

Verbetes Metodológico

# Regressão

Dalson Figueiredo<sup>1</sup> 

<sup>1</sup>Professor do Departamento de Ciência Política da Universidade Federal de Pernambuco, DCP/UFPE, Recife, PE, Brasil.

E-mail: dalsonjus@gmail.com

DOI: 10.17666/ebcs.2026.042

## PREÂMBULO

Regressão é o nome genérico de uma família de modelos estatísticos que estimam como o valor esperado de uma variável de interesse, a variável dependente, se altera diante da variação de uma ou mais variáveis explicativas. Um exemplo simples é a relação entre a intensidade da chuva e o preço de uma corrida por aplicativo de transporte. Quanto mais chuva, maior o valor da corrida, em média. A regressão consegue capturar a direção e a magnitude dessa associação. Os modelos de regressão compartilham uma estrutura comum: um preditor linear ligado à média da variável dependente por uma função de ligação; e diferem sobretudo pela distribuição assumida para essa variável, que acompanha seu nível de mensuração (nominal, ordinal, de contagem ou contínua). A regressão serve para descrever associações, para prever e, sob pressupostos adicionais que dependem do desenho de pesquisa, para estimar efeitos causais. O verbete apresenta a origem da técnica e seus fundamentos, a lógica da estimação e de seus pressupostos, suas vantagens e limites, três aplicações nas ciências sociais e os debates atuais sobre identificação causal, aprendizado de máquina e reprodutibilidade.

## SUMÁRIO

1. Histórico e fundamentos epistemológicos
2. Procedimentos principais
3. Vantagens, limites de validade e críticas conhecidas
4. Exemplos de aplicação em ciências sociais
5. Debates contemporâneos e variações
6. Referências bibliográficas

Aprovado em: 28/09/2026

Editor(a) responsável: Adriano Codato



Este é um artigo publicado em acesso aberto (Open Access) sob a licença Creative Commons Attribution (<https://creativecommons.org/licenses/by/4.0/>), que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

Enciclopédia Brasileira de Ciências Sociais (EBCS). ANPOCS, 042 - 2026

## 1. Histórico e fundamentos epistemológicos

A regressão nasceu de um problema de hereditariedade. Em 1875, Francis Galton, primo de Charles Darwin, distribuiu a amigos pacotes de sementes de ervilha, cada pacote com sementes de mesmo peso, mas com variação significativa entre os pacotes. Recolhidas as sementes da geração seguinte, Galton representou graficamente o peso das sementes-filhas em relação ao peso das sementes-mães e percebeu que a linha que unia as médias condicionais tinha inclinação inferior a uma unidade, em torno de 0,33 em seu primeiro estudo documentado, apresentado em uma conferência de 1877 (Stanton, 2001). O nome do método deriva desse achado substantivo: descendentes de progenitores extremos tendiam a “regredir” em direção à média da população. Galton aplicou a mesma ideia à altura de pais e filhos em estudo publicado em 1886, no periódico do *Anthropological Institute*, e o mesmo padrão de regressão à média foi detectado (Galton, 1886). Karl Pearson, colaborador e depois biógrafo de Galton, converteu a intuição gráfica em álgebra ao formular o método do produto-momento, do qual derivam tanto o coeficiente de correlação (o  $r$  de Pearson) quanto a generalização do modelo para múltiplos preditores, a regressão múltipla (não confundir com regressão multivariada que exige mais de uma variável dependente ao mesmo tempo).

A difusão do método para as Ciências Sociais foi rápida e carregou desde o início a ambição causal, tema retomado adiante nas seções 3 e 5. Já em 1899, George Udny Yule, conhecido por estudar associação em tabelas de contingência e por antecipar o problema hoje chamado de paradoxo de Simpson, usou um modelo de regressão para sustentar que a concessão de auxílio fora das casas de trabalho aumentava o número de pobres na Inglaterra (Freedman, 1991). Desde então, a regressão tornou-se ferramenta estatística padrão do campo. Na Ciência Política, por exemplo, um levantamento de 1.756 artigos publicados na *American Political Science Review*, no *American Journal of Political Science* e no *Journal of Politics*, entre 1990 e 2005, mostra que os mínimos quadrados ordinários (MQO) apareciam em 30,8% dos trabalhos e que apenas 37,1% dos artigos que recorriam ao MQO empregavam um segundo método, como registram James Krueger e Michael Lewis-Beck em “*Is OLS Dead?*”, de 2008. Um levantamento mais recente e mais amplo, apresentado por Andrés Castro Torres e Aliakbar Akbaritabar em “*The Hegemonic Use of Linear Models in Quantitative Social Sciences*”, de 2025, indica que modelos lineares respondem por 70% a 80% da pesquisa quantitativa nas subdisciplinas das Ciências Sociais e Humanas. Dadas a versatilidade e a relativa simplicidade do método, a regressão é frequentemente descrita como o “cavalo de batalha” (*workhorse*) da pesquisa quantitativa.

