

REGRESSÃO DE POISSON PARA ESTIMAR O COMPARECIMENTO ÀS INTERVENÇÕES PELO TREINAMENTO FÍSICO OU À ESCOLA DE FUTEBOL

BRITO, Diogo de Freitas^{1,2}; BRASIL, Roxana Macedo³; BARRETO, Ana Cristina Lopes y Glória²; JUNIOR, Homero da Silva Nahum^{2,4}

286

Resumo

O estudo corrente objetivou modelar o comparecimento de clientes às intervenções em uma escola esportiva. O banco de dados utilizado tinha 96 registros sobre sexo, nível cronológico do cliente, mês do ano e número diário de pessoas nas sessões de treinamento. O modelo estimado foi de Poisson, codificado no RStudio. A estimativa com Nível como única variável explicativa foi significativo, $\alpha < 0,04$, sem violar a regra de igualdade entre esperança e variância, Verdadeira Dispersão = 0,93. Todavia, não houve diferença estatisticamente significativa com relação ao modelo nulo, valor-p = 0,18. A conclusão foi de que o modelo foi válido, mas não tece capacidade de reduzir a incerteza na tomada de decisão.

Palavras-chave: Econometria. Modelagem. Negócio. Administração. Planejamento.

Abstract

The current study aimed to model client attendance at interventions in a sports school. The database used had 96 records on gender, client chronological level, month of the year and daily number of people in training sessions. The estimated model was Poisson, coded in RStudio. The estimate with Level as the only explanatory variable was significant, $\alpha < 0.04$, without violating the rule of equality between expectation and variance, True Dispersion = 0.93. However, there was no statistically significant difference in relation to the null model, p-value = 0.18. The conclusion was that the model was valid, but it does not have the capacity to reduce uncertainty in decision-making.

Keywords: Econometrics. Modeling. Business. Administration. Planning.

Introdução

O modelo de regressão de Poisson se alicerçaria na distribuição de probabilidade homônima, cuja teoria de probabilidade foi originalmente desenvolvida para aplicações na área de ciências humanas, mais precisamente em fenômenos cíveis e criminais (Poisson, 1837). A área social seria rica em exemplos como a relação entre racismo, e ambiente e insegurança alimentares de adultos residentes em determinado centro urbano (Vargas,

¹ Docente do Curso de Gestão Desportiva e do Lazer do Centro Universitário Celso Lisboa;

² Docentes do Curso de Educação Física do Centro Universitário Celso Lisboa;

³ Docente Ph.D. em Educação Física;

⁴ Docente da Escola de Saúde da Universidade Cândido Mendes.

2024), e a investigação da associação entre futebol e violência doméstica, modelada pelo quantitativo de homicídios femininos (Schäffer, 2024).

Estudos na interação entre humanas e saúde seriam desenvolvidos para avaliar a associação entre homofobia internalizada e os respectivos domínios, e a depressão em indivíduos homossexuais e bissexuais (Batista *et al.*, 2024), por exemplo. Assim como para estimar a prevalência, em adolescentes, de *cyberbullying*, considerando características demográficas, familiares, comportamentais e saúde psíquica (Malta *et al.*, 2024). A ocorrência de sífilis gestacional e congênita em certa Unidade da Federação (Nogueira *et al.*, 2021), ou a coligação ente o uso de serviços de saúde e a existência de comorbidades em sulistas com diagnóstico de Covid-19 (Delpino *et al.*, 2024).

No domínio das políticas públicas, as Parcerias Público-Privadas (PPP) tiveram os respectivos fatores críticos identificados e testados (Thamer e Lazzarini, 2015), fornecendo referencial ao processo anterior à celebração do contrato. A modelagem de Poisson detetou o conjunto de determinantes às PPP para América Latina e Caribe, adicionalmente estimou a probabilidade de realização de contratos de *Greenfield* (terreno sem edificação) e Alienação (Fernandez *et al.*, 2017).

Especificamente no domínio da Administração e dos negócios, a violação de conformidades industriais foi pesquisada com o uso da representação gráfica dos resíduos de Poisson (Russo, Camargo e Samohyl, 2008). Possível foi demonstrar para os segmentos de vestuários feminino e unissex e calçados femininos que a diversificação de marcas franqueadas dependia da quantidade unidades que o franqueado possuía (Silva, 2014), ou seja, o comportamento do investidor tendia à ausência de fidelidade à marca.

A defasagem na divulgação das demonstrações contábeis por Organizações brasileiras de capital aberto teve os determinantes levantados por modelagem de Poisson, aspecto particularmente relevante porque as ações daquelas instituições comporiam o Índice da Bolsa de Valores de São Paulo (Kirch *et al.*, 2012). Os exemplos apresentados teriam por interseção a característica discreta da variável de interesse, sem a possibilidade de ocorrência de valores negativos, portanto se ajustando à distribuição de Poisson dada por (Gujarati e Porter, 2011):

$$f(Y_i) = \frac{\mu^Y e^{-\mu}}{Y!}; Y = 0, 1, 2, \dots$$

Na qual: $f(Y_i)$: probabilidade da variável Y assumir valores inteiros não negativos;

$Y_i!$: fatorial de Y_i .

