ME 720 - Modelos Lineares Generalizados
Primeiro semestre de 2016
Prova I - segunda chamada
Data: 26/05/2016

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.**
- **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 10h às 12h, improrrogáveis.**

Faça uma excelente Prova!!

1. Sejam $Y_i \overset{ind.}{\sim} \text{NI}(\mu_i, \phi), \ln(\mu_i) = \beta x_i, \beta \in (-\infty, \infty)$ e x_i (não aleatórias e conhecidas), $i = 1, 2, ..., n$ (NI corresponde à distribuição normal inversa, veja formulário). 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. (300 pontos)

   b) Obtenha a expressão que representa o incremento multiplicativo na média 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. (200 pontos)

2. O conjunto de dados analisado corresponde ao número de bactérias sobreviventes em amostras de um produto alimentício segundo o tempo (em minutos) de exposição do produto a uma temperatura de $300^o F$. Nessas amostras de alimentos foram feitas 12 medições, a cada minuto, contabilizando a quantidade de bactérias vivas (do total original) sobreviventes. A variável resposta corresponde ao número (contagem) de bactérias sobreviventes (Y_i) enquanto que a variável explicativa (x_i) é o tempo de exposição $i = 1, 2, ..., 12$. Para isso quatro modelos foram ajustados (veja a descrição deles abaixo) em que $\overline{x} = \dfrac{1}{12}\displaystyle\sum_{i=1}^{12} x_i$. 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 nas Tabelas 1 e 2 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. A despeito da sua decisão, quais seriam as limitações/problemas dos quatro modelos em questão? Seus comentários não podem ultrapassar 20 linhas. (200 pontos)

   b) Para o modelo 3 interprete o parâmetro $e^{\beta_0+\beta_1}$, em termos do problema. Repita o procedimento para o modelo 4, considerando o parâmetro $e^{\beta_0+\beta_1+\beta_2}$. (150 pontos)

   c) Encontre a distribuição assintótica do emv de $e^{\beta_1+\beta_2(x_i-\overline{x})+\beta_2(x_i-\overline{x})^2}$, para um dado x_i, 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	107,63	109,08	10,00	0,4405
2	96,71	98,65	9,00	0,4373
3	78,39	79,84	12,02	0,2834
4	77,86	79,80	12,02	0,2122

Tabela 2: Estimativas (pontuais e intervalares) e testes de hipótese de nulidade dos parâmetros dos modelos: Questão 2

Modelo	Par.	Est.	EP	IC(95%)	Estat. Z_t	p-valor
1	β_0	61,08	5,28	[50,73 ; 71,44]	11,56	$<$ 0,0001
	β_1	-12,48	1,53	[-15,48 ; -9,48]	-8,16	$<$ 0,0001
2	β_0	45,76	4,90	[36,15 ; 55,37]	9,33	$<$ 0,0001
	β_1	-12,48	0,94	[-14,32 ; -10,63]	-13,25	$<$ 0,0001
	β_2	1,29	0,31	[0,68 ; 1,89]	4,17	0,0024
3	β_0	3,81	0,03	[3,74 ; 3,88]	109,22	$<$ 0,0001
	β_1	-0,24	0,01	[-0,26 ; -0,22]	-23,38	$<$ 0,0001
4	β_0	3,86	0,05	[3,76 ; 3,96]	77,39	$<$ 0,0001
	β_1	-0,24	0,01	[-0,26 ; -0,22]	-24,69	$<$ 0,0001
	β_2	-0,0045	0,0031	[-0,0106 ; 0,0017]	-1,43	0,1867

