

Educação e as Origens da Desigualdade Regional no Brasil

Versão para Submissão ao 44º Encontro Anual de Economia

Bruno Komatsu

Naercio Menezes Filho

Pedro Oliveira

Leonardo Viotti

Centro de Políticas Públicas do Insper e USP

Resumo

Esse artigo busca entender as origens da desigualdade regional no Brasil, enfatizando o papel da educação. Combinamos diversas fontes de dados, como os Censos Demográficos e os Anuários Estatísticos do Brasil do IBGE com dados do final do século XIX e início do século XX, para construir medidas de estoque de pessoas com diferentes níveis de escolaridade de 1900 a 2000 para diferentes estados brasileiros e para o Brasil como um todo. A partir dessas medidas, construímos um índice de Gini educacional que mostra que a desigualdade era muito elevada no início do século passado, permaneceu elevada até 1920, declinou lentamente entre 1920 e 1950 e mais rapidamente a partir de então. Em 1900 a desigualdade educacional era a mesma em São Paulo e Bahia, mas a partir de então declinou muito mais rapidamente em São Paulo o que faz com que a desigualdade de renda seja maior na Bahia do que em São Paulo hoje em dia.

Palavras-Chave: Educação, Desigualdade, Índice de Gini.

Abstract

This article aims to understand the origins of regional inequality in Brazil, emphasizing the role of education. We combine various data sources such as the Demographic Census and Brazil's Statistical Yearbooks of IBGE with the late nineteenth century and early twentieth century data, to build measures of people with different levels of education from 1900 to 2000 for different Brazilian states and Brazil as a whole. From these measurements, we build an educational Gini index, which shows that inequality was very high at the beginning of the last century, remained high until 1920, declined slowly between 1920 and 1950 and reduced faster thereafter. In 1900, educational inequality was the same in São Paulo and Bahia, but from then declined much faster in São Paulo, what causes income inequality to be greater in Bahia than in São Paulo nowadays.

Key Words: Education, Inequality, Gini Index.

Área Anpec: Área 3 – História Econômica

JEL: I24 – Education and Inequality, I25 – Education and Economic Development

1. Introdução

Apesar das importantes transformações ocorridas no período recente e que levaram à acentuada redução da desigualdade de renda no Brasil na última década, o país figura ainda hoje entre os países com maior nível de desigualdade no mundo¹. As causas que levaram a este resultado são bastante controversas, e alvo de um grande número de pesquisas. Na literatura referente ao assunto, é possível enxergar algumas vertentes principais, que observam uma desigualdade histórica e a associam ao processo de colonização e às instituições resultantes, ou que entendem que a desigualdade, historicamente baixa, se elevou principalmente entre os séculos XIX e XX.

Analisando o continente latino-americano, Acemoglu *et al.* (2001) colocam as instituições como principal fator explicativo para a alta desigualdade observada na região. Neste sentido, os autores argumentam que diferentes formas de colonização adotadas pelos europeus levaram a diferentes configurações institucionais, que geraram resultados socioeconômicos também diferenciados. Em especial, nas colônias de exploração, a qualidade das instituições herdadas do período colonial seria menor, gerando um maior nível de desigualdade. Em uma visão similar, Engerman e Sokoloff (1994) destacam o papel da dotação inicial de fatores e recursos naturais e sua influência sobre a formação e o funcionamento das instituições. Em contrapartida, autores como Williamson (2015) argumentam que, a desigualdade não esteve sempre em níveis altos no continente latino-americano, mas teria se elevado como consequência de transformações ocorridas entre os séculos XIX e XX, tendo como ponto de partida o *boom das commodities* do final do século XIX.

Além dessas duas abordagens principais, outra importante linha de explicação, muito presente na literatura, relaciona a alta desigualdade e o relativo atraso do continente latino-americano com a questão educacional (Camps e Engerman, 2014). Tendo em vista que a educação se desenvolveu tardiamente e de forma muito desigual na região, priorizando o acesso das elites, e considerando que aqueles mais escolarizados são melhor remunerados devido à escassez de trabalho qualificado, é possível compreender a razão de a renda ser tão desigualmente distribuída. Musachio, Martinez e Viarengo (2014) abordam as três hipóteses para o caso brasileiro, dando destaque à educação e à influência das instituições coloniais sob o desenvolvimento desta.

Tendo como base essa literatura, o presente trabalho tem como intuito analisar de forma detalhada o quanto grande era a desigualdade educacional do Brasil no período que vai do início do século XX até a década de 1960. Para tanto, utilizamos dados históricos dos AEBs (Anuários Estatísticos do Brasil) e dos Censos Demográficos. Com isso, foi possível estimar um índice Gini Educacional, que nos dá um panorama mais claro da evolução e do comportamento dessas desigualdades educacionais, e como esta respondeu às reformas empreendidas pelo governo, na tentativa de expandir o sistema educacional. Além disso, nos permite avaliar como a educação se comportou em algumas regiões do país e em que medida esta é responsável pela desigualdade regional existente ainda hoje.

O artigo está dividido em sete partes, além desta breve introdução. Na segunda seção apresentaremos o debate da literatura econômica acerca das causas da desigualdade no Brasil e na América Latina. Depois disso, realizamos uma breve comparação dos níveis de instrução do Brasil com alguns outros países e entre as regiões brasileiras. Na quarta parte, faremos uma descrição dos dados utilizados, bem como de suas fontes, e das especificidades de cada um destes. Em seguida, explicaremos detalhadamente a metodologia utilizadas no trabalho. Partiremos então para a apresentação dos resultados obtidos e por fim apresentaremos algumas conclusões sobre o assunto.

¹ The World Factbook disponível em: <<https://www.cia.gov/library/publications/the-world-factbook/fields/2172.html>>.

2. Revisão Bibliográfica

A problemática da desigualdade social vem ganhando destaque nas mais diversas áreas de pesquisa. Isto, porque estudos recentes apontam um aumento da desigualdade em todo o mundo (Dabla-Norris, 2015). Observando regionalmente, vemos que mesmo com as transformações ocorridas nos últimos anos, como mudanças políticas, maior crescimento do PIB e políticas de combate à extrema pobreza, a América Latina permanece sendo uma das regiões mais desiguais (OCDE, 2014). Há, na literatura sobre o assunto, uma série de hipóteses a fim de explicar as possíveis causas para a alarmante desigualdade na região.

Acemoglu *et al.* (2001) mostram que a qualidade das instituições fundadas durante o período colonial estão fortemente associadas à qualidade daquelas existentes hoje em dia. Neste sentido, o tipo de colonização adotado pelos europeus é um fator importante para explicar a desigualdade atual. Basicamente, duas formas de colonização são apresentadas: uma primeira, fundamentada no estabelecimento dos colonizadores europeus nas colônias, conhecida como “colônia da povoamento”, e, uma segunda, baseada na exploração, caso em que as colônias atendiam apenas a fins extrativos, chamadas de “colônias de exploração”. Vale frisar que o tipo de colonização não é arbitrário, mas sim condicionado às características de cada região. Em regiões com taxas de mortalidade mais elevadas, colonizadores enfrentavam maiores dificuldades pra estabelecer-se e, com consequência, tinham uma maior probabilidade de estabelecer instituições extrativas

Cada forma de colonização legaria diferentes configurações institucionais às suas respectivas localidades. Nos locais em que se estabeleceram, isto é nas “colônias de povoamento”, os europeus fundaram instituições melhores, mais desenvolvidas e mais democráticas, que favoreciam investimentos e crescimento econômico, ao passo que nas chamadas “colônias de exploração”, instituições piores foram criadas.

Por sua vez, Williamson (2015) apresenta uma visão bastante diferente. O autor discorda da ideia de que a desigualdade de renda sempre foi elevada no continente latino-americano, ou que seria simplesmente um produto da colonização ibérica. No lugar desse argumento, argumenta que a desigualdade teria se elevado entre os séculos XIX e XX, como consequência de alguma transformação ocorrida no período e apresenta evidências empíricas de que a desigualdade na América Latina era baixa no período colonial – e mesmo antes deste –, se comparada aos países industriais. De fato, o autor encontra um índice de Gini menor em alguns países latino-americanos naquele período do que os observados atualmente. Isto pode ser explicado pelo fato de que a América Latina era uma região pobre, de forma que o excedente que poderia ser absorvido pelas elites era muito baixo.

Nessa visão, o *boom* das *commodities* do final do século XIX teria sido o ponto inicial para o aumento da desigualdade na América Latina. Após a Primeira Guerra Mundial, um forte sentimento de antiglobalização teria gerado um movimento isolacionista no continente. Entre as décadas de 1920 e 1970, enquanto praticamente todas as outras regiões do globo passaram por um processo de nivelamento de renda, a desigualdade latino-americana aumentou.

É possível identificar, na hipótese proposta por Williamson (2015), uma referência à curva de Kuznets. Isto é, de acordo com as evidências apresentadas no estudo, uma curva de Kuznets pré-industrial se adequa a situação observada na América Latina. No início da fase pré-industrial, de baixo desenvolvimento, a desigualdade era baixa, porém se elevou continuamente até alcançar um pico, caindo daí em diante, durante uma fase de crescimento econômico moderno.

Há uma clara dicotomia entre essas visões, e ambos os autores consideram que apenas estes fatores seriam fundamentais na explicação da alta desigualdade observada no continente latino-americano. Camps e Engerman (2014), no entanto, abordam a questão a partir de um outro ângulo, colocando os baixos níveis educacionais da região como o principal gerador de desigualdade e atraso em comparação a países mais desenvolvidos. Entretanto, não descartam a importância das características iniciais e da dotação de fatores de cada local para explicar a desigualdade. Em sua visão, a forma de produção que predominava no

continente – baseada em *plantations* – era amplamente concentradora, e esta concentração não se restringia à riqueza e propriedade, mas chegava também ao poder político e principalmente à educação.

Desse modo, apenas uma pequena proporção da população era escolarizada, o que implicava em baixas taxas de alfabetização. Por volta do fim do século XIX, a necessidade de uma reforma que expandisse o sistema educacional e atendesse às demandas sociais era evidente. Todavia, esta reforma só veio a ocorrer de forma significativa no início do século XX. Ainda, quando a escolarização em massa chegou aos países latino-americanos, beneficiou especialmente a população branca composta em parte por imigrantes, relegando à segundo plano nativos americanos, negros e mestiços.