Em termos técnicos, toda regressão é uma estimativa da esperança condicional, visto que informa o valor médio da variável dependente ( $y$ ) para cada configuração das variáveis explicativas ( $x$ ). Nas Ciências Sociais, a regressão serve tanto para descrever relações entre variáveis quanto para tentar identificar relações causais e para fazer previsões (Figueiredo Filho et al., 2011). Berk (2010) organiza esses usos em três níveis. No nível I, a regressão é ferramenta de descrição de dados, não exige pressuposto sobre o processo que gerou as observações e visa resumir associações entre variáveis. No nível II, acrescenta-se a inferência estatística (erros-padrão, testes de significância e intervalos de confiança), que exige

pressupostos adicionais sobre o processo de amostragem ou sobre o processo gerador dos dados. No nível III, acrescenta-se a inferência causal, que exige pressupostos de identificação, como ausência de variáveis de confundimento, exogeneidade e comparabilidade entre grupos. Nenhum teste estatístico ou algoritmo permite passar de um nível a outro: a passagem depende da teoria, do desenho de pesquisa e dos pressupostos que se está disposto a assumir.

## 2. Procedimentos principais

O modelo linear clássico expressa a variável dependente como soma de um intercepto, de um conjunto de termos que multiplicam as variáveis independentes por coeficientes desconhecidos e de um termo de erro:  $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$ . A estimação por mínimos quadrados ordinários escolhe os valores dos coeficientes que minimizam a soma dos quadrados dos resíduos, isto é, das distâncias verticais entre os valores observados e os valores previstos pelo modelo, o que pode ser visualizado no mesmo tipo de gráfico de dispersão usado por Galton. Cada coeficiente é lido como a variação esperada em Y associada ao aumento de uma unidade na respectiva variável independente, mantidas constantes as demais variáveis do modelo (*ceteris paribus*). “Manter constante” é aqui uma operação algébrica sobre o conjunto de variáveis que o pesquisador decidiu incluir no modelo: o que ficou de fora não é mantido constante por nenhum procedimento e a omissão de uma variável relevante pode comprometer a estimação dos coeficientes, principalmente se a variável que ficou de fora for correlacionada com as variáveis que foram incluídas no modelo.

A validade das estimativas depende de um conjunto de pressupostos, sistematizados por William Berry em “*Understanding Regression Assumptions*”, de 1993, e cada violação tem consequências de gravidade variável. A linearidade nos parâmetros e a exogeneidade dizem respeito ao próprio coeficiente: se forem violadas, a estimativa é enviesada, isto é, tende a superestimar ou subestimar sistematicamente o valor do parâmetro de interesse, e nenhum aumento da amostra corrige o problema. A homocedasticidade (variância constante do erro) e a ausência de autocorrelação dizem respeito à precisão: quando violadas, os coeficientes permanecem não enviesados, mas os erros-padrão ficam incorretos, o que compromete os testes de significância e os intervalos de confiança. A colinearidade elevada entre preditores não produz viés, mas infla a variância das estimativas e dificulta a identificação da contribuição individual de cada variável; erros-padrão inflados tendem a produzir coeficientes não significativos e, assim, conclusões equivocadas sobre as relações de interesse.

Na etapa de inferência, o valor-p informa a compatibilidade dos dados com a hipótese nula sob o modelo assumido (Wasserstein & Lazar, 2016). Por essa razão, a análise deve considerar também o tamanho do efeito e o grau de incerteza das estimativas, e não apenas os indicadores de significância.

