

MLG para respostas positivas (assimétricas): parte 3

Prof. Caio Azevedo

Exemplo 7: Dados sobre a pesca do peixe batata

- Veja Paula (2013) e Paula e Oshiro (2001).
- O espinhel de fundo é definido como um método de pesca passivo (são aquelas que envolvem capturas de peixes ou outros animais aquáticos por enredamento, aprisionamento ou pescaria anzol, que não são movidos ativamente pelo homem ou por máquinas enquanto os organismos são capturados, Lagler (1978), http://redeetec.mec.gov.br/images/stories/pdf/eixo_rec_naturais/aquicultura/181012_tec_pes.pdf), sendo utilizado em todo o mundo em operações de pesca de diferentes magnitudes, da pesca artesanal a modernas pescarias mecanizadas.

- É adequado para capturar peixes com distribuição dispersa ou com baixa densidade, além de ser possível utilizá-lo em áreas irregulares ou em grandes profundidades.
- É um dos métodos que mais satisfazem às premissas da pesca responsável, com alta seletividade de espécies e comprimentos, alta qualidade do pescado, consumo de energia baixo e pouco impacto sobre o fundo oceânico.
- A espécie de peixe considerada neste exemplo é o peixe-batata pela sua importância comercial e ampla distribuição espacial.
- Uma amostra de $n = 156$ embarcações foi analisada no período de 1995 a 1999 sendo 39 da frota de Ubatuba e 117 da frota de Santos.
As variáveis consideradas para cada embarcação são as seguintes:

- frota (Santos ou Ubatuba), ano (95 a 99), trimestre (1 ao 4), latitude (distância ao Equador medida ao longo do meridiano de Greenwich, de $23,25^\circ$ a $28,25^\circ$), longitude (distância ao meridiano de Greenwich medida ao longo do Equador, de $41,25^\circ$ a $50,75^\circ$), dias de pesca (quantidade de dias que o barco dispendeu na pesca), captura (quantidade de peixes batata capturados, em kg) e cpue (captura por unidade de esforço, kg/dias de pesca).
- Um dos objetivos desse estudo é tentar explicar a cpue pelas variáveis frota, ano, trimestre, latitude e longitude.
- Consideraremos frota, ano e trimestre como fatores qualitativos nominais.

- Estudos similares realizados em outros países verificaram que é bastante razoável supor que a c_{pue} tem distribuição assimétrica à direita, como é o caso da distribuição gama (vide, por exemplo, Goni, Alvarez e Adlerstein, 1999).
- A abordagem aqui adotada é ajustar um modelo inicial com somente os fatores principais. Posteriormente, dos fatores principais retidos no modelo, exploraremos a existência de possíveis interações de primeira ordem.

Medidas resumo:cpue

Frota		
Medida resumo	Santos	Ubatuba
Média	215,76	134,90
DP	125,50	81,64
Var.	15751,38	6665,16
CV(%)	58,17	60,52
CA	1,01	0,87
Mín.	50,00	43,75
Máx.	600,00	350,00
n	117	39

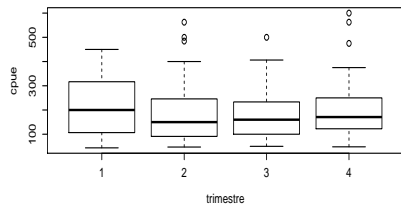
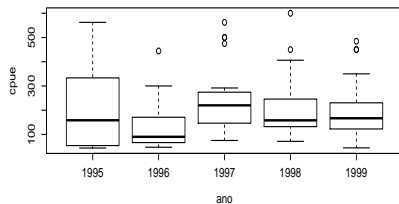
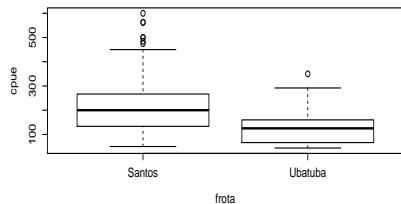
Medidas resumo:cpue

