

Análise de gorjetas recebidas em restaurantes utilizando uma abordagem GAMLSS

Elias M. Sabe^{†1}, Viviane C. Silva¹, Luiz R. Nakamura², Andréa C. Konrath³, Thiago G. Ramires⁴

¹Programa de Pós-Graduação em Estatística e Experimentação Agropecuária. Universidade Federal de Lavras (UFLA).

²Departamento de Estatística. Universidade Federal de Lavras (UFLA).

³Departamento de Informática e Estatística. Universidade Federal de Santa Catarina (UFSC).

⁴Departamento de Matemática. Universidade Tecnológica Federal do Paraná (UTFPR).

Resumo: Este estudo visa analisar a quantidade de gorjeta, em dólares, que um garçom recebe em um restaurante, dadas algumas características coletadas no estabelecimento. No total, foram consideradas 244 observações e, além da gorjeta recebida (variável resposta), outras seis variáveis explicativas foram obtidas, sendo elas uma variável numérica (conta total, em dólares) e cinco fatores: sexo do pagador (dois níveis: masculino; feminino), fumante (dois níveis: sim; não), dia da semana (quatro níveis: quinta; sexta; sábado; domingo), tipo de refeição (dois níveis: almoço; jantar) e quantidade de pessoas na mesa (três níveis: 1 ou 2; 3; 4 ou mais). Para a modelagem, foram considerados os modelos aditivos generalizados para localização, escala e forma (GAMLSS), devido à sua grande flexibilidade. Por conta da característica assimétrica da variável gorjeta, foram consideradas três distribuições para explicar a resposta, a saber: gama, inversa Gaussiana e Box-Cox Cole e Green (BCCG). Para o processo de seleção das covariáveis em cada um dos parâmetros das distribuições foi utilizado um procedimento baseado no stepwise e os melhores modelos baseados em cada distribuição foram comparados por meio do critério de informação de Akaike (AIC). O modelo baseado na distribuição BCCG foi o que apresentou menor AIC e, após uma análise de resíduos, verificamos que ele é adequado para explicar o conjunto de dados em estudo.

Palavras-chave: Garçom; Regressão distribucional; Variabilidade.

Analysis of tips received in restaurants using a GAMLSS approach

Abstract: The main aim of this paper is to analyse the amount of tips obtained by a waiter in a restaurant, in dollars, while taking into account other collected variables within the establishment. In total, 244 observations were analysed, and six extra covariates were obtained in addition to the received tips (target variable), one numerical covariate (total bill in dollars) and five factors: payer gender (two levels: male; female), smoking status (two levels: yes; no), day of the week (four levels: Thursday; Friday; Saturday; Sunday), meal type (two levels: lunch; dinner), and number of people at the table (three levels: 1 or 2; 3; 4 or more). The generalised additive models for location, scale, and shape (GAMLSS) were considered due to its high flexibility. Given the asymmetric nature of the tip variable, three distributions were considered to explain the response: gamma, inverse Gaussian, and Box-Cox Cole and Green (BCCG). A stepwise-based procedure was utilised in the covariate selection process for each of the distribution parameters, and the best models for each distribution were compared using the Akaike information criterion (AIC). The model based on the BCCG distribution returned the lowest AIC and based on a residual analysis, it was found to be suitable for explaining the dataset under study.

Keywords: Distributional regression; Variability; Waiter.

[†]Autor correspondente: elias.sabe@estudante.ufla.br.

Introdução

A importância da gorjeta no contexto sociocultural e econômico do mundo contemporâneo é um tema de considerável relevância. Por questões culturais, em certos estabelecimentos comerciais, como bares, restaurantes e similares, a prática da gorjeta varia de país para país. Por exemplo, na Inglaterra, Estados Unidos, Canadá e França, a contribuição aos garçons e bartenders, apesar de não obrigatória por lei, trata-se de uma norma cultural (RIBEIRO, 2017). Nos Estados Unidos, a gorjeta é uma prática comum e pode representar um montante até maior que o próprio salário dos funcionários (SANVOZO, 2023). Por outro lado, de acordo com Ribeiro (2017), na maioria dos países europeus, os salários possuem valores adequados de modo que os garçons não dependam das gorjetas.

Ainda conforme Ribeiro (2017), esta prática acontece em países africanos e do Oriente Médio, entretanto não há um valor específico a ser entregue, variando de acordo com a vontade do cliente, bem como do ambiente em que ele se encontra. Em alguns desses países, é cobrado 10% da conta total de forma opcional, assim como no Brasil. Há ainda países, como o Japão, por exemplo, em que esta prática é praticamente inexistente.

