

ANÁLISE ESTATÍSTICA DE DADOS COMPOSICIONAIS LONGITUDINAIS

Taciana Kisasi Oliveira SHIMIZU¹
Jorge Alberto ACHCAR²
Mário Hissamitsu TARUMOTO¹

- RESUMO: Neste artigo, introduzimos uma análise Bayesiana para dados composicionais considerando a transformação logaritmo da razão aditiva (ALR) e Box-Cox aplicadas a dados de um experimento de medidas repetidas. Tratando-se de dados pareados, foram introduzidos efeitos aleatórios para capturar a dependência entre os dados longitudinais. Além disso, comparamos os resultados obtidos através da estimação Bayesiana com a proposta de Hijazi e Jernigan (2009), que analisaram dados composicionais utilizando a transformação ALR e o modelo de regressão Dirichlet sob o enfoque clássico. A metodologia proposta foi ilustrada considerando um conjunto de dados reais.
- PALAVRAS-CHAVE: Dados composicionais; inferência Bayesiana; modelo de regressão Dirichlet clássico; erros não correlacionados.

1 Introdução

A teoria de dados composicionais desenvolveu-se na década de 80 com as contribuições de Aitchison e Shen (1980) e Aitchison (1982). Desde então, há aplicações em várias áreas do conhecimento, como por exemplo em geologia, ciências do solo, medicina, entre outras.

Dados composicionais são definidos como um vetor \boldsymbol{x} com elementos positivos x_1, \dots, x_G que representam proporções de um “todo”, tais que a soma de todos os elementos do vetor é igual a um, ou seja, $x_1 + \dots + x_G = 1$, em que G representa o número total de componentes dos dados composicionais (Aitchison, 1986).

¹Universidade Estadual Paulista – UNESP, Faculdade de Ciências e Tecnologia, CEP: 19060-900, Presidente Prudente, SP, Brasil. E-mail: taci_kisaki@yahoo.com.br / tarumoto@fct.unesp.br

²Universidade de São Paulo – USP, Faculdade de Medicina de Ribeirão Preto - FMRP, Departamento de Medicina Social, CEP: 14049-900, Ribeirão Preto, SP, Brasil. E-mail: achcar@fmrp.usp.br

A modelagem estatística para dados composicionais torna-se complicada pelo fato da ausência de independência e de classes paramétricas no Simplex, espaço amostral definido para dados composicionais. Nesse contexto, um modelo natural introduzido para a análise desse tipo de dados é dado pela distribuição Dirichlet. Apesar disso, o uso da distribuição Dirichlet tem uma limitação na análise de dados composicionais pois sua estrutura de correlação é totalmente negativa, tornando-a inadequada para o padrão de dados composicionais, que pode possuir correlações positivas.

Aitchison (1982) propôs uma alternativa às restrições existentes na análises de dados composicionais através de transformações que levam do espaço amostral restrito Simplex para o espaço bem definido dos reais.

Mais especificamente, em Aitchison e Shen (1980) desenvolveu-se a classe de distribuições Normal-logística que utiliza a transformação do espaço amostral Simplex para o espaço dos reais, denominando-a como transformação logaritmo da razão aditiva (ALR). Rayens e Srinivasan (1991) consideraram a transformação Box-Cox como uma generalização da transformação ALR. Alternativamente, Campbell e Mosimann (1987) propuseram o modelo de regressão Dirichlet para a modelagem de dados composicionais.

Neste estudo, foi realizado um estudo comparativo entre as transformações ALR e Box-Cox sob o enfoque Bayesiano de forma similar ao estudo realizado por Hijazi e Jernigan (2009), que realizaram uma comparação entre o modelo de regressão Dirichlet com modelos usando a transformação ALR utilizando métodos clássicos para a inferência.

Como motivação e ilustração para o presente estudo, consideramos um conjunto de dados composicionais referentes a um experimento de medidas repetidas, em que os componentes da composição do leite de trinta vacas foram avaliados antes e depois de uma dieta controlada em um período de oito semanas. Esses dados são apresentados em Aitchison (2005) e estão dispostos na Tabela 1.

2 Análise Bayesiana para dados composicionais longitudinais

Nesta seção, assumimos modelos composicionais longitudinais para a análise dos dados da Tabela 1. Mais especificamente, aplicamos as transformações ALR e Box-Cox nos dados composicionais e os métodos Bayesianos são utilizados para a estimação dos parâmetros do modelo ajustado.

O uso de métodos Bayesianos é uma boa alternativa para a análise de dados composicionais (ver por exemplo, Iyengar e Dey, 1996, 1998)), especialmente considerando métodos de Monte Carlo via cadeias de Markov (MCMC) (ver por exemplo, Gelfand e Smith, 1990 ou Smith e Roberts, 1993).

