

REGRESSÃO LINEAR — Conceitos, análise de pressupostos e aplicação com Excel

A econometria é a disciplina que une economia e estatística, com o objetivo de analisar dados econômicos e quantificar as relações entre diferentes variáveis. Uma das ferramentas mais poderosas da econometria é a **regressão linear**.

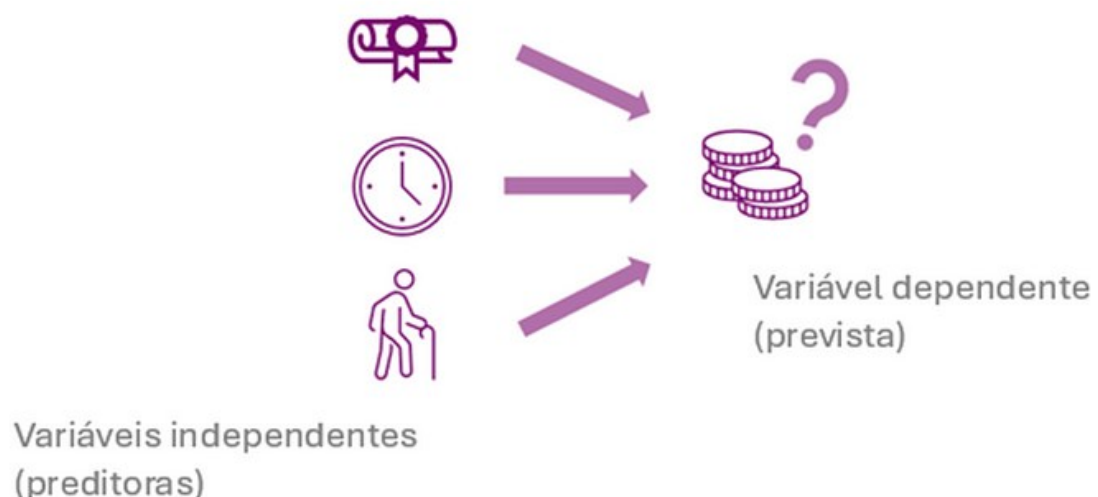
Imagine que você, como economista, queira entender como a taxa de juros influencia o investimento das empresas. A regressão linear permite construir um modelo matemático que descreve essa relação. Ao coletar dados sobre investimentos e taxas de juros em diferentes períodos, você pode estimar uma equação que mostre, por exemplo, que a cada aumento de 1% na taxa de juros, o investimento das empresas tende a cair em X%.

O que é método de regressão linear?

A regressão é um método estatístico que permite modelar relações entre uma variável dependente e uma ou mais variáveis independentes.

Uma análise de regressão permite inferir ou prever outra variável com base em uma ou mais variáveis.

Por exemplo, você pode estar interessado em saber como influenciar o salário dos funcionários de uma empresa e em prever o valor desses salários. Para isso, você pode considerar os dados sobre o nível de educação, o tempo de experiência e a idade dos funcionários dessa empresa.



Uma análise de regressão permitirá que você investigue essas três variações que têm influência nos valores dos atrasos dos funcionários dessa empresa. Se houver influência, você poderá prever o salário dos funcionários dessa empresa usando dados de nível educacional, tempo de experiência e idade.

Qual o objetivo da regressão?

Uma análise de regressão pode ser realizada quando se deseja atingir dois objetivos. Um deles é estudar a influência de uma ou mais variáveis (preditoras) sobre outra variável (a ser prevista), pois com a regressão é possível mensurar a influência do conjunto de variáveis preditoras sobre a variável a ser prevista. Outro objetivo é a própria previsão da variável de interesse a partir de uma ou mais variáveis preditoras, o que será possível por meio de uma solução de regressão (modelo de regressão).

Resumidamente uma análise de regressão servirá para quando você tiver como objetivos:

- *Prever uma variável de interesse*
- *Avaliar a influência de uma ou mais variáveis preditoras sobre a variável de interesse (a ser prevista).*

Quando usar a regressão linear?

Quando há manifestação linear entre as variáveis independentes e a variável dependente, e seu objetivo envolve fazer a variação do seu interesse, e/ou avaliar se o conjunto das variáveis escolhidas como preditoras tem influência significativa sobre a variável a ser determinada .

Quando usar Regressão Simples ou Múltipla?

- A regressão Linear Simples está considerando uma única variável independente (preditora) para estimar o valor da variável de interesse
- A regressão Linear Múltipla se leva em conta duas ou mais variáveis independentes (preditoras) para estimar o valor da variável de interesse

Observe que essa influência conjunta das variáveis independentes sobre a variável dependente não é captada por uma análise de especulação entre as variáveis, pois a análise conjunta só mede o relacionamento de cada par de variáveis , isoladamente, e não supõe quem são as variáveis independentes ou dependentes.