Musachio *et al.* (2014) acabam por relacionar os três argumentos expostos anteriormente. Analisando o Brasil, demonstram como as instituições originadas no período colonial geraram desigualdade na distribuição da renda e da riqueza. Contudo, uma mudança no quadro institucional, em conjunto com um choque de comércio no século XIX (*boom das commodities*), alterou a estrutura educacional do país, e elevou a desigualdade.

O ponto chave deste argumento é de que as regiões brasileiras se especializaram na produção de bens diferentes. Primeiramente, é importante salientar que o *boom das commodities* não beneficiou de forma igualitária todos os produtos. Na realidade, o preço do café – produto do qual o Brasil era um dos principais produtores mundiais – foi o que mais se elevou. Também uma descentralização fiscal ocorrida no período passou para as unidades da federação a responsabilidade de investir em educação e as receitas de arrecadação com exportação, favorecendo estados produtores de *commodities* com alta demanda.

Regiões exportadoras de café – como é o caso da região Sudeste – dispunham de um volume maior de recursos, o que possibilitava um maior investimento em educação. Já onde se exportava açúcar, algodão e tabaco, as receitas eram limitadas, o que inviabilizava grandes investimentos em educação. Assim, Musachio *et al.* (2014) encontram uma relação positiva entre o preço dos produtos exportados, receitas tributárias destes produtos e o gasto *per capita* em educação. Com isso, se observa uma expansão educacional nas primeiras décadas do século XX regionalmente extremamente desigual, e que segregava indivíduos por classe social e raça. Além disso, os autores mostram que não importava apenas qual *commodity* a região exportava, mas sim em qual sistema de produção esta estava inserida, e o tipo de instituições que estabeleceu durante o século XIX.

O papel fundamental da educação para explicação da alta desigualdade existente no Brasil também é colocado por Rocha, Ferraz e Soares (2015). Observando o período pós 1850, o estudo foca na influência que o capital humano e as imigrações tiveram sob o desenvolvimento brasileiro. Com a interrupção do tráfico de escravos, o governo lançou mão de políticas para atrair imigrantes europeus para determinados locais do país. Esse “povoamento” tinha objetivo não só de suprir a demanda por mão-de-obra, mas também de branquear a população, e ocupar o território. Além disso, esses imigrantes eram em geral mais escolarizados.

As regiões que receberam maior influxo de imigrantes sofreram um choque de capital humano, São Paulo sendo o principal exemplo. Essas regiões também passaram a demandar maior investimento em educação, o que fez com que a desigualdade regional de educação se perpetuasse. Os autores associam um maior nível de capital humano ao progresso industrial ao longo do tempo, e consequentemente a um maior nível de renda *per capita*.

As implicações que esta desigualdade educacional trouxe para a desigualdade de renda é observada por Langoni (1973), que analisa as causas para o aumento da desigualdade de renda ocorrido entre as décadas de 1960 e 1970, e mostra que por volta de 33% desse aumento pode ser explicado pela variável educação. Também, Paes de Barros e Reis (1991) investigam a desigualdade salarial de acordo com escolaridade em várias regiões do Brasil. Utilizando como medida o índice de Theil, os autores observam

que o diferencial salarial por grau de escolaridade é maior nas regiões menos desenvolvidas do país, a saber a região Nordeste, o que significa que a importância da educação para a desigualdade é maior nestes locais.

Como exposto por Paes de Barros e Reis (1991), a desigualdade inter-regional no Brasil também é bastante significativa. Leff (1972) analisa a desigualdade entre as regiões Sudeste e Nordeste, remontando ao século XIX, a fim de explicar o surgimento da desigualdade regional em meio ao processo de desenvolvimento econômico. O autor nota diferenças relevantes já na década de 1850. Em sua concepção, a desigualdade regional pode ser explicada pelo diferencial no crescimento das exportações de cada região. Isso se deve a mudança na pauta de exportações brasileira ao longo do século XIX. Como mencionado anteriormente, a participação do café no total das exportações cresce cada vez mais, de forma que já na década de 1870 este havia substituído o açúcar e o algodão como principal produto exportado pelo Brasil.

Todavia, apenas algumas regiões, em especial o Sudeste, possuíam o tipo de terra adequado para o plantio do café, e por este motivo a região tornou-se, naturalmente, a principal zona produtora de café do país. Ademais, o café demandava uma estrutura produtiva diferente daquela empregada na produção do açúcar, com grande utilização de mão-de-obra estrangeira e assalariada. Isso permitiu a criação de um mercado interno, e a aceleração da urbanização e do desenvolvimento industrial nas zonas cafeeiras. Já no Nordeste, onde se produzia principalmente algodão e açúcar, o que se observou no período foi uma queda da renda *per capita* em termos relativos e absolutos (Furtado, 2005).

Outro ponto ressaltado por Leff é a quase estagnação, se não declínio, das exportações brasileiras de açúcar e algodão, frente a um aumento da demanda desses produtos no mercado internacional. Simonsen (2005) coloca como uma possível causa para isto, restrições da demanda impostas pelos países importadores destes bens, que priorizavam importações de suas próprias colônias. Outros argumentos que buscam explicar o resultado das exportações desses bens durante o século XIX se baseiam em: a) Políticas coloniais; b) condições da oferta; c) aspectos socioculturais; e d) atraso tecnológico do modo de produção. Entretanto, para o autor nenhum destes motivos é forte o suficiente.

Ao longo do século XIX a lucratividade do açúcar decresceu, a do algodão se manteve em patamares baixos e a do café se elevou. Essa diferença deveria ser compensada, ao menos em parte, pela deslocamento de fatores de produção – neste caso mão-de-obra, uma vez que o fator terra é fixo – em direção às regiões com maiores rendimentos. Entretanto, observou-se apenas um fluxo de mão-de-obra escrava em direção à região Sudeste, ao passo que a alocação de mão-de-obra livre pouco se alterou. Essa limitada movimentação de fatores pode ser explicada pelos altos custos de transporte do período.

Além disso, outro fator limitativo das exportações de açúcar e algodão foi a taxa de câmbio valorizada. Essa valorização ocorreu, pois o café brasileiro podia ser comercializado no mercado internacional à taxas de câmbio inferiores a que produtos como açúcar e algodão demandavam, e à medida que a participação do café na pauta de exportações brasileira se elevou, a taxa de câmbio passou a refletir cada vez mais o peso deste produto. Em resumo, a política cambial e externa era direcionada para o café, e por isso priorizava os lucros e benefícios deste.

Assim, o deslocamento das vantagens comparativas brasileiras do açúcar para o café parece ter sido a causa inicial para o aumento das desigualdades regionais. A partir disto, e da política cambial focada no café, os preços do açúcar e do algodão sofreram queda real, o que implicou em menor renda na região Nordeste. Ainda, devido às dificuldades de deslocamento dos fatores de produção, os que permaneciam em atividades de menor renda acabavam por ser sub-remunerados, gerando não só maior desigualdade inter-regional, mas também intra-regional.

É possível que a formulação de políticas públicas fosse capaz de minimizar as disparidades regionais. Entretanto, já em meados do século XIX o governo era dependente político das oligarquias cafeeiras, e ao longo deste século e no início do seguinte essa dependência só aumentou. De fato, durante boa parte da República Velha (1889 – 1930) foi a oligarquia cafeeira que governou o país, no período chamado de

“República das Oligarquias” e da “Política de café com leite”. Dessa forma, de acordo com Leff, o maior crescimento do Sudeste implicou em estagnação ou declínio do Nordeste.

É possível notar muitas semelhanças entre os argumentos expostos por Leff (1972) e por Musachio *et al.* (2014). A principal divergência entre os dois é que, como mencionado anteriormente, o segundo considera a educação. Nessa linha, Barros (2011) apresenta uma explicação para as desigualdades regionais, que dialoga não só com o estudo de Musachio *et al.* (2014), mas também com o de Rocha *et al.* (2015) e Paes de Barros e Reis (1991), mencionados anteriormente.

Em sua visão, as desigualdades regionais se devem à má distribuição de um fator de produção, o capital humano. O autor coloca que essa distribuição desigual acaba por impedir a convergência do nível de renda *per capita*, que se esperaria a partir de um modelo neoclássico (Barro e Sala-i-Martin, 1992). A distribuição desigual de capital humano tende a persistir, uma vez que há desigualdade de capital humano também entre as famílias. Nesse sentido, as famílias podem atuar como um fator concentrador ou distributivo, uma vez que a acumulação de capital humano dos pais afeta positivamente a dos filhos. Assim, na ausência de políticas públicas que busquem redistribuir educação de forma mais igualitária entre as regiões, a desigualdade regional tende a se manter por tempo indeterminado.

Barros (2011) pontua ainda as externalidades positivas associadas ao capital humano, relacionando-as tecnologia. A primeira se associa ao ganho de eficiência e aumento da produtividade dos indivíduos com maior nível de instrução. A segunda é relacionada também à produtividade, dado que a tecnologia é desenvolvida respeitando as restrições dos outros fatores de produção. Portanto, regiões com maior disponibilidade de capital humano terão tecnologias que usam mais intensivamente este fator, e em consequência serão mais produtivas.

Em síntese, embora não haja na literatura uma total concordância quanto às origens e às explicações para a alta desigualdade existente no Brasil, a relevância da educação para compreender e explicar a desigualdade de renda é inegável. Nesse sentido, nossa principal contribuição é construir índices de Gini educacionais desde o início do século XX para o Brasil e alguns Estados selecionados, para documentar a evolução da composição educacional brasileira e seu impacto nas desigualdades regionais.

