

**Condati Consultoria LTDA.**

**RELATÓRIO METODOLÓGICO**  
**DADOS HISTÓRICOS DE EDUCAÇÃO NO BRASIL**  
**1900-2015**

**Abril de 2023**

## Introdução

Este documento apresenta o relatório metodológico dos seguintes produtos fornecidos pela Condati Consultoria LTDA. para a FGV IBRE em dezembro de 2022.

1. Matrículas no ensino primário, 1871-2010
2. Matrículas e taxas brutas de matrícula por cor, 1960-2010
3. Anos médios de estudo, 1925-2015
4. Matrículas e taxas brutas de matrícula (ensino fundamental, ensino médio e ensino superior), 1933-2010
5. Despesas públicas em educação (ensino fundamental, ensino médio e ensino superior), 1933-2010
6. Matrículas e taxas brutas de matrícula por unidade da federação, 1955-2010
7. Razão de distribuição por série (*grade distribution ratio – GDR*), 1955-2010

A listagem acima segue o estabelecido em contrato. Esse relatório metodológico está dividido em três grandes blocos:

1. Matrícula
2. Despesa
3. Anos de escolaridade

Alguns dos aspectos metodológicos mencionados aqui podem ser encontrados nos seguintes artigos ou *working papers*, que serão citados ao longo deste documento:

- Thomas H. Kang, Luis F. Z. Paese e Nilson F. A. Felix (2021), “Late and unequal: measuring enrolments and retention in Brazilian education, 1933-2010”, *Revista de História Económica - Journal of Iberian and Latin American Economic History*, 39 (2), pp. 191-218, 2021.
- Thomas H. Kang e Isabela Menetrier (2023), *Políticas elitistas e despesas em educação no Brasil, 1933-2010*, working paper.
- Julia Walter e Thomas H. Kang (2022), *A new dataset of average years of schooling in Brazil, 1925-2015*, working paper.

# 1. Matrícula

Esta seção apresenta cinco produtos relacionados a dados de matrícula:

- Matrículas no ensino primário, Brasil, 1871-2010;
- Matrículas e taxas brutas de matrícula (ensino fundamental, médio e superior), Brasil, 1933-2010;
- Matrículas e taxas brutas de matrícula (ensino fundamental), Estados, 1955-2010;
- Matrículas e taxas brutas de matrícula por cor, Brasil, 1960-2010.
- Razão de distribuição por série (*grade distribution ratio* – GDR), 1955-2010.

## ***Matrículas no ensino primário, Brasil, 1871-2010 (Produto 1)***

As matrículas no ensino primário foram obtidas por meio de documentos oficiais, principalmente edições do Anuário Estatístico do Brasil (AEB). Tratam-se de estatísticas de matrícula bruta, ou seja, pessoas matriculadas nos seguintes níveis: (i) antigo ensino primário (de quatro a seis séries) antes de 1971; (ii) primeira etapa do ensino de primeiro grau (quatro séries) entre 1971 e 1996; e (iii) anos iniciais do ensino fundamental (quatro séries para ingressantes antes de 2007, cinco séries para ingressantes de 2007 em diante) de 1996 em diante.

## ***Matrículas e taxas brutas de matrícula (ensino fundamental, médio e superior), Brasil, 1933-2010 (Produto 4)***

As matrículas nos níveis de ensino fundamental, médio e superior provêm das mesmas fontes dos dados anteriores. A base de dados utiliza as denominações pós-1996: ensino fundamental (EF – oito ou nove séries), ensino médio (EM - três séries) e ensino superior (ES). O ensino fundamental aparece por vezes dividido em anos iniciais (EF1 – quatro a seis séries, chamado de elementar ou primário antes de 1971) e anos finais (EF2 – quatro séries, chamado de primeiro ciclo do ensino médio antes de 1971).

Como essas informações podem ser confusas para quem não está familiarizado com as mudanças históricas da legislação educacional, a Tabela 1 mostra a correspondência entre os níveis de ensino conforme a legislação antes e depois da reforma do ensino de 1971.