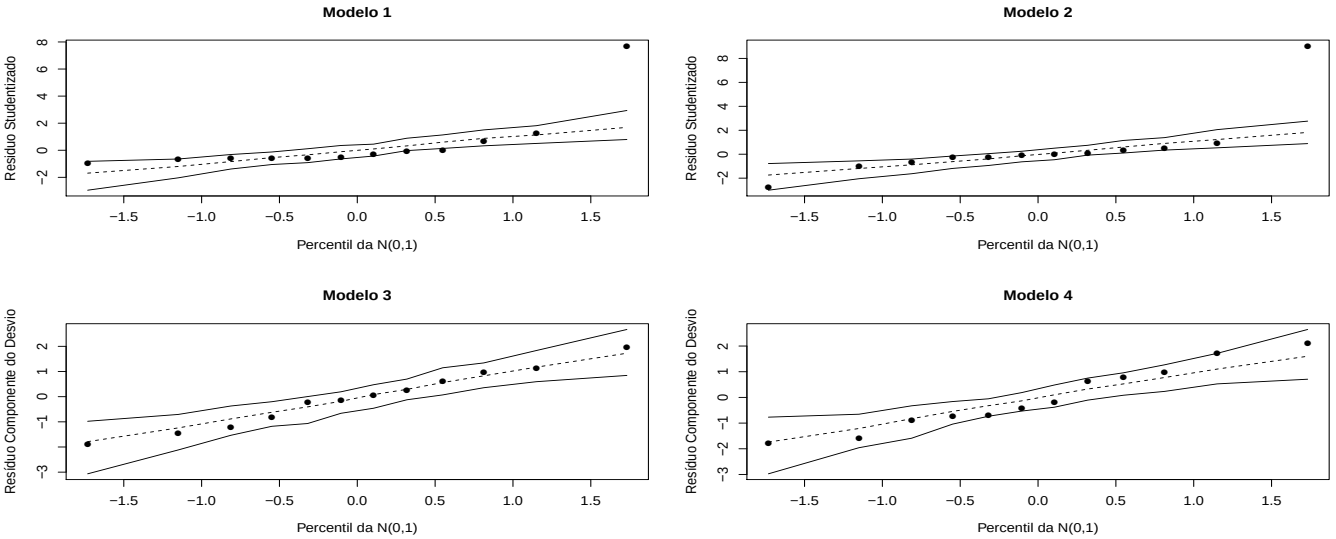

Figura 1: Gráficos de envelope para os quatro modelos: Questão 2

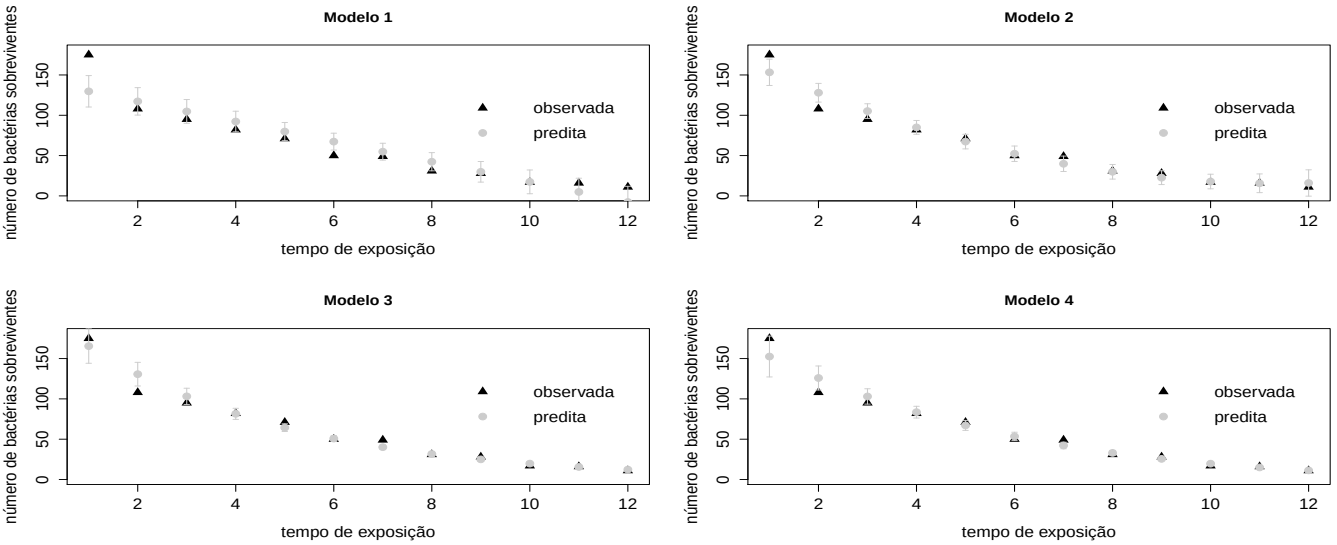

Figura 2: Valores observados e preditos (com IC's (95%)) pelos quatro modelos: Questão 2

3. O conjunto de dados analisado estão relacionados a um grupo de gestantes fumantes classificadas segundo alguns fatores a saber, idade: < 30 (menos que 30 anos) ou 30+ (30 ou mais) e número de cigarros consumidos por dia: < 5 (menos do que 5) ou 5+ (5 ou mais). Para cada um desses quatro grupos foram contados, de um total (coluna "Total") de recém-nascidos, a quantidade daqueles que sobrevieram (coluna "Não"). A Tabela 3 apresenta os dados do experimento. O objetivo consiste em modelar a proporção de recém-nascidos que vieram à óbito em função dos fatores de interesse (idade e número de cigarros). Sejam Y_{ij}: número de recém-nascidos que vieram à óbito, de mães que fumam uma quantidade j de cigarros por dia e que pertencem ao grupo i da idade e m_{ij} : número total de recém-nascidos de mães que fumam uma quantidade j de cigarros por dia e que pertencem ao grupo i da idade. Para analisar os dados considerou-se o seguinte modelo de regressão :

$$\begin{aligned} Y_{ij} &\overset{ind.}{\sim} \text{binominal}(m_{ij}, \mu_{ij}) \\ \ln\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) &= \alpha + \beta_i + \gamma_j + (\beta\gamma)_{ij}, i = 1, 2, j = 1, 2, \beta_1 = \gamma_1 = (\beta\gamma)_{i1} = (\beta\gamma)_{1j} = 0, \forall i, j. \end{aligned}$$