Então, $E(Y) = \text{var}(Y) = \mu$, logo o modelo de regressão de Poisson seria $Y_i = E(Y_i) + v_i = \mu_i + v_i$, nele os Y seriam distribuídos independentemente como variáveis aleatórias com média μ_i , consequentemente cada ocorrência ou evento do fenômeno estudado seria explicado por variáveis independentes (X) e dada por:

$$\mu_i = E(Y_i) = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$$

Pela explicação matemática, a estimação do modelo de Poisson envolveria parâmetros não lineares, tendo $\mu = \mu_i$ e Y_i dada por:

$$Y_i = \frac{\mu^Y e^{-\mu}}{Y!} + v_i$$

Com base no exposto, o estudo corrente objetivou modelar o comparecimento de clientes às intervenções em uma escola esportiva.

Metodologia

Empregou-se o banco de dados de uma escola esportiva composto por 96 registros, e as informações sobre sexo (Sexo, dicotômica), nível cronológico do cliente (Nível, politômica), mês do ano (Mês, politômica) e número diário de pessoas presentes nas intervenções em razão das variáveis anteriores (Presentes, discreta). Essa foi a variável dependente, as duas primeiras explicativas apresentavam as codificações constantes no Quadro I, enquanto que Mês adotou a numeração padrão.

Quadro I: Codificação das Variáveis Sexo e Nível.

Variável	Código	Semântica
Sexo	0	Feminino
	1	Masculino
Nível	1	6 a 8 anos
	2	9 a 11 anos
	3	12 a 14 anos
	4	15 a 17 anos

Fonte: Os Autores (2025).

A regressão de Poisson, utilizando como função de ligação log para garantir os valores ajustados no intervalo $[0; \infty]$, foi implementada no RStudio 2022.12.0+353 para

Windows®, utilizando as bibliotecas AER 1.2-14 – *Applied Econometrics with R* (Kleiber e Zeileis, 2008), broom 1.0.7 – *Convert Statistical Objects into Tidy Tibbles* (Robinson, Hayes e Couch, 2024), car 3.1-3 – *Companion to Applied Regression* (Fox e Weisberg, 2019), ggplot2 3.5.1 – *Create Elegant Data Visualizations Using the Grammar of Graphics* (Wickham, 2016), ISwR 2.0-9 – *Introductory Statistics with R* (Dalgaard, 2024), readxl 1.4.3 – *Read Excel Files* (Wickham e Bryan, 2023) e tidyverse 2.0.0 – *Easily install and Load the ‘Tidyverse’* (Wickham et al., 2019).

Resultados e Discussão

Considerando a quantidade de variáveis explicativas ($n = 3$) foram estimados todos os modelos possíveis (Total) para prever Presentes, o que poderia ser feito com as três variáveis ($C(3,3)$), duas dentre as três ($C(3,2)$) e a utilização de uma única preditora ($C(3,1)$). Então, a quantidade de combinações possíveis (C) refletiria o número de modelos gerados, considerando o quantitativo de variáveis independentes utilizadas (k), o estatisticamente foi estimado por (Meyer, 2022):

$$C(n, k) = \frac{n!}{k!(n-k)!}, \text{ logo}$$

$$Total = C(3,3) + C(3,2) + C(3,1) = \frac{3!}{3!(3-3)!} + \frac{3!}{2!(3-2)!} + \frac{3!}{1!(3-1)!} = 3 + 3 + 1 = 7$$

Para cada modelo desenvolvido necessário foi verificar a acurácia (precisão dos valores estimados) e não violação das características da Distribuição de Poisson, $E(Y) = \text{var}(Y)$. À acurácia foi estimado o Critério de Informação de Akaike (AIC – *Akaike’s Information Criterion*), porque seria indicador matematicamente robusto (Akaike, 1974; Shibata, 1981; Burnham e Anderson, 2004), e amplamente empregado nos domínios acadêmico e científico (Soares, Guimarães e Souza, 2013; Ferro, Lima e Trentin, 2019; Bleidorn et al., 2024). O modelo adequado seria aquele com o menor AIC, então elegíveis foram A e B (Tabela 1).

O Desvio Residual (DR) indicaria a qualidade do ajuste ao modelo, em outras palavras, o quanto os dados se ajustariam à Distribuição de Poisson, então a razão entre DR e a quantidade de Graus de Liberdade (GL) informaria a não violação daquela distribuição de probabilidade. Assim, $DR/DL \cong 1,00$ significaria que $E(Y) \cong \text{var}(Y)$, o que foi

constatado em todos os modelos (Tabela 1), ratificando que o fenômeno estudado, número de Presentes, se aproximaria da Distribuição de Poisson, e a seleção dos modelos A e B.

Tabela 1: Estimativas AIC, Desvio Residual (DR) e Graus de Liberdade (GL) para os Possíveis Modelos, tendo $\alpha = 0,05$.

Modelo	Variáveis Explicativas	AIC	DR	GL	DR/GL
A	Nível	546,16	89,69	92	0,97
B	Sexo	546,42	93,95	94	1,00
C	Nível e Sexo	547,51	89,04	91	0,98
D	Mês	552,84	80,36	94	0,85
E	Nível e Mês	553,92	75,45	81	0,93
F	Sexo e Mês	554,18	79,71	83	0,96
G	Nível, Sexo e Mês	555,27	74,80	80	0,94

Fonte: Os Autores (2025).

O sumário do Modelo B, utilizando somente Sexo como variável explicativa, revelou que em relação ao Sexo Feminino, o Masculino não apresentava diferença estatisticamente significativa, valor-p = 0,42 (Tabela 2). O Intercepto seria o valor de Y quando X = 0 (Fernandes *et al.*, 2020), em vista disso, seria carente de valor pragmático, dado o resultado da variável Sexo. Em última análise, o modelo em tela não deteve aplicabilidade, mesmo conquistando significância estatística o Coeficiente do Intercepto.

