

# Modelos para dados de contagem: parte 3

Prof. Caio Azevedo

## Exemplo 12: número de faltas de estudantes australianos

- **Venables e Ripley** (1999, Caps. 6 e 7) apresentam os resultados de um estudo sociológico desenvolvido na Austrália com 146 estudantes da 8ª série até o 3º ano do ensino médio, com o objetivo de comparar a ausência na escola segundo os seguintes fatores:
- Ano que o estudante está cursando (1: 8ª série, 2: 1º ano do ensino médio, 3: 2º ano do ensino médio, 4: 3º ano do ensino médio), etnia (aborígine, não aborígine), desempenho escolar (insuficiente, suficiente) e sexo (masculino, feminino).
- Não há, na base de dados, alunos do 3º com desempenho insuficiente.

## Medidas Resumo: etnia

medida resumo	etnia	
	aborígene	não-aborígene
Média	21,23	12,18
DP	17,72	13,56
Var.	313,95	183,89
CA	1,08	2,11
Curt	3,63	8,59
Min.	0	0
Max.	81	69
CV(%)	83,45	111,32

## Medidas Resumo: sexo

medida resumo	sexo	
	feminino	masculino
Média	15,23	17,96
DP	15,93	16,63
Var.	253,80	276,70
CA	1,77	1,20
Curt	6,31	4,03
Min.	0	0
Max.	81	69
CV(%)	104,64	92,65

## Medidas Resumo: ano escolar

medida resumo	ano escolar			
	8ª série	1º ano	2º ano	3º ano
Média	14,85	11,15	21,05	19,62
DP	14,80	11,64	20,14	15,97
Var.	218,90	135,51	405,49	255,18
CA	1,84	2,22	1,05	0,84
Curt	7,14	8,48	3,41	3,78
Min.	0	0	0	0
Max.	67	54	81	69
CV(%)	99,62	104,38	95,66	81,48

## Medidas Resumo: desempenho

medida resumo	desempenho	
	suficiente	insuficiente
Média	15,82	17,30
DP	13,91	18,99
Var.	193,50	360,63
CA	1,17	1,53
Curt	4,34	4,59
Min.	0	0
Max.	69	81
CV(%)	87,93	109,76

## Medidas resumo: todos os grupos

etnia	sexo	ano-escolar	desempenho	Média	Var.	CA	Curt.
A	F	8S	suficiente	21,25	313,58	0,56	1,80
A	F	8S	insuficiente	3,00	-	-	-
A	F	1A	suficiente	11,40	42,80	0,09	1,31
A	F	1A	insuficiente	22,60	349,16	0,75	2,15
A	F	2A	suficiente	2,00	-	-	-
A	F	2A	insuficiente	36,38	702,55	0,44	1,84
A	F	3A	suficiente	14,56	220,53	0,77	2,11

(\*)“-” não há alunos, na base de dados, do grupo em questão.

## Medidas resumo: todos os grupos

etnia	sexo	ano-escolar	desempenho	Média	Var.	CA	Curt.
A	M	8S	suficiente	13,00	64,50	0,03	1,29
A	M	8S	insuficiente	9,00	39,00	-0,53	1,50
A	M	1A	suficiente	10,50	24,50	0,00	1,00
A	M	1A	insuficiente	9,00	27,00	0,71	1,50
A	M	2A	suficiente	27,43	215,95	0,32	1,17
A	M	2A	insuficiente	37,00	547,33	-0,54	1,71
A	M	3A	suficiente	27,14	107,48	-0,79	2,65



