

4

Resultados – O Modelo

Neste capítulo é apresentada a modelagem da média e da variância da variável resposta utilizando modelos lineares mistos. Os dados utilizados foram os resultantes do experimento apresentado no Capítulo 3. Neste Capítulo 4 é descrito como foi selecionado o modelo para a média e para a variância, incluindo as estruturas utilizadas para a matriz de covariâncias e correlação, necessárias nos modelos mistos. Por fim, é apresentada a estrutura do modelo para a média da espessura e para a variância da espessura do material utilizado na fabricação das latas de duas peças, assim como a análise dos resíduos.

4.1

Motivo de utilizar a modelagem mista para o tratamento dos dados

O delineamento utilizado no presente estudo, foi um desenho fatorial fracionado, com cinco fatores, sendo utilizados dois níveis em cada fator. Porém, não houve aleatorização nas rodadas do experimento, assim como nas replicações realizadas. Além disso, em nenhum momento os níveis dos fatores foram reinicializados. Dessa forma, analisar os dados por meio do método de mínimos quadrados ordinários não seria viável, dado que a premissa de independência das observações não é observada por definição.

A fim de utilizar um método de análise quando temos um desenho fatorial fracionado com restrição na aleatorização, optou-se pela modelagem mista. O *software* utilizado na análise foi o S-Plus, versão 6.0. Na subseção 4.2, a seguir, serão descritas as características do modelo utilizado para a média da espessura do material em questão.

4.2

Modelagem da média da espessura do material

Para esta modelagem foi utilizada a média da espessura como variável resposta (Y), e trabalhando-se com modelos lineares mistos, foram incluídos efeitos fixos para os fatores A, B, C, D e E, descritos no Capítulo 3, e suas interações. Foram consideradas apenas interações de segunda ordem.

Como efeitos aleatórios, foram considerados o intercepto e o *slope*. Nesse caso, foi utilizado o fator D, uma vez que foram utilizados apenas dois diâmetros do cilindro no experimento – e cada bobina recebeu apenas um dos níveis desse fator.

Ainda sobre os efeitos aleatórios, foi considerada a variável de grupamento (bloco). Analisando as condições em que foi realizado o experimento, optou-se por usar como bloco as ‘rodadas’ (de 1 a 16) do experimento, uma vez que elas foram realizadas sequencialmente para as quatro replicações. É válido mencionar que blocagem e aleatorização estão diretamente relacionadas, tendo em vista que a primeira é também considerada uma restrição no processo de aleatorização do experimento. Dessa forma, o bloco se refere às 16 combinações de níveis dos fatores.

Como mencionado no Capítulo 3, a base de dados contém 64 observações para as variáveis resposta – média e variância da espessura do material utilizado na fabricação das latas de duas peças.

4.2.1

Detalhes em relação à modelagem da média da espessura do material

No modelo misto é possível especificar a estrutura de correlação dentro do grupo, a estrutura de heterocedasticidade e o método usado para ajustar o modelo. O padrão (*default*) é que os erros dentro do grupo sejam independentes e tenham variâncias iguais (homocedásticos).

No S-Plus, a estrutura de covariância do erro dentro do grupo (*within-group*) pode ser modelada por meio da combinação de estruturas de correlação e funções de variância. As estruturas de correlação são usadas para modelar as correlações *within-group* que não são capturadas pelos efeitos aleatórios. Elas estão geralmente associadas à dependência temporal ou espacial. Já as funções de variância são usadas para modelar a heterocedasticidade nos erros *within-group*. Como estrutura de correlação, foi utilizada a *compound symmetry*, pois foi a que forneceu os melhores resultados no ajuste – com base na estatística AIC. Em relação à função da variância, foi utilizado o tipo ‘*Identity*’, que estima diferentes variâncias por nível de um fator. As demais funções requerem a presença de covariantes, e, portanto não são possíveis de serem usadas com os dados do experimento.

