Universidade Estadual de Campinas Instituto de Matemática, Estatística e Computação Cientifica Departamento de Estatística

Relatório - Parte I Exercício 2

Guilherme Pazian RA:160323 Henrique Capatto RA:146406 Hugo Calegari RA:155738 Leonardo Uchoa Pedreira RA:156231

Professor: Caio Lucidius Naberezny Azevedo

Campinas-SP, 05 de Junho de 2017

Exercício 2

1.Introdução

Os dados a serem analisados são provenientes de um experimento desenvolvido para avaliar a germinação de um determinado tipo de semente de acordo com três condições experimentais: nível da temperatura: 21°C, 42°C e 62°C; nível de umidade: baixo, médio e alto; e temperatura de germinação: 11°C e 21°C. Por exemplo, a uma temperatura de 11°C, com nível de umidade baixo e temperatura de umidade de 21°C, 100 sementes foram plantadas e observou-se que 98 delas germinaram depois de 5 dias.

A tabela que contém os dados apresenta o número de sementes que germinaram após 5 dias para cada 100 sementes submetidas a cada condição experimental. Cada casela que compõem a tabela assume contagens Y_{ijk} tais que representam variáveis aleatórias binomiais independentes com tamanho fixo de n = 100 e probabilidade de sucesso (neste caso, estimada como a proporção de sementes que germinaram depois de 5 dias).

O objetivo é estimar a probabilidade de sementes germinadas depois de 5 dias, baseado em efeitos principais (nível de temperatura, nível de umidade e temperatura de germinação), verificar quais interações são significativas (combinações dos efeitos principais) e fornecer a combinação entre os níveis de fatores que gera a melhor taxa de germinação.

2. Análise Descritiva

Nota-se que para cada casela pode-se obter a proporção de sementes que germinaram depois de 5 dias para cada 100 sementes submetidas a cada condição experimental. Abaixo tem-se as tabelas com os dados e com as proporções estimadas, respectivamente.

	21°C	42°C	62°C
11°C e BAIXO	98	96	62
11℃ e MÉDIO	94	79	3
11℃ e ALTO	92	41	1
21°C e BAIXO	94	93	65
21°C e MÉDIO	94	71	2
21°C e ALTO	91	30	1

	21°C	42°C	62°C
11℃ e BAIXO	0.98	0.96	0.62
11℃ e MÉDIO	0.94	0.79	0.03
11℃ e ALTO	0.92	0.41	0.01
21°C e BAIXO	0.94	0.93	0.65
21℃ e MÉDIO	0.94	0.71	0.02
21°C e ALTO	0.91	0.30	0.01

Pela tabela de proporções estimadas, consegue-se notar que para uma dada temperatura de germinação e temperatura de umidade, à medida em que o nível de umidade aumenta (por exemplo, de baixo para médio ou de médio para alto) a proporção de sementes germinadas após cinco dias diminui. Fixado a temperatura de germinação e o nível de umidade, quando se aumenta a temperatura de umidade (por exemplo, de 21°C para 42°C) a proporção de sementes germinadas também diminui.

Nota-se ainda que fixado um nível de umidade (por exemplo, baixo ou médio), para diferentes temperaturas de germinação, a proporção de sementes germinadas diminui com o aumento da temperatura de umidade. Finalmente, fixado a temperatura de umidade, a proporção de sementes germinadas para as diferentes temperaturas de germinação e diferentes níveis de umidade também é reduzida.

Assim, baseado nessas informações, à princípio pode-se conjecturar de que ao se utilizar uma temperatura de germinação mais baixa, neste caso 11°C, com o menor nível de umidade, no contexto umidade baixa, e com menor temperatura de umidade, neste caso 21°C, fornecerá uma combinação razoável para a maior proporção de sementes germinadas depois de 5 dias.

