

MLG para respostas positivas (assimétricas): parte 3

Prof. Caio Azevedo

Exemplo 7: Dados sobre a pesca do peixe batata

- Veja Paula (2024) e Paula e Oshiro (2001) (veja referência completa em Paula (2024)).
- 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 \(1978\)](#)).
- É utilizado em todo o mundo em operações de pesca de diferentes magnitudes, da pesca artesanal a modernas pescarias mecanizadas.

Cont.

- 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.

Cont.

- 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 $50,75^\circ$).

Cont.

- (Cont.)
 - dias de pesca (quantidade de dias que o barco dispendeu na pesca).
 - captura (quantidade de peixes batata capturados, em kg).
 - 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, por simplicidade, apenas as covariáveis frota e ano e estas, como fatores qualitativos nominais.
- Exercício: refazer as análises considerando todas as covariáveis.

Cont.

- 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](#)).
- Consideraremos um modelo completo, com a presença dos efeitos principais e da interação entre eles.
- Algumas discussões sobre interação (do ponto de vista Estatístico) podem ser vistas [aqui](#) e nas respectivas referências.

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	1,03	0,90
Curtose	3,59	2,96
Mínimo	50,00	43,75
Máximo	600	350

Medidas resumo:cpue

	Ano				
Medida resumo	1995	1996	1997	1998	1999
Média	204,51	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	9972,638
CV(%)	73,99	77,31	55,45	58,04	53,44
CA	0,64	1,60	0,96	1,61	0,99
Curtose	2,39	4,95	3,02	5,33	3,97
Mínimo	43,75	47,08	75,00	71,43	44,44
Máximo	562,50	444,44	562,50	600,00	484,50



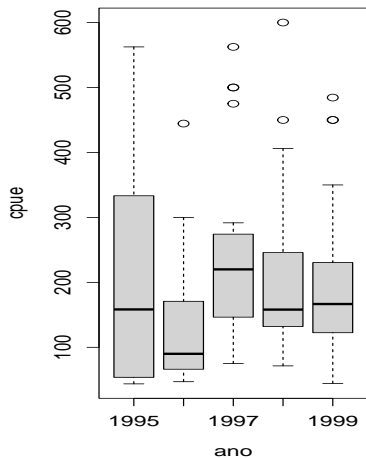
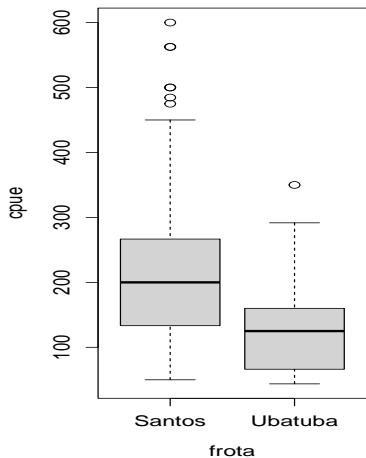
Medidas resumo:cpue (Santos)

	Ano				
Medida resumo	1995	1996	1997	1998	1999
Média	229,37	193,19	262,67	210,39	199,88
DP	148,07	132,55	153,60	122,95	103,45
Var.	21924,60	17569,11	23593,58	15117,82	10702,10
CV(%)	64,56	68,61	58,48	58,44	51,76
CA	0,48	0,81	0,77	1,57	0,91
Curtose	2,33	2,47	2,35	5,17	3,68
Mínimo	50,00	80,00	75,00	71,43	53,00
Máximo	562,50	444,44	562,50	600,00	484,50