Em seguida, apresentam-se as taxas brutas de matrícula, que são obtidas pela divisão do número total de matrículas sobre a faixa etária considerada adequada para aquela faixa de ensino, como mostra a Tabela 2.

Tabela 1 - Níveis de ensino conforme legislação antes e depois da Reforma do Ensino em 1971.

Antes de 1971		Depois de 1971	
Nível de ensino	Subdivisões	Nível de ensino	Subdivisões
Elementar ou primário	(4 a 6 anos)	1º grau ou fundamental (EF) (8 ou 9 anos)*	Anos iniciais (EF1) (4 anos ou 5 anos)*
Médio (7 anos)	1º ciclo (4 anos)		Anos finais (EF2) (4 anos)
	2º ciclo (3 anos)	2º grau ou médio (EM) (3 anos)	(3 anos)
Superior		3º grau ou superior (ES)	

Nota: \* Foi acrescido um ano nos anos iniciais do ensino fundamental em 2007.

Tabela 2 – Faixas etárias consideradas adequadas para cada nível de ensino

Nível de ensino	Faixa etária	Subdivisão	Faixa etária
<i>Ensino fundamental</i>	7-14	Anos iniciais	7-10 ou 7-11
		Anos finais	11-14
<i>Ensino médio</i>		15-17	
<i>Ensino superior</i>		18-22	

De 1940 em diante, utilizamos os dados de idade simples dos Censos Demográficos (inclusive os dados de 1940 e 1950 foram digitados manualmente, uma vez que o site do IBGE não disponibiliza dados de idade simples em formato eletrônico para esses censos). Em anos entre os Censos, recorreu-se a interpolações. Para o período 1933-1939, obtivemos as estimativas da tese de Maduro (2007).<sup>1</sup>

### ***Matrículas e taxas brutas de matrícula (ensino fundamental), Estados, 1955-2010 (Produto 6)***

As matrículas por UF entre 1955 e 2010 estão disponíveis para o ensino fundamental (séries 1-8). A principal fonte até o final da década de 1980 é um relatório de Goldenberg (1990).<sup>2</sup> A partir de 1995, a fonte são dados do INEP.

<sup>1</sup> Maduro, P. (2007). Taxas de matrícula e gastos em educação no Brasil. Dissert. Mestrado. EPGE - Fundação Getúlio Vargas.

<sup>2</sup> Goldenberg, M. (1990). Ensino Regular de 1 Grau: Séries Temporais 1955/87. Relatório técnico. Brasília: MEC.

Para consistência estatística ao longo do tempo, utilizamos os estados brasileiros como definidos em 1940. A Tabela 3 mostra as correspondências com os estados em 2020. Mais detalhes sobre as fontes e a metodologia podem ser encontrados em Kang et al. (2021).

Tabela 3 – Estados brasileiros, 1940 and 2020

Região	Estados em 1940	Estados em 2020
Norte	Amazonas (AM)	Acre (AC)
		Amazonas (AM)
		Roraima (RR)
		Rondônia (RO)
	Pará (PA)	Amapá (AP)
		Pará (PA)
Nordeste	Alagoas (AL)	Alagoas (AL)
	Bahia (BA)	Bahia (BA)
	Ceará (CE)	Ceará (CE)
	Maranhão (MA)	Maranhão (MA)
	Paraíba (PB)	Paraíba (PB)
	Pernambuco (PE)	Pernambuco (PE)
	Piauí (PI)	Piauí (PI)
	Rio Grande do Norte (RN)	Rio Grande do Norte (RN)
	Sergipe (SE)	Sergipe (SE)
Sudeste	Espírito Santo (ES)	Espírito Santo (ES)
	Minas Gerais (MG)	Minas Gerais (MG)
	Rio de Janeiro (RJ)*	Rio de Janeiro (RJ)
	São Paulo (SP)	São Paulo (SP)
Sul	Paraná (PR)	Paraná (PR)
	Rio Grande do Sul (RS)	Rio Grande do Sul (RS)
	Santa Catarina (SC)	Santa Catarina (SC)
Centro-Oeste	Goias (GO)	Goiás (GO)
		Tocantins (TO)**
	Mato Grosso (MT)	Mato Grosso (MT)
		Mato Grosso do Sul (MS)

Notas: \* Até 1960, a capital era o Rio de Janeiro e, portanto, o antigo Distrito Federal localizava-se na Região Sudeste. Após a construção de Brasília, a nova capital, o Distrito Federal mudou-se para a Região Centro-Oeste.

