

Análise de resíduos em modelos hierárquicos/multiníveis: parte 2

Prof. Caio Azevedo (grande parte do material apresentado foi extraído do livro Modelos de regressão com apoio computacional do

Prof. Gilberto A. Paula

[link](#) e da Dissertação de Mestrado do Prof. Juvêncio Nobre: [link](#))

Recapitulando

- Estrutura norteadora da verificação da qualidade de ajuste de um modelo hierárquico :
 - Verificar o comportamento dos RCM: normalidade / homocedasticidade / independência entre as UAE's do nível 2 (comportamento do erro condicional ξ). Eventuais problemas podem estar relacionados à má especificação de ξ e/ou do preditor ($Z\gamma$).

Cont.

- Verificar o comportamento dos RCP e RMP: normalidade / homocedasticidade / independência (difícil separar a influência das observações dos níveis 1 e 2).
 - Comportamento dentro do esperado (ξ): eventuais problemas podem estar relacionados à má especificação da estrutura de $\mathbf{X}u$ e/ou do preditor do modelo $\mathbf{Z}\gamma$.
 - Comportamento fora do esperado (ξ): eventuais problemas podem estar relacionados à má especificação da estrutura de $\mathbf{X}u$ e/ou do preditor do modelo $\mathbf{Z}\gamma$ e/ou de (ξ).

Cont.

- Verificar o comportamento dos efeitos aleatórios (Xu):
 - Comportamento dentro do esperado (ξ): eventuais problemas podem estar relacionados à má especificação da estrutura de Xu e/ou do preditor do modelo $Z\gamma$.
 - Comportamento fora do esperado (ξ): eventuais problemas podem estar relacionados à má especificação da estrutura de Xu e/ou do preditor do modelo $Z\gamma$ e/ou de (ξ).

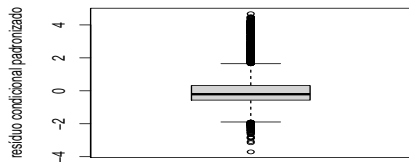
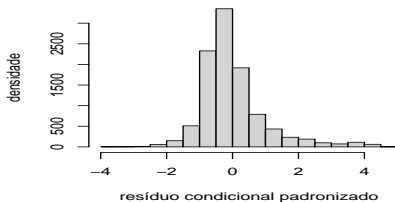
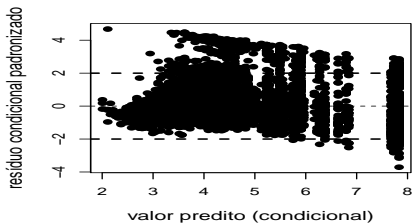
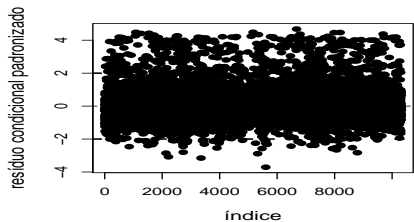
Voltando ao Exemplo 3: Achievement data ([link](#))

- Já vimos (aqui) que há efeito (positivo) do “vocabulário” na “leitura” e que tal efeito varia entre as escolas.
- Percebemos também uma aparente ausência de normalidade de ambos os efeitos aleatórios (intercepto e coeficiente angular).
- Ao se utilizar a função “ana.res.MH” (a qual usa a função “residdiag.nlme”) houve problemas de memória (uma solução seria usar pacotes que auxiliam na manipulação de grandes bases de dados).

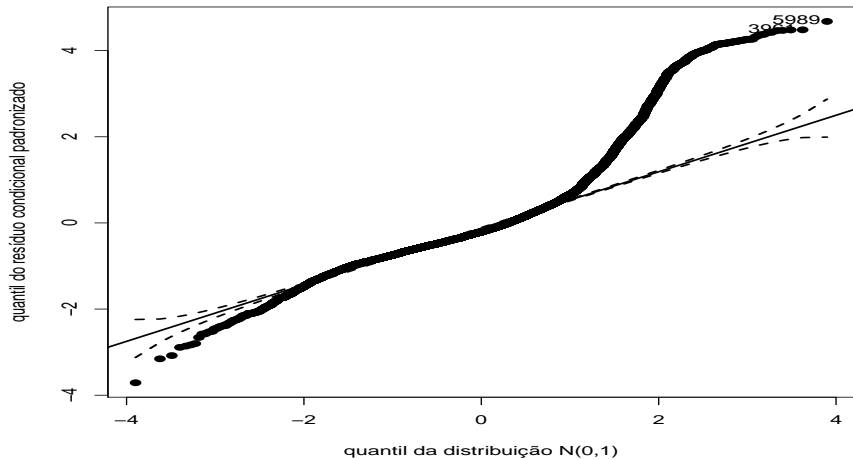
Voltando ao Exemplo 3: Achievement data ([link](#))

- A função “lme” calcula somente o RCP, enquanto que a função “lmer” calcula o resíduo de Pearson.
- Alternativas à função “ana.res.MH” (“residdiag.nlm”):
 - Usar os outputs de pelo menos uma das funções acima e gerar, manualmente, os resíduos.
 - Utilizar pacotes que lidam com bases de dados grandes junto com a função “ana.res.MH” (“residdiag.nlm”).
 - Pacote do R “HLMdiag”: [manual](#), [site](#), [artigo](#).
 - Pacote do R “DHARMa”: [manual](#), [site](#), [vinhetas](#).
- Vamos realizar uma análise de diagnóstico (slides seguintes) usando somente o RCP via função “lme”.

RCP: diagnóstico



RCP: qqplot



Comentários

- Os resíduos apresentam assimetria positiva, heterocedasticidade e presença de outliers.
- Os resultados referentes aos efeitos aleatórios ([link](#)) também indicam não normalidade deles.
- Há, portanto, indicações de que o modelo não está bem ajustado aos dados. Como não se conseguiu realizar análises dos RCM e do RMP, devido a natureza do problema e dos resultados da análise descritiva, conjecturamos que, ao menos inicialmente, devemos mudar a distribuição dos erros condicionais ou utilizar um **MLG** hierárquicos com distribuição do tipo binomial.

Voltando ao Exemplo 4: Produção de aveia ([link](#))

- Vamos analisar os dados desse experimento feito em [parcela dividida](#) (split-plot, um tipo particular de [experimento com vários fatores](#)).
- Vimos duas possibilidades em termos de como considerar a estrutura hierárquica dos dados, quais sejam : 2 e 3 níveis ([aqui](#)).
- Entretanto, vamos considerar uma estrutura e, conseqüentemente, um modelo, de dois níveis, diferente daquele apresentando.

Voltando ao Exemplo 4: Produção de aveia ([link](#))

- Resposta: produção de aveia (produção)
- Fatores:
 - Variedade (qualitativo): Golden Rain (GR), Marvellous (M) e Victory (V).
 - Manjar (quantitativo): 0,0 wct; 0,2 wct; 0,3 wct; 0,4 wct.
 - Bloco (qualitativo/quantitativo): fator de controle (I, II, III, IV, V e VI).
- Vamos (re)ver a estrutura do problema e os dados nos próximos dois slides (respectivamente, “croqui de campo” e tabela com os dados).