Tabela 1 - Composições do leite de 30 vacas antes e depois da nova dieta

Antes						Depois					
pr	gl	c	Ca	S	p	pr	gl	c	Ca	S	p
0,1389	0,2278	0,1553	0,1699	0,1400	0,1680	0,1753	0,1459	0,1552	0,2122	0,1642	0,1473
0,1377	0,1661	0,1540	0,2066	0,1549	0,1807	0,2090	0,0937	0,1313	0,2341	0,1717	0,1603
0,1464	0,1525	0,1672	0,1976	0,1572	0,1792	0,2387	0,1207	0,1497	0,1832	0,1652	0,1426
0,1950	0,1564	0,1562	0,1993	0,1163	0,1768	0,2398	0,1345	0,1726	0,2310	0,0896	0,1326
0,1988	0,1423	0,1507	0,1869	0,1470	0,1742	0,1173	0,1647	0,1535	0,2482	0,1577	0,1586
0,1498	0,1979	0,1858	0,1782	0,1237	0,1645	0,1701	0,1063	0,1524	0,2508	0,1826	0,1379
0,1467	0,1552	0,1828	0,1778	0,1564	0,1812	0,2018	0,1109	0,1166	0,2800	0,1471	0,1436
0,1109	0,2690	0,1466	0,2046	0,0996	0,1693	0,2142	0,0944	0,1472	0,2488	0,1536	0,1418
0,1198	0,2005	0,1351	0,1984	0,1654	0,1807	0,1890	0,1622	0,2066	0,2182	0,0965	0,1274
0,2164	0,1624	0,1687	0,1818	0,1111	0,1597	0,2097	0,1431	0,1706	0,2082	0,1435	0,1249
0,1792	0,1585	0,1358	0,1907	0,1645	0,1713	0,1562	0,1611	0,1901	0,2452	0,1126	0,1349
0,1650	0,1836	0,1401	0,1896	0,1477	0,1740	0,1292	0,2046	0,1977	0,2104	0,1313	0,1268
0,1744	0,1999	0,1742	0,1608	0,1310	0,1597	0,2538	0,1314	0,1499	0,1665	0,1655	0,1328
0,1319	0,1689	0,1338	0,2055	0,1713	0,1886	0,1959	0,1289	0,1612	0,2324	0,1370	0,1448
0,1482	0,2426	0,1544	0,1524	0,1476	0,1549	0,2154	0,1707	0,1713	0,2456	0,0875	0,1095
0,1857	0,1891	0,1810	0,1829	0,1042	0,1571	0,1748	0,1715	0,1458	0,1959	0,1795	0,1326
0,1497	0,1552	0,1419	0,2027	0,1623	0,1883	0,1446	0,1634	0,1757	0,2265	0,1375	0,1523
0,1518	0,1703	0,1412	0,1656	0,2064	0,1646	0,1690	0,1918	0,1625	0,2510	0,0992	0,1264
0,1582	0,1437	0,1682	0,1852	0,1733	0,1713	0,1791	0,1607	0,1792	0,1990	0,1221	0,1599
0,1683	0,1832	0,1618	0,1659	0,1643	0,1565	0,2149	0,1210	0,1446	0,2589	0,1090	0,1516
0,1394	0,2128	0,1999	0,1619	0,1336	0,1523	0,1799	0,1545	0,1605	0,2272	0,1408	0,1371
0,1687	0,1570	0,1399	0,1883	0,1632	0,1829	0,1723	0,1566	0,1638	0,2460	0,1299	0,1314
0,1988	0,1436	0,1529	0,1809	0,1537	0,1700	0,1778	0,1285	0,1905	0,2468	0,1161	0,1403
0,1870	0,1770	0,1561	0,1754	0,1424	0,1622	0,2045	0,1670	0,1612	0,2124	0,1248	0,1301
0,1243	0,2008	0,1520	0,2043	0,1346	0,1840	0,2063	0,1206	0,1428	0,2287	0,1461	0,1555
0,1686	0,2286	0,1465	0,1641	0,1373	0,1549	0,2709	0,1018	0,1207	0,2491	0,1226	0,1349
0,1512	0,1692	0,1658	0,1865	0,1508	0,1766	0,2099	0,1188	0,1450	0,2617	0,1047	0,1598
0,2033	0,2042	0,1676	0,1460	0,1344	0,1445	0,2046	0,1370	0,1325	0,2779	0,1111	0,1369
0,1455	0,1817	0,1783	0,1519	0,1798	0,1628	0,2808	0,1252	0,1390	0,1813	0,1328	0,1408
0,1451	0,2350	0,1886	0,1696	0,1195	0,1422	0,1245	0,1871	0,1554	0,2084	0,1573	0,1672

*pr=proteína, gl=gordura do leite, c=carboidrato, Ca=cálcio, S=sódio, p=potássio.

2.1 Transformação ALR considerando efeitos aleatórios ou variáveis latentes no modelo

Considerando os dados composicionais dos dois grupos de vacas (antes e depois de receber a dieta controlada) da Tabela 1 por $x_{i1} = pr$, $x_{i2} = gl$, $x_{i3} = c$, $x_{i4} = Ca$, $x_{i5} = S$, $x_{i6} = p$, $x_{i1} + x_{i2} + x_{i3} + x_{i4} + x_{i5} + x_{i6} = 1$, $i = 1, \dots, 60$, podemos assumir a transformação ALR dada por $y_{i1} = \log(x_{i1}/x_{i6})$, $y_{i2} = \log(x_{i2}/x_{i6})$, $y_{i3} = \log(x_{i3}/x_{i6})$, $y_{i4} = \log(x_{i4}/x_{i6})$ e $y_{i5} = \log(x_{i5}/x_{i6})$.

Para a modelagem dos dados composicionais da Tabela retdados com a transformação ALR dada por y_{ij} , assumimos o modelo de regressão (ver por exemplo, Iyengar e Dey (1996)) dado por

$$\mathbf{y}_i = \beta_0 + \beta_1 z_i + w_i + \epsilon_i, \quad (1)$$

em que \mathbf{y}_i é o vetor de variáveis resposta, para $j = 1, \dots, 5$ e $i = 1, \dots, 60$ ($n = n_1 + n_2$, em que $n_1 = 30$ para o grupo 1 e $n_2 = 30$ para o grupo 2), z_i representa a covariável associada ao i -ésimo indivíduo (variável indicadora em que $z_i = 0$ para o grupo 1 e $z_i = 1$ para o grupo 2), w_i é um efeito aleatório (variável latente não observável) que captura a dependência entre as medidas repetidas para cada indivíduo assumindo $w_i \sim N(0, \sigma_w^2)$ e ϵ_i é o vetor de erros aleatórios independentes com distribuição Normal $N(0, \sigma_j^2)$.

Denominamos o modelo da transformação ALR (Bayesiano) como “modelo 1”.

Além disso, podemos obter os estimadores das proporções verdadeiras dos componentes do leite $\alpha_{i1}, \dots, \alpha_{i6}$, sendo $\alpha_{i1} + \dots + \alpha_{i6} = 1$, com $\alpha_{i1} > 0, \dots, \alpha_{i6} > 0$.

Partindo do modelo (1) e da transformação ALR, temos as seguintes relações

$$\begin{aligned} \alpha_{ij} &= \frac{e^{\beta_{0j} + \beta_{1j}z_i + w_i}}{(1 + e^{\beta_{01} + \beta_{11}z_i + w_i} + e^{\beta_{02} + \beta_{12}z_i + w_i} + \dots + e^{\beta_{05} + \beta_{15}z_i + w_i})}, \\ &\vdots \\ \alpha_{i6} &= \frac{1}{(1 + e^{\beta_{01} + \beta_{11}z_i + w_i} + e^{\beta_{02} + \beta_{12}z_i + w_i} + \dots + e^{\beta_{05} + \beta_{15}z_i + w_i})}, \end{aligned} \quad (2)$$

em que $i = 1, \dots, 60$ ($n = n_1 + n_2$, onde $n_1 = 30$ para o grupo 1 e $n_2 = 30$ para o grupo 2) e $j = 1, \dots, 5$.

A parametrização $\alpha_{i1} > 0, \dots, \alpha_{i6} > 0$ pode ser usada para obter inferências sobre as composições em cada valor da covariável (Achcar e Obage, 2005).