Medida resumo	Ano				
	1995	1996	1997	1998	1999
Média	204,52	134,93	249,08	204,77	186,88
Dp	151,31	104,32	138,10	118,85	99,86
Var.	22894,22	10883,03	19071,81	14124,25	9972,64
CV(%)	73,99	77,31	55,45	58,04	53,44
CA	0,59	1,48	0,90	1,53	0,960
Mín.	43,75	47,08	75,00	71,43	44,44
Máx.	562,5	444,44	562,50	600,00	484,50
n	22	20	23	32	59

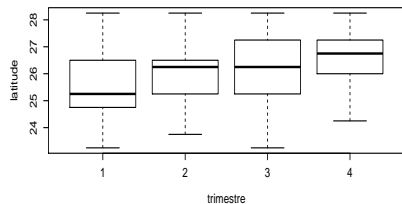
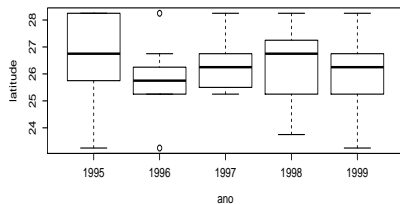
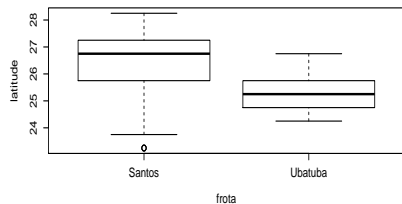
Medidas resumo:cpue

Medida resumo	Trimestre			
	1	2	3	4
Média	217,52	187,80	180,48	204,50
DP	136,92	130,16	97,28	125,57
Var.	18745,80	16941,60	9463,89	15767,49
CV(%)	62,94	69,31	53,90	61,40
CA	0,35	1,20	1,12	1,35
Mín.	43,75	47,08	50,00	48,00
Máx.	450,00	562,50	500,00	600,00
n	27	40	45	44

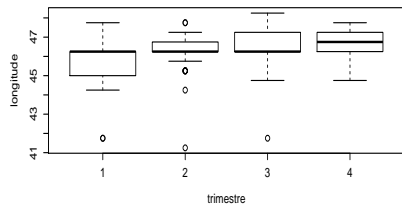
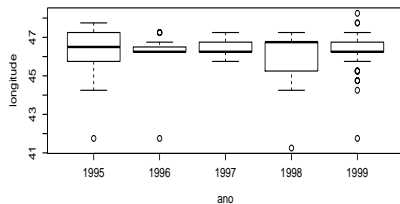
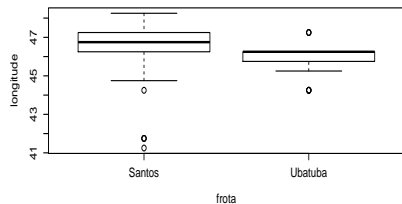
Boxplot - cpue



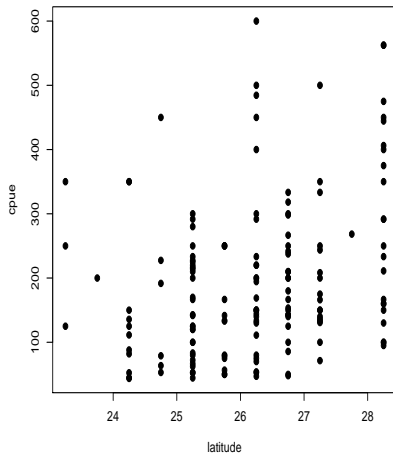
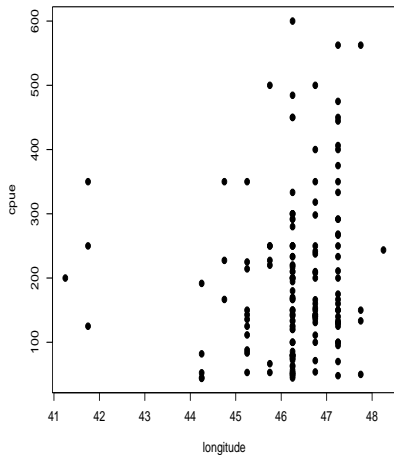
Boxplot - latitude



