

BREVE NOTA SOBRE O NÚMERO DE USUÁRIOS DA RSP SEGUNDO O GOOGLE ANALYTICS

Victor Schmidt Comitti

IF Sudeste MG - Campus Bom Sucesso – MG, Brasil

O número de visitantes do site de um periódico pode ser utilizado como uma *proxy* de seu impacto. Nesta nota utilizamos o *Product Partition Dynamic Generalized Linear Model* para localizar possíveis pontos de mudança na série de visitantes do site da *Revista do Serviço Público* (RSP) ao longo do período que vai de janeiro de 2017 até abril de 2022. Nossos resultados apontam para a existência de doze quebras estruturais na série: seis positivas e seis negativas. As quebras positivas, associadas a um aumento no fluxo de visitantes, podem ser atribuídas ao lançamento de novas edições. Verificamos que duas delas geraram uma mudança permanente no número de visitantes da página da RSP. Já os pontos de mudança negativo, em geral, representam um movimento de retorno à média depois de um aumento inesperado de acessos no site da revista. Esperamos que a determinação dos marcos editoriais da RSP possa auxiliar o corpo editorial da revista a identificar temas e autores de maior interesse para os leitores.

Palavras-chave: séries temporais, pontos de mudança, *Product Partition Models*, *Dynamic Generalized Linear Models*, marcos editoriais



BREVE NOTA SOBRE EL NÚMERO DE USUARIOS DE RSP SEGÚN GOOGLE ANALYTICS

El número de visitantes al sitio web de una revista se puede utilizar como indicador de su impacto. En esta nota, utilizamos el *Product Partition Dynamic Generalized Linear Model* para ubicar posibles puntos de inflexión en la serie de visitantes del sitio web de la *Revista do Serviço Público* (RSP) durante el período comprendido entre enero de 2017 y abril de 2022. Nuestros resultados apuntan a la existencia de doce rupturas estructurales en la serie: seis positivas y seis negativas. Las rupturas positivas, asociadas a un aumento en el flujo de visitantes, se pueden atribuir al lanzamiento de nuevas ediciones. Encontramos que dos de ellos generaron un cambio permanente en el número de visitantes a la página RSP. Por otro lado, los puntos de cambio negativos, en general, representan un movimiento de regreso al promedio luego de un aumento inesperado en las visitas al sitio web de la revista. Esperamos que la determinación de los hitos editoriales de RSP pueda ayudar al consejo editorial de la revista a identificar temas y autores de mayor interés de los lectores.

Palabras-clave: series temporales, puntos de inflexión, *Product Partition Models*, *Dynamic Generalized Linear Models*, hitos editoriales

BRIEF NOTE ON THE NUMBER OF RSP USERS ACCORDING TO GOOGLE ANALYTICS

The number of visitors to a journal's website can be used as a proxy for its impact. In this note we use the *Product Partition Dynamic Generalized Linear Model* to find possible change-points in the time series of visitors to the *Revista do Serviço Público* (RSP) website over the period from January 2017 to April 2022. Our results point to the existence of twelve structural breaks in the series: six positives and six negatives. The positive breaks, that is, associated with an increase in the flow of visitors, can be attributed to the release of new editions. We also verified that two of these breaks generated a permanent change in the number of visitors to the RSP's page. On the other hand, the negative change points, in general, represent a mean reversion after an unexpected increase in hits on the magazine's website. We hope that the determination of RSP's editorial milestones can help the journal's editorial board to identify themes and authors of greatest interest to readers.

Keywords: time series, change-points, *Product Partition Models*, *Dynamic Generalized Linear Models*, editorial milestones

1. INTRODUÇÃO

Uma das métricas utilizadas para se avaliar o impacto de um periódico científico é o número de acessos em sua página na internet ao longo do tempo (MINKERS; LEYDESDORFF, 2015). Esse número pode oscilar devido a diversos fatores tais como o lançamento de uma edição especial ou a publicação de um artigo de grande repercussão. Conhecer e monitorar esse comportamento pode ajudar os gestores do periódico a entender as preferências de seu público-alvo de modo a permitir que a linha editorial da publicação seja ajustada caso se identifique a necessidade.

O objetivo principal desta nota é identificar os marcos editoriais da *Revista do Serviço Público* (RSP) no passado recente e verificar se esses eventos levaram a uma mudança permanente no nível de interesse pela RSP ou se houve apenas um crescimento transitório do número de usuários do site seguido por um retorno à média anterior.