Na modelagem mista, também é possível especificar opções de otimização para as funções de ajuste do modelo. Para o modelo em questão, optou-se por utilizar o método da máxima verossimilhança restrita (REML), uma vez que tal método fornece uma correção para o viés encontrado nas estimativas por máxima verossimilhança.

4.2.2

Resultados da modelagem da média da espessura do material

Para a seleção dos fatores, e/ou interações de segunda ordem entre os fatores mais significativos na determinação da espessura média do material utilizado na fabricação das latas para bebidas carbonatadas, procedeu-se ao método *backward elimination*. Para isso, foram usados os seguintes fatores e interações: A, B, C, D, E, AB, AC, AD, AE, BC, BD, BE, CD, CE, DE, descritos no Capítulo 3.

Considerou-se as especificações *Identity*, para a variância *within-group* (função de variância) na qual utiliza diferentes desvios-padrão por estratos, e *compound symmetry* para a correlação *within-group* (estrutura de correlação), ambas referentes aos resíduos, e a especificação *General Symmetric* para estrutura de covariância dos efeitos aleatórios. Sendo assim, a ordem de exclusão das variáveis foi, respectivamente: AD, AC, BE, E, A, DE, BC, CE, AB, CD, B, BD.

As três variáveis significativas, ao nível de significância de 1%, foram: C, D e AE. Vale ressaltar que o ‘bloco’ foi considerado como ‘*Group variable*’ tanto na função de variância *within-group* quanto na correlação *within-group* dos resíduos. O resultado encontra-se a seguir, na Tabela 4.

Tabela 4– Estimativa dos parâmetros para o ajuste da média da espessura do material, utilizando modelos lineares mistos

	Coefficiente	Erro-padrão	gl	t	valor-p
Intercept	0,2476317	0,00002745497	48	9019,558	<0,0001
C	0,0000405	0,00000971077	12	4,175	0,0013
D	0,0001855	0,00003115567	12	5,953	0,0001
AE	0,0000603	0,00001856725	12	3,246	0,0070

Nota 1: Variável dependente: Y_m

Figura 1- Diagrama de dispersão dos resíduos

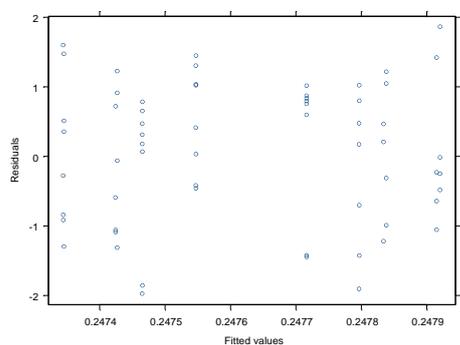
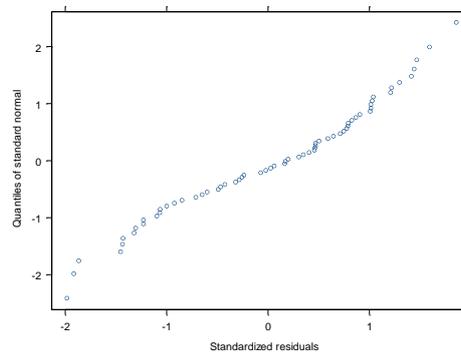


Figura 2- QQ Plot



Observando o valor-p, na Tabela 4, existe uma interação significativa entre os fatores A e E. E uma alta significância dos fatores C e D. Sendo assim, o modelo tem as três variáveis significativas e os gráficos dos resíduos (Figuras 1 e 2) satisfazem às premissas de normalidade e homocedasticidade.

Em seguida, foi testado se o intercepto e o coeficiente angular aleatórios eram necessários. Para isso, omitiu-se cada um deles do modelo e calculou-se a estatística razão de verossimilhança, contrastando o modelo reajustado com o modelo original. Os resultados estão apresentados a seguir (Tabela 5 e Tabela 6), e mostram que ambos o intercepto e a inclinação aleatórios não são necessários para o modelo (valor-p igual a 0,998).

