

Explicando Decisões: as Aplicações da Análise por Regressão Logística (*Logit*) no Estudo do Comportamento Judicial¹

JOSÉ MÁRIO WANDERLEY GOMES NETO

Doutor em Ciência Política pela Universidade Federal de Pernambuco (2015), Professor Assistente da Universidade Católica de Pernambuco.

LUIS FELIPE ANDRADE BARBOSA

Mestrado em Ciência Política pela Universidade Federal de Pernambuco (2012), Professor da Associação Caruaruense de Ensino Superior.

JORGE LUIZ GONZAGA VIEIRA

Doutorado em Letras pela Pontifícia Universidade Católica de Minas Gerais (2014), Professor da Faculdade de Tecnologia de Alagoas.

RESUMO: Como testar empiricamente hipóteses baseadas na influência de variáveis sobre uma decisão entre duas alternativas? A literatura oferece, para tanto, a regressão logística (*logit*) como o método apropriado para conduzir a análise de dependência quando as variáveis dependentes são de natureza categórica discreta binária, ainda mais quando estas são resultado de um processo de escolha. Este artigo apresenta essa ferramenta relevante de pesquisa, dirigindo-se principalmente para o público de Direito, enfrentando detalhes como identificação de problemas de pesquisa, codificação das variáveis, interpretação dos resultados e revisão de casos em que foi aplicada.

PALAVRAS-CHAVE: Estudos empíricos em Direito; regressão logística; *logit*; comportamento judicial.

ABSTRACT: How to empirically test hypotheses based on the influence of variables on a decision between two alternatives? The literature offers logit regression as the appropriate method to conduct dependency analysis when the dependent variables are of binary categorical nature, especially when they are the result of a choice process. This article presents this relevant research tool, aimed mainly at the legal research public, addressing details such as identification of research problems, coding of variables, interpretation of results and review of cases in which it was applied.

KEYWORDS: Empirical studies in Law; logistic regression; logit; behavior.

SUMÁRIO: Introdução; 1 O que é a regressão logística (*logit*)?; 2 Da coleta dos dados à organização do modelo: estratégia empírica da análise por regressão logística; 3 Como interpretar os resultados?; 4 Quais as aplicações mais comuns da ferramenta?; 4.1 Aplicação no estudo do comportamento judicial: literatura estrangeira; 4.2 Aplicação no estudo do comportamento judicial: literatura brasileira; Conclusões; Referências.

1 Este artigo é dedicado ao Prof. Dr. Enivaldo Carvalho Rocha (UFPE). Os autores agradecem ao Prof. Dr. Flávio Rezende (UFPE), à Profª Dra. Mariana Batista da Silva (UFPE) e ao Prof. Dr. Manoel Leonardo Santos (UFMG) pela leitura de versões iniciais do trabalho e pelas suas inestimáveis contribuições. Da mesma maneira, agradecem as sugestões e as contribuições dos pareceristas anônimos, ressaltando que qualquer imprecisão remanescente é de responsabilidade exclusiva dos autores.

For the rational study of the law the blackletter man may be the man of the present, but the man of the future is the man of statistics and the master of economics. (Oliver Wendell Holmes)

INTRODUÇÃO

Quais fatores influenciam na concessão ou na revogação de uma prisão preventiva? O que pesa mais na hora de um juiz de primeiro grau decidir sobre um pedido de liminar para fornecimento de medicamento? Em que situações há maior ou menor probabilidade de sucesso na interposição de um agravo de instrumento? Quando são maiores as chances de um cidadão obter a concessão de um *habeas corpus* ou de um mandado de segurança?

Essas questões possuem um ponto comum: todas elas estão relacionadas ao grau de associação entre a presença de diversos fatores (variáveis numéricas, ordinais ou, principalmente, categóricas) e a probabilidade de uma decisão binária (um processo de escolha entre um sim e um não; entre negar ou conceder a liminar; entre prover ou desprover o recurso interposto; entre declarar constitucional ou inconstitucional uma lei etc.), ou de uma decisão complexa (em que há um processo de escolha entre três ou mais alternativas).

Para tanto, modelos formais explicativos do comportamento judicial (atitudinais, legalistas, estratégicos, behavioristas, entre outros) oferecem hipóteses testáveis, derivadas de sólidos fundamentos teóricos, que se traduzem em informações sobre cada caso, passíveis de codificação e, conseqüentemente, de transformação em variáveis categóricas explicativas, cuja provável relação com a tomada de decisão (variável dependente) é objeto de análise quantitativa e/ou qualitativa.

Neste sentido, a análise estatística por regressão logística (*logit*) aparece como uma adequada ferramenta de pesquisa para problemas que envolvam a relação entre variáveis categóricas – predominantes nos estudos sobre o Direito e sobre as instituições judiciais – e a provável influência destas em processos decisórios (nada mais que escolhas entre alternativas).

Este artigo tem por escopos apresentar as características da ferramenta, descrever suas potenciais aplicações e orientar o pesquisador atuante em estudos empíricos em Direito, desejoso de saber se a presença de determinada variável categórica exerce influência sobre o resultado, aumentando ou diminuindo as chances da tomada de certas decisões, tal como esperado no modelo formal adotado na pesquisa.