Além disso, algumas medidas resumo podem ser avaliadas como quantidade de sementes germinadas em diferentes temperaturas de germinação para os diferentes níveis de umidade e temperatura de umidade. Ou seja, é possível obter informações do comportamento da quantidade média de sementes germinadas em diferentes temperaturas de germinação, para cada nível de umidade e observar se as diferenças dos valores médios são muito discrepante.

Temperatura germinação	Nível umidade	Média	Variância	DP	MAX	MIN
11°C	alto	44.67	2080.33	45.61	92	1
11°C	baixo	85.33	409.33	20.23	98	62
11°C	médio	58.67	2380.33	48.79	94	3
21°C	alto	40.67	2110.33	45.94	91	1
21°C	baixo	84.00	271.00	16.46	94	65
21°C	médio	55.67	2292.33	47.88	94	2

Pela tabela anterior observa-se que os valores médios de sementes germinadas após 5 dias, para uma dada temperatura de germinação e nível de umidade, desconsiderando-se a temperatura de umidade, diferem entre si. Por exemplo, a maior diferença nos valores médios está associada a seguinte combinação: temperatura de germinação de 11°C e 21°C, nível de umidade alta, com diferença média de sementes germinadas de 4. Essa diferença deve ser avaliada em conjunto com o pesquisador, pois deve-se levar em conta se essa diferença é significativa na prática, ou seja, se de fato em média essas quatro sementes teriam uma produtividade em média maior.

Foi possível elaborar, também, gráficos de perfis que avalie o comportamento da umidade em uma dada temperatura de germinação, para os diferentes níveis de temperatura da umidade, isto é, as possíveis interações existentes.

Informação relevante é obtida com o gráfico de perfil das proporções de sementes germinadas depois de 5 dias. Nota-se que para cada nível de temperatura de (11°C e 21°C) o comportamento é muito semelhante das proporções estimadas, isto é, para o aumento da temperatura de umidade há diminuição das proporção estimadas de sementes germinadas. Com isso, pode-se cogitar em uma possível interação entre as variáveis nível de umidade e temperatura de umidade.

Mais do que isso, para os diferentes valores de temperatura de germinação e diferentes níveis de temperatura, num mesmo nível de umidade, o comportamento de decaimento das proporções estimadas de sementes germinadas são muito semelhantes. Isto pode ser observado pelos gráficos de perfis representados abaixo.

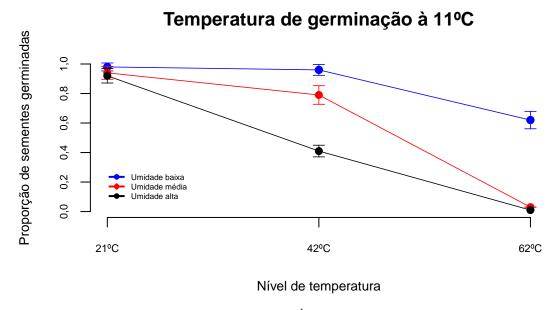


Figura 1: Gráficos de perfil para temperatura de germinação à 11°C. É perceptível que, para este perfil, com o aumento do nível de temperatura de umidade reduz-se a proporção estimada de sementes germinadas.

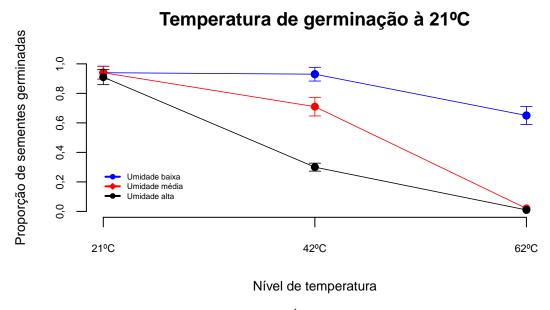


Figura 2: Gráfico de perfil para temperatura de germinação à 21°C. É perceptível que, para este perfil, com o aumento do nível de temperatura de umidade reduz-se a proporção estimada de sementes germinadas.