A função de verossimilhança para os parâmetros $\beta_0 = (\beta_{01}, \dots, \beta_{05})$, $\beta_1 = (\beta_{11}, \dots, \beta_{15})$, $\sigma^2 = (\sigma_1^2, \dots, \sigma_5^2)$ e σ_w^2 do modelo (1) é dada por

$$L(\beta_0, \beta_1, \sigma^2, \sigma_w^2) = \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2)^{-n/2} \exp\left(-\frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^{60} \epsilon_{ij}^2\right), \quad (3)$$

em que $\sum_{i=1}^{60} \epsilon_{ij}^2 = \sum_{i=1}^{60} (y_{ij} - (\beta_{0j} + \beta_{1j}z_i))^2$, $j = 1, \dots, 5$.

Para uma análise Bayesiana hierárquica, assumimos as seguintes distribuições *a priori* para $\beta_0 = (\beta_{01}, \dots, \beta_{05})$, $\beta_1 = (\beta_{11}, \dots, \beta_{15})$, $\sigma^2 = (\sigma_1^2, \dots, \sigma_5^2)$ para a estimação dos parâmetros do modelo (1),

$$\begin{aligned} \beta_{0j} &\sim N(a_{0j}, b_{0j}^2), \\ \beta_{1j} &\sim N(a_{1j}, b_{1j}^2), \\ \sigma_j^2 &\sim IG(c_j, d_j), \end{aligned} \quad (4)$$

em que $N(a, b^2)$ denota uma distribuição Normal com média a e variância b^2 , $IG(c, d)$ denota uma distribuição Gama inversa com média $d/(c-1)$ e variância $d^2/[(c-1)^2(c-2)]$, $c > 2$; e $a_{0j}, b_{0j}, a_{1j}, b_{1j}, c_j$ e d_j são hiperparâmetros conhecidos, para $j = 1, \dots, 5$.

Para o segundo estágio da análise Bayesiana hierárquica, assumimos a seguinte distribuição *a priori* para a variância σ_w^2 da variável latente w_i ,

$$\sigma_w^2 \sim IG(c_w, d_w), \quad (5)$$

em que c_w e d_w são hiperparâmetros conhecidos. Além disso, assumimos independência *a priori* entre os parâmetros.

Um segundo modelo pode ser generalizado através do modelo (1) com a introdução de dois efeitos aleatórios extras: v_{1i} que captura a dependência entre as

medidas repetidas para o grupo 1 (antes de receber a dieta) para $i = 1, \dots, 30$ e v_{2i} que captura a dependência entre as medidas repetidas para o grupo 2 (depois de receber a dieta) para $i = 1, \dots, 30$. Desta forma, teremos três efeitos aleatórios no modelo, que é dado por

$$\mathbf{y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 z_i + w_i + a_j v_{1i} + (1 - a_j) v_{2j} + \boldsymbol{\epsilon}_i, \quad (6)$$

em que $a_j = 1$ (Grupo 1) e $a_j = 0$ (Grupo 2) para $j = 1, \dots, 5$. Assumimos distribuições Normais $N(0, \sigma_{v1}^2)$ e $N(0, \sigma_{v2}^2)$, para v_{1i} e v_{2i} , respectivamente. Também assumimos as mesmas suposições consideradas para o modelo (1) e independência entre as quantidades aleatórias w_i , v_{1i} , v_{2i} e ϵ_{ji} , $i = 1, \dots, 60$; $j = 1, \dots, 5$. Denotamos o modelo definido em (6) como “modelo 2”.

De acordo com o resultado obtido em (2) dos estimadores das proporções verdadeiras dos componentes do leite, podemos estendê-lo para o modelo 2 que considera a transformação ALR com três efeitos aleatórios.

A função de verossimilhança para os parâmetros $\boldsymbol{\beta}_0 = (\beta_{01}, \dots, \beta_{05})$, $\boldsymbol{\beta}_1 = (\beta_{11}, \dots, \beta_{15})$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_5^2)$, σ_w^2 , σ_{v1}^2 e σ_{v2}^2 do modelo (6) é dada por

$$L(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\sigma}^2, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2) = \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)^{-n/2} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)} \sum_{i=1}^{30} \epsilon_{ij}^2 \right] \\ \times \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)^{-n/2} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)} \sum_{i=1}^{30} \epsilon_{ij}^2 \right], \quad (7)$$

em que $\sum_{i=1}^{30} \epsilon_{ij}^2 = \sum_{i=1}^{30} [y_{ij} - (\beta_{0j} + \beta_{1j} z_i)]^2$, $j = 1, \dots, 5$.

Consideramos as mesmas distribuições *a priori* dadas em (4) e (5), e para σ_{vl} assumimos a distribuição *a priori*

$$\sigma_{vl}^2 \sim IG(a_{vl}, b_{vl}), \quad (8)$$

em que a_{vl} e b_{vl} são hiperparâmetros conhecidos, $l = 1, 2$. Assumimos independência *a priori* entre os parâmetros.

Como vamos utilizar o *software* OpenBugs na geração das quantidades aleatórias de interesse, não apresentamos as distribuições a posteriori condicionais requeridas para o algoritmo Gibbs sampling ou Metropolis-Hastings, pois esse *software* só requer a especificação da distribuição dos dados e as distribuições a priori para as quantidades aleatórias de interesse.

2.2 Transformação Box-Cox considerando um efeito e três efeitos aleatórios no modelo

Considerando a transformação Box-Cox para os dados composicionais introduzidos na Tabela 1, obtemos

$$y_{i1}^{(\lambda_1)} = \begin{cases} \frac{(\frac{x_{i1}}{x_{i6}})^{\lambda_1 - 1}}{\lambda_1} & \text{se } \lambda_1 \neq 0, \\ \log\left(\frac{x_{i1}}{x_{i6}}\right) & \text{se } \lambda_1 = 0 \end{cases}, y_{i2}^{(\lambda_2)} = \begin{cases} \frac{(\frac{x_{i2}}{x_{i6}})^{\lambda_2 - 1}}{\lambda_2} & \text{se } \lambda_2 \neq 0, \\ \log\left(\frac{x_{i2}}{x_{i6}}\right) & \text{se } \lambda_2 = 0 \end{cases}$$

$$y_{i3}^{(\lambda_3)} = \begin{cases} \frac{(\frac{x_{i3}}{x_{i6}})^{\lambda_3 - 1}}{\lambda_3} & \text{se } \lambda_3 \neq 0, \\ \log\left(\frac{x_{i3}}{x_{i6}}\right) & \text{se } \lambda_3 = 0 \end{cases}, y_{i4}^{(\lambda_4)} = \begin{cases} \frac{(\frac{x_{i4}}{x_{i6}})^{\lambda_4 - 1}}{\lambda_4} & \text{se } \lambda_4 \neq 0, \\ \log\left(\frac{x_{i4}}{x_{i6}}\right) & \text{se } \lambda_4 = 0 \end{cases}$$

$$y_{i5}^{(\lambda_5)} = \begin{cases} \frac{(\frac{x_{i5}}{x_{i6}})^{\lambda_5 - 1}}{\lambda_5} & \text{se } \lambda_5 \neq 0, \\ \log\left(\frac{x_{i5}}{x_{i6}}\right) & \text{se } \lambda_5 = 0 \end{cases}$$

A transformação Box-Cox é um caso especial da transformação ALR quando $\lambda_j = 0$ para $j = 1, \dots, 5$.