3. Uma Breve Comparação Internacional e Inter-regional

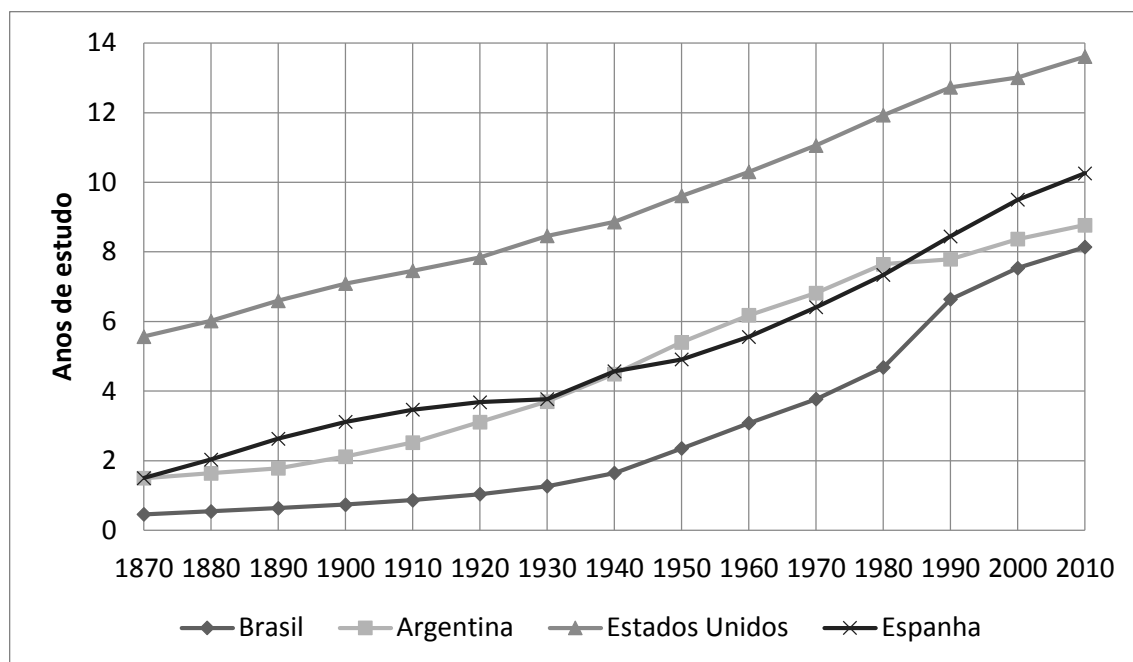
Enquanto colônia ibérica, a América Latina viu pouco progresso educacional. Alguns países europeus iniciaram importantes reformas educacionais no século XVIII. Os Estados Unidos tiveram no século XIX o seu “Século da Escola”. Em contrapartida, apenas na metade final do século XIX, algumas tentativas de reformas educacionais – não muito bem sucedidas – ocorreram nos países latino-americanos, mais especificamente na Argentina e na Colômbia. O Brasil, como veremos detalhadamente mais adiante, passou por reformas educacionais significativas apenas no século XX (Marcílio, 2005).

Essa descrição histórica indica que o Brasil pode ter algum atraso educacional em comparação a outros países, mesmo em relação a outros países latino-americanos. Mas qual é o tamanho dessa defasagem? A Figura 1 compara a evolução da média de anos de estudo para quatro países: Brasil, Argentina, Estados Unidos e Espanha. Percebe-se que a média de anos de estudo no Brasil era muito baixa no final do século XIX (0,7 anos de estudo em 1900) e assim permaneceu durante as três primeiras décadas do século XX. A partir de 1940 este indicador passou a apresentar crescimento mais acelerado no Brasil, tornando a curva mais inclinada. Ainda assim, a média de anos de escolaridade do brasileiro permaneceu abaixo de quatro anos até a década de 1960, o que significa que boa parte da população não concluía nem mesmo o ensino primário. Isso se deve em parte ao fato de que, ainda em 1960, a maior parte da população brasileira vivia em zona rural, onde havia pouco acesso à escola. Somente na década de 1990 é que a média de anos de estudo do Brasil chega ao nível observado nos Estados Unidos no início do século XX, o que explicita o grande atraso educacional do país. Apesar disso, a diferença em relação ao país que estaria na fronteira, os Estados Unidos, não se mantém constante ao longo do tempo. De 1870 a 1970 a diferença passa de 5,1 anos para 7,3 anos, porém entre 1970 e 2010 a diferença diminui, voltando ao patamar de 5,5

anos de estudo. Essa evolução provavelmente resulta das reformas educacionais adotadas pelo Brasil ao longo do século XX.

As trajetórias do indicador na Argentina e na Espanha foram bem semelhantes à do Brasil, porém num patamar mais elevado. A Argentina apresenta crescimento comparativamente mais equilibrado ao longo de todo o período, exceto por uma década de estagnação entre 1980 e 1990, década em que o Brasil mais avançou nesse indicador. A diferença entre o Brasil e a Argentina também aumenta entre 1870 e 1970, de 1 ano para 3 anos, se reduzindo a 0,6 anos em 2010. As diferenças em relação à Espanha são semelhantes, exceto pelo fato de que elas aumentam ligeiramente entre 1990 e 2010.

Figura 1 - Média de anos de estudo para países selecionados



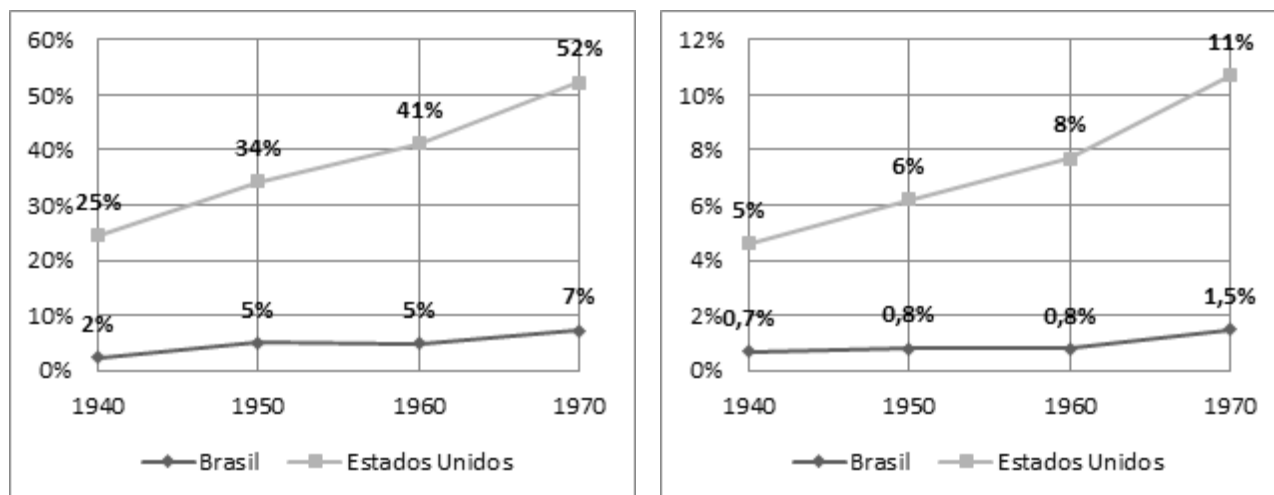
Fonte: Clio-Infra (Leeuwen et al., 2013). Elaboração própria. Dados para população de 15 anos ou mais.

Na Figura 2, mostramos uma comparação entre o percentual de indivíduos com ensino médio e superior no Brasil e nos Estados Unidos. Apesar do período relativamente mais curto, vemos que a evolução da escolaridade dos brasileiros está muito aquém à dos norte-americanos. Somente em 2010 é que a proporção de indivíduos com ensino médio no Brasil se igualou a que já era observada nos Estados Unidos na década de 1940 (IBGE, 2012). A mesma conclusão pode ser tirada quando se observa o ensino superior, sendo que somente na década de 2010 se atingiu no Brasil níveis observados nos Estados Unidos em 1970.

Figura 2 – Evolução da Escolaridade – EUA e Brasil

a) Ensino Médio

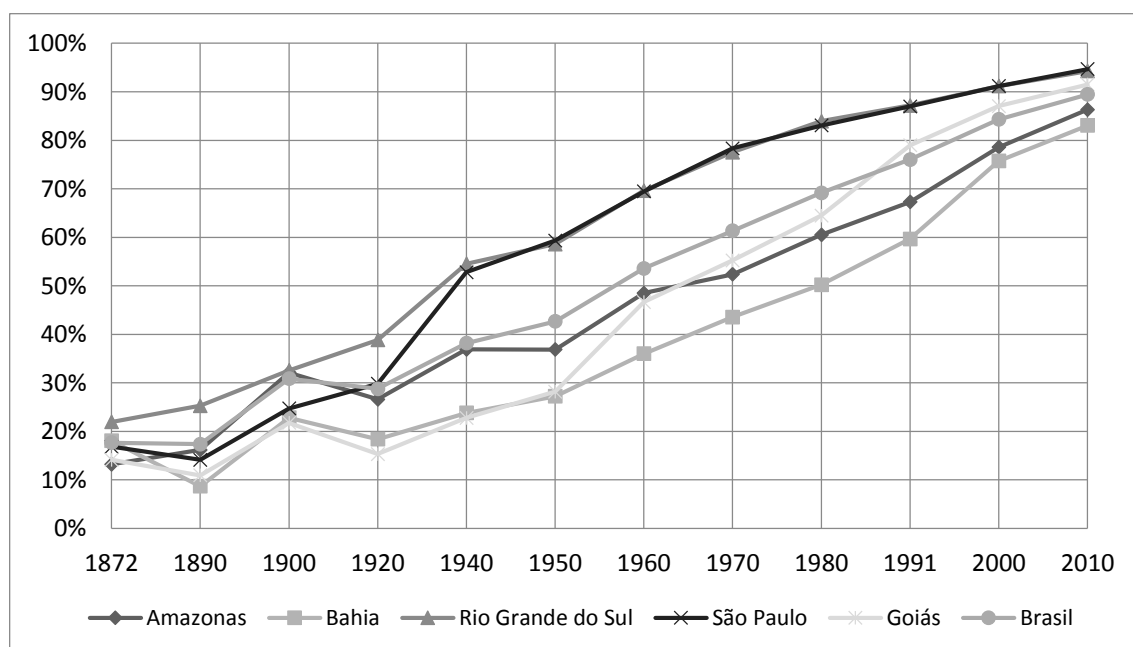
b) Ensino Superior



Fonte: Censos Demográficos/IBGE para o Brasil e Census Bureau para o EUA. Elaboração própria. Dados para população de 25 anos ou mais. Obs: Para o Brasil, dados de 1940 e 1950 são aproximados.

As diferenças no nível educacional observadas entre países podem ocorrer também internamente a um país, entre suas diferentes regiões, contribuindo para geração de desigualdade social. A Figura 3 ilustra esse ponto. Apesar do grande aumento da proporção de alfabetizados em todo o país, aqueles estados que possuíam menor proporção de indivíduos nesta condição no início do século XX, permanecem na pior posição nos dias atuais, como é o caso da Bahia e do Amazonas. A principal divergência entre essas regiões parece ter surgido entre as décadas de 1920 e 1940, quando São Paulo e Rio Grande do Sul despontaram com maior proporção de alfabetizados, posição que mantêm até hoje. Estes dois últimos estados estão, no presente, entre os mais ricos do país, o que indica uma possível correlação positiva entre taxa de alfabetização e PIB *per capita*, que pode se dar por diversas razões.