Tabela 5- Resultado do teste de Razão de Verossimilhança, para a verificação do fator ‘D’ como aleatório

Modelo	gl	AIC	BIC	logLik	Teste	L.Ratio	valor-p
1	24	-767,8570	-717,5927	407,9285			
2	22	-771,8609	-725,7853	407,9305	1 vs 2	0,00397	0,998

Nota 1: Modelo 1: existe o fator D como fator aleatório, Intercepto e Bloco

Nota 2: Modelo 2: o fator D foi retirado dos fatores aleatórios

Tabela 6 - Resultado do teste de Razão de Verossimilhança, para a verificação do ‘intercepto’ como aleatório

Modelo	gl	AIC	BIC	logLik	Teste	L.Ratio	valor-p
1	24	-767,8570	-717,5927	407,9285			
2	22	-771,8609	-725,7853	407,9305	1 vs 2	0,00397	0,998

Nota 1: Modelo 1: existe o fator D como fator aleatório, Intercepto e Bloco

Nota 2: Modelo 2: o ‘intercepto’ foi retirado dos fatores aleatórios

Decidiu-se então testar se o ‘Bloco’ seria significativo. Caso contrário, só existiriam efeitos fixos. Desta vez, vemos pelo valor-p ($< 0,0001$), apresentado na Tabela 7, que o ‘Bloco’ é necessário no modelo. Além disso, se o ‘Bloco’ for excluído como ‘*Group Variable*’ da função de variância e da estrutura de correlação, os gráficos dos resíduos sinalizam padrões problemáticos (não normalidade, *outliers* e heterocedasticidade), como apresentado nos Gráficos 3 e 4, a seguir.

Tabela 7 - Resultado do teste de Razão de Verossimilhança, para a verificação do ‘Bloco’ como aleatório

Modelo	gl	AIC	BIC	logLik	Teste	L.Ratio	valor-p
1	24	-767,8570	-717,5927	407,9285			
2	09	-714,1367	-695,2876	366,0684	1 vs 2	83,72024	$<0,0001$

Nota 1: Modelo 1: existe o fator D como fator aleatório, Intercepto e Bloco

Nota 2: Modelo 2: o ‘Bloco’ foi retirado dos fatores aleatórios

Figura 3 - Diagrama de dispersão dos resíduos sem a presença de blocos'

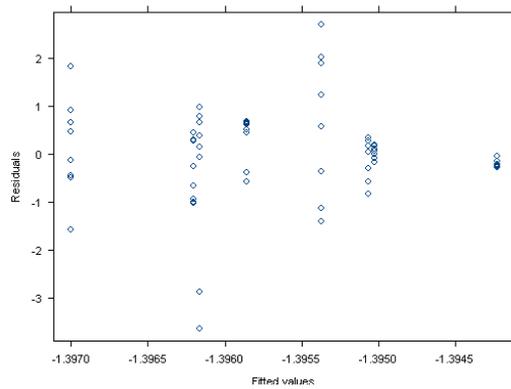
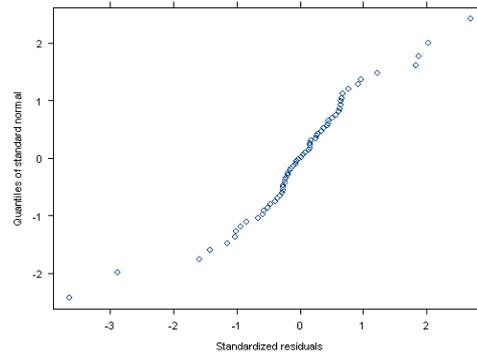


Figura 4- QQ Plot sem a presença de 'blocos'



Dessa forma, optou-se por retirar o fator D dos efeitos aleatórios, mas manter o 'Bloco'. O próximo passo foi testar se a variância *within-group* (função de variância) era necessária na especificação do modelo. Por meio da razão de verossimilhança, temos valor-p $< 0,0001$, indicando que a função de variância é necessária. Este resultado é apresentado na Tabela 8.

