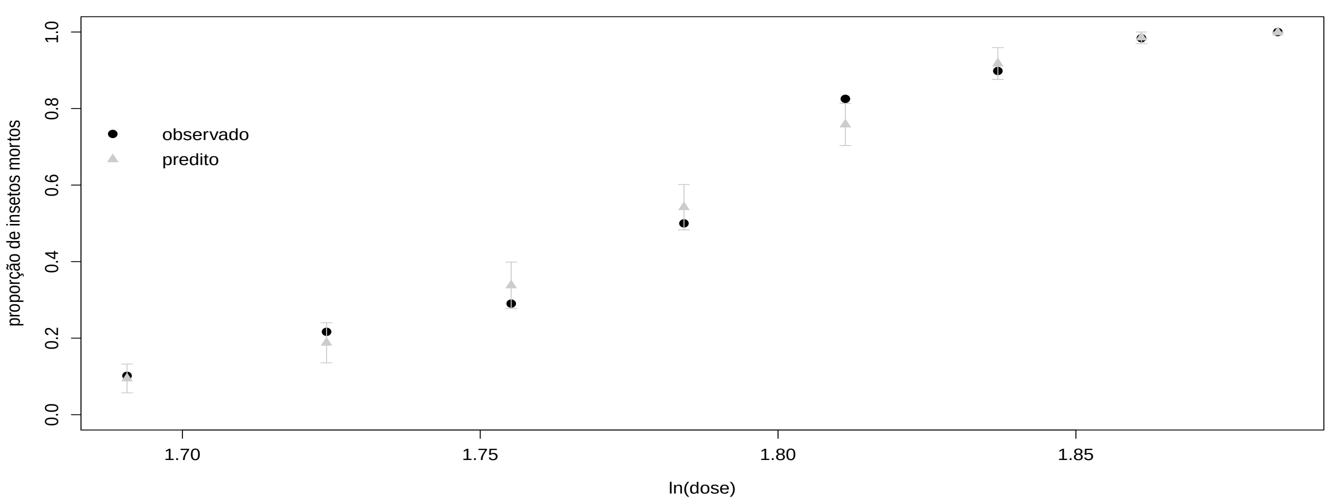

Figura 4: Valores observados e preditos (com IC's (95%)) pelo modelo ajustado: Questão 3

4. Sejam Y_i, i = 1,...,n, variáveis aleatórias independentes com distribuição gama de média $\mu_i = g^{-1}(\eta_i)$ e parâmetro de precisão ϕ, em que $g(.)$ é uma função de ligação apropriada. Responda os itens:

    a) Mostre que a veersão assintótica da estatística observada (ou seja, em função de y_i) do teste da razão de verossimilhanças para testar $H_0 : \phi = 1$ contra $H_1 : \phi \neq 1$ é igual à $\lambda = 2n\left[\ln(\widehat{\phi}) - \ln(\Gamma(\widehat{\phi})) - (\widehat{\phi} - 1)\left\{1 - \Psi(\widehat{\phi})\right\}\right]$, em que $\Gamma(.)$ é a função gama, $\Psi(.)$ é a função digama e $\Gamma(1) = 1$. Use o fato de que $\ln(\widehat{\phi}) - \Psi(\widehat{\phi}) = \dfrac{\overline{D}}{2}$, em que $\overline{D} = \sum_{i=1}^{n} D(y_i, \widehat{\mu}_i)/n = D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}})/n$ é o desvio não escalonado médio do modelo. Além disso, utilize os resultados sobre estimação dos MLG que você julgar necessários, sem se esquecer de citá-los. (400 pontos)

    b) Em termos da distribuição de probabilidade da variável resposta, a que corresponde tais hipóteses? (100 pontos)

Formulário

1. Se $Y \sim \text{Bernoulli}(\mu), \mu \in (0, 1)$, então $f(y) = \mu^y(1-\mu)^{1-y}\mathbb{1}_{\{0,1\}}(y)$.

2. Se $Y \sim \text{gama}(\mu, \phi)$, então $f(y) = \dfrac{1}{\Gamma(\phi)}\mu^{-\phi}\phi^{\phi}y^{\phi-1}\exp\left\{-\dfrac{\phi y}{\mu}\right\}$
e o desvio, para o respectivo MLG, é dado por $D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n}\{-\ln(y_i/\widehat{\mu}_i) + (y_i - \widehat{\mu}_i)/\widehat{\mu}_i\}$.

3. A estatística do teste da razão de verossimilhanças (na sua versão assintótica) para testar: $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ vs $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ é dada por $\lambda = -2(l(\widehat{\boldsymbol{\theta}}_0) - l(\widehat{\boldsymbol{\theta}}))$, em que: $l(.)$ representa a logverossimilhança do modelo, $\widehat{\boldsymbol{\theta}}_0$ o estimador de máxima verossimilhança (emv) de $\boldsymbol{\theta}$ sob H_0 e $\widehat{\boldsymbol{\theta}}$ o emv de $\boldsymbol{\theta}$ irrestrito.

4. Sob certas condições, $D(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}) \approx \chi^2_{(n-p)}$, para n suficientemente grande, em que n é o tamanho da amostra, p é o número de parâmetros e $D(\boldsymbol{y}; \widehat{\boldsymbol{\mu}})$ representa a função desvio (ou simplesmente desvio) do modelo.

5. Seja $g(x) = e^{f(x)}$, então $\dfrac{\partial g(x)}{\partial x} = e^{f(x)}\dfrac{\partial f(x)}{\partial x}$. Regra do quociente, seja $f(x) = \dfrac{g(x)}{h(x)}$, então $\dfrac{\partial f(x)}{\partial x} = \dfrac{g'(x)h(x) - g(x)h'(x)}{h^2(x)}$, em que $f'(x) = \dfrac{\partial f(x)}{\partial x}$ representa o operador derivada.

6. Método delta univariado: Seja $\widehat{\beta}_1$ de sorte que, para n suficientemente grande, $\widehat{\beta}_1 \approx N_1(\beta_1, \sigma)$ e defina $\widehat{\tau} = g(\widehat{\beta}_1)$. Então, para n suficientemente grande, $\widehat{\tau} \approx N(\tau, \psi^2\sigma)$, em que $\tau = g(\beta_1)$ e $\psi = \left[\dfrac{dg(\beta_1)}{d\beta_1} \right]$.

7. Método delta bivariado: Seja $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \widehat{\beta}_2)'$ de sorte que, para n suficientemente grande, $\widehat{\boldsymbol{\theta}} \approx N_2(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ e defina $\widehat{\tau} = g(\widehat{\boldsymbol{\beta}})$. Então, para n suficientemente grande, $\widehat{\tau} \approx N(\tau, \boldsymbol{\Psi}\boldsymbol{\Sigma}\boldsymbol{\Psi}')$, em que $\tau = g(\boldsymbol{\beta})$ e $\boldsymbol{\Psi} = \left[\begin{array}{cc} \dfrac{\partial g(\boldsymbol{\beta})}{\partial \beta_1} & \dfrac{\partial g(\boldsymbol{\beta})}{\partial \beta_2} \end{array} \right]$ (vetor linha).