

# Regressão Linear e Mínimos Quadrados aplicados às notas do IDEB

Álgebra Linear

---

Hugo de Araújo, Marcéli Melchiors

Novembro, 2021

Fundação Getúlio Vargas - Escola de Matemática Aplicada (FGV/EMAp)

- O trabalho consiste em um modelo linear para as notas do IDEB de cada unidade federativa brasileira.
- Uso de regressão linear feita pelo método de mínimos quadrados.
- Cálculo do coeficiente de determinação ( $R^2$ ) para verificar se o modelo se ajusta bem aos dados.

# Sobre os métodos

- Modelo:  $b_i = Dt_i + C$ , onde  $b_i$  é a nota em função de  $t_i$ , o ano.  
Tendo os pontos  $(t_i, b_i)$ , a solução utilizando mínimos quadrados, sendo erro = e, é dada por:

- $$e = \sum_{i=1}^n (y(t_i) - b_i)^2 = \sum_{i=1}^n (C + Dt_i - b_i)^2$$

- $$A = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \quad \hat{x} = \begin{bmatrix} C \\ D \end{bmatrix}$$

- $$e = \|Ax - b\|^2 \Rightarrow \hat{x} = (A^T A)^{-1} A^T b \Rightarrow A^T A \hat{x} = A^T b$$

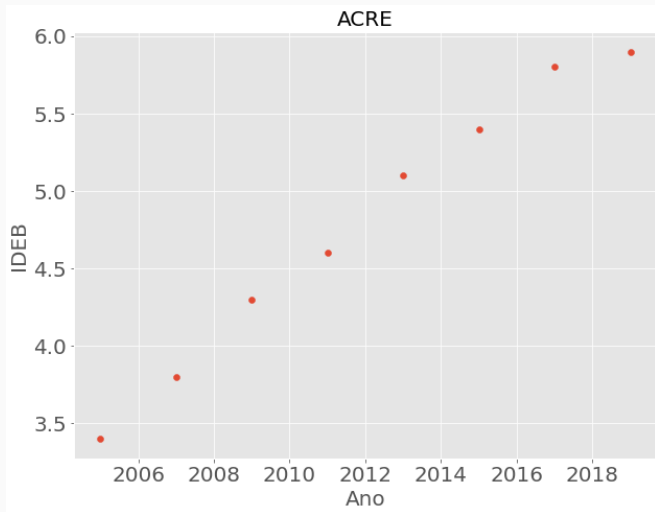
- $$A^T A = \begin{bmatrix} m & \sum_{i=1}^n t_i \\ \sum_{i=1}^n t_i & \sum_{i=1}^n t_i^2 \end{bmatrix}, \quad A^T b = \begin{bmatrix} \sum_{i=1}^n b_i \\ \sum_{i=1}^n t_i b_i \end{bmatrix}$$

- $$R^2 = \frac{SEQ_m - SEQ}{SEQ_m}.$$

- O Índice de Desenvolvimento da Educação Básica (IDEB) é um indicador social da qualidade da educação brasileira criado em 2005 pelo Ministério da Educação.
- As notas variam de 0 a 10 e são computadas com base no fluxo escolar e nas provas do Saeb (Sistema de Avaliação da Educação Básica).
- O IDEB serve para nortear as políticas públicas e encontrar exemplos de sucesso na educação brasileira.
- Para este trabalho, utilizou-se as notas do IDEB para o 5º ano do Ensino Fundamental de 2005 a 2019, no nível estadual.

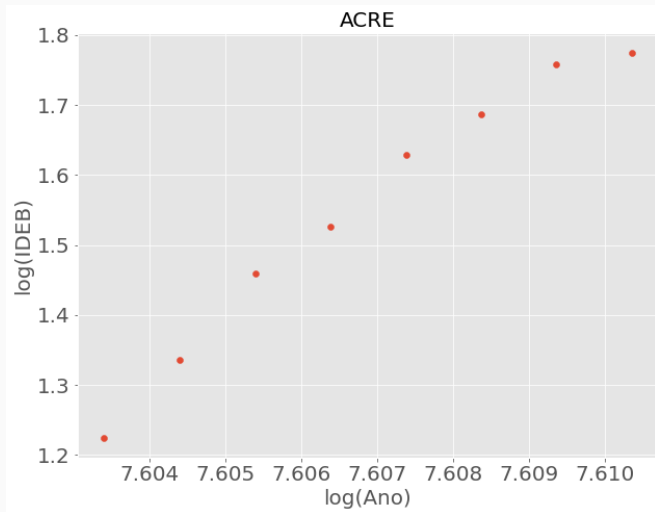
- Primeiramente, plotou-se um gr fico de dispers o entre o IDEB do estado e o tempo, em anos.
- Pela diferen a da escala das vari veis, o logaritmo natural foi aplicado nas vari veis e mais um plot foi feito.
- Por fim, plotou-se um terceiro gr fico contendo a regress o encontrada pelo m todo dos m nimos quadrados e outra encontrada pela biblioteca Scikit-Learn.
- As regress es foram feitas para as 27 unidades federativas.