	Variedade					
Bloco	Marvellous		Golden rain		Victory	
I	0,0 wct	0,2 wct	0,0 wct	0,2 wct	0,0 wct	0,2 wct
	0,4 wct	0,6 wct	0,4 wct	0,6 wct	0,4 wct	0,6 wct
II	0,0 wct	0,2 wct	0,0 wct	0,2 wct	0,0 wct	0,2 wct
	0,4 wct	0,6 wct	0,4 wct	0,6 wct	0,4 wct	0,6 wct
III	0,0 wct	0,2 wct	0,0 wct	0,2 wct	0,0 wct	0,2 wct
	0,4 wct	0,6 wct	0,4 wct	0,6 wct	0,4 wct	0,6 wct
IV	0,0 wct	0,2 wct	0,0 wct	0,2 wct	0,0 wct	0,2 wct
	0,4 wct	0,6 wct	0,4 wct	0,6 wct	0,4 wct	0,6 wct
V	0,0 wct	0,2 wct	0,0 wct	0,2 wct	0,0 wct	0,2 wct
	0,4 wct	0,6 wct	0,4 wct	0,6 wct	0,4 wct	0,6 wct
VI	0,0 wct	0,2 wct	0,0 wct	0,2 wct	0,0 wct	0,2 wct
	0,4 wct	0,6 wct	0,4 wct	0,6 wct	0,4 wct	0,6 wct

Bloco	Variedade	manjar			
		0.0cwt	0.2cwt	0.4cwt	0.6cwt
I	Victory	111	130	157	174
	Golden rain	117	114	161	141
	Marvellous	105	140	118	156
II	Victory	61	91	97	100
	Golden rain	70	108	126	149
	Marvellous	96	124	121	144
III	Victory	68	64	112	86
	Golden rain	60	102	89	96
	Marvellous	89	129	132	124
IV	Victory	74	89	81	122
	Golden rain	64	103	132	133
	Marvellous	70	89	104	117
V	Victory	62	90	100	116
	Golden rain	80	82	94	126
	Marvellous	63	70	109	99
VI	Victory	53	74	118	113
	Golden rain	89	92	86	104
	Marvellous	97	99	119	121

Estrutura hierárquica

- À rigor podemos considerar estruturas de dois e três níveis.
- Dois níveis:
 - Excluindo blocos: nível 2 - parcelas maiores (variedade) ; nível 1 - parcelas menores (manjar).
 - Considerando os blocos: nível 2 - parcelas maiores (variedade) ; nível 1 - parcelas menores (manjar). Os blocos funcionam como grupos (à exemplo dos [dados de Potthoff and Roy](#)).
 - Nesse dois casos a estrutura de regressão varia entre as parcelas maiores (estrutura 1) ou parcelas maiores ao longo dos grupos (estrutura 2).

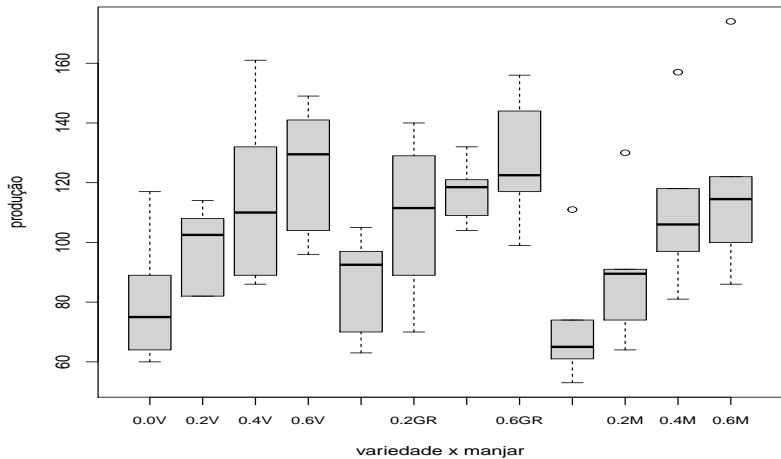
Estrutura hierárquica

- Três níveis (considerando os blocos): nível 3 - blocos; nível 2 - parcelas maiores (variedade); nível 1 - parcelas menores (manjar).
- Nesse caso, teríamos os coeficientes de regressão (CR) do nível 1, variando ao longo dos elementos do nível 2, e os CR's do nível 2, variando ao longo dos elementos do nível 3 (veremos ao longo do curso, modelos com essa estrutura).

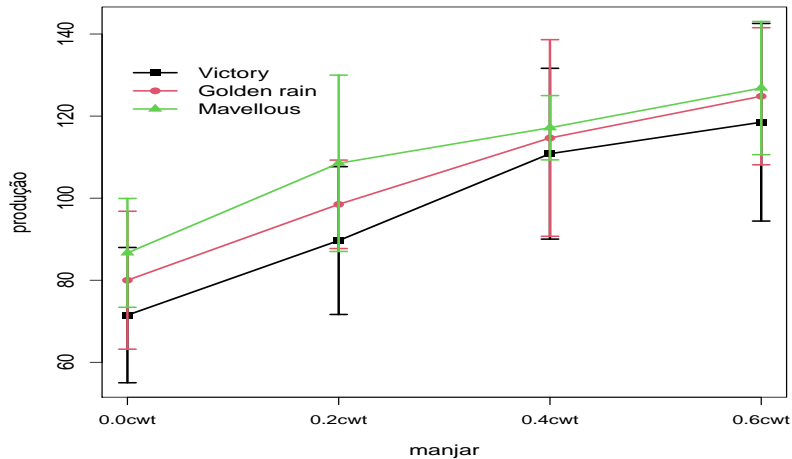
Medidas resumo

V	M	media	dp	var.	cv(%)	min.	med.	max.	ca	kur	n
GR	0,0cwt	80,00	21,00	441,20	26,26	60	75,00	117	0,91	2,61	6
	0,2cwt	98,50	13,47	181,50	13,68	82	102,50	114	-0,38	1,56	6
	0,4cwt	114,67	29,94	896,67	26,11	86	110,00	161	0,46	1,79	6
	0,6cwt	124,83	20,88	435,77	16,72	96	129,50	149	-0,34	1,63	6
M	0,0cwt	86,67	16,57	274,67	19,12	63	92,50	105	-0,47	1,66	6
	0,2cwt	108,50	26,85	721,10	24,75	70	111,50	140	-0,25	1,65	6
	0,4cwt	117,17	9,79	95,77	8,35	104	118,50	132	0,12	2,18	6
	0,6cwt	126,83	20,29	411,77	16,00	99	122,50	156	0,19	2,01	6
V	0,0cwt	71,50	20,60	424,30	28,81	53	65,00	111	1,33	3,43	6
	0,2cwt	89,67	22,51	506,67	25,10	64	89,50	130	0,86	2,94	6
	0,4cwt	110,83	26,01	676,57	23,47	81	106,00	157	0,85	2,82	6
	0,6cwt	118,50	30,09	905,50	25,39	86	114,50	174	1,05	3,14	6

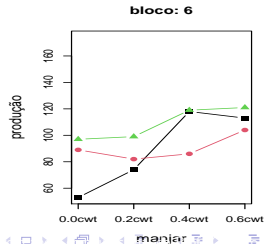
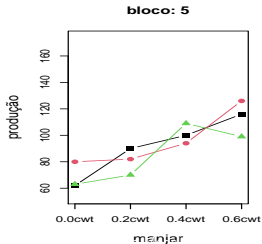
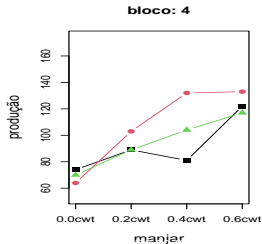
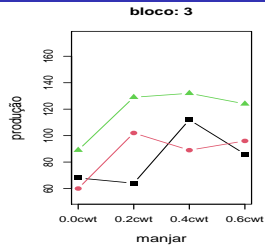
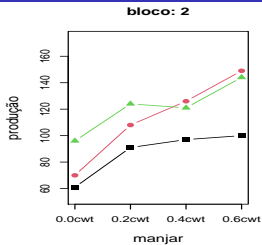
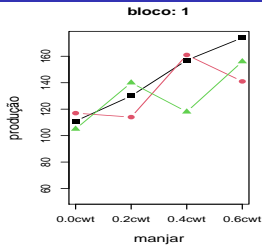
Box-plot



Gráficos de perfis médios observados



Gráficos de perfis individuais observados



Comentários

- Observamos, através das medidas resumo, uma tendência crescente na produção média, à medida que o nível do manjar “aumenta”, bem como, em ordem crescente (V, GR e M).
- Presença de heterocedasticidade. Assimetria e curtose diferente da distribuição normal, para a distribuição da “produção” em função das combinações dos níveis de variedade e manjar.