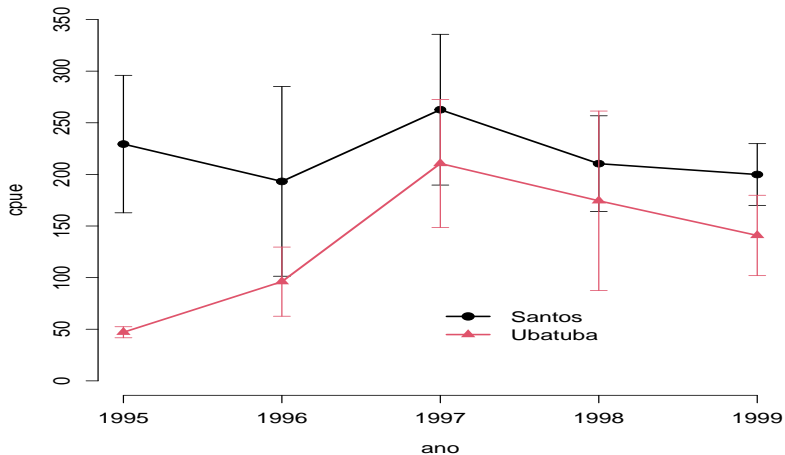
Medidas resumo:cpue (Ubatuba)

	Ano				
Medida resumo	1995	1996	1997	1998	1999
Média	47,08	96,09	210,56	174,43	140,85
DP	4,73	59,19	77,51	99,16	71,59
Var.	22,40	3503,45	6007,41	9833,27	5125,16
CV(%)	10,05	61,60	36,81	56,85	50,83
CA	0,65	1,53	-0,35	1,42	0,62
Curtose	1,50	4,75	1,60	3,15	2,67
Mínimo	43,75	47,08	100,00	111,43	44,44
Máximo	52,50	250,00	291,67	350,00	291,67

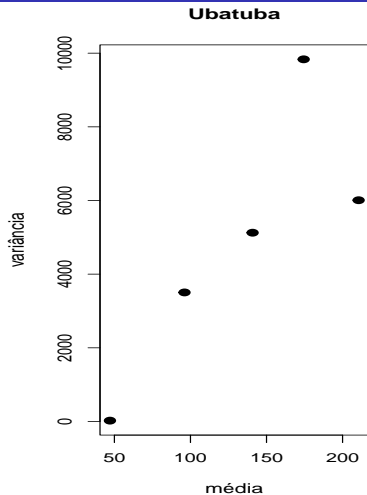
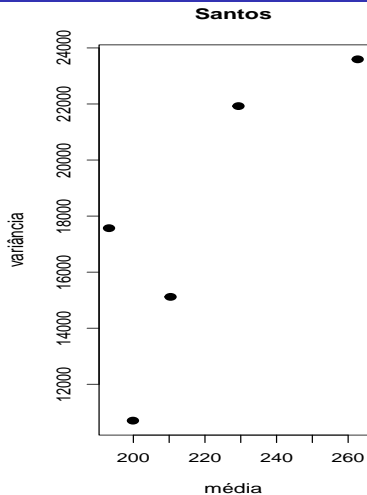
Boxplot - cpue (ano e frota)



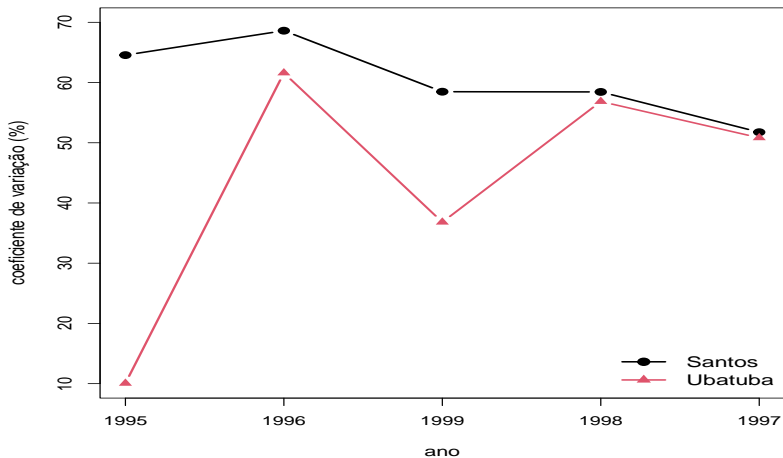
Gráficos de perfis médios - frota x ano



Média \times variância



Coefficiente de variação ao longo dos anos



Comentários (análise descritiva)

- Há indícios de existência de interação entre os fatores (Frota e Ano), pois os perfis não sugerem paralelismo.
- Há indícios de que o modelo gama possa não ser apropriado, pois algumas distribuições (cpue ao longo das combinações entre os níveis de Frota e Ano) apresentam curtose menor do que 3 e os CV's não parecem ser constantes ao longo daquelas.
- Aparentemente, as cpue's de cada frota apresentam comportamento oscilante ao longo dos anos, com a frota de Santos apresentando cpue uniformemente maior do que aquela de Ubatuba.

