

Modelo de Regressão Simplex Aplicado a Dados Longitudinais de Consumo de Energia Metabolizável com Sobredispersão

Lucimary Afonso dos Santos ¹, Vanderly Janeiro ², Ricardo Souza Vasconcellos ³, Terezinha Aparecida Guedes⁴

1 Introdução

A classe de modelos de dispersão (MD) abrange um grande número de distribuições não normais como, por exemplo, a distribuição simplex, que é adequada para a análise de dados contínuos pertencentes ao intervalo (0,1). A distribuição simplex, proposta por Barndorff-Nielsen e Jørgensen (1991), foi utilizada por diversos autores em situações variadas. Song e Tan (2000) a utilizaram para analisar dados longitudinais na forma de proporção, sob a suposição de homogeneidade do parâmetro de dispersão. Song, Qiu e Tan (2004), modificaram essa abordagem, considerando o mesmo tipo de dados, porém, considerando heterogeneidade do parâmetro de dispersão. Fizeram o uso de um estudo de simulação, para demonstrar que a inferência fica comprometida pela suposição de homogeneidade do parâmetro de dispersão. Lopez (2013) apresenta técnicas Bayesianas para estimação dos parâmetros do modelo de regressão simplex considerando variância constante e heterogeneidade de variâncias.

Neste trabalho, procurou-se aplicar a metodologia a dados de alimentação de felinos avaliados ao longo do tempo com objetivo de determinar o consumo de energia metabolizável (ME) por eles após um regime de perda de peso, considerando sobredispersão dos dados.

2 Material e métodos

2.1 Material

O presente estudo refere-se a um conjunto de dados oriundos de um experimento para avaliar o efeito da perda de peso de dois grupos de gatos alimentados com diferentes níveis de proteína na dieta sobre o consumo de energia metabolizável (consumo relativo à quantidade preconizada para a espécie, em porcentagem) após o término do regime, na fase de manutenção do peso.

Para isso, um grupo de gatos, composto por 8 machos e 8 fêmeas foram realocados em dois grupos contendo ambos os gêneros. Um grupo foi submetido a uma dieta seca com alto teor de proteína (42% de proteína bruta, HP) e o outro grupo, controle, com uma dieta com menor teor de proteína (29% de proteína bruta, CO).

Cada gato, ao perder 20% de seu peso corporal inicial, foi administrado um alimento seco comercial de baixa energia para gatos adultos, em quantidade suficiente para manter seu peso constante. Todos os gatos de ambos os grupos experimentais foram tratados com o mesmo alimento durante a fase experimental de manutenção, que durou um período de 120 dias.

¹UNESPAR. e-mail: *lucimaryafonso@hotmail.com*

²DES/UEM. e-mail: *vjaneiro@uem.br*

³DZO/UEM. e-mail: *rsvasconcellos@uem.br*

⁴DES/UEM. e-mail: *taguedes@uem.br*

O período de manutenção foi subdividido em três fases: fase um (0 a 40 dias), fase dois (41 a 80 dias) e fase três (81 a 120 dias).

Durante as fases os gatos foram pesados uma vez por semana, sempre pela manhã antes da primeira refeição e a quantidade de alimento foi ajustada caso tenha ocorrido uma variação de peso corporal maior que 0,2% em relação ao peso inicial da fase.

Para determinar a ingestão calórica, a quantidade de alimento consumida foi multiplicada pela quantidade de ME da dieta. A variável resposta para o experimento foi a média, por fase, da quantidade de ME ($kcal \cdot kg^{-0,40}d^{-1}$) consumida por cada gato e dividida pela ingestão preconizada para a espécie ($130kcal \cdot kg^{-0,40}d^{-1}$) gerando, desta forma, um consumo relativo em porcentagem. Para maiores detalhes do experimento, consultar Vasconcellos *et al* (2009).

2.2 Métodos

2.2.1 Distribuição Simplex

A distribuição simplex, proposta por Barndorff-Nielsen e Jørgensen (1991) é utilizada para modelar dados de proporção contínuos. A distribuição simplex univariada com parâmetros $\mu \in (0, 1)$ e $\sigma^2 > 0$, denotada por $S^-(\mu, \sigma^2)$, tem a função densidade de probabilidade,

$$p(y; \mu, \sigma^2) = \{2\pi\sigma^2[y(1-y)]^3\}^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}d(y; \mu)\right\} \quad 0 < y < 1, \quad (1)$$

sendo $d(y; \mu) = \frac{(y-\mu)^2}{y(1-y)\mu^2(1-\mu)^2}$.