Comentários

- Vê-se duas possibilidades:
 - Presença de interação e dos efeitos dos fatores principais ou
 - ausência de interação com efeitos somente de manjar.
- Efeito de bloco: os perfis de produção em função das combinações dos níveis de variedade x manjar diferem ao longo dos blocos.

Modelo inicial

- Nesse caso, um modelo inicial apropriado é :

$$Y_{jik} = \mu_{jk} + \alpha_j + \beta_i + (\alpha\beta)_{ji} + \delta_k + \xi_{jik} \text{ [nível 1, sub-parcelas(manjar)]}$$

$$\mu_{jk} = \mu + u_{jk} \text{ [nível 2, parcelas maiores(variedades) ao longo dos blocos]}$$

$$j = 1(\text{Golden Rain}), 2(\text{Marvellous}), 3(\text{Victory}) \text{ [variedade],}$$

$$i = 1(0,0\text{cwt}), 2(0,2\text{cwt}), 3(0,4\text{cwt}), 4(0,6\text{cwt}) \text{ [manjar],}$$

$$k = 1, \dots, 6 \text{ (bloco), } \alpha_1 = \beta_1 = (\alpha\beta)_{1j} = (\alpha\beta)_{i1}, \forall i, j$$

- $\xi_{jik} \stackrel{iid}{\sim} N(0, \sigma^2)$, $u_{jk} \stackrel{iid}{\sim} N(0, \psi)$ e $\xi_{jik} \perp u_{jk}, \forall j, i, k$.

- $\mathcal{E}(Y_{jik}|u_{jk}) = \mu_{jk} + \alpha_j + \beta_i + (\alpha\beta)_{ji} + \delta_k$.

- $\mathcal{V}(Y_{jik}|u_{jk}) = \sigma^2$.

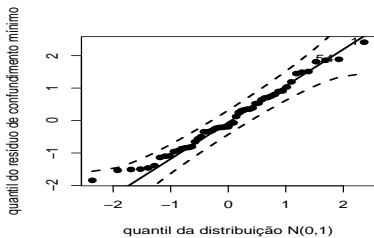
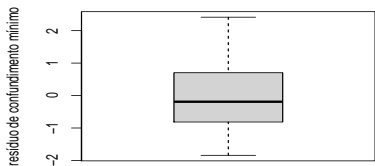
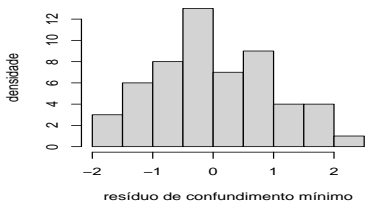
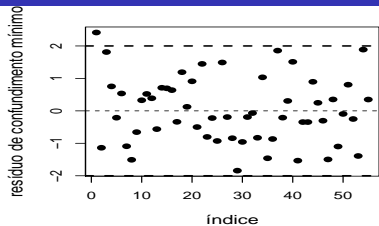
Modelo inicial

- $\mathcal{E}(Y_{jik}) = \mu + \alpha_j + \beta_i + (\alpha\beta)_{ji} + \delta_k$. $\mathcal{V}(Y_{jik}) = \sigma^2 + \psi$.
- $\text{Cov}(Y_{jik}, Y_{ji'k}) = \psi$ e $\text{Corr}(Y_{jik}, Y_{ji'k}) = \frac{\psi}{\psi + \sigma^2}, i \neq i'$ (
 $\text{Cov}(Y_{jik}, Y_{jik'}) = \text{Corre}(Y_{jik}, Y_{j'i'k'}) = 0, \forall j \neq j', k, k', i, i'$).
- Como estamos considerando interação entre os fatores principais, as interpretações dos efeitos fixos tornam-se mais “específicas” (veja tabela seguinte).
- Os parâmetros $(\mu, \alpha, \beta, \delta)'$, respectivamente, estão relacionados com (o grupo de referência, variedade, manjar, bloco) mas, seus impactos nas médias podem depender dos parâmetros associados à interação $(\alpha\beta)$.

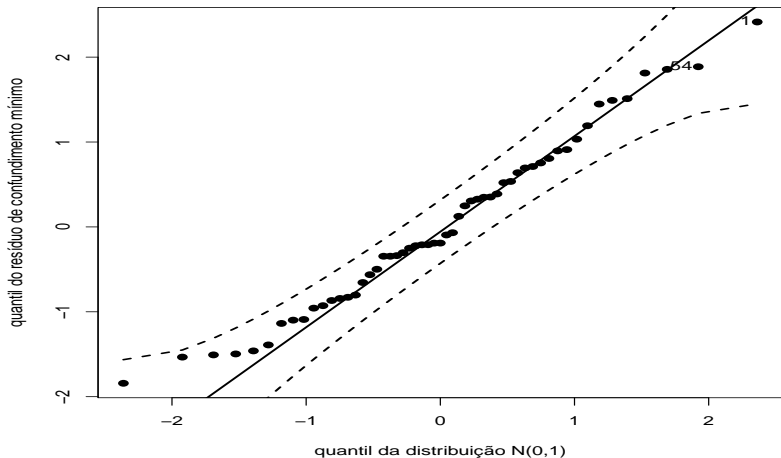
Médias marginais (MM), em relação aos blocos

Variedade	Manjar	MM
GR	0,0 wct	$\mu + \sum_{k=1}^6 \delta_k/6$
	0,2 wct	$\mu + \beta_2 + \sum_{k=1}^6 \delta_k/6$
	0,4 wct	$\mu + \beta_3 + \sum_{k=1}^6 \delta_k/6$
	0,6 wct	$\mu + \beta_4 + \sum_{k=1}^6 \delta_k/6$
M	0,0 wct	$\mu + \alpha_2 + \sum_{k=1}^6 \delta_k/6$
	0,2 wct	$\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22} + \sum_{k=1}^6 \delta_k/6$
	0,4 wct	$\mu + \alpha_2 + \beta_3 + (\alpha\beta)_{23} + \sum_{k=1}^6 \delta_k/6$
	0,6 wct	$\mu + \alpha_2 + \beta_4 + (\alpha\beta)_{24} + \sum_{k=1}^6 \delta_k/6$
V	0,0 wct	$\mu + \alpha_3 + \sum_{k=1}^6 \delta_k/6$
	0,2 wct	$\mu + \alpha_3 + \beta_2 + (\alpha\beta)_{32} + \sum_{k=1}^6 \delta_k/6$
	0,4 wct	$\mu + \alpha_3 + \beta_3 + (\alpha\beta)_{33} + \sum_{k=1}^6 \delta_k/6$
	0,6 wct	$\mu + \alpha_3 + \beta_4 + (\alpha\beta)_{34} + \sum_{k=1}^6 \delta_k/6$