Utilizando os dados transformados $y_{i1}^{(\lambda_1)}, \dots, y_{i5}^{(\lambda_5)}$, $i = 1, \dots, 60$, e assumindo o modelo (1), o modelo 3 é dado por

$$\mathbf{y}_i^{(\boldsymbol{\lambda})} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 z_i + w_i + \boldsymbol{\epsilon}_i, \quad (9)$$

em que \mathbf{y}_i é o vetor de variáveis resposta, para $j = 1, \dots, 5$ e $i = 1, \dots, 60$ ($n = n_1 + n_2$, em que $n_1 = 30$ para o grupo 1 e $n_2 = 30$ para o grupo 2), z_i representa a covariável associada ao i -ésimo indivíduo (variável indicadora em que $z_i = 0$ para o grupo 1 e $z_i = 1$ para o grupo 2), w_i é um efeito aleatório (variável latente não observável) que captura a dependência entre as medidas repetidas para cada indivíduo e $\boldsymbol{\epsilon}_i$ é o vetor de erros aleatórios independentes com distribuição Normal $N(0, \sigma_j^2)$.

Podemos observar que os estimadores das proporções verdadeiras dos componentes do leite considerando o modelo (9) e a transformação Box-Cox são dados por

$$\alpha_{ij} = \frac{[\lambda_j(\beta_{0j} + \beta_{1j}z_i + w_i) + 1]^{\frac{1}{\lambda_j}}}{\left\{1 + [\lambda_1(\beta_{01} + \beta_{11}z_i + w_i) + 1]^{\frac{1}{\lambda_1}} + \dots + [\lambda_5(\beta_{05} + \beta_{15}z_i + w_i) + 1]^{\frac{1}{\lambda_5}}\right\}},$$

$$\vdots$$

$$\alpha_{i6} = \frac{1}{\left\{1 + [\lambda_1(\beta_{01} + \beta_{11}z_i + w_i) + 1]^{\frac{1}{\lambda_1}} + \dots + [\lambda_5(\beta_{05} + \beta_{15}z_i + w_i) + 1]^{\frac{1}{\lambda_5}}\right\}}, \quad (10)$$

em que $i = 1, \dots, 60$ ($n = n_1 + n_2$, em que $n_1 = 30$ para o grupo 1 e $n_2 = 30$ para o grupo 2) e $j = 1, \dots, 5$.

A função de verossimilhança para $\beta_0 = (\beta_{01}, \dots, \beta_{05})$, $\beta_1 = (\beta_{11}, \dots, \beta_{15})$, $\sigma^2 = (\sigma_1^2, \dots, \sigma_5^2)$, σ_w^2 e $\lambda = (\lambda_1, \dots, \lambda_5)$ do modelo (9) é dada por

$$L(\beta_0, \beta_1, \sigma^2, \sigma_w^2, \lambda) = \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2)^{-n/2} \left(\prod_{i=1}^n y_{ij}^{\lambda_j - 1} \right) \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^{60} \epsilon_{ij}^2 \right], \quad (11)$$

em que $\left(\prod_{i=1}^n y_{i1}^{\lambda_1 - 1} \right), \dots, \left(\prod_{i=1}^n y_{i5}^{\lambda_5 - 1} \right)$ são os produtos do Jacobiano, e $\sum_{i=1}^{60} \epsilon_{ij}^2 = \sum_{i=1}^{60} \left[y_{ij}^{(\lambda_j)} - (\beta_{0j} + \beta_{1j} z_i) \right]^2$, $j = 1, \dots, 5$.

Para uma análise Bayesiana hierárquica com os dados transformados através da transformação Box-Cox, assumimos as mesmas distribuições *a priori* para $\beta_0 = (\beta_{01}, \dots, \beta_{05})$, $\beta_1 = (\beta_{11}, \dots, \beta_{15})$, $\sigma^2 = (\sigma_1^2, \dots, \sigma_5^2)$ e σ_w^2 dadas em (4) e (5) para a estimação dos parâmetros do modelo (1).

Para $\lambda = (\lambda_1, \dots, \lambda_5)$, assumimos a distribuição *a priori*,

$$\lambda_j \sim N(e_j, f_j^2),$$

em que $j = 1, \dots, 5$ e e_j e f_j são hiperparâmetros conhecidos.

Um quarto modelo pode ser generalizado através do modelo (9) com a introdução de dois efeitos aleatórios extras: v_{1i} que captura a dependência entre as medidas repetidas para o grupo 1 (antes de receber a dieta) para $i = 1, \dots, 30$ e v_{2i} que captura a dependência entre as medidas repetidas para o grupo 2 (depois de receber a dieta) para $i = 1, \dots, 30$. Desta forma, teremos três efeitos aleatórios no modelo 4, que é dado por

$$y_i^{(\lambda)} = \beta_0 + \beta_1 z_i + w_i + a_j v_{1i} + (1 - a_j) v_{2j} + \epsilon_i, \quad (12)$$

em que $a_j = 1$ (Grupo 1) e $a_j = 0$ (Grupo 2) para $j = 1, \dots, 5$. Assumimos a distribuição Normal $N(0, \sigma_{v1}^2)$ e $N(0, \sigma_{v2}^2)$, para v_{1i} e v_{2i} , respectivamente. Também assumimos as mesmas suposições consideradas para o modelo (9) e independência entre as quantidades aleatórias w_i , v_{1i} , v_{2i} e ϵ_{ji} , $i = 1, \dots, 60$; $j = 1, \dots, 5$.

Podemos generalizar através de (10) para a obtenção dos estimadores das proporções verdadeiras dos componentes do leite considerando o modelo (12) com três efeitos aleatórios e a transformação Box-Cox.

A função de verossimilhança para os parâmetros $\beta_0 = (\beta_{01}, \dots, \beta_{05})$, $\beta_1 = (\beta_{11}, \dots, \beta_{15})$, $\sigma^2 = (\sigma_1^2, \dots, \sigma_5^2)$, $\lambda = (\lambda_1, \dots, \lambda_5)$, σ_w^2 , σ_{v1}^2 e σ_{v2}^2 do modelo (12) é dada por

$$\begin{aligned}
L(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\sigma}^2, \boldsymbol{\lambda}, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2) &= \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)^{-n/2} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)} \sum_{i=1}^{30} \epsilon_{ij}^2 \right] \\
&\times \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)^{-n/2} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)} \sum_{i=1}^{30} \epsilon_{ij}^2 \right] \\
&\times \left(\prod_{i=1}^n y_{ij}^{\lambda_j - 1} \right), \tag{13}
\end{aligned}$$

em que $\sum_{i=1}^{30} \epsilon_{ij}^2 = \sum_{i=1}^{30} \left(y_{ij}^{(\lambda_j)} - (\beta_{0j} + \beta_{1j} z_i) \right)^2$.