Podemos identificar esses pontos de mudança (ou quebras estruturais) com auxílio das técnicas oferecidas pela econometria de séries temporais. Neste artigo utilizamos o *Product Partition Dynamic Generalized Linear Model* (daqui em diante denotado por PPDGLM) proposto em Comitti *et al.* (2021) para estimar probabilidades a posteriori da ocorrência de quebras estruturais na série temporal de visitantes da página na internet da *Revista do Serviço Público* (RSP). Os resultados obtidos indicam ao menos doze pontos de mudança na série de visitantes do site da RSP nos últimos cinco anos. Além disso, verifica-se uma tendência de crescimento no número de usuários da revista.

Essa nota está estruturada da seguinte maneira: a Seção 2 apresenta a base de dados e introduz a técnica econométrica utilizada para se detectar pontos de mudança. A discussão sobre o PPDGLM é sucinta e detalhes técnicos a respeito do modelo são omitidos uma vez que o foco está na análise exploratória da série temporal de visitantes do site da RSP. A seção 3 mostra os principais resultados obtidos e as discussões pertinentes. A Seção 4 fica reservada para as considerações finais.

2. BASE DE DADOS E ESTRATÉGIA EMPÍRICA

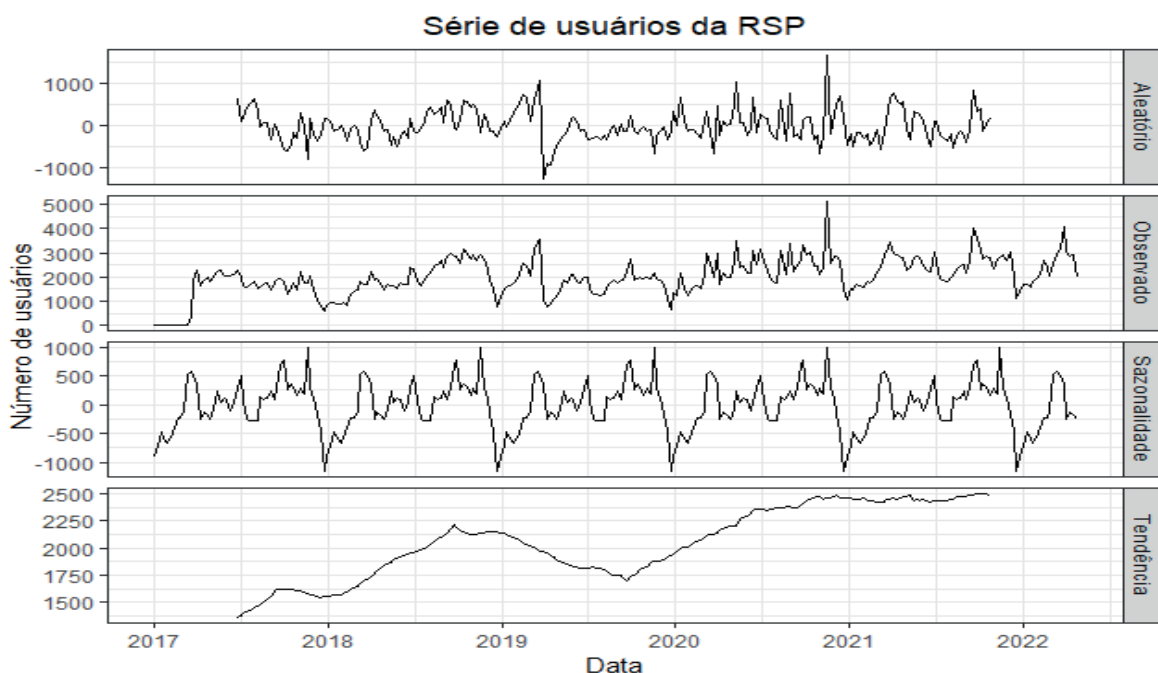
Neste exercício utilizamos observações diárias do número de visitantes no site da RSP coletados através do *Google Analytics*. Para diminuir a variabilidade dos dados, eles foram agregados por semana. A base final utilizada é composta por 278 observações que formam uma série temporal com início em 01/01/2017 e fim em 24/04/2022.

Uma análise descritiva da série temporal mostra que o site RSP teve 2015 usuários por semana em média durante período analisado. Para verificar se existe uma tendência de crescimento desse número ao longo do tempo podemos decompor a série em suas componentes constituintes. A Figura 1 mostra o resultado de tal decomposição. A implementação foi feita em

R¹ através da função *decompose* considerando um modelo aditivo (para mais detalhes sobre o método ver Kendall & Stuart (1983)).