Nos países onde o costume da gorjeta é usual, mas não há uma porcentagem da conta total pré-definida para tal prática, surge um interesse legítimo em investigar quais características dos clientes ou do ambiente impactam diretamente sobre o montante da gratificação oferecida.

Com isto em mente, neste artigo utilizamos os modelos aditivos generalizados para locação, escala e forma (GAMLSS; RIGBY e STASINOPOULOS, 2005), também descritos como modelos de regressão distribucional (HELLER et al., 2022), que generalizam outros modelos famosos como os modelos lineares generalizados (NELDER e WEDDERBURN, 1972) e os modelos aditivos generalizados (HASTIE e TIBSHIRANI, 1990), com o intuito de explicar as gorjetas recebidas por um garçom a partir de algumas variáveis explicativas.

Material e métodos

Conjunto de dados

O conjunto de dados utilizado está disponível em uma coletânea de estudos de caso para estatística empresarial e foi publicado, originalmente, em Bryant e Smith (1995). A base de dados contém 244 observações e sete variáveis que são apresentadas na Tabela 1 juntamente com suas abreviações que serão utilizadas ao longo deste artigo, bem como seu tipo (numérica ou fator) e a unidade de medida ou categorias (níveis) que foram utilizadas.

Tabela 1: Variáveis, abreviação utilizada, classificação e unidades de medida, em estudo

Variável	Abreviação	Tipo	Unidade de medida/Níveis
Gorjeta recebida (resposta)	Gorjeta	numérica	US\$
Valor da conta total	Conta	numérica	US\$
Sexo do pagador	Sexo	fator	masculino; feminino
Há fumantes na mesa	Fumante	fator	sim; não
Dia da semana da refeição	Dia	fator	quinta-feira; sexta-feira; sábado; domingo
Tipo de refeição	Refeição	fator	almoço; jantar
Número de pessoas na mesa	Pessoas	fator	1 ou 2; 3; 4 ou mais

Fonte: Autores.

Estes dados já foram estudados por alguns autores e, do ponto de vista dos modelos estatísticos de predição, Mirugwe (2020) apresenta uma solução a partir de um modelo de regressão linear múltipla e, posteriormente, lasso. Todavia, fica claro no trabalho do autor que o modelo pode não ser o mais adequado, uma vez que claramente há grandes desvios, em especial

na cauda superior, no *qq-plot* dos resíduos, indicando a sua não-normalidade e, assim, violando os pressupostos básicos deste modelo e, conseqüentemente, potencialmente prejudicando suas inferências.

Com o intuito de flexibilizar as restrições impostas pelo modelo de regressão linear, a proposta deste artigo é utilizar os GAMLSS para explicar a resposta, a partir das variáveis explicativas disponíveis na Tabela 1.

Modelos aditivos generalizados para locação, escala e forma (GAMLSS)

Formalmente, os GAMLSS são modelos de regressão semi-paramétricos, em que uma distribuição para a variável resposta é assumida e todos os seus parâmetros podem ser modelados a partir de uma estrutura linear ou por funções de suavização. Nesta abordagem, todos os parâmetros $\theta_k = (\theta_1, \dots, \theta_p)^\top$, $k = 1, \dots, p$, de uma distribuição com função (densidade) de probabilidade $f_Y(y|\theta_k)$, em que Y é a variável resposta (valor de gorjeta recebida, por exemplo), são modelados utilizando-se diferentes variáveis explicativas disponíveis. Matematicamente, podemos escrever o modelo como

$$g_k(\theta_k) = \eta_k = X_k\beta_k + \sum_{j=1}^{J_k} s_{jk}(x_{jk}), \quad (1)$$

em que $g_k(\cdot)$ é uma função de ligação usualmente selecionada de acordo com o suporte do parâmetro (DE BASTIANI et al., 2018), η_k é o preditor do k -ésimo parâmetro da distribuição da variável resposta, X_k é uma matriz de delineamento, β_k é um vetor de parâmetros, $s_{jk}(\cdot)$ são funções de suavização, como um P-spline (EILERS e MARX, 1996), por exemplo, e x_{jk} são as covariáveis que necessitam da função de suavização para explicar o k -ésimo parâmetro de maneira adequada. Conforme indicado por Ramires et al. (2019), caso uma função de suavização seja introduzida em uma estrutura de regressão, não realizamos testes de hipóteses acerca da função ou de seus coeficientes, mas sim, observamos e interpretamos seu comportamento graficamente.

Distribuições para a variável resposta

Dada a característica assimétrica da variável resposta que será apresentada na seção de Resultados e Discussão, há necessidade de selecionarmos uma distribuição para representar a variável resposta que seja capaz de capturar e modelar adequadamente tal característica. Dentre as distribuições mais utilizadas para este fim, em especial nos modelos lineares generalizados, estão a gama (GA) e a inversa Gaussiana (IG).