Assumimos as mesmas distribuições *a priori* dadas em (4), (5) e (8) para os parâmetros do modelo (12).

3 Análise clássica para dados composicionais longitudinais

Nesta seção assumimos modelos composicionais longitudinais para os dados da Tabela 1 sob o enfoque clássico, conforme proposto em Hijazi e Jernigan (2009), em que compararam as análises realizadas assumindo um modelo de regressão Dirichlet e um modelo assumindo a transformação ALR.

3.1 Modelo de regressão Dirichlet

Campbell e Mosimann (1987) abordaram uma extensão da distribuição Dirichlet para uma classe de modelos de regressão Dirichlet.

Assim, podemos assumir $\boldsymbol{x} = (x_{i1}, \dots, x_{iG})$ um vetor positivo ($1 \times G$) com distribuição Dirichlet com parâmetros positivos $(\lambda_1, \dots, \lambda_G)$ em que sua função densidade de probabilidade é dada por

$$f(\boldsymbol{x}) = \left(\Gamma(\boldsymbol{\lambda}) / \prod_{j=1}^G \Gamma(\lambda_j) \right) \prod_{j=1}^G x_j^{\lambda_j - 1},$$

em que $\sum_{j=1}^G x_j = 1$ e $\boldsymbol{\lambda} = \sum_{j=1}^G \lambda_j$.

O modelo de regressão Dirichlet é facilmente obtido permitindo que os parâmetros de uma distribuição Dirichlet mudem com covariáveis associadas (Camargo, 2012). Para um dado vetor de covariável \boldsymbol{z}_i , $i = 1, \dots, n$, cada parâmetro λ_j pode ser escrito como uma combinação linear $\lambda(\boldsymbol{z}_i)$ da covariável \boldsymbol{z}_i , ou seja,

$$\lambda_j(\boldsymbol{z}_i) = \sum_i \sum_j z_{i,1} \beta_{1,j} = \boldsymbol{z}_i \boldsymbol{\beta}_j, \tag{14}$$

em que $i = 1, \dots, n$ e $j = 1, \dots, G$.

Consideramos a função de verossimilhança (3) do modelo 1 para a estimação dos parâmetros para a inferência clássica.

4 Análise dos dados composicionais da Tabela 1

4.1 Análise Bayesiana

Para uma análise Bayesiana dos dados composicionais da Tabela 1 para o modelo 1, assumimos distribuições *a priori* não informativas para $\beta_{0j}, \beta_{1j}, \sigma_j^2, \sigma_w^2$ com hiperparâmetros cujos valores foram $a_{0j} = 1, a_{1j} = 0, b_{0j} = b_{1j} = 1.000.000, c_j = d_j = a_w = b_w = 1.000$ e simulamos amostras da distribuição a posteriori conjunta de interesse usando o software OpenBugs. Dessa forma simulamos 205.000 iterações, nas quais foram descartadas as 5.000 iterações iniciais (“burn-in”). Após o período de “burn-in”, consideramos saltos de tamanho 20 para obtermos amostras não correlacionadas, na qual resultou uma amostra final de tamanho 10.000 para cada parâmetro. A convergência do algoritmo foi verificada pelo diagnóstico de Geweke (Geweke, 1992).

Nas Tabelas 2 e 3 apresentamos os sumários *a posteriori* dos parâmetros dos modelos (1) e (6) e os valores correspondentes da estatística de teste para o diagnóstico de convergência Geweke.

Tabela 2 - Resumos *a posteriori* e Estatística de Teste para o Diagnóstico de Geweke – Modelo 1

Parâmetro	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	z_G
β_{01}	-0,060	0,042	(-0,143; 0,022)	-0,374
β_{02}	0,079	0,042	(-0,003; 0,162)	-0,979
β_{03}	-0,058	0,031	(-0,119; 0,003)	-0,473
β_{04}	0,070	0,022	(0,027; 0,114)	0,593
β_{05}	-0,150	0,034	(-0,150; -0,082)	0,452
β_{11}	0,365	0,059	(0,247; 0,482)	0,111
β_{12}	-0,083	0,060	(-0,206; 0,035)	0,946
β_{13}	0,170	0,044	(0,082; 0,258)	-1,006
β_{14}	0,415	0,032	(0,353; 0,478)	0,405
β_{15}	0,090	0,048	(-0,003; 0,182)	-0,267
σ_1	0,050	0,010	(0,034; 0,073)	0,163
σ_2	0,052	0,011	(0,035; 0,077)	0,623
σ_3	0,027	0,006	(0,018; 0,041)	-0,760
σ_4	0,012	0,003	(0,008; 0,019)	0,426
σ_5	0,033	0,007	(0,022; 0,048)	-0,785
σ_w	0,002	0,001	(0,001; 0,006)	-0,156

Podemos observar na Tabela 2, que há evidências de diferenças significativas nos grupos antes e depois da dieta nos seguintes componentes do leite: proteína, carboidrato e cálcio, pois os intervalos de credibilidade para β_{11}, β_{13} e β_{14} não incluem o valor zero. Além disso, observa-se que os valores da estatística do teste z_G estão no intervalo (-1,96; 1,96) indicando convergência para todos os parâmetros do modelo (1).

Para o modelo 2, assumimos os mesmos valores dos hiperparâmetros para as distribuições *a priori* considerados no modelo 1, sendo que os valores dos hiperparâmetros dos dois efeitos aleatórios extras foram:

Tabela 3 - Resumos *a posteriori* e Estatística de Teste para o Diagnóstico de Geweke
– Modelo 2

Parâmetro	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	z_G
β_{01}	-0,060	0,041	(-0,143; 0,021)	1,929
β_{02}	0,079	0,042	(-0,003; 0,159)	1,715
β_{03}	-0,058	0,030	(-0,177; 0,002)	1,582
β_{04}	0,070	0,020	(0,030; 0,111)	1,323
β_{05}	-0,150	0,036	(-0,221; -0,080)	-0,438
β_{11}	0,366	0,060	(0,249; 0,486)	-0,771
β_{12}	-0,082	0,060	(-0,199; 0,036)	-0,263
β_{13}	0,170	0,045	(0,081; 0,259)	-1,481
β_{14}	0,416	0,032	(0,354; 0,479)	-0,543
β_{15}	0,091	0,053	(-0,015; 0,196)	1,737
σ_1	0,049	0,010	(0,033; 0,072)	0,630
σ_2	0,050	0,010	(0,034; 0,073)	-0,255
σ_3	0,025	0,005	(0,016; 0,037)	-0,812
σ_4	0,010	0,003	(0,005; 0,017)	-0,499
σ_5	0,036	0,008	(0,024; 0,053)	0,761
σ_w	0,001	0,0009	(0,0003; 0,004)	0,635
σ_{v1}	0,001	0,0009	(0,0003; 0,004)	1,884
σ_{v2}	0,006	0,004	(0,001; 0,016)	-0,143

$a_{v1} = b_{v1} = a_{v2} = b_{v2} = 1.000$, e também foram geradas 205.000 iterações em que descartamos as 5.000 iterações iniciais (“burn-in”).