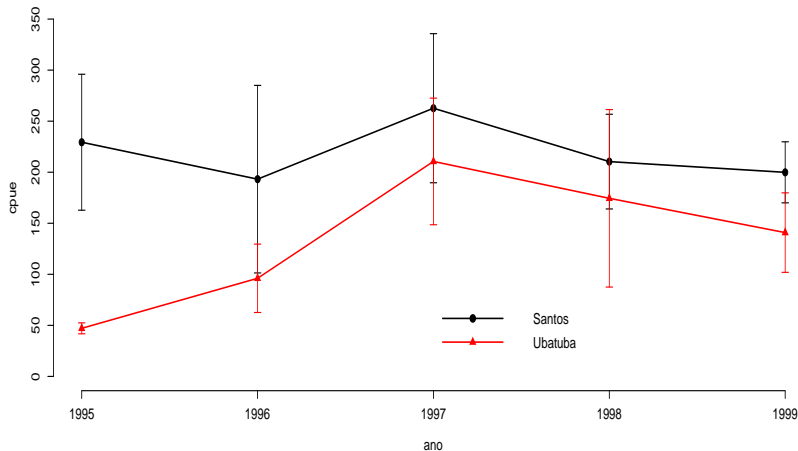
Boxplot - longitude



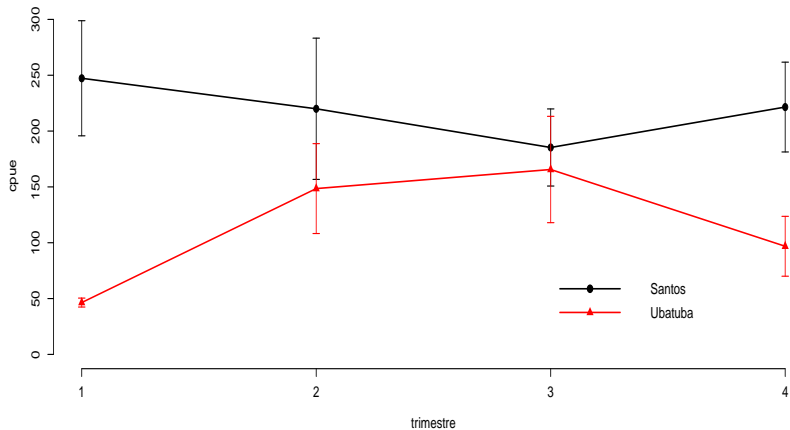
Gráficos de dispersão



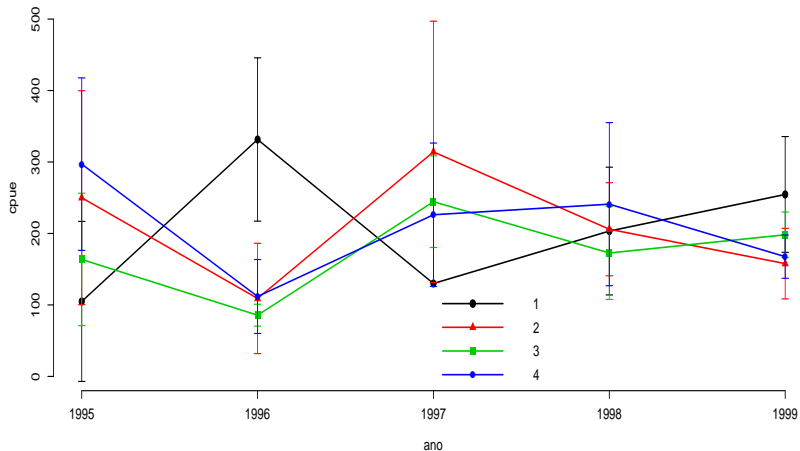
Gráficos de perfias médios - frota x ano