Tabela 8 - Resultado do teste de Razão de Verossimilhança, para a verificação da ‘função de variância’ dos resíduos

Modelo	gl	AIC	BIC	logLik	Teste	L.Ratio	valor-p
1	24	-767,8570	-717,5927	407,9285			
2	07	-718,4714	-703,8110	366,2357	1 vs 2	83,38557	<0,0001

Nota 1: Modelo 1: com a função de variância dos resíduos

Nota 2: Modelo 2: sem a função de variância dos resíduos

Por fim, restava testar se a correlação *within-group* era significativa. Uma vez que a autocorrelação estimada (-0.1538) foi baixa, supunha-se que não havia autocorrelação serial significativa. O teste de razão de verossimilhança estabelece a significância estatística da autocorrelação do erro. Por meio do valor-p (0,1845), vemos que a autocorrelação não é significativa. Com isso não é necessário especificar a correlação *within-group*. Este resultado encontra-se na Tabela 9.

Tabela 9 - Resultado do teste de Razão de Verossimilhança, para a verificação da ‘correlação *within-group*’

Modelo	gl	AIC	BIC	logLik	Teste	L.Ratio	valor-p
1	24	-767,8570	-717,5927	407,9285			
2	21	-772,1018	-728,1205	407,0509	1 vs 2	1,755201	0,6247

Nota 1: Modelo 1: com a correlação *within-group*

Nota 2: Modelo 2: sem a correlação *within-group*

Como a especificação do modelo foi alterada, o termo aleatório (D) foi testado novamente. O resultado encontra-se a seguir (Tabela 10), mostrando, pelo valor-p, que o fator D não é realmente necessário no modelo como um efeito aleatório.

Tabela 10 - Resultado do teste de Razão de Verossimilhança, para a verificação do fator ‘D’ como aleatório após a alteração na estrutura do modelo

Modelo	gl	AIC	BIC	logLik	Teste	L.Ratio	valor-p
1	23	-768,0984	-719,9285	407,0492			
2	21	-772,1018	-728,1205	407,0509	1 vs 2	0,003305024	0,9983

Nota 1: Modelo 1: fator D como aleatório, na nova estrutura obtida

Nota 2: Modelo 2: sem o fator D como aleatório, na nova estrutura obtida

Após a alteração nas especificações do modelo, a interação AE deixou de ser significativa ao nível de 1% (valor-p = 0,0371). Sendo assim, tal interação foi retirada do modelo. Dessa forma, o modelo, final, ajustado para a média da espessura do material utilizado na fabricação das latas para bebidas carbonatadas tem como variável dependente a média da espessura do material, e como efeitos fixos os fatores C e D. É usada como variável de grupamento (Bloco) as 16 rodadas do experimento. Tal variável também faz parte da função de variância (variância *within-group*), como variável de grupamento. Na parte dos efeitos aleatórios é considerado o intercepto e a variável de grupamento, fundamental para a correta especificação do modelo. A estimativa dos parâmetros está apresentada a seguir, na Tabela 11, juntamente com o gráfico dos resíduos (Figuras 5 e 6), onde se verificam as premissas de normalidade e homocedasticidade. Os demais resultados obtidos nesta modelagem encontram-se no Anexo 1.

Tabela 11 – Estimativa dos parâmetros para o ajuste da média da espessura do material, utilizando modelos lineares mistos

	Coefficiente	Erro-padrão	gl	t	valor-p
Intercept	0,2476336	0,00003807525	48	6503,794	<0,0001
C	0,0000466	0,00001246532	13	3,739	0,0025
D	0,0002348	0,00003785850	13	6,203	<0,0001