\*\* Tocantins é atualmente parte da Região Norte.

### ***Razão de distribuição por série (grade distribution ratio – GDR), 1955-2010 (Produto 7)***

A metodologia de razão de distribuição por série (grade distribution ratio ou GDR) vem de Frankema (2009). Em sua utilização mais comum, utilizam-se os dados de matrícula das seis primeiras séries. Nesse caso, designamos a estatística como  $GDR_{1-6}$ , que se trata da razão entre a soma das matrículas das séries 4 a 6 e a soma das matrículas das séries 1 a 3:

$$GDR_{1-6} = \frac{\sum_{i=4}^6 g_i}{\sum_{i=1}^3 g_i}$$

em que  $g_i$  é a matrícula na série  $i$ .

Índices altos indicam menores níveis de retenção, enquanto indicadores baixos estão normalmente relacionados a altos índices de repetência ou evasão. Como o número de ingressantes não é o uniforme ao longo do tempo, esse indicador poderia ser em teoria maior do que 1. Explicação mais detalhada da aplicação do GDR para o caso brasileiro pode ser obtida em Kang et al. (2021).

### ***Matrículas e taxas brutas de matrícula por cor, Brasil, 1960-2010 (Produto 2)***

As estimativas de matrícula e taxas brutas de matrícula por cor entre 1960 e 2010 foi realizada por meio de procedimentos estatísticos, uma vez que os AEB não fornecem dados de matrícula por raça. A seguir apresenta-se o passo-a-passo das estimações realizadas.

#### *Estimativa da população por idade e raça de 1950:*

- 1) A população por faixa etária e raça para o ano censitário de 1950 ( $\hat{P}_{i=1950}^{f,r}$ ) – contemplando as faixas etárias  $f$  de 7 a 14 anos, 15 a 17 anos e 18 a 24 anos, foi estimada utilizando-se a mesma proporção da população por cor existente na população total dentro de cada faixa etária.

#### *Estimativa da população por idade e raça entre 1950 e 2010:*

- 1) A população por raça e faixa etária  $P_{i=1960,1980,1991,2000,2010}^{f,r}$ , para cada ano  $i$ , conforme a faixa etária  $f$  (7-14, 15-17 e 18-24) e a raça  $r$  (branca, parda, preta e amarela), para os anos censitários, foi obtida através dos microdados dos Censos 1960, 1980, 1991, 2000 e 2010.
- 2) Estimou-se a população para cada faixa etária e raça,  $\hat{P}_{i=[1950;2010]}^{f,r}$  para todos os anos, aplicando interpolações polinomiais, passando pelos pontos das populações do ano de 1950 (estimada na seção anterior) e dos anos censitários de 1960, 1980, 1991, 2000 e 2010.

#### *Estimativa do número de matrículas por idade e raça entre 1960 e 2010:*

- 1) O número total de matrículas  $M_i^g$  para cada ano  $i$ , conforme o grau  $g$  (1º, 2º e 3º), foi obtido de Kang et al. (2021) para todos os anos.
- 2) Posteriormente, obteve-se a frequência bruta à escola para cada grau e raça  $FB_{i=1960,1980,1991,2000,2010}^{g,r}$  nos anos censitários de 1960, 1980, 1991, 2000 e 2010.
- 3) Para os anos de 1960, 1980, 1991, 2000 e 2010, estimou-se o número de matrículas por grau e raça ( $\hat{M}_i^{g,r}$ ) utilizando-se o total de matrículas  $M_i^g$  obtido por Walter e Kang (2022) e

posteriormente aplicando-se a mesma proporção obtida na frequência bruta em cada raça nos respectivos Censos, de forma pró rata.

