

Modelo linear misto bivariado na análise de dados de concentração de caseínas durante o armazenamento de leite UAT

Simone Silmara Werner Gurgel do Amaral ^{1 3}

César Gonçalves de Lima ²

1 Introdução

Em trabalhos com armazenamento de produtos agropecuários é comum a mensuração de diversas variáveis respostas nas mesmas unidades experimentais no decorrer do tempo, o que produz um conjunto de dados longitudinais multivariado, sendo de interesse do pesquisador conhecer a evolução conjunta dessas variáveis que irão determinar a qualidade do produto. Neste caso, conhecer o comportamento das variáveis resposta isoladamente pode não ser suficiente, pois no processo de envelhecimento dos produtos as variáveis interagem. Além da covariância existente entre observações feitas na mesma unidade experimental ao longo do tempo tem-se a covariância entre as diferentes variáveis respostas mensuradas (WEISS, 2005).

De acordo com FIEUWS et al. (2007), quando mais de uma variável resposta está presente no estudo pode-se empregar um modelo misto para cada variável resposta e uni-los em uma classe de modelos mistos multivariados especificando a distribuição conjunta para os efeitos aleatórios. Tal abordagem é particularmente útil quando o interesse não está em descrever o comportamento individual de cada uma das variáveis respostas ou descrever o comportamento de uma das variáveis em função de outras (utilizando como variáveis explicativas), mas sim verificar a evolução conjunta das variáveis no decorrer do tempo.

A proposta deste trabalho é avaliar a aplicação da metodologia de análise conjunta de dados longitudinais multivariados em um estudo que envolve a avaliação da concentração de proteínas do α_{s1} e β -caseína durante o armazenamento de Leite UAT.

2 Material e métodos

Os dados utilizados nesse trabalho foram obtidos da tese de Fernandes (2007) que avaliou os efeitos da contagem de células somáticas (CCS) nas frações de caseína do leite UAT (Ultra Alta Temperatura), ao longo do tempo de armazenamento.

¹FAPEAGRO. e-mail: simoneswerner@gmail.com

³Agradecimento ao CNPq pelo apoio financeiro.

²FZEA - USP.

Quinze lotes de leite UAT foram categorizados em grupos de acordo com a CCS: o grupo com baixa CCS (Grupo 1) constituiu-se de lotes com contagens entre 197.000 e 316.000 CS/mL; o grupo com CCS intermediária (Grupo 2), por lotes com contagens entre 379.000 e 560.000 CS/mL e o grupo com alta CCS (Grupo 3) constituiu-se de lotes com contagem entre 600.000 e 800.000 CS/mL.

Cinco amostras por lote do produto (ou seja, cinco embalagens originais de leite UAT contendo, cada uma, 1 litro do produto) foram avaliadas 8, 30, 60, 90 e 120 dias após a fabricação, quanto a quantidade de α_{s1} -caseína e β -caseína (mg/mL).

Na análise dos dados empregou-se o modelo de efeitos aleatórios bivariado, para o qual considerou-se y_{rij} a j -ésima mensuração obtida na amostra de leite proveniente do i -ésimo lote para a r -ésima variável resposta e \mathbf{y}_{ri} , o vetor das medidas da variável r tomadas no indivíduo i . Matricialmente tem-se:

$$\mathbf{Y}_{ri}|\mathbf{b}_{ri} \sim N_{ri}(\mathbf{X}_{ri}\boldsymbol{\beta}_r + \mathbf{Z}_{ri}\mathbf{b}_{ri}, \boldsymbol{\Sigma}_{ri})$$

em que N_{ri} representa a distribuição normal r_i -dimensional, \mathbf{X}_{ri} e \mathbf{Z}_{ri} são as matrizes do delineamento, conhecidas, para os efeitos fixos e aleatórios respectivamente, e $\boldsymbol{\beta}_r$ é o vetor de parâmetros fixos desconhecidos associado à variável r , $r = 1, 2$. $\boldsymbol{\Sigma}_{ri}$ representa a matriz de variância e covariância associada aos componentes do resíduo, e considerando o modelo de independência condicional é dada por $\sigma^2\mathbf{I}_{ri}$.