Tabela 2: Sumário dos Modelos com Coeficientes (Coef), Exponencial dos Coeficientes (Exp(Coef)), Limites Inferior (LI), Superior (LS), Erro Padrão (S), Estatística Teste (Z) e Valor-p, $\alpha = 0,05$.

Modelo	Variável	Coef	Exp(Coef)	LI	LS	S	Z	Valor-p
A	Intercepto	278,81				0,05	55,06	0,00
	Nível 2	0,04	1,04	0,97	1,11	0,07	0,50	0,62
	Nível 3	0,07	1,08	0,93	1,25	0,07	1,05	0,29
	Nível 4	0,15	1,16	0,87	1,54	0,07	2,10	0,04
B	Intercepto	287,29				0,03	83,71	0,00
	Sexo Masculino	-0,04	0,96	1,04	0,89	0,05	-0,81	0,42

Fonte: Os Autores (2025).

No Modelo A (Tabela 2), a significância se apresentou para o Intercepto, o qual isoladamente seria desprovido de utilidade. A estimativa para o Nível 4, também, conquistou relevância estatística, valor-p = 0,04, indicando que, em relação ao Nível 1, aquele teria 1,16 vezes mais chance de se fazer presente às intervenções. Essa constatação convergiria ao esperado, pois a faixa etária 15 a 17 anos (Nível 4) teria, a princípio, menor dependência dos responsáveis para se deslocar, enquanto os indivíduos entre seis e oito anos seriam plenamente dependentes.

Tais considerações confirmariam a adequação do modelo estimado, pois a elevação da faixa etária foi acompanhada pelas estimativas dos coeficientes e a redução do valor-p. A razão de chance (*odds ratio* = 1,16) expressaria a chance de ocorrência do efeito (presença) sobre a chance de não ocorrer o efeito (falta), mas poderia ser entendida como a razão entre as chances do efeito em um grupo sobre a de outro (Francisco *et al.*, 2008). No estudo corrente, a chance de presença do Nível 4 sobre a do Nível 1.

Valeria destacar que investigações transversais e epidemiológicas, comumente utilizariam razão de chances e razão de prevalências, estimativa de associação entre duas variáveis qualitativas (Sousa e Araújo, 2024). Essas medidas dependeriam do modelo de regressão, pois dependeriam dos coeficientes estimados, consequentemente, o intervalo de confiança influenciaria a validade, independentemente da utilização da modelagem de Cox, log-binomial, Poisson ou logístico, por exemplo (Coutinho, Scazufca e Menezes, 2008; Hoffmann, 2016). De forma geral, a opção pela medida deveria considerar a prevalência do fenômeno ou efeito, caso fosse elevada, o emprego da razão de prevalência seria mais adequado, porque os intervalos de confiança tenderiam a ser menores. Do contrário, prevalência baixa, também a seriam as diferenças entre razões de chances e de prevalência e erros padrão (Francisco *et al.*, 2008).

Em fenômenos raros, a razão de chance se aproximaria do risco relativo (probabilidade de ocorrência do efeito em indivíduos expostos em relação aos não expostos), entretanto quando a raridade não seria manifestaria como característica, o risco relativo seria sobrestimado pela razão de chance, particularmente quando do emprego da modelagem logística múltipla e a existência de associação espúria pela influência de Y e X por determinada variável, denominada efeito de confusão, confundidor, variável de confusão ou confundimento (Fumo-dos-Santos e Ferreira, 2023; Souza *et al.*, 2024). Não obstante, a regressão de Poisson tenderia a fornecer estimativas mais robustas ao risco, porém com intervalos de confiança maiores. Consequentemente, imperativo seria considerar inadequada a interpretação da razão de chance como risco relativo ou razão de prevalência, salvo se o efeito desejado for raro (Aguiar e Nunes, 2013).

Respeitando a ausência de aplicabilidade do Modelo B, as próximas análises serão limitadas ao A. Inicialmente, buscou-se ratificar a não violação de $E(Y) \cong \text{var}(Y)$, para tanto foi implementado o Teste de Equidispersão (Tabela 3), para garantir que os dados tinham dispersão similar (Tadano, Ugaya e Franco, 2009), tendo por desenho: H0: equidispersão e H1: superdispersão (variância > média) ou subdispersão (variância < média), portanto

o resultado de interesse era a ausência de diferença estatística, tal como ocorrido, valor-p = 0,56. Valeria ressaltar que similaridade de dispersão entre $E(Y)$ e $Var(Y)$ havia sido investigada na Tabela, $DR/GL = 0,97$, estimativa sutilmente diferente da Verdadeira Dispersão = 0,93, porém guardando a interpretação semântica. A pormenorização dos resultados no contexto inferencial dotaria de robustecimento essa escolha em detrimento da razão anterior.

Tabela 3: Resultados do Teste de Equidispersão com Média ($\bar{x}$), Variância (σ^2), Estatística Teste (Z), Valor-p e Verdadeira Dispersão (Dispersão), $\alpha = 0,05$.

Estatística	$\bar{x}$	σ^2	Z	valor-p	Dispersão
Estimativa	17,34	16,96	-0,58	0,56	0,93

Fonte: Os Autores (2025).