Nos GAMLSS, é comum utilizarmos parametrizações não-usuais de distribuições, como a GA e a IG, com o intuito de facilitar a interpretação das estruturas de regressão construídas a partir do modelo apresentado em (1) (RIGBY et al., 2019). Se $Y > 0$ segue uma distribuição GA com parâmetros μ e σ , então sua função densidade de probabilidade pode ser escrita como

$$f_Y(y|\mu, \sigma) = \frac{y^{\frac{1}{\sigma^2}-1} \exp\left(-\frac{y}{\mu\sigma^2}\right)}{(\mu\sigma^2)^{\frac{1}{\sigma^2}} \Gamma\left(\frac{1}{\sigma^2}\right)},$$

em que $\mu > 0$ é exatamente a média e $\sigma > 0$ é o coeficiente de variação. Ainda, se $Y > 0$ segue uma distribuição IG com parâmetros μ e σ , então sua função densidade de probabilidade é definida como

$$f_Y(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2y^3}} \exp\left[-\frac{(y-\mu)^2}{2\mu^2\sigma^2y}\right],$$

em que $Y > 0$, $\mu > 0$ é exatamente a média e $\sigma > 0$.

Como podemos observar, uma vez que ambas as distribuições supracitadas possuem apenas dois parâmetros, não há um parâmetro que modele diretamente a assimetria da resposta, isto é, que considere uma estrutura de regressão que relacione covariáveis com esta característica. Neste sentido, compararemos ambas as distribuições GA e IG com a Box-Cox Cole e Green (BCCG; COLE e GREEN, 1992), que possui um de seus parâmetros associado diretamente ao coeficiente de assimetria. A distribuição BCCG no contexto dos GAMLSS, e outras distribuições pertencentes à família Box-Cox, vêm ganhando grande notoriedade em diferentes aplicações na literatura, como, por exemplo, nos trabalhos de Bann et al. (2022), Nakamura et al. (2022a), Nakamura et al. (2022b), Ramires et al. (2021), Ribeiro et al. (2021), Stasinopoulos et al. (2022) e Wada et al. (2022).

Se $Y > 0$ segue uma distribuição BCCG com parâmetros μ , σ e ν , então sua função densidade de probabilidade é definida como

$$f_Y(y|\mu, \sigma, \nu) = \frac{y^{\nu-1} \exp(-0,5 z^2)}{\mu^\nu \sigma \sqrt{\pi} \Phi\left(\frac{1}{\sigma|\nu|}\right)}$$

em que $Y > 0$, $\mu > 0$ é exatamente a mediana, $\sigma > 0$ é o coeficiente de variação, $-\infty < \nu < \infty$ é um parâmetro relacionado à assimetria, Z segue uma distribuição normal padrão e $\Phi(\cdot)$ é a função distribuição acumulada de uma normal padrão.

Seleção do modelo final

Para o processo de seleção de covariáveis em cada uma das estruturas de regressão do modelo apresentado em (1), neste trabalho, consideramos um procedimento baseado no *stepwise*, denominado Estratégia A (NAKAMURA et al., 2017; STASINOPOULOS et al., 2017; RAMIRES et al., 2021). Basicamente, é realizado um procedimento *forward* no primeiro parâmetro (μ), em sequência um outro passo *forward* no segundo parâmetro (σ) e assim sucessivamente, até o último parâmetro da distribuição. Posteriormente, um procedimento *backward* é aplicado com o intuito de retirar variáveis nas estruturas de regressão, caso necessário.

A escolha dos melhores modelos baseados nas distribuições GA, IG e BCCG, selecionados a partir da Estratégia A, foi realizada com base no critério de informação de Akaike (AKAIKE, 1974). O modelo que apresenta o menor AIC é considerado como o melhor ajustado.

Para verificar a adequação dos modelos ajustados, conduzimos uma análise dos resíduos quantílicos normalizados (DUNN e SMYTH, 2012) que, no caso de respostas contínuas, são definidos como