Assume-se que o vetor \mathbf{b}_i de todos os efeitos aleatórios para o i -ésimo indivíduo, é proveniente de uma distribuição normal multivariada de dimensão $q_1 + q_2$ (em que q_r representa o número de efeitos aleatórios associados à r -ésima variável), com média zero e matriz de covariâncias \mathbf{D} , ou seja:

$$\mathbf{b}_i = \begin{pmatrix} \mathbf{b}_{1i} \\ \mathbf{b}_{2i} \end{pmatrix} \sim N_{q_1+q_2} \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{D}_{21} & \mathbf{D}_{22} \end{pmatrix} \right]$$

em que \mathbf{D}_{11} representa a matriz de variâncias e covariâncias associadas aos efeitos aleatórios da variável α_{s1} -caseína, \mathbf{D}_{22} representa a matriz de variâncias e covariâncias associadas aos efeitos aleatórios da variável β -caseína e \mathbf{D}_{12} representa a matriz de covariâncias entre \mathbf{b}_{1i} e \mathbf{b}_{2i} . Esta matriz permite que os níveis médios dos dois efeitos sejam correlacionados; caso eles não sejam correlacionados \mathbf{D}_{12} será uma matriz nula (WEISS, 2005).

Para ajuste do modelo, foram consideradas diferentes estruturas, tanto para parte fixa como para a parte aleatória. Realizou-se primeiramente a seleção dos efeitos aleatórios utilizando o modelo maximal para a parte fixa (considerando todos os efeitos de grupo e as possíveis interações e o tempo como fator qualitativo) e o método REML de estimação. Em seguida realizou-se a seleção da parte fixa (método da máxima verossimilhança - ML) sendo posteriormente realizada uma nova seleção da parte aleatória, empregando o modelo selecionado para a parte fixa e o método de máxima verossimilhança restrita - REML. Na seleção dos modelos não encaixados empregou-se os Critérios de Informação de Akaike (AIC) e Bayesiano (BIC),

e o teste da razão de verossimilhanças para comparação de modelos encaixados (FIEUWS e VERBEKE, 2004).

Para todos os testes empregou-se o nível de 5% de significância. Para ajuste dos modelos aos dados utilizou-se o *software* estatístico SAS (SAS Institute Inc., 2008) e a análise de diagnóstico foi realizada no software R (R Development Core Team, 2012)

3 Resultados e discussões

Dentre os modelos testados, o modelo selecionado (com o menor valor do AIC e BIC) apresenta quatro componentes de variância (efeito aleatório no intercepto compartilhado pelas variáveis, efeito de coeficiente angular semelhante em cada uma das variáveis e variância residual distinta para cada variável) e quatro parâmetros de efeito fixo: interceptos semelhantes entre os grupos 1 e 2 para $alpha_{s1}$ -caseína e para $beta$ -caseína, interceptos semelhantes para o grupo 3 em relação a $alpha_{s1}$ e $beta$ -caseína e coeficientes angulares semelhantes para as duas variáveis e todos os grupos.

As estimativas dos parâmetros de efeito fixo para o modelo selecionado estão apresentadas na expressão (1).

$$\hat{y}_{rij} = 10,94G12_{1i} + 11,83G12_{2i} + 9,86G3_i - 0,0314t_j \quad (1)$$

Observa-se na expressão (1) que os grupos 1 e 2 apresentam concentrações iniciais de $beta$ -caseína superiores às concentrações de $alpha_{s1}$ -caseína, o que não ocorre no grupo 3, para o qual as concentrações iniciais de $alpha_{s1}$ -caseína e $beta$ -caseína foram consideradas iguais e estimadas em 9,8 mg/mL. O grupo 3 possui a menor concentração das caseínas em estudo, no entanto a redução da concentração no decorrer do armazenamento ocorre de forma semelhante para os três grupos e para as duas caseínas estudadas, a uma taxa estimada em 0,0314 mg/mL por dia de armazenamento. A Figura 1 apresenta o gráfico dos perfis individuais e perfil médio estimados para cada um dos grupos.