Alguns análises se encontram na Tabela 4 e na Figura 3. Responda os itens:

Tabela 3: Proporção de recém-nascidos que vieram à obito e sobreviveram: Questão 3

idade	N. de cigarros	Sobrevivência		
		Não	Sim	Total
<30	< 5	74	4327	4401
	5+	15	499	514
30+	< 5	55	1741	1796
	5+	5	135	140

a) Escreva cada um dos parâmetros μ_{ij} em função dos parâmetros $\boldsymbol{\beta} = (\alpha, \beta_2, \gamma_2, (\beta\gamma)_{22})'$ e forneça uma interpretação para cada um desses parâmetros ($\boldsymbol{\beta}$) (100 pontos).

b) O que você pode afirmar acerca da existência de interação entre os fatores idade e número de cigarros? E em relação à existência de efeitos de cada um deles? Suas conclusões eram esperadas? Justifique, adequadamente, sua resposta (150 pontos).

c) Você utilizaria o modelo em questão para analisar os dados, ou ajustaria algum outro? Se for o caso, escreva esse outro modelo, justificando o porque de sua escolha (de ajustar ou não outro modelo e porque, se for o caso, considerar o modelo que você está propondo) (100 pontos).

d) Obtenha a expressão da razão de chances (RC), entre as idades, para uma quantidade de cigarros consumida de < 5, e a denote por τ. Com base nos resultados apresentados, obtenha uma estimativa pontual para esse parâmetro. (150 pontos)

Tabela 4: Estimativas (pontuais e intervalares) e testes de hipótese de nulidade dos parâmetros do modelo ajustado: Questão 3