Nota: Variável dependente: Y_m

Figura 5 - Diagrama de dispersão dos resíduos

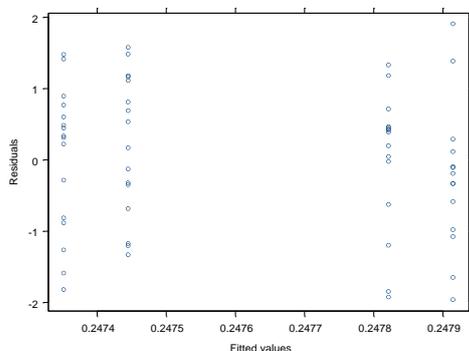
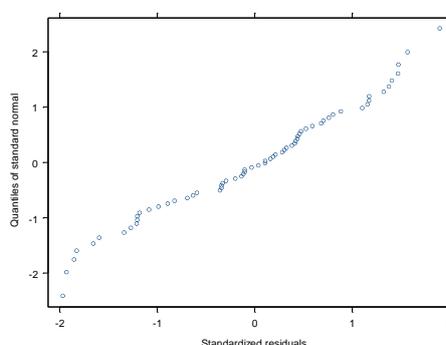


Figura 6 - QQ Plot



Sendo assim, o modelo estimado para a média da espessura tem a seguinte forma:

$$\hat{Y}_m = 0,2476 + 0,0000466 C + 0,0002348 D$$

Ao comparar tal modelo com o encontrado por Gomes (2007), utilizando a mesma base de dados, vemos uma estrutura diferente. Gomes (2007) utilizou inicialmente o método dos mínimos quadrados ordinários. Em sua modelagem, o único fator que apresentava valor-p < 5% era o fator D. Além disso, a modelagem por mínimos quadrados não se ajustou bem aos dados, que apresentavam comportamento heterocedástico. Dessa forma, foi necessário modelar a variância usando modelos lineares generalizados (GLM), para que se conseguisse uma correta estrutura do modelo. Aqui, o que obtivemos foi a modelagem da mesma variável por modelos lineares mistos, que permitiu que se atingisse uma correta estrutura sem a necessidade da modelagem da variância da média. Vale destacar que a estimativa do valor médio no modelo encontrado por Gomes (2007) foi de 0,2478877, enquanto que o resultado gerado por modelos mistos foi 0,247915, sendo o valor-alvo 0,248.

4.3

Modelagem da variância da espessura do material

O modelo utilizado para descrever a variância da espessura do material utilizou, em sua construção, como variável resposta, o logaritmo neperiano da variância da espessura. Como efeitos fixos foram testados os fatores A, B, C, D e E, e suas interações de segunda ordem. Como efeito aleatório foi utilizado o fator D. Além disso, em relação aos efeitos aleatórios, foi considerada a variável de agrupamento (bloco). O bloco se refere às 16 combinações de níveis dos fatores.

4.3.1

Detalhes em relação à modelagem da variância da espessura do material

Para tal modelagem, a estrutura de correlação utilizada foi a *compound symmetry*. Em relação à função da variância, foi utilizado o tipo ‘*Identity*’, que é aquela que estima diferentes variâncias por nível de um fator. O método de otimização utilizado foi o método da máxima verossimilhança restrita (REML).

4.3.2

Resultados da modelagem da variância da espessura do material

Para a seleção dos fatores, e/ou interações de segunda ordem entre os fatores mais significativos, procedeu-se ao método *backward elimination*. Para isso, foram usados os seguintes fatores e interações: A, B, C, D, E, AB, AC, AD, AE, BC, BD, BE, CD, CE, DE.

Considerando as especificações *Identity*, para a variância *within-group*, e *compound symmetry* para a correlação *within-group*, e a especificação *General Symmetric* para estrutura de covariância dos efeitos aleatórios, as cinco variáveis significativas foram: C, D, E, AD e BD.

Vale destacar que a variável ‘bloco’ foi considerada como ‘*Group variable*’ tanto na função de variância *within-group* quanto na correlação *within-group* dos resíduos. O resultado gerado pelo S-Plus está demonstrado a seguir, na Tabela 12.