Figura 3 - Alfabetização nos principais Estados brasileiros



Fonte: Censos Demográficos/IBGE. Elaboração própria. Dados para a população de 5 anos ou mais. Obs: O Censo de 1900 omitiu alguns distritos da pesquisa.

4. Dados

A fim de estimar as estatísticas apresentadas neste *paper*, foram combinadas diversas fontes de dados históricos. Foram necessárias informações populacionais e educacionais para o total do Brasil e desagregadas por estados.

Para estatísticas como taxa de alfabetização, faixa de renda ou grau de escolaridade da população, utilizamos os Censos Demográficos, divulgados aproximadamente a cada dez anos pelo Instituto Brasileiro de Geografia e Estatística (IBGE). O primeiro recenseamento nacional ocorreu em 1872, abrangendo todas as regiões do território, exceto o atual estado do Acre.

Como fonte de dados históricos, utilizamos os Anuários Estatísticos do Brasil (AEB) também publicados pelo Instituto Brasileiro de Geografia e Estatística (IBGE), na época chamado Diretoria Geral de Estatística (DGE). O primeiro AEB foi divulgado em 1916, contendo dados para os anos entre 1907-1912. Apenas a partir de 1936 a publicação passou a ser anual. Os Anuários reúnem uma grande quantidade de informações sobre educação, saúde, gasto público, trabalho, política, entre outros temas, entretanto, estes dados não são sempre comparáveis. Utilizamos as informações disponíveis acerca de matrículas, conclusões de curso e números de instituições de ensino.

Com intuito de obter um percentual de indivíduos matriculados no nível escolar correspondente a cada faixa etária, cruzamos os dados dos AEBs com as estimativas populacionais dos Censos Demográficos. Porém, devido às faixas etárias dos censos não serem exatamente compatíveis às de cada nível escolar, foi necessário realizar uma aproximação. Assim, para o ensino primário foi considerada a população de cada estado, entre 5 e 14 anos. Para o ensino secundário e técnico, o total de pessoas por estado entre 10 e 19 anos, e para o ensino superior indivíduos entre 20 e 24 anos.

Para 1930 utilizamos estimativas populacionais feitas pelo DGE, uma vez que nesta década não foi realizado Censo Demográfico. Entretanto, estas estimativas eram apenas para o total da população de cada estado. Para chegar à população por faixa etária, foi feita uma média entre o percentual da população em cada faixa etária por estado dos Censos de 1920 e 1940. Em seguida, este resultado foi multiplicado pela população total do estado correspondente em 1930.

Dados de matrículas no ensino primário, secundário e técnico no final do século XIX foram obtidas dos Relatórios Anexos ao Ministério dos Negócios do Império, da Diretoria Geral de Estatística (Directoria Geral de Estatística, 1875; Secretaria d'Estado dos Negocios no Imperio, 1887).

Utilizamos dados de mortalidade do Brasil como um todo novamente dos AEBs. Para o estado de São Paulo, utilizamos dados de óbitos do projeto Memória das Estatísticas Demográficas (SEADE, 2010).

5. Metodologia

Em linhas gerais, nossa metodologia consiste em calcular o índice de Gini com estimativas de estoques populacionais por escolaridade ao longo do tempo. Nossa maior contribuição é de combinar uma série de dados de fontes diferenciadas e chegar às estimativas dos estoques populacionais, com estimações separadas para alguns estados brasileiros. Para isso, calculamos estoques iniciais por educação e, a partir dele e de outros dados, os estoques ao longo do século XX.

5.1 Metodologia das Estimativas da Distribuição da População por Escolaridade em 1900

Dados de pesquisas censitárias eram raros no início do século XIX no Brasil. De forma mais aguda, as informações sobre escolaridade relevantes para a estimativa de um índice de Gini educacional (relativas aos estoques por grau escolar) eram inexistentes, uma vez que o interesse dos recenseamentos sobre os estoques da população por escolaridade incidiam sobre a divisão entre aqueles que sabiam ler e escrever e

aqueles que não sabiam (que representavam uma parcela muito relevante). Em consequência, é preciso estimar a distribuição da população por escolaridade a partir de algumas hipóteses.

Nós nos inspiramos na metodologia de inventário perpétuo descrito por Foldvári e Leeuwen (1914) para chegar a valores aproximados da participação relativa dos grupos de escolaridade na população de 10 anos ou mais em 1900. Nós trabalhamos com quatro categorias de escolaridade: sem escolaridade, ensino primário completo, ensino secundário completo (incluindo o ensino profissional), e ensino superior completo. Em linhas gerais, a metodologia consiste em obter percentuais de matrículas em dado grau de ensino em uma determinada faixa etária em algum período anterior, e aplicar esse percentual para a faixa etária correspondente em 1900. Por exemplo, se em 1880 a proporção das pessoas com entre 5 e 10 anos de idade matriculadas no primário fosse de 10%, então supomos que o mesmo percentual das pessoas com entre 25 e 30 anos de idade em 1900 possivelmente possuem o ensino primário. Nas estimativas populacionais do século XIX, nós consideramos também a população escrava.

Há duas hipóteses implícitas para a realização dessa estimativa. Em primeiro lugar, supomos que todos os matriculados em um grau escolar de uma geração irão se formar (taxa de evasão nula). Essa hipótese será relaxada mais adiante, porém os resultados não se alteraram de forma relevante. A segunda hipótese é de que a taxa de mortalidade e migração daqueles é homogênea entre pessoas com diferentes graus escolares. Apesar de forte, não é possível estimar taxas de mortalidade e migração diferenciadas por grau escolar com os dados disponíveis. Somente estudos utilizando dados muito recentes são capazes de realizar esse tipo de estimação.

Para obter as proporções de matriculados por faixas etárias, nós obtivemos as proporções de crianças e jovens matriculados no primário e no secundário em relação ao total de indivíduos nas faixas de idade consideradas adequadas para cada grau. Essa estimação foi possível para o início dos anos 1870 (com dados do Recenseamento de 1872 e dos Relatórios e Trabalhos Estatísticos de 1875) e dos anos 1880 (com dados dos Censos de 1872 e 1890, e dos Trabalhos da Seção de Estatística de 1886). Para a década de 1880, também foi possível estimar a proporção de adultos matriculados no ensino superior. Exceto para três escolas de nível superior, não conseguimos encontrar dados de matrículas nesse nível de ensino para o século XIX. Na ausência dessas informações, nós procedemos de duas formas diferentes (que no entanto levaram a resultados semelhantes). Em primeiro lugar, os Anuários Estatísticos fornecem dados de matrículas no ensino superior em 1907 (Anuário Estatístico do Brasil de 1953) por estado e, além disso, o número de escolas de nível superior no final do século XIX é conhecido (Anísio Teixeira, 1989). Com essas duas informações, tentamos aproximar o número de matriculados existentes em 1890, supondo que em 1907 as escolas possuíam número semelhante de alunos entre si. Dessa forma, nós ajustamos o número de alunos de 1907 às escolas existentes em 1890, supondo adicionalmente que não houve crescimento relevante do número de vagas por instituição. Apesar de se tratar de uma aproximação, a comparação com os dados para as três escolas disponíveis mostra que os números calculados não estão muito longe da realidade. A partir desses números, nós então calculamos o número de matrículas como proporções da população entre 20 e 24 anos.

Uma crítica que poderia ser feita ao método anterior seria de que nós não temos estimativas de migração dos graduados entre os estados, de modo que só conseguiríamos realizar estimativas nos estados com alguma escola de nível superior. Não é possível conhecer a priori a direção do viés devido ao método em todos os estados, uma vez que é possível que estados sem escolas de nível superior tenham graduados que migraram para lá, e estados com escolas de nível superior também pode atrair graduados de outros estados. Na tentativa de lidar com esse problema, tentamos obter uma estimativa do estoque de pessoas com nível superior a partir das informações de profissão do Censo de 1872. Para isso, consideramos que todas as pessoas que exerciam algumas profissões liberais eram formadas no nível superior. A partir dessas informações, obtivemos a proporção de pessoas com ensino superior na população.

Com as estimativas de taxas de matrículas em mãos, nós aplicamos essas proporções de pessoas com primário completo, secundário completo e ensino superior completo para a população de 1900 da seguinte forma. As proporções de pessoas com primário completo por estado em 1880 foram aplicadas para as faixa

com 10 a 29 anos em 1900, enquanto para as coortes mais velhas no mesmo ano, aplicamos as proporções de 1870. As proporções de pessoas com secundário completo de 1880 foram aplicadas às pessoas entre 15 e 29 anos em 1900, e para os mais velhos, aplicamos as proporções de 1870. Por fim, as proporções de pessoas com ensino superior foram aplicadas à população com 25 anos ou mais em 1900. Como aqueles que fizeram um grau escolar mais elevado passaram obrigatoriamente pelos graus mais elementares, nós evitamos que eles fossem contados duplamente nos graus escolares anteriores. Todos aqueles que não foram designados a algum grau escolar, foram considerados sem escolaridade.

No caso da estimativa de formados no ensino superior a partir do Censo de 1872, nós aplicamos a proporção obtida à população com 25 anos ou mais em 1900 e, em seguida, aplicamos as demais proporções ao restante da população sem o ensino superior.

Como mencionamos anteriormente, é possível relaxar a hipótese de que todos os matriculados se formam. Se tivermos a proporção dos concluintes em relação aos matriculados, podemos obter uma estimativa mais real do número de formados em um grau escolar. Dessa forma, tentamos lidar com o problema aplicando a taxa de conclusão aos matriculados, e multiplicando o valor resultante (número de concluintes em um ano) pelo tamanho da faixa etária que frequentaria aquele grau escolar.

Os primeiros dados de concluintes para todos os níveis de ensino estão disponíveis somente a partir da década de 1930. No entanto, como a taxa de conclusão por matriculado é relativamente estável entre 1932 e 1949, nós utilizamos essa taxa para estimar a proporção de concluintes por faixa etária em 1870 e 1880. Em seguida, calculamos novamente as proporções de formados em 1900, como descrito anteriormente.

Apesar de cada alteração implementada para lidar com um problema diferente gerar modificações na distribuição de pessoas por escolaridade, essas distribuições não se alteram significativamente quando fazemos uma comparação.