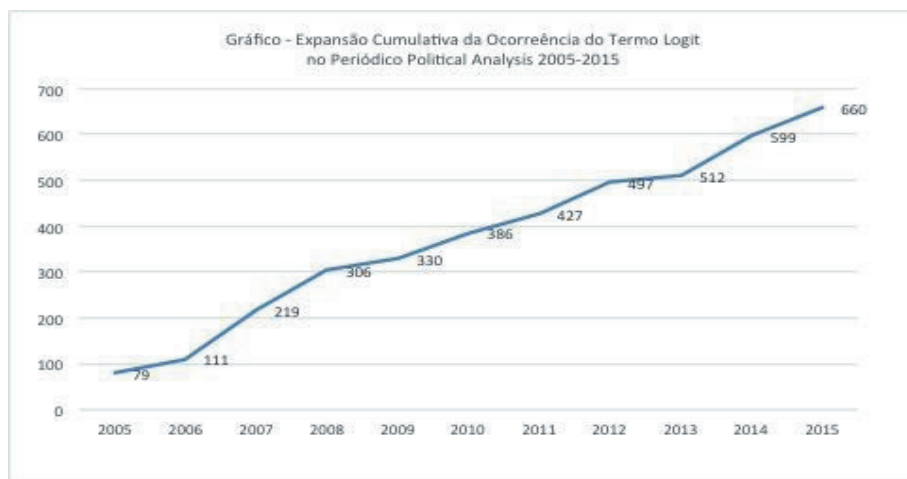
1 O QUE É A REGRESSÃO LOGÍSTICA (*LOGIT*)?

Logit modeling represents a breakthrough in the methodology of social science research because it offers ways to more efficient estimation of models with multiple, categorical variables. [...] [P]roper ap-

plication of logit techniques can strengthen inference. (Lewis-Beck, 1992, p. v)

A análise de dados por regressão logística – ou simplesmente *logit* – é apresentada como uma ferramenta adequada² e relevante para a compreensão de certos fenômenos sociais: quando se trata do estudo da relação entre variáveis categóricas por essência (como é o caso de quase todas as variáveis passíveis de extração e de codificação a partir da legislação, da jurisprudência, enfim, das instituições jurídicas e judiciais, de modo geral) e um resultado dicotômico (também chamado discreto na terminologia usual da ciência estatística), seja ele *sucesso* ou *insucesso*, *deferido* ou *indeferido*, *julgado* ou *não julgado*.

Rezende (2017), ao tratar do pluralismo inferencial em pesquisas empíricas, mostra a evolução cumulativa, no período compreendido entre 2005-2015, da ocorrência do termo “logit” na revista *Political Analysis*, que é a mais importante publicação metodológica da Ciência Política, servindo o dado para mostrar, de maneira mais ampla, como essa técnica tem adquirido maior latITUDE ao longo do tempo na pesquisa empírica de alto nível, alcançando relevantes problemas de pesquisa acerca das probabilidades de ocorrência de diversos fenômenos sociais e institucionais.



Fonte: Figura gerada por uma consulta aos dados de contagem realizada por Rezende no banco de dados da pesquisa sobre o Pluralismo Inferencial (2017).

2 Para outras situações que envolvam decisões mais complexas, entre três ou mais alternativas, e as possíveis associações entre estas e um conjunto de variáveis explicativas, há a previsão da regressão logística múltipla, outra interessante ferramenta, de aplicações relevantes, mas que excede o objeto deste trabalho (Levin et. al, 2012).

Tal método foi descoberto e primeiramente empregado, ainda no século XIX, pelo matemático belga Pierre-François Verhulst (1804-1849) em trabalhos de matemática pura, descrevendo o que viria a ser a sua função logística. Em seguida, foi empregado em 1920 em estudos empíricos sobre o crescimento populacional nos EUA, por Pearl e Reed, a partir de corolários da função logística extraída do trabalho de Verhulst. Posteriormente, a partir de 1980, por influência de estudos da Economia, tal ferramenta passou a ser utilizada em pesquisas sociais aplicadas: na Ciência Política, na Administração, na Sociologia e, ainda em caráter de desenvolvimento inicial, nos estudos empíricos em Direito (Demaris, 1992; Batistela *et al.*, 2009).

A regressão logística foi descoberta no século XIX para descrever o crescimento das populações e as reações químicas de autocatálise. A ideia básica do desenvolvimento logístico é simples e usada nos dias atuais para modelar o crescimento populacional e muitas outras técnicas, tais como descrever a chance de ocorrência de um determinado evento, ou seja, a chance de desistência de um aluno, a chance de um cliente de um banco vir a se tornar inadimplente, entre outras. (Batistela *et al.*, 2009)

Trata-se de uma espécie do gênero análise por regressão, conjunto de ferramentas estatísticas usuais em estudos empíricos, empregada quando a variável resposta for uma variável categórica binária (quando p tende a 0, $\text{logit}(p)$ tende a $-\infty$, e quando p tende a 1, $\text{logit}(p)$ tende a ∞), baseada na transformação logística (logarítmica) ou *logit* de uma proporção, sendo uma ferramenta metodológica empregada para estimar os valores empíricos dos parâmetros em modelos de resposta qualitativa (Everitt, 2006).

Todo processo de escolha possui, em si, uma característica aleatória, que se soma à racionalidade envolvida na apreciação das alternativas; assim, o modelo estatístico calcula as chances de o fenômeno pesquisado ocorrer e se essa ocorrência estaria, ou não, associada à presença das variáveis que se pretende testar na pesquisa.

Assim, exemplificativamente, em um modelo baseado em uma análise *logit* binária, se o resultado estatístico encontrado tiver sinal positivo (+), significa que a variável explicativa (independente) está associada ao aumento das chances de a primeira alternativa (1) ocorrer; por outro lado, se o resultado obtido tiver sinal negativo (-), a respectiva variável explicativa estaria associada à diminuição das chances do evento (1) ocorrer; destarte, se o resultado obtido for muito próximo a 0 (zero), significa ausência de associação ou uma associação muito baixa, sem qualquer interferência nas chances de ocorrer o evento sob predição no modelo.

Segundo Demaris (1992), a análise *logit* provê um modelo linear interpretável para uma resposta categórica, de vantajoso emprego para aqueles pesquisadores que lidam com o aludido tipo de variável na codificação de seus

dados: na modelagem *logit*, podemos expressar o logaritmo natural das chances de um fenômeno ocorrer, isto é, deve-se entender por *chances* a relativa probabilidade de o resultado recair sobre uma entre duas categorias da variável dependente (resposta).