# Gráfico do Acre



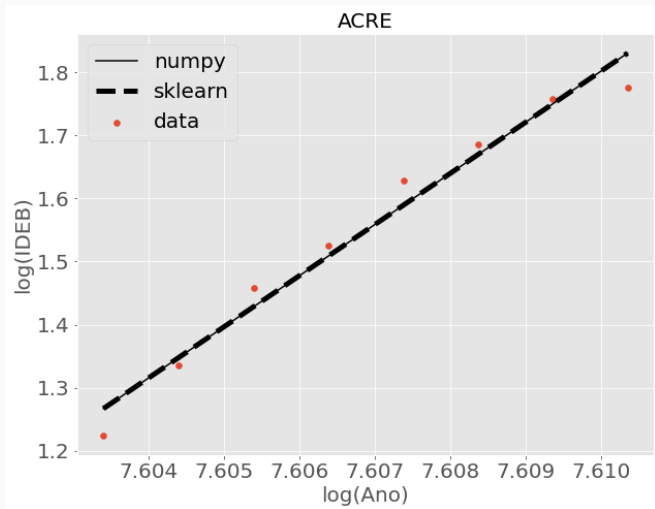
**Figura 1:** Gráfico de Dispersão para o estado do Acre.

# Gráfico modelado do Acre



**Figura 2:** Gráfico de Dispersão para o estado do Acre contendo o logaritmo natural.

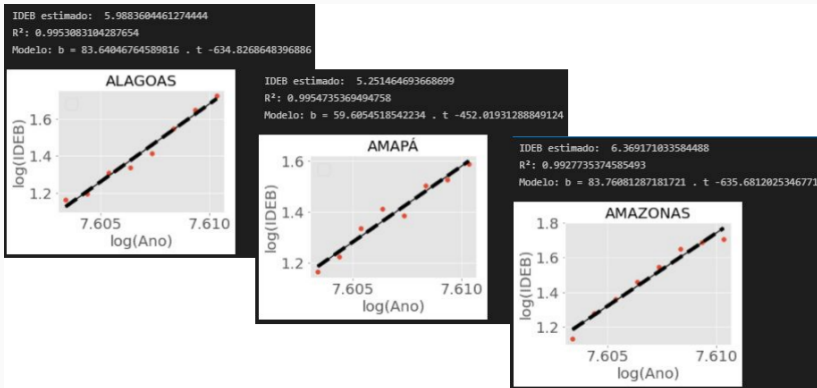
# Gráfico do Acre com eixos do NumPy e Scikit-Learn