Modelo inicial (modelo 1)

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

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

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

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

$$\ln(\mu_{ijk}) = \alpha + \beta_i + \gamma_j + (\beta\gamma)_{ij},$$

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

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

Modelo

- Y_{ijkl} : é a cpue obtida pelo l -ésimo barco, da i -ésima frota, no j -ésimo ano, no k -ésimo trimestre.
- Parâmetros:

Média	representação	descrição
μ_{11}	e^{α}	cpue média: Santos em 1995
μ_{12}	$e^{\alpha+\gamma_2}$	cpue média: Santos em 1996
μ_{13}	$e^{\alpha+\gamma_3}$	cpue média: Santos em 1997
μ_{14}	$e^{\alpha+\gamma_4}$	cpue média: Santos em 1998
μ_{15}	$e^{\alpha+\gamma_5}$	cpue média: Santos em 1999

Modelo

■ (Cont.) Parâmetros:

Média	representação	descrição
μ_{21}	$e^{\alpha+\beta_2}$	cpue média: Ubatuba em 1995
μ_{22}	$e^{\alpha+\beta_2+\gamma_2+(\beta\gamma)_{22}}$	cpue média: Ubatuba em 1996
μ_{23}	$e^{\alpha+\beta_2+\gamma_3+(\beta\gamma)_{23}}$	cpue média: Ubatuba em 1997
μ_{24}	$e^{\alpha+\beta_2+\gamma_4+(\beta\gamma)_{24}}$	cpue média: Ubatuba em 1998
μ_{25}	$e^{\alpha+\beta_2+\gamma_5+(\beta\gamma)_{25}}$	cpue média: Ubatuba em 1999

Modelo (algumas interpretações)

- $\frac{\mu_{1j}}{\mu_{11}} = e^{\gamma_j}, j = 2, 3, 4, 5$, parâmetro que diferencia a cpue média, para a frota de Santos, dos anos de 1996 a 1999 em relação ao ano de 1995.
- $\frac{\mu_{2j}}{\mu_{21}} = e^{\beta_2 + \gamma_j + (\beta\gamma)_{2j}}, j = 2, 3, 4, 5$, parâmetro que diferencia a cpue média, para a frota de Ubatuba, dos anos de 1996 a 1999 em relação ao ano de 1995.
- $\frac{\mu_{2j}}{\mu_{1j}} = e^{\beta_2 + (\beta\gamma)_{2j}}, j = 2, 3, 4, 5$, parâmetro que diferencia a cpue média, entre as frotas, para o mesmo ano.

Modelo (algumas interpretações)

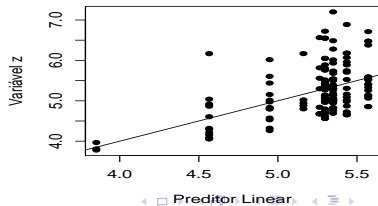
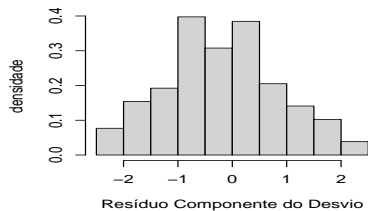
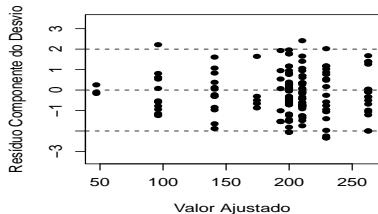
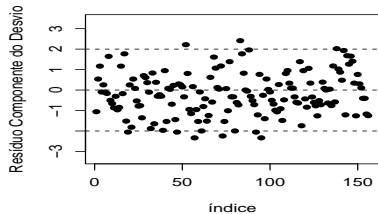
- Uma forma de definir interação entre dois fatores, é quando a diferença das médias entre os níveis de um dos fatores é a mesma ao longo dos níveis do outro fator.
- Em nosso problema, equivale a dizer que a diferença das médias da cpue entre as froas se mantêm constante ao longo dos anos, ou seja:

$$\mu_{21} - \mu_{11} = \mu_{22} - \mu_{12} = \mu_{23} - \mu_{13} = \mu_{24} - \mu_{14} = \mu_{25} - \mu_{15}$$