$$\hat{r} = \Phi^{-1} \left[F(y|\hat{\theta}) \right],$$

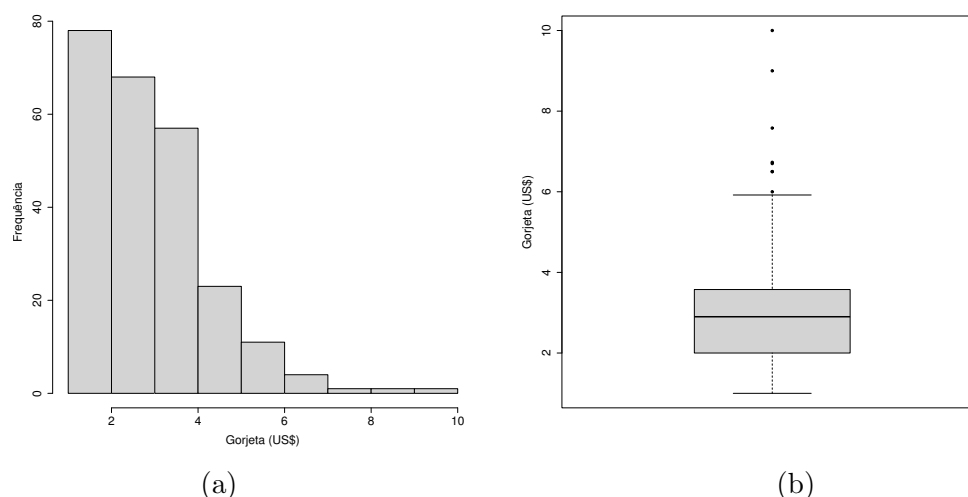
em que $\Phi^{-1}(\cdot)$ é a inversa da função distribuição acumulada de uma normal padrão e $F(\cdot)$ é a função distribuição acumulada ajustada relacionada à resposta. A grande vantagem da utilização deste tipo de resíduo é que, caso o modelo ajustado seja adequado, independentemente da distribuição considerada para explicar a resposta, os resíduos seguirão uma distribuição normal padrão (STASINOPOULOS et al., 2017).

Resultados e discussão

Análise descritiva e exploratória

Inicialmente foi realizada uma análise descritiva acerca da variável resposta em estudo. Para tal, a Figura 1 apresenta o histograma (Painel (a)) e o gráfico de caixa (Painel (b)) das gorjetas recebidas, onde podemos observar claramente sua característica assimétrica, justificando assim a escolha das distribuições GA, IG e BCCG como potenciais candidatas em seu processo de modelagem.

Figura 1: Gorjetas recebidas em um restaurante: (a) histograma e (b) gráfico de caixa



Fonte: Autores.

O coeficiente de assimetria observado é de 1,44 (Tabela 2), corroborando assim com o observado na Figura 1. As gorjetas média e mediana recebidas foram de US\$ 3,00 e US\$ 2,90, respectivamente. A menor gorjeta foi de, exatamente, US\$ 1,00, ao passo que a maior foi de US\$ 10,00. Ainda, podemos observar que a distribuição marginal das gorjetas recebidas é leptocúrtica (possui caudas pesadas), uma vez que o coeficiente de curtose observado foi de 3,50.

Tabela 2: Estatísticas descritivas acerca da resposta gorjeta recebida (US\$)

	Desvio Padrão	Primeiro Quartil	Mediana	Terceiro Quartil	Máximo	Assimetria	Curtose
Média	1,38	2,00	2,90	3,56	10,00	1,44	3,50

Fonte: Autores.

Com o intuito de compreender as características dos clientes em estudo, seguimos com uma análise dos cinco fatores candidatos a explicar o valor de gorjeta recebida pelo garçom. De acordo com o conjunto de dados em estudo, das 244 gorjetas, 157 (64,34%) são provenientes de clientes do sexo masculino e 87 (35,66%) do sexo feminino conforme apresentado na Tabela 3. A maioria dos clientes não era fumante (61,89%), bem como estavam sozinhos ou acompanhados de apenas uma pessoa (65,57%). Ainda, o maior número de gorjetas obtidas foi observado no sábado (35,65%) e em jantares (72,13%).

Em relação à única variável explicativa quantitativa candidata, observamos pela Tabela 4 que os valores médio e mediano das contas foi de US\$ 19,79 e US\$ 17,80, respectivamente. A conta mais barata refere-se à soma de US\$ 3,07 e a mais cara de US\$ 50,81. Ainda, a distribuição desta variável é assimétrica (coeficiente igual a 1,12) e leptocúrtica (coeficiente igual a 1,14).

A Figura 2 apresenta o relacionamento da resposta, gorjetas recebidas, com cada uma das covariáveis candidatas. De acordo com o Painel (a), observamos que há uma tendência quase

Tabela 3: Frequência observada dos níveis dos fatores candidatos a explicar a gorjeta recebida pelo garçom

Fator	Nível	Frequência	Porcentagem
Sexo	Masculino	157	64,34
	Feminino	87	35,66
Fumante	Sim	93	38,11
	Não	151	61,89
Dia	Quinta-feira	62	25,41
	Sexta-feira	19	7,79
	Sábado	87	35,65
	Domingo	76	31,15
Refeição	Almoço	68	27,87
	Jantar	176	72,13
Pessoas	1 ou 2	160	65,57
	3	38	15,57
	4 ou mais	46	18,86

Fonte: Autores.