**Figura 3:** Gráfico de Dispersão para o estado do Acre comparando as regressões pela fórmula e pelo Scikit-Learn.

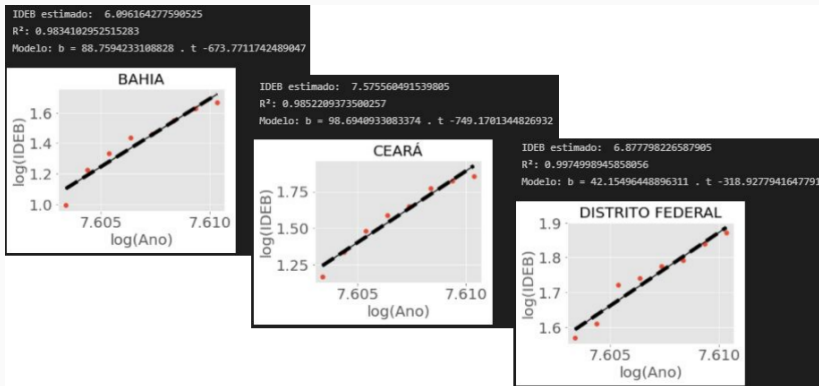


# Gráficos: Alagoas, Amapá e Amazonas



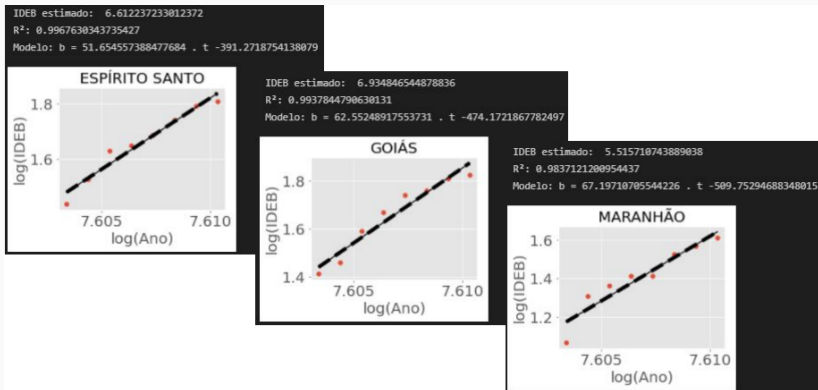
**Figura 4:** Gráfico de Dispersão para os estados de Alagoas, Amapá e Amazonas.

# Gráficos: Bahia, Ceará e Distrito Federal



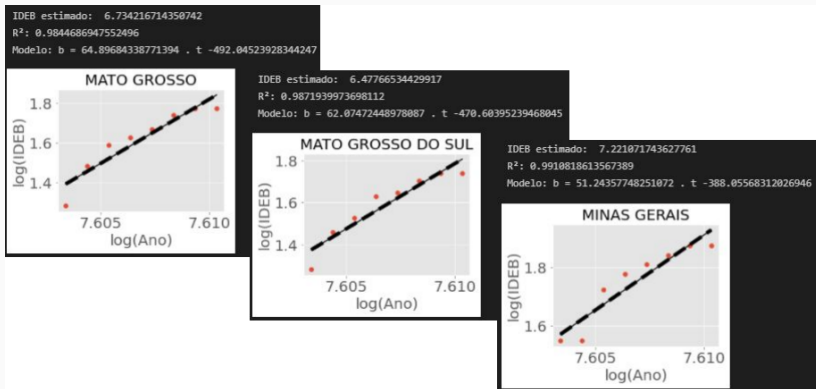
**Figura 5:** Gráfico de Dispersão para os estados da Bahia, Ceará e o Distrito Federal.

# Gráficos: Espírito Santo, Goiás e Maranhão



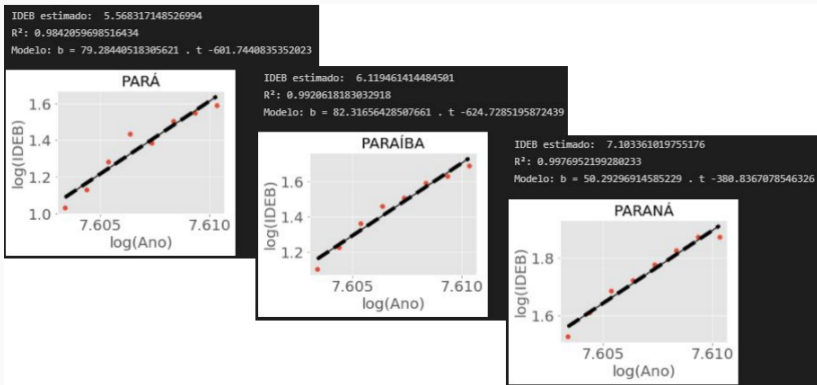
**Figura 6:** Gráfico de Dispersão para os estados do Espírito Santo, Goiás e Maranhão.

# Gráficos: Mato Grosso, Mato Grosso do Sul e Minas Gerais



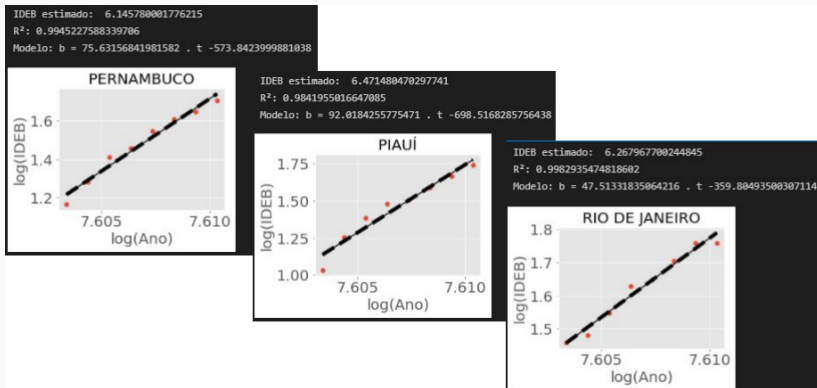
**Figura 7:** Gráfico de Dispersão para os estados do Mato Grosso, Mato Grosso do Sul e Minas Gerais.