Percebe-se que para os componentes do leite, proteína, carboidrato e cálcio há diferenças significativas nos grupos antes e depois da dieta, sendo que os intervalos de credibilidade dos parâmetros β_{11} , β_{13} e β_{14} não incluem o valor zero (Tabela 3). Além disso, mais uma vez houve indicação de convergência das cadeias dos parâmetros estimados do modelo (6).

Considerando o modelo 3, que utilizou a transformação Box-Cox e um efeito aleatório no modelo, assumimos as distribuições *a priori* não informativas para β_{0j} , β_{1j} , σ_j^2 , σ_w^2 e λ_j com hiperparâmetros cujos valores foram $a_{0j} = 1$; $b_{0j} = b_{1j} = f_j = 1.000.000$, $a_{1j} = e_j = 0$, $c_j = d_j = a_w = b_w = 1.000$, geramos no *software* OpenBugs 205.000 iterações, nas quais foram descartadas as 5.000 iterações iniciais (“burn-in”).

Nas Tabelas 4 e 5 apresentamos os sumários *a posteriori* dos parâmetros dos modelos (9) e (12) e os valores correspondentes da estatística de teste para o diagnóstico de convergência Geweke.

De acordo com os resultados da Tabela 4, observamos que as diferenças entre os dois grupos (antes e depois da dieta) dos componentes proteína, carboidrato e cálcio foram significativas, pois os intervalos de credibilidade para β_{11} , β_{13} e β_{14} não incluem o valor zero. Além disso, apenas o λ_4 apresentou significância, indicando que para a maioria dos λ 's a transformação ALR, que é um caso especial da transformação Box-Cox, seria mais adequada para os casos em que o intervalo de credibilidade incluem o valor zero ($\lambda = 0$).

Tabela 4 - Resumos *a posteriori* e Estatística de Teste para o Diagnóstico de Geweke
- Modelo 3

Parâmetro	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	z_G
β_{01}	-0,060	0,043	(-0,144; 0,026)	0,037
β_{02}	0,076	0,046	(-0,013; 0,166)	1,514
β_{03}	-0,069	0,032	(-0,132; -0,004)	1,636
β_{04}	0,064	0,013	(0,039; 0,092)	1,207
β_{05}	-0,156	0,038	(-0,234; -0,085)	-0,180
β_{11}	0,373	0,065	(0,249; 0,503)	0,860
β_{12}	-0,084	0,062	(-0,208; 0,036)	-1,067
β_{13}	0,166	0,044	(0,079; 0,252)	-1,868
β_{14}	0,274	0,040	(0,206; 0,365)	-1,346
β_{15}	0,092	0,049	(-0,004; 0,188)	0,308
σ_1	0,054	0,012	(0,035; 0,082)	-0,281
σ_2	0,055	0,011	(0,037; 0,081)	0,759
σ_3	0,028	0,006	(0,019; 0,041)	-0,441
σ_4	0,004	0,002	(0,002; 0,009)	-0,814
σ_5	0,035	0,009	(0,022; 0,057)	-0,399
σ_w	0,001	0,0007	(0,0003; 0,003)	-0,770
λ_1	0,075	0,381	(-0,681; 0,822)	-0,228
λ_2	-0,184	0,506	(-1,170; 0,829)	-0,676
λ_3	-0,723	0,610	(-1,940; 0,466)	0,635
λ_4	-1,523	0,483	(-2,446; -0,518)	-1,132
λ_5	-0,205	0,627	(-1,440; 1,012)	-0,119

Tabela 5 - Resumos *a posteriori* e Estatística de Teste para o Diagnóstico de Geweke
- Modelo 4

Parâmetro	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	z_G
β_{01}	-0,059	0,044	(-0,145; 0,029)	0,369
β_{02}	0,075	0,045	(-0,013; 0,165)	1,762
β_{03}	-0,067	0,032	(-0,129; -0,004)	-0,751
β_{04}	0,066	0,015	(0,037; 0,097)	0,166
β_{05}	-0,158	0,040	(-0,241; -0,083)	-0,986
β_{11}	0,371	0,066	(0,244; 0,502)	-1,649
β_{12}	-0,084	0,062	(-0,205; 0,038)	-0,776
β_{13}	0,166	0,045	(0,079; 0,254)	-0,638
β_{14}	0,304	0,047	(0,218; 0,406)	-0,097
β_{15}	0,091	0,053	(-0,012; 0,197)	-0,629
σ_1	0,053	0,012	(0,034; 0,081)	-0,240
σ_2	0,053	0,011	(0,036; 0,079)	1,346
σ_3	0,027	0,006	(0,018; 0,040)	0,812
σ_4	0,004	0,002	(0,001; 0,010)	0,764
σ_5	0,039	0,010	(0,023; 0,063)	-0,873
σ_w	0,0009	0,0006	(0,0002; 0,002)	-0,042
σ_{v1}	0,0009	0,0006	(0,0002; 0,002)	0,720
σ_{v2}	0,003	0,002	(0,0005; 0,009)	-0,125
λ_1	0,066	0,383	(-0,684; 0,826)	0,130
λ_2	-0,163	0,507	(-1,143; 0,839)	1,713
λ_3	-0,608	0,604	(-1,818; 0,549)	0,542
λ_4	-1,114	0,491	(-2,143; -0,170)	0,294
λ_5	-0,328	0,647	(-1,623; 0,920)	-0,209

Para o modelo 4, utilizamos os mesmos valores dos hiperparâmetros para as distribuições *a priori* não informativas considerados no modelo 2, que contém também três efeitos aleatórios. Sendo que, para os parâmetros λ , consideramos os valores dos hiperparâmetros aplicados no modelo 3.

Observa-se na Tabela 5, que os resultados são semelhantes aos da Tabela 4 em que utilizou apenas um efeito aleatório.