Tabela 4: Estatísticas descritivas acerca do valor da conta total (US\$)

	Desvio Padrão	Primeiro Mínimo	Primeiro Quartil	Mediana	Terceiro Quartil	Máximo	Assimetria	Curtose
Média	8,90	3,07	13,35	17,80	24,13	50,81	1,12	1,14

Fonte: Autores.

linear de que, quanto maior o valor da conta total, maior a gorjeta recebida. Fica evidente ainda a grande dispersão das observações, bem como a presença de alguns valores discrepantes. A partir do Painel (b), podemos observar que a mediana das gorjetas recebidas é praticamente a mesma independentemente do sexo do cliente, sendo a variabilidade observada em clientes do sexo masculino um pouco superior, bem como a quantidade de valores discrepantes e o valor máximo observado. Em relação ao Painel (c), aparentemente não há diferenças entre a gorjeta mediana recebida pelo garçom por clientes fumantes, apesar de uma maior variabilidade apresentada neste grupo. No que tange às gorjetas recebidas nos diferentes dias de semana, Painel (d), domingo apresentou maior mediana, ao passo que quinta-feira a menor. Finalmente, a maior mediana de gorjetas é observada no jantar (Painel (e)) e em grupos de quatro ou mais pessoas (Painel (f)).

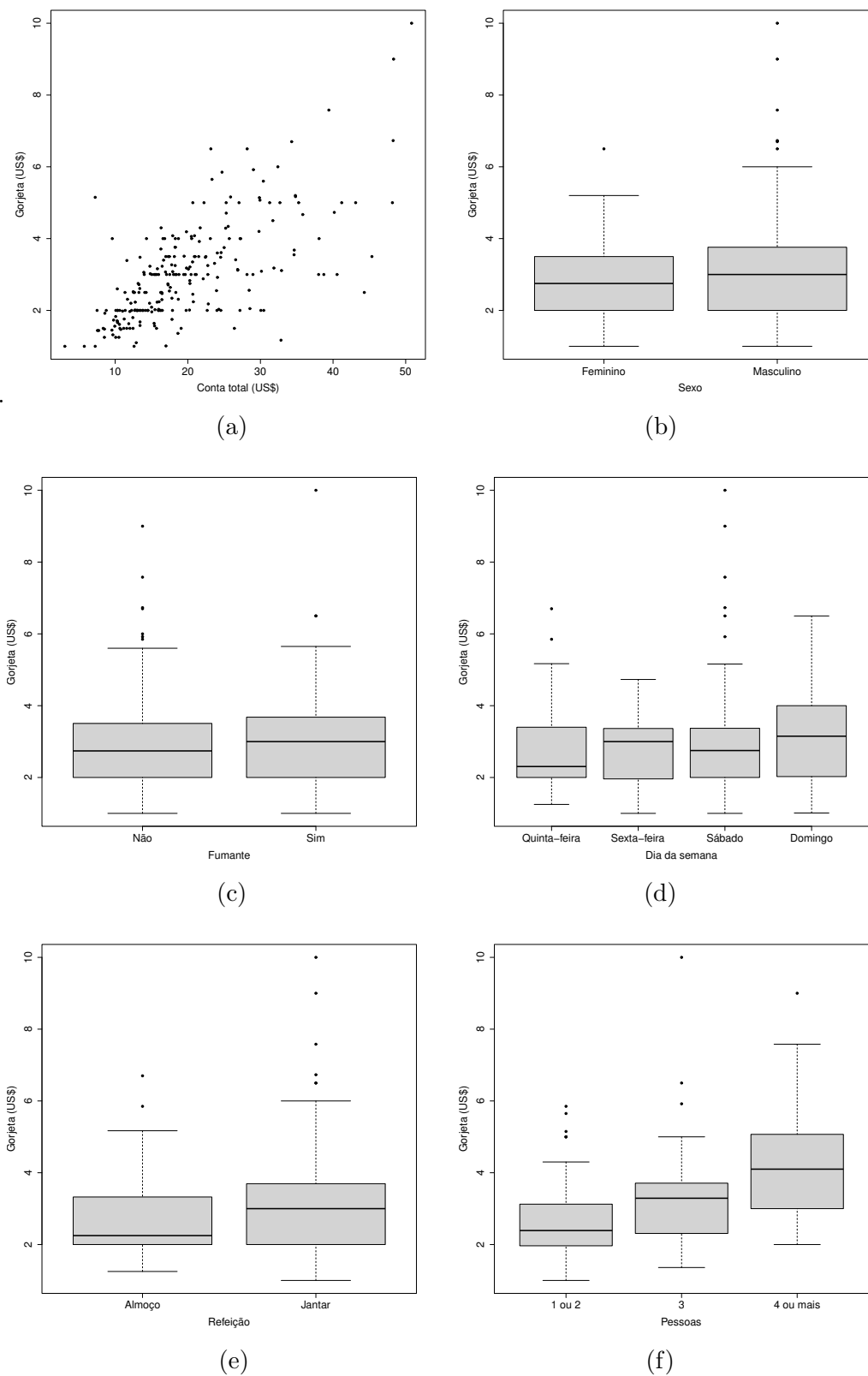
Modelagem

Realizada a análise descritiva e exploratória, conforme descrito na seção de Metodologia, foram ajustados os GAMLSS baseados em três distribuições distintas (GA, IG e BCCG). Após o processo de seleção de variáveis empregado (Estratégia A), foram comparados os AIC obtidos dos