A escolha do modelo dentro da família de regressão depende do nível de mensuração da variável dependente. Quando Y é contínua e não tem limites naturais, o modelo linear é o ponto de partida. Quando Y é binária, como reeleger-se ou não, migrar ou não, o modelo linear pode gerar probabilidades previstas fora do intervalo [0,1] e a regressão logística torna-se a alternativa mais adequada; seus coeficientes são estimados na escala do logaritmo das

chances e geralmente traduzidos em razões de chance ou em probabilidades preditas para facilitar a interpretação (Fernandes et al., 2020). Quando  $Y$  é uma contagem, como o número de homicídios, a regressão de Poisson é o modelo de referência, sob o pressuposto de que média e variância são iguais; violado esse pressuposto, recorre-se à regressão binomial negativa, em caso de superdispersão, e a modelos inflacionados de zeros, quando o excesso de zeros resulta de um processo distinto do que gera as contagens positivas (Coxe et al., 2009). Quando  $Y$  é uma taxa ou proporção restrita ao intervalo  $(0,1)$ , como a participação eleitoral, a prática recorrente era transformar a variável e aplicar regressão linear; a regressão beta foi proposta justamente para essa situação, acomodando a assimetria e a heterocedasticidade características dessas variáveis (Ferrari & Cribari-Neto, 2004).

Em comum, esses modelos têm um preditor linear e uma função de ligação que o conecta à média da variável dependente. Muda a distribuição assumida para  $Y$  e, com ela, a escala em que os coeficientes são expressos e a forma de interpretá-los; permanece o objetivo de decompor a variação de uma variável em função de outras. A última etapa de validação envolve a análise de resíduos, a identificação de casos influentes (*outliers*) e testes de robustez, mais difundidos na Economia do que na Ciência Política e na Sociologia. Mais recentemente, o movimento de ciência aberta passou a exigir também a disponibilização de dados e códigos que permitam a terceiros reproduzir os resultados.

### 3. Vantagens, limites de validade e críticas conhecidas

As vantagens da regressão explicam sua popularidade. É um método parcimonioso, que resume relações entre muitas variáveis em poucos números. É transparente, no sentido de que cada escolha da pesquisadora fica inscrita na especificação e pode ser contestada, sobretudo quando dados e códigos são compartilhados. É comparável, porque produz quantidades que podem ser confrontadas entre estudos. É flexível, porque acomoda variáveis dependentes com diferentes níveis de mensuração. É relativamente robusta a violações moderadas de alguns pressupostos, como os relativos à distribuição dos erros, embora não às que afetam a exogeneidade (seção 2). Por fim, funciona como língua franca da pesquisa quantitativa: resultados apresentados sob a forma de modelos de regressão são legíveis por comunidades disciplinares inteiras, dentro e fora das Ciências Sociais, e seu domínio é condição para ler boa parte da produção científica contemporânea. Um discente que termina a pós-graduação sem o conhecimento básico de regressão vai ter muitas dificuldades para consumir e interagir com o que se produz na ciência, independentemente da área de formação.

Entre as limitações, uma das mais graves é o viés de variável omitida. Se uma variável associada simultaneamente a  $X$  e a  $Y$  fica fora do modelo, o coeficiente estimado de  $X$  incorpora parte do efeito da variável omitida. Na estimação do efeito do policiamento ( $X$ ) sobre a criminalidade ( $Y$ ), por exemplo, a omissão da vulnerabilidade social do bairro é problemática: bairros mais vulneráveis podem receber mais policiamento e, ao mesmo tempo, apresentar mais criminalidade. Nesse caso, o coeficiente do policiamento absorve parte do efeito da vulnerabilidade e fica enviesado para cima.

Outra limitação, que o sociólogo Hubert Blalock considerava o principal problema da área, diz respeito à mensuração. Dada a complexidade dos fenômenos sociais, as variáveis das Ciências Sociais tendem a conter mais erro de mensuração do que as das Ciências Naturais. É muito mais fácil medir a concentração de sal em uma substância do que a ideologia de um deputado, por exemplo. Erros de mensuração nas variáveis independentes afetam a consistência dos coeficientes estimados, que tendem a ser menores do que seriam na ausência desses erros; por isso, a ausência de erro de mensuração nas variáveis independentes figura entre os pressupostos do modelo.