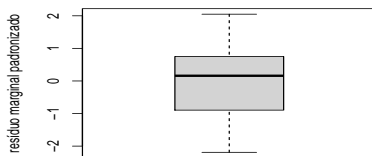
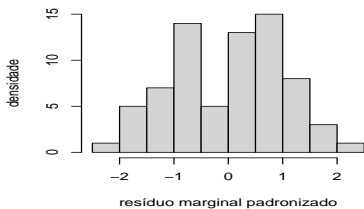
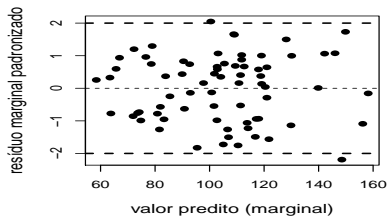
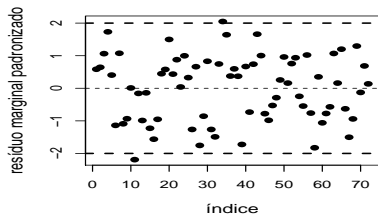
Gráficos de diagnóstico (RCM): Modelo 1



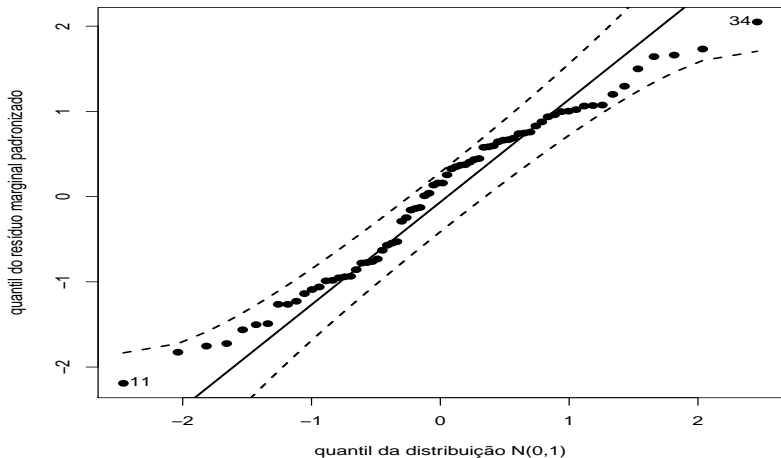
QQplot (RCM): Modelo 1



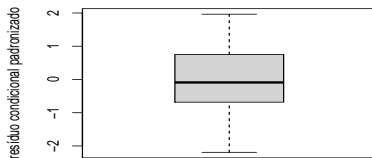
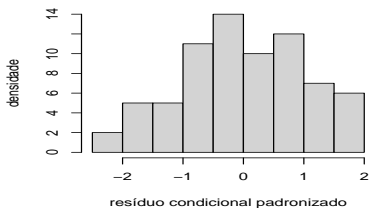
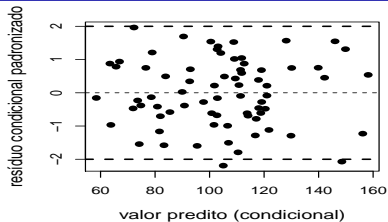
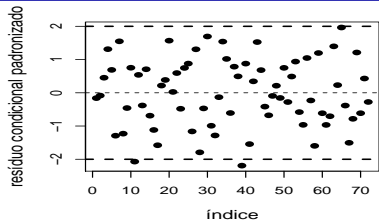
Gráficos de diagnóstico (RMP): Modelo 1