melhores modelos obtidos a partir de cada distribuição e, conforme pode ser visualizado na Tabela 5, o GAMLSS baseado na distribuição BCCG foi o mais adequado para explicar o conjunto de dados em estudo (AIC equivalente a 597,92).

Assim, o modelo final ajustado é dado por

$$\begin{aligned}
 \hat{\mu} &= \exp \{0,29 + s_{11}(\text{Conta})\} \\
 \hat{\sigma} &= \exp \{-1,42 + 0,41[\text{Fumante} = \text{Sim}]\} \\
 \hat{\nu} &= 0,16 + s_{11}(\text{Conta}) - 1,04[\text{Sexo} = \text{Masculino}].
 \end{aligned}
 \tag{2}$$

Figura 2: Relacionamento entre a gorjeta recebida em um restaurante com cada uma das co-variáveis candidatas: (a) valor total da conta; (b) sexo do pagador; (c) há fumantes na mesa; (d) dia da semana da refeição; (e) tipo de refeição; e (f) número de pessoas na mesa



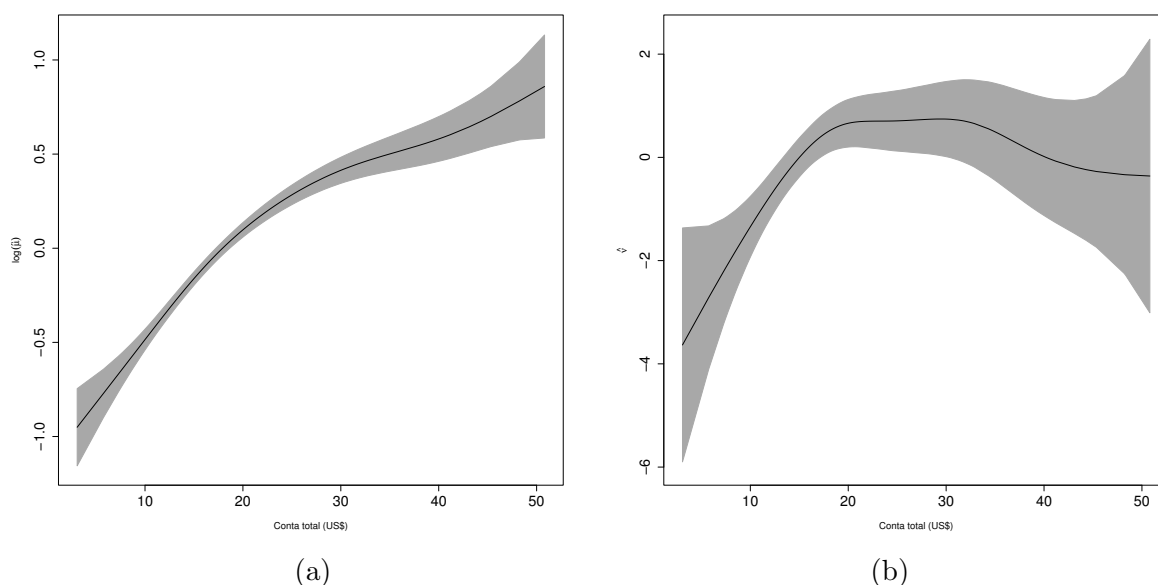
Fonte: Autores.

Tabela 5: Valores obtidos de AIC pelos melhores modelos baseados em cada distribuição

Modelo	Parâmetros	AIC
Box-Cox Cole e Green	3	597,92
Gama	2	609,78
Inversa Gaussiana	2	615,84

Fonte: Autores.

Podemos observar que a única variável selecionada para explicar a mediana (μ) das gorjetas recebidas foi o valor da conta total, por meio de uma função de suavização. A Figura 3 apresenta este relacionamento, onde podemos observar que há um aumento das gorjetas recebidas conforme o valor da conta total aumenta. Isto é, a função de suavização capturou exatamente o comportamento mediano discutido previamente no gráfico de dispersão apresentado na Figura 2.