## Medidas resumo: todos os grupos

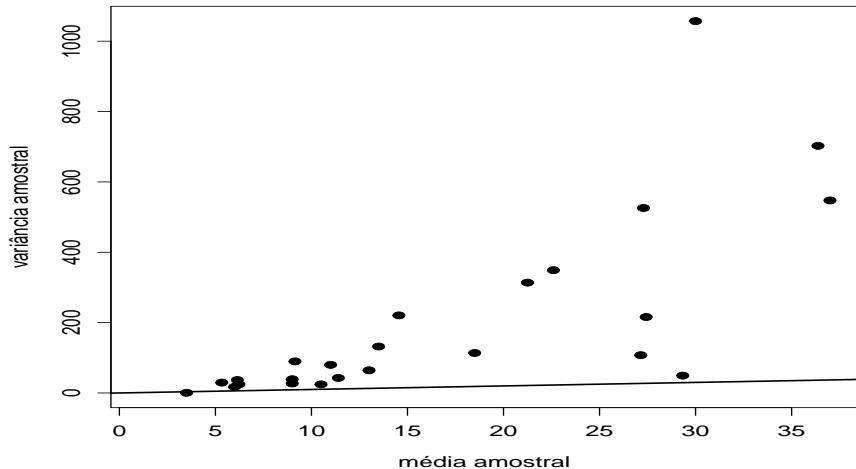
etnia	sexo	ano-escolar	desempenho	Média	Var.	CA	Curt.
N	F	8S	suficiente	18,50	113,67	0,64	1,81
N	F	8S	insuficiente	25,00	-	-	-
N	F	1A	suficiente	11,00	80,00	1,35	3,24
N	F	1A	insuficiente	6,00	17,40	0,74	3,31
N	F	2A	suficiente	1,00	-	-	-
N	F	2A	insuficiente	6,22	24,69	0,32	1,67
N	F	3A	suficiente	13,50	132,06	0,72	2,62

(\*)“-” não há alunos, na base de dados, do grupo em questão.

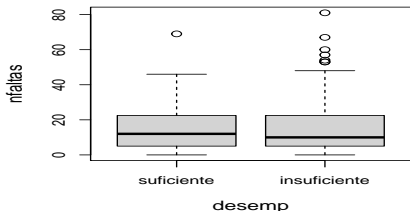
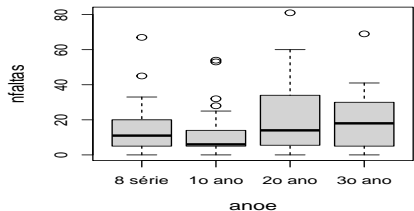
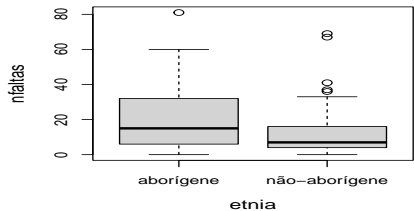
## Medidas resumo: todos os grupos

etnia	sexo	ano-escolar	desempenho	Média	Var.	CA	Curt.
N	M	8S	suficiente	5,33	29,47	0,20	1,31
N	M	8S	insuficiente	30,00	1057,00	0,62	1,50
N	M	1A	suficiente	3,50	0,50	0,00	1,00
N	M	1A	insuficiente	6,14	36,81	0,75	2,48
N	M	2A	suficiente	9,14	89,81	0,98	2,79
N	M	2A	insuficiente	29,33	49,33	-0,17	1,50
N	M	3A	suficiente	27,29	525,90	0,71	2,63

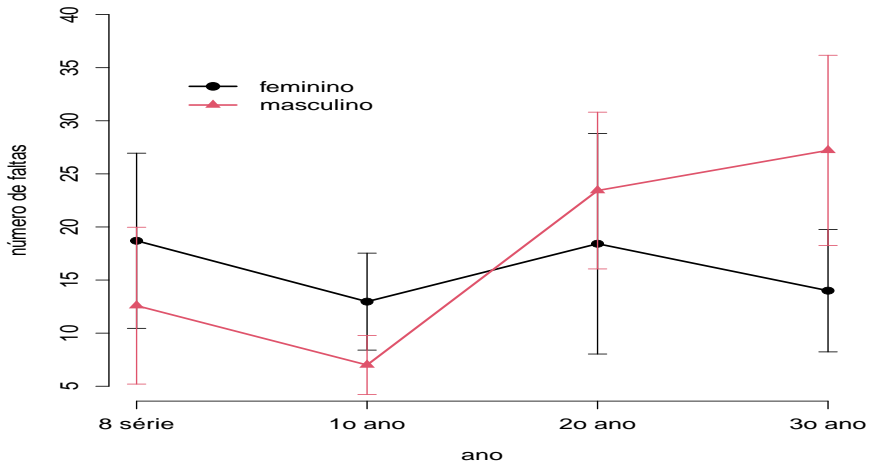
# Médias e variâncias amostrais de cada grupo