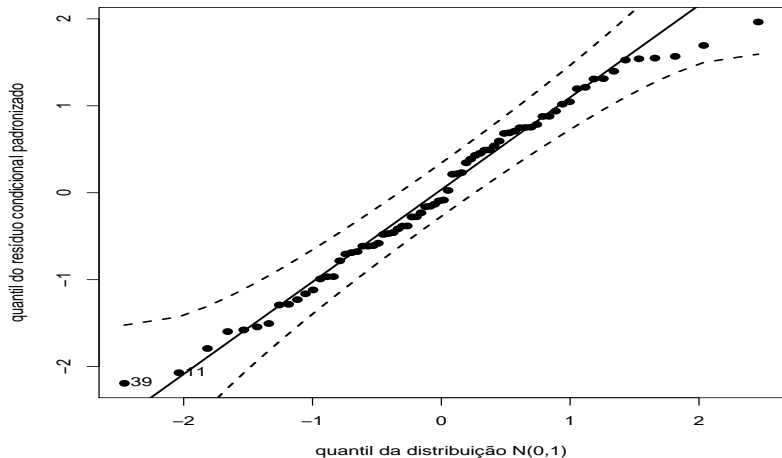
QQplot (RMP): Modelo 1



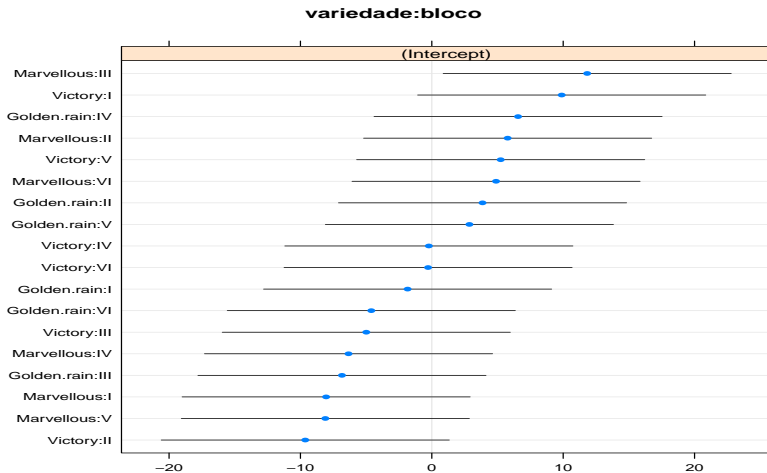
Gráficos de diagnóstico (RCP): Modelo 1



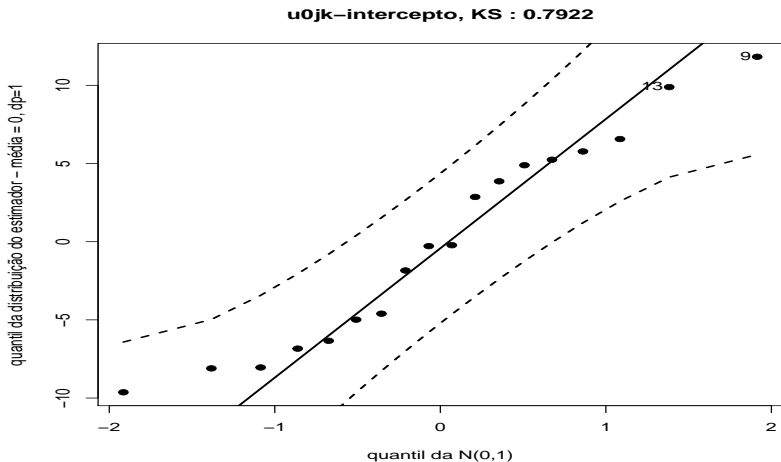
QQplot (RCP): Modelo 1



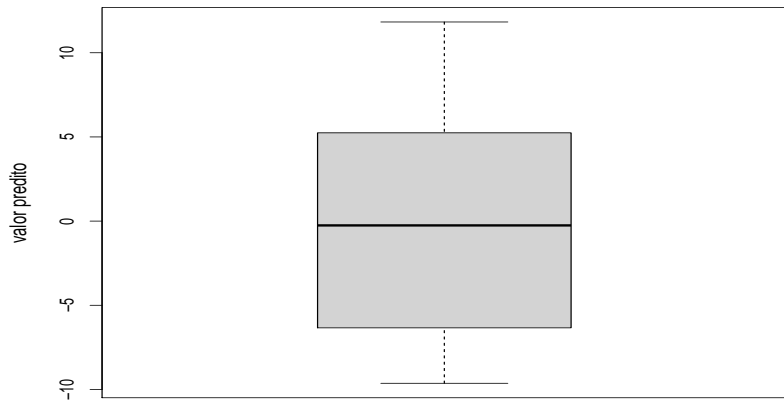
Estimativas e IC dos EA's: Modelo 1



QQplot dos EA's: Modelo 1



Boxplot dos EA's: Modelo 1



Comentários

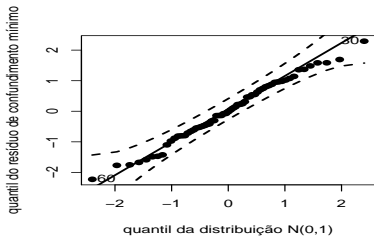
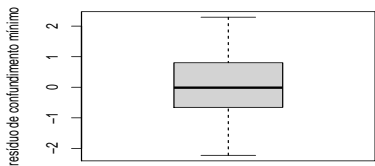
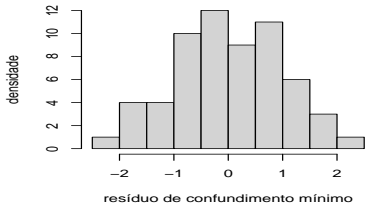
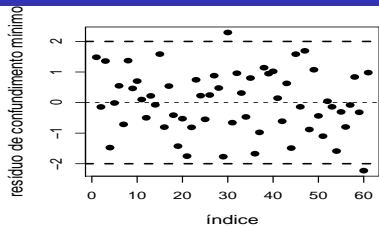
- Em geral, os RCM e os RMP apresentam leve assimetria negativa, os RCP assimetria positiva.
- Indicações leve de heterocedasticidade dos resíduos.
- Os efeitos aleatórios apresentam leve assimetria positiva e indicam diferença entre os perfis individuais (variedade x manjar) ao longo dos blocos.

Parâmetro	Estim.	EP	IC(95%)	Estat. z	p-valor
μ	111,36	9,43	[92,88 ; 129,85]	11,81	0,00
α_2	6,67	9,72	[-12,37 ; 25,71]	0,69	0,49
α_3	-8,50	9,72	[-27,54 ; 10,54]	-0,87	0,38
β_2	18,50	7,68	[3,44 ; 33,56]	2,41	0,02
β_3	34,67	7,68	[19,61 ; 49,72]	4,51	0,00
β_4	44,83	7,68	[29,78 ; 59,89]	5,84	0,00
δ_2	-28,08	10,01	[-47,70 ; -8,46]	-2,81	0,01
δ_3	-39,42	10,01	[-59,04 ; -19,80]	-3,94	0,00
δ_4	-37,17	10,01	[-56,79 ; -17,55]	-3,71	0,00
δ_5	-44,42	10,01	[-64,04 ; -24,80]	-4,44	0,00
δ_6	-39,08	10,01	[-58,70 ; -19,46]	-3,90	0,00
$(\alpha\beta)_{22}$	3,33	10,87	[-17,96 ; 24,63]	0,31	0,76
$(\alpha\beta)_{32}$	-0,33	10,87	[-21,63 ; 20,96]	-0,03	0,98
$(\alpha\beta)_{23}$	-4,17	10,87	[-25,46 ; 17,13]	-0,38	0,70
$(\alpha\beta)_{33}$	4,67	10,87	[-16,63 ; 25,96]	0,43	0,67
$(\alpha\beta)_{24}$	-4,67	10,87	[-25,96 ; 16,63]	-0,43	0,67
$(\alpha\beta)_{34}$	2,17	10,87	[-19,13 ; 23,46]	0,20	0,84

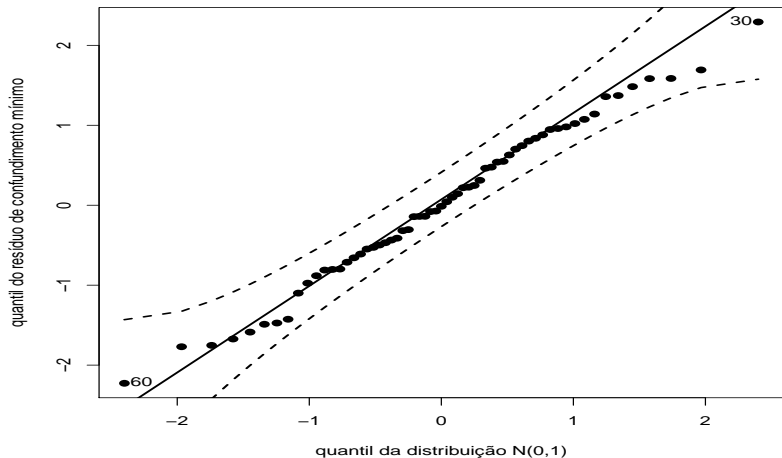
ψ	σ^2	CCI
106,06	177,08	0,37

- O teste $C\beta = M$ par a ausência de interação resultou em: 1,82 (0,9357).
- Assim, optou-se por ajustar um modelo sem interação.
- Modelo 2 (modelo sem interação). Ou seja, o modelo 1 com $((\alpha\beta)_{ij} = 0, \forall i, j)$.

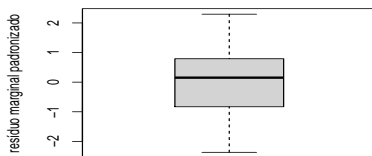
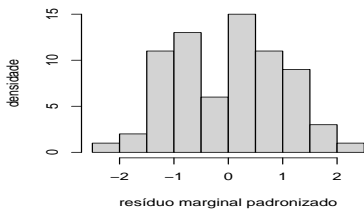
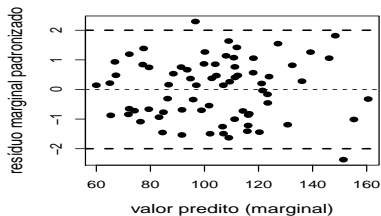
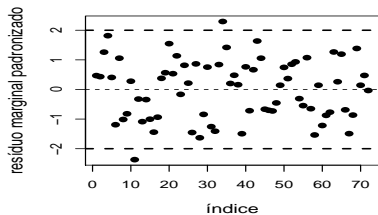
Gráficos de diagnóstico (RCM): Modelo 2



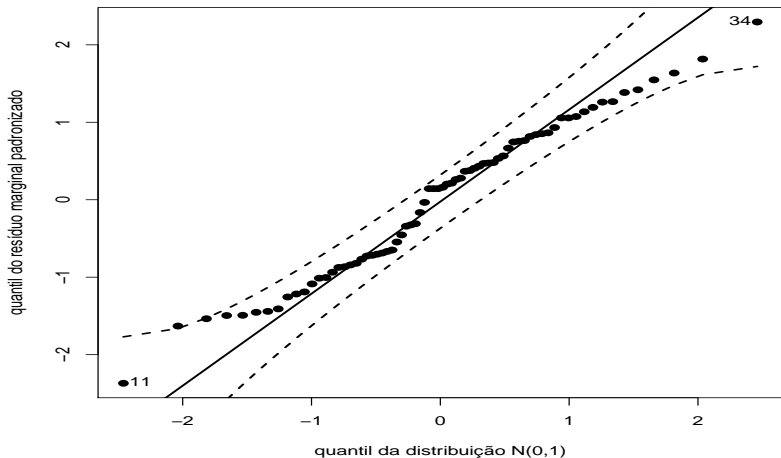
QQplot (RCM): Modelo 2



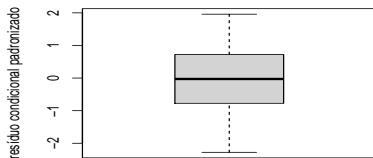
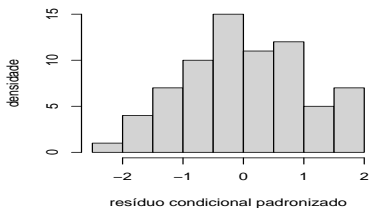
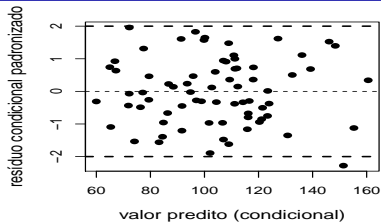
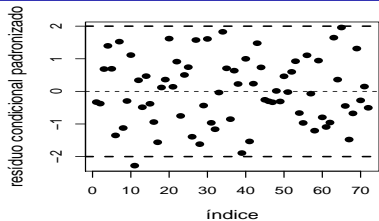
Gráficos de diagnóstico (RMP): Modelo 2