Tabela 12 – Estimativa dos parâmetros para o ajuste do ln da variância da espessura do material, utilizando modelos lineares mistos

	Coefficiente	Erro-padrão	gl	t	valor-p
(Intercept)	-7,261054	0,02205210	48	-329,2682	<0,0001
C	-0,092812	0,02567200	10	-3,6153	0,0047
D	-0,091885	0,02006713	10	-4,5789	0,0010
E	-0,129932	0,02087384	10	-6,2246	0,0001
AD	-0,156840	0,02494760	10	-6,2868	0,0001
BD	-0,143941	0,02276703	10	-6,3223	0,0001

Nota: Variável dependente: LN.Yv

Figura 7 - Diagrama de dispersão dos resíduos

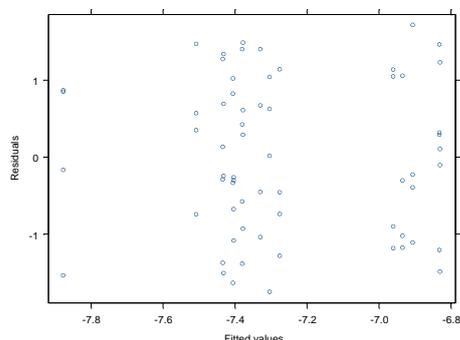
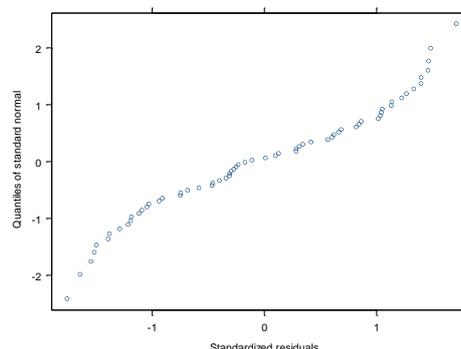


Figura 8 - QQ Plot



Observando o valor- $p < 1\%$, na Tabela 12, existem duas interações significativas (AD e BD) e uma alta significância dos fatores C, D e E. Além disso, é observado nos gráficos dos resíduos que as premissas básicas são atendidas.

O próximo passo, então, foi testar a presença de autocorrelação serial dos resíduos, por meio da estatística de razão de verossimilhança. Segundo o resultado apresentado na Tabela 13, a seguir, vemos que não existe autocorrelação serial nos resíduos (valor- $p = 11,42\%$).

Tabela 13 - Resultado do teste de Razão de Verossimilhança, para a verificação da ‘correlação *within-group*’

Modelo	gl	AIC	BIC	logLik	Teste	L.Ratio	valor-p
1	26	191,1753	244,7468	-69,58766			
2	25	191,6710	243,1821	-70,83550	1 vs 2	2,495684	0,1142

Nota 1: Modelo 1: com a correlação *within-group*

Nota 2: Modelo 2: sem a correlação *within-group*

Testou-se, então, o fator D como aleatório, e em seguida, testou-se a presença da variável de grupamento. Os resultados encontram-se nas Tabelas 14 e 15, respectivamente. Por meio deles, vemos que o fator D não tem efeito significativo como aleatório, porém, existe efeito do ‘Bloco’ (valor- $p < 1\%$).

Tabela 14 - Resultado do teste de Razão de Verossimilhança, para a verificação do fator 'D' como aleatório

Modelo	gl	AIC	BIC	logLik	Teste	L.Ratio	valor-p
1	26	191,1753	244,7468	-69,58766			
2	24	187,1753	236,6260	-69,58766	1 vs 2	7,11e-6	1

Nota 1: Modelo 1: existe o fator D como fator aleatório

Nota 2: Modelo 2: o fator D foi retirado dos fatores aleatórios

Tabela 15 - Resultado do teste de Razão de Verossimilhança, para a verificação do 'Bloco' como aleatório

Modelo	gl	AIC	BIC	logLik	Teste	L.Ratio	valor-p
1	26	191,1753	244,7468	-69,58766			
2	11	197,8848	220,5496	-87,94238	1 vs 2	36,70944	0,0014

Nota 1: Modelo 1: existe o Bloco nos fatores aleatórios

Nota 2: Modelo 2: o 'Bloco' foi retirado dos fatores aleatórios

Finalmente, testou-se a presença de heterocedasticidade. O resultado apresentado na Tabela 16 comprova que a especificação da função de variância é relevante (valor-p < 1%).