Os painéis da Figura 1 mostram, respectivamente, o ruído aleatório, os dados observados, a sazonalidade e a tendência da série temporal. O padrão de sazonalidade indica que o site da RSP perde usuários nos meses finais do ano e os recupera ao longo do primeiro trimestre do ano seguinte. Já o painel de tendência mostra um crescimento da base de usuários da revista no período 2017-2022. Esse crescimento ocorreu de maneira mais acentuada em dois períodos específicos: nos três primeiros trimestres de 2018 e do segundo semestre de 2019 até o final de 2020. Entre o final de 2018 e o primeiro semestre de 2019 observa-se uma tendência de diminuição do número de visitantes da RSP. Já de 2021 até o final do período analisado o nível da série temporal permaneceu estável.

Figura 1 – Decomposição da série temporal de usuários da RSP



Fonte: elaboração própria.

Uma inspeção visual do painel com os dados observados sugere alguns picos de visitação do site da revista. Para identificar esses possíveis pontos de mudança na série utilizamos o PPDGLM. Esse modelo combina os *Product Partition Models* (PPM) propostos em Barry & Hartigan (1992) com o *Dynamic Generalized Linear Model* (DGLM) de West, Harrison & Migon (1985). Essa formulação mantém a estrutura inferencial Bayesiana do DGLM – com passos sequenciais de evolução e atualização dos vetores de estado do modelo – ao mesmo tempo em que permite partições aleatórias dos dados. Esse modelo admite observações pertencentes às

¹ Os códigos podem ser disponibilizados pelo autor, a pedido.

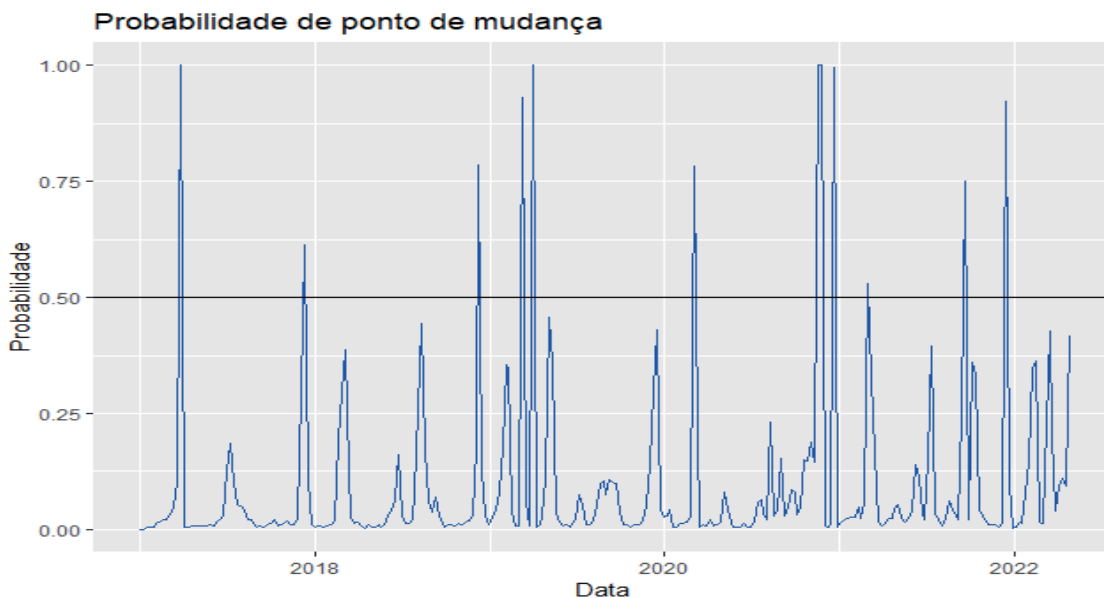
distribuições da Família Exponencial e possibilita ao usuário fazer tanto a análise retrospectiva do processo como previsões ‘k’ passos à frente.

Uma discussão aprofundada do PPDGLM está fora do escopo deste trabalho. O leitor interessado pode consultar Comitti *et al.* (2021). Nesta nota utilizamos um PPDGLM-Normal com distribuição *a priori* Beta (1,1) para o fator de desconto e coesão *a priori* definida de acordo com Yao (1984) com *priori* Beta (1,500) para a probabilidade instantânea de ocorrência de um ponto de mudança. Para a estimação foi utilizado o esquema de amostragem de Gibbs proposto em Loschi & Cruz (2002) com um passo de *Adaptive Rejection Metropolis Sampling* (ARMS) para o cálculo do fator de desconto ótimo do modelo.