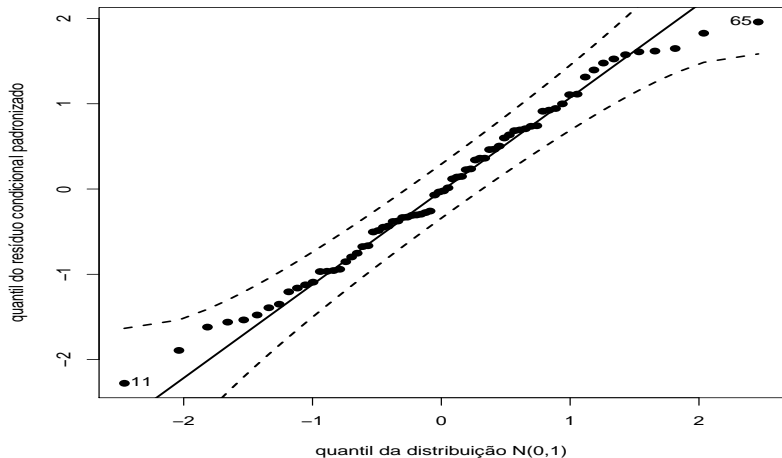
QQplot (RMP): Modelo 2



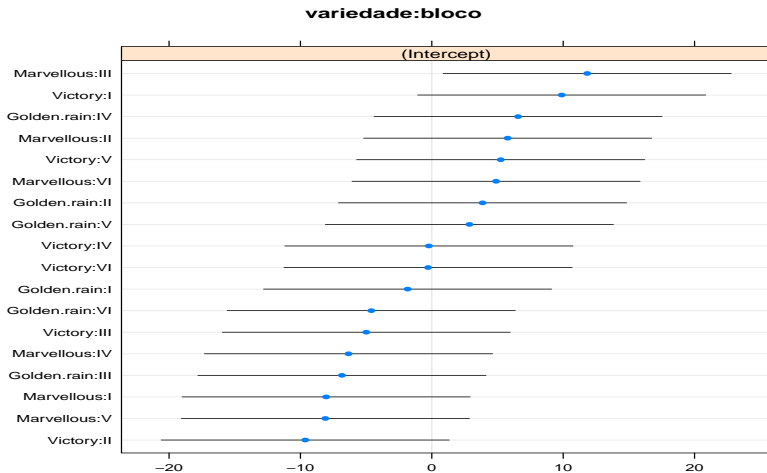
Gráficos de diagnóstico (RCP): Modelo 2



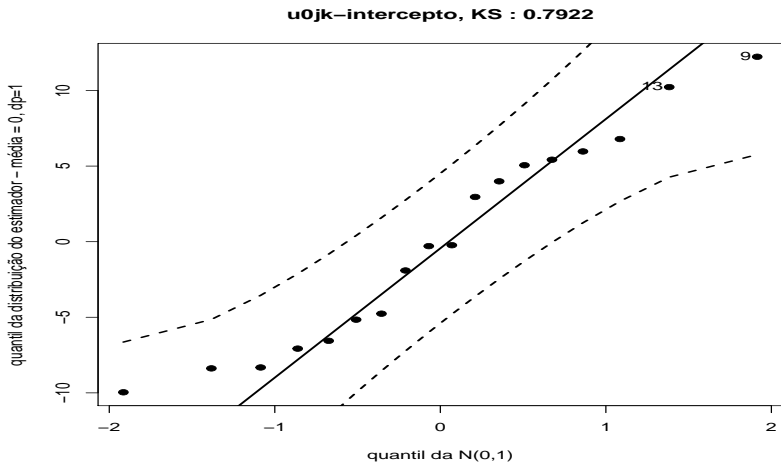
QQplot (RCP): Modelo 2



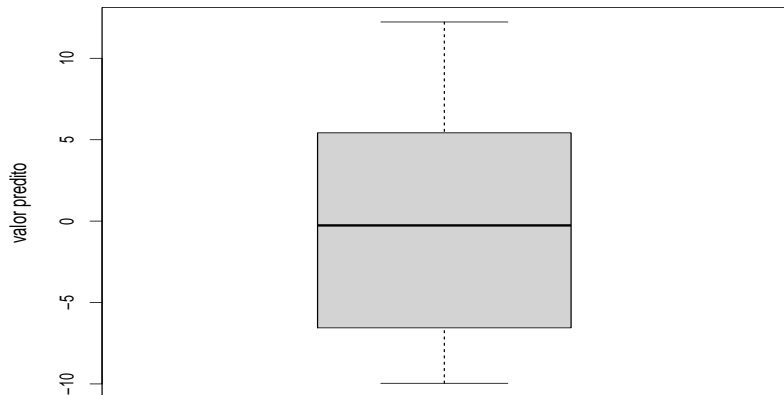
Estimativas e IC dos EA's: Modelo 2



QQplot dos EA's: Modelo 2



Boxplot dos EA's: Modelo 2



Comentários

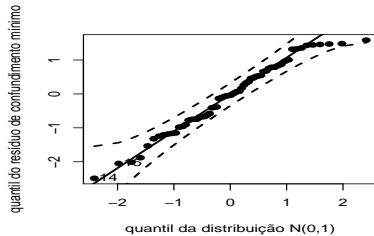
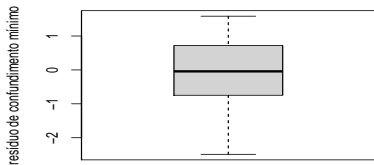
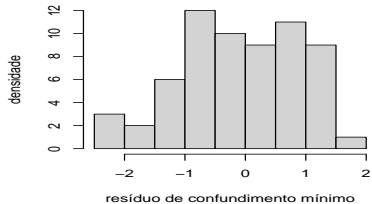
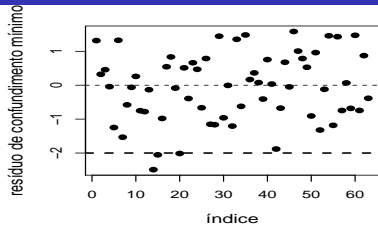
- Em geral, os RCM, RMP e os RCP apresentam assimetria negativa, os RCP assimetria positiva.
- Indicações leve de heterocedasticidade dos resíduos.
- Os efeitos aleatórios apresentam leve assimetria positiva e indicam diferença entre os perfis individuais (variedade x manjar) ao longo dos blocos.