Numa regressão logística com resposta binária, parâmetros são estimados pela maximização da função probabilística dos efeitos da amostra sobre a variável dependente. Esta função expressa a distribuição condicional da variável dependente, dado o conjunto de preditores [variáveis explicativas]. (*Idem*, p. 67)

Epstein e Martin (2014, p. 213-214) afirmam que, em se tratando da tarefa de explicar uma variável dicotômica (discreta), os modelos de regressão logística (*logit*) obtêm melhores resultados explicativos em comparação com os demais: “Ao invés de modelar o valor de Y_i , como se faz nas regressões lineares, [é modelada] a probabilidade de Y_i assumir o valor (1)”, assumindo que a relação entre as variáveis não é linear.

Já Lattin *et al.* (2011, p.13) sustentam que a regressão logística (*logit*) é “o método apropriado para conduzir a análise de dependência quando as variáveis dependentes são de natureza discreta [binária]”, pois, quando estas “são resultado de um processo de escolha”, é possível construir “um modelo de escolha para criar uma função não linear que descreva a probabilidade da escolha”, sendo possível calcular as chances de cada escolha, através da função logarítmica.

[A] regressão logística fornece uma extensão muito útil para o caso de previsão de dependentes dicotômicas. A regressão logística se concentra nas chances de um evento, e em até que ponto a extensão de diversas variáveis prognosticadoras aumentam ou reduzem as chances de um resultado sobre outro. (Levin *et al.*, 2012, p. 353)

Nas situações envolvendo os modelos formais de análise de fenômenos judiciais ou jurídicos, os pesquisadores, ao utilizarem a regressão logística, buscam traçar relações entre a presença das variáveis explicativas categóricas inerentes a cada modelo e as prováveis chances dos eventos dicotômicos (discretos), também categóricos, ocorrerem ou não. Em outras palavras, é estatisticamente testado se as variáveis ofertadas nos modelos estão associadas ao resultado, seja no aumento ou na redução das chances de o evento predito acontecer.

2 DA COLETA DOS DADOS À ORGANIZAÇÃO DO MODELO: ESTRATÉGIA EMPÍRICA DA ANÁLISE POR REGRESSÃO LOGÍSTICA

Mas como extrair as informações dos casos e codificá-las de uma forma que possam ser interpretadas e processadas pela ferramenta metodológica da

análise por regressão logística (*logit*)? Para tanto, é necessário saber quais as variáveis integrantes do modelo e qual a sua natureza.

Inicialmente, é necessário identificar no problema de pesquisa a variável resposta (também chamada dependente, ou simplesmente Y), indicativa do efeito, resultado ou fenômeno, que se pretende explicar e compreender. Nos modelos que utilizam a *logit* (Levin *et al.*, 2012), a variável dependente é dicotômica (o mesmo que afirmar ser binária ou discreta, pois expressa em termos de 0 ou 1), isto é, traduz a redução da resposta à pergunta de pesquisa a dois resultados categóricos (embora expressos quantitativamente) possíveis (0) ou (1), onde (0) expressa a ausência do evento esperado (v.g., negativa da liminar, não interposição do recurso, absolvição do acusado etc.) e (1) expressa a presença do evento (v.g., concessão da liminar, interposição do recurso, condenação do acusado etc.).

Em seguida, é preciso identificar quais as variáveis explicativas (independentes, ou simplesmente X) e sua respectiva natureza (King *et al.*, 1994; Levin *et al.*, 2012; Lattin *et al.*, 2011), cuja provável associação com Y pretende-se testar através do modelo em que se emprega a *logit*, pois a natureza das variáveis explicativas é determinante para o procedimento de codificação das variáveis e para o preenchimento da matriz (que, em linguagem coloquial, alguns poderiam chamar de planilha).

“Dados” são elementos de informação sobre o mundo sistematicamente coletados. *Eles podem ser de natureza qualitativa ou quantitativa.* Na maioria das vezes, dados são coletados para avaliar especificamente uma teoria, mas com alguma frequência pesquisadores coletam dados antes de saber precisamente o que estão interessados em encontrar. (King *et al.*, 1994, p. 23 – grifos dos autores).

Em primeiro lugar, as *variáveis independentes podem ser de natureza quantitativa discreta*, isto é, números que são resultados apenas da contagem de determinadas informações, como, a título de exemplos, o número de juízes em um dado território, a quantidade de servidores lotados em uma comarca, o valor da remuneração dos juízes etc. Tome-se, por paradigma, um estudo hipotético em que se procure estimar se variáveis estritamente numéricas (valor da causa e renda do autor) estariam associadas à redução ou ao aumento das chances de se levar o conflito a juízo.

QUADRO 1 – EXEMPLO DE MATRIZ LOGIT PARA VARIÁVEIS INDEPENDENTES QUANTITATIVAS DISCRETAS

CASOS	DECISÃO PELA PROPOSITURA DA AÇÃO (1/0)	VALOR DA CAUSA (REAIS)	RENDA MENSAL DO AUTOR (REAIS)
Conflito 1	1	10.000,00	1.800,00
Conflito 2	0	790,00	900,00
Conflito N	1	8.500,00	10.000,00

Fonte: elaboração dos autores.

Por sua vez, *variáveis explicativas quantitativas e contínuas* são aquelas que são resultado de medições a partir de escalas convencionais como tempo, massa, altura, idade, v.g., a quantidade de tempo que o condenado cumpriu em regime fechado em sua pena anterior (medida em dias) ou a quantidade de entorpecente (medida em gramas) encontrada com a pessoa indiciada por tráfico. Imagine-se um estudo hipotético em que se desejasse estimar as chances de reincidência criminal por tráfico de entorpecentes em uma determinada cidade, a partir de características estritamente quantitativas contínuas, sendo a matriz preenchida da seguinte maneira:

QUADRO 2 – EXEMPLO DE MATRIZ LOGIT PARA VARIÁVEIS INDEPENDENTES QUANTITATIVAS CONTÍNUAS

CASOS	REINCIDÊNCIA (1/0)	TEMPO ANTERIOR EM REGIME FECHADO (DIAS)	QUANTIDADE DE ENTORPECENTE EM SUA POSSE (GRAMAS) NA CONDENÇÃO ANTERIOR
Pessoa 1	1	384	2048
Pessoa 2	0	158	234
Pessoa N	1	786	4789

Fonte: elaboração dos autores.