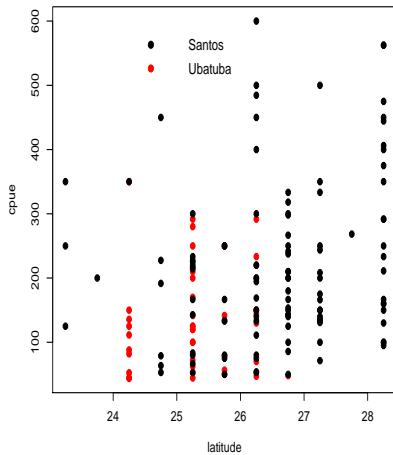
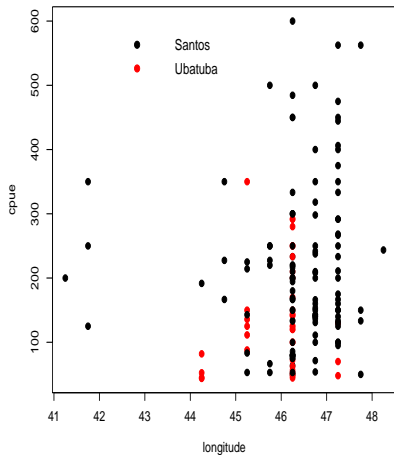
Gráficos de perfias médios - frota x trimestre



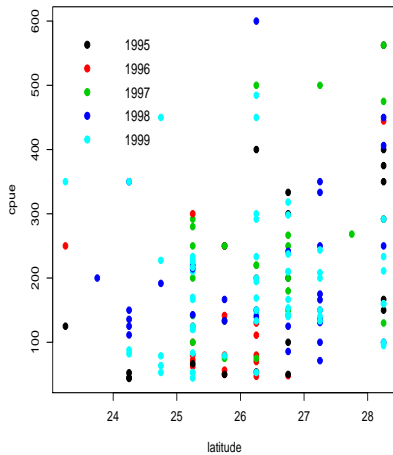
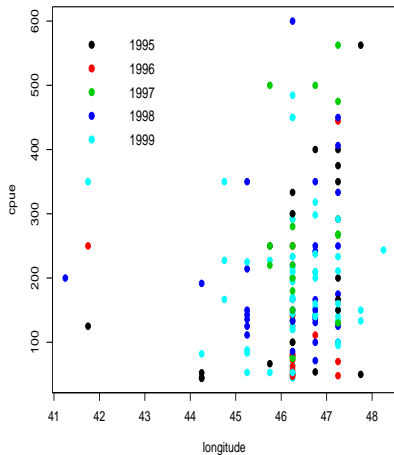
Gráficos de perfias médios - ano x trimestre



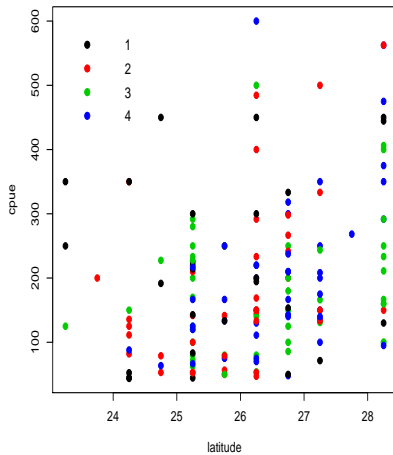
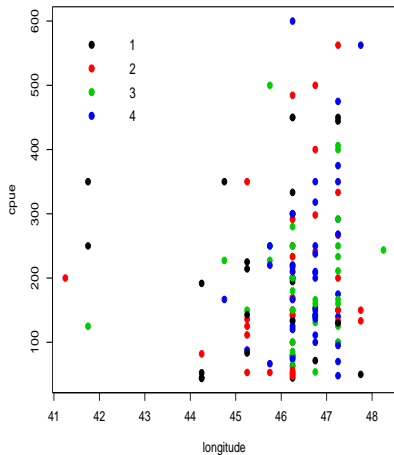
Gráficos de dispersão por frota