- 4) Para estimar o número de matrículas por grau e raça entre 1960 e 2010, para todos os anos,  $(\widehat{M}_{i=[1960,2010]}^{g,r})$  aplicou-se interpolações polinomiais entre os anos censitários acima citados.
- 5) Para estimar a taxa bruta de matrícula  $\widehat{TBM}_{i=[1960,2010]}^{g,r}$ , para cada grau e raça, para o período de 1960 a 2010, dividiu-se  $\widehat{M}_{i=[1960,2010]}^{g,r}$  pela população obtida na seção anterior  $(\widehat{P}_{i=[1960;2010]}^{f,r})$ .
- 6) De forma pró-rata,  $\widehat{M}_{i=[1960,2010]}^{g,r}$  foi reestimada, com ligeiras alterações, de tal modo que a soma das matrículas por cor em cada grau totalizasse o número de matrículas em cada grau fornecidos por Kang et al. (2021).

## 2. Despesa

### *Despesas públicas em educação (total e por nível de ensino), Brasil, 1933-2010 (Produto 5)*

As despesas públicas em educação foram obtidas por duas fontes principais: (i) AEB e (ii) Finanças do Brasil. Adicionalmente, utilizamos informações de outros documentos de entidades estatais (MEC, IPEA e STN), como Kang e Menetrier (2022) mostram em detalhes. Maduro (2007) é a principal fonte secundária.

Em resumo, a metodologia foi a seguinte:

- 1) Até 1975, os gastos totais em educação e cultura de cada nível de governo foram retirados de edições das FINBRA;
- 2) Quando o FINBRA não estava disponível, procuramos fontes que eram em princípio compatíveis com o FINBRA (principalmente os AEB e outros documentos do MEC);
- 3) Outras fontes determinaram os *shares* de cada nível de escolaridade sobre as despesas totais em educação e cultura;
- 4) A soma das despesas de cada nível de escolaridade (primário, secundário e terciário) é igual ao total de gastos.

O procedimento (1), no entanto, não foi empregado no período 1975-1993 por conta de mudanças no cômputo da estatística oficial. Melchior (1980, p. 52) relata que, em 1975, “várias contas foram deslocadas também para a função educação e cultura”, incluindo verbas gastas em ciência e tecnologia que antes não constavam nesta rubrica.<sup>3</sup> Para evitar essas estimativas viesadas para cima, utilizou-se uma estimativa já tratada de despesas em educação para a década de 1980 que consta no AEB 1993, conforme feito também por Maduro (2007). Para a segunda metade da década de 1970, recorreremos a interpolações na variável despesa total em educação, ainda que tenhamos continuado a usar os *shares* de nossas fontes do MEC.

A partir das despesas públicas em educação, calcularam-se os seguintes indicadores:

- Despesas públicas em educação como proporção do PIB (total, EF, EF1, EF2, EM e ES)
- Despesas públicas em educação por estudante como proporção do PIB per capita (total, EF, EF1, EF2, EM e ES)

---

<sup>3</sup> Em recente livro, Antônio Gois menciona o texto de Melchior em recente livro sobre a história da educação no Brasil. Ver Gois, Antônio (2022), *O ponto a que chegamos: duzentos anos de atraso educacional e seu impacto nas políticas do presente*, São Paulo: FGV Editora. Ver também Melchior, J. C. A. (1980), “Financiamento da educação no Brasil numa perspectiva democrática”, *Cadernos de Pesquisa*, 34, p. 39-83.

- Razão dupla de despesa por estudante (ES/EF1 e ES/(EF+EM)).

As duas primeiras variáveis não necessitam de explicação conceitual, ao contrário da razão dupla de despesa por estudante, que pode ser definido pela seguinte fórmula conforme Lindert (2004):

$$rd_{i/j} = \frac{desp_i}{matr_i} / \frac{desp_j}{matr_j}$$

em que *desp* é a despesa pública e *matr* é o número de matrículas. Como os índices *i* e *j* são os níveis de ensino, *rd<sub>i/j</sub>* representa quantas vezes a despesa por estudante no nível de ensino *i* é maior do que a despesa por estudante no nível *j*. Disponibilizamos duas séries com esse formato: a razão dupla superior/EF1 e a razão dupla superior/(EF+EM).