Já as *variáveis explicativas qualitativas ordinais* são de natureza categórica e os respectivos números expressam uma escala crescente ou decrescente (v.g., 1, 2, 3, 4), na qual existe uma *ordenação* entre as categorias, a exemplo do estágio da carreira do Magistrado julgador (probatório, substituto ou titular), da qualidade da democracia em estudos comparativos entre nações (autocracia, semidemocracia, democracia ou poliarquia), mês da observação (janeiro, fevereiro, março, abril) ou períodos democráticos dentro de um mesmo país (ditadura, transição, democracia).

Um estudo hipotético poderia investigar que variáveis categóricas ordinais (preenchidas a partir de suas escalas pré-definidas) estariam associadas às chances de emendamento constitucional por deputados integrantes do Congresso Nacional.

QUADRO 3 – EXEMPLO DE MATRIZ LOGIT PARA VARIÁVEIS INDEPENDENTES QUALITATIVAS ORDINAIS

CASOS	EMENDAMENTO CONSTITUCIONAL (1/0)	PERÍODO DE TEMPO (NÚMERO EQUIVALENTE AO MÊS DO ANO)	ANO DO MANDATO PARLAMENTAR – DEPUTADO (1, 2, 3, 4)
Caso 1	1	3	2
Caso 2	0	12	4
Caso N	1	7	3

Fonte: elaboração dos autores.

Por fim, restam as variáveis qualitativas nominais, de natureza categórica estrita, a espécie de variável explicativa mais comum nos estudos sobre o Poder Judiciário ou sobre questões jurídicas, em que não existe qualquer ordenação entre categorias. Nelas, cada categoria é investigada isoladamente, correspondendo cada uma a uma coluna da matriz, e codificadas através de variáveis artificiais (*dummies*), nas quais (1) equivale à presença da categoria e (0) expressa a ausência da categoria, no respectivo caso analisado. Essa transformação permite que dados de natureza qualitativa e estritamente categórica possam ser submetidos a análises computacionais, situações usualmente classificadas no Brasil como pesquisas quali-quantitativas.

Assim são as variáveis estratégicas referentes aos atores autorizados a propor ações de controle concentrado de constitucionalidade (PGR, União, Mesas do Legislativo, partidos políticos, OAB, associações) ou as variáveis legalistas referentes à espécie de legislação impugnada (lei federal ou lei estadual). Considere-se uma nova pesquisa hipotética, em que se pretende testar a associação entre o gênero e a raça do(a) acusado(a) e as chances de ser concedida a medida de liberdade em *habeas corpus*: cada categoria, se presente, será preenchida com o valor (1); se ausente, será preenchida com o valor (0).

QUADRO 4 – EXEMPLO DE MATRIZ LOGIT PARA VARIÁVEIS INDEPENDENTES QUALITATIVAS NOMINAIS

CASOS	CONCESSÃO DA MEDIDA DE LIBERDADE (1/0)	RAÇA (NEGROS)	RAÇA (PARDOS)	RAÇA (BRANCOS)	GÊNERO (MASCULINO)	GÊNERO (FEMININO)
HC 1	1	1	0	0	1	0
HC 2	0	0	1	0	0	1
HC N	1	0	0	1	1	0

Fonte: elaboração dos autores.

A identificação correta da natureza das variáveis independentes (explicativas) influi na maneira como serão codificados os dados e no respectivo preenchimento da matriz, ajudando a evitar indesejados vieses e resultados estatísticos que, confrontados à teoria sobre o problema, não façam sentido.

3 COMO INTERPRETAR OS RESULTADOS?

Codificadas as informações extraídas dos casos integrantes da amostra pesquisada e preenchida a respectiva matriz, os dados serão submetidos à análise estatística por regressão logística, através de cálculos de função logarítmica relacionados às referidas variáveis. Tais cálculos são, nos dias de hoje, realizados por aplicativos de processamento de dados e de análise estatística (Microsoft Excel, SAS, Systat, R, Stata, SPSS, Python, BioStat, entre outros), cujos resultados (*outputs*) são apresentados através de expressões numéricas, a serem

devidamente interpretadas (e compreendidas) à luz dos modelos, da teoria e das hipóteses inerentes às questões de pesquisa.

Após rodar a análise por regressão logística no aplicativo, isto é, submeter a matriz produzida no item anterior a uma série de operações matemática pré-definidas, serão obtidos os seguintes resultados (*outputs*) a serem interpretados pelo pesquisador:

B (β β)	S. E.	Wald	Sig.	Exp (B)	O. R.
-----------------------	-------	------	------	---------	-------

Em primeiro lugar, entre os resultados (*outputs*) da análise está o coeficiente (β β ou *Estimativa*) que aponta a intensidade da associação entre a chance de se verificar a variável dependente (resposta) e a presença das variáveis independentes (explicativas), bem como o sinal do resultado da função logarítmica, que permite saber se a dita variável está associada ao incremento das chances da alternativa (1) ou à sua redução, intuitivamente associada a favor da alternativa (0) – (Demaris, 1992; Peng *et al.*, 2002).

O desvio padrão (*standard error* ou S.E.) indica, por sua vez, o grau de variabilidade da distribuição analisada, isto é, avaliando a colocação do resultado dentro de toda a distribuição, tem-se uma interessante medida acerca da confiabilidade da variável para a explicação do fenômeno. Um desvio padrão muito alto indica uma variável de baixa confiabilidade, que possivelmente pode estar prejudicando o modelo (Everitt, 2006).