Gráficos de dispersão por ano



Gráficos de dispersão por trimestre



Modelo inicial (só efeitos principais) (modelo 1)

$$Y_{ijkl} \stackrel{ind.}{\sim} \text{gama}(\mu_{ijkl}, \phi),$$

$$i = 1, 2 \text{ (frota : 1 - Santos; 2 - Ubatuba)}$$

$$j = 1, 2, 3, 4, 5 \text{ (ano : 1 - 1995, 2 - 1996, 3 - 1997, 4 - 1998, 5 - 1999)}$$

$$k = 1, 2, 3, 4 \text{ (trimestre : 1 - 1, 2 - 2, 3 - 3, 4 - 4)}$$

$$l = 1, 2, \dots, n_{ijk} \text{ (barco - unidade experimental)}$$

$$\ln(\mu_{ijkl}) = \alpha + \beta_i + \gamma_j + \delta_k + \lambda_1(x_{1ijkl} - 23, 25) + \lambda_2(x_{2ijkl} - 41, 25),$$

$$\beta_1 = \gamma_1 = \delta_1 = 0$$

$$\mathcal{E}(Y_{ijkl}) = \mu_{ijkl} \quad ; \quad \mathcal{V}(Y_{ijkl}) = \frac{\mu_{ijkl}^2}{\phi}$$

Modelo

- Y_{ijkl} : é a cpue obtida pelo l -ésimo barco, da i -ésima frota, no j -ésimo ano, no k -ésimo trimestre.
- x_{1ijkl} : é a latitude em que a cpue obtida pelo l -ésimo barco, da i -ésima frota, no j -ésimo ano, no k -ésimo trimestre fora medida.
- x_{2ijkl} : é a longitude em que a cpue obtida pelo l -ésimo barco, da i -ésima frota, no j -ésimo ano, no k -ésimo trimestre fora medida.
- e^α : é o valor esperado da cpue para barcos da frota de Santos, que pescaram no ano de 1995, no trimestre 1, na latitude de 23,25 e longitude de 41,25.

Modelo

- $e^{\beta_2} = \frac{\mu_{2jkl}}{\mu_{1jkl}}$: é o incremento multiplicativo no valor esperado da cpue quando se compara a frota de Santos com a frota de Ubatuba (mantendo-se todas as outras variáveis fixas) .
- $e^{\gamma_j} = \frac{\mu_{ijk1}}{\mu_{i1kl}}$: é o incremento multiplicativo no valor esperado da cpue quando se compara o ano j com o ano de 1995 (mantendo-se todas as outras variáveis fixas) .
- $e^{\delta_k} = \frac{\mu_{ijk1}}{\mu_{ij1l}}$: é o incremento multiplicativo no valor esperado quando se compara o trimestre k com o trimestre 1 (mantendo-se todas as outras variáveis fixas) .

Modelo

- e^{γ_1} : é o incremento (positivo ou negativo) no valor esperado da cpue para o aumento em uma unidade da latitude.
- e^{γ_2} : é o incremento (positivo ou negativo) no valor esperado da cpue para o aumento em uma unidade da longitude.

Ajuste do modelo

- Utilizando o método de seleção de stepwise via estatística AIC o modelo selecionado corresponde ao modelo 1 sem a variável trimestre, doravante modelo 2.
- Segundo Paula (2013) a estimativa de ϕ é mantida constante ao longo do ajuste dos modelos quando se usa a função *stepAIC* (é utilizada a estimativa default do R ; neste exemplo as estimativas de MV e pelo MM de ϕ não diferem muito entre si, $\tilde{\phi}_M = 3,14$ e $\tilde{\phi}_{MV} = 3,47$). Sendo assim, adicionalmente utilizamos a análise do desvio para testar

$$H_0 : \delta_2 = \delta_3 = \delta_4 = 0 \text{ vs } H_1 : \delta_k \neq 0, \text{ para pelo menos um } k.$$