Não raramente, ocorreria a ofensa à propriedade básica, quando da utilização de dados experimentais (Ribeiro, 2006), então a modelagem poderia, de acordo com o objetivo, utilizar a *double* Poisson se houvesse coletas longitudinais (Batista, 2020). Na presença de superdispersão, o modelo binomial negativo poderia ser empregado (Riberiro Júnior, 2016), assim como as distribuições compostas seriam adequadas, especialmente com modelos hierárquicos ou em dois estágios, porque associariam uma distribuição à resposta condicionada à média e outra ao parâmetro de média, o que poderia ser implementado pelas distribuições Poisson-Gama, Poisson-Inversa Gaussiana e a Poisson-Lindley Generalizada (Duarte, 2021).

O modelo Conway-Maxwell-Poisson poderia ser empregado na constatação de sub ou superdispersão, exigindo a adição de parâmetro, logo garantiria a estimativa paramétrica. Outrossim, a existência de correlação entre as variáveis ou grande quantidade de valores nulos não inviabilizaria a adoção da modelagem ora proposta, todavia exigiria atenção à seleção de efeitos fixos ou aleatórios (Ribeiro Júnior, 2016).

A Análise de Desvios (Tabela 4) demonstrou que o Modelo A seria similar ao Nulo, valor-p = 0,18. Nesse não haveria qualquer preditor, então avaliado foi se a distribuição marginal na variável dependente seria estatisticamente possível, dito de outra maneira, o Modelo Nulo assumiu que o valor de Presentes em determinado instante seria estabelecido pela média de todos os momentos do banco de dados (Paes e Blinder, 1995; César e Soares, 2001).

Essencialmente, apesar de necessária, não seria suficiente a análise limitada às métricas do modelo, estudar o ganho de informação conquistados com os resultados

estimados seria fundamental à prospecção (Martins *et al.*, 2024; Carvalho, Dantas e Hernandez, 2024; Lima, Silva e Alencar, 2024). No estudo corrente o ganho seria nulo, portanto não alicerçaria a tomada de decisão, o que equivaleria a não reduzir a incerteza sobre o futuro.

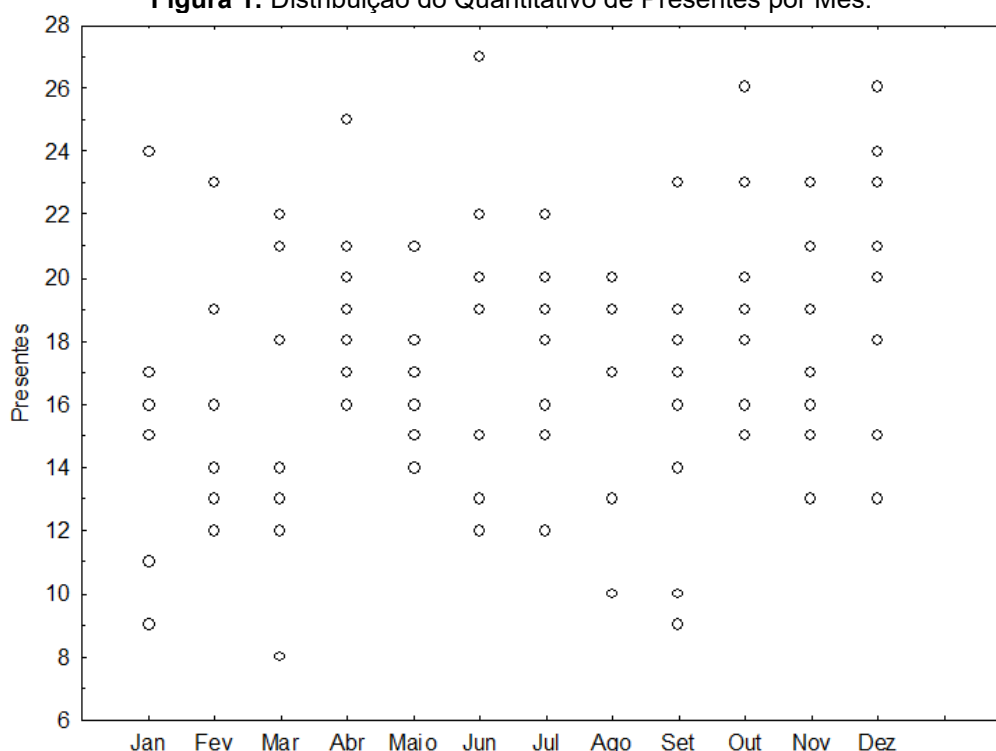
Tabela 4: Resultados da Análise de Desvios, $\alpha = 0,05$.

Fonte	GL	Desvio Residual	GL	Estatística Teste	Valor-p
Modelo Nulo			95,00	94,60	
Nível	3,00	4,91	92,00	89,69	0,18

Fonte: Os Autores (2025).