A função de variância, a esperança e a variância são respectivamente denotadas por $V(\mu) = \mu^3(1-\mu)^3$, $E(Y) = \mu$ e $\text{Var}(Y) = \mu(1-\mu) - \frac{1}{\sqrt{2\phi-1}} \exp\left(\frac{\phi}{\mu^2(1-\mu)^2}\right) \Gamma\left(\frac{1}{2}, \frac{\phi}{2\mu^2(1-\mu)^2}\right)$, sendo que $\Gamma(.,.)$ é a função gama incompleta medida nos pontos a e b e definida por $\Gamma(a, b) = \int_b^\infty y^{a-1}e^{-y}dy$, (JØRGENSEN, 1997).

2.2.2 Modelo Marginal Simplex para Análise de Dados Longitudinais

Seja y_{ij} , $j = 1, \dots, n_i$, uma sequência de medidas repetidas observadas no i -ésimo sujeito dentre m , t_{ij} , $j = 1, \dots, n_i$, os tempos correspondentes às medidas tomadas em cada sujeito, e x_{ijk} , $k = 1, \dots, p$, as p variáveis explicativas. Assumindo-se que y_{ij} são respostas para uma variável aleatória Y_{ij} que segue distribuição simplex $Y_{ij} \sim S^-(\mu_{ij}, \sigma_{ij}^2)$, com $\mu_{ij} \in (0, 1)$ parâmetros de média e $\sigma_{ij}^2 > 0$ parâmetros de dispersão.

Ainda, sejam $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$, $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})^T$. Considere $\mathbf{Y}_i, \dots, \mathbf{Y}_m$ independentes.

O modelo simplex marginal consiste de três componentes, sendo que a primeira componente é um modelo que descreve os efeitos da média populacional, μ_{ij} , que depende das covariáveis x_{ij} , variando no tempo por meio de um modelo linear generalizado da forma,

$$\eta_{ij} = g(\mu_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta} = \beta_0 + \beta_1 S_i + \beta_2 T_i + \beta_3 F_i, \quad (2)$$

sendo g uma função de ligação conhecida (no caso, função *logit*) e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ os coeficientes de regressão a serem estimados. As covariáveis sexo, tratamento e fase, denotadas, por S, T e F respectivamente. A segunda componente descreve o padrão do parâmetro de dispersão, σ_{ij}^2 , como função das covariáveis z_{ij} dado como,

$$h(\sigma_{ij}^2) = \mathbf{z}_{ij}^T \boldsymbol{\gamma} = \gamma_0 + \gamma_1 S_i + \gamma_2 T_i + \gamma_3 F_i, \quad (3)$$

com função de ligação, h , conhecida (no caso, função *logarítmica*) e $\boldsymbol{\gamma} = (\gamma_0, \dots, \gamma_{r-1})^T$, sendo γ_0 o intercepto.

A estrutura de correlação é modelada pela terceira componente. A correlação existente entre Y_{ij} e Y_{ik} é uma função dos parâmetros de locação e, talvez, de parâmetros adicionais $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_q)^T$, dada por:

$$COR(Y_{ij}, Y_{ik}) = \rho(\mu_{ij}, \mu_{ik}, \boldsymbol{\alpha}),$$

sendo $\rho(\cdot)$ uma função conhecida (ZHANG, QIU e SHI, 2016).

Vários tipos de estrutura de correlação podem ser utilizados para ρ .

3 Resultados

Observando-se as Figuras (1) e (2) percebe-se que tanto para sexo quanto para tratamento houve aumento na porcentagem de energia metabolizada (ME) para cada animal. Considerando-se o tratamento, constata-se que a quantidade de ME, para o grupo que recebeu a dieta com alto teor de proteínas foi superior ao grupo controle.