A causalidade reversa é outro obstáculo, pois a álgebra da regressão não identifica a direção da causalidade e a endogeneidade daí resultante tende a produzir coeficientes viesados e inconsistentes. Os manuais também apontam o risco de extrapolação para regiões do espaço das variáveis em que não há observações. Se uma regressão entre anos de escolaridade e renda é estimada com uma amostra de pessoas que têm entre 8 e 18 anos de estudo, o modelo pode prever matematicamente a renda de quem tem 2 ou 25 anos de escolaridade, mas essas previsões não são sustentadas por observações. Em modelos para variáveis binárias, as estimativas tornam-se instáveis quando o evento de interesse é raro ou a amostra é pequena, situações que exigem correções específicas.

Uma das críticas mais influentes ao uso da regressão foi formulada por Freedman (1991), que ressalta a primazia do desenho de pesquisa. O contraexemplo que ele mobiliza é o de John Snow, o médico que, na Londres da década de 1850, estabeleceu que a cólera se transmitia pela água contaminada comparando a mortalidade em domicílios servidos por companhias com formas distintas de captação de água. A credibilidade da inferência vem da qualidade da comparação e da inventividade do desenho, não da complexidade do estimador. Berk (2010) expande essa crítica e formula recomendações práticas. Na ausência de amostragem probabilística, testes de hipótese e intervalos de confiança perdem o fundamento que justifica sua interpretação usual, e a análise deve ser lida como descrição. Na ausência de um modelo aproximadamente correto do processo gerador dos dados, condição raramente satisfeita em pesquisas observacionais, a interpretação causal dos coeficientes não se sustenta. Nessa perspectiva, a maior parte das regressões com dados observacionais deveria ser interpretada como descrição de associações entre variáveis.

## 4. Exemplos de aplicação em ciências sociais

O estudo de Card e Krueger (1994) sobre salário mínimo e emprego ilustra a regressão aplicada a um problema clássico. A teoria econômica convencional prevê que elevar o salário mínimo reduz o emprego. Em 1º de abril de 1992, Nova Jersey elevou seu salário mínimo de US\$ 4,25 para US\$ 5,05 por hora, enquanto a vizinha Pensilvânia o manteve constante. A decisão metodológica foi usar os estabelecimentos da Pensilvânia como grupo de comparação dos de Nova Jersey, entrevistando 410 restaurantes de *fast-food* antes e depois da mudança e comparando a variação do emprego entre os dois estados, estratégia conhecida como diferenças em diferenças que, sob determinados pressupostos, permite sustentar inferências causais. O estudo não encontrou redução do emprego em Nova Jersey; ao contrário, estimou

um ganho relativo de 2,76 empregados em tempo integral por estabelecimento, cerca de 13%. O exemplo mostra como, a depender do desenho e do ajuste por covariáveis, a regressão pode ser útil para detectar relações causais.

Fernandes et al. (2020) replicam o estudo de Mônica Mata Machado de Castro e Felipe Nunes, “Candidatos corruptos são punidos? Accountability na eleição brasileira de 2006”, de 2014, para exemplificar o uso da regressão logística. A pergunta substantiva é se o envolvimento em escândalos de corrupção afeta a chance de reeleição de deputados federais no Brasil. Reanalizando dados de 451 candidatos na eleição de 2006, os autores reportam que 60,53% dos deputados foram reconduzidos, mas apenas 17,86% (10 de 56) entre os citados nos esquemas do mensalão e dos sanguessugas, contra 66,58% entre os não citados; os não citados tiveram chance de reeleição cerca de oito vezes maior. O estudo também exemplifica boas práticas de pesquisa, pois dados e códigos de replicação foram disponibilizados publicamente, o que aumenta a transparência dos resultados e facilita o reuso das informações.

Schaefer e Figueiredo Filho (2023) mostram o que muda quando a regressão é precedida pelo pré-processamento dos dados. Ao investigar o efeito das escolas militares sobre a aprendizagem, os autores parearam estudantes com base no desempenho anterior, medido pela nota de matemática no 5º ano, antes de estimar a diferença de proficiência no 9º ano. A comparação direta, sem pareamento, indicava vantagem de 26,34 pontos para as escolas militares; após o pareamento, a diferença cai para 18,59 pontos. Ambas as estimativas são estatisticamente significantes, mas a magnitude do efeito é substancialmente superestimada sem o pareamento.