Modelo 2

$$Y_{ijl} \stackrel{ind.}{\sim} \text{gama}(\mu_{ijl}, \phi),$$

$$i = 1, 2 \text{ (frota : 1 - Santos; 2 - Ubatuba)}$$

$$j = 1, 2, 3, 4, 5 \text{ (ano : 1 - 1995, 2 - 1996, 3 - 1997, 4 - 1998, 5 - 1999)}$$

$$l = 1, 2, \dots, n_{ij} \text{ (barco - unidade experimental)}$$

$$\ln(\mu_{ijl}) = \alpha + \beta_i + \gamma_j + \lambda_1(x_{1ijl} - 23, 25) + \lambda_2(x_{2ijl} - 41, 25),$$

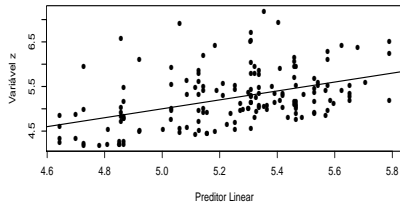
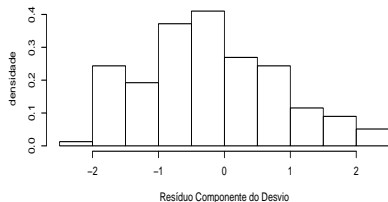
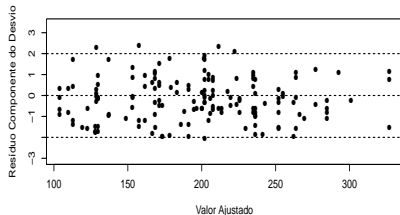
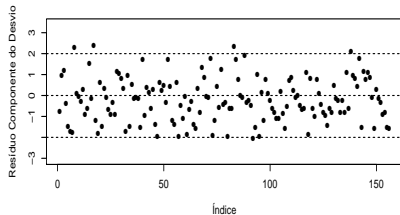
$$\beta_1 = \gamma_1 = 0$$

$$\mathcal{E}(Y_{ijl}) = \mu_{ijl} \quad ; \quad \mathcal{V}(Y_{ijl}) = \frac{\mu_{ijl}^2}{\phi}$$

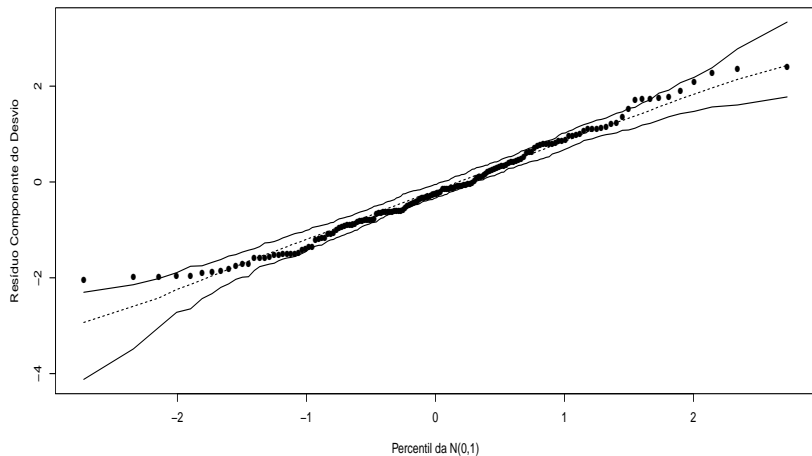
Ajuste do modelo

- Obtivemos o seguinte resultado para a análise do desvio:
 $f_c(p - valor) = 0,51(0,8469)$ o que confirma a não significância do fator trimestre.
- Exercício: escrever as hipóteses acima na forma $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{M}$ vs $H_1 : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{M}$.
- Para o modelo 2 o desvio (p-valor) = 163,62 (0,1405).
Considerando-se a aproximação pela $\chi^2_{(148)}$ válida, esse resultado indica um bom ajuste do modelo.