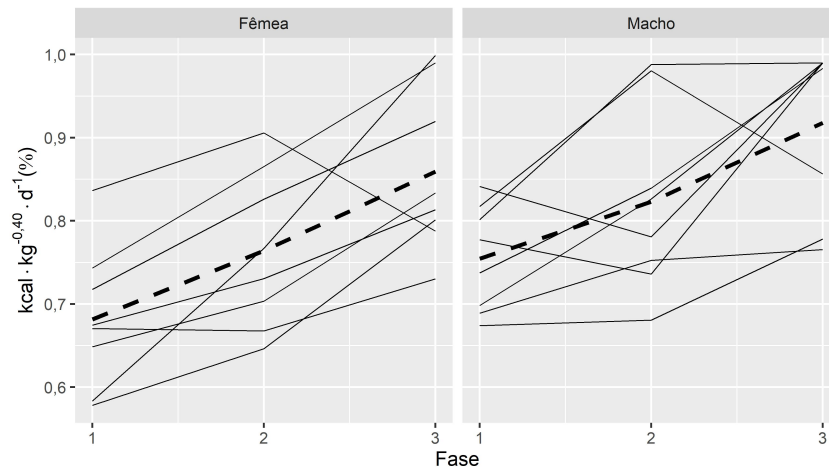


Figura 1: Gráfico de perfis da quantidade de energia metabolizada por sexo.

Fonte: O próprio autor.

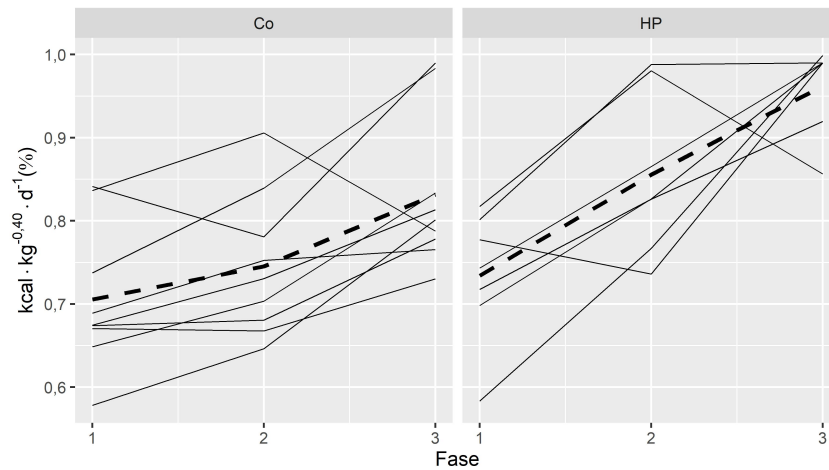


Figura 2: Gráfico de perfis da porcentagem de energia metabolizada por tratamento.

Fonte: O próprio autor.

Tabela 1: Correlação entre as fases

	Fase 1	Fase 2	Fase 3
Fase 1	1,0000	0,7073	0,2787
Fase 2	0,7073	1,0000	0,4317
Fase 3	0,2787	0,4317	1,0000

Fonte: O próprio autor

Observando-se a Tabela (1) percebe-se que a correlação existente entre as fases diminui ao longo do tempo. Tem-se ainda que a correlação entre fases próximas é maior que entre fases mais distantes sugerindo assim uma estrutura de correlação do tipo AR1.

Na Tabela (2) são apresentadas as estimativas dos parâmetros dos modelos (2) e (3) ajustados. Observa-se ao considerar o modelo completo para o ajuste, que a variável sexo não interfere no aumento percentual médio do consumo de energia (valor-p=0,1669) e também não contribui para a sobredispersão dos dados (valor-p=0,8873). Dessa forma, o fator sexo foi removido da primeira e da segunda componentes do modelo e procedeu-se um novo ajuste, o que não gerou grandes alterações às estimativas dos parâmetros mantidos no modelo. Para alguns casos houve um pequeno aumento na precisão da estimativa. Observa-se ainda que a correlação (ρ) entre as fases é significativa para os dois modelos ajustados.

Tabela 2: Resumo estatístico das estimativas do modelo

Parâmetro	Modelo completo			Modelo incompleto		
	Estimativa	Erro Padrão	valor-p	Estimativa	Erro Padrão	valor-p
β_0	0,2529	0,2135	0,2364	0,3667	0,1661	0,0272
β_1 (Macho)	0,2795	0,2022	0,1669	—	—	—
β_2 (HP)	0,2518	0,1484	0,0898	0,2791	0,1699	0,1005
β_3	0,4884	0,0789	<0,0000	0,4939	0,0744	<0,0000
γ_0	-2,7842	1,0176	0,0062	-2,5815	0,7967	0,0012
γ_1 (Macho)	0,0803	0,5667	0,8873	—	—	—
γ_2 (HP)	1,6059	0,5659	0,0045	1,6212	0,5346	0,0024
γ_3	2,0317	0,3589	<0,0000	1,9756	0,2896	<0,0000
α	-1,6197	0,4499	0,0003	-1,5036	0,3563	<0,0000
ρ	0,1979	0,0891	0,0262	0,2223	0,0792	0,0050

Fonte: o próprio autor

Nota: — indica que o parâmetro foi removido do modelo

De acordo com o gráfico apresentado na Figura 3 (a), a suposição de que os dados seguem distribuição simplex se verifica. Já na Figura 3(b) a maioria dos pontos está contida dentro das bandas de confiança levando-nos a concluir que a função de ligação *logit* é adequada aos dados.