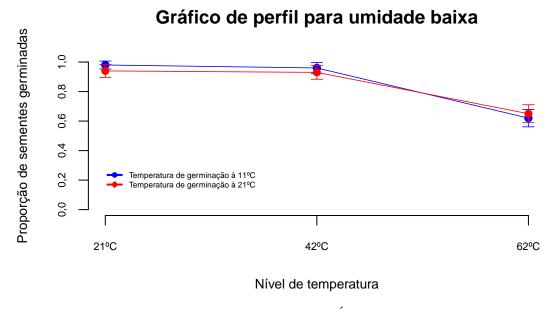


Figura 3: Gráficos de perfil para temperatura de germinação à 11°C e 21°C. É perceptível que, para este perfil, com o aumento do nível de temperatura de umidade reduz-se a proporção de sementes grminadas. Note que o comportamento entre as temperaturas de germinação é semelhante. Com exceção do nível de temperatura de 62°C, à uma temperatura de germinação de 21°C, observa-se que a estimativa pontual da proporção de sementes germinadas é inferior em relação aos demais níveis de temperatura.

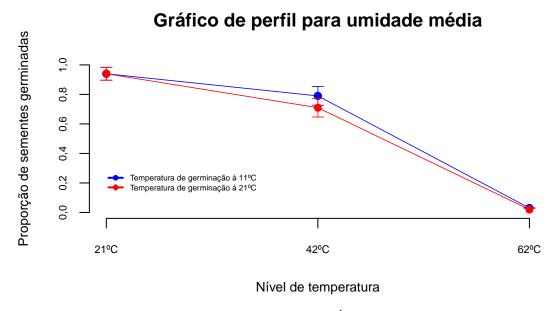


Figura 4: Gráfico de perfil para temperatura de germinação à 11°C e 21°C. É perceptível que, para este perfil, com o aumento do nível de temperatura de umidade reduz-se a proporção estimada de sementes grminadas. Note que o comportamento entre as temperaturas de germinação é semelhante. Observa-se que, em geral, a proporção estimada de sementes germinadas à temperatura de germinação de 21°C é inferior em relação à 11°C.

Gráfico de perfil para umidade alta

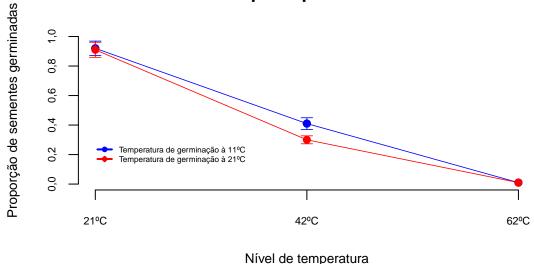


Figura 5: Gráfico de perfil para temperatura de germinação à 11°C e 21°C. É perceptível que, para este perfil, com o aumento do nível de temperatura de umidade reduz-se a proporção estimada de sementes grminadas. Note que o comportamento entre as temperaturas de germinação é semelhante. Observa-se que, em geral, a proporção estimada de sementes germinadas à temperatura de germinação de 21°C é inferior em relação à 11°C.

3. Análise Inferencial

Descrição dos Modelos

Ao considerar o problema posto e a análise descritiva realizada, foram propostos os seguintes modelos

• Modelo 1

$$Y_{ijk} \stackrel{ind}{\sim} Binomial(100, p_{ik})$$

$$ln\left(\frac{p_{ijk}}{1-p_{ijk}}\right) = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$$
 onde $i=1,2,\ j=1,2,3,\ k=1,2,3,\ \alpha_1=\beta_3=\gamma_1=(\beta\gamma)_{j1}=(\beta\gamma)_{3k}=(\alpha\beta)_{1j}=(\alpha\beta)_{23}=(\alpha\gamma)_{1k}=(\alpha\gamma)_{21}=(\alpha\beta\gamma)_{1kj}=(\alpha\beta\gamma)_{23k}=(\alpha\beta\gamma)_{2j1}=0$

observação: este modelo é saturado (ou seja, a quantidade de parâmetros é igual ao número de observações). Logo, não é possível realizar análise de resíduos e cálculos de desvios.