A estatística de *Wald* (W.) trata da avaliação individual do Chi-quadrado de cada variável independente (Peng *et al.*, 2002) e auxilia na determinação da significância do modelo, sendo utilizada para avaliar quando a hipótese nula de não efeito na população pode, ou não, ser rejeitada (Levin *et al.*, 2012), permitindo avaliar a sua solidez estrutural e a revisão acerca da presença, ou não, de certas variáveis no modelo. Se o teste do modelo não é capaz de rejeitar, estatisticamente e dentro de parâmetros pré-estabelecidos, a hipótese nula, é necessário reconstruir o modelo com novos parâmetros explicativos.

Em seguida, a *significância* (Sig. e/ou *p-valor*) demonstra a robustez do modelo para explicar a associação entre as variáveis testadas, isto é, se é cabível realizar a inferência estatística a partir dos resultados obtidos ou se tal resultado é produto, dentre diversas causas possíveis, de um viés ou de um erro amostral (Everitt, 2006). Por convenção, considera-se estatisticamente significativos os resultados de um modelo quando o valor deste coeficiente é igual ou inferior a 0.050, o que equivale a afirmar que a explicação é válida para ao menos 95% (noventa e cinco por cento) dos casos contidos na amostra e apenas 5% dos casos estariam fora da capacidade explicativa do modelo (Levin *et al.*, 2012).

Entretanto, mesmo modelos com menor significância estatística podem eventualmente trazer relevantes explicações sobre a ocorrência dos eventos, ainda que as inferências deixem de explicar parcelas maiores da amostra colhida, contrariando a convenção dominante entre os estudos empíricos quantitativos, o que dá espaço para a replicação do trabalho e para a colheita de novas amostras de casos sobre o problema estudado (Vianna, 2001).

Os *coeficientes exponenciais* ($\text{Exp}^{\beta \beta}$) podem ser interpretados no que diz respeito aos prováveis efeitos de uma variável sobre as chances de a variável dependente ocorrer, isto é, apresentar o resultado (1), uma vez que tais valores indicam o multiplicador pelo qual o aumento de uma unidade da variável explicativa (independente) altera (aumentando ou reduzindo) as chances previstas para a variável resposta (dependente), mantidas inalteradas todas as demais variáveis explicativas do modelo, permitindo interpretar o efeito de cada variável prognosticadora (Levin *et al.*, 2012). Desta forma, *v.g.*, se o resultado do exponencial de $\beta \beta$ é 1.500 e o sinal do coeficiente é positivo, significa que o aumento de uma unidade da variável explicativa X1 influi positivamente num aumento de uma vez e meia nas chances de ocorrer o evento esperado (1).

Por fim, tem-se as *razões de chances* (*odds ratio* ou simplesmente OR), expressas pela razão entre as chances, observadas no modelo, de ocorrer cada alternativa, o que permite comparar as chances de verificação dos fenômenos, entre os grupos de alternativas, identificando quais delas prevalecem sobre as outras (Demaris, 1992; Levin *et al.*, 2012).

Neste sentido, em termos estatísticos, a razão “[d]as chances de um evento ocorrer são definidas como a probabilidade de que esse evento ocorrerá dividida pela probabilidade de que ele nunca ocorrerá (*Chances = P/1-P*)” (Levin *et al.*, 2012, p. 349). Esse indicador avalia a chance de um objeto exposto a uma variável explicativa possuir a condição de interesse esperada, em comparação a um objeto não exposto (Francisco *et al.*, 2008).

Considere-se o seguinte exemplo de interpretação: Lima *et al.* (2016) utilizaram a análise estatística por regressão logística (*logit*) para testar se a natureza dos atores responsáveis por ajuizar as ações diretas de inconstitucionalidade (ADIns), conjunto de variáveis extraído do modelo estratégico de explicação do comportamento judicial, estaria associada à ocorrência do resultado “*perda de objeto*”, quando, pela demora da Corte em julgar o conflito, verifica-se, em caráter superveniente, a revogação do ato impugnado, a reforma do texto constitucional ou a perda posterior de legitimidade do ator político. Este tipo de julgamento está presente em 13,4% dos casos e está fortemente associada ao comportamento autorrestrito dos ministros do Supremo Tribunal Federal, no qual se observa a estratégia de evitar, de modo geral, o julgamento de conflitos

constitucionais instaurados por *partidos políticos*, notadamente acerca de regras eleitorais.

QUADRO 5: MODELO *LOGIT* PARA O RESULTADO “PERDA DE OBJETO” NAS AÇÕES DIRETAS DE INCONSTITUCIONALIDADE (AJUSTADO PELO MÉTODO *FORWARD STEPWISE*)

	B	S. E.	Wald	Sig.	Exp (B)
Partidos Políticos	.821	.204	16,252	.000	2,272
Constant	-1.058	.097	117,961	.000	.347

U = 5546 casos; amostra aleatória sorteada pelo pacote SPSS = 681.

Variável dependente: resultado (1 = perda de objeto, 0 = ausência de perda de objeto).

Nível de confiança da amostra em 95% e intervalo de confiança amostral em 3.52%.

Fonte: Lima *et al.*, 2016.

Na pesquisa, cujos resultados do modelo ajustado estão apresentados acima, podem-se inferir alguns pontos relevantes para compreensão do comportamento judicial, na hipótese dos processos de controle concentrado de constitucionalidade encerrados por “*perda superveniente de objeto*”.

O resultado do *coeficiente* é positivo, indicando que a presença do ator partido político como responsável por propor a ação direta de inconstitucionalidade está associada ao aumento das chances de ocorrer a perda de objeto. Por sua vez, o baixo desvio padrão auferi confiabilidade à variável estratégica testada.

O modelo é estatisticamente *significante* para 100% ($p = 0.000$) dos casos integrantes da amostra, bem como o valor do coeficiente exponencial de β aponta que a eventual presença da variável categórica estratégica *partido político* em um caso – isto é, em uma ação direta de inconstitucionalidade –, afeta a verificação do resultado testado. O *exponencial de β* obtido no modelo mostra que as ações diretas de inconstitucionalidade propostas pelos partidos políticos possuem aproximadamente duas vezes (2,272) mais chances de serem encerradas por “*perda superveniente de objeto*” que aquelas propostas pelos demais legitimados.