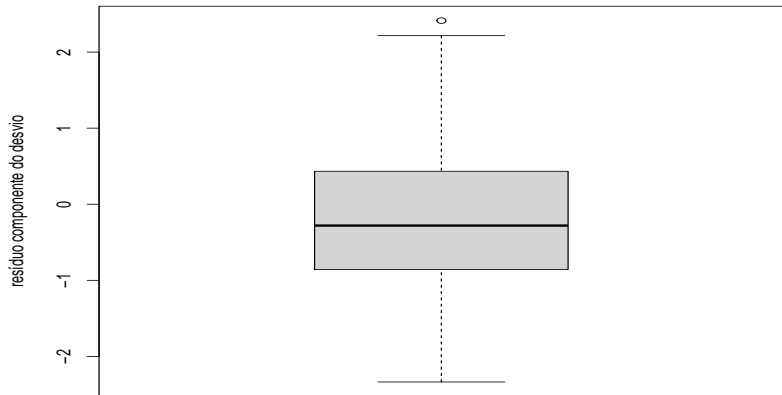
- A igualdade acima equivale à (provar)

$$(\beta\gamma)_{22} = (\beta\gamma)_{23} = (\beta\gamma)_{24} = (\beta\gamma)_{25} = 0$$

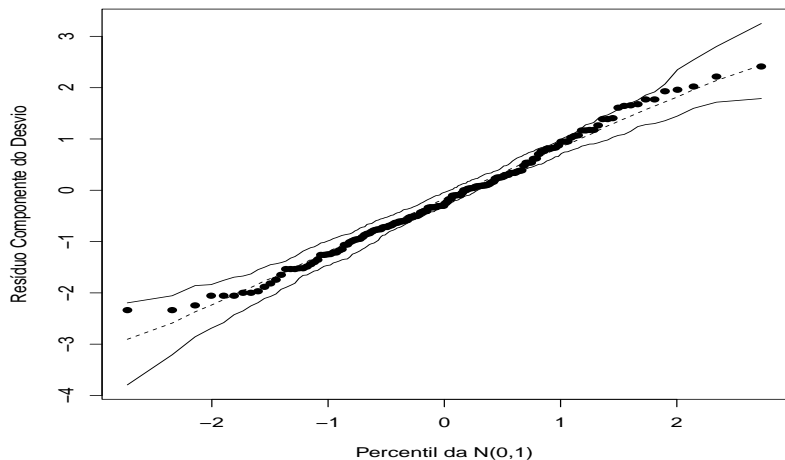
Gráficos de diagnóstico: modelo 1



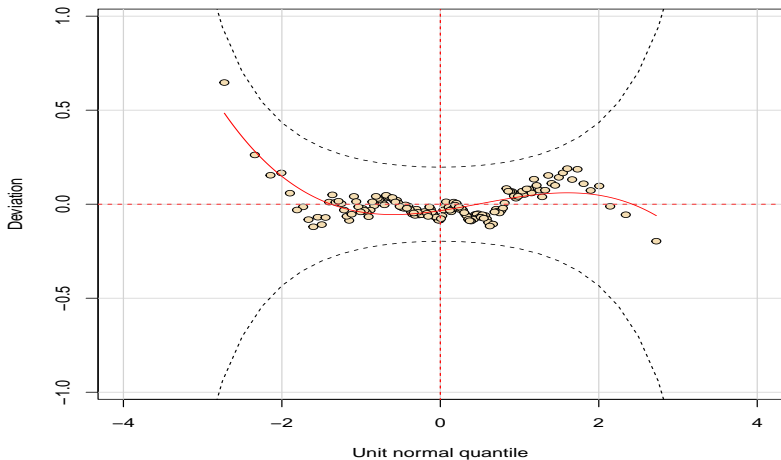
Boxplot para os resíduos: modelo 1



Envelope para os resíduos: modelo 1



Worm plot para os resíduos: modelo 1



Modelo 1

- Teste de Shapiro Wilks para o RCD: $p = 0,3259$ o que sugere normalidade.
- Desvio : $p = 0,15413$.
- Embora os resultados indiquem normalidade tanto do RCD quanto do RQ, há uma certa heterocedasticidade não contemplada pelo modelo.
- Alternativas: considerar as outras covariáveis e/ou um modelo com dispersão variável (veja Paula, 2024).
- Contudo, continuaremos com o modelo em questão.

Estimativas dos parâmetros: Modelo gama 1

Parâmetro	Estimativa	EP	IC(95%)	Estat. Z	p-valor
α	5,434	0,12	[5,20 ; 5,68]	44,243	< 0,0001