3. RESULTADOS E DISCUSSÃO

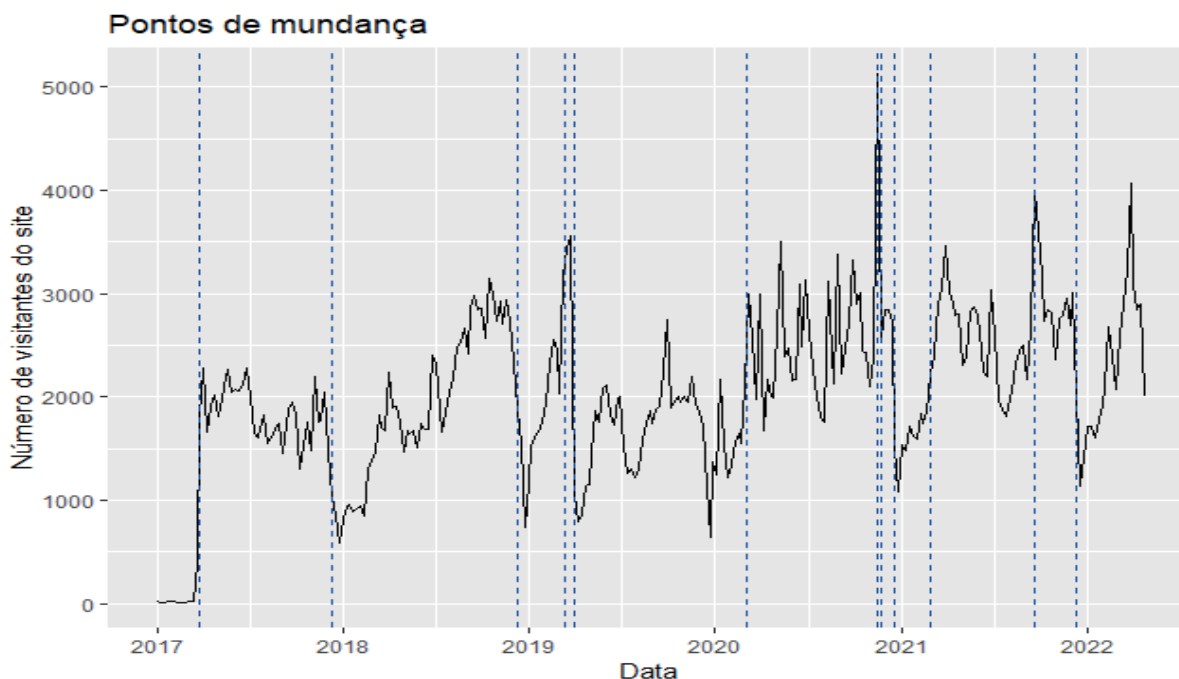
Utilizando o PPDGLM com as especificações descritas na seção anterior foi possível estimar probabilidades *a posteriori* da ocorrência de uma quebra estrutural para cada ponto da série temporal analisada. Os resultados obtidos podem ser vistos na Figura 2. Serão consideradas quebras estruturais apenas aqueles pontos nos quais a probabilidade estimada de mudança ultrapassa 50%. Sendo assim, nosso método detectou doze possíveis datas marcantes na história recente da RSP, assinaladas com uma linha tracejada na Figura 3.

Figura 2 – Probabilidade de ponto de mudança. A linha horizontal indica probabilidade de 50%



Fonte: elaboração própria.

Figura 3 – Pontos de mudança identificados. As linhas tracejadas indicam as datas em que as quebras ocorreram



Fonte: elaboração própria.

Dos pontos de mudança identificados, seis refletem impactos positivos na série temporal; ou seja: estão associados a um aumento no número de leitores da RSP. Destes seis marcos, é possível apontar ao quatro deles como sendo resultado do lançamento de uma nova edição da revista. Observa-se ainda que os impactos negativos ocorrem no mês de dezembro – seguindo o padrão de sazonalidade identificado na Figura 1 -- ou algumas semanas após o lançamento de um novo número, sugerindo que há um movimento rápido de reversão à média da série temporal. Analisando o painel de tendência da Figura 1, é possível perceber que, ao longo do período analisado, as únicas quebras estruturais que parecem ter causado uma mudança permanente no nível de visitantes do site da RSP são aquelas ocorridas nas semanas de 26/03/2017 e 01/03/2020. A primeira se refere ao lançamento da primeira edição de 2017 e a segunda ao lançamento de uma edição especial sobre governança pública.