4.2 Análise clássica

Os dados composicionais da Tabela 1 foram analisados através de métodos clássicos aplicando a transformação ALR em que utilizou-se a rotina *PROC NLMIXED* do *software SAS* para a estimação dos parâmetros do modelo (1) com o objetivo de comparação entre os métodos de estimação Bayesiano e clássico.

As estimativas dos parâmetros e seus respectivos intervalos de confiança são apresentados na Tabela 6.

Tabela 6 - Estimativas dos parâmetros do modelo 1 - Método Clássico

Parâmetro	Estimativa	Desvio Padrão	Intervalo de Confiança (95%)
β_{01}	-0,026	0,017	(-0,062; 0,009)
β_{02}	0,034	0,018	(-0,002; 0,071)
β_{03}	-0,025	0,013	(-0,053; 0,002)
β_{04}	0,030	0,009	(0,012; 0,049)
β_{05}	-0,065	0,014	(-0,093; -0,037)
β_{11}	0,159	0,025	(0,109; 0,210)
β_{12}	-0,036	0,025	(-0,088; 0,016)
β_{13}	0,074	0,019	(0,035; 0,113)
β_{14}	0,180	0,013	(0,155; 0,206)
β_{15}	0,039	0,019	(0,0001; 0,078)
σ_1	0,095	0,009	(0,077; 0,114)
σ_2	0,098	0,009	(0,079; 0,118)
σ_3	0,073	0,007	(0,059; 0,087)
σ_4	0,049	0,005	(0,039; 0,058)
σ_5	0,074	0,007	(0,060; 0,088)
σ_w	0,012	0,008	(-0,004; 0,027)

Na Tabela 6, observa-se que as diferenças antes e depois da dieta para os componentes proteína, carboidrato e cálcio foram significativas; além disso, o componente sódio também mostrou diferença significativa após a dieta quando modelamos os dados através da inferência clássica. Percebe-se que o limite inferior do intervalo de confiança do σ_w é negativo, indicando problemas na estimação clássica para esse parâmetro.

O ajuste do modelo de regressão Dirichlet foi feito sob o enfoque clássico, em que as estimativas de máxima verossimilhança para os parâmetros do modelo

proposto (14) foram obtidas usando o *software* R, através do pacote *DirichletReg*. Os resultados da estimação dos parâmetros do modelo estão dispostos na Tabela 7.

Tabela 7 - Estimativas dos parâmetros do modelo de Regressão Dirichlet

Parâmetro	Estimativa	Desvio Padrão	Intervalo de Confiança (95%)
β_{01}	3,818	0,118	(3,587; 4,048)
β_{02}	3,957	0,117	(3,727; 4,187)
β_{03}	3,821	0,118	(3,590; 4,051)
β_{04}	3,948	0,117	(3,718; 4,178)
β_{05}	3,73	0,118	(3,499; 3,960)
β_{06}	3,878	0,117	(3,648; 4,108)
β_{11}	-0,225	0,167	(-0,552; 0,101)
β_{12}	-0,668	0,167	(-0,996; -0,341)
β_{13}	-0,419	0,167	(-0,746; -0,091)
β_{14}	-0,177	0,166	(-0,503; 0,149)
β_{15}	-0,497	0,167	(-0,825; -0,168)
β_{16}	-0,586	0,167	(-0,914; -0,259)

Na Tabela 7, podemos verificar que para os componentes gordura do leite, carboidrato, sódio e potássio, os coeficientes de regressão foram significativos, ou seja, houve diferença nesses componentes do leite após a aplicação da dieta. Podemos notar que esses coeficientes apresentaram valores negativos, indicando que a proporção para esse componentes no leite diminuiu após a dieta.

Para a comparação dos modelos ajustados aos dados composicionais da Tabela 1, calculou-se a soma do quadrado das diferenças entre os valores observados e os valores estimados das médias dos modelos(ver Tabela 8).

Tabela 8 - Soma do quadrado das diferenças entre os valores observados e os valores ajustados.

Modelo	$\sum_{i=1}^n \sum_{j=1}^6 (y_{ij} - \hat{y}_{ij})^2$
Modelo 1 (transformação ALR com 1 efeito aleatório - Bayesiano)	0,220812
Modelo 2 (transformação ALR com 3 efeitos aleatórios - Bayesiano)	0,220705
Modelo 3 (transformação Box-Cox com 1 efeito aleatório - Bayesiano)	0,220743
Modelo 4 (transformação Box-Cox com 3 efeitos aleatórios - Bayesiano)	0,220877
Modelo 5 (transformação ALR com 1 efeito aleatório - Análise Clássica)	0,300242
Modelo 6 (Regressão Dirichlet clássico)	0,220714

Analisando os resultados obtidos, percebe-se que há pouca diferença entre os valores da Tabela 8. Por outro lado, podemos afirmar que dentre os modelos ajustados, o modelo 2 (ALR com 3 efeitos aleatórios) se ajusta melhor aos dados da Tabela 1 (diferença mínima em relação aos outros modelos). É importante salientar que os menores valores da Tabela 8 foram obtidos para os modelos em que os parâmetros foram estimados através da inferência Bayesiana.

As Figuras 1, 2 e 3 apresentam os valores ajustados e observados dos componentes do leite (proteína, gordura do leite, carboidrato, cálcio, sódio e potássio) para cada modelo proposto. Podemos observar que os valores ajustados através da inferência clássica são os mais distantes em relação à média observada.

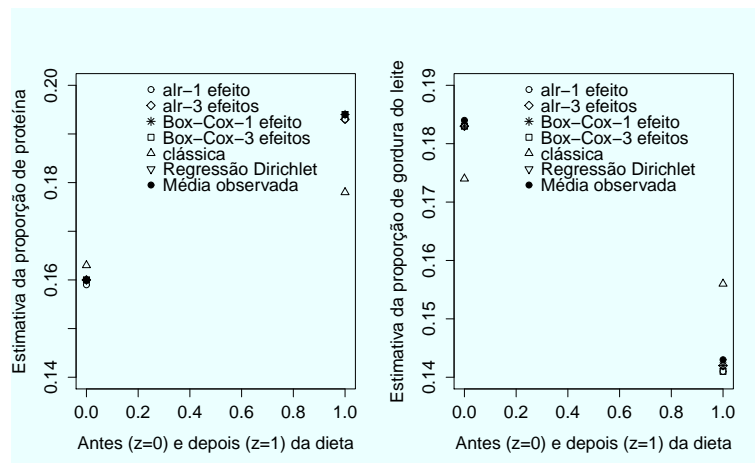


Figura 1 - Gráficos dos valores observados e ajustados dos componentes do leite: proteína e gordura do leite de acordo com o modelo.