A metodologia de estimação de despesas públicas em proporção do PIB para os anos iniciais (EF1), anos finais do ensino fundamental (EF2) e ensino médio (EM) exige detalhamento por contas das mudanças legislativas. Entre 1933 e 1971, existe uma série de dados de despesa para o antigo ensino primário (equivalente ao EF1) e para o antigo ensino médio (equivalente à soma de EF2 com o atual ensino médio). A situação muda a partir de 1972, em que os dados são apresentados separados em primeiro grau (equivalente ao EF atual) e segundo grau (equivalente ao EM atual). Para termos uma série contínua por ciclo (EF1, EF2 e EM), procedemos da seguinte maneira:

- 1) Parte-se de estimativas de estudantes matriculados na rede pública de ensino, uma vez que não há informações da matrícula por rede para todos os anos;
- 2) No período 1933-1970, existem dados divididos em ensino primário (quatro a seis séries) e ensino médio (sete séries). Assume-se que a despesa por aluno no segundo ciclo do ensino médio (*colegial*) seja 25% maior do que a despesa por aluno no ensino médio (ambos os ciclos);
- 3) De 1971 em diante, os dados estão divididos em ensino de primeiro grau (oito séries) e segundo grau (três séries). No período 1971-1989, pressupõe-se que a despesa por estudante no primeiro ciclo do primeiro grau (1ª a 4ª série) seja 20% menor do que primeiro grau considerando ambos os ciclos (1ª a 8ª série);
- 4) De 1990 a 1993, aplicam-se interpolações devido a falta de dados;
- 5) A partir de 1994, parte-se da hipótese de que a despesa por estudante no primeiro ciclo do ensino fundamental (1ª a 4ª série) seja igual ao do primeiro grau considerando ambos os ciclos (1ª a 8ª série/9º ano). Tenta-se refletir a mudança de ênfase havida na política pública, que passou a priorizar a universalização do ensino fundamental.

Detalhes desses procedimentos podem ser encontrados em Kang e Menetrier (2022).

### 3. Anos de escolaridade

*Anos médios de estudo da população entre 15 e 64 anos de idade, Brasil, 1925-2015 (total e por cor e gênero) e 1950-2015 (por UF e região).*

A série de anos médios de escolaridade foi obtida de Walter e Kang (2022), onde pode se consultar as seções de fontes dos dados e metodologia. As fontes principais de dados são os Censos Demográficos e pesquisas domiciliares do IBGE. Assim como o conhecido estudo de Barro e Lee (2013), também se utilizam *benchmarks* como parâmetros com uma adaptação da metodologia de Földvári e Van Leeuwen (2009) para estimação dos anos intercensitários. As estimativas obtidas são mais precisas que as de Barro e Lee (2013) para o Brasil, que apresentam uma queda implausível no estoque de escolaridade nas décadas de 1970 e 1980.<sup>4</sup>

As estatísticas obtidas se referem à população entre 15 e 64 anos (ou seja, aquela referente à força de trabalho). Foram obtidos também estatísticas para subgrupos por cor e gênero (1925-2015) e por UF e região (1950-2015). Como se trata de uma metodologia complexa e longa, recomenda-se a leitura do estudo de Walter e Kang (2022) para maiores detalhes.

---

<sup>4</sup> Barro, R. e Lee J. H. (2013), "A New Data Set of Educational Attainment in the World , 1950 – 2010." *Journal of Development Economics* 104: 184–98. <https://doi.org/10.1016/j.jdeveco.2012.10.001>.

Földvári, P. e Van Leeuwen, B. (2009). "Average Years of Education in Hungary: Annual Estimates, 1920-2006." *Eastern European Economics* 47 (2): 5–20. <https://doi.org/10.2753/EEE0012-8775470201>.