## 5. Debates contemporâneos e variações

O debate metodológico contemporâneo deslocou o eixo da especificação do modelo (a inclusão das variáveis relevantes e a adequação da forma funcional) para a estratégia de identificação (o desenho de pesquisa que sustenta inferências válidas). A pergunta que organiza a avaliação de um trabalho empírico deixou de ser “quais variáveis de controle foram incluídas?” e passou a ser “que comparação sustenta a afirmação causal?”. A regressão não perdeu espaço nesse movimento; mudou de função. Se antes se esperava extrair efeitos causais de dados observacionais por meio do controle estatístico, hoje a regressão entra como instrumento de estimação subordinado a um desenho com estratégia de identificação definida. Os “cinco furiosos” (*Furious Five*) descritos por Joshua Angrist e Jörn-Steffen Pischke em “*Mastering 'Metrics: The Path from Cause to Effect*”, de 2015 (experimentos aleatorizados, regressão, diferenças em diferenças, regressão descontínua e variáveis instrumentais), passam a constituir a base da inferência causal.

O caso das diferenças em diferenças ilustra o ritmo dessa revisão. A forma canônica, com dois grupos e dois períodos, é bem compreendida e amplamente utilizada, inclusive na avaliação de impacto de políticas públicas. A literatura recente demonstrou que sua extensão para múltiplos períodos com adoção escalonada do tratamento, estimada por efeitos fixos bidirecionais, pode produzir estimativas enviesadas porque incorpora comparações entre unidades já tratadas. Daí a proliferação de estimadores alternativos e de testes de

sensibilidade a violações do pressuposto de tendências paralelas (Roth et al., 2023). Na regressão descontínua, em que o efeito é estimado por regressão local em torno de um ponto de corte que determina o tratamento, o amadurecimento da literatura produziu dois arcahouços distintos de interpretação, o de continuidade e o de aleatorização local, além de procedimentos padronizados para escolha da janela, inferência robusta ao viés e testes de falsificação (Cattaneo e Titiunik, 2022). Em ambos os casos, consolidou-se a exigência de que decisões antes tratadas como técnicas sejam justificadas e submetidas a verificação.

Uma segunda frente é a incorporação de métodos de aprendizado de máquina, que recoloca a distinção entre predição e explicação: modelos flexíveis frequentemente superam a regressão na previsão fora da amostra, mas prever bem não é o mesmo que estimar o efeito de intervir sobre uma variável. Um algoritmo pode prever com elevada precisão quais estudantes têm maior chance de abandonar a escola ao combinar notas, frequência, renda familiar, idade e histórico escolar, sem que isso signifique que alguma dessas variáveis tenha efeito causal sobre a evasão. A previsão responde à pergunta “quem tem maior risco de evadir?”; a inferência causal procura responder “o que aconteceria com a evasão se a frequência escolar fosse intencionalmente alterada?”. Uma terceira frente é a da reprodutibilidade, que depende da disponibilização de dados e códigos, hoje exigida por um número crescente de periódicos. A abertura desses materiais permite que a comunidade acadêmica avalie criticamente as decisões metodológicas tomadas na produção de uma estimativa e facilita que novos estudos reanalisem os mesmos dados com outros métodos e correções.

A produção metodológica brasileira e latino-americana participa desse debate em duas posições. Como formuladora, a região deu contribuições incorporadas ao repertório internacional: a regressão beta, proposta por pesquisadores da Universidade de São Paulo e da Universidade Federal de Pernambuco, tornou-se referência padrão para modelar taxas e proporções (Ferrari & Cribari-Neto, 2004). Como difusora, consolidou-se no Brasil uma tradição de artigos didáticos em português que apresentam métodos com dados e códigos de replicação abertos, voltados ao ensino de pós-graduação (Chein, 2019; Fernandes et al., 2020; Schaefer & Figueiredo Filho, 2023). Essa produção responde ao diagnóstico de Soares (2005) sobre a fragilidade da formação metodológica na Ciência Política brasileira e sobre a resistência de parte da área aos métodos quantitativos. A dor em nosso calcanhar parece estar menor.