Não foram encontrados pelos autores daquele estudo resultados estatísticos significativos nas espécies legislativas questionadas, nas atitudes dos juízes,

bem como no tempo necessário ao julgamento, que permitissem inferir probabilisticamente as chances de ocorrência do resultado perda de objeto (1)³. Do mesmo modo, não foram encontradas associações estatisticamente significativas entre o resultado “perda de objeto” e a presença dos demais atores legitimados a propor as ações diretas de inconstitucionalidade, o que não resulta no descarte destas variáveis, passíveis de testes através de outras ferramentas empíricas, v.g., métodos experimentais.

4 QUAIS AS APLICAÇÕES MAIS COMUNS DA FERRAMENTA?

Uma das aplicações mais interessantes da ferramenta em apreço, para a área jurídica, versa particularmente em relação à compreensão da atuação e do comportamento do Poder Judiciário, a partir da consideração da teoria sobre os modelos decisórios. Nesse panorama, foram desenvolvidos diversos modelos específicos sobre o comportamento judicial, entre os quais se destacam três: legalista, atitudinal e estratégico.

Em linhas gerais, o modelo legalista é relacionado a uma abordagem normativa e argumenta que o comportamento judicial restringe-se pela lei e pelo Direito, baseando-se sua análise em méritos exclusivamente jurídicos (Baum, 2009).

Já para o modelo atitudinal, os juízes decidem de acordo com as suas preferências políticas individuais, sendo identificáveis através de preferências quantificáveis, a partir de valores e ideologias categorizadas e arrançadas matematicamente ao longo de uma escala convencional (Gilman e Clayton, 1999).

Por fim, para o modelo estratégico, o juiz adota estratégias para chegar aos seus objetivos, uma vez que considera a existência de várias restrições externas ao seu comportamento. Neste sentido, como clara decorrência da teoria da escolha racional, os Magistrados levam em consideração a reação de todos os agentes no processo jurídico-político, seja no âmbito interno ou externo dos Tribunais, adotando uma conduta estratégica para atingir seus objetivos (Posner, 2008).

Todos esses modelos oferecem uma vasta gama de variáveis explicativas (independentes), discretas, contínuas, ordinais ou nominais, passíveis de testes

3 Há hoje nos estudos empíricos dos fenômenos sociais um debate muito forte sobre os limites dos estudos baseados em ferramentas estatísticas, ilustrado pela máxima “*correlation is not causation*” (correlação não é causalidade): é comum uma pesquisa encontrar variáveis estatisticamente correlacionadas, mas que não possuam relação de causalidade (*correlação espúria*), nada explicando sobre o fenômeno, bem como situações causais que não expressam correlações estatísticas entre si. Há fenômenos cuja explicação está fora das inferências estatísticas tradicionais, demandando abordagens multimétodo e/ou desenhos de pesquisa experimentais ou quase-experimentais para a compreensão correta da complexidade do evento e das relações de causalidade (King *et al.*, 1994).

por meio de modelagem empírica, a partir da aplicação da ferramenta da análise por regressão logística.

4.1 APLICAÇÃO NO ESTUDO DO COMPORTAMENTO JUDICIAL: LITERATURA ESTRANGEIRA

Conforme sinalizado, verifica-se que a literatura jurídica estrangeira apresenta uma gama diversificada de trabalhos empíricos que voltam sua atenção para o estudo do comportamento judicial, valendo-se da ferramenta da regressão logística.

Waltenburg e Swinford (1999) analisaram as demandas dos Estados no âmbito da Suprema Corte norte-americana, em confronto com o desenho federal norte-americano. Avalia-se no trabalho a quantidade de demandas encaminhadas pelos Estados à Suprema Corte, envolvendo questões referentes à política nacional, bem como uma aparente atuação da Corte no sentido de reorientar o federalismo norte-americano para um panorama de maior autonomia estatal.

Por sua vez, Vanberg (2001) cria um modelo que se propõe a analisar o comportamento estratégico dos juízes, demonstrando-se sob quais circunstâncias os Magistrados serão restringidos ou não por considerações estratégicas, a partir de resultados gerados através de regressão *logit* que tomam como variável dependente decisões do Tribunal Constitucional alemão, entre 1983 a 1995.

Neste sentido, verificou que o Tribunal alemão é mais suscetível a anular um estatuto quando há probabilidade maior de atuar em um ambiente político mais transparente. Portanto, conclui que a natureza das relações jurídico-judiciais depende significativamente do ambiente político em que o tribunal e o legislador se inserem, assim como, inegavelmente, das preferências judiciais.

Em um dos trabalhos que discutem o comportamento judicial nas Cortes de Apelação norte-americanas, Giles *et al.* (2006, p. 28) identificaram pouca evidência de que os fatores ideológicos desempenham papel significativo no comportamento judicial dos Magistrados, confirmando-se, através da *logit*, um panorama legalista na atuação daquelas Cortes. Ademais, sinalizaram para uma importante observação que deve nortear os estudos empíricos sobre o tema, ao destacar a necessidade de se voltar os olhares para as particularidades da formação da agenda do Tribunal, tendo em vista a complexidade do processo decisório.

Ao associar a teoria da influência dos grupos de interesse sobre a legislatura e suas implicações no controle do Congresso sobre o Poder Judiciário, Iaryczower *et al.* (2006) apresentaram resultados empíricos significativos, através da *logit*, para as interações entre tribunais, sindicatos e bancada legislativa na Argentina. Verificaram, assim, que os tribunais argentinos são mais sensíveis

às demandas dos sindicatos quanto mais fortes estes são, instrumentalizando tal poder por meio do exercício das greves.