Por exemplo, nos casos em que seu objetivo seja:

1) Medir a influência de uma ou mais variáveis sobre outra variável, em contextos como:

- O nível de concentração de adolescentes (Y) é influenciado pelo tempo que utilizam dispositivos eletrônicos diariamente (X)?
- A renda familiar (X1) e o local de residência (X1) afetaram o desempenho educacional das crianças (Y)?

2) Previsão de uma variável por meio de uma ou mais variáveis conjuntamente, contextos como:

- Quanto tempo um paciente pode permanecer internado no hospital (Y), a partir de dados de idade (X1) e quantidade de comorbidades (X2).
- Na faixa de preço de um produto a pessoa tem mais probabilidade de comprar em uma loja online (Y) a partir de dados de renda familiar (X).

Quais são os modelos de regressão?

As análises de regressão são divididas em regressão linear simples, regressão linear múltipla e regressão logística. O tipo de análise de regressão que deve ser usado depende do número de variáveis independentes e da escala de medição da variável dependente.

	Número de variáveis independentes	Escala de medição da variável dependente	Escala de medição das variáveis independentes
Regressão linear simples	uma	quantitativa	quantitativa, ordinal, nominal
Regressão Linear Múltipla	múltiplas	quantitativa	quantitativa, ordinal, nominal
Regressão Logística	múltiplas	ordinal, nominal	quantitativa, ordinal, nominal

A **regressão logística** pode ser utilizada quando a variável depende do nominal, como por exemplo a satisfação ou não de funcionários com seu trabalho.

Podemos realizar uma **regressão linear simples** que prevê o salário dos funcionários em função do tempo de experiência; uma **regressão linear múltipla** que estime faixas de funcionários (Y) em função do nível de escolaridade (X1), do tempo de experiência (X2), e de sua

idade (X3), conjuntamente; e uma **regressão logística** que indica a satisfação ou não dos funcionários (Y) em função das variações independentes de nível de escolaridade (X1), do tempo de experiência (X2), e de sua idade (X3).

Quais são as premissas da regressão linear?

Os pressupostos básicos da regressão linear são condições que precisam ser atendidas para garantir que os resultados do modelo sejam confiáveis e válidos. Esses pressupostos incluem:

1. **Linearidade** : A relação entre as variáveis independentes (preditoras) e a variável dependente deve ser linear. Isso significa que a variação na variável dependente pode ser explicada de forma linear pelas variáveis independentes.
2. **Independência dos Erros (Autocorrelação)**: Os erros (ou resíduos) do modelo devem ser independentes uns dos outros. Em séries temporais, por exemplo, os valores sucessivos não devem estar correlacionados.
3. **Homoscedasticidade** : A variância dos erros deve ser constante ao longo de todas as observações. Isso implica que os erros não devem se expandir ou contrair conforme o valor das variáveis preditoras muda.
4. **Normalidade dos Erros** : Os resíduos devem seguir uma distribuição normal. Isso é particularmente importante para testes de hipóteses, como os testes de significância dos coeficientes.

O atendimento a esses pressupostos contribui para validação do modelo de regressão linear como um modelo eficaz para prever e inferir a relação entre as variações verificadas.

Como fazer regressão linear no Excel?

O Excel é uma ferramenta poderosa e de fácil acesso para uso pessoal e amplamente difundida no mercado de trabalho. A maioria das empresas usa planilhas com diferentes objetivos nos processos de trabalho, e dessa forma a habilidade no uso do Excel torna-se uma competência básica para profissionais de diversas áreas e níveis.

No [artigo completo sobre Regressão Linear no Excel](#) , você acompanhará um **passo a passo de uma análise de Regressão Múltipla no Excel** , desde a execução até a verificação de suposições, passando pela montagem do modelo de regressão (equação) e simulação de uma previsão usando o modelo. Além disso, um dos vídeos contém um tutorial rápido sobre como **habilitar a Ferramenta de Análise de Dados do Excel** , que é um complemento do Excel, e precisa ser habilitado para que possamos utilizar.

Referências

1. GUJARATI, Damodar N. Econometria básica. São Paulo: Bookman, 2010.
2. WOOLDRIDGE, Jeffrey M. Introdução à econometria: uma abordagem moderna. São Paulo: Cengage Learning, 2016.
3. HILL, R. Carter; GRIFFITHS, William E.; LIM, Guay C. Princípios de econometria com aplicações. São Paulo: Editora Saraiva, 2018.

Referências

FERREIRA, Rafael Gastão Coimbra... [et al.] **Preparação e análise exploratória de dados**; revisão técnica: Jean Paul Barddal. Porto Alegre: SAGAH, 2021.

GABRIEL, Martha. **Inteligência artificial**: do zero ao metaverso / Martha Gabriel. - 1. ed. - Barueri [SP] : Atlas, 2022.

SANTOS, Marcelo Henrique dos. **Introdução à Inteligência Artificial** / Marcelo Henrique dos Santos, – Londrina: Editora e Distribuidora Educacional S.A. 2021.

GOODFELLOW, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep Learning (Vol. 1). MIT press Cambridge.