Figura 3: Funções de suavização utilizadas para explicar o relacionamento da covariável conta total e: (a) gorjeta mediana recebida – μ ; (b) parâmetro de assimetria – ν 

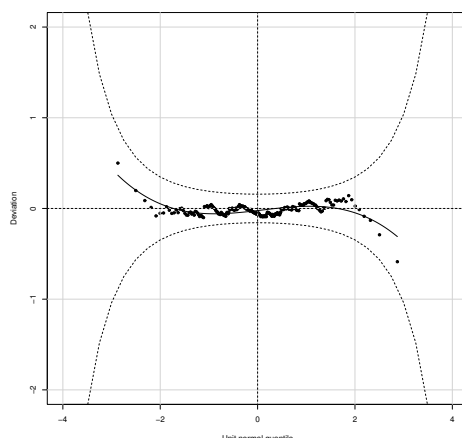
Fonte: Autores.

Apenas a variável fumante foi significativa ao nível de 5% para explicar o coeficiente de variação (σ) das gorjetas recebidas. Podemos observar que há uma maior variabilidade das gorjetas recebidas de clientes fumantes, mais especificamente, o coeficiente de variação para este grupo é de 36,42%, ao passo que para clientes não fumantes este coeficiente equivale a 24,17%.

A estrutura de regressão do parâmetro ν , relacionado à assimetria, foi composta por duas variáveis, conta (a partir de uma função de suavização) e sexo (significativa ao nível de 5%). Conforme podemos observar na Figura 3, há um crescimento no valor de $\hat{\nu}$ no intervalo das contas de US\$ 1,00 até, aproximadamente, US\$ 20,00. Após este limiar, o coeficiente é praticamente constante até, aproximadamente, US\$ 30,00 e, há um leve declínio após esta cifra. Apesar do comportamento observado, o valor de $\hat{\nu}$ permaneceu inferior a um em toda a amplitude das contas totais. Conforme Rigby et al. (2019), para a distribuição BCCG, $\nu < 1$ indica assimetria positiva da resposta, fato este corroborado pelas análises descritiva e exploratória apresentadas na Figura 1 e Tabela 2. A assimetria observada é ainda mais forte em clientes do sexo masculino.

A Figura 4 apresenta o *worm plot* (VAN BUUREN e FREDRIKS, 2001) dos resíduos obtidos para o modelo final apresentado em (3). Uma vez que todos os pontos encontram-se dentro das bandas de 95% de confiança, podemos afirmar que o GAMLSS baseado na distribuição BCCG ajustado forneceu um ajuste razoável e é adequado para o conjunto de dados em estudo.

Figura 4: *Worm plot* obtido a partir do modelo final ajustado com base na distribuição BCCG



Fonte: Autores.

Considerações finais

O uso dos modelos aditivos generalizados para locação, escala e forma (GAMLSS) com base na distribuição Box-Cox Cole e Green (BCCG) foi adequado para modelar os dados referentes às gorjetas recebidas num restaurante registradas por um garçom. O modelo não só foi vantajoso para modelar o parâmetro de locação (mediana da gorjeta) como também os parâmetros de escala (coeficiente de variação) e forma (assimetria). Com esta abordagem, foi possível descrever e interpretar a natureza da variável resposta de uma forma mais objetiva, corrigindo eventuais problemas nas pressuposições de modelos de regressão mais usualmente aplicados na literatura.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Referências

- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19, n. 6, p. 716–723, 1974.
- BANN, D.; WRIGHT, L.; COLE, T. J. Risk factors relate to the variability of health outcomes as well as the mean: a GAMLSS tutorial. *eLife*, v. 11, p. e72357, 2022.
- BRYANT, P. G.; SMITH, M. A. *Practical Data Analysis*. Chicago: Irwin. 1995.
- COLE, T. J.; GREEN, P. J. Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in Medicine*, v. 11, n. 10, p. 1305–1319, 1992.
- DE BASTIANI, F.; RIGBY, T. A.; STASINOPOULOS, D. M.; CYSNEIROS, A. H. M. A.; URIBE-OPAZO, M. A. Gaussian Markov random field spatial models in GAMLSS. *Journal of Applied Statistics*, v. 45, n. 1, p. 168–186, 2018.
- DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, v. 5, n. 3, p. 236–244, 1996.