# Gráficos: Pará, Paraíba e Paraná



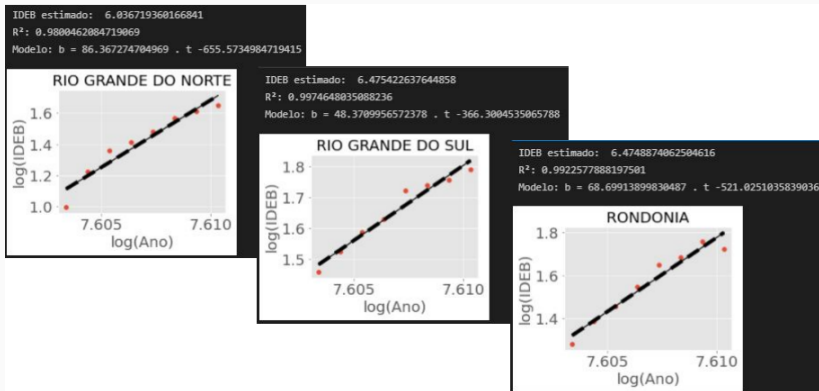
**Figura 8:** Gráfico de Dispersão para os estados do Pará, Paraíba e Paraná.

# Gráficos: Pernambuco, Piauí e Rio de Janeiro



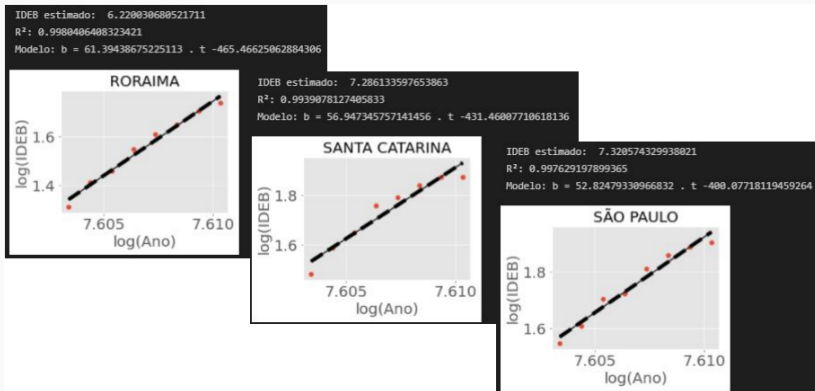
**Figura 9:** Gráfico de Dispersão para os estados de Pernambuco, Piauí e Rio de Janeiro.

# Gráficos: Rio Grande do Norte, Rio Grande do Sul e Rondônia



**Figura 10:** Gráfico de Dispersão para os estados do Rio Grande do Norte, Rio Grande do Sul e Rondônia.

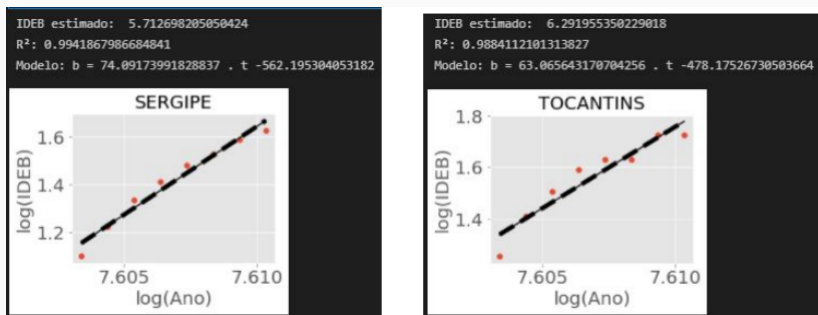
# Gráficos: Roraima, Santa Catarina e São Paulo



**Figura 11:** Gráfico de Dispersão para os estados de Roraima, Santa Catarina e São Paulo.



# Gráficos: Sergipe e Tocantins



**Figura 12:** Gráfico de Dispersão para os estados de Sergipe e Tocantins.

# Conclusão

- A regressão pelo Scikit-Learn coincide com a feita pelo NumPy, confirmando a funcionalidade do método dos mínimos quadrados.
- Pelo cálculo do  $R^2$ , que variou entre 0.980 (correspondente ao Rio Grande do Norte) e 0.998 (correspondente ao Rio de Janeiro), podemos concluir que o modelo se ajusta bem aos dados.
- Os valores estimados pela fórmula do Rio de Janeiro, Distrito Federal e Rio Grande do Sul se assemelharam à meta estipulada pelo governo para 2021.
- Por outro lado, o Rio Grande do Norte e o Ceará apresentaram grande diferenças entre esses valores.
- Isso mostra que, para estipular uma meta para o IDEB, não basta apenas observar sua regressão, mas também o contexto de cada localidade.