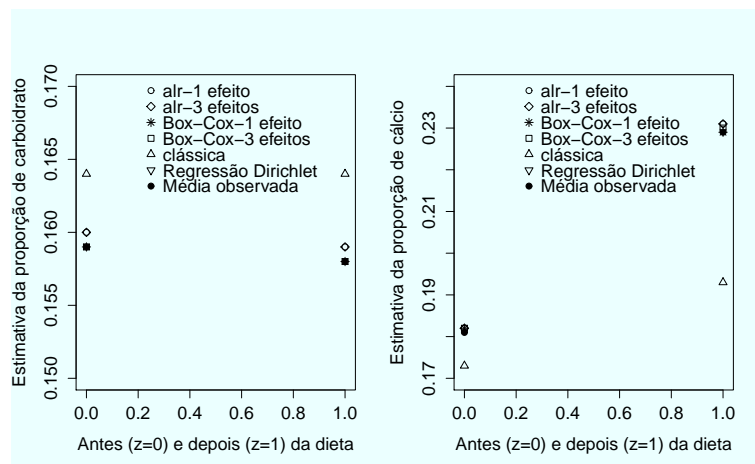


Figura 2 - Gráficos dos valores observados e ajustados dos componentes do leite: carboidrato e cálcio de acordo com o modelo.

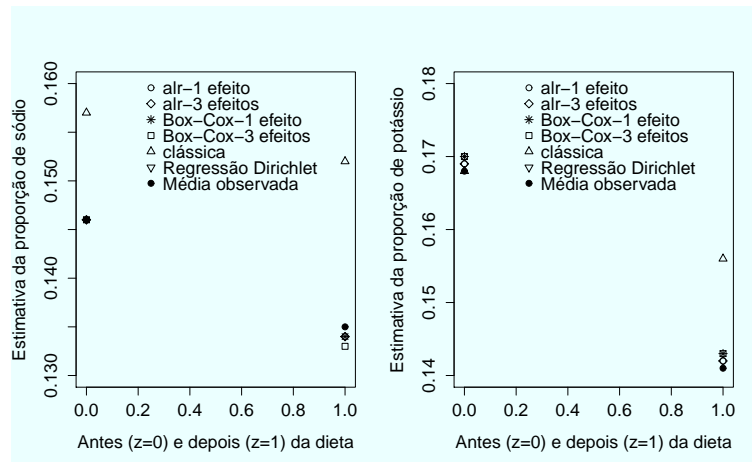


Figura 3 - Gráficos dos valores observados e ajustados dos componentes do leite: sódio e potássio de acordo com o modelo.

Conclusões

Neste artigo, apresentamos um estudo de dados composicionais longitudinais aplicados em modelos de regressão sob o enfoque Bayesiano considerando as transformações ALR e Box-Cox, e também usando o método clássico (proposto em Hijazi e Jernigan, 2009). Observar que a restrição dos dados composicionais, em que a soma das proporções ou frações é igual a um, não foi ignorada. A importância prática da metodologia proposta para dados composicionais foi ilustrada através de um conjunto de dados. Como assumimos um experimento de medidas repetidas, foram introduzidos efeitos aleatórios nos modelos para capturar a dependência dos dados longitudinais. Os resultados obtidos através da estimação Bayesiana, mostraram-se satisfatórios quanto à convergência das cadeias geradas na simulação e suas estimativas. Por outro lado, a estimação clássica (modelo 5) apresentou algum problema na estimação da variância do efeito aleatório (σ_w) do modelo.

Agradecimentos

Os autores agradecem à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo apoio financeiro.

SHIMIZU, T.K.O.; ACHCAR, J.A.; TARUMOTO, M.H. Statistical Analysis of longitudinal compositional data. *Rev. Bras. Biom.*, São Paulo, v.32, n.1, p.42-58, 2014.

- **ABSTRACT:** In this paper, we introduce a Bayesian analysis for compositional data considering additive log-ratio (ALR) and Box-Cox transformations applied in an experiment of repeated measures. In the case of paired data, random effects were introduced to capture the dependence between for the longitudinal data. Furthermore, we compare the obtained results under a Bayesian approach with the proposal of Hijazi and Jernigan (2009), who analyzed compositional data using ALR transformation and Dirichlet regression model under the classical approach. The proposed methodology was illustrated considering a real data set.
- **KEYWORDS:** Compositional data; Bayesian inference; Dirichlet regression model; uncorrelated errors.

Referências

- ACHCAR, J. A.; OBAGE, S. C. Uma abordagem Bayesiana para dados composicionais considerando erros correlacionados. *Revista de Matemática e Estatística*. v. 23, n. 2, p. 95-107, 2005.
- AITCHISON, J. The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 44, n. 2, p. 139-177, 1982.
- AITCHISON, J. *The statistical analysis of compositional data*. Chapman & Hall, 1986.
- AITCHISON, J. *A Concise Guide to Compositional Data Analysis*. CDA Workshop, Girona, 2005.
- AITCHISON, J.; SHEN, S. M. Logistic-normal distributions: Some properties and uses. *Biometrika*. v. 67, n. 2, p. 261-272, 1980.
- BOX, G. E. P.; COX, D. R. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 26, n. 2, p. 211-252, 1964.
- CAMARGO, A. P.; STERN, J. M.; LAURETTO, M. S. Estimation and Model Selection in Dirichlet Regression. In: *AIP Conference Proceedings*, v. 1443, p. 206-213, 2012.
- CAMPBELL, G.; MOSIMANN, J. Multivariate methods for proportional shape. *ASA Proceedings of the Section on Statistical Graphics*, p. 10-17, 1987.
- GELFAND, A. E.; SMITH, A. F. M. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, v. 85, n. 410, p. 398-409, 1990.
- GEWEKE, J. *Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments*. In Bayesian Statistics 4, Oxford University Press, 1992.
- HIJAZI, R.; JERNIGAN, R. W. Modelling compositional data using Dirichlet regression models. *Journal of Applied Probability & Statistics*, v. 4, n. 1, p. 77-91, 2009.
- IYENGAR, M.; DEY, D. K. *Bayesian analysis of compositional data*. Department of Statistics, University of Connecticut, Storrs, CT 06269-3120, 1996.

IYENGAR, M.; DEY, D. K. *Box-Cox transformations in Bayesian analysis of compositional data. Environmetrics*, v. 9, n. 6, p. 657-671, 1998.

RAYENS, W. S.; SRINIVASAN, C. Box-Cox transformations in the analysis of compositional data. *Journal of Chemometrics*, v. 5, n. 3, p. 227-239, 1991.

SMITH, A.F.M.; ROBERTS, G.O. Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society, B*, v. 55, p. 3-23, 1993.

Recebido em 10.03.2014.

Aprovado após revisão em 11.04.2014.