- EILERS, P. H. C.; MARX, B. D. Flexible smoothing with *B*-splines and penalties. *Statistical Science*, v. 11, n. 2, p. 89–121, 1996.
- HASTIE, T. J.; TIBSHIRANI, R. J. *Generalized Additive Models*. Boca Raton: CRC Press. 1990.
- HELLER, G. Z.; ROBLEDO, K. P.; MARSCHNER, I. C. Distributional regression in clinical trials: treatment effects on parameters other than the mean. *BMC Medical Research Methodology*, v. 22, p. 56, 2022.
- MIRUGWE, A. *Restaurant Tipping Linear Regression Model*. Disponível em: <https://ssrn.com/abstract=3719232>. 2020. Acesso em: 30/09/2023.
- NAKAMURA, L. R.; RAMIRES, T. G.; RIGHETTO, A. J.; PESCIM, R. R.; ROQUIM, F. V.; SAVIAN, T. V.; STASINOPOULOS, D. M. Cattle reference growth curves based on centile estimation: a GAMLSS approach. *Computers and Electronics in Agriculture*, v. 192, p. 106572, 2022a.
- NAKAMURA, L. R.; RAMIRES, T. G.; RIGHETTO, A. J.; SILVA, V. C.; KONRATH, A. C. Using the Box-Cox family of distributions to model censored data: a distributional regression approach. *Brazilian Journal of Biometrics*, v. 40, n. 4, p. 407–414, 2022b.
- NAKAMURA, L. R.; RIGBY, R. A.; STASINOPOULOS, D. M.; LEANDRO, R. A.; VILLEGAS, C.; PESCIM, R. R. Modelling location, scale and shape parameters of the Birnbaum-Saunders generalized *t* distribution. *Journal of Data Science*, v. 15, n. 2, p. 221–237, 2017.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society: Series A*, v. 135, n. 3, p. 370–384, 1972.
- R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2023. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- RAMIRES, T. G.; NAKAMURA, L. R.; RIGHETTO, A. J.; KONRATH, A. C.; PEREIRA, C. A. B. Incorporating clustering techniques into GAMLSS. *Stats*, v. 2021, n. 4, p. 916–930, 2021.
- RAMIRES, T. G.; NAKAMURA, L. R.; RIGHETTO, A. J.; PESCIM, R. R.; MAZUCHELI, J.; CORDEIRO, G. M. A new semiparametric Weibull cure rate model: fitting different behaviors within GAMLSS. *Journal of Applied Statistics*, v. 46, n. 15, p. 2744–2760, 2019.
- RAMIRES, T. G.; NAKAMURA, L. R.; RIGHETTO, A. J.; PESCIM, R. R.; MAZUCHELI, J.; RIGBY, R. A.; STASINOPOULOS, D. M. Validation of stepwise-based procedure in GAMLSS. *Journal of Data Science*, v. 19, n. 1, p. 96–110, 2021.
- RIBEIRO, L. F. N. *Gorjetas: uma Breve Análise Jurídica sob o Prisma do Direito Consumerista e Trabalhista*. Trabalho de Conclusão de Curso (Graduação em Direito) – Faculdade de Direito do Recife, Universidade Federal de Pernambuco, p. 60. 2017.
- RIBEIRO, T. F.; SEIDEL, E. J.; GUERRA, R. R.; PEÑA-RAMÍREZ, F. A.; SILVA, A. M.

- Soybean production value in the Rio Grande do Sul under the GAMLSS framework. *Communications in Statistics: Case Studies, Data Analysis and Applications*, v. 7, n. 2, p. 146–165, 2021.
- RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society – Series C: Applied Statistics*, v. 54, n. 3, p. 507–554, 2005.
- RIGBY, R. A.; STASINOPOULOS, M. D.; HELLER, G. Z.; DE BASTIANI, F. *Distributions for Modeling Location, Scale, and Shape: Using GAMLSS in R*. Boca Raton: CRC Press. 2019.
- SANVOZO, C. C. Discurso polêmico em torno da lei da gorjeta. *Discursividades*, v. 12, n. 1, p. e-121230, 2023.
- STASINOPOULOS, D. M.; RIGBY, R. A.; DE BASTIANI, F. Principal component regression in GAMLSS applied to Greek–German government bond yield spreads. *Statistical Modelling*, v. 22, n. 1–2, p. 127–145, 2022.
- STASINOPOULOS, M. D.; RIGBY, R. A.; HELLER, G. Z.; VOUDOURIS, V.; DE BASTIANI, F. *Flexible Regression and Smoothing: Using GAMLSS in R*. Boca Raton: CRC Press. 2017
- VAN BUUREN, S.; FREDRIKS, M. Worm plot: a simple diagnostic device for modelling growth reference curves. *Statistics in Medicine*, v. 20, n. 8, p. 1259–1277, 2001.
- WADA, Y.; GOTO, N.; KITAGUCHI, Y.; YASUO, M.; HANAOKA, M. Referential equations for pulmonary diffusing capacity using GAMLSS models derived from Japanese individuals with near-normal lung function. *PLOS ONE*, v. 17, n. 7, p. e0271129, 2022.