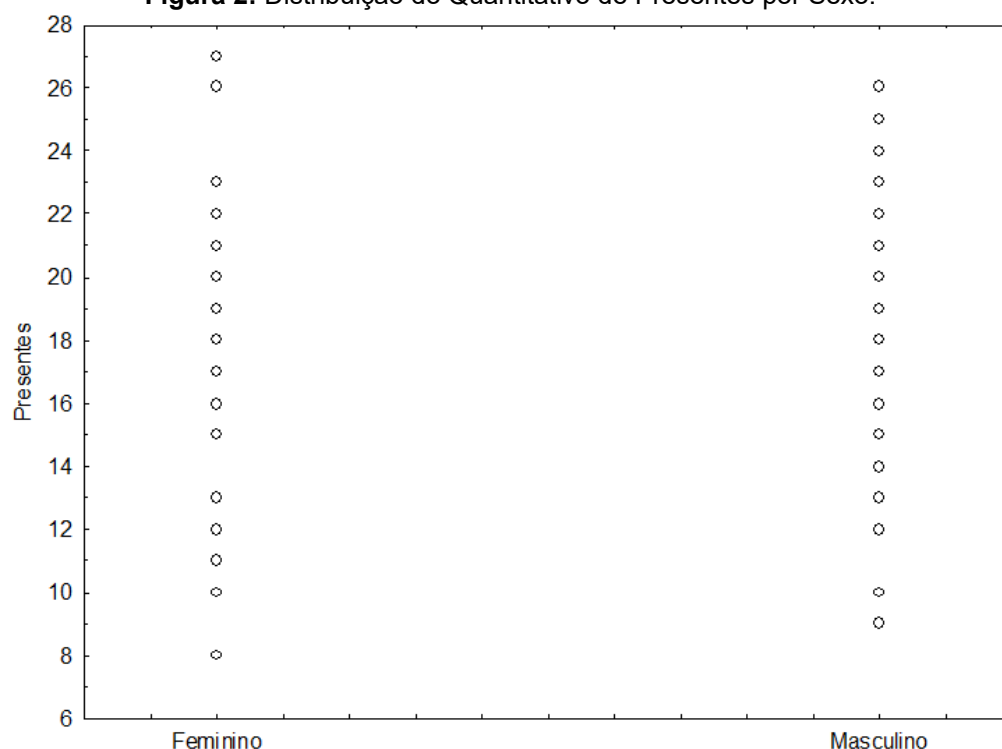
Isso pode ter ocorrido, pois considerando: $E(Y) \pm 2\sqrt{VAR(Y)} \cong 17,34 \pm 2.4,12 \cong [9,11; 25,58]$, nesse intervalo haveria 90 ocorrências (93,75%), excedendo-o somente os pontos: {(março, 8), (janeiro, 9), (setembro, 9), (outubro, 26), (dezembro, 26), (junho, 27)}, demonstrando uniformidade de presença ao longo dos meses (Figura 1). Característica igualmente estabelecida em Sexo (Figura 2): {(Feminino, 8), (Masculino, 9), (Masculino, 9), (Masculino, 26), (Feminino, 26), (Feminino, 27)}. E, divergindo da expectativa, presente em Nível (Figura 3): {(1, 8), (2, 9), (2, 9), (1, 26), (3, 26), (3, 27)}.

Figura 1: Distribuição do Quantitativo de Presentes por Mês.



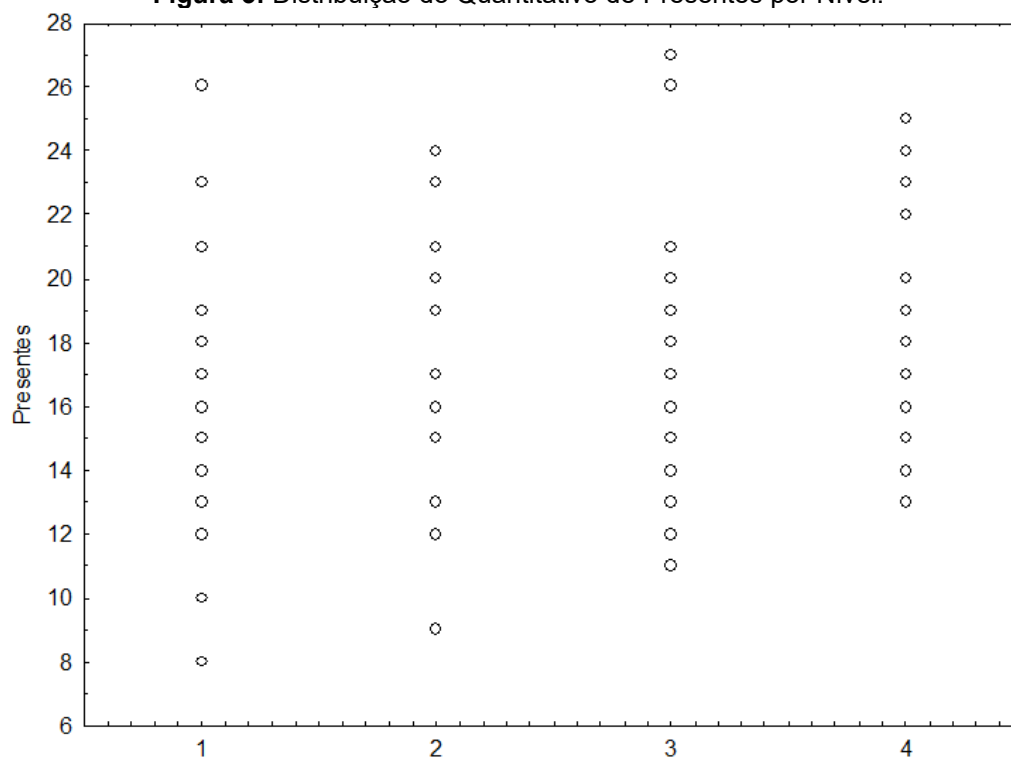
Fonte: Os Autores (2025).

Figura 2: Distribuição do Quantitativo de Presentes por Sexo.



Fonte: Os Autores (2025).

Figura 3: Distribuição do Quantitativo de Presentes por Nível.



Fonte: Os Autores (2025).