5.2 Composição Educacional e População ao Longo do Século XX

Para o cálculo do índice de Gini educacional ao longo do século XX é preciso obter estimativas dos estoques por grau escolar a cada ano. Como os dados disponíveis são heterogêneos ao longo do tempo, nós estimamos esses estoques por partes, da seguinte forma.

Em primeiro lugar, nossa população de referência foi aquela com 5 anos ou mais, excluindo as crianças que não estariam em idade escolar. A partir do Censo de 1940 há informações do número de pessoas por grau escolar no Brasil e nas Unidades Federativas, de modo que podemos obter os estoques facilmente. Nosso maior desafio foi reconstruir a série para a primeira metade do século XX. Para esses anos, lidamos separadamente com os graus escolares, de acordo com a disponibilidade das informações.

Em relação à população de referência e à parcela desta que não sabia ler, há dados disponíveis em todos os Censos Demográficos, desde 1872, para o Brasil como um todo e para as Unidades Federativas. Para preencher a população total e as subpopulações nos anos dentro de cada década foi utilizada uma interpolação linear simples.

Com o intuito de construir a composição educacional da população brasileira na primeira metade do séc. XX também foram utilizados dados do número de concluintes no ensino primário, secundário, profissional e superior disponíveis nos Anuários Estatísticos. A disponibilidade dessas informações varia bastante ao longo dos anos, mas estão disponíveis dados para quase todas as Unidades da Federação entre os anos de 1908 a 1912, 1932 a 1937 e 1951 a 1962. Assumindo que a quantidade de matrículas e as taxas de conclusão pouco mudam entre os anos de 1900 e 1908, as séries de concluintes para cada nível educacional foram interpoladas para os anos em que os Anuários não foram publicados, ou em que estes não apresentam essa informação. Dessa forma, é possível construir uma série de concluintes de cada nível educacional ininterrupta entre 1900 e 1960.

Para estimar o número de pessoas com cada nível de ensino completo, aplicou-se a soma cumulativa do número de concluintes de cada ano subtraindo do número de concluintes do nível seguinte. Por exemplo, o estoque total de pessoas com ensino primário em um determinado ano é calculado como a soma entre o estoque do ano anterior e o número de concluintes do ensino primário daquele ano menos os concluintes do ensino secundário do mesmo ano. Partindo dos estoques iniciais estimados na seção anterior, é possível prever a composição educacional para todos os anos entre 1900 e 1960. Para que as estimativas fossem comparáveis com os censos de 1940 e 1950, os concluintes dos níveis secundário e profissional foram agrupados para estimar o estoque de pessoas com ensino médio.

Também foram incorporadas na estimação algumas informações sobre migração e taxa de mortalidade. Há poucos dados de migração para o a primeira metade do século, mas Ferrari (2005) apresenta dados sobre o total de imigrantes para o estado de São Paulo nos anos de 1950 e 1960 e o número de imigrantes por estado de procedência em trânsito pela Hospedaria de Imigrantes de São Paulo, ano a ano, na década de 1950. As informações sobre a taxa de mortalidade foram retiradas dos Censos Demográficos (IBGE, 2007) e assumiu-se que a taxa é constante entre todos os níveis educacionais e Unidades da Federação, com exceção do São Paulo, cuja taxa de mortalidade para o período analisado é divulgada pela Fundação SEADE (SEADE, 2010).

5.3 Gini Educacional

Para calcular a desigualdade educacional na primeira metade do séc. XX, considerando a disponibilidade de informações, a população com 5 anos ou mais foi categorizada em 5 níveis de escolaridade: pessoas sem escolaridade (analfabetos), primário incompleto, primário completo, médio completo e superior. O número de pessoas com ensino primário incompleto foi calculado como resíduo da previsão dos demais níveis de ensino, ou seja, subtraindo da população com 5 anos todos os demais níveis de ensino

Para cada categoria, atribuiu-se a quantidade de anos de educação formal equivalente à duração de todos os níveis concluídos pelos indivíduos presentes no grupo. Dessa forma, como os anos de educação são constantes para cada grupo, não é possível mensurar a dispersão de escolaridade intra-grupo. Para analfabetos, definimos uma quantidade de anos de estudo próxima, mas distinta de, zero (0,01 anos) e para aqueles com o nível primário incompleto atribuiu-se o valor de 1 ano de estudo.

Optou-se pelo índice de Gini para medir a desigualdade na distribuição de educação. A metodologia usada é uma versão da fórmula usual de cálculo da medida (Cowell, 2011) adaptada ao tipo de agregação utilizada. O Gini educacional (G) para um determinado ano e unidade da federação é calculado como a média da diferença entre todos os possíveis pares de indivíduos, dividida pela média de anos de escolaridade (\bar{X}). Como há apenas cinco tipos distintos de indivíduos a formula pode ser simplificada da seguinte maneira:

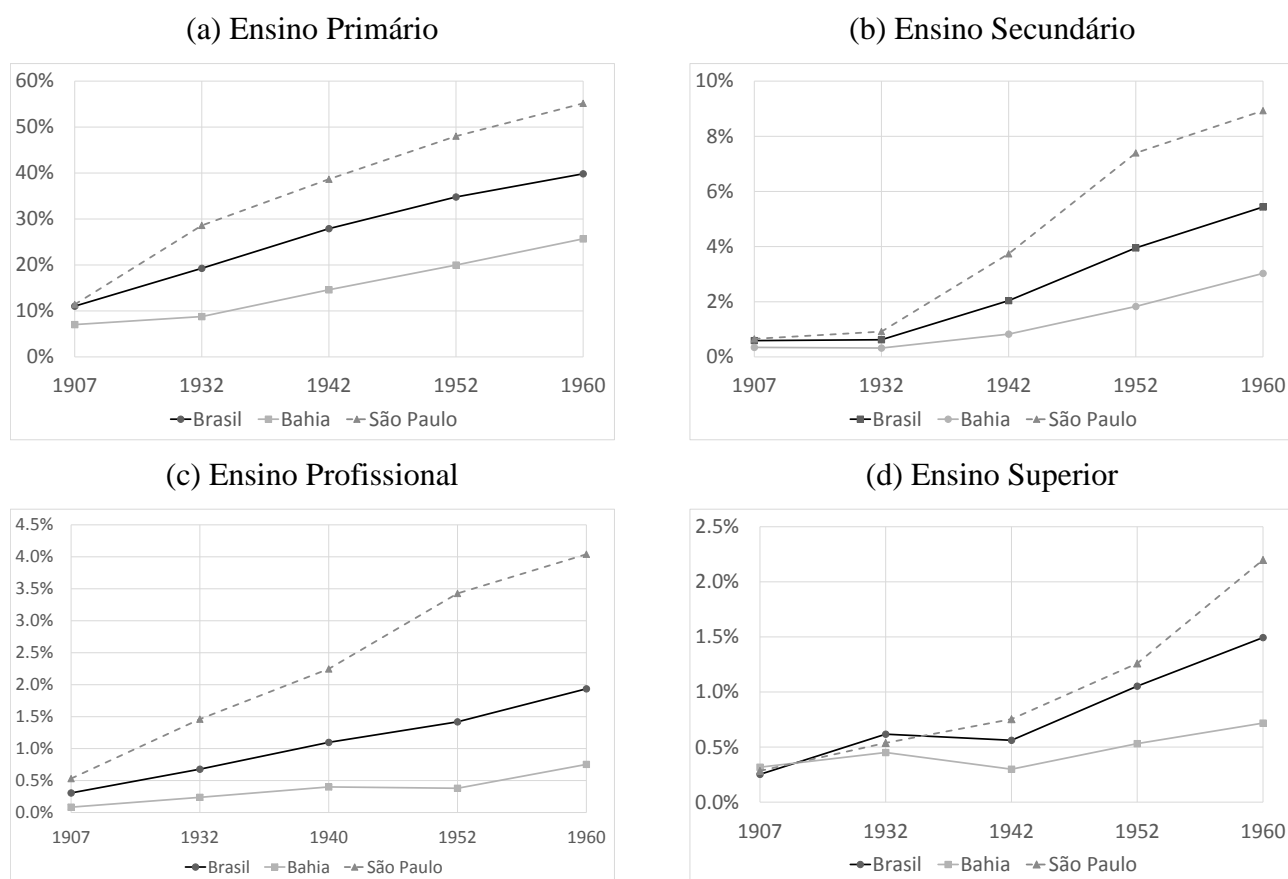
$$G = \frac{\sum_{i=1}^5 \sum_{j=1}^5 Q_i Q_j d_{ij}}{2N^2 \bar{X}}$$

Onde N é a população com 5 anos ou mais; \bar{X} é a média de anos de estudo; Q_i é a quantidade de pessoas do grupo i ; de maneira análoga Q_j é a quantidade de pessoas do grupo j ; e d_{ij} é a diferença de anos de estudo entre os grupos i e j .

6. Resultados

Nessa seção, apresentaremos os resultados das nossas estimações de taxas de matrícula e do índice de Gini educacional. Em primeiro lugar, a Figura 4 mostra as razões entre os matriculados em cada grau escolar e a população na faixa etária considerada ideal para o mesmo. Podemos observar que o Brasil como um todo apresenta um crescimento das taxas de matrícula em todos os graus escolares, ocorrendo primeiro e de forma mais acentuada no ensino primário. Para esse grau de ensino, o Brasil apresentava uma taxa de

matrícula de 10% em 1907, passando a 20% em 1932, com aumentos de pouco menos de 10 pontos percentuais a cada década subsequente. O ensino secundário e o profissional possuem níveis iniciais muito mais baixos (de respectivamente 0,6% e 0,3%) e o primeiro só começou a apresentar crescimento a partir de 1932, momento em que o governo estabeleceu um sistema nacional de ensino. No caso do ensino superior, só há uma tendência de crescimento sistemático da taxa a partir de 1942.