Tabela 16 - Resultado do teste de Razão de Verossimilhança, para a verificação da 'função de variância' dos resíduos (*weight*)

Modelo	gl	AIC	BIC	logLik	Teste	L.Ratio	valor-p
1	26	191,1753	244,7468	-69,58766			
2	11	196,9323	219,5972	-87,46617	1 vs 2	35,75702	0,0019

Nota 1: Modelo 1: com a função de variância dos resíduos

Nota 2: Modelo 2: sem a função de variância dos resíduos

Após testar as especificações do modelo, rodou-se o modelo final novamente, com a especificação da função de variância, mas sem a estrutura de correlação. Além disso, o fator D foi removido da parte de efeitos aleatórios. O resultado está apresentado na Tabela 17. Com isso, o fator C deixa de ser significativo (valor-p = 8,5%).

Tabela 17 - Estimativa dos parâmetros para o ajuste do ln da variância da espessura do material, utilizando modelos lineares mistos

	Coefficiente	Erro-padrão	gl	t	valor-p
(Intercept)	-14,54601	0,07251025	48	-200,6062	<0,0001
C	-0,15962	0,08351429	10	-1,9113	0,0850
D	-0,17074	0,06856019	10	-2,4903	0,0320
E	-0,28260	0,06958453	10	-4,0612	0,0023
AD	-0,34774	0,08109466	10	-4,2880	0,0016
BD	-0,26121	0,07648874	10	-3,4150	0,0066

Nota: Variável dependente: LN.Yv

Após ele ser retirado do modelo, temos o seguinte resultado (Tabela 18), mostrando que o fator D deixa de ser significativo (valor-p > 11%).

Tabela 18 - Estimativa dos parâmetros para o ajuste do ln da variância da espessura do material, utilizando modelos lineares mistos

	Coefficiente	Erro-padrão	gl	t	valor-p
(Intercept)	-14,62157	0,06771857	48	-215,9168	<,0001
D	-0,11581	0,06670149	11	-1,7363	0,1104
E	-0,32622	0,06895907	11	-4,7306	0,0006
AD	-0,45213	0,06564365	11	-6,8877	<,0001
BD	-0,23404	0,07498335	11	-3,1213	0,0097

Nota: Variável dependente: LN.Yv

Porém, como o fator D está presente nas interações AD e BD, optou-se por deixá-lo no modelo. O diagrama de dispersão dos resíduos, e o QQ-Plot estão apresentados a seguir.

Figura 9 - Diagrama de dispersão dos resíduos

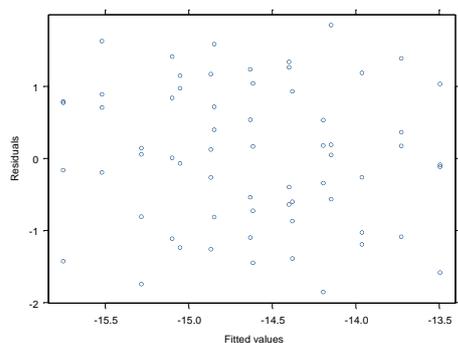
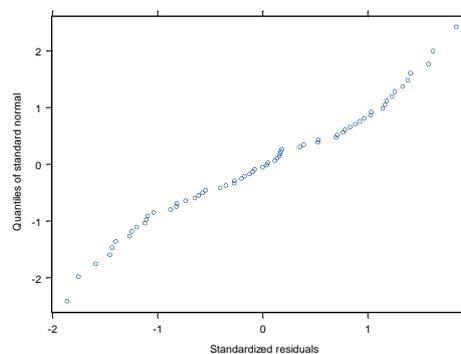


Figura 10 -: QQ Plot



Por fim, decidiu-se incluir no modelo os efeitos simples A e B, por fazerem parte das interações que são significativas ao nível de 1%. O resultado encontra-se na Tabela 19¹.