• Modelo 2

$$Y_{ijk} \stackrel{ind}{\sim} Binomial(100, p_{ik})$$

$$ln\left(\frac{p_{ijk}}{1 - p_{ijk}}\right) = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$$
 onde $i = 1, 2, j = 1, 2, 3, k = 1, 2, 3, \alpha_1 = \beta_3 = \gamma_1 = (\beta\gamma)_{j1} = (\beta\gamma)_{3k} = (\alpha\beta)_{1j} = (\alpha\beta)_{23} = (\alpha\gamma)_{1k} = (\alpha\gamma)_{21} = 0.$

Modelo 3

$$Y_{ijk} \stackrel{ind}{\sim} Binomial(100, p_{ik})$$

$$ln\left(\frac{p_{ijk}}{1 - p_{ijk}}\right) = \mu + \alpha_i + \beta_j + \gamma_k + (\beta\gamma)_{jk}$$

onde
$$i = 1, 2, j = 1, 2, 3, k = 1, 2, 3, \alpha_1 = \beta_3 = \gamma_1 = (\beta \gamma)_{j1} = (\beta \gamma)_{3k} = 0$$
 e

Todos os modelos considerados foram ajustados através da metodologia de máxima verossimilhança (veja Paula (2013)).

Avaliação dos Ajustes

Modelo 1

Ao observar a tabela ??, que mostra as características gerais sobre este modelo, pode-se perceber que vários parâmetros não são estatísticamente significantes. Este curioso fato é difícil explicação

Modelo 2

Para a avaliação deste modelo, considere a figura abaixo. Podemos notar que até o ponto 0 do eixo horizontal, todos os pontos estão acima da linha de base (linha tracejada), enquanto que o oposto ocorre após este ponto (ou seja, não existe independência entre os resíduos). Isto pode ser um indicativo de que alguma característica dos dados ainda resta a ser explicada (este argumento é suportado pelo fato de que existem, relativamente, bastante precisão em cada observação). Para isto, pode-se considerar um estudo com outros fatores experimentais. Entretanto, ao considerar que todos os pontos estão dentro do envelope e que seguem a linha de base, não existem indicativos de mal ajuste do modelo.

Ao considerar a tabela $\ref{eq:constraints}$, pode-se perceber alta significância de vários parâmetros, ao contrário do modelo 1. Todavia, as interações relativas à Temperatura de Germinação são todas não significativas. Isto é um indicativo de que estes efeitos podem podem não ser estatísticamente nulos. Para verificar esta hipótese, utilizou o teste H_0 : $C\beta = M$ vs H_1 : $C\beta \neq M$ para verificar se a hipótese de que $(\alpha\beta)_{21} = (\alpha\beta)_{21} = (\alpha\gamma)_{22} = 0$ (para mais detalhe, ver Azevedo(2017)). O valor-p resultante do teste é

Gráfico Quantíl-Quantíl

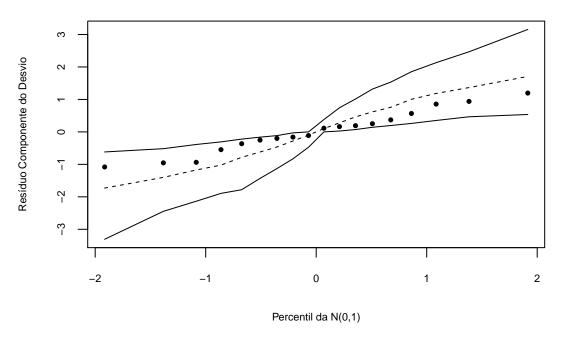


Figura 6: Este gráfico mostra uma comparação entre os quantís do RCD (ver Paula, 2013). Quanto mais pontos estiverem dentro do envelope (linha cheia) e perto da reta pontilhada, mais indicativos temos de um bom ajuste. Note, porém, todos os pontos estão acima da linha de base (linha tracejada). até o ponto 0 do eixo horizontal, enquanto que o oposto ocorre após este ponto.