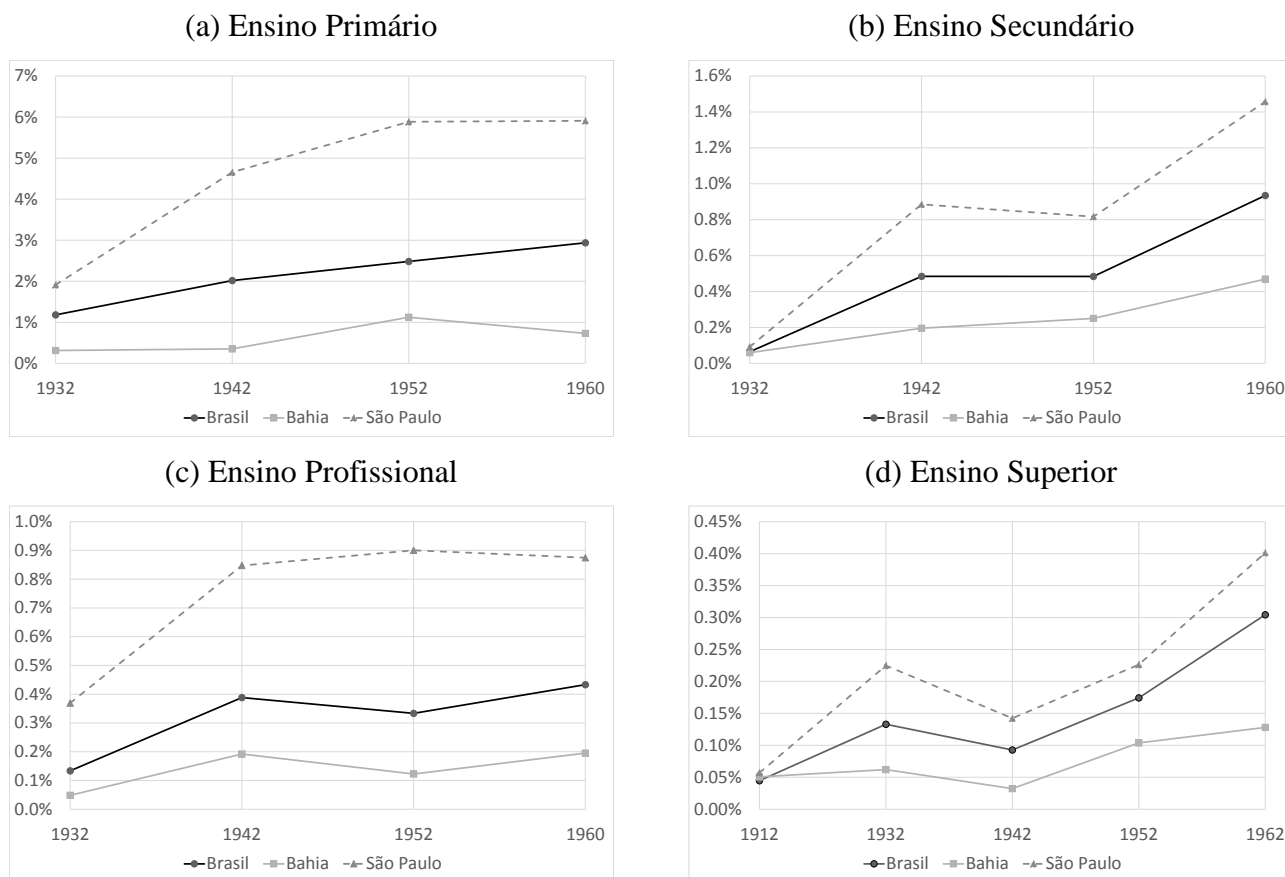
Figura 4 – Taxas de Matrícula

Fonte: Censos Demográficos; Anuários Estatísticos. Elaboração própria.

Quando analisamos os estados de São Paulo e da Bahia, fica claro que em quase todos os casos (exceto para o ensino superior em 1932) o primeiro apresenta taxas acima da média brasileira, enquanto o último apresenta taxas abaixo dela. No ensino primário e no profissional, São Paulo apresenta um rápido crescimento da taxa entre 1907 e 1932, enquanto o percentual permanece estagnado na Bahia. Note que o nível atingido pelo estado paulista só é alcançado pelo Brasil em 1942, enquanto que a Bahia só alcança o mesmo nível após 1960. Em relação ao secundário e ao superior, as taxas dos dois estados demoram mais a crescer, porém novamente o estado paulista apresenta crescimento mais acelerado a partir de 1932.

O patamar das matrículas no ensino primário, muito superior aos demais níveis, pode ser explicado pelo fato de que desde a Proclamação da República no final do século XIX, a necessidade de expandir a educação básica tornou-se evidente, de forma que o intuito do governo era expandir a educação elementar – independente de qualidade, o que importava era quantidade – em detrimento dos demais níveis educacionais.

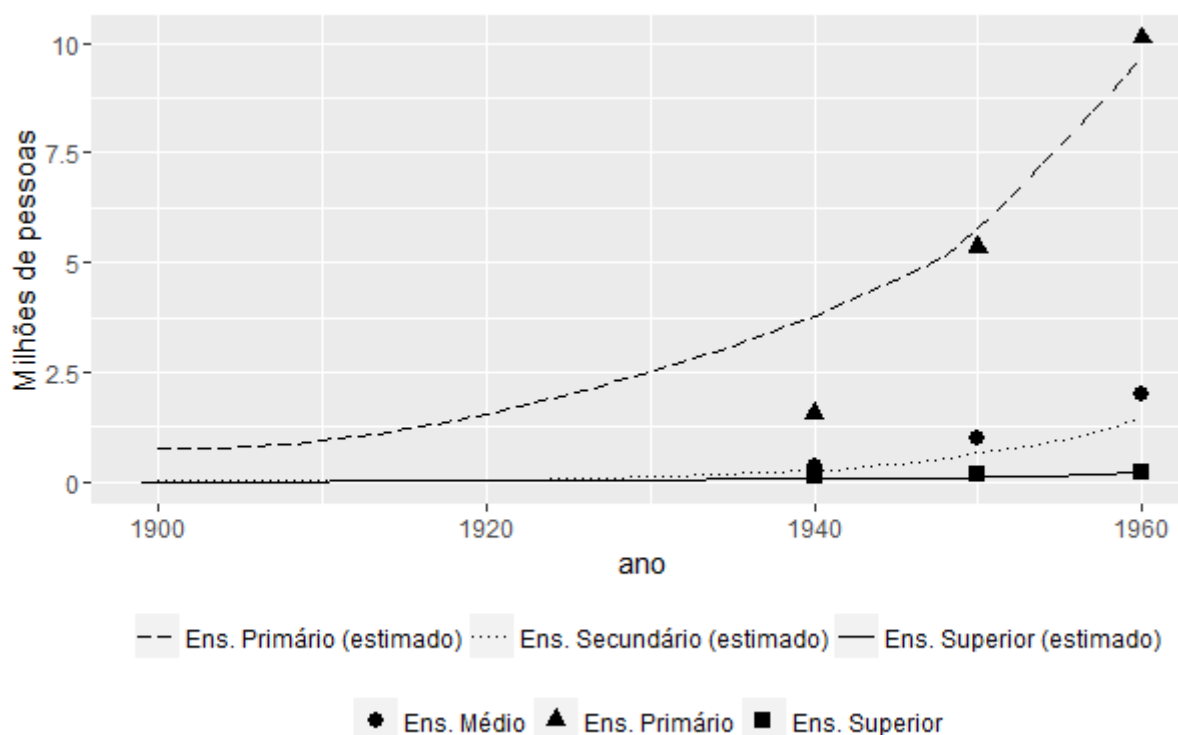
Na Figura 5 apresentamos as taxas de conclusão, calculadas como a proporção de concluintes em relação à população na faixa de idade correspondente ao grau escolar. Note que a taxa de conclusão reflete de forma mais precisa o incremento de um grau escolar em uma faixa de idade. Exceto pelo ensino superior, cuja série se inicia em 1912, os demais graus escolares possuem dados disponíveis somente a partir de 1932. Podemos observar no painel (a) que o crescimento da taxa de conclusão do primário cresce consistentemente para o Brasil em todo o período. Os padrões de crescimento para os demais graus escolares (painéis (b) a (d)) são menos claros do que o das taxas de matrícula, apresentando estagnação entre 1932 e 1942 no ensino superior, entre 1942 e 1952 no ensino secundário, e a partir de 1932 no profissional. Após esse período, o ensino superior apresenta crescimento acelerado até 1962. Na comparação entre estados, novamente São Paulo aparece acima da média do país, enquanto a Bahia apresenta percentuais abaixo dela.

Figura 5 – Taxas de Conclusão

Fonte: Censos Demográficos; Anuários Estatísticos. Elaboração própria.

As taxas de matrícula e conclusão já dão uma pista em relação à evolução da composição da população por escolaridade e do índice de Gini educacional. A Figura 6 mostra a composição educacional estimada para o Brasil a partir dos níveis iniciais calculados para o ano de 1900. As linhas representam os níveis estimados de pessoas em cada grau escolar (primário, secundário e ensino superior), enquanto os pontos mostram os totais obtidos a partir dos Censos Demográficos de 1940 a 1960. É possível verificar que a quantidade de pessoas com o primário completo é muito superior àquele dos demais níveis educacionais, o que é consistente com o intuito dos governantes do período, que desejavam expandir a educação elementar, e apresenta crescimento explosivo a partir de 1940. Apesar de a nossa estimativa estar acima do valor observado em 1940, os níveis estimados são semelhantes aos observados nos anos posteriores. Essas diferenças são relativamente menores no caso do secundário e do ensino superior.