Parâmetro	Estimativa	EP	IC(95%)	Estat. Z_t	p-valor
α	-4,07	0,12	[-4,30 ; -3,84]	-34,70	$<$ 0,0000
β_2	0,56	0,29	[0,00 ; 1,13]	1,96	0,0495
γ_2	-3,45	0,14	[-3,72 ; -3,19]	-25,23	$<$ 0,0000
$(\beta\gamma)_{22}$	0,16	0,48	[-0,77 ; 1,09]	0,33	0,7381

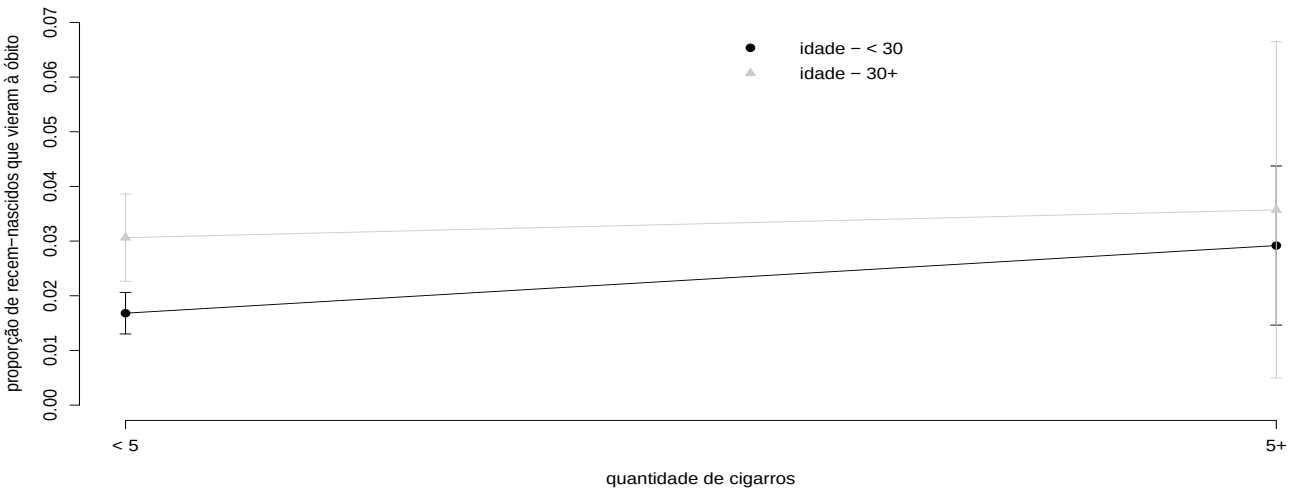

Figura 3: Gráficos de perfis médios amostrais: Questão 3

4. Sejam $Y_i \overset{ind.}{\sim} (m_i, \mu_i), i = 1, 2$, $\mu_i = g^{-1}(\beta_i), \beta_1 = 0$, em que $g(.)$ é uma função de ligação apropriada e β_2 um parâmetro desconhecido. Responda os itens:

   a) Obtenha, simplificando-a o máximo possível, a versão assintótica da estatística observada (ou seja, em função de y) do teste da razão de verossimilhanças para testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$. Utilize os resultados sobre estimação dos MLG que você julgar necessários, sem se esquecer de citá-los. (400 pontos)

   b) A que corresponde tais hipóteses? (100 pontos)

Formulário

1. Se $Y \sim$ binomial$(m, \mu), m \in \{1, 2, ...\}, \mu \in (0, 1)$ então $f(y) = \binom{m}{y}\mu^y(1-\mu)^{m-y}\mathbb{1}_{\{0,1,...,m\}}(y)$. Se m = 1, então $Y \sim$ Bernoulli(μ).

2. Se $Y \sim NI(\mu, \phi)$ (normal inversa), então: $f_Y(y) = \dfrac{\phi^{1/2}}{\sqrt{2\pi y^3}} \exp\left\{-\dfrac{\phi(y-\mu)^2}{2\mu^2 y}\right\}$. Além disso $\mathcal{E}(Y) = \mu$ e $\mathcal{V}(Y) = \dfrac{\mu^3}{\phi}$.

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).