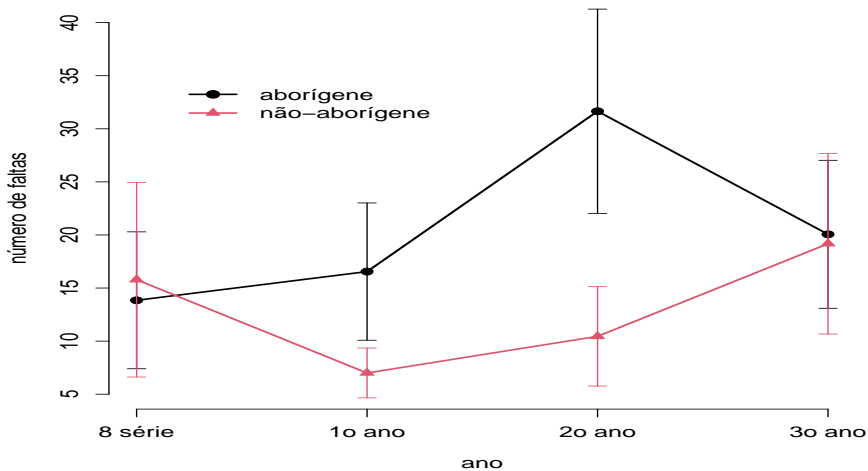
# Boxplots por cada um dos fatores



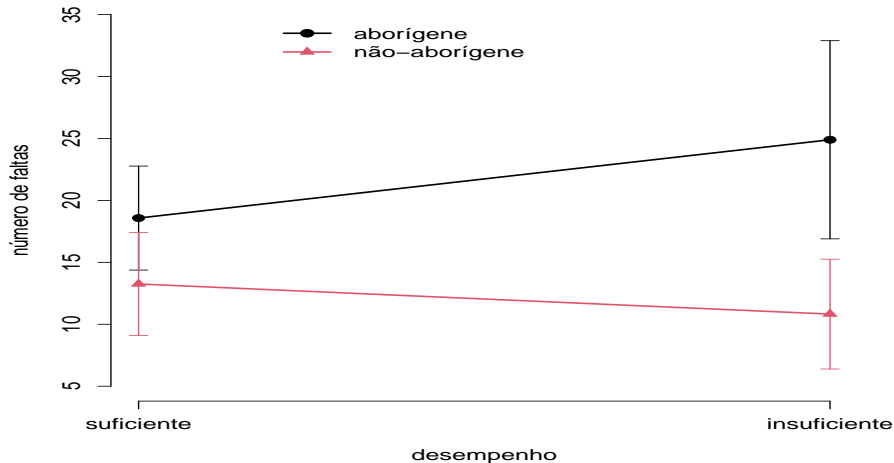
# Gráficos de perfis médios



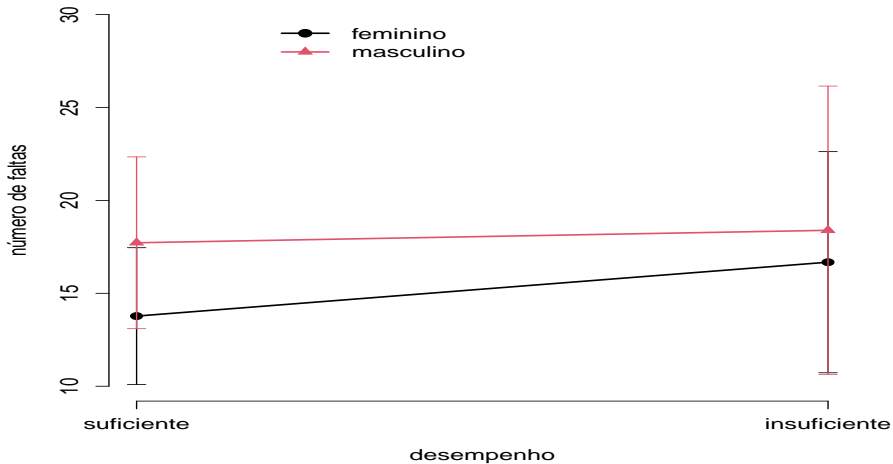
# Gráficos de perfis médios



# Gráficos de perfis médios

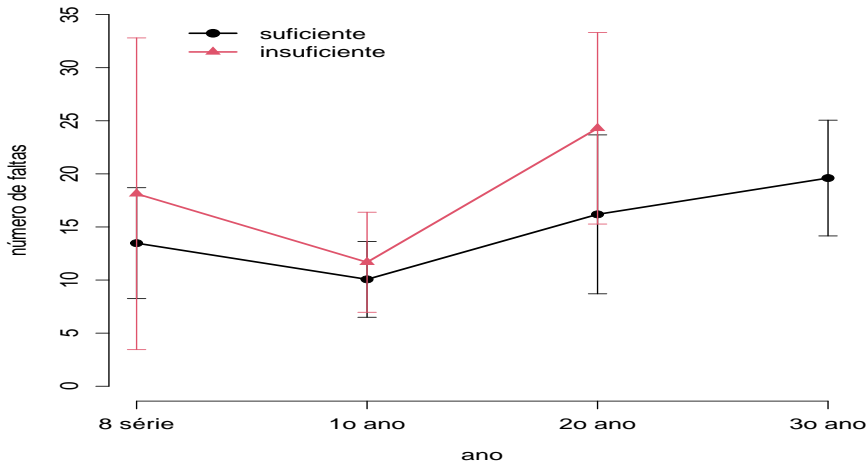


# Gráficos de perfis médios



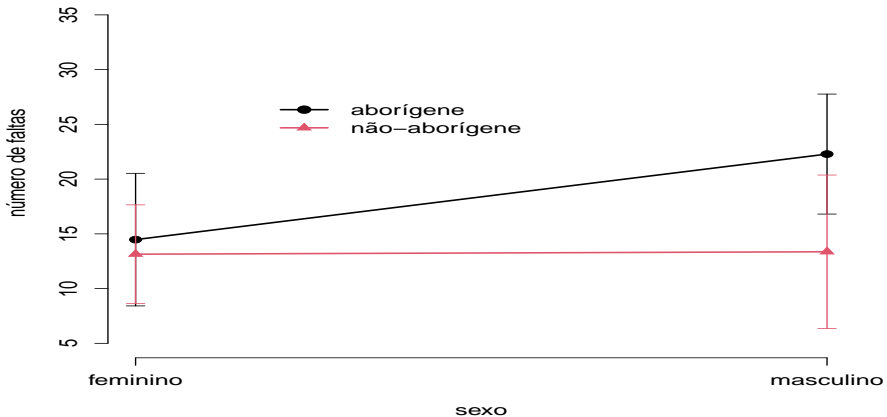


# Gráficos de perfis médios



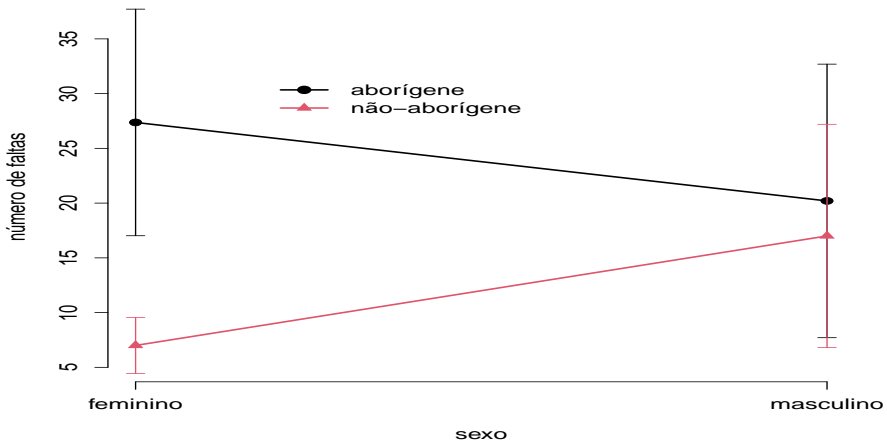
# Gráficos de perfis médios

## Desempenho suficiente



# Gráficos de perfis médios

## Desempenho insuficiente



# Comentários

- Outro gráficos de perfis médios não foram feitos devido à pouca (ou nenhuma) quantidade de observações presentes em alguns grupos (cruzamentos entre níveis de dois ou mais fatores).
- Há evidências da existência de pelo menos algumas interações de primeira e segunda ordens e, conseqüentemente, do efeito de fatores principais.
- Há indícios da presença de superdispersão.