β_2	-1,58	0,33	[-2,24 ; -0,93]	-4,760	< 0,0001
γ_2	-0,17	0,23	[-0,61 ; 0,27]	-0,760	0,4470
γ_3	0,14	0,18	[-0,22 ; 0,49]	0,758	0,4482
γ_4	-0,09	0,16	[-0,40 ; 0,23]	-0,539	0,5901
γ_5	-0,14	0,15	[-0,42 ; 0,15]	-0,942	0,3461

Estimativas dos parâmetros: Modelo gama 1

(Cont.)

Parâmetro	Estimativa	EP	IC(95%)	Estat. Z	p-valor
$(\beta\gamma)_{22}$	0,89	0,41	[0,08 ; 1,69]	2,144	0,0320
$(\beta\gamma)_{23}$	1,36	0,42	[0,54 ; 2,18]	3,253	0,0011
$(\beta\gamma)_{24}$	1,40	0,42	[0,57 ; 2,22]	3,303	0,0010
$(\beta\gamma)_{25}$	1,23	0,37	[0,50 ; 1,96]	3,309	0,0009
ϕ	3,49	0,38	[2,75 ; 4,23]	-	-

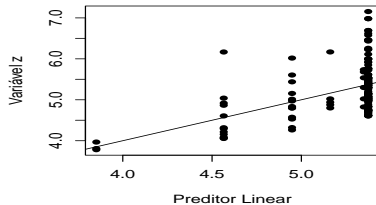
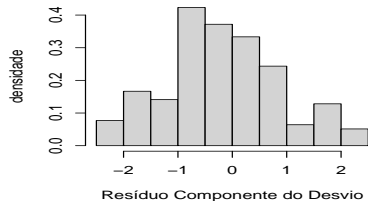
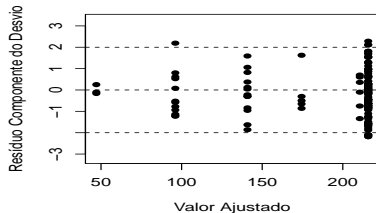
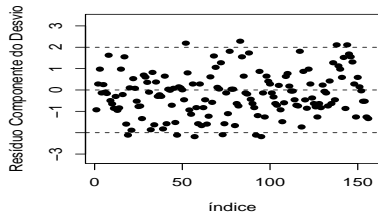
Aparentemente, embora haja indícios de presença de interação (nesse caso $(\beta\gamma)_{1j} \neq 0, j = 2, \dots, 5$), nenhum dos parâmetros γ_j são diferentes de zero (além disso, modelo gama é preferível ao exponencial).