0,8143, que indica a não rejeição da hipótese de que todos estes coeficientes são, simultaneamente, nulos. Isto forneceu evidências para uma redução do modelo, que levou ao modelo 3.

Modelo 3

Este modelo é derivado, como citado, do modelo 2. Ao observar a figura abaixo, pode-se perceber um comportamento muito similar ao visto no modelo 2, que sugere que alguma característica dos dados ainda resta a ser explicada (devido à não independência dos resíduos). Logo, pode ser benéfico considerar outras condições experimentais, para explicar esta característica.

Os gráficos de diagnósticos para o modelo 3 possuem características semelhantes aqueles do modelo 2. Percebe-se para os gráficos 1 e 2, no entanto, que existem valores dos resíduos que estão mais próximos limites inferior e superior. O gráfico 3 representa, assim como para o modelo 2, que a função de ligação utilizada e o preditor linear são razoáveis. No caso do preditor linear, possivelmente, as covariáveis utilizadas não nos indicam a falta de informação.

Seleção dos Modelos Ajustados

A tabela a seguir fornece algumas informações sobre os modelos ajustados.

pode-se considerar um estudo com outros fatores experimentais

Dela, pode-se depreender que de acordo com os critérios de informação, o modelo 3 é superior, pois possui os menores valores.

Gráfico Quantíl-Quantíl

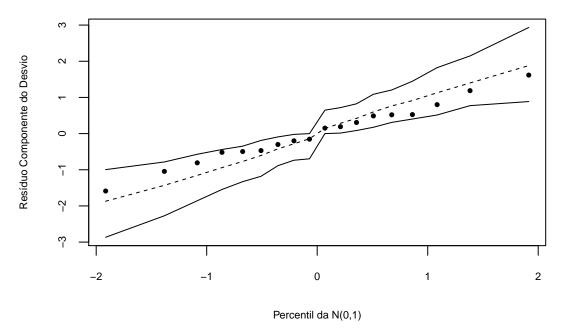


Figura 7: Este gráfico mostra uma comparação entre os quantís do RCD (ver Paula, 2013). Quanto mais pontos estiverem dentro do envelope (linha cheia) e perto da reta pontilhada, mais indicativos temos de um bom ajuste. Note, porém, todos os pontos estão acima da linha de base (linha tracejada), enquanto que o oposto ocorre após este ponto. Perceba, também, que este comportamento é similar ao que foi visto no modelo 2.

Por outro lado, de acordo com os Desvios, o modelo 2 é superior, pois apresenta menor Desvio. Porém, basear-se somente nos desvios é insuficiente, pois estes valores desconsideram o preditor linear e função de ligação (para mais detalhes, ver Paula, 2013). Por outro, considerar somente os critérios de informação não é suficiente, mas também desconsidera a função de ligação.

Adicionalmente, ao utilizar o Teste da Razão de Verossimilhanças para testar H_0 : $(\alpha\beta)_{21} = (\alpha\gamma)_{21} = (\alpha\gamma)_{21} = (\alpha\gamma)_{22} = 0$, obtêm-se o valor-p de 0,5994, o que leva à não rejeição da hipótese nula, o que fornece suporte ao resultado do teste citado ao final da análise do modelo 3 e nos leva a preferir o modelo 3. Portanto, optou-se por manter o modelo 3, como definitivo.

Para este modelo, com probabilidades

$$p_{ijk} = \frac{\exp(\mu + \alpha_i + \beta_j + \gamma_k + (\beta \gamma)_{jk})}{1 + \exp(\mu + \alpha_i + \beta_j + \gamma_k + (\beta \gamma)_{jk})}$$

abaixo estão as interpretações dos parâmetros para tal modelo.