Com o modelo utilizado, tanto a covariância entre as duas variáveis respostas como a covariância entre medidas da mesma variável reposta e a heterocedasticidade foram modeladas marginalmente por meio dos efeitos aleatórios considerados.

A correlação marginal entre $alpha_{s1}$ -caseína e $beta$ -caseína, considerando $t = 0$, obtida por meio das estimativas dos componentes de variância e covariância é dada por:

$$r_M = \frac{\hat{\sigma}_{b_{01}b_{02}}}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_{b_{01}}^2} \sqrt{\hat{\sigma}_2^2 + \hat{\sigma}_{b_{02}}^2}}$$

$$r_M = \frac{\sqrt{0,5218}}{\sqrt{0,5091 + 0,5218} \sqrt{0,7387 + 0,5218}}$$

$$r_M = 0,6337$$

No modelo considerado, a covariância entre as variáveis respostas é constante, no entanto, a correlação entre elas diminui ao longo do tempo devido ao aumento da variância com o passar do tempo.

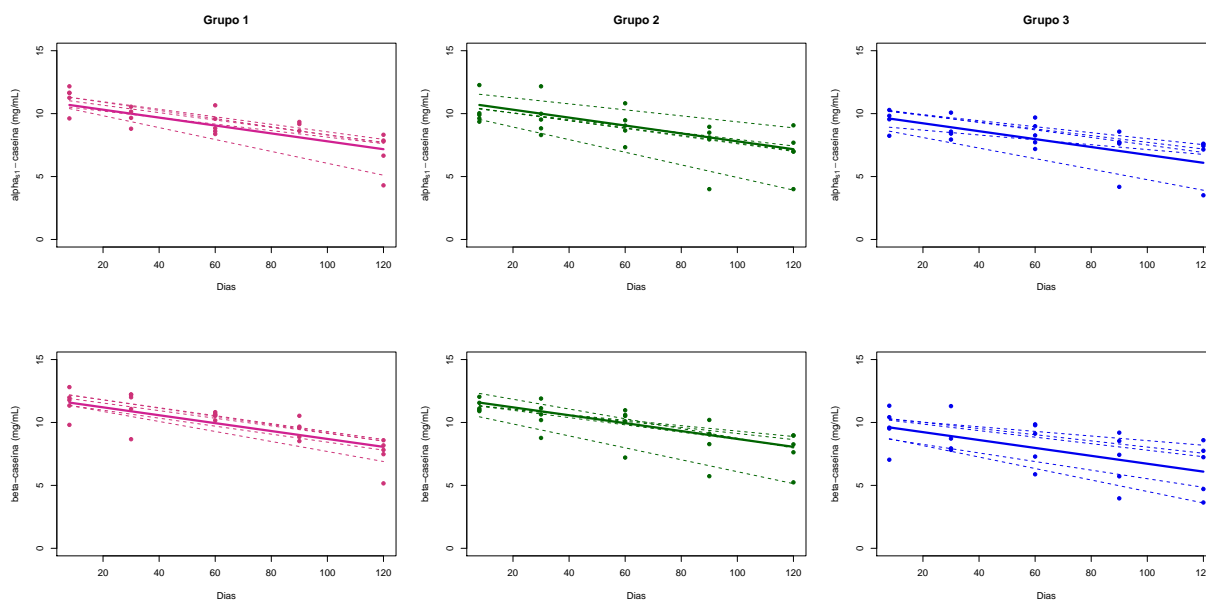


Figura 1: Gráfico de perfis estimados segundo o modelo selecionado para as variáveis α -caseína e β -caseína

A Figura 2 apresenta os gráficos dos resíduos do modelo selecionado. Pode-se visualizar apenas um ponto fora da elipse no *boxplot* bivariado, que não está fora do envelope simulado no gráfico quantil-quantil para a distância de Mahalanobis versus os quantis da quiquadrado com 2 graus de liberdade. Observa-se ainda que o ângulo entre os segmentos de reta no *boxplot* bivariado está próximo de 90° , indicando que a maior parte da correlação entre as variáveis foi explicada pelo modelo utilizado.