Identificaram, ainda, que os sindicatos são mais ativos quando se deparam com legislaturas mais unificadas, pressionando-as para imprimir na agenda legislativa seus interesses, o que repercute em decisões judiciais favoráveis aos próprios sindicatos. Diante desse panorama, o trabalho teceu considerações sobre a importância de atentar para a “natureza real do processo de formulação de políticas”, particularmente no que se refere à interação da burocracia, dos tribunais e do Legislativo.

Sunstein *et al.* (2006) utilizaram um modelo *logit* para demonstrar que, entre os juízes federais estadunidenses, as categorias referentes aos perfis de escolha, a partir da indicação partidária republicana ou democrata, influenciam, dependendo da matéria, positiva ou negativamente, quanto ao resultado do julgamento.

Já Taylor (2008) aplicou um modelo *logit* para estudar o comportamento do Supremo Tribunal Federal brasileiro quanto ao julgamento de liminares em ações diretas de inconstitucionalidade, descobrindo que alguns atores legitimados possuem maiores chances de obter a liminar, em comparação com os demais, bem como que, na categoria relativa ao assunto do conflito, alguns tópicos, como legislações envolvendo benefícios judiciais ou tributos, possuem maiores chances de serem apreciadas em decisões liminares, seja negando ou concedendo a medida.

Com o objetivo de explicar o comportamento dos juízes constitucionais em Cortes do tipo kelseniano, em relação à sua independência frente aos interesses dos partidos políticos, Amaral-Garcia *et al.* (2009) avaliaram, utilizando-se da *logit* no estudo da Corte constitucional portuguesa, que os juízes constitucionais em Portugal são significativamente sensíveis à sua afiliação política, o que repercute em sua atuação decisória, quando da presença de seus partidos políticos originais no governo, assim como atuando como parte nas demandas judiciais.

Em paralelo, também identificaram que a pressão dos demais Magistrados na atividade judicial é significativa, especialmente em relação a votações unânimes, além do fato de que a reforma legislativa visando à independência judicial não surtiu efeitos estatisticamente robustos quanto às chances de ocorrer o resultado pretendido.

Epstein *et al.* (2013) também utilizaram um modelo de regressão logística para, em relação aos juízes integrantes dos Tribunais Regionais Federais norte-americanos, testar modelos atitudinais e estratégicos, obtendo resultados que demonstram, com significância estatística, haver associação entre as categorias

que expressam as preferências individuais dos juízes e as chances de ser escolhida esta ou aquela alternativa de julgamento, afastando hipóteses de natureza legalista.

Dessa forma, através da análise *logit*, verificaram que as adições republicanas à composição do Tribunal tiveram um efeito positivo sobre a probabilidade de sucesso das demandas propostas pelos Estados. Em paralelo, observam que a presença de melhorias na capacidade das procuradorias estaduais não teve efeito significativo sobre a taxa de sucesso das demandas estaduais na Suprema Corte.

4.2 APLICAÇÃO NO ESTUDO DO COMPORTAMENTO JUDICIAL: LITERATURA BRASILEIRA

Em que pese a identificação de uma série de estudos na literatura estrangeira que se valem da *logit* para a análise do comportamento judicial, são ainda tímidos os trabalhos que se utilizam dessa ferramenta estatística para tal finalidade no Brasil. Partindo-se dessa identificação, destacam-se abaixo alguns trabalhos publicados na literatura brasileira.

Ao utilizar um modelo de regressão *logit* para explicar os fatores sociojurídicos que influenciam a aplicação da medida socioeducativa ao adolescente autor de ato infracional, Cunha e Silva (2008) testaram um modelo *logit* que leva em consideração as variáveis gênero, local de residência, reincidência e ato infracional, como categorias explicativas do comportamento dos juízes na aplicação, ou não, da medida socioeducativa ao adolescente que cometeu ato infracional, confirmando-se a hipótese de que “a medida aplicada ao adolescente autor de ato infracional é influenciada por suas características sociodemográficas”.

Por seu turno, em trabalho sobre a natureza jurídica e a efetividade da razoável duração do processo, Mourão (2011) analisou os hábitos de judiciabilidade dos respondentes da pesquisa SIPS Justiça, em função de seus atributos sociodemográficos e dos principais problemas por eles enfrentados, através de análise de regressão *logit*. Nesse contexto, identificou que a probabilidade de os respondentes buscarem a justiça (variável dependente binária) é maior nos casos criminais e menor nos casos envolvendo as seguintes categorias: “empresas com as quais fez negócio”, “cobrança de impostos ou outros conflitos com o Fisco”, “vizinhança”, “pessoas com as quais já fez negócios”, ou “previdência, assistência social ou demandas por direitos sociais”.

Ao avaliar a possibilidade de existência de um possível viés favorável aos hipossuficientes pelos Magistrados em decisões judiciais, valendo-se de regressão logística aplicada a partir de dados coletados junto ao portal institucional do Tribunal de Justiça do Paraná, Castro (2013) identifica que os resultados da

estimação apontam para “uma correlação negativa e estatisticamente significativa entre ser autor pobre e ganhar o processo judicial”.

Em trabalho recente sobre o comportamento decisório do Ministro Presidente do Supremo Tribunal Federal diante da instrumentalização de pedidos de suspensão pelo Procurador Geral ou pelos representantes do Poder Executivo, Gomes Neto (2015) verificou, ao testar hipóteses formuladas à luz dos modelos decisórios, por um lado, que as variáveis jurídicas pouco ou nada influenciam sobre a decisão em sede de pedido de suspensão, bem como que a identidade institucional não possui relevância para a decisão.

Em contrapartida, o estudo destaca que a identidade partidária ou orientação ideológica do Ministro possui resultados significativos. Nesse sentido, o autor destaca que os achados reforçam a tese de que o modelo atitudinal caracteriza-se como ferramenta explicativa do processo decisório judicial no Brasil.