- Y_{ijk}: reflete a quantidade de sementes germinadas, após 5 dias, de um total de 100, para a i-ésima temperatura de germinação,
 o j-ésimo nível de umidade e k-ésima temperatura de umidade.
- $e^{\beta_1} = \frac{\left(\frac{P_{111}}{1-P_{111}}\right)}{\left(\frac{P_{131}}{1-P_{131}}\right)}$ é a razão de chances que nos indica o quanto maior ou menor é a proporção de sementes germinadas depois de 5 dias a uma temperatura de germinação de 11°C ou 21°C, para umidade baixa em relação a alta.

Tabela 1: Critérios de Seleção para os Modelos Ajustados

	Desvios	Valor-p	AIC	BIC
modelo 1	0	0	103.10	119.13
modelo 2	1.65	0.80	96.76	109.22
modelo 3	4.43	0.82	91.53	100.44

Tabela 2: Nesta tabela, estão relatados os Desvios e seus valores-p associados (de forma que os graus de liberdade são dados pela diferença entre as quantidade de observações e parâmetros ajustados), AIC (Critério de Informação de Akaike), BIC (Critério de Informação Bayesiana) à cada modelo. Pode-se perceber que o terceiro modelo possui os menores critérios de informação. Entretanto, modelo 2 apresenta desvio menor do que o modelo 3. Note, também, que o modelo 1 apresenta desvio nulo, isto deve-se à sua saturação, o que não fornece muita informação. (Para mais detalhes, veja (Paula, 2013))

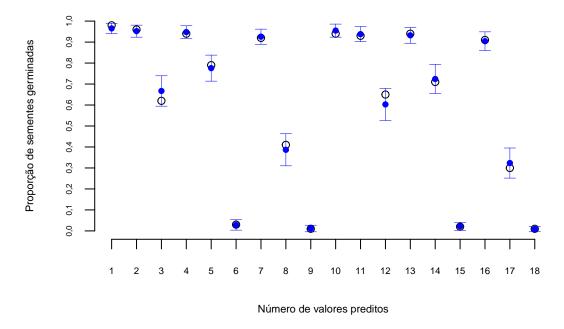


Figura 8: Gráfico da análise preditiva. O eixo horizontal deste gráfico representa as observações (casela), enquanto que o eixo vertical representa a proporção de sementes germinadas. Os pontos azuis representam as estimativas pontuais preditas pelo modelo, com seus respectivos intervalos de 95 porcento de confiança. Além disso, o círculo vazio representa a proporção observada nos dados. Note que todos os valores observados estão conditos nos intervalos.

- $e^{\beta_2} = \frac{\left(\frac{p_{121}}{1-p_{121}}\right)}{\left(\frac{p_{131}}{1-p_{131}}\right)}$ é a razão de chances que nos indica o quanto maior ou menor é a proporção de sementes germinadas depois de 5 dias a uma temperatura de germinação de 11°C ou 21°C, para umidade média em relação a alta.
- $e^{\gamma_2} = \frac{\left(\frac{P_2 32}{1-p_2 32}\right)}{\left(\frac{P_2 31}{1-p_2 32}\right)}$ é a razão de chances que nos indica o quanto maior ou menor é a proporção de sementes germinadas depois de 5 dias a uma temperatura de germinação de 11°C ou 21°C, para temperatura de 21°C em relação à 42°C.
- $e^{\gamma_3} = \frac{\left(\frac{P_{233}}{1-p_{233}}\right)}{\left(\frac{P_{231}}{1-p_{231}}\right)}$ é a razão de chances que nos indica o quanto maior ou menor é a proporção de sementes germinadas depois de 5 dias a uma temperatura de germinação de 11°C ou 21°C, para temperatura de 21°C em relação à 62°C.
- $e^{\alpha_2} = \frac{\left(\frac{P_{213}}{1-P_{213}}\right)}{\left(\frac{P_{113}}{1-P_{113}}\right)}$ é a razão de chances que nos indica o quanto maior ou menor é a proporção de sementes germinadas depois de 5 dias a uma temperatura de germinação de 11°C ou 21°C, para temperatura de germinação de 21°C em relação à 11°C.