O Quadro 1, abaixo, apresenta um resumo das datas em que o modelo detectou um ponto de mudança juntamente com o tipo de mudança observada (positiva ou negativa) e a indicação relativa ao lançamento, ou não, de uma nova edição naquela semana.

Quadro 1 – Pontos de mudança identificados

Semana da quebra	Tipo de mudança	Nova edição?
26/03/2017	Positiva	Sim
10/12/2017	Negativa	Não
09/12/2018	Negativa	Não
10/03/2019	Positiva	Sim
31/03/2019	Negativa	Não
01/03/2020	Positiva	Sim
15/11/2020	Positiva	Não
22/11/2020	Negativa	Não
20/12/2020	Negativa	Não
28/02/2021	Positiva	Não
19/09/2021	Positiva	Sim
12/12/2021	Negativa	Não

Fonte: elaboração própria.

Os dois pontos de mudança positiva que não estão associados ao lançamento de uma nova edição ocorreram nas semanas de 15/11/2020 e 28/02/2021. O primeiro pode estar relacionado ao lançamento, nos dois meses imediatamente anteriores, da terceira edição da RSP em 2020 e de um número especial sobre desenvolvimento; já o segundo não pode ser associado a nenhuma decisão editorial da RSP. Dentre todos os pontos de mudança detectados, esse é o que possui a menor probabilidade a posteriori (53%). A especificação do PPDGLM é sensível à coesão *a priori* escolhida e, em especial, aos hiperparâmetros da distribuição Beta utilizada como *priori* para a probabilidade instantânea de ocorrência de um ponto de mudança. Dessa forma, uma hipótese é que de essa observação seja um falso positivo do modelo e não represente uma quebra estrutural na série temporal.

4. CONSIDERAÇÕES FINAIS

Essa nota teve como objetivo identificar e entender as quebras estruturais na série temporal de visitas do site da RSP. Foram encontrados doze possíveis pontos de mudança nessa série: seis negativos e seis positivos. Verificamos que os pontos de mudança negativos, na maior parte das vezes, estão relacionados a um movimento de retorno à média depois de um pico inesperado de visitantes na página da revista. Parece haver ainda um padrão de sazonalidade que leva a uma queda no fluxo de visitantes do site em dezembro. Em alguns anos, o modelo utilizado entende esse padrão como uma quebra estrutural da série.

Em relação aos pontos positivos, nosso estudo mostra que o lançamento de novas edições da revista gera um aumento não esperado nas visualizações da página da RSP. Esse efeito parece ser relevante também quando há o lançamento de uma edição especial (os pontos de mudança de 19/09/2021 e 01/03/2020, por exemplo). Esses picos de visitação tendem a durar pouco tempo e, em geral, não resultam em uma mudança permanente no nível da série. As exceções são as quebras identificadas nas semanas de 26/03/2017 e 01/03/2020.

De maneira geral, os resultados obtidos podem ser utilizados para identificar os assuntos e autores mais relevantes para os leitores da RSP e, eventualmente, preparar edições especiais com temas que despertam maior interesse do público da revista.

REFERÊNCIAS

- BARRY, D.; HARTIGAN, J. A. Product Partition Models for Change Point Problems. *The Annals of Statistics*, v.20, n.1, p.260–279, 1992.
- COMITTI, V. S.; DEMARQUI, F. N.; DOS SANTOS, T. R.; ALMEIDA, J. da A. *Product Partition Dynamic Generalized Linear Models*. 2021. Disponível em: <<http://arxiv.org/abs/2103.02470>>. Acesso em: 11/3/2021.
- KENDALL, M.; STUART, A. *The Advanced Theory of Statistics*. Distribution Theory. Griffin, London, v. 3, p. 410-414, 1983.
- LOSCHI, R.; CRUZ, F. Applying the Product Partition Model to the Identification of Multiple Change Points. *Advances in Complex Systems (ACS)*. 05. p.371-387, 2002.
- MINKERS, J.; LEYDESDORFF, L. A review of theory and practice in scientometrics. *European journal of operational research*, v.246, n.1, p.1-19, 2015.
- WEST, M.; HARRISON, P. J.; MIGON, H. S. Dynamic Generalized Linear Models and Bayesian Forecasting. *Journal of the American Statistical Association*, v. 80, n.389, p. 73–83, 1985.
- YAO, Y. Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches. *The Annals of Statistics* v.12, p.1434-1447, 1984.

Victor Schmidt Comitti

 <https://orcid.org/0000-0002-9734-3792>

Doutor em estatística pela UFMG. Professor do Instituto Federal Sudeste de Minas Gerais

– Campus Bom Sucesso

victorsch2@gmail.com