Figura 6 – Composição Educacional do Brasil

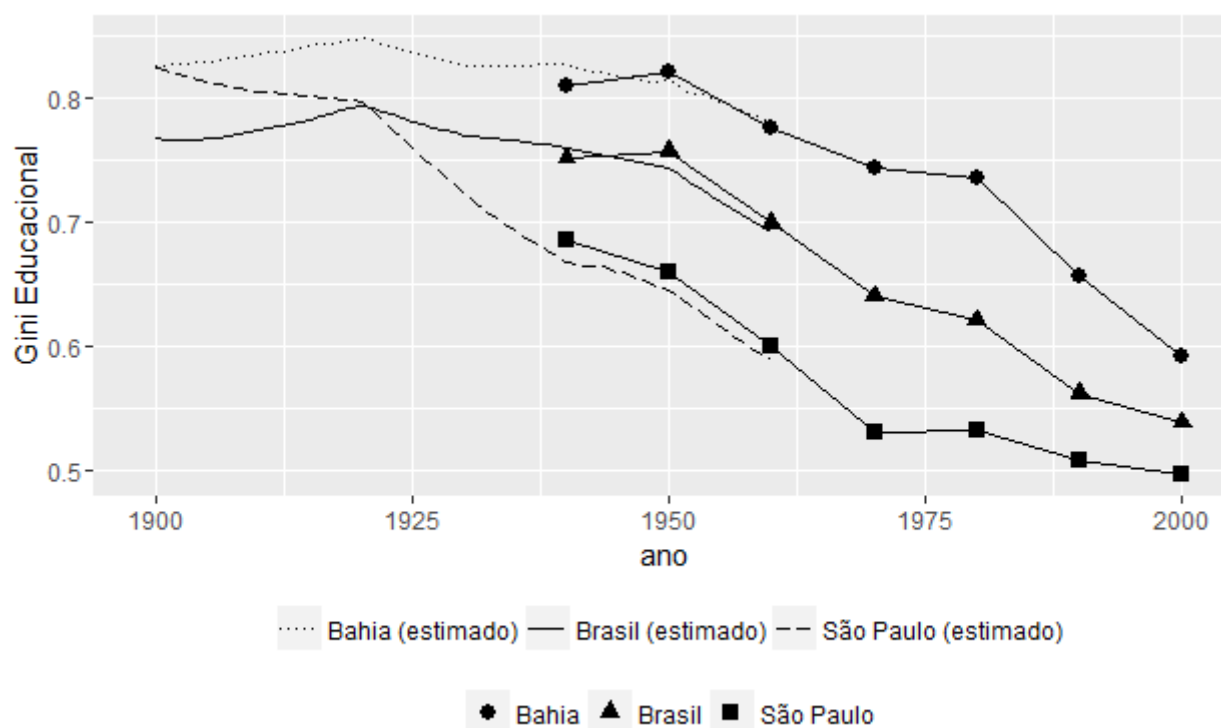


Fonte: Elaboração própria.

A Figura 5 apresenta a estimaco principal do Gini Educacional para a primeira metade do sculo XX. Adicionalmente,  utilizada a mesma metodologia de agregao em 5 grupos de escolaridade para os dados dos Censos de 1940 a 2000, para fins de comparao.  possvel notar que no comeo do sculo os estados da Bahia e So Paulo tm nveis de desigualdade educacional muito prximos, acima da mdia do Brasil. Alm disso  possvel perceber que a desigualdade educacional brasileira aumenta nas primeiras dcadas do sculo. Esse resultado pode ser explicado por um contingente muito pequeno da populao com educao formal. Conforme a oferta de educao surge, apenas uma parcela da populao total tem acesso ao sistema e a desigualdade inicialmente aumenta. Com a expanso do sistema de ensino e o aumento do acesso, uma quantidade mais significativa da populao brasileira passa a adquirir educao formal e a desigualdade cai.

Entre as dcadas de 1900 e 1920, enquanto a desigualdade da Bahia e a mdia nacional aumentam, a de So Paulo apresenta reduo. De 1920 em diante, todas as sries apresentam diminuio. Contudo, o estado de So Paulo tem declive mais acentuado do que a mdia entre as dcadas de 1920 e 1940.

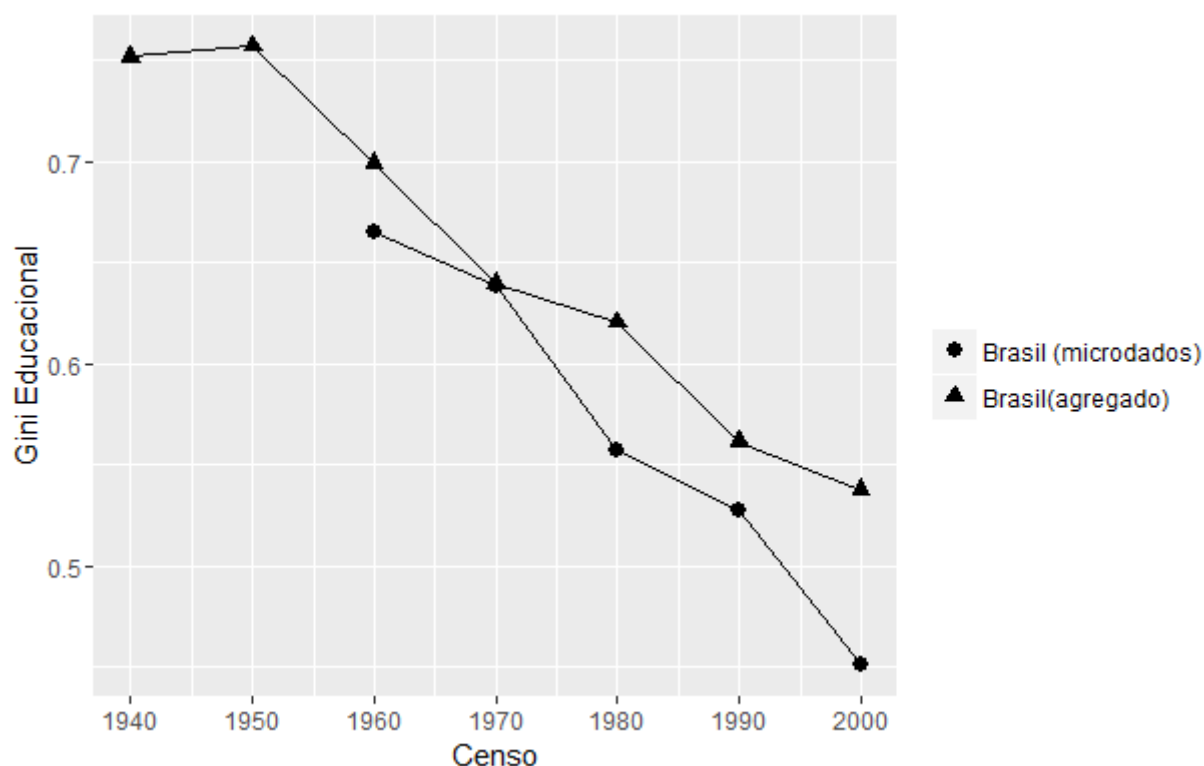
Figura 7 – Gini Educacional Estimado vs Real



Fonte: Elaboração própria.

A Figura 6 apresenta a comparação entre o cálculo do índice de Gini usando a metodologia de agregação por níveis educacionais (i.e. sem variação intra-grupo) e o cálculo utilizando dados em nível individual. Ambas as séries fazem uso dos dados do Censo Demográfico, porém apenas a partir de 1960 é possível utilizar os micro dados. A estimação com dados agregados situa-se acima daquela de nível individual em todos os anos. Como para definir os 5 grupos de nível de ensino foi considerado apenas o último nível concluído por cada indivíduo, ao desconsiderar a variação intra-grupos, a agregação provavelmente subestima os anos de estudo da parcela educada da população e superestima o índice. No entanto, é possível observar que as tendências de longo prazo das duas séries são muito semelhantes entre si, o que nos dá maior confiança sobre as séries estimadas para a primeira metade do século XX.

Figura 8 – Gini Educacional Microdados vs Agregados



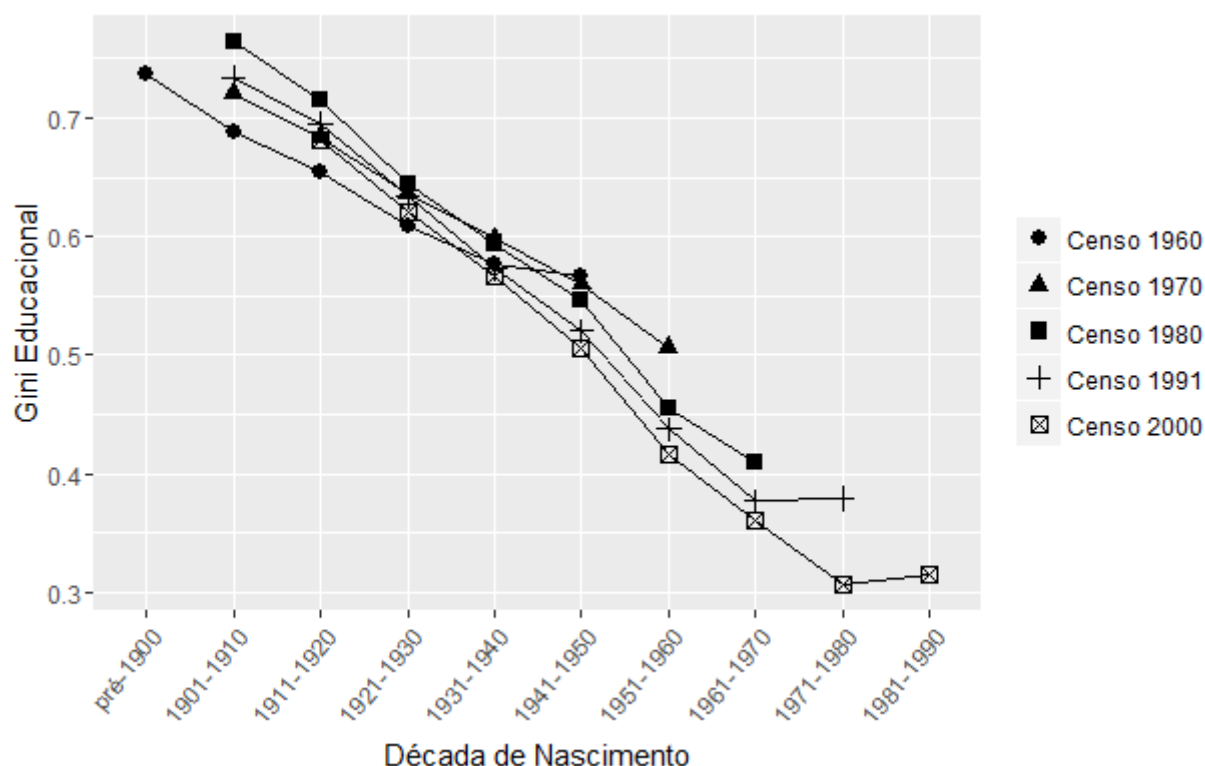
Fonte: Censos Demográficos/IBGE. Elaboração própria.

É possível perceber, entretanto, que a diferença é menor nos anos de 1960 e 1970. Dado que a proporção de pessoas com educação formal aumenta ao longo do tempo, a diferença entre as duas séries tende a aumentar com o passar do tempo. Assim, a superestimação do índice tende a ser menos expressiva do começo à metade do século, período de interesse dessa análise.