Ajuste do modelo

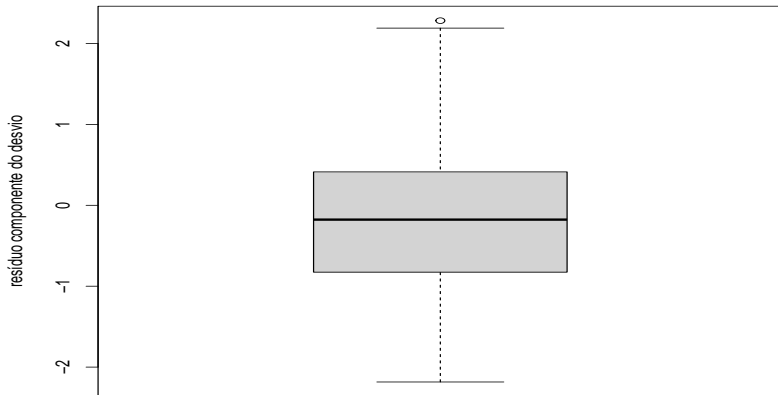
- Com efeito obtivemos o teste de nulidade dos parâmetros $\gamma_j, j = 2, \dots, 5$ resultou em: análise do desvio ($p = 0,7326$), teste $C\beta = M$ (Wald) ($p = 0,4239$).
- Exercício: escrever as hipóteses acima na forma $H_0 : C\beta = M$ vs $H_1 : C\beta \neq M$.
- Critérios de Informação (M:modelo):

M	AIC	BIC	AICc	SABIC	HQCIC	CAIC
1	1871,36	1904,91	1873,20	1870,09	1884,99	1915,91
2	1867,32	1888,67	1868,08	1866,52	1875,99	1895,67

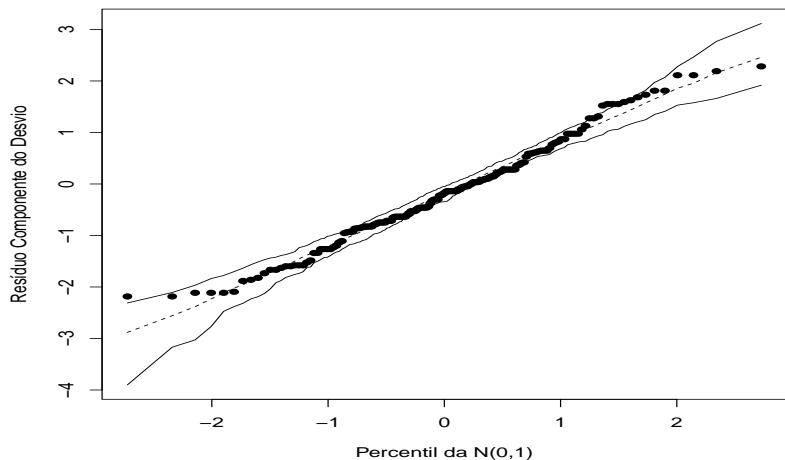
Gráficos de diagnóstico: modelo 2



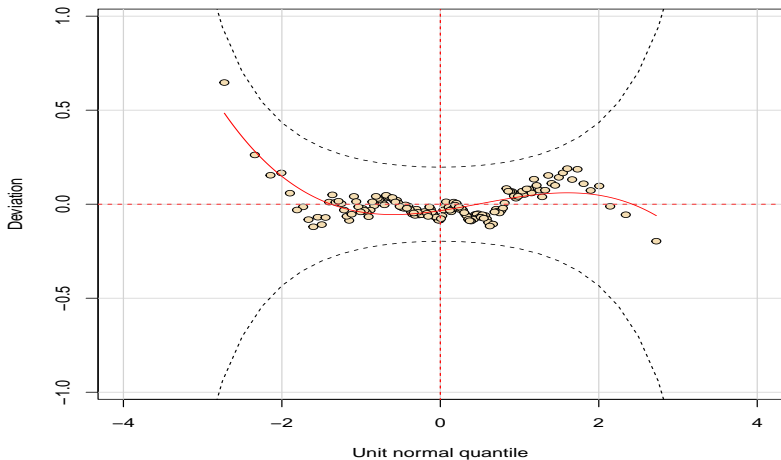
Boxplot para os resíduos: modelo 2



Envelope para os resíduos: modelo 2



Worm plot para os resíduos: modelo 2



Modelo 2

- Teste de Shapiro Wilks para o RCD: $p = 0,0729$ o que sugere normalidade (marginal).
- Desvio : $p = 0,2796$, o que indica que o modelo está bem ajustado (supondo que a aproximação pela χ^2_{146} seja apropriada).
- Há uma tendência nas caudas inferior e superior do RCD.
- Embora os resultados indiquem normalidade (ainda que marginal) tanto do RCD quanto do RQ, há uma certa heterocedasticidade não contemplada pelo modelo.