Parâmetro	Estim.	EP	IC(95%)	Estat. z	p-valor
μ	111.28	8.58	[94.46 ; 128.09]	12.97	0.00
α_2	5.29	7.08	[-8.58 ; 19.17]	0.75	0.45
α_3	-6.88	7.08	[-20.75 ; 7.00]	-0.97	0.33
β_2	19.50	4.25	[11.17 ; 27.83]	4.59	0.00
β_3	34.83	4.25	[26.50 ; 43.16]	8.20	0.00
β_4	44.00	4.25	[35.67 ; 52.33]	10.35	0.00
δ_2	-28.08	10.01	[-47.70 ; -8.46]	-2.81	0.01
δ_3	-39.42	10.01	[-59.04 ; -19.80]	-3.94	0.00
δ_4	-37.17	10.01	[-56.79 ; -17.55]	-3.71	0.00
δ_5	-44.42	10.01	[-64.04 ; -24.80]	-4.44	0.00
δ_6	-39.08	10.01	[-58.70 ; -19.46]	-3.90	0.00

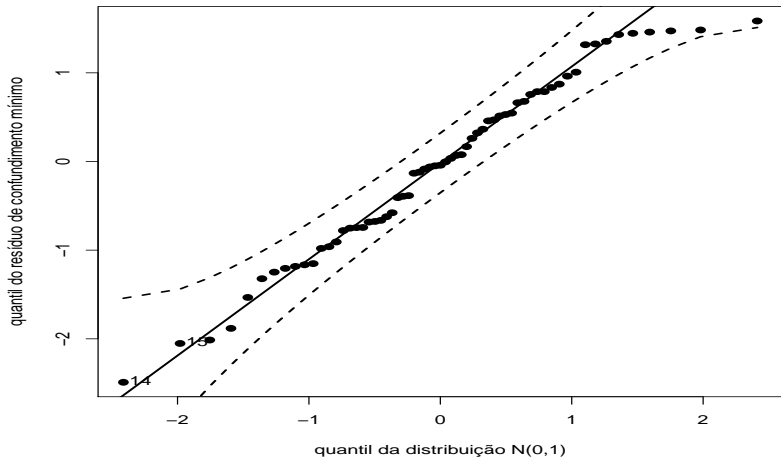
ψ	σ^2	CCI
109,69	162,56	0,40

Pelos resultados anteriores, vamos ajustar um modelo sem efeito principal de variedade. Ou seja, o modelo 2, com $\alpha_2 = \alpha_3 = 0$

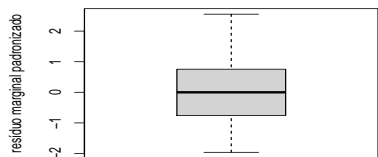
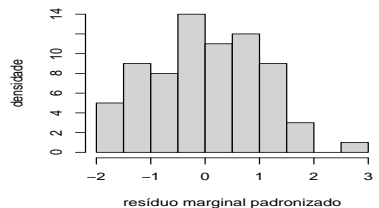
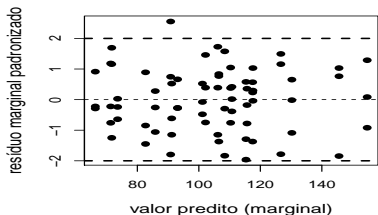
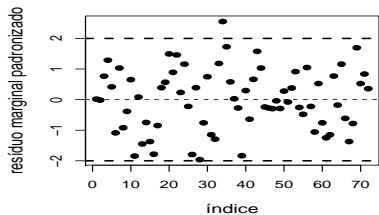
Gráficos de diagnóstico (RCM): Modelo 3



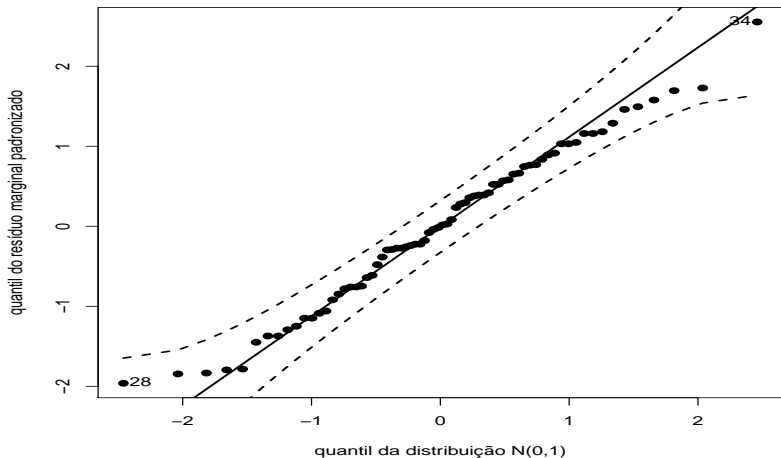
QQplot (RCM): Modelo 3



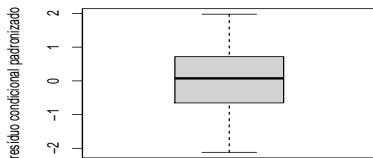
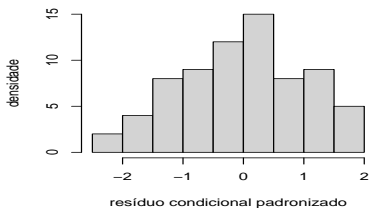
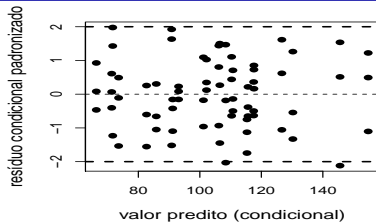
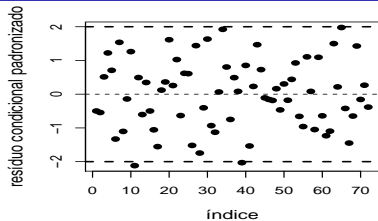
Gráficos de diagnóstico (RMP): Modelo 3