# Modelo

$$Y_{ijklr} \stackrel{ind.}{\sim} \text{Poisson}(\mu_{ijkl}),$$

$i = 1, 2$  (etnia: 1 - aborígene, 2 - não aborígene),

$j = 1, 2$  (sexo : 1- feminino, 2 - masculino),

$k = 1, 2, 3, 4,$  (ano escolar: 1 - 8ª série, 2 - 1º ano, 3 - 2º ano,  
4 - 3ºano)

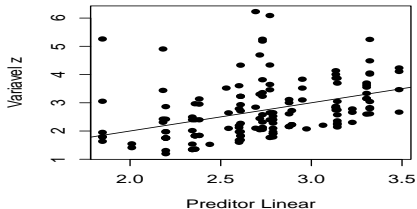
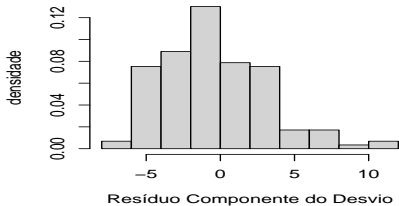
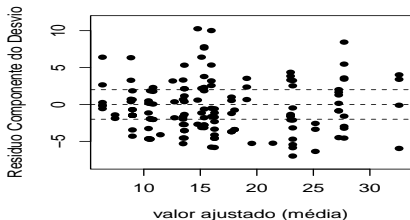
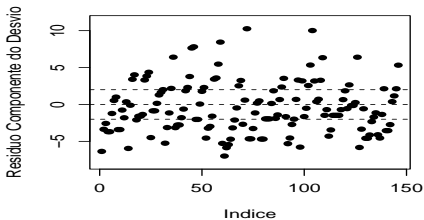
$l = 1, 2$  (desempenho: 1- suficiente, 2 - insuficiente,

$r = 1, 2, \dots, n_{ijkl}$ (unidades amostrais)

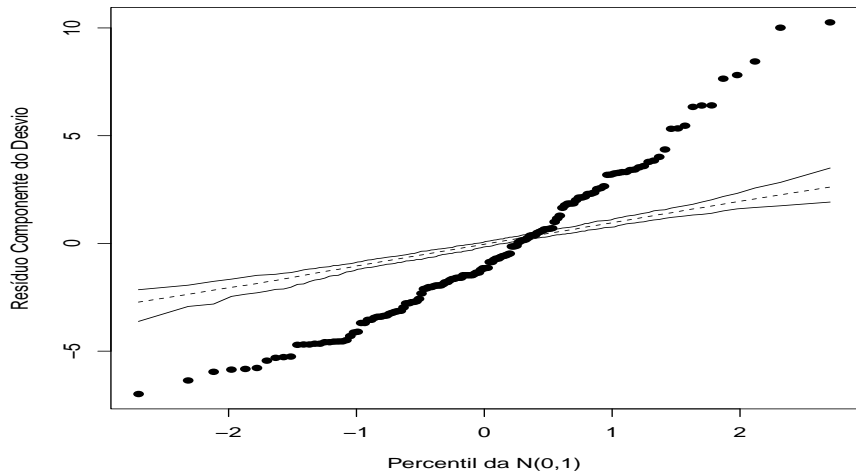
$$\ln \mu_{ijkl} = \alpha + \beta_i + \gamma_j + \delta_k + \lambda_l$$