Em suma, os dados eram uniformes, ratificando as métricas expostas por DR/GL (Tabela 1), sempre próximas à unidade, e a Verdadeira Dispersão = 0,93 (Tabela 3). Resultados esses ratificadores da: 1) não violação de requisitos e 2) alta qualidade do ajuste, mas sugestivos de não aplicabilidade do modelo, tal como ratificado pela Análise de Resíduos (Tabela 4).

Pelo discutido, razoável seria supor que a assiduidade dos clientes pudesse estar relacionada a variáveis não aferidas como o objetivo da instituição, o qual influenciaria na cobrança por resultados ou rendimentos. Assim como o método de ensino ou a dinâmica de treinamento (Casarin e Cella, 2008). Particularmente, no referente ao Nível 4, a possibilidade de existência de metas de médio ou longo prazo poderia influenciar a consistência dos clientes nas sessões de treino, desde que essas fossem entendidas como o meio ao alcance das primeiras (Manzini, Freitas e Brandão, 2024).

De forma geral, a presença às intervenções poderia ser influenciada por variáveis não associadas ao treinamento e, portanto, distantes do controle da escola esportiva. O deslocamento do cliente envolveria o meio de transporte, a distância a ser percorrida e o tempo demandado. A mobilidade urbana não seria uniforme, ou seja, os diversos modais apresentariam distribuição e condições de acesso distintos, assim como o desempenho estaria intensamente associado à conjuntura de tráfego (condições atmosféricas, estado das vias e intervalo entre composições, por exemplo). O transporte particular, independentemente de motorização, poderia se submeter à similar conjunto de intervenientes, além da condição de estacionamento (Olekszechen e Kuhnen, 2016; Murta, 2017; Spignardi, 2019; Souza, Carvalho e Castañon, 2022; Moura e Rodrigues, 2023; Hannas, Moreira e Cruz, 2024; Alvim, Izaga e Claps, 2024; Reis Vêras, 2024).

A distância percorrida poderia caracterizar dificultador, dado que não necessariamente a geodésia significaria percorrer trajeto curto (Blitzkow *et al.*, 2011; Guimarães e Blitzkow, 2011) e, somente por coincidência, representaria a braquistócrona (Andrade e Ferreira Filho, 2015; Reis, 2020). Então, a análise da distância e do tempo do trajeto configurariam variáveis distintas e, na realidade tridimensional, independentes.

Considerações Finais

Objetivando prever o comparecimento de clientes às intervenções em uma escola esportiva, desenvolveu-se o modelo de Poisson, o qual foi significativo, porém sem

capacidade de elevar o ganho de informação. Então, concluiu-se que para o banco de dados utilizado não foi possível estimar modelo que reduzisse a incerteza.

Aos estudos futuros recomenda-se replicar a investigação, todavia utilizando a modelagem hierárquica, que poderá ter função prospectiva. Utilizar-se banco de dados com pelo menos dois anos de coleta poderá contribuir a diferenciação frente ao modelo nulo. Por último, a aplicação de métodos de aprendizagem de máquina tenderia a favorecer a acurácia, consequentemente o afastamento do modelo nulo.

Referências

- AGUIAR, P; NUNES, B. Odds ratio: reflexão sobre a validade de uma medida de referência em epidemiologia. **Acta Médica Portuguesa**, v. 26, n. 5, p. 505-510, 2013.
- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control AC**, v. 19, n. 6, p. 716-723, 1974.
- ALVIM, ATB; IZAGA, FG; CLAPS, RF. Mobilidade urbana em perspectiva: novos olhares sobre as dinâmicas da cidade contemporânea. **Cadernos Metrópole**, v. 26, n. 60, p. 413-421, 2024.
- ANDRADE, MA; FERREIRA FILHO, LG. Uma abordagem geométrica ao problema da braquistócrona. **Revista Brasileira de Ensino de Física**, v. 37, n. 2, 2309, 2015. DOI: <http://dx.doi.org/10.1590/S1806-11173721786>.
- BATISTA, DT. Modelos para dados de contagem não equidispersos com aplicação à ecologia e em estudos longitudinais. Tese (Doutorado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura “Luiz de Queiroz”. Universidade de São Paulo. Piracicaba (SP), 2020.
- BATISTA, TS *et al.* Homofobia internalizada e depressão em mulheres e homens homossexuais e bissexuais: Inquérito de saúde LGBT+, 2020. **Ciência & Saúde Coletiva**, v. 29, e05412023, 2024.
- BLEIDORN, MT *et al.* Investigação do efeito do uso de metodologias de imputação de dados faltantes no desempenho do modelo SARIMA: aplicação para vazões médias mensais. **Revista Brasileira de Recursos Hídricos**, v. 29, e18, 2024.
- BLITZKOW, D *et al.* O conceito atual dos referenciais usados em geodésia. **Revista Brasileira de Cartografia**, v. 63, n. 5, p. 633-648, 2011.
- BURNHAM, KP; ANDERSON, DR. Multimodel inference: understanding AIC and BIC in model selection. **Sociological Methods and Research**, 33, p. 261–304, 2004.
- CARVALHO, R; DANTAS, J; HERNANDEZ, J. Oldenburg Burnout Inventory com Policiais Militares de UPPs do Rio de Janeiro: Propriedades Psicométricas. **Revista Ciência & Polícia**, v. 10, n. 2, p. 243-264, 2024.
- CASARIN, RV; CELLA, MT. **Análise das sessões de treinamento de futebol entre categorias de base e escolinhas**. Relatório de Pesquisa (Graduação em Educação Física) – Universidade do Oeste de Santa Catarina. Xanxerê (SC), 2008.
- CÉSAR, CC; SOARES, JF. Desigualdades acadêmicas induzidas pelo contexto escolar. **Revista Brasileira de Estudos de População**, v. 18, n. 1/2, p. 97–110, 2001.
- COUTINHO, LMS; SCAZUFCA, M; MENEZES, PR. Métodos para estimar razão de prevalência em estudos de corte transversal. **Revista de Saúde Pública**, v. 42, n. 6, p. 992-998, 2008.