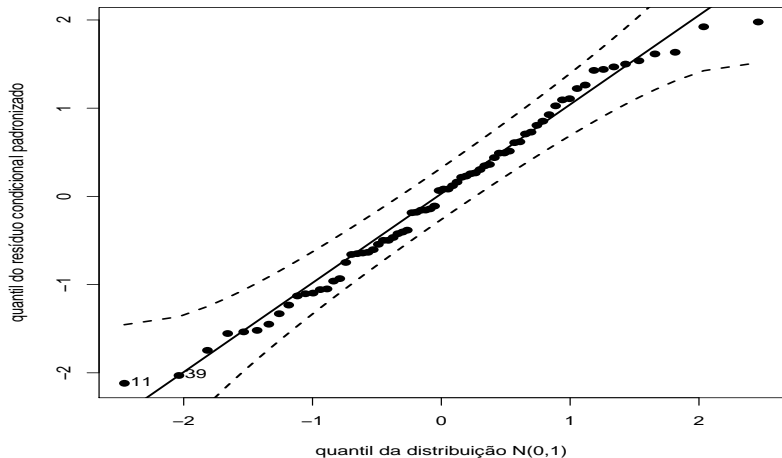
QQplot (RMP): Modelo 3



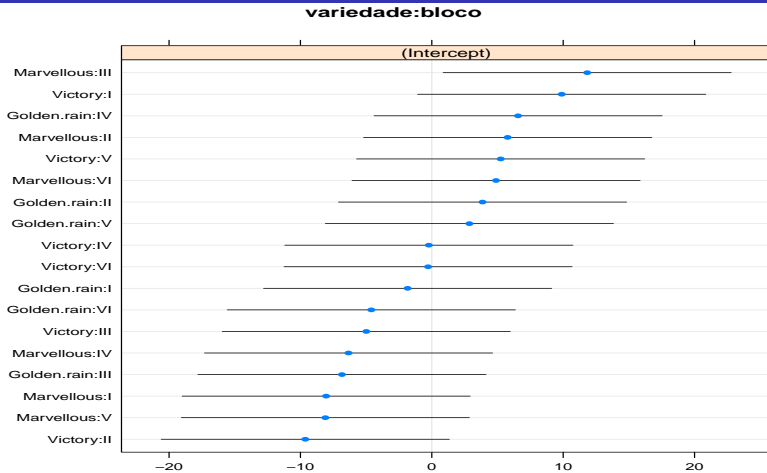
Gráficos de diagnóstico (RCP): Modelo 3



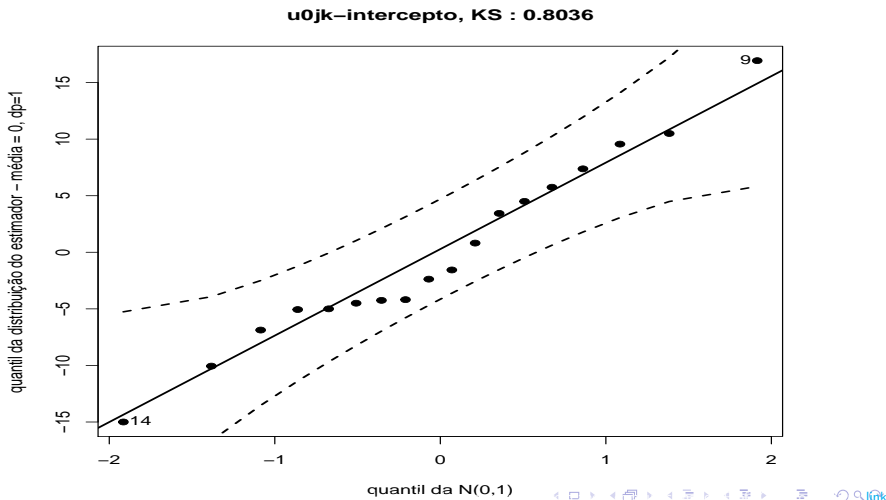
QQplot (RCP): Modelo 3



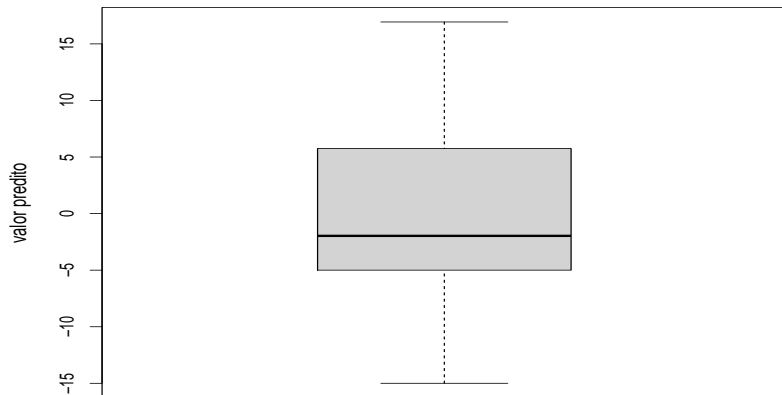
Estimativas e IC dos EA's: Modelo 3



QQplot dos EA's: Modelo 3



Boxplot dos EA's: Modelo 3



Comentários

- Em geral, os RCM, RMP e os RCP apresentam assimetria negativa, os RCP assimetria positiva.
- Indicações leve de heterocedasticidade dos resíduos.
- Os efeitos aleatórios apresentam leve assimetria positiva e indicam diferença entre os perfis individuais (variedade x manjar) ao longo dos blocos.

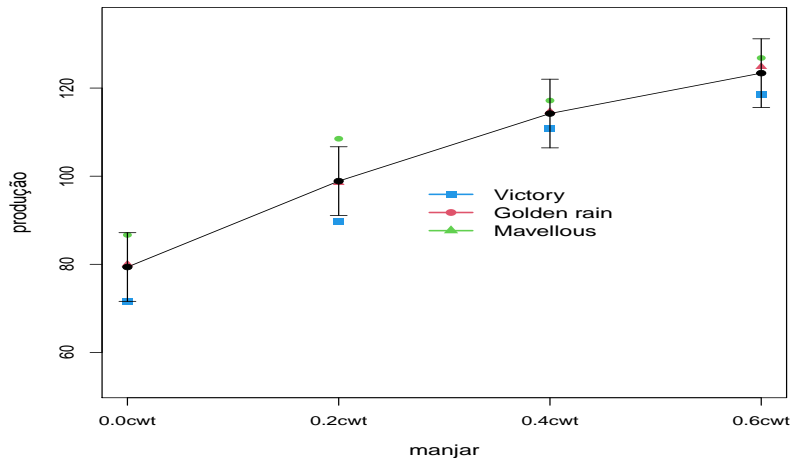
Parâmetro	Estim.	EP	IC(95%)	Estat. z	p-valor
μ	110.75	7.81	[95.45 ; 126.05]	14.19	0.00
β_2	19.50	4.25	[11.17 ; 27.83]	4.59	0.00
β_3	34.83	4.25	[26.50 ; 43.16]	8.20	0.00
β_4	44.00	4.25	[35.67 ; 52.33]	10.35	0.00
δ_2	-28.08	10.41	[-48.48 ; -7.68]	-2.70	0.01
δ_3	-39.42	10.41	[-59.82; -19.02]	-3.79	0.00
δ_4	-44.42	10.41	[-64.82 ; -24.02]	-4.27	0.00
δ_5	-37.17	10.41	[-57.57 ; -16.77]	-3.57	0.00
δ_6	-39.08	10.41	[-59.48; -18.68]	-3.76	0.00

ψ	σ^2	CCI
121,85	162,56	0,43

MM (modelo reduzido), em relação aos blocos

Variedade	Manjar	MM
GR	0,0 wct	$\mu + \sum_{k=1}^6 \delta_k/6$
	0,2 wct	$\mu + \beta_2 + \sum_{k=1}^6 \delta_k/6$
	0,4 wct	$\mu + \beta_3 + \sum_{k=1}^6 \delta_k/6$
	0,6 wct	$\mu + \beta_4 + \sum_{k=1}^6 \delta_k/6$
M	0,0 wct	$\mu + \sum_{k=1}^6 \delta_k/6$
	0,2 wct	$\mu + \beta_2 + \sum_{k=1}^6 \delta_k/6$
	0,4 wct	$\mu + \beta_3 + \sum_{k=1}^6 \delta_k/6$
	0,6 wct	$\mu + \beta_4 + \sum_{k=1}^6 \delta_k/6$
V	0,0 wct	$\mu + \sum_{k=1}^6 \delta_k/6$
	0,2 wct	$\mu + \beta_2 + \sum_{k=1}^6 \delta_k/6$
	0,4 wct	$\mu + \beta_3 + \sum_{k=1}^6 \delta_k/6$
	0,6 wct	$\mu + \beta_4 + \sum_{k=1}^6 \delta_k/6$

Perfis médios observados e ajustados



Comentários gerais

- Há efeito, somente, de manjar (positivo) e negativo de blocos (em relação bloco 1). Não há interação nem efeito de variedade.
- A produção ótima é obtida com manjar 0,6cwt da ordem de 123,39 [115,60;131,18], (a tabela abaixo apresenta estimativas pontuais intervalares para as médias de interesse).

manjar	Estimativa	IC(95%)
0,0cwt	79,39	[71,60 ; 87,18]
0,2cwt	98,89	[91,10 ; 106,68]
0,4cwt	114,22	[106,43 ; 122,01]
0,6cwt	123,39	[115,60 ; 131,18]