A Figura 7 mostra a evolução do índice de Gini educacional por coorte de nascimento de 1900 a 1990. De forma consistente com a evolução do índice de Gini estimado para os estoques populacionais ao longo do século XX, fica clara a redução do índice ao longo das coortes. Apesar de uma seleção possivelmente diferenciada por escolaridade ao longo do tempo (devido às diferenças das taxas de mortalidade), é possível notar que as linhas não se distanciam umas das outras, o que significa que essa seleção é pequena em comparação com a variação entre as coortes.

Com a expansão do sistema de ensino ao longo do tempo, a tendência de redução da desigualdade educacional significa que a cada geração, mais indivíduos foram incorporados ao sistema educacional, de modo a torna-los mais homogêneos em termos de escolaridade. Apesar dessas mudanças nas coortes de nascimento, as transformações ocorridas para o estoque populacional são mais lentas, uma vez que, a cada instante, há várias coortes sobrepostas, com diferentes níveis de desigualdade educacional. No entanto, esse resultado mostra que a tendência de queda observada para os estoques de fato ocorre.

Figura 9 – Gini Educacional por Coortes de Nascimento



Fonte: Censos Demográficos/IBGE. Elaboração própria.

7. Conclusões

Nesse artigo tivemos como objetivo medir a desigualdade educacional no Brasil e nos Estados de São Paulo e da Bahia, com o intuito de contribuir para o debate acerca dos determinantes da elevada desigualdade social observada no país. Apesar da recente tendência de queda da desigualdade de renda, o Brasil ainda é um dos países mais desiguais do mundo, e parte dessa situação parece ser em parte resultado do processo diferenciado de expansão educacional nos estados brasileiros. Para verificar essa hipótese, construímos séries de um índice de Gini educacional desde o início do século XX para o Brasil e os Estados analisados.

Nossos resultados mostram, em primeiro lugar, que a desigualdade educacional era extremamente elevada em 1900. De acordo com as nossas estimativas, naquele ano a grande maioria da população (69,1%) era composta por analfabetos ou por pessoas sem o primário completo (25,5%), o que gerava um índice de Gini de 0,76. O sistema educacional pouco acessível à população em geral também contribuía para a manutenção da desigualdade, de modo que a geração nascida entre 1900 e 1910 manteve uma desigualdade também relativamente elevada (com índice de Gini de 0,75).

Para o Brasil como um todo, entre 1900 e 1920 a desigualdade educacional aumentou, com um crescimento maior daqueles com nenhuma escolaridade ou primário incompleto, em comparação aos graus mais elevados. Ao longo das décadas, no entanto, as coortes foram se tornando cada vez menos desiguais, o que é medido pelo declínio sistemático do Gini educacional das coortes para cerca de 0,3 entre os nascidos na década de 1980. Esse movimento foi concretizado pelo aumento de matrículas inicialmente no ensino primário, e posteriormente no secundário e no ensino superior, uma vez que na situação inicial a grande maioria da população tinha escolaridade muito próxima a zero. Dessa forma, a redução da desigualdade educacional coincide com um aumento do nível médio de escolaridade. A introdução sucessiva das novas

coortes menos desiguais no estoque foi reduzindo a disparidade geral de educação, resultando em um movimento de redução das desigualdade no total da população entre 1920 e 2000.

São Paulo era um dos Estados com menor desigualdade educacional no Brasil, em contraste com alguns Estados da região Nordeste, como o Ceará e a Bahia, o que pode ajudar a esclarecer as diferenças de desenvolvimento regional. Por esse motivo, nós realizamos análises separadas para São Paulo e Bahia. Nossas estimativas mostram que os níveis de desigualdade desses dois eram semelhantes em 1900, e extremamente elevados (0,825 e 0826 respectivamente), ficando acima da média brasileira. Com a evolução diferenciada das matrículas (a educação era de responsabilidade do governo estadual) e do influxo de migrantes estrangeiros (que possuíam escolaridade média maior), o estado de São Paulo foi capaz de reduzir mais rapidamente as desigualdades educacionais. Isso ocorreu desde o início do século, enquanto o estado da Bahia acompanhou o movimento geral do país, com uma pequena elevação da desigualdade entre 1900 e 1920. Do lado da oferta educacional, ainda que em escala relativamente limitada, o estado de São Paulo foi capaz de expandir as matrículas em todos os níveis de ensino antes que o Brasil no geral e o estado da Bahia, especialmente no ensino primário. A proporção de jovens no ensino primário em São Paulo em 1932 (28%) só é alcançado pelo Brasil em 1942, e pela Bahia após 1960.

Essa evolução apresentada da desigualdade escolar ajuda a esclarecer em parte a desigualdade social entre as regiões. O aumento da escolaridade, associado à redução da desigualdade, parece ter uma papel relevante para explicar as diferenças de renda observadas ao longo do século XX e que ainda persistem atualmente.

8. Referências Bibliográficas

- ACEMOGLU, D.; JOHNSON, S.; ROBINSON, J. A. The Colonial Origins of Comparative Development: An Empirical Investigation. **American Economic Association**, v. 91, n. 5, p. 1369-1401, dez. 2001.
- BARROS, A. R. **Desigualdades Regionais no Brasil: Natureza, causas, origens e soluções**. Rio de Janeiro: Elsevier, 2011, 336 p.
- CAMPS, E.; ENGERMAN, S. L. **The Impact of Race and Inequality on Human Capital Formation in Latin America During the Nineteenth and Twentieth Centuries**. Economics Working Papers 1436, Departament of Economics and Business, Universitat Pompeu Fabra, 2014.
- COWELL, F. A. **Measuring Inequality**. Orford University Press, 2009, 236 p.
- DABLA-NORRIS, E.; KOCHHAR, K.; RICKA, F.; SUPHAPHIPHAT, N.; TSOUNTA, E. **Causes and Consequences of Income Inequality: A Global Perspective**. International Monetary Fund (Discussion Note), jun. 2015.
- DIRECTORIA GERAL DE ESTATISTICA. Relatorio Anexo ao do Ministerio dos Negocios do Imperio, de 1875. Rio de Janeiro: Typographia de Pinto, Brandão e Comp., 1875.
- ENGERMAN, S. L.; SOKOLOFF, K. L. Factor Endowments, Institutions, and Differential Paths of Growth Among New World Economies: A View from Economic Historians of the United States. **National Bureau of Economic Research**, n. 66, dez. 1994.
- FOLDVARI, P.; LEEUWEN, B. Educational and Income Inequality in Europe, ca. 1870-2000. **Cliometrica**, v. 8, p. 271-300, 2014.
- FUNDAÇÃO SISTEMA ESTADUAL DE ANÁLISE DE DADOS (SEADE). Memória das Estatísticas Demográficas. Disponível em: <<http://produtos.seade.gov.br/produtos/500anos/index.php?tip=prir>>. Acesso em julho de 2016.
- FURTADO, C. **Formação Econômica do Brasil**. 32. ed. São Paulo: Companhia Editora Nacional, 2005.

IBGE. Anuário Estatístico do Brasil. Rio de Janeiro: IBGE, 1938-1962.

_____. Estatísticas do Século XX. Rio de Janeiro: IBGE, 2007.

_____. Trabalho e Rendimento Educação e Deslocamento. Rio de Janeiro: IBGE, 2012.

INSTITUTO NACIONAL DE ESTATISTICA. Anuário Estatístico do Brasil. Rio de Janeiro: Tip do Departamento de Estatística de Publicidade, 1936-1937.

LANGONI, C. **A Distribuição de Renda e Desenvolvimento Econômico no Brasil**. Rio de Janeiro: Editora Expressão e Cultura, 1973.

LEEUWEN, B. V.; LEEUWEN-LI, J. V.; FOLDVARI, P. Average Years of Education (Average, total Population 15 years and older), 1850-2010. Disponível em: <<https://www.clio-infra.eu/>>. Acesso em julho de 2016.

LEFF, N. H. Desenvolvimento Econômico e Desigualdade Regional: origens do caso brasileiro. **Revista Brasileira de Economia**, vol. 26, n. 1, 1972.

MARCÍLIO, M. L. **História da Escola em São Paulo e no Brasil**. São Paulo: Imprensa Oficial do Estado de São Paulo: Instituto Fernand Braudel, 2005. 485 p.

MINISTÉRIO DA AGRICULTURA, INDÚSTRIA E COMÉRCIO. DIRECTORIA GERAL DE ESTATÍSTICA. Anuario Estatistico do Brazil – 1º Anno (1908-1912). Rio de Janeiro: Typographia, vol. 1, 1916.

MUSACHIO, A.; MARTINEZ, A.; VIARENGO, M. Colonial Institutions, Commodity Boom, and the Diffusion of Elementary Education in Brazil, 1889-1930. National Bureau of Economic Research (NBER Working Paper n. 20029), jan. 2015. Disponível em: <<http://www.nber.org/papers/w20029>>. Acesso em maio de 2016.

OCDE. Fomentando o Crescimento Inclusivo da Produtividade na América Latina. Paris: (Série Melhores Políticas), 2014, p. 19.

REIS, J. G. A.; BARROS, R. P. de. Wage Inequality and the Distribution of Education: a study of the evolution of the regional differences in inequality metropolitan Brazil. **Journal of Development Economics**, n. 36, p. 117-143, 1991.

ROCHA, R.; FERRAZ, C.; SOARES, R. R. Human Capital Persistence and Development. The Institute for the Study of Labor (Discussion Paper n. 9101), jun. 2015. Disponível em: <<http://ftp.iza.org/dp9101.pdf>>. Acesso em maio de 2016.

SECRETARIA D'ESTADO DOS NEGÓCIOS DO IMPÉRIO. 3ª DIRECTORIA. SECÇÃO DE ESTATÍSTICA. Trabalhos da Secção de Estatística Anexa a 3ª Directoria da Secretaria d'Estado dos Negócios do Império, Anno de 1886. Rio de Janeiro: Imprensa Nacional, 1887.

SIMONSEN, R. **História Econômica do Brasil**. 4. Ed. Brasília: Edições do Senado Federal, 2005.

U. S. Census Bureau. A Half-Century of Learning: Historical Census Statistics on Educational Attainment in the United States, 1940 to 2000. Disponível em: <https://www.census.gov/hhes/socdemo/education/data/census/half-century/files/US.pdf> . Acesso em: abr. 2016.

WILLIAMSON, J. Latin America Inequality: Colonial Origins, Commodity Booms, or a Missed 20th Century Leveling? National Bureau of Economic Research (NBER Working Paper n. 20915), jan. 2015. Disponível em: <<http://www.nber.org/papers/w20915>>. Acesso em maio de 2016.