Gráficos de diagnóstico: Modelo 2



Envelope para os resíduos: Modelo 2



Estimativas dos parâmetros: Modelo gama 2

Parâmetro	Estimativa	EP	IC(95%)	Estat. Z	p-valor
α	5,29	0,25	[4,81 ; 5,78]	21,34	< 0,0001
β_2	-0,28	0,12	[-0,52 ; -0,03]	-2,24	0,0253
γ_2	-0,15	0,18	[-0,50 ; 0,19]	-0,87	0,3864
γ_3	0,33	0,16	[0,01 ; 0,65]	2,02	0,0430
γ_4	0,12	0,15	[-0,18 ; 0,41]	0,77	0,4421
γ_5	0,07	0,14	[-0,20 ; 0,34]	0,51	0,6114
λ_1	0,17	0,07	[0,03 ; 0,30]	2,45	0,0143
λ_2	-0,11	0,07	[-0,25 ; 0,03]	-1,55	0,1203
ϕ	3,38	0,37	[2,67 ; 4,10]	-	-

Estimativas dos parâmetros: Modelo gama 2

- Modelo gama é preferível ao modelo exponencial. Aparentemente a variável longitude não é significativa. Vamos ajustar um modelo sem esse variável mas contemplando a interação entre ano e frota, doravante modelo 3.

Modelo 3

$$Y_{ijl} \stackrel{ind.}{\sim} \text{gama}(\mu_{ijl}, \phi),$$

$$i = 1, 2 \text{ (frota : 1 - Santos; 2 - Ubatuba)}$$

$$j = 1, 2, 3, 4, 5 \text{ (ano : 1 - 1995, 2 - 1996, 3 - 1997, 4 - 1998, 5 - 1999)}$$

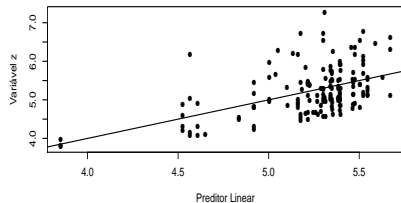
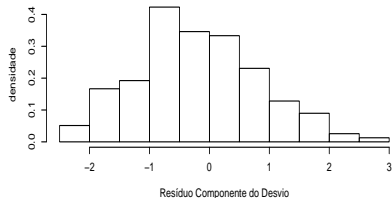
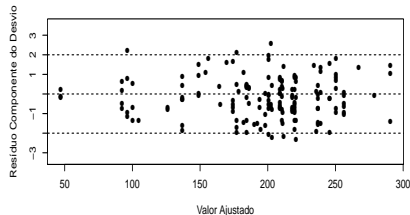
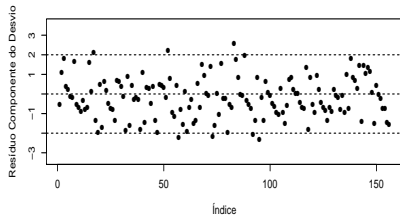
$$l = 1, 2, \dots, n_{ij} \text{ (barco - unidade experimental)}$$

$$\ln(\mu_{ijl}) = \alpha + \beta_i + \gamma_j + (\beta\gamma)_{ij} + \lambda_1(x_{1ijl} - 23, 25),$$

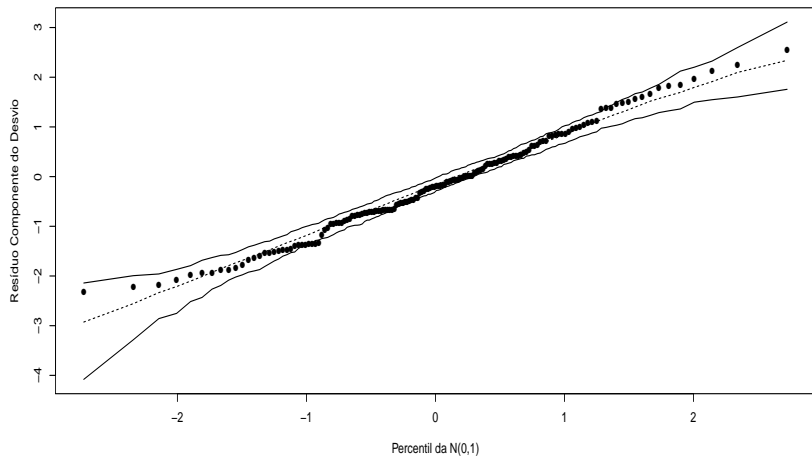
$$\beta_1 = \gamma_1 = (\beta\gamma)_{1j} = (\beta\gamma)_{i1} = 0, \forall i, j$$

$$\mathcal{E}(Y_{ijl}) = \mu_{ijl} \quad ; \quad \mathcal{V}(Y_{ijl}) = \frac{\mu_{ijl}^2}{\phi}$$

Gráficos de diagnóstico: Modelo 3



Envelope para os resíduos: Modelo 3



Estimativas dos parâmetros: Modelo gama 3

Parâmetro	Estimativa	EP	IC(95%)	Estat. Z	p-valor
α	5,11	0,19	[4,73 ; 5,48]	26,89	0,0000
β_2	-1,34	0,34	[-2,01 ; -0,66]	-3,88	0,0001
γ_2	-0,07	0,23	[-0,51 ; 0,38]	-0,30	0,7676
γ_3	0,15	0,18	[-0,20 ; 0,49]	0,84	0,3985
γ_4	-0,05	0,16	[-0,36 ; 0,26]	-0,29	0,7716
γ_5	-0,05	0,15	[-0,34 ; 0,23]	-0,37	0,7080
λ_1	0,08	0,04	[0,00 ; 0,16]	2,05	0,0408

Estimativas dos parâmetros: Modelo gama 3 (cont.)

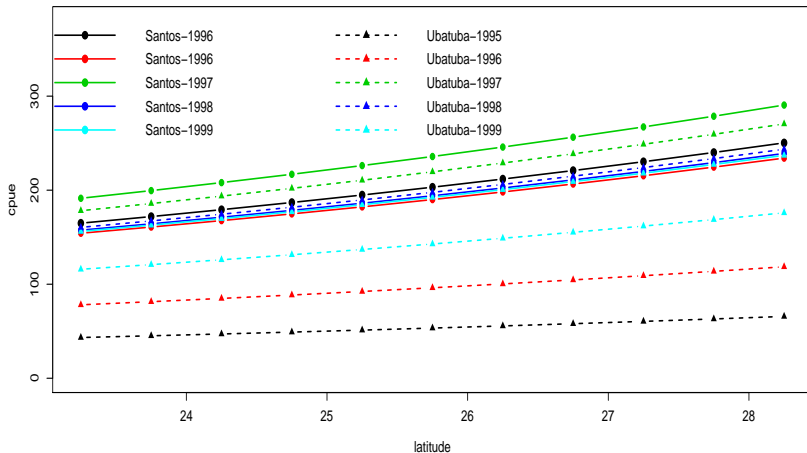
Parâmetro	Estimativa	EP	IC(95%)	Estat, Z	p-valor
$(\beta\gamma)_{22}$	0,66	0,42	[-0,17 ; 1,48]	1,56	0,1186
$(\beta\gamma)_{23}$	1,27	0,42	[0,45 ; 2,08]	3,05	0,0023
$(\beta\gamma)_{24}$	1,36	0,42	[0,54 ; 2,17]	3,25	0,0012
$(\beta\gamma)_{25}$	1,04	0,37	[0,31 ; 1,77]	2,78	0,0055
ϕ	3,58	0,39	[2,82 ; 4,34]	-	-

Estimativas dos parâmetros: Modelo gama 3 (cont.)

- Para o modelo 3 o desvio (p-valor) = 163,20 (0,1432).

Considerando-se a aproximação pela $\chi^2_{(145)}$ válida, esse resultado indica um bom ajuste do modelo. Modelo gama é preferível ao modelo exponencial. Poderíamos eliminar todos os parâmetros relativos ao efeito principal de ano (γ_i) e o parâmetro $(\beta\gamma)_{22}$, doravante modelo 4 (esse modelo apresentou limitações em termos de predição).

Médias previstas pelo modelo



Médias previstas (com IC's) pelo modelo