Para este modelo, a figura acima mostra o resultado da análise preditiva. Pode-se perceber que o modelo fornece um resultado satisfatório, pois o intervalo de 95% confiança contempla todos os valores observados.

Por fim, sob este modelo, a melhor combinação de fatores experimentais é dada por baixa temperatura de umidade, temperatura de germinação de 11°C e temperatura de umidade de 21°C.

4. Conclusões

1. descritiva

- a quantidade de 100 aparanetemente é perquena p/ ver se o efeito da temperatura de germinacao é alto
- ao aumentar nivel de umidade, temperatura de germinacao e umidade, pior fica a cosia

2. inferencial

- nao sabemos explicar pq o modelo saturado apresenta aqueles valores-p
- o envelope mostra um efeito nao capturado pelo modelo com tais variaveis (dependencia nos residuos)
- modelo preferido é o 3
- as interacoes tem impacto mt alto
- a analise preditiva mostra que o modelo é razoavel

•

Bibliografia

- Azevedo, C. L. N. (2017). Notas de aula sobre análise de dados discretos, (http://www.ime.unicamp.br/~cnaber/Material_ ADD_1S_2017.htm)
- 2. Paula, G. A. (2013). Modelos de regressão com apoio computacional, versão pré-eliminar, (http://www.ime.usp.br/~giapaula/texto_2013.pdf)

Anexos

Tabelas com estimativas pontuais e intervalares, erros padrão e valores-p. As tabelas a seguir são referentes aos modelos 1,2 e 3, respectivamente

Graficos

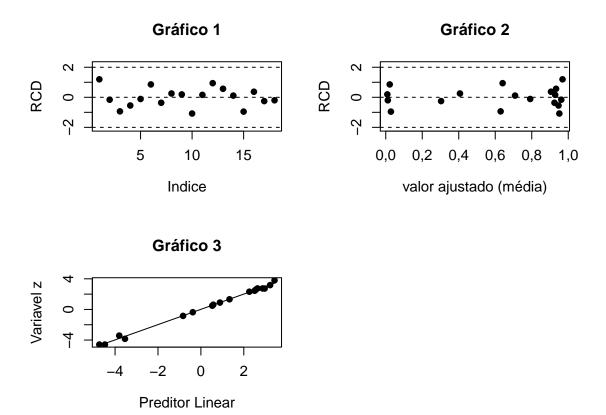


Figura 9: Gráficos para verificação da adequabilidade do ajuste do modelo 2.

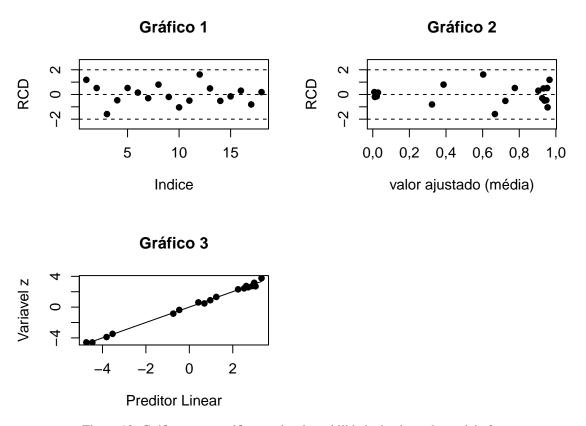


Figura 10: Gráficos para verificação da adequabilidade do ajuste do modelo 3.