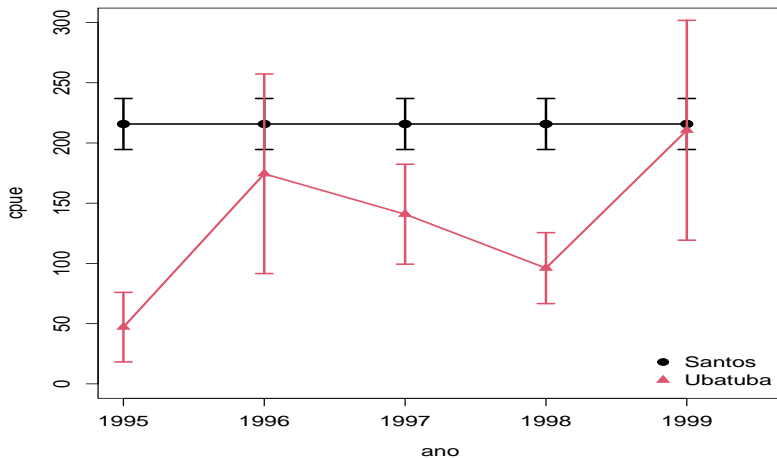
Modelo 2

- Alternativas: considerar as outras covariáveis e/ou um modelo com dispersão variável (veja Paula, 2024).
- Contudo, continuaremos com o modelo em questão.

Estimativas dos parâmetros: Modelo gama 2

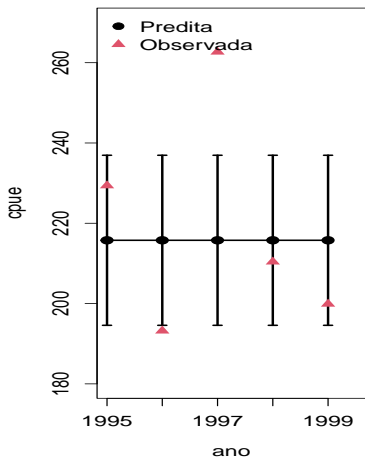
Parâmetro	Estimativa	EP	IC(95%)	Estat. Z	p-valor
α	5,37	0,05	[5,28;5,47]	107,30	<0,0001
β_2	-1,52	0,32	[-2,14;-0,90]	-4,81	<0,0001
$(\beta\gamma)_{22}$	0,71	0,35	[0,03;1,40]	2,04	0,0413
$(\beta\gamma)_{23}$	1,50	0,38	[0,75;2,25]	3,91	0,0001
$(\beta\gamma)_{24}$	1,31	0,40	[0,53;2,09]	3,31	0,0009
$(\beta\gamma)_{25}$	1,10	0,35	[0,42;1,78]	3,16	0,0016
ϕ	3,41	0,37	[2,69;4,13]	-	-

Médias preditas \times observadas

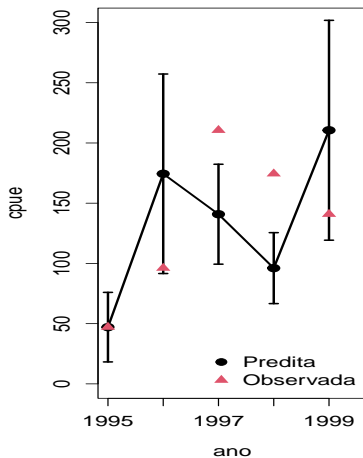


Médias previstas \times observadas

Santos



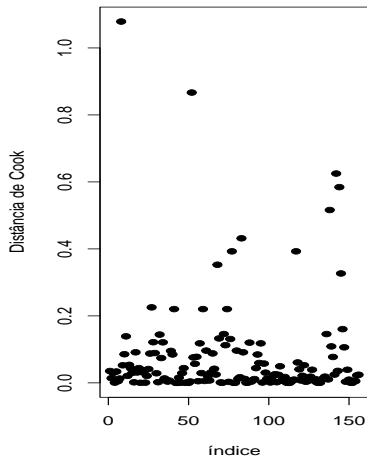
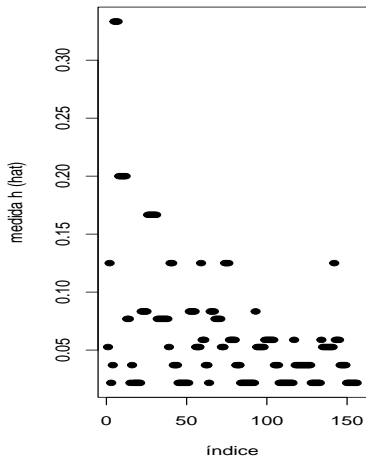
Ubatuba