Tabela 19 - Estimativa dos parâmetros para o ajuste do ln da variância da espessura do material, utilizando modelos lineares mistos

	Coefficiente	Erro-padrão	gl	t	valor-p
(Intercept)	-14,60499	0,06979241	48	-209,2633	<,0001
D	-0,10782	0,06834414	9	-1,5777	0,1491
E	-0,32931	0,08824760	9	-3,7316	0,0047
AD	-0,40990	0,07731232	9	-5,3019	0,0005
BD	-0,27993	0,08775556	9	-3,1899	0,0110
A	-0,08761	0,08034974	9	-1,0903	0,3039
B	0,09275	0,08565853	9	1,0827	0,3071

Nota: Variável dependente: LN.Yv

Figura 11 - Diagrama de dispersão dos resíduos

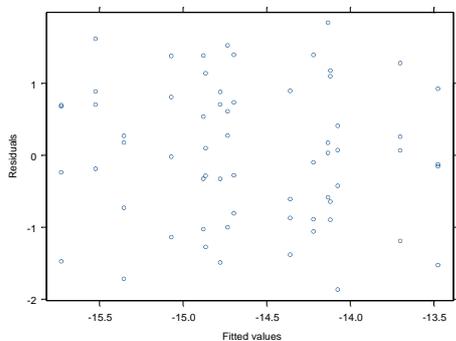
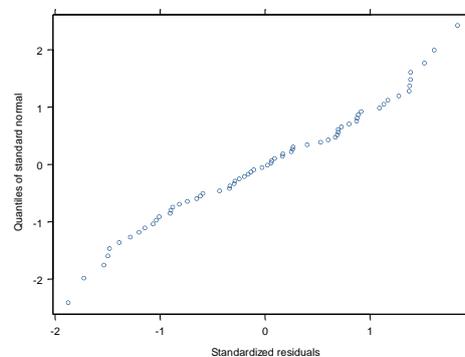


Figura 12 - QQ Plot



Dessa forma, o modelo para a variância da espessura do material utilizado na fabricação das latas para bebidas carbonatadas tem como variável dependente o logaritmo neperiano da variância da espessura do material, e como efeitos fixos os fatores A, B, D e E, sendo que o único significativo ao nível de 1% é o fator E. Os demais fatores fazem parte do modelo, pois eles estão presentes nas interações. As

¹ Os demais resultados encontram-se no Anexo 2.

interações AD e BD foram significativas (valor-p < 1% e valor-p < 5%, respectivamente). É usada como variável de grupamento (bloco) as 16 rodadas do experimento. Tal variável (bloco) também faz parte da função de variância (variância *within-group*), como variável de grupamento. Na parte dos efeitos aleatórios é considerado o intercepto e a variável de grupamento, fundamental para a correta especificação do modelo. Por meio do gráfico dos resíduos verifica-se o atendimento às premissas básicas do modelo.

Sendo assim, o modelo estimado para a variância da espessura do material tem a seguinte forma:

$$\hat{Y}_v = \exp [-14,60499 + (- 0,08761 - 0,40990D)A + (0,09275 - 0,27993D)B - 0,10782D - 0,32931E]$$

Ao comparar o modelo encontrado, utilizando-se modelagem mista, com o modelo apresentado por Gomes (2007), observamos uma estrutura diferente no conjunto de variáveis selecionadas. No modelo encontrado por Gomes (2007), os fatores que possuíam efeito significativo foram C, D, E, assim como a interação AD. Vale citar que Gomes (2007) utilizou GLM para o ajuste da variância da espessura. No modelo encontrado por Gomes (2007) a variância estimada era igual a 1,95 E-07, enquanto que utilizando modelos mistos a variância estimada foi de 1,5 E-07. Em ambos os modelos a variância mínima ocorre quando os fatores estão no nível alto.

Neste capítulo foi apresentada a modelagem para a média e para a variância da espessura do material utilizado na fabricação das latas de duas peças. Foi descrito o método de seleção dos modelos da média e da variância, incluindo as estruturas utilizadas para a matriz de covariâncias. E, finalmente, foi apresentada a estrutura do modelo para a média da espessura e para a variância da espessura do material, além da análise dos resíduos.