DALGAARD, P. **ISwR: Introductory Statistics with R**. R package version 2.0-9, 2024. <https://CRAN.R-project.org/package=ISwR>.

DELPINO, FM *et al.* Multimorbidade e uso de serviços de saúde em população diagnosticada com Covid-19 em município da região Sul do Brasil, 2020-2021: estudo transversal. **Epidemiologia e Serviços de Saúde**, v. 33, e2023915, 2024.

DUARTE, CG. **Uso das distribuições Poisson, Poisson-Gama, Poisson-Inversa Gaussiana e Poisson-Lindley generalizada para dados de contagem**. Trabalho de Conclusão de Curso (Bacharelado em Estatística) – Departamento de Estatística e Matemática Aplicada. Centro de Ciências. Universidade Federal do Ceará. Fortaleza (CE), 2021.

FERNANDES, AAT *et al.* Read this paper if you want to learn logistic regression. **Revista de Sociologia e Política**, v. 28, n. 74, e006, 2020.

FERNANDEZ, RN *et al.* Determinantes das Parcerias Público-Privadas na América Latina e no Caribe. **Perspectiva Econômica**, v. 13, n. 2, p. 86-99, 2017.

FERRO, WA; LIMA, JD; TRENTIN, MG. Combinações de Métodos Quantitativos na Previsão de Demanda de Vendas de Eletrodomésticos. **GEPROS. Gestão da Produção, Operações e Sistemas**, v. 14, n. 5, p. 67 - 88, 2019.

FOX, J; WEISBERG, S. **An R companion to applied regression**. Thousand Oaks CA (USA): Sage, 2019.

FRANCISCO, PMSB *et al.* Medidas de associação em estudo transversal com delineamento complexo: razão de chances e razão de prevalência. **Revista Brasileira de Epidemiologia**, v. 11, n. 3, p. 347-355, 2008.

FUMO-DOS-SANTOS, C; FERREIRA, JC. Lidando com fatores de confusão em estudos observacionais. **Jornal Brasileiro de Pneumologia**, v. 49, n. 4, e20230281, 2023.

GUIMARÃES, GN; BLITZKOW, D. Problema de valor de contorno da geodésia: uma abordagem conceitual. **Boletim de Ciências Geodésicas**, v. 17, n. 4, p.607-624, 2011.

GUJARATI, DN; PORTER, DC. **Econometria básica**. São Paulo: McGraw-Hill, 2011.

HANNAS, ASCD; MOREIRA, GC; CRUZ, A. Mobilidade urbana e desigualdades de gênero na cidade de São Paulo. **urbe. Revista Brasileira de Gestão Urbana**, v. 16, e20230141, 2024.

HOFFMANN, R. Odds ratio versus razão de prevalências ou modelo de logística versus regressão de Poisson. **Segurança Alimentar e Nutricional**, v. 23, n. 1, p. 816-825, 2016.

KIRCH, G. Determinantes da defasagem na divulgação das demonstrações contábeis das companhias abertas brasileiras. **Revista de Contabilidade & Finanças – USP**, v. 23, n. 60, p. 173-186, 2012.

KLEIBER, C; ZEILEIS, A. **Applied Econometrics with R**. New York (USA): Springer-Verlag, 2008.

LIMA, LS; SILVA, FR; ALENCAR, RCM. Desenvolvimento sustentável como fator impulsionador do desempenho empresarial. **Revista Pensamento Contemporâneo em Administração**, v. 18, n. 3, p. 180-195, 2024.

MALTA, DC *et al.* Cyberbullying entre escolares brasileiros: dados da Pesquisa Nacional de Saúde do Escolar, 2019. **Ciência & Saúde Coletiva**, v. 29, n. 9, e19572023, 2024.

MANZINI, MFP; FREITAS, IO; BRANDÃO, MRF. Relação entre garra e assiduidade nos treinamentos: um estudo de caso. **Revista Foco**, v. 17, n. 1, e4059, 2024.

MARTINS, T *et al.* A influência da proximidade com as matas ciliares e a distância das áreas urbanas nos atropelamentos de vertebrados em uma área fragmentada do Cerrado brasileiro. **Austral Ecology**, v. 49, n. 1, e13415, 2024.

MEYER, PL. **Probabilidade** – aplicações à estatística. Rio de Janeiro: LTC, 2022.

MOURA, LSL; RODRIGUES, SFF. **Análise dos fatores determinantes na escolha do meio de transporte na cidade de Marizópolis – PB**. Trabalho de Conclusão de Curso (Bacharelado em Engenharia Civil) - Instituto Federal de Educação, Ciência e Tecnologia da Paraíba. Cajazeiras (PB), 2023.

MURTA, IBD. Ressignificação do espaço urbano: a bicicleta como meio de transporte e lazer. **Revista Brasileira de Estudos do Lazer**, v. 4, n. 3, p. 116-138, 2017.