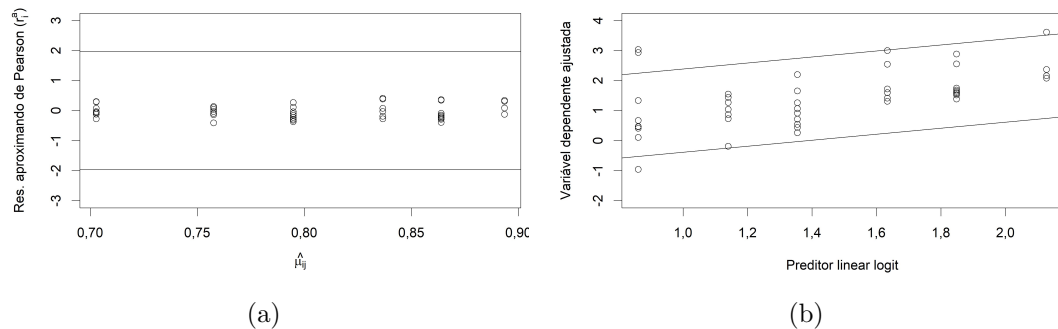


Figura 3: Gráficos de diagnóstico: (a) Verificação da suposição da distribuição; (b) Verificação da função de ligação.

Fonte: o próprio autor.

4 Considerações

De acordo com o modelo ajustado, o aumento percentual médio do consumo de energia se deve principalmente ao tempo/fases $\hat{\beta}_3 = 0,4939$ e os animais que receberam a dieta de ração seca com alto teor de proteína durante o período de emagrecimento contribuem 1,3 ($\exp\{0,2791\}$) vezes mais do que o grupo controle para esse aumento. Percebeu-se ainda que tanto a variável explicativa tratamento quanto o tempo de observação contribuem para ocorrência de sobredispersão dos dados.

Referências Bibliográficas

- BARNDORFF-NIELSEN, O.E.; JØRGENSEN, B. Some parametric models on the simplex. *Journal of Multivariate Analysis*, New York, v.39, n.1, p. 106-116, 1991.
- JØRGENSEN, B. *The theory of dispersion models*. London: Chapman & Hall, 1997.
- LOPEZ, F. O. A Bayesian Approach to Parameter Estimation in Simplex Regression Model: A Comparison with Beta Regression. *Revista Colombiana de Estadística*, v.13, n.1, p.1-21, 2013. Disponível em: <http://emis.impa.br/EMIS/journals/RCE/ingles/V36/v36n1a01.pdf>. Acesso em: 03 mar.2019.
- R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2017. ISBN 3-900051-07-0, <http://www.R-project.org/>.
- SONG, P.X.-K.; TAN, M. Marginal models for longitudinal continuous proportional data, *Biometrics*, Washington, v.56, n.2, p.496-502, 2000.
- SONG, P.X.-K.; QIU, Z.; TAN, M. Modelling heterogeneous dispersion in marginal models for longitudinal proportional data. *Biometrical Journal*, Berlin, v.46, p.540-553, 2004.
- VASCONCELLOS, R. S.; BORGES, N. C.; GONÇALVES, K. N. V.; CANOLA, J. C.; PAULA, F. J. A.; MALHEIROS, E. B.; BRUNETTO, M. A.; CARCIOFI, A. C. Protein Intake during Weight Loss Influences the Energy Required for Weight Loss and Maintenance in Cats. *The Journal of Nutrition*, v.139, p.540-553, 2009. Disponível em: <https://academic.oup.com/jn/article/139/5/855/4670357>. Acesso em: 10 mar. 2019.
- ZHANG, P.; QIU, Z.; SHI, C. Simplexreg: An R package for regression analysis of proportional data using the simplex distribution. *Journal of Statistical Software*, v.71, 2016. Disponível em: <https://www.jstatsoft.org/article/view/v071i11>. Acesso em: 20 set. 2016.