Tabela 3: Tabela Resumo para Modelo 1

	Estimativa	Erro Padrão	2.5 %	97.5 %	Estatística Z	Valor-p
μ	2.442	0.369	1.720	3.165	6.626	0.000
$oldsymbol{eta}_1$	1.450	0.804	-0.126	3.025	1.803	0.071
eta_2	0.309	0.560	-0.788	1.406	0.552	0.581
γ_1	-2.806	0.421	-3.631	-1.981	-6.666	0.000
γ_2	-7.037	1.071	-9.136	-4.939	-6.574	0.000
$lpha_2$	-0.129	0.508	-1.124	0.867	-0.253	0.800
$(\beta \gamma)_{11}$	2.092	0.974	0.184	4.001	2.149	0.032
$(\beta \gamma)_{21}$	1.380	0.644	0.117	2.642	2.142	0.032
$(\beta \gamma)_{12}$	3.635	1.303	1.081	6.190	2.789	0.005
$(\beta \gamma)_{22}$	0.810	1.291	-1.721	3.340	0.627	0.530
$(\alpha\beta)_{21}$	-1.012	0.972	-2.917	0.894	-1.040	0.298
$(\alpha\beta)_{22}$	0.129	0.783	-1.405	1.663	0.165	0.869
$(\alpha \gamma)_{21}$	-0.355	0.589	-1.509	0.800	-0.602	0.547
$(\alpha\gamma)_{22}$	0.129	1.509	-2.830	3.087	0.085	0.932
$(\alpha\beta\gamma)_{211}$	0.903	1.204	-1.455	3.262	0.751	0.453
$(\alpha\beta\gamma)_{221}$	-0.075	0.900	-1.839	1.690	-0.083	0.934
$(\alpha\beta\gamma)_{212}$	1.141	1.747	-2.283	4.565	0.653	0.514
$(\alpha\beta\gamma)_{222}$	-0.544	1.867	-4.204	3.115	-0.292	0.771

Tabela 4: Tabela Resumo para Modelo 2

	Estimativa	Erro Padrão	2.5 %	97.5 %	Estatística Z	Valor-p
μ	2.506	0.334	1.852	3.161	7.508	0.000
$oldsymbol{eta}_1$	0.926	0.526	-0.105	1.957	1.759	0.079
eta_2	0.371	0.443	-0.496	1.239	0.839	0.402
γ_1	-2.884	0.365	-3.599	-2.169	-7.907	0.000
γ_2	-7.239	0.811	-8.829	-5.650	-8.929	0.000
$lpha_1$	-0.247	0.394	-1.019	0.524	-0.628	0.530
$(\beta \gamma)_{11}$	2.678	0.562	1.576	3.779	4.765	0.000
$(\beta \gamma)_{21}$	1.341	0.450	0.458	2.223	2.978	0.003
$(\beta \gamma)_{12}$	4.338	0.850	2.672	6.003	5.103	0.000
$(\beta \gamma)_{22}$	0.556	0.929	-1.265	2.377	0.599	0.549
$(\alpha\beta)_{21}$	-0.212	0.486	-1.165	0.741	-0.436	0.662
$(\alpha\beta)_{22}$	0.008	0.375	-0.728	0.744	0.021	0.983
$(\alpha \gamma)_{21}$	-0.207	0.406	-1.002	0.589	-0.509	0.611
$(\alpha\gamma)_{22}$	0.507	0.530	-0.531	1.545	0.957	0.339

Tabela 5: Tabela Resumo para Modelo 3

	Estimativa	Erro Padrão	2.5 %	97.5 %	Estatística Z	Valor-p
μ	2.522	0.267	1.999	3.046	9.447	0.000
eta_1	0.803	0.441	-0.062	1.667	1.819	0.069
eta_2	0.376	0.391	-0.391	1.143	0.960	0.337
γ_1	-2.984	0.294	-3.560	-2.408	-10.149	0.000
γ_2	-6.989	0.755	-8.468	-5.509	-9.258	0.000
α_2	-0.276	0.149	-0.569	0.016	-1.853	0.064
$(\beta \gamma)_{11}$	2.650	0.559	1.553	3.746	4.736	0.000
$(\beta \gamma)_{21}$	1.328	0.449	0.447	2.208	2.955	0.003
$(\beta \gamma)_{12}$	4.358	0.849	2.693	6.023	5.130	0.000
$(\beta \gamma)_{22}$	0.556	0.929	-1.265	2.377	0.598	0.550