$$\beta_1 = \gamma_1 = \delta_1 = \lambda_1 = 0$$

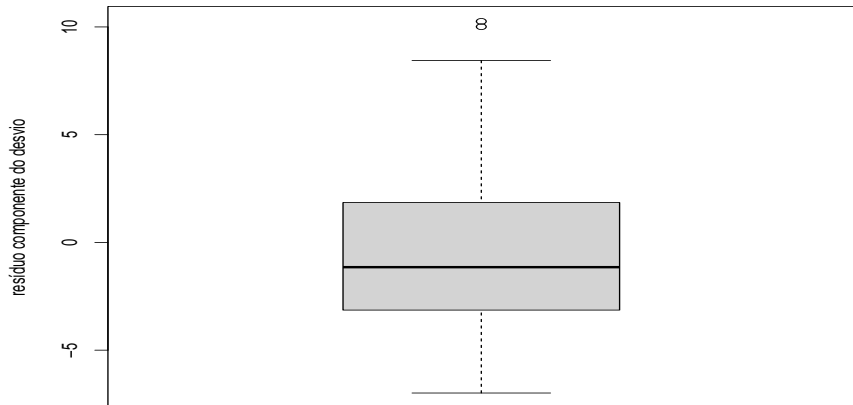
# Gráficos de diagnóstico do RCD



# Gráfico de envelope do RCD

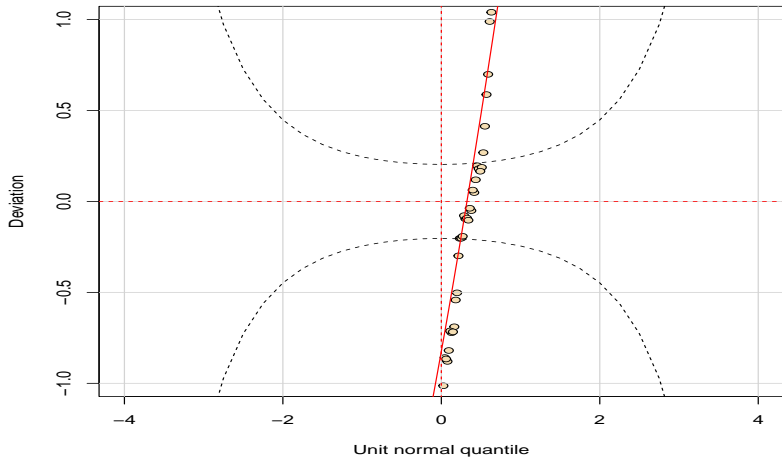


# Boxplot do RCD

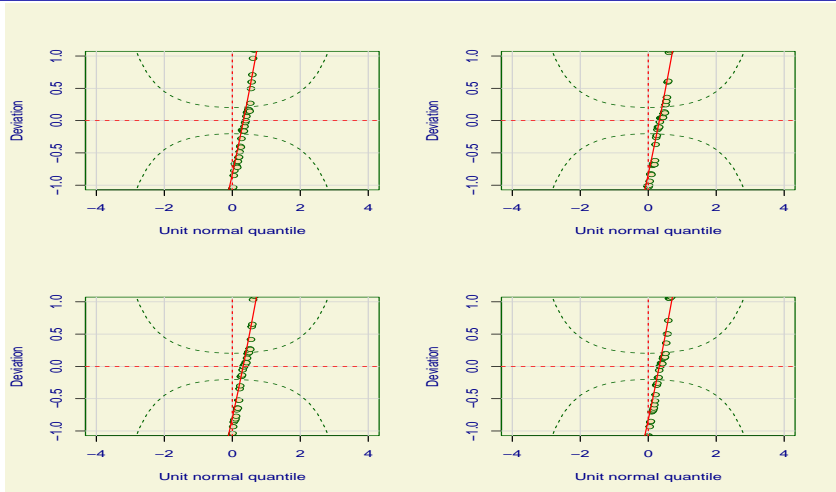




## Worm plot para os resíduos: Exemplo 4



# Worm plots para os resíduos: Exemplo 4



## Comentários

- $D(\mathbf{y}, \tilde{\boldsymbol{\mu}}) = 1696,71$  ( $p = < 0,0001$ ) (considerando-se a aproximação pela distribuição  $\chi^2_{(139)}$  adequada), o que indica que o modelo não se ajustou bem aos dados.
- A análise residual indica a presença de superdispersão e heterocedasticidades, apesar dos resíduos apresentarem simetria.
- Se o problema, em relação ao mal ajuste, estiver sendo causado por superdispersão (o que parece ser o caso), os erros-padrão estão sendo subestimados.
- O modelo não está bem ajustado e, assim, faz-se mister utilizar alguma alternativa apropriada.

# Estimativas

Par.	Est.	EP	IC(95%)	Estat. $Z_t$	p-valor
$\alpha$	2,72	0,06	[2,59 ; 2,84]	41,98	< 0,0001
$\beta_2$ (não-aborígene)	-0,53	0,04	[-0,62 ; -0,45]	-12,74	< 0,0001
$\gamma_2$ (masculino)	0,16	0,04	[0,08 ; 0,24]	3,80	0,0001
$\delta_2$ (1º ano)	-0,33	0,07	[-0,47 ; -0,20]	-4,76	< 0,0001
$\delta_3$ (2º ano)	0,26	0,06	[0,14 ; 0,38]	4,13	< 0,0001
$\delta_4$ (3º ano)	0,43	0,07	[0,30 ; 0,56]	6,32	< 0,0001
$\lambda_2$ (insuficiente)	0,35	0,05	[0,25 ; 0,45]	6,70	< 0,0001

É possível que algumas covariáveis não sejam, realmente, significativas.