## 6. Referências bibliográficas

### Textos fundadores e genealogia

Galton, Francis. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland*, 15, 246–263. <https://doi.org/10.2307/2841583>.

Stanton, Jeffrey M. (2001). Galton, Pearson, and the peas: a brief history of linear regression for statistics instructors. *Journal of Statistics Education*, 9(3). <https://doi.org/10.1080/10691898.2001.11910537>.

### Manuais e guias de referência

Chein, Flávia. (2019). *Introdução aos modelos de regressão linear: um passo inicial para compreensão da econometria como uma ferramenta de avaliação de políticas públicas*. Enap. <https://bibliotecadigital.gestao.gov.br/handle/123456789/525564>.

Coxe, Stefany, West, Stephen G., & Aiken, Leona S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of Personality Assessment*, 91(2), 121–136. <https://doi.org/10.1080/00223890802634175>.

Fernandes, Antônio Alves Tôrres, Figueiredo Filho, Dalson Britto, Rocha, Enivaldo Carvalho da, & Nascimento, Willber da Silva. (2020). Leia este artigo se você quiser aprender regressão logística. *Revista de Sociologia e Política*, 28(74), e006. <https://doi.org/10.1590/1678-987320287406>.

Ferrari, Sílvia Lopes de Paula, & Cribari-Neto, Francisco. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7), 799–815. <https://doi.org/10.1080/0266476042000214501>.

Figueiredo Filho, Dalson Britto, Nunes, Felipe, Rocha, Enivaldo Carvalho da, Santos, Manoel Leonardo, Batista, Mariana, & Silva Júnior, José Alexandre. (2011). O que fazer e o que não fazer com a regressão: pressupostos e aplicações do modelo linear de mínimos quadrados ordinários (MQO). *Revista Política Hoje*, 20(1), 44–99. <https://periodicos.ufpe.br/revistas/politica hoje/article/view/3808>.

Wasserstein, Ronald L., & Lazar, Nicole A. (2016). The ASA statement on p-values: context, process, and purpose. *The American Statistician*, 70(2), 129–133. <https://doi.org/10.1080/00031305.2016.1154108>.

### **Crítica metodológica**

Berk, Richard. (2010). What you can and can't properly do with regression. *Journal of Quantitative Criminology*, 26(4), 481–487. <https://doi.org/10.1007/s10940-010-9116-4>.

Freedman, David A. (1991). Statistical models and shoe leather. *Sociological Methodology*, 21, 291–313. <https://doi.org/10.2307/270939>

### **Debates contemporâneos**

Cattaneo, Matias Damian, & Titiunik, Rocío. (2022). Regression discontinuity designs. *Annual Review of Economics*, 14, 821–851. <https://doi.org/10.1146/annurev-economics-051520-021409>.

Roth, Jonathan, Sant'Anna, Pedro Henrique Costa Gomes de, Bilinski, Alyssa, & Poe, John. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2), 2218–2244. <https://doi.org/10.1016/j.jeconom.2023.03.008>.

Soares, Gláucio Ary Dillon. (2005). O calcanhar metodológico da ciência política no Brasil. *Sociologia, Problemas e Práticas*, (48), 27–52. <https://repositorio.iscte-iul.pt/handle/10071/192>.

### **Aplicações empíricas**

Card, David, & Krueger, Alan Bennett (1994). Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4), 772–793. <https://www.jstor.org/stable/2118030>.

Schaefer, Bruno Marques, & Figueiredo Filho, Dalson Britto. (2023). Deu match? Uma introdução às técnicas de pareamento. *Revista Brasileira de Ciências Sociais*, 38(111), e3811039. <https://doi.org/10.1590/3811039/2023..>

## **VER TAMBÉM:**

Inferência causal / modelos causais | Estudos Quase-Experimentais | Variáveis instrumentais | Mensuração | Amostragem | Replicabilidade | Ciência Aberta | Pareamento (*matching*)

## **COMO CITAR ESTE VERBETE:**

Figueiredo, Dalson. (2026). Regressão. In *Enciclopédia Brasileira de Ciências Sociais (EBCS)*. ANPOCS. 10.17666/ebcs.2026.042.