A realização de testes comparando os modelos bivariados para as taxas de redução das caseínas possibilitou verificar que as concentrações iniciais de α_{s1} e β -caseína foram diferentes apenas para os grupos com CCS baixa ou intermediária, sendo que as taxas de redução foram consideradas semelhantes nos três grupos.

4 Conclusões

A utilização de modelos mistos bivariados para dados longitudinais permite estudar a inclusão de componentes de variância e covariância, que servem para explicar a variabilidade das respostas ao longo do tempo e também a correlação entre as variáveis respostas. Além disso, permite a realização de testes estatísticos visando à escolha de parâmetros comuns às duas variáveis em estudo, o que é particularmente interessante quando as variáveis têm a mesma uni-

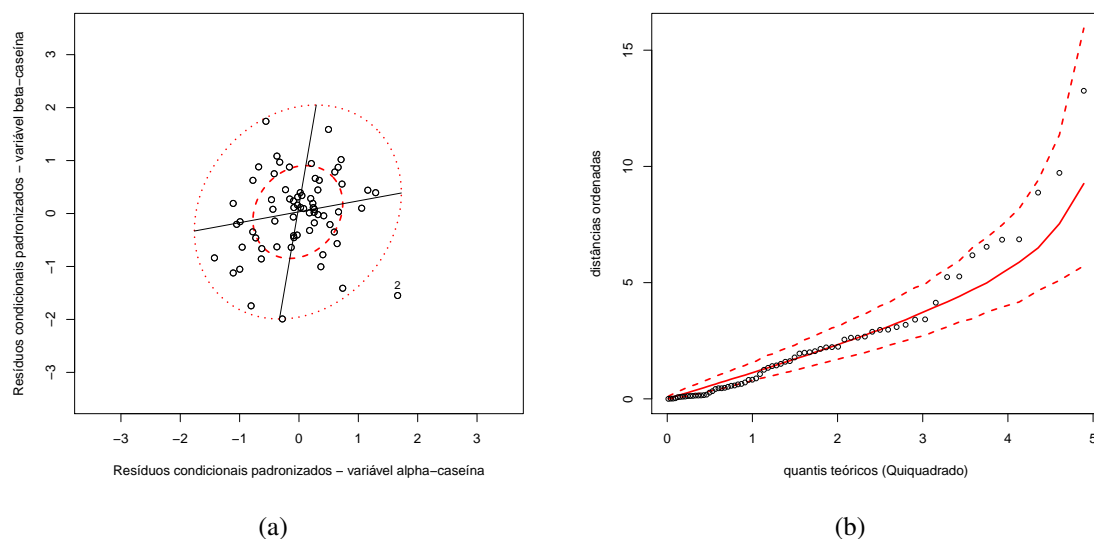


Figura 2: (a) Gráfico boxplot bivariado para os resíduos da variável $alpha_{s1}$ -caseína versus resíduos da variável $beta$ -caseína; (b) Gráfico quantil-quantil para a distância de mahalanobis versus os quantis da qui quadrado

dade de medida. Neste caso, as conclusões tiradas com o uso dos modelos mistos bivariados não seriam possíveis com a utilização de modelos mistos univariados.

Referências

- [1] FERNANDES, A.M. **Efeitos dos níveis de células somáticas sobre a qualidade do leite integral obtido por processo UAT direto**. 2007. 109 p. Tese (Doutorado em Zootecnia) - Faculdade de Zootecnia e Engenharia de Alimentos, Pirassununga, 2007.
- [2] FIEUWS, S. and VERBEKE, G. and MOLENBERGHS, G., Random-effects models for multivariate repeated measures. **Statistical methods in medical research**, n. 5, v. 16, 2007, p. 387-397
- [3] FIEUWS, S. and VERBEKE, G., Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. **Statistics in Medicine**, v. 23, 2004, p. 3093-3104
- [4] R Development Core Team, **R: A Language and Environment for Statistical Computing**, R Foundation for Statistical Computing. Vienna, Austria, 2012. <http://www.R-project.org/>
- [5] SAS Institute Inc. (2008), **SAS/STAT 9.2 Users Guide**, Cary, NC: SAS Institute Inc.
- [6] WEISS, R. **Modeling Longitudinal Data**, New York: Springer-Verlag, 2005.