NOGUEIRA, ICR *et al.* Sífilis gestacional e congênita no Estado do Tocantins: análise de tendência por regressão de poisson, 2005 a 2018. **Brazilian Journal of Development**, v. 7, n. 8, p. 78091-78110, 2021.

OLEKSZECHEN, N; KUHNEN, A. Barreiras e facilitadores no uso da bicicleta como meio de transporte entre universitários. **INTERthesis: Revista Internacional Interdisciplinar**, v. 13, n. 2, p. 119-140, 2016.

PAES, ET; BLINDER, PB. Modelos nulos e processos de aleatorização: algumas aplicações em ecologia de comunidades. **Oecologia Brasiliensis**, v. 2, n. 1, p. 119-139, 1995.

POISSON, SD. **Recherches sur la probabilité des jugements en matière criminelle et en matière civile, precedees des règles générales du calcul des probabilités**. Paris (France): Bachelier, imprimeur-Libraire, 1837.

REIS, GM. **Sobre as curvas tautócrona e braquistócrona, teorema da desigualdade isoperimétrica e teorema dos quatro vértices**. Dissertação (Mestrado Profissional em Matemática) — Universidade de Brasília, Brasília, 2020.

RIBEIRO JÚNIOR, EE. **Extensões e aplicações do modelo de regressão Conway-Maxwell-Poisson para modelagem de dados de contagem**. Trabalho de Conclusão de Curso (Bacharelado em Estatística) - Universidade Federal do Paraná. Curitiba (PR), 2016.

RIBEIRO, AJF. **Um estudo sobre mortalidade dos aposentados por invalidez do regime geral da previdência social (RG PS)**. Tese (Doutorado em Demografia) - Centro de Desenvolvimento e Planejamento Regional. Faculdade de Ciências Econômicas. Universidade Federal de Minas Gerais. Belo Horizonte (MG), 2006.

ROBINSON, D; HAYES, A; COUCH, S. **broom: Convert Statistical Objects into Tidy Tibbles**. R package version 1.0.7, 2024. <https://CRAN.R-project.org/package=broom>.

RUSSO, S; CAMARGO, ME; SAMOHYL, RW. Gráficos de controle baseado nos resíduos do modelo de regressão Poisson. **Revista Produção On Line**, v. VIII, n. IV, 2008.

SCHÄFFER, YB. **Dia de jogo: uma análise da relação entre futebol e violência doméstica no brasil entre 2015 e 2019**. Dissertação (Mestrado em Economia) - Programa de Pós-Graduação em Economia do Desenvolvimento. Escola de Negócios. Pontifícia Universidade Católica do Rio Grande do Sul. Porto Alegre (RS), 2024.

SHIBATA, R. An optimal selection of regression variables. **Biometrika**, v. 68, n. 1, p. 45-54, 1981.

SILVA, CS. **Efeitos da diversificação e da experiência do franqueado na expansão de suas lojas: evidências do varejo de vestuário e calçados**. Dissertação (Mestrado Profissional em Administração de Empresas) – Insper Instituto de Ensino e Pesquisa, São Paulo, 2014.

SOARES, PM; GUIMARÃES, DA; SOUZA, RAA. Estimação do número de fontes utilizando métodos baseados na teoria da informação. **Revista Telecomunicações**, v. 15, n. 02, p. 93-102, 2013.

SOUZA, CC; ARAÚJO, TM. Efeitos combinados de gênero, raça e estressores ocupacionais na saúde mental. **Revista Brasileira de Saúde Ocupacional**, v. 49, p. edepi12, 2024.

SOUZA, DM *et al.* Prática da imersão em água quente no alívio da dor em neonatologia: revisão integrativa. **Revista Brasileira de Enfermagem**, v. 77, n. 1, p. e20230260, 2024.

SOUZA, LM; CARVALHO, YM; CASTAÑON, JAB. A popularidade do skate durante os Jogos Olímpicos refletirá no seu uso como meio de transporte urbano no Brasil? **Research, Society and Development**, v. 11, n. 3, e14411326203, 2022.

SPIGNARDI, MCM. **A influência do microclima urbano na decisão pelo uso da bicicleta como meio de transporte no centro expandido da Cidade de São Paulo**. Dissertação (Mestrado em Arquitetura e Urbanismo) - Faculdade de Arquitetura e Urbanismo. Universidade de São Paulo. São Paulo, 2019.

TADANO, YS; UGAYA, CML; FRANCO, AT. Método de regressão de Poisson: metodologia para avaliação do impacto da poluição atmosférica na saúde populacional. **Ambiente & Sociedade**, v. XII, n. 2, p. 241-255, 2009.

THAMER, R; LAZZARINI, SG. Projetos de parceria público-privada: fatores que influenciam o avanço dessas iniciativas. **Revista de Administração Pública**, v. 49, n. 4, p. 819-846, 2015.

VARGAS, EB. **Racismo, ambiente alimentar e insegurança alimentar em adultos de uma capital do sul do Brasil**. Trabalho de Conclusão de Curso (Bacharelado em Nutrição) – Faculdade de Medicina. Universidade Federal do Rio Grande do Sul. Porto Alegre (RS), 2024.

WICKHAM, H *et al.* Welcome to the tidyverse. **Journal of Open Source Software**, v. 4, n. 43, article 1686, 2019. <https://doi.org/10.21105/joss.01686>.

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. New York (USA): Springer-Verlag, 2016.

WICKHAM, H; BRYAN, J. **readxl: Read Excel Files**. R package version 1.4.3, 2023. <https://CRAN.R-project.org/package=readxl>.