CONCLUSÕES

O estudo sobre os fatores que influenciam o comportamento judicial e dos demais atores, que interferem e participam da judicialização das questões e de conflitos politicamente relevantes, notadamente no que diz respeito aos processos decisórios por eles realizados, ainda é um vasto, fértil e pouco explorado campo para as pesquisas empíricas em Direito ou em Ciência Política.

A literatura oferece para tanto a regressão logística (*logit*) como o método apropriado para conduzir a análise de dependência quando as variáveis dependentes são de natureza categórica discreta binária, ainda mais quando estas são resultado de um processo de escolha entre alternativas categóricas, tal como descrito acima.

Essa técnica tem adquirido maior latitude, ao longo do tempo, na pesquisa empírica, alcançando relevantes problemas de pesquisa acerca das probabilidades de ocorrência de diversos fenômenos sociais e institucionais.

Tal ferramenta é empregada para testar as hipóteses previstas nos diversos modelos formais, ao verificar se existe associação entre a presença, ou ausência, de variáveis extraídas dos modelos e os resultados binários esperados, baseados na referida escolha entre alternativas, de modo a indicar, no campo da probabilidade, se a presença das variáveis explicativas aumenta ou diminui as chances do evento predito ocorrer. Na seara jurídica, a título de exemplificação, este evento pode corresponder ao deferimento de um despejo, à condenação por um júri, a uma decisão colegiada em controle concentrado de constitucionalidade ou ao julgamento de um mandado de segurança.

Este artigo apresenta esta ferramenta relevante de pesquisa para o público em geral, em especial aos pesquisadores que pretendam realizar estudos empíricos de natureza quali-quantitativa, amparados no tratamento computacional de dados sobre o fenômeno investigado, enfrentando detalhes como identificação de problemas de investigação, codificação das variáveis, interpretação dos resultados e revisão de casos em que foi aplicada.

REFERÊNCIAS

- AMARAL-GARCIA, S. *et al.* (2009). Judicial independence and party politics in the Kelsenian Constitutional Courts: The Case of Portugal. Cornell Law School and Wiley Periodicals, Inc. *Journal of Empirical Legal Studies*, v. 6, Issue 2, 381-404, June 2009.
- BATISTELA, G. C. *et al.* (2009). Estudo sobre a evasão escolar usando regressão logística: análise dos alunos do curso de administração da Fundação Educacional de Ituverava. *Tékhnê – Lógos*, v. 1, n. 1.
- BAUM, L. (2009). *Judges and their audiences: a perspective on judicial behavior*. Princeton University Press.
- DEMARIS, A. (1992). *Logit modeling: practical applications*. Newbury Park: Sage.
- EPSTEIN, L.; MARTIN, A. (2014). *An introduction to empirical legal research*. Oxford: Oxford University Press.
- EPSTEIN, L. *et al.* (2013). *The behavior of Federal Judges: a theoretical and empirical study of rational choice*. Cambridge: Harvard University Press.
- EVERITT, B. S. (2006). *The Cambridge Dictionary of Statistics*. Cambridge: Cambridge University Press.
- FRANCISCO, P. M. S. B. *et al.* (2008). 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.
- GILES, M. W. *et al.* (2006). Setting a Judicial Agenda: The Decision to Grant En Banc Review in the U.S. Courts of Appeals. Chicago, The Chicago University Press. *The Journal of Politics*, v. 68, No. 4, p. 852-866, Nov. 2006.
- GILLMAN, H.; CLAYTON, C. W. (1999). Beyond Judicial Attitudes: Institutional approaches to Supreme Court Decision-Making. In: CLAYTON, C. W.; GILLMAN, H. (Ed.). *Supreme Court Decision-Making: new institutionalist approaches*. Chicago: University of Chicago Press.
- GOMES NETO, J. M. W. (2015). *Pretores estratégicos: por que o Judiciário decide a favor do Executivo e contra suas próprias decisões? Análise empírica dos pedidos de suspensão apresentados ao STF (1993-2012)*. Tese de Doutorado. Recife, UFPE.
- IARYCZOWER, M. *et al.* (2006). Judicial lobbying: the politics of labor law constitutional interpretation. *American Political Science Review*, v. 100, No. 1, February 2006.
- LATTIN J. *et al.* (2011). *Análise de dados multivariados*. São Paulo: Cengage Learning.
- LEVIN, J. *et al.* (2012). *Estatística para ciências humanas*. São Paulo: Pearson.

- LEWIS-BECK, M. S. (1992). Series Editor's Introduction. In: DEMARIS, A. (1992). *Logit modeling: practical applications*. Newbury Park: Sage.
- LIMA, F. D. S. *et al.* (2016). Exploring the wonderful mystery of time: "lack of grounds" in constitutional judicial review as an evidence of passive virtues in Brazilian Supreme Court (STF). Paper presented at 24th IPSA World Congress of Political Science. Poznan, Poland.
- PENG, C. J. *et al.* (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, v. 96, n. 1, p. 3-14.
- POSNER, R. (2008). *How judges think*. Cambridge: Harvard University Press.
- REZENDE, F. (2017). O pluralismo inferencial na ciência política pós-KKV (2005-2015): argumento e evidências. *Revista Política Hoje*, 26(1), 241-278.
- SUNSTEIN, C. R. *et al.* (2006). *Are judges political? An empirical analysis of the Federal Judiciary*. Washington DC: The Brookings Institution.
- TAYLOR, M. M. (2008). *Judging Policy: Courts and Policy Reform in Democratic Brazil*. Stanford: Stanford University Press.
- VANBERG, G. (2001). Legislative-judicial relations: a game-theoretic approach to constitutional review. *American Journal of Political Science*, V. 45, No. 2, p. 346-361, Apr. 2001.
- VIANNA, I. O. A. (2001). *Metodologia do trabalho científico: um enfoque didático da produção científica*. São Paulo: EPU.
- WALTENBURG, E. N.; SWINFORD, B. (1999). *Litigating federalism: the states before the U.S. Supreme Court*. London: Greenwood Press.