Comentários

- Os resultados indicam que, em média, a cpue da frota de Santos é constante, ao longo dos anos e maior do que a da frota de Ubatuba, com exceção dos anos de 1996 e 1999.
- Por outro lado, há uma tendência crescente da cpue da frota de Ubatuba.
- Em geral, a predição das cpue's médias não foram boas.

Medidas de alavancagem e influência



Análise de sensibilidade

Parâmetro (α)

Observações	Est.	EP	IC(95%)	z value	p-valor
todas	5,37	0,05	[5,28;5,47]	107,30	< 0,0001
#-5	5,37	0,05	[5,28;5,47]	106,99	< 0,0001
#-6	5,37	0,05	[5,28;5,47]	106,99	< 0,0001
#-8	5,37	0,05	[5,28;5,47]	107,98	< 0,0001

Análise de sensibilidade

Parâmetro (β_2)

Observações	Est.	EP	IC(95%)	z value	p-valor
todas	-1,52	0,32	[-2,14;-0,90]	-4,81	< 0,0001
#-5	-1,58	0,39	[-2,34;-0,82]	-4,08	< 0,0001
#-6	-1,58	0,39	[-2,34;-0,82]	-4,08	< 0,0001
#-8	-1,52	0,31	[-2,14;-0,91]	-4,84	< 0,0001

Análise de sensibilidade

Parâmetro $((\beta\gamma)_{22})$

Observações	Est.	EP	IC(95%)	z value	p-valor
todas	0,71	0,35	[0,03;1,40]	2,04	0,0413
#-5	0,77	0,41	[-0,04;1,59]	1,86	0,0626
#-6	0,77	0,41	[-0,04;1,59]	1,86	0,0625
#-8	0,71	0,35	[0,03;1,39]	2,05	0,0401

Análise de sensibilidade

Parâmetro $((\beta\gamma)_{23})$

Observações	Est.	EP	IC(95%)	z value	p-valor
todas	1,50	0,38	[0,75;2,25]	3,91	0,0001
#-5	1,56	0,44	[0,69;2,43]	3,51	0,0004
#-6	1,56	0,44	[0,69;2,43]	3,51	0,0004
#-8	1,50	0,38	[0,75;2,24]	3,93	0,0001

Análise de sensibilidade

Parâmetro $((\beta\gamma)_{24})$

Observações	Est.	EP	IC(95%)	z value	p-valor
todas	1,31	0,40	[0,53;2,09]	3,31	0,0009
#-5	1,37	0,45	[0,48;2,26]	3,01	0,0026
#-6	1,37	0,45	[0,48;2,26]	3,01	0,0026
#-8	1,02	0,41	[0,21;1,83]	2,48	0,0131

Análise de sensibilidade

Parâmetro $((\beta\gamma)_{25})$

Observações	Est.	EP	IC(95%)	z value	p-valor
todas	1,10	0,35	[0,42;1,78]	3,16	0,0016
#-5	1,16	0,41	[0,35;1,96]	2,80	0,0051
#-6	1,16	0,41	[0,35;1,96]	2,80	0,0051
#-8	1,10	0,34	[0,42;1,77]	3,18	0,0015

Comentários

- O mau desempenho da predição, provavelmente, é devido ao mal ajuste do modelo.
- Aparentemente, não houve mudanças significativas dos parâmetros de regressão, com a retirada individual, de cada um dos três pontos destacados (dois como candidatos a alavanca e outro pela distância de Cook).
- Alternativas: considerar as outras covariáveis e/ou um modelo com